

ALIBABA CLOUD

Alibaba Cloud

Apsara Stack Enterprise Product Introduction

Product Version: 1907, Internal: V3.8.0

Document Version: 20200901

 Alibaba Cloud

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).
5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.
6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
 Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 Danger: Resetting will result in the loss of user configuration data.
 Warning	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 Warning: Restarting will cause business interruption. About 10 minutes are required to restart an instance.
 Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: If the weight is set to 0, the server no longer receives new requests.
 Note	A note indicates supplemental instructions, best practices, tips, and other content.	 Note: You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings> Network> Set network type .
Bold	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click OK .
Courier font	Courier font is used for commands	Run the <code>cd /d C:/window</code> command to enter the Windows system folder.
<i>Italic</i>	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	This format is used for a required value, where only one item can be selected.	<code>switch {active stand}</code>

Table of Contents

1. Apsara Stack introduction	24
1.1. What is Alibaba Cloud Apsara Stack?	24
1.2. The reasons to choose Apsara Stack	25
1.2.1. Hyper-scale distributed cloud operating system	25
1.2.2. Deployment and control system of Apsara Infrastruc...	27
1.2.3. High-reliability disaster recovery solutions	27
1.2.4. Unified operations management and automated ope...	28
1.2.5. Open cloud service interface	29
1.3. Product architecture	29
1.3.1. Types of private cloud architectures	29
1.3.2. System architecture	29
1.3.3. Network architecture	31
1.3.3.1. Network architecture overview	31
1.3.3.2. Business service area	32
1.3.3.3. Integrated access area	33
1.3.3.4. VPC leased line access	35
1.3.4. Security architecture	36
1.3.5. Base assembly	37
1.4. Product panorama	38
1.5. Scenarios	39
1.6. Compliance security solution	40
1.6.1. Overview	40
1.6.2. Interpretation on key points	41
1.6.3. Cloud-based classified protection compliance	42
1.6.4. Classified protection implementation process	45
1.6.5. Security and compliance architecture	45

1.6.6. Solution benefits	46
2.Elastic Compute Service (ECS)	48
2.1. What is ECS?	48
2.2. Benefits	48
2.3. Architecture	50
2.4. Features	52
2.4.1. Instances	52
2.4.1.1. Overview	52
2.4.1.2. Instance type families	52
2.4.1.3. Instance types	62
2.4.1.4. UserData	86
2.4.1.5. Instance lifecycle	86
2.4.1.6. EBM Instances	88
2.4.2. Block storage	89
2.4.2.1. Overview	89
2.4.2.2. Block storage performance	90
2.4.2.2.1. Overview	90
2.4.2.2.2. Elastic block storage performance	90
2.4.2.2.3. Local disk storage performance	92
2.4.2.2.4. Test disk performance	92
2.4.2.3. Elastic block storage	93
2.4.2.3.1. Overview	94
2.4.2.3.2. Cloud disks	94
2.4.2.3.3. Shared block storage	94
2.4.2.3.4. Triplicate storage	95
2.4.2.4. ECS disk encryption	96
2.4.2.5. Local storage	96
2.4.3. Images	97

2.4.4. Snapshots	98
2.4.4.1. Overview	98
2.4.4.2. Mechanisms	99
2.4.4.3. Specifications of ECS Snapshot 2.0	99
2.4.4.4. Technical comparison	100
2.4.5. Deployment sets	101
2.4.6. Network and security	103
2.4.6.1. IP addresses of ECS instances of VPC type	103
2.4.6.2. Elastic network interfaces	104
2.4.6.3. Intranet	105
2.4.6.4. Security group rules	106
2.5. Scenarios	106
2.6. Limits	106
2.7. Terms	108
3.Container Service	110
3.1. What is Container Service?	110
3.2. Benefits	110
3.3. Architecture	112
3.4. Features	113
3.5. Scenarios	114
3.6. Limits	119
3.7. Terms	120
4.Auto Scaling (ESS)	123
4.1. What is ESS?	123
4.2. Benefits	124
4.3. Architecture	124
4.4. Features	126
4.5. Scenarios	127

4.6. Limits	128
4.7. Terms	128
5.Object Storage Service (OSS)	130
5.1. What is OSS?	130
5.2. Advantages	130
5.3. OSS architecture	131
5.4. Functions	132
5.5. Scenarios	134
5.6. Limits	134
5.7. Terms	135
6.Table Store	137
6.1. What is Table Store?	137
6.2. Benefits	137
6.3. Architecture	138
6.4. Features	140
6.5. Scenarios	140
6.6. Limits	146
6.7. Terms	147
7.Network Attached Storage (NAS)	150
7.1. What is NAS?	150
7.2. Benefits	150
7.3. Architecture	150
7.4. Features	151
7.5. Scenarios	152
7.6. Limits	152
7.7. Terms	153
8.Apsara File Storage for HDFS	155
8.1. What is Apsara File Storage for HDFS?	155

8.2. Benefits	155
8.3. Architecture	155
8.4. Features	156
8.5. Scenarios	157
8.6. Limits	157
8.7. Terms	158
9. ApsaraDB for RDS	159
9.1. What is ApsaraDB for RDS?	159
9.2. Benefits	160
9.2.1. Ease of use	160
9.2.2. High performance	160
9.2.3. High security	161
9.2.4. High reliability	162
9.3. Architecture	162
9.4. Features	162
9.4.1. Data link service	162
9.4.2. High-availability service	164
9.4.3. Backup and recovery service	166
9.4.4. Monitoring service	167
9.4.5. Scheduling service	168
9.4.6. Migration service	168
9.5. Scenarios	169
9.5.1. Diversified data storage	169
9.5.2. Read/write splitting	170
9.5.3. Big data analysis	171
9.6. Usage limits	172
9.6.1. Limits on ApsaraDB RDS for MySQL	172
9.6.2. Usage limits of ApsaraDB RDS for PostgreSQL	173

9.6.3. Usage limits of ApsaraDB RDS for PPAS	174
9.7. Terms	175
9.8. Instance types	175
10.KVStore for Redis	180
10.1. What is KVStore for Redis?	180
10.2. Benefits	180
10.3. Architecture	181
10.4. Features	182
10.5. Scenarios	183
10.6. Limits	185
10.7. Terms	186
10.8. Instance types	186
11.ApsaraDB for MongoDB	192
11.1. What is ApsaraDB for MongoDB?	192
11.2. Benefits	192
11.3. Architecture	193
11.4. Features	194
11.5. Scenarios	195
11.6. Limits	195
11.7. Terms	196
11.8. Instance specifications	197
12.KVStore for Memcache	198
12.1. What is KVStore for Memcache?	198
12.2. Benefits	198
12.3. Architecture	199
12.4. Features	200
12.5. Scenarios	201
12.6. Limits	201

12.7. Terms	202
12.8. Instance types	202
13. AnalyticDB for PostgreSQL	205
13.1. What is AnalyticDB for PostgreSQL?	205
13.2. Benefits	205
13.3. Architecture	206
13.4. Features	207
13.4.1. Distributed architecture	208
13.4.2. High-performance data analysis	208
13.4.3. High-availability service	209
13.4.4. Data synchronization and tools	209
13.4.5. Data security	209
13.4.6. Supported SQL features	209
13.5. Scenarios	212
14. Data Transmission Service (DTS)	215
14.1. What is DTS?	215
14.2. Benefits	215
14.3. Architecture	216
14.4. Features	220
14.4.1. Data migration	220
14.4.2. Data synchronization	223
14.4.3. Data subscription	227
14.5. Scenarios	229
14.6. Concepts	236
15. Data Management Service (DMS)	238
15.1. What is Data Management Service?	238
15.2. Benefits	238
15.3. Architecture	238

15.4. Features	239
15.5. Scenarios	240
15.5.1. Convenient data operations	240
15.5.2. Prohibiting data export	240
15.5.3. SQL statement reuse	241
15.6. Limits	241
16. Server Load Balancer (SLB)	243
16.1. What is Server Load Balancer?	243
16.2. Benefits	244
16.3. Architecture	244
16.4. Features	246
16.5. Scenarios	247
16.6. Limits	247
16.7. Terms	247
17. Virtual Private Cloud (VPC)	249
17.1. What is VPC?	249
17.2. Benefits	250
17.3. Architecture	251
17.4. Features	252
17.5. Scenarios	253
17.6. Terms	254
17.7. Limits	254
18. Log Service	256
18.1. What is Log Service?	256
18.2. Benefits	256
18.3. Product architecture	257
18.4. Features	259
18.4.1. Core features	259

18.4.2. Other features	260
18.4.2.1. Log	260
18.4.2.2. Project	262
18.4.2.3. Logstore	263
18.4.2.4. Shard	263
18.4.2.5. Log topic	265
18.5. Scenarios	266
18.6. Limits	267
18.7. Terms	268
19.Apsara Stack Security	270
19.1. What is Apsara Stack Security?	270
19.2. Advantages	270
19.3. Architecture	272
19.4. Features	273
19.5. Restrictions	278
19.6. Terms	278
20.Key Management Service (KMS)	280
20.1. What is KMS?	280
20.2. Benefits	280
20.3. Architecture	281
20.4. Features	282
20.5. Scenarios	282
20.6. Limits	286
20.7. Terms	286
21.Apsara Stack DNS	288
21.1. What is Apsara Stack DNS?	288
21.2. Benefits	288
21.3. Architecture	289

21.4. Features	289
21.5. Scenarios	290
21.6. Limits	291
21.7. Basic concepts	291
22.API Gateway	293
22.1. What is API Gateway?	293
22.2. Features	293
22.3. Benefits	294
22.4. Concepts	294
23.MaxCompute	297
23.1. What is MaxCompute?	297
23.2. Benefits	297
23.3. Architecture	298
23.4. Features	300
23.4.1. Tunnel	300
23.4.1.1. Terms	300
23.4.1.2. Tunnel features	300
23.4.1.3. Data upload and download through Tunnel	300
23.4.2. SQL	301
23.4.2.1. Terms	301
23.4.2.2. SQL characteristics	302
23.4.2.3. Comparison with open source products	302
23.4.3. MapReduce	303
23.4.3.1. Terms	303
23.4.3.2. MapReduce characteristics	303
23.4.3.3. MaxCompute MapReduce process	303
23.4.3.4. Hadoop MapReduce VS MaxCompute MapReduce	304
23.4.4. Graph	305

23.4.4.1. Terms	305
23.4.4.2. Graph characteristics	305
23.4.4.3. Graph relational network models	305
23.4.5. Unstructured data processing (integrated computin...	306
23.4.6. Unstructured data processing in MaxCompute	306
23.4.7. Enhanced features	307
23.4.7.1. Spark on MaxCompute	307
23.4.7.1.1. Terms	307
23.4.7.1.2. Features of Spark on MaxCompute	307
23.4.7.1.3. Spark features	308
23.4.7.1.4. Spark architecture	308
23.4.7.1.5. Benefits of Spark on MaxCompute	309
23.4.7.2. Elasticsearch on MaxCompute	309
23.4.7.2.1. Overview	310
23.4.7.2.2. Features of Elasticsearch on MaxCompute	310
23.4.7.2.3. Elasticsearch features	311
23.4.7.2.4. Elasticsearch architecture	311
23.4.7.2.5. Elasticsearch benefits	312
23.5. Scenarios	313
23.5.1. Scenario 1: Migrate data to the cloud cost-effectivel...	313
23.5.2. Scenario 2: Improve development efficiency and red...	313
23.5.3. Scenario 3: Use mass data to achieve precision ma...	314
23.5.4. Scenario 4: Achieve precision marketing with big da...	314
23.6. Limits	315
23.7. Terms	315
24.DataWorks	317
24.1. What is DataWorks?	317
24.2. Benefits	317

24.2.1. Powerful computing capabilities	317
24.2.2. End-to-end platform	317
24.2.3. Data integration from heterogeneous data sources	317
24.2.4. Web-based software	318
24.2.5. Multi-tenant architecture	318
24.2.6. Intelligent monitoring and alerts	318
24.2.7. Easy-to-use SQL editor	318
24.2.8. Comprehensive data quality monitoring	318
24.2.9. Convenient API development and management	318
24.2.10. Secure data sharing	318
24.3. Architecture	318
24.3.1. Services	318
24.3.2. System architecture	319
24.3.3. Security architecture	319
24.3.4. Multi-tenant architecture	319
24.4. Services	319
24.4.1. Data Integration	319
24.4.1.1. Overview	319
24.4.1.2. Various data sources	320
24.4.1.2.1. Metadata synchronization	320
24.4.1.2.2. Relational database	320
24.4.1.2.3. NoSQL database	320
24.4.1.2.4. MPP database	320
24.4.1.2.5. Big data product	320
24.4.1.2.6. Unstructured data storage	320
24.4.1.3. Inbound data control	320
24.4.1.4. High transmission rate	320
24.4.1.5. Accurate flow control	320

24.4.1.6. Synchronization agents	321
24.4.1.7. Transmission in complicated networks	321
24.4.2. DataStudio	321
24.4.2.1. Overview	321
24.4.2.2. Business flows	321
24.4.2.2.1. Description	321
24.4.2.2.2. Nodes	321
24.4.2.2.3. Configure nodes	321
24.4.2.2.4. Versions	322
24.4.2.2.5. Publish business flows	322
24.4.2.3. Solutions	322
24.4.2.4. Code editor	322
24.4.2.4.1. ODPS SQL nodes	322
24.4.2.4.2. ODPS MR nodes	322
24.4.2.4.3. Resources	322
24.4.2.4.4. Register user-defined functions	322
24.4.2.4.5. Shell nodes	322
24.4.2.5. Code repository and team collaboration	323
24.4.3. Administration	323
24.4.3.1. Overview	323
24.4.3.2. Overview	323
24.4.3.3. Task instances	323
24.4.3.4. Monitor	323
24.4.4. Workspace Management	324
24.4.4.1. Description	324
24.4.4.2. Organization	324
24.4.4.3. Workspaces	324
24.4.4.4. Members	324

24.4.4.5. Authorizations	324
24.4.5. Realtime Analysis	325
24.4.6. Data Service	325
24.4.7. Data Asset Management	326
24.4.8. Data Protection	326
24.4.8.1. Overview	326
24.4.8.2. Terms	327
24.4.8.2.1. Organization	327
24.4.8.2.2. Workspace	327
24.4.8.2.3. Regular expression	327
24.4.8.2.4. Data classification	327
24.4.8.2.5. Data recognition	327
24.4.8.2.6. Data masking	327
24.4.8.2.7. MaxCompute	327
24.5. Scenarios	328
24.5.1. Cloud-based data warehouse	328
24.5.2. Business intelligence	328
24.5.3. Data-driven management	328
24.6. Limits	328
24.7. Terms	328
25.Apsara Bigdata Manager (ABM)	330
25.1. What is Apsara Bigdata Manager?	330
25.2. Benefits	330
25.3. Architecture	330
25.3.1. System architecture	330
25.4. Features	332
25.4.1. Dashboard	332
25.4.2. Business	334

25.4.3. O&M	335
25.4.4. Management	340
25.5. Scenarios	341
25.6. Limits	341
25.7. Concepts	341
26. Realtime Compute	343
26.1. What is Realtime Compute?	343
26.2. End-to-end real-time computing	344
26.3. Differences between real-time computing and batch co... ..	345
26.3.1. Overview	345
26.3.2. Batch computing	345
26.3.3. Real-time computing	346
26.3.4. Comparison between real-time computing and batc... ..	347
26.4. Benefits	348
26.5. Product architecture	349
26.5.1. Business process	349
26.5.2. Business architecture	351
26.5.3. Technical architecture	352
26.6. Features	353
26.7. Product positioning	355
26.8. Scenarios	356
26.8.1. Overview	356
26.8.2. Management of e-commerce activities	356
26.8.3. Multidimensional analysis of data from IoT sensors	357
26.8.4. Big screen service for the Tmall Double 11 Shoppin... ..	362
26.8.5. Mobile data analysis	363
26.9. Restrictions	364
26.10. Terms	364

27.DataQ - Smart Tag Service	366
27.1. What is DataQ - Smart Tag Service?	366
27.2. Benefits	367
27.3. Architecture	367
27.4. Features	368
27.4.1. Tag center	368
27.4.2. Tag factory (DataQ - Smart Tag Service Advanced E...	371
27.4.3. Tag statistic (DataQ - Smart Tag Service Advanced E...	373
27.4.4. Analysis APIs	374
27.4.5. Dashboards (DataQ - Smart Tag Service Advanced E...	375
27.4.6. Tag sync	375
27.5. Scenarios	376
27.5.1. Overview	376
27.5.2. Analysis APIs	376
27.5.3. Device records analysis	376
27.5.4. Geographic analysis	376
27.6. Limits	376
27.7. Terms	376
28.E-MapReduce (EMR)	378
28.1. What is EMR?	378
28.2. Architecture	378
28.3. Benefits	378
28.4. Features	379
28.5. Scenarios	379
28.6. Limits	381
28.7. Terms	381
29.Quick BI	383
29.1. What is Quick BI?	383

29.2. Benefits	383
29.3. Architecture	383
29.4. Features	385
29.5. Scenarios	386
29.5.1. Instant data analysis and effective decision-making	386
29.5.2. Integration with existing systems	387
29.5.3. Permission control of transaction data	388
29.6. Limits	389
29.7. Terms	389
30. Graph Analytics	391
30.1. What is Graph Analytics	391
30.2. Benefits	391
30.3. Product architecture	392
30.3.1. System architecture	392
30.3.2. OLEP model	393
30.4. Features	394
30.4.1. Features	394
30.4.2. Search module	394
30.4.3. Relationship networks	395
30.4.4. Link engine	397
30.5. Scenarios	397
30.5.1. Scenario overview	397
30.5.2. Intelligent relationship networks	398
30.5.3. Industrial risk control	399
30.5.4. Public security protection	400
30.6. Restrictions	401
30.7. Terms	401
31. Machine Learning Platform for AI	404

31.1. What is machine learning?	404
31.2. Benefits	404
31.3. Architecture	406
31.4. Features	407
31.4.1. Visualized modeling	407
31.4.2. All-in-one experience	408
31.4.3. Multiple templates on the homepage	410
31.4.4. Data visualization	410
31.4.5. Model management	411
31.4.6. Multiple algorithm components	411
31.5. Scenarios	419
31.6. Limits	420
31.7. Terms	421
32. Dataphin	423
32.1. What is Dataphin?	423
32.2. Benefits	423
32.3. Features	424
32.4. Functions	425
32.4.1. Overview	425
32.4.2. Resolved issues	426
32.4.3. Platform management	428
32.4.4. Global design	429
32.4.5. Data ingestion	430
32.4.6. Data standardization	430
32.4.6.1. Dimensions	431
32.4.6.2. Business processes	432
32.4.6.3. Atomic metrics	432
32.4.6.4. Business filters	433

32.4.6.5. Derived metrics	433
32.4.7. Modeling	434
32.4.7.1. Overview	434
32.4.7.2. Logical dimension tables	435
32.4.7.3. Logical fact tables	435
32.4.7.4. Logical aggregate tables	436
32.4.7.5. Coding automation	436
32.4.8. Coding	436
32.4.8.1. Overview	436
32.4.8.2. Code editor	437
32.4.8.3. Task scheduling configuration and publishing	437
32.4.8.4. Code management	437
32.4.8.5. Collaborative programming	438
32.4.9. Resource and function management	438
32.4.9.1. Overview	438
32.4.9.2. Resource management	438
32.4.9.3. Function management	439
32.4.10. Scheduling and management	439
32.4.11. Metadata warehouse	440
32.4.12. Data asset management	441
32.4.13. Security management	441
32.4.13.1. Overview	441
32.4.13.2. Permission types	442
32.4.13.3. Permission management	442
32.4.14. Ad hoc query	443
32.5. Scenarios	444
32.6. Limits	444
32.7. Concepts	444

33.Elasticsearch	446
33.1. What is Elasticsearch?	446
33.2. Benefits	446
33.3. Architecture	447
33.4. Features	448
33.5. Scenarios	448
33.6. Limits	449
33.7. Terms	452
34.DataHub	454
34.1. What is DataHub?	454
34.2. Benefits	454
34.3. Architecture	455
34.4. Features	456
34.4.1. Data queue	456
34.4.2. Checkpoint-based data restoration	456
34.4.3. Data synchronization	456
34.4.4. Scalability	457
34.5. Scenarios	457
34.5.1. Overview	458
34.5.2. Data uploading	458
34.5.3. Data collection	458
34.5.4. Realtime Compute	459
34.5.5. Data utilization	459
34.5.6. Data archiving	459
34.6. Limits	459
34.7. Terms	460

1.Apsara Stack introduction

1.1. What is Alibaba Cloud Apsara Stack?

Private cloud

Private cloud is a cloud computing system that is built within enterprises by cloud computing service providers. It places cloud infrastructures, software, and hardware resources within firewalls to allow departments within an organization or an enterprise to share resources in their data centers. It can be managed by an organization or a third party and located within the organization or outside the organization. Compared with public cloud, private cloud provides better privacy and exclusivity.

Private cloud is divided into two types by the enterprise scale or business requirements:

- **Multi-tenant comprehensive private cloud for industries and large groups:** A full stack cloud system created in a top-down manner, which takes the hyper-scale digital applications as a business driver and satisfies IT requirements, such as the continuous integration and development of DevOps applications and operation support of production environments.
- **Single-tenant basic private cloud for small- and medium-sized enterprises and scenarios:** A cloud system that hosts technical systems including large-scale Software as a Service (SaaS) applications, industrial clouds, and large group clouds. It also performs local computing tasks.

Alibaba Cloud Apsara Stack

During the evolution from IT architecture to clouds, more and more enterprises want to have the service experience that is brought by large-scale cloud computing in their own data centers, which is based on the construction requirements, such as security compliance, reuse of existing data centers, and experience localization.

Alibaba Cloud Apsara Stack is an extension of Alibaba Cloud public cloud, which brings the technologies of public cloud to Apsara Stack. By helping enterprises deliver complete and customizable Alibaba Cloud software solutions in their own data centers, Apsara Stack allows you to experience the same characteristics as the hyper-scale cloud computing and big data products provided by Alibaba Cloud public cloud in the local environment. Apsara Stack also provides enterprises with the consistent hybrid cloud experience where you can obtain IT resources as required and guarantee the business continuity.

Service values

Supported by various products and services, based on successful digital practice cases of Alibaba Group, and integrated with the mature solutions and rich experience in various industries, Apsara Stack helps governments and enterprises digitally transform their businesses and services. Apsara Stack provides service values in the following four aspects:

- **Elastic**
Combines all resources into a supercomputer and flexibly scales out resources to minimize costs and maximize performance and stability.
- **Agile**
Integrates business by using Internet and microservice to speed up the innovation of traditional enterprises.
- **Data**
Uses digitalization to allow data to flow between vertical businesses and forms a data shared service to deal with large amounts of data.
- **Smart**
Allows smart transformation of businesses globally and helps reinvent business models.

Platform features

As an enterprise-level cloud platform, Apsara stack has the following three features:

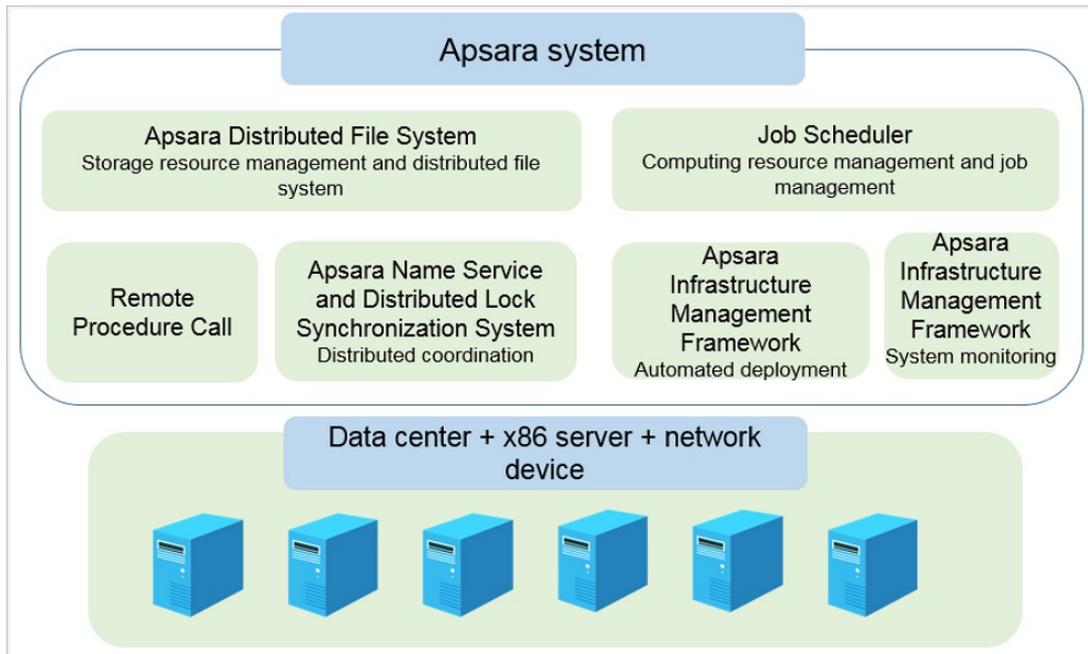
- **Software-defined platform:** masks underlying hardware differences, enables resources to scale up or out as required, and does not affect the performance of upper-layer applications.
- **Production-level reliability and security compliance:** guarantees the continuity and security of enterprise data.
- **Unified access management:** isolates permissions of different roles for easy subsequent operations management.

1.2. The reasons to choose Apsara Stack

1.2.1. Hyper-scale distributed cloud operating system

Apsara Stack is based on the same underlying architecture (large-scale distributed computing system kernel of Apsara) as Alibaba Cloud public cloud. It provides underlying support for upper-layer services in terms of storage, computing, and scheduling. It is a hyper-scale and universal computing operating system that is independently developed by Alibaba Cloud for the global market. Apsara can connect millions of servers all over the world into a supercomputer and provide the community with computing capabilities in the form of online public services. The computing capabilities provided by Apsara are powerful, universal, and beneficial to everyone.

Apsara system kernel architecture



The modules of the Apsara platform kernel have the following primary functions:

- Underlying services for distributed systems

The modules provide the underlying services required in a distributed environment, such as coordination, remote procedure call, security management, and resource management. These services provide support for the upper-layer modules, such as the distributed file system and job scheduling.

- Distributed file system

The modules aggregate storage capabilities from different nodes in a cluster to construct a massive, reliable, and scalable data storage service. The modules also protect against software and hardware faults automatically to guarantee uninterrupted data access. With the support for incremental expansion and automatic data balancing, the modules provide APIs that are similar to Portable Operating System Interfaces (POSIX) of UNIX for accessing the files in the user space. The modules also perform random read/write and append write operations.

- Job scheduling

The modules schedule jobs in cluster systems and support online services that rely heavily on the response speed and offline jobs that require high data processing throughput. The modules detect faults and hot spots in systems automatically and guarantee a stable and reliable job completion in various methods, such as error retries and issuing concurrent backup jobs for long-tail jobs.

- Cluster monitoring and deployment

The modules monitor the status of clusters and also monitor the running status and performance metrics of upper-layer application services to send alert notifications of exception events and keep a record. The modules enable the operations personnel to manage the deployment and configuration of Apsara platform and upper-layer applications. The modules also support online cluster scaling and online update of application services.

1.2.2. Deployment and control system of Apsara Infrastructure Management Framework

Apsara Infrastructure Management Framework provides the cloud services with basic support by supplying unified deployment, authentication, authorization, and control capabilities for cloud service products.

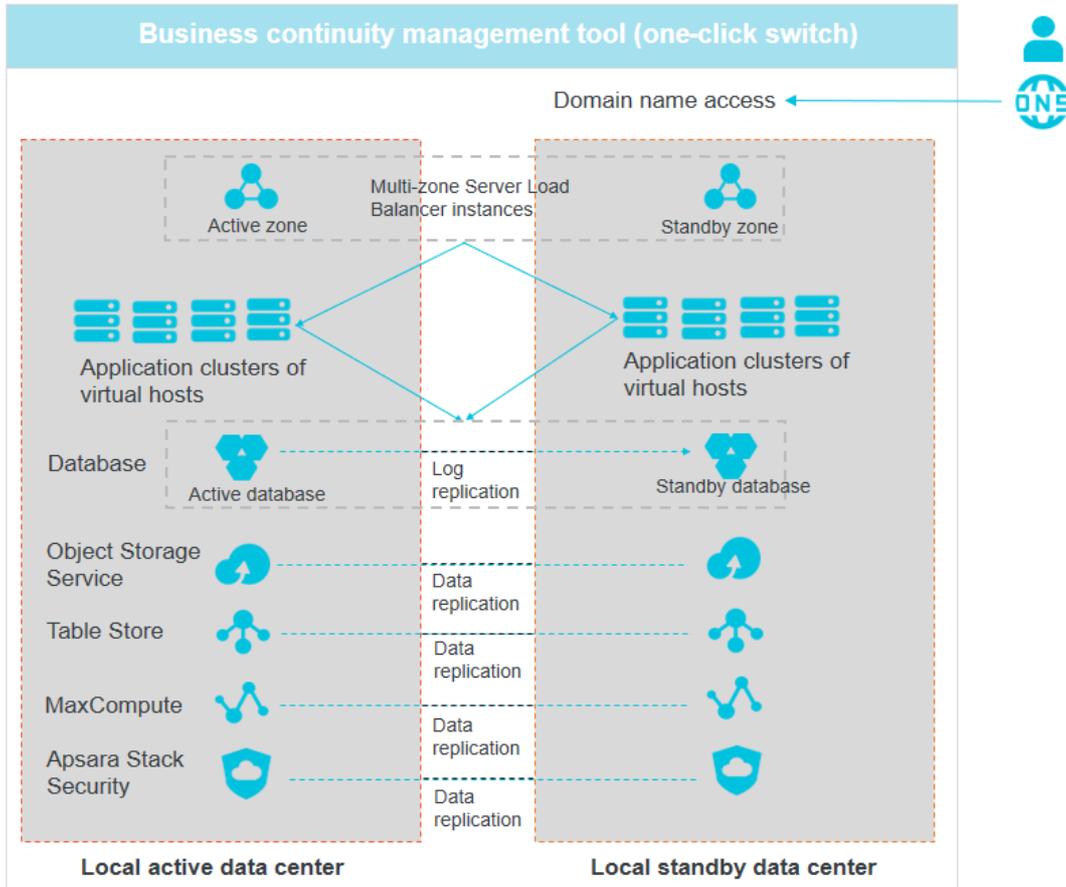
Apsara Infrastructure Management Framework contains various modules, including deployment framework, resource library, meta database, authentication and authorization component, interface gateway, Log Service, and control service module.

- The deployment framework provides all cloud services with unified functions that can deploy access platform and manage the dependencies among services.
- The resource library stores the execution files of all cloud services and their dependent components.
- The authentication and authorization component provides access control for cloud services and supports isolation of multiple tenants.
- The interface gateway provides a unified API management console for all cloud services.
- Log Service stores, retrieves, and obtains logs of cloud services.
- The control service module monitors the basic health status of cloud services and supports the operations system of the cloud platform.

1.2.3. High-reliability disaster recovery solutions

Apsara Stack disaster recovery solutions are designed and developed based on the cloud computing capabilities of Alibaba Cloud. The solutions comply with common international disaster recovery standards. The standby data center must be within a 50-kilometer radius from the active data center in the same city, with a network latency of less than 0.6ms. The Apsara Stack platform deploys the network access layer and user application layer in an active-active mode and the data persistence layer in an active-standby mode.

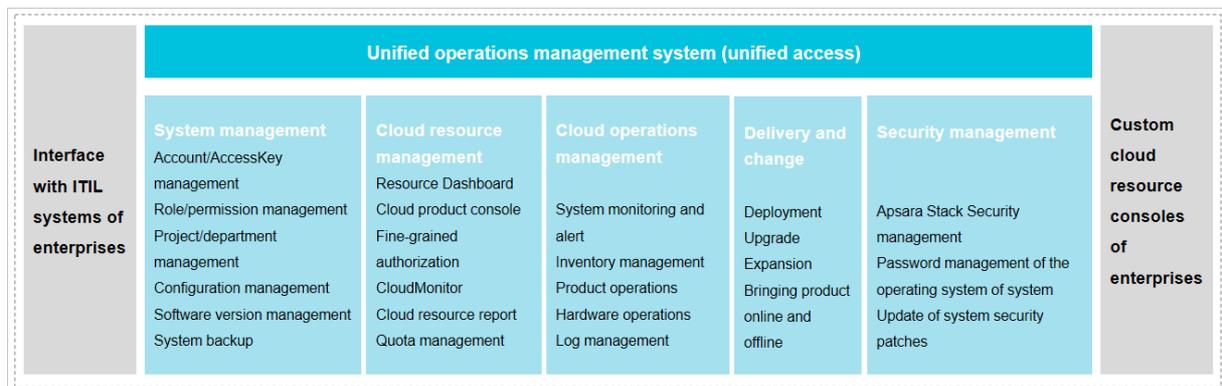
Local disaster recovery



1.2.4. Unified operations management and automated operations capabilities

Apsara Stack provides a unified entry for operations management system to configure different management permissions for different user roles. You can gain operations management capabilities by using APIs and customize your own cloud resource consoles. Apsara Stack also provides the capability to interface with the Information Technology Infrastructure Library (ITIL) systems of enterprises, which aims to interface and integrate synchronously with the existing IT systems of various enterprises.

Unified operations management



1.2.5. Open cloud service interface

Cloud services provide a wide variety of SDKs and RESTful APIs on an OpenAPI platform. You can use these APIs to flexibly access various cloud services provided by Apsara Stack. You can also obtain basic control information about the cloud platform by using these APIs and connect the Apsara Stack platform to your unified control system.

1.3. Product architecture

1.3.1. Types of private cloud architectures

Private cloud architectures have two types: native cloud architecture and integrated cloud architecture.

- Native cloud architecture

The native cloud architecture evolves from the open architecture of Internet and is based on the distributed system framework. It is initially designed to handle big data and host Web applications, and subsequently expands to run basic services.

- Integrated cloud architecture

The integrated cloud architecture focuses on virtualization of computing services. As a breakthrough from the traditional architecture, it is open-sourced by the OpenStack and becomes the mainstream private cloud architecture.

Apsara Stack adopts the native cloud architecture and is based on self-developed distributed technologies and products of Alibaba Cloud. The single system supports all cloud products and services, and enables complete openness of the cloud platform. It comes with comprehensive service features for enterprises, a complete backup capability, and full autonomous control capability.

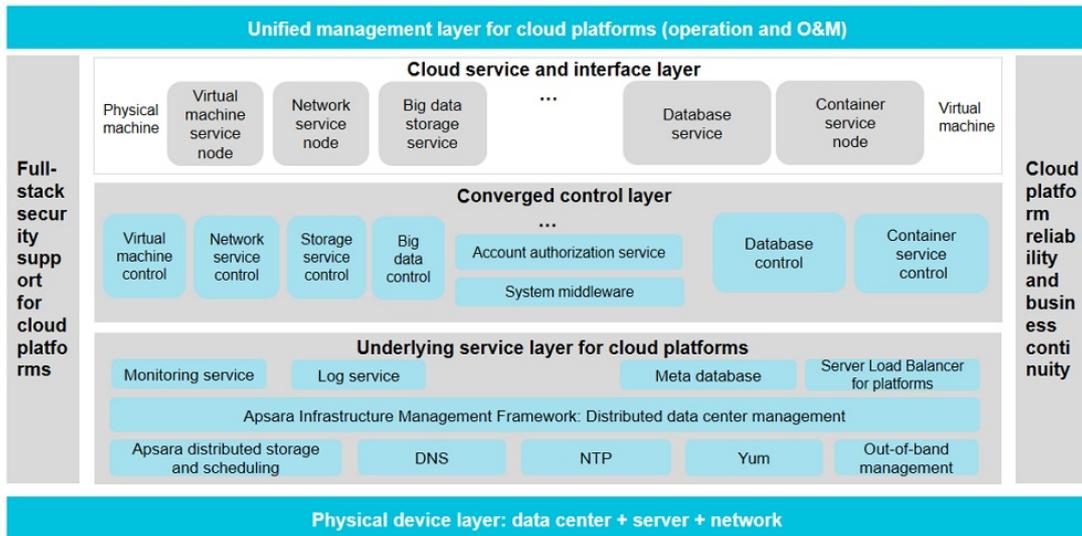
1.3.2. System architecture

The system architecture of Apsara Stack consists of the following parts, as shown in [The system architecture of Apsara Stack](#):

- Physical device layer: includes hardware devices for cloud computing, such as physical data centers, servers, and network.
- Underlying service layer for cloud platforms: bases on the underlying physical environment to provide underlying services for upper-layer applications.
- Hyper-converged control layer: provides unified scheduling for upper-layer applications or services by using the hyper-converged control architecture.
- Cloud service and interface layer: provides unified management and Operation and Maintenance (O&M) for virtual machines and physical machines by using converged service nodes management, and uses the API platform to unify the interfaces and support custom development.
- Unified management layer for cloud platforms: provides unified operation and O&M management.

Apsara Stack also provides full-stack security support and guarantees the reliability of cloud platforms and business continuity.

The system architecture of Apsara Stack

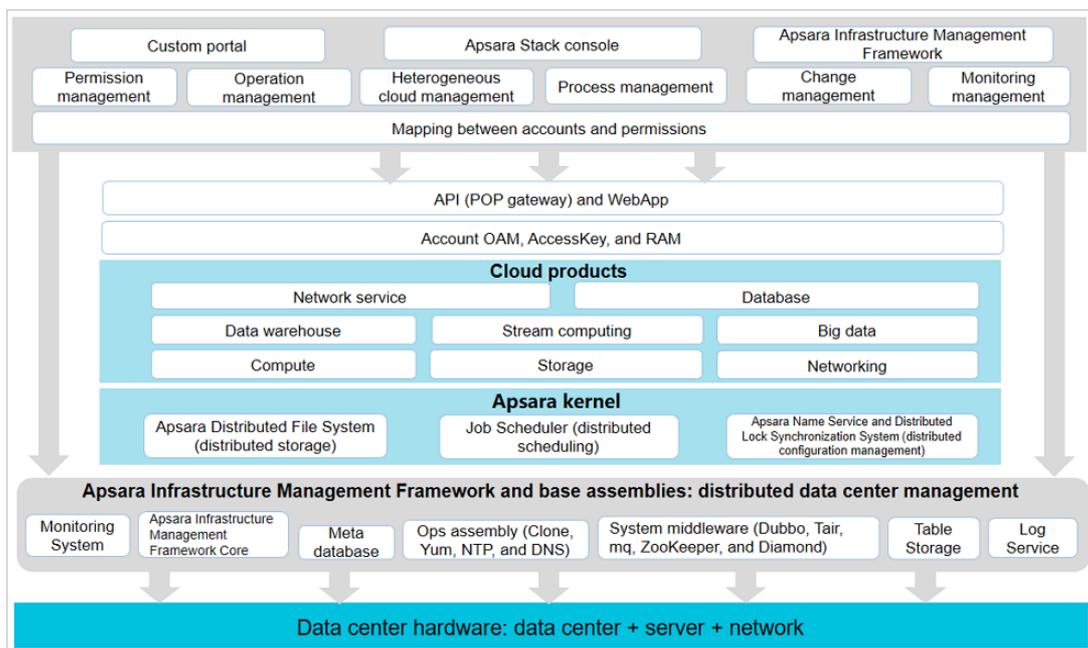


Logical architecture

Apsara Stack virtualizes the computing and storage capabilities of physical servers and network devices to achieve virtual computing, distributed storage, and software-defined networks. On this basis, Apsara Stack provides ApsaraDB and big data processing services. Apsara Stack also provides the supporting capabilities of underlying IT services for your applications, and can be interconnected with your existing account systems and monitoring operations systems. The logical architecture of Apsara Stack has the following characteristics:

- With data center + x86 server + network device as the hardware basis
- Based on the Apsara kernel (distributed engine) to provide various cloud products
- All cloud products are required to follow a unified API framework, O&M (accounts, authorization, monitoring, and logs) and management system, and security system.
- Make sure that all cloud products provide a consistent user experience.

The logical architecture of Apsara Stack



1.3.3. Network architecture

1.3.3.1. Network architecture overview

The network architecture of Apsara Stack defines two logical areas, namely the business service area and the integrated access area, as shown in [Logical areas](#).

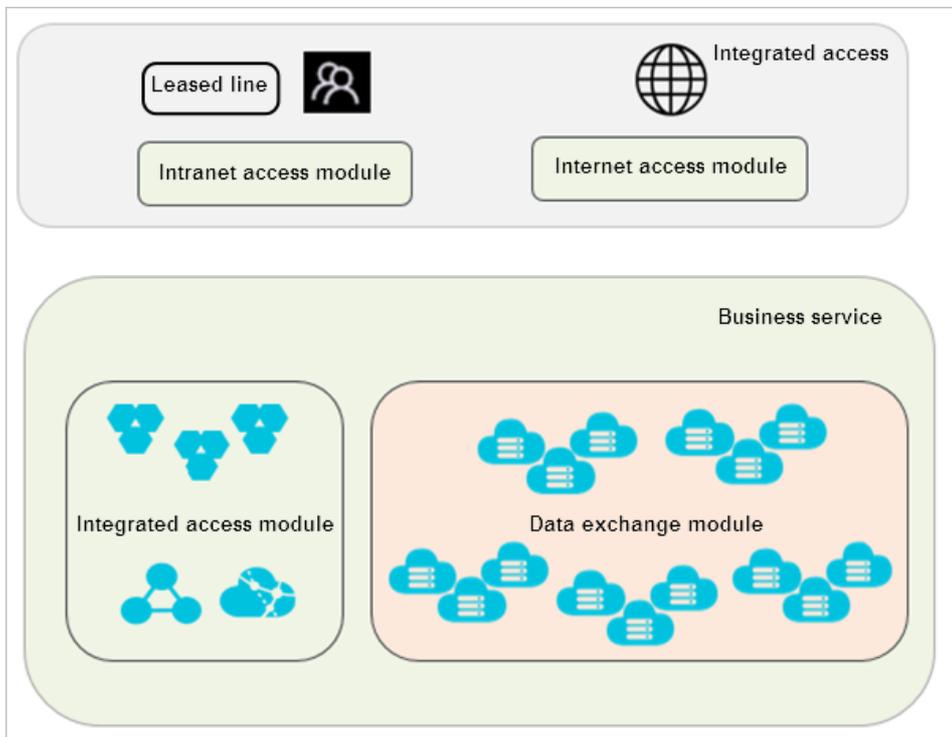
- **Business service area**

This area provides the networks of all cloud services and all cloud service systems exchange internal traffic in this area. This is the core area of Apsara Stack networks.

- **Integrated access area**

This area can be tailored based on the actual deployment requirements. As an extension of the business service area, the integrated access area provides a channel for user management, and the access to Apsara Stack networks by using Internet and user private networks.

Logical areas



The roles and purposes of the switches in each area are as follows:

Role	Module	Function
ISW (interconnected switch)	Internet access module	ISW is an egress switch and provides access to Internet service providers (ISPs) or users' backbone networks.

Role	Module	Function
CSW (intranet access switch)	Intranet access module	CSW facilitates the access to users' internal backbone networks, including the access to Virtual Private Cloud (VPC) by using leased lines. It performs route distribution and interaction between the inside and outside of cloud networks.
DSW (distributed layer switch)	Data exchange module	DSW functions as a core switch to connect each access switch (ASW).
ASW (access switch)	Data exchange module	ASW provides access to Elastic Compute Service (ECS) and is uplinked with the core switch DSW.
LSW (integrated access switch)	Integrated access module	LSW provides access to cloud product services, such as VPC and Server Load Balancer (SLB).

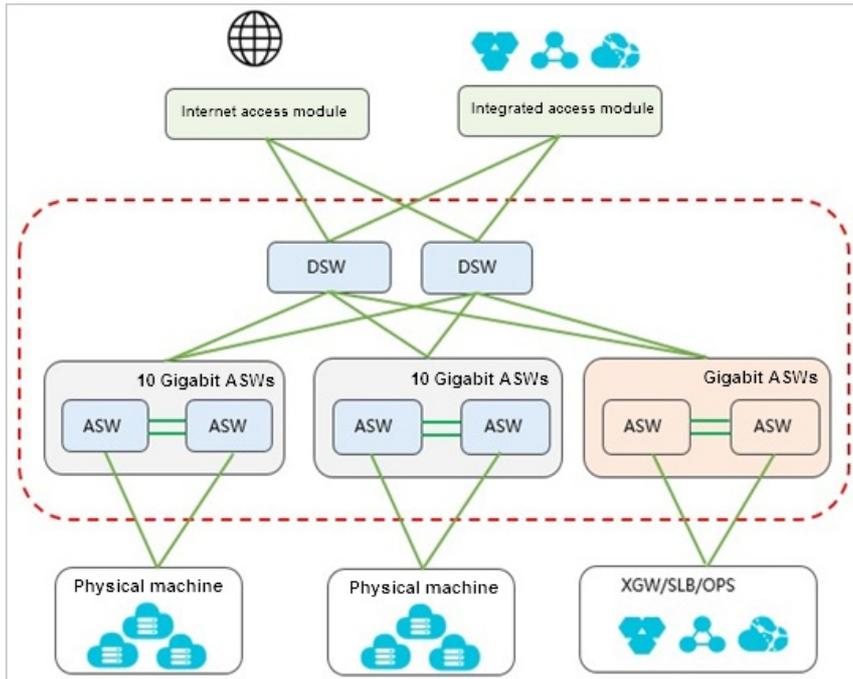
1.3.3.2. Business service area

The business service area consists of the data exchange module and the integrated service module.

- Data exchange module

The data exchange module has a typical layer-2 CLOS architecture that consists of DSWs and ASWs. Each ASW pair forms a stack as a leaf node. According to the network sizes, this node can select data exchange models that have different applicable scopes. All cloud service servers are uplinked with the devices on the ASW stacks. ASWs are connected to DSWs by using External Border Gateway Protocol (EBGP). The DSWs are isolated from each other. The data exchange module is connected to other modules by using EBGP, receives the Internet routes from ISWs, and releases the Classless Inter-Domain Routing (CIDR) block of cloud products to the ISWs.

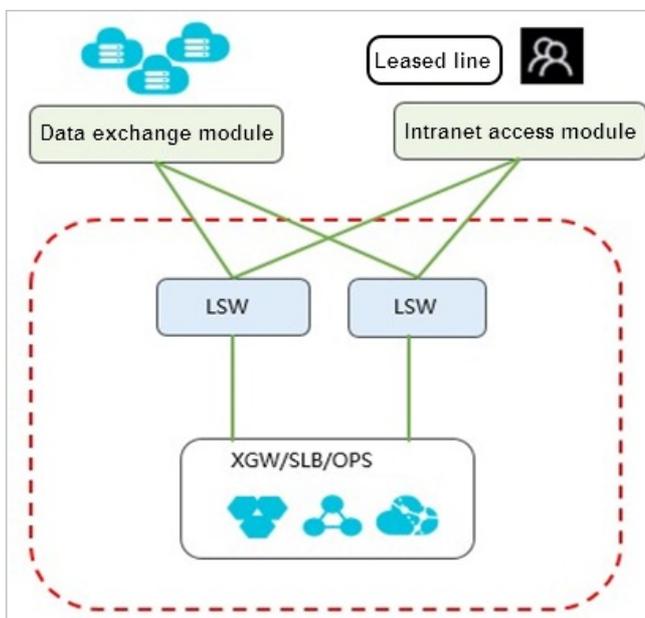
Data exchange module



- Integrated service module

Each cloud service server (XGW/SLB/OPS) is connected to two LSWs. These servers exchange routing information by using Open Shortest Path First (OSPF). The two LSWs exchange routing information between each other by using Internal Border Gateway Protocol (IBGP), and LSW exchange routing information with DSWs and CSWs by using EBGP.

Integrated service module



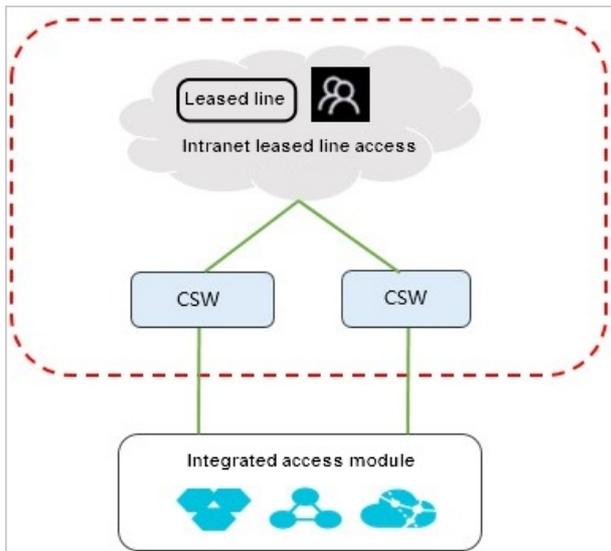
1.3.3.3. Integrated access area

The integrated access area consists of the intranet access module and Internet access module.

- Intranet access module

In the intranet access module, two CSWs provide internal users with access to VPC (Virtual Private Cloud) and general cloud services. For access to VPC, CSWs set up a map from internal users to VPCs and import these users into different VPCs. Different user groups are isolated from each other on CSWs. For access to general cloud services, CSWs are connected to the integrated service module by using External Border Gateway Protocol (EBGP) and allow direct access to all resources in the business service area.

Intranet access module



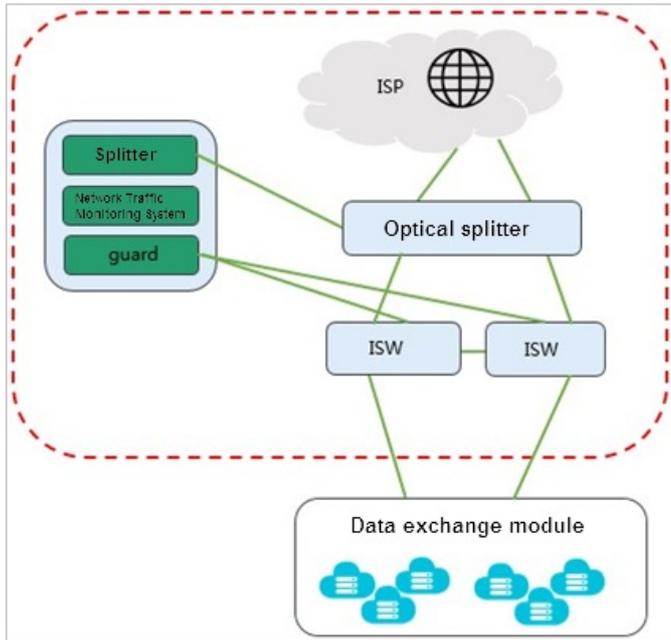
- Internet access module

The Internet access module consists of two ISWs. It facilitates the access to ISPs or users' public backbone networks and performs route distribution and interaction between the inside and outside of cloud networks. The two ISWs run Internal Border Gateway Protocol (IBGP) to back up routes between each other. Based on actual conditions, ISWs can use static routing or EBGP to uplink with Internet service providers (ISPs) or users' public backbone networks. The link bandwidth is defined based on the size of users' Alibaba Cloud networks and the bandwidth of their public backbone networks. We recommend that ISWs can use BGP to connect with multiple carriers to improve the reliability. Each carrier has 2×10 GE lines.

The Internet access module also uses EBGP to exchange routes with the data exchange module, releases relevant Internet routes to the data exchange module, and receives the internal cloud service routes that are sent by the data exchange module to implement the interaction between the inside and outside of cloud networks.

The Internet access module is parallel to an Alibaba Cloud security protection system. The traffic generated by the Internet to cloud networks is diverted to Network Traffic Monitoring System by using an optical splitter. When Network Traffic Monitoring System detects malicious traffic, it releases the corresponding route by using Apsara Stack Security to divert the malicious traffic to Apsara Stack Security for scrubbing. The scrubbed traffic is injected back into the Internet access module.

Internet access module

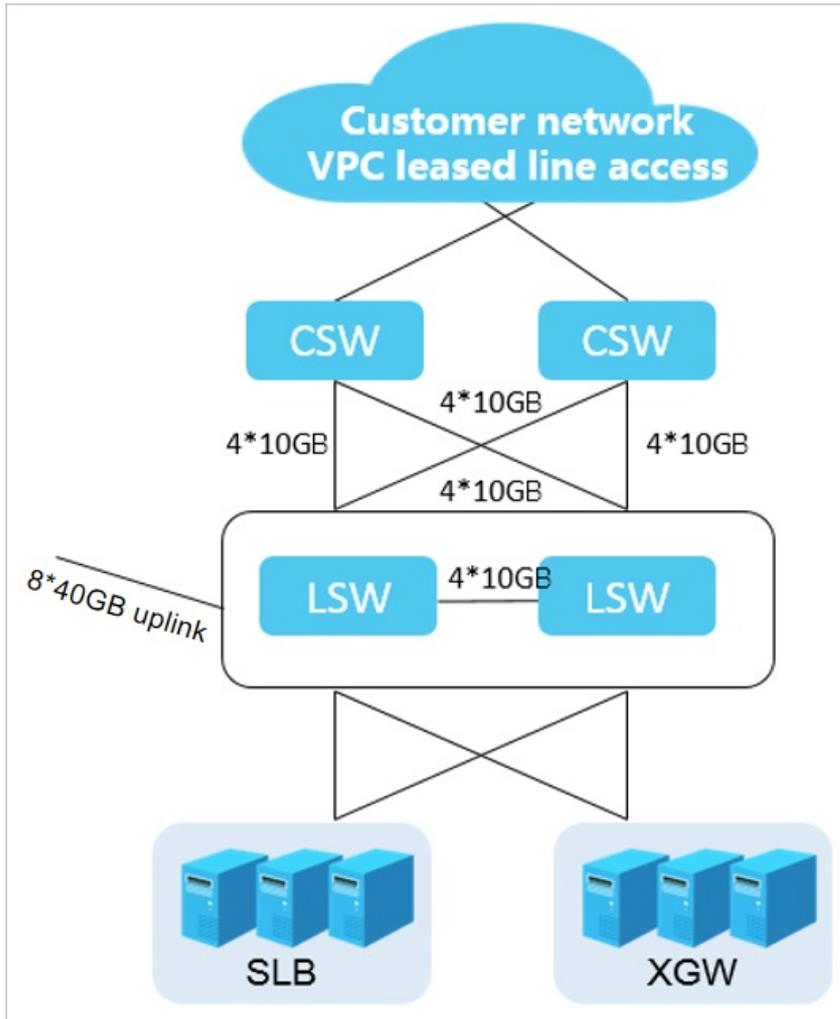


1.3.3.4. VPC leased line access

The leased line access solution of Virtual Private Cloud (VPC) allows you to control over your own virtual networks, such as selecting your own IP address ranges and configuring route tables and gateways. You can also connect your VPC to a traditional data center by using leased lines or VPN connections to create a customized network environment. This enables smooth migration of applications to the cloud.

Each cloud service server (XGW/SLB) is connected to two LSWs. These servers exchange routing information by using Open Shortest Path First (OSPF). Two LSWs exchange routing information between each other by using Internal Border Gateway Protocol (IBGP), and LSWs exchange routing information with CSWs by using External Border Gateway Protocol (EBGP).

VPC leased line access

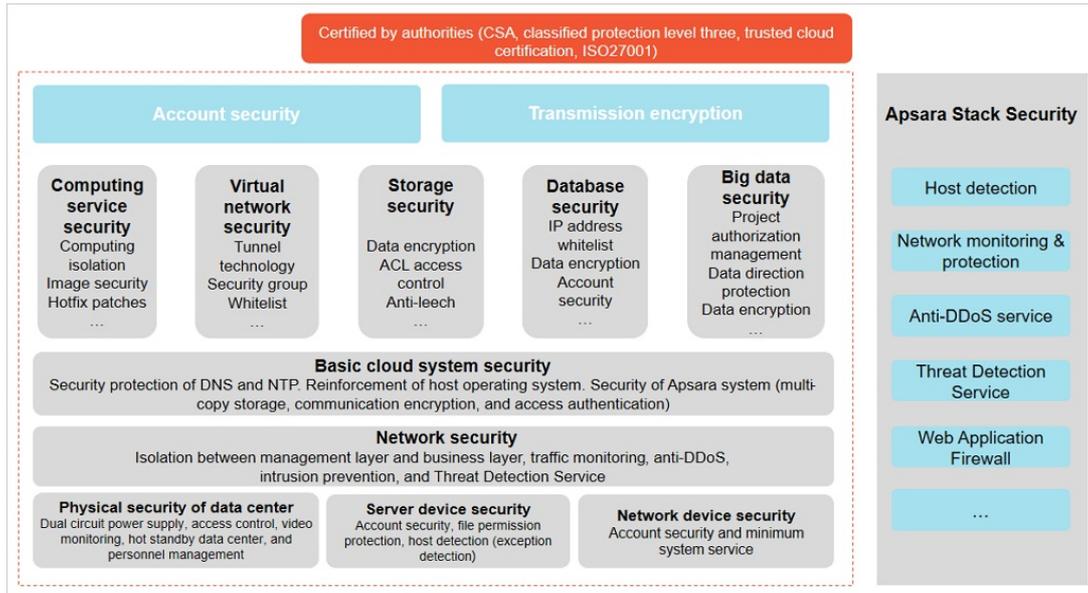


1.3.4. Security architecture

Apsara Stack provides all-around security capabilities from underlying communication protocols to upper-layer applications to guarantee the security of your access and data. Access to every console in Apsara Stack is allowed only with HTTPS certificates. Apsara Stack provides a comprehensive role authorization mechanism to guarantee a secure and controllable access to resources in multi-tenant mode. It supports different security roles, such as security administrator, system administrator, and security auditor.

Apsara Stack has incorporated Apsara Stack Security since the V3 version and provides you with a multi-level and integrated cloud security protection solution.

Hierarchical security architecture of Apsara Stack



1.3.5. Base assembly

Apsara Stack base consists of three types of assemblies, which provides support for the deployment and operations of the cloud platform.

Base assembly

Assembly		Function
Ops assemblies	Yum	<p>Install package</p> <p>The software source is deployed during the initial installation phase. This package is mainly used to install the operating system and deploy application software packages and their dependent components of Apsara Stack, such as the Apsara platform and Elastic Compute Service (ECS), on physical machines.</p>
	Clone	Machine cloning service
	NTP	<p>Clock source service</p> <p>The physical machines deployed on Apsara Stack synchronize time from a standard NTP time source and provide the time to other hosts.</p>
	DNS	<p>Domain name resolution service</p> <p>DNS provides forward and reverse resolution of domain names for the internal Apsara Stack environment. It runs a bind instance on each of the two OPS machines and uses keepalived to provide high-availability services. When one machine fails, the other machine automatically takes over its work.</p>

Assembly		Function
Base middleware	Dubbo	Distributed RPC service
	Tair	Cache service
	mq	Message Queue service
	ZooKeeper	Distributed collaboration
	Diamond	Configuration management service
	SchedulerX	Timing task service
Basic base assemblies	Apsara Infrastructure Management Framework	Data center management
	Monitoring System	Data center monitoring
	OTS-inner	Table Store service
	SLS-inner	Log Service of cloud platform
	Meta database	Meta database
	POP	APIs on the cloud platform
	OAM	Account system
	RAM	Authentication and authorization system
WebApps	Support for the Apsara Stack Operations console	

1.4. Product panorama

Apsara Stack provides a variety of products to meet requirements of different users.

Infrastructure

Apsara Stack provides a wide variety of basic virtual resources, such as virtual computing, virtual network, and virtual scheduling. The main products include Elastic Compute Service (ECS), Virtual Private Cloud (VPC), Server Load Balancer (SLB), Container Service, Auto Scaling, and Key Management Service (KMS).

Storage products

Apsara Stack provides various storage products for different storage objects. The main products include Object Storage Service (OSS), Network Attached Storage (NAS), Table Store, and Apsara File Storage for HDFS (HDFS).

Internet middleware and applications

Apsara Stack provides middleware services and can host various customer applications. This facilitates the conversion of applications to services and encourages applications to evolve into a microservice architecture. The main products include API Gateway, Log Service, and Domain Name System (DNS).

Database

Apsara Stack provides diversified data engines. These data engines can interoperate with each other. The main products include ApsaraDB for RDS, KVStore for Redis, KVStore for Memcache, ApsaraDB for MongoDB, Data Transmission Service (DTS), and Data Management Service (DMS).

Big Data Processing

Apsara Stack provides various big data capabilities of analysis, application, and visualization, which gives value to data. The main products include MaxCompute, DataWorks, Realtime Compute, Quick BI, E-MapReduce (EMR), Graph Analytics, DataQ-Smart Tag Service, Dataphin, and Apsara Bigdata Manager (ABM).

Artificial intelligence

Apsara Stack provides a machine learning algorithm platform based on the distributed computing engine developed by Alibaba Cloud, such as Machine Learning (PAI).

Security

Apsara Stack provides an all-around protection from underlying communication protocols to upper-layer applications, which guarantees the security of your access and data. For example, the product Apsara Stack Security.

1.5. Scenarios

Apsara Stack provides flexible and scalable industrial solutions for users who are from different scales and different sectors. Based on the business traits of different sectors, such as industry, agriculture, transportation, government, finance, and education, Apsara Stack creates custom solutions to provide users with one-stop products and services. This topic focuses on introducing the following two scenarios.

City Brain

Urban management is a field that involves one of the largest volumes of data in China. This marks the transition of governmental information from a closed-flow model to an open-flow online model. With more time and space to flow in, urban data has a higher value. Cloud computing becomes an urban infrastructure, data becomes a new means of production and a strategic resource, and AI technology becomes the nerve center of a smart city. All of these forms the City Data Brain.

The values and features are as follows:

- A breakthrough of urban governance mode. With the urban data as a resource, City Brain improves the government management capabilities, resolves prominent issues of urban governance, and achieves an intelligent, intensive, and humane form of governance.
- A breakthrough of urban service mode. City Brain provides more accurate and convenient services for enterprises and individuals, makes the urban public services more efficient, and saves more public resources.
- A breakthrough of urban industrial development. City Brain lays down an industrial AI layout, takes open urban data as an important fundamental resource, drives the development of industries, and promotes the transformation and upgrade of traditional industries.

Finance Cloud

Finance Cloud is an industrial cloud that serves financial organizations, such as banks, security agencies, insurance companies, and funds. It relies on a cluster of independent data centers to provide cloud products that meet the regulatory requirements of the People's Bank of China, China Banking Regulatory Commission (CBRC), China Securities Regulatory Commission (CSRC), and China Insurance Regulatory Commission (CIRC). It also provides more professional and comprehensive services for financial customers. Enterprises can build Finance Cloud independently or with Alibaba Cloud. Finance Cloud meets the requirements of large- and medium-sized financial organizations for independent cloud data centers that are completely physically isolated. It can also output the cloud computing and big data platform to customers' data centers.

The values and features are as follows:

- Independent resource clusters
- Stricter data center management
- Better disaster recovery capability
- Stricter requirements for network security isolation
- Stricter access control
- Compliance with the security supervision requirements and compliance requirements of banks
- Dedicated security operation team, security compliance team, and security solution team of the Finance Cloud sector
- Dedicated account managers and cloud architects of Finance Cloud
- Stricter user access mechanism

1.6. Compliance security solution

1.6.1. Overview

On June 1, 2017, the *Cybersecurity Law of the People's Republic of China* was officially implemented, which has made clear provisions for classified protection compliance. Drawing on its technical advantages on Apsara Stack Security products, Alibaba Cloud builds a classified protection compliance ecosystem to help you quickly align with the provisions for classified protection compliance. Alibaba Cloud works with its cooperative assessment agencies and security consulting providers based worldwide to offer one-stop classified protection assessment services. It offers complete attack protection, data auditing, encryption, and security management that make it easier for you to quickly pass the classified protection compliance assessment.

1.6.2. Interpretation on key points

Network and communication security

Interpretation on clauses

- Divide the network into different security domains according to different server roles and server importance.
- Set access control policies at the security domain boundary between the intranet and Internet, which must be configured on specific ports.
- Deploy intrusion prevention measures at the network boundary to prevent against and record intrusion behaviors.
- Record and audit the user behavior logs and security events in the network.

Coping strategies

- We recommend that you use Virtual Private Cloud (VPC) and security group of Alibaba Cloud to divide a network into different security domains and perform reasonable access control.
- You can use Web Application Firewall (WAF) to prevent network intrusion.
- You can use the log feature to record, analyze, and audit user behavior logs and security events.
- If the system is frequently threatened by DDoS attacks, you can use Anti-DDoS Pro to filter and scrub abnormal traffic.

Device and computing security

Interpretation on clauses

- Avoid account sharing, record, and audit operations actions, which is an elementary security requirement.
- Secure system layer with necessary security measures and prevent servers from intrusions.

Coping strategies

- You can audit the server and data actions, and create an independent account for each operations personnel to avoid account sharing.
- You can use Server Guard to conduct complete vulnerability management, baseline check and intrusion prevention on servers.

Application and data security

Interpretation on clauses

- An application directly implements specific business and is not like the network and system with relative standard characteristics. The functions of most applications such as identity authentication, access control and operation audit are difficult to be replaced by third-party products.
- Encryption is the most effective method to secure data integrity and confidentiality except security prevention methods at other levels.
- Remote data backup is one of the most important requirements that distinguishes the third level of classified protection from the second level. It is also the most basic technical safeguard measure for business sustainability.

Coping strategies

- At the beginning of the application development, application functions such as identity authentication, access control, and security audit must be considered.
- For online systems, you can add functions such as account authentication, user permission identification, and log auditing to satisfy classified protection requirements.
- For data security, HTTPS can be used to guarantee that data remains encrypted in the transmission process.
- For data backup, we recommend that you can use remote disaster recovery instance of ApsaraDB for RDS to automatically back up data and manually synchronize backup files of database to Alibaba cloud servers in other regions.

Security management policies

Interpretation on clauses

- Security policy, regulation, and management personnel are significant bases for sustainable security. Policy guides the security direction. Regulation specifies the security process. Management personnel fulfills the security responsibilities.
- Classified protection requirements provide a methodology and best practice. You can perform continuous security construction and management according to the classified protection methodology.

Coping strategies

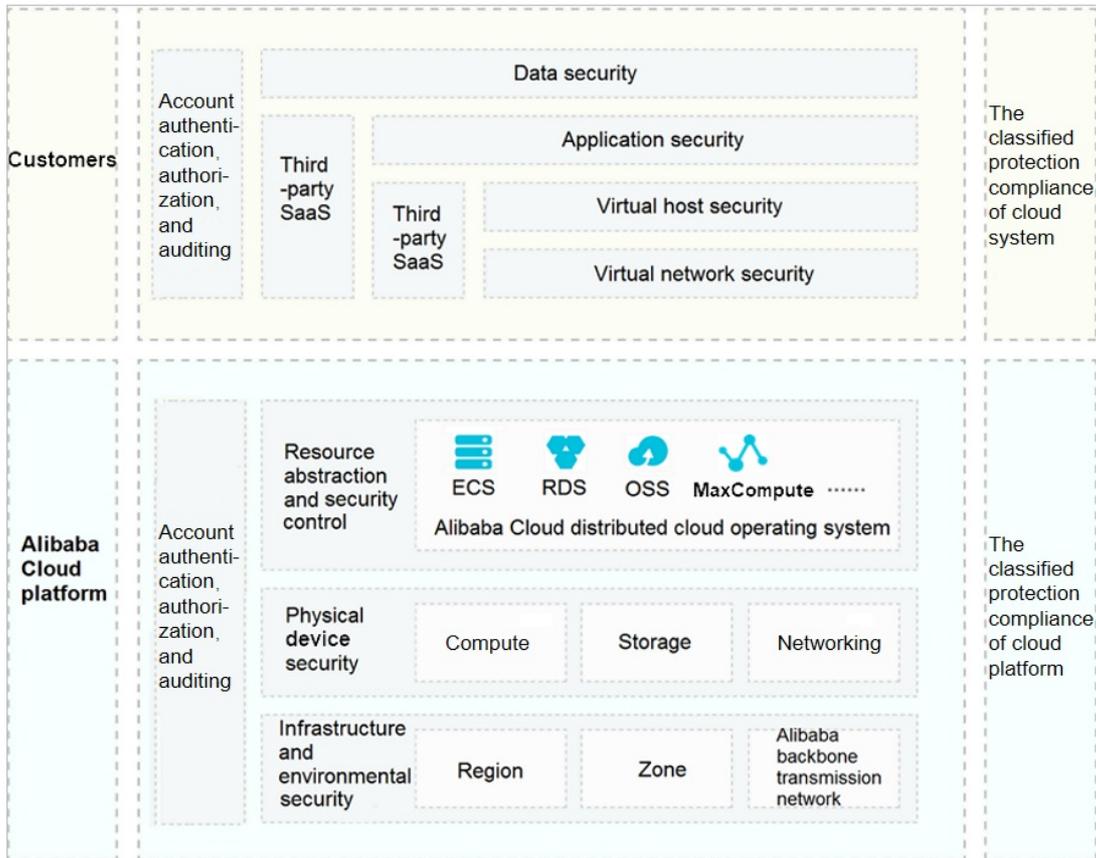
- The customer management staff can arrange, prepare, and fulfill the security policy, regulation, and management personnel according to the actual condition of enterprise and form specialized documents.
- For the technical means required in the process of vulnerability management, we recommended that you can use Alibaba Cloud Server Guard to quickly detect the vulnerabilities of cloud system and resolve them in time.

1.6.3. Cloud-based classified protection compliance

Shared compliance responsibilities

The Alibaba Cloud platform and the cloud tenant systems are classified and assessed respectively. Assessment conclusions of the Alibaba Cloud platform can be used for the tenant systems.

Shared compliance responsibilities



Alibaba Cloud provides the following contents:

- Classified protection filing certification of the Alibaba Cloud platform
- Key pages of Alibaba Cloud assessment report
- Sales license of Apsara Stack Security
- Description of partial assessment items of Alibaba Cloud

More details about shared responsibilities are as follows:

- Alibaba Cloud is the unique cloud service provider in China that participates in and passes the pilot demonstration of cloud computing classified protection standard. Public Cloud and E-Government Cloud pass the filing and assessment of the third level of classified protection. Finance Cloud passes the filing and assessment of the fourth level of classified protection.
- According to regulatory authority, assessment conclusions of physical security, partial network security, and security management can be reused for the tenant systems on Alibaba Cloud for classified protection assessment, and Alibaba Cloud can provide supporting details.
- With the complete security technology and management architecture, and the protection system of Apsara Stack Security, Alibaba Cloud platform makes it easy for tenants to pass

classified protection assessment.

Classified protection compliance ecology

Current conditions of cloud-based classified protection are as follows:

- Most tenants do not know classified protection.
- Most tenants do not know how to start with classified protection.
- Most tenants are not good at communicating with supervision authorities.
- Security systems lag behind business development.

Alibaba Cloud establishes Classified Protection Compliance Ecology to provide one-stop classified protection compliance solutions for cloud-based systems to quickly pass classified protection assessment.

Classified protection compliance ecology



Work division of classified protection:

- Alibaba Cloud: integrates capabilities of service agencies and provides security products
- Consulting firm: provides technical support and consulting services in the whole process
- Assessment agency: provides assessment services
- Public security organ: reviews filing and supervises services

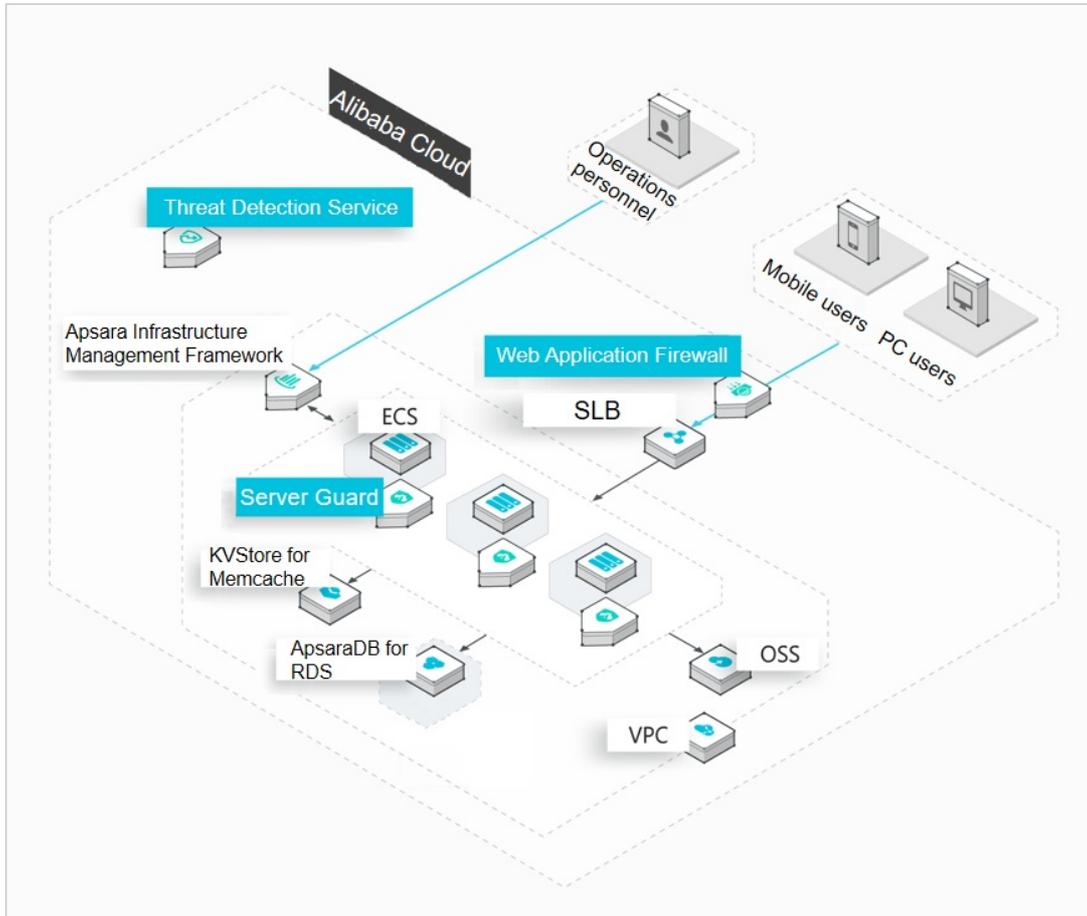
1.6.4. Classified protection implementation process

	Operating unit	Alibaba Cloud	Consulting or assessment agency	Public security organ
System rating	Determine the class of security protection and write rating report	Coordinate the third party agency to provide counseling services for operating unit	Counseling the operating unit to prepare the rating materials and organize expert review (level three)	None
System filing	Prepare and present the filing materials to the local public security organ	Coordinate the third party agency to provide counseling services for operating unit	Counseling the operating unit to prepare the filing materials and to issue filing	None
Construction rectification	Construct the security technology and management system in line with class requirements	Provide the obligatory security products and services that meet the class requirements	Counseling the operating unit to carry out system security reinforcement and develop security management regulation	The local public security organ reviews and accepts the filing materials
Rating assessment	Prepare for and accept the assessment from the assessment agency	Provide the cloud service provider's security qualification and the proof that the cloud platform has passed the classified protection	The assessment agency assesses the system class conformity	None
Supervision & inspection	Accept the regular inspection of public security organ	None	None	Supervise and inspect the operating unit to carry out the class protection work

1.6.5. Security and compliance architecture

Rapid connection to Apsara Stack Security and fast debugging. Compliant with basic technical requirements for classified protection at minimal security costs.

Security compliance architecture



Basic requirements of classified protection are as follows:

- **Physical and environmental security:** includes data center power supply, temperature and humidity control, and prevention of wind, rain, and lightning. Assessment conclusions of Alibaba Cloud can be reused.
- **Network and communication security:** includes network architecture, boundary protection, access control, intrusion prevention, and communication encryption.
- **Device and computing security:** includes intrusion prevention, malicious code prevention, identity authentication, access control, centralized control, and security auditing.
- **Application and data security:** includes security auditing, data integrity, and data confidentiality.

1.6.6. Solution benefits

One-stop assessment service of classified protection

Select high-performance consulting and assessment partners to provide one-stop compliance support throughout, allowing the operators to achieve significant cost savings.

- Eliminates multi-level communication and work redundancy to help the operators reduce investment.
- Improves efficiency by shortening the assessment cycle to as short as two weeks.
- Alibaba Cloud provides best practices of security and compliance on the cloud.

A complete security protection system

With a complete Apsara Stack Security architecture, operators can locate corresponding products on Alibaba Cloud, rectify non-conformances, and meet all requirements of classified protection.

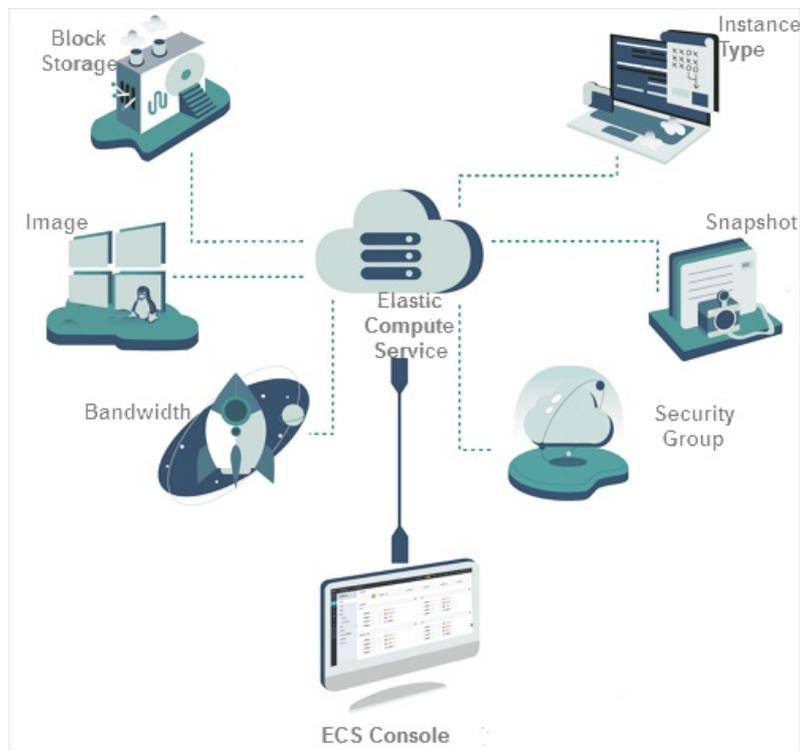
2. Elastic Compute Service (ECS)

2.1. What is ECS?

Elastic Compute Service (ECS) is a computing service that features elastic processing capabilities. Compared with physical servers, ECS instances are more user-friendly and can be managed more efficiently. You can create instances, resize disks, and add or release any number of ECS instances at any time based on your business needs.

An ECS instance is a virtual computing environment that contains the most basic components of computers such as the CPU, memory, and storage. Users perform operations on ECS instances. Instances are core components of ECS, and operations can be performed on instances through the ECS console. Other resources, such as block storage, images, and snapshots, can only be used after they are integrated with ECS instances. For more information, see [ECS components](#).

ECS components



2.2. Benefits

ECS has the following benefits compared with the services provided by other service vendors and common IDCs:

- **High availability**
- **Security**
- **Scalability**

High availability

Compared with the services provided by common IDCs and server vendors, ECS adopts more stringent IDC standards, server access standards, and O&M standards to ensure data reliability, and high availability at the infrastructure and instance levels.

Apsara Stack provides you with the following support services:

- Products and services to improve availability. These include ECS, Server Load Balancer (SLB), multi-backup database, and Data Transformation.
- Industry partners and ecosystem partners that help you build a more advanced and stable architecture and ensure service continuity.
- Diverse training services to help you achieve high availability from the business end to the underlying basic service end.

Security

Security and stability are two of the primary concerns for any cloud service user. Alibaba Cloud has recently passed a host of international information security certifications which demand strict confidentiality of user data and user information and user privacy protection, including ISO27001 and MTCS.

- **With a simple configuration to connect your business environment to global IDCs, Apsara Stack Virtual Private Cloud (VPC) can increase the flexibility, scalability, and stability of your business.**
- **You can connect your own IDC to Apsara Stack VPC through a leased line to implement a hybrid cloud. You can use a variety of hybrid cloud architectures to provide network services and robust networking. A superior business ecosystem is made possible with the ecosystem of Apsara Stack.**
- **VPCs are more stable and secure.**
 - **Stability:** After constructing your VPC, you can update your network architecture and obtain new functions daily to constantly evolve your network infrastructure and ensure your business is always running smoothly. VPCs allow you to divide, configure, and manage your network as needed.
 - **Security:** VPCs feature traffic isolation and attack isolation to protect your services against endless attacks on the Internet. The first line of defense against malicious attacks and traffic is established when you build your VPC.

VPCs provide a stable, secure, controllable, and fast-deliverable network environment. The capability and architecture of VPC hybrid cloud bring the technical advantages of cloud computing to enterprises in traditional industries not engaged in cloud computing.

Elasticity

Elasticity is a key benefit of cloud computing.

- **Elastic computing**

- **Vertical elasticity involves modifying the configurations of a server.** In a traditional IDC, it is difficult to change the configuration of a single server. However, you can change the capacity of your ECS instance or storage service based on the actual needs of your business.
- **Transverse elasticity allows for re-division of resources between applications.** For example, a traditional IDC may not be able to immediately provide sufficient resources for online gaming or live video streaming applications during peak hours. The elasticity of cloud computing makes it possible to provide the resources required in peak hours. When the load returns to normal levels, you can release unnecessary resources to reduce operation costs. The combination of ECS vertical and transverse elasticity enables resources to scale up and down by specified quantity as scheduled or against business load.

- **Elastic storage**

Apsara Stack provides elastic storage. For a traditional IDC, you must add servers to increase the storage space. However, the number of servers that you can add is limited. Cloud computing can provide you with large-capacity storage. You can purchase the storage you need at any time.

- **Elastic network**

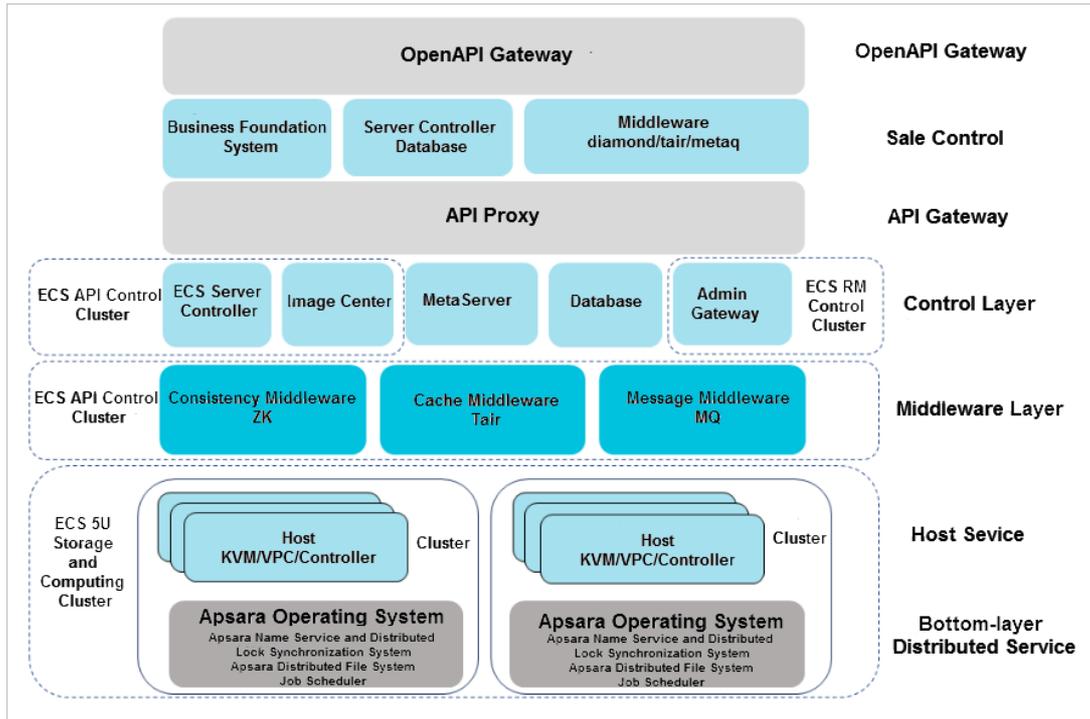
Apsara Stack features elastic network as well. Apsara Stack VPCs can be configured to match the specifications of your own IDCs. In addition, VPCs provide the following benefits: The IDCs can communicate with each other while being isolated through security domains. VPC configurations and planning are flexible.

To sum up, Apsara Stack provides elastic computing, storage, networking, and business architecture planning and allows you to combine your businesses as needed.

2.3. Architecture

ECS is built on the Apsara system that was developed by Alibaba Cloud. The individual ECS instances are virtualized by using KVM while storage is implemented on Apsara Distributed File System.

ECS architecture



Architecture description

Component	Description
Apsara Name Service and Distributed Lock Synchronization System	A basic module that provides services related to distribution consistency in Apsara Stack. As a key distributed coordination system of Apsara Stack, it provides three types of basic services: distributed lock services, subscription and notification services, and lightweight metadata storage services.
Apsara Distributed File System	A distributed storage system developed by Alibaba Cloud. As of 2017, hundreds of clusters and hundreds of thousands of storage nodes using Apsara Distributed File System have been deployed in the production environments. Apsara Distributed File System manages tens of exabytes of disk space.
Job Scheduler	A distributed resource scheduler that manages and allocates resources in the distributed systems.
Server Controller	The ECS scheduling system that schedules storage, network, and computing resources in a unified manner and produces virtual machines (ECS instances) deliverable to users.
Scheduling process	API > Business layer > Control system > Host service
OpenAPI Gateway	Provides services such as authentication and request forwarding.

Component	Description
Business Foundation System	Creates and releases instances and snapshot policies, processes sales requests, and provides APIs to users.
API Proxy	Forwards requests to services of the region corresponding to region_id.
Server Controller databases	Stores control data and status data.
Server Controller	A center for scheduling storage, network, and computing resources.
Tair	Provides cache services for Server Controller.
Zookeeper	Provides the distributed lock service for Server Controller.
MQ	Virtual machine status message queue service.
Image Center	Provides image management services such as import and copy.
MetaServer	Provides metadata management services for ECS instances.
Host service	Provides services such as KVM for computing virtualization, VPC for network virtualization, and control through interaction with Libivrt.
Admin Gateway (AG)	Functions as the bastion host used to log on to an NC during O&M management.
ECS Decider	Determines on which NC to deploy ECS.

2.4. Features

2.4.1. Instances

2.4.1.1. Overview

An ECS instance is a virtual machine that contains basic computing components such as the CPU, memory, operating system, and network. You can fully customize and modify all configurations of an ECS instance. After logging on to Apsara Stack console, you can manage resources and configure the environment of your ECS instances.

2.4.1.2. Instance type families

An instance is the smallest unit in ECS that can provide computation services. The computing capabilities of an instance are determined by its instance type. ECS instances can be divided into multiple families based on their configurations and application scenarios.

Note The instance type families and instance types described in this topic are for reference only. The specific configurations of your instances are determined by the physical servers that the instances are hosted on.

Instance type family	Description	Scenario
N4	<p>N4 instances are commonly shared instances with the following features:</p> <ul style="list-style-type: none"> • 1:2 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2680 v3 (Broadwell) processors. • Latest DDR4 memory. • I/O-optimized by default. 	<p>Small and medium Web servers, batch processing, advertising services, and distributed analysis.</p>
MN4	<p>MN4 instances are balanced and shared instances with a greater processor to memory ratio. mn4 instances have the following features:</p> <ul style="list-style-type: none"> • 1:4 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2680 v3 (Broadwell), Intel Xeon E5-2680 v4 (Haswell), Intel Xeon E5-2682 v4 (Broadwell), or Intel Xeon E5-2650 v2 (Haswell) processors. • Latest DDR4 memory. • I/O-optimized by default. 	<p>Medium Web servers, batch processing, advertising services, distributed analysis, and Hadoop clusters.</p>
XN4	<p>XN4 instances are compact and shared instances with the following features:</p> <ul style="list-style-type: none"> • 1:1 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2680 v4 (Haswell) or Intel Xeon E5-2682 v4 (Broadwell) processors. • Latest DDR4 memory. • I/O-optimized by default. 	<p>Minisite Web applications, small databases, development or testing environments, and code storage servers.</p>

Instance type family	Description	Scenario
E4	<p>E4 instances are memory-optimized instances with the following features:</p> <ul style="list-style-type: none"> • 1:8 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2680 v4 (Broadwell), Intel Xeon E5-2680 v3 (Broadwell), Intel Xeon E5-2650 v2 (Haswell), or Intel Xeon E5-2682 v4 (Broadwell) processors. • I/O-optimized by default. 	<p>Applications that involve a large number of memory operations, queries, and computations, such as Cache/Redis, search applications, and in-memory databases.</p>
SN1NE, compute-optimized type family with enhanced network performance	<ul style="list-style-type: none"> • 1:2 processor to memory ratio. • Ultra high PPS rates. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors with stable computing performance. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Applications that require high PPS rates, such as live commenting on videos and telecom service forwarding. • Web front-end servers. • Front ends of massively multiplayer online (MMO) games. • Data analysis, batch computation, and video encoding. • High-performance scientific and engineering applications.
SN2NE, general-purpose type family with enhanced network performance	<ul style="list-style-type: none"> • 1:4 processor to memory ratio. • Ultra high PPS rates. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors with stable computing performance. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Applications that require high PPS rates, such as live commenting on videos and telecom service forwarding. • Enterprise applications of various types and scales. • Large and medium database systems, caches, and search clusters. • Data analysis and computation. • Computing clusters and data processing that depend on memory.

Instance type family	Description	Scenario
<p>SE1NE, memory-optimized type family with enhanced network performance</p>	<ul style="list-style-type: none"> • 1:8 processor to memory ratio. • Ultra high PPS rates. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors with stable computing performance. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Applications that require high PPS rates, such as live commenting on videos and telecom service forwarding. • High performance databases and large memory databases. • Data analysis and mining, and distributed memory cache. • Hadoop, Spark, and other enterprise-level applications with large memory requirements.
<p>SE1, memory-optimized type family</p>	<ul style="list-style-type: none"> • Stable computing performance. • 1:8 processor to memory ratio. • 2.5 GHz Intel Xeon, E5-2682 v4 (Broadwell), or E5-2680 v4 (Haswell) processors. • Latest DDR4 memory. • Higher instance offering matching higher network performance. • I/O-optimized by default. 	<p>SE1 instances are memory-dedicated instances with a greater memory-to-CPU ratio. se1 instances are applicable to computing scenarios with fixed performance. These scenarios involve Cache, Redis, searching, memory databases, databases with high I/O (Oracle and MongoDB), Hadoop clusters, and involve large-volume data processing.</p>
<p>EBMG5, general-purpose ECS Bare Metal (EBM) Instance type family</p>	<ul style="list-style-type: none"> • 1:4 processor to memory ratio. • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96-core vCPU, up to 2.9 GHz Turbo Boost. • High network performance with 4,500,000 PPS. • Support for SSD cloud disks and ultra cloud disks. 	<ul style="list-style-type: none"> • Deploy Apsara Stack services such as OpenStack and ZStack. • Deploy services such as Docker Container. • Applications that require high PPS rates, such as live commenting on videos and telecom service forwarding. • Enterprise applications of various types and scales. • Large and medium database systems, caches, and search clusters. • Data analysis and computation. • Computing clusters and data processing that depend on memory.

Instance type family	Description	Scenario
I2, a type family with local SSDs	<ul style="list-style-type: none"> • High-performance local NVMe SSDs with high IOPS, high I/O throughput, and low latency. • 1:8 processor to memory ratio, designed for high-performance databases. • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors. • Higher instance offering matching higher network performance. 	<ul style="list-style-type: none"> • OLTP and high-performance relational databases. • NoSQL databases such as Cassandra and MongoDB. • Search scenarios, such as elastic search.
D1, big data type family	<ul style="list-style-type: none"> • High-capacity local SATA HDDs with high throughput and 17 Gbit/s of maximum bandwidth between instances. • 1:4 processor to memory ratio, designed for big data scenarios. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Hadoop MapReduce, HDFS, Hive, and HBase. • Spark in-memory computing and MLlib. • Enterprises that require big data computing and storage analysis, such as those in the Internet and finance industries, to store and compute large volumes of data. • Scenarios involving elastic search and logs.
SCCG5IB, general-purpose Super Computing Cluster (SCC) type family	<ul style="list-style-type: none"> • 1:8 processor to memory ratio. • Ultra high packet forwarding rate. • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors with stable computing performance. • 100G IB network with ultra high bandwidth and ultra low latency. 	<ul style="list-style-type: none"> • Data analysis and computation. • Artificial intelligence computation. • Manufacturing simulation. • High-performance computing clusters. • Genetic analysis. • Pharmaceutical analysis.
SCCG5IB, SCC instance type family with a high operating frequency	<ul style="list-style-type: none"> • 1:6 processor to memory ratio. • Ultra high packet forwarding rate. • 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors with stable computing performance. • 100G IB network with ultra high bandwidth and ultra low latency. 	<ul style="list-style-type: none"> • Data analysis and computation. • Artificial intelligence computation. • Manufacturing simulation. • High-performance computing clusters. • Genetic analysis. • Pharmaceutical analysis.

Instance type family	Description	Scenario
RE5, memory-optimized instance type family with enhanced performance	<ul style="list-style-type: none"> • Optimization for high-performance databases, high memory databases, and other memory-intensive enterprise applications. • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors with stable computing performance. • 1:16.5 processor to memory ratio, up to 2,970 GiB of memory is supported. 	<ul style="list-style-type: none"> • High-performance databases and large memory databases. • Memory-intensive applications. • Big data processing engines such as Apache Spark or Presto.
SN1, a compute-optimized instance type family	<ul style="list-style-type: none"> • 1:2 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors with stable computing performance. • Higher instance offering matching higher network performance. 	<ul style="list-style-type: none"> • Web front-end servers. • Front ends of MMO games. • Data analysis, batch computation, and video encoding. • High-performance scientific and engineering applications.
SN2, general-purpose type family	<ul style="list-style-type: none"> • 1:4 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors with stable computing performance. • Higher instance offering matching higher network performance. 	<ul style="list-style-type: none"> • Enterprise applications of various types and scales. • Small and medium database systems, caches, and search clusters. • Data analysis and computation. • Computing clusters and data processing that depend on memory.

Instance type family	Description	Scenario
F1, compute-optimized instance type family with FPGA	<ul style="list-style-type: none"> • Intel ARRIA 10 GX 1150 compute card. • 1:7.5 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Deep learning and reasoning. • Genomics research. • Financial analysis. • Image transcoding. • Computational workloads, such as real-time video processing and security.

Instance type family	Description	Scenario
F3, compute-optimized instance type family with FPGA	<ul style="list-style-type: none"> • The self-developed VU9P compute card based on Xilinx Virtex UltraScale +. • 1:4 processor to memory ratio. • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Deep learning and reasoning. • Genetic computation. • Video encoding and decoding. • Chip prototype verification. • Accelerate database operations.

Instance type family	Description	Scenario
GN5, compute-optimized instance type family with GPU	<p>GN5 instances use GPU processing and have the following features:</p> <ul style="list-style-type: none"> • NVIDIA P100 GPU compute card. • A variety of processor to memory ratios. • High-performance local NVMe SSDs. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Deep learning. • Scientific computations, such as computational fluid dynamics, computational finance, genomics research, and environmental analysis. • Server-end GPU computational workloads, such as high-performance computation, rendering, and multi-media coding and decoding.
GN4, compute-optimized instance type family with GPU	<p>GN4 instances use GPU processing and have the following features:</p> <ul style="list-style-type: none"> • NVIDIA M40 GPU compute card. • A variety of processor to memory ratios. • 2.5 GHz Intel Xeon E5-2680 v4 (Haswell) processors. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Deep learning. • Scientific computations, such as computational fluid dynamics, computational finance, genomics research, and environmental analysis. • Server-end GPU computational workloads, such as high-performance computing, rendering, and multi-media encoding and decoding.

Instance type family	Description	Scenario
GA1, visualized compute type family with GPU	<p>GA1 instances use GPU processing and have the following features:</p> <ul style="list-style-type: none"> • AMD S7150 GPU compute card. • 1:2.5 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors. • High-performance NVMe SSDs. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Rendering and multimedia encoding and decoding. • Machine learning, high-performance computing, and high-performance databases. • Other server-end business that requires powerful concurrent floating-point compute capabilities.
GN5I, compute-optimized type family with GPU	<p>GN5I instances are provided with a GPU device and have the following features:</p> <ul style="list-style-type: none"> • NVIDIA P4 GPU compute card. • 1:4 processor to memory ratio. • 2.5 GHz Intel Xeon, E5-2682 v4 (Broadwell), or E5-2680 v4 (Haswell) processors. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Deep learning and reasoning. • Server-end GPU computational workloads such as multimedia encoding and decoding.
GN5E, compute-optimized type family with GPU	<ul style="list-style-type: none"> • I/O-optimized instances. • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors with stable computing performance. • Nvidia P4 GPU compute card. • Higher specifications offering higher network performance. 	<ul style="list-style-type: none"> • Deep learning and reasoning. • Video and image processing, such as noise reduction, encoding, and decoding.

The following instance types are only applicable to environments that have been upgraded from Apsara Stack V2 to V3.

Instance type family	Description	Scenario
----------------------	-------------	----------

Instance type family	Description	Scenario
N1, entry-level instances for shared computing	<ul style="list-style-type: none"> • 1:2 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors. • Higher instance offering matching higher network performance. • I/O-optimized instances. • Support for SSD cloud disks and ultra cloud disks. 	<ul style="list-style-type: none"> • Small and medium Web servers. • Batch processing. • Distributed analysis. • Advertisement services.
N2, shared general-purpose entry-level instances	<ul style="list-style-type: none"> • 1:4 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors. • Higher instance offering matching higher network performance. • I/O-optimized instances. • Support for SSD cloud disks and ultra cloud disks. 	<ul style="list-style-type: none"> • Medium Web servers. • Batch processing. • Distributed analysis. • Advertisement services. • Hadoop clusters.
E3, shared memory-optimized entry-level instances	<ul style="list-style-type: none"> • 1:8 processor to memory ratio. • 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors. • Higher instance offering matching higher network performance. • I/O-optimized instances. • Support for SSD cloud disks and ultra cloud disks. 	<ul style="list-style-type: none"> • Cache/Redis. • Search applications. • Memory databases. • Databases with high I/O requirements, such as Oracle and MongoDB. • Hadoop clusters. • Computing scenarios that involve large-volume data processing.
Generation I instance type C1	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Choose from I/O-optimized and non-I/O-optimized instances. • I/O-optimized instances support both SSD and ultra cloud disks. • Non-I/O-optimized instances only support basic cloud disks. 	<p>All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.</p>

Instance type family	Description	Scenario
<p>Generation I instance type C2</p>	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Choose from I/O-optimized and non-I/O-optimized instances. • I/O-optimized instances support both SSD and ultra cloud disks. • Non-I/O-optimized instances only support basic cloud disks. 	<p>All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.</p>
<p>Generation I instance type M1</p>	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Choose from I/O-optimized and non-I/O-optimized instances. • I/O-optimized instances support both SSD and ultra cloud disks. • Non-I/O-optimized instances only support basic cloud disks. 	<p>All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.</p>
<p>Generation I instance type M2</p>	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Choose from I/O-optimized and non-I/O-optimized instances. • I/O-optimized instances support both SSD and ultra cloud disks. • Non-I/O-optimized instances only support basic cloud disks. 	<p>All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.</p>

Instance type family	Description	Scenario
Generation I instance type S1	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Non-I/O-optimized instances. • Support for basic cloud disks only. 	All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.
Generation I instance type S2	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Choose from I/O-optimized and non-I/O-optimized instances. • I/O-optimized instances support both SSD and ultra cloud disks. • Non-I/O-optimized instances only support basic cloud disks. 	All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.
Generation I instance type S3	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Choose from I/O-optimized and non I/O-optimized instances. • I/O-optimized instances support both SSD and ultra cloud disks. • Non-I/O-optimized instances only support basic cloud disks. 	All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.
Generation I instance type T1	<ul style="list-style-type: none"> • Intel Xeon E5-2420 processors with the minimum operating frequency of 1.9 GHz. • Latest DDR3 memory. • Non-I/O-optimized instances. • Support for basic cloud disks only. 	All of these instance types are legacy types of shared instances. They are still categorized by the number of cores they are assigned, ranging from 1 to 16 cores, and are not sensitive to type family.

2.4.1.3. Instance types

An instance is the smallest unit in ECS that is capable of providing computing services. The computing capabilities vary with instance type.

The instance type defines the CPU and memory configuration of an instance and determines properties such as CPU model and processing frequency. However, you must combine the instance type with disk category, image, and network type to uniquely identify an instance. The following table describes all instance type families.

Instance type family	Type code	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues	ENIs (including one primary ENI)
n4	ecs.n4.small	1	2.0	None	0.5	5	1	1
	ecs.n4.large	2	4.0	None	0.5	10	1	1
	ecs.n4.xlarge	4	8.0	None	0.8	15	1	2
	ecs.n4.2xlarge	8	16.0	None	1.2	30	1	2
	ecs.n4.4xlarge	16	32.0	None	2.5	40	1	2
	ecs.n4.8xlarge	32	64.0	None	5.0	50	1	2
mn4	ecs.mn4.small	1	4.0	None	0.5	5	1	1
	ecs.mn4.large	2	8.0	None	0.5	10	1	1
	ecs.mn4.xlarge	4	16.0	None	0.8	15	1	2
	ecs.mn4.2xlarge	8	32.0	None	1.2	30	1	2
	ecs.mn4.4xlarge	16	64.0	None	2.5	40	1	2
	ecs.mn4.8xlarge	32	128.0	None	5.0	50	2	8
xn4	ecs.xn4.small	1	1.0	None	0.5	5	1	1
	ecs.e4.small	1	8.0	None	0.5	5	1	1

Instance type family	Type code	CP U (co re)	Memory (GiB)	Local stora ge (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queu es	ENIs (includin g one primary ENI)
e4	ecs.ce4.xlarge	2	16.0	None	0.5	10	1	1
	ecs.ce4.xlarge	4	32.0	None	0.8	15	1	2
	ecs.e4.2xlarge	8	64.0	None	1.2	30	1	3
	ecs.e4.4xlarge	16	128.0	None	2.5	40	1	8
sn1ne	ecs.sn1ne.large	2	4.0	None	1.0	30	2	2
	ecs.sn1ne.xlarge	4	8.0	None	1.5	50	2	3
	ecs.sn1ne.2xlarge	8	16.0	None	2.0	100	4	4
	ecs.sn1ne.3xlarge	12	24.0	None	2.5	130	4	6
	ecs.sn1ne.4xlarge	16	32.0	None	3.0	160	4	8
	ecs.sn1ne.6xlarge	24	48.0	None	4.5	200	6	8
	ecs.sn1ne.8xlarge	32	64.0	None	6.0	250	8	8
	ecs.sn2ne.large	2	8.0	None	1.0	30	2	2
	ecs.sn2ne.xlarge	4	16.0	None	1.5	50	2	3

Instance type family	Type code	CP U (co re)	Memory (GiB)	Local stora ge (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queu es	ENIs (includin g one primary ENI)
sn2ne	ecs.sn2ne.2xlarge	8	32.0	None	2.0	100	4	4
	ecs.sn2ne.3xlarge	12	48.0	None	2.5	130	4	6
	ecs.sn2ne.4xlarge	16	64.0	None	3.0	160	4	8
	ecs.sn2ne.6xlarge	24	96.0	None	4.5	200	6	8
	ecs.sn2ne.8xlarge	32	128.0	None	6.0	250	8	8
	ecs.sn2ne.14xlarge	56	224.0	None	10.0	450	14	8
se1ne	ecs.se1ne.large	2	16.0	None	1.0	30	2	2
	ecs.se1ne.xlarge	4	32.0	None	1.5	50	2	3
	ecs.se1ne.2xlarge	8	64.0	None	2.0	100	4	4
	ecs.se1ne.3xlarge	12	96.0	None	2.5	130	4	6
	ecs.se1ne.4xlarge	16	128.0	None	3.0	160	4	8
	ecs.se1ne.6xlarge	24	192.0	None	4.5	200	6	8

Instance type family	Type code	CP U (co re)	Memory (GiB)	Local stora ge (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queu es	ENIs (includin g one primary ENI)
	ecs.se1 ne.8xlarge	32	256.0	None	6.0	250	8	8
	ecs.se1 ne.14xlarge	56	480.0	None	10.0	450	14	8
se1	ecs.se1. large	2	16.0	None	0.5	10	1	2
	ecs.se1. xlarge	4	32.0	None	0.8	20	1	3
	ecs.se1. 2xlarge	8	64.0	None	1.5	40	1	4
	ecs.se1. 4xlarge	16	128.0	None	3.0	50	2	8
	ecs.se1. 8xlarge	32	256.0	None	6.0	80	3	8
	ecs.se1. 14xlarge	56	480.0	None	10.0	120	4	8
ebmg5	ecs.eb mg5.24xlarge	96	384.0	None	10.0	400	8	32
i2	ecs.i2.xlarge	4	32.0	1 × 894	1.0	50	2	3
	ecs.i2.2 xlarge	8	64.0	1 × 1,788	2.0	100	2	4
	ecs.i2.4 xlarge	16	128.0	2 × 1,788	3.0	150	4	8
	ecs.i2.8 xlarge	32	256.0	4 × 1,788	6.0	200	8	8
	ecs.i2.1 6xlarge	64	512.0	8 × 1788	10.0	400	16	8
	ecs.d1.2 xlarge	8	32.0	4 × 5,500	3.0	30	1	4

Instance type family	Type code	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues	ENIs (including one primary ENI)
d1	ecs.d1.3xlarge	12	48.0	6 × 5,500	4.0	40	1	6
	ecs.d1.4xlarge	16	64.0	8 × 5,500	6.0	60	2	8
	ecs.d1.6xlarge	24	96.0	12 × 5,500	8.0	80	2	8
	ecs.d1-c8d3.8xlarge	32	128.0	12 × 5,500	10.0	100	4	8
	ecs.d1.8xlarge	32	128.0	16 × 5,500	10.0	100	4	8
	ecs.d1-c14d3.14xlarge	56	160.0	12 × 5,500	17.0	180	6	8
	ecs.d1.14xlarge	56	224.0	28 × 5,500	17.0	180	6	8
sccg5ib	ecs.sccg5ib.24xlarge	96	384.0	None	10.0	450	8	32
scch5ib	ecs.sccch5ib.16xlarge	64	192.0	None	10.0	450	8	32
re5	ecs.re5.15xlarge	60	990.0	None	10.0	100	16	8
	ecs.re5.30xlarge	120	1980.0	None	15.0	200	16	15
	ecs.re5.45xlarge	180	2,970.0	None	30.0	450	16	15
	ecs.sn1.medium	2	4.0	None	0.5	10	1	2
	ecs.sn1.large	4	8.0	None	0.8	20	1	3

Instance type family	Type code	CP U (co re)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues	ENIs (including one primary ENI)
	ecs.sn1.xlarge	8	16.0	None	1.5	40	1	4
	ecs.sn1.3xlarge	16	32.0	None	3.0	50	2	8
	ecs.sn1.7xlarge	32	64.0	None	6.0	80	3	8
sn2	ecs.sn2.medium	2	8.0	None	0.5	10	1	2
	ecs.sn2.large	4	16.0	None	0.8	20	1	3
	ecs.sn2.xlarge	8	32.0	None	1.5	40	1	4
	ecs.sn2.3xlarge	16	64.0	None	3.0	50	2	8
	ecs.sn2.7xlarge	32	128.0	None	6.0	80	3	8
	ecs.sn2.14xlarge	56	224.0	None	10.0	120	4	8

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues	ENIs (including one primary ENI)	FP GA
	ecs.f1-c8f1.2xlarge	8	60.0	None	3.0	40	4	4	Intel ARRIA 10 GX 1150

Instance type family	Instance type	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues	ENIs (including one primary ENI)	FP GA
f1	ecs.f1-c8f1.4xlarge	16	120.0	None	5.0	100	4	8	2 × Intel ARRIA 10 GX 1150
	ecs.f1-c28f1.7xlarge	28	112.0	None	5.0	200	8	8	Intel ARRIA 10 GX 1150
	ecs.f1-c28f1.14xlarge	56	224.0	None	10.0	200	14	8	2 × Intel ARRIA 10 GX 1150
f3	ecs.f3-c16f1.4xlarge	16	64.0	None	5.0	100	4	8	1 × Xilinx VU 9P
	ecs.f3-c16f1.8xlarge	32	128.0	None	10.0	200	8	8	2 × Xilinx VU 9P
	ecs.f3-c16f1.16xlarge	64	256.0	None	20.0	200	16	8	4 × Xilinx VU 9P

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn5-c4g1.xlarge	4	30.0	440	3.0	30	1
	ecs.gn5-c8g1.2xlarge	8	60.0	440	3.0	40	1

Instance type family	Instance type	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn5-c4g1.2xlarge	8	60.0	880	5.0	100	2
	ecs.gn5-c8g1.4xlarge	16	120.0	880	5.0	100	4

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
gn5	ecs.gn5-c28g1.7xlarge	28	112.0	440	5.0	100	8
	ecs.gn5-c8g1.8xlarge	32	240.0	1,760	10.0	200	8

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn5-c28g1.14xlarge	56	224.0	880	10.0	200	14
	ecs.gn5-c8g1.14xlarge	54.	480.0	3,520	25.0	400	14

Instance type family	Instance type	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn4-c4g1.xlarge	4	30.0	None	3.0	30	1
	ecs.gn4-c8g1.2xlarge	8	30.0	None	3.0	40	1

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn4.8x large	32	48.0	None	6.0	80	3
	ecs.gn4-c4g1.2xlarge	8	60.0	None	5.0	50	1

gn4							
Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn4-c8g1.4xlarge	16	60.0	None	5.0	50	1
	ecs.gn4.14xlarge	56	96.0	None	10.0	120	4

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.ga1.xl arge	4	10.0	1 × 87	1.0	20	1
	ecs.ga1.2x large	8	20.0	1 × 175	1.5	30	1

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
ga1	ecs.ga1.4x large	16	40.0	1 × 350	3.0	50	2
	ecs.ga1.8x large	32	80.0	1 × 700	6.0	80	3

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.ga1.14xlarge	56	160.0	1 × 1,400	10.0	120	4
	ecs.gn5i-c2g1.large	2	8.0	None	1.0	10	2

Instance type family	Instance type	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn5i-c4g1.xlarge	4	16.0	None	1.5	20	2
	ecs.gn5i-c8g1.2xlarge	8	32.0	None	2.0	40	4

gn5i Instance type family	Instance type	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn5i-c16g1.4xlarge	16	64.0	None	3.0	80	4
	ecs.gn5i-c16g1.8xlarge	32	128.0	None	6.0	120	8

Instance type family	Instance type	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn5i-c28g1.14xlarge	56	224.0	None	10.0	200	14
	ecs.gn5e-c11g1.3xlarge	10	58.0	None	2.0	15	1

Instance type family	Instance type	CP U (co re)	Memory (GiB)	Local storag e (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
gn5e	ecs.gn5e-c11g1.5xlarge	22	116.0	None	4.0	30	1
	ecs.gn5e-c11g1.11xlarge	44	232.0	None	6.0	60	2

Instance type family	Instance type	CPU (core)	Memory (GiB)	Local storage (GiB)	Bandwidth (inbound or outbound in Gbit/s)	Network packet processing (inbound or outbound in 10,000 PPS)	NIC queues
	ecs.gn5-c11g1.22xlarge	88	464.0	None	10.0	120	4

The following instance types are only applicable to environments that are upgraded from Apsara Stack V2 to V3.

Instance type family	Instance type	CPU (core)	Memory (GiB)
n1	ecs.n1.tiny	1	1.0
	ecs.n1.small	1	2.0
	ecs.n1.medium	2	4.0
	ecs.n1.large	4	8.0
	ecs.n1.xlarge	8	16.0

Instance type family	Instance type	CPU (core)	Memory (GiB)
	ecs.n1.3xlarge	16	32.0
	ecs.n1.7xlarge	32	64.0
n2	ecs.n2.small	1	4.0
	ecs.n2.medium	2	8.0
	ecs.n2.large	4	16.0
	ecs.n2.xlarge	8	32.0
	ecs.n2.3xlarge	16	64.0
	ecs.n2.7xlarge	32	128.0
e3	ecs.e3.small	1	8.0
	ecs.e3.medium	2	16.0
	ecs.e3.large	4	32.0
	ecs.e3.xlarge	8	64.0
	ecs.e3.3xlarge	16	128.0
c1	ecs.c1.small	8	8.0
	ecs.c1.large	8	16.0
c2	ecs.c2.medium	16	16.0
	ecs.c2.large	16	32.0
	ecs.c2.xlarge	16	64.0
m1	ecs.m1.medium	4	16.0
	ecs.m1.xlarge	8	32.0
m2	ecs.m2.medium	4	32.0
s1	ecs.s1.small	1	2.0
	ecs.s1.medium	1	4.0
	ecs.s1.large	1	8.0
s2	ecs.s2.small	2	2.0
	ecs.s2.large	2	4.0
	ecs.s2.xlarge	2	8.0

Instance type family	Instance type	CPU (core)	Memory (GiB)
	ecs.s2.2xlarge	2	16.0
s3	ecs.s3.medium	4	4.0
	ecs.s3.large	4	8.0
t1	ecs.t1.small	1	1.0

2.4.1.4. UserData

UserData allows you to customize the startup behavior of instances and import data to ECS instances. It is the basis for ECS instance customization.

UserData is implemented through different types of scripts. Before UserData is implemented on an instance, all ECS instances will have the same initial environment and configurations when started for the first time. After enterprises or individuals enter valid UserData information based on their scenarios and needs, required ECS instances are provided after the first startup.

Methods

- **UserData-Scripts:** are applicable to users who need to initialize instances by executing the shell scripts. The UserData-Scripts begin with `#!/bin/sh`. A review of user data shows that most users input UserData by running UserData-Scripts. UserData-Scripts are also suitable for complicated deployment scenarios.
- **Cloud-Config:** is a special script supported by cloud-init. It packs frequently-used personalized configurations into YAML files, which enable you to complete the frequently-used configurations more conveniently. The script starts with `# Cloud-config` in the first line and is followed by an array containing `ssh_authorized_keys`, `hostname`, `write_files`, and `manage_etc_hosts`.

Scenarios

- SSH authentication
- Software source updates and configuration
- DNS configuration
- Application installation and configuration

2.4.1.5. Instance lifecycle

The lifecycle of an ECS instance begins when it is created and ends when it is released. This topic describes the instance status, status attributes, and corresponding API status.

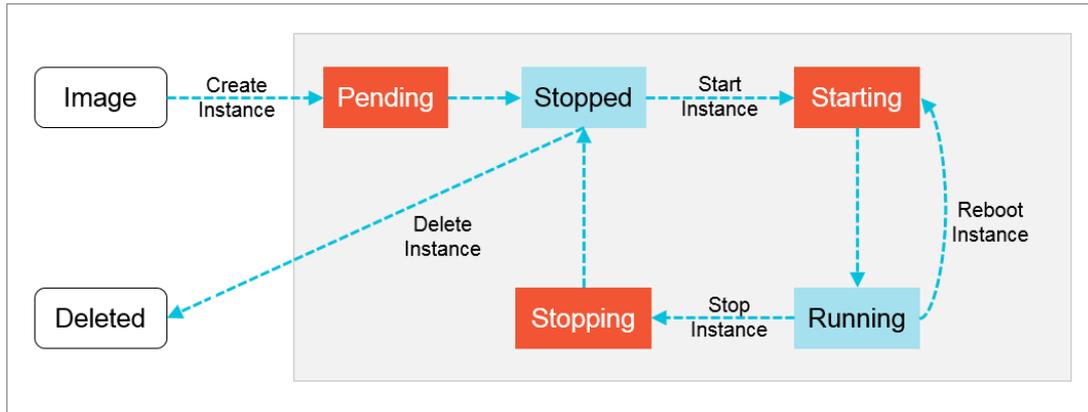
An instance has several inherent states throughout its lifecycle, as shown in [Lifecycle description](#).

Lifecycle description

Status	Status attribute	Description	Corresponding API status
Instance being created	Intermediate	The instance is being created and is waiting to be enabled. If an instance remains in this status for a long period of time, an exception occurs.	Pending
Starting	Intermediate	After an instance is restarted or started from the console or through APIs, the instance enters the starting state before entering the running state. If an instance remains in the starting state for a long period of time, an exception occurs.	Starting
Running	Stable	Indicates that the instance is running normally and can accommodate your business needs.	Running
Stopping	Intermediate	After an instance is stopped from the console or through APIs, the instance enters the stopping state before entering the stopped state. If an instance remains in the stopping state for a long period of time, an exception occurs.	Stopping
Stopped	Stable	Indicates that an instance has been stopped. An instance in the stopped state cannot provide external services.	Stopped
Reinitializing	Intermediate	After the system disk or data disk is reinitialized from the console or through APIs the instance enters the reinitializing state before entering the running state. If an instance remains in the reinitializing state for a long period of time, an exception occurs.	Stopped
Changing system disk	Intermediate	After the system disk is changed from the console or through APIs, the instance enters the changing system disk state before entering the running state. If an instance remains in the changing system disk state for a long time, an exception occurs.	Stopped

Lifecycle description describes corresponding relationship between instance states in the console and instance states in APIs. **Instance status in APIs** shows the instance states in APIs.

Instance status in APIs



2.4.1.6. EBM Instances

ECS Bare Metal (EBM) Instance is a new computing service that combines the elasticity of virtual machines with the performance and features of physical machines. EBM Instances are designed based on the state-of-the-art virtualization technology developed by Alibaba Cloud.

The virtualization used by EBM Instances is optimized to support common ECS instances and nested virtualization, maintaining elastic performance with the user experience of physical servers.

Benefits

EBM Instances provides the following benefits through technological innovation:

- **Exclusive computing resources**

As a cloud-based elastic computing service, EBM Instances surpass the performance and isolation of physical servers, enabling you to exclusively occupy computing resources without virtualization performance overhead or feature loss. EBM Instances support ultrahigh-frequency instances and can contain 8, 16, 32, or 96 CPU cores. An EBM Instance with eight CPU cores supports ultrahigh frequency processing from 3.7 to 4.1 GHz, providing better performance and response for gaming and finance businesses than peer services.

- **Encrypted computing**

For security, EBM Instances use a chip-level trusted execution environment (Intel® SGX) in addition to physical server isolation to ensure that encrypted data can only be computed within a secure and trusted environment. This chip-level hardware security protection provides a safe box for the data of cloud users and allows users to control all data encryption and key protection processes.

- **Any Stack on Alibaba Cloud**

An EBM Instance combines the performance strengths and complete features of physical machines and the ease-of-use and cost-effectiveness of cloud servers. It can effectively meet the demands of high-performance computing and help you build new hybrid clouds. Thanks to the flexibility, elasticity, and all the other strengths inherited from both virtual and physical machines, EBM Instances are endowed with re-virtualization ability. Offline private clouds can be seamlessly migrated to Alibaba Cloud without the performance overhead that may arise from nested virtualization, giving you a new approach for moving businesses onto the cloud.

- **Heterogeneous instruction set processor support**

The virtualization 2.0 technology used by EBM Instances is developed independently by Alibaba Cloud and supports ARM and other instruction set processors at no additional cost.

Configuration features

The following table lists the configuration features of EBM Instances.

Features

Item	Description
CPU configuration	Only supports the ebmg5 general-purpose EBM Instance type family.
Memory configuration	Supports expansion from 32 GiB to 384 GiB as needed. The ratio of CPU to memory is 1:2 or 1:4 to provide better computing performance.
Storage configuration	Supports startup from virtual machine images or cloud disks, achieving delivery in several seconds.
Network configuration	Supports Virtual Private Clouds (VPCs), and interoperability with ECS, GPU, and other cloud services. Delivers performance and stability comparable to physical machine networks.
Image configuration	Supports ECS images.
Security configuration	Maintains the same security policies and flexibility as existing ECS instances.

2.4.2. Block storage

2.4.2.1. Overview

This topic describes the diverse types of block storage. This includes elastic block storage services based on a distributed storage architecture and local storage services based on the local hard disks of physical machines.

Description of elastic block storage and local storage:

- **Elastic block storage** provides ECS with persistent and highly reliable random block-level data storage with low latency. Data is stored in triplicate over a distributed system to ensure data reliability. Elastic block storage can be created, released, and expanded at any time.
- **Local storage**, also known as local disks, refers to temporary disks mounted on the physical machine where an ECS instance resides. Local storage is designed for business scenarios that require high storage I/O performance. Local storage provides block-level data access for instances with low latency, high random IOPS, and high throughput.

Block storage, OSS, and NAS

Currently, Apsara Stack provides three types of data storage services: block storage, Network Attached Storage (NAS), and Object Storage Service (OSS).

Storage service comparison

Type	Description	Scenario
Block storage	A high-performance and low-latency block-level storage device provided by Apsara Stack for ECS instances. It supports random reads and writes. You can format the block storage and create a file system on it as you would with a hard disk.	Block storage can be used for data storage in most common business scenarios.
OSS	A storage space suited for high volume unstructured data such as images, video, audio, and other data generated online. Data stored in OSS can be accessed anytime and anywhere through APIs.	OSS is suited for business scenarios such as website construction, separation between dynamic and static resources, and CDN acceleration.
NAS	NAS is a storage system similar to OSS, and as such is also suited for storing high volume unstructured data. Data must be accessed through standard file access protocols, such as the Network File System (NFS) protocol for Linux systems and the Common Internet File System (CIFS) protocol for Windows systems. You can set permissions to allow different clients to access the same file at the same time.	NAS is suited for business scenarios such as file sharing across departments, non-linear editing of radio and television data, high-performance computing, and Docker.

2.4.2.2. Block storage performance

2.4.2.2.1. Overview

This topic describes the key performance indicators of different block storage.

2.4.2.2.2. Elastic block storage performance

The performance of elastic block storage includes the performance of its component cloud disks and shared block storage.

Cloud disks

 **Note** Cloud disks with standard configurations are compared by using the standard test method. The following table lists the comparison data.

Disk performance

Block storage	SSD cloud disk	Ultra cloud disk	Basic cloud disk
Maximum capacity	32,768 GiB	32,768 GiB	2,000 GiB
Maximum IOPS	20,000	3,000	Several hundreds
Maximum throughput	300 Mbit/s	80 Mbit/s	20-40 Mbit/s
Performance formula	$\text{IOPS} = \min \{30 \times (\text{capacity}, 20,000)\}$ $\text{Throughput} = \min \{50 + 0.5 \times (\text{capacity}, 300)\} \text{ Mbit/s}$	$\text{IOPS} = \min \{1000 + 6 \times (\text{capacity}, 3,000)\}$ $\text{Throughput} = \min \{50 + 0.1 \times (\text{capacity}, 80)\} \text{ Mbit/s}$	N/A
API name	cloud_ssd	cloud_efficiency	cloud
Typical scenario	<ul style="list-style-type: none"> I/O-intensive applications Large and medium relational databases NoSQL databases 	<ul style="list-style-type: none"> Small and medium databases Large development and testing Web server logs 	Infrequently-accessed or low-I/O applications

Shared block storage

Disk performance

Parameter	SSD shared block storage	Ultra shared block storage
Maximum capacity	<ul style="list-style-type: none"> • Single disk: 32,768 GiB • Single instance: 128 TiB 	<ul style="list-style-type: none"> • Single disk: 32,768 GiB • Single instance: 128 TiB
Maximum random IOPS	30,000	5,000
Maximum sequential throughput	512 Mbit/s	160 Mbit/s
Performance formula for a single disk	IOPS = $\min \{40 \times (\text{capacity}, 30,000)\}$	IOPS = $\min \{1,000 + 6 \times (\text{capacity}, 5,000)\}$
	Throughput = $\min \{50 + 0.5 \times (\text{capacity}, 512)\}$ Mbit/s	Throughput = $\min \{50 + 0.15 \times (\text{capacity}, 160)\}$ Mbit/s
Typical scenario	<ul style="list-style-type: none"> • Oracle RAC • SQL Server • Failover cluster • High-availability architecture of servers 	<ul style="list-style-type: none"> • High-availability architecture of servers • High-availability architecture of development and testing databases

2.4.2.2.3. Local disk storage performance

This topic describes the storage performance of local disks. For more information, see [Local storage](#).

2.4.2.2.4. Test disk performance

Depending on which OS an instance is being run on, you can use the following tools to test disk performance:

- For Linux instances, we recommend that you use the DD, fio, or sysbench tools.
- For Windows instances, we recommend that you use the fio or iometer tools.

This topic describes how to use the fio tool to test disk performance, using a Linux OS instance as an example. Before testing the disk, you must make sure the disk is 4 KB aligned.

 **Note** The performance of a block storage disk can be obtained by testing raw, unformatted disks. However, data loss may occur when testing raw disks directly. You must back up the data of raw disks before they are tested.

- Run the following command to test random write IOPS:


```
fio -direct=1 -iodepth=128 -rw=randwrite -ioengine=libaio -bs=4k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Rand_Write_Testing
```
- Run the following command to test random read IOPS:

```

fiio -direct=1 -iodepth=128 -rw=randread -ioengine=libaio -bs=4k -size=1G -numjobs=1 -
runtime=1000 -group_reporting -filename=iotest -name=Rand_Read_Testing

```

- Run the following command to test write throughput:

```

fiio -direct=1 -iodepth=64 -rw=write -ioengine=libaio -bs=64k -size=1G -numjobs=1 -
runtime=1000 -group_reporting -filename=iotest -name=Write_PPS_Testing

```

- Run the following command to test read throughput:

```

fiio -direct=1 -iodepth=64 -rw=read -ioengine=libaio -bs=64k -size=1G -numjobs=1 -
runtime=1000 -group_reporting -filename=iotest -name=Read_PPS_Testing

```

The following table describes the fio parameters used in the preceding tests.

Parameter	Description
-direct=1	Indicates that I/O cache is ignored in the test. Data is written directly.
-rw=randwrite	Read and write policies. Available options: randread (random read), randwrite (random write), read (sequential read), write (sequential write), and randrw (random read and write).
-ioengine=libaio	Use libaio (Linux asynchronous I/O) as the testing tool. Applications usually use one of two methods to implement I/O operations: synchronous and asynchronous. In synchronous I/O mode, a thread sends only one I/O request to the kernel and waits for the I/O operation to complete. In this case, the iodepth of a single thread is always less than 1, but anywhere from 16 to 32 concurrent threads can be used to fill up the iodepth. In asynchronous I/O mode, a job uses the libaio method to send multiple I/O requests to the kernel and waits for the I/O operations to complete. Asynchronous I/O reduces the number of interactions and increases operation efficiency.
-bs=4k	Indicates that a single I/O operation accesses block files of 4 KB. By default, this parameter is set to 4 KB.
-size=1G	Indicates that the tested file is 1 GB in size.
-numjobs=1	Indicates that the number of tested jobs is 1.
-runtime=1000	Indicates that the test duration is 1,000s. If this parameter is not configured, the file of the size specified by -size is written in blocks of the size specified by -bs.
-group_reporting	Indicates the test result display mode. group_reporting summarizes the statistics of each process, but does not show information of different jobs.
-filename=iotest	Indicates the path and name of the output files. After the test is completed, delete unnecessary files to release disk space.
-name=Rand_Write_Testing	Indicates the name of the testing task.

2.4.2.3. Elastic block storage

2.4.2.3.1. Overview

Elastic block storage can be divided into the following types based on whether it can be attached to multiple ECS instances.

- **Cloud disks:** A cloud disk can be attached to a single ECS instance that resides in the same zone and region.
- **Shared block storage:** A shared block storage can be attached to up to four ECS instances that belong to the same zone and region.

2.4.2.3.2. Cloud disks

Cloud disks can be classified in either of the following ways:

- **Performance-based classification**

Cloud disks are divided by performance into basic cloud disks, ultra cloud disks, and SSD cloud disks.

- Basic cloud disks are ideal for the minimally I/O-intensive scenarios, and only provide several hundreds of IOPS for ECS instances.
- Ultra cloud disks are ideal for medium I/O load scenarios and provide a storage performance of up to 3,000 random IOPS for ECS instances.
- SSD cloud disks are ideal for I/O-intensive scenarios and provide stable and high random IOPS performance.

- **Function-based classification**

Cloud disks can be divided by their functions into system disks and data disks.

- **System disks:** has a lifecycle that is tied to the ECS instance to which it is attached. A system disk is created and released along with the instance. Shared access is not allowed.
- **Data disks:** can be created separately or together with ECS instances. Shared access is not allowed. A data disk created with an ECS instance has a lifecycle that is tied to that of the instance, and is created and released along with the instance. Data disks that are created independently can be released independently or in conjunction with the ECS instance to which it is attached. The capacity of a data disk is determined by its type.

2.4.2.3.3. Shared block storage

Shared block storage is a block-level data storage service that supports concurrent read/write operations to multiple ECS instances, giving it a high level of performance and reliability.

A single shared block storage device can be attached to a maximum of four ECS instances. Shared block storage can only be used as data disks and must be created individually. Shared access is allowed. You can configure a shared block storage device to be released when its associated ECS instance is released.

Shared block storage can be divided into the following types based on performance:

- **SSD shared block storage:** adopts an SSD storage medium to provide stable and high-performance storage with enhanced random I/O and data reliability.
- **Ultra shared block storage:** adopts a hybrid SSD and HDD storage medium.

When used as data disks, shared block storage allows up to 16 data disks to be attached to each ECS instance.

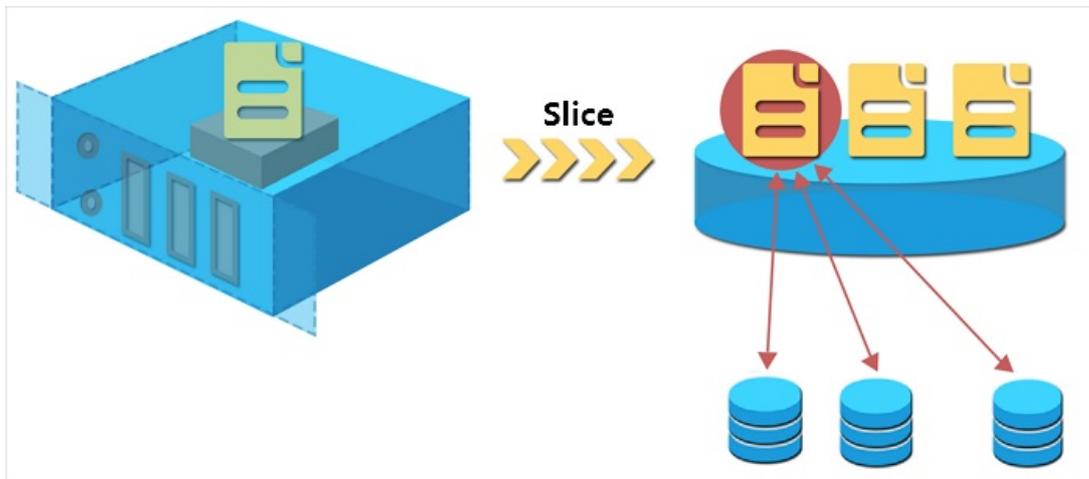
2.4.2.3.4. Triplicate storage

Apsara Distributed File System provides stable, efficient, and reliable data access to ECS instances.

Chunks

When ECS users perform read and write operations on virtual disks, the operations are translated into the corresponding processes on the files stored in Apsara Stack data storage system. Apsara Stack uses a flat design in which a linear address space is divided into slices called chunks. Each chunk is replicated into three copies. Each copy is stored on a different node in the cluster, which ensures data reliability.

Triplicate backup



How triplicate technology works

Triplicate storage is made up of three components: master, chunk server, and client. Each write operation performed by an ECS user is converted into an operation executed by the client. The execution process is as follows:

1. The client determines the location of a chunk corresponding to the write operation.
2. The client sends a request to the master to query the chunk servers where the three chunk replicas are each stored.
3. The client sends write requests to the chunk servers based on the results returned from the master.
4. If the three replicas of the chunk are all successfully written as requested, the client returns a message to indicate the success of the operation. If the write operation fails, a failure message is returned.

The master component distributes chunks based on the disk usage, rack distribution, power supply, and machine workloads of chunk servers. This ensures that chunk replicas are each distributed to chunk servers on different racks and that data does not become unavailable due to the failure of a single server or rack.

Data protection mechanism

When a data node is damaged or disk faults occur on a data node, the total number of valid replicas of some chunks in a cluster becomes less than three. In these cases, the master replicates data between chunk servers to ensure that there are always three valid replicas of chunks in the cluster.

Automatic replication



All user-level operations for data on cloud disks are synchronized across the three chunk replicas at the underlying layer. Operations that are synchronized include adding, modifying, and deleting data. This mode ensures the reliability and consistency of user data.

To prevent data losses caused by viruses, accidental deletion, or malicious attacks, we recommend that you use other protection methods such as backing up data and taking snapshots in addition to triplicate storage. Implement all appropriate measures to ensure the security and availability of your data.

2.4.2.4. ECS disk encryption

ECS disk encryption is a simple and secure encryption method that can be used to encrypt new cloud disks.

With ECS disk encryption, there is no need to create or maintain your own key management infrastructure, change existing applications and maintenance procedures, or add additional encryption operations. Disk encryption does not have any negative impact on your business processes. After an encrypted ECS disk is created and attached to an ECS instance, the following types of data can be encrypted:

- Data on the cloud disk.
- Data transmitted between the cloud disk and instance. Data in the instance operating system is not encrypted again.
- All snapshots created from the encrypted cloud disk. These snapshots are called encrypted snapshots.

The data transmitted from the ECS instance to the cloud disk is encrypted on the host where the ECS instance resides.

Disk encryption is supported on all available cloud disks (basic cloud disks, ultra cloud disks, and SSD cloud disks) and shared block storage (ultra and SSD) for all Apsara Stack instances.

2.4.2.5. Local storage

Local storage refers to a local disk attached to a physical host where the ECS resides. As a form of temporary block storage, local disks are designed for scenarios that require extremely high I/O performance.

Local storage enables block-level data access to provide instances with low latency, high random IOPS, and high throughput storage. Because a local disk is attached to a single physical server, the reliability of data depends on the physical server, which may lead to single points of failure within your architecture. We recommend that you implement data redundancy at the application layer to guarantee the data remains available.

Note Storing data on local disks carries risks to data persistence, such as when the host server is down. We recommend that you never use local disks for data storage scenarios that require long-term persistence. If no data reliability architecture is available for your application, we recommend that you use cloud disks or shared block storage for your ECS instances.

Local disk types

Currently, Apsara Stack provides two types of local disks:

- **NVMe SSDs:** are used together with GN5 and GA1 instances.
- **SATA HDDs:** are used together with D1NE and D1 instances. These type of storage disks are suited for customers from Internet, finance, and other industries that require large storage capacity with storage analysis and offline computing. SATA HDDs satisfy the performance, capacity, and bandwidth requirements of distributed computing models such as Hadoop.

Performance of local SATA HDDs

The following table lists the performance parameters of a local SATA HDD attached to a D1NE or D1 instance.

Disk Performance

Performance parameter	Local SATA HDD
Capacity	<ul style="list-style-type: none"> • Maximum capacity of a single disk: 5,500 GiB. • Total capacity per instance: 154,000 GiB.
Throughput	<ul style="list-style-type: none"> • Single disk throughput: 190 Mbit/s. • Total throughput per instance: 5,320 Mbit/s.
Latency	In milliseconds.

2.4.3. Images

An image is a template for running environments in one or more ECS instances, and generally includes an operating system and preinstalled software.

An image works as a copy of all of the data stored on one or more disks. These disks can be a single system disk, or a combination of system disk and data disks. You can use an image to create an ECS instance or change the system disk of an ECS instance.

Image types

ECS provides a variety of image types to easily access image resources.

Image description

Type	Description
Public image	<p>Public images officially provided by Apsara Stack support Windows and most popular versions of Linux operating systems, including:</p> <ul style="list-style-type: none"> • Windows • CentOS • CoreOS • Debian • Gentoo • FreeBSD • OpenSUSE • SUSE Linux • Ubuntu
Custom image	<p>Custom images created based on your existing physical servers, virtual machines, or cloud hosts. This image type is flexible enough to meet all of your specific business needs.</p>

Obtain an image

ECS allows you to obtain images through the following methods:

- Create a custom image based on an existing ECS instance.
- Choose an image shared by another Apsara Stack tenant account.
- Import an offline image file to an ECS cluster to generate a custom image.
- Copy a custom image to another region to achieve consistent environment or application deployment across regions.

Image formats

Currently, ECS supports VHD and RAW images. Images in other formats must be converted to VHD or RAW images before they can be run in ECS. For more information about format conversion, see *Convert image format in ECS User Guide*.

2.4.4. Snapshots

2.4.4.1. Overview

A snapshot is a copy of data on a disk created at a specific point in time. When using a disk, you may encounter the following scenarios:

- When writing or storing data to a disk, you may want to use snapshot data from another disk as the basis for the target disk.
- While cloud disks represent a secure way to store data, their data may be subject to errors caused by application errors or malicious read and write operations, and require additional

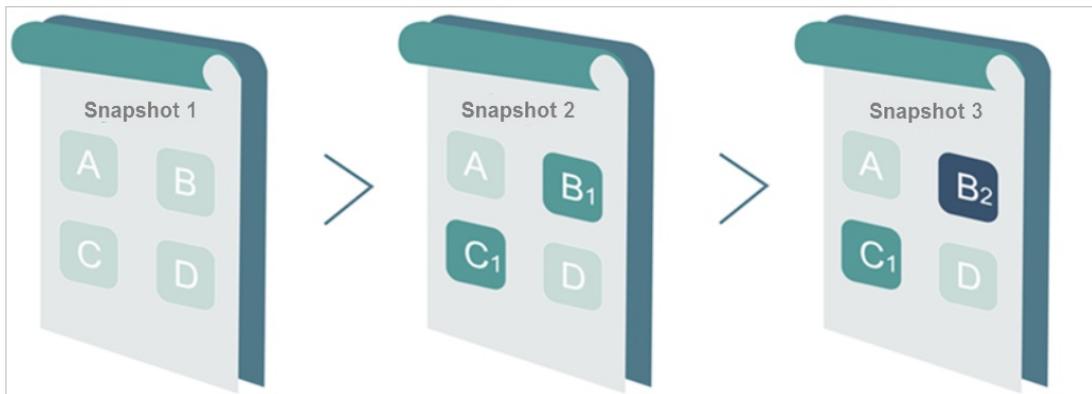
safeguard mechanisms. For this reason, you may want to use snapshots to restore data to a previous point in time in case of data errors.

2.4.4.2. Mechanisms

This topic describes snapshots. Snapshots retain a copy of data stored on a disk at a certain point in time. You can schedule disk snapshots to be created periodically to ensure continuous operation of your business.

Snapshots are created incrementally such that only data changes between two snapshots are copied instead of all of the data, as shown in [Snapshots](#).

Snapshots



Snapshot 1, Snapshot 2, and Snapshot 3 are the first, second, and third snapshots of a disk. When a snapshot is created, the file system checks each block of data stored on the disk, and only copies the blocks of data that differ from those on the previous snapshots. The changes between snapshots in the preceding figure are described as follows:

- All data on the disk is copied to Snapshot 1 because it is the first disk snapshot.
- The changed blocks B1 and C1 are copied to Snapshot 2. Blocks A and D are referenced from Snapshot 1.
- The changed block B2 is copied to Snapshot 3. Blocks A and D are referenced from Snapshot 1, and block C1 is referenced from Snapshot 2.
- When the disk needs to be restored to the status of Snapshot 3, snapshot rollback will copy blocks A, B2, C1, and D to the disk, which will be restored to the status at the time of Snapshot 3.
- If Snapshot 2 is deleted, block B1 in the snapshot is deleted, but block C1 is retained because it is referenced by other snapshots. When you roll back a disk to Snapshot 3, block C1 is recovered.

Note Snapshots are stored on the Object Storage Service (OSS), but are hidden from users. Snapshots do not consume bucket space in OSS. Snapshot operations can only be performed from the ECS console or through APIs.

2.4.4.3. Specifications of ECS Snapshot 2.0

Built on the features of the original snapshot function, ECS Snapshot 2.0 data backup service provides a higher snapshot quota and a more flexible automatic task policy, further reducing its impact on business I/O.

Comparison of snapshot specifications

Item	Traditional snapshot specifications	Snapshot 2.0 specifications	User benefits	Example
Snapshot quota	The snapshot quota is limited to: the number of disks $\times 6 + 6$.	Each disk can have up to 64 snapshots.	Longer protection cycle, and smaller protection granularity.	<ul style="list-style-type: none"> A snapshot is created for the data disks of non-core business at 00:00 every day. Snapshots for the last two months are retained. A snapshot is created for the data disks of core business every four hours. Snapshots for the last ten days are retained.
Automatic task policy	By default, the task is scheduled to be triggered once a day and cannot be modified manually.	You can customize the time of day and days of the week that snapshots are scheduled to be created and the retention period of snapshots. The disk quantity and related details associated with an automatic snapshot policy can be queried.	More flexible protection policy	<ul style="list-style-type: none"> You can schedule snapshots to be created on the hour several times in a single day. You can choose which days of the week to recur snapshot creation. You can specify the snapshot retention period or choose to retain it permanently. When the number of automatic snapshots reaches the upper limit, the oldest automatic snapshot will be automatically deleted.
Implementation	Copy-on-write (COW)	Redirect-on-write (ROW)	Mitigates the impact of snapshot tasks on I/O write performance.	Snapshots can be taken at any time without interruptions to your business.

2.4.4.4. Technical comparison

Alibaba Cloud ECS Snapshot 2.0 has many advantages over the snapshot feature of traditional storage products.

Comparison of technical advantages

Item	ECS Snapshot 2.0	Traditional snapshot
Capacity	Unlimited capacity, meeting the data protection needs of extra-large businesses.	Capacity limited by the initial storage device capacity, merely meeting the data protection needs for a few core services.

2.4.5. Deployment sets

A deployment set is a tool that allows you to view the physical topology of hosts, racks, and switches and select a deployment policy that best suits the reliability and performance requirements of your business.

There may be increased reliability or performance requirements when you use multiple ECS instances in the same zone.

- **Improve business reliability**

To avoid the impacts caused by the failure of physical hosts, racks, or Switches, multiple copies of application instances must be distributed across different physical hosts, racks, or Switches.

- **Improve network performance**

For scenarios that involve frequent network interactions between instances, lower latency and higher bandwidth can be achieved by aggregating corresponding instances onto a single Switch.

Deployment granularities and policies

- **Deployment granularities**
 - Host: indicates physical-server-level scheduling.
 - Rack: indicates rack-level scheduling.
 - Switch: indicates Switch-level scheduling.
- **Deployment policies**
 - LooseAggregation
 - StrictAggregation
 - LooseDispersion
 - StrictDispersion

LooseAggregation and StrictAggregation are intended for higher performance, while LooseDispersion and StrictDispersion are intended for higher reliability.

Granularities and policies lists the deployment policies and business scenarios corresponding to each deployment granularity.

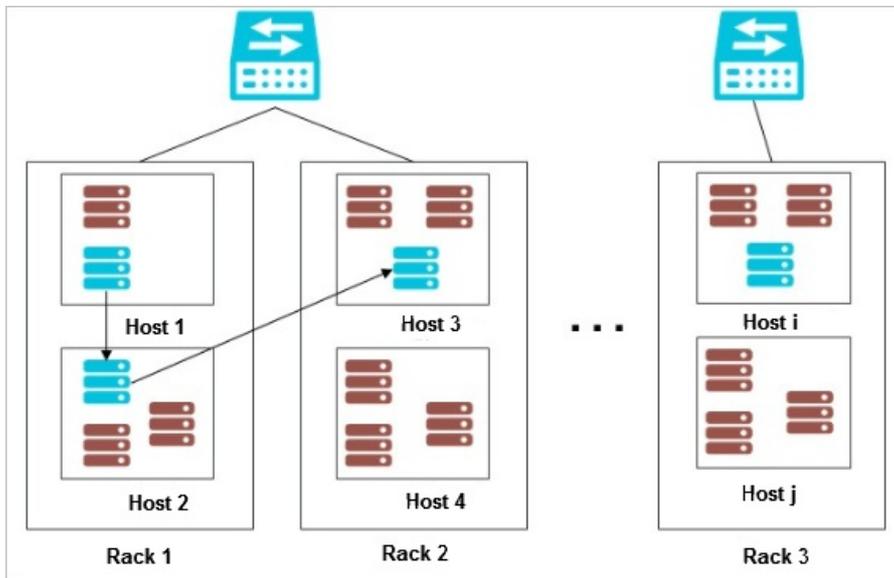
Granularities and policies

Deployment granularity	Deployment policy	Business scenario
Host	StrictDispersion	General purposes
	LooseDispersion	
Rack	StrictDispersion	Big data and databases
	LooseDispersion	Game customers
Switch	StrictDispersion	VPN
	LooseDispersion	Game customers
	StrictAggregation	Big data and databases
	LooseAggregation	Game customers

Typical examples

The following figure shows a typical case where business reliability is improved by using deployment sets. Three ECS instances of a tenant are distributed on three different physical hosts, which are distributed on at least two different racks.

Typical example



Note For more information about the deployment set APIs, see *Deployment sets in ECS Developer Guide*.

2.4.6. Network and security

2.4.6.1. IP addresses of ECS instances of VPC type

This topic describes the IP address types supported by ECS instances and the corresponding scenarios.

IP address types

ECS instances have the following IP address types:

- **Private IP addresses**

When you create an ECS instance, a private IP address is assigned based on the VPC and the CIDR block of the VSwitch to which the instance belongs.

- **Elastic IP (EIP)**

An EIP is a public IP address. You can apply for an EIP as necessary.

Scenarios

- **Private IP:** A private IP address is used to access the intranet. When creating an instance, you can directly configure the private IP address.

 **Note** If the private IP address is not configured, the system automatically allocates a private IP address for the instance.

- **EIP:** An EIP is used to access the Internet. You can separately bind an EIP to an instance after it has been created. For more information, see [EIP](#) in *VPC User Guide*. EIPs can be applied for and retained long-term. You can bind and unbind an EIP to and from an instance, delete the EIP, or modify its bandwidth.

2.4.6.2. Elastic network interfaces

This topic describes Elastic Network Interfaces (ENIs) and their application scenarios.

An ENI is a virtual network card that can be attached to an ECS instance on a VPC network. ENIs help you implement high-availability clusters, low-cost failover, and refined network management. ENIs are supported in all regions.

ENIs can be used in the following scenarios:

- **High-availability clusters**

An ENI can meet the demands for multiple NICs on a single instance in a high-availability architecture.

- **Cost-effective failover**

You can detach an ENI from a failed ECS instance and reattach it to another instance to quickly redirect traffic intended for a failed instance to a backup instance and immediately recover service.

- **Lean network management**

You can configure multiple ENIs for an instance. For example, you can use some ENIs for internal management and others for Internet business access, so as to isolate management data from business data. You can also configure precisely-targeted security group rules for each ENI based on the source IP address, protocols, and ports, so as to achieve traffic control.

ENI types

ENIs are classified into two types:

- **Primary ENI**

The ENI created by default upon the creation of an instance in a VPC is called the **primary ENI**. The lifecycle of the primary ENI is tied to that of the instance, and the primary ENI cannot be removed from the instance.

- **Secondary ENI**

You can create a secondary ENI and attach it to or detach it from the instance. The maximum number of ENIs that can be attached to a single instance varies with the instance type. For more information, see [Instance families](#).

ENI attributes

The following table displays the attributes of an ENI.

Attribute description

Attribute	Quantity
Primary private IP addresses	1
MAC addresses	1
Security groups	1 to 5
Description	1
ENI name	1

Limits

ENIs have the following limits:

- A single account can own up to 100 ENIs in a single region.
- The ECS instance must belong to the same zone and same region as the ENI, but does not have to use the same VSwitch.
- For instance types that support ENI attaching and the number of ENIs supported by each instance type, see [Instance types](#).
- Attaching multiple ENIs does not increase the instance bandwidth.

 **Note** The instance bandwidth varies according to the instance type.

2.4.6.3. Intranet

ECS instances communicate through the intranet. Non-I/O-optimized instances share 1 GiB of bandwidth and I/O-optimized instances share 10 GiB of bandwidth. The intranet is a shared network, so the bandwidth may fluctuate.

 **Note** Currently, most mainstream instances are I/O-optimized instances, and the actual bandwidth is related to the physical hardware.

If you need to transmit data between two ECS instances in the same region, we recommend that you transmit data through an intranet connection. Intranet connections can also be implemented between ApsaraDB for RDS, SLB, and OSS services. Intranets can share up to 1 GiB of bandwidth.

ECS can communicate with RDS, SLB, and OSS in the same region through the intranet.

The following rules apply to VPC-type ECS instances in the intranet:

- Intranet communication is permitted by default for instances in the same security group of the same account in the same region. If instances with the same account in the same region are in different security groups, intranet communication can be implemented by authorizing mutual access between the two security groups.
- For instances that belong to the same account and same region but do not belong to the same

VPC network, you can use Express Connect to implement their intranet communication.

- The intranet IP address of an instance can be modified or changed as needed.
- Virtual IP (VIP) addresses cannot be configured as the intranet or Internet addresses of instances.
- Instances of different network types cannot communicate with each other over an intranet.

2.4.6.4. Security group rules

Security group rules permit or deny Internet or intranet traffic to or from the ECS instances associated with the security group.

You can add or delete security group rules at any time. Changes in security group rules are automatically applied to ECS instances associated with the security group.

Be sure to configure concise security group rules. If you associate an instance with multiple security groups, hundreds of rules may apply to the instance. This may cause connection errors when you access the instance.

2.5. Scenarios

ECS can be used either independently as a simple Web server or with other Apsara Stack services such as OSS, to provide advanced flexible solutions. ECS is typically applied in the following scenarios:

Official websites for enterprises and simple Web applications

Initially, official websites for enterprises do not have high volumes of traffic and only require low-configuration ECS instances to run applications, databases, and store files. As your website develops, you can increase the ECS specifications and the number of ECS instances at any time without the need to worry about low resources during traffic spikes.

Multimedia and high-traffic applications or websites

When ECS is used with OSS, static images, videos, and downloaded packages can be stored on OSS to reduce storage costs. In addition, ECS can be used with Server Load Balancer (SLB) to greatly shorten user response time, reduce bandwidth fees, and improve availability.

Applications or websites with large traffic fluctuations

Some applications and websites may encounter large fluctuations in traffic within a short period. ECS also features elastic processing capabilities. The number of ECS instances automatically increases or decreases in response to spikes and dips in traffic for the purpose of meeting resource requirements and preserving cost efficiency. ECS can be used with SLB to achieve a high availability architecture.

Databases

Databases with high I/O requirements are supported. A high-configuration I/O-optimized ECS instance can be used with an SSD cloud disk to support high I/O concurrency with higher data reliability. Alternatively, multiple lower-configuration I/O-optimized ECS instances can be used with SLB to achieve a high availability architecture.

2.6. Limits

The limits of ECS are as follows:

- ECS instances with 4 GiB or higher memory must use a 64-bit operating system. 32-bit operating systems have a maximum of 4 GiB of memory addressing.
- A 32-bit Windows operating system can use a maximum of 4 cores in its CPU.
- Windows operating systems support a maximum of 64 vCPUs in their instance specifications.
- Virtualization software installation and subsequent virtualization such as VMware are not supported.
- Currently, sound card applications are not supported. Only GPU instances support virtual sound cards. External hardware devices, such as hardware dongles, USB flash drives, external hard disks, and bank U keys, cannot be directly connected to ECS instances.
- ECS does not support multicast protocols. If multicasting services are required, we recommend that you use unicast instead.

The following table lists additional limits to ECS.

Other limits

Type	Description
Instance type	For more information, see Instance families and Instance types .
Block storage	<p>Specification limits</p> <ul style="list-style-type: none"> • Number of system disks per instance: 1. • Number of data disks per instance: 16. • Default quota of instances to which one shared block storage can be attached: 4. • System disk capacity: 40-500 GiB. • Capacity of one basic cloud disk: 5-2,000 GiB. • Capacity of one SSD disk: 20-32,768 GiB. • Capacity of one ultra cloud disk: 20-32,768 GiB. • Total capacity of one ultra block storage: 32,768 GiB. <p>Limits</p> <ul style="list-style-type: none"> • Only data disks can be encrypted. System disks cannot be encrypted. • Unencrypted disks cannot be directly converted into encrypted disks. • Encrypted disks cannot be directly converted into unencrypted disks. • Unencrypted snapshots cannot be directly converted into encrypted snapshots. • Encrypted snapshots cannot be directly converted into unencrypted snapshots. • Images with encrypted snapshots cannot be shared. • Images with encrypted snapshots cannot be exported.
Snapshot quota	The number of disks × 64.
Image	<ul style="list-style-type: none"> • Maximum number of users that a single image can be shared with: 50. • Instances with 4 GiB of memory or higher memory do not support 32-bit images.

Type	Description
Security group	<ul style="list-style-type: none"> • A single security group cannot contain more than 1,000 instances. If more than 1,000 instances need to access each other over the intranet, you can distribute them to different security groups and authorize mutual access among the security groups. • Each instance can belong to a maximum of five security groups. • Each user can have a maximum of 100 security groups. • Each security group can have a maximum of 100 security group rules. • Adjusting security groups will not affect the continuity of services. • Security groups are stateful. If a security group permits outbound traffic over a link, it also permits inbound traffic over this link.
Elastic network interface	The number of elastic network cards that can be bound to different instance type families. For details, see Instance types .
Instance user data	Currently, ECS UserData supports VPC + I/O optimized instances. You can use the UserData function when creating such instances. Because UserData depends on the cloud-init service, cloud-init must be installed in the image.

2.7. Terms

ECS

A simple and efficient cloud computing service that provides elastic processing capabilities and supports operating systems such as Linux and Windows.

instance

An independent resource entity that contains basic resource elements.

security group

A virtual firewall that provides status detection and packet filtering functions and is used to control the network access of one or more ECS instances. Instances in the same security group are able to communicate with each other, while instances in different security groups are isolated from each other. You can configure the rules of two security groups to authorize mutual access between them.

image

A running environment template for ECS instances. An image includes an operating system and preinstalled software. Images can be divided into public images and custom images. You can use an image to create an ECS instance or change the system disk of an ECS instance.

snapshot

Data backup of a disk at a certain point in time. Includes automatic snapshots and manual snapshots.

cloud disk

An independent disk that can be attached to any ECS instance in the same zone of the same region. Cloud disks are divided by performance into ultra disks, SSD disks, and basic disks.

block storage

A low-latency and high-reliability persistent random block-level data storage service provided by Apsara Stack for ECS.

throughput

The amount of data successfully transmitted through a network, device, port, virtual circuit, or another facility within a given period of time.

performance test

A world-leading SaaS performance test platform, with powerful distributed stress test capability. It can simulate real business scenarios with large amounts of users to locate all application performance problems.

Virtual Private Cloud (VPC)

A virtual private cloud built and customized based on Apsara Stack. Full logical isolation is achieved between VPCs. Users can create and manage cloud services, such as ECS instances, Intranet Server Load Balancer (SLB) instances, and RDS instances in their own VPCs.

intranet IP address

A service connection address for a client that uses a private IP address as the source.

GPU instance

A GPU-based computing service used in scenarios such as video decoding, graphics rendering, deep learning, and scientific computation. GPU instances feature real-time and high-speed computation and provide powerful concurrent and floating point computing capabilities.

3. Container Service

3.1. What is Container Service?

Container Service provides high-performance, enterprise-class management for scalable Kubernetes-based containerized applications throughout the application lifecycle.

Container Service simplifies the creation and scaling of container management clusters. It integrates Apsara Stack virtualization, storage, network, and security capabilities, providing the optimal environment to run Kubernetes-based containerized applications in the cloud. Alibaba Cloud is a Kubernetes certified service provider, with Container Service being among the first services to pass the Certified Kubernetes Conformance Program. Container Service provides professional container support and services.

3.2. Benefits

Overview

Easy to use

- You can easily create Kubernetes clusters in the Container Service console.
- You can easily upgrade Kubernetes clusters in the Container Service console.

When you use custom Kubernetes clusters, you may need to handle clusters of different versions. Currently, each time you upgrade the clusters, you need to make major adjustments and high operation and maintenance costs are incurred. Container Service allows you to perform rolling upgrades based on images and supports full metadata backups. You can easily roll back clusters to previous versions.

- Allows you to easily scale Kubernetes clusters in the Container Service console.

Kubernetes clusters enable you to quickly scale up or down applications to handle traffic fluctuations in a timely manner.

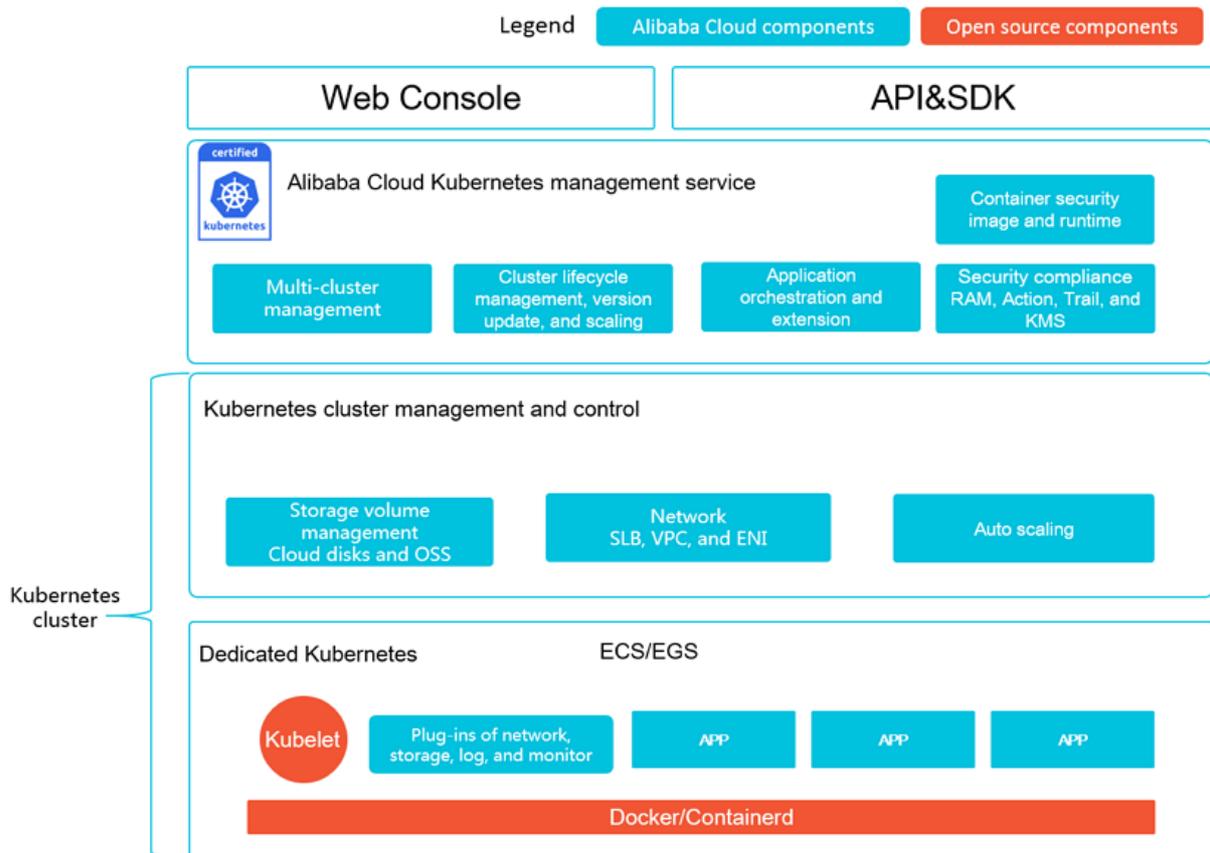
Features

Feature	Description
Network	Supports continuous network integration to optimize network performance.

Feature	Description
Load balancing	<p>Allows you to create public and internal SLB instances.</p> <p>If you use an Ingress to control access to your Kubernetes cluster, frequent service releases may negatively affect the performance of the Ingress and increase the error rate. Container Service allows you to create SLB instances, which provide high availability load balancing and can automatically modify network configurations to suit your business needs. This solution is adopted by a large number of users and has been proven to be a more stable and reliable alternative to Ingresses.</p>
Storage	<p>Supports Apsara Stack cloud disks, Network Attached Storage (NAS), and Block Storage, and provides FlexVolume drivers.</p> <p>Supports seamless integration with cloud storage services for custom Kubernetes clusters that cannot use cloud storage resources.</p>
O&M	<ul style="list-style-type: none"> • Supports integration with Apsara Stack Log Service. • Supports automatic scaling.
Image repository	<ul style="list-style-type: none"> • Provides high availability and high concurrency. • Supports accelerated image retrieval. • Supports peer-to-peer image distribution. <p>Custom image repositories may stop responding when millions of clients attempt to pull images at the same time. Container Service provides an image repository system that offers enhanced reliability and reduces O&M and upgrade costs.</p>
Stability	<ul style="list-style-type: none"> • Dedicated support teams guarantee the stability of containers. • All Linux and Kubernetes versions must pass rigorous testing before they are available to the public. <p>Container Service supports Docker CE and provides a Docker community to help you communicate with other Docker enthusiasts and solve problems. Best practices are provided to help you address issues, such as network interruptions, kernel incompatibilities, or Docker crashes.</p>

Feature	Description
Technical support	<ul style="list-style-type: none"> Allows you to quickly upgrade Kubernetes clusters to the latest version. Provides professional technical support services to help you solve the issues that may occur when you use containers.

3.3. Architecture



Container Service is adapted and enhanced on the basis of native Kubernetes. This service simplifies cluster creation and scaling and integrates Apsara Stack virtualization, storage, network, and security capabilities, providing the optimal environment to run Kubernetes-based containerized applications in the cloud.

Feature	Description
Dedicated Kubernetes mode	Integrated with Apsara Stack virtualization technologies, the service allows you to create dedicated Kubernetes clusters. ECS, Elastic GPU Service (EGS), and ECS Bare Metal instances can all be used as cluster nodes. Instances support a wide range of plug-ins and can be flexible configured to different specifications.

Feature	Description
Alibaba Cloud Kubernetes cluster management and control service	The service provides powerful network, storage, cluster management, scaling, and application extension features.
Alibaba Cloud Kubernetes management service	The service supports secure images and is highly integrated with Apsara Stack Resource Access Management (RAM), Key Management Service (KMS), and logging and monitoring services to provide a secure and compliant Kubernetes solution.
Convenient and efficient use	Container Service for Kubernetes provides services through the Web console, APIs, and SDKs.

3.4. Features

Features

Cluster management

- With the Container Service console, you can easily create a classic dedicated Kubernetes cluster supporting GPU servers within 10 minutes.
- Provides container-optimized OS images as well as Kubernetes and Docker versions that have undergone stability testing and security enhancement.
- Supports multi-cluster management, cluster upgrades, and cluster scaling.

Provides end-to-end container lifecycle management

- **Network**

Provides high performance VPC and elastic network interface (ENI) plug-ins optimized for Apsara Stack, boasting 20% increased performance compared with regular network solutions.

Supports container access and throttling policies.

- **Storage**

Container Service is integrated with Apsara Stack disks and OSS, and provides the standard FlexVolume drive.

Supports real-time creation and migration of volumes.

- **Logs**

Provides high-performance log collection integrated with Apsara Stack Log Service.

Supports the integration with third-party open-source logging solutions.

- **Monitoring**

Supports both container-level and VM-level monitoring. Integration with third-party open-source monitoring solutions is supported.

- **Permissions**

Supports cluster-level Resource Access Management (RAM).

Supports application-level permission configuration management.

- **Application management**

Supports phased release and blue-green release.

Supports application monitoring and scaling.

High-availability scheduling policies that allow you to easily handle upstream and downstream delivery processes

- Supports service-level affinity policies and scale-out.
- Provides high availability and disaster recovery across zones.
- Provides cluster and application management APIs to easily implement continuous integration and private system deployment.

3.5. Scenarios

DevOps continuous delivery

Optimized continuous delivery pipeline

Container Service works with Jenkins to automate the DevOps pipeline, from code submission to application deployments. The service ensures that code is only submitted for deployment after passing automated testing, and provides a better alternative to traditional delivery models that involve complex deployments and slow iterations.

Benefits

- DevOps pipeline automation

Automates the DevOps pipeline, from code updates to code builds, image builds, and application deployments.

- Consistent environment

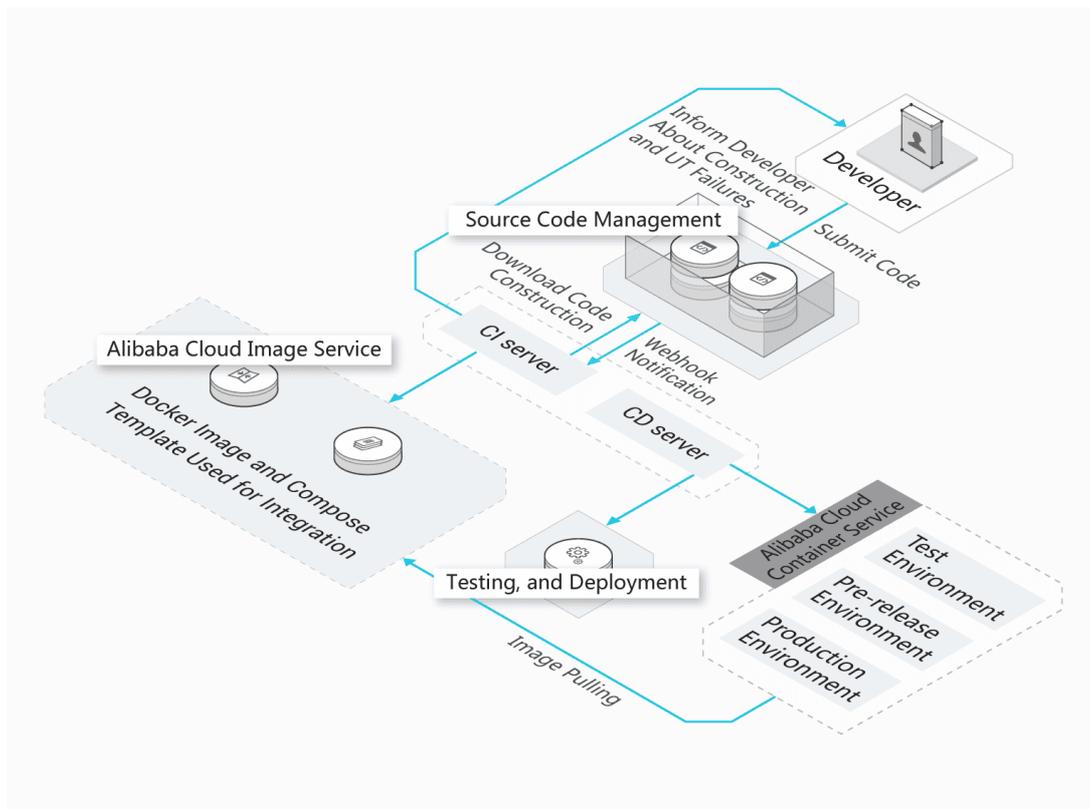
Allows you to deliver code and runtime environments based on the same architecture.

- Continuous feedback

Provides immediate feedback on each integration or delivery.

Related products and services

ECS + Container Service



Machine learning based on cloud-native technology

Enables rapid application developments with a focus on machine learning

Container Service allows data engineers to easily develop and deploy machine learning applications in heterogeneous computing clusters. Integrated with multiple distributed storage systems, the service supports faster read and write speeds to facilitate the testing, training, and release of data models. You can focus on your core business operations instead of worrying about the deployment and maintenance process.

Benefits

- Ecosystem support

Supports mainstream deep learning frameworks, such as TensorFlow, Caffe, MXNet, and Pytorch, and offers optimized features of these frameworks.

- Quick start and elastic scaling

Provides machine learning services for development, training, and inference. Supports the startup of training and inference tasks within seconds, and elastic scaling of GPU resources.

- Easy to use

Allows you to easily create and manage large-scale GPU clusters and monitor core metrics, such as GPU utilization.

- Deep integration

Seamless integration with Apsara Stack storage, logging and monitoring, and security infrastructure capabilities.

Related products and services

ECS/EGS/HPC + Container Service + OSS/NAS/CPFS

Microservices architecture

Agile development and deployment to speed up the evolution of business models

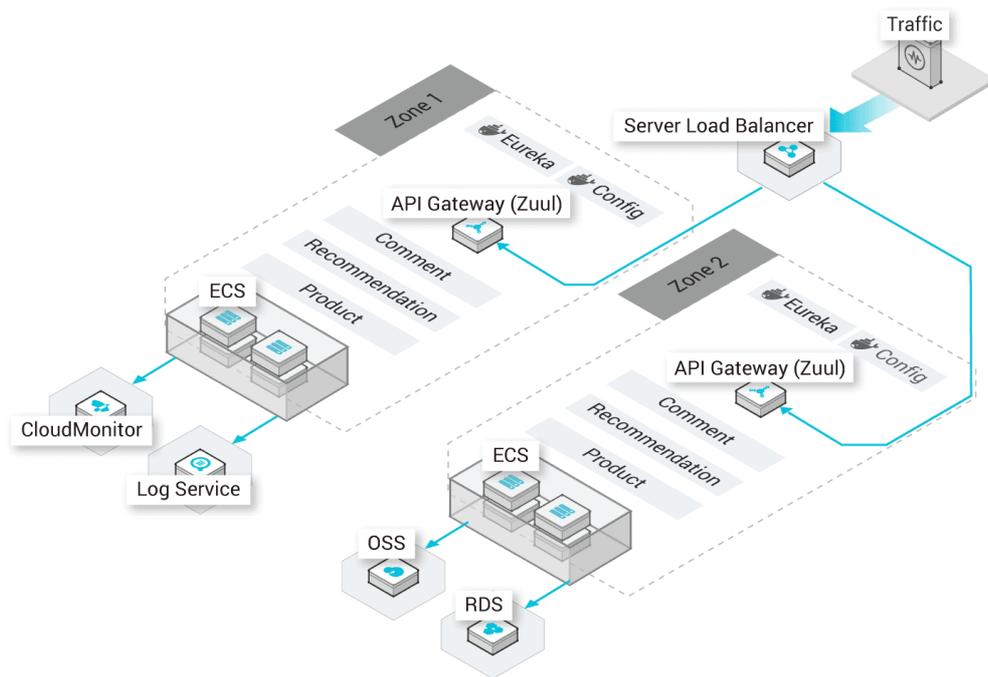
In the production environment, you can split your system into microservices and use Apsara Stack image repositories to store these microservice applications. Apsara Stack can schedule, orchestrate, deploy, and implement phased releases of microservice applications while you focus on feature updates.

Benefits

- **Load balancing and service discovery**
Forwards layer 4 and layer 7 requests and binds the requests to backend containers.
- **Multiple scheduling and disaster recovery policies**
Supports different levels of affinity scheduling policies, and cross-zone high availability and disaster recovery.
- **Microservices monitoring and auto scaling**
Supports microservice and container monitoring, and microservice auto scaling.

Related products and services

ECS + ApsaraDB RDS + OSS + Container Service



Hybrid cloud architecture

Unified O&M of cloud resources

You can centrally manage cloud and on-premises resources in the Container Service console. Containers hide the differences between infrastructures. This enables you to use the same images and orchestration templates to deploy applications in the cloud and on premises.

Benefits

- Application scaling in the cloud

During peak hours, Container Service can scale up applications in the cloud and forward traffic to the scaled-up resources.

- Disaster recovery in the cloud

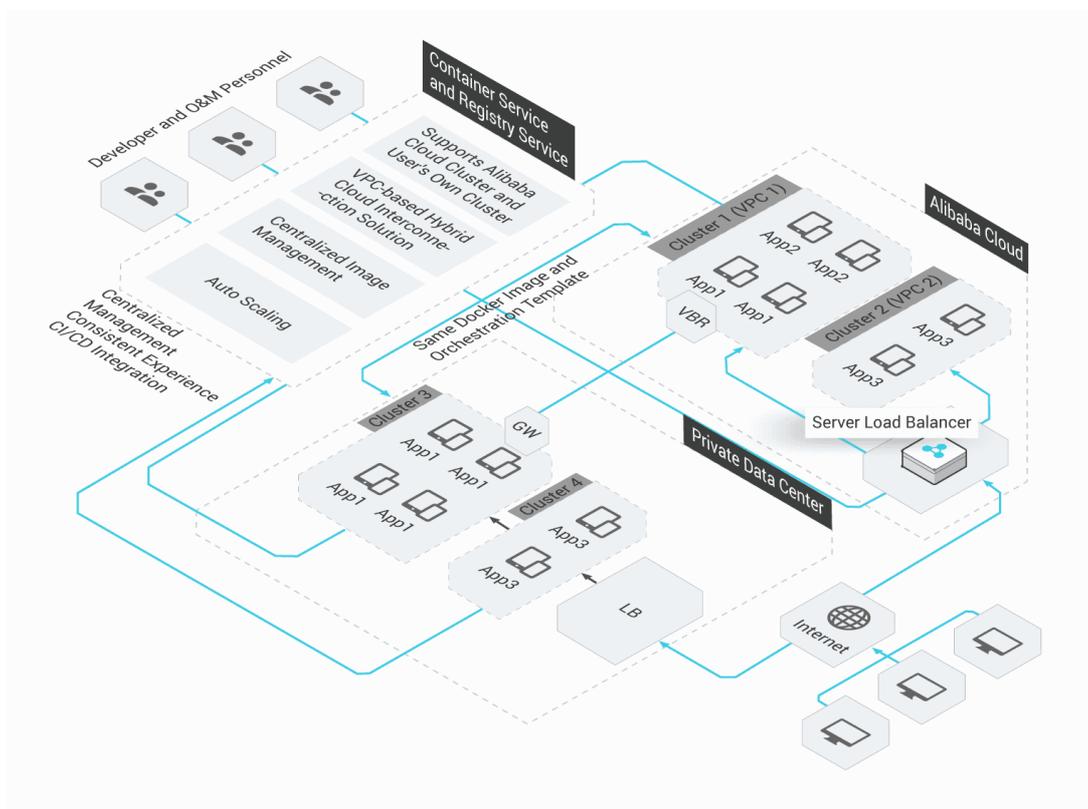
Business systems can be deployed on premises for service provisioning and in the cloud for disaster recovery.

- On-premises development and testing

Applications that are developed and tested on premises can be seamlessly released to the cloud.

Related products and services

ECS + VPC + Express Connect



Automatic scaling architecture

Traffic-based scalability

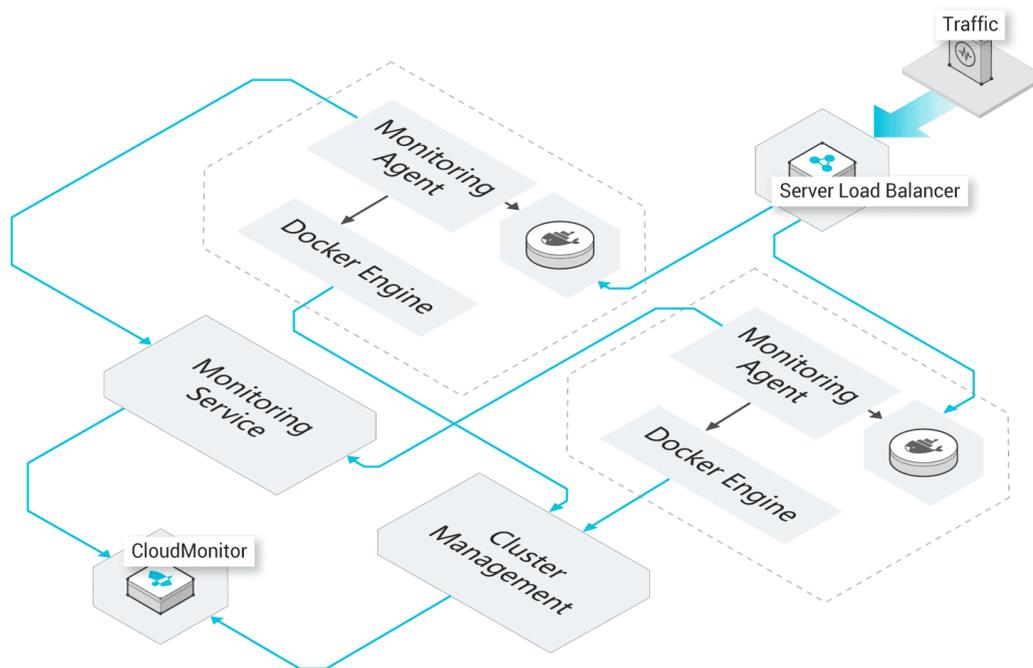
Container Service enables businesses to auto-scale their resources based on traffic. This prevents traffic spikes from bringing down your system and eliminates idle resources during off-peak hours.

Benefits

- Quick response
Container scale-out can be triggered within seconds when traffic reaches the scale-out threshold.
- Auto scaling
The scaling process is fully automated without human interference.
- Low cost
Containers are automatically scaled in when traffic decreases to avoid resource waste.

Related products and services

ECS + CloudMonitor



3.6. Limits

Limits for Kubernetes clusters

Limit	Description
-------	-------------

Limit	Description
Cluster	<ul style="list-style-type: none"> You can create up to 50 clusters across all regions for each account. A cluster can contain up to 40 nodes. To create more clusters or nodes, submit a ticket. Kubernetes clusters only support Linux containers. Kubernetes clusters only support VPCs. When creating a Kubernetes cluster, you can either create a new VPC or use an existing one.
ECS instance	<ul style="list-style-type: none"> Only the CentOS operating system is supported. Limits for adding an existing ECS instance: <ul style="list-style-type: none"> The ECS instance to be added must be in the same VPC as the cluster. The ECS instance to be added must belong to the same account as the cluster.
Cluster scale-in and scale-out	<ul style="list-style-type: none"> The number of worker nodes must be within the range of 1 to 5. Clusters must be manually scaled in or out. Automatic scaling is not supported. Master nodes in a Kubernetes cluster cannot be scaled out automatically. Based on the rules of Resource Orchestration Service (ROS), nodes that were created automatically during cluster creation and nodes that were manually added to a cluster will not be removed when you scale in the cluster. Only nodes you added when scaling out the cluster will be removed. Nodes are removed from the cluster in reverse order to when they were added to the cluster during cluster scale-out. Newly added nodes are reclaimed first.

3.7. Terms

cluster

A collection of cloud resources that are required to run containers. Several cloud resources, such as ECS instances, SLB instances, and VPCs, are associated together to form a cluster.

node

A server that has a Docker engine installed and is used to deploy and manage containers. A node can be either an ECS instance or a physical server. The Container Service Agent program is installed on a node and registered to a cluster. The number of nodes in a cluster can be scaled based on your requirements.

container

A runtime instance created from a Docker image. A single node can run multiple containers.

image

A standard packaging format of a containerized application in Docker. An image from the Docker Hub, Alibaba Cloud Container Registry, or your own private registry can be specified to deploy its packaged containerized application. image ID An image ID is a unique identifier composed of the image repository URI and image tag. The latest image tag is used for the image ID by default.

Kubernetes terms

node

A worker server in a Kubernetes cluster. A node can be either a virtual server or a physical server. Pods always run on nodes. kubelet runs on each node in a cluster to manage containers in a pod and ensure that they are running properly.

namespace

A method used in Kubernetes to divide cluster resources between multiple users. By default, Kubernetes starts with three initial namespaces: default, kube-system, and kube-public. Administrators can also create new namespaces as required.

pod

The smallest deployable computing unit that can be created and managed in Kubernetes. A pod is a group of one or more containers that share storage and network resources and a common set of specifications for how to run the containers.

Replication Controller (RC)

A feature that monitors running pods to ensure that a specified number of pod replicas are running at any given time. One or more pod replicas can be specified. If the number of pod replicas is smaller than the specified value, an RC starts new pod replicas. If the number of pod replicas exceeds the specified value, the RC stops the redundant pod replicas.

Replica Set (RS)

The upgraded version of RC. Compared with RCs, RSs support more selector types. RS objects are not used independently, but are used as deployment parameters under ideal conditions.

deployment

An update operation performed on a Kubernetes cluster. Deployment is more widely applied than RS. You can use deployments to create, update, or perform rolling updates for services. A new RS is created when you perform a rolling update for a service. A compound operation is carried out to increase the number of replicas in the new RS to the desired value while decreasing the number of replicas in the original RS to zero. This kind of compound operation is better carried out by a deployment than through RS. We recommend that you do not manage or use the RS created by a deployment.

service

The basic operation unit of Kubernetes. It is an abstraction of real application services. Each service has multiple containers that support it. The Kube-Proxy port and service selector determine whether the service request is forwarded to the back-end container, and a single access interface is displayed externally. Back-end operations are invisible to users.

label

A collection of key-value pairs attached to resource objects. Labels are intended to specify identifying attributes of objects that are meaningful and relevant to users, but do not directly imply semantics to the core system. Labels can be attached to objects at creation time, and subsequently added and modified at any time. Each object can have a set of key/value labels, and each key must be unique for a specified object.

volume

Volumes in Kubernetes clusters are similar to Docker volumes. However, they are different in one key aspect. Docker volumes are used to persist data in Docker containers, while Kubernetes volumes share the same lifetime as the pods that enclose them. The volumes declared in each pod are shared by all containers in the pod. The actual back-end storage technology used is irrelevant when you use Persistent Volume Claim (PVC) logical storage. The specific configurations for Persistent Volume (PV) are completed by storage administrators.

PV and PVC

PVs and PVCs allow Kubernetes clusters to provide a logical abstraction over the storage resources, so that the actual configurations of back-end storage can be ignored by the pod configuration logic, and instead completed by the PV configurators. The relationship between PVs and PVCs is similar to that of nodes and pods. PVs and nodes are resource providers which can vary by cluster infrastructure, and are configured by the administrators of a Kubernetes cluster. PVCs and pods are resource consumers that can vary based on service requirements, and are configured by either the users or service administrators of a Kubernetes cluster.

Ingress

A collection of rules that allow inbound access to cluster services. An Ingress can be configured to provide services with externally-reachable URLs, load balance traffic, terminate SSL, and offer name-based virtual hosting. You can request the Ingress by posting Ingress resources to API servers. An Ingress controller is responsible for fulfilling an Ingress, usually with a load balancer. It can also be used to configure your edge router or additional frontends to help handle the traffic.

Related documents

- [Docker glossary](#)
- [Kubernetes concepts](#)

4.Auto Scaling (ESS)

4.1. What is ESS?

Auto Scaling (ESS) is a management service that automatically adjusts the number of elastic computing resources based on your business demands and strategies.

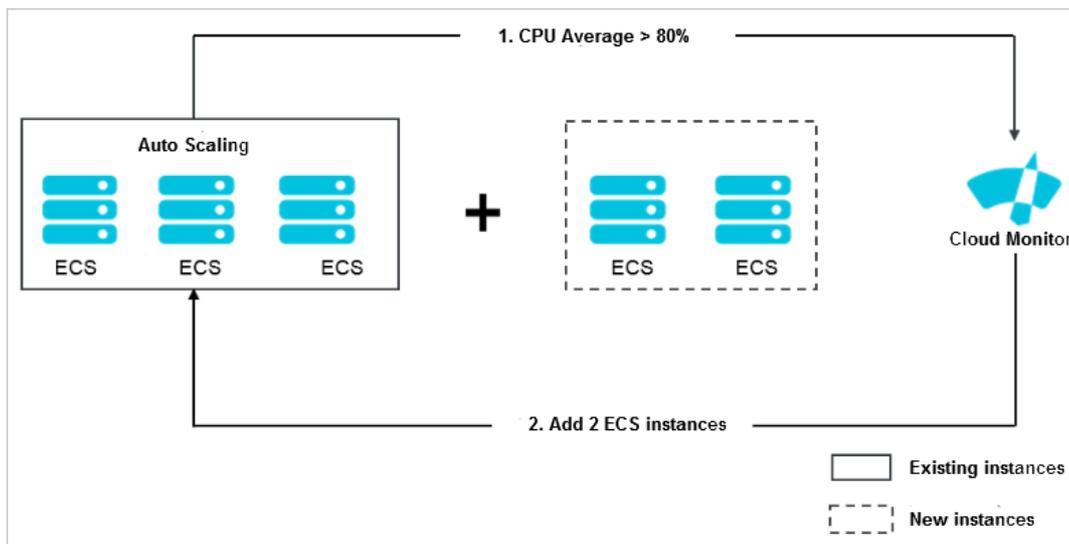
Based on user-defined scaling rules, ESS automatically adds ECS instances as business loads increase to ensure sufficient computing capabilities. When your business loads decrease, ESS automatically removes ECS instances to reduce running costs.

ESS provides the following functions:

- Elastic scale-out

When business loads surge, ESS automatically increases underlying resources. This helps maintain access speed and ensure that resources are not overloaded. For example, if the CPU utilization of ECS instances exceeds 80%, ESS scales out ECS resources based on the rules you defined. During the scale-out process, ESS automatically creates and adds ECS instances to a scaling group, and adds the new instances to the SLB instance and RDS whitelist. **Elastic scale-out** shows the process.

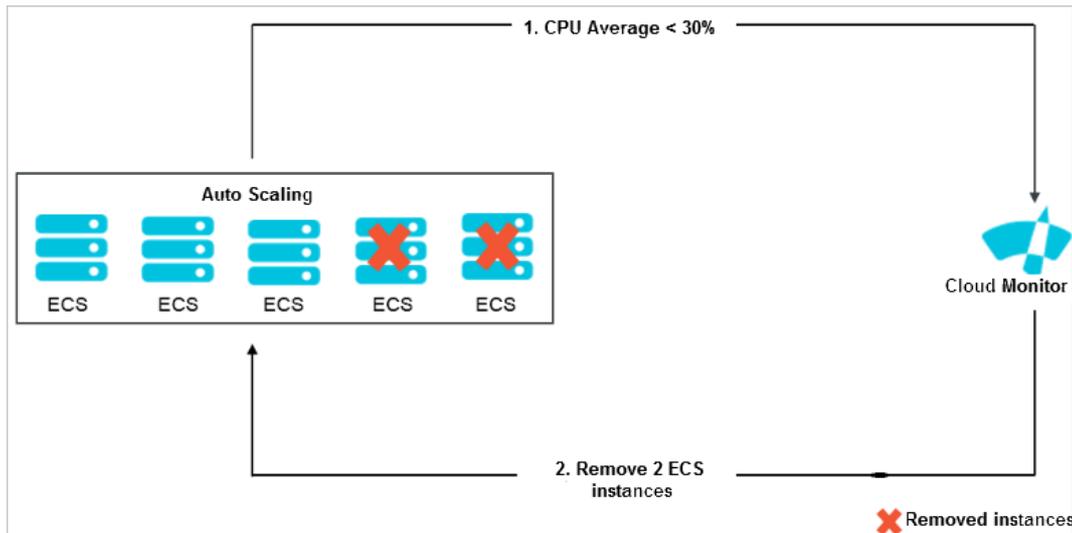
Elastic scale-out



- Elastic scale-in

When business loads decrease, ESS automatically releases underlying resources. This prevents resource wastage and helps to reduce cost. For example, if the CPU utilization of ECS instances in a scaling group falls below 30%, ESS scales in ECS resources based on the rules you defined. During the scale-in process, ESS removes the ECS instances from the scaling group, the SLB instance, and RDS whitelist. **Elastic scale-in** shows the process.

Elastic scale-in



- **Elastic recovery**

The health status of ECS instances in a scaling group is determined based on the life cycle of the instances. If an ECS instance is in an unhealthy state, ESS automatically releases the instance and creates a new one. ESS adds the new instance to the SLB instance and RDS whitelist. This process is called elastic recovery. It ensures that the number of healthy ECS instances in a scaling group will not fall below the threshold that you defined.

4.2. Benefits

ESS has the following benefits:

- **Automatic scaling of instances on-demand**

ESS can automatically add ECS instances during peak traffic hours, and remove ECS instances during off-peak hours to scale with actual business needs. This helps to lower infrastructure costs because you only pay for what you actually use.

- **Real-time instance monitoring and automatic replacement of unhealthy instances**

ESS performs real-time monitoring on instances and automatically replaces unhealthy instances that are discovered, reducing operations and maintenance (O&M) overheads.

- **Intelligent whitelist management and control, no user intervention required**

ESS is integrated with Server Load Balancer (SLB) and ApsaraDB for Relational Database Service (RDS). It automatically manages SLB backend servers and RDS whitelists, eliminating the need to perform manual O&M.

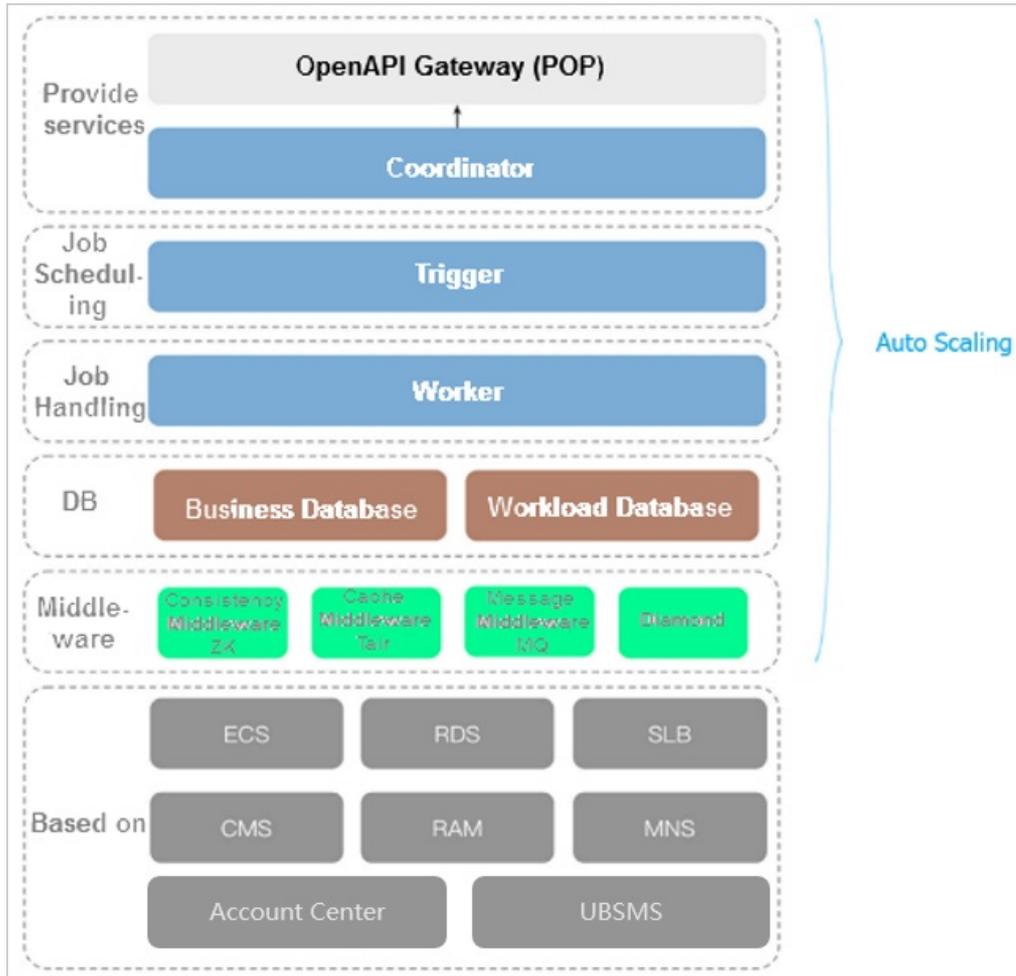
- **Various scaling modes for you to mix and match**

ESS allows you to schedule, customize, fix the minimum number of instances, and configure automatic replacement of unhealthy instances. It also provides APIs to allow you to monitor instances through external monitoring systems.

4.3. Architecture

[Auto Scaling architecture](#) shows the Auto Scaling architecture details.

Auto Scaling architecture



Architecture description

Component	Description
OpenAPI Gateway	Provides basic services such as authentication and parameter passthrough.
Coordinator	Serves as the ingress of the ESS architecture. It provides external management and control for services, processes API calls, and triggers tasks.
Trigger	Obtains information from the health checks of instances and scaling groups, scheduled tasks, and CloudMonitor to perform tasks scheduling.
Worker	Functions as the core part of ESS. After receiving a task, it handles the entire life cycle of the task, including splitting, executing, and returning the execution results.
Database	Includes the business database and workload database.
	ZooKeeper: ensures consistency by implementing distributed locks for Server Controller.

Component	Description
Middleware layer	Tair: provides caching services for Server Controller.
	Message Queue (MQ): provides message queuing services of VM statuses.
	Diamond: manages persistent configurations.

4.4. Features

ESS has the following features:

- **Automatically adding or removing ECS instances based on your business demands**

You can use the following scaling modes to adjust the number of ECS instances:

- **Scheduled mode:** Configure periodic tasks to add or remove ECS instances at a specified point in time, such as 13:00 every day.
 - **Custom mode:** Call APIs to manually adjust the number of ECS instances based on monitoring system statistics.
 - You can manually implement scaling rules.
 - You can manually add or remove existing ECS instances.
 - After you have manually adjust MinSize (the minimum number of instances) and MaxSize (the maximum number of instances), ESS automatically creates or releases ECS instances to ensure that the number of instances remains within the MinSize and MaxSize range.
 - **Fixed-number mode:** Maintain a fixed number of healthy ECS instances by specifying the MinSize attribute. This mode can be used to ensure day-to-day business availability.
 - **Health mode:** Automatically remove or release ECS instances when they are detected as unhealthy (such as they are not in the running state).
 - **Multimode:** Combine any of the preceding modes to meet your own business requirements. For example, if you predict that business peak hours are between 13:00 to 14:00, you can configure a scaling mode that creates 20 ECS instances at the scheduled time. If you are not sure whether the actual demand during peak hours will exceed the number of scheduled resources (for example, the actual load requires 40 ECS instances), another scaling mode can be configured to handle unexpected business loads.
- **Automatically adding or removing ECS instances to or from the SLB backend server group**

The health status of an ECS instance in a scaling group is determined based on the life cycle of the instances. If an ECS instance is in an unhealthy state, ESS automatically removes the instance and creates a new one. ESS then adds the new instance to the SLB instance and RDS whitelist.

 **Note** ECS instances used for automatic scaling can be removed. Therefore, these instances cannot be used to store application status information (such as sessions) and related data (such as databases and logs). If applications deployed on these ECS instances require data to be saved, you can save the status information to independent ECS instances, databases to RDS, and logs to Log Service.

- **Automatically adding or removing IP addresses of ECS instances to or from the RDS whitelist**

When an ECS instance is automatically added to or removed from an SLB backend server group, the IP address of the ECS instance is also automatically added or removed from the RDS whitelist. This mechanism automatically maintains the RDS whitelist and effectively controls access to the RDS instance.

4.5. Scenarios

ESS can be used in the following scenarios:

- **Video streaming:** Traffic loads surge during holidays and festivals. Cloud computing resources must be automatically scaled out to meet the increased demands.
- **Live streaming and broadcast:** Traffic loads are ever-changing and difficult to predict. Cloud

computing resources must be scaled based on CPU utilization, application load, and bandwidth usage.

- Gaming: Traffic loads increase at 12:00 and from 18:00 to 21:00. Cloud computing resources must be scaled out on a regular basis.

4.6. Limits

Auto Scaling has the following limits:

- Applications on ECS instances deployed in a scaling group must be stateless and horizontally scalable.
- Instances created by Auto Scaling cannot be automatically added to the instance access whitelist of KVStore for Memcache. You must manually add the instances to the whitelist. For more information, see *KVStore for Memcache Product Introduction*.
- Auto Scaling does not support vertical scaling. It can only scale the number of ECS instances. The CPU, memory, and bandwidth configurations of the ECS instances cannot be automatically adjusted.
- Scaling configurations, scaling rules, and scaling activities are dependent on the lifecycle of a scaling group. If a scaling group is deleted, all scaling group configurations, rules, and activities associated with this group are also deleted.
- Scheduled tasks are independent from scaling groups. Deleting a scaling group does not affect the scheduled tasks.
- Each user can create a limited number of scaling groups, scaling configurations, scaling rules, ECS instances for scaling, and scheduled tasks. For more information, see [Quantity restrictions](#).

Quantity restrictions

Item	Description
Scaling group	You can create up to 20 scaling groups.
Scaling configuration	You can create up to 10 scaling configurations in a scaling group.
Scaling rule	You can create up to 10 scaling rules in a scaling group.
ECS instance for scaling	<p>You can configure up to 100 ECS instances for automatic scaling in a scaling group.</p> <div style="background-color: #e0f2f1; padding: 5px; border: 1px solid #ccc;"> <p> Note The limit applies to the ECS instances that are automatically created, but does not apply to manually added ones.</p> </div>
Scheduled task	You can create up to 20 scheduled tasks.

4.7. Terms

Auto Scaling

Auto Scaling (ESS) is a management service that automatically adjusts the number of elastic computing resources based on your business demands and strategies. It automatically creates ECS instances during high business loads, and automatically releases ECS instances during low business loads.

Scaling group

A scaling group is a group of ECS instances that is dynamically scaled based on the configured scenario. You can specify the maximum and minimum number of ECS instances in a scaling group, as well as SLB and RDS instances associated with the group.

Scaling configuration

A scaling configuration specifies the configurations of ECS instances in ESS.

Scaling rule

A scaling rule defines the specific scaling activity, for example, the number of ECS instances to be added or removed.

Scaling activity

After a scaling rule is triggered, a scaling activity is performed. A scaling activity shows the changes to the ECS instances in a scaling group.

Scaling trigger task

A scaling trigger task is a task that triggers a scaling rule, such as scheduled tasks.

Cooldown period

The cooldown period indicates a period of time after the completion of a scaling activity in a scaling group. During this period, no other scaling activities can be executed.

5.Object Storage Service (OSS)

5.1. What is OSS?

Alibaba Cloud Object Storage Service (OSS) is a massive, secure, low-cost, and highly reliable cloud storage service provided by Alibaba Cloud.

It can be considered as an out-of-the-box storage solution with unlimited storage capacity. Compared with the user-created server storage, OSS has many outstanding advantages in reliability, security, cost, and data processing capabilities. Using OSS, you can store and retrieve a variety of unstructured data files, such as text files, images, audios, and videos, over the network at any time.

OSS uploads data files as objects to buckets. OSS is an object storage service that uses a key-value pair format. You can retrieve object content based on unique object names (keys).

On OSS, you can:

- Create a bucket and upload objects to the bucket.
- Obtain an object URL from OSS to share or download an object.
- Complete the ACL settings of a bucket or object by modifying its properties or metadata.
- Perform basic and advanced OSS tasks through the OSS console.
- Perform basic and advanced OSS tasks using the Alibaba Cloud SDKs or directly calling the RESTful APIs in your application.

5.2. Advantages

Advantages of OSS over user-created server storage

Item	OSS	User-created server storage
Reliability	<ul style="list-style-type: none"> • The capacity is automatically expanded without affecting external services. • Offers automatic redundant data backup. 	<ul style="list-style-type: none"> • Prone to errors due to low hardware reliability. If a disk has a bad sector, data may be irretrievably lost. • Manual data restoration is complex and requires a lot of time and technical resources.
Security	<ul style="list-style-type: none"> • Provides hierarchical security protection for enterprises. • User resource isolation mechanisms and local disaster recovery • Provides various authentication and authorization mechanisms, as well as whitelisting, hotlinking protection, and RAM. It also provides Security Token Service (STS) for temporary access. 	<ul style="list-style-type: none"> • Additional scrubbing and black hole equipment is required. • A separate security mechanism is required.
Data processing	Image processing capabilities	Image processing capabilities must be purchased and deployed separately.

More benefits of OSS

- Ease of use

Provides standard RESTful APIs (some compatible with Amazon S3 APIs), a wide range of SDKs and client tools, and a management console. You can easily upload, download, retrieve, and manage large amounts of data for websites and applications, similar to regular files systems.

- There is no limit on the number and size of objects. Therefore, you can easily expand your buckets in OSS as required.
- Supports streaming writing and reading, which is suitable for business scenarios where you need to simultaneously read and write videos and other large objects.
- Supports lifecycle management. You can delete expired data in batches.

- Powerful and flexible security mechanisms

Flexible authentication and authorization mechanisms are available. OSS provides STS and URL authentication and authorization, as well as whitelisting, hotlinking protection, and RAM.

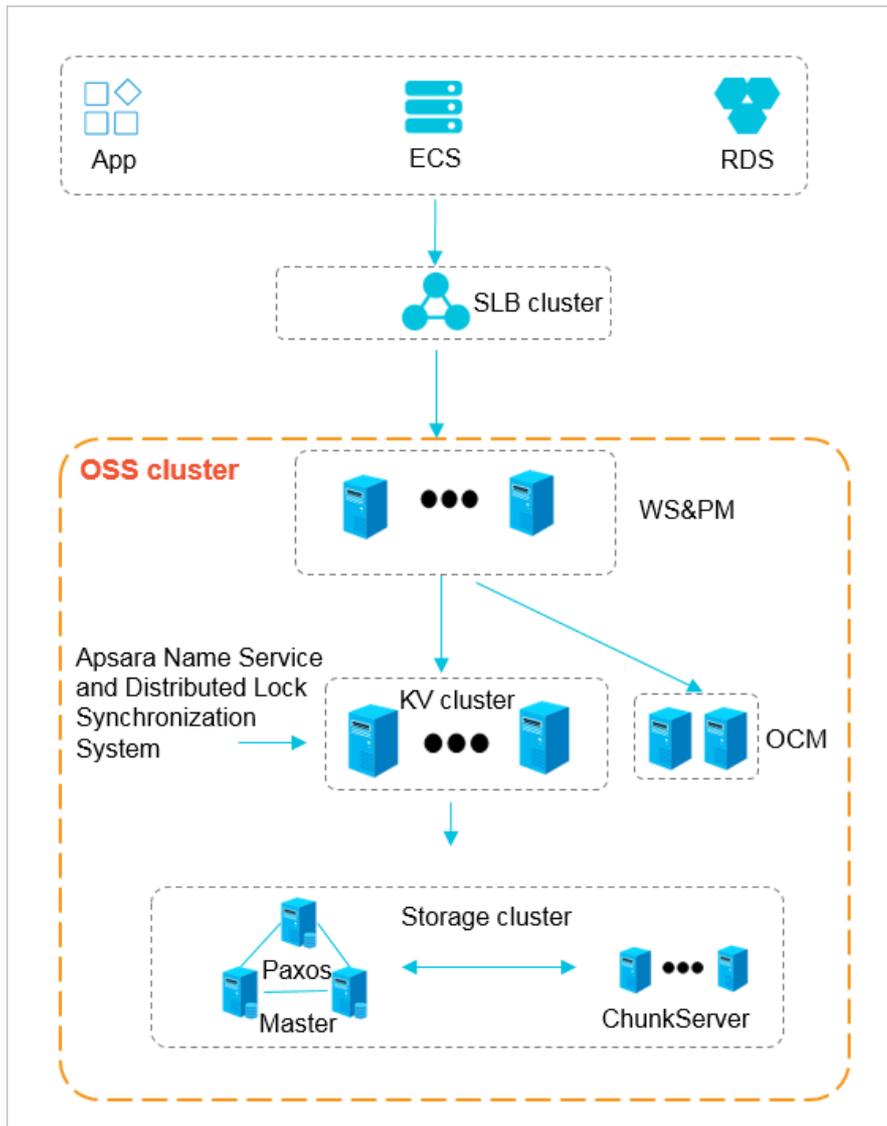
- Rich image processing functions

Supports format conversion, thumbnails, cropping, watermarking, resizing for object formats such as JPG, PNG, BMP, GIF, WEBP, and TIFF.

5.3. OSS architecture

Object Storage Service (OSS) is a storage solution built on the Alibaba Cloud Apsara platform. It is based on infrastructure such as Apsara Distributed File System and SchedulerX. Such infrastructure provides OSS and other Alibaba Cloud services with distributed scheduling, high-speed networks, and distributed storage features. The following figure shows the OSS architecture.

OSS architecture



- **WS & PM (the protocol layer):** is used for receiving users' requests sent through the REST protocol and performing authentication. If the authentication succeeds, users' requests are forwarded to the key-value engine for further processing. If the authentication fails, an error message is returned.
- **KV cluster:** is used for processing structured data, including reading and writing data based on keys. The KV cluster also supports large-scale concurrent requests. When a service has to operate on a different physical server due to a change in the service coordination cluster, the KV cluster can quickly coordinate and find the access point.
- **Storage cluster:** Metadata is stored on the master node. A distributed message consistency protocol (Paxos) is adopted between master nodes to ensure the consistency of metadata. This ensures efficient distributed storage and access of objects.

5.4. Functions

OSS offers the following functions:

OSS functions

Category	Function	Description
Bucket	Create a bucket	Before uploading an object to OSS, you need to create a bucket to store the objects.
	Delete a bucket	If you are no longer using a bucket, delete it to avoid incurring further fees.
	Modify ACL settings for a bucket	OSS provides an ACL for permission control. You can configure an ACL when creating a bucket and modify it after creating the bucket.
	Configure static website hosting	You can configure static website hosting for your bucket and access this static website through the bucket endpoint.
	Configure hotlinking protection	To prevent fees incurred by hotlinked OSS data, OSS supports hotlinking protection based on the referer field in the HTTP header.
	Manage CORS	OSS provides Cross-Origin Resource Sharing (CORS) settings in the HTML5 protocol to help you achieve cross-region access.
	Configure lifecycle	You can define and manage the lifecycle of all objects in a bucket or a subset of objects. Lifecycle is generally set for batch object management and automatic part deletion.
Object	Upload an object	You can upload all types of objects to a bucket.
	Create a folder	You can manage OSS folders in the same way you manage Windows folders.
	Search for objects	You can search for objects with the same prefix in a bucket or folder.
	Obtain an object URL	You can obtain an object URL from OSS to share or download an object.
	Delete an object	You can delete a single object or several objects in batches.
	Delete a folder	You can delete a folder or delete folders in batches.
	Modify ACL settings for an object	You can configure ACL settings when uploading an object and modify them after uploading the object.
	Manage parts	You can delete all or some parts from a bucket.
Image processing	Image processing	You can perform operations such as format conversion, cropping, scaling, rotating, watermarking, style encapsulation on images stored in OSS.

Category	Function	Description
OSS access control for VPC	Single tunnel	You can establish single tunnels to access OSS resources from VPC.
API	API	Provides RESTful API operations supported by OSS and relevant examples.
SDK	SDK	Provides SDK development operations and relevant examples in commonly used languages.

5.5. Scenarios

Massive storage for image, audio, and video applications

OSS can be used to store large amounts of data, such as images, audios, videos, and logs. OSS supports various devices. Websites and mobile applications can directly read or write OSS data. OSS supports file writing and streaming writing.

Dynamic and static content separation for websites and mobile applications

OSS leverages the BGP bandwidth to achieve ultra-low latency of direct data download.

Offline data storage

OSS is cheap and highly available, enabling enterprises to store data that needs to be archived offline for a long time to OSS.

5.6. Limits

Item	Description
Bucket	<ul style="list-style-type: none"> You can create a maximum of 10 buckets. After a bucket is created, its name and region cannot be modified.
Object upload	<ul style="list-style-type: none"> Objects uploaded through the Apsara Stack console, simple upload, form upload, and append upload cannot exceed 5 GB in size. To upload an object greater than 5 GB, you must use multipart upload. The size of each object uploaded through multipart upload cannot be greater than 48.8 TB. You can upload objects with the same name, but the existing objects are overwritten.
Object deletion	<ul style="list-style-type: none"> Deleted objects cannot be restored. You can delete up to 50 objects simultaneously through the Apsara Stack console. To delete more objects simultaneously, you must use APIs or SDKs.

Item	Description
Lifecycle	You can configure up to 1,000 lifecycle rules for each bucket.
Image processing (IMG)	<ul style="list-style-type: none"> • For a source image: <ul style="list-style-type: none"> ◦ Only JPG, PNG, BMP, GIF, WebP, and TIFF files are supported. ◦ The file size cannot exceed 20 MB. ◦ For image rotation, the width or height of the image cannot exceed 4,096 px. • For a thumbnail: <ul style="list-style-type: none"> ◦ The product dimensions cannot exceed 4,096 × 4,096 px. ◦ The length of each side cannot exceed 4,096 px.

5.7. Terms

This topic describes several basic terms used in OSS.

object

Files that are stored in OSS. They are the basic unit of data storage in OSS. An object is composed of Object Meta, object content, and a key. An object is uniquely identified by a key in the bucket. Object Meta defines the properties of an object, such as the last modification time and the object size. You can also specify User Meta for the object.

The lifecycle of an object starts when it is uploaded, and ends when it is deleted. Throughout the lifecycle of an object, Object Meta cannot be changed. Unlike the file system, OSS does not allow you to modify objects directly. If you want to modify an object, you must upload a new object with the same name as the existing one to replace it.

 **Note** Unless otherwise stated, objects and files mentioned in OSS documents are collectively called objects.

bucket

A container that stores objects. Objects must be stored in the bucket they are uploaded to. You can set and modify the properties of a bucket for object access control and lifecycle management. These properties apply to all objects in the bucket. Therefore, you can create different buckets to implement different management functions.

- OSS does not have the hierarchical structure of directories and subfolders as in a file system. All objects belong to their corresponding buckets.
- You can have multiple buckets.
- A bucket name must be globally unique within OSS and cannot be changed after a bucket is created.
- A bucket can contain an unlimited number of objects.

strong consistency

A feature of operations in OSS. Object operations in OSS are atomic, which indicates that operations are either successful or failed. There are no intermediate states. OSS never writes corrupted or partial data.

Object operations in OSS are strongly consistent. For example, after you receive a successful upload (PUT) response, the object can be read immediately, and the data is already written in triplicate. Therefore, OSS avoids the situation where no data is obtained when you perform the read-after-write operation. An object also has no intermediate states when you delete the object. After you delete an object, that object no longer exists.

Similar to traditional storage devices, modifications are immediately visible in OSS while consistency is guaranteed.

Comparison between OSS and the file system

OSS is a distributed object storage service that uses a key-value pair format. You can retrieve object content based on unique object names (keys). Although you can use names like test1/test.jpg, this does not necessarily indicate that the object is saved in a directory named test1. In OSS, test1/test.jpg is only a string, which is no different from a.jpg. Therefore, similar resources are consumed when you access objects that have different names.

A file system uses a typical tree index structure. Before accessing a file named test1/test.jpg, you must access directory test1 and then locate test.jpg. This makes it easy for a file system to support folder operations, such as renaming, deleting, and moving directories, because these operations are only directory node operations. System performance depends on the capacity of a single device. The more files and directories that are created in the file system, the more resources are consumed, and the lengthier your process becomes.

You can simulate similar functions in OSS, but this operation is costly. For example, if you want to rename test1 directory test2, the actual OSS operation would be to replace all objects whose names start with test1/ with copies whose names start with test2/. Such an operation would consume a large amount of resources. Therefore, try to avoid such operations when using OSS.

You cannot modify objects stored in OSS. A specific API must be called to append an object, and the generated object is of a different type from that of normally uploaded objects. Even if you only want to modify a single Byte, you must re-upload the entire object. A file system allows you to modify files. You can modify the content at a specified offset location or truncate the end of a file. These features make file systems suitable for more general scenarios. However, OSS supports sporadic bursts of access, whereas the performance of a file system is subject to the performance of a single device.

Therefore, mapping OSS objects to file systems is inefficient, which is not recommended. If attaching OSS as a file system is required, we recommended that you perform only the operations of writing data to new files, deleting files, and reading files. You can make full use of OSS capabilities. For example, you can use OSS to store and process large amounts of unstructured data such as images, videos, and documents.

6. Table Store

6.1. What is Table Store?

Table Store is a NoSQL database service independently developed by Alibaba Cloud. Table Store is a proprietary software program that is certified by the relevant authority in China. Table Store is built on the Apsara system of Alibaba Cloud, and can store large amounts of structured data and allow real-time access to these data.

Table Store provides the following features:

- Table Store offers schema-free data structure storage. You do not need to define attribute columns before you use them. You do not require table-level changes to add or reduce attribute columns. You can enable time to live (TTL) on a table to delete expired data from the table.
- Adopts the triplicate technology to keep three copies of data on three servers across three different racks. Each cluster supports either pure SSD instances or mixed storage instances (SSD and SATA) to meet different budget and performance requirements.
- Adopts a fully redundant architecture that prevents single point of failures (SPOFs). With support for online smooth upgrades, hot cluster upgrades, and automatic data migration, you can dynamically add or remove nodes without service interruptions for maintenance. The concurrent read/write throughput and storage capacity can be linearly scaled. Each cluster can have no less than 500 hosts.
- Supports highly concurrent read/write operations. Concurrent read/write capabilities can be scaled out with the increase of hosts. The read/write performance is indirectly related to the amount of data in a single table.
- Supports identity authentication and multi-tenancy. Comprehensive access control and isolation mechanisms are provided to safeguard your data. VPC and access over HTTPS are supported. Provides multiple authentication and authorization mechanisms so that you can define access permissions for individual tables and operations.

6.2. Benefits

Scalability

There is no upper limit to the amount of data that can be stored in Table Store tables. As data increases, Table Store adjusts partitions to provide more storage space for tables and improve the capability of handling access request bursts.

High performance

If you use a high-performance instance, its average access latency of single rows is measured in single-digit milliseconds. The read/write performance is not affected by the size of data in a table.

Reliability

Table Store provides high data reliability. It stores multiple data copies and restores data when some copies become invalid.

High availability

Through automatic failure detection and data migration, Table Store shields applications from host- and network-relevant hardware faults to achieve high availability.

Ease of management

Table Store automatically performs complex O&M tasks, such as the management of data partitions, software and hardware upgrades, configuration updates, and cluster scale-out.

Access security

Table Store provides multiple permission management mechanisms. It verifies and authenticates the identity of the request to prevent unauthorized data access, improving the data security.

Strong consistency

Table Store ensures strong data consistency for data writes. A successful write operation indicates that the data is written to three copies and stored in disks. Applications can read the latest data immediately.

Flexible data models

Table Store tables do not require a fixed format. Each row can contain a different number of columns. Table Store supports multiple data types, such as Integer, Boolean, Double, String, and Binary.

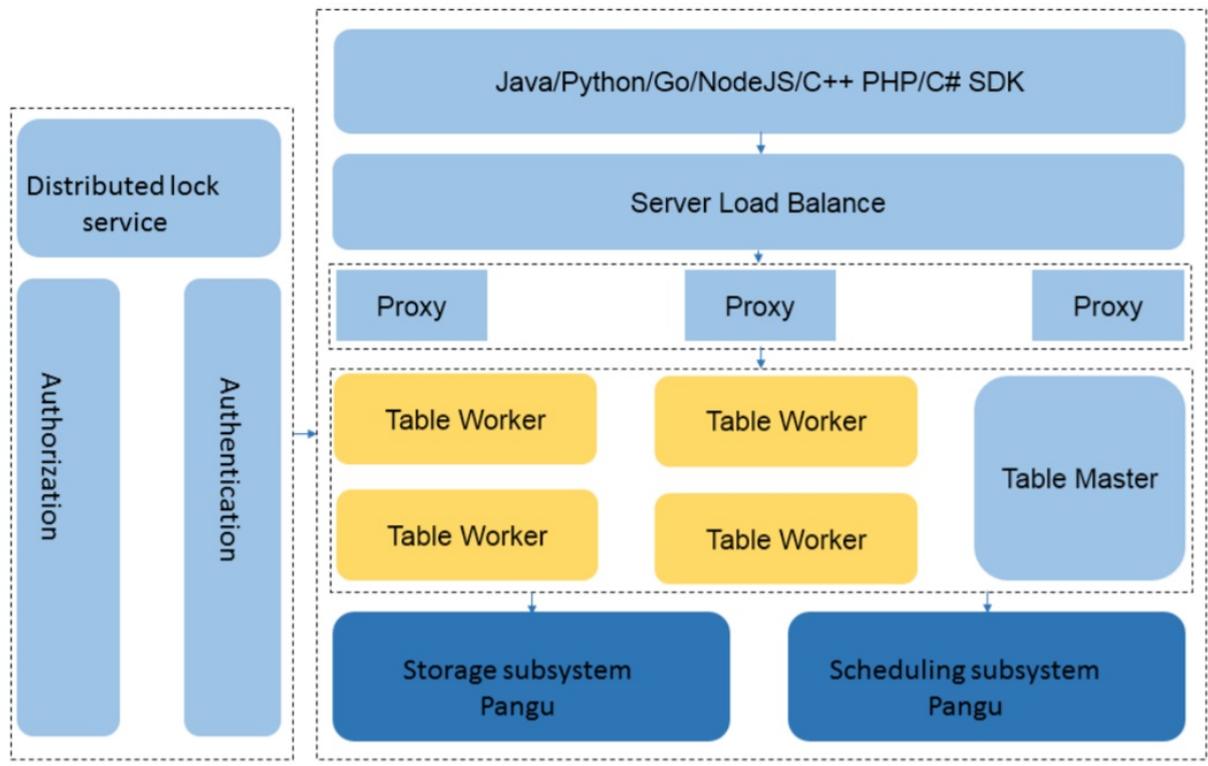
Monitoring integration

You can log on to the Table Store console to obtain monitoring information in real time, including the requests per second and average response latency.

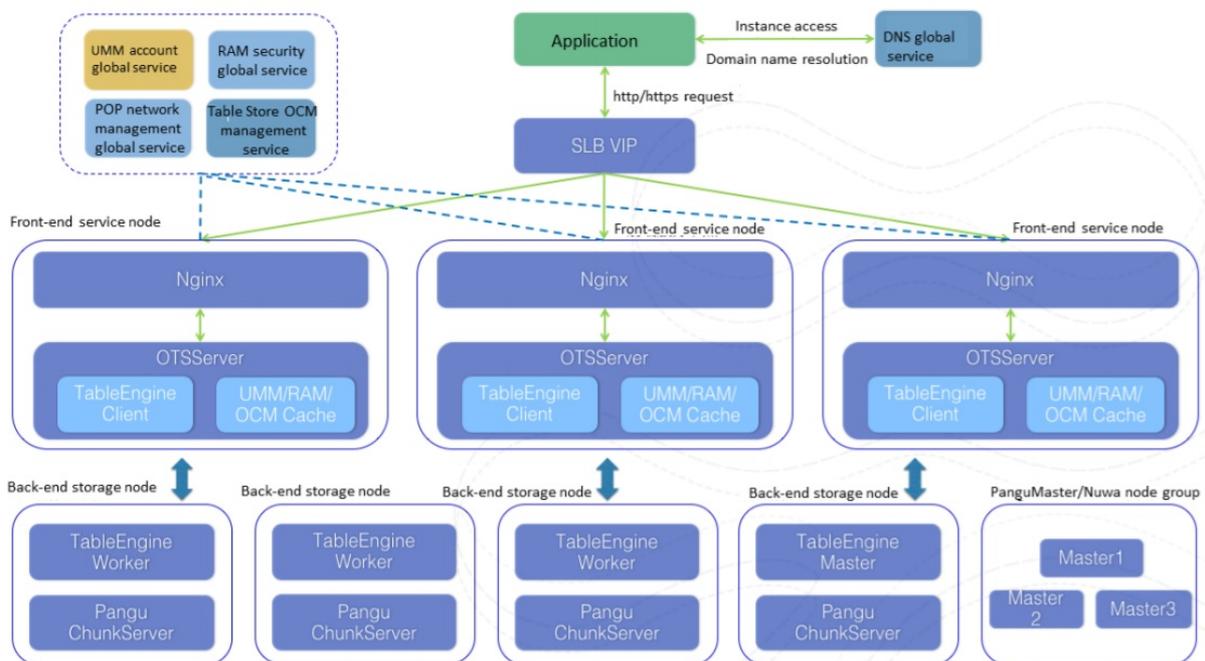
6.3. Architecture

The architecture of Table Store is referenced from Bigtable (one of the three core technologies of Google) and uses the log-structured merge-tree (LSM) storage engine to provide high performance writes. The performance of primary key-based single-row queries and range queries is stable and predictable. The performance is not affected by the volume of data and access concurrency.

The following figure shows the basic architecture of Table Store.



The following figure shows the detailed architecture of Table Store.



- The top layer is the protocol access layer. SLB distributes user requests to various proxy nodes. The proxy nodes receive requests that are sent through the RESTful protocol and implement security authentication. If the authentication succeeds, the user requests are forwarded to the corresponding data engine based on the value of the first primary key column for further operations. If the authentication fails, an error message is directly returned to the user.
- Table Worker is the data engine layer that processes structured data. It uses a primary key to

search for or store data. Table Worker supports large-scale access request bursts.

- The bottom layer is the storage layer. Apsara Distributed File System is deployed at this layer. Metadata is stored in Master server roles. The distributed message consistency protocol Paxos is adopted between Master service roles to ensure metadata consistency. In this scenario, efficient distributed file storage and access are achieved. This method guarantees three copies of data stored in the system and system recovery from any hardware or software faults.

6.4. Features

- **Data partition and load balancing**

The first column of a primary key in each row of a table is called the partition key. The system partitions a table into multiple partitions based on the range of the partition key. These partitions are evenly scheduled to different storage nodes. When the data in a partition exceeds a certain size, the partition is automatically split into two smaller partitions. The data and access loads are distributed to these two partitions. The partitions are scheduled to different nodes. As a result, access loads are scattered to different nodes. Eventually, the linear scalability of the single-table data scale and access loads is achieved. A partition is a logical organization of data based on the shared storage mechanism. No migration of physical data is involved when a partition is split. The theoretical impact of load balancing on the partition is that the partition fails to provide services within 100 milliseconds.

- **Automatic recovery after a single node failure**

In the storage engine of Table Store, each node serves a number of data partitions in different tables. The Master service role manages partition distribution and scheduling, and monitors the health of each service node. If a service node fails, the Master service role migrates data partitions from this faulty node to other healthy nodes. The migration is logically performed, and does not involve physical entities, so services can recover from the single point of failure (SPOF) within several minutes.

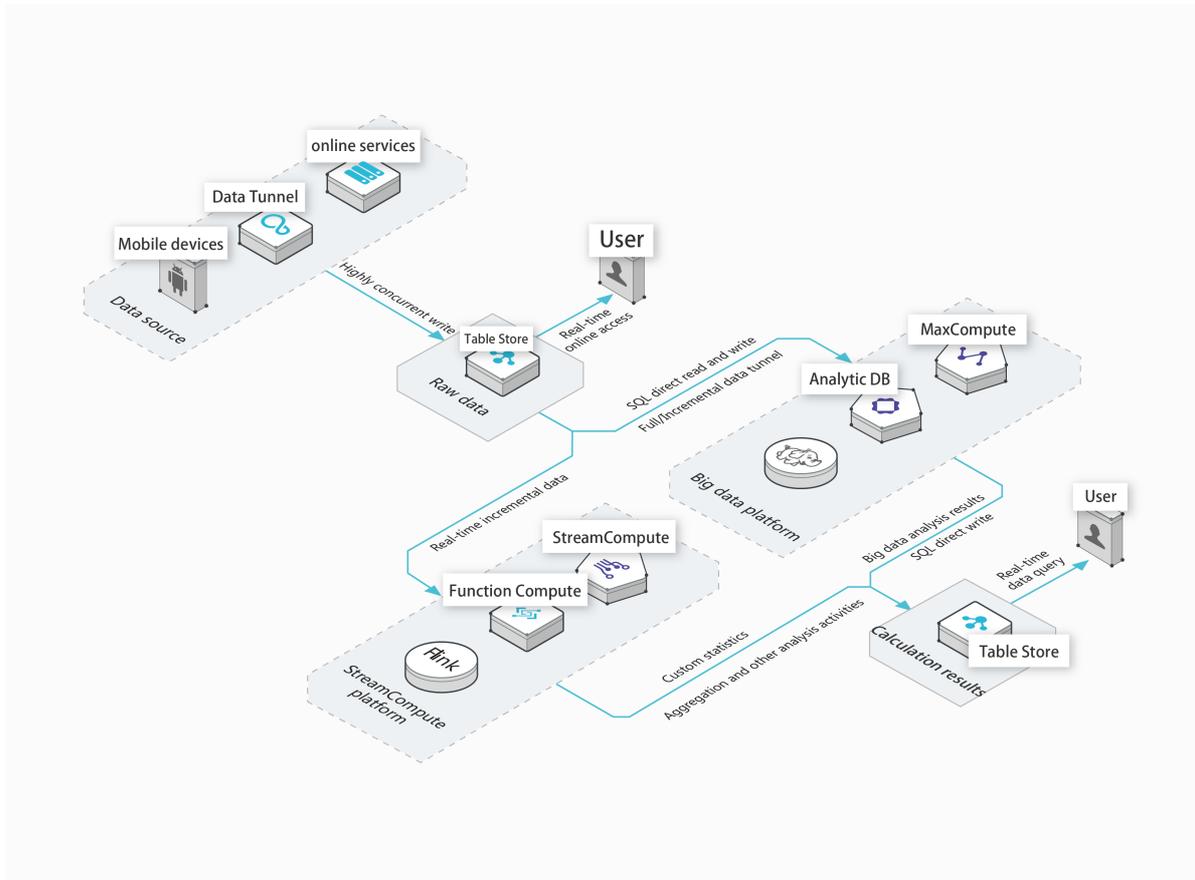
6.5. Scenarios

- **Scenario 1: Big data storage and analytics**

Table Store provides cost-friendly, highly-concurrent, short latency storage, and online access to large amounts of data. It provides full and incremental data tunnels and supports direct SQL read and write operations on various big data analytics platforms such as MaxCompute. An efficient incremental streaming read operation is provided for easy computing of real-time data streams.

In this scenario, Table Store provides the following features:

- Table Store supports various big data computing platforms, stream computing services, and real-time computing services that are provided by Alibaba Cloud.
- Table Store provides instances with high performance and high capacity to meet the requirements of different businesses.

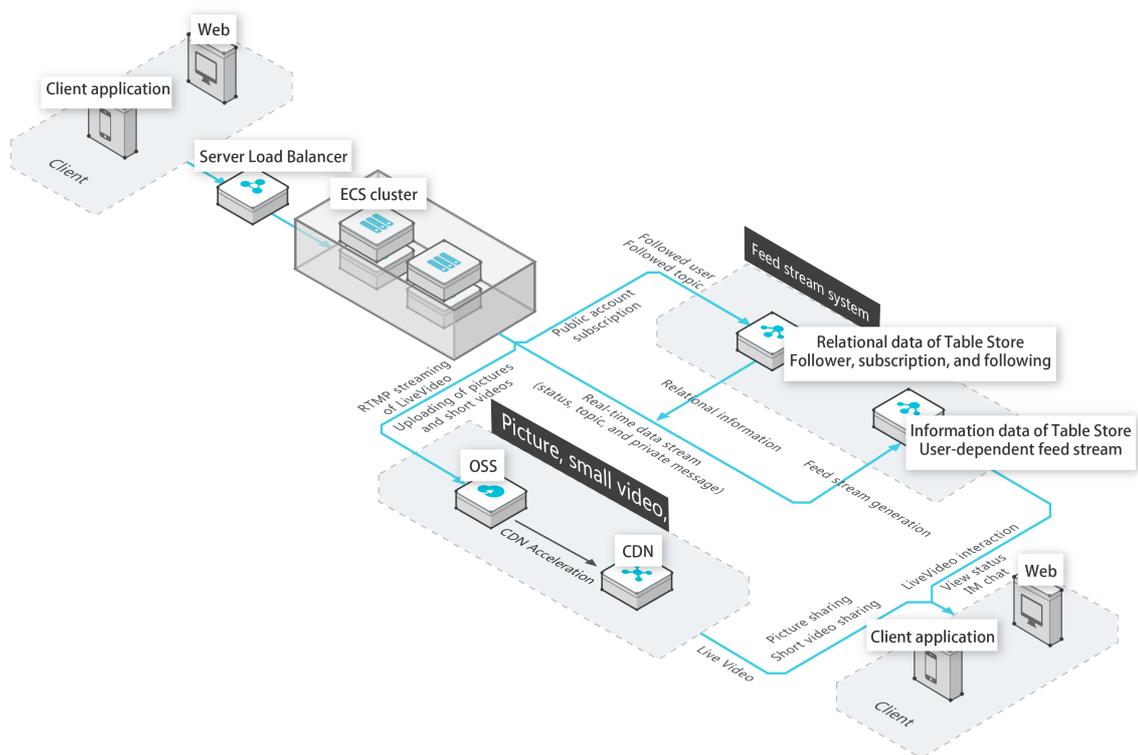


• Scenario 2: Social media feeds on the Internet

You can use Table Store to store large amounts of instant messaging (IM) messages, and social media feed information such as comments, posts, and likes. The elastic resources available for Table Store can meet application requirements at relatively low costs, such as significant traffic fluctuations, high concurrency, and short latency.

In this scenario, Table Store provides the following features:

- Built-in auto-increment primary key columns simplify external system dependencies.
- Average read and write performance of high-performance instances is not affected by volumes.
- High availability storage of large amounts of messages, and multi-terminal message synchronization are supported.

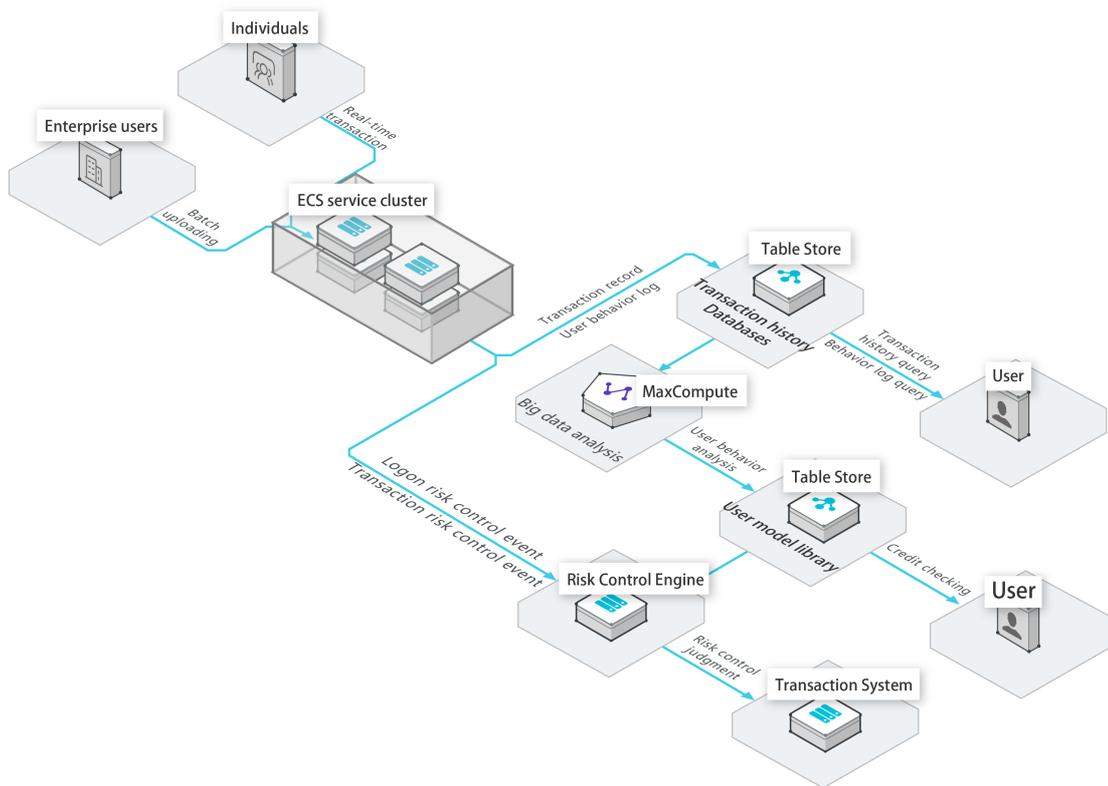


- **Scenario 3: Storage and real-time queries of large amounts of transaction records and user models**

Elastic resources can meet access requirements of short latency and high concurrency, allowing your risk control system to always operate in optimal conditions. You can strictly control transaction risks. Furthermore, the flexible data structure allows your business model to rapidly evolve to meet market demands.

In this scenario, Table Store provides the following features:

- A table can contain 1 trillion records and easily store full historical transaction records.
- Three copies are used to ensure high consistency and data security.
- Rapid service development is made available by a schema-free model and attribute columns that can be added as required.

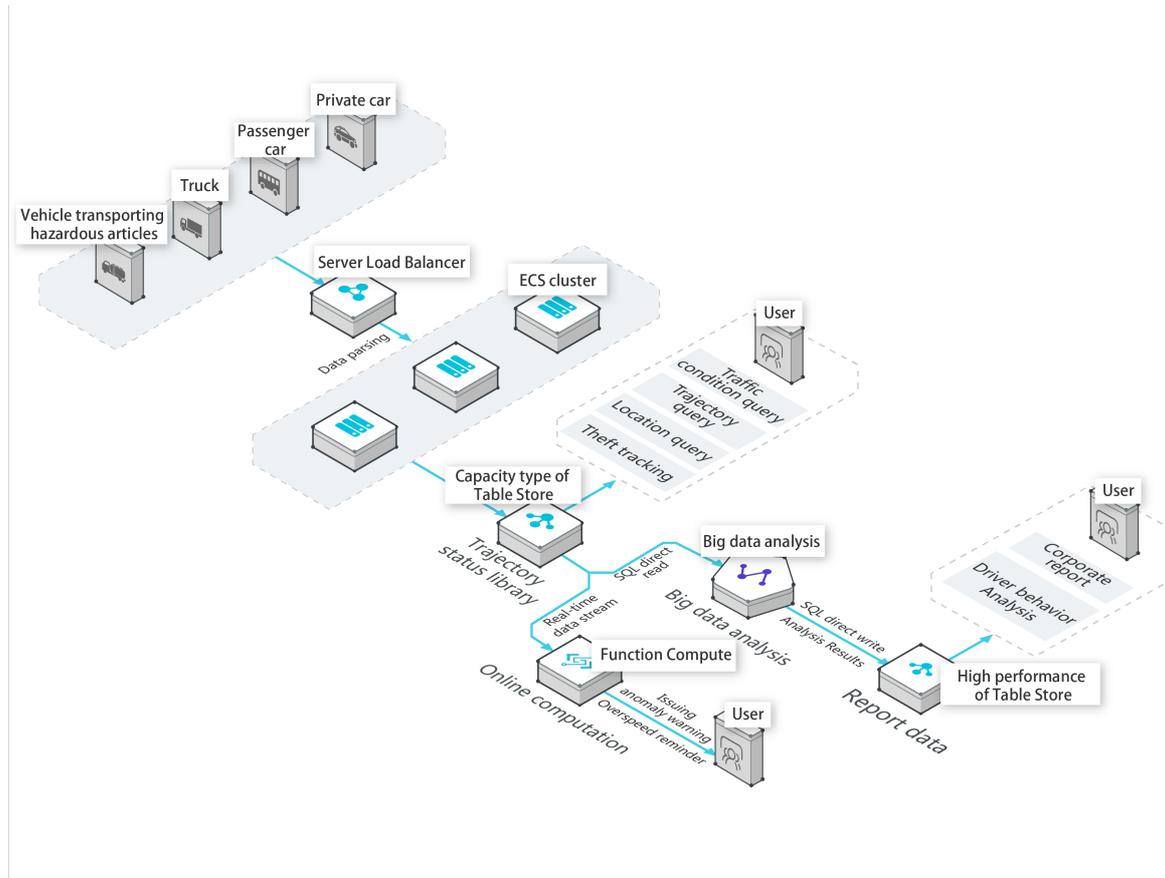


- Scenario 4: Efficient and flexible storage of large amounts of IoV data

The schema-free data model enables easy access to the data collected from different vehicle-mounted devices. Table Store can be seamlessly integrated with multiple big data analytics platforms and real-time computing services for ease of real-time online queries and business report analysis.

In this scenario, Table Store provides the following features:

- The query performance for vehicle conditions and routes is stable and predictable.
- The schema-free model allows you to easily store data collected from different vehicle-mounted devices.

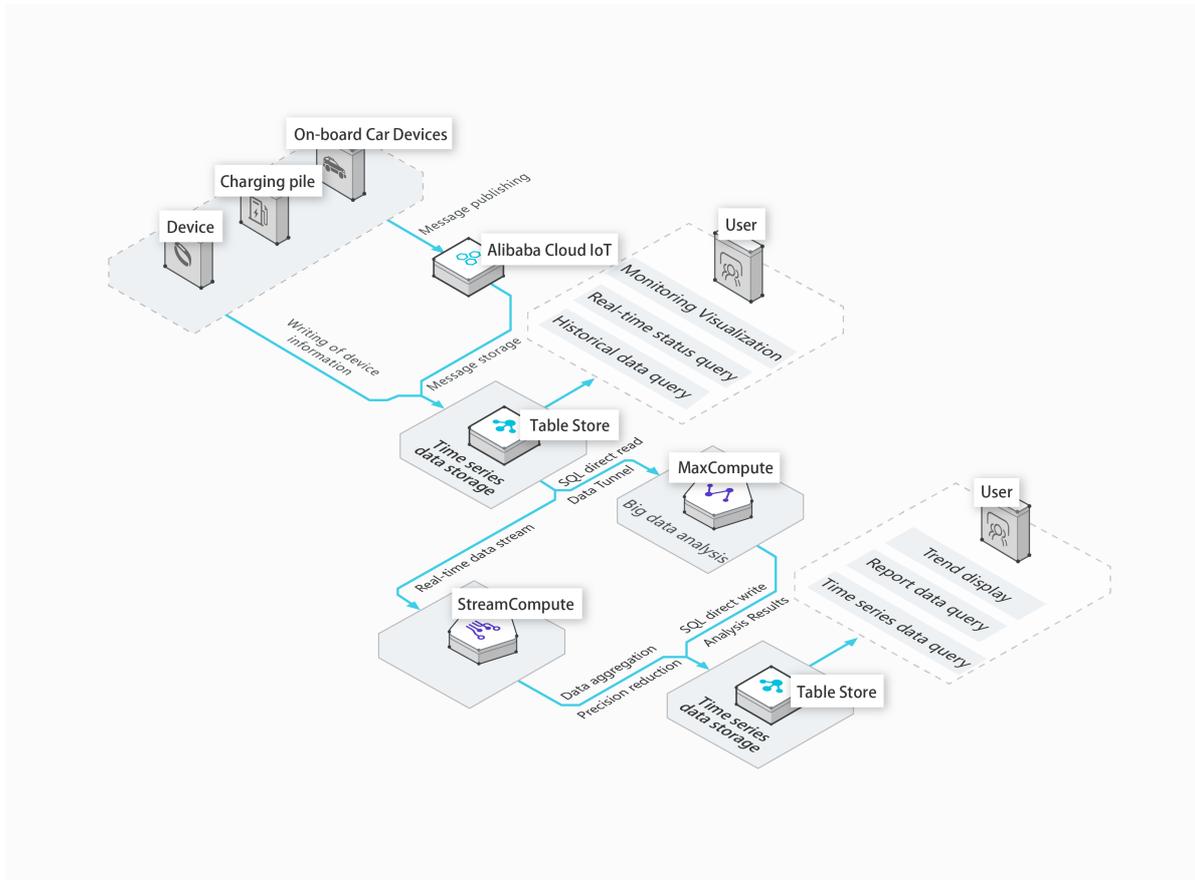


- **Scenario 5: Storage of large amounts of IoT data and efficient queries and analysis**

Table Store can easily store time series data from IoT devices and monitoring systems. The direct SQL read for big data analytics and the efficient incremental streaming read API allow easy offline data analysis and real-time stream computing.

In this scenario, Table Store provides the following features:

- Table Store can meet the data write and storage requirements of ultra-large-scale IoT devices and monitoring systems.
- Table Store can integrate with a variety of offline or stream data analysis platforms. This allows you to use a single piece of data for multiple analysis and computing operations.
- Table Store supports data lifecycle management.

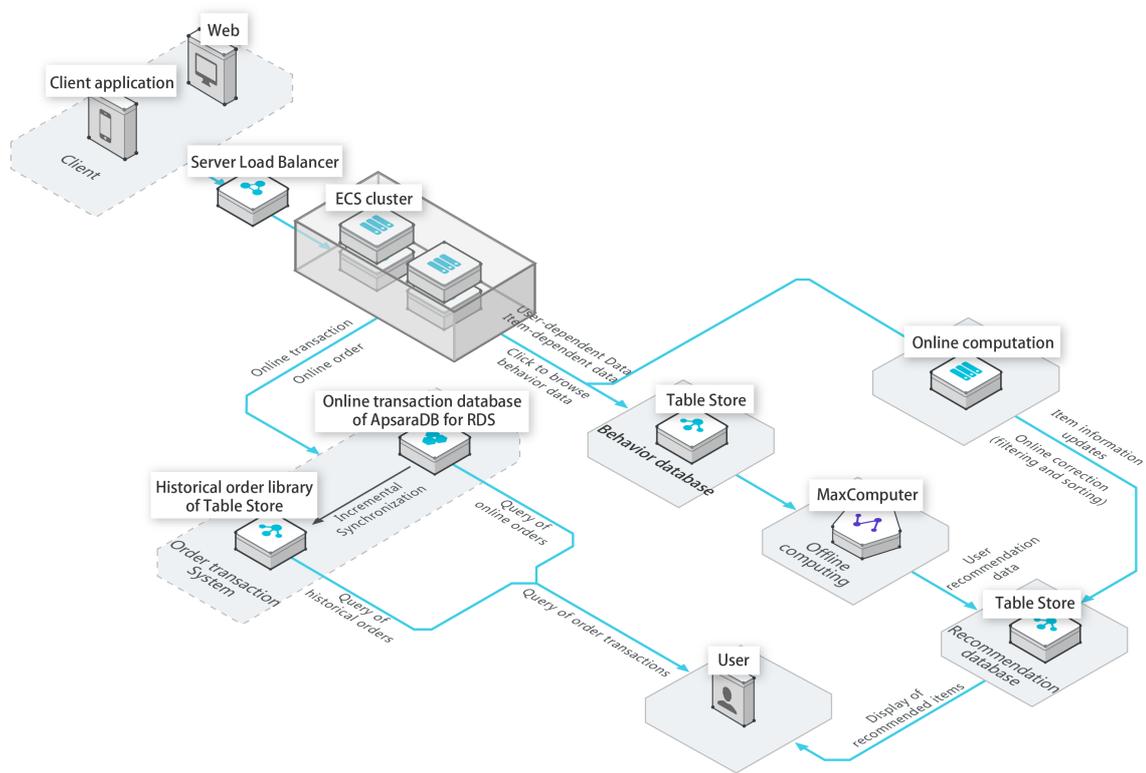


- **Scenario 6: Large-scale e-commerce transaction orders and user-specific database recommendations**

Table Store can easily manage large amounts of historical transaction data and improve access performance. Combined with MaxCompute, Table Store enables precision marketing and elastic resource storage. This scenario allows you to handle service requests during peak hours when all users go online.

In this scenario, Table Store provides the following features:

- Elastic scaling of resources based on data volumes and access concurrency meets the requirements of scenarios with access fluctuations during various periods.
- Various big data analytics platforms are supported for direct analysis of user behavior.
- The query latency of large amounts of transaction data is reduced to milliseconds.



6.6. Limits

Limits lists the limits for Table Store. Some limit ranges indicate the maximum values that can be used rather than the suggested values. Table structures and row data sizes can be tailored to improve performance.

Limits

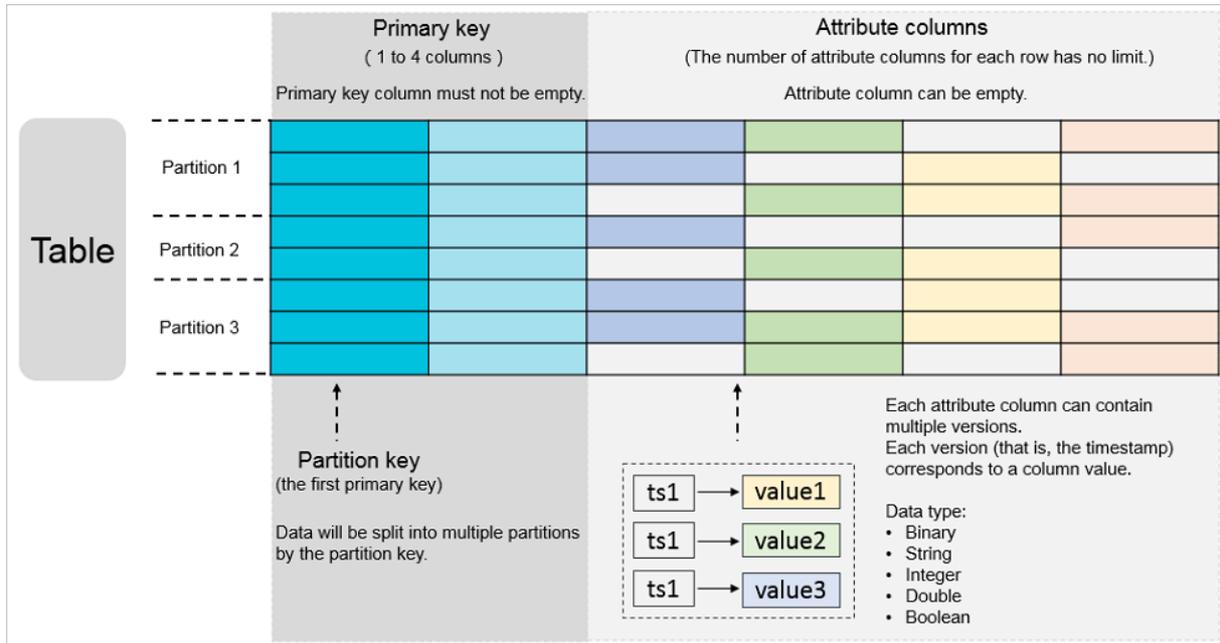
Limit	Limit range	Description
The number of instances created under an Apsara Stack tenant account	1024	If you need to increase the maximum number of instances, contact an administrator.
The number of tables in an instance	1024	If you need to increase the maximum number of tables, contact an administrator.
The number of columns of a primary key	1-4	A primary key can contain one to four columns.
The size of the value in a string type primary key column	1 KB	The size of the value in a string type primary key column cannot exceed 1 KB.
The size of the value in a string type attribute column	2 MB	The size of the value in a string type attribute column cannot exceed 2 MB.

Limit	Limit range	Description
The size of the value in a binary type primary key column	1 KB	The size of the value in a binary type primary key column cannot exceed 1 KB.
The size of the value in a binary type attribute column	2 MB	The size of the value in a binary type attribute column cannot exceed 2 MB.
The number of attribute columns in a single row	Unlimited	A single row can contain an unlimited amount of attribute columns.
The number of attribute columns written by one request	1024	The number of attribute columns written by a PutRow, UpdateRow, or BatchWriteRow request to a row cannot exceed 1,024.
The size of a single row	Unlimited	The total size of all column names and column value data for a row is unlimited.
The number of columns involved in columns_to_get in a read request	0-128	The maximum number of columns obtained from a row of data in a read request cannot exceed 128.
The number of UpdateTable operations for a table	Upper limit: unlimited Lower limit: unlimited	The frequency of UpdateTable operations for a table is limited.
UpdateTable frequency for a table	Once every 2 minutes	The reserved read/write throughput for a table can be adjusted once every two minutes at most.
The number of rows read by one BatchGetRow request	100	N/A
The number of rows written by one BatchWriteRow request	200	N/A
The size of data written by one BatchWriteRow request	4 MB	N/A
Data returned by one GetRange request	5,000 rows or 4 MB	The data returned by a request cannot exceed 5,000 rows or 4 MB. When either of the limits is exceeded, data that exceeds the limits is truncated at the row-level. The data primary key information in the next row is returned.
The data size of an HTTP request body	5 MB	N/A

6.7. Terms

data model

A model that consists of tables, rows, primary keys, and attributes, as shown in the following figure.



max versions

A data table attribute that indicates the maximum number of data versions that can be stored in each attribute column of a data table. If the number of versions in an attribute column exceeds the max versions value, the oldest version is deleted. This operation is performed asynchronously.

time To live (TTL)

A data table attribute measured in seconds. It indicates the validity period of data. To save space and reduce costs for data storage, the Table Store automatically clears any data that exceeds its TTL.

max version offset

A data table attribute that describes the maximum allowable difference between the version to be written and the current time, in seconds.

To prevent the writing of unexpected data, a server checks the versions of attribute columns when processing writing requests. Writing data to a row fails if the row has an attribute column in which: its version is earlier than the current writing time minus the max version offset value, or its version is later than or equal to the current writing time plus the max version offset value.

The valid version range for an attribute column is calculated based on the formula: Valid version range = [Data write time - Max version offset value, Data write time + Max version offset value). Data write time is the number of seconds that have elapsed since 1970-01-01 00:00:00 UTC. The versions of the attribute columns are stored in milliseconds, and must fall within the valid version range after it is converted to seconds (divide by 1,000).

primary key and attribute

A primary key is the unique identifier of each row in a table. It can consist of one to four primary key columns. When you create a table, you must define a primary key. Specifically, you must specify the name, data type, and sequence of each primary key column. Data types of values in primary key columns include String, Integer, and Binary. For a primary key column of the String or Binary type, the size of the column value must be smaller than 1 KB.

An attribute is a Table Store attribute that stores data in a row. You can create an unlimited number of attribute columns for each row.

read/write throughput

A Table Store attribute that is measured by read/write capacity units (CUs).

region

An Apsara Stack physical data center. Table Store is deployed across multiple Apsara Stack regions. Select a region that suits your business requirements.

instance

A logical entity in Table Store. It is used to manage tables, which are equivalent to databases in traditional relational databases. An instance is the basic unit of the Table Store resource management system. Table Store allows you to control access and meter resources at the instance level.

endpoint

The connection URL (also known as an endpoint) for each instance. You need to specify the endpoint before you perform any operations on Table Store tables and data.

stream

A data table attribute used for real-time analysis of incremental data streams and incremental data synchronization.

Serial ATA (SATA)

A disk that is based on serial connections and provides stronger error-correcting capabilities. It aims to improve the reliability of data during transmission.

7. Network Attached Storage (NAS)

7.1. What is NAS?

Alibaba Cloud NAS provides file storage services to compute nodes, such as ECS instances and Container Service nodes. With standard file system access protocols, NAS enables you to use a distributed file system with a variety of features. These features include unlimited capacity, expandable performance, unique namespace, parallel shared access, high reliability, and high availability.

After creating a NAS file system and mount point, you can mount the file system on compute nodes, such as ECS instances and Container Service nodes, through standard NFS protocols. You can access the file system through POSIX-based interfaces. A file system can be mounted on multiple compute nodes to share files and folders.

7.2. Benefits

Parallel shared access

A file system can be simultaneously mounted on multiple compute nodes to share access. It reduces a large number of data replication and synchronization costs.

High reliability

Alibaba Cloud NAS provides reliable data storage. Compared with user-created NAS file systems, you can reduce a large number of maintenance costs and avoid data security risks if you use Alibaba Cloud NAS.

Elastic scalability

A single file system has a maximum capacity of 10 PB. At any time (or on demand), this file system can be scaled up to quickly adapt to business changes.

High performance

The throughput of a single NAS file system increases with growing capacity. Alibaba Cloud NAS helps you to reduce a large number of costs by removing the upfront investment in high-end NAS storage devices.

Easy-to-use

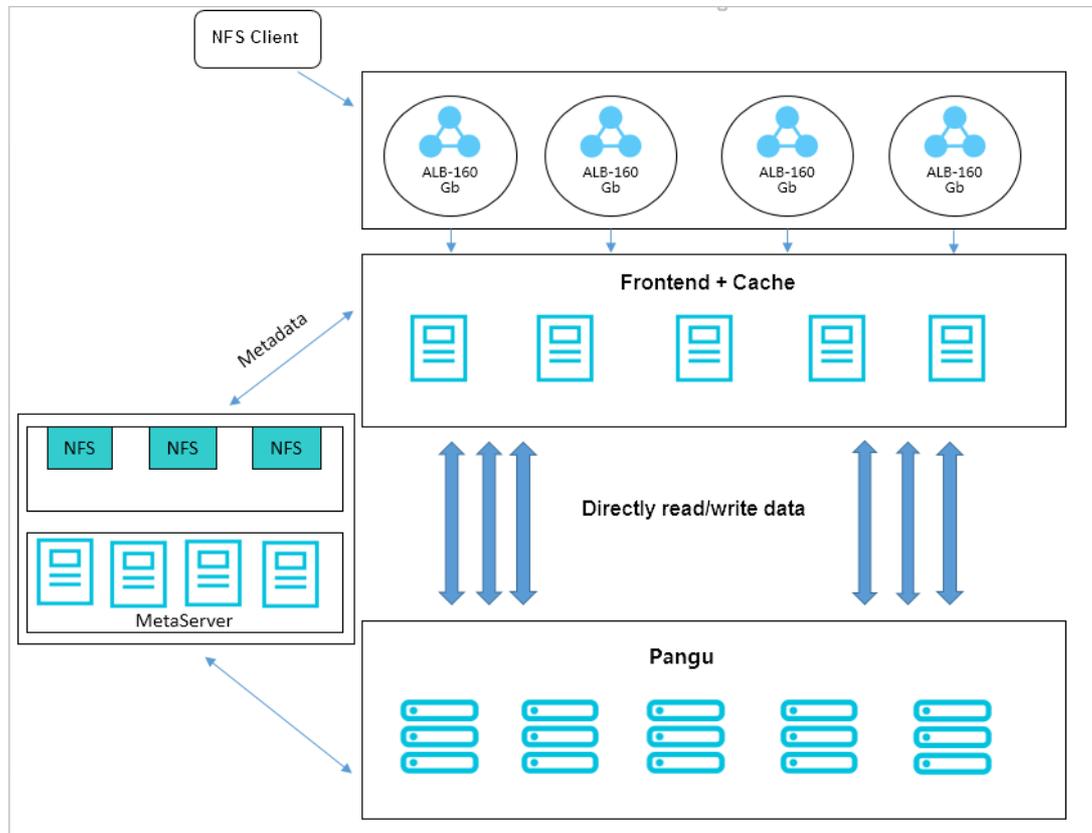
Alibaba Cloud NAS supports the NFSv3 and NFSv4 protocols. You can access file systems through standard POSIX-based interfaces, regardless of the types of compute nodes and where file systems are located.

7.3. Architecture

NAS is based on Apsara Distributed File System. It maintains three copies of each data across multiple storage nodes. The frontend nodes receive connection requests from NFS clients and provide the cache function. These nodes are stateless and distributed to ensure high availability of the frontend. The metadata of NAS instances is stored in MetaServer. When the frontend nodes receive I/O requests, they access the MetaServer to obtain the metadata. Then, the frontend nodes access the backend data nodes for user data.

Both the frontend and backend can expand elastically, ensuring high availability, high throughput, and low latency.

System architecture



7.4. Features

Seamless integration

Network Attached Storage (NAS) supports the NFSv3 and NFSv4 protocols and provides access through standard file system interfaces. Most applications and workloads can seamlessly work with NAS without any change.

Shared access

A NAS file system can be accessed by multiple compute nodes. NAS supports simultaneous access from multiple compute nodes. Therefore, a NAS file system is well suited if your application is deployed across multiple ECS instances or Container Service nodes.

Access control

NAS provides multiple security control mechanisms to ensure data security of its file systems. These mechanisms include but are not limited to: network isolation by VPCs, user isolation in classic networks, standard permission control for file systems, security group based access control, and RAM user authorization.

Linear performance

NAS allows your applications to achieve optimal storage performance of high throughput and IOPS with consistent low latency. The storage performance linearly improves as the storage capacity increases. This meets the high requirements imposed by business growth on both storage capacity and storage performance.

7.5. Scenarios

Scenario 1: shared storage and high availability for SLB

Your SLB instance is connected to multiple ECS instances. You can store the data of the applications on these ECS instances on a shared NAS instance. This implements data sharing and ensures high availability of the SLB servers.

Scenario 2: file sharing within an enterprise

The employees of an enterprise need to access the same datasets for work purposes. The administrator can create a NAS instance and configure different file or directory permissions for users or user groups.

Scenario 3: data backup

You want to back up the data stored in the data center to the cloud and use a standard interface to access the cloud storage service. You can back up the data in the data center to a NAS instance.

Scenario 4: server log sharing

You want to store the application server logs of multiple compute nodes on the shared file storage. You can store these server logs on a NAS instance for centralized log processing and analysis.

7.6. Limits

- NAS supports the NFSv3 and NFSv4 protocols.
- The following table lists the attributes that are not supported by NFSv4.0 and NFSv4.1, and their client errors.

Protocol	Unsupported attribute	Client error
NFSv4.0	FATTR4_MIMETYPE, FATTR4_QUOTA_AVAIL_HARD, FATTR4_QUOTA_AVAIL_SOFT, FATTR4_QUOTA_USED, FATTR4_TIME_BACKUP, and FATTR4_TIME_CREATE	NFS4ERR_ATTRNOTSUPP

Protocol	Unsupported attribute	Client error
NFSv4.1	FATTR4_DIR_NOTIF_DELAY, FATTR4_DIRENT_NOTIF_DELAY, FATTR4_DACL, FATTR4_SACL, FATTR4_CHANGE_POLICY, FATTR4_FS_STATUS, FATTR4_LAYOUT_HINT, FATTR4_LAYOUT_TYPES, FATTR4_LAYOUT_ALIGNMENT, FATTR4_FS_LOCATIONS_INFO, FATTR4_MDSTHRESHOLD, FATTR4_RETENTION_GET, FATTR4_RETENTION_SET, FATTR4_RETENT_EVT_GET, FATTR4_RETENT_EVT_SET, FATTR4_RETENTION_HOLD, FATTR4_MODE_SET_MASKED, and FATTR4_FS_CHARSET_CAP	NFS4ERR_ATTRNOTSUPP

- NFSv4 does not support the following OPs: OP_DELEGPURGE, OP_DELEGRETURN, and NFS4_OP_OPENATTR. The client displays an NFS4ERR_NOTSUPP error.
- NFSv4 does not support Delegation.
- About UID and GID:
 - For NFSv3, if the file UID or GID exists in a Linux local account, the corresponding username and group name is displayed based on the mapping between the local UID and GID. If the file UID or GID does not exist in the local account, the UID and GID is displayed.
 - For NFSv4, if the version of the local Linux kernel is earlier than 3.0, the UIDs and GIDs of all files are displayed as "nobody." If the kernel version is later than 3.0, the display rule is the same as that of NFSv3.

 **Notice** If you use NFSv4 to mount a NAS instance and the Linux kernel version is earlier than 3.0, we recommend that you do not change the owner or group of local files or directories. Such changes can cause the UIDs and GIDs of the files or directories to become "nobody."

- You can mount a NAS instance to up to 10,000 compute nodes.

7.7. Terms

mount point

A mount point is the access address of a NAS instance in a VPC or classic network. Each mount point corresponds to a domain name. To mount a NAS instance to a local directory, you must specify the domain name of the mount point.

permission group

The permission group is a whitelist mechanism provided by NAS. You can add rules to a permission group of a NAS instance to allow users from specified IP addresses or address segments to access the NAS instance with different permissions.

 **Note** Each mount point must be bound with a permission group.

authorized object

An authorized object is an attribute of the permission group rule. It specifies the IP address or address segment to which the permission group rule is applied. In a VPC, an authorized object can be a single IP address or an address segment. In a classic network, an authorized object must be an IP address, generally the intranet IP address of an ECS instance.

8. Apsara File Storage for HDFS

8.1. What is Apsara File Storage for HDFS?

Apsara File Storage for HDFS is a file storage service for computing resources such as Alibaba Cloud ECS instances and Container Service. It supports standard HDFS access protocols. You can use Apsara File Storage for HDFS without modifying existing big data applications. Apsara File Storage for HDFS offers various features such as unlimited capacity, performance expansion, single namespace, multi-party sharing, high reliability, and high availability.

Apsara File Storage for HDFS is applicable to the Internet, finance, and other businesses that require big data computing capabilities and are required to store large amounts of data and perform offline computation. It fully meets the needs of distributed computing business models represented by Hadoop in multiple aspects, such as distributed storage performance, capacity, and reliability.

After creating an Apsara File Storage for HDFS instance, you can access the file system through standard HDFS protocol interfaces in computing resources, such as ECS and Container Service instances. In addition, multiple compute nodes can simultaneously access the same Apsara File Storage for HDFS to share files and directories.

8.2. Benefits

This topic describes the benefits of Apsara File Storage for HDFS in terms of data reliability and ease of use.

High reliability

Data is stored in three copies to improve reliability. Compared with user-created HDFS, Apsara File Storage for HDFS minimizes O&M costs and data security risks.

Scalability

The capacity of a single file system is unlimited, which allows services to scale-up and scale-down at any time.

High performance

Throughput performance is optimized for small files. Compared with user-created HDFS, Apsara File Storage for HDFS significantly improves the throughput performance of small files.

Multi-tenancy

Multiple Apsara File Storage for HDFS instances can be created in the storage system for centralized permission and capacity management.

Ease of use

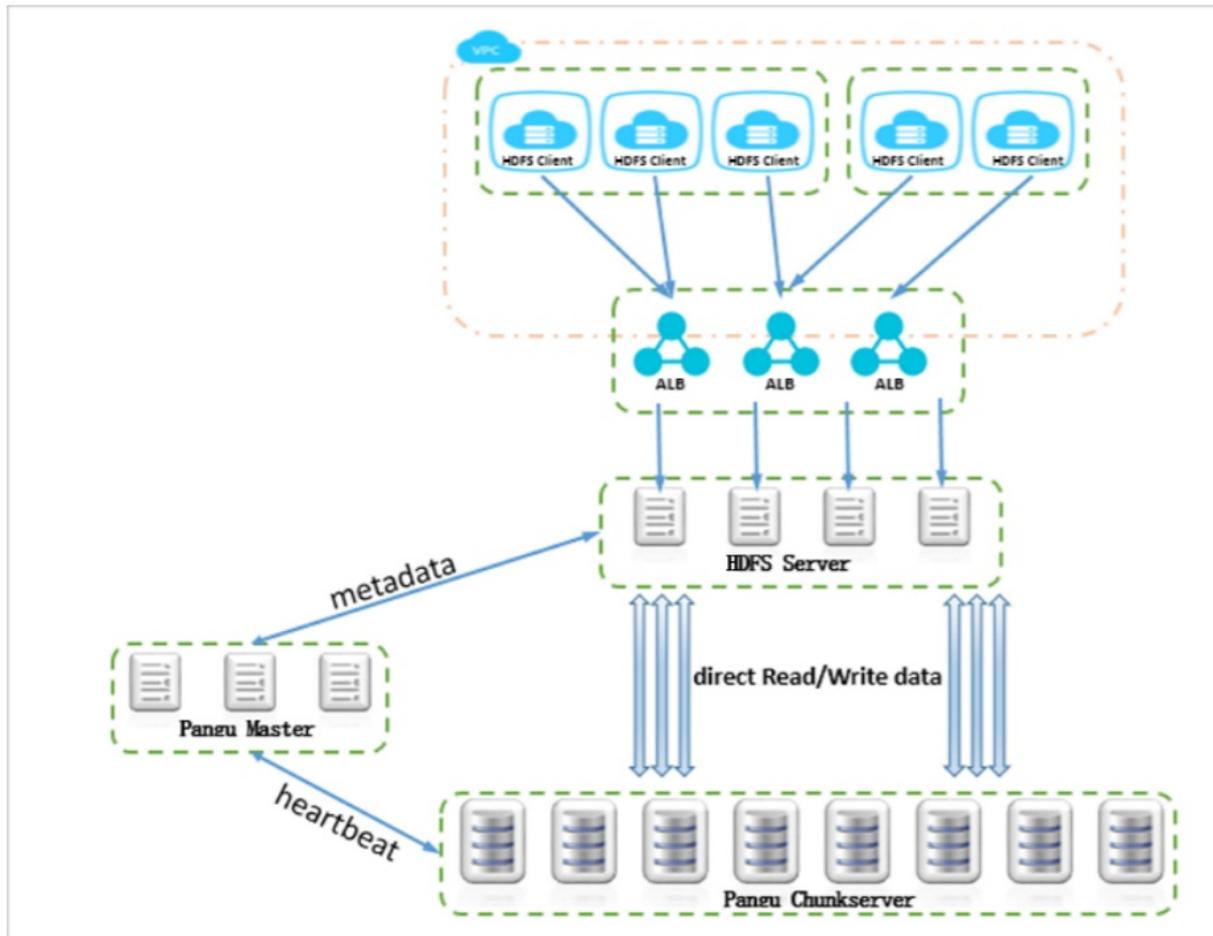
Apsara File Storage for HDFS provides automated O&M functions, which reduces O&M workloads and human errors. It also ensures the security of the system.

8.3. Architecture

The architecture of Apsara File Storage for HDFS is divided into two parts: front end and back end.

The back end is based on Apsara Distributed File System. Data is stored in Apsara Distributed File System as multiple copies. The front-end access nodes of Apsara File Storage for HDFS receive connection requests from ECS (for example, MapReduce, Spark, and other Hadoop computing applications) or Container Service instances, and cache data. Apsara Distributed File System also manages metadata and data of Apsara File Storage for HDFS.

The overall architecture of Apsara File Storage for HDFS is as follows.



8.4. Features

This topic describes features of Apsara File Storage for HDFS such as shared access and security control.

Seamless integration

Apsara File Storage for HDFS supports Hadoop 2.7.X protocols and accesses data through standard HDFS syntax. Mainstream Hadoop applications and workloads can be integrated seamlessly with Apsara File Storage for HDFS.

Shared access

A single Apsara File Storage for HDFS instance can be accessed by multiple compute nodes simultaneously. Apsara File Storage for HDFS is suitable for scenarios where applications deployed across multiple ECS or Container Service instances access the same data source.

Security control

Multiple security mechanisms are implemented to secure data. These security mechanisms include network isolation (such as the use of VPCs), standard permission control of file systems, permission groups, and RAM.

Linear scalability

Apsara File Storage for HDFS stores application workloads with high throughput, high IOPS, and low latency. Additionally, the linear relationship between performance and capacity is sufficient to overcome business requirements for capacity and storage performance when the business volume increases.

8.5. Scenarios

This topic describes the usage scenarios of Apsara File Storage for HDFS.

Scenario 1: shared storage and high availability

Apsara File Storage for HDFS supports standard HDFS protocols. You can store files in real time or in batches to Apsara File Storage for HDFS by using standard HDFS interfaces.

If you want to share files or have higher availability requirements on files, we recommend that you use Apsara File Storage for HDFS to store files.

Scenario 2: big data analytics and machine learning

In big data analytics and machine learning scenarios, applications require high throughput performance and short latency for data access. Apsara File Storage for HDFS supports high-throughput and low-latency access. You do not need to migrate data to local computing resources. Therefore, Apsara File Storage for HDFS is recommended in this scenario.

After data is stored in Apsara File Storage for HDFS, ECS instances or other computing resources can directly access the data. You can deploy Hadoop or other machine learning applications on multiple computing resources so that applications can access data directly through the HDFS interfaces to perform online or offline computation. You can also export the calculation results to an Apsara File Storage for HDFS instance and store them permanently.

8.6. Limits

This topic describes the service limits of Apsara File Storage for HDFS.

Hadoop FileSystem or AbstractFileSystem

- Does not support the setting of directory modification time (mtime) and access time (atime), or the setting of file mtime and atime through setTimes.
- Does not support symbolic links.
- Does not support file truncation (truncate).
- Does not support file concatenation (concat).
- Does not support extended attributes (XAttrs).
- Does not support snapshot operations.
- Does not support delegation token operations.
- Does not support checksum operations such as setWriteChecksum and setVerifyChecksum.
- Does not support ACL operations.
- Does not support file block locations.

Hadoop fs command line tool

- Does not support snapshot commands such as createSnapshot, deleteSnapshot, or renamesnapshot.
- Does not support ACL commands such as setfacl or getfacl.
- Does not support XAttr commands such as setfattr or getfattr.
- Does not support file truncation commands such as truncate.

8.7. Terms

This topic introduces several terms used in Apsara File Storage for HDFS, so that you can better understand Apsara File Storage for HDFS.

mount point

The access address of the file system in a VPC or classic network. Each mount point is mapped to a domain name. You need to modify the *core-site.xml* configuration to access files in an Apsara File Storage for HDFS instance.

permission group

An access control mechanism provided by Apsara File Storage for HDFS. You can add rules to a permission group to allow access from different IP addresses or segments to access an Apsara File Storage for HDFS instance based on different permissions.

 **Note** Each mount point must have a specified permission group.

authorized object

An attribute of a permission group rule that specifies the IP address or segment to which the permission group rule is applied. In a VPC, an authorized object can be a single IP address or an IP address segment. In a classic network, an authorized object can only be a single IP address (generally the internal IP address of an ECS instance).

9. ApsaraDB for RDS

9.1. What is ApsaraDB for RDS?

Apsara Stack ApsaraDB for Relational Database Service (RDS) is a stable, reliable, and automatically scaling online database service.

Based on the distributed file system and high-performance storage, ApsaraDB for RDS allows you to easily perform database operations and maintenance with its complete set of solutions for disaster recovery, backup, restoration, monitoring, and migration.

ApsaraDB for RDS supports four storage engines: MySQL, PostgreSQL, and PPAS. These storage engines can help you create database instances suitable to your business needs.

ApsaraDB RDS for MySQL

Originally based on a branch of MySQL, ApsaraDB RDS for MySQL has proven its performance and throughput during the high-volume concurrent traffic of Double 11. ApsaraDB RDS for MySQL provides whitelist configuration, backup and restoration, transparent data encryption, data migration, and management for instances, accounts, and databases. It also provides the following advanced features:

- Read-only instance

In scenarios where RDS reads a small number of write requests but a large number of read requests, you can enable read/write splitting to distribute read requests away from the primary instance. Read-only instances allow ApsaraDB RDS for MySQL 5.6 to automatically scale the reading capability and increase the application throughput when a large amount of data is being read.

- Read/write splitting

The read/write splitting feature provides an extra read/write splitting address. This address enables an automatic link for the primary instance and all its read-only instances. An application can use this method to read and write data by connecting to the read/write splitting address. Write requests are automatically distributed to the primary instance while read requests are distributed to read-only instances based on their weights. To scale up the reading capacity of the system, you can add more read-only instances.

- Data compression

ApsaraDB RDS for MySQL 5.6 allows you to compress data by using the TokuDB storage engine. Data transferred from the InnoDB storage engine to the TokuDB storage engine can be reduced by 80% to 90% in volume. 2 TB of data in InnoDB can be compressed to 400 GB or less in TokuDB. In addition to data compression, TokuDB supports transaction and online DDL operations. TokuDB is compatible with MyISAM and InnoDB applications.

ApsaraDB RDS for PostgreSQL

ApsaraDB RDS for PostgreSQL is an advanced open source database system that is fully compatible with SQL and supports a diverse range of data formats such as JSON, IP, and geometric data. In addition to support for features such as transactions, subqueries, multi-version concurrency control (MVCC), and data integrity check, ApsaraDB RDS for PostgreSQL integrates a series of features including high availability, backup, and restoration to ease operations and maintenance loads.

ApsaraDB RDS for PostgreSQL provides basic features such as whitelist configuration, backup and restoration, data migration, and management for instances, accounts, and databases.

ApsaraDB RDS for PPAS

ApsaraDB RDS for Postgres Plus Advanced Server (PPAS) is a stable, secure, and scalable enterprise-class relational database. Based on ApsaraDB RDS for PostgreSQL, ApsaraDB RDS for PPAS features enhanced performance, application solutions, and compatibility. It is able to directly run Oracle applications. You can run enterprise-class applications on PPAS in a stable and cost-effective manner.

ApsaraDB RDS for PPAS provides basic features such as whitelist configuration, backup and restoration, data migration, and management for instances, accounts, and databases.

9.2. Benefits

9.2.1. Ease of use

ApsaraDB for RDS is a ready-to-use service featuring on-demand upgrades, convenient management, high transparency, and high compatibility.

Ready-to-use

You can use the API to create instances of any specified RDS instance type.

On-demand upgrade

When the database load or data storage capacity changes, you can upgrade the RDS instance by changing its type. The upgrades do not interrupt the data link service.

Transparency and compatibility

ApsaraDB for RDS is used in the same way as the native RDS database engine, allowing it to be adopted easily without the need to learn new database engines. ApsaraDB for RDS is compatible with existing programs and tools. Data can be migrated to ApsaraDB for RDS through ordinary import and export tools.

Easy management

Alibaba Cloud is responsible for the routine maintenance and management tasks for ApsaraDB for RDS such as troubleshooting hardware and software issues or issuing database patches and updates. You can also manually add, delete, restart, back up, and restore databases through the Apsara Stack console.

9.2.2. High performance

ApsaraDB for RDS implements parameter optimization, SQL optimization, and high-end back-end hardware to achieve high performance.

Parameter optimization

All RDS instance parameters have been optimized over their several years of production. Professional database administrators continue to optimize RDS instances over their lifecycles to ensure that ApsaraDB for RDS runs at peak efficiency.

SQL optimization

ApsaraDB for RDS locks inefficient SQL statements and provides recommendations to optimize code.

High-end back-end hardware

All servers used by ApsaraDB for RDS are evaluated by multiple parties to ensure stability.

9.2.3. High security

ApsaraDB for RDS implements anti-DDoS protection, access control, system security, and transparent data encryption (TDE) to guarantee the security of your databases.

DDoS attack prevention

 **Note** You must activate Alibaba Cloud security services to use this feature.

When you access an ApsaraDB for RDS instance from the Internet, the instance is vulnerable to DDoS attacks. When a DDoS attack is detected, the RDS security system first scrubs the inbound traffic. If traffic scrubbing is not sufficient or if the traffic exceeds a specified threshold, black hole filtering is triggered.

Access control

You can configure an IP address whitelist for ApsaraDB for RDS to allow access for specified IP addresses and deny access for all others.

Each account can only view and operate their own respective database.

System security

ApsaraDB for RDS is protected by several layers of firewalls capable of blocking a variety of attacks to secure data.

ApsaraDB for RDS servers cannot be logged onto directly. Only the ports required for specific database services are provided.

ApsaraDB for RDS servers cannot initiate an external connection. They can only receive access requests.

TDE

Transparent Data Encryption (TDE) can be used to perform real-time I/O encryption and decryption on instance data files. Data is encrypted before it is written to disks and decrypted before it is read from disks to the memory. TDE will not increase the size of data files. Developers do not need to modify their applications before using the TDE feature.

9.2.4. High reliability

ApsaraDB for RDS provides hot standby, multi-copy redundancy, data backup, and data recovery to achieve high reliability.

Hot standby

ApsaraDB for RDS adopts a hot standby architecture. If the primary server fails, services will fail over to the secondary server within seconds. Applications running on the servers are not affected by the failover process and will continue to run normally.

Multi-copy redundancy

ApsaraDB for RDS servers implement a RAID architecture to store data. Data backup files are stored on OSS.

Data backup

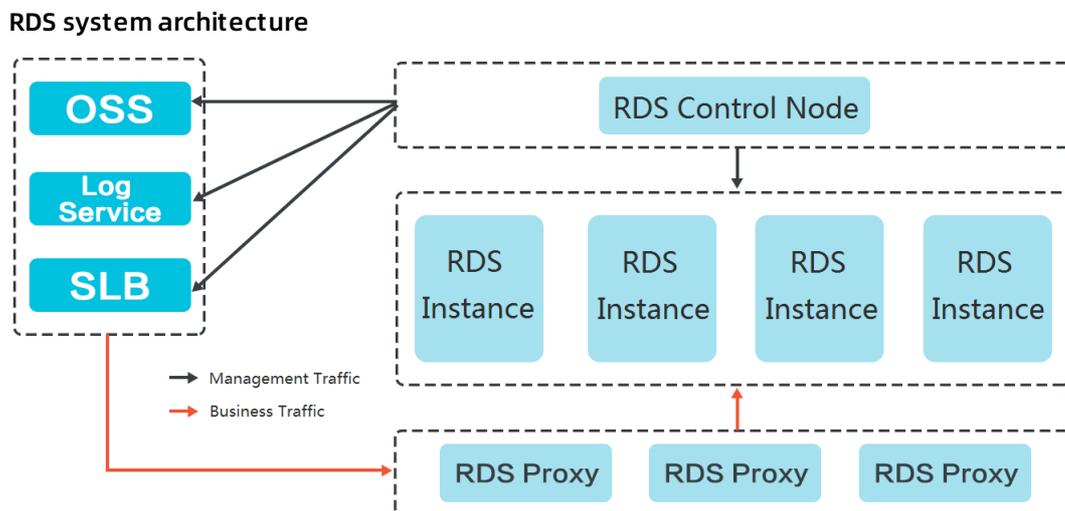
ApsaraDB for RDS provides an automatic backup mechanism. You can schedule backups to be performed periodically, or manually initiate temporary backups as necessary to meet your business needs.

Data recovery

Data can be restored from backup sets or cloned instances created at previous points in time. After data is verified, the data can be migrated back to the primary RDS instance.

9.3. Architecture

The following figure shows the system architecture of ApsaraDB for RDS.

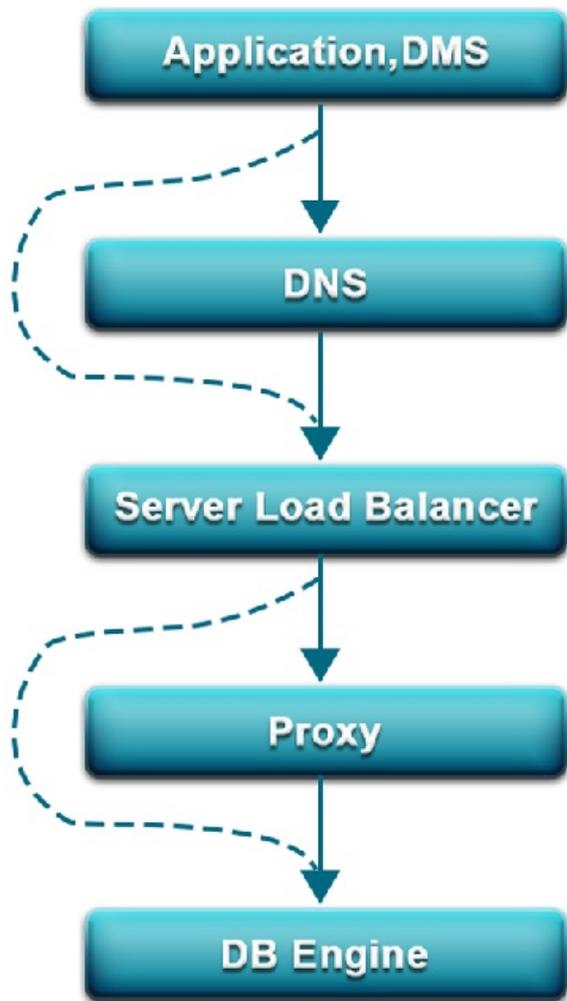


9.4. Features

9.4.1. Data link service

ApsaraDB for RDS provides all data link services, including DNS, Server Load Balancer (SLB), and Proxy.

ApsaraDB for RDS uses native database engines with similar database operations to minimize learning costs and facilitate database access.



DNS

The DNS module can dynamically resolve domain names to IP addresses. Therefore, IP address changes do not affect the performance of RDS instances. After the domain name of an RDS instance is configured in the connection pool, the RDS instance can be accessed even if its corresponding IP address changes.

For example, the domain name of an ApsaraDB for RDS instance is `test.rds.aliyun.com`, and its corresponding IP address is `10.10.10.1`. The instance can be accessed when either `test.rds.aliyun.com` or `10.10.10.1` is configured in the connection pool of a program.

After a zone migration or version upgrade is performed for this ApsaraDB for RDS instance, the IP address may change to `10.10.10.2`. If the domain name `test.rds.aliyun.com` is configured in the connection pool, the instance can still be accessed. However, if the IP address `10.10.10.1` is configured in the connection pool, the instance will no longer be accessible.

SLB

The SLB module provides both the internal IP address and public IP address of an ApsaraDB for RDS instance. Therefore, server changes do not affect the performance of the instance.

For example, the internal IP address of an RDS instance is 10.1.1.1, and the corresponding Proxy or DB Engine runs on 192.168.0.1. The SLB module typically redirects all traffic destined for 10.1.1.1 to 192.168.0.1. If 192.168.0.1 fails, another server in hot standby status with the IP address 192.168.0.2 will take over for the initial server. In this case, the SLB module will redirect all traffic destined for 10.1.1.1 to 192.168.0.2, and the RDS instance will continue to provide services normally.

Proxy

The Proxy module provides a number of features including data routing, traffic detection, and session persistence.

- **Data routing:** aggregates the distributed complex queries found in big data scenarios and provides the corresponding capacity management capabilities.
- **Traffic detection:** reduces SQL injection risks and supports SQL log backtracking when necessary.
- **Session persistence:** prevents database connection interruptions when faults occur.

DB Engine

The following table describes the mainstream database protocols supported by RDS.

RDS database protocols

RDBMS	Version
MySQL	5.6 or 5.7 (including read-only instances)
PostgreSQL	9.4
PPAS	9.3 or 9.6

9.4.2. High-availability service

The high-availability (HA) service consists of modules such as the Detection, Repair, and Notice.

The HA service guarantees the availability of data link services and processes internal database exceptions.

Detection

The Detection module checks whether the primary and secondary nodes of the DB Engine are providing services normally. The HA node uses heartbeat information taken at 8 to 10 second intervals to determine the health status of the primary node. This information, along with the health status of the secondary node and heartbeat information from other HA nodes, provides a reference for the Detection module. All this information helps the module avoid misjudgment caused by exceptions such as network jitter. Failover can be completed within 30 seconds.

Repair

The Repair module maintains the replication relationship between the primary and secondary nodes of the DB Engine. It can also correct errors that occur on either node during normal operations.

For example:

- It can automatically restore primary/secondary replication after a disconnection.
- It can automatically repair table-level damage to the primary or secondary node.
- It can save and automatically repair the primary or secondary node in case of crashes.

Notice

The Notice module informs the SLB or Proxy module of status changes to the primary and secondary nodes to ensure that you always access the correct node.

For example, imagine that the Detection module discovers problems with the primary node and instructs the Repair module to resolve these problems. If the Repair module fails to resolve a problem, it instructs the Notice module to perform traffic switchover. The Notice module forwards the switching request to the SLB or Proxy module, and then all traffic is redirected to the secondary node. Meanwhile, the Repair module creates a new secondary node on a different physical server and synchronizes this change back to the Detection module. The Detection module rechecks the health status of the instance.

HA policy

Each HA policy defines a combination of service priorities and data replication modes defined to meet the needs of your business.

There are two service priorities:

- **Recovery time objective (RTO):** The database preferentially restores services to maximize the availability time. Use the RTO policy if you require longer database uptime.
- **Recovery point objective (RPO):** The database preferentially ensures data reliability to minimize data loss. Use the RPO policy if you require high data consistency.

There are three data replication modes:

- **Asynchronous replication (Async):** When an application initiates an update request such as add, delete, or modify operations, the primary node responds to the application immediately after the primary node completes the operation. The primary node then replicates data to the secondary node asynchronously. This means that the operation of the primary database is not affected if the secondary node is unavailable. Data inconsistencies may occur if the primary node is unavailable.
- **Forced synchronous replication (Sync):** When an application initiates an update request such as add, delete, or modify operations, the primary node replicates data to the secondary node immediately after the primary node completes the operation. The primary node then waits for the secondary node to return a success message before the primary node responds to the application. The primary node replicates data to the secondary node synchronously. Unavailability of the secondary node will affect the operation on the primary node. Data will remain consistent even when the primary node is unavailable.
- **Semi-synchronous replication (Semi-Sync):** Data is typically replicated in Sync mode. When trying to replicate data to the secondary node, if an exception occurs causing the primary and secondary nodes to be unable to communicate with each other, the primary node will suspend response to the application. If the connection cannot be restored, the primary node will degrade to Async mode and restore response to the application after the Sync replication times out. In a situation such as this, the primary node becoming unavailable will lead to data inconsistency. After the secondary node or network connection is recovered, data replication between the two nodes is resumed, and the data replication mode will change from Async to Sync.

You can select different combinations of service priorities and data replication modes to improve availability based on the business features.

9.4.3. Backup and recovery service

This service supports data backup, dump, and recovery functions.

ApsaraDB for RDS can initiate database backup at any time. It can also restore databases to the status of any point in time based on backup policy, improving the traceability of data.

Backup

The Backup module compresses and uploads data and logs on both the primary and secondary nodes. ApsaraDB for RDS uploads backup files to OSS by default and dumps the backup files to a more cost-effective and persistent Archive Storage system. When the secondary node is operating properly, backup is always initiated on the secondary node. This will not affect the services on the primary node. When the secondary node is unavailable or damaged, the Backup module initiates backup on the primary node.

Recovery

The Recovery module restores backup files stored on OSS to a destination node.

- **Primary node rollback:** when an operation error occurs, rolls back the primary node to a specified point in time.
- **Secondary node repair:** when an irreparable fault occurs on the secondary node, creates a new secondary node to reduce risk.
- **Read-only instance creation:** creates a read-only instance from backup files.

Dump

The Dump module uploads, dumps, and downloads backup files. Currently, all backup data is uploaded to OSS for storage. You can obtain temporary links to download data as needed. In certain scenarios, the Dump module allows you to dump backup files from OSS to Archive Storage for more cost-effective and longer-term offline storage.

 **Note** PPAS cannot support the download of backup files. It must back up data through `pg_dump`.

9.4.4. Monitoring service

ApsaraDB for RDS provides multilevel monitoring services across the physical, network, and application layers to ensure service availability.

Service

The Service module tracks the status of services that RDS depends on, such as SLB, OSS, log service, and Archive Storage, to ensure they are operating properly. Monitored metrics include functionality and response time. The Service module also uses logs to determine whether the internal RDS services are operating properly.

Network

The Network module tracks statuses at the network layer. It monitors the connectivity between ECS and RDS and between physical RDS servers. It also monitors the rates of packet loss on the VRouter and VSwitch.

OS

The OS module tracks the status of hardware and the OS kernel. The monitored items include:

- **Hardware maintenance:** The OS module constantly checks the operating status of the CPU, memory, motherboard, and storage device. It can predict faults in advance and automatically submit repair reports when it determines a fault is likely to occur.
- **OS kernel monitoring:** The OS module tracks all database calls and analyzes the causes of slow calls or call errors based on the kernel status.

Instance

The Instance module collects the following information on RDS instances:

- Instance availability information
- Instance capacity and performance metrics
- Instance SQL execution records

9.4.5. Scheduling service

The Resource module implements the scheduling of resources and services.

Resource

The Resource module allocates and integrates underlying RDS resources when you activate and migrate instances. When you use the RDS console or API to create an instance, the Resource module calculates the most suitable host to carry the traffic to and from the instance. This module also allocates and integrates the underlying resources required to migrate RDS instances. After repeated instance creation, deletion, and migration operations, the Resource module calculates the degree of resource fragmentation. It also regularly integrates resources to improve the service carrying capacity.

9.4.6. Migration service

RDS provides Data Transmission Service (DTS) to help you migrate databases quickly.

The migration service helps you migrate data from the on-premises database to ApsaraDB for RDS, or migrate data from an instance to another instance in ApsaraDB for RDS.

DTS

DTS enables data migration from on-premises databases to RDS instances or between different RDS instances. DTS supports three database engines: MySQL and PostgreSQL.

DTS provides three migration methods: schema migration, full migration, and incremental migration.

- Schema migration

DTS migrates the schema definitions of migration objects to the destination instance. Tables, views, triggers, stored procedures, and stored functions can be migrated in this mode.

- Full migration

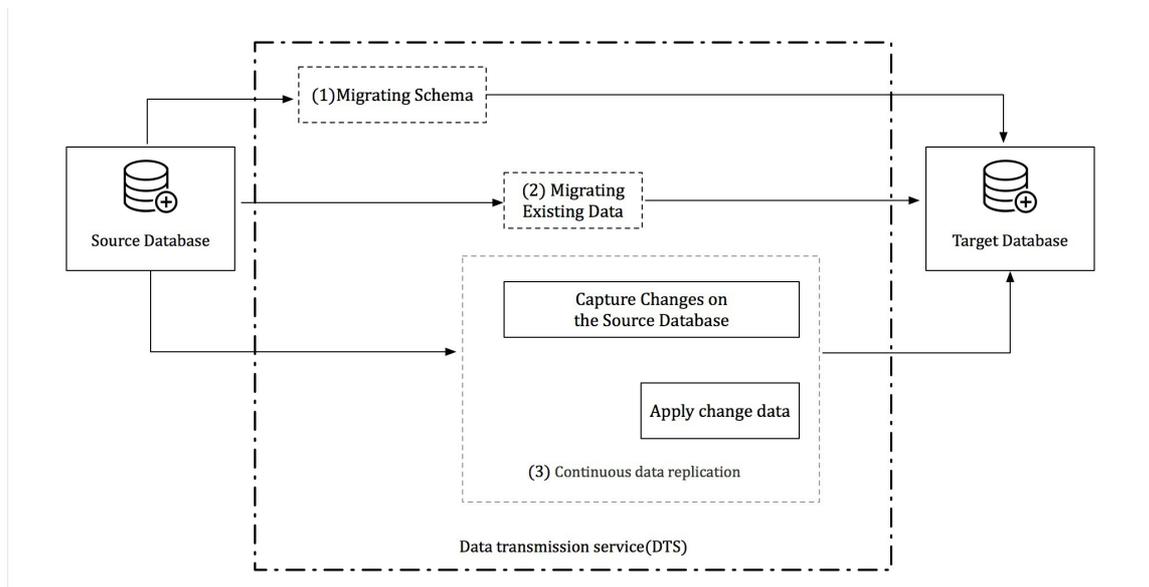
DTS migrates all data of migration objects from the source database to the destination instance.

Notice To ensure data consistency, non-transaction tables that do not have primary keys will be locked when performing a full migration. Locked tables cannot be written to. The lock duration depends on the amount of data in the tables. The tables will be unlocked only after they are fully migrated.

- Incremental migration

DTS synchronizes data changes made in the migration process to the destination instance.

Notice If a DDL operation is performed during data migration, schema changes will not be synchronized to the destination instance.



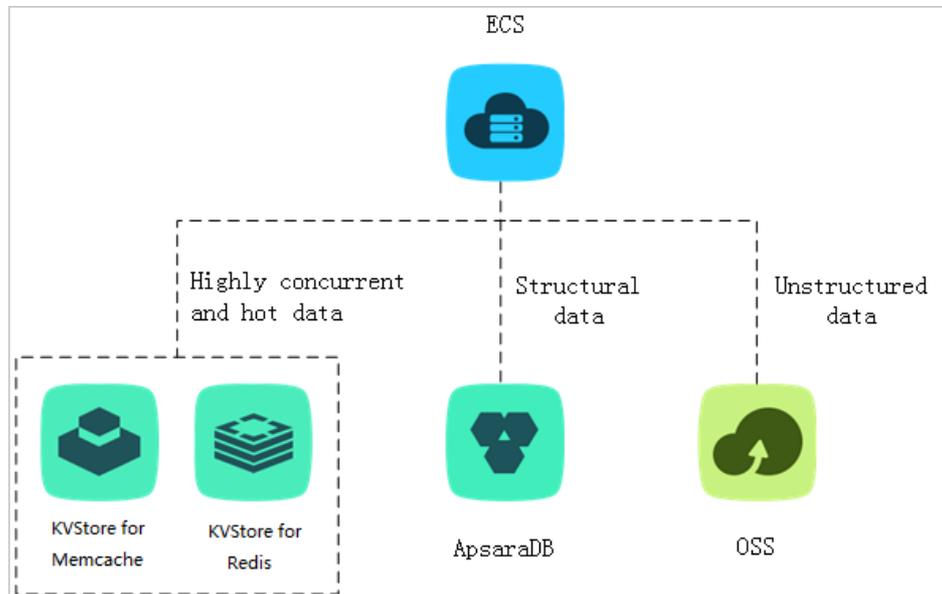
9.5. Scenarios

9.5.1. Diversified data storage

ApsaraDB for RDS provides cache data persistence and multi-structure data storage.

You can diversify the storage capabilities of ApsaraDB for RDS through services such as KVStore for Memcache, KVStore for Redis, and OSS, as shown in **Diversified data storage**.

Diversified data storage



Cache data persistence

ApsaraDB for RDS can be used with KVStore for Memcache and KVStore for Redis to form a high-throughput and low-latency storage solution. ApsaraDB cache services have the following benefits over ApsaraDB for RDS:

- High response speed: The request latency of KVStore for Memcache and KVStore for Redis is only a few milliseconds.
- The cache area supports a higher number of queries per second (QPS) than ApsaraDB for RDS.

Multi-structure data storage

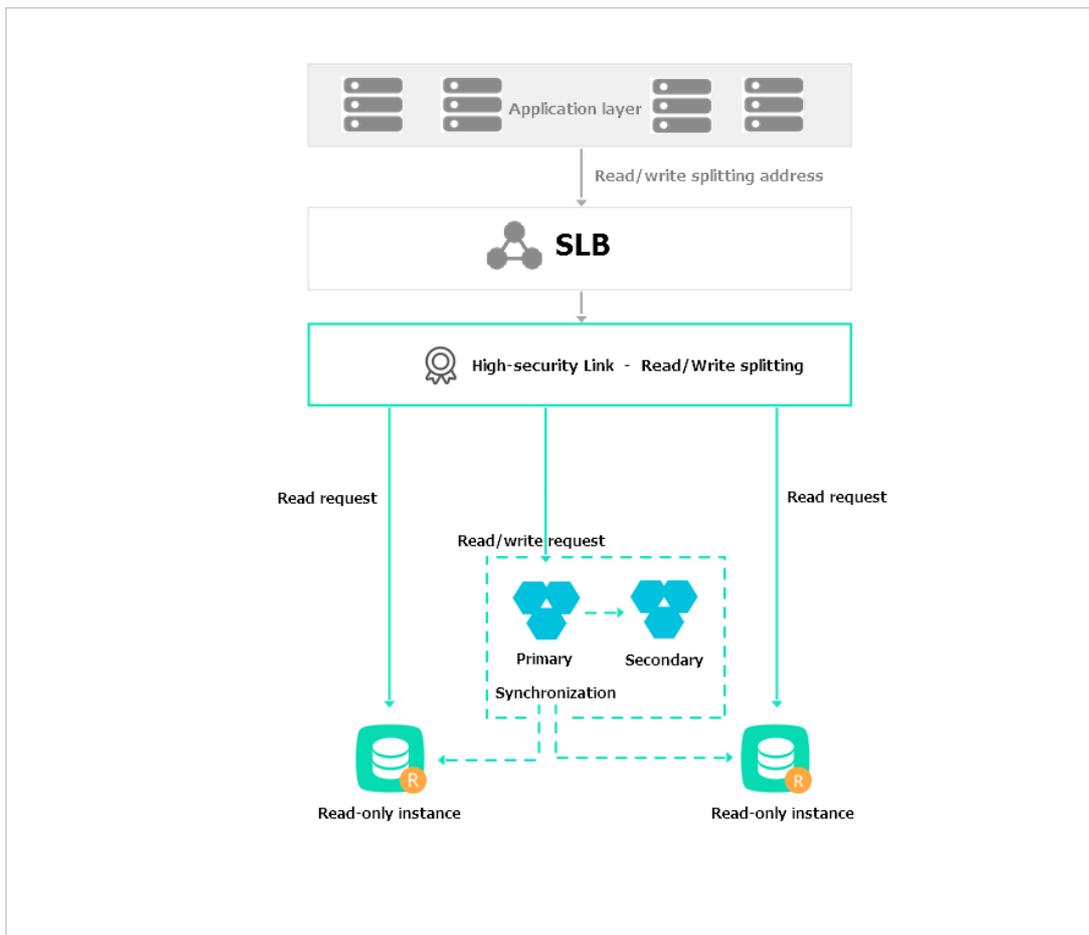
OSS is a secure, reliable, low-cost, and high-capacity storage service from Alibaba Cloud. ApsaraDB for RDS can be used with OSS to implement a multi-type data storage solution. For example, imagine ApsaraDB for RDS and OSS are used together to implement an online forum. Resources such as the images of registered users and posts on the forum can be stored in OSS to reduce storage needs on ApsaraDB for RDS.

9.5.2. Read/write splitting

This feature allows you to split read requests and write requests across different instances to expand the processing capability of the system.

ApsaraDB RDS for MySQL allows you to directly attach read-only instances to ApsaraDB for RDS to reduce read pressure on the primary instance. The primary instance and read-only instances of ApsaraDB RDS for MySQL each have their own connection addresses. The system also offers an extra read/write splitting address after read/write splitting is enabled. This address associates the primary instance with all of its read-only instances for automatic read/write splitting, allowing applications to send all read and write requests to a single address. Write requests are automatically routed to the primary instance, and read requests are routed to each read-only instance based on their weights. You can scale out the processing capability of the system by adding more read-only instances. There is no need to modify applications, as shown in [Read/write splitting](#).

Read/write splitting

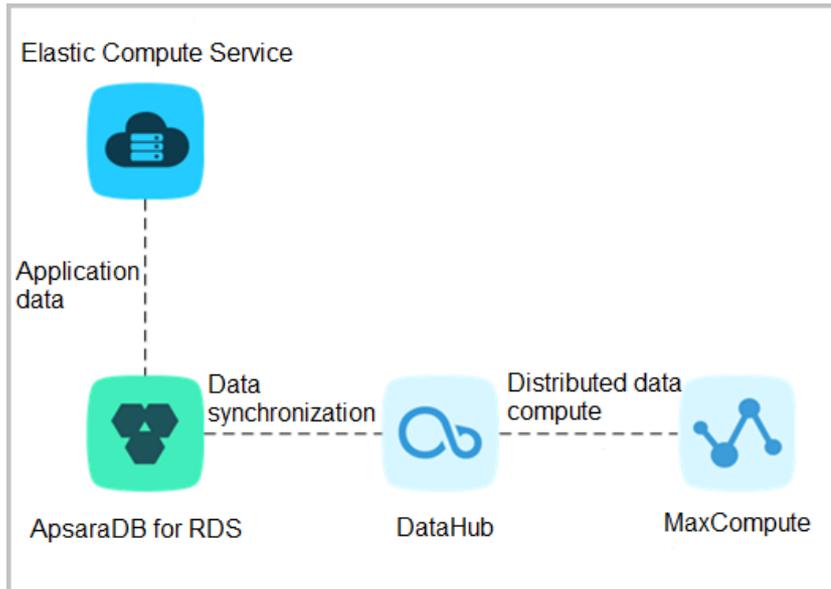


9.5.3. Big data analysis

You can import data from RDS to MaxCompute to enable large-scale data computing.

MaxCompute is used to store and compute batches of structured data. It provides various data warehouse solutions as well as big data analysis and modeling services, as shown in [Big data analysis diagram](#).

Big data analysis diagram



9.6. Usage limits

9.6.1. Limits on ApsaraDB RDS for MySQL

Before you use ApsaraDB RDS for MySQL, you must understand its limits and take precautions.

To guarantee instance stability and security, ApsaraDB RDS for MySQL has some service limits, as listed in [Limits on ApsaraDB RDS for MySQL](#).

Limits on ApsaraDB RDS for MySQL

Operation	Description
Database parameter modification	Database parameters can only be modified through the RDS console or API operations. Due to security and stability considerations, only specific parameters can be modified.
Root permissions of databases	The root and SA permissions are not provided.
Database backup	<ul style="list-style-type: none"> Logical backup can be performed through the command line interface (CLI) or graphical user interface (GUI). Physical backup can only be performed through the RDS console or API operations.
Database restoration	<ul style="list-style-type: none"> Logical restoration can be performed through the CLI or GUI. Physical restoration can only be performed through the RDS console or API operations.
Data import	<ul style="list-style-type: none"> Logical import can be performed through the CLI or GUI. Data can only be imported through the MySQL CLI.

Operation	Description
ApsaraDB RDS for MySQL storage engine	<ul style="list-style-type: none"> • Only InnoDB and TokuDB are supported. Due to the inherent shortcomings of the MyISAM engine, some data may be lost. Only some stock instances use the MyISAM engine. MyISAM engine tables in newly created instances will be automatically converted to InnoDB engine tables. • For safety performance and security considerations, we recommend that you use the InnoDB storage engine. • The Memory engine is not supported. Newly created Memory tables will be automatically converted into InnoDB tables.
Database replication	ApsaraDB RDS for MySQL provides dual-node clusters based on a primary/secondary replication architecture. The secondary instances in this replication architecture are hidden and cannot be accessed directly.
RDS instance restart	Instances must be restarted through the RDS console or API operations.
Account and database management	ApsaraDB RDS for MySQL uses the RDS console to manage accounts and databases. ApsaraDB RDS for MySQL also allows you to create a privileged account to manage users, passwords, and databases.
Standard account	<ul style="list-style-type: none"> • Custom authorization is not supported. • The account management and database management interfaces are provided in the RDS console. • Instances that support standard accounts also support privileged accounts.
Privileged account	<ul style="list-style-type: none"> • Custom authorization is supported. • The RDS console does not provide interfaces to manage accounts or databases. Relevant operations can only be performed through code or DMS. • The privileged account cannot be reverted back to a standard account.

9.6.2. Usage limits of ApsaraDB RDS for PostgreSQL

Before you use ApsaraDB RDS for PostgreSQL, you need to understand its limits and take precautions against them.

To guarantee instance stability and security, ApsaraDB RDS for PostgreSQL has some service limits, as listed in [Limits on ApsaraDB RDS for PostgreSQL](#).

Limits on ApsaraDB RDS for PostgreSQL

Operation	Description
Database parameter modification	Not supported.
Root permission of databases	Superuser permissions are not provided.

Operation	Description
Database backup	Data can only be backed up by using <code>pg_dump</code> .
Data migration	Only PostgreSQL can be used to restore data that was backed up by using <code>pg_dump</code> .
Database replication	<ul style="list-style-type: none"> The system automatically builds HA databases based on PostgreSQL streaming replication without user input. PostgreSQL standby nodes are hidden and cannot be accessed directly.
RDS instance restart	RDS instances must be restarted from the RDS console or through APIs.
Network settings	For instances that are operating in safe mode, <code>net.ipv4.tcp_timestamps</code> cannot be enabled in SNAT mode.

9.6.3. Usage limits of ApsaraDB RDS for PPAS

Before you use ApsaraDB RDS for PPAS, you must understand its limits and take precautions against them.

To guarantee instance stability and security, ApsaraDB RDS for PPAS has some service limits, as listed in [Limits on ApsaraDB RDS for PPAS](#).

Limits on ApsaraDB RDS for PPAS

Operation	Description
Database parameter modification	Not supported.
Root permission of databases	Superuser permissions are not provided.
Database backup	Data can only be backed up by using <code>pg_dump</code> .
Data migration	Only PostgreSQL can be used to restore data that was backed up by using <code>pg_dump</code> .
Database replication	<ul style="list-style-type: none"> The system automatically builds HA databases based on PPAS streaming replication without user input. PPAS standby nodes are hidden and cannot be accessed directly.
RDS instance restart	RDS instances must be restarted from the RDS console or through APIs.
Network settings	For instances that are operating in safe mode, <code>net.ipv4.tcp_timestamps</code> cannot be enabled in SNAT mode.

9.7. Terms

Term	Description
Region	The geographical location where the server of your RDS instance resides. You must specify a region when you create an RDS instance. The region of an instance cannot be changed after instance creation. RDS must be used together with ECS and only supports intranet access. Because of this, RDS instances must be located in the same region as their corresponding ECS instances.
Zone	The physical area with an independent power supply and network in a region. Zones in a region can communicate through the intranet. Network latency for resources within the same zone is lower than for those across zones. Faults are isolated between zones. Single zone refers to the case where the three nodes in the RDS instance replica set are all located in the same zone. Network latency is reduced if an ECS instance and its corresponding RDS instance are both deployed in the same zone.
Instance	The most basic unit of RDS. An instance is the operating environment of ApsaraDB for RDS and works as an independent process on a host. You can create, modify, or delete an RDS instance from the RDS console. Instances are mutually independent and their resources are isolated. They do not compete for resources such as CPU, memory, or I/O. Each instance has its own features, such as database type and version. RDS controls instance behavior by using corresponding parameters.
Memory	The maximum amount of memory that can be used by an ApsaraDB for RDS instance.
Disk capacity	The amount of disk space selected when creating an ApsaraDB for RDS instance. Instance data that occupies disk space includes aggregated data as well as data required for normal instance operations such as system databases, database rollback logs, redo logs, and indexing. Ensure that the disk capacity is sufficient for the RDS instance to store data. Otherwise, the RDS instance may be locked. If the instance is locked due to insufficient disk capacity, you can unlock the instance by expanding the disk capacity.
IOPS	The maximum number of read/write operations performed per second on block devices at a granularity of 4 KB.
CPU core	The maximum computing capability of the instance. A single Intel Xeon series CPU core has at least 2.3 GHz of computational power with hyper-threading capabilities.
Number of connections	The number of TCP connections between a client and an RDS instance. If the client uses a connection pool, the connection between the client and RDS instance is a persistent connection. Otherwise, it is a transient connection.

9.8. Instance types

Instances of different editions, versions, and types each perform differently from one another.

ApsaraDB RDS for MySQL instance types

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment	
High-availability Edition	5.6 or 5.7	Dedicated instance (X8)	mysql.x8.medium.2	2 cores, 16 GB	2,500	4,500	250 GB	Single-data center deployment	
			mysql.x8.large.2	4 cores, 32 GB	5,000	9,000	500 GB		
			mysql.x8.xlarge.2	8 cores, 64 GB	10,000	18,000	1 TB		
			mysql.x8.2xlarge.2	16 cores, 128 GB	20,000	36,000	2 TB		
		Dedicated instance (X4)	mysql.x4.large.2	4 cores, 16 GB	2,500	4,500	250 GB		Dual-data center deployment
			mysql.x4.xlarge.2	8 cores, 32 GB	5,000	9,000	500 GB		
			mysql.x4.2xlarge.2	16 cores, 64 GB	10,000	18,000	1 TB		
			mysql.x4.4xlarge.2	32 cores, 128 GB	20,000	36,000	2 TB		
		Dedicated host	rds.mysql.st.d13	30 cores, 220 GB	64,000	20,000	3 TB		
		MySQL Finance Edition (three-node)	5.6	Dedicated instance (high memory)	mysql.x8.medium.4	2 cores, 16 GB	2,500	4,500	250 GB
mysql.x8.large.4	4 cores, 32 GB				5,000	9,000	500 GB		
mysql.x8.xlarge.4	8 cores, 64 GB				10,000	18,000	1 TB		
mysql.x8.2xlarge.4	16 cores, 128 GB				20,000	36,000	2 TB		
mysql.x8.4xlarge.4	32 cores, 256 GB				40,000	72,000	3 TB		

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment	
Read-only instance	5.6 or 5.7	Dedicated instance (X8)	mysqlro.x8.medium.1	2 cores, 16 GB	2,500	4,500	250 GB	Single-data center deployment	
			mysqlro.x8.large.1	4 cores, 32 GB	5,000	9,000	500 GB		
			mysqlro.x8.xlarge.1	8 cores, 64 GB	10,000	18,000	1 TB		
			mysqlro.x8.2xlarge.1	16 cores, 128 GB	20,000	36,000	2 TB		
		Dedicated instance (X4)	mysqlro.x4.large.1	4 cores, 16 GB	2,500	4,500	250 GB		Dual-data center deployment
			mysqlro.x4.xlarge.1	8 cores, 32 GB	5,000	9,000	500 GB		
			mysqlro.x4.2xlarge.1	16 cores, 64 GB	10,000	18,000	1 TB		
			mysqlro.x4.4xlarge.1	32 cores, 128 GB	20,000	36,000	2 TB		
		Dedicated host	rds.mysql.st.d13	30 cores, 220 GB	64,000	20,000	3 TB		

ApsaraDB RDS for PostgreSQL instance types

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
---------	---------	-----------------	---------------	----------------	---------------------	--------------	---------------	-----------------------------------

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
High-availability Edition	9.4	Dedicated instance (high memory)	pg.x8.medium.2	2 cores, 16 GB	2,500	4,500	250 GB	Single-data center deployment Dual-data center deployment
			pg.x8.large.2	4 cores, 32 GB	5,000	9,000	500 GB	
			pg.x8.xlarge.2	8 cores, 64 GB	10,000	18,000	1 TB	
			pg.x8.2xlarge.2	16 cores, 128 GB	12,000	36,000	2 TB	
		Dedicated instance (high CPU)	pg.x4.large.2	4 cores, 16 GB	2,500	4,500	250 GB or 500 GB	
			pg.x4.xlarge.2	8 cores, 32 GB	5,000	9,000	500 GB or 1 TB	
			pg.x4.2xlarge.2	16 cores, 64 GB	10,000	18,000	1 TB or 2 TB	
			pg.x4.4xlarge.2	32 cores, 128 GB	12,000	36,000	2 TB or 3 TB	
		Dedicated host	rds.pg.sdt.d13	30 cores, 220 GB	4,000	20,000	3 TB	

ApsaraDB RDS for PPAS instance types

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
---------	---------	-----------------	---------------	----------------	---------------------	--------------	---------------	-----------------------------------

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
High-availability Edition	9.3 or 9.6	Dedicated instance	ppas.x4.small.2	1 core, 4 GB	200	5,000	250 GB	Single-data center deployment
			ppas.x4.medium.2	2 cores, 8 GB	400	10,000	250 GB	
			ppas.x8.medium.2	2 cores, 16 GB	2,500	15,000	250 GB	
			ppas.x4.large.2	4 cores, 16 GB	2,500	20,000	250 GB or 500 GB	
			ppas.x8.large.2	4 cores, 32 GB	5,000	30,000	250 GB or 500 GB	
			ppas.x4.xlarge.2	8 cores, 32 GB	5,000	40,000	500 GB or 1 TB	
			ppas.x8.xlarge.2	8 cores, 64 GB	10,000	60,000	500 GB or 1 TB	
			ppas.x4.2xlarge.2	16 cores, 64 GB	10,000	80,000	1 TB or 2 TB	
			ppas.x8.2xlarge.2	16 cores, 128 GB	12,000	120,000	1 TB or 2 TB	
			ppas.x4.4xlarge.2	32 cores, 128 GB	12,000	160,000	2 TB or 3 TB	
		ppas.x8.4xlarge.2	32 cores, 256 GB	12,000	240,000	2 TB or 3 TB		
		Dedicated host	rds.ppas.st.d13	30 cores, 220 GB	4,000	20,000	3 TB	Dual-data center deployment

10.KVStore for Redis

10.1. What is KVStore for Redis?

KVStore for Redis is an online key-value storage service compatible with open-source Redis protocols. KVStore for Redis supports various types of data, such as strings, lists, sets, sorted sets, and hash tables. The service also supports advanced features, such as transactions, message subscription, and message publishing. Based on the hybrid storage of memory and hard disks, KVStore for Redis can provide high-speed data read/write capability and support data persistence.

As a cloud computing service, KVStore for Redis works with hardware and data deployed in the cloud, and provides comprehensive infrastructure planning, network security protections, and system maintenance services. This service allows you to focus on business innovation.

10.2. Benefits

High performance

- Supports cluster features and provides cluster instances of 128 GB or higher to meet large capacity and high performance requirements.
- Provides primary/secondary instances of 32 GB or smaller to meet general capacity and performance requirements.

Elastic scaling

- Easy scaling of storage capacity: you can scale instance storage capacity in the KVStore for Redis console based on business requirements.
- Online scaling without interrupting services: you can scale instance storage capacity on the fly. This does not affect your business.

Resource isolation

Instance-level resource isolation provides enhanced stability for individual services.

Data security

- Persistent data storage: based on the hybrid storage of memory and hard disks, KVStore for Redis can provide high-speed data read/write capability and support data persistence.
- Dual-copy backup and failover: KVStore for Redis backs up data on both a primary node and a secondary node and supports the failover feature to prevent data loss.
- Access control: KVStore for Redis requires password authentication to ensure secure and reliable access.
- Data transmission encryption: KVStore for Redis supports encryption based on Secure Sockets Layer (SSL) and Secure Transport Layer (TLS) to secure data transmission.

High availability

- Primary/secondary structure: each instance runs in this structure to eliminate the possibility of single points of failure (SPOFs) and guarantee high availability.
- Automatic detection and recovery of hardware faults: the system automatically detects hardware faults and performs the failover operation within several seconds. This can minimize your business losses caused by unexpected hardware faults.

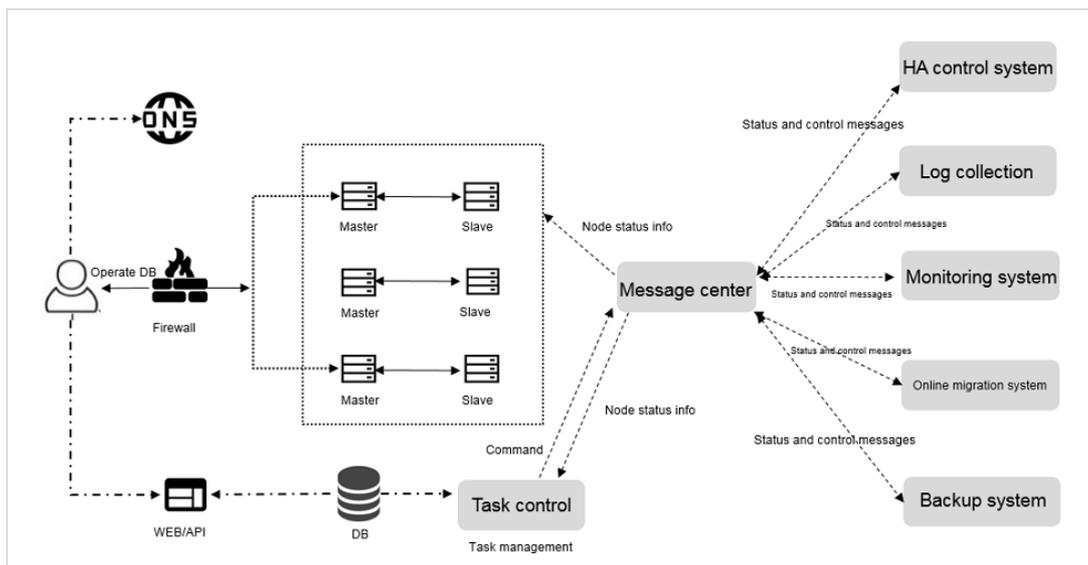
Easy to use

- Out-of-the-box service: KVStore for Redis requires no setup or installation. You can use the service immediately after purchase to ensure efficient business deployment.
- Compatible with open-source Redis: KVStore for Redis is compatible with Redis commands. You can use any Redis clients to easily connect to KVStore for Redis and perform data operations.

10.3. Architecture

The architecture of KVStore for Redis is as shown in [Architecture diagram](#).

Architecture diagram



KVStore for Redis automatically builds a primary/secondary structure. You can use this structure directly.

- **HA control system**

A high-availability (HA) detection module is used to detect and monitor the operating status of KVStore for Redis instances. If this module determines that a primary node is unavailable, the module automatically performs the failover operation to ensure high availability of KVStore for Redis instances.

- **Log collection**

This module collects instance operation logs, including slow query logs and access control logs.

- **Monitoring system**

This module collects performance monitoring information of KVStore for Redis instances, including basic group monitoring, key group monitoring, and string group monitoring.

- **Online migration system**

When an error occurs on the physical server that hosts a KVStore for Redis instance, this module recreates an instance on the fly based on the backup files stored in the backup system. This ensures high availability of your business.

- **Backup system**

This module generates backup files of KVStore for Redis instances, and stores the backup files in Object Storage Service (OSS). The backup system allows you to customize the backup settings, and retains backup files for up to seven days.

- **Task Control**

KVStore for Redis instances support various management and control tasks, including instance creation, specifications changes, and instance backups. The task system flexibly controls and tracks tasks and manages errors according to your instructions.

10.4. Features

- **High-availability technology ensures service stability**

The system synchronizes data between the primary node and the secondary node in real time. If the primary node fails, the system automatically performs the failover operation and restores services within a few seconds. The secondary node takes over services. This process does not affect your business, and ensures high availability of system services.

Cluster instances run in a distributed architecture. Each node uses a primary/secondary high-availability structure to automatically perform failover and disaster recovery and ensure high availability of system services.

- **Easy backup and recovery support custom backup policies**

You can back up data in the console and customize automatic backup policies. The system automatically retains backup data for seven days. You can easily restore data in the case of accidental data operations to minimize your business losses.

- **Multiple network security protections secure your data**

A Virtual Private Cloud (VPC) isolates network transmission at the transport layer. The Anti-Distributed-Denial-of-Service (DDoS) protection service monitors and protects against DDoS attacks. The system supports a whitelist that contains a maximum of 1,000 IP addresses or CIDR blocks to prevent malicious login attempts.

- **Kernel optimization avoids vulnerability exploits**

The experts of Alibaba Cloud have performed in-depth kernel optimization for the Redis source code to effectively prevent running out of memory, fix security vulnerabilities, and protect your business.

- **Elastic scaling eliminates capacity and performance bottlenecks**

KVStore for Redis supports multiple memory types. You can upgrade the memory type based on your service requirements.

The cluster architecture allows you to elastically scale the storage space and throughput performance of the database system. This eliminates the performance bottlenecks.

- **Multiple instance types support flexible specifications changes**

The single-node cache architecture and two-node storage architecture are applicable to various service scenarios. You can flexibly change instance specifications.

- **Monitoring and alerts allow you to check instance status in real time**

KVStore for Redis provides monitoring and alerts of instance information, such as CPU usage, connections, and disk utilization. You can check instance status anywhere and at any time.

- **Visual management simplifies operations and maintenance**

The KVStore for Redis console, a visual management platform, allows you to easily perform frequent and risky operations, such as instance cloning, backup, and data restoration.

- **Automatic engine version upgrades prevent software flaws**

The system automatically upgrades engine versions and efficiently fixes flaws so that you can easily manage database versions.

- **Custom parameters support individual requirements**

You can set parameters in the KVStore for Redis console to make full use of system resources.

10.5. Scenarios

Game industry applications

KVStore for Redis can be an important part of the business architecture for deploying a game application.

Scenario 1: KVStore for Redis works as a storage database

The architecture for deploying a game application is simple. You can deploy a main program on an ECS instance and all business data on a KVStore for Redis instance. The KVStore for Redis instance works as a persistent storage database. KVStore for Redis supports data persistence, and stores redundant data on primary and secondary nodes.

Scenario 2: KVStore for Redis works as a cache to accelerate connections to applications

KVStore for Redis can work as a cache to accelerate connections to applications. You can store data in a Relational Database Service (RDS) database that works as a backend database.

Reliability of the KVStore for Redis service is vital to your business. If the KVStore for Redis service is unavailable, the backend database is overloaded when processing connections to your application. KVStore for Redis provides a two-node hot standby architecture to ensure high availability and reliability of services. The primary node provides services for your business. If this node fails, the system automatically switches services to the secondary node. The complete failover process is transparent.

Live video applications

In live video services, KVStore for Redis works as an important measure to store user data and relationship information.

Two-node hot standby ensures high availability

KVStore for Redis uses the two-node hot standby method to maximize service availability.

Cluster editions eliminate the performance bottleneck

KVStore for Redis provides cluster instances to eliminate the performance bottleneck that is caused by Redis single-thread mechanism. Cluster instances can effectively handle traffic bursts during live video streaming and support high-performance requirements.

Easy scaling relieves pressure at peak hours

KVStore for Redis allows you to easily perform scaling. The complete upgrade process is transparent. Therefore, you can easily handle traffic bursts at peak hours.

E-commerce industry applications

In the e-commerce industry, the KVStore for Redis service is widely used in the modules such as commodity display and shopping recommendation.

Scenario 1: rapid online sales promotion systems

During a large-scale rapid online sales promotion, a shopping system is overwhelmed by traffic. A common database cannot properly handle so many read operations.

However, KVStore for Redis supports data persistence, and can work as a database system.

Scenario 2: counter-based inventory management systems

In this scenario, you can store inventory data in an RDS database and save count data to corresponding fields in the database. In this way, the KVStore for Redis instance reads count data, and the RDS database stores count data. KVStore for Redis is deployed on a physical server. Based on solid-state drive (SSD) high-performance storage, the system can provide a high-level data storage capacity.

10.6. Limits

Item	Description
List data type	The number of lists is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a list is less than 8,192. The value length is 1 MB or less.
Set data type	The number of sets is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a set is less than 8,192. The value length is 1 MB or less.
Sorted set data type	The number of sorted sets is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a sorted set is less than 8,192. The value length is 1 MB or less.
Hash data type	The number of fields is not limited. The size of each element in a hash table is 512 MB or less. We recommend that the number of elements in a hash table is less than 8,192. The value length is 1 MB or less.
Number of databases (DBs)	Each instance supports 256 DBs.
Supported Redis commands	For more information, see the "Supported Redis commands" topic of <i>KVStore for Redis User Guide</i> .
Monitoring and alerts	KVStore for Redis does not provide capacity alerts. You have to configure this feature in CloudMonitor. We recommend that you set alerts for the following metrics: instance faults, instance failover, connection usage, failed operations, capacity usage, write bandwidth usage, and read bandwidth usage.
Expired data deletion policies	<ul style="list-style-type: none"> Active expiration: the system periodically detects and deletes expired keys in the background. Passive expiration: the system deletes expired keys when you access these keys.

Item	Description
Idle connection recycling mechanism	KVStore for Redis does not actively recycle idle connections to KVStore for Redis. You can manage the connections.
Data persistence policy	KVStore for Redis uses the AOF_FSYNC_EVERYSEC policy, and runs the fsync command at a one-second interval.

10.7. Terms

Redis

A high-performance key-value storage system that works as a cache and store and that is compatible with BSD open-source protocols.

Instance ID

An instance corresponds to a user space, and serves as the basic unit of using Redis.

Redis has limits on instance specifications, such as connections, bandwidth, and CPU processing capacity. These limits vary according to different instance types. You can view the list of instance identifiers that you have purchased in the console. KVStore for Redis instances are classified into primary/secondary instances and high-performance cluster instances.

Primary/secondary instance

A KVStore for Redis instance that contains a primary/secondary structure. The primary/secondary instance provides limited capacity and performance.

High-performance cluster instance

A KVStore for Redis instance that runs in a scalable cluster architecture. Cluster instances provide better scalability and performance, but they still have limited features.

Connection address

The host address for connecting to KVStore for Redis. The connection address is displayed as a domain name. To obtain the connection address, go to the **Instance Information** tab page, and check the address in the **Connection Information** field.

Eviction policy

The policy that KVStore for Redis uses to delete earlier data when the memory of KVStore for Redis reaches the upper limit as specified in maxmemory. Eviction policies of KVStore for Redis are consistent with Redis eviction policies. For more information, see [Using Redis as an LRU cache](#).

DB

The abbreviation of the word "database" to indicate a database in KVStore for Redis. Each KVStore for Redis instance supports 256 databases numbered DB 0 to DB 255. The system writes data to DB 0 by default.

10.8. Instance types

 **Note** The maximum bandwidth includes the maximum upstream bandwidth and the maximum downstream bandwidth.

Standard dual-copy edition

Standard plan

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description	Zone-disaster recovery deployment
1 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb1g.1db.0rodb.4proxy.default	10,000	10	Single-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
2 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb2g.1db.0rodb.4proxy.default	10,000	16	Single-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
4 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb4g.1db.0rodb.4proxy.default	10,000	24	Single-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
8 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb8g.1db.0rodb.4proxy.default	10,000	24	Single-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description	Zone-disaster recovery deployment
16 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb16g.1db.0rodb.4proxy.default	10,000	32	Single-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
32 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb32g.1db.0rodb.4proxy.default	10,000	32	Single-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region

Premium plan

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description	Zone-disaster recovery deployment
1 GB advanced primary/secondary edition	redis.master.small.special2x	20,000	48	Single-core	Primary/secondary instance	Deployed in one zone
2 GB advanced primary/secondary edition	redis.master.mid.special2x	20,000	48	Single-core	Primary/secondary instance	Deployed in one zone
4 GB advanced primary/secondary edition	redis.master.stand.special2x	20,000	48	Single-core	Primary/secondary instance	Deployed in one zone
8 GB advanced primary/secondary edition	redis.master.large.special1x	20,000	48	Single-core	Primary/secondary instance	Deployed in one zone

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description	Zone-disaster recovery deployment
16 GB advanced primary/secondary edition	redis.master.2xlarge.special1x	20,000	48	Single-core	Primary/secondary instance	Deployed in one zone
32 GB advanced primary/secondary edition	redis.master.4xlarge.special1x	20,000	48	Single-core	Primary/secondary instance	Deployed in one zone

Cluster edition

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description
16 GB cluster edition	redis.sharding.small.default	80,000	384	4-core	High-performance cluster instance
32 GB cluster edition	redis.sharding.mid.default	80,000	384	8-core	High-performance cluster instance
64 GB cluster edition	redis.sharding.large.default	80,000	384	8-core	High-performance cluster instance
128 GB cluster edition	redis.sharding.2xlarge.default	160,000	768	16-core	High-performance cluster instance
256 GB cluster edition	redis.sharding.4xlarge.default	160,000	768	16-core	High-performance cluster instance
512 GB cluster edition	redis.logic.sharding.16g.32db.0rodb.32proxy.default	320,000	1,536	8-core	
1 TB cluster edition	redis.sharding.16xlarge.default	640,000	3,072	8-core	

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description
2 TB cluster edition	redis.sharding.32xlarge.default	1,280,000	6,144	8-core	
4 TB cluster edition	redis.logic.sharding.16g.256db.0rodb.256proxy.default	2,560,000	12,288	16-core	

Cluster edition for zone-disaster recovery

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description
16 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredis.mdb16g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
32 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredis.mdb32g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
64 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredis.mdb64g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
128 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredis.mdb128g.16db.0rodb.16proxy.default	160,000	768	16-core	Cluster instance for zone-disaster recovery
256 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredis.mdb256g.16db.0rodb.16proxy.default	160,000	768	16-core	Cluster instance for zone-disaster recovery

Instance type	Service code	Maximum number of connections	Maximum bandwidth (MB)	Processing capability	Description
256 GB cluster edition for zone-disaster recovery	redis.logic.sharding. 16g. 32db. 0rodb. 32proxy.default	320,000	1,536	8-core	Cluster instance for zone-disaster recovery
1 TB cluster edition for zone-disaster recovery	redis.sharding. 16xlarge.default	640,000	3,072	8-core	Cluster instance for zone-disaster recovery
2 TB cluster edition for zone-disaster recovery	redis.sharding. 32xlarge.default	1,280,000	6,144	8-core	Cluster instance for zone-disaster recovery
4 TB cluster edition for zone-disaster recovery	redis.logic.sharding. 16g. 256db. 0rodb. 256proxy.default	2,560,000	12,288	16-core	Cluster instance for zone-disaster recovery

QPS performance reference

QPS performance reference

Instance type (GB)	Maximum number of connections	Maximum internal network bandwidth (MB)	CPU processing capacity	QPS reference value
8	10,000	24	Single-core	80,000

 **Note** QPS reference values of non-cluster instances range from 80,000 to 100,000. QPS reference values of cluster instances are the number of nodes multiplied by the value from 80,000 to 100,000.

11. ApsaraDB for MongoDB

11.1. What is ApsaraDB for MongoDB?

ApsaraDB for MongoDB is a high-performance distributed data storage service. It is a stable, reliable, and resizable database service fully compatible with MongoDB protocols. ApsaraDB for MongoDB offers a full range of database solutions, such as disaster recovery, backup, restoration, monitoring, and alerts.

ApsaraDB for MongoDB supports the following features:

- Automatically creates a three-node MongoDB replica set that encapsulates advanced functions such as disaster recovery and failover.
- Supports quick database backup and restoration. You can easily perform standard database backup and database rollback operations in the ApsaraDB for MongoDB console.
- Provides over 20 performance monitoring metrics and sends alerts. This helps you learn about the performance status of your database.
- Provides visual data management tools for convenient operations and maintenance.

11.2. Benefits

- **High availability**
 - The three-node replica set high-availability architecture that delivers extremely high service availability.

ApsaraDB for MongoDB uses a high-availability architecture that features a three-node replica set. These three data nodes are located on different physical servers and automatically synchronize data. Services are provided by the primary and secondary nodes. When the primary node fails, the system automatically selects a new primary node. When the secondary node is unavailable, the standby node takes over the services.
 - Automatic backup and quick data restoration.

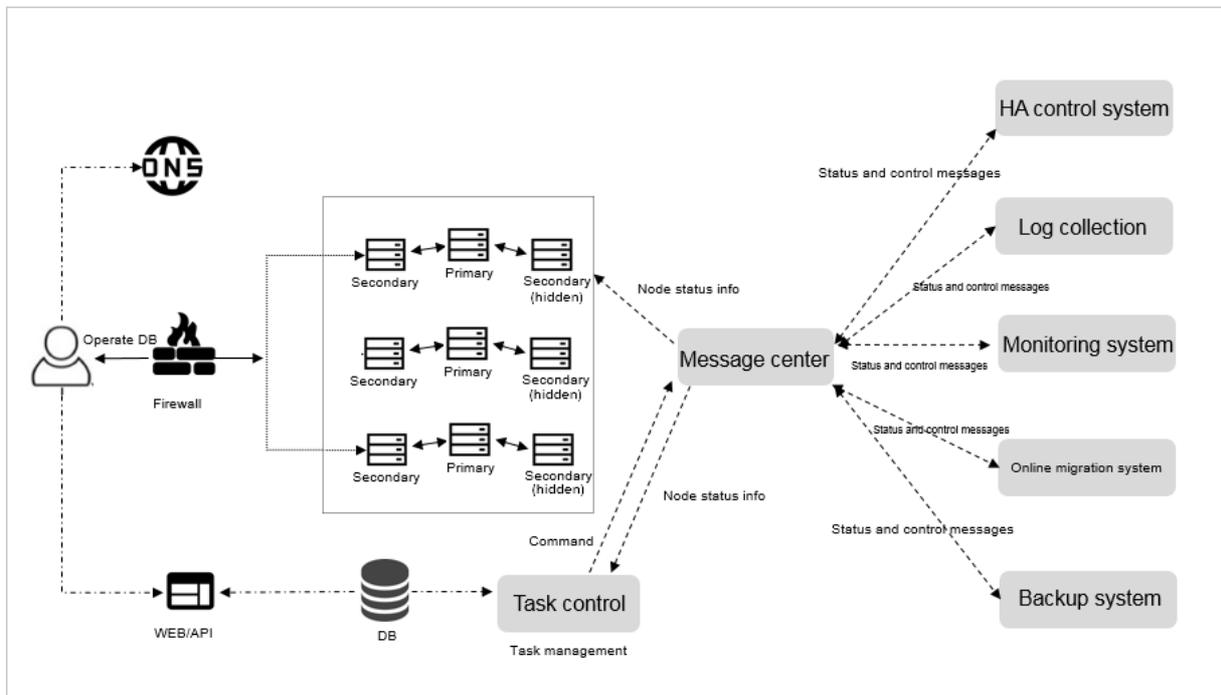
Data is automatically backed up and uploaded to Object Storage Service (OSS) each day. This improves data disaster recovery capabilities while effectively reducing disk space consumption. You can use the backup files to restore the instance data to the original instance. This effectively prevents irreversible effects on service data caused by incorrect operations or other reasons.
- **High security**
 - Anti-DDoS protection: Filters inbound traffic. When DDoS attacks are identified, the source IP addresses will be scrubbed. If scrubbing fails, the black hole mechanism is triggered.
 - IP whitelist: You can whitelist up to 1,000 IP addresses for an ApsaraDB for MongoDB instance. This helps you reduce attack risks.
- **Ease of use**

Excellent performance monitoring The monitoring platform provides real-time monitoring information about the CPU usage, connections, and disk utilization, and sends alerts. This allows you to learn about instance status.
- **Scalability**

ApsaraDB for MongoDB supports three-node replica sets that can be elastically resized. You can change the configuration of your instance if the current configuration cannot meet performance requirements or is unsuitable for your business needs. The configuration change process is completely transparent and will not affect your business.

11.3. Architecture

ApsaraDB for MongoDB provides a three-node replica set. You can directly use the primary or secondary node. The following figure shows the system architecture.



- HA control system:** This module checks the high availability of an instance. You can use this module to detect and monitor the running status of ApsaraDB for MongoDB instances. If the system detects that the primary node instance is unavailable, it switches over from the primary node to the secondary node to ensure the high availability of MongoDB instances.
- Log collection:** This module collects MongoDB running logs, including the slow query log and access log of an instance.
- Monitoring system:** This module collects performance monitoring information about MongoDB instances, including basic metrics, disk capacity, network requests, and the number of operations that you have performed.
- Online migration system:** When an error occurs with the physical server that hosts the instances, this module recreates an instance on the fly based on the backup files stored in the backup system. This ensures the continuity of your business.
- Backup system:** This module generates ApsaraDB for MongoDB instance backups and stores the backup files in OSS. Currently, this module allows you to configure custom backup settings and create temporary backups. Files are retained for seven days.
- Task control:** ApsaraDB for MongoDB supports multiple management operations, such as creating instances, changing instance configurations, and backing up instances. This module controls and tracks these tasks and troubleshoots errors based on your commands.

11.4. Features

Flexible architecture

ApsaraDB for MongoDB provides a three-node replica set. You can directly use the primary and secondary nodes. If the system detects that the primary node instance is unavailable, it switches over from the primary node to the secondary node to ensure the high availability of MongoDB instances.

Elastic scaling

- **Quick scaling of storage capacity:** You can adjust the storage capacity of an instance in the ApsaraDB for MongoDB console based on business requirements.
- **Storage capacity adjustment on the fly:** You can adjust storage capacity of an instance on the fly. This ensures the continuity of your business.

Data security

- **Automatic backup:** ApsaraDB for MongoDB allows you to set a time interval based on which backups are periodically created. You can flexibly set the backup start time based on off-peak hours of your business. Backup files are retained for free for up to seven days.
- **Temporary backup:** You can use this feature to create temporary backups as needed. Backup files are retained for free for up to seven days.
- **Data restore:** You can use backup files to directly overwrite existing data and restore an instance to a previous status.
- **Backup file download:** ApsaraDB for MongoDB retains your backup files for free for up to seven days. During this period, you can log on to the ApsaraDB for MongoDB console and download the backup files to a local device.
- **Create instances from backup sets:** You can create an instance in the ApsaraDB for MongoDB console by using backup files. This helps you quickly complete the deployment process.
- **IP whitelist:** ApsaraDB for MongoDB can filter IP addresses that access your instance. You can log on to the ApsaraDB for MongoDB console and configure a whitelist of up to 1,000 IP addresses. This provides a highly secure access environment.
- **Multi-layer network security protection**

VPC networks are isolated at the TCP layer. Anti-DDoS can monitor and block DDoS attacks in real time. You can add up to 1,000 IP addresses to the whitelist.

Intelligent operations and maintenance

- **Monitoring platform**

This platform provides real-time monitoring information about the CPU usage, connections, and disk utilization, and sends alerts. This allows you to learn about instance status.

- **Graphical O&M platform**

This platform allows you to quickly perform frequent and risky operations, such as instance cloning, backup, and data restoration.

- **Database kernel version management**

This feature proactively performs upgrades and quickly fixes exceptions. It also optimizes ApsaraDB for MongoDB parameter configurations and maximizes the utilization of system resources.

11.5. Scenarios

- **Businesses that require read/write splitting**

ApsaraDB for MongoDB uses a high-availability architecture that features a three-node replica set. These three data nodes are located on different physical servers. The secondary and standby nodes automatically synchronize data from the primary node. Services are provided by the primary and secondary nodes. These two nodes have separate domain names and collaborate with MongoDB drivers to distribute read requests.

- **Businesses that require flexibility**

As a schema-free database, ApsaraDB for MongoDB is particularly suitable for startup businesses because it does not require you to change table schema. You can store data with fixed structures in ApsaraDB for RDS databases, business data with flexible structures in ApsaraDB for MongoDB databases, and frequently accessed data in KVStore for Memcache databases or KVStore for Redis databases. This helps you store data efficiently and reduce costs.

- **Mobile applications**

ApsaraDB for MongoDB supports two-dimensional space indexes. Therefore, it can provide support for location-based mobile application services. ApsaraDB for MongoDB adopts a dynamic storage method that is suitable for storing heterogeneous data from multiple systems. This satisfies the needs of mobile applications.

- **IoT applications**

ApsaraDB for MongoDB provides excellent performance and an asynchronous data writing function. In special scenarios, it can provide in-memory database performance. This makes it extremely suitable for IoT writing scenarios with high concurrency. The MapReduce feature of ApsaraDB for MongoDB can aggregate and analyze large amounts of data.

- **Core log systems**

In scenarios where data is asynchronously written to disks, ApsaraDB for MongoDB can provide excellent data insertion performance and processing capabilities of an in-memory database. ApsaraDB for MongoDB allows you to create secondary indexes for dynamic queries. It can use the MapReduce aggregation framework to perform multidimensional data analysis.

11.6. Limits

Procedure	Limit
Create a database replica	The system automatically creates a three-node replica set. ApsaraDB for MongoDB provides a primary node, a secondary node, and a hidden standby node for each replica set. You cannot create secondary nodes.
Restart a database	You must restart instances in the ApsaraDB for MongoDB console.

11.7. Terms

Term	Description
Region	The geographical location of the server on which the ApsaraDB for MongoDB instance runs. You must specify a region when you create an ApsaraDB for MongoDB instance. The region cannot be changed after the instance is created. When you create an ApsaraDB for MongoDB instance, you must use it with an Alibaba Cloud ECS instance. You can access ApsaraDB for MongoDB instances through internal networks. Make sure that the region of an ApsaraDB for MongoDB instance is the same as that of the corresponding ECS instance.
Instance	An ApsaraDB for MongoDB instance. An instance is the basic unit of ApsaraDB for MongoDB services that you create. An instance is the operating environment for ApsaraDB for MongoDB and exists as a separate process on a host. You can create, modify, and delete an instance in the ApsaraDB for MongoDB console. Instances are independent and their resources are isolated. An instance does not consume resources such as CPU, memory, or I/O of another instance. Each instance has its own features, such as database type and version. The system has parameters to control instance behaviors.
Memory	The maximum memory that an instance can use.
Disk capacity	The disk size that you select when you create an instance. The disk capacity occupied by the instance includes datasets and the space required for normal instance operations, such as the system database, database rollback log, redo log, and indexes. Make sure that the ApsaraDB for MongoDB instance has sufficient disk space to store data. Otherwise, the instance may be locked. If an instance is locked due to insufficient disk capacity, you can purchase a larger disk to unlock the instance.
IOPS	The maximum number of read or write operations performed on block devices per second. Each operation consumes 4 KB.
CPU core	The maximum computing capability of the instance. A single core CPU has a minimum of 2.3 GHz hyper-threading (Intel Xeon series Hyper-Threading) computing power.
Connections	The TCP connections between clients and ApsaraDB for MongoDB instances. If a client uses a connection pool, the connections between the client and instance are persistent connections. Otherwise, they are short connections.

Term	Description
Mongos	The routing service that processes requests. All requests must be coordinated through mongos that serves as a request distribution center and forwards data requests to the corresponding shard server. You can use multiple mongos to process requests. If one fails, other mongos can continue to process the requests.
Config server	The servers that store all database metadata configurations, including routers and shards. mongos does not store but caches shard server information and data routing information in memory. The information is stored on config servers. When you start mongos for the first time or shut it down and then restart it, it automatically loads configuration information from config servers. mongos updates the cache when there are metadata changes. This ensures that mongos can always obtain the correct routing information. Config servers store metadata of shards and routers and have high requirements for service availability and data reliability. Therefore, ApsaraDB for MongoDB uses a three-node replica set to ensure the reliability of the config servers.

11.8. Instance specifications

ApsaraDB for MongoDB replica set specifications

Type	Specification	Code	Maximum connections	Maximum IOPS
General specifications	1 Core - 2 GB	dds.mongo.mid	500	1,000
	2 Core - 4 GB	dds.mongo.standard	1,000	2,000
	4 Core - 8 GB	dds.mongo.large	2,000	4,000
	8 Core - 16 GB	dds.mongo.xlarge	4,000	8,000
	8 Core - 32 GB	dds.mongo.2xlarge	8,000	14,000
	16 Core - 64 GB	dds.mongo.4xlarge	16,000	16,000
Dedicated specifications	2 Core - 16 GB	mongo.x8.medium	2,500	4,500
	4 Core - 32 GB	mongo.x8.large	5,000	9,000
	8 Core - 64 GB	mongo.x8.xlarge	10,000	18,000
	16 Core - 128 GB	mongo.x8.2xlarge	20,000	36,000
	32 Core - 256 GB	mongo.x8.4xlarge	40,000	72,000
Dedicated physical machine	60 Core - 440 GB	dds.mongo.2xmonopolize	100,000	100,000

12.KVStore for Memcache

12.1. What is KVStore for Memcache?

KVStore for Memcache is a memory-based cache service for high-speed access to large amounts of small-size data. KVStore for Memcache can reduce the load on back-end storage services and speed up website and application responses.

KVStore for Memcache supports data in the key-value structure. It can communicate with memcached-compatible clients.

KVStore for Memcache supports out-of-the-box deployment. It also relieves the load on databases from dynamic Web applications and improves website response speed by using the cache service.

Similar to user-created memcached databases, KVStore for Memcache is also compatible with the memcached protocol and user environments. The difference is that the data, hardware infrastructure, network security, and system maintenance services used by KVStore for Memcache are all deployed on the cloud.

12.2. Benefits

Ease of use

- **Out-of-the-box deployment:** Instances are available immediately after creation, facilitating fast business deployment.
- **Compatible with open-source memcached:** KVStore for Memcache is compatible with the memcached binary protocol. All clients that support this protocol and SASL can connect to KVStore for Memcache.
- **Visualized management and monitoring panel:** The console provides several monitoring metrics to facilitate management for Memcache instances.

Cluster features

KVStore for Memcache supports super large capacity and provides super high performance. The cluster output utilizes super large cluster instances to meet demands for large capacity and high performance.

Elastic scalability

- **Scale-out of storage capacity with a single click:** You can adjust the storage capacity of an instance in the console based on your business requirements.
- **Online scale-out without service interruption:** You can adjust the instance capacity without suspending your services or affecting your business.

Resource isolation

Instance-level resource isolation provides enhanced stability for individual services.

High security and reliability

- **Password authentication ensures secure and reliable access.**
- **Persistent data storage:** The use of memory and hard disks can provide high-speed data reading and writing and meet data persistence demands.

High availability

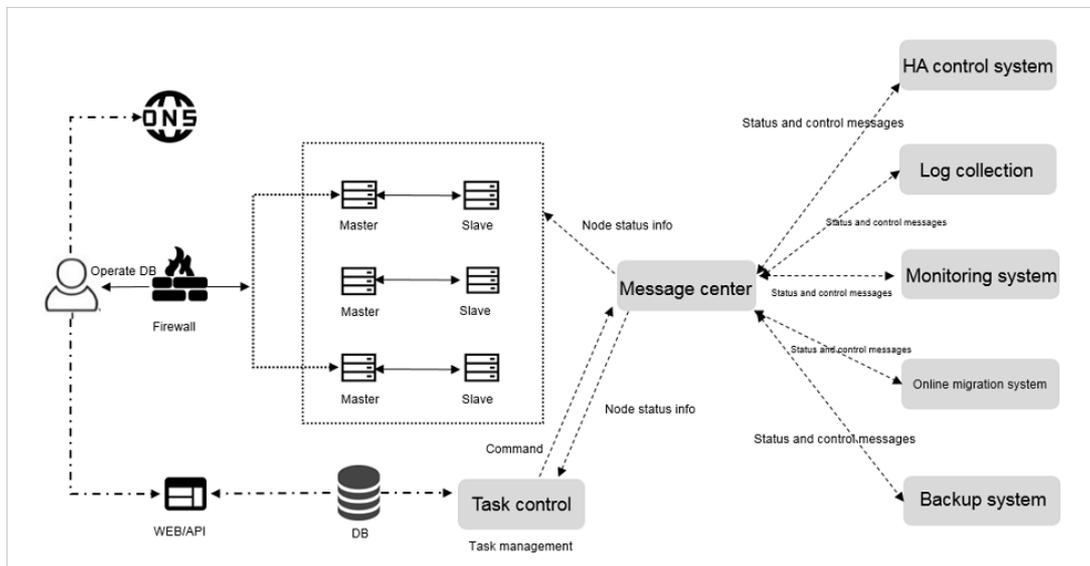
- Each instance has a primary node and a secondary node. This prevents service interruption caused by single point of failures (SPOFs).
- Automatic detection and recovery of hardware faults: KVStore for Memcache automatically detects hardware faults and fails services over within seconds to recover services.

12.3. Architecture

KVStore for Memcache uses a cluster-based architecture embedded with data sharding and reading algorithms. These algorithms can be used to streamline your R&D and O&M processes. Each shard uses a primary/secondary architecture to ensure high availability of services.

KVStore for Memcache consists of three components: proxy servers, partitioning servers, and configuration servers.

Memcache architecture



Proxy server

Each proxy server is configured as a single node. A cluster structure may contain multiple proxy servers, and the system implements load balancing and failover for the proxy servers.

Partitioning server

Each partitioning server is in a high availability dual-node architecture. When the primary node becomes faulty, the system automatically implements the primary/secondary switchover to ensure high availability of services.

Configuration server

Configuration servers are used to store cluster configuration information and partitioning policies. They ensure high availability with their dual-node architecture.

The numbers and specifications of the three components depend on the specifications that you select when creating the cluster instance. The following table describes the specifications.

Specifications	Number of proxy servers	Number of partitioning servers	Memory size of a partitioning server
1 GB	1	1	1 GB
2 GB	1	1	2 GB
4 GB	1	1	4 GB
8 GB	1	1	8 GB
16 GB	2	2	8 GB
32 GB	4	4	8 GB
64 GB	8	8	8 GB
128 GB	16	16	8 GB
256 GB	16	16	16 GB
512 GB	32	32	16 GB

 **Notice** A Memcache cluster provides a uniform domain name for access. You can use this domain name to access and perform data operations on Memcache. The proxy, partitioning, and configuration servers do not provide domain name access and cannot be accessed directly.

12.4. Features

Distributed architecture, freeing businesses from the impact of single point of failure (SPOF)

- KVStore for Memcache uses a distributed cluster architecture. Each node is capable of automatic disaster tolerance and failover and is composed of two servers for hot backup.
- Many different types of KVStore for Memcache with different service requirements and stresses are all able to expand database performance without limits.
- KVStore for Memcache supports data persistence and backup recovery policies. It ensures data reliability and mitigates the impact of physical node faults on back-end databases.

A multi-level security defense system to resist more than 90% of network attacks

- **Anti-DDoS:** monitors inbound traffic in real time. When a large amount of malicious traffic is identified, it scrubs traffic through IP filtering. If traffic scrubbing is ineffective, it triggers the black hole process.
- **IP address whitelist configuration:** A maximum of 1,000 IP addresses can be configured in the whitelist to access an instance, restricting risks from outside of the permitted source.
- **VPC:** KVStore for Memcache is fully compatible with VPCs and can be used to build an isolated network environment on Alibaba Cloud.
- **SASL authentication:** SASL-enabled user identity authentication secures data access.

12.5. Scenarios

Frequently-accessed businesses

Frequently-accessed businesses include social networks, e-businesses, games, and advertisements. Frequently-accessed data can be stored in KVStore for Memcache, while underlying data can be stored in RDS.

Large promotion businesses

Large promotional events and flash sales place systems under high access pressure. Average databases cannot withstand such high read/write pressure, so KVStore for Memcache can act as a viable alternative.

Inventory systems with counters

ApsaraDB for RDS and KVStore for Memcache can be used in combination. RDS stores the specific data and database fields store the specific statistics. KVStore for Memcache reads the statistics, while RDS stores the statistics.

Data analysis businesses

KVStore for Memcache can be used in combination with MaxCompute to analyze and process big data in a distributed manner. It is suitable for big data processing scenarios such as business analysis and data mining. The Data Integration service can simplify data operations by synchronizing data between KVStore for Memcache and MaxCompute.

12.6. Limits

KVStore for Memcache has the following limits.

Item	Limit
Data type	KVStore for Memcache only supports the key-value data format. Complex data types such as array, map, and list are not supported.
Data reliability	The data of KVStore for Memcache is stored in the memory. Cached data cannot be guaranteed against data loss. KVStore for Memcache is not suitable to store the data that requires high consistency.

Item	Limit
Data amount	The maximum size of a key in KVStore for Memcache is 1 KB. The maximum value of a single piece of cached data in KVStore for Memcache is 1 MB. KVStore for Memcache is not suitable to store data of large sizes.
Transaction support	KVStore for Memcache does not support transactions. Data that requires transactions should be written directly to the database.
Scenario	When data access traffic is evenly distributed and there are no obvious hot or cold spots, many access requests cannot reach the cached data of KVStore for Memcache. Therefore, KVStore for Memcache does not effectively function as a database cache. Before you select a database cache, you must consider the data access requirements of the business model.

12.7. Terms

memcached

A high-performance, distributed caching system for memory objects. For more information, see the [official introduction to memcached](#). KVStore for Memcache is compatible with the memcached binary protocol and text protocol.

instance ID

The basic unit used by KVStore for Memcache. KVStore for Memcache imposes different QPS and traffic limits on different instances based on their capacity specifications. You can view your instance IDs in the console.

connection address

The host address used to connect to KVStore for Memcache. It is displayed as a domain name. You can view this address on the [Instance Information](#) page.

connection password

The password used to connect to KVStore for Memcache. You can set the password during instance creation, or reset the password later after the instance is created.

hit ratio

The number of successful reads divided by the total number of reads.

12.8. Instance types

KVStore for Memcache uses a cluster-based architecture. The following table describes the instance types of KVStore for Memcache.

Specifications	Type code	CPU	Number of nodes	Maximum connections	Maximum internal bandwidth	Description
----------------	-----------	-----	-----------------	---------------------	----------------------------	-------------

Specifications	Type code	CPU	Number of nodes	Maximum connections	Maximum internal bandwidth	Description
1 GB	memcache.master.small.default	1 core	1	10,000	10	Primary/secondary dual-node architecture
2 GB	memcache.master.mid.default	1 core	1	10,000	16	Primary/secondary dual-node architecture
4 GB	memcache.master.stand.default	1 core	1	10,000	24	Primary/secondary dual-node architecture
8 GB	memcache.master.large.default	1 core	1	10,000	24	Primary/secondary dual-node architecture
16 GB	memcache.sharding.small.default	2 cores	2	10,000	96	High-performance computing cluster architecture
32 GB	memcache.sharding.mid.default	4 cores	4	40,000	192	High-performance computing cluster architecture
64 GB	memcache.sharding.large.default	8 cores	8	80,000	384	High-performance computing cluster architecture

Specifications	Type code	CPU	Number of nodes	Maximum connections	Maximum internal bandwidth	Description
128 GB	memcache.sharding.2xlarge.default	16 cores	16	160,000	768	High-performance computing cluster architecture
256 GB	memcache.sharding.4xlarge.default	16 cores	16	160,000	768	High-performance computing cluster architecture

13. AnalyticDB for PostgreSQL

13.1. What is AnalyticDB for PostgreSQL?

AnalyticDB for PostgreSQL (formerly known as HybridDB for PostgreSQL) is a distributed analytic database that adopts a massive parallel process (MPP) architecture and consists of multiple compute nodes. AnalyticDB for PostgreSQL provides MPP warehousing services, supports horizontal scaling of storage and compute capabilities, online analysis of petabytes of data, and offline extract, transform, and load (ETL) task processing.

AnalyticDB for PostgreSQL is developed based on the PostgreSQL kernel and has the following features:

- Supports the standard query syntax of SQL 2008, OLAP aggregate functions, views, Procedural Language for SQL (PL/SQL), user-defined functions (UDF), and triggers. AnalyticDB for PostgreSQL is partially compatible with the Oracle syntax.
- Uses the MPP architecture that can be horizontally scaled and supports range and list partitioning.
- Supports row store, column store, and multiple indexes. Supports multiple compression strategies based on column store. This reduces storage costs.
- Supports standard database isolation levels and distributed transactions. This ensures data consistency.
- Provides the vector computing engine and the CASCADE-based SQL optimizer. This ensures high-performance SQL analysis capabilities.
- Supports the primary/secondary architecture. This ensures dual-replica data storage.
- Provides online scaling, monitoring, and disaster recovery. This helps reduce O&M costs.

13.2. Benefits

 Real-time analysis	<p>Supports the MPP architecture that can be horizontally scaled, allowing you to query petabytes of data in seconds. AnalyticDB for PostgreSQL supports the leading vector computing feature and intelligent indexes of column store. It also supports the CASCADE-based SQL optimizer to make complex queries without the need for tuning.</p>
 Stability and reliability	<p>Provides ACID properties for distributed transactions. Transactions are consistent among nodes and all data are synchronized between primary and secondary nodes. It supports distributed deployment and provides transparent monitoring, switching, and restoration to secure your data infrastructure.</p>
 Easy to use	<p>Supports a large number of SQL syntax and functions, Oracle functions, stored procedures, user-defined functions (UDF), and isolation levels of transactions and databases. You can use popular BI software and ETL tools online.</p>
 Ultra-high performance	<p>Supports row store, column store, and multiple indexes. The vector engine provides high-performance analysis and computing capabilities. The CASCADE-based SQL optimizer enables complex queries without the need for tuning. It supports high-performance parallel import of data from OSS.</p>



Enables you to scale up segments, CPU, memory, and storage resources on demand to improve OLAP performance.

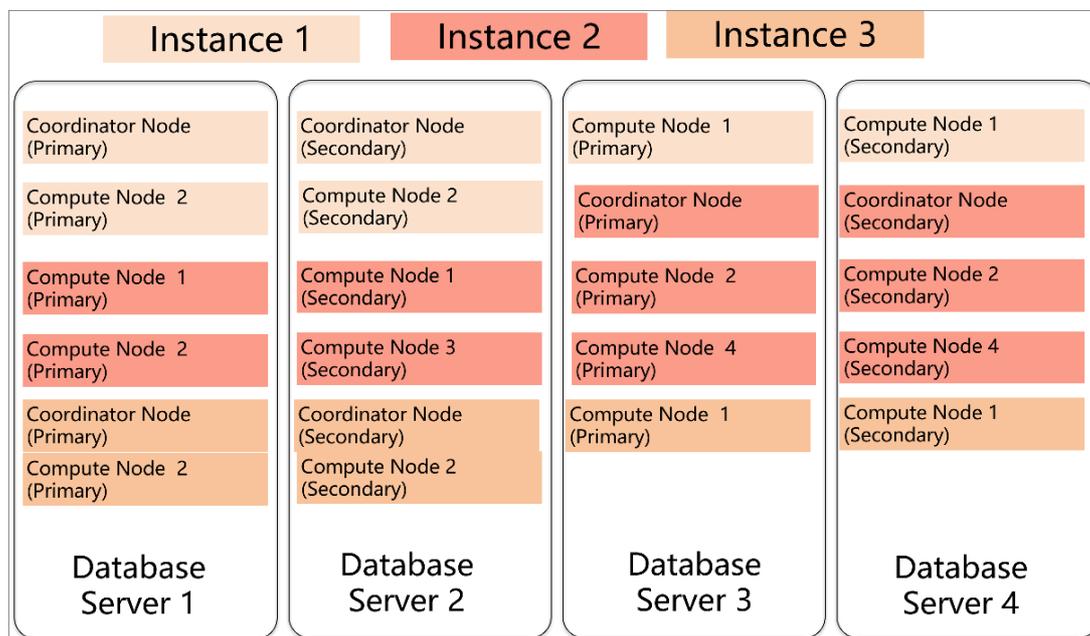
Supports transparent OSS operations. OSS offers a larger storage capacity for cold data that does not require online analysis.

13.3. Architecture

Physical cluster architecture

The following figure shows the physical cluster architecture of AnalyticDB for PostgreSQL.

Physical cluster architecture



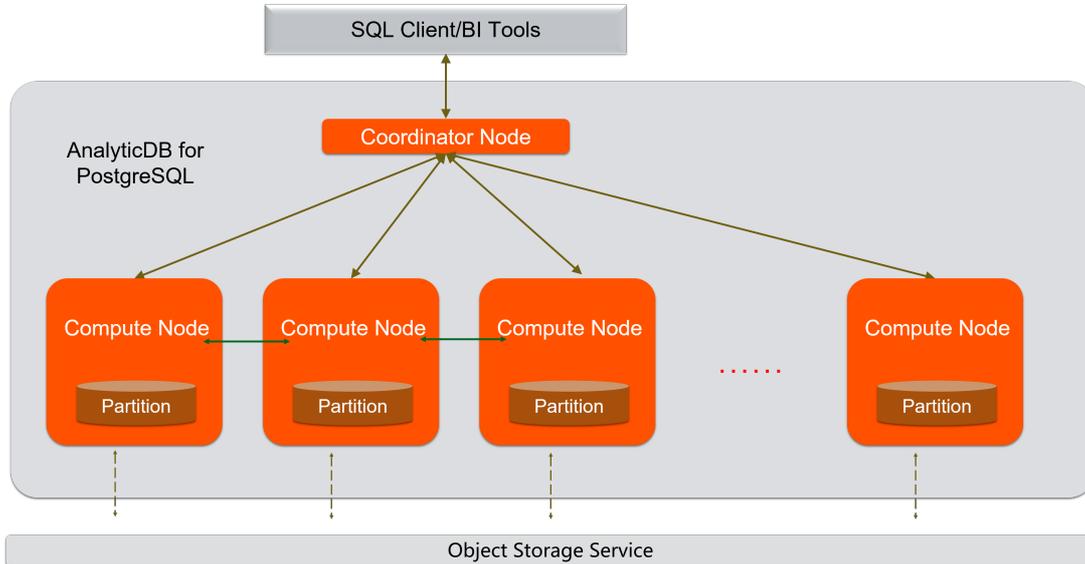
You can create multiple instances in a physical cluster of AnalyticDB for PostgreSQL. Each cluster includes two components: the master and the segment.

- The master is used to access applications. It receives connection requests and SQL query requests from clients and dispatches computing tasks to segments. The cluster deploys a secondary node of the master on an independent physical server and replicates data from the primary node to the secondary node for failover. The secondary node does not accept external connections.
- Segments are independent instances in AnalyticDB for PostgreSQL. Data is evenly distributed among segments by hash value or RANDOM function, and is analyzed and computed among segments in parallel. Each segment consists of a primary node and a secondary node for automatic failover.

Logical architecture of an instance

You can create multiple instances in a cluster of AnalyticDB for PostgreSQL. The following figure shows the logical architecture of an instance.

Logical architecture of an instance



Data is distributed among segments by hash value or RANDOM function of a specified distributed column. Each segment consists of a primary node and a secondary node to ensure dual-replica storage. High-performance network communication is supported among nodes. When the master receives a request from the application, the master parses and optimizes SQL statements to generate a distributed execution plan. After the master sends the execution plan to the segments, the segments perform an MPP execution of the plan.

13.4. Features

Distribution

- **MPP architecture**

AnalyticDB for PostgreSQL is based on the Massively Parallel Processing (MPP) architecture. The storage is extended linearly and computing capabilities are enhanced by adding more compute groups, which leverage the OLAP computing performance of each compute group.

- **Distributed transactions**

Supports distributed SQL OLAP and window functions, distributed PL/pgSQL stored procedures and triggers, and enables databases to support distributed computing.

Learning and analysis

- **MADlib machine learning**

Provides a large number of SQL-based machine learning tools for data science users and is built in with more than 50 machine learning algorithms.

- **GIS-based geographic analysis**

Supports hybrid geographic data analysis that complies with the OpenGIS specifications, and enables you to use a single SQL statement to analyze a large amount of geographic data, such as population flow, area statistics, and traces.

Data interconnection

- **Heterogeneous data import**

Imports data from MySQL databases by using the mysql2pgsql tool. You can use popular ETL tools to import data to AnalyticDB databases through the ETL process.

- **OSS heterogeneous data storage**

Uses standard SQL syntax to query structured files stored in OSS by using external tables in real time.

- **Transparent data replication**

Replicates data transparently from RDS for PostgreSQL or RDS for PPAS without the need to program for consecutive incremental replication. This feature simplifies maintenance, and allows high-performance internal modeling and data cleansing for the imported data.

Security

- **IP address whitelist**

You can add up to 1,000 IP addresses to the whitelist of AnalyticDB for PostgreSQL. This feature allows you to control risks from sources of access.

- **Anti-DDoS**

Monitors inbound traffic in real time, scrubs large amounts of malicious traffic by filtering source IP addresses, and throws affected servers into a black hole when traffic scrubbing becomes inefficient.

13.4.1. Distributed architecture

AnalyticDB for PostgreSQL is based on the MPP architecture. Data is distributed evenly among nodes by hash value or RANDOM function, and is analyzed and computed among nodes in parallel. The storage and computing capacities are scaled horizontally as more nodes are added. This ensures a quick response when the data volume increases.

AnalyticDB for PostgreSQL supports distributed transactions to ensure data consistency among nodes. It supports three transaction isolation levels: SERIALIZABLE, READ COMMITTED, and READ UNCOMMITTED.

13.4.2. High-performance data analysis

AnalyticDB for PostgreSQL supports column store and row store for tables. Row store provides high update performance and column store provides high OLAP aggregate analysis performance for tables. AnalyticDB for PostgreSQL supports the B-tree index, bitmap index, and hash index that enable high-performance analysis, filtering, and query.

AnalyticDB for PostgreSQL adopts the CASCADE-based SQL optimizer. AnalyticDB for PostgreSQL combines the cost-based optimizer (CBO) and the rule-based optimizer (RBO) to provide SQL optimization features such as automatic subquery decorrelation. These features enable complex queries without the need for tuning.

13.4.3. High-availability service

AnalyticDB for PostgreSQL builds a system for automatic monitoring, diagnosis, and error handling based on the Apsara platform of Alibaba Cloud, which helps to reduced O&M costs.

The master stores database metadata and receives query requests from clients to compile and optimize SQL statements. The master adopts a primary/secondary architecture to ensure strong consistency of metadata. If the primary master fails, the service is automatically switched to the secondary master.

All segments adopt a primary/secondary architecture to ensure strong data consistency between primary and secondary nodes when data is written into or updated. If the primary segment fails, the service is automatically switched to the secondary segment.

13.4.4. Data synchronization and tools

You can use Data Transmission Service (DTS) or DataWorks to synchronize data from MySQL or PostgreSQL databases to AnalyticDB for PostgreSQL. Popular extract, transform, and load (ETL) tools can import ETL data and schedule jobs on AnalyticDB for PostgreSQL databases. You can also use standard SQL syntax to query data from formatted files stored in OSS by using external tables in real time.

AnalyticDB for PostgreSQL supports Business Intelligence (BI) reporting tools, including Quick BI, DataV, Tableau, and FineReport. It also supports ETL tools, including Informatica and Kettle.

13.4.5. Data security

AnalyticDB for PostgreSQL supports IP whitelist configuration. You can add IP addresses of up to 1,000 servers that are allowed to access your instance to the whitelist. This enables you to control risks from the access source. AnalyticDB for PostgreSQL also supports Anti-DDoS that monitors inbound traffic in real time. When a large amount of malicious traffic is identified, it scrubs traffic through IP filtering. If traffic scrubbing is ineffective, it triggers the black hole process.

13.4.6. Supported SQL features

- Supports row store and column store.
- Supports multiple indexes, including the B-tree index, bitmap index, and hash index.
- Supports distributed transactions and standard isolation levels, which ensure data consistency among nodes.
- Supports character, date, and arithmetic functions.
- Supports stored procedures, user-defined functions (UDF), and triggers.
- Supports views.

- Supports range partitioning, list partitioning, and the definition of multi-level partitions.
- Supports multiple data types. The following table provides a list of data types and their information.

Data type	Alias	Storage	Range	Description
bigint	int8	8 bytes	- 922337203685477 5808 to 922337203685477 5807	Large-range integer
bigserial	serial8	8 bytes	1 to 922337203685477 5807	Large auto- increment integer
bit [(n)]	N/A	n bits	Bit string constant	Fixed-length bit string
bit varying [(n)]	varbit	Variable-length bit string	Bit string constant	Variable-length bit string
boolean	bool	1 byte	true/false, t/f, yes/no, y/n, 1/0	Boolean value (true/false)
box	N/A	32 bytes	((x1,y1),(x2,y2))	A rectangular box on a plane, not allowed in distribution key columns
bytea	N/A	1 byte + binary string	Sequence of octets	Variable-length binary string
character [(n)]	char [(n)]	1 byte + n	String up to n characters in length	Fixed-length, blank-padded string
character varying [(n)]	varchar [(n)]	1 byte + string size	String up to n characters in length	Variable length with limit
cidr	N/A	12 or 24 bytes	N/A	IPv4 and IPv6 networks
circle	N/A	24 bytes	<(x,y),r> (center and radius)	A circle on a plane, not allowed in distribution key columns
date	N/A	4 bytes	4713 BC to 294,277 AD	Calendar date (year, month, day)

Data type	Alias	Storage	Range	Description
decimal [(p, s)]	numeric [(p, s)]	variable	No limit	User-specified precision, exact
double precision	float8	8 bytes	15 decimal digits of precision	Variable precision, inexact
	float			
inet	N/A	12 or 24 bytes	N/A	IPv4 and IPv6 hosts and networks
integer	int or int4	4 bytes	-2.1E+09 to +2147483647	Typical choice for integer
interval [(p)]	N/A	12 bytes	-178000000 years to 178000000 years	Time span
json	N/A	1 byte + json size	JSON string	Unlimited variable length
lseg	N/A	32 bytes	((x1,y1),(x2,y2))	A line segment on a plane, not allowed in distribution key columns
macaddr	N/A	6 bytes	N/A	Media Access Control (MAC) addresses
money	N/A	8 bytes	-92233720368547758.08 to +92233720368547758.07	Currency amount
path	N/A	16+16n bytes	[(x1,y1),...]	A geometric path on a plane, not allowed in distribution key columns
point	N/A	16 bytes	(x,y)	A geometric point on a plane, not allowed in distribution key columns

Data type	Alias	Storage	Range	Description
polygon	N/A	40+16n bytes	((x1,y1),...)	A closed geometric path on a plane, not allowed in distribution key columns
real	float4	4 bytes	6 decimal digits of precision	Variable precision, inexact
serial	serial4	4 bytes	1 to 2147483647	Auto-increment integer
smallint	int2	2 bytes	-32768 to +32767	Small-range integer
text	N/A	1 byte + string size	Variable-length string	Unlimited variable length
time [(p)] [without time zone]	N/A	8 bytes	00:00:00[.000000] to 24:00:00[.000000]	Time of day (without time zone)
time [(p)] with time zone	timetz	12 bytes	00:00:00+1359 to 24:00:00-1359	Time of day (with time zone)
timestamp [(p)] [without time zone]	N/A	8 bytes	4713 BC to 294,277 AD	Date and time
timestamp [(p)] with time zone	timestamptz	8 bytes	4713 BC to 294,277 AD	Date and time (with time zone)
xml	N/A	1 byte + XML size	Variable-length XML string	Unlimited variable length

13.5. Scenarios

AnalyticDB for PostgreSQL is applicable to the following OLAP data analysis services.

- ETL for offline data processing

AnalyticDB for PostgreSQL provides the following benefits that make it ideal to optimize complex SQL queries and aggregate and analyze huge amounts of data:

- Supports standard SQL, OLAP window functions, and stored procedures.
- Provides the CASCADE-based SQL optimizer to make complex queries without the need for tuning.
- Uses the MPP architecture that can be horizontally scaled and can process petabytes of data in seconds.
- Provides column store-based high-performance storage and aggregation of large tables and high compression ratio to save storage space.

- **Online high-performance query**

AnalyticDB for PostgreSQL provides the following benefits for real-time exploration, warehousing, and updating of data:

- Allows you to write and update high-throughput data through INSERT, UPDATE, and DELETE operations.
- Allows you to query data based on row store and multiple indexes (B-tree, bitmap, and hash) to obtain results in milliseconds.
- Supports distributed transactions, standard database isolation levels, and HTAP.

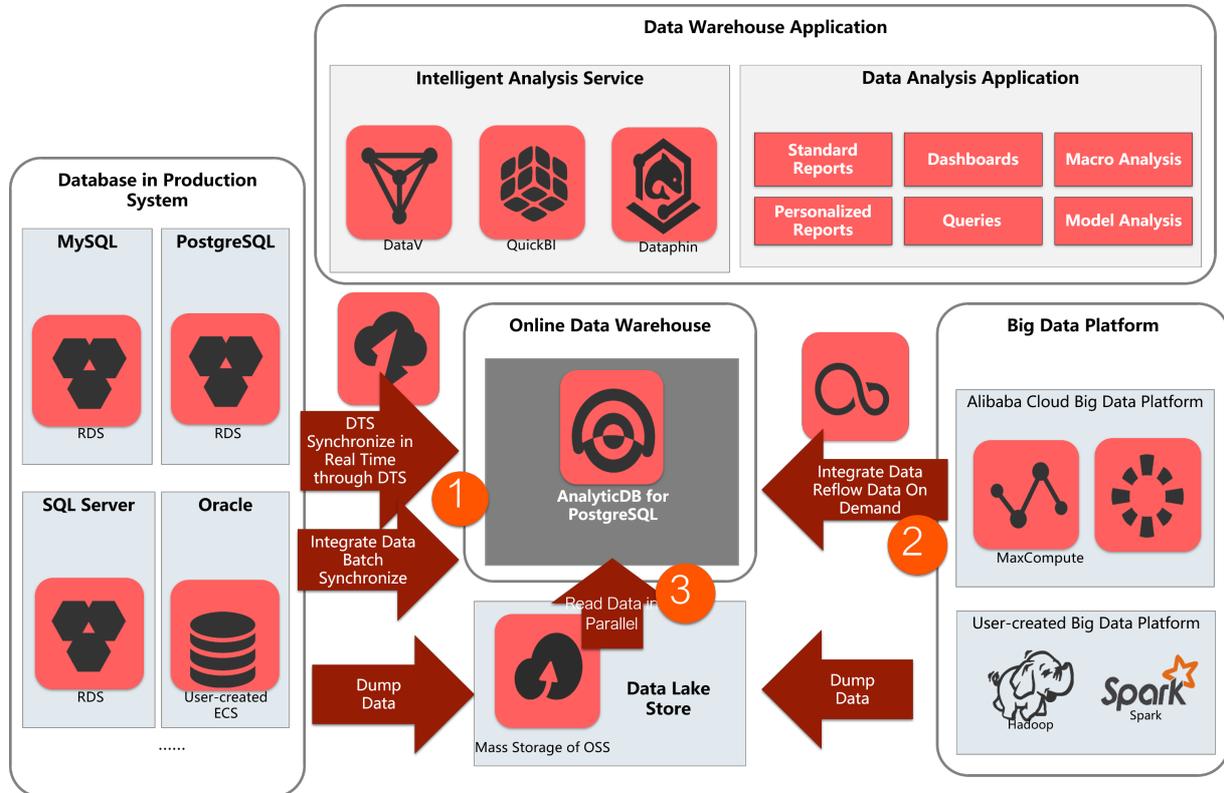
- **Multi-model data analysis**

AnalyticDB for PostgreSQL provides the following benefits for processing a variety of unstructured data sources:

- Supports the PostGIS extension for geographic data analysis and processing.
- Takes advantage of the MADlib extension, a library of in-database machine learning algorithms, to implement an AI-native database.
- Provides high-performance retrieval and analysis of unstructured data such as images, speeches, and texts through vector retrieval.
- Supports formats such as JSON. It can also process and analyze semi-structured data such as logs.

Typical scenarios

AnalyticDB for PostgreSQL is applicable to the three following scenarios:



- **Data warehousing service**

Data Transmission Service (DTS) can synchronize data in real time in production system databases such as ApsaraDB RDS for MySQL, ApsaraDB RDS for PostgreSQL, ApsaraDB for POLARDB, and traditional databases such as Oracle and SQL Server. Data can also be synchronized in batches to AnalyticDB for PostgreSQL through the data integration service (DataX). AnalyticDB for PostgreSQL supports Extract, Transform, and Load (ETL) operations on large amounts of data. You can also use DataWorks to schedule these tasks. AnalyticDB for PostgreSQL also provides high-performance online analysis capabilities and can use Quick BI, DataV, Tableau, and FineReport for report presentation and real-time query.

- **Big data analytics platform**

To perform high-performance analysis, processing, and exploration, you can import huge amounts of data from MaxCompute, Hadoop, and Spark to AnalyticDB for PostgreSQL through DataX or OSS.

- **Data lake analytics**

AnalyticDB for PostgreSQL can use an external table mechanism to access the huge amounts of data stored in OSS in parallel and build an Alibaba Cloud data lake analytics platform.

14.Data Transmission Service (DTS)

14.1. What is DTS?

Data Transmission Service (DTS) is a data service provided by Alibaba Cloud that supports data exchange between relational databases, OLAP databases, and other data sources.

DTS supports data migration, real-time data subscription, and real-time data synchronization. DTS can be used in multiple business scenarios, including interruption-free data migration, geo-disaster recovery, cross-border data synchronization, and cache update policies, helping you build a secure, scalable, and highly available data architecture.

- DTS aims to help you with complex data interactions so that you can focus on upper-layer service development.
- DTS supports the following data sources:
 - Relational databases: MySQL and Oracle
 - OLAP databases: MaxCompute

14.2. Benefits

DTS supports transmitting data between data sources such as relational databases and OLAP databases. DTS provides you with multiple data transmission methods such as data migration, real-time data subscription, and real-time data synchronization. Compared with other third-party data migration and synchronization tools, DTS provides multiple transmission channels with high performance, security, and reliability. DTS also makes it easy to create and manage transmission channels.

Diverse transmission methods

DTS supports multiple data transmission features, including data migration, data subscription, and data synchronization. In data subscription and data synchronization, data is transmitted in real time.

Data migration enables you to migrate data between databases without interrupting application operations. The application service downtime during data migration is reduced to minutes.

High performance

DTS uses servers with high specifications to ensure high data transmission performance for each synchronization or migration channel.

At the underlying layer, multiple measures are taken to improve DTS performance.

Compared with traditional data synchronization tools, the real-time synchronization feature of DTS enables you to concurrently transmit transactions. It also allows you to synchronize table data you want to update at a time. This greatly improves synchronization performance.

High security and reliability

DTS is implemented using clusters. If a node in a cluster is down or faulty, the control center quickly moves all tasks from this node to another healthy node in the cluster.

DTS provides a 24 x 7 mechanism for validating data accuracy in some transmission channels to quickly locate and correct incorrect data. This helps ensure reliable data transmission.

Secure transmission protocols and tokens are used for authentication across DTS modules to ensure reliable data transmission.

Easy-to-use

The DTS console is a visual management interface that provides a wizard-like process to assist you in creating data transmission channels.

You can also view data transmission information in the DTS console, including the transmission status, progress, and performance, to better manage the transmission channels.

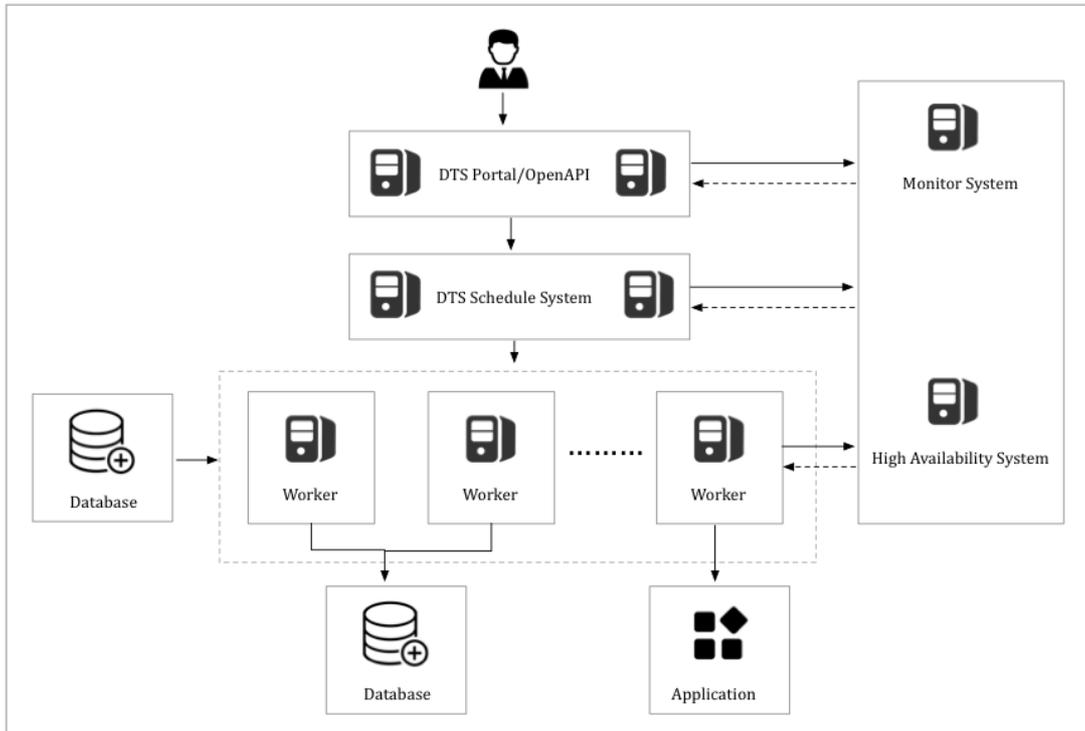
DTS supports resumable transmission, and regularly monitors channel status to avoid interruptions resulting from network or system exceptions. When DTS detects a channel exception, it automatically repairs or restarts the channel. In cases where manual operations are needed, you can directly repair the channel and restart it in the DTS console.

14.3. Architecture

System architecture

System architecture shows the system architecture of DTS.

System architecture



- **High availability**

Each DTS module comes with a primary-secondary architecture to ensure high availability of the system. The disaster recovery module runs a health check on each node in real time. Once a node exception is detected, the module switches the channel to another healthy node within seconds.

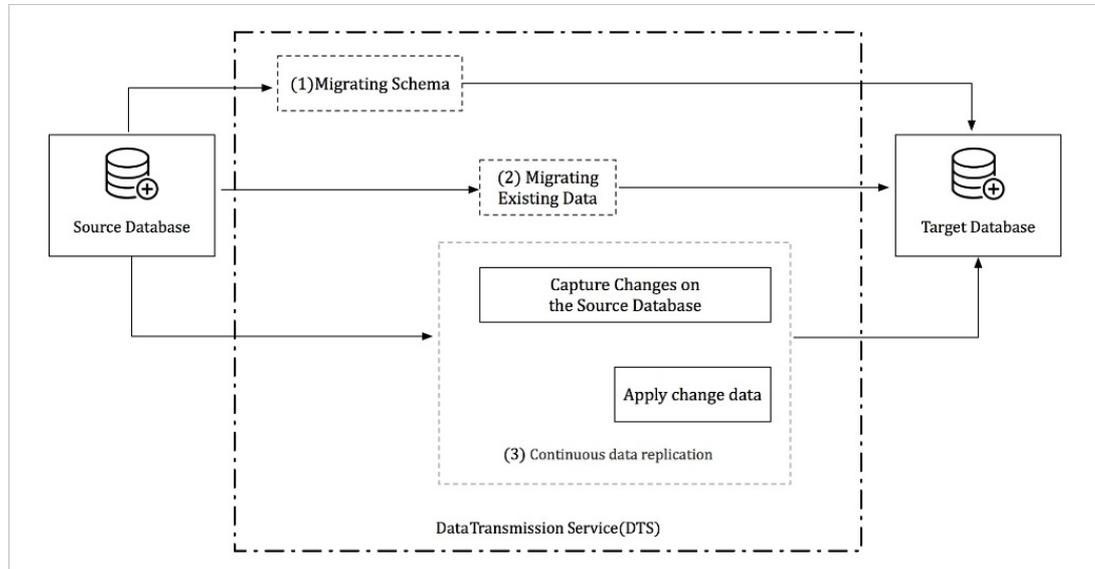
- **Monitor changes in the data source IP address**

For data subscription and synchronization channels, the disaster recovery module checks for any changes. For example, once it detects a change in the data source address, the module dynamically changes the method for connecting to the data source to ensure channel stability.

Data migration process

Data migration workflow shows how data migration works.

Data migration workflow



Data migration supports schema migration, real-time full data migration, and real-time incremental data migration. To implement migration without service interruption, follow these steps:

1. Schema migration
2. Full data migration
3. Incremental data migration

For migration between heterogeneous databases, DTS reads the schema using the syntax of the source database, translates the schema into the syntax of the destination database, and then imports the schema to the destination instance.

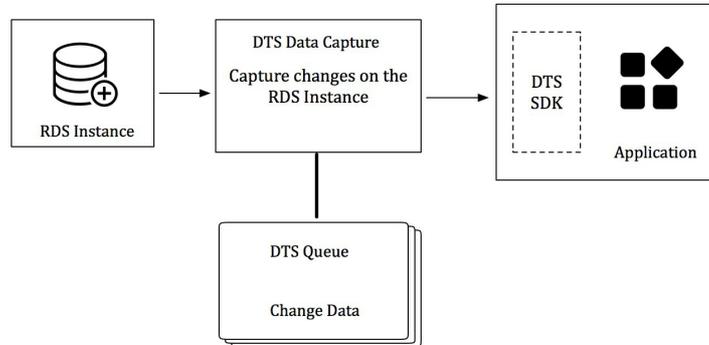
Full data migration takes a longer time. In this process, new data is continuously written into the source instance. To ensure data consistency, DTS starts the incremental data pulling module before full data migration. This module pulls the incremental data from the source instance and then parse, encapsulate, and store the data locally.

When full data migration is complete, DTS starts the incremental data playback module. The module retrieves the incremental data from the incremental data pulling module. After reverse parsing, filtering, and encapsulation, the data is synchronized to the destination instance. Eventually, data is synchronized between the source and destination instances in real time.

Data subscription process

Data subscription workflow shows how data subscription works.

Data subscription workflow



Data subscription supports pulling incremental data from the RDS instance in real time. You can subscribe to the incremental data on the data subscription server using the DTS SDK. You can also customize data consumption based on business requirements.

The data pulling module of the DTS server captures raw data from the data source, and makes the incremental data locally persistent by parsing, filtering, and formatting it.

The data capturing module connects to the source instance using the database protocol and pulls the incremental data from the source instance in real time. For example, the data capturing module connects to an RDS for MySQL instance using the binlog dump command.

DTS guarantees the high availability of the data pulling module and downstream consumption SDKs.

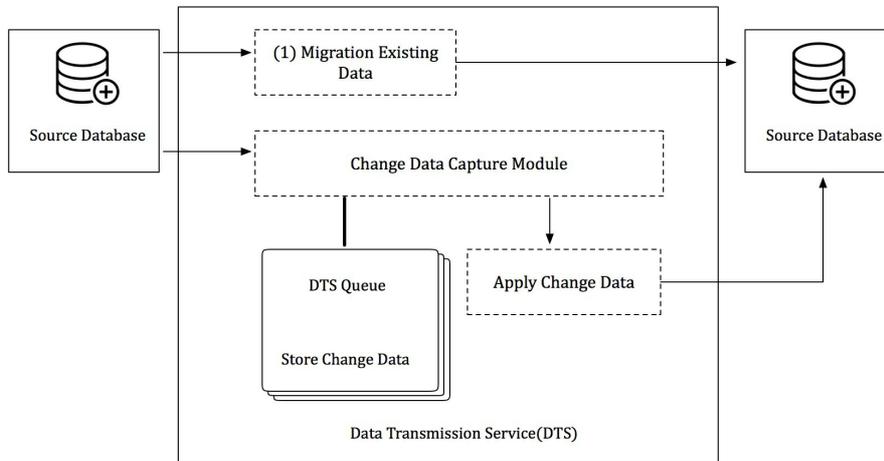
To ensure the high availability of the data pulling module, the DTS disaster recovery module restarts the data pulling module on a healthy service node once an exception is detected in the data pulling module.

The DTS server ensures the high availability of downstream consumption SDKs. If you start multiple consumption SDKs for the same subscription channel, the server pushes the incremental data to only one SDK at a time. If the consumption encounters an exception, the service end selects another consumption process from other healthy downstream nodes to push data to that consumption process. In this way, the high availability of downstream consumption processes can be guaranteed.

Real-time synchronization workflow

Real-time synchronization workflow shows how real-time synchronization works.

Real-time synchronization workflow



The data synchronization feature in DTS enables real-time synchronization of incremental data between any two RDS instances.

To create a synchronization channel, follow these steps:

- **Initial synchronization:** The existing data in the source instance is synchronized to the destination instance.
- **Incremental data synchronization:** After initial synchronization, the incremental data starts to be synchronized between the source instance and destination instance in real time. During this phase, data is eventually synchronized between the source and destination instances.

DTS provides the following underlying modules for real-time incremental data synchronization:

- **Data reading module**
The data reading module reads raw data from the source instance and makes the data locally persistent by parsing, filtering, and formatting it. The data reading module connects to the source instance using the database protocol and reads the incremental data from the source instance. For example, the data reading module connects to an RDS for MySQL instance using the binlog dump command.
- **Data playback module**
The data playback module requests incremental data from the data reading module, filters data based on the objects to be synchronized, and then synchronizes the data to the destination instance without compromising the transaction sequence and consistency.

DTS ensures the high availability of the data reading module and data playback module. When a channel exception is detected, the disaster recovery module restarts the channel on a healthy service node. In this way, the high availability of the synchronization channels is guaranteed.

14.4. Features

14.4.1. Data migration

Data migration allows you to migrate data between multiple data sources. Typical scenarios include data migration to the cloud, data migration between instances within Apsara Stack, and database sharding and scaling. DTS supports data migration between homogeneous and heterogeneous data sources. It also supports ETL features such as data mapping at three levels (databases, tables, and columns) and data filtering.

Data sources

DTS supports migrating data between multiple data sources. [Data sources supported by each data migration type](#) lists the data sources supported by each data migration type.

Data sources supported by each data migration type

Data source	Schema migration	Full data migration	Incremental data migration
MySQL > RDS for MySQL	Supported	Supported	Supported
MySQL > Oracle	Not supported	Supported	Supported
Oracle > RDS for MySQL	Supported	Supported	Supported

DTS supports migrating data from the following data sources:

- RDS instances
- Oracle databases
- On-premises databases

DTS supports migrating data to the following data sources:

- RDS instances

Online migration

DTS uses online migration. You only need to configure the source instance, destination instance, and objects to be migrated. DTS automatically completes the entire data migration process. To minimize the impact of online data migration on your services, you must ensure that the DTS server can connect to both the source and destination instances.

Data migration types

DTS supports schema migration, full data migration, and incremental data migration.

- **Schema migration:** migrates schemas from the source instance to the destination instance.
- **Full data migration:** migrates historical data from the source instance to the destination instance.
- **Incremental data migration:** migrates incremental data generated during migration from the source instance to the destination instance in real time. You can select schema migration, full data migration, and incremental migration to migrate data with minimal downtime.

ETL features

Data migration supports the following ETL features:

- Object name mappings of databases, tables, and columns. With this feature, you can migrate data between two databases, tables, or columns that have different names.
- Data filtering. With this feature, you can use SQL conditions to filter the required data in a specific table. For example, you can specify a time range to migrate only the latest data.

Alerts

Data migration supports sending alerts when errors occur. When an error occurs, DTS immediately sends an SMS alert to the task owner, allowing the owner to promptly handle the exception.

Migration task

A migration task is a basic unit of data migration. To migrate data, you must create a data migration task in the DTS console. To create a migration task, you must configure the required information such as the connection type of the source and destination instances, the migration type, and the objects to be migrated. You can create, manage, stop, and delete data migration tasks in the DTS console.

Task statuses describes possible statuses of a migration task.

Task statuses

Task status	Description	Available operations
Not started	The migration task configurations have been completed, and the precheck is not performed.	<ul style="list-style-type: none"> • Perform the precheck • Delete the migration task.
Prechecking	A precheck is being performed before the migration task starts.	Delete the migration task.
Precheck passed	The migration task has passed the precheck and has not started.	<ul style="list-style-type: none"> • Start the migration task. • Delete the migration task.

Task status	Description	Available operations
Migrating	Data is being migrated.	<ul style="list-style-type: none"> • Pause the migration task. • Stop the migration task. • Delete the migration task.
Migration failed	An error occurred during migration. You can determine the specific phase in which the migration failed based on the progress of the migration task.	Delete the migration task.
Paused	The migration task is paused.	<ul style="list-style-type: none"> • Start the migration task. • Delete the migration task.
Completed	The migration task is complete, or you have clicked End to stop data migration.	Delete the migration task.

14.4.2. Data synchronization

Real-time data synchronization enables you to synchronize data between two data sources in real time.

This feature applies to multiple scenarios, such as active geo-redundancy, geo-disaster recovery, local disaster recovery, cross-border data synchronization, data query, data streaming for reports, cloud BI systems, and real-time data warehousing.

Synchronization features

Synchronization features lists the synchronization features supported by DTS.

Synchronization features

Source instance	Destination instance	One-way/two-way synchronization
MySQL	MySQL	<ul style="list-style-type: none"> One-way synchronization Two-way synchronization
MySQL	MaxCompute	One-way synchronization
MySQL	DataHub	One-way synchronization

Synchronization objects

- Data synchronization objects include databases, tables, and columns. You can specify one or more tables that you want to synchronize.
- Destination database names, table names, and column names can be different from those on the source side. This enables you to synchronize data between two different databases or tables.
- To meet special business requirements, you can specify the columns that you want to synchronize.

Synchronization tasks

A synchronization task is a basic unit of real-time data synchronization. To synchronize data between two instances, you must create a synchronization task in the DTS console.

Task statuses describes possible statuses of a synchronization task.

Task statuses

Task status	Description	Available operations
Prechecking	A precheck is being performed before the synchronization task starts.	<ul style="list-style-type: none"> View configurations for synchronization. Delete the synchronization task. Replicate configurations for synchronization. Set monitoring and alerting.

Task status	Description	Available operations
Precheck failed	The synchronization task has failed to pass the precheck.	<ul style="list-style-type: none"> • Perform the precheck. • View configurations for synchronization. • Modify synchronization objects. • Modify the synchronization speed. • Delete the synchronization task. • Replicate configurations for synchronization. • Set monitoring and alerting.
Not started	The synchronization task has passed the precheck and has not started.	<ul style="list-style-type: none"> • Perform the precheck. • Start the synchronization task. • Modify synchronization objects. • Modify the synchronization speed. • Delete the synchronization task. • Replicate configurations for synchronization. • Set monitoring and alerting.
Performing initial synchronization	The initial synchronization is in progress.	<ul style="list-style-type: none"> • View configurations for synchronization. • Delete the synchronization task. • Replicate configurations for synchronization. • Set monitoring and alerting.
Initial synchronization failed	Data migration has failed during initial synchronization.	<ul style="list-style-type: none"> • View configurations for synchronization. • Modify synchronization objects. • Modify the synchronization speed. • Delete the synchronization task. • Replicate configurations for synchronization. • Set monitoring and alerting.

Task status	Description	Available operations
Synchronizing	The task is synchronizing data.	<ul style="list-style-type: none"> • View configurations for synchronization. • Modify synchronization objects. • Modify the synchronization speed. • Pause the synchronization task. • Delete the synchronization task. • Replicate configurations for synchronization. • Set monitoring and alerting.
Synchronization failed	An error occurred during synchronization.	<ul style="list-style-type: none"> • View configurations for synchronization. • Modify synchronization objects. • Modify the synchronization speed. • Start the synchronization task. • Delete the synchronization task. • Replicate configurations for synchronization. • Set monitoring and alerting.
Paused	The synchronization task is paused.	<ul style="list-style-type: none"> • View configurations for synchronization. • Modify synchronization objects. • Modify the synchronization speed. • Start the synchronization task. • Delete the synchronization task. • Replicate configurations for synchronization. • Set monitoring and alerting.

Advanced features

You can use the following advanced features to facilitate data synchronization:

- Add and remove synchronization objects

You can add and remove synchronization objects during data synchronization.

- View the synchronization performance

Data synchronization provides diagrams for analyzing the latency, RPS, and traffic statistics of synchronization tasks. This allows you to easily view the performance trend of a synchronization channel.

- Set the monitoring threshold

Data synchronization monitors the status and latency of synchronization tasks and sends an alert if the predefined threshold is reached. You can set the threshold for synchronization latency alerts based on the sensitivity of your businesses to data synchronization latency.

14.4.3. Data subscription

Real-time data subscription is designed to help users retrieve incremental data from Relational Database Service (RDS) in real time. In business scenarios such as cache update, asynchronous decoupling, real-time data synchronization between heterogeneous data sources, and real-time data synchronization with complex ETL, you may choose to consume the incremental data as required.

Features

- Supports data subscription for RDS for MySQL instances in classic and VPC networks.

Data source types

Real-time data subscription supports the following data sources:

- RDS for MySQL

Objects to be subscribed

Objects to be subscribed include databases and tables. You can subscribe to the incremental data of specified tables as needed.

In data subscription, incremental data is further divided into data updates (DML) and schema updates (DDL). When you configure data subscription, you can select a data change type as needed.

Subscription channels

Subscription channels are used for incremental data subscription and consumption. To subscribe to incremental data of an RDS instance, you must create a subscription channel in the DTS console for this instance. The subscription channel reads incremental data in the RDS instance in real time and stores the most recent increments. You can use the SDK provided by DTS to subscribe to and consume the incremental data in the channel. You can also create, manage, and delete subscription channels in the DTS console.

Data in a subscription channel can only be subscribed and consumed using one SDK. To subscribe to an RDS instance for multiple downstream SDKs, you must create an equivalent number of subscription channels. RDS instances subscribed by these channels share the same instance ID.

[Subscription channel status and descriptions](#) shows the different status of a subscription channel during its lifecycle.

Subscription channel status and descriptions

Channel status	Description	Available actions
Prechecking	The subscription channel has been configured and a preliminary check is being performed.	Delete the subscription channel.
Not started	The migration task has passed the precheck, but is not started.	<ul style="list-style-type: none"> Start the subscription channel. Delete the subscription channel.
Initial subscription	Initial subscription is being enabled and takes about one minute.	Delete the subscription channel.
Running	The subscription channel is reading incremental data from an RDS instance.	<ul style="list-style-type: none"> View sample code. View the subscribed data. Delete the subscription channel.
Error	An exception occurs when the subscription channel reads incremental data from an RDS instance.	<ul style="list-style-type: none"> View sample code. Delete the subscription channel.

Advanced features

The following features are used in data subscription:

- Add and remove objects to be subscribed as required

You can add or remove the objects to be subscribed during a data subscription.

- View the subscribed data online

You can view the subscribed incremental data in the DTS console.

- Modify the data consumption checkpoint

You can modify the time for data consumption at any time.

- Complete monitoring system

Data subscription monitors the subscription channel status and reports an alert when the threshold for downstream consumption delays is reached. You can set the alert threshold according to business sensitivity.

14.5. Scenarios

DTS supports multiple features including data migration, real-time data subscription, and real-time data synchronization to meet the following scenarios.

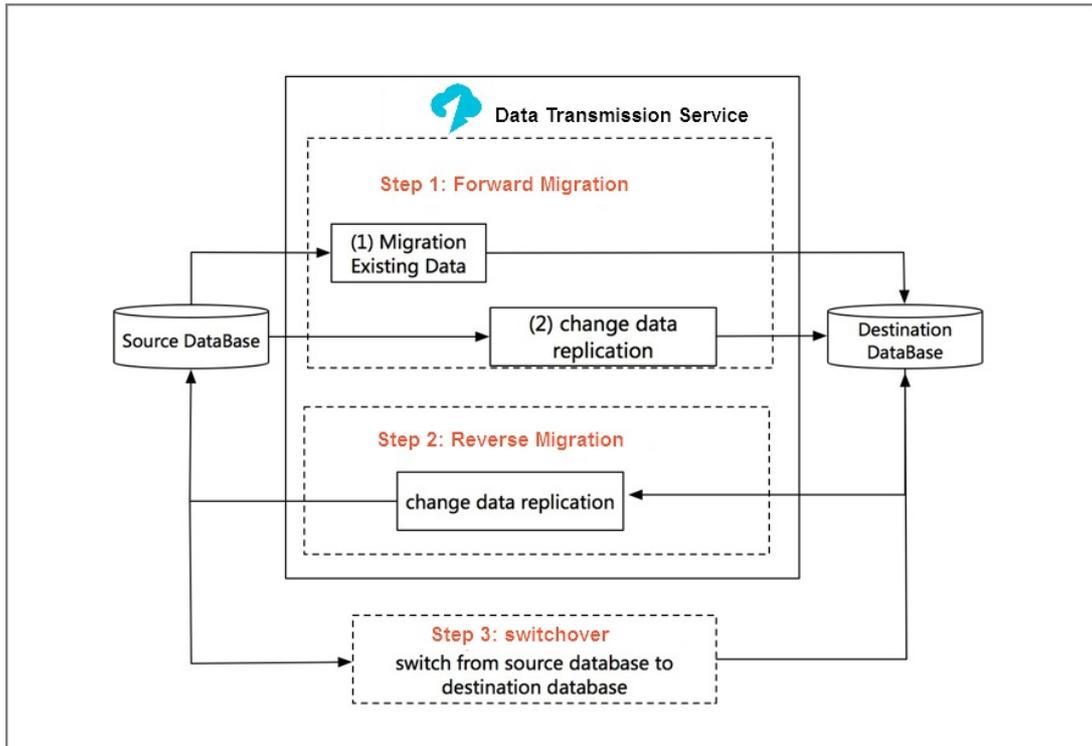
Migration with service downtime reduced to minutes

Many users seek for a way to migrate systems without affecting their services. However, data changes if services are not suspended during the migration. To ensure data consistency, many third-party migration tools require that the service be suspended during data migration. It may take hours or even days throughout the migration and result in a significant loss in service availability.

To reduce the barrier of database migration, DTS provides an interruption-free migration solution that minimizes the service downtime to minutes.

[Interruption-free migration](#) shows how interruption-free migration works.

Interruption-free migration



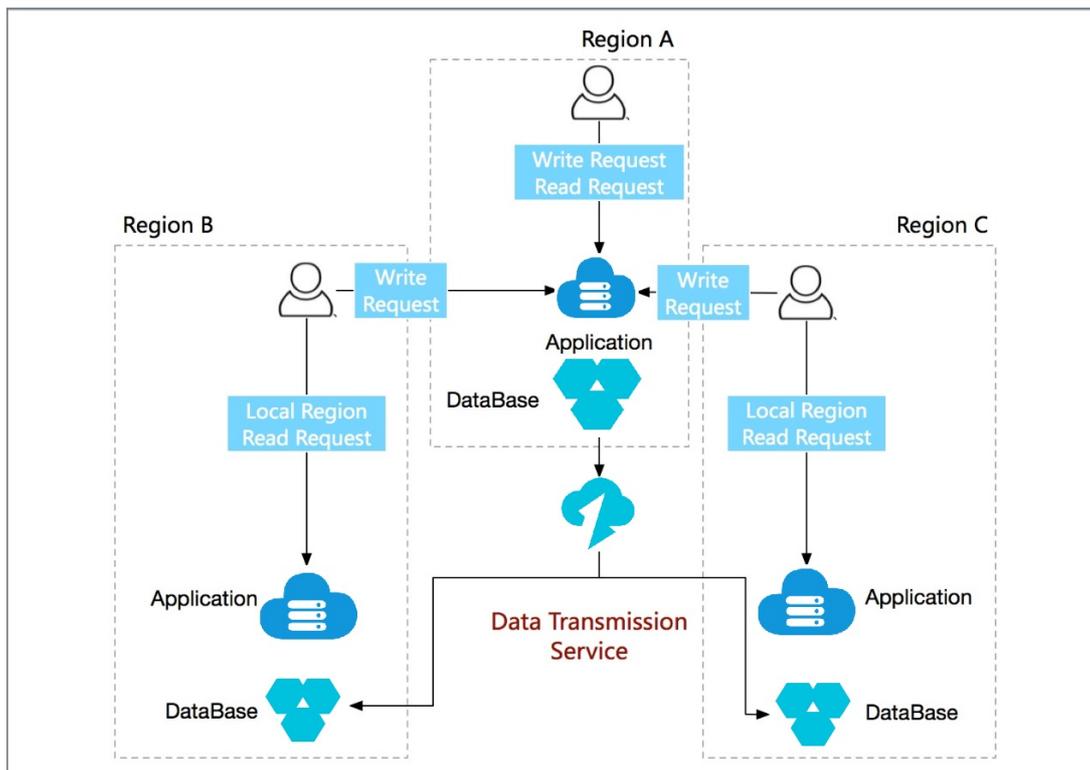
The interruption-free data migration process involves schema migration, full data migration, and incremental data migration. In the incremental data migration phase, data is synchronized between the source and destination instances in real time. You can validate the service in the destination database. After the validation is complete, the service is migrated to the destination database. The entire system is then eventually migrated.

Throughout the migration process, the service experiences interruptions only when it is switched from the source instance to the destination instance.

Accelerated access to global services to empower cross-border businesses

If services with widely distributed users, such as global services, are deployed only in one region, users in other regions have to access them remotely, resulting in high access latency and poor user experience. To accelerate the access to global services and improve access experience, you can adjust the architecture, as shown in **Reduced cross-region access latency**.

Reduced cross-region access latency



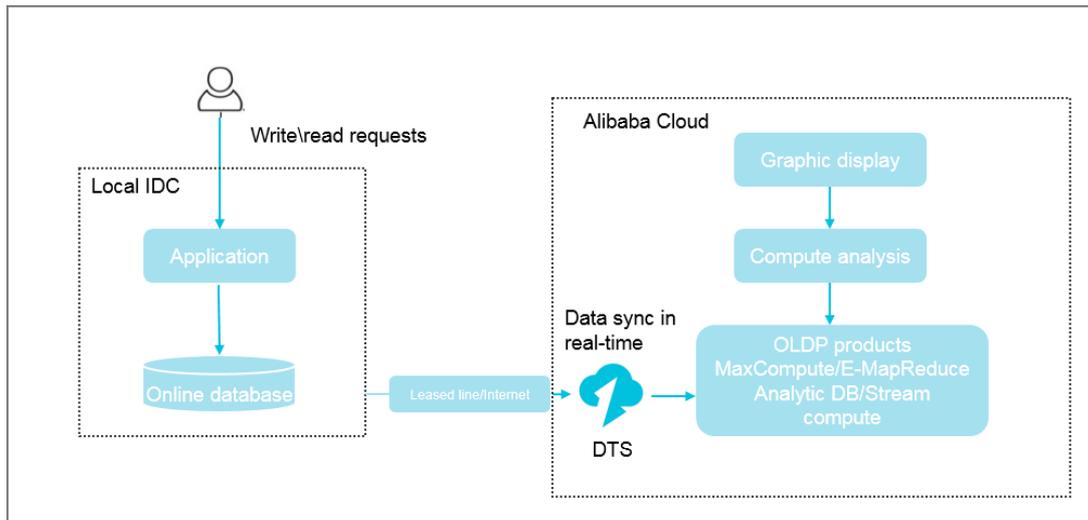
This architecture consists of one center and multiple units. Write requests of users in all regions are routed back to the center. DTS synchronizes data in the center to all units. Read requests of users in different regions can be routed to nearby units to avoid remote access and reduce access latency. In this way, access to global services is accelerated.

Custom cloud BI system built with more efficiency

User-created business intelligence (BI) systems cannot meet the increasing demand for real-time performance and are difficult to manipulate. With the Apsara Stack BI architecture, you can quickly build a BI system without affecting the current architecture. For this reason, more and more users choose to build BI systems that meet their own business requirements on Apsara Stack.

DTS can help you synchronize data stored in local databases to an Apsara Stack BI system (such as MaxCompute or StreamCompute) in real time. You can then perform subsequent data analysis with various compute engines while viewing the computing results in real time with a visualization tool. You can also synchronize those results back to the local IDC with a migration tool. **Cloud BI architecture** shows the implementation architecture.

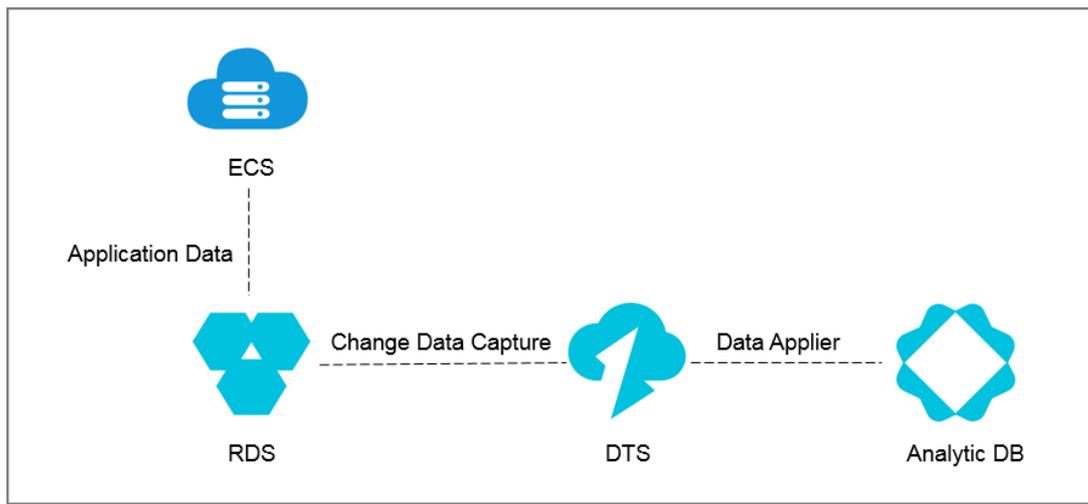
Cloud BI architecture



Real-time data analysis to rapidly respond to market conditions

Data analysis is essential in improving enterprise insights and user experience. Real-time data analysis enables enterprises to adjust marketing strategies more quickly and flexibly so that they can adapt to the rapidly changing marketing conditions and demands for higher user experience. To implement real-time data analysis without affecting online services, service data needs to be synchronized to the analysis system in real time. For this reason, acquiring service data in real time becomes essential. In DTS, the data subscription feature can help you acquire real-time incremental data without affecting online services and synchronize the data to the analysis system using the SDK for real-time data analysis, as shown in **Real-time data analysis**.

Real-time data analysis

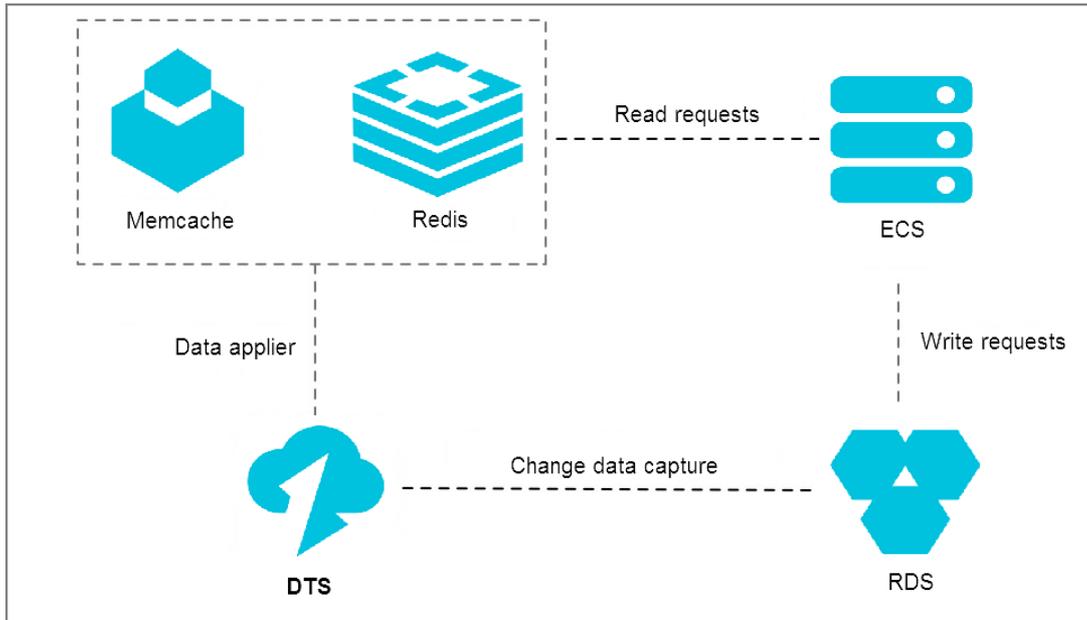


Lightweight cache update policies to make core services more simple and reliable

To accelerate service access and improve concurrent read performance, many enterprises introduce the caching layer to the service architecture. In this architecture, all the read requests are routed to the caching layer, and the memory reading mechanism greatly improves read performance. Cached data cannot persist. If caching ends abnormally, data in the cache memory is lost. To ensure data integrity, the updated service data is kept in a persistent storage medium, such as a database.

In this condition, the service data is inconsistent between the cache and the persistent databases. The data subscription feature can help asynchronously subscribe to the incremental data in those databases and update the cached data to implement lightweight cache update policies. **Cache update policies** shows the architecture of these policies.

Cache update policies



Cache update policies offer the following benefits:

- Quick update with low latency

Cache invalidation is an asynchronous process, and the service returns data directly after the database update is complete. For this reason, you do not need to consider the cache invalidation process, and the entire update path is short with low latency.

- Simple and reliable applications

The complex doublewrite logic is not required for the application. You only need to start the asynchronous thread to monitor the incremental data and update the cached data.

- Application updates without extra performance consumption

Because data subscription acquires incremental data by parsing incremental logs in the database, the acquisition process does not damage the performance of services and databases.

Asynchronous service decoupling to make core services simpler and more reliable

Data subscription optimizes intensive coupling to asynchronous coupling by using real-time message notifications. This makes the core service logic simpler and more reliable. This application has been widely implemented in Alibaba. Tens of thousands of downstream services in the Taobao ordering system acquire real-time data updates through data subscription to trigger the business logic every day.

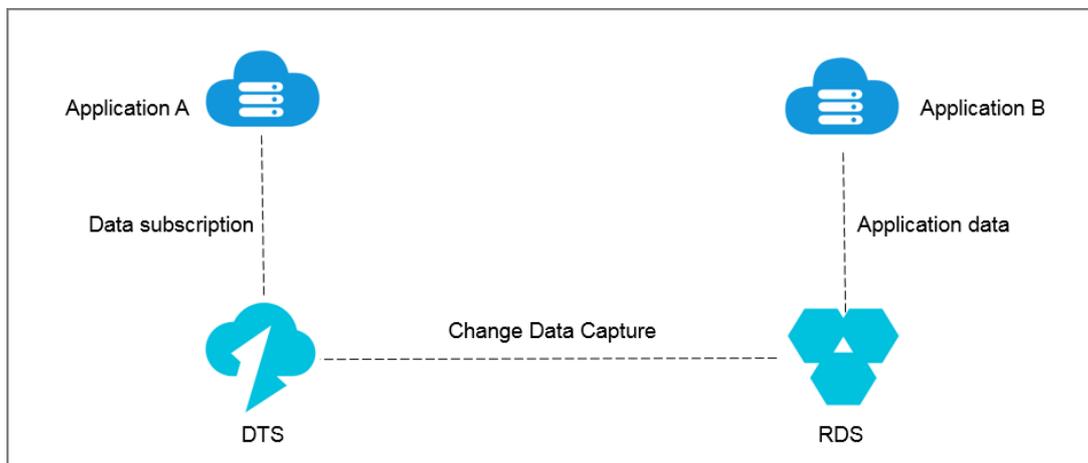
The following uses a simple example to describe the benefits of implementing data subscription in this scenario.

The e-commerce industry involves multiple services including the order management system, inventory management, and the shipping of goods. An ordering process with all of those services included is as follows: After a user places an order, downstream services including seller inventory notification and goods shipping are modified. When all logic modifications are complete, the order result is returned to the user. However, this ordering logic has the following issues:

- The lengthy ordering process results in poor user experience.
- The system is unstable and any downstream fault directly affects the availability of the ordering system.

To improve user experience of core applications, you can decouple the core applications and the dependent downstream services so that they can work asynchronously. In this way, the core applications become more stable and reliable. **Asynchronous service decoupling** shows how to adjust the logic.

Asynchronous service decoupling



The ordering system returns the order result directly after order placement. With DTS, the underlying layer acquires the updated data from the ordering system in real time. Then, the downstream service subscribes to the modified data using the SDK and triggers the service logic such as inventory and shipping. In this way, the ordering system becomes simpler and more reliable.

Horizontal scaling to improve read performance and quickly adapt to business growth

A single RDS instance may not be able to support a large number of read requests, which may affect the main service process. To elastically improve the read performance and reduce database workload, you can create read-only instances using the real-time synchronization feature of DTS. These read-only instances take on large amounts of the database reading workload and expand the throughput of applications.

14.6. Concepts

Precheck

Precheck is an essential stage before a migration task starts. It mainly checks the prerequisites that may affect a successful migration, such as the connectivity of the source and destination instances and the permissions of the migration accounts. If the precheck fails, you can fix the problems as instructed and run the precheck again.

Schema migration

Schema migration is a type of migration tasks. In database migration, it refers to migrating the schema syntax, including tables, views, triggers, stored procedures, stored functions, and synonyms. For migration between heterogeneous databases, data types are mapped during schema migration, and the schema syntax is adjusted according to the schema syntax of the source and destination instances.

Full data migration

Full data migration is a type of migration task. It refers to migrating all the data except the schema syntax from the source instance to the destination instance. If you select Full Data Migration only and leave Schema Migration unselected, new data generated in the source instance will not be migrated to the destination instance.

Incremental data migration

Incremental data migration is a type of migration tasks. It refers to synchronizing the new data written to the source instance to the destination instance during the migration. When creating a migration task, if you select both Full Data Migration and Incremental Data Migration, DTS will first perform a static snapshot on the source instance, migrate the snapshot data to the destination instance, and then synchronize the new data from the source instance to the destination instance during the migration. Incremental data migration is a process of synchronizing data between the source and destination instances in real time. This process does not automatically end. If you want to stop migrating data, you must manually disable the task in the console.

Initial synchronization

Initial synchronization refers to synchronizing the historical data of the objects to be synchronized to the destination instance before synchronizing the incremental data through the synchronization channel.

Initial synchronization includes initial schema synchronization and initial full data synchronization. Initial schema synchronization refers to synchronizing the required schema syntax in the initial stage. Initial full data synchronization refers to synchronizing the data of the objects for the first time.

Synchronization performance

Synchronization performance is measured based on the number of records that are synchronized to the destination instance per second. The measurement unit is records per second (RPS).

Synchronization delay

Synchronization delay refers to the duration between the timestamp when the latest data in the destination instance is starting to be synchronized from the source instance and the current timestamp of the source instance. It reflects the time difference between the data in the source and destination instances. If the synchronization delay is zero, data in the source instance is in sync with that in the destination instance.

Subscription channel ID

The subscription channel ID is a unique identifier of a subscription channel. After you purchase a subscription channel, DTS automatically generates a subscription channel ID. To consume the incremental data using the SDK, you must configure a correct subscription channel ID. You can find the ID that corresponds to each subscription channel in the subscription list of the DTS console.

Data update

In DTS, you can update data or its schema. A data update only modifies the data. The schema syntax is not changed. Operations including INSERT, UPDATE, and DELETE fall into this category.

Schema update

In DTS, you can update data or its schema. Schema update modifies the schema syntax. Operations including CREATE TABLE, ALTER TABLE, and DROP VIEW fall into this category. You can choose whether to subscribe to schema update when you create a subscription channel.

Data range

Data range refers to the range of timestamps of incremental data stored in the subscription channel. The timestamp of a piece of incremental data is the time when the incremental data is applied and written to the transaction log in the database instance. By default, only data generated on the most recent day is retained in the subscription channel. DTS regularly cleans the expired incremental data and updates the data range of the subscription channel.

Consumption checkpoint

The consumption checkpoint is the timestamp of the latest consumed incremental data that is subscribed using the downstream SDK. The SDK sends an ACK message to DTS for every piece of data that is consumed. The server updates and saves the consumption checkpoint corresponding to the SDK. When the SDK encounters an exception, the server restarts and automatically pushes the data at the latest consumption checkpoint.

15. Data Management Service (DMS)

15.1. What is Data Management Service?

Data Management Service (DMS) is a Web-based data management service used to manage relational databases such as MySQL and PostgreSQL, as well as OLAP databases.

It integrates data management with structure management.

15.2. Benefits

Support for multiple data sources

- Relational databases such as MySQL and PostgreSQL.

Data analysis

- Intuitive analysis of the numbers of read, inserted, deleted and updated rows in business tables.

Efficient development

- Table structure comparison.
- Automatic completion of SQL statements.
- Reuse of custom SQL statements and templates.
- Automatic recovery of operation environments.
- Dictionary and document export.

15.3. Architecture

Apsara Stack Data Management Service (DMS) consists of the business layer, scheduling layer, and connection layer. It processes real-time data access and schedules data-related back-end tasks for relational databases.

Business layer

- The DMS business layer provides online GUI-based database operations, and can be extended linearly to improve the general service capabilities of DMS.
- DMS supports stateless failover, ensuring 24/7 availability.

Scheduling layer

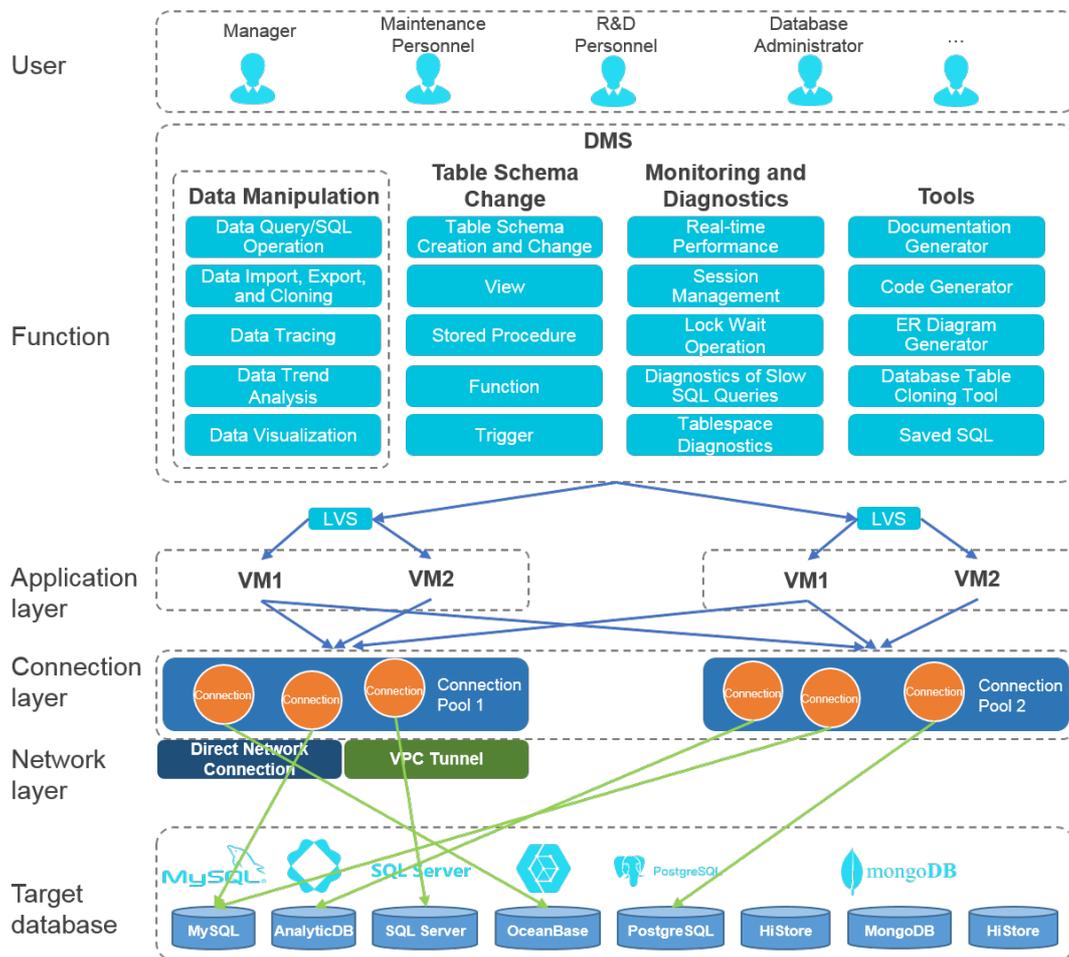
- The scheduling layer allows you to import and export tables and compare table structures. This layer uses the thread pool to schedule tasks. There are two modes of scheduling: real-time scheduling and background periodic scheduling.
- Real-time scheduling allows you to quickly schedule and execute a task in the front end. After you submit a task, DMS automatically executes the task in the background. After the task is completed, you can download or view the execution result.
- Background periodic scheduling allows you to periodically obtain specified data, such as data trends. DMS collects business data in the background based on scheduled tasks, allowing you to query and analyze the collected data.

Connection layer

The connection layer is the core component for data access in DMS. It has the following features:

- Processes requests from MySQL and PostgreSQL databases.
- Isolates sessions and provides session persistence. You can open multiple SQL windows in DMS, and the SQL window sessions are isolated from each other. The session in each SQL window is persistent to simulate the client experience.
- Controls the number of instance sessions to avoid establishing a large number of connections to a single instance.
- Provides different connection release policies for different functions. This improves user experience and reduces the number of connections to the databases.

DMS system architecture



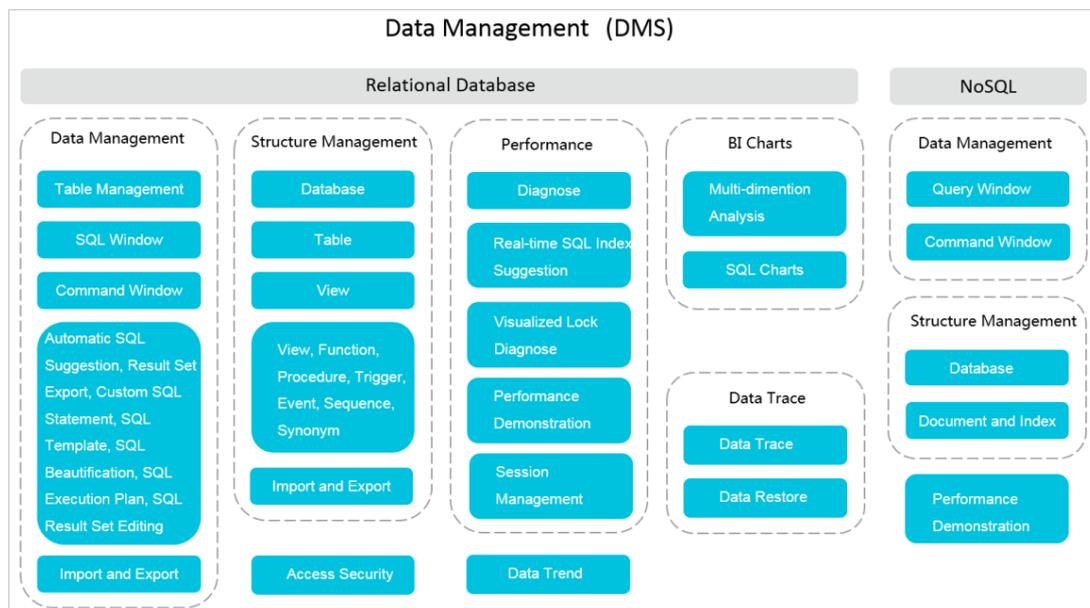
15.4. Features

Relational database management

- **Data management:** includes functions such as SQL windows, SQL command lines, table data, intelligent SQL prompts, SQL formatting, custom SQL statements, SQL templates, SQL execution plans, and import and export operations.
- **Structure management:** includes functions such as table structure comparison, and management of objects (databases, tables, views, functions, storage procedures, triggers, events, series, and synonyms).

Feature diagram

Feature diagram



15.5. Scenarios

15.5.1. Convenient data operations

Pain point

You need a lightweight product that features full functionality to create SQL statements, save frequently used SQL statements, and use these statements in your business.

Solution

- You can open a table in DMS and perform operations on table data as you would in an Excel worksheet. You can add, delete, change, query, and make statistical analysis of table data without understanding SQL.
- You can customize SQL statements, save frequently used SQL statements, and apply these SQL statements to databases or instances.

15.5.2. Prohibiting data export

Paint point

When cooperating with a partner, an enterprise manages data and its partner develops functions. The partner needs to have access to view the enterprise's data but cannot have the ability to export data to ensure data security.

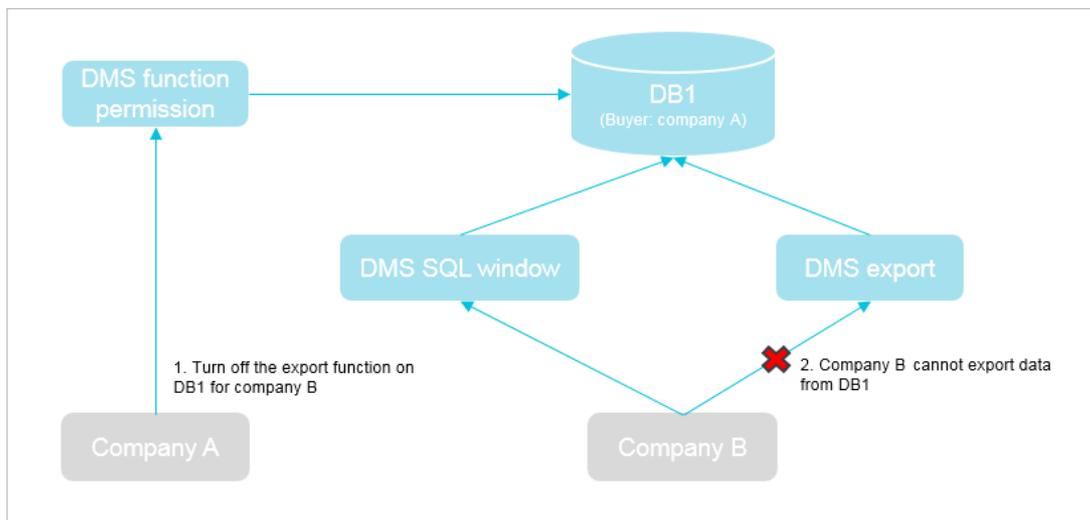
Solution

Enterprise users can log on to the DMS console to grant their partners access permissions on the corresponding database instances, disabling data exporting to protect their data.

Partners are permitted only to query and view data, eliminating the risk of data leakage.

DMS – Function-based authorization shows how to use the function-based authorization feature to prohibit partners from exporting data.

DMS – Function-based authorization



15.5.3. SQL statement reuse

Pain point

SQL statements are used when you access a database. While simple queries are easy to use, rewriting SQL queries for complex data analysis or SQL queries that contain service logic is time-consuming. Even if you save these SQL queries to files, you have to maintain the files and you cannot use them without access to the files.

Solution

You can use the **My SQL** function provided to save frequently used SQL statements to DMS. As the SQL statements are not saved locally, they can be reused in any databases or instances.

15.6. Limits

Relational databases

Support for relational databases

Module	Function	MySQL	PostgreSQL
Data management	Table data management	√	√
	SQL windows	√	√
	SQL command lines	√	√
	SQL templates	√	
	SQL formatting	√	√
	Custom SQL statements	√	
	Intelligent SQL prompts	√	
	SQL execution plans	√	√
Structure management	Database management	√	√
	Table management	√	√
	Management of objects such as indexes, views, stored processes, functions, triggers, and events	√	√
	Entity relationship diagram display	√	
	Data dictionaries	√	
Import and export	Basic import and export functions	√	√
	Export of large volumes of data	√	√

16. Server Load Balancer (SLB)

16.1. What is Server Load Balancer?

Server Load Balancer (SLB) is a traffic distribution service that distributes inbound traffic to backend Elastic Compute Service (ECS) instances based on configured forwarding rules. SLB improves the service capability and availability of applications.

You can use SLB to virtualize multiple ECS instances in the same region into an application server pool. Then, you can distribute client requests to the ECS instances based on forwarding rules.

SLB checks the health status of the ECS instances and automatically isolates abnormal ones in the server pool to eliminate single points of failure (SPOFs), improving the overall service capability of applications. SLB is also well equipped to defend against DDoS attacks.

SLB consists of three components:

- **SLB instances:** An SLB instance is a running load-balancing service that receives and distributes inbound traffic to backend servers.

To use the SLB service, you must create an SLB instance with at least one listener and two ECS instances configured.

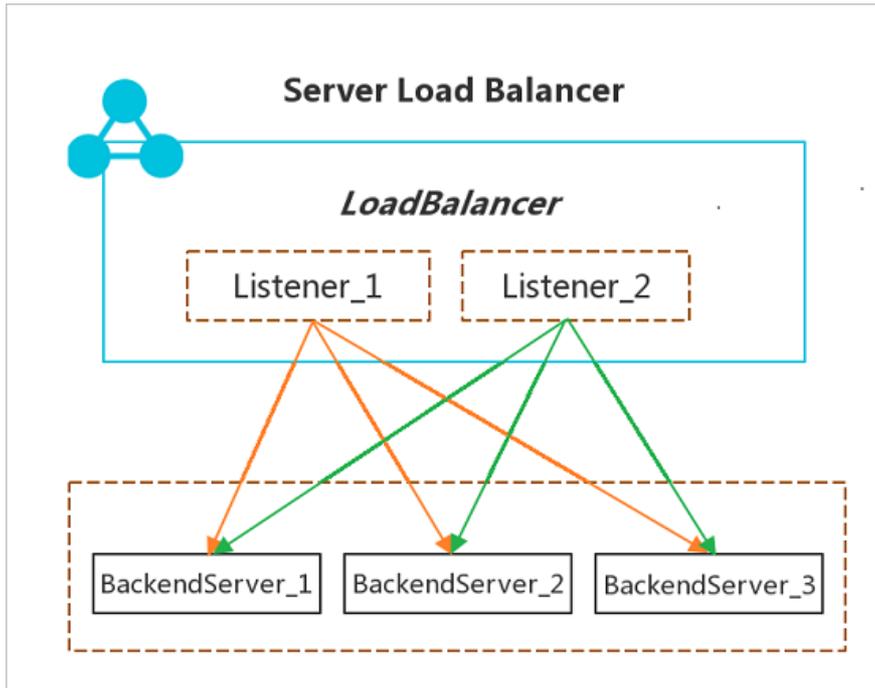
- **Listeners:** A listener checks client requests and forwards them to backend servers. It also performs health checks on the backend servers.

You can create Layer-4 (TCP/UDP) or Layer-7 (HTTP/HTTPS) listeners to suit your needs. You can create domain-based and URL-based forwarding rules for Layer-7 listeners.

- **Backend servers:** Backend servers are ECS instances attached to an SLB instance to receive and process the distributed requests. You can divide ECS instances running different applications or functioning different roles into different server groups.

As shown in the following figure, after the SLB instance receives a client request, the listener forwards the request to the corresponding backend ECS instances based on the configured forwarding rules.

SLB components



16.2. Benefits

High availability

SLB is designed with full redundancy to avoid SPOFs, and supports zone-level disaster recovery. SLB can be scaled based on the application load, without interrupting external services in the event of traffic fluctuations.

Low cost

SLB is more cost-efficient than traditional hardware load-balancing systems. By giving you free access to VPC-connected instances without generating any O&M cost, the SLB service removes your need to purchase expensive load-balancing devices.

Security

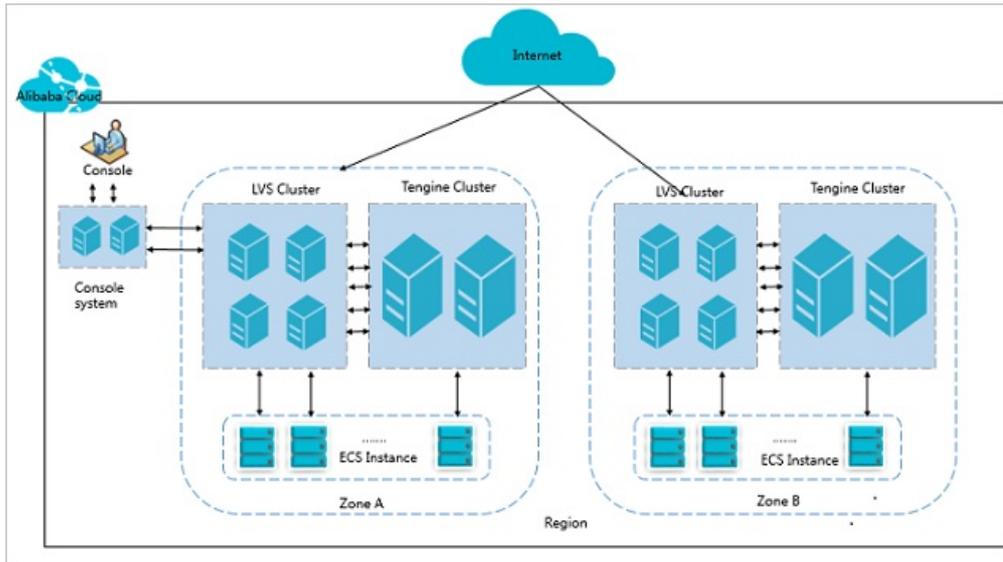
Working with Alibaba Cloud Security, SLB is shielded from DDoS attacks such as CC and SYN flood attacks.

16.3. Architecture

SLB is deployed in clusters to achieve session synchronization. This can eliminate SPOFs of backend servers, improve redundancy, and ensure service stability. Apsara Stack provides Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) load-balancing services.

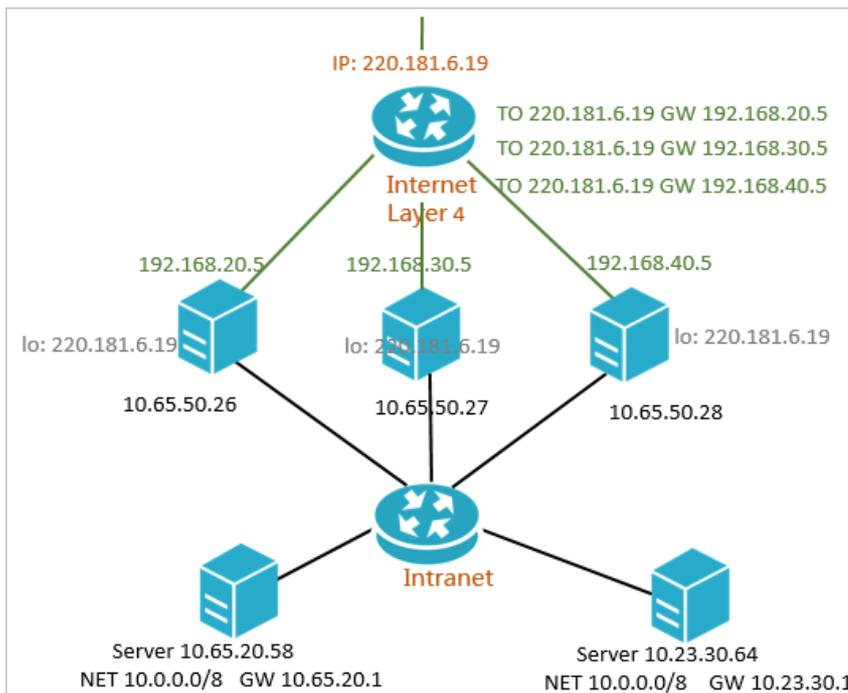
- Layer-4 SLB combines the open-source Linux Virtual Server (LVS) with Keepalived to balance loads, and implements customized optimizations to meet cloud computing requirements.
- Layer-7 SLB uses Tengine to balance loads. Tengine is a Web server project launched by Taobao. Based on NGINX, Tengine has a wide range of advanced features enabled for high-traffic websites. For more information, see *CreateLoadBalancer* in *API reference*.

SLB architecture



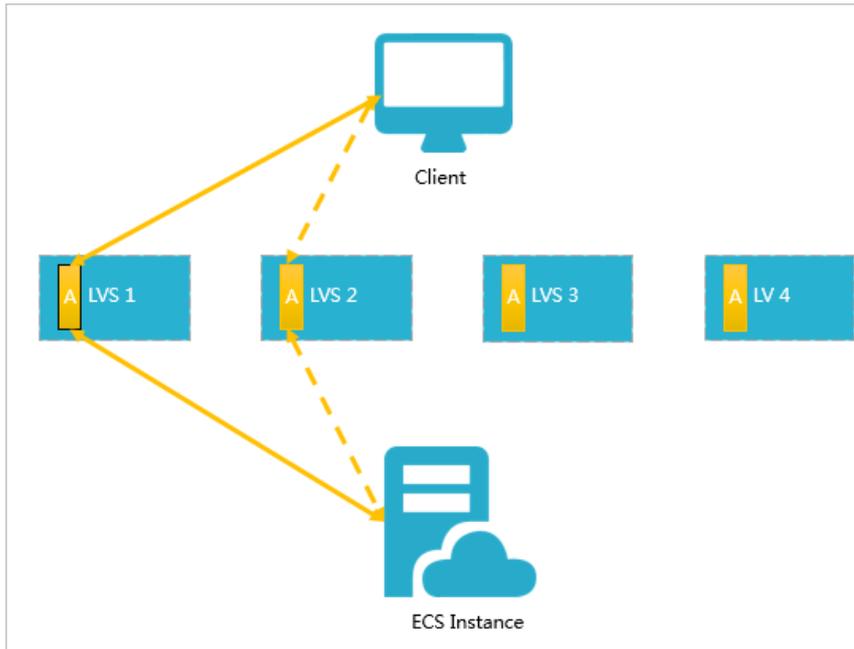
As is shown in the following figure, Layer-4 SLB runs in a cluster of LVS machines. The cluster deployment model strengthens the availability, stability, and scalability of load-balancing services in abnormal circumstances.

Cluster deployment



Each machine in the LVS cluster uses multicast packets to synchronize sessions with the other LVS machines. As shown in the following figure, Session A established on LVS1 is synchronized to other LVS machines after the client transfers three data packets to the server. Solid lines indicate the current active connections, while dotted lines indicate that the session requests will be sent to other normally working machines if LVS1 fails or is being maintained. In this way, you can perform hot updates, machine failure maintenance, and cluster maintenance without affecting your services.

Session synchronization



16.4. Features

Supported protocols

SLB currently supports Layer-4 load balancing over TCP or UDP and Layer-7 load balancing over HTTP or HTTPS.

Health checks

SLB checks the health of backend ECS instances. SLB stops forwarding requests to abnormal ECS instances until they recover.

Session persistence

SLB supports session persistence. You can forward requests from a client to the same backend ECS instance during the session lifecycle.

Scheduling algorithms

SLB supports the following scheduling algorithms:

- Round robin: Requests are distributed across backend servers sequentially.
- Least connections: More requests are distributed to backend servers with fewer connections.

Access control

SLB supports whitelist-based access control. You can configure a whitelist to control which IP addresses can access SLB.

Certificate management

SLB provides centralized certificate management for HTTPS listeners. You do not need to upload certificates to backend ECS instances because decryption is performed on SLB. This feature reduces the CPU usage of backend ECS instances.

Virtual server groups

A virtual server group is a group of ECS instances. You can divide ECS instances running different applications into several virtual server groups. For each group running the same application, you create a specific listener that takes in only certain types of requests.

16.5. Scenarios

SLB applies to the following scenarios:

Scenario 1: Balance your application loads

You can configure listening rules to distribute heavy traffic among ECS instances. You can also use the session persistence feature to forward requests from a client to the same backend ECS instance to enhance access efficiency.

Scenario 2: Scale your applications

You can extend the service capability of your applications by adding or removing backend ECS instances to suit your business needs. SLB can be used for both Web and application servers.

Scenario 3: Eliminate SPOFs

You can add multiple ECS instances to an SLB instance. When ECS instances malfunction, SLB automatically isolates them and distributes inbound requests to healthy ECS instances, ensuring the normal operation of the application system.

16.6. Limits

- For Layer-4 (TCP) load-balancing service, a backend ECS instance cannot act both as a backend server and a client that sends requests to an SLB instance. The returned packets will only be forwarded among the backend ECS instances but not through SLB. Therefore, you cannot access the SLB instance from backend ECS instances.
- Before using SLB to provide services, make sure all applications on backend ECS instances have been correctly configured and are accessible to the IP addresses of these ECS instances.
- SLB cannot synchronize data among ECS instances. If the applications deployed on backend ECS instances are stateless, you can store data in independent ECS instances or Relational Database Service (RDS). If these applications are stateful, make sure data is synchronized among these ECS instances.
- After a domain name is resolved into the IP address of an SLB instance, do not delete the SLB instance while it is providing services externally. If the SLB instance is deleted, its IP address is released, which will interrupt services.

16.7. Terms

Server Load Balancer (SLB)

A traffic distribution service that distributes inbound traffic to backend Elastic Compute Service (ECS) instances based on the configured forwarding rules. SLB distributes traffic to ECS instances to extend the service capabilities of application systems. This also enhances the availability of the application systems by eliminating single points of failure (SPOFs).

SLB instance

A running entity of the SLB service. To use the load-balancing service, you must create an SLB instance.

SLB IP address

An IP address allocated to an SLB instance. The IP address can be either public or private, depending on the instance type.

listener

A listener defines how to forward inbound requests to backend servers. A listener has listener ports, SLB policies, and health check configurations. Each listener corresponds to a backend application.

backend server

An ECS instance receiving SLB-distributed requests. SLB forwards requests to backend ECS instances based on configured rules.

17.Virtual Private Cloud (VPC)

17.1. What is VPC?

A Virtual Private Cloud (VPC) is a private network established in Apsara Stack. VPCs are logically isolated from each other.

You have full control over your VPC. For example, you can select its IP address range and configure routing tables and gateways. You can also use Alibaba Cloud resources such as ECS, RDS, and SLB in your own VPCs. You can connect a VPC to other VPCs or a local network to form an on-demand customizable network environment. This allows you to smoothly migrate applications to the cloud.

Components

Each VPC consists of a private Classless Inter-Domain Routing (CIDR) block, a VRouter, and at least a VSwitch.

- **CIDR block**

A CIDR block is a private IP address range in a VPC. The IP addresses of all cloud resources deployed in the VPC are within the specified CIDR block. When creating a VPC or a VSwitch, you must specify the private IP address range in the form of a CIDR block.

You can use any of the following standard CIDR blocks and their subnets as the IP address range of the VPC.

CIDR block	Number of available private IP addresses (system reserved ones excluded)
192.168.0.0/16	65,532
172.16.0.0/12	1,048,572
10.0.0.0/8	16,777,212

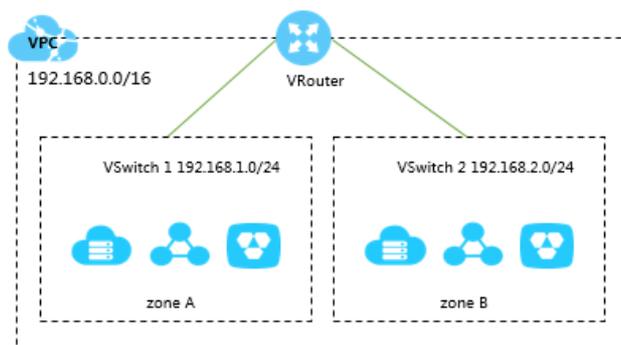
- **VRouter**

The VRouter is the hub of a VPC. As an important component of a VPC, the VRouter connects the VSwitches in a VPC and serves as the gateway connecting the VPC with other networks. After you create a VPC, the system automatically creates a VRouter, which is associated with a routing table.

- **VSwitch**

A VSwitch is a basic network device in a VPC and is used to connect different cloud product instances. After creating a VPC, you can further divide the VPC to one or more subnets by creating VSwitches. The VSwitches within a VPC are interconnected. You can deploy applications in VSwitches of different zones to improve the service availability.

VPC



17.2. Benefits

VPC features high security and flexible configurations, and supports multiple connection methods.

Secure

Each VPC is identified by a unique tunnel ID. VPCs are completely isolated from one another. You can use security groups or whitelists to control access to cloud resources in the VPC.

Easy to use

You can create and manage a VPC in the VPC console. After a VPC is created, the system automatically creates a VRouter and a routing table for it.

Scalable

You can create multiple subnets in a VPC to deploy different services. Additionally, you can connect a VPC to a local data center or other VPCs to extend the network architecture.

17.3. Architecture

VPCs are isolated virtual networks achieved by using tunneling technology. Each VPC is identified by a unique tunnel ID.

Background information

The continuous development of cloud computing technologies leads to increasing virtual network requirements such as scalability, security, reliability, privacy, and performance. This scenario has hastened the birth of a variety of network virtualization technologies.

Earlier solutions combined virtual and physical networks to form a flat network architecture, such as large layer-2 networks. As the scale of virtual networks grew, earlier solutions faced more serious problems. A few notable problems include ARP spoofing, broadcast storms, and host scanning. Various network isolation technologies emerged to resolve these problems by completely isolating the physical networks from the virtual networks. One of the technologies utilized VLAN to isolate users, but due to VLAN limitations, it could only support up to 4096 nodes. It is insufficient to support the huge amount of users in the cloud.

VPC basis

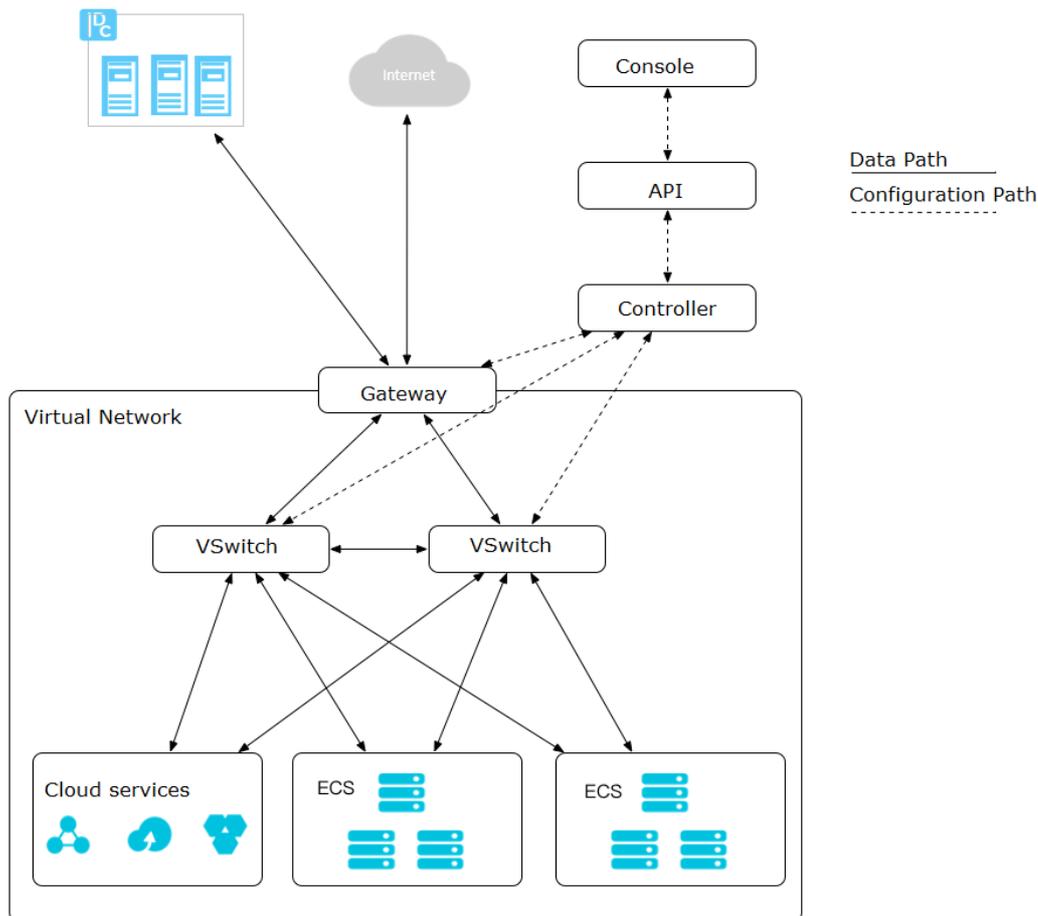
VPCs are isolated virtual networks achieved by using tunneling technology. Each VPC is identified by a unique tunnel ID. A unique tunnel ID is generated when tunnel encapsulation is performed on each data packet transmitted between the ECS instances within a VPC. Then, the data packet is transmitted over the physical network. ECS instances in different VPCs cannot communicate with each other. They have different tunnel IDs and therefore are on different routing planes.

Alibaba Cloud developed technologies such as VSwitch, Software Defined Network (SDN), and hardware gateway based on tunneling technology. These technologies serve as the basis for VPC.

Logical architecture

As shown in the following figure, the VPC architecture contains three main components: VSwitches, gateway, and controller. VSwitches and gateways form the key data path. Controllers use the protocol developed by Alibaba Cloud to forward the forwarding table to the gateway and VSwitches, completing the key configuration path. In the overall architecture, the configuration path and data path are separated from each other. VSwitches are distributed nodes, the gateway and controller are deployed in clusters, backup for disaster recovery is supported by multiple data centers, and all links have redundant disaster recovery. This improves the overall availability of the VPC.

VPC architecture



17.4. Features

Custom VPCs

You can customize VPCs. When you create VPCs or VSwitches, you can specify private CIDR blocks for them. You can divide a VPC into multiple subnets and deploy services in different subnets to improve service availability.

Custom routes

You can add custom routes to the VPC routing table to forward traffic to the specified next hop. The routing table uses the longest prefix matching rule for traffic routing. The routing entry with the longest subnet mask will be used because it is the most specific route.

Varied connection methods

A VPC provides you with varied connection methods. You can connect a VPC to the public network, a local IDC, or another VPC.

- Connect a VPC to the public network

You can connect a VPC to the public network by binding an EIP to the VPC or configuring NAT Gateway, so that cloud services in the VPC can communicate with the public network.

- Connect a VPC to another VPC

You can connect a VPC to another VPC by creating a pair of router interfaces to build high speed and secure intranet communication.

- Connect a VPC to a local IDC

You can connect a VPC to a local IDC by using a leased line to smoothly migrate local applications to the cloud.

17.5. Scenarios

VPCs are applicable to scenarios that require high communication security and service availability.

Host applications

You can host applications that provide external services in a VPC and control access to these applications from the public network by creating security group rules and access control whitelists. You can also isolate application servers from databases to implement access control. For example, you can deploy Web servers in a subnet that can access the public network, and deploy their application databases in a subnet that cannot access the public network.

Host applications that require public network access

You can host an application that requires access to the public network in a subnet of a VPC and route the traffic through NAT. After you configure SNAT rules, instances in the subnet can access the public network without exposing their private IP addresses, which can be changed to public IP addresses any time to avoid external attacks.

Zone-disaster recovery

You can divide a VPC into one or multiple subnets by creating VSwitches. Different VSwitches within the same VPC can communicate with each other. Resources can be deployed to VSwitches of different zones to achieve zone-disaster recovery.

Isolate business systems

VPCs are logically isolated from each other. To isolate multiple business systems, such as the production and test environments, you can create a VPC for each environment. When the VPCs need to communicate with each other, you can create a peering connection between them.

Extend the local network architecture

To extend the local network architecture, you can connect the local data center to a VPC. You can also seamlessly migrate local applications to the cloud without changing the application access method.

17.6. Terms

Virtual Private Cloud (VPC)

A private network established in Apsara Stack. VPCs are logically isolated from each other. You can create and manage cloud service instances in your VPC, such as ECS instances, SLB instances, and RDS instances.

VRouter

A hub in a VPC. It connects all VSwitches in the VPC and serves as a gateway that connects the VPC to other networks. A VRouter routes the network traffic to their destinations based on the configured routing entries.

VSwitch

A basic network device of a VPC. It is used to connect different cloud service instances. When creating a cloud service instance in a VPC, you must specify the VSwitch that is used by the instance.

Routing table

A list of routing entries in a VRouter.

Routing entry

An entry in a routing table. A routing entry specifies the next hop address for the network traffic destined to a CIDR block. There are two types of entries, system routing entry and custom routing entry.

17.7. Limits

VPC

Resource	Default limit
Maximum number of VRouters in a VPC	1
Maximum number of routing tables in a VPC	1
Maximum number of VSwitches in a VPC	24
Maximum number of routing entries in a routing table	48

VRouter and VSwitch

Resource	Default limit
VRouter	<ul style="list-style-type: none">• Each VPC can have only one VRouter.• Each VRouter can have only one routing table.• Dynamic routing protocols such as BGP and OSPF are not supported.
VSwitch	<ul style="list-style-type: none">• Layer-2 broadcasting and multicasting are not supported.

18. Log Service

18.1. What is Log Service?

Log Service is a one-stop service designed to manage log data. You can use Log Service to perform operations on log data such as collection, query, analysis, and consumption.

Log Service has been used in various big data scenarios within Alibaba Group. You can use Log Service to collect, consume, query, and analyze log data without performing any programming. It helps increase O&M efficiency and build capabilities to process large-volume logs in the data technology (DT) era.

Log Service provides you with the following features:

- **Log collection:** Log Service allows you to collect various formats of log data such as events, binary logs, and text logs in real time through multiple methods, such as Logtail and JS.
- **Query and analysis:** Log Service provides real-time query and analysis for the collected log data, and allows you to create visual charts and dashboards based on analysis results.
- **Status alert:** Log Service allows you to regularly execute, query, and analyze statements based on query and analysis features. When query results meet alert conditions, real-time alerts are reported based on pre-configured alert tasks.
- **Real-time consumption:** Log Service provides real-time consumption interfaces for log data collected to the server.

18.2. Benefits

Fully managed service

- Log Service can be set up in five minutes, and is easy to use.
- LogHub provides all of the features of Kafka, such as data monitoring and alerts. It also scales automatically to handle PBs of data each day, saving more than 50% of costs compared with self-built systems.
- LogSearch/Analytics provides dashboards, saved searches and alerts to help reduce costs by over 80% compared with self-built systems.
- You can use more than 30 methods to collect log data. Log Service can be seamlessly integrated with open-source software such as Storm and Spark.

Rich ecosystem

- LogHub supports over 30 types of data sources. It can be interfaced with embedded devices, webpages, servers, and programs. LogHub can also interface with consumption systems such as Storm and Spark Streaming.
- LogSearch/Analytics is compatible with SQL-92, has complete query and analysis syntax, and can interface with Grafana by using the JDBC protocol.

Strong real-timeliness

- Data can be used immediately after it is written. Logtail (the data collection agent) collects and transfers data in real time.
- LogSearch/Analytics: Data can be queried and analyzed after it is written. Billions of pieces of data can be queried per second with multiple search conditions. Hundreds of millions of pieces of data can be analyzed per second with multiple aggregate conditions.

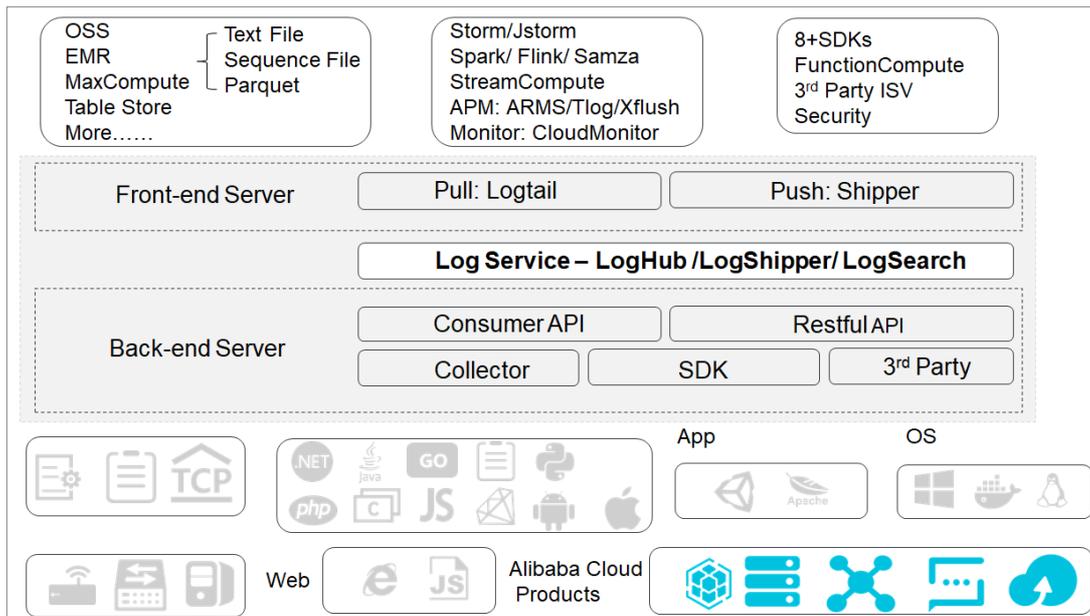
Complete APIs and SDKs

- Log Service supports user-defined management and secondary development.
- All Log Service functions can be implemented by using APIs and SDKs. SDKs in multiple programming languages are provided, which allows you to easily manage services and millions of devices.
- The syntax for query and analysis is simple and compatible with SQL-92. User-friendly interfaces are integrated into the software environment.

18.3. Product architecture

Product architecture shows the architecture of Log Service.

Product architecture



Logtail

The Logtail agent collects logs. It has the following characteristics:

- **Non-intrusive file-based log collection**
 - Logtail reads only files.
 - The log collection is not intrusive.
- **High security and reliability**
 - Logtail can rotate files without data loss.
 - Logtail supports local caching.
 - Logtail retries when network exceptions occur.
- **Convenient management**
 - Logtail can be accessed through a Web client.
 - Logtail supports UI-based configuration.
- **Comprehensive self-protection**
 - Logtail monitors CPU and memory usage of its processes in real time.
 - Logtail allows you to set an upper limit on the resource usage of its processes.

Frontend servers

Frontend servers are built on LVS and NGINX. They have the following characteristics:

- **Support for HTTP and REST**
- **Scale-out**
 - The processing capabilities can be increased quickly when traffic rises.
 - Frontend servers can be added.
- **High throughput and low latency**
 - **Asynchronous processing:** If an exception occurs when a single request is sent, other requests are not affected.
 - **LZ4 compression:** The processing capabilities of individual servers are increased while network bandwidth consumption is reduced.

Backend servers

The backend service is a distributed process deployed on multiple servers. The service performs storage, indexing, and queries on Logstore data in real time. The overall characteristics of the backend service are as follows:

- High data security
 - Each log is saved to three copies stored on different servers.
 - Data can be automatically recovered in the cases of disk damage or server downtime.
- Stable service
 - Logstores are automatically migrated in the cases of process crashes or server downtime.
 - Automatic load balancing ensures that traffic is distributed evenly among different servers.
 - Strict quotas prevent incorrect or unexpected operations of a single user from affecting other users.
- Scale-out
 - A shard is the basic unit for scale-out.
 - You can add shards as needed to increase throughput.

18.4. Features

18.4.1. Core features

LogHub

LogHub supports a variety of methods for lossless log collection such as clients, webpages, protocols, SDKs, and APIs (for mobile terminals and gaming), as well as consumption ways such as SDKs, Storm Sprout, and Spark Client. By supporting multiple formats of real-time log collection and consumption, LogHub helps you streamline the processing of multi-device and multi-source log collection and consumption.

Features:

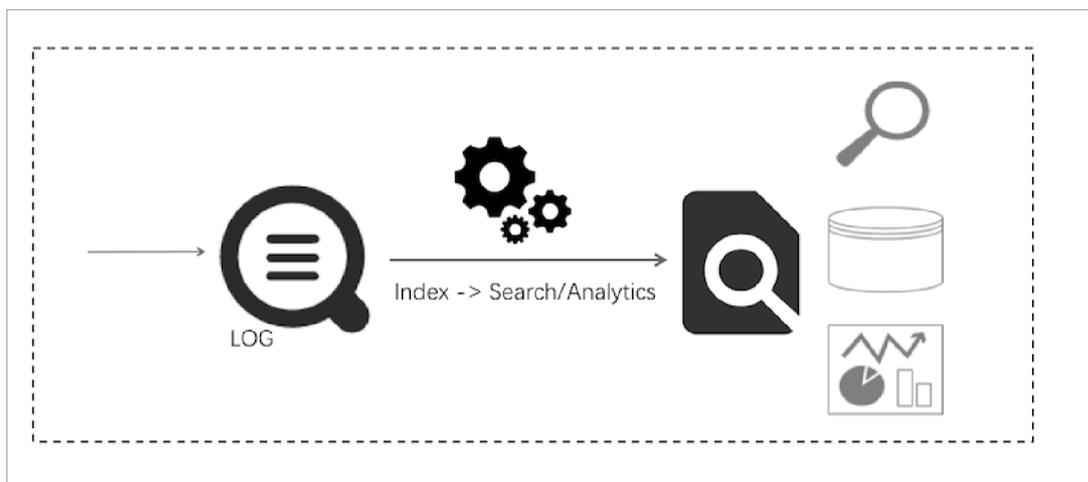
- LogHub collects real-time log data such as metrics, events, binary logs, text logs, and clicks from Elastic Compute Service (ECS), containers, mobile terminals, open-source software, and JS.
- A real-time consumption interface is provided to interconnect with real-time computing and service.

LogSearch/Analytics

LogSearch/Analytics can index, query, and analyze log data collected to the server in real time and generate dynamic data reports based on query and analysis results. It supports visual analysis of log data in multiple scenarios.

- Query: keyword, fuzzy, context, and range queries
- Statistics: a variety of query methods including SQL aggregate queries
- Visualization: dashboards and reports
- Interconnection: Grafana, JDBC, and SQL-92

LogSearch/Analytics



18.4.2. Other features

18.4.2.1. Log

Logs in Log Service

A log is an abstraction of changes that happen in a system. A log is a sequence of records ordered by time, which contains information about operations and results of specific objects. Log files, events, binary logs, and metrics are all different types of logs. Each log file is composed of one or more log entries. A log entry is the smallest unit of data that can be processed in Log Service. Each log entry describes a single system event.

Log Service uses a semi-structured data model to define a log. This model is composed of four fields: Topic, Time, Content, and Source.

Log Service has different format requirements for different log fields, as described in the following table:

Field	Description	Format
Topic	This is a user-defined field used to mark a batch of logs. For example, access logs can be marked based on sites.	Any string up to 128 characters in length, including empty strings. This field is an empty string by default.

Field	Description	Format
Time	This is a reserved field in a log and is used to indicate the time when a log is generated. In most cases, it is generated directly based on the time in a log.	Integer. It must be in Unix timestamp format. The Unix timestamp is the number of seconds that have elapsed since 00:00:00 Thursday, January 1, 1970 UTC.
Content	This field is used to record the specific content of a log. The content consists of one or more content items, and each content item is a key-value pair.	A key is a UTF-8 encoded string up to 128 characters in length. It can contain letters, underscores, and digits. It cannot start with a digit. The following keywords cannot be used in the key: <code>__time__</code> , <code>__source__</code> , <code>__topic__</code> , <code>__partition_time__</code> , <code>__extract_others__</code> , and <code>__extract_others__</code> . The value can be any string up to 1024 × 1024 bytes.
Source	This field indicates the source of a log. For example, the IP address of the machine where a log is generated.	Any string up to 128 characters in length. This field is null by default.

Various log formats are used in actual usage scenarios. The following example shows you how to map an NGINX access log to the log data model of Log Service. Assume that the IP address of your NGINX server is `10.249.201.117`. An original log of this server is as follows:

```
10.1.168.193 - - [01/Mar/2012:16:12:07 +0800] "GET /Send? AccessKeyId=8225105404 HTTP/1.1" 200 5 "-" Mozilla/5.0 (X11; Linux i686 on x86_64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2"
```

The following example shows how to map the original log to the log data model of Log Service.

Field	Content	Description
Topic	<code>""</code>	The default value (empty string) is used.
Time	<code>1330589527</code>	The exact timestamp when the log is generated. It is the number of seconds that have elapsed since 00:00:00 Thursday, January 1, 1970 UTC. The time is converted from the timestamp of the original log.
Content	Key-value pair	The specific content of a log.

Field	Content	Description
Source	"10.249.201.117"	The IP address of the server is used as the log source.

You can decide how to extract the original content of a log and combine the extracted content into key-value pairs. The table below is shown as an example.

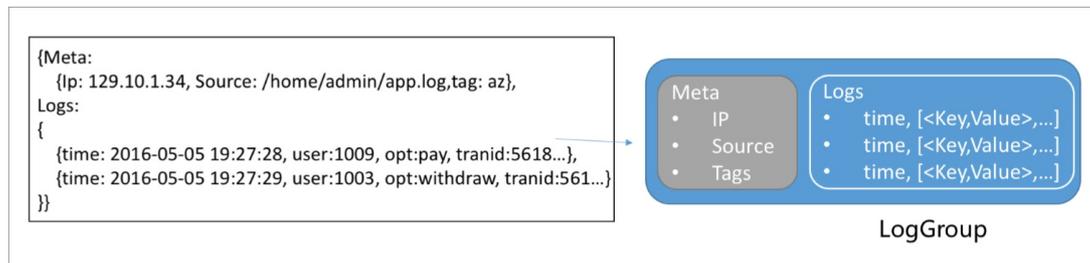
Key	Value
ip	"10.1.168.193"
method	"GET"
status	"200"
length	"5"
ref_url	"_ "
browser	"Mozilla/5.0 (X11; Linux i686 on x86_64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2"

Log group

A log group is a collection of logs. These groups are the basic units used for read and write operations.

The maximum capacity of a log group is 4096 logs or 10 MB.

Log group



18.4.2.2. Project

A project is the management unit for resources in Log Service and is used to isolate and control resources. You can use a project to manage all the logs and related log sources of an application. A project is used to manage Logstores of a user and server configurations for log collection. A project also serves as the portal for a user to access the resources of Log Service.

Projects provide the following features:

- Help you to organize and manage different Logstores. You can use Log Service to collect and store the logs of different projects, services, or environments. You can classify different logs for management in different projects to facilitate subsequent log consumption, exporting, or indexing. In addition, projects are the carriers for log access control.
- Provide you with a portal to access Log Service resources. Log Service allocates an exclusive

access portal to each project. The access portal allows you to write, read, and manage logs through the network.

18.4.2.3. Logstore

A Logstore is the unit used in Log Service for log data collection, storage and query. Each Logstore can belong to only one project, but multiple Logstores can be created for a single project. You can create multiple Logstores for a project as needed. Typically, an independent Logstore is created for each type of log in an application. For example, you have a game called big-game, and it stores three types of logs on the server: `operation_log`, `application_log`, and `access_log`. You can create a project named big-game, and then create three Logstores under this project for the three types of logs to collect, store, and query those logs.

Whether writing or querying logs, you must specify a Logstore for the operation. When you transfer log data to MaxCompute for offline analysis, the data is transferred in units of Logstores. The data in each Logstore is synchronized to separate MaxCompute tables.

Logstores provide the following features:

- Log collection: Logstores support real-time logging.
- Log storage: Logstores support real-time consumption.
- Index creation: Logstores support real-time log query.

18.4.2.4. Shard

Logstore read/write logs must be saved in a shard. A shard is the unit to compose Logstores. Each shard is represented by a non-overlapping, left-closed, right-open interval of MD5 values. The range of the Logstore is represented by the entire range of MD5 values of the shards in the Logstore.

Range

You must specify the number of shards when creating a Logstore. The entire MD5 range is automatically divided evenly based on the specified number of shards. Each shard has a certain range, which can be represented in MD5 and must be within the following value range: [00000000000000000000000000000000,fffffffffffffffffffffffffffffffff).

All of the shard ranges are left-closed and right-open intervals. They are composed of the following keys:

- **BeginKey:** indicates the start of the shard. This key is included in the shard range.
- **EndKey:** indicates the end of the shard. This key is excluded from the shard range.

With the shard range, you can write logs by specifying the hash key, as well as splitting or merging shards. When reading data from a shard, you must specify the corresponding shard. When writing a shard, you can use Server Load Balancer or specify the Hash Key. When Server Load Balancer is used, each data packet is randomly written to any available shard. When a hash key is specified, data is written to the shard whose range is within the specified key.

For example, a Logstore has four shards and the MD5 value range of this Logstore is [00, FF). The following table lists shard ranges.

Shard No.	Range
Shard0	[00,40)
Shard1	[40,80)
Shard2	[80, C0)
Shard3	[C0, FF)

If you specify the MD5 key as 5F by specifying the hash key when writing logs, the log data is written to Shard1 that contains MD5 key 5F. If you specify the MD5 key as 8C, the log data is written to Shard2 that contains MD5 key 8C.

Shard read/write capacities

The read/write capacities of a shard are as follows:

- Write: 5 Mbit/s, 2000 times/s
- Read: 10 Mbit/s, 100 times/s

We recommend that you plan the number of shards based on the actual data traffic. If the traffic exceeds the read/write capacities, increase the number of shards by splitting the shard in a timely manner to achieve greater read/write capacities. If the traffic is far less than the maximum read/write capacities of shards, reduce the number of shards by merging the shards to save costs.

For example, assume that you have two shards in the read/write state and can write data at 10 Mbit/s at maximum. If you write data at 14 Mbit/s in real time, we recommend that you split a shard to increase the number of shards in the read/write state to three. If you write data at 3 Mbit/s, we recommend that you merge two shards into one, because one shard is sufficient for your needs.



Note

- If 403 or 500 errors are reported constantly during log writing, view the traffic and HTTPS status code and determine whether you need to increase the number of shards.
- For read/write operations that exceed the service capacities of shards, the system attempts to provide the needed services. However, the service quality cannot be guaranteed.

Status

The shard status includes:

- read/write: supports reading and writing data.
- read-only: only supports reading data.

When a shard is created, it is in the read/write state. Split or merge operations change the shard status to read-only and generate a new shard in the read/write state. The status of the shard does not affect its read performance. Shards in the read/write state support data writes, while shards in the read-only state does not support data writes.

When splitting a shard, you must specify a ShardId in the read/write state and an MD5. The MD5 must be greater than the shard BeginKey and smaller than the shard EndKey. A split operation can split two other shards from one. After a split operation, two more shards are added. After the split, the status of the original shard is changed from read/write to read-only. Data can still be consumed, but new data cannot be written. The two new shards are in the read/write state and arranged behind the original shard. The MD5 ranges of these two shards are within the range of the original shard.

When merging shards, you must specify a shard in the read/write state. The server automatically finds the adjacent shard at the right of the specified shard and merges these two shards. Make sure the specified shard is not the last shard in the read/write state. After the merge, the merged shards are in the read-only state. Data can still be consumed, but new data cannot be written. A new shard in the read/write state is generated and its MD5 range covers the total range of the original two shards.

18.4.2.5. Log topic

A log topic is used to classify logs in a Logstore. Topics can be specified when logs are written and serve as a filter when logs are queried. For example, you can use your user ID as the log topic when writing logs. In this way, you can choose to only view your own logs based on the log topic when querying logs. If you do not need to classify logs in a Logstore, use the same topic for all logs.

 **Note** An empty string is a valid log topic and is the default log topic when you are writing and querying logs. If you do not need to use log topics, use the empty string to write or query logs.

18.5. Scenarios

Log Service is applicable to the following scenarios: data collection, real-time computing, data warehousing and offline analysis, product operation and analysis, operations and maintenance, and management.

Data collection and consumption

LogHub provides low cost access to large amounts of real-time log data such as metrics, events, binary logs, text logs, and clicks.

Benefits:

- **Easy to use:** More than 30 real-time data collection methods are provided for you to quickly set up your platform and reduce O&M workload.
- **Automatically scalable:** Log Service scales based on traffic and business requirements, helping you handle traffic spikes and respond to growing business demands.

ETL and stream processing

LogHub can interconnect with various real-time computing and services to provide features such as complete progress monitoring and alerting. LogHub also can achieve SDK- and API-based custom consumption.

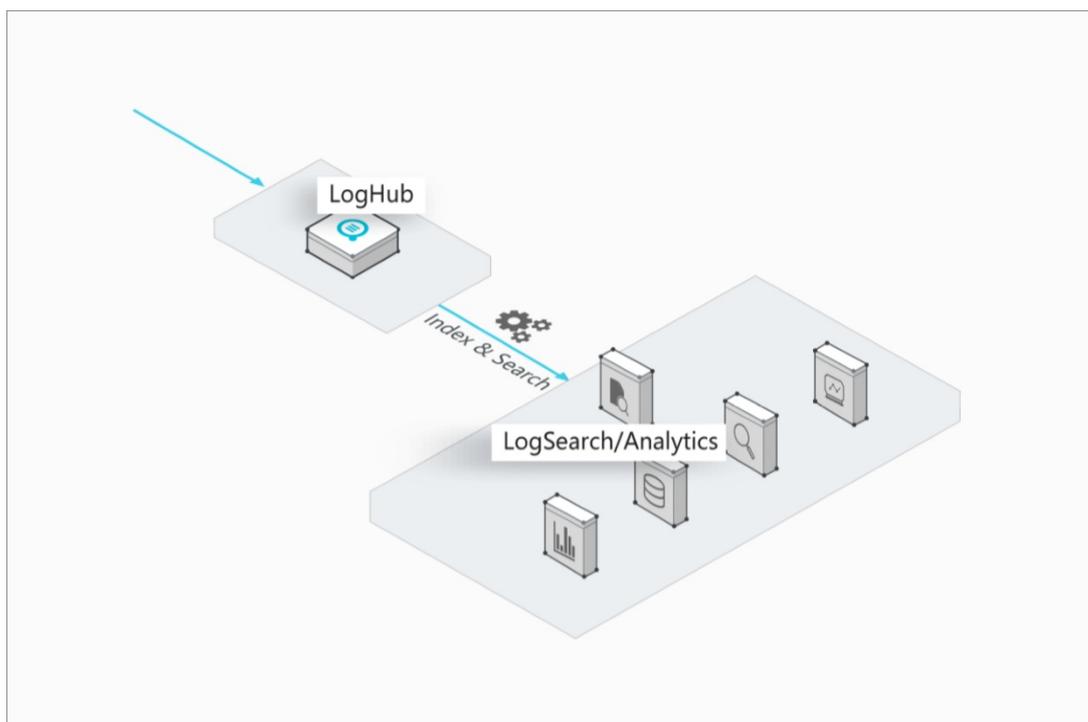
- **Easy operations:** LogHub provides SDKs in multiple programming languages and programming frameworks. It can interconnect with various stream computing engines.
- **Comprehensive features:** LogHub supports alert mechanisms and provides large amounts of monitoring data.
- **Elastic scaling:** PB-grade elasticity and zero latency.

LogSearch/Analytics

LogSearch/Analytics allows you to index Loghub data in real time and provides a variety of query methods such as keyword, fuzzy, context, range, and SQL aggregate queries.

- Strong real-timeliness: Data can be queried immediately after it is written.
- High efficiency at low cost: LogSearch/Analytics is able to index PBs of data each day. Costs are 85% lower compared with self-built systems.
- Strong analysis capability: LogSearch/Analytics supports multiple query methods and SQL for aggregation analysis. It also provides visualization and alerting capabilities.

LogSearch/Analytics



18.6. Limits

Resource limits

Item	Description	Remarks
Project	Up to 10 projects can be created in each department.	If you require additional quotas, submit a ticket.
Logstore	Up to 100 Logstores can be created for a project.	If you require additional quotas, submit a ticket.
Shard	<ul style="list-style-type: none"> Up to 10 shards can be created for a Logstore. However, you can split the shards to increase the number of shards. Up to 100 shards can be created for a project. 	If you require additional quotas, submit a ticket.
Dashboard	<ul style="list-style-type: none"> Up to five dashboards can be created for each project. Each dashboard can contain up to 10 charts. 	If you require additional quotas, submit a ticket.
Saved search	Up to 10 saved searches can be created for each project.	If you require additional quotas, submit a ticket.
Logtail configuration	Up to 100 Logtail configurations can be created in each project.	If you require additional quotas, submit a ticket.
Consumer group	Up to 10 consumer groups can be created for each project.	If you require additional quotas, submit a ticket.
Machine group	Up to 100 machine groups can be created for each project.	If you require additional quotas, submit a ticket.
Log retention time	Logs that are collected to the server can be kept for up to 365 days.	If you require additional quotas, submit a ticket.

18.7. Terms

Log

A log is an abstraction of changes that happen in a system. A log is a sequence of records ordered by time, which contains information about operations and results of specific objects. Log files, events, binary logs, and metrics are all different types of logs. Each log file is composed of one or more log entries. A log entry is the smallest unit of data that can be processed in Log Service. Each log entry describes a single system event.

Log group

A log group is a collection of logs. These groups are the basic units used for read and write operations.

Log topic

A log topic is used to classify logs in a Logstore. Topics can be specified when logs are written, and serve as a filter when logs are queried.

Project

A project is the management unit for resources in Log Service and is used to isolate and control resources. You can use a project to manage all the logs and related log sources of an application. A project is used to manage Logstores of a user and server configurations for log collection. A project also serves as the portal for a user to access the resources of Log Service.

Logstore

A Logstore is the unit used in Log Service for log data collection, storage and query. Each Logstore can belong to only one project, but multiple Logstores can be created for a single project.

Shard

A shard is the unit to compose Logstores. Each shard is represented by a non-overlapping, left-closed, right-open interval of MD5 values. The range of the Logstore is represented by the entire range of MD5 values of the shards in the Logstore.

19. Apsara Stack Security

19.1. What is Apsara Stack Security?

Apsara Stack Security is a solution that provides Apsara Stack with a full set of security features, such as network security, server security, application security, data security, security management, and security operations services.

In today's cloud computing environment, new technologies are developed every day. Border security protection methods that use traditional detection technologies are insufficient to secure cloud businesses. Apsara Stack Security combines the powerful data analysis capabilities of Alibaba Cloud with the expertise of the Alibaba Cloud security operations team. It provides integrated security protection services at the network layer, application layer, and server layer.

Apsara Stack Security protects core business applications that provide services for the Internet. It provides real-time protection capabilities, including distributed denial of service (DDoS) detection and prevention, Web attack detection and prevention, Web vulnerability detection and fix, server vulnerability detection and fix, and server intrusion prevention. Using a large amount of local security data and the intelligence collected from the cloud, this service performs big data analysis in the security data analysis engine cluster. It then presents security administrators with the overall security situation and intrusion tracing results, including targeted attack detection, staff intelligence leak alerts, and intrusion cause analysis. Based on this core security information, security administrators can understand the security status and use the custom analysis interface provided by the security data analysis engine to perform scenario-based analysis on security data for flexible customization of security analysis capabilities.

19.2. Advantages

As a pioneer of cloud security, Apsara Stack Security has received a variety of authoritative certifications. Through mature security systems and advanced security technologies, Apsara Stack Security can fully protect the security of the Apsara Stack environment.

Pioneer of cloud security

The Apsara Stack Security team has accumulated a wealth of security experience by protecting all internal business systems of Alibaba Group since 2005. Since its release in 2011, Apsara Stack Security has become a pioneer in providing comprehensive protection for cloud security.

Apsara Stack Security protects more than 40% of all websites in China. It prevents more than 50% of all distributed denial of service (DDoS) attacks and blocks up to 3.6 billion attacks every day. It has fixed over 6.13 million vulnerabilities over the last year.

Mature systems and advanced technologies

Apsara Stack Security is a product born from ten years of protection experience. After a decade of experience in providing security operations services for the internal businesses of Alibaba Group, Alibaba has obtained considerable security research achievements, security data, and security operations methods, and has built a professional cloud security team. Apsara Stack Security brings together the rich experience of these experts to develop the sophisticated systems that provide enhanced security for cloud computing platforms. This product can protect the cloud platform, cloud network environments, and cloud business systems of Apsara Stack users.

Comparison with traditional security products

Feature	Traditional security product	Apsara Stack Security
Comprehensive industry-leading security capabilities among Internet enterprises	A traditional security service provider only has limited products and features and cannot provide a comprehensive security protection system.	Alibaba has accumulated a large number of intelligence sources through years of attack prevention experience. This has allowed it to detect common Internet attacks including zero-day exploits, and provide comprehensive security capabilities.
Early risk detection	Traditional security service providers cannot detect risks due to a lack of complete monitoring systems.	Apsara Stack Security can detect and respond to critical vulnerabilities and security events quickly to prevent security issues.
Security big data modeling analysis	Traditional security service providers cannot detect threats through signature scanning. The traditional log analysis feature only provides data collection and reporting. It does not provide data modeling analysis.	Big data modeling analysis enables Apsara Stack Security to detect threats in the entire network and display the security data. More than 30 algorithmic models are used to analyze the historical data, network data, and server data. This enables security situation awareness.
Scalability and decoupling with hardware	Traditional security products are developed based on the existing hardware devices. Security product software also relies on the virtual machines on virtual platforms.	<ul style="list-style-type: none"> • Hardware and software decoupling: All modules are developed based on the cloud computing architecture and the common x86 hardware platform, and therefore do not rely on specific hardware. • Scalability: You can simply increase the amount of hardware for higher performance without the need to change the network architecture.
Collaboration between the network and servers	Traditional security service providers increase security features by adding devices. The devices can only collect device logs and status data and display the data on the management platform. They cannot collaborate to provide more features.	Apsara Stack Security provides complete Internet protection to ensure the security of networks, applications, and servers. The security modules interact with each other to form a comprehensive protection system that blocks attacks effectively.

Feature	Traditional security product	Apsara Stack Security
Compatibility with all data center environments and decoupling with specific cloud platforms	Most traditional security products are provided in hardware appliances. This makes the product incompatible with the cloud platforms based on Software Defined Network (SDN) technology.	Based on the interactions between servers and the operating system, Apsara Stack Security detects threats at the network perimeter through data analysis. This enables the service compatibility with all data center environments by avoiding the complex network topology inside the data centers.

19.3. Architecture

Apsara Stack Security consists of Apsara Stack Security Standard Edition and optional security services.

Apsara Stack Security Standard Edition

- **Traffic Security Monitoring:** This module is deployed on the network perimeter of Apsara Stack. It allows you to inspect and analyze each inbound or outbound packet of an Apsara Stack network by traffic mirroring. The analysis results are used by other Apsara Stack Security modules.
- **Server Intrusion Detection:** This module collects information and performs detection through the client deployed on physical servers. It detects file tampering, suspicious processes, suspicious network connections, suspicious port listening, and other suspicious activities on all servers in the Apsara Stack environment. This helps you detect server security risks in time.
- **Server Guard:** This module provides security protection features such as vulnerability management, baseline check, intrusion detection, and asset management for Elastic Compute Service (ECS) instances through log monitoring, file analysis, and signature scanning.
- **Web Application Firewall (WAF):** This module protects Web applications against common Web attacks defined by Open Web Application Security Project (OWASP), such as Structured Query Language (SQL) injections, cross-site scripting (XSS), exploit of Web server plugin vulnerabilities, Trojan upload, and unauthorized access. It blocks a large number of malicious visits to avoid website data leaks. This ensures the security and availability of your website.
- **Threat Detection Service (TDS):** This service collects traffic data and server information and detects potential intrusions or attacks through machine learning and data modeling. It detects vulnerability exploitation and new virus attacks launched by advanced attackers, and shows you the information of ongoing attacks, enabling business security visualization and awareness.

Apsara Stack Security Standard Edition also provides on-premises security operations services. On-premises security operations services help you make better use of the features of Apsara Stack products and Apsara Stack Security to ensure your application security.

Security operations services include pre-release security assessment, access control policy management, Apsara Stack Security product configuration, periodic security check, routine security inspection, and urgent event handling. These services cover the entire lifecycle of your businesses in Apsara Stack. On-premises security operations services help you create a security operations system for cloud businesses. This system enhances the security of application systems and ensures the security and stability of your businesses.

Optional security services

You can also choose the following optional service modules based on your own business needs to enhance your system security.

- **DDoS Traffic Scrubbing:** This module detects and filter out Distributed Denial of Service (DDoS) attack traffic to block DDoS attacks.

19.4. Features

Apsara Stack Security is developed based on the Apsara Stack environment and adopts a cloud security architecture that enables in-depth defense and multi-module collaboration. Our product is unlike traditional software and hardware security products. Apsara Stack Security provides comprehensive and integrated cloud security protection capabilities that protect the network layer, application layer, server layer, and other layers.

Apsara Stack Security Standard Edition features

Apsara Stack Security Standard Edition features describes the features provided by Apsara Stack Security Standard Edition.

Apsara Stack Security Standard Edition features

Module	Feature	Description
Traffic Security Monitoring	Traffic data collection and analysis	Uses a bypass in traffic mirroring mode to collect inbound and outbound traffic through the interconnection switch (ISW) and generates a traffic diagram.
	Malicious server detection	Detects attacks that are launched by malicious servers within the Apsara Stack network.
	Unusual traffic detection	Uses a bypass in traffic mirroring mode to detect the unusual traffic that has exceeded the scrubbing threshold.
	Web application protection	Uses a bypass to block common Web attacks at the network layer based on default Web attack detection rules.
Server Intrusion Detection	Key directory integrity check	Checks the integrity of files in a specific system directory such as <i>/etc/init.d</i> to detect file tampering and generate alerts.
	Suspicious process alerts	Detects suspicious processes and generates alerts.
	Suspicious port alerts	Detects new port listening tasks in time and generates alerts.
	Suspicious network connection alerts	Detects active connections with the public network and generates alerts.

Module	Feature	Description
Server Guard	Baseline check	Checks the security baselines of Elastic Compute Service (ECS) instances, including the account security, weak passwords, and at-risk configuration items. This ensures that the ECS instances comply with the security standards for enterprise servers.
	Vulnerability management	<ul style="list-style-type: none"> Scans for vulnerabilities in the software of ECS instances, and provides suggestions on vulnerability fixes. Provides quick fixes for critical vulnerabilities in applications and the operating system in your ECS instance, such as Web application vulnerability fixes and system file repair.
	Webshell detection and removal	Accurately detects and removes webshells based on rule matching, and allows you to manually quarantine webshells.
	Brute-force attack blocking	Detects and blocks brute-force attacks in real time.
	Unusual logon alerts	Detects unusual logons based on the approved logon settings and generates alerts.
	Suspicious server detection	Detects suspicious process activities such as reverse shells, java processes running CMD commands, and unusual file downloads using bash.
	Asset fingerprints	Collects information about the servers, including ports, accounts, processes, and software, to learn the server running status and perform event tracing.
	Log retrieval	Centrally manages server logs on processes, networks, and system logons. This allows you to quickly locate the cause of a problem by log retrieval.

Module	Feature	Description
Web Application Firewall (WAF)	Protection against common Web attacks	<p>Detects Structured Query Language (SQL) injections, cross-site scripting (XSS), intelligence, cross-site request forgery (CSRF), server-side request forgery (SSRF), Hypertext Preprocessor (PHP) deserialization, Java deserialization, Active Server Pages (ASP) code injections, file inclusion attacks, file upload attacks, PHP code injections, command injections, crawlers, and server responses.</p> <p>WAF provides five built-in protection templates, including the template with default protection policies, monitoring mode template, anti-DDoS template, template for financial customers, and template for Internet customers. WAF allows you to customize the decoding algorithms in the templates, enable or disable each attack detection module separately, and set the detection granularity.</p>
	HTTP flood mitigation	<p>Allows you to set access frequency control rules for domain names and URLs to restrict the access frequency of IP addresses or sessions that meet the criteria or block these IP addresses or sessions.</p> <p>Restricts the access frequency of known IP addresses or sessions or block these IP addresses or sessions.</p> <p>The HTTP flood mitigation rules do not apply to IP addresses or sessions that have been added to the whitelist.</p>
	Custom and precise access control	<p>Supports precise access control based on the following HTTP contents or their combinations: URI, GET parameters, decoded path, HOST header, complete cookie, POST parameters, complete body, HTTP status code, and response content.</p>
	Security situation overview	<p>Provides the overall security information, including the number of emergencies, attacks on the current day, flaws on the current day, attack trend, latest threat analysis, latest intelligence, and protected assets information.</p>

Module	Feature	Description
Threat Detection Service (TDS)	Access analysis	Analyzes all information about the access to the protected Web services, including the top 10 accessed services, number of normal source IP addresses, number of malicious source IP addresses, number of crawler source IP addresses, and detailed access samples.
	Screens	Provides map-based traffic data screens and server security screens.
	Security event analysis	Uses big data algorithms and models to detect zombies, brute-force attacks, backdoors, distributed denial of service (DDoS) attacks, hacking tools, suspicious network connections, unusual traffic, and other security events.
	Traffic data collection and analysis	Collects the traffic data in the monitored IP range, including traffic on the current day, traffic in the last 30 days, traffic in the last 90 days, and the queries per second (QPS). TDS also allows you to view the traffic data of a specific IP address.
	Malicious server detection	Detects attacks launched by internal malicious servers, such as HTTP flood and DDoS attacks, and identifies the controlled malicious servers.
	Web attack detection	Detects Web vulnerability exploitation, malicious scanning tools, webshell uploads and connections, SQL injections, XSS attacks, local and remote file inclusion attacks, code or command execution attacks, and other attacks.
	Server vulnerability exploitation detection	Converts packet feature characters into binary strings and matches these strings with signatures to detect security events such as the exploitation of Redis server vulnerabilities.
	Application vulnerability analysis	Detects Web application vulnerabilities and provides suggestions on vulnerability fixes and vulnerability fix verification. TDS periodically and automatically scans network address translation (NAT) assets and servers for application vulnerabilities, verifies the detected vulnerabilities, and updates the vulnerability status.
	Server vulnerability analysis	Detects server vulnerabilities and provides the scanning results and suggestions on vulnerability fixes.

Module	Feature	Description
	Weak password analysis	Detects weak passwords of accounts in common systems such as Web, SSH, FTP, MySQL, and SQL Server and allows you to customize weak password policies. TDS automatically scans NAT assets and servers at a scheduled time each day, verifies the detected weak passwords, and updates the weak password detection time.
	At-risk configuration detection	Scans the access of the external service pages, generates alerts on leaks of configuration items on webpages, verifies the detected leaks of configuration items at a scheduled time every day, and updates the detection time.

On-premises security operations services

Apsara Stack Security Standard Edition provides on-premises security operations services that ensure the security of your business systems. **On-premises security operations services** describes the included on-premises security operations services.

On-premises security operations services

Category	Service	Description
User business security operations	User asset research	Periodically researches your cloud businesses with your authorization and develops a business list containing information such as the business system name, ECS, Relational Database Service (RDS), IP address, domain name, and owner.
	New business security assessment	<ul style="list-style-type: none"> Detects system vulnerabilities and application vulnerabilities in the new business system by using both the automation tools and manual operations before you migrate a new business system to the cloud. Provides suggestions and verification on vulnerability fixes.
	Periodic business security assessment	<ul style="list-style-type: none"> Periodically uses automation tools to detect system vulnerabilities, application vulnerabilities, and security risks in running businesses. Provides suggestions on handling detected risks, including but not limited to security policy settings, patch updates, and application vulnerability handling.
	Access control management	Provides inspection and guidance on applying access control policies when a new business is migrated to the cloud.

Category	Service	Description
	Access control routine inspection	Periodically checks for access control risks of your businesses.
	Security risk routine inspection	Monitors and inspects security events in Apsara Stack Security, informs you of the verified events, and provides suggestions on event handling.
Apsara Stack Security maintenance	Rule update	Periodically updates the rules repository of Apsara Stack Security products.
	Product integration	<ul style="list-style-type: none"> Provides support for integrating Apsara Stack Security products with your application systems. Helps you customize and optimize security policies.
Security event response	Event alerts	Synchronizes security events information from Alibaba Cloud, and helps you remove the risks.
	Event handling	Handles urgent events such as attacker intrusions.

Optional services

The following table describes the optional services provided by Apsara Stack Security.

Optional services

Module	Feature	Description
DDoS Traffic Scrubbing	Traffic scrubbing against DDoS attacks	Detects and prevents attacks such as SYN flood, ACK flood, ICMP flood, UDP flood, NTP flood, DNS flood, and HTTP flood.
	DDoS attack display	Allows you to search for DDoS attacks by IP address, status, and event information.
	DDoS traffic analysis	Allows you to monitor and analyze the traffic of a DDoS attack, and view the attack traffic protocol and the top 10 IP addresses that have launched most attacks.

19.5. Restrictions

None

19.6. Terms

DDoS attacks

An attacker combines multiple computers by using the client-server model to form an attack platform and initiates a large number of valid requests to one or more targets from this platform to cause network failures. Distributed denial of service (DDoS) attacks are much stronger than common denial of service (DoS) attacks.

SQL injections

An attacker makes the server run malicious Structured Query Language (SQL) commands by inserting these commands in Web tables or inserting malicious strings in URL requests.

Traffic scrubbing

The traffic scrubbing service monitors the inbound traffic of a data center in real time and detects unusual traffic that may be from DDoS attacks and other attacks. This service scrubs the unusual traffic without affecting businesses.

Brute-force attacks

Brute-force attacks work by iterating through all possible combinations that can make up a password.

Webshells

A webshell is a script written in languages such as Active Server Pages (ASP) and Hypertext Preprocessor (PHP). Attackers can run a webshell on a Web server to perform risky operations. This enables attackers to obtain sensitive information or control the server through server penetration or privilege escalation.

Server intrusion detection

By analyzing server logs, Apsara Stack Security can detect attacks, such as system password cracking and logons from unusual IP addresses, and generate real-time alerts.

20. Key Management Service (KMS)

20.1. What is KMS?

Key Management Service (KMS) is a secure and easy-to-use key management service provided by Apsara Stack. KMS allows you to create and manage CMKs with ease and use a DEKs to encrypt your data.

KMS integrates many Alibaba Cloud products and services to help protect your data in the cloud.

KMS solutions describes how KMS provides solutions for a variety of concerns and issues.

KMS solutions

Role	Requirement	Solution
Application or website developer	<ul style="list-style-type: none"> • My program needs keys or certificates for encryption or signature, and I want secure and independent key management services. • I want to securely access keys regardless of where my application is deployed, and cannot take the risk of deploying plaintext keys elsewhere. 	KMS provides envelope encryption, allowing you to store the Customer Master Key (CMK) in KMS and deploy only the EDKs. You can simply call a KMS API to decrypt DEKs only when necessary.
Service developer	<ul style="list-style-type: none"> • I do not want to be responsible for securing users keys and data. • I want users to manage their own keys. I want to use specified keys to encrypt their data after obtaining their authorization. In this way, I can focus on developing service features. 	Envelop encryption and KMS APIs allow service developers to use specified CMKs to encrypt and decrypt DEKs. Plaintexts are not directly stored in a storage device. This method helps service developers manage CMKs.
Chief security officer (CSO)	<ul style="list-style-type: none"> • There are compliance requirements that I expect our key management activities to meet. • I need to ensure that keys are reasonably authorized and that the use of any keys is audited. 	KMS can connect to RAM for unified authorization management.

20.2. Benefits

Cost-friendly

Traditional key management solutions require the purchase of secure key management equipment to construct a secure physical environment, as well as the design and implementation of key management solutions and specifications. This mode leads to high costs in hardware and software.

KMS enables you to manage your keys on the cloud platform in a unified manner while minimizing hardware and software investment.

Ease of use

KMS uses the unified APIs and standard HTTPS for ease-of-use.

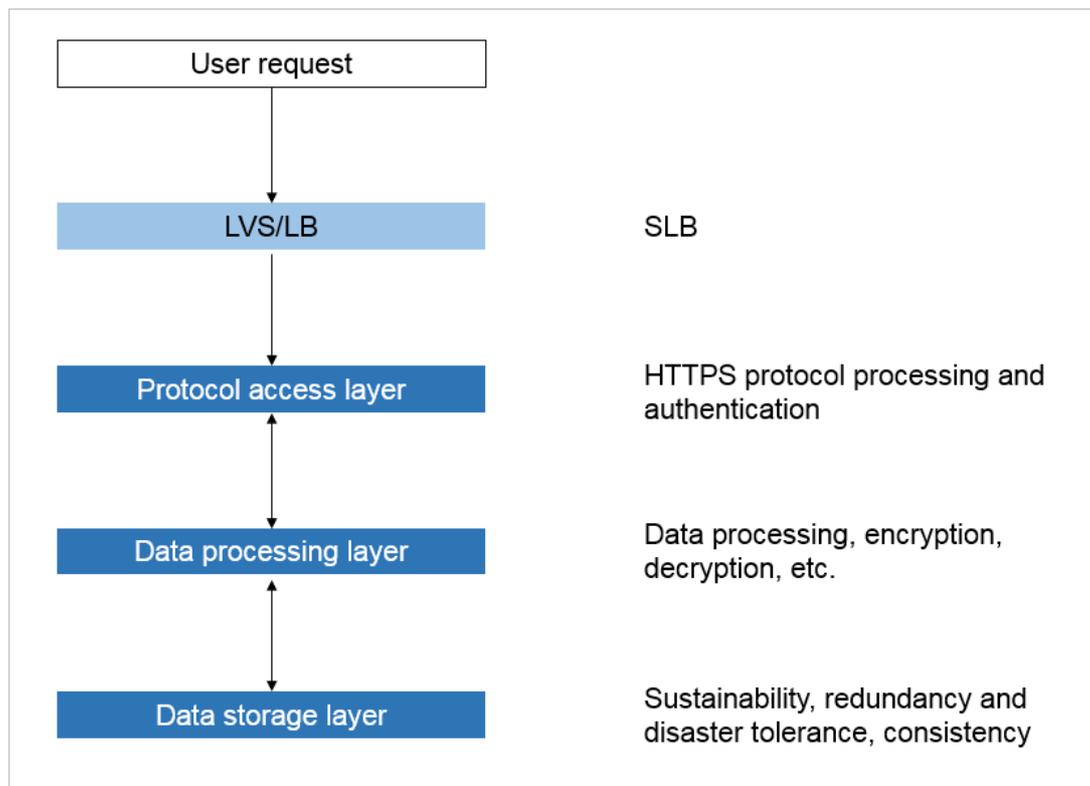
Reliability

KMS uses a distributed system to enhance reliability.

20.3. Architecture

The following figure shows the KMS architecture.

Architecture



- The protocol access layer of KMS receives HTTPS requests sent from a user to KMS, verifies the user identity, and authenticates the permission. After the verification and authentication succeed, the user request is forwarded to the data processing layer. The data processing layer receives the processing result and sends it to the user. If the verification and authentication fail, no data is processed and error information is returned.
- The data processing layer of KMS processes requests. Data processing in KMS involves cryptography-relevant operations such as encryption and decryption. The protocol access layer

and data processing layer communicate with each other based on RPC of TLS. The data processing layer adopts distributed deployment. The nodes are independent of each other. Requests sent from the protocol access layer can be properly processed on any node at the data processing layer.

- The storage layer of KMS stores core root keys, uses Raft to ensure data consistency, and uses TPM to implement persistent encrypted storage.

20.4. Features

The following table describes the features of KMS.

Feature	Description
Create a CMK	You need to create at least one CMK before KMS can be used. CMKs can be used to encrypt a small amount of data (less than 4 KB). However, in most cases, CMKs are used to call the <code>GenerateDataKey</code> API to generate DEKs.
Create a DEK	You must use a specified CMK to create a DEK when using KMS for envelop encryption. You can use the DEK to encrypt local data.
Encrypt data	You can use a specified CMK to encrypt a small amount of data (less than 4 KB) such as RSA keys, database passwords, or other sensitive user data.
Decrypt data	You can decrypt data encrypted through KMS.
View CMKs	You can obtain IDs of all CMKs that belong to the current region in your account.
View CMK details	You can view detailed information about a specified CMK, such as its creation date and time, description, globally unique identifier, CMK status, purpose, scheduled deletion time, creator, source of the CMK material, and expiration time of the CMK material.
Enable a CMK	You can change the status of a CMK from disabling to enabling.
Disable a CMK	If the status of a CMK is changed from enabling to disabling, the disabled CMK cannot be used to encrypt or decrypt data.
Schedule a CMK to be deleted	You can schedule a CMK to be automatically deleted after a specified deletion period.
Cancel scheduled key deletion	You can cancel the scheduled deletion of a CMK to change the CMK status back to enabling.
API	You can call HTTPS APIs to use KMS.
SDK	You can use SDKs in mainstream languages.

20.5. Scenarios

This topic describes the following typical scenarios of KMS:

- Use KMS to encrypt and decrypt data.
- Use envelope encryption to encrypt and decrypt data locally.

Example description

Example	Description	Example	Description
	CMK		The ciphertext key.
	The plaintext certificate.		The plaintext file.
	The ciphertext certificate.		The ciphertext file.
	The plaintext key.	-	-

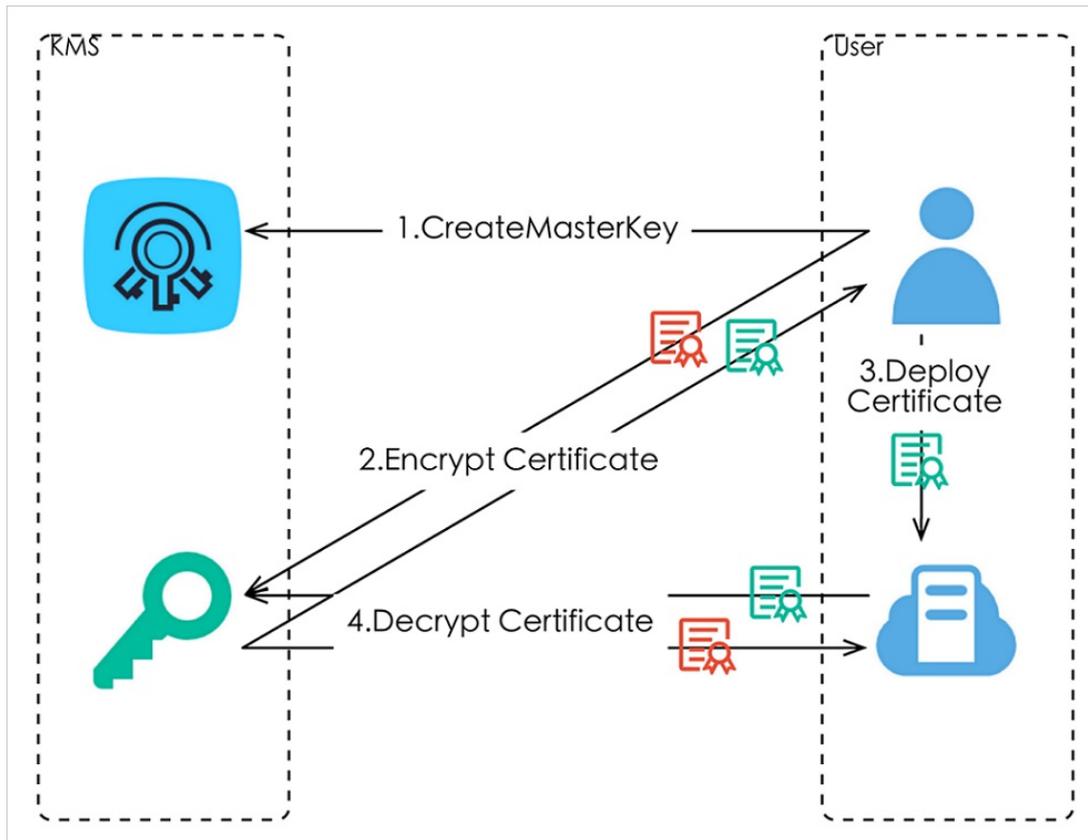
Directly use KMS to encrypt and decrypt data

You can directly call KMS APIs to encrypt and decrypt data with a specified CMK.

This scenario applies to the encryption and decryption of a small amount of data (less than 4 KB). Data is transmitted to and from, and encrypted or decrypted on the KMS server over secure channels.

Example: Encrypt the HTTPS certificate on the server, as shown in [Encrypt the HTTPS certificate on the server](#).

Encrypt the HTTPS certificate on the server



The procedure is as follows:

1. Create a CMK.
2. Call the Encrypt API to encrypt the plaintext certificate.
3. Deploy the ciphertext certificate on the server.
4. When the server has been started and needs the plaintext certificate, call the Decrypt API to decrypt the ciphertext certificate.

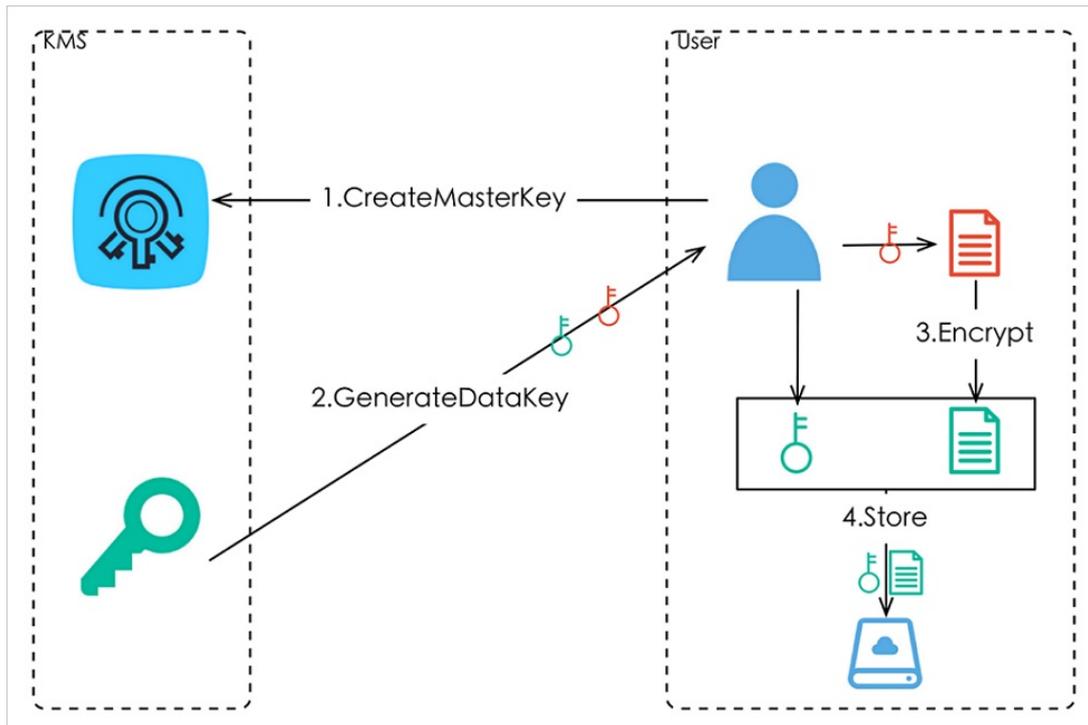
Use envelope encryption to encrypt and decrypt data locally

You can directly call a KMS API to use a specified CMK to generate and decrypt a DEK, and use the DEK to encrypt and decrypt data locally.

This scenario applies to encryption and decryption of large amounts of data that does not need to be transmitted over the network, which minimizes costs.

Example: Encrypt a local file, as shown in [Encrypt a local file](#).

Encrypt a local file

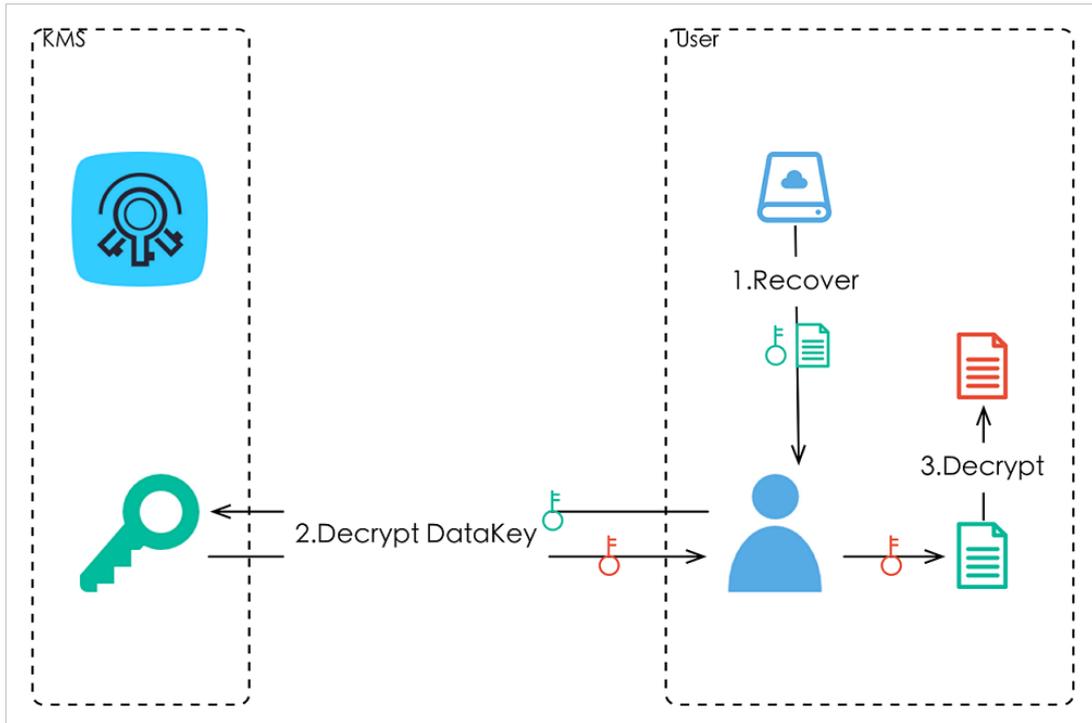


The encryption process is as follows:

1. Create a CMK.
2. Call the GenerateDataKey API to generate a DEK.
You can obtain a DEK and an EDK.
3. Use the DEK to encrypt the file and generate a ciphertext file.
4. Save the EDK and the ciphertext file to a persistent storage device or service.

Decryption process shows how to decrypt the encrypted file.

Decryption process



The decryption process is as follows:

1. Read the EDK and the ciphertext file from the persistent storage device or service.
2. Call the Decrypt API to decrypt the EDK and obtain the DEK.
3. Use the DEK to decrypt the file.

Notes

1. You must authenticate the Alibaba Cloud server HTTPS certificate to prevent phishers from stealing your information.
2. We recommend that you assign different permissions to users based on their CMKs.

20.6. Limits

A maximum of 200 CMKs can be created for a department.

20.7. Terms

envelope encryption

The practice of encrypting plaintexts by using a unique DEK, which is then encrypted with CMK. The EDK is stored and transferred directly over unsecured communication processes. You need to retrieve the EDK only when you need it.

customer master key (CMK)

A master key created by a user in Apsara Stack KMS, which is used to encrypt DEKs and generate EDKs. It can also be used to encrypt a small amount of data.

enveloped data key (EDK)/data encryption key (DEK)

EDK: the ciphertext key generated by using envelop encryption. DEK: the plaintext key used to encrypt data.

21. Apsara Stack DNS

21.1. What is Apsara Stack DNS?

Apsara Stack DNS is a service that runs on Apsara Stack and translates domain names. Based on the rules you have set, Apsara Stack DNS translates domain names that you have requested and direct requests from the client to the corresponding cloud services, business systems in enterprise internal networks, and services provided by Internet service providers.

Apsara Stack DNS provides basic domain name translation and scheduling services for VPC environments. You can perform the following operations through Apsara Stack DNS in your VPC:

- Access other ECS servers deployed in VPCs.
- Access cloud service instances provided by Apsara Stack.
- Access custom enterprise business systems.
- Access Internet services and business.
- Establish network connections between Apsara Stack DNS and user-created DNS through a leased line.

21.2. Benefits

Domain name management for enterprise domains

Apsara Stack DNS provides domain name management and translation services for enterprise domains.

- Apsara Stack DNS supports DNS resolution and reverse DNS resolution for domain names of cloud service instances, including ECS instance domain names.
- It also supports DNS resolution and reverse DNS resolution for your internal domain names.
- You can add, modify, and delete DNS records, including A, AAAA, CNAME, NS, MX, TXT, SRV, and PTR.
- You can add multiple DNS records, including A, AAAA, and PTR, for one host. By default, the resolution finds all matching records. The records can be randomly rotated to balance the load.

Flexible networking

Apsara Stack DNS provides the domain name forwarding service for enterprise domains, which allows you to flexibly create or combine networks.

- Supports forwarding all domain names.
- Supports forwarding specific domain names.

Access the Internet from your server

When the public network is accessible, Apsara Stack DNS supports recursive queries for public domain names and Internet domain names. This service allows your servers to access the Internet.

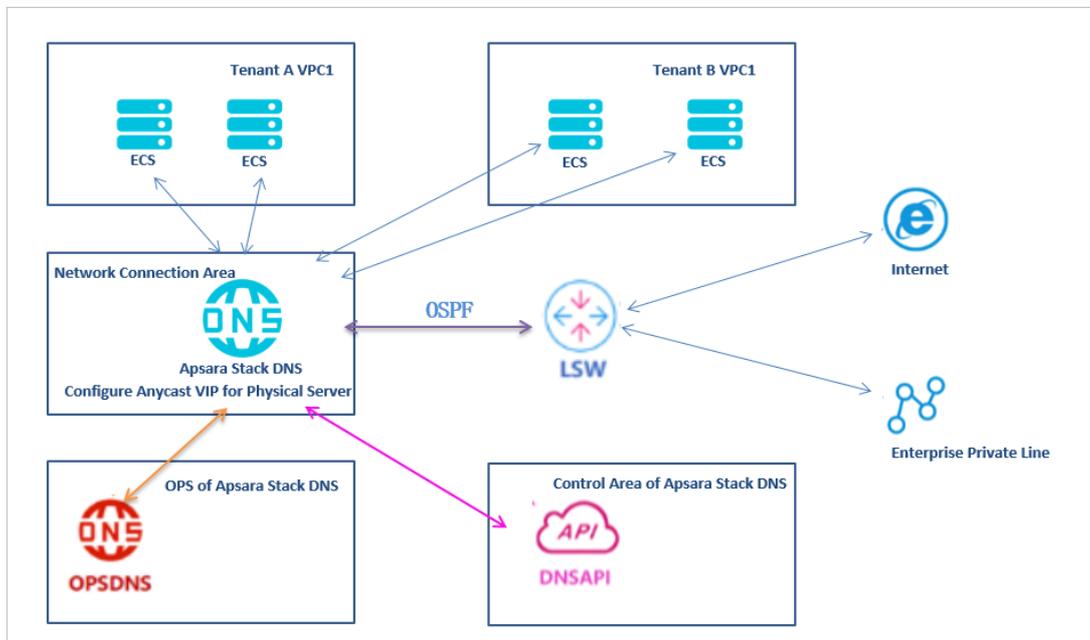
A unified management platform

The management system of Apsara Stack DNS is built on the unified management platform of Apsara Stack. You can use one account to manage all services. Apsara Stack DNS has the following benefits:

- Data management and service management support Web actions, which are easy to learn and operate.
- Apsara Stack DNS is deployed on clusters. You can add more clusters based on your needs.
- You can deploy Apsara Stack DNS in multiple zones. Apsara Stack DNS supports active-active deployment in the same city and disaster recovery deployment in the same city.
- Apsara Stack DNS is deployed based on anycast. High availability and disaster recovery can be automatically enabled.

21.3. Architecture

DNS architecture



The architecture of Apsara Stack DNS

- Deploys two physical servers for network connections and you can add more servers based on your needs.
- Uses two control interfaces for bond, which is uplinked to the ASW. The gateway is the default gateway of the internal network.
- Two service interfaces are uplinked to the LSW (ECMP is supported). These interfaces support OSPF to advertise anycast VIP routes, and are connected to the Internet.
- The control system is deployed in a container in the control area.

21.4. Features

Internal domain name management

Apsara Stack DNS provides data management for internal domain names. You can register, search, and delete internal domain names and add remarks. You can also add, delete, and modify DNS records. Supported DNS record types include A, AAAA, CNAME, NS, MX, TXT, SRV, and PTR.

Internal domain name management can translate internal domain names for servers deployed in a VPC. The DNS server addresses are deployed based on anycast, which ensures the continuity of services if errors occur.

Domain name forwarding management

Apsara Stack DNS can forward a specific domain name to other DNS servers for translation.

The domain name forwarding feature includes two forwarding modes: forward all requests (with recursion) and forward all requests (without recursion).

- Forward all requests (without recursion): Uses the target DNS server to translate domain names. If the domain names cannot be translated, or the request is timed out, a message is returned to the DNS client indicating that the query fails.
- Forward all requests (with recursion): Uses the target DNS server to translate domain names. If the domain names cannot be translated, then uses the local DNS server to translate them.

Recursive query management

Apsara Stack DNS supports recursive queries, which enables your servers to access the Internet.

Option configuration

You can enable, modify, or disable global default forwarding for Apsara Stack DNS.

21.5. Scenarios

Scenario A: Access cloud resources from a VPC environment

Apsara Stack DNS allows VPC-connected ECS or Docker instances to access Alibaba Cloud instances such as RDS, SLB, and OSS instances.

Scenario B: Access ECS hostnames from a VPC environment

If you need to define hostnames for your VPC-connected ECS and Docker instances according to your own rules, then use Apsara Stack DNS to remotely access and control the ECS instances and Docker instances using their hostnames.

Scenario C: Access the service domain name in the internal network from a VPC environment

If you need to develop your own SaaS service on Apsara Stack and assign a domain name that only allows internal access, Apsara Stack DNS helps you access the SaaS service through the domain name in a VPC environment.

Scenario D: Perform round-robin traffic redistribution for the internal network services provided by Apsara Stack

If you need to develop your own SaaS service on Apsara Stack and assign a domain name that only allows internal access, and this service is deployed in multiple zones or regions, Apsara Stack DNS helps you access your SaaS service in a VPC environment and redistribute traffic to different nodes.

Scenario E: Access the Internet from a VPC environment

You can use Apsara Stack DNS to access the Internet from a VPC environment.

Scenario F: Establish network connections among multiple networks on Apsara Stack

You can use Apsara Stack DNS to establish network connections between your internal network and Apsara Stack networks.

21.6. Limits

Apsara Stack DNS clusters have the following restrictions:

Restriction

Cluster	Module	Server type	Configuration requirement	Quantity requirement
Service cluster	Basic edition-resolution module	Q46S1.2B	Minimum configuration: 16-core CPU + 96 GB of memory + two GE ports + two 10-GE ports + 600 GB of hard disk	2
Control cluster	Basic edition-resolution module	Base container	Minimum configuration: 4-core CPU + 8 GB of memory + 60 GB of hard disk + network connection support	2

21.7. Basic concepts

DNS

Domain Name System (DNS) is a distributed database used for TCP/IP applications. It translates domain names into IP addresses, and selects paths for emails.

Domain name resolution

This is a process that translates domain names into IP addresses based on the DNS system. Domain name resolution includes authoritative DNS and recursive DNS.

Recursive DNS

Recursive DNS queries domain names cached on the local DNS server or sends a request to the authoritative resolver to obtain the corresponding IP addresses. You can use recursive DNS to translate Internet domain names.

Authoritative DNS

Authoritative DNS translates root domains, top-level domains, and other levels of domains.

Authoritative domain names

Authoritative domain names are domain names translated by the local DNS server. You can configure and manage DNS records on the local DNS server.

DNS forwarding

DNS forwarding uses two local DNS servers to provide DNS resolution services. One DNS server is used to configure and manage the domain name resolution data. The other DNS server is used to translate domain names.

Default forwarding

DNS queries for authoritative domain names are forwarded to another DNS server for resolution if they are not translated by the local DNS server.

22.API Gateway

22.1. What is API Gateway?

API Gateway provides a comprehensive suite of API hosting services that help you share capabilities, services, and data with partners in the form of APIs.

- API Gateway provides multiple security mechanisms to secure APIs and reduce the risks introduced by open APIs. These mechanisms include protection against replay attacks, request encryption, identity authentication, permission management, and throttling.
- API Gateway provides API lifecycle management that allows you to create, publish, and unpublish APIs, and improve API management and iteration efficiency.

API Gateway allows enterprises to reuse and share their capabilities with each other so that they can focus on their core business.

API Gateway



22.2. Features

API lifecycle management

- API lifecycle management enables you to manage APIs throughout their full lifecycle, including publishing, testing, and unpublishing APIs.
- API lifecycle management supports maintenance features such as routine management, version management, and quick rollback.

Comprehensive security protection

- API Gateway supports multiple authentication methods and HMAC (SHA-1 and SHA-256) algorithms.
- API Gateway supports HTTPS and SSL encryption.
- API Gateway provides multiple security mechanisms to prevent injections, replay attacks, and tempering.

Flexible access control

- Applications are used to make API requests. API Gateway implements access control for applications.
- If an application attempts to call an API, the application must first be authorized.
- API providers can authorize applications to call APIs.

Precise throttling

- You can use throttling to control API access frequency, application request frequency, and user request frequency.
- The unit of time for throttling can be set to minute, hour, or day.

Request validation

API Gateway validates parameter types and values by using range, enumeration, and regular expression. When an API request fails to be validated, API Gateway immediately rejects the request. This helps reduce the amount of back-end resources wasted on invalid requests and significantly lowers the processing costs of back-end services.

Data conversion

API Gateway enables you to configure mapping rules to translate front-end and back-end data.

- API Gateway supports data conversion for front-end requests.

22.3. Benefits

Easy maintenance

After you create APIs in API Gateway, API Gateway performs all the other API management functions. This significantly reduces routine maintenance costs.

Large scale and high performance

API Gateway uses a distributed deployment and automatic scaling model to respond to a large number of API access requests at very low latencies. It provides highly secure and efficient gateway functions for your backend services.

Security and stability

You can securely open your services to API Gateway on the intranet. API Gateway also provides enhanced permission management functions, and precise request throttling functions. It makes your services secure, stable, and controllable.

22.4. Concepts

It is important to familiarize yourself with the following basic concepts when you use API Gateway.

Application

An application defines the identity of an API caller. To call an API, you must first create an application.

AppKey and AppSecret

Each application has an AppKey and AppSecret pair. This pair is encrypted and attached to a request as the signature.

Encrypted signature

An encrypted signature is attached to each API request and is authenticated by API Gateway.

Authorize

The API service provider can open an API to an application by granting authorization to the application. Only authorized applications can call the specified API.

API lifecycle

The API service provider manages an API by stages, including creating an API, testing the API, publishing the API, unpublishing the API, and changing the version.

API definition

An API definition is a set of rules defined by the API service provider when creating an API. The API definition specifies the backend service, request format, received format, and returned format.

Parameter mapping

Parameter mapping is configured by the API service provider. It is used when the parameters in a request are inconsistent from those of the API backend service.

Parameter verification

Parameter verification is performed based on a set of rules defined by the API service provider. API Gateway filters out invalid requests based on these rules.

Constant parameter

API users do not have to input the constant parameters. The constant parameters are always received by the backend service.

System parameter

You can configure API Gateway to add certain system parameters such as CaClientIP (request IP address) to the requests sent to you by the backend service.

API group

An API group is a group of APIs that are managed by the API service provider as a whole. Before you create an API, you must first create an API group.

Second-level domain name

A second-level domain name is a domain name that you bind to an API group when creating the group. The second-level domain name is used to test API calling.

Independent domain name

An independent domain name is a domain name that you bind to an API group when opening an API in the group. Users must access the independent domain name to call the API.

Signature key

A signature key is created by the API service provider and bound to an API. The signature is added to each request sent from API Gateway to the backend service. The backend service checks the signature for security purposes.

Throttling policy

The API service provider can configure a throttling policy to limit the maximum number of requests for an API, and the maximum number of API requests that can be initiated by a user or an application. The throttling granularity can be day, hour, or minute.

23. MaxCompute

23.1. What is MaxCompute?

MaxCompute is a highly efficient, highly available, and low-cost EB-level big data computing service independently developed by Alibaba Cloud. This service is used within Alibaba Group to process large volumes (EBs) of data each day. MaxCompute is a distributed system oriented towards big data processing. As one of the core products in the Alibaba Cloud computing solution, the service is used mainly to store and compute structured data.

MaxCompute is designed to support multiple tenants, and provide data security and horizontal scaling. Based on an abstract job processing framework, the service provides centralized programming interfaces for various data processing tasks of different users.

MaxCompute is used to store and compute large volumes of structured data. It provides various data warehouse solutions as well as big data analysis and modeling services. MaxCompute is designed to provide an easier approach to analyze and process large amounts of data. You can analyze big data without deep knowledge about distributed computing.

MaxCompute has the following features:

- Uses a distributed architecture that can be scaled as needed.
- Provides an automatic storage and fault tolerance mechanism to ensure high data reliability.
- Allows all computing tasks to run in sandboxes to ensure high data security.
- Uses RESTful APIs to provide services.
- Can upload or download high-concurrency, high-throughput data.
- Supports two service models: the offline computing model and the machine learning model.
- Supports data processing methods based on programming models such as SQL, MapReduce, Graph, and MPI.
- Supports multiple tenants, allowing multiple users to collaborate on data analysis.
- Provides user permission management based on ACLs and policies, allowing you to configure flexible data access control policies to prevent unauthorized access to data.
- Provides Spark on MaxCompute for enhanced application.
- Provides Elasticsearch on MaxCompute for enhanced application.
- Supports access and processing of unstructured data.

23.2. Benefits

China's only big data cloud service and real data sharing platform

- Warehousing, mining, analysis, and sharing of data can all be performed on the same platform.
- Alibaba Group implements this unified data processing platform in several of its own products such as Aliloan, Data Cube, DMP (Alimama), and Yu'e Bao.

Support for large numbers of clusters, users, and concurrent jobs

- A single cluster can contain more than 10,000 servers and maintain 80% linear scalability.
- A single MaxCompute instance can support more than 1 million servers in multiple clusters without restrictions (linear expansion is slightly affected). It supports the local multi-IDC mode.
- A single MaxCompute instance supports over 10,000 users, over 1,000 projects, and over 100 departments (of multi-tenants).
- A single MaxCompute instance supports more than 1 million jobs (daily submitted jobs on average) and more than 20,000 concurrent jobs.

Big data computing at your fingertips

You do not have to worry about the storage difficulties and the prolonged computing time caused by the increasing data volume. MaxCompute automatically expands the storage and computing capabilities of clusters based on the volume of data to process, allowing you to focus on maximizing the efficiency of data analysis and mining.

Out-of-the-box service

You do not have to worry about cluster construction, configuration, and O&M. Only a few simple steps are required to upload data, analyze data, and obtain analysis results in MaxCompute.

Secure and reliable data storage

User data is protected against loss, theft, and exposure by the multi-level data storage and access security mechanisms. These mechanisms include multi-replica technology, read/write request authentication, and application and system sandboxes.

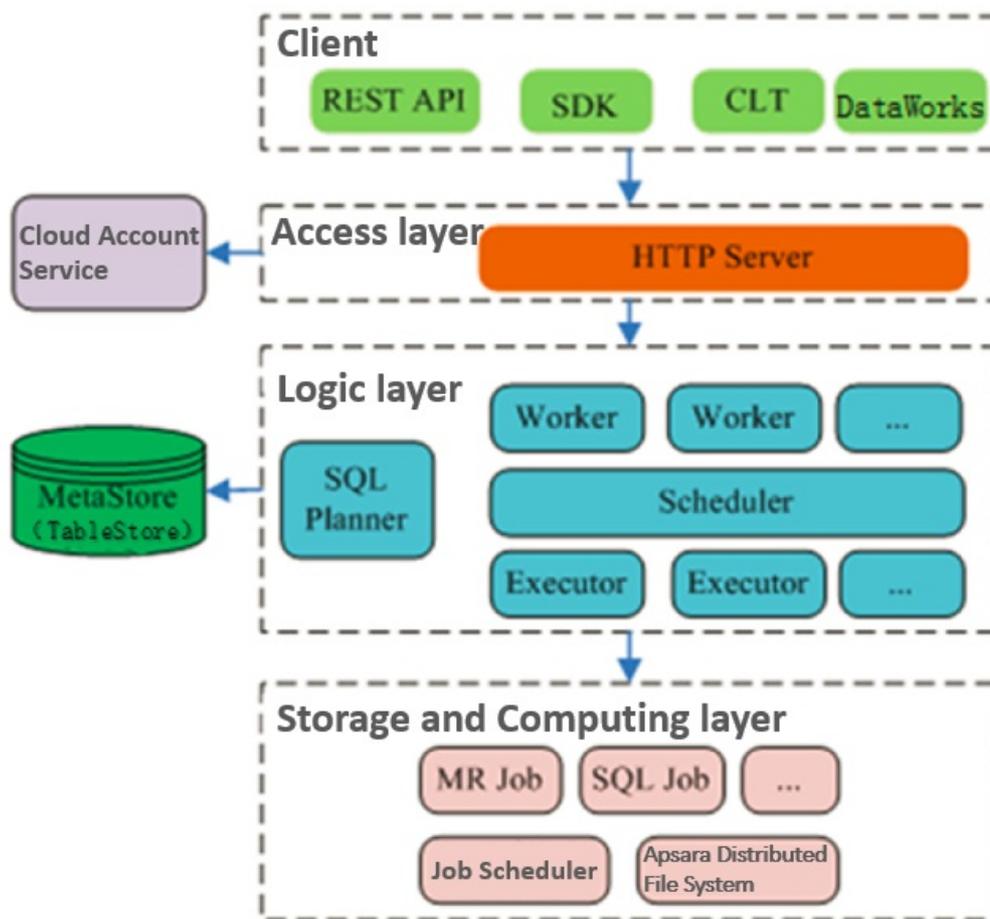
Multi-tenancy for multi-user collaboration

You can have multiple data analysts in your organization to work together by configuring different data access policies, while ensuring that each analyst can only access data within their own permissions. This maximizes work efficiency while ensuring data security.

23.3. Architecture

MaxCompute architecture shows the MaxCompute architecture.

MaxCompute architecture



The MaxCompute service is divided into four parts: client, access layer, logic layer, and storage and computing layer. Each layer can be scaled out.

The following methods can be used to implement the functions of a MaxCompute client:

- **API:** RESTful APIs are used to provide offline data processing services.
- **SDK:** RESTful APIs are encapsulated in SDKs. SDKs are currently available in programming languages such as Java.
- **Command line tool (CLT):** This client-side tool runs on Windows and Linux. CLT allows you to submit commands to manage projects and use DDL and DML.
- **DataWorks:** DataWorks provides upper-layer visual ETL and BI tools that allow you to synchronize data, schedule tasks, and create reports.

The access layer of MaxCompute supports HTTP, HTTPS, load balancing, user authentication, and service-level access control.

The logic layer is at the core of MaxCompute. It supports project and object management, command parsing and execution logic, and data object access control and authorization. The logic layer is divided into control and compute clusters. The control cluster manages projects and objects, parses queries and commands, and authorizes access to data objects. The compute cluster executes tasks. Both control and compute clusters can be scaled out as required. The control cluster is comprised of three different roles: Worker, Scheduler, and Executor. These roles are described as follows:

- **The Worker role processes all RESTful requests and manages projects, resources, and jobs.** Workers forward jobs that need to launch Fuxi tasks (such as SQL, MapReduce, and Graph jobs)

to the Scheduler for further processing.

- The Scheduler role schedules instances, splits instances into multiple tasks, sorts tasks that are pending for submission, and queries resource usage from FuxiMaster in the compute cluster for throttling. If there are no idle slots in Job Scheduler, the Scheduler stops processing task requests from Executors.
- The Executor role is responsible for launching SQL and MapReduce tasks. Executors submit Fuxi tasks to FuxiMaster in the compute cluster and monitor the operating status of these tasks.

In summary, when you submit a job request, the Web server at the access layer queries the IP addresses of registered Workers and sends API requests to randomly selected Workers. The Workers then send these requests to the Scheduler for scheduling and throttling. Executors actively poll the Scheduler queue. If the necessary resources are available, the Executors start executing tasks and return the task execution status to the Scheduler.

The storage and computing layer of MaxCompute is a core component of the proprietary cloud computing platform developed by Alibaba Cloud. The architecture diagram illustrates only major modules.

23.4. Features

23.4.1. Tunnel

23.4.1.1. Terms

Tunnel is the data tunnel service provided by MaxCompute. You can use Tunnel to import data from various heterogeneous data sources into MaxCompute or export data from MaxCompute. As the unified channel for MaxCompute data transmission, Tunnel provides stable and high-throughput services.

Tunnel provides RESTful APIs and Java SDKs to facilitate programming. You can upload and download only table data (excluding view data) through Tunnel.

23.4.1.2. Tunnel features

- The channel through which data flows in to and out of MaxCompute
- Highly concurrent upload and download
- Horizontal expansion of service capabilities
- Up to 1 PB throughput per day
- Batch and real-time upload modes
- Support for publishing and subscription models in real-time mode
- Tools based on MaxCompute Tunnel, such as TT, CDP, Flume, and Fluentd
- Support for reads and writes of tables (excluding views)
- Support for data writes in append mode
- Concurrency capabilities to improve total throughput
- Avoiding of frequent submissions
- Support for data upload only when target partitions exist
- Real-time upload mode

23.4.1.3. Data upload and download through Tunnel

Tunnel commands

```
odps@ > tunnel upload log.txt test_project.test_table/p1="b1",p2="b2 ";
```

```
odps@ > tunnel download test_project.test_table/p1="b1",p2="b2" log.txt;
```

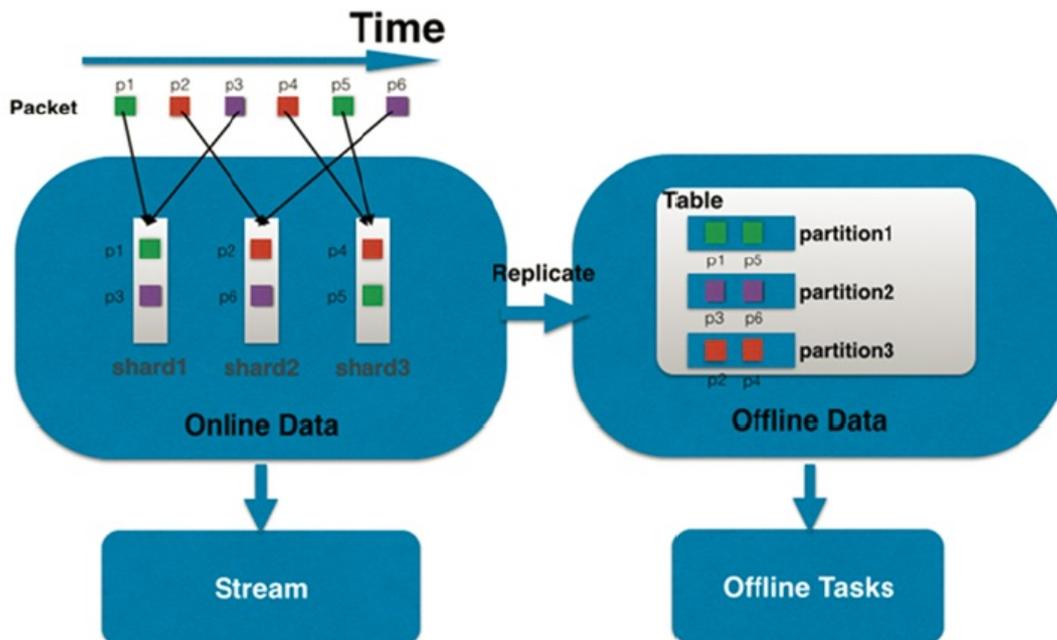
Notes

- Tunnel is a CLT based on the Tunnel SDK and can be used to upload local text files to MaxCompute or download data tables to your local device.
- You must create table partitions before using Tunnel.
- DataX, CDP, and TT provide enhanced Tunnel-based tools, which are used to exchange data between MaxCompute and relational databases.
- You can import log data by using the Flume and Fluentd tools.
- In some scenarios, you can develop custom tools based on Tunnel.

Real-time upload

- Upload in small batches
- High QPS performance
- Latency within milliseconds
- Subscription available

Real-time upload



23.4.2. SQL

23.4.2.1. Terms

The syntax of MaxCompute SQL is similar to SQL. It can be considered as a subset of standard SQL. However, MaxCompute SQL is not equivalent to a database, because it does not possess many characteristics that a database has, such as transactions, primary key constraints, and indexes. The maximum SQL statement size currently allowed in MaxCompute is 2 MB.

MaxCompute SQL offline computing is applicable to scenarios that have a large amount of data (measured in TBs) and that do not have high real-time processing requirements. It takes a relatively long time to prepare and submit each job. Therefore, MaxCompute SQL is not optimal for services that need to process thousands of transactions per second. MaxCompute SQL online computing is applicable to scenarios that require near-real-time processing.

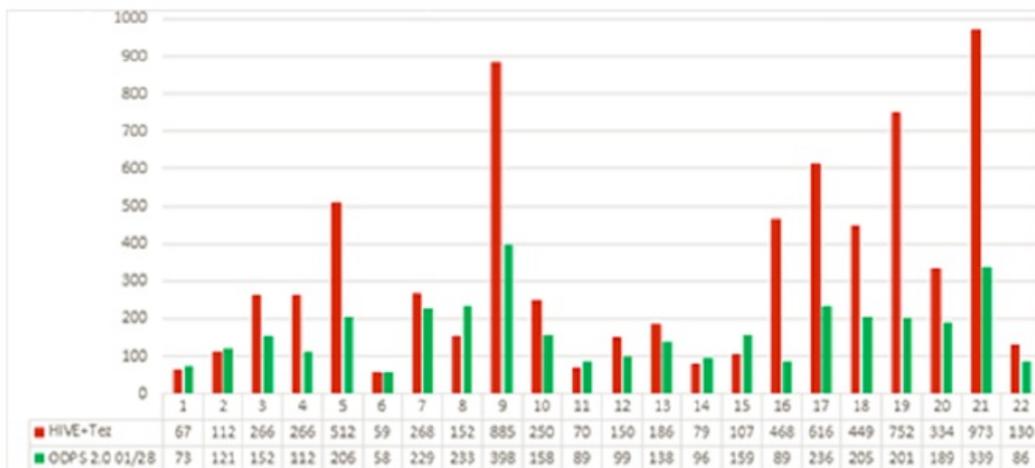
23.4.2.2. SQL characteristics

- It is suitable for processing large volumes of data (TBs or PBs).
- It has relatively high latency. The runtime of each SQL statement ranges from dozens of seconds to several hours.
- Its syntax is similar to that for Hive HQL. It is extended based on standard SQL syntax.
- It does not involve transactions or primary keys.
- It does not support UPDATE and DELETE operations.

23.4.2.3. Comparison with open source products

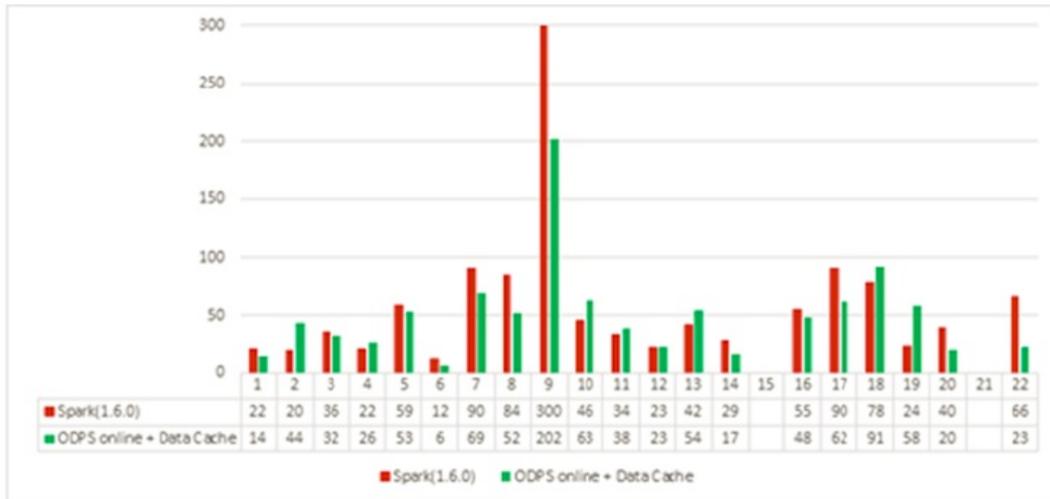
- TPC-H 1 TB data benchmark: Compared with Hive (Apache Hive-1.2.1-bin + Tez-UI-0.7.0 with CBO), MaxCompute has a 95.6% improvement in performance.

MaxCompute 2.0 VS Hive



- TPC-H 450 GB data benchmark: Compared with Spark SQL V1.6.0 (the latest release), MaxCompute has a 17.8% improvement in performance.

MaxCompute 2.0 VS Spark SQL



23.4.3. MapReduce

23.4.3.1. Terms

MapReduce is a programming model, which is basically equivalent to Hadoop MapReduce. The model is used for parallel MaxCompute operations on large-scale data sets (measured in TBs).

MaxCompute provides a MapReduce programming interface. You can use Java APIs, which is provided by MapReduce, to write MapReduce programs for processing data in MaxCompute.

? **Note** All data in MaxCompute is stored as tables. The inputs and outputs of MaxCompute MapReduce can only be tables. Custom output formats are not supported, and no interface, such as a file system, is provided.

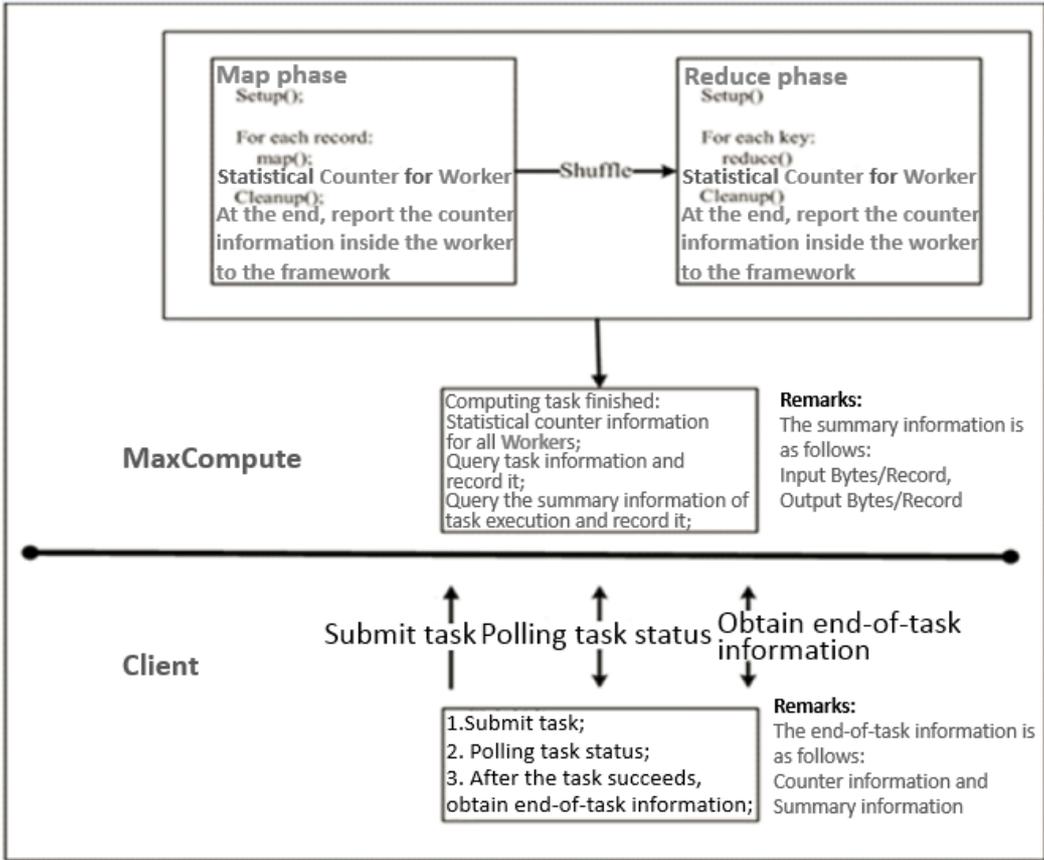
23.4.3.2. MapReduce characteristics

- It only supports the input and output of MaxCompute built-in data types.
- It supports the input and output of multiple tables to different partitions.
- It reads resources.
- It does not support using views as data inputs.
- It supports MapReduce programming only in the JDK 1.8 environment.
- It provides a limited sandbox security environment.

23.4.3.3. MaxCompute MapReduce process

The following figure shows the MapReduce process in MaxCompute:

MapReduce process



23.4.3.4. Hadoop MapReduce VS MaxCompute MapReduce

The following table describes the comparison between Hadoop MapReduce and MaxCompute MapReduce.

Mapper/Reducer

Mapper/Reducer	
Hadoop MapReduce	MaxCompute MapReduce
Map (InKey key, InputValue value, OutputCollector<OutKey, OutValue> output, Reporter reporter)	Map (long key, Record record, TaskContext context)
Reduce (InKey key, Iterator<InValue> values, OutputCollector<OutKey, OutValue> output, Reporter reporter)	Reduce (lRecord key, Iterator<Record> values, TaskContext context)

MapReduce

```
@Override
public void map(long recordNum, Record record, TaskContext context)
    throws IOException {
    for (int i = 0; i < record.getColumnCount(); i++) {
        word.set(new Object[] { record.get(i).toString() });
        context.write(word, one);
    }
}
```

23.4.4. Graph

23.4.4.1. Terms

Graph is the computing framework of MaxCompute designed for iterative graph processing. It provides programming interfaces similar to Pregel, allowing you to use Java SDKs to develop efficient machine learning and data mining algorithms.

Graph jobs use graphs to build models. This process outputs a result after performing iterative graph editing and evolution.

23.4.4.2. Graph characteristics

- It is a graphic computing programming model (similar to Google Pregel).
- It loads data to the memory, which is superior in multiple iteration scenarios.
- It can be used to develop machine learning algorithms.
- It can support 10 billion vertices and 150 billion edges.
- Its typical applications include:
 - PageRank
 - K-means clustering
 - Level 1 and level 2 relationships and shortest path
- Graph jobs process graph data.
- The original data is stored in tables. The user-defined Graph Loader loads data in the table as vertexes and edges.
- It supports iterative computing.

23.4.4.3. Graph relational network models

A relational network engine provides a variety of business-oriented relational network models. It helps you quickly implement relational data mining at finer granularities.

Community discovery

- Input to the engine: relational data.
- Engine output: IDs and community IDs.
- Computing logic: locates N communities with the optimal global network connection. The communities are close enough internally, and sparse enough in between.

Semi-supervised category

- Input to the engine: problematic IDs.
- Engine output: potentially problematic IDs and weights.
- Computing logic: uses existing problematic IDs (of one or more categories) to determine potential problematic IDs of the same or multiple categories and corresponding weights based on the entire network connection relationships.

Isolated point detection

- Input to the engine: relational data.
- Engine output: isolated points and weights.
- Computing logic: determines whether there are relatively isolated nodes using the connection relationships in a relational network, and generates the result.

Key point mining

- Input to the engine: relational data.
- Engine output: key point IDs and categories.
- Computing logic: calculates the key type nodes in a computing network using the connection relationships (such as centrality, influence, and betweenness centrality) in a relational network.

Level N relationships

- Input to the engine: relational data.
- Engine output: retrievable relational networks.
- Computing logic: manages multi-dimensional relationships using the connection relationships in the relational network, and creates indexes to facilitate the query for specific associations of an ID.

23.4.5. Unstructured data processing (integrated computing scenarios)

Alibaba Cloud introduced the MaxCompute-based unstructured data processing framework so that MaxCompute SQL commands can directly process external user data, such as unstructured data from OSS. You are no longer required to first import data into MaxCompute tables.

You can run a simple DDL statement to create an external table in MaxCompute and associate MaxCompute tables with external data sources. This table can then act as an interface between MaxCompute and external data sources. The external table can be accessed in the same way as a MaxCompute table, and computed by MaxCompute SQL.

MaxCompute allows you to process the following data sources by creating external tables:

- Internal data sources: OSS, Table Store, AnalyticDB, ApsaraDB for RDS, HDFS (Alibaba Cloud), and TDDL.
- External data sources: HDFS (Open Source), ApsaraDB for MongoDB, and Hbase.

23.4.6. Unstructured data processing in MaxCompute

MaxCompute has the following problems when processing unstructured data: MaxCompute stores data as volumes and must export generated unstructured data to an external system for processing.

To alleviate these problems, MaxCompute uses external tables to enable connections between MaxCompute and various data types. MaxCompute uses external tables to read and write data volumes as well as process unstructured data from external sources such as OSS.

23.4.7. Enhanced features

23.4.7.1. Spark on MaxCompute

23.4.7.1.1. Terms

Spark on MaxCompute is a solution developed by Alibaba Cloud to enable seamless use of Spark on the MaxCompute platform, extending the functions of MaxCompute.

Spark on MaxCompute provides a native Spark user experience with its native Spark components and APIs. It allows access to MaxCompute data sources and better security for multi-tenant scenarios. It also offers a management platform enabling Spark jobs to share resources, storage, and user systems with MaxCompute jobs. This guarantees high performance and low costs. Spark can work with MaxCompute to create better and more efficient data processing solutions. Spark Community applications can run seamlessly in Spark on MaxCompute.

Spark on MaxCompute has an independent data development node in DataWorks and supports data development in DataWorks.

23.4.7.1.2. Features of Spark on MaxCompute

Processing of data from MaxCompute and unstructured data sources

- Processes MaxCompute tables through APIs based on Scala, Python, Java, and R programming languages.
- Processes MaxCompute tables through components such as Spark SQL, Spark MLlib, GraphX, and Spark Streaming.
- Can process unstructured data from Alibaba Cloud OSS.

User-friendly experience and management functions

- Supports job submission in a way similar to Spark on YARN. Spark on MaxCompute is compatible with YARN and HDFS APIs.
- Supports components including Spark SQL, Spark MLlib, GraphX, and Spark Streaming.
- Can work with SQL and Graph components of MaxCompute to form optimized solutions.
- Can connect to the native Spark UI.
- Allows you to directly use the powerful management functions of MaxCompute.
- Supports not only Spark Community but also tools such as client, Livy, and Hue.

Scalability

Spark and MaxCompute share cluster resources. Spark resources can be scaled from large-scale MaxCompute clusters.

23.4.7.1.3. Spark features

The following table describes Spark on MaxCompute features.

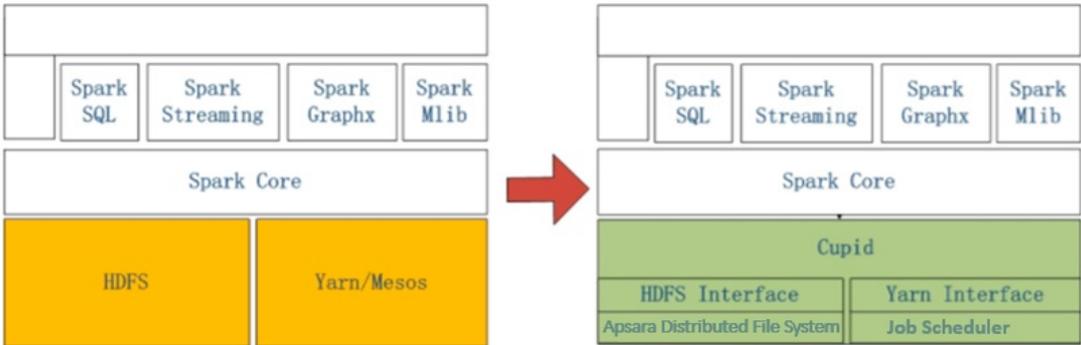
Features

Type	Feature	Description
Distributed cluster	Cluster deployment Cluster monitoring	Provide an O&M platform to monitor clusters and nodes.
Data processing component	Support for components such as Spark SQL, Spark MLlib, GraphX, and Spark Streaming	Provide native Spark components.
Job management	Centralized resource management, life cycle management, and authentication	The features are available through compatible YARN APIs.
Data sources	Unstructured data Table data sources in MaxCompute	Provide data processing capabilities of SQL and MapReduce on MaxCompute.
Security management	User identification, data authentication, and multi-tenant job isolation	Harden Spark security through authentication and sandboxes.

23.4.7.1.4. Spark architecture

The following figure shows the architectural comparison between Spark on MaxCompute and native Spark.

Architectural comparison between Spark on MaxCompute and native Spark



Note On the left is the native Spark architecture and on the right is the Spark on MaxCompute architecture.

As shown in the figure, Spark on MaxCompute has the computing capabilities of native Spark and the functions related to management, O&M, scheduling, security, and data interconnection. The management function of Spark is implemented by starting a Cupid Task instance of MaxCompute. The resource application function is realized through layer-1 YARN APIs provided by MaxCompute. The security function is offered through the sandbox mechanism of MaxCompute. The processing of and interconnection between data and metadata are also made available. The module details are described as follows:

- The MaxCompute control cluster starts a Spark driver by using the Cupid Task instance. The Spark driver uses YARN APIs to apply for resources from FuxiMaster, the central resource manager.
- The MaxCompute control cluster manages user quota consumed by running Spark instances, life cycles of Spark instances, and permissions on accessible data sources.
- The MaxCompute computing cluster starts a Spark driver and Executor as parent and child processes and executes Spark code in the sandbox of MaxCompute, ensuring security in multi-tenant scenarios.
- MaxCompute allows you to use the native Spark UI through its Proxy Server and manage job information through its management components.

23.4.7.1.5. Benefits of Spark on MaxCompute

Support for the complete Spark ecosystem

Provides consistent user experience with that of open source Spark.

Full integration with MaxCompute

Implements centralized management of resources, data, and security features for both Spark and MaxCompute.

Combination of Spark and the Apsara system

Combines the flexibility and ease of use of Spark with the high availability, scalability, and stability of the Apsara system.

Support for multi-tenancy

Reduces costs by centrally scheduling resources in large-scale clusters and ensuring high performance of physical machines.

Support for cross-cluster scheduling

Maximizes the efficiency of cluster resources by effectively allocating clusters and scheduling resources across clusters.

Support for real-time scaling of Spark resources

Scales resources in Spark Community in real time to better utilize resources and avoid waste.

 **Note** Real-time Spark resource scaling is not enabled on all MaxCompute clusters. To use this function, contact the MaxCompute team.

23.4.7.2. Elasticsearch on MaxCompute

23.4.7.2.1. Overview

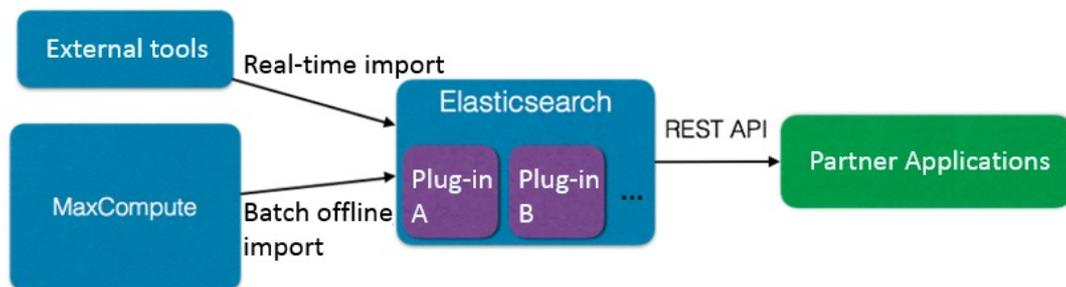
Elasticsearch on MaxCompute is an enterprise-class full-text retrieval system developed by Alibaba Cloud to retrieve large volumes of data with near-real-time search performance.

Elasticsearch on MaxCompute provides elastic full-text retrieval and supports native Elasticsearch APIs. You can import data from heterogeneous data sources and perform O&M for clusters and services. The centralized scheduling and management capabilities of MaxCompute allow Elasticsearch to provide more efficient core services for data retrieval at large volumes.

Elasticsearch on MaxCompute can also work with plug-ins available from the Elasticsearch open source community to enhance retrieval functions.

Elasticsearch on MaxCompute allows you to use tools to import data from external sources in real time. You can also import offline data from MaxCompute. After the imported data is indexed, Elasticsearch on MaxCompute provides retrieval services through RESTful APIs. The following figure shows its usage.

Elasticsearch on MaxCompute usage



23.4.7.2.2. Features of Elasticsearch on MaxCompute

Distributed cluster architecture

- Improves retrieval and reliability of data with a distributed architecture.
- Supports elastic scaling.
- Supports dynamic scaling.
- Supports service-level O&M and monitoring.

Robust full-text retrieval

- Performs full-text retrieval at the word, phrase, sentence, and section levels.
- Available in languages such as Chinese and English.
- Provides precise word segmentation with 100% recall for Chinese information retrieval.
- Supports complex searching methods, such as Boolean retrieval, proximity search, and fuzzy search.
- Sorts search results by relevance, field, and custom weight, and allows for secondary sorting.
- Performs statistical classification and analysis of search results.
- Allows real-time indexing and retrieval, so that inserted data can be retrieved immediately.
- Allows an index to be used multiple times after it is created.
- Allows you to modify the index structure in real time or rebuild the index to re-distribute data.

Support for multiple data sources

- Imports data from native Elasticsearch interfaces.
- Provides data import tools for MaxCompute.
- Supports full and incremental update.

Reliability

- Stores data in multiple copies, preventing user data from being lost during the downtime of machines.
- Implements a high availability architecture and comprehensive failover for nodes and services.
- Provides comprehensive O&M and monitoring functions.
- Authenticates access to protect data from malicious operations and ensure security.

23.4.7.2.3. Elasticsearch features

Elasticsearch on MaxCompute features are described as follows:

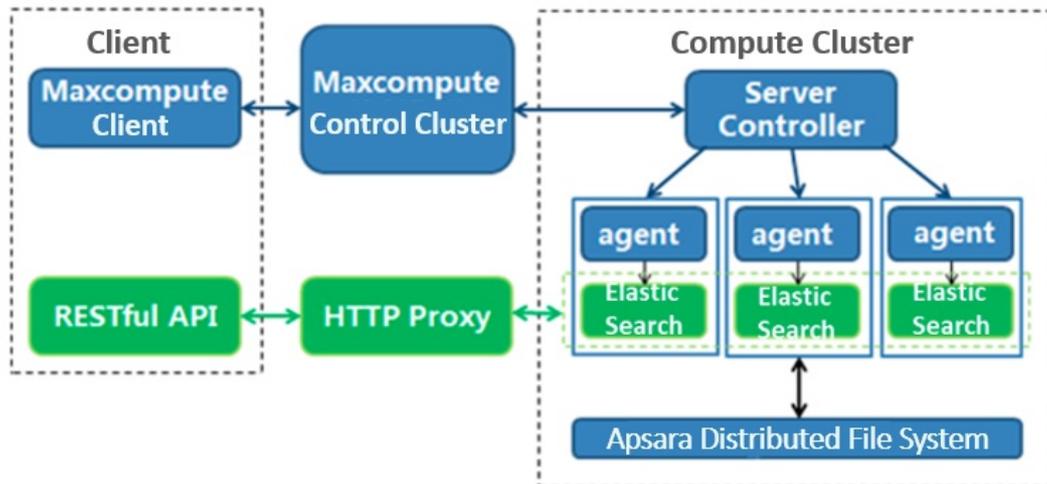
Features

Type	Feature	Description
Distributed cluster	Cluster deployment Cluster monitoring	Provide an O&M platform to monitor clusters, nodes, and indexes.
Retrieval management	Index configuration management Structure definition and index rebuilding	Provide a retrieval management platform and support configuration.
Full-text retrieval	Retrieval Sorting Statistical analysis	The features are provided through RESTful APIs.
Data collection	Elasticsearch data import APIs MaxCompute data import tools Full and incremental collection	Support a variety of interfaces to collect native data. Provide integrated tools to import MaxCompute data.
Service authentication	Service-level user authentication	Allow you to configure user authentication in a centralized manner.

23.4.7.2.4. Elasticsearch architecture

Elasticsearch on MaxCompute provides core search engine services, management platforms for O&M and indexes, MaxCompute management system, and MaxCompute data import tools. It can work with universal data import interfaces and data retrieval SDKs of Elasticsearch, enabling you to retrieve applications and perform full-text retrieval of large volumes of data. The following figure shows the overall architecture.

Overall architecture



An Elasticsearch cluster corresponds to a MaxCompute Server Task instance in MaxCompute. You can quickly and flexibly deploy, operate, and expand Elasticsearch clusters on a MaxCompute client. In the overall architecture,

- The MaxCompute control cluster starts Server Controller and forwards control requests from a client.
- Server Controller is the core component for Elasticsearch cluster management. It applies for resources, starts each Elasticsearch node, and responds to the control requests that a client forwards through the control cluster. It also returns the running status of Elasticsearch clusters or adjusts the clusters.
- An agent starts Elasticsearch node processes, monitors node running status, handles failover events, and executes tasks distributed by Server Controller.
- Elasticsearch on MaxCompute stores its data in Apsara Distributed File System. Once a node is started successfully, Elasticsearch on MaxCompute can provide services through HTTP Proxy and allow users to use its functions through RESTful APIs.

23.4.7.2.5. Elasticsearch benefits

Integration of big data computation and data retrieval for resource sharing

It can access and import MaxCompute data to an Elasticsearch cluster for full-text retrieval, and manages and uses data in a centralized manner.

Centralized management of computing and storage resources

It solves storage problems or prolonged computing tasks caused by increased data volumes. It supports automatic scaling of your cluster storage and retrieval capacities based on the amount of data you have. In this way, you can focus on data analysis and mining to maximize your data value.

Provision of services such as Elasticsearch cluster deployment and O&M

It helps you deploy, configure, operate, and maintain clusters. You can upload data, analyze data, and obtain corresponding results by performing a few steps in the offline analysis service.

Secure and reliable data storage

It uses multi-replica technology to store user data at multiple layers, preventing data from being lost, leaked, or intercepted.

Open service interfaces

It provides Elasticsearch SDKs which are native and open, allowing data import, indexing, and data retrieval with Elasticsearch on MaxCompute.

23.5. Scenarios

23.5.1. Scenario 1: Migrate data to the cloud cost-effectively and quickly

Usage scenario: The customer is a data and information service provider focusing on the new energy power sector. The customer's target is to build a cloud platform for Internet big data application services of the new energy industry.

Results: The customer's entire business system has been migrated to the cloud within three months. The data processing time is decreased to less than one third when compared with the customer-built system. Cloud data security is ensured through multiple security mechanisms.

Customer benefits:

- **More focus on its core business:** The entire business system is migrated to the cloud within three months, which enables the customer to use a variety of cloud resources to improve the business.
- **Low investment and O&M costs:** The cloud platform helps to significantly lower the costs of infrastructure construction, O&M personnel, and R&D when compared with a customer-built big data platform.
- **Security and stability:** Alibaba Cloud's comprehensive service and stable performance guarantee data security on the cloud.

23.5.2. Scenario 2: Improve development efficiency and reduce storage and computing costs

Usage scenario: Massive log analysis services for weather query and advertising business are provided to meet the business needs of an emerging mobile Internet company aiming for an excellent weather service provider.

Results: After the Internet company's log analysis business is migrated to MaxCompute, the development efficiency is improved by more than five times, the storage and computing costs are reduced by 70%, and 2 TB of log data is processed and analyzed every day. This more efficiently empowers its personalized marketing strategies.

Customer benefits:

- **Improved work efficiency:** All log data is analyzed by using SQL, and the work efficiency is increased by more than 5 times.
- **Improved storage usage:** The overall storage and computing cost is reduced by 70%, and the performance and stability are also improved.
- **Personalized service:** Machine learning algorithms on MaxCompute are used to perform in-depth data mining and provide personalized services for users.
- **Easy use of big data:** MaxCompute provides plugins for a variety of open-source software to easily migrate data to the cloud.

23.5.3. Scenario 3: Use mass data to achieve precision marketing for millions of users

Usage scenario: To meet the business needs of a community-oriented vertical e-commerce app that focuses on the manicure industry, you can use MaxCompute to build a big data platform for the app. It is mainly used in four aspects: business monitoring, business analysis, precision marketing, and recommendation.

Results: This e-commerce app uses the big data platform built based on MaxCompute to achieve precision marketing for millions of users through the computing capability of MaxCompute, making e-commerce business more agile, intelligent, and insightful. The platform can quickly respond to the data and analysis needs of new business.

Customer benefits:

- **Improved business insights:** Through the computing capabilities of MaxCompute, precision marketing for millions of users is achieved.
- **Data-driven business:** The platform improves the business data analysis capability and effectively monitors business data to better empower businesses.
- **Fast response to business needs:** The MaxCompute ecosystem can quickly respond to changing business data analysis needs.

23.5.4. Scenario 4: Achieve precision marketing with big data

Usage scenario: MaxCompute is used to meet the business needs of an Internet company that focuses on precision marketing and advertising technologies and services. A core big data-based precision marketing platform will be built for the company.

Results: Based on MaxCompute, the company builds a core big data-based precision marketing platform. All log data is stored in MaxCompute, and offline scheduling and analysis are performed through DataWorks.

Customer benefits:

- **Efficient and low-cost analysis of massive data:** Statistical analysis of massive data can reduce expenditures by half to meet the same business needs, effectively saving costs and helping

startup enterprises grow rapidly.

- **Real-time data query and analysis:** MaxCompute helps the enterprise establish technical advantages, overcoming the technical bottleneck of massive data processing and analysis, and real-time query and analysis. MaxCompute collects, analyzes, and stores more than 2 billion visitor activities every day. At the same time, it performs millisecond-level queries in hundreds of millions of log tables based on user requirements.
- **Machine learning platform with low entry barrier:** As for a precision marketing and advertising provider, the quality of algorithm models is directly linked to its final revenue. Therefore, selecting the ease-of-use MaxCompute machine learning platform with low entry barrier can get twice the result with half the effort.

23.6. Limits

None.

23.7. Terms

Project

A project is the basic unit of operation in MaxCompute. It is similar to the concept of database or schema in traditional databases, and sets the boundary for MaxCompute multi-user isolation and access control. A user can have permissions on multiple projects.

 **Note** After security authorization, a user can access objects such as tables, resources, functions, and instances in a project from another project.

Table

A table is the data storage unit in MaxCompute. A table is a two-dimensional data structure composed of rows and columns. Each row represents a record, and each column represents a field of the same data type. One record can contain one or more columns. The column names and data types comprise the schema of a table.

 **Note** There are two types of MaxCompute tables: external tables and internal tables.

Partition table

A partition tables refer to the partition space specified during table creation. This specifies certain fields in a table as partition columns. If you specify the name of a partition to access for data use, the system reads data only from the specified partition. This avoids full table scan, thereby improving processing efficiency and reducing costs.

Type

Columns in MaxCompute tables must be one of the following types: tinyint, smallint, int, bigint, string, float, boolean, double, datetime, decimal, varchar, binary, timestamp, array, map, and struct.

Resource

A resource is a unique concept of MaxCompute. User-defined functions (UDFs) and MapReduce of MaxCompute are implemented depending on resources.

 **Note** Resource types in MaxCompute include file, MaxCompute table, JAR (compiled Java JAR package), and archive. Compressed files are identified by the extensions of resource names. Supported file types include .zip, .tgz, .tar.gz, .tar, and .jar.

Function

MaxCompute provides you with SQL computing capabilities. In MaxCompute SQL, you can use built-in functions for computing and calculation. When the built-in functions are not sufficient to meet your requirements, you can use the Java programming interface provided by MaxCompute to develop UDFs.

 **Note** UDFs can be further divided into scalar-valued functions, user-defined aggregate functions (UDAFs), and user-defined table functions (UDTFs).

Task

A task is the basic computing unit in MaxCompute. Computing jobs such as those involving SQL and MapReduce functions are completed by using tasks.

Task instance

In MaxCompute, some tasks are converted into instances when being executed and subsequently exist as MaxCompute instances.

Resource quota

There are two types of quotas: storage and computation. MaxCompute allows you to set an upper limit of storage for a project. When the used storage space approaches to the upper limit, MaxCompute triggers an alarm. The computing quota limits the use of memory and CPU resources. The memory usage and CPU usage of running processes in a project cannot exceed the specified upper limit.

24.DataWorks

24.1. What is DataWorks?

DataWorks is an end-to-end big data platform that uses MaxCompute as its compute engine. It integrates all processes from data collection to data display and from data analysis to application running. DataWorks provides various features to help you quickly and effectively complete the entire research and development (R&D) process.

DataWorks supports batch processing, analysis, and mining for large volumes of data. It integrates services such as Data Integration, DataStudio, Administration, Real-Time Analysis, Data Asset Management, Data Quality, Data Protection, and Data Service to enable core data-related processes. It also provides an online API development and management platform and interworks seamlessly with Machine Learning Platform for AI.

In 2018, Forrester, a globally recognized market research company, named Alibaba Cloud DataWorks and MaxCompute as a world-leading cloud-based data warehouse solution. This solution is currently the only product from a Chinese company to receive such an acknowledgment. Building on the success of the previous version, DataWorks V2.0 includes several new additions, such as business processes and components. DataWorks V2.0 also isolates development and production environments, adopts standard development processes, and uses a specific mechanism to reduce errors in code.

24.2. Benefits

24.2.1. Powerful computing capabilities

DataWorks uses MaxCompute as its compute engine, which can process large amounts of data.

- Supports join operations for trillions of records, millions of concurrent jobs, and I/O throughput of up to multiple petabytes (PB) each day.
- Supports batch scheduling of millions of tasks, real-time monitoring, and alerts.
- Provides powerful and easy-to-use SQL and MapReduce engines, and supports the majority of standard SQL syntax.
- Uses triplicate data storage and multiple access control mechanisms, including read/write request authentication, application sandboxing, and system sandboxing. All these mechanisms prevent data loss, leakage, and breaches.

24.2.2. End-to-end platform

- Integrates all processes from data integration, processing, management, and monitoring to output.
- Provides a visualized workflow design tool.
- Adopts a multi-user architecture and enables online job processing. You can create and assign roles with varying permissions to different users.

24.2.3. Data integration from heterogeneous data sources

DataWorks supports batch synchronization between data sources at custom intervals in minutes, days, hours, weeks, or months. More than 400 heterogeneous data source pairs are supported.

24.2.4. Web-based software

DataWorks is an out-of-the-box service. You can use it over the Internet or any internal network without the need of installation and deployment.

24.2.5. Multi-tenant architecture

Data is isolated among different tenants. Tenants independently control permissions, process data, allocate resources, and manage members.

24.2.6. Intelligent monitoring and alerts

You can specify thresholds for the task duration in DataWorks to monitor the task progress. DataWorks provides you with the progress of each task and an overview of the progress.

24.2.7. Easy-to-use SQL editor

The editor supports automatic code and metadata completion, code formatting and folding, and pre-compilation. It also offers two editor themes. These features help to ensure a good user experience.

24.2.8. Comprehensive data quality monitoring

DataWorks supports validation, notification, and management for data sources, batches of data, and real-time data.

24.2.9. Convenient API development and management

The API Gateway service and the Data Service service allow you to easily develop and publish APIs for data sharing.

24.2.10. Secure data sharing

DataWorks enables you to mask sensitive data before sharing them to other tenants, which ensures the security of your big data assets while maximizing their value.

24.3. Architecture

DataWorks is an end-to-end big data platform launched by Alibaba Group, which supports big data processing, management, analysis, mining, sharing, and transmission. It releases you from cluster deployment and management. DataWorks adopts MaxCompute (formerly known as ODPS) as the compute engine to process large volumes of data.

DataWorks is developed based on MaxCompute. DataWorks provides a management console and supports functions such as data processing, management, analysis, and mining.

24.3.1. Services

DataWorks provides the following services:

- Data Integration: big data integration between heterogeneous data sources
- DataStudio: data warehouse design and whole extract, transform, load (ETL) procedure design

- Administration: online ETL job management and monitoring
- Real-Time Analysis: ad hoc query and analysis
- Data Asset Management: metadata management, overall table management, data lineage, and data asset dashboard
- Data Quality: data quality check, monitoring, verification, and grading
- Data Protection: permission management, data management based on security levels, data masking, and data auditing
- Data Service: data sharing and transmission through APIs

24.3.2. System architecture

DataWorks is an end-to-end big data platform that enables you to process data by using Data Integration, DataStudio, Data Management, Data Service, and other services. It serves as a basis for upper-layer applications, which satisfies all user requirements.

24.3.3. Security architecture

The security architecture of DataWorks is composed of error proofing, data security, and optional security tools.

- Error proofing ensures proper running of DataWorks during coding, deployment, and configuration.
- Data security ensures basic security of DataWorks by using features such as resource isolation among tenants, user identity verification, authentication, and log auditing.
- Optional security tools, integrated in DataWorks, enable you to customize security policies for protection and management on your system and data.

24.3.4. Multi-tenant architecture

DataWorks adopts a multi-tenant architecture. The following items are isolated among tenants to ensure resource and data security.

- Storage and compute resources, which are scalable
- Data, permissions, users, and roles

24.4. Services

24.4.1. Data Integration

24.4.1.1. Overview

Data Integration provides stable, efficient, and scalable services for data synchronization. Data Integration supports fast and stable data transmission and synchronization between various heterogeneous data stores in complex networks.

Data Integration provides you with an overview of all data stores, you can monitor and read data in your data stores. Data Integration supports the following types of data stores: relational and NoSQL databases, big data platforms, and File Transfer Protocol (FTP) servers. Data Integration also supports synchronization between heterogeneous data stores in diverse complex networks. Data Integration supports batch, full, and incremental synchronization. You can also synchronize data at intervals of minutes, hours, days, weeks, or months.

24.4.1.2. Various data sources

24.4.1.2.1. Metadata synchronization

Data Integration collects metadata from more than 20 types of common data stores, such as MySQL, Oracle, and MaxCompute. It generates a clear view of all data assets from the collected metadata. It also helps users to track assets and synchronize data.

24.4.1.2.2. Relational database

Data Integration supports read/write operations on relational databases such as MySQL, Oracle, PostgreSQL, IBM Db2, and ApsaraDB for PPAS.

24.4.1.2.3. NoSQL database

Data Integration supports read/write operations on NoSQL databases such as MongoDB and Table Store.

24.4.1.2.4. MPP database

Data Integration supports read/write operations on Massively Parallel Processor (MPP) databases such as AnalyticDB for PostgreSQL.

24.4.1.2.5. Big data product

Data Integration supports read/write operations on MaxCompute and Hadoop Distributed File System (HDFS).

24.4.1.2.6. Unstructured data storage

Data Integration supports read/write operations on Object Storage Service (OSS) and FTP servers.

 **Note** Data Integration supports data synchronization for more than 400 pairs of source and target data stores.

24.4.1.3. Inbound data control

Data Integration supports conversion between various data types. It accurately identifies, filters, collects, and displays dirty data to facilitate inbound data control. Data Integration enables you to view important statistics such as data volume, data throughput, and job duration. It also can detect dirty data for each job.

24.4.1.4. High transmission rate

Data Integration enhances the performance of one-way synchronization agents and makes full use of the network interface card (NIC) on each server. DataWorks adopts separate reader and writer agents, which ensures transmission of multiple GB or TB data within a short time.

24.4.1.5. Accurate flow control

Data Integration implements accurate flow control on channels, record streams, and byte streams. It also supports fault tolerance. With Data Integration, you can rerun specific threads, processes, and jobs.

24.4.1.6. Synchronization agents

Data Integration provides you with synchronization agents. You can install the agents on data source servers to facilitate data synchronization.

24.4.1.7. Transmission in complicated networks

DataWorks enables data to be transmitted in complicated networks, for example, between local private networks, between VPCs, and across air gaps.

 **Note** Transmission of large amounts of data over a long distance is accelerated by adopting certain protocols, which ensures high stability and efficiency.

24.4.2. DataStudio

24.4.2.1. Overview

DataStudio is an integrated development environment (IDE) that allows you to develop ETL and data mining algorithms, and build data warehouses in DataWorks.

Before using DataStudio, you need to add data stores by using Data Integration. Then, you can use DataStudio to process the data retrieved from the data stores.

24.4.2.2. Business flows

24.4.2.2.1. Description

In DataStudio, you can organize nodes in a business flow. The following node types are supported: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL component, and virtual node.

DataStudio provides you with a directed acyclic graph (DAG) for nodes in each business flow. DataStudio also provides professional tools and supports administrative operations for business flows, which promotes intelligent development and management.

You can configure dependencies between nodes within the same business flow or across different business flows. You can schedule a whole business flow or specific nodes.

24.4.2.2.2. Nodes

The following node types are supported: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL component, and virtual node. Tasks are initiated based on either node dependencies or task schedules.

24.4.2.2.3. Configure nodes

After you double-click a node either in the left-side navigation pane or in a DAG, the editor of the node appears. You can configure the node in the editor, for example, entering SQL statements for ODPS SQL nodes and configuring synchronization rules for data integration nodes. If you need to view earlier versions or modify the schedule, dependencies, lineage, and other settings of the node, click the corresponding right-side buttons in the editor.

24.4.2.2.4. Versions

You can view earlier versions of ODPS SQL, ODPS MR, shell, and other nodes. If required, you can roll back a node to an earlier version.

24.4.2.2.5. Publish business flows

In workspaces that adopt standard mode, you can easily submit business flows that have passed checks to the production environment.

24.4.2.3. Solutions

In a DataWorks workspace, you can group multiple business flows in a solution.

You can also add a business flow to multiple solutions. In addition, business flows can be used repeatedly in different solutions, allowing you to assess your solutions from a business perspective.

24.4.2.4. Code editor

24.4.2.4.1. ODPS SQL nodes

DataWorks provides a Web-based SQL editor. The editor supports a variety of features such as automatic SQL statement completion, code formatting and highlighting, and debugging.

24.4.2.4.2. ODPS MR nodes

You can upload Java Archive (JAR) files that contain MapReduce code to DataWorks as resources and insert them into ODPS MR nodes.

24.4.2.4.3. Resources

DataWorks supports the following resource types:

- **JAR:** You can upload JAR files to DataWorks as JAR resources. These JAR resources can then be called by user-defined functions and ODPS MR nodes.
- **Python:** You can upload Python files to DataWorks as Python resources. These Python resources can then be called by user-defined functions.
- **File:** You can upload shell scripts, XML and TXT configuration files, and other files to DataWorks as file resources.
- **Archive:** You can upload compressed files to DataWorks as archive resources. The following formats are supported: .zip, .tgz, .tar.gz, .tar, and .jar. DataWorks automatically identifies the file format based on the extension of uploaded files.

24.4.2.4.4. Register user-defined functions

You can define functions and use them for data processing. Before using user-defined functions, upload JAR files and Python files to DataWorks as resources and register related functions in DataWorks.

24.4.2.4.5. Shell nodes

DataWorks supports online editing and debugging of shell scripts.

24.4.2.5. Code repository and team collaboration

DataWorks enables multiple users to simultaneously work on the same workspace. DataWorks allows you to create workflows and adopts a lock mechanism, which ensures that each node is edited by only one user at the same time. To edit a node locked by another user, you can force unlock the node and then lock the node yourself. This operation is called steal lock. After you steal the lock of the node, the system sends a notification to the other user.

In addition, DataWorks records each submitted version of your node and workflow. You can compare two versions of a node, and roll back a node to an earlier version.

24.4.3. Administration

24.4.3.1. Overview

Due to the volume, diversity, and complexity of data used in DataWorks, it is necessary to use a scheduling system that supports high concurrency, multiple cycles, and various data processing procedures.

To meet this requirement, Operation Center is developed, which is a centralized data operation and management platform for data developers and administration experts. With Operation Center, you can control and monitor the running of instances, and set node priorities. In Operation Center, you can trace all the nodes committed to the scheduling system, view alerts when nodes do not run as scheduled or fail, and view daily reports of node statistics.

24.4.3.2. Overview

The Overview page displays statistics of nodes, tasks, and task instances. The following statistics items are involved: the trend of task instances run today and in past days, the sorting of tasks by duration, by number of errors, and by number of overtime task instances within 30 days, and the distribution of tasks by status and by type.

24.4.3.3. Task instances

You can perform the following operations on task instances:

- Rerun, terminate, suspend, and set task instances to successful. Set alerts to monitor task execution.
- View task instances in a list or DAG. The DAG clearly shows the relationships between nodes.
- View the status of periodic task instances, ad hoc task instances, and test instances.
- View runtime logs of tasks instances and the code and attributes of corresponding nodes.

24.4.3.4. Monitor

The Monitor feature sends alerts based on specified rules, time, methods, and recipients. The Monitor feature:

- Reduces your operating expenditure (OPEX).
- Prevents invalid alerts.
- Automatically enables alerts for tasks that are configured with baselines.

The Monitor feature automatically creates alert rules for tasks that are configured with baselines. You can also customize alert rules by completing basic settings.

24.4.4. Workspace Management

24.4.4.1. Description

Workspace Management enables administrators to manage their organizations and workspaces.

Workspaces are the organizational unit for code, member, role, and permission management in DataWorks. Workspaces are isolated from each other. You can view and change code in a workspace only if you are a member of the workspace and have been assigned the required permissions.

 **Note** You can add a user to multiple workspaces. The user's permission varies according to the role assigned to the user in each workspace.

24.4.4.2. Organization

The Organization page displays the account, AccessKey ID, and AccessKey Secret of the organization owner. On this page, you can manage all members in an organization.

24.4.4.3. Workspaces

On the Workspaces page, administrators can create, modify, activate, or disable workspaces.

24.4.4.4. Members

The member list shows the name, account, role, and other information of each member.

- Fuzzy match is supported for member search.
- Only administrators can add and remove members from workspaces.

 **Note** When you add a user to a workspace, you must assign at least one role to the user.

Only administrators can remove members from workspaces.

 **Note** After a user is removed from a workspace, all permissions that have been granted to the user within the workspace are revoked.

24.4.4.5. Authorizations

On the Authorizations page, you can manage roles and specific permissions for all users. The **Roles and permissions** table describes the permissions of each role in DataWorks.

Roles and permissions

Role	Permissions
Administrator	An administrator can manage the basic information, data sources, compute engine configurations, and members for the workspace. The administrator can also assign users with the administrator, developer, administration expert, deployment expert, and visitor roles.

Role	Permissions
Developer	A developer can create workflows, script files, resources, and functions. The developer can also create and delete tables, and create publish tasks. However, the developer cannot perform publish operations.
Administration expert	The administration permissions of an administration expert is assigned by an administrator. An administration expert can perform publish and administrative operations, but does not have the permissions of a developer.
Deployment expert	The permissions of a deployment expert is similar to those of an administration expert. However, the deployment expert is not authorized to perform administrative operations.
Visitor	A visitor can only view information in the workspace, and cannot change any workflow, code, or other configurations.

24.4.5. Realtime Analysis

Realtime Analysis provides two core features: ad hoc query and private table management. It expedites the analysis process by using the data collection tools of MaxCompute in near real-time mode.

Benefits

The near real-time mode is used by default.

You can run the `set ODPS.service.mode=[all|off|limited]` command to change the configuration.

The advantages of near real-time mode over standard mode are described as follows:

- In near real-time mode, Data Analysis preallocates thread pools based on the job size. The near real-time mode eliminates the need for Job Scheduler to plan jobs and reduces the preparation time to run jobs.
- In near real-time mode, Data Analysis shuffles data from Mappers to Reducers, without transferring the data to Apsara Distributed File System.

Principle

- You can switch to the near real-time mode by setting the `ODPS.service.mode` parameter to `all`. However, if resources in MaxCompute are insufficient to run SQL nodes, Data Analysis switches to the standard mode, and then Job Scheduler is responsible for resource allocation. For example, this occurs if available workers are insufficient for creating instances.
- The near real-time mode includes a complex scheduling mechanism, which is improved based on the mechanism adopted by Job Scheduler to reduce the preparation time for running jobs.
- If the near real-time mode is used, Realtime Analysis first tries to run jobs in near real-time mode. Realtime Analysis uses the standard mode if system resources are insufficient, or if known issues or unknown exceptions occur in near real-time mode.

24.4.6. Data Service

With Data Service, you can manage all your table APIs after you create new APIs or register existing APIs. You can also easily publish your APIs to API Gateway. Together with API Gateway, Data Service provides a secure, stable, cost-effective, and easy-to-use API development and management service. Data Service adopts a serverless architecture and allows you to develop table APIs without thinking about infrastructure such as compute resources. Data Service supports automatic scaling for compute resources, which significantly reduces your OPEX.

Data Service serves the government as a secure, flexible, and reliable platform for data sharing across departments and networks within the government. It also enables the government to share data with the public.

Create an API

You can create APIs for tables in relational databases, NoSQL databases such as Table Store, and analytical database. You can quickly create an API within a few minutes by simply following the wizard provided, and immediately call the API operation after the creation is complete. You can also create APIs by specifying SQL scripts. The script mode supports advanced functions such as associative tables, complex criteria, and aggregate functions.

Register an API

You can register existing Restful APIs to Data Service for unified API management. Four request methods and three data formats are supported. The four request methods are GET, POST, PUT, and DELETE. The three data formats are tables, JSON, and XML.

API Gateway

API Gateway provides API lifecycle management services, including API publishing, management, maintenance, and monetization. It enables low-risk, simple, cost-effective, and fast microservice integration, front and back end separation, and system integration. You can use API Gateway to share functions and data with your partners and third-party developers. API Gateway supports authorization, authentication, flow control, and billing for Data Service.

24.4.7. Data Asset Management

Data Asset Management provides you with an overview of your data assets. Data Asset Management feature requires that data be synchronized by using Data Integration and processed by using DataStudio before you manage your tables and APIs stored in your business system and DataWorks.

24.4.8. Data Protection

24.4.8.1. Overview

Data Protection is a data security management platform for data asset and sensitive data recognition, classification, masking, risk behavior monitoring, and auditing.

It provides security management services for MaxCompute.

Data Protection provides the following features.

- Sensitive data recognition

Data Protection can recognize sensitive data in enterprises based on self-training algorithms to show the types, distribution, and volume of sensitive data. It also recognizes custom data types.

- Custom data classification
Data Protection allows you to classify data by creating custom levels for better data management.
- Flexible data masking
Data Protection provides diverse and configurable methods for dynamic data masking.
- Risk behavior monitoring and auditing
Data Protection uses various correlation analysis algorithms to identify risk behavior. Data Protection provides alerts and supports visualized auditing for identified risks.

24.4.8.2. Terms

24.4.8.2.1. Organization

An organization refers to all system settings and resources owned by a single tenant in DataWorks. The system settings and resources include RAM configurations, permission settings, and custom applications.

24.4.8.2.2. Workspace

Workspaces are the organizational unit in DataWorks. Similar to databases in a relational database management system (RDBMS), workspaces isolate resources among different users and offer boundaries for access control. User-defined tables, resources, functions, and nodes are isolated among different workspaces.

24.4.8.2.3. Regular expression

A regular expression is a sequence of characters that define filter criteria. You can use regular expressions to identify sensitive data.

 **Note** A regular expression consists of metacharacters and literal characters such as letters from a to z.

24.4.8.2.4. Data classification

Data is classified based on value, sensitivity, related risks, legal and regulatory requirements, and the potential impact of data breaches.

24.4.8.2.5. Data recognition

The Data Protection service recognizes sensitive data on end user side based on user-defined rules.

24.4.8.2.6. Data masking

The Data Protection service masks sensitive data based on user-defined rules.

24.4.8.2.7. MaxCompute

MaxCompute is a data processing platform developed by Alibaba Cloud for large-scale data warehousing. It supports storage and compute for batches of structured data, and meets the requirements for big data modeling and analysis in most scenarios.

24.5. Scenarios

24.5.1. Cloud-based data warehouse

Enterprise customers can create large data warehouses by using DataWorks in Apsara Stack.

DataWorks can integrate petabytes (PB) of data for enterprise customers.

- **Storage:** DataWorks provides a scalable data warehouse for petabytes and exabytes of data.
- **Data integration:** DataWorks supports data synchronization and integration across heterogeneous stores to eliminate data silos.
- **Data processing:** DataWorks uses MaxCompute as the compute engine. It provides a visualized workflow designer and programming framework for SQL and MapReduce.
- **Data management:** DataWorks supports metadata-based data resource management and permission-based resource access control.
- **Batch scheduling:** DataWorks provides the scheduling of recurring nodes at different intervals, and supports scheduling of millions of concurrent nodes, error alerts, and real-time monitoring of running node instances.

24.5.2. Business intelligence

This topic describes how to create reports by using DataWorks.

You can analyze the following items based on network logs:

- Page views, unique visitors, and device types such as Android devices, iPads, iPhones, and personal computers. You can also create a daily report based on these statistics.
- Locations of visitors.

A sample log entry is provided as follows:

```
xx.xxx.xx.xxx - - [12/Feb/2014:03:15:52 +0800] "GET /articles/4914.html HTTP/1.1" 200 37666  
"http://xxx.cn/articles/6043.html" "Mozilla/5.0 (Windows NT 6.2; WOW64)  
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/xx.x.xxxx.xxx Safari/537.36" -
```

Create a table: Before importing data to MaxCompute, you need to create a target table with the name `ods_log_tracker` in MaxCompute.

24.5.3. Data-driven management

- **Innovative business:** Data mining, data modeling, and real-time decision making can be implemented based on big data analytics results provided by DataWorks.
- **Small and medium enterprises:** With DataWorks, data can be quickly analyzed and put into commercial use, which helps enterprises to generate operational strategies.

24.6. Limits

None.

24.7. Terms

node

Nodes define operations on data. You can create nodes of following types: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL, and virtual nodes.

instance

When the scheduling system or user triggers a node, an instance is created. The instance is a snapshot of the running node. In Operation Center of DataWorks, you can find the schedule, status, and runtime log of the instance.

commit

 **Note** You can commit nodes in the development environment to the scheduling system. The scheduling system runs the code specified in the committed nodes according to the node configurations.

The scheduling system only initiates nodes after you commit them.

manually triggered workflow

You need to manually trigger nodes in this type of workflows, and cannot schedule the nodes. Therefore, you cannot configure parent nodes and outputs for nodes in manually triggered workflows.

25. Apsara Bigdata Manager (ABM)

25.1. What is Apsara Bigdata Manager?

Apsara Bigdata Manager (ABM) is an O&M platform for big data products, including MaxCompute, DataWorks, StreamCompute, Quick BI, Graph Analytics, Elasticsearch, Dataphin, DataHub, and Machine Learning Platform for AI.

ABM supports O&M on the business, services, clusters, and hosts of these big data products. Besides, you can upgrade big data products, customize alert configurations, and view the O&M history in ABM.

By using ABM, on-site Apsara Stack engineers can easily manage big data products, such as viewing resource usage, checking alerts and fix methods, and modifying configurations.

25.2. Benefits

Cluster health monitoring

Supports status monitoring and configuration management of devices, resources, and services in clusters of big data products, and collects cluster operating status in real time for dynamic display.

Resource usage analysis

Collects cluster operating status in real time, including the status of devices, resources, and services, and supports data aggregation and analysis to help evaluate the health status of clusters. If the evaluation result shows potential risks in a cluster, responsible engineers can be notified immediately.

Graphical management interface

Provides a graphical user interface for displaying system operating status and supports common O&M operations.

25.3. Architecture

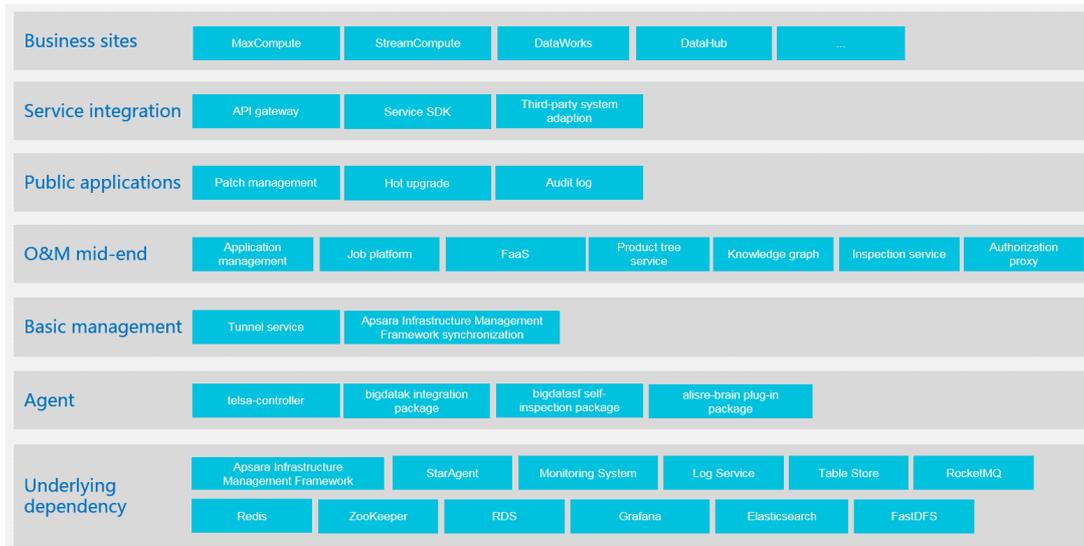
25.3.1. System architecture

This topic describes the system architecture of ABM and the functions of each component.

ABM adopts a microservice architecture that enables data integration, interface integration, and feature integration through a unified platform, and provides standard service interfaces. This architecture enables a consistent user interface, which means the O&M operations are the same for all products. This reduces training costs and lowers O&M risks.

The ABM system consists of the following components: underlying dependency, agent, basic management, O&M mid-end, public applications, service integration, and business sites.

Architecture



Underlying dependency

ABM depends on open-source systems from Alibaba and third parties.

- Uses StarAgent and Monitoring System of Alibaba to run remote commands and remote data collection instructions.
- Uses ZooKeeper to coordinate primary and secondary services. This ensures high service availability.
- Uses RDS to store metadata, Redis to store cache data, and Table Store to store large amounts of self-test data. This improves service throughput.

Agent

The agent provides client SDKs, scripts, and monitoring packages to be deployed on each management host.

O&M mid-end and basic management

The O&M mid-end and basic management components form the base of ABM. Each service in the two components provides its own capabilities for business sites. This enables quick construction of business sites and makes the capabilities of each business site complete.

Public applications

Based on the O&M mid-end, ABM provides multiple public applications. These applications are designed with special purposes and adaptive to all big data products supported by ABM.

Service integration

Service integration functions as a link between business sites and underlying components. It integrates interfaces of all internal services, adapts to various third-party systems, and provides a unified SDK for users.

Business sites

Business sites are constructed based on the O&M mid-end of ABM and cover all big data products, including MaxCompute, StreamCompute, DataWorks, and DataHub. A business site functions as a one-stop O&M portal of a product.

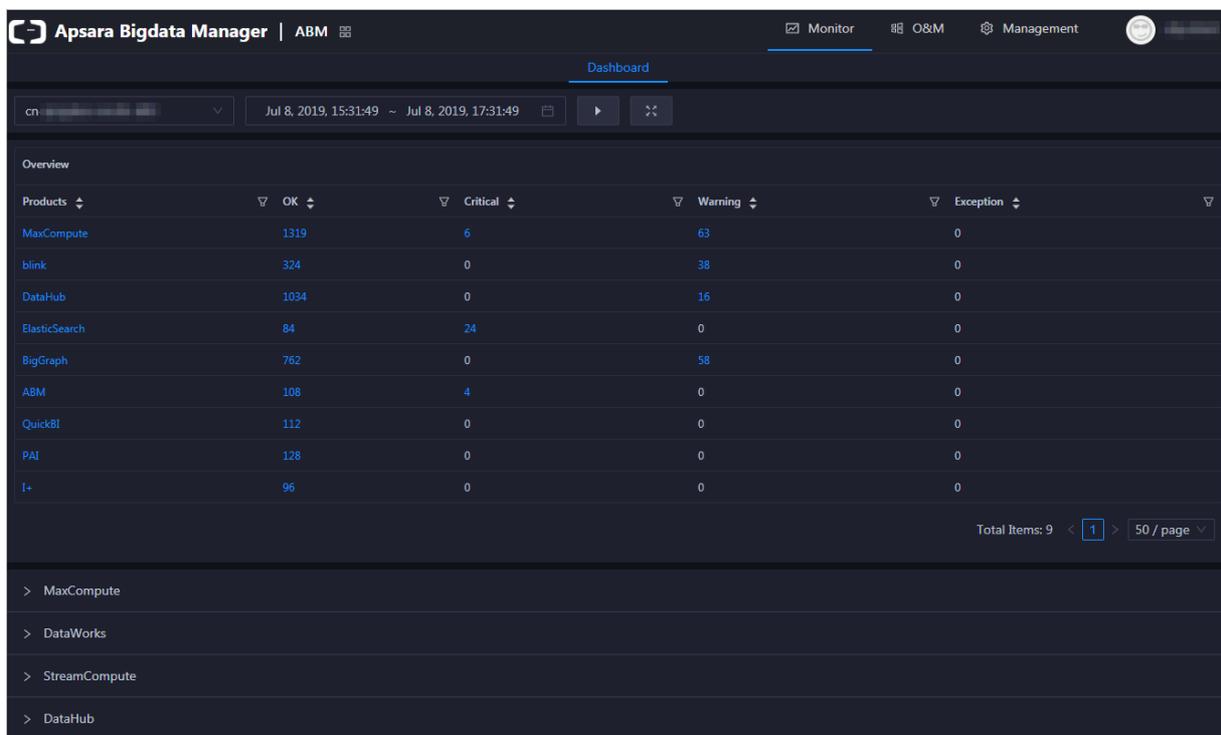
25.4. Features

25.4.1. Dashboard

The dashboard is the home page of ABM. It displays key operating metrics for the following key big data products: MaxCompute, DataWorks, StreamCompute, and DataHub. Besides, you can view alerts for all big data products in the dashboard. This allows you to understand the overall operating status of all big data products.

Dashboard page

When you log on to ABM, the **Dashboard** page appears by default. To return to the **Dashboard** page from any other page, click  in the upper-left corner and select **ABM**.



Products	OK	Critical	Warning	Exception
MaxCompute	1319	6	63	0
blink	324	0	38	0
DataHub	1034	0	16	0
ElasticSearch	84	24	0	0
BigGraph	762	0	58	0
ABM	108	4	0	0
QuickBI	112	0	0	0
PAI	128	0	0	0
I+	96	0	0	0

On the **Dashboard** page, you can select a region from the **Dashboard** drop-down list in the upper-left corner. Then, you can view the cluster operating status of big data products in the specified region.

Alerts for big data products

You can view the numbers of alerts generated for all big data products in the **Overview** area. Pay attention to **Critical** and **Warning** alerts. They need to be handled in time.

Products	OK	Critical	Warning	Exception
MaxCompute	1319	6	63	0
blink	324	0	38	0
DataHub	1034	0	16	0
ElasticSearch	84	24	0	0
BigGraph	762	0	58	0
ABM	108	4	0	0
QuickBI	112	0	0	0
PAI	128	0	0	0
I+	96	0	0	0

Total Items: 9 < 1 > 50 / page

In the **Overview** area, you can click a product name or a number of alerts to go to the **O&M** page of the product.

Overview of operating metrics for MaxCompute

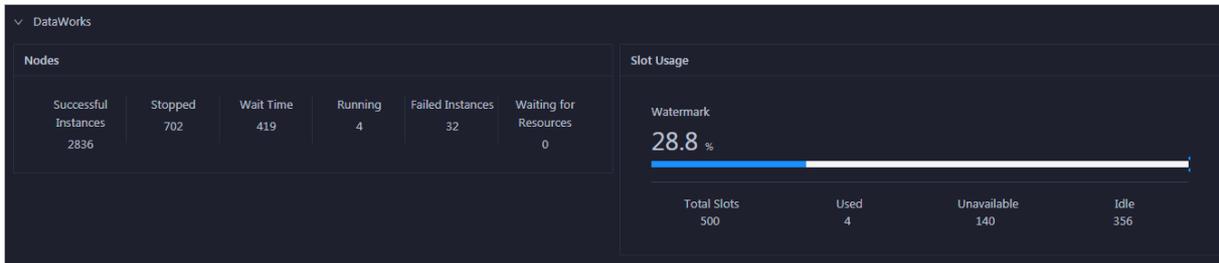
The **Dashboard** page displays key operating metrics for MaxCompute. To view these metrics, click **MaxCompute** under the **Overview** area.



In the **MaxCompute** area, you can view the allocation of CPU and memory, CPU and memory usage tendency charts, job operating status, and storage usage of the MaxCompute cluster.

Overview of operating metrics for DataWorks

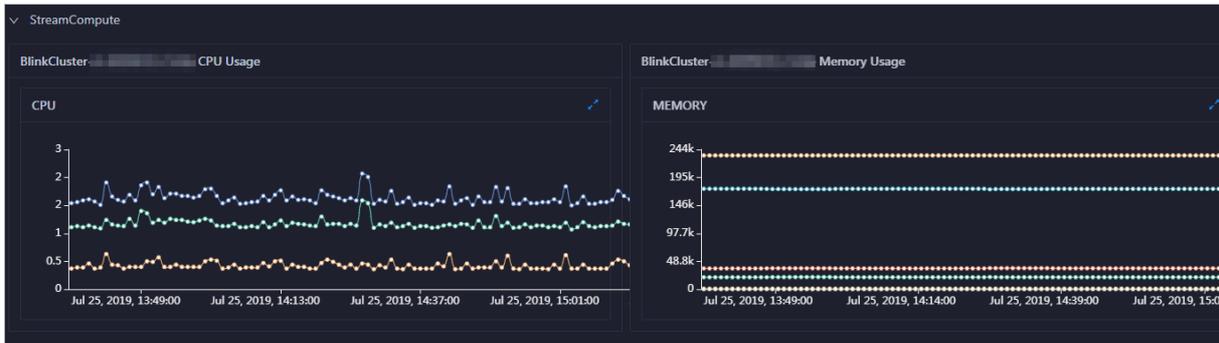
The **Dashboard** page displays key operating metrics for DataWorks. To view these metrics, click **DataWorks** under the **Overview** area.



In the **DataWorks** area, you can view the node scheduling and slot usage of the DataWorks cluster.

Overview of operating metrics for StreamCompute

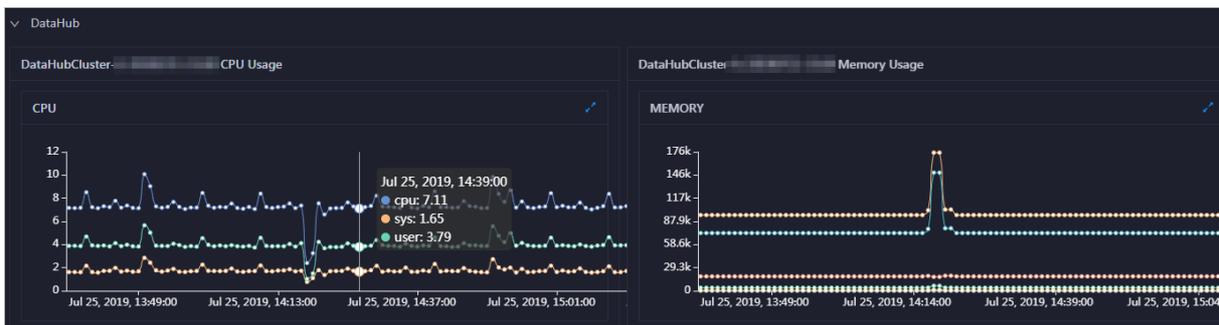
The **Dashboard** page displays key operating metrics for StreamCompute. To view these metrics, click **StreamCompute** under the **Overview** area.



In the **StreamCompute** area, you can view the CPU and memory usage tendency charts of the StreamCompute cluster.

Overview of operating metrics for DataHub

The **Dashboard** page displays key operating metrics for DataHub. To view these metrics, click **DataHub** under the **Overview** area.



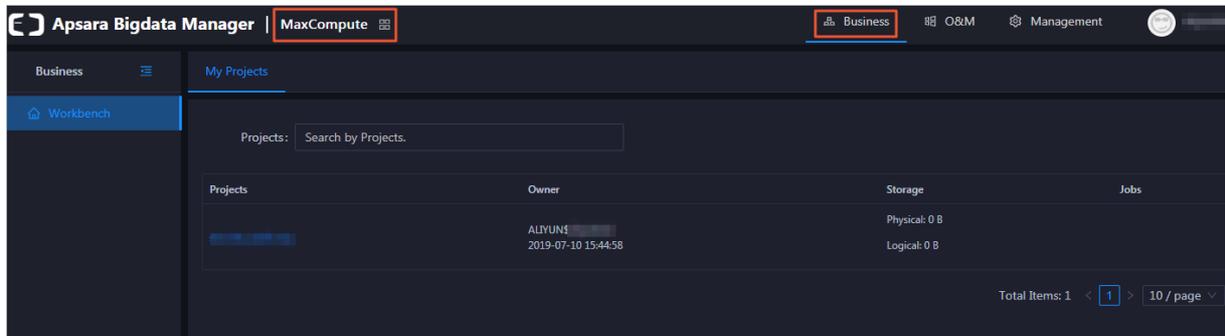
In the **DataHub** area, you can view the CPU and memory usage tendency charts of the DataHub cluster.

25.4.2. Business

The business feature is exclusive for MaxCompute. It provides a workbench so that you can view the MaxCompute project list and details, and configure project parameters.

Workbench

By default, the **My Projects** tab appears in the workbench. You can click a project name to view its details.



Project details

Project details include the project overview and information about jobs, storage, configuration, and quota groups.

- **Overview:** displays the basic information and the CPU and memory resources of the project.
- **Job:** allows you to view snapshots of jobs in the recent seven days on a day basis and operational logs of jobs. This helps you locate job failures.
- **Storage:** displays the storage usage percent, used storage space, storage quota, and free storage space.
- **Configuration:** allows you to configure properties and the encryption algorithm of the project. The properties include general properties, sandbox, SQL, MapReduce, access control, and resource recycling.
- **Quota group:** displays the quota group information about the project.

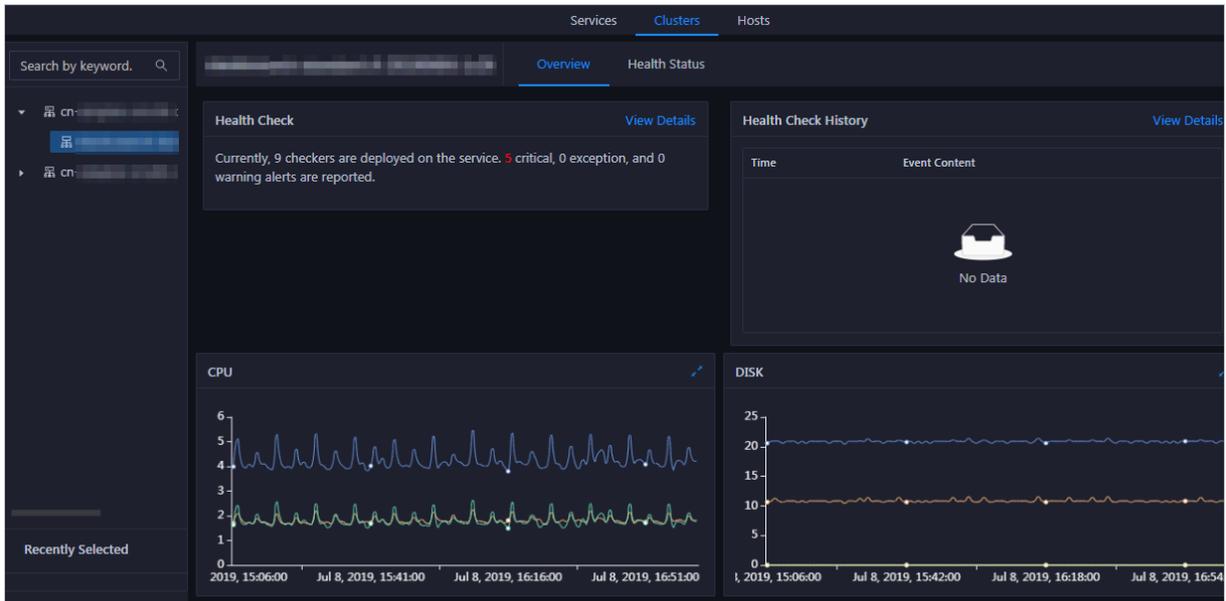
25.4.3. O&M

ABM supports O&M on over ten big data products, including MaxCompute, DataWorks, StreamCompute, and DataHub. ABM supports O&M on clusters, services, and hosts of these big data products except for DataWorks. Besides, ABM provides custom O&M features for some products.

Note ABM integrates the O&M feature for DataWorks into that for MaxCompute. To perform O&M on DataWorks, go to the O&M page for MaxCompute in ABM, and then choose **Services > DataWorks**.

Cluster O&M

ABM allows you to perform O&M on clusters of big data products. ABM provides two general cluster O&M features: cluster overview and cluster health.



- **Cluster overview:** displays general operating information about a cluster, including the host, service, CPU, disk, memory, load, package, and health check status, and health check history.
- **Cluster health:** displays the status of all health check items performed on a cluster. Check items can be in the Critical, Warning, Exception, or OK status.

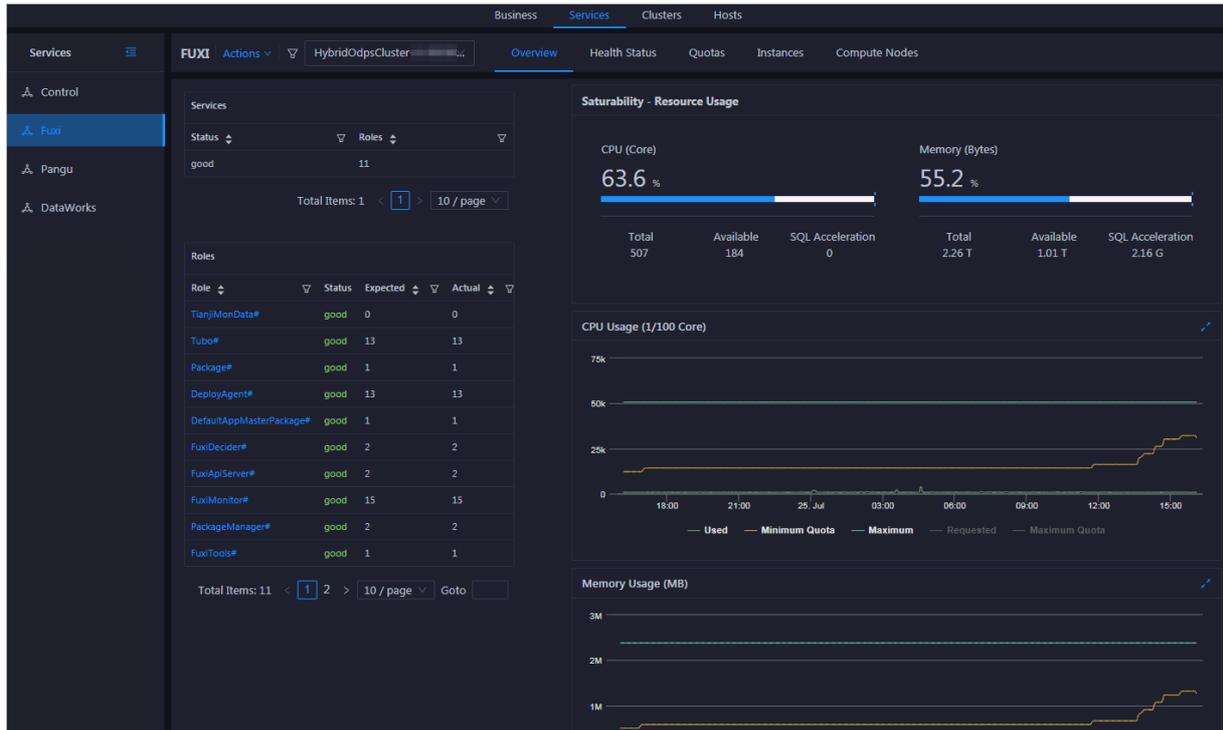
In addition to the preceding features, ABM provides the following custom cluster O&M features for MaxCompute:

- **Server list:** displays a list of all hosts in a MaxCompute cluster and the host information, including the CPU usage, memory usage, disk root directory usage, packet loss rate, and packet error rate.
- **Cluster scaling:** allows you to scale in or out a MaxCompute cluster by removing or adding physical hosts.

Service O&M

ABM allows you to perform O&M on services of big data products. ABM provides custom service O&M features for MaxCompute and StreamCompute, which are different from the service O&M features for other big data products.

For MaxCompute, ABM supports O&M on the control service, DataWorks, Job Scheduler, and Apsara Distributed File System.

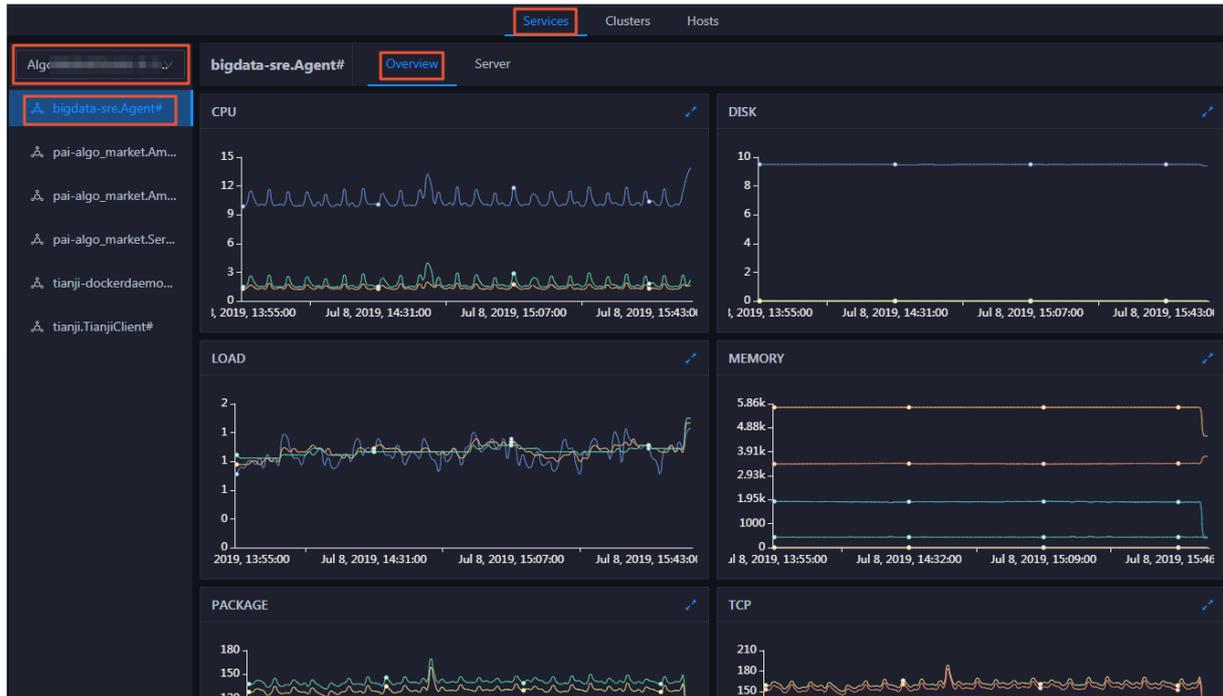


- **Control service:** displays general information about the MaxCompute control service, health check items, and service instances. In addition, you can configure the control service at the cluster level and enable or disable service roles.
- **DataWorks:** displays general information about DataWorks, health check items, service instances, slots, and tasks. In addition, you can modify configurations for DataWorks service roles, and scale in or out DataWorks clusters by removing or adding physical hosts.
- **Job Scheduler:** displays general information about Job Scheduler, health check items, and service instances. In addition, you can manage quota groups, set compute nodes to read-only or read-write, add compute nodes to or remove compute nodes from the blacklist, and enable or disable SQL acceleration.
- **Apsara Distributed File System:** displays general information about Apsara Distributed File System, health check items, and service instances. In addition, you can set the status of storage nodes to disabled or normal, set the status of disks to error or normal, change the primary Master node, empty the recycle bin, enable or disable data rebalancing, and perform the checkpoint operation on Master nodes.

For StreamCompute, ABM supports O&M on Druid and YARN.

- **Druid:** displays the numbers of Master nodes and Worker nodes of Druid.
- **YARN:** displays information about YARN queue API.

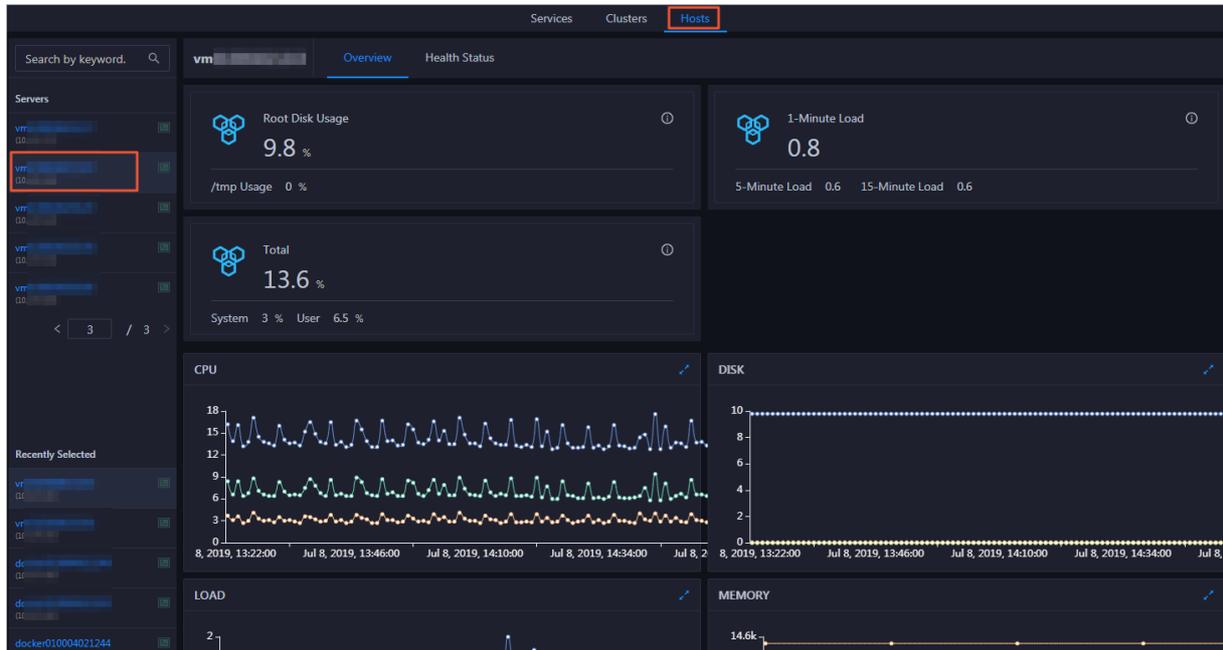
For other big data products, ABM displays all service roles in a cluster and resource usage tendency charts for each service role.



You can select a service from the left-side service list and view the CPU, disk, memory, load, package, TCP, and disk root directory usage of this service.

Host O&M

ABM allows you to perform O&M on hosts of big data products. ABM provides two general host O&M features: host overview and host health.



- **Host overview:** displays information about hosts in a MaxCompute cluster, including the host information, service role status, health check status, health check history, and tendency charts of the CPU, memory, storage, load, and packet loss rate metrics.
- **Host health:** displays the status of all health check items performed on a host. Check items can be in the Critical, Warning, Exception, or OK status.

In addition to the preceding features, ABM provides the following custom host O&M features for MaxCompute:

- **Charts:** displays enlarged charts of the CPU, memory, storage, load, and package metrics. They are the same as the charts displayed in host overview.
- **Host service:** displays the cluster, service instances, and service roles of a host.

Business O&M

Business O&M is a custom O&M feature exclusive for MaxCompute and DataHub. For MaxCompute, ABM provides the following business O&M features: project management, job management, and business optimization.

Project	Cluster	Quota Group	Physical Sto...	Logical Stor...	File Count	Jobs	Owner	Created At	Description	Actions
admin	HYBRIDODPSCLUST	odps_quota	0 B	0 B			ALYUN\$	2019-07-10 15:44:51		Modify Copy-Reso
admin_task_project	HYBRIDODPSCLUST	odps_quota	0 B	0 B	0		ALYUN\$	2019-07-10 15:39:21		Modify Copy-Reso
ads	HYBRIDODPSCLUST	odps_quota	0 B	0 B			ALYUN\$	2019-07-10 15:44:51		Modify Copy-Reso
adsmr	HYBRIDODPSCLUST	odps_quota	2.5 M	856.21 K	6		ALYUN\$	2019-07-10 15:44:51		Modify Copy-Reso
algo_market	HYBRIDODPSCLUST	odps_quota	0 B	0 B	0		ALYUN\$	2019-07-10 15:44:51		Modify Copy-Reso
algo_public	HYBRIDODPSCLUST	odps_quota	0 B	0 B	0		ALYUN\$	2019-07-10 15:44:51		Modify Copy-Reso
base_meta	HYBRIDODPSCLUST	odps_quota	8.58 G	2.86 G	12517		ALYUN\$	2019-07-10 15:44:51		Modify Copy-Reso
base_yunduntest	HYBRIDODPSCLUST	QuotaGroup7a9b05	2.05 M	702.17 K	4		ALYUN\$	2019-07-24 10:26:21		Modify Copy-Reso
biggraph_internal_j	HYBRIDODPSCLUST	biggraph_quota	8.47 M	2.82 M	3	1	ALYUN\$	2019-07-10 15:53:30		Modify Copy-Reso
cosmo_pully	HYBRIDODPSCLUST	odps_quota	0 B	0 B	0		ALYUN\$	2019-07-11 20:51:11		Modify Copy-Reso

- **Project management:**
 - **Project list:** displays all projects deployed on a MaxCompute cluster. You can filter, search, and sort these projects, modify their quota groups, and quickly copy project resources.
 - **Project migration:** allows you to create, manage, and execute project migration tasks, and view task details.
- **Job management:** displays information about jobs deployed on a MaxCompute cluster. You can filter and search these jobs, view their operational logs, and terminate ongoing jobs.
- **Business optimization:** allows you to create, manage, search, and filter cluster merge tasks and project merge tasks.

For DataHub, ABM displays read and write metrics of DataHub projects and topics.

Read and write metrics of projects and topics are displayed by time in a chart. Curves of different colors represent different metrics. You can hide or display a metric by clicking the metric name under the curve chart. A highlighted metric name indicates that the metric is displayed in the chart. Otherwise, the metric is hidden in the chart.

25.4.4. Management

The Management feature is a comprehensive configuration management module provided by Apsara Bigdata Manager (ABM). Its supported functions include job management, patch management, hot upgrade, health configuration, and audit log.

Job management

ABM executes jobs to perform O&M operations on big data products. Jobs are divided into two types: cron jobs and ordinary jobs. Cron jobs are executed automatically on schedule or manually. Ordinary jobs are manually executed.

ABM offers multiple O&M schemes to cover most scenarios. A scheme is a job template. You can easily create and execute jobs by using schemes.

Apart from schemes, ABM also provides an atom library that contains most common O&M operations. An atom is a template of an atomic step. When you create a job by using a scheme, you can directly use atoms as steps to quickly create the job.

Patch management

Patch management allows you to apply patches to Docker containers of each big data product. Docker is an application container engine. It allows you to quickly upgrade product software by replacing only files that need to be updated.

Hot upgrade

Hot upgrade allows you to upgrade checkers in ABM without interrupting services.

Health configuration

ABM provides a variety of built-in checkers for each big data product. These checkers are used to check product faults and generate alerts. In this way, you can detect and rectify faults in time.

Scheduling: You can run checkers on all hosts of a specified Apsara Infrastructure Management Framework role as scheduled to generate raw alert data. The raw alert data includes the checker, host, alert severity, and alert information. ABM stores the raw alert data in its database.

Monitoring: You can mount checkers to product pages in ABM. When mounting a checker to a product page, you can set a filter policy to display only required alerts.

ABM allows you to customize the execution interval, execution parameters, and mount point for a checker, and enable or disable a checker.

Audit log

ABM records the O&M history and details of each O&M operation. This allows you to view the O&M history and locate faults when required.

25.5. Scenarios

If you have deployed Apsara Stack Enterprise and any big data products, you need to use ABM to perform O&M operations on these big data products.

Apsara Stack Enterprise + Big data products

If you have deployed Apsara Stack Enterprise and any big data products, such as MaxCompute, DataWorks, StreamCompute, and DataHub, you need to use ABM to perform O&M operations on these big data products.

25.6. Limits

None.

25.7. Concepts

This topic describes basic concepts of ABM.

Product

A group of clusters. A product provides services for users.

Cluster

A group of physical hosts. A cluster provides services logically and is used to deploy software of a product. A cluster belongs to only one product. You can deploy multiple services on a cluster.

Service

A group of software used to provide an independent feature. A service contains one or more service roles. You can deploy a service on multiple clusters.

Service role

One or multiple indivisible function units of a service. A service role contains one or more applications. If you deploy a service on a cluster, you must deploy all service roles of the service on hosts in the cluster.

Service role instance

A service role on a specific host. A service role can be deployed on multiple hosts. The service role on a specific host is called a service role instance.

Application

A software entity, which is the minimum unit for starting software. Generally, an application is an executable file or a Docker container. If you deploy a service role on a host, you must deploy all applications of the service role on the host.

Service tree

The overall organizational structure of a product. Each product is an independent entity consisting of a certain number of services. The hierarchy of a product's services forms a service tree.

Workflow

A packaged framework that consists of a sequence of processes predetermined based on specific rules. A workflow supports automatic execution. You can use workflows to perform repetitive tasks.

Job

A product O&M task created by users.

Atom

A template of an atomic step. Atoms can be used to create jobs.

Atomic step

An atom that is directly included as a step when you use schemes to create jobs.

Scheme

A job template. You can use schemes to create jobs.

26. Realtime Compute

26.1. What is Realtime Compute?

Alibaba Cloud Realtime Compute is an advanced stream processing platform that provides real-time computations over data streams.

Background

We are seeing an increasing demand for high timeliness and operability of information, which requires software systems to process more data in less time. In traditional models for big data processing, online transaction processing (OLTP) and offline data analysis are separately performed at different times. These models cannot satisfy the growing demand for real-time big data processing.

Realtime Compute comes from the strict demand for the timeliness of data processing. The business value of data decreases as time passes by. Therefore, data must be computed and processed as soon as possible after it is generated. The traditional models for big data processing follow the scheduled processing mode, that is, accumulating and processing data with hours or even days as the computing cycle. This processing mode cannot satisfy the growing demand for computing data streams. Batch (or offline) processing is inapplicable to delay-sensitive scenarios such as real-time big data analytics, risk control and alerting, real-time prediction, and financial transactions. Realtime Compute enables real-time computing over data streams. With Realtime Compute, you can achieve a short data processing delay, easily implement real-time computational logic, and greatly reduce computing costs. This helps you meet the business needs for real-time processing of big data.

Streaming data

Broadly speaking, big data can be viewed as a series of discrete events. These discrete events form event streams or data streams along a timeline. Unlike traditional offline data, streaming data is continuously generated by thousands of data sources. Streaming data is usually sent in the form of data records. Compared with offline data, streaming data is on a smaller scale. Streaming data is generated from endless event streams, including:

- Log files
- Online shopping data
- In-game player activity information
- Social network information
- Financial transaction information
- Geospatial service information
- Telemetry data from devices or instruments

Features

Realtime Compute has the following features:

- Real-time and unbounded data streams

Realtime Compute can compute directly on a real-time, streaming data source. Realtime Compute subscribes to and consumes streaming data in order of time. Data streams are continuously and permanently collected into the Realtime Compute system as long as data is constantly generated. For example, in scenarios where Realtime Compute processes data streams from website visit logs, the log data streams continuously enter the Realtime Compute system before the website is shut down. Therefore, the data in the Realtime Compute system is in real time and unbounded.

- Continuous and efficient computing

Realtime Compute is an *event-driven* system where unbounded event or data streams continuously trigger real-time computations. Once new streaming data enters Realtime Compute, Realtime Compute immediately initiates and performs a computing job. In this regard, the real-time computing of Realtime Compute is an ongoing process that never stops.

- Real-time integration of streaming data

Once a real-time computing job is triggered by streaming data, the computing result is directly written to sinks. For example, you can directly write the computed report data to an RDS system for report display. Realtime Compute can continuously write the computing result of streaming data to sinks, in the same way as data is written to streaming data sources.

26.2. End-to-end real-time computing

Unlike offline or batch computing, end-to-end real-time computing of Alibaba Cloud runs real-time computations over data streams, including real-time data collection, computing, and integration. The real-time computational logic of Realtime Compute ensures a short processing delay.

1. Data collection

You can use data collection tools to collect and send streaming data in real time to a publish-subscribe system for big data analysis. This publish-subscribe system continuously produces events for Realtime Compute in the downstream to trigger stream processing jobs.

2. Stream processing

Data streams continuously enter Realtime Compute for real-time computing. At least one data stream must enter the Realtime Compute system to trigger a real-time computing job. Each batch of incoming data records initiates a stream processing procedure in Realtime Compute. The computing results for each batch of data records are then instantly provided.

3. Data integration

Realtime Compute allows you to write the result data of stream processing to sinks, such as tables of data stores and message delivery systems. You can also integrate Realtime Compute with the alerting system that is connected to your business applications. This enables you to easily receive alerts if the specified business rules for alerting are satisfied. Unlike batch computing products such as MaxCompute and open source Apache Hadoop, Realtime Compute inherently comes with data integration modules that allow you to write result data to sinks.

4. Data consumption

After the result data of stream processing is written to sinks, the data consumption phase is decoupled from real-time computing. You can use data stores, data transmission systems, or alerting systems to access the result data, send and receive the result data, or send alerts, respectively.

26.3. Differences between real-time computing and batch computing

26.3.1. Overview

Compared with batch computing, real-time computing has made groundbreaking progress in the field of big data computing. This section describes the differences between batch computing and real-time computing from two aspects: users and products.

 **Note** For more detailed theoretical analysis, see Wikipedia: [Stream processing](#).

26.3.2. Batch computing

Batch computing models have been used for most traditional data computing and analysis services. In batch computing models, extract-transform-load (ETL) or online transaction processing (OLTP) systems are used to load data into data stores. The loaded data is then used for online data services, such as ad-hoc queries and dashboard services, based on SQL statements. You can also use SQL statements to obtain results from the analysis.

Batch computing models are widely accepted along with the evolution of relational databases in diversified industries. However, in the era of big data, with the increasing number of human activities being converted to information and then data, more and more data requires real-time and stream processing. The current processing models are facing great challenges in real-time processing.

A typical batch computing model is described as follows:

1. An ETL or OLTP system is used to build data stores and provides raw data for computing and analysis. The batch computing model where users load the data and the batch computing system optimizes queries on the loaded data using multiple methods, such as creating indexes, based on its storage and computing capabilities. In batch computing models, data must be loaded into the batch computing system. Newly arriving data records are collected into a batch and the entire batch is then processed after all data in the batch is loaded.
2. A user or system initiates a computing job, such as a MaxCompute SQL job or Hive SQL job, and submits requests to the ETL or OLTP system. The batch computing system then schedules computing nodes to perform computations on large amounts of data. This may take several minutes or even hours. The mechanism of batch computing determines that the data to be processed is the accumulated historical data. As a result, the data processing may not be in real time. In batch computing, you can change computational logic using SQL at any time to meet your needs. You can also perform ad-hoc queries instantly after changing the logic.
3. The computing results are returned in the form of data sets when a computing job is completed. If the size of result data is excessively large, the result data is stored in the batch computing system. In this scenario, you can integrate the batch computing system with another system to view the result data. Large amounts of result data lead to a lengthy process of data integration. The process may take several minutes or even hours.

Batch computing jobs are initiated by users or systems and are processed with a long delay. The batch computing procedure is described as follows:

1. You load data into the data processing system.
2. You submit computing jobs. In this phase, you can change computing jobs to meet your business needs, and publish the changed jobs.
3. The batch computing system returns the computing results.

26.3.3. Real-time computing

Unlike batch computing, real-time computing runs real-time computations over data streams and allows for a low processing delay. The differences between real-time computing and batch computing are described as follows:

1. **Data integration.** For real-time computing, data integration tools are used to send streaming data in real time to streaming data stores such as DataHub. For batch computing, large amounts of data are accumulated and then processed. In contrast, streaming data is sent in micro batches in real time, which ensures a short delay for data integration.

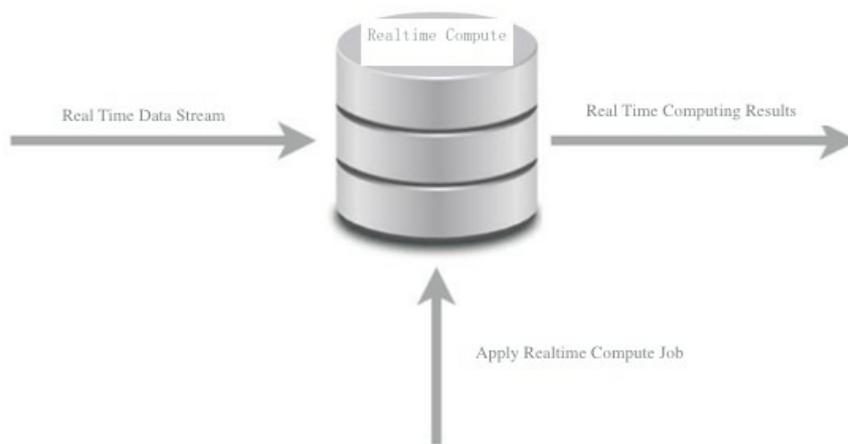
The streaming data is continuously written to data stores in real time. You do not need to preload data for processing. Realtime Compute does not store real-time data that is continuously processed. The real-time data is discarded instantly after it has been processed.

2. **Data computing.** For batch computing, data is processed only after large amounts of data have been accumulated. In contrast, a real-time computing job is resident in the system and waits to be triggered by events once it is started. Each incoming micro batch of streaming data records initiates a real-time computing job. The computing results are instantly provided by Realtime Compute. Realtime Compute also divides large batches of data records into smaller batches for incremental computing. This effectively shortens the processing delay.

For real-time computing, you must predefine the computational logic in Realtime Compute. You cannot change the computational logic when real-time computing jobs are running. If you terminate a running job and publish the job after changing the computational logic, the streaming data that has been processed before the change cannot be processed again.

3. **Writing result data to target systems.** For batch computing, result data can be written to online systems by batch only after all accumulated data has been processed. In contrast, real-time computing allows for writing result data to online and offline systems instantly after each micro batch of data records has been processed. This allows you to view the computing results in real time.

Realtime computing



Realtime Compute runs real-time computations over data streams, which are continuously generated from data sources, based on an event-driven mechanism. Realtime Compute allows you to process data streams with a short delay. The real-time computing procedure is described as follows:

1. You publish real-time computing jobs.
2. Streaming data triggers real-time computing jobs.
3. Realtime Compute constantly returns the computing results.

26.3.4. Comparison between real-time computing and batch computing

Comparison between real-time computing and batch computing shows the differences between real-time computing and batch computing.

Comparison between real-time computing and batch computing

Item	Batch computing	Real-time computing
Data integration	You load data into the data processing system.	Data is loaded and processed in real time.
Computational logic	The computational logic can be changed, and data can be reprocessed.	After the computational logic is changed, data cannot be reprocessed. This is because streaming data is processed in real time.
Data scope	You can query and process all or most of the data in the data set.	You can query and process the latest data record or the data within the tumbling window.
Data size	It processes large batches of data.	It processes individual records or micro batches consisting of a few records.
Performance	It achieves a processing delay of several minutes or hours.	It achieves a processing delay of several seconds and even milliseconds.

Item	Batch computing	Real-time computing
Analysis	You can perform complex data analysis.	You can perform simple analysis, such as simple response functions, aggregates, and rolling metrics.

Realtime Compute uses a simple computing model. Real-time computing of Realtime Compute makes significant improvements to batch computing in most scenarios of big data computing. In particular, in scenarios where event streams need to be processed with an extremely low processing delay, real-time computing is a valuable service for big data computing.

26.4. Benefits

Realtime Compute provides competitive advantages in stream processing, which allows you to easily handle the demand for real-time big data analysis. Realtime Compute offers the following benefits:

Powerful real-time computing functions

Realtime Compute simplifies the development process by integrating a wide range of functions. These functions are described as follows:

- A powerful engine is used. This engine offers the following advantages:
 - Provides the standard Flink SQL that enables automatic data recovery from failures. This ensures accurate data processing when failures occur.
 - Supports multiple types of built-in functions, such as text functions, date and time functions, and statistics functions.
 - Enables an accurate control over computing resources. This ensures complete isolation of each tenant's jobs.
- The key performance metrics of Realtime Compute are three to four times higher than those of Apache Flink. For example, in Realtime Compute, the data processing delay is reduced to seconds or even to sub-second level. The throughput of a job reaches millions of data records per second. A cluster can contain thousands of nodes.
- Realtime Compute integrates cloud-based data stores such as MaxCompute, DataHub, Log Service, ApsaraDB for RDS, Table Store, and AnalyticDB for MySQL. With Realtime Compute, you can read data from and write data to these systems with the least efforts in data integration.

Managed real-time computing services

Unlike open source or user-developed stream processing services, Realtime Compute is a fully managed stream processing engine. You can query streaming data without deploying or managing any infrastructure. With Realtime Compute, you can use streaming data processing services with a few clicks. Realtime Compute integrates services such as development, administration, monitoring, and alerting. This allows you to use cost-effective streaming data services for trial and migrate your data for deployment.

Realtime Compute also enables complete isolation between tenants. This isolation and protection extends from the top application layer to the underlying infrastructure layer. This helps to ensure the security and privacy of your data.

Excellent user experience during development

Realtime Compute provides a standard SQL engine: Flink SQL. It also provides many built-in functions, such as the text functions, date and time functions, and statistics functions. The application of these functions greatly simplifies and accelerates the Flink-based development. With Flink SQL, even users with limited development knowledge, such as business intelligence (BI) analysts and marketers, can easily perform real-time analysis and processing of big data.

Realtime Compute provides an end-to-end solution for stream processing, including development, administration, monitoring, and alerting. On the Realtime Compute development platform, only three steps are required to publish a job.

Low costs in labors and compute clusters

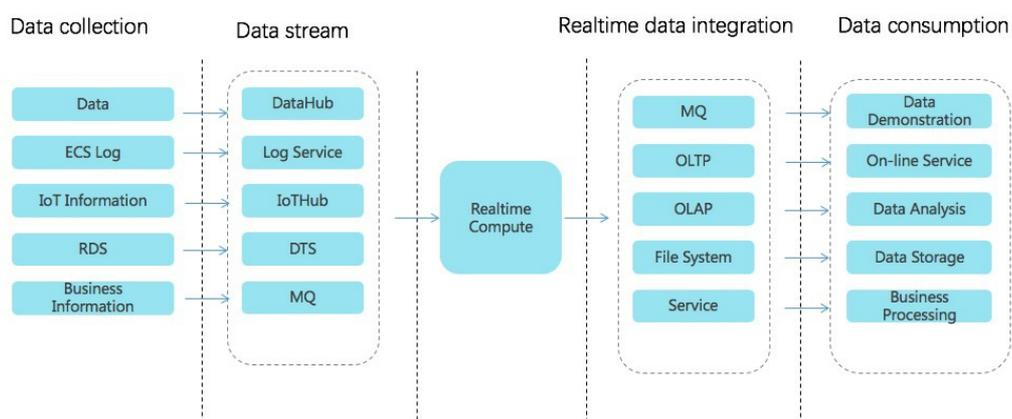
We have made many improvements to the SQL execution engine, allowing you to create jobs more cost-effectively than to create Flink jobs. Realtime Compute is more cost-effective than open source stream frameworks in both development and production costs. To create an Apache Storm job with complex computational logic, you have to incur high costs and devote a lot of effort, such as writing enormous lines of Java code, debugging, testing, performance tuning, publishing, and long-term administration of open source software applications like Apache Storm and Zookeeper. Realtime Compute allows you to offload the heavy lifting of handling these issues, which helps you focus on your business strategies and rapidly achieve market goals.

26.5. Product architecture

26.5.1. Business process

We recommend that you have a general knowledge about the stream processing architecture of StreamCompute before using this service. This helps you create effective plans for the design of stream processing systems. **Architecture** shows the stream processing architecture of StreamCompute.

Architecture



- Data collection

You can use data collection tools to collect and send streaming data in real time to a publish-subscribe system for big data analysis. This publish-subscribe system continuously produces events for Realtime Compute in the downstream to trigger real-time computing jobs. The big data ecosystem of Alibaba Cloud offers a wide range of publish-subscribe systems to process streaming data in diversified scenarios. Realtime Compute integrates many of these systems, as shown in the preceding figure. This allows you to easily integrate multiple streaming data stores. To enable compatibility between the computing model of Realtime Compute and that of certain data stores, another data store may be required for data processing. Realtime Compute is seamlessly connected to the following data stores:

- DataHub

DataHub allows you to upload data into its system using a wide range of tools and interfaces. For example, you can easily upload logs, binary log files, and IoT streaming data into the DataHub system. DataHub also integrates open source business software applications. For more information about the data collection tools of DataHub, see DataHub documentation.

- Log Service

Log Service is a one-stop logging service that has been developed by Alibaba Group based on years of experience in addressing challenges involving large amounts of big data experienced by Alibaba Group. Log Service allows you to quickly collect, transfer, query, consume, and analyze log data.

- IoT Hub

IoT Hub is a service that enables developers of IoT applications to implement two-way communications between devices (such as sensors, final control elements, embedded devices, and smart home appliances) and the cloud by creating secure data channels.

You can use the IoT Hub rule engine to easily send IoT data to DataHub, and use Realtime Compute and MaxCompute to process and perform computations on data.

- Data Transmission Service (DTS)

DTS supports data transmission between structured data stores represented by databases. DTS is a data exchange service that streamlines data migration, data synchronization, and data subscription. You can use the data transmission function of DTS to easily parse binary log files such as RDS logs and send data to DataHub. Realtime Compute and MaxCompute allow you to run computations over the data.

- MQ

Message Queue (MQ) is a key service that provides messaging capabilities, such as message publishing and subscription, message tracing, scheduled, and delayed messages, resource statistics, monitoring, and alerting. MQ offers a complete set of enterprise-level messaging functions powered by high-availability (HA) distributed systems and clusters.

- Realtime computing

Data streams continuously enter Realtime Compute for real-time computing. At least one data stream must enter the Realtime Compute system to trigger a real-time computing job. In complex business scenarios, Realtime Compute allows you to perform association queries for static data from data stores and streaming data. For example, you can perform JOIN operations on DataHub and RDS tables based on the primary key of streaming data. You can then perform association queries on DataHub streaming data and RDS static data. Realtime Compute also enables you to associate multiple data streams. With Flink SQL, you can easily handle large amounts of data and complex business scenarios, such as those experienced by Alibaba Group.

- Realtime data integration

To minimize the data processing delay and simplify data transmission links, Realtime Compute directly writes the result data of real-time computing to data sinks. Realtime Compute allows for a larger Alibaba Cloud ecosystem by integrating the following systems: online transaction processing (OLTP) systems such as ApsaraDB for RDS, NoSQL database services such as Table Store, online analytical processing (OLAP) systems such as AnalyticDB, message queue systems such as DataHub and RocketMQ, and mass storage systems such as Object Storage Service (OSS) and MaxCompute.

- Data consumption

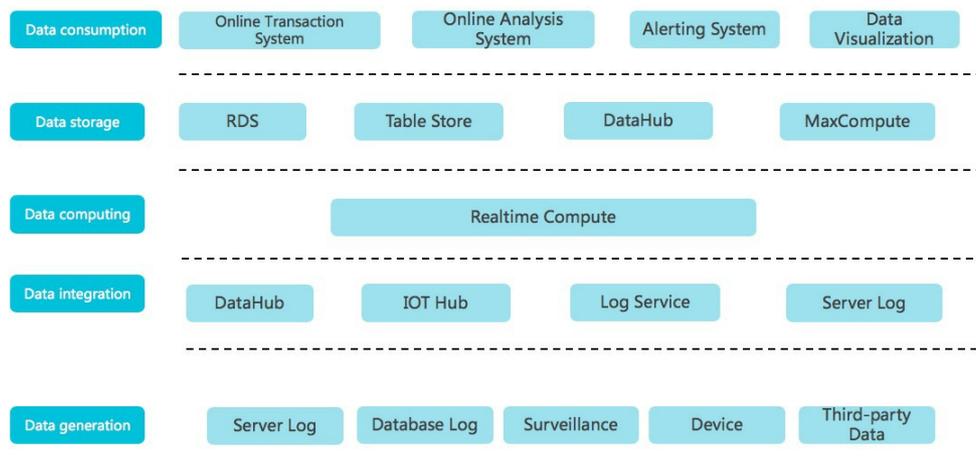
After the result data of real-time computing is written to the sinks, you can consume the data using custom applications.

- You can use data stores to access the result data.
- You can use data transfer systems to send and receive the result data.
- You can use alerting systems to send alerts.

26.5.2. Business architecture

Realtime Compute is a lightweight SQL-enabled streaming engine for real-time processing and analysis of data streams.

Business architecture



- Data generation

In this phase, streaming data is generated from sources such as server logs, database logs, sensors, and third-party systems. The generated streaming data moves on to the next phase for data integration to drive real-time computing.

- Data integration

In this phase, the streaming data is integrated. You can subscribe to and publish the integrated streaming data. The following Alibaba Cloud products can be used in this phase: DataHub for big data computing, IoT Hub for connecting IoT devices, and Log Service for integrating ECS logs.

- Data computing

In this phase, the streaming data, which has been subscribed to in the data integration phase, acts as inputs to drive real-time computing in Realtime Compute.

- Data store

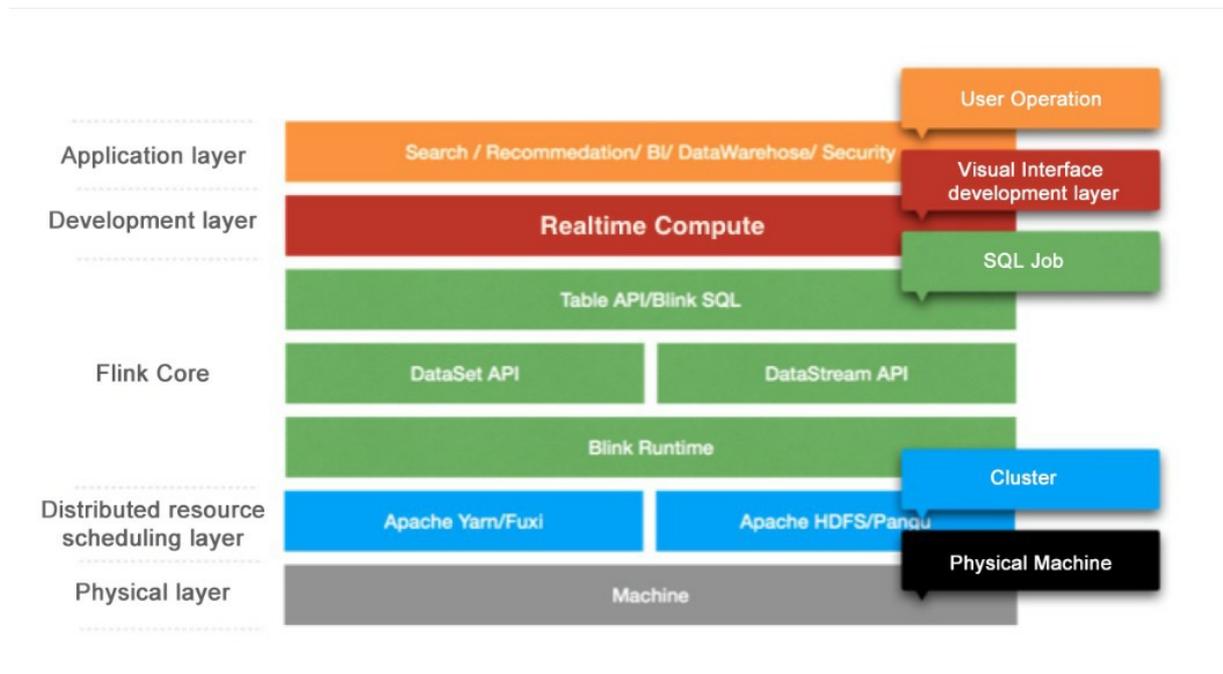
Realtime Compute does not provide built-in data stores. Instead, it writes computing results to external data stores, such as relational databases, NoSQL databases, and online analytical processing (OLAP) systems.

- Data consumption

Realtime Compute supports multiple data store types, which allows you to consume data in various ways. For example, data stores for message queues can be used to report alerts, and relational databases can be used to provide online support.

26.5.3. Technical architecture

Realtime Compute is a real-time data analysis platform for incremental computing. This platform provides statements that are similar to SQL statements and uses the MapReduceMerge (MRM) computing model for incremental computing. Realtime Compute offers a failover mechanism to ensure data accuracy when errors occur.



The Realtime Compute architecture consists of the following five layers.

- Application layer

This layer allows you to create SQL files and publish jobs for real-time data processing based on a development platform. With a well-designed monitoring and alerting system, you would be notified of a processing delay for each job in a timely manner. You can also use systems like Flink UI to view the running information of published jobs and analyze performance bottlenecks. This allows you to quickly and effectively improve job performance.

- Development layer

This layer parses Flink SQL and generates logical and physical execution plans. The execution plans are then conceptualized as executable directed acyclic graphs (DAGs). Based on these DAGs, directed graphs that consist of various models are obtained. Directed graphs are used to implement specific business logic. A model usually contains the following three modules:

- **Map:** Operations such as data filtering, distribution (GROUP), and join (MAPJOIN) are performed.
- **Reduce:** Realtime Compute processes streaming data by batch, and each batch contains multiple data records.
- **Merge:** You can update the state by merging the computing results of the batch, which are produced from the Reduce module, with the previous state. Checkpoints are created after N (configurable) batches have been processed. In this way, the state is stored persistently in a data store, such as Tair and Apache HBase.

- **Flink Core**

This layer provides a wide range of computing models, Table API, and Flink SQL. You can use DataStream API and DataSet API at the lower sublayer. At the bottom sublayer is Flink Runtime, which schedules resources to ensure that jobs can run properly.

- **Distributed resource scheduling layer**

Realtime Compute clusters run based on the Gallardo scheduling system. This system ensures that Realtime Compute runs effectively and fault tolerance is provided for recovery.

- **Physical layer**

This layer provides powerful hardware devices for clusters.

26.6. Features

Realtime Compute has the following features:

- **Data collection and storage**

The premise of running a big data analysis system is that data has been collected into the system. To make full use of your existing streaming data store, Realtime Compute supports integration with multiple upstream streaming data stores, such as DataHub, Log Service, IoT Hub, Table Store, and MQ. You can use streaming data in existing data stores without operations of data collection and data integration.

You can register data stores on the Realtime Compute development platform. This enables you to leverage the advantages of the one-stop Realtime Compute development platform. Realtime Compute provides the UI for managing different data stores, such as ApsaraDB for RDS, AnalyticDB, and Table Store. Realtime Compute allows you to manage cloud-based data stores in one stop.

- **Data development**

- Realtime Compute provides a fully managed online development platform that integrates a wide range of SQL coding assistance features, such as Flink SQL syntax checking, intelligent code completion, and syntax highlighting.
 - **Syntax checking**

On the Development page of Realtime Compute, the revised script is automatically saved. When the script is saved, an SQL syntax check is automatically performed. If a syntax error is detected, the Development page shows the row and column where the error is located, and the cause of the error.
 - **Intelligent code completion**

When you enter SQL statements on the Development page of Realtime Compute, auto-completion prompts about keywords, built-in functions, and SQL statements are automatically displayed.
 - **Syntax highlighting**

Flink SQL keywords are highlighted in different colors to differentiate data structures.
- The Realtime Compute development platform allows you to manage different versions of SQL code.

Realtime Compute provides key features that help you complete development tasks, such as coding assistance and code version management. On the data development platform, you can manage SQL code versions. Each time you commit code, the system generates a code version, which can be used for version tracking, modification, and rollback.
- The Realtime Compute development platform allows you to register data stores on its Development page for effective data store management, such as data preview and auto DDL generation.
 - **Data preview**

The Development page of Realtime Compute allows you to preview the data of multiple data store types. Data preview helps you efficiently analyze upstream and downstream data, identify key business logic, and complete development tasks.
 - **Auto DDL generation**

In most cases, the DDL statements for data stores are manually translated into the DDL statements for real-time computing. Therefore, the DDL generation process includes a large number of repetitive tasks. Realtime Compute provides an auto DDL generation feature. This feature simplifies the way that you edit SQL statements for stream processing jobs, reduces the possibility of encountering errors when you manually enter SQL statements, and also improves efficiency.
- Realtime Compute allows you to implement real-time data cleansing, statistics, and analysis using standard SQL. Realtime Compute also supports common aggregation functions, and association queries for streaming data and static data.
- The Realtime Compute development platform provides a simulated running environment where you can customize uploaded data, simulate operations, and check output results.
- **Data operation**

Realtime Compute allows you to manage stream processing jobs on the following tabs under the Administration page: Overview, Curve Charts, FailOver, CheckPoints, JobManager, TaskExecutor, Data Lineage, and Properties and Parameters.
- **Performance tuning**

- Improve performance by automatic configuration

The automatic configuration function of Realtime Compute helps you address performance issues, such as a low throughput of jobs and data piling up in the upstream.

- Improve performance by manual configuration

You can manually configure resources to improve job performance using one of the following methods:

- Optimize resource configuration. You can modify the resources to improve performance by reconfiguring parameters, such as parallelism, core, and heap_memory.
- Improve performance based on job parameter settings. You can specify the job parameters such as miniBatch to improve performance.
- Improve upstream and downstream data stores based on parameter settings. You can specify related parameters to optimize the upstream and downstream data stores for a job.

- **Monitoring and alerting**

This allows you to collect the performance metrics of cloud resources or other custom performance metrics, view service availability, and specify alerts based on the performance metrics. In this way, you can easily view the cloud resource usage and running information of jobs. You can also receive and respond to alerts in a timely manner to ensure that applications can run properly. With Realtime Compute, you can specify alerts for the following performance metrics:

- Processing delay
- Input RPS
- Output RPS
- Failover rate

26.7. Product positioning

Realtime Compute offers Flink SQL to support standard SQL semantics and help you easily implement the computational logic of stream processing. Realtime Compute also provides full-featured UDFs for some authorized users, helping you customize business-specific data processing logic in scenarios where SQL code functions cannot meet your business needs. In the field of streaming data analysis, you can directly use Flink SQL and UDFs to enable most of the streaming data analysis and processing logic. Realtime Compute focuses on the analysis, statistics, and processing of streaming data. It is less applicable to non-SQL businesses, such as complex iterative data processing and complex rule engine alerts.

Realtime Compute is applicable to the following scenarios:

- Collects the data about page views (PVs) and unique visitors (UVs) in real time.
- Collects the data about the average traffic flow at a traffic checkpoint every 5 minutes.
- Collects and displays the pressure data of hydroelectric dams.
- Reports alerts for financial thefts in online payment services based on fixed rules.

Realtime Compute is inapplicable to the following scenarios for now:

- Replacing Oracle stored procedures with Realtime Compute: Realtime Compute cannot implement all the functions of Oracle stored procedures, because they are designed to handle issues in different fields.
- Seamlessly migrating Spark jobs to Realtime Compute: Currently, you cannot seamlessly migrate Spark jobs to Realtime Compute. However, you can change the stream processing of

Apache Spark and migrate this part to Realtime Compute. This eliminates various Apache Spark administration tasks and Spark-based development costs.

- **Complex rule engines for alerting:** Realtime Compute cannot handle scenarios where multiple complex alerting rules are specified for each data record, and the rules continue to change when the system is running. Specific rule engines need to be used to resolve these issues.

Realtime Compute provides a full set of development tools for streaming data analysis, statistics, and processing based on UDFs and Flink SQL. It allows you to devote the least efforts in developing the underlying code and simply write SQL statements to analyze streaming data. This makes Realtime Compute a good choice for users such as data warehouse developers and data analysts.

26.8. Scenarios

26.8.1. Overview

Realtime Compute uses Flink SQL to provide solutions for streaming data analysis.

- **Real-time extract-transform-load (ETL)**

Realtime Compute allows you to cleanse, aggregate, and sort streaming data in real time by leveraging the advantages of multiple data channels and flexible data processing capabilities of SQL. Realtime Compute serves as an effective supplement and optimization of offline data warehouses and provides a computing channel for real-time data transmission.

- **Real-time reports**

Realtime Compute allows you to collect and process streaming data, monitor performance metrics of the business, and view corresponding reports in real time. This enables real-time data administration.

- **Monitoring and alerting**

Realtime Compute allows you to monitor systems and analyze user behavior in real time, which helps to identify faults and risks in real time.

- **Online systems**

Realtime Compute allows you to run real-time computations over data streams and view performance metrics in real time. You can shift strategies for online systems in a timely fashion. Realtime Compute can be widely used in various content delivery and intelligent mobile push scenarios.

26.8.2. Management of e-commerce activities

Realtime Compute has evolved into a reliable stream processing platform from Alibaba Group's big data architecture in the e-commerce industry. Realtime Compute is suitable for analyzing various streaming data and providing report support in the e-commerce industry. The e-commerce industry needs to process streaming data in real time in the following scenarios:

- **Real-time analysis of user behavior,** for example, display of transaction data and user data on big screens. In traditional batch processing models, large amounts of data are processed inefficiently with a long delay. The size of the result data may be excessively large, which poses considerable challenges for online systems that are used for displaying the result data. This may compromise the stability of the online systems.
- **Real-time monitoring of users, services, and systems.** For example, marketers and engineers can have knowledge of the transactions on the platform over a specified period by viewing the

corresponding curve chart. If abnormal fluctuations occur, such as a sharp decrease in transactions, alerts must be instantly triggered and sent to users. This helps users effectively respond to abnormal fluctuations and reduces negative impacts on the business.

- Real-time monitoring of major promotional events. For example, marketers need to monitor the metrics of promotional events in real time, such as the Double 11 Shopping Festival created by Alibaba Group and 618 mid-year shopping festival started by JD.com, Inc. This helps marketers effectively decide whether to change strategies.

Integrating with Alibaba Cloud computing and storage systems, Realtime Compute allows you to meet your custom needs for streaming data analysis. Realtime Compute not only satisfies diverse business needs but also simplifies the business development process by using Flink SQL.

26.8.3. Multidimensional analysis of data from IoT sensors

Background

With the economic tidal wave of globalization sweeping over the world, industrial manufacturers are facing increasingly fierce competition. To increase competitiveness, manufacturers in the automotive, aviation, high-tech, food and beverage, textile, and pharmaceutical industries must innovate and replace the existing infrastructure. These industries have to address many challenges during the innovation process. For example, the existing traditional devices and systems have been used for decades, which results in high maintenance costs. However, replacing these systems and devices may slow down the production process and compromise the product quality.

These industries face two additional challenges, which are high security risks and the urgent need for complex process automation. The manufacturing industry has prepared to replace the existing traditional devices and systems. In this industry, high reliability and availability systems are needed to ensure the safety and stability of real-time operations. A manufacturing process involves a wide range of components, such as robotic arms, assembly lines, and packaging machines. This requires remote applications that can seamlessly integrate each stage of the manufacturing process, including the deployment, update, and end-of-life management of devices. The remote applications also need to handle failover issues.

Another requirement for these next-generation systems and applications is that they be able to capture and analyze the large amounts of data generated by devices, and respond appropriately in a timely manner. To increase competitiveness and accelerate development, manufacturers need to optimize and upgrade their existing systems and devices. The application of Realtime Compute and Alibaba Cloud IoT solutions allows you to analyze device running information, detect faults, and predict yield rates in real time. This topic describes a use case as an example. In this use case, a manufacturer uses Realtime Compute to analyze the large amounts of data collected from sensors in real time. Realtime Compute is also used to cleanse and aggregate data in real time, write data to an online analytical processing (OLAP) system in real time, and monitor the key metrics of devices in real time.

Scenario description

In this use case, the manufacturer has more than 1,000 devices from multiple factories in many cities. Each device is equipped with 10 types of sensors. These sensors send the collected data every 5 seconds to Log Service. The data collected from each sensor follows the format described in the following table.

s_id	s_value	s_ts
The ID of the sensor.	The current value from the sensor.	The time when the data was sent.

The sensors are distributed across devices from multiple factories. The manufacturer creates an RDS dimension table to display the distribution of sensors across devices and factories.

s_id	s_type	device_id	factory_id
The ID of the sensor.	The type of the sensor.	The ID of the device.	The ID of the factory.

The information included in this dimension table is stored in the RDS system. The manufacturer needs to organize the data from sensors based on this dimension table, and sort the data by device. To meet this need, Realtime Compute provides a summary table where the data sent from sensors is logically aggregated by device every minute.

ts	device_id	factory_id	device_temp	device_pres
The time when the data was sent.	The ID of the device.	The ID of the factory.	The temperature of the device.	The pressure of the device.

Assume that there are only two types of sensors in this use case: temperature and pressure. The computational logic is described as follows:

1. Realtime Compute identifies the devices whose temperatures are higher than 80°C and triggers alerts at the downstream nodes. In this use case, Realtime Compute sends the data of the identified devices to MQ. MQ then triggers alerts that the manufacturer has specified in the downstream alerting system.
2. Realtime Compute writes the data to an OLAP system. In this use case, the manufacturer uses HybridDB for MySQL. To integrate with HybridDB for MySQL, the manufacturer has developed a set of business intelligence (BI) applications for multidimensional data display.

FAQ

- How can I aggregate data into a summary table?

In most cases, each sensor only collects the IoT data of one dimension. This poses challenges for subsequent data processing and analysis. To create a summary table, Realtime Compute aggregates data based on windows and organizes data by dimension.

- Why is MQ used to trigger alerts?

Realtime Compute allows you to write data to any type of storage system. We recommend that you use message storage systems like MQ for sending alerts and notifications. This is because the application of these systems helps to prevent the errors encountered by user-defined alerting systems. These errors may cause failures to report certain alerts and notifications.

Code description

Send the data uploaded from sensors to Log Service. The data format of a row is shown as follows:

```
{
  "sid": "t_xxsfdsad",
  "s_value": "85.5",
  "s_ts": "1515228763"
}
```

Define a Log Service source table `s_sensor_data`.

```
CREATE TABLE s_sensor_data (
  s_id VARCHAR,
  s_value VARCHAR,
  s_ts VARCHAR,
  ts AS CAST(FROM_UNIXTIME(CAST(s_ts AS BIGINT)) AS TIMESTAMP),
  WATERMARK FOR ts AS withOffset(ts, 10000)
) WITH (
  TYPE='sls',
  endPoint ='http://cn-hangzhou-corp.sls.aliyuncs.com',
  accessId ='*****',
  accessKey ='*****',
  project ='ali-cloud-streamtest',
  logStore ='stream-test',
);
```

Create an RDS dimension table `d_sensor_device_data`. This dimension table stores the mappings between sensors and devices.

```
CREATE TABLE d_sensor_device_data (
  s_id VARCHAR,
  s_type VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  PRIMARY KEY(s_id)
) WITH (
  TYPE='RDS',
  url="",
  tableName='test4',
  userName='test',
  password='*****'
);
```

Create an MQ result table `r_monitor_data`. This table specifies the logic for triggering alerts.

```
CREATE TABLE r_monitor_data (
  ts VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  device_TEMP DOUBLE,
  device_PRES DOUBLE
) WITH (
  TYPE='MQ'
);
```

Create a HybridDB for MySQL result table `r_device_data`.

```
CREATE TABLE r_device_data (
  ts VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  device_temp DOUBLE,
  device_pres DOUBLE,
  PRIMARY KEY(ts, device_id)
) WITH (
  TYPE='HybridDB'
);
```

Aggregate the data collected from sensors by minute and create a summary table based on the aggregated data. To clearly view the code structure and facilitate subsequent administration, we create views in this use case.

```
// Create a view to obtain the device and factory mapping each sensor.
CREATE VIEW v_sensor_device_data
AS
SELECT
s.ts,
s.s_id,
s.s_value,
s.s_type,
s.device_id,
s.factory_id
FROM
s_sensor_data s
JOIN
d_sensor_device_data d
ON
s.s_id = d.s_id;

// Aggregate the data collected from sensors.
CREATE VIEW v_device_data
AS
SELECT
// Specify the start time of a tumbling window as the time for the record.
CAST(TUMBLE_START(v.ts, INTERVAL '1' MINUTE) AS VARCHAR) as ts,
v.device_id,
v.factory_id,
CAST(SUM(IF(v.s_type = 'TEMP', v.s_value, 0)) AS DOUBLE)/CAST(SUM(IF(v.s_type = 'TEMP', 1, 0)) AS DOUBLE) device_temp, // Compute the average temperature by minute.
CAST(SUM(IF(v.s_type = 'PRES', v.s_value, 0)) AS DOUBLE)/CAST(SUM(IF(v.s_type = 'PRES', 1, 0)) AS DOUBLE) device_pres // Compute the average pressure by minute.
FROM
v_sensor_device_data v
GROUP BY
TUMBLE(v.ts, INTERVAL '1' MINUTE), v.device_id, v.factory_id;
```

In the preceding core computational logic, the average temperature and pressure by minute are computed as the output. Tumbling windows are used in this use case. A new window is started every minute, and a new set of data is generated every minute. The generated data is then filtered and written to the MQ result table and HybridDB result table.

```
// Identify the sensors whose temperatures are higher than 80°C and write the data to the MQ result table to trigger alerts.
INSERT INTO r_monitor_data
SELECT
ts,
device_id,
factory_id,
device_temp,
device_pres
FROM
v_device_data
WHERE
device_temp > 80.0;
// Write the result data to the HybridDB for MySQL result table for analysis.
INSERT INTO r_device_data
SELECT
ts,
device_id,
factory_id,
device_temp,
device_pres
FROM
v_device_data;
```

26.8.4. Big screen service for the Tmall Double 11 Shopping Festival

The annual Tmall Double 11 Shopping Festival has become the largest sales event for online shopping in the world. A large number of netizens demonstrate a strong desire to purchase products during the sales event each year. One of the key highlights of this event has been the increase in the overall turnover that is displayed on the Tmall big screen in real time. The real-time display of turnover on the big screen is a result of our senior engineers' hard work over several months. The big screen service excels in key performance metrics. For example, the end-to-end delay has been reduced within 5 seconds, from placing orders on the Tmall platform, to data collection, processing, verification, and to displaying the sales data on the big screen. As for the processing capability, hundreds of thousands of orders can be processed during the peak hours around 00:00 on November 11. Additionally, to ensure fault tolerance, multiple channels have been used to back up data.

Realtime Compute provides key support for the big screen service. The stream processing of the big screen service was previously based on the open source Apache Storm. The Storm-based development process took around one month. The application of Flink SQL shortened the development process of the big screen service to one week. The underlying layer of Realtime Compute removes the Apache Storm modules that are designed for execution optimization and troubleshooting. This enables higher efficiency and faster processing for Realtime Compute jobs.

- Online shopping rush

During the Double 11 Shopping Festival, an enormous number of netizens join the online shopping rush on the Tmall platform. During the peak hours when "seckill" activities occur, such as 00:00 on November 11, hundreds of thousands of sales orders need to be processed in real time. The word "seckill" vividly describes fighting among buyers, which means that a buyer wins or loses all in a matter of seconds.

- Real-time data collection

The data collection system collects and sends the logs of database changes to the DataHub system. With the application of Data Transmission Service (DTS), the data from online transaction processing databases can be written to DataHub tables within seconds at the peak hours around 00:00 on November 11.

- Real-time data computing

Realtime Compute subscribes to the DataHub streaming data, continuously analyzes the streaming data, and calculates the total turnover up to the current time. In Realtime Compute, a cluster can contain up to thousands of nodes. The throughput of a job reaches millions of data records per second, fully meeting the system requirements of processing hundreds of thousands of transactions per second in Tmall. Realtime Compute subscribes to data and writes the result data to an online RDS system in real time.

- Frontend data visualization

We also provide advanced data visualization components for the Tmall Double 11 Shopping Festival. These components allow you to view the total turnover on a dashboard, and the distribution of global transaction activities across the world in real time. To achieve astounding visual effects for the big screen, the frontend server performs periodic polling operations on the RDS system, and advanced web frontend applications are used.

26.8.5. Mobile data analysis

Realtime Compute allows you to analyze the data of mobile apps in real time. With Realtime Compute, you can analyze performance metrics of mobile apps, such as crash detection and distribution, and distribution of app versions. Mobile Analytics is a product provided by Alibaba Group to analyze the data of mobile apps. This product allows you to analyze user behavior and logs from multiple dimensions. It also helps mobile developers implement fine-grained operations based on big data analysis, improve product quality and customer experience, and enhance customer stickiness. The underlying big data computing of Mobile Analytics is implemented based on big data products of Alibaba Cloud, such as Realtime Compute and MaxCompute. Mobile Analytics uses Realtime Compute as the underlying engine for streaming data analysis. This allows Mobile Analytics to offer a wide range of real-time data analysis and reporting services for mobile apps.

- Data collection

To collect data, developers can include the software development kit (SDK) provided by Mobile Analytics into an app installation package. This SDK offers data collection components based on mobile operating systems. These components collect and send the data about mobile phones and user behavior to the backend of Mobile Analytics for analysis.

- Data reporting

The backend of Mobile Analytics offers a data reporting system, which allows you to collect the data reported by mobile phones using the specified SDK. The data reporting system preliminarily removes dirty data, and sends the processed data to DataHub.

 **Note** In the future, DataHub provides an SDK for mobile phones to directly report data. The removal of dirty data is performed in Realtime Compute instead of Mobile Analytics, reducing the host costs of Mobile Analytics.

- Stream processing

Realtime Compute continuously subscribes to the DataHub streaming data. It also continuously reads and runs computations over the data about the performance metrics of mobile apps. Realtime Compute then writes the result data of stream processing during each period to an online RDS or Table Store system.

- Data display

Mobile Analytics provides a complete set of performance metrics that allow you to quickly view the running information and usage of mobile apps. For example, you can quickly know user locations, visited pages, browsing duration, end devices and network environments, and slow responses or crashes. With Mobile Analytics, you can also analyze crashes by device, and view the details of crashes. The data display is based on the result data that is obtained in the stream processing phase.

26.9. Restrictions

None

26.10. Terms

Project

In Realtime Compute, a project is a basic unit for managing clusters, jobs, resources, and users. Project administrators can create projects, or add users to other existing projects. Realtime Compute projects can be collaboratively managed by Apsara Stack tenant accounts and RAM users.

Job

Similar to a MaxCompute or Hadoop job, a Realtime Compute job implements the computational logic of stream processing. A job is a basic unit for stream processing.

CU

In Realtime Compute, a compute unit (CU) defines the minimum capabilities of stream processing for a job with the specified CPU cores, memory, and input/output capacities. A Realtime Compute job can use one or more CUs.

Currently, a CU is assigned with one CPU core and 4 GB memory .

Flink SQL

Unlike most open source stream processing systems that provide basic APIs, Realtime Compute offers Flink SQL that includes standard SQL semantics and advanced semantics for stream processing. Flink SQL is designed to satisfy diversified business needs, and it allows developers to perform stream processing by using standard SQL. With Realtime Compute, even users with limited technological skills, such as data analysts, can quickly and easily process and analyze streaming data.

UDF

Realtime Compute allows you to use user-defined functions (UDFs) that are similar to Apache Hive UDFs. We recommend that you use UDFs to implement your custom computational logic. UDFs are a supplement to Flink SQL that can be used for standard stream processing. Currently, Realtime Compute only supports Java UDFs.

Resource

Currently, Realtime Compute only supports Java UDFs. A JAR file uploaded by a user is defined as a resource.

Data collection

During a typical data collection process, data is collected from sources and ingested into a big data processing engine. The data collection process of Realtime Compute focuses on the phases where data is collected from the source and then transferred into a data bus.

Data store

Realtime Compute is a lightweight computing engine without built-in data stores. Data sources and sinks of Realtime Compute are based on external data stores. For example, you can use RDS to store result tables for Realtime Compute.

Data development

During the data development process, you edit Flink SQL statements to create a Realtime Compute job. Realtime Compute offers an online integrated development environment (IDE) where you can edit SQL statements and debug data before publishing a Realtime Compute job.

Data administration

The data administration page of the Realtime Compute platform allows for online management of jobs. Realtime Compute helps you easily and effectively manage stream processing jobs.

27.DataQ - Smart Tag Service

27.1. What is DataQ - Smart Tag Service?

With the rapid development of Internet and big data technologies in recent years, various data products have emerged. The development of big data applications has the following challenges:

- To cope with the rapid growth of data volumes, various types of distributed data computing and storage technologies are developed to solve many difficulties across diverse application scenarios. In a non-traditional IT architecture, only a single database is required to support data analysis reports for the entire enterprise. The way for integrating and managing various types of data, merging various business databases, and managing the distribution of multiple computing and storage resources has become a major challenge.
- Big data is used in various industries, such as digital advertising, Internet finance, e-commerce, and online security and risk control. A data application includes report analysis, behavior prediction, real-time monitoring, credit scoring, personalized recommendations, text mining, and spatiotemporal data. It integrates various big data technologies rather than only generating report statistics for enterprise operations.
- Currently, the target data users are not limited to professional data analysts and data warehouse engineers, but also include the business personnel who have limited technical knowledge. This requires a system to help them perform data exploration in an easy and cost-effective way.

Therefore, if you want make good use of big data, you must consider the design of the enterprise IT infrastructure and enhance the comprehensive abilities of technical engineers. It is necessary to understand the characteristics of various types of distributed computing and storage resources. Design capable architectures for these resources for various application scenarios, such as data analysis and algorithm services is also required. The technical engineers also need to understand several different types of the data usage scenarios of business personnel and then help develop business-oriented data products.

Alibaba Cloud DataQ - Smart Tag Service provides a data IDE to accelerate the development and implementation of big data applications. This product helps developers integrate various big data products based on their business needs, which reduces most of the engineering workloads that are necessary for building big data applications. By using the product together with relevant industry application solutions, developers who are less experienced in development of big data applications can quickly build big data applications. This can help realize the true value of big data over a relatively short period of time.

DataQ - Smart Tag Service can help data developers to build models based on data tables. This helps extract the business data and convert the data into objects that can be understood by business personnel to accelerate application development. For example, doctors can understand patients, doctors, diseases, medical records, and other real objects, rather than multiple tables of unrelated facts and figures, such as a user table. DataQ - Smart Tag Service tags allows you to build business-oriented model and convert data into objects that can be understood in the medical industry. For example, if a patient is considered as an object, the object must be assigned with age, gender, blood type, pregnancy, and other tag information. Meanwhile, a link can be established between the patient and their past medical records. The object that has tags can be understood and analyzed by the doctor.

DataQ - Smart Tag Service helps you to build objects and provides actual content for applications to analyze. Meanwhile, application developers can intuitively understand the data objects and directly process, derive, and call the business-oriented objects and tags.

DataQ - Smart Tag Service provides the following benefits:

- Simplifies the integration with complex systems because application developers do not need to have a deep understanding of multiple underlying computing and storage resources.
- Helps the IT team to manage data usage by providing data service APIs, avoiding duplication and redundancy of resources.

The IT team can share tags that are frequently used in business scenarios. You can apply for using these shared tags. After obtaining authorization to use these tags, you can perform corresponding computations by calling APIs. You can also generate code that can be independently deployed by configuring parameters in the console. This helps provide an easy way to build the corresponding big data product.

27.2. Benefits

The next-generation and enterprise-level big data application platform

DataQ - Smart Tag Service aims to accelerate business data-driven processes and provides logical abstract models (object-link-tag model) for data management. It also provides basic data application engines such as profile analysis, marketing engine, rule warning, recommendation engine, and visualization.

Various application scenarios

- Analysis based on behaviors and other detailed data

The detailed data of various kinds of behaviors is analyzed in a free-style way, and the cross-relationships that are between dimensions exist in the analyzed content. Therefore, calculating such a diverse plethora of information in advance is a very difficult task.

- Extracting features from semi-structured data

Flexible analysis also requires the integration with leading edge practices that include prediction, scoring, text feature extraction, and other algorithm-based technologies. This helps you to perform extensive and in-depth analyses. Algorithms that support preference calculation and text mining can be used to help perform an in-depth analysis and extract user features from the semi-structured data.

- Interactive search and analysis

Analysis is often used to explore and exploit useful information derived from the data. You can adjust filtering conditions, dimension combinations, and drill-down aggregation until results are returned as expected. This process requires a quick response during querying.

An open platform

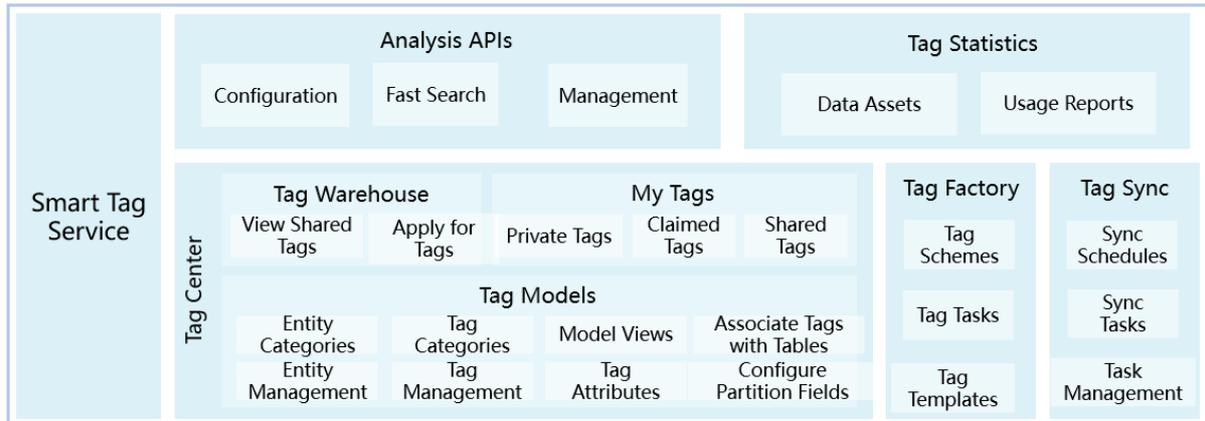
DataQ - Smart Tag Service not only provides UI interaction, but also provides a set of powerful APIs to integrate systems and reduce application development costs.

A web-based software

DataQ - Smart Tag Service can be deployed in both public network and internal network environments.

27.3. Architecture

DataQ - Smart Tag Service includes the following features: tag statistics on the homepage, tag center, analysis APIs, tag factory, dashboards, and tag sync.



27.4. Features

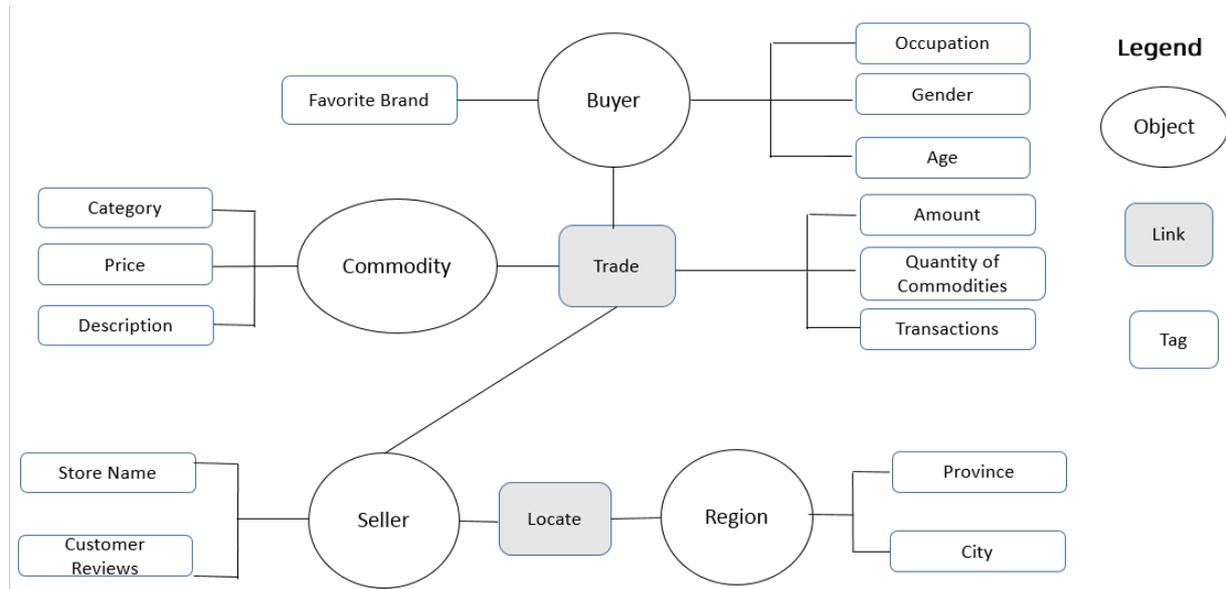
27.4.1. Tag center

The tag center is cross-computing storage that supports logical and dynamic modeling based on physical models (object-link-tag model). It integrates with data services to provide data modeling and data management tools for big data application and development. The data model view of an enterprise can be presented through visual methods. It is easy for business personnel, developers, and database administrators to gain a deeper insight into enterprise data assets.

The tag center is used to build a logic model across computing and storage resources based on existing data tables. This allows you to manage, process, and query data at the tag model layer without interacting with underlying big data computing and storage resources. The tag center is more important when the data architecture is complex and the combination of multiple computing and storage resources is required.

The tag modeling method is widely used in precision marketing, personalized recommendation, user profiling, credit scoring, and other big data applications based on detailed data computing. A tag is the minimum unit of description for a user object and represents an abstract expression of a specific objective fact of an object. The abstract expression, such as attributes, behaviors, and interests, is a data modeling method from the business perspective. For example, attributes include gender (the tag value is male or female) and age (the tag value is the actual age). Behaviors include turnover, bookmarks, and location. Interests include preference for multiple keywords. A tag can be a column consisting of values, enumerated values, and multiple key values, or a fact table consisting of multiple fields (such as subjects, time, predicates, and objects).

In terms of conceptual model, the tag system is a tag-based description methodology. This is built around multiple objects (such as buyer, seller, commodity, enterprise, and equipment) and the links between objects (such as transactions).



In traditional modeling, the concept and logic model are first designed according to business needs, and then the physical data tables are processed and sorted based on the logic model. In tag modeling, the logic model is directly built based on existing physical data or models. With the parsing of different data service agents, you can perform various computations on the model view without preprocessing a large amount of physical data.

Tags are created based on data of physical tables. In a cross-computing context, you may experience differences in query languages and performance between multiple computations. Therefore, tags created on logical requests may not be computed. In general, each tag you defined still needs to be associated with the corresponding physical table. However, in Smart Tag Service, you can define a computing logic of a query as a temporary tag in the corresponding data service. When a computing logic is related to cross-computing, it needs to be converted into data of physical tables to avoid errors.

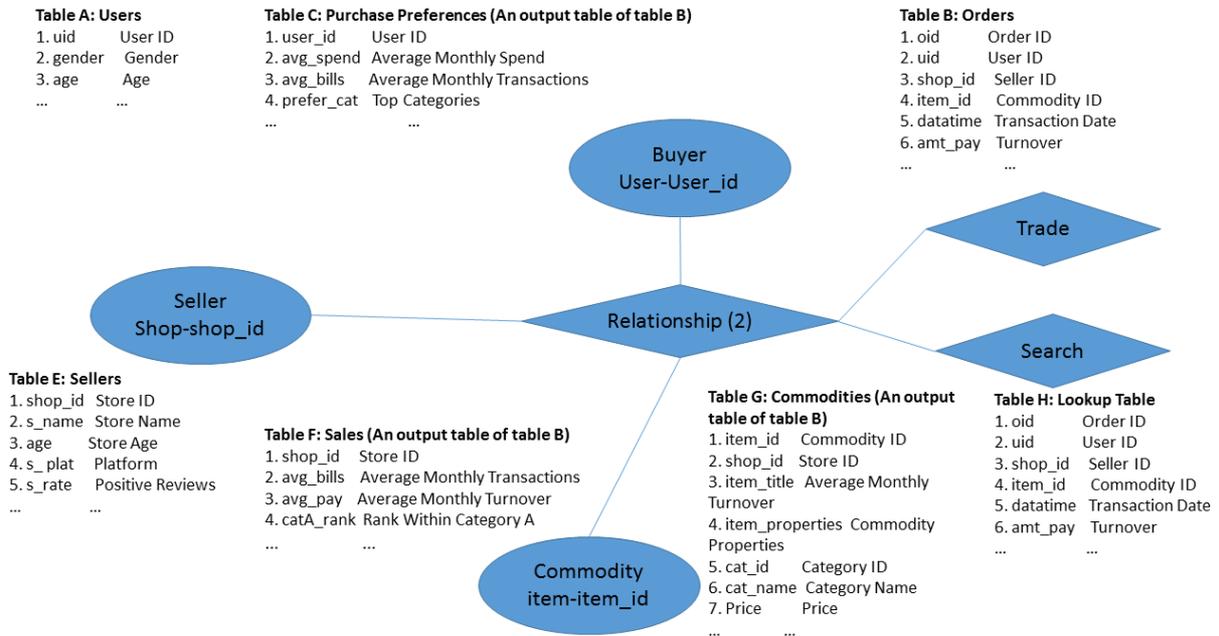
Tag models

The tag model is network-based modeling of data distributed in different databases based on three elements of an OLT model: object, link, and tag.

An object is used to describe a real object such as device, personnel, and address, which corresponds to physical data tables (usually property tables). In this table, the primary key represents the object and other columns are tags (namely, the properties of the described object).

Links are relationships, events, and actions between objects. They correspond to physical tables, which are usually fact tables. For example, a deal, repair, and ride.

Compared with the metric-dimension system, this modeling method is more suitable for the description and expression of detailed data. Most of the detailed data is a fact table. The concept of links corresponding to the fact table is introduced in a better way to show a clear representation of the relationships between multiple objects. This concept is conducive to management and expression during analysis. On the business side, it is also closer to the conceptual model design and easier to understand.



After modeling, you can convert the model in the preceding table to the logical relationships that are shown in the preceding figure. Transaction tables are mapped to the links, while the amount and time are the tags of the links. The user tables and commodity tables are mapped to buyer and commodity respectively, while gender and age are the tags of the buyer. This modeling method is ideal for scenarios where an analysis is performed based on detailed behaviors and relational data.

The Tag Center page consists of seven modules: tag warehouse, my tags, tag models, overview chart, model views, schemas, and data import.

Object-link model management is the main function of tag center to configure the logic model. It can read metadata from different database sources and integrate the metadata as an object or link. Multiple tables describing the same object (primary key) can be accumulated into a large wide table at the logic layer. The composite primary key tables can be considered as links during creation and used to associate multiple objects. Other descriptive fields are defined as tags as required.

My tags

On the My Tags page, it displays private tags, claimed tags, and shared tags.

As a department member, you can view, search, modify, and share private tags. You can also perform fast search, share multiple tags, revoke tag sharing, and detach private tags on the Private Tags tab page.

Tag warehouse

Tag warehouse stores shared tags. You can view and apply for shared tags in tag warehouse.

The tag warehouse has the following functions:

- **View shared tags:** You can view shared tags by workspace, filter tags by tag category, and search for shared tags by keywords.
- **Apply for tags:** select the required tags and apply for permissions to use the tags. You can apply for multiple tags at a time, and check the application status in the approval process.

The overview chart and model views

On the Overview page, you can view all objects, the relationships, and attributes between these objects, and tags attached to objects in a two-dimensional graphical manner. You can view and analyze the entire tag model through the overview chart.

When the number of business models is large, the relationships between business models are difficult to be analyzed in the overview chart. On the Model View page, you can drag and drop an object or a link from the search entities to create an intuitive model view. This model view is a sub chart of the overview chart and allows you to quickly find the required data in a complex model.

Schemas

Schema management supports communications between multiple computing and storage resources to obtain metadata.

Currently, DataQ - Smart Tag Service allows you to manage the following computing and storage resources.

- ApsaraDB for RDS
- MaxCompute
- AnalyticDB
- Table Store
- DataHub
- Realtime Compute

Data import

Currently, the tag center supports uploading files of the following types: TXT and CSV. You can import tasks into schemas (only MaxCompute schema is currently supported) to create tags.

27.4.2. Tag factory (DataQ - Smart Tag Service Advanced Edition)

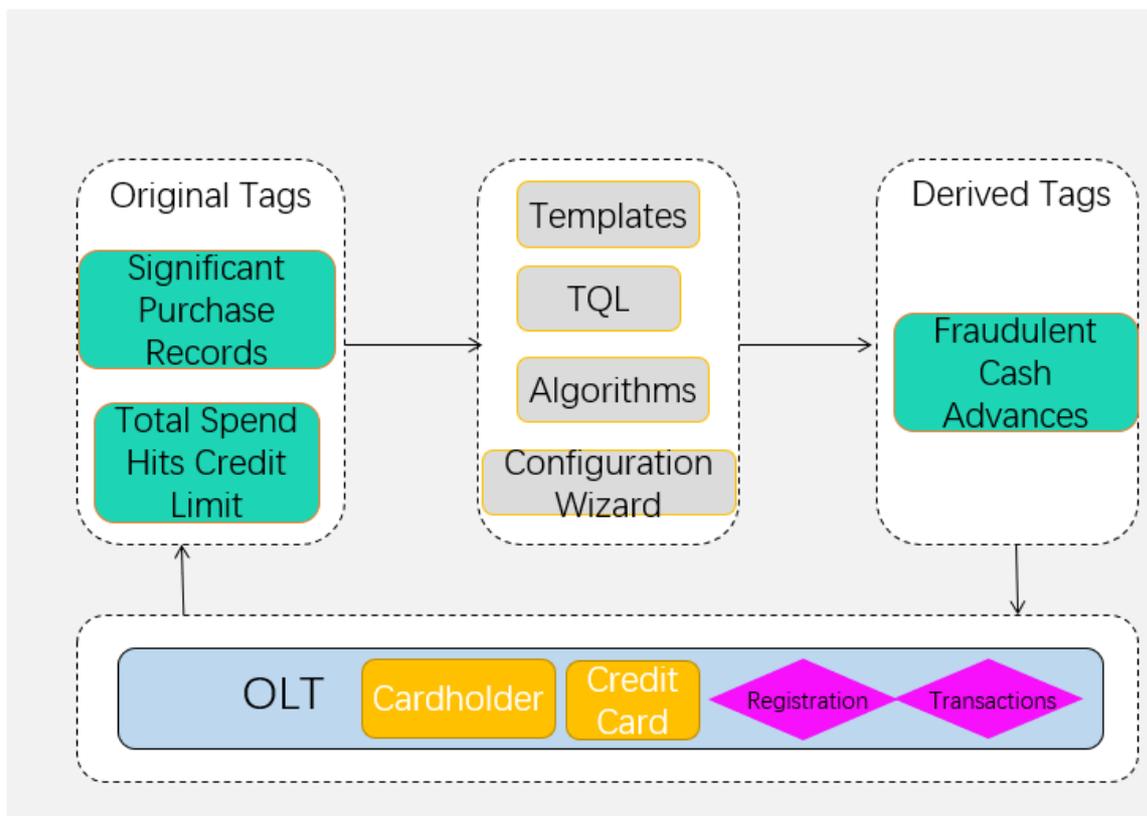
The tag factory module provides the processing of business tags, including the processing of tag schemes and tag tasks.

On the Tag Schemes page, you can create tag schemes by configuring TQL and algorithms. On the Tag Tasks page, you can run the tag tasks that are generated by the configured tag schemes.

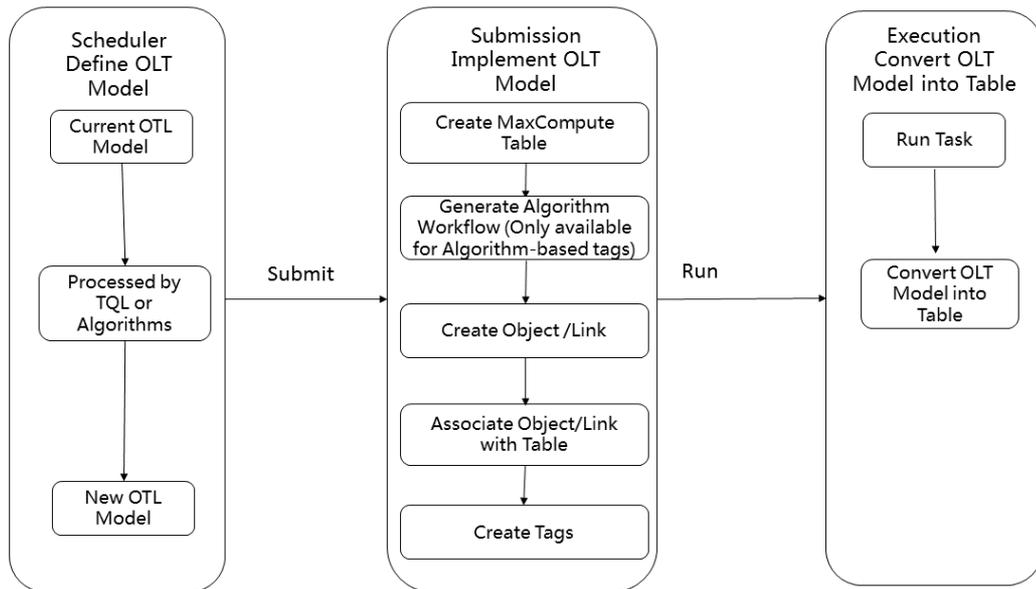
The tag factory is applicable to scenarios involving flexibly generated new tags and human resources of data processing that are insufficient. This reduces the requirements for developing derived tags and allows business personnel who have the ability of configuring simple TQL statements to configure derived tags as expected.

Based on existing tags, tag construction is used to carry out feature engineering of derived classes for common tags, or to extract structured tags from unstructured text data.

Feature engineering is an assistant tool to develop derived methods for existing tags. For example, when analyzing an individual's consumption behaviors, more information will be calculated based on the original transaction details of the individual. This information includes the monthly average consumption amount, category preferences, and purchase frequency. Alternatively, the combination of conditions that are used frequently in the market will be configured as tags for extra convenience, such as consumer groups with a high consumption in baby products. Feature engineering helps you to generate multiple tags at a time. This reduces redundant expressions when you use tags in application modules. In terms of resource utilization, configuring these frequently-used conditions as tags can reduce the pressure on online computing and save costs.



The workflow of the tag factory is shown in the following figure.



Tag schemes

A tag scheme is used to define the logic of derived tags, including the type, tag configuration, scheduling configuration, and parameter configuration of the tag scheme.

When creating derived tags, you must define the tag generation logic, the algorithms based on existing tags, the result fields corresponding to the new tags, and tag objects to be associated. If the result is a MaxCompute partitioned table, you must also configure which output field is used as the partitioning field.

The task includes the following two scheduling types: one-time schedule and recurring schedule. The tag generation logic supports TQL statements, common functions, and logical expressions.

Tag tasks

After the tag scheme is configured, you need to run it to generate a derived tag. On the Tag Tasks page, you can manage the tasks generated by task schemes and schedule these tasks. You can run a tag scheme after it is modified and submitted.

A task will be generated each time when you click Start to run a tag scheme. You can set execution parameters for a recurring task. After the task starts scheduling, you can view the task instances, running status, and logs for each scheduled task.

27.4.3. Tag statistic (DataQ - Smart Tag Service Advanced Edition)

As the middle layer for business-based data, tags will accumulate a large number of physical tables associated with tags and collect lots of logs. Tag statistic provides the capabilities of exploring and metering these tables and logs.

Tag statistic also provides the statistics of the following smart data assets:

- The links and total number of tags that are created in the object-link model on the platform.

- The number of analysis APIs configured and the number of times these APIs are called.
- The top tags and objects that are calculated by frequently used tags and their values.
- The derived tags that are mined by analyzing the tag query expressions.

27.4.4. Analysis APIs

The APIs are exported for special calls by various applications through the TQL capability provided by Analysis APIs. This is based on the behavior track and tag construction of relationships between objects and links, and is designed to help achieve rapid application development.

On the Analysis API page, you can create, view, debug APIs, and manage the API categories. API factory is a business function module built on a tag-based view. You can use tags as dimensions to perform unified computations for data across multiple computing resources by configuring APIs or calling API operations.

The combination of API factory and logic modeling reduces the workload and offers high scalability. Especially when the big data environment needs to consolidate data from multiple systems, it is difficult to design a single plan to meet all the data requirements. This dynamic logical modeling method has high scalability.

From the perspective of applications, the tag model allows you to calculate and query detailed data in the intuitive tag system without using the complex data structure. This also helps you to optimize the process of data development and application development.

Analysis APIs provides analysis query (fast search) and data service APIs.

- Analysis query (fast search)

Automatically synchronizes data across multiple data sources from tags to tables or indexes of AnalyticDB or relational databases. You can debug the analysis APIs on the Fast Search page. The entire analysis API is expressed by querying the TQL created on the tag model layer. Attributes related to the same object you are querying can be considered as a wide table, but actually the data may be distributed across multiple physical tables.

The API factory allows you to debug analytic statements and encapsulate the data analysis APIs. The query expressions of analysis APIs are built on an object-link model.

By debugging the API, you can view query results, runtime errors, time spent parsing syntax for each step, and parsed SQL statements.

- Data service APIs

You can directly generate APIs for analysis query. Application developers can analyze and query data through APIs. You can also manage the API categories, debug APIs, and publish APIs in the analysis API module.

The Analysis API page consists of three modules: API factory, API list, and fast search.

- API factory

The API Factory page allows you to query tags of business objects, shared tags, and object-link models using TQL. You can also view the list of query results in JSON format. By debugging the API on the API Factory page, you can check each process of TQL and its time performance.

- APIs

The APIs page allows you to query and analyze the generated APIs and configure parameters for analysis and query. It also allows you to manage APIs, view details, and test APIs. If you have deployed API Service Bus, it allows you to publish APIs to API store by one click.

- Fast search

The Fast Search page allows you to set filtering conditions and parameters to generate charts and data. You can also generate APIs based on the data of fast search and view the results on the APIs page.

27.4.5. Dashboards (DataQ - Smart Tag Service Advanced Edition)

The Dashboards page provides an intuitive data view and analysis reports for business personnel by displaying the tag query results. After a report is generated, it can be published to authorized users for viewing. This provides an easy way for authorized users to view and share data analysis reports.

The Dashboards page consists of three modules: datasets, report configurations, and report permissions.

- Datasets

Choose **Dashboards > Datasets** to open the corresponding page. You can view, test, unpublish, and delete the existing datasets.

- Report configurations

Choose **Dashboards > Report Configurations** to open the corresponding page, you can not only create a group and report but also edit or unpublish an existing report.

- When a report is saved but not published, move the pointer over the report and the Edit button appears. You can click Edit to open the editing page and modify the report.
- When a report has been published, move the pointer over the report and the View button appears. You can click View to open the report, and click **Modify** to unpublish the report and modify it offline.

- Report permissions

On the Role Management tab page, you can create roles for users and set different role permissions. On the User Authorization tab page, you can modify the user roles.

27.4.6. Tag sync

Tag sync is one of the most important functions of processing cross-computing data flow for DataQ - Smart Tag Service. When data is required by the corresponding data service, the tag center can collect the data distributed across multiple storage systems. The data is then subscribed to the location where the data service needs to compute.

In scenarios that require quick response during synchronization, you need to first subscribe to the API factory.

Tag sync includes sync schedules, sync tasks, and task O&M.

- Sync schedules

When you plan to synchronize data from a source schema to a target schema, you need to select a scheduling mode.

- Sync tasks

After a sync schedule is started, a sync task is generated. You can view the task status, scheduling mode, and operations.

- Task O&M

Select a sync schedule, and you can view the detailed logs of the sync schedules instance based on the scheduling mode.

27.5. Scenarios

27.5.1. Overview

DataQ - Smart Tag Service is applicable to the following scenarios:

- The total amount of structured data is greater than 200 GB.
- Multiple business systems or small data warehouses need to be integrated.
- The analysis scenario is based on a single object (such as a person, equipment, and vehicle).
- Number of data dimensions (such as derived dimensions) is greater than 10.
- Business personnel want to analyze data by themselves rather than just viewing reports.

27.5.2. Analysis APIs

Integrates user favorites, turnover, clicks, registering information, locations, derived tags, and other data to perform a comprehensive analysis of the correlations between various user behaviors. This allows you to design more effective cross-selling, marketing content, target groups, and other strategies.

27.5.3. Device records analysis

DataQ collects all device data throughout the lifecycle of a device, including purchase, maintenance, servicing, scrapping, and technical transformation. This data helps analyze device assets and study the effects of various types of external data on device conditions. This improves the efficiency of device assets management.

27.5.4. Geographic analysis

As this service integrates geographical information, you do not need to first calculate the distributions for each geographic grid. You can apply a variety of filters and perform summary analysis with the detailed points of interest (POI) data. This increases the flexibility of spatiotemporal data analysis.

27.6. Limits

None

27.7. Terms

Object

Objects are the subjects of analysis and the abstractions of physical things in the real world, such as citizens, trains, hotels, Internet cafes, and cases involving security scenarios.

Link

Links are the relationships between two or more objects. For example, in a security scenario, citizens have many activities, such as travel (by train), hotel check-in, and Internet access in an Internet cafe. Links are the relationships between these citizens, activities, and the criminal case.

Tag

Tags are attached to an object or a link.

- The basic tags of a citizen such as name, age, and nationality belong to the citizen object.
- The travel time, departure station, and destination of a train are the tags that are attached to the link of a citizen's travel.
- The check-in time and the length of stay in a hotel are the tags that are attached to the accommodation link between the hotel and citizens.

TQL

Tag Query Language (TQL) is a query language supported by Analysis API. The query syntax is based on an object-link-tag (OLT) model. TQL syntax is similar to SQL syntax, which supports SELECT statements. However, TQL does not support UPDATE, INSERT, or DELETE statements.

JQL

JSON Query Language (JQL) is a query language supported by Analysis API. The syntax is in JSON format and based on an object-link-tag (OLT) model. The JQL syntax has the same semantics as the TQL syntax.

28.E-MapReduce (EMR)

28.1. What is EMR?

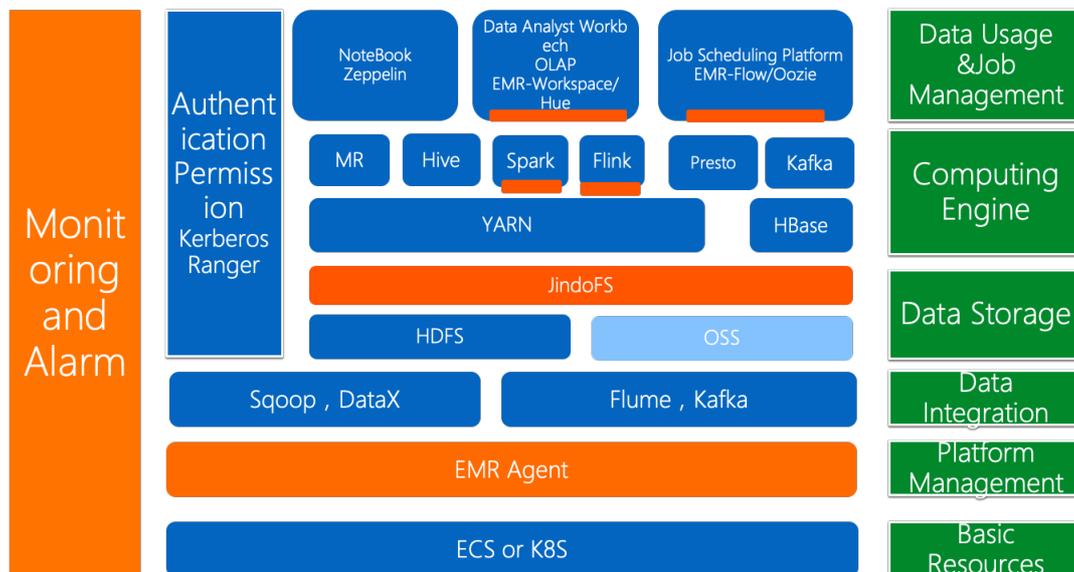
Elastic- MapReduce (EMR) is a managed cluster platform that simplifies running big data frameworks, such as Hadoop, Spark, Kafka, and Storm. EMR provides you with one-stop big data processing and analysis services to manage clusters, jobs, and data.

EMR is a service that is based on ECS and uses open-source Apache Hadoop and Spark to process and analyze vast amounts of data. You can use components in the Hadoop and Spark ecosystems, such as Apache Hive, Apache Pig, and HBase to process and analyze data. You can also use EMR to import and export data from Alibaba Cloud data stores and databases, such as OSS and ApsaraDB for RDS.

28.2. Architecture

EMR architecture shows the architecture of EMR.

EMR architecture



Dependent on the Hadoop ecosystem, EMR clusters are created based on Alibaba Cloud ECS instances. Clusters work seamlessly with cloud services, such as OSS and ApsaraDB for RDS to exchange data. This facilitates data transit and sharing between services to meet specific business requirements. EMR provides diverse API operations for you to perform actions on clusters, jobs, and execution plans.

For more information about components in the Hadoop ecosystem, see *Terms of Product Introduction*.

28.3. Benefits

Compared with user-created clusters, EMR provides you with easy and well-organized methods to manage your clusters. EMR also offers the following benefits:

- Deep integration

EMR works seamlessly with other Alibaba Cloud services, such as Object Storage Service (OSS), Message Service (MNS), Apsara for RDS, and MaxCompute. This allows data of these services to be used as the input or output of the Hadoop or Spark services of EMR.

- Security

With Resource Access Management (RAM), RAM user accounts are authorized to access different EMR resources.

28.4. Features

EMR provides the following features:

- Supports a variety of jobs, such as Spark, Hadoop, Hive, Pig, Sqoop, Spark SQL, and Shell. You can use these jobs to meet specific requirements, such as log analysis, data warehousing, business intelligence, machine learning, and scientific simulation.

After selecting a job type, you can specify the required commands and actions to follow after a job failure. You can also copy, modify, and delete a job.

- Supports creating flexible execution plans.

An execution plan includes a set of jobs. You can run an execution plan on an existing cluster or create a temporary cluster to run an execution plan. With scheduling policies, you can run a job at one time or on a regular basis. The major benefit of an execution plan is that you can use as many resources as possible to maximize resource utilization. A flexible execution plan has the following benefits:

- An execution plan can include different types of jobs in any combination. These types of jobs include Hadoop, Spark, Hive, and Pig.
- You can manually run an execution plan or create a schedule to periodically run an execution plan.

- Provides an interactive workbench

The interactive workbench allows you to write and run Spark, Spark SQL, and Hive SQL tasks in the EMR console. After a task is complete, you can view the results in the workbench. You can use the workbench to process different types of tasks, such as short-term, real-time results-oriented, and debugging tasks. We recommend that you use jobs and execution plans to process long-term scheduled tasks.

- Supports alerts.

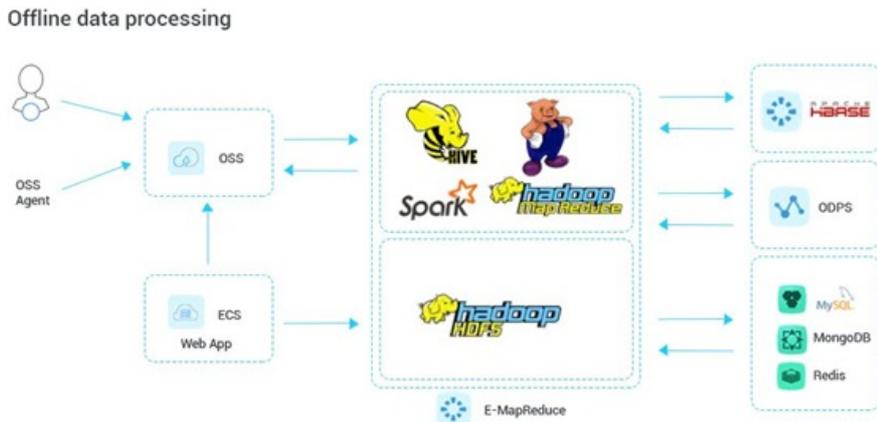
You can associate execution plans with alert groups by using EMR. After you configure the Alert Notification setting on the Execution Plan page, contacts included in the specified alert contact group will receive SMS alerts after each execution plan is complete. An SMS message includes the name of an execution plan, cluster name, duration, status of a job, number of successful tasks, and number of failed tasks.

28.5. Scenarios

- Batch data processing

You can replicate a large number of logs from application servers, such as games, web applications, and mobile apps to EMR nodes. Then, you can use mainstream compute frameworks, such as Hive, Spark, and Presto to quickly retrieve data and obtain forecasts with tools such as Hue. You can also use tools such as Sqoop to retrieve data from multiple distributed RDS instances or other data stores. After analyzing data that is retrieved you can replicate refined data to RDS instances as the data source for data visualization services.

Batch data processing

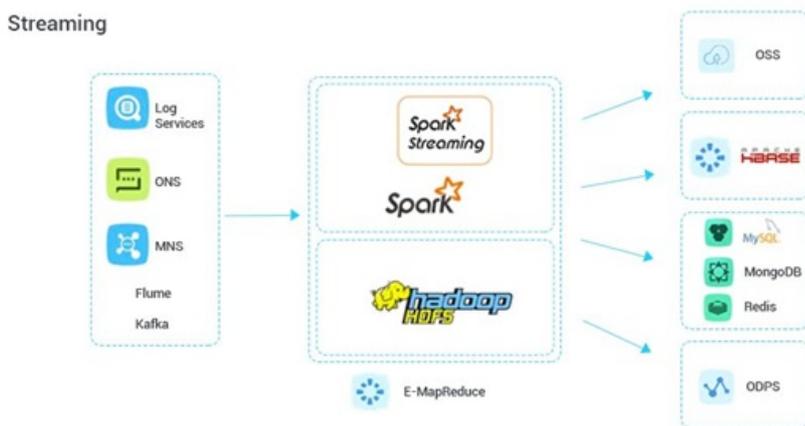


- **Streaming data processing**

With Spark Streaming and Storm jobs, you can use and process real-time data from data streams, such as Log Service, Message Queue (ONS), Message Service (MNS), and Apache Kafka.

EMR processes streaming data in a fault-tolerant manner. After data processing is complete, EMR writes results to Object Storage Service (OSS) or HDFS.

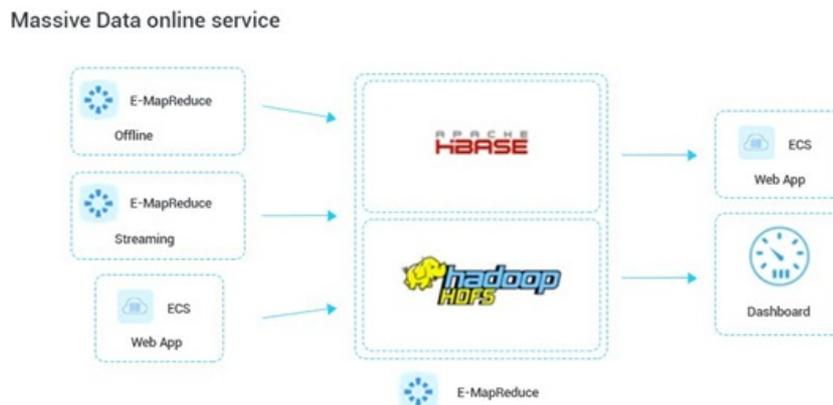
Streaming data processing



- **Massive cloud data processing**

EMR analyzes petabytes of structured, semi-structured, and unstructured data created by your Web and mobile apps. This allows Web applications or data visualization services to visualize data in real time based on analytic results from EMR.

Massive cloud data processing



28.6. Limits

None.

28.7. Terms

job

Similar to MaxCompute or Hadoop jobs, an E-MapReduce (EMR) job is the basic unit to process and analyze big data.

Hadoop

- YARN

EMR uses YARN to manage cluster resources and schedule jobs.

- HDFS

Hadoop Distributed File System (HDFS) is a distributed file system for data storage.

Hive

Hive is a Hadoop-based open-source data warehouse software that provides an SQL-like interface for data processing and analysis. Hive uses tables to store and manage data.

Spark

Spark is a memory-based distributed computing framework that supports batch and real-time computing, SQL statements, and machine learning.

Hue

Hue is an open-source user interface for visualizing data. Hue supports multiple components, such as Hadoop, Hive, Oozie, and HBase.

Oozie

Oozie is a job scheduler that supports workflow orchestration by building a directed acyclic graph (DAG). Oozie supports multiple types of jobs.

Presto

Presto is a distributed SQL query engine for retrieving large datasets from one or more data sources.

Zeppelin

Zeppelin is a Web-based notebook that enables interactive data analytics and collaborative documents with SQL and Scala.

ZooKeeper

ZooKeeper is an open-source and distributed service for coordinating applications. ZooKeeper is a close clone of Google Chubby and an important component of Hadoop and HBase. ZooKeeper is a centralized service that provides consistent services for distributed applications. These services include configuration maintenance, naming, distributed synchronization, and group services.

Sqoop

Sqoop is a tool designed for migrating data between HDFS and relational databases.

Kafka

Kafka is a high-throughput messaging system with a variety of features, such as high-throughput, scalability, high reliability, and high performance. Example applications of Kafka include real-time compute, log processing, and data aggregation.

HBase

HBase is an open-source, distributed, and column-oriented data store. HBase is a component of the Apache Hadoop project. Different from typical relational databases, HBase is a data store that is designed to store unstructured data. HBase is a column-oriented rather than row-oriented data store.

Phoenix

Phoenix provides SQL-like statements that allow you to perform data analysis on HBase data.

MetaService

With MetaService, you can access Alibaba Cloud resources in EMR clusters without using an AccessKey pair.

Metastore

Metastore is the central repository of Hive metadata. You can use a metastore to organize, store, and manage data by using data schema.

Kerberos authentication

Kerberos is a third-party authentication protocol that is designed for TCP/IP networks. Kerberos uses symmetric cryptography based on the data encryption standard (DES).

29. Quick BI

29.1. What is Quick BI?

Quick BI is a flexible and lightweight self-service BI platform based on cloud computing.

Quick BI supports various data sources, such as MaxCompute (formerly known as ODPS) and AnalyticDB for PostgreSQL. Quick BI can connect to user-created MySQL databases that are hosted on ECS. It can also connect to data sources in VPC networks. Quick BI provides real-time online analysis for a large amount of data. It significantly reduces data retrieval costs and is easy to use with the support of intelligent data modeling tools. Drag-and-drop operations and various visual charts allow you to create reports and use pivot charts and tables, downloads, data exploration, and BI portals to easily analyze data.

Quick BI enables everyone to be both a data viewer and a data analyst to achieve data-based operation of enterprises.

29.2. Benefits

Benefits of Quick BI can be summarized as follows:

High compatibility

Supports various data sources, such as MaxCompute and AnalyticDB for PostgreSQL.

Fast response

Responds in seconds to hundreds of millions of data queries.

Powerful capabilities

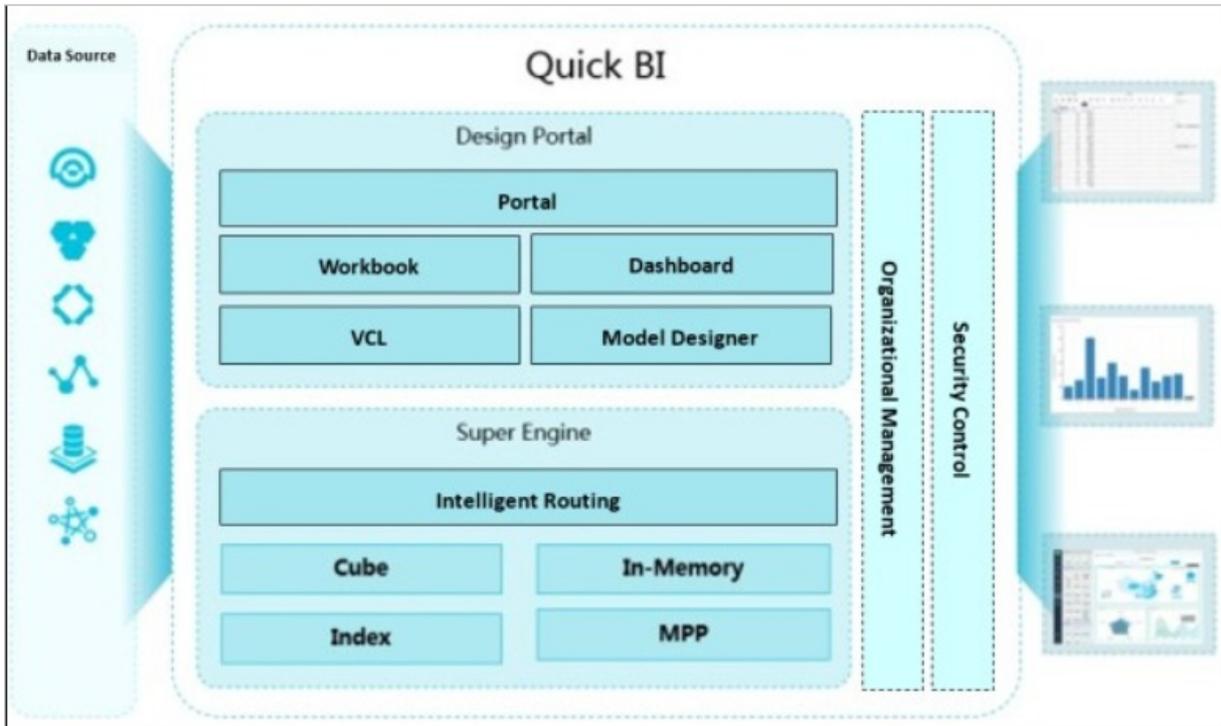
Allows users to easily create complex reports by using workbooks.

User-friendliness

Provides various data visualization functions and automatically identifies data properties to generate the most appropriate charts for users.

29.3. Architecture

The following figure shows the architecture of Quick BI.



Modules and features of Quick BI:

- **Data connection module**

Connects to various data sources, such as MaxCompute and AnalyticDB for PostgreSQL. This module provides APIs to query metadata or data from data sources.

- **Data pre-processing module**

Provides lightweight ETL processing for data sources. Currently, Quick BI supports custom SQL of MaxCompute. Quick BI will support data pre-processing for more data sources in the future.

- **Data modeling**

Takes charge of OLAP modeling of data sources and transforms data sources into multi-dimensional analysis models. It supports standard semantics such as dimensions (such as Date type dimensions and Geo type dimensions), measures, and galaxy schemas. It also supports calculated fields, and allows you to process dimensions and measures by using SQL syntax for existing data sources.

- **Workbooks and classic workbooks**

Provides workbook functions. This module enables data analysis, such as row and column filtering, standard and advanced filtering, subtotal and total calculation, and conditional formatting. It also supports data export, text processing, sheet processing, and other operations.

- **Dashboards**

Assembles visual charts into dashboards. The various types of charts supported by dashboards include line chart, pie chart, vertical bar chart, funnel chart, hierarchy chart, bubble map, colored map, Kanban, and others. Moreover, it supports inter-chart field dependency and five basic widgets including filter bar, tab, iFrame, picture, and text box.

- **BI portals**

Allows you to display internal content and external content. Internal content includes dashboards, workbooks, forms, and downloads. External content refers to the content of external link pages.

- **Query engine**

Queries data that is stored in data sources.

- **Permission management in organizations**

Sets permissions based on organizations, workspaces, and workspace-specific user roles. Quick BI allows you to grant different permissions for a same report to your members as needed.

- **Row-level permission management**

Controls row-level permissions of data. Employees can only view part of a comprehensive report based on the permissions they are granted.

- **Share and publish**

Shares workbooks, classic workbooks, dashboards, and BI portals with other organization members. You can also publish dashboards to the Internet to share with the public.

29.4. Features

Quick BI provides the following features:

Seamless integration with cloud databases

Supports various Alibaba Cloud data sources, such as MaxCompute and AnalyticDB for PostgreSQL.

Charts

Provides diverse options for data visualization. To meet data presentation needs in different scenarios, Quick BI supports various built-in visual charts, such as vertical bar charts, line charts, pie charts, radar charts, and scatter charts. Quick BI automatically identifies data properties and intelligently recommends appropriate visualization solutions.

Analysis

Enables multi-dimensional data analysis. Quick BI is a web-based data analysis system that provides simple data import methods. It supports drag-and-drop operations in workbooks and supports real-time analysis. You can use Quick BI to analyze data from different perspectives without repetitive modeling.

Quick building of data portals

Provides drag-and-drop operations, powerful data modeling capabilities, and multiple visual charts to help you build BI portals in a short time.

Real-time analysis

Supports online analysis for a large amount of data without the need of pre-processing, significantly improving the efficiency of data analysis.

Data permissions

Supports member management and row-level data permission control. This enables users of different roles to view different reports and to view different data of the same report.

29.5. Scenarios

29.5.1. Instant data analysis and effective decision-making

Business goals:

- A convenient method to retrieve data

Quick BI eliminates the reliance on IT professionals to write SQL statements for multidimensional data analysis.

- An easier report making process

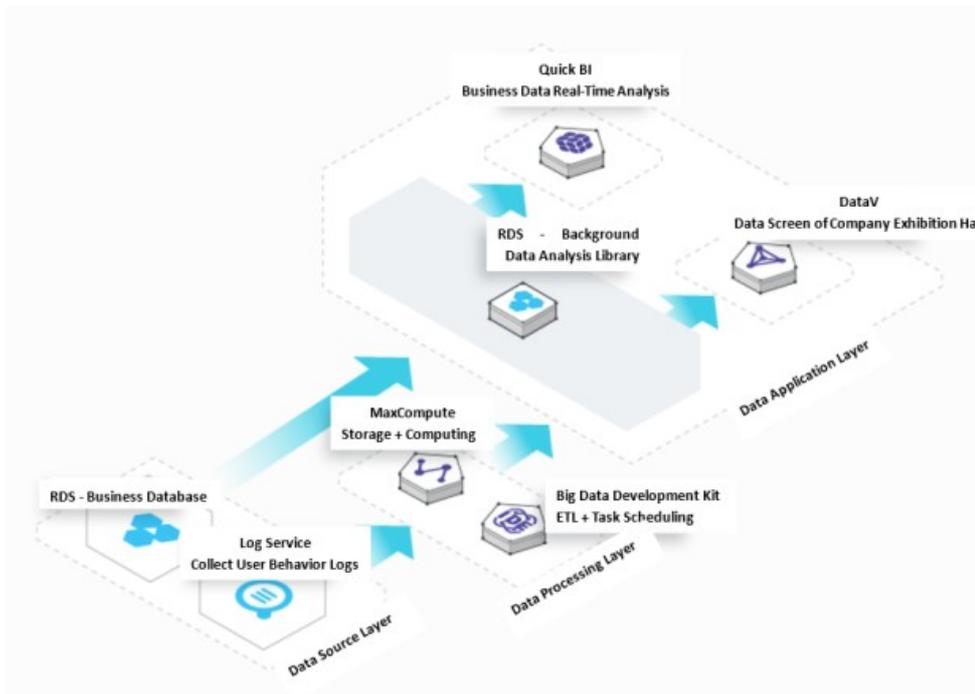
Quick BI simplifies and shortens the process of delivering updates and new code to an analytics system.

- A reduction of human resource costs

Quick BI provides an easy-to-use user interface and reduces your maintenance costs.

Recommended combination: relational database and Quick BI

Instant data analysis and effective decision-making



29.5.2. Integration with existing systems

Business goals:

- Easy adoption

Quick BI is a user-friendly and easy-to-use service for users from different backgrounds, which satisfies data analysis needs of personnel in various departments.

- High efficiency for data visualization

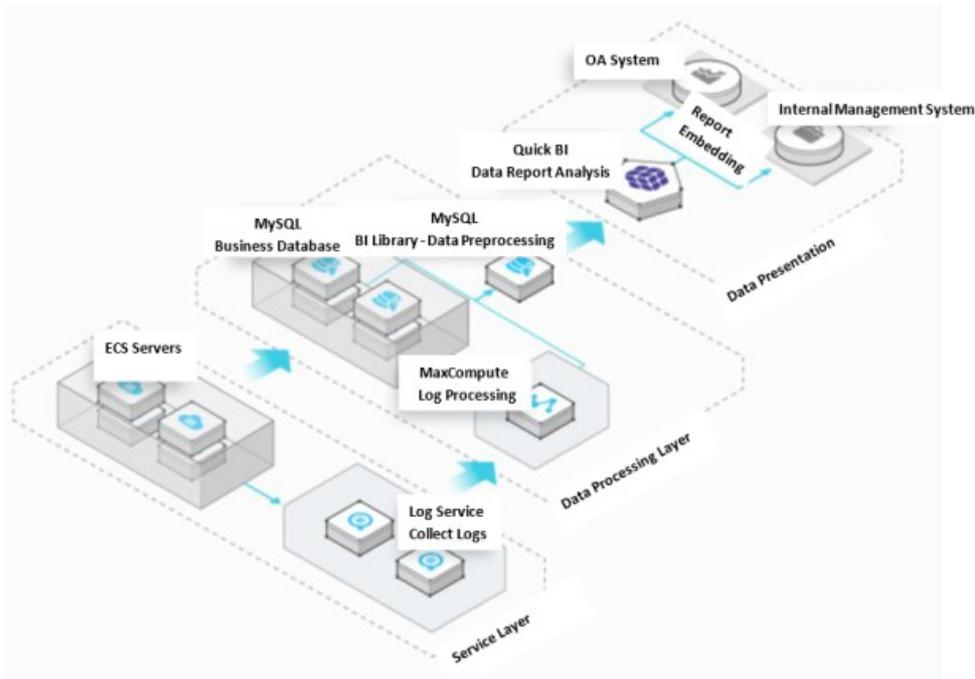
Integration with existing systems allows for quick data analysis and improves the efficiency of viewing data.

- Unified management platform

You can access and manage data by using a unified platform instead of multiple systems.

Recommended combination: relational database and Quick BI

Integration with existing systems



29.5.3. Permission control of transaction data

Business goals:

- Row-level permission control

You can easily create a comprehensive report for all members, but one member can only view data related to their marketplace.

- Dynamic business requirements

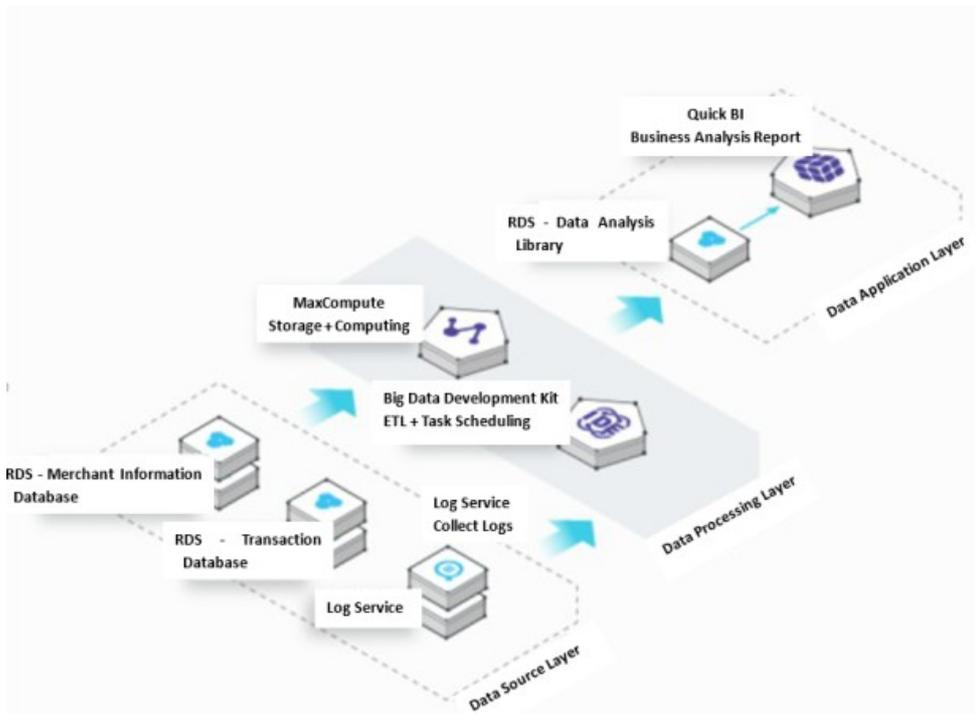
Quick BI responds quickly to a large number of changes in statistical indicators as business grows.

- Consistent computing performance across multiple data sources

Quick BI provides powerful cloud infrastructure and data analysis platforms, and enables data analysis across data sources without compromising performance.

Recommended combination: Log Service, relational database, Quick BI, and MaxCompute

Permission control of transaction data



29.6. Limits

None.

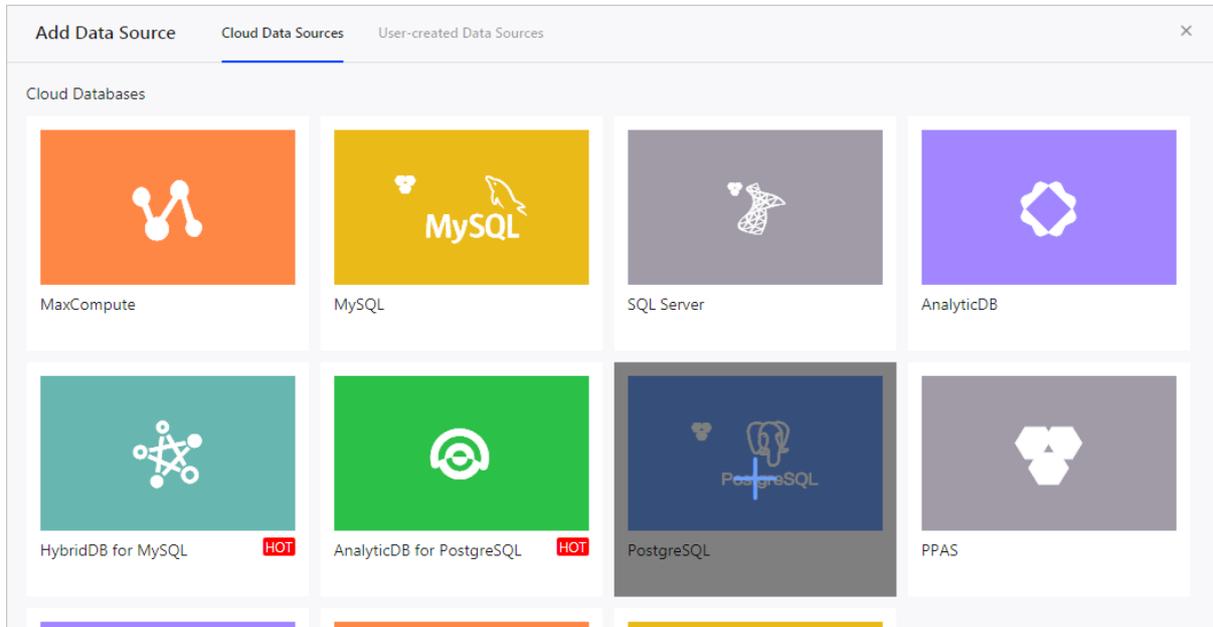
29.7. Terms

This topic describes the commonly used terms and concepts in Quick BI.

data source

When you use Quick BI for data analysis, you must first specify the data source of your raw data. A data source is where data is stored. You can add data sources by using either of the following methods:

- Add data sources from cloud databases
- Add data sources from user-created databases



dataset

You can use tables in data sources to create datasets. You can edit, move, or delete a dataset from the dataset list.

dashboard

Dashboards employ a flexible tile layout to allow you to create interactive reports. Moreover, dashboards support data filtering and data query functions, and adopt multiple charts to display data.

You can drag and drop fields or double-click them to add the data of the fields to charts in dashboards and view the data more clearly. Dashboards provide a user-friendly and efficient experience when you perform data analysis tasks.

workbook

Workbooks display analyzed and processed data in a dataset. You can use workbooks in both personal and group workspaces. To analyze data in a workbook, you can select the dataset where the data is located and perform required operations.

BI portal

A BI portal is a set of dashboards organized in the form of menus. You can build data analysis systems by using BI portals, such as a business analysis system. A BI portal references analyzed data from Quick BI and supports external links.

30. Graph Analytics

30.1. What is Graph Analytics

Graph Analytics is a visual analysis platform for relationship networks. Graph Analytics is widely used in Alibaba Group and Ant Financial for risk control including anti-fraud, anti-theft, and anti-money laundering solutions. Graph Analytics provides solutions for multiple industries, including public security protection, taxation, customs, banking, insurance, and the Internet.

Graph Analytics is designed to facilitate multi-source data integration, computing applications, visual analytics, and intelligent businesses. Based on relationship networks, Graph Analytics can visualize the properties of objects and reveal the relationship among objects.

Graph Analytics provides multiple features, including relationship networks, search networks, information cubes, intelligent judgment, collaboration and sharing, and dynamic modeling. Graph Analytics integrates machine computing capabilities with human cognition to provide users with data insight and help users obtain information and knowledge more efficiently.

30.2. Benefits

Performs massive data mining in real time

Graph Analytics can perform relationship mining and computing on petabytes of data. Graph Analytics can handle tens of billions of nodes, hundreds of billions of links, and trillions of records based on real-time commands.

Understands the connectivity of things using the OLEP model

Graph Analytics understands the connectivity of things using the OLEP model. The OLEP model studies objects, links, and events, and integrates heterogeneous data based on data properties.

Flexible service scenarios

Based on the OLEP model, Graph Analytics provides suitable business configurations and detection features to enable human-machine interaction. Applicable scenarios include public security protection, anti-fraud solutions, financing, and taxing.

Efficient visual analyses

Graph Analytics works on key issues to be improved in user experience and pain points in the data analysis business. Based on the analysis results, Graph Analytics provides iterative, visual analysis and collaborative analysis services for users to build traceable links and paths among objects.

User-friendly intelligence

Graph Analytics helps business users analyze, scrutinize, and handle challenges with ease. Graph Analytics provides deep training models, including the intimacy degree model, terror degree model, and the drug involvement model.

Robust analysis systems

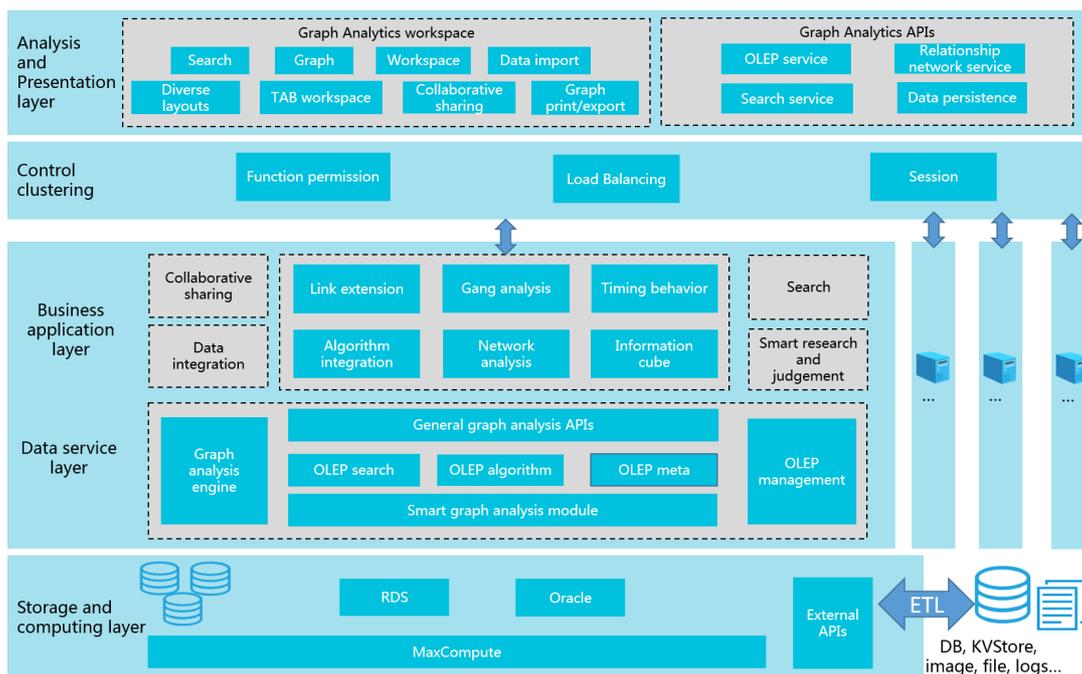
Tested by multiple key national projects, Graph Analytics is considered an important product and has impressed customers with its application in public security protection, anti-terror, and tariff services.

30.3. Product architecture

30.3.1. System architecture

Graph Analytics provides built-in components, user-friendly designs, and a multi-layer structure for you to extract insight from data with ease. The data storage and compute platform of Graph Analytics is built on DTplus Platform, a basic big data service platform developed by Alibaba Cloud. Featuring powerful capabilities in data integration, data processing, data analysis, and computing, this data storage and compute platform can process and store up to several exabytes of data.

System architecture



The entire system is divided into four layers: the storage and compute layer, the data service layer, the application layer, and the analysis and presentation layer.

- **Storage and compute layer**

Based on the Alibaba Cloud big data platform, Graph Analytics supports multiple open data sources. The compute platform consists of an offline platform and several online platforms. The offline compute platform refers to MaxCompute, a platform that supports data integration and processing. The online compute platforms, including StreamCompute, can support computing in real time.

- **Data service layer**

Graph Analytics extracts the object-property-link model from relationship categories, relationship types, and relationship items to model the business logic. Based on these models, Graph Analytics integrates data from different categories and resources, and provides flexible management and maintenance of logical models. The data service engine provides a standard query language of the business logic for the application layer and performs complex network queries and analyses using algorithms.

- **Business application layer**

Graph Analytics provides business APIs to call applications at the analysis layer. These applications include relationship networks, search networks, information cubes, intelligent judgement, collaboration and sharing, and dynamic modeling.

- Analysis and presentation layer

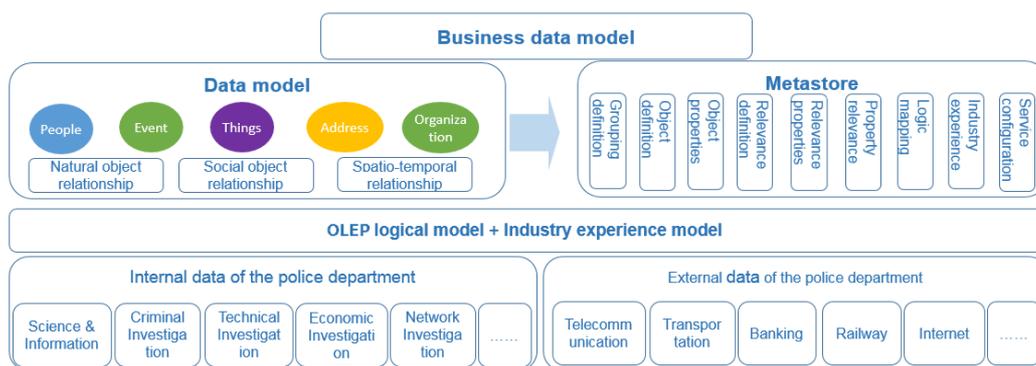
Graph Analytics provides a diversified, visual, and interactive analysis interface and supports various terminals. Graph Analytics provides visual components and external APIs and supports third-party system integration.

30.3.2. OLEP model

OLEP model analyzes objects, links, real-world events, and integrates heterogeneous data based on its properties. As the foundation of Graph Analytics, the OLEP model is the key to connecting correlated links and objects and building elaborate relationship networks.

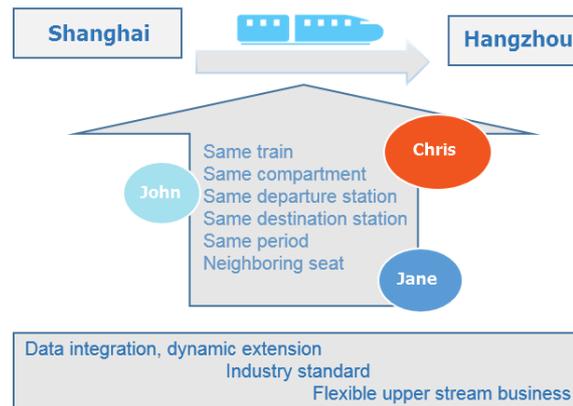
For example, the public security industry uses both data from within the security industry and external security data. In this scenario, Graph Analytics can leverage the physical data to build OLEP models and industry models, and map elements in these models to metadata definitions, including object definitions, object properties, link definitions, and link properties, as shown in [OLEP model](#).

OLEP model



Take high-speed rail as an example: Three people, John, Jane, and Chris are taking a high-speed train from Shanghai to Hangzhou. Using the OLEP model, you can analyze the travel data and determine whether they are on the same train or in the same carriage. You can also tell whether they are from the same source station or heading to the same destination, as shown in [High-speed rail as an example](#).

High-speed rail as an example



30.4. Features

30.4.1. Features

Graph Analytics includes the following two feature modules.

Search

Search provides the features of object information retrieval. It helps user to fast locate the information, and play a fundamental role for relationship network and map analysis. Search introduces the retrieval of the object information into the relationship network and map analysis to further analyze.

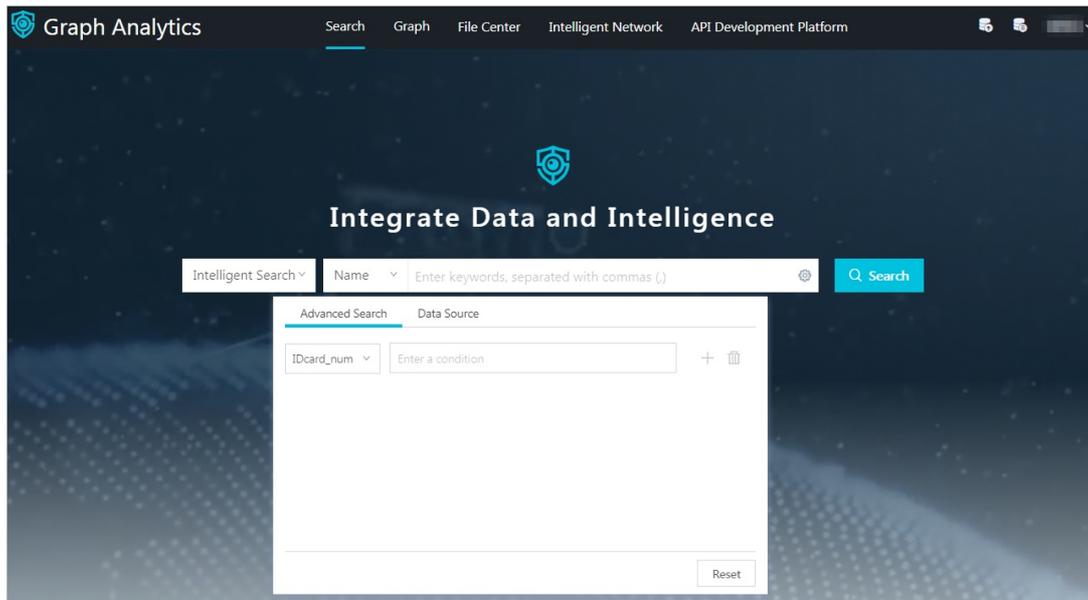
Relationship network

Relationship network is the core module of the I+. Relationship network consists of the relationship topology of all objects, business computing, visualization layout, and the interactive operation. Relationship network has the ability of relationship investigation, grouping analysis, neighborhood analysis, route analysis, kinship analysis, behavior analysis, data cubes, and so forth.

30.4.2. Search module

As one of the two independent modules of Graph Analytics, the search model can help analysts quickly locate and view specific objects, such as mobile phones and ID cards. Meanwhile, the searched objects can be viewed as independent object nodes and added to the relationship analysis or GIS analysis as the starting point.

Search module



30.4.3. Relationship networks

Graph Analytics provides multiple analysis methods for you to easily obtain useful intelligence from complex networks. The features of Graph Analytics include link lookup, group analysis, common neighbor analysis, backbone analysis, lineage analysis, information cube, group statistics, and label statistics.

Link extension

An infinitely extended analysis that begins with any single object or a group of objects. Link lookup helps to build infinite information associations. The key to intelligence analysis is to discover related clues and intelligence from a large amount of unrelated information and convert the information into useful and actionable intelligence. Graph Analytics provides simple link lookup services and advanced link lookup services.

Group analysis

Analyzes the direct and indirect relationships between a group of objects of the same type or of different types.

Common neighbor analysis

Analyzes the objects that are commonly associated with two groups of objects, including groups of objects of the same type or of different types.

Path analysis

Analyzes the link path between two objects.

Backbone analysis

Locates the core backbone nodes in a group network using smart algorithms.

Lineage analysis

Displays the lineage relationship among people based on families (family IDs).

Information cube

- Behavior analysis
Displays the frequency of an event in a chronological order.
- Chronology analysis
Displays the details of each event in a chronological order.
- Behavior details
Displays the details of events. The original data records are filtered according to specific rules.
- Object information
Aggregates objects in a relationship network and classifies the objects by type.
- Statistics information
Analyzes the relationships and objects in a relationship network, including object properties, link properties, and the distribution of objects.

Group statistics

Analyzes the distribution of groups in a network. A group consists of multiple object nodes, with any two object nodes connected topologically. Nodes within a merged node are connected topologically.

Label statistics

Collates the label information of object nodes in a relationship network. Graph Analytics supports two types of labels: system labels and user labels. System labels, such as whitelists and blacklists, are defined by the service system for specific nodes. User labels are added to specific nodes by users on the Graph Analytics platform.

Graph layouts

Supports matrix layouts, ring layouts, horizontal layouts, vertical layouts, force-directed layouts, and hierarchical layouts.

Right-click operations

The information in the graph area includes objects, relationship details, and images mapped in a network structure. The key elements of Graph Analytics are nodes and links, and the analysis in Graph Analytics is based on nodes and links. You can use the right-click shortcut menu to edit and analyze nodes and relationships in Graph Analytics.

Directory management

You can use this feature to manage and operate directories and analyses. Supported features include:

- Create analyses and delete, rename, and create directories.
- Open, delete, and rename analyses.
- Locate specific content using the search function.
- Share personal analyses to others for further analysis.

Collaboration and sharing

Collaboration and sharing is a new analysis model provided by Graph Analytics. You can use this model to share your personal analyses with other users, pass on your ideas and experience to them, and integrate others' experience and discoveries to achieve team collaboration.

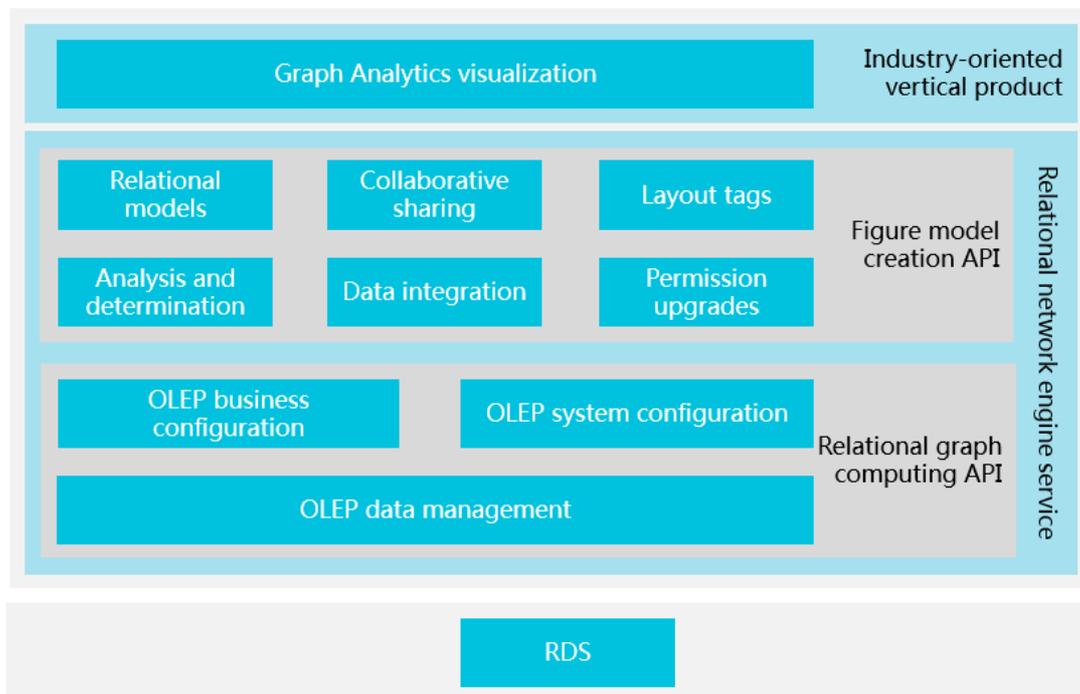
In the collaboration and sharing process, key roles include the initiator and collaborators. The process is as follows: The initiator shares a personal analysis with the specified members. The collaborators can continue the analytics work after receiving the analysis, save the analysis results, and generate a new version of the analysis.

Meanwhile, you can use Graph Analytics to manage collaborative analyses, including deleting analyses, renaming analyses, and managing history versions.

30.4.4. Link engine

The logic and features of the link engine are as follows.

Link engine



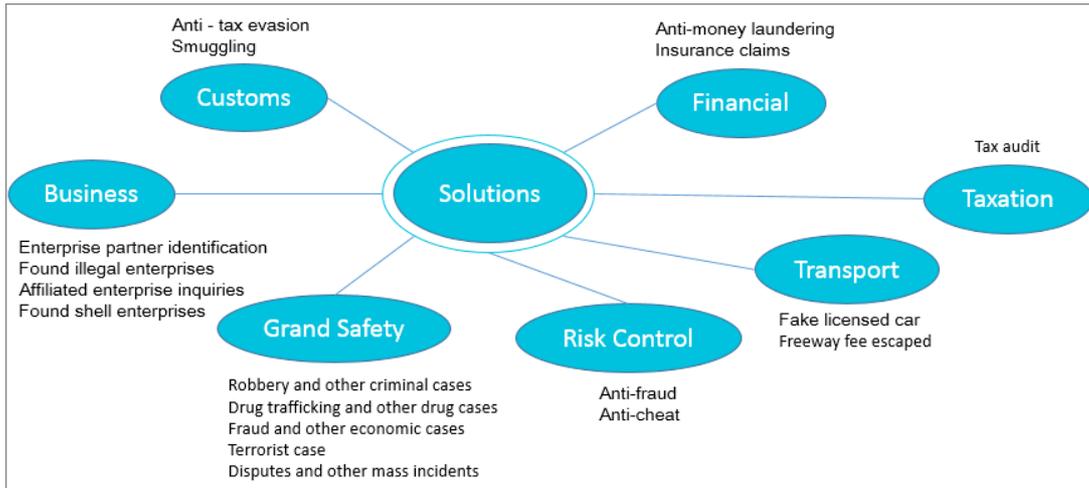
- The relationship network and the search module APIs are open.
- Graph Analytics supports basic computing on nodes and links.

30.5. Scenarios

30.5.1. Scenario overview

This topic describes the main scenarios for Graph Analytics.

Scenarios



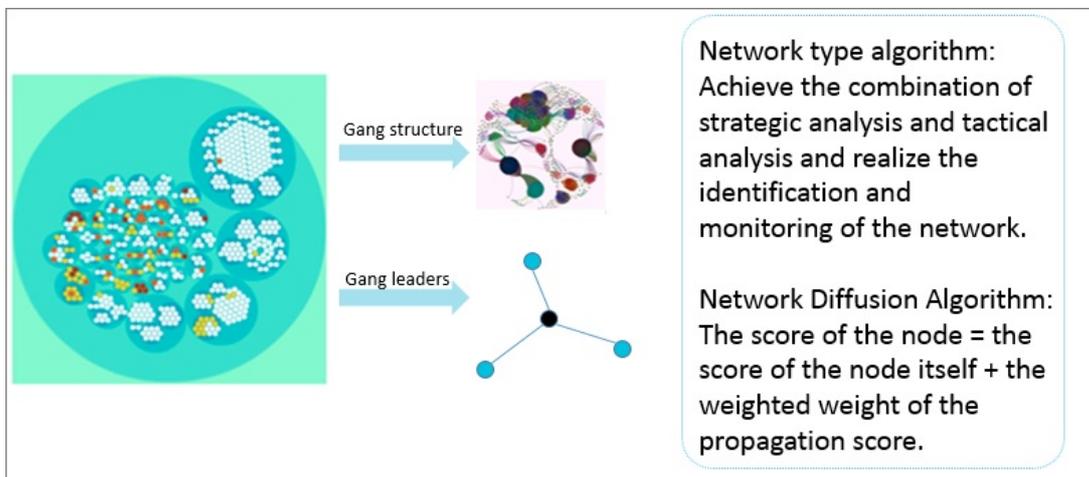
30.5.2. Intelligent relationship networks

Graph Analytics provides intelligent relationship networks to help you quickly analyze the relationships among multiple objects. The following are examples of gang relationship analysis and transaction analysis.

Gang relationship analysis

Graph Analytics can analyze the relationships among gang members, and illustrate the structure of the gang. Graph Analytics can use network topologies to locate key gang members in the relationship network, as shown in the following figure.

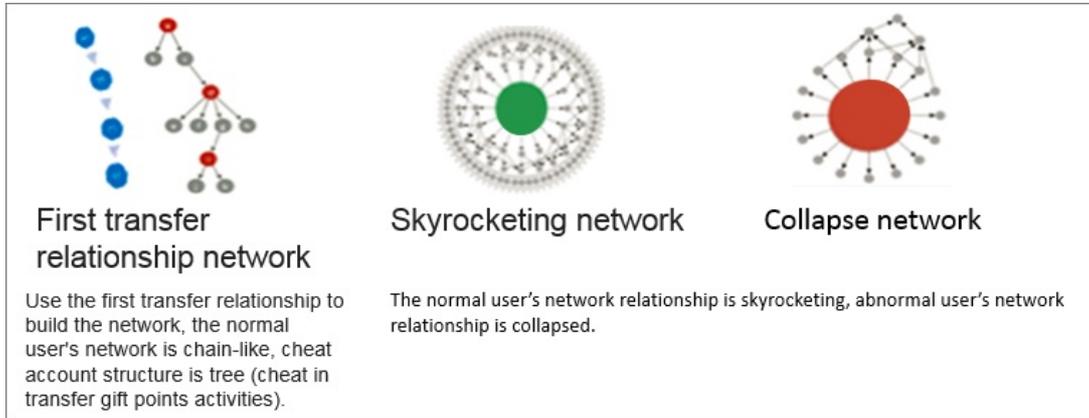
Gang relationship analysis



Transaction analysis

Graph Analytics can detect potential abnormal transactions by analyzing the transactions between accounts. For example, Graph Analytics can detect market manipulation, as shown in the following figure.

Transaction analysis



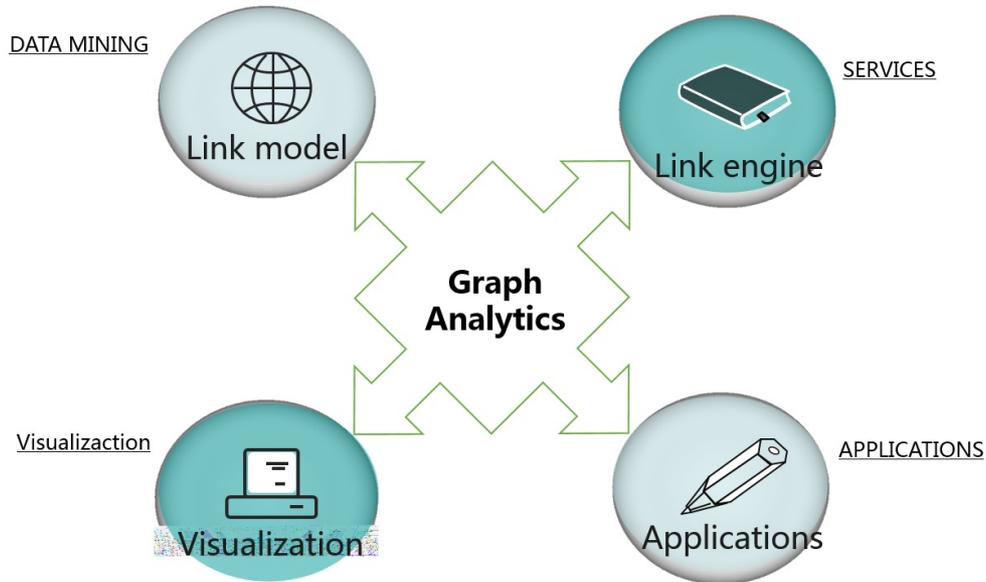
The initial transaction network is generated the first time a transaction relationship is established. In normal cases, the network is linear, but during any market manipulation, the initial transaction network is very intricate. For example, in the case of special offers where users can obtain extra points after completing a transaction, manipulation activities will generate an intricate transaction network. For transactions during market manipulation, you can select the buyer and the seller to build the initial transaction network. In this network, you can analyze the size and growth rate of the network and the proportions of nodes.

Upward-trend networks and downward-trend networks: You can start from some of the most heavily funded nodes in the network and move down along the funding path to check the growth trend of the network. The growth trend of a normal network is upward, while the growth of a marketing cheating network is downward and all paths will eventually go to one account.

30.5.3. Industrial risk control

The applications of Graph Analytics in industrial risk control are as follows.

Industrial risk control graph



- **Link model:** Graph Analytics creates a link model among humans, accounts, equipment, and the environment. Graph Analytics uses the data mining algorithm to identify the properties of each link, such as the strength, influence, and type of the link. Graph Analytics also identifies the key characters and studies their sub-groups.
- **Link engine:** Graph Analytics converts relationship data to standardized engine and interface services to benefit more businesses.
- **Visualization:** Graph Analytics displays the relationships among objects in an intuitive, user-friendly manner.
- **Applications:** Graph Analytics has gained insights from its application in multiple scenarios, including risk control and relationship network recommendation.

30.5.4. Public security protection

Customers in the public security industry can use Graph Analytics to build their own information systems, query and analyze security information, and display the visualized analysis results, as shown in the following graph.

Public security protection graph



30.6. Restrictions

None.

30.7. Terms

Object

Object refers to entities and things that exist in the real world. For example, people, mobile phone numbers, and cars. In Graph Analytics, each object needs a primary key as a unique identifier. For example, the primary keys of people, mobile phones, and cars are ID cards, mobile phone numbers, and license plate numbers, respectively.

Link

Link describes the interaction among multiple objects. In Graph Analytics, a link refers to the relationship built among objects. For example, the link between two mobile phone numbers can be phone calls and text messages. The direct link between a person and a mobile phone number can be that the person is the owner of this mobile phone number.

Event

Events are things that have an impact on specific entities. In Graph Analytics, an event refers to the behavior of an object. For example, people choosing to travel by car is an event.

Property

Properties of objects or links. In Graph Analytics, properties cannot be separated from objects or links. For example, properties of a person include height, weight, birthplace, and name. Properties of a mobile phone number include the registration location and the telecommunications operator of this phone number. Primary keys are also properties. For example, an ID card number is one of the properties of a person, and a mobile phone number is one of the properties of a mobile phone.

OLEP data

This module parses data into objects, properties, events, and links between objects to build a highly abstract OLEP model for relationship analysis.

Link lookup

An infinitely extended analysis that begins with any single object or a group of objects. Link lookup helps to build infinite information associations. The key to intelligence analysis is to discover related clues and intelligence from a large amount of unrelated information and convert the information into useful and actionable intelligence. Graph Analytics provides simple link lookup services and advanced link lookup services.

Group analysis

Analyzes the direct and indirect relationships between a group of objects of the same type or of different types.

Common neighbor analysis

Analyzes the objects that are commonly associated with two groups of objects, including groups of objects of the same type or of different types.

Path analysis

Analyzes the link path between two objects.

Backbone analysis

Locates the core backbone nodes in a group network using smart algorithms.

Lineage analysis

Displays the lineage relationship among people based on families (family IDs).

Information cube

- Behavior analysis
Displays the frequency of an event in chronological order.
- Chronology analysis
Displays the details of each event in chronological order.
- Behavior details
Displays the details of events. The original data records are filtered according to specific rules.
- Object information
Aggregates objects in a relationship network and classifies the objects by type.
- Statistics information
Analyzes the relationships and objects in a relationship network, including object properties, link properties, and the distribution of objects.

Group statistics

Analyzes the distribution of groups in a network. A group consists of multiple object nodes, with any two object nodes connected topologically. Nodes within a merged node are connected topologically.

Label statistics

Collates the label information of object nodes in a relationship network. Graph Analytics supports two types of labels: system labels and user labels. System labels, such as whitelists and blacklists, are defined by the service system for specific nodes. User labels are added to specific nodes by users on the Graph Analytics platform.

Pattern

A pattern is the relationship graph structure model that is predefined in Intelligent Network. Patterns are divided into private patterns and public patterns.

- Private pattern: Only administrators and creators can use private patterns to create private tasks. Private patterns can be set to public patterns, but this is an irreversible operation.
- Public pattern: All users can use public patterns to create public or private tasks. Public patterns cannot be set to private patterns.

Task

Intelligent Network allows you to query subgraphs with the same graph structure as a task specified in a predefined pattern. Tasks are created based on the pattern and used to query data with the same graph structure as the task in the data source. You can modify the graph structure, filter conditions, and other information of the task. Tasks are divided into private tasks and public tasks.

- **Private task:** Only administrators and creators can use private tasks. Private tasks created based on public patterns can be set to public tasks, but this is an irreversible operation.
- **Public tasks:** All users can use public tasks. No public tasks can be converted to private tasks.

31. Machine Learning Platform for AI

31.1. What is machine learning?

Machine learning is a process of using statistical algorithms to learn large amounts of historical data and generate an empirical model to provide business strategies.

Apsara Stack Machine Learning Platform for AI is a set of data mining, modeling, and prediction tools. It is developed based on MaxCompute (also known as ODPS). Machine Learning Platform for AI supports the following functions:

- Provides an all-in-one algorithm service covering algorithm development, sharing, model training, deployment, and monitoring.
- Allows you to complete the entire procedure of an experiment either through the GUI or by running PAI commands. This function is typically intended for data mining personnel, analysts, algorithm developers, and data explorers.
- In Apsara Stack, Machine Learning Platform for AI runs on MaxCompute. Machine Learning Platform for AI allows you to call algorithms to decouple the applications and compute engines after you have deployed algorithm packages in MaxCompute clusters.
- Provides various algorithms and reliable technical support, providing more options to resolve service issues. In the Data Technology (DT) era, you can use Machine Learning Platform for AI to implement data-driven services.

Machine Learning Platform for AI can be applied in the following scenarios:

- Marketing: commodity recommendations, user profiling, and precise advertising.
- Finance: loan delivery prediction, financial risk control, stock trend prediction, and gold price prediction.
- Social network sites (SNSs): microblog leader analysis and social relationship chain analysis.
- Text: news classification, keyword extraction, text summarization, and text analysis.
- Unstructured data processing: image classification and image text extraction through OCR.
- Other prediction cases: rainfall forecast and football match result prediction.

Machine learning can be divided into three types:

- Supervised learning: Each sample has an expected value. You can create a model and map input feature vectors to target values. Typical examples of this learning mode include regression and classification.
- Unsupervised learning: No samples have a target value. This learning mode is used to discover potential regular patterns from data. Typical examples of this learning mode include simple clustering.
- Reinforcement learning: This learning mode is complex. A system constantly interacts with the external environment to obtain external feedback and determines its own behavior to achieve a long-term optimization of targets. Typical examples of this learning mode include AlphaGo and driverless vehicles.

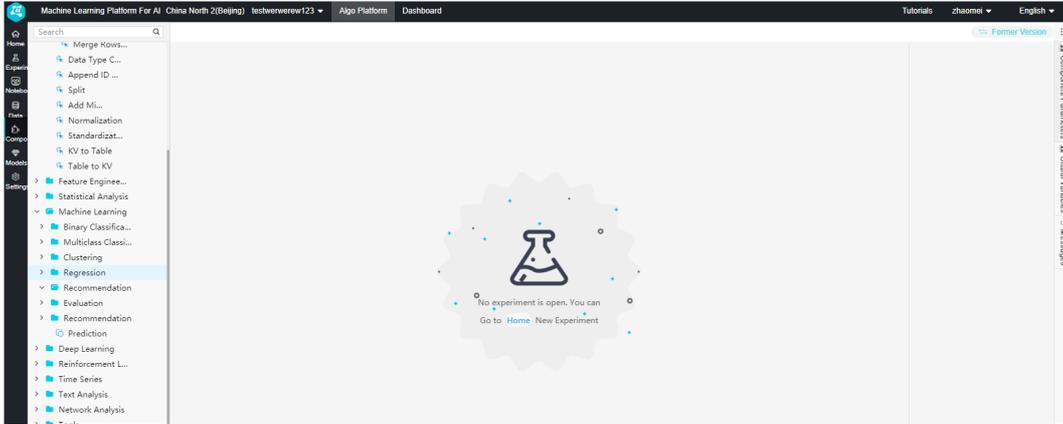
31.2. Benefits

Alibaba Cloud Machine Learning Platform for AI has the following benefits:

All-in-one visual user interface

- Machine Learning Platform for AI provides a Web interface for you to mine data by dragging and dropping components without programming, like piling up blocks, as shown in **User interface**.

User interface



- Machine Learning Platform for AI provides the data model visualization function. It allows you to use charts to view data analysis results and algorithm evaluation.
- Machine Learning Platform for AI provides an all-in-one solution for data processing, model training, prediction, evaluation, model deployment, service building, and task scheduling.
- In addition to the Web interface, Machine Learning Platform for AI also provides command line tools to easily integrate algorithms into your projects.

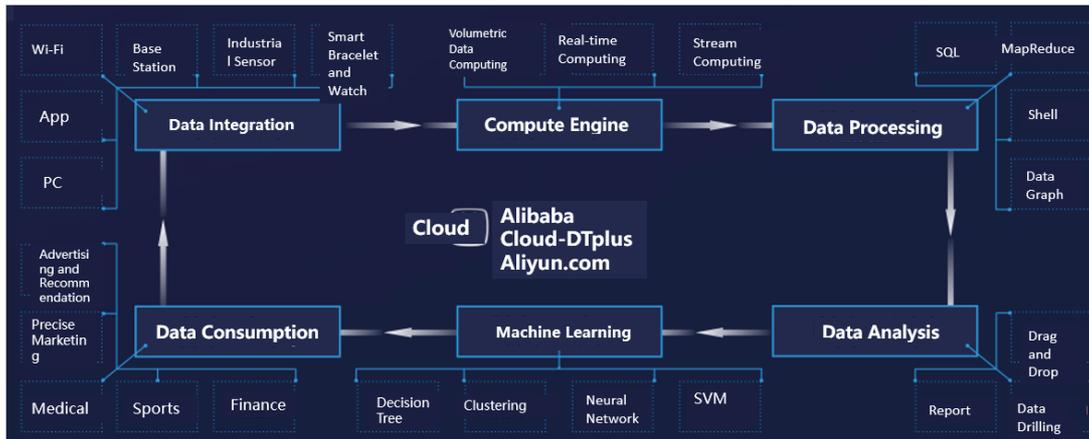
Multiple high-performance machine learning algorithms

- Machine Learning Platform for AI provides nearly 100 machine learning algorithms that can be applied to multiple business scenarios, such as data preprocessing, clustering, regression, text analysis, and feature processing algorithms.
- Compared with traditional software, Machine Learning Platform for AI adopts the latest and optimal algorithms in the machine learning industry to improve the computing capability and accuracy.
- Machine Learning Platform for AI supports deep learning and GPU job scheduling. Machine Learning Platform for AI integrates and completely optimizes the TensorFlow framework. You can get started with TensorFlow for model training.
- Machine Learning Platform for AI provides open-source algorithms that are developed based on years of experience of Alibaba Cloud in big data mining and utilization. This greatly shortens the data modeling, model deployment, and model utilization period.

Full compatibility with Alibaba Cloud services

- Apsara Stack has established a big data ecosystem, such as Machine Learning Platform for AI. All services are ready for use after you activate them.
- Machine Learning Platform for AI runs on MaxCompute and is integrated with DTplus DataWorks to help data mining, parent and child node data collection, experiment scheduling, and data utilization, as shown in [Alibaba Cloud DTplus services](#).
- Based on the MPI, PS, graph algorithms, and MapReduce computing frameworks and distributed algorithms, Machine Learning Platform for AI easily handles a large amount of data.

Alibaba Cloud DTplus services



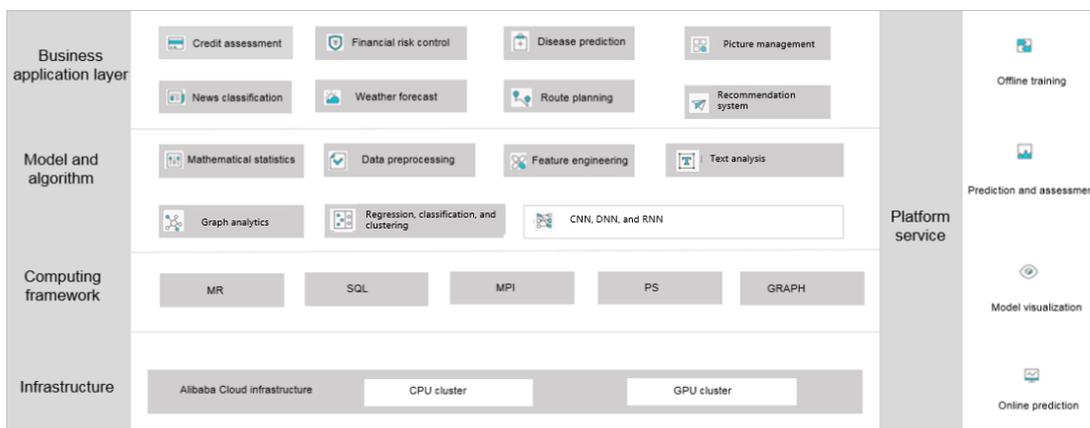
High-quality technical support

Machine Learning Platform for AI is supported by Alibaba algorithm scientists and Apsara Stack technical support. If you have any issues, submit a ticket through the ticket system or contact Apsara Stack technical personnel.

31.3. Architecture

[Basic architecture of Machine Learning Platform for AI](#) shows the basic architecture of Machine Learning Platform for AI.

Basic architecture of Machine Learning Platform for AI



The architecture is composed of the following layers:

- Infrastructure layer: provides the cluster resources for computing. You can choose CPU or GPU

computing clusters based on the algorithm type.

- The CPU cluster runs machine learning algorithms and provide computing resources such as CPU and memory resources. Computing resources are centrally managed by an algorithm framework. After jobs are submitted, the algorithm framework schedules compute nodes in the CPU cluster and dispatches jobs to the compute nodes.
- The GPU cluster runs deep learning framework jobs and provides computing resources such as GPU and graphics memory. Computing resources are centrally managed by an algorithm framework. After jobs are submitted, the algorithm framework schedules compute nodes in the GPU cluster. For a task that requires multiple workers and GPUs, a virtual network is automatically created to dispatch the jobs to the compute nodes in the virtual network.
- Computing framework layer: manages the CPU resources, GPU resources, and a basic runtime environment for algorithms, such as the MapReduce runtime library, MPI runtime library, PS runtime library, and TensorFlow framework.

The deep learning framework TensorFlow supports the open-source version 1.4. The computing framework layer also optimizes the performance and I/O interfaces. You can use TensorFlow to read files from and write models to OSS buckets. When TensorFlow is running, you can start TensorBoard to display the status of parameter convergence during convolution.

- Model and algorithm layer: provides basic components such as data preprocessing, feature engineering, and machine learning algorithm components. All algorithm components come from the Alibaba Group algorithm system and have been tested on petabytes of service data.
- Business application layer: The Alibaba search system, recommendation system, Ant Financial, and other projects use Machine Learning Platform for AI for data mining. Machine Learning Platform for AI can be applied to industries such as finance, medical care, education, transportation, and security.

Based on this architecture, Machine Learning Platform for AI provides a Web-based visual algorithm experiment console. The Web GUI allows you to perform offline training, prediction, and evaluation, visualize models, deploy online prediction services, or release experiments to the scheduling system of DataWorks.

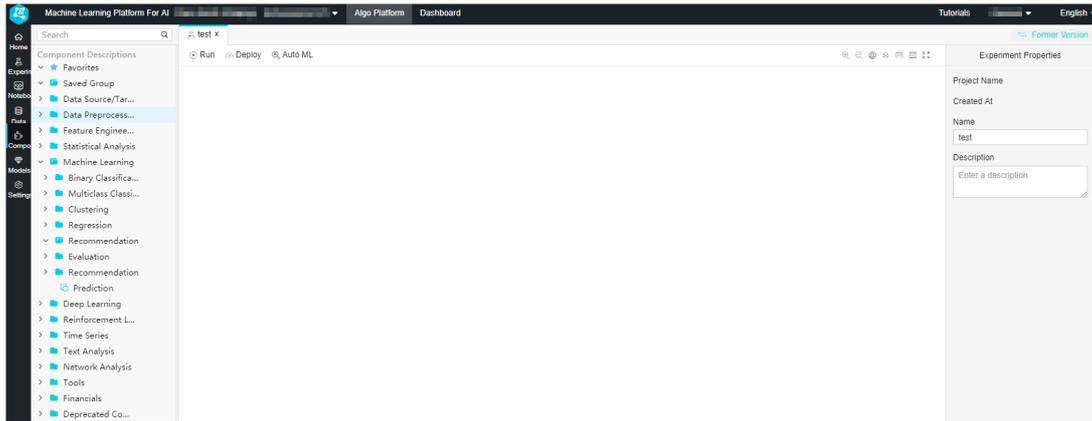
31.4. Features

31.4.1. Visualized modeling

Machine Learning Platform for AI provides the easy-to-use visual modeling feature, which allows you to view the logic of the procedure.

The visual modeling pages include the algorithm platform page and online model service page. [Algorithm platform page](#) shows the function section on the algorithm platform page. [Online model service page](#) shows the online model service page.

Algorithm platform page



Sections on the algorithm platform page

The algorithm platform page includes the following sections:

- **Features section:** displays machine learning features and information, such as experiments, components, data sources, and models, in a tree structure.
- **Canvas section:** You can drag and drop components to the canvas to build a directional workflow in order to complete data mining tasks, such as the metadata collection, data processing, modeling, and model deployment.
- **Properties section:** You can configure component parameters in this section.

Features section

The features section on the algorithm platform page provides the following menus:

- **Search:** You can search data, tables, and experiments.
- **Experiments:** After you double-click the name of an experiment, the canvas displays the directional flowchart of the experiment. You can continue modifying the experiment.
- **Data sources:** allows you to view and manage all data tables.
- **Components:** provides multiple key features of machine learning, such as machine learning components.
- **Models:** allows you to manage all models.
- **Developer tool:** allows you to view the experiment runtime log and troubleshoot experiment issues based on returned error messages and alerts.

Online model service

In the upper section of the page, select Online Model Service to go to the online model service page. This page displays user-created online services. You can monitor or select an action for these algorithm services.

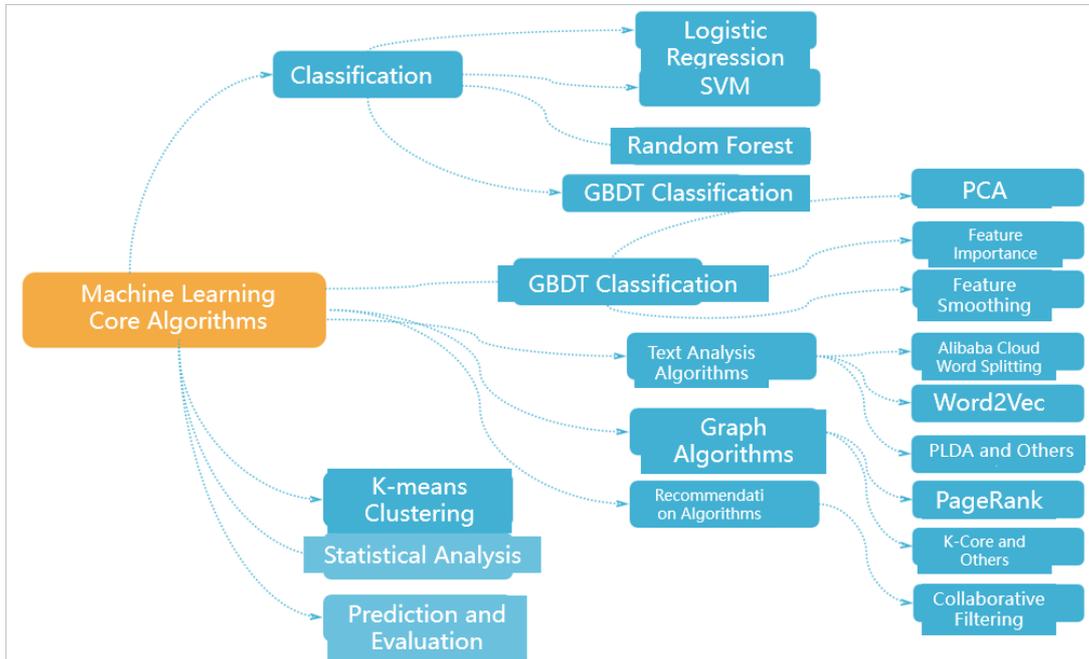
Online model service page

Model ID	Model Name	Created By	Monitor	Current Version	Status	Resources Used	Runtime	Updated At	Source	Actions
10	test1234			v1.0.0	Running	1Process 1Queue	97758*	11/28/2019 14:19:31		Update Delete Distribute Traffic

31.4.2. All-in-one experience

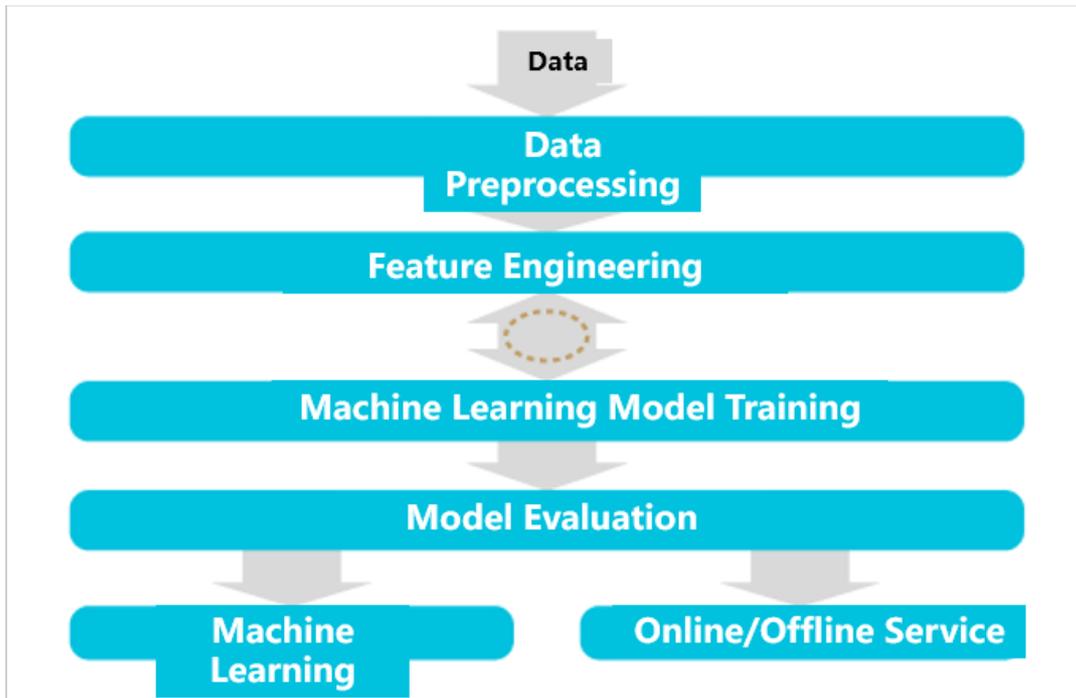
Machine Learning Platform for AI has a complete algorithm library, as shown in [Algorithm library of Machine Learning Platform for AI](#).

Algorithm library of Machine Learning Platform for AI



Typically, you need to perform many operations to complete data mining or model training such as data extract, transform, and load (ETL), data preprocessing, feature engineering, model training, evaluation, and deployment, as shown in [Algorithm development process of atypical model](#). Machine Learning Platform for AI provides an all-in-one development environment with a complete set of components and tools for you to complete the entire data mining or model training task, such as metadata processing and model deployment. With these basic components, you can import data to the platform, create an experiment, and create solutions to resolve issues in different scenarios, and save costs on environment switching.

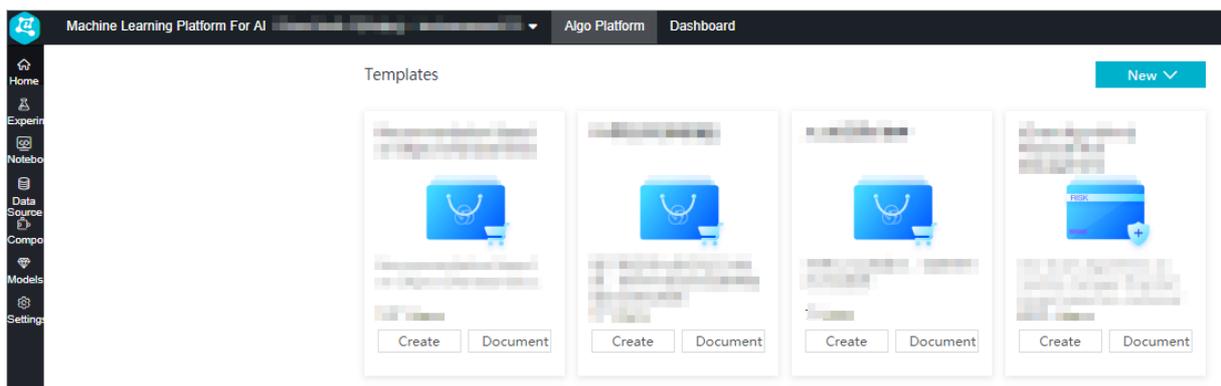
Procedure of developing algorithms for typical machine learning models



31.4.3. Multiple templates on the homepage

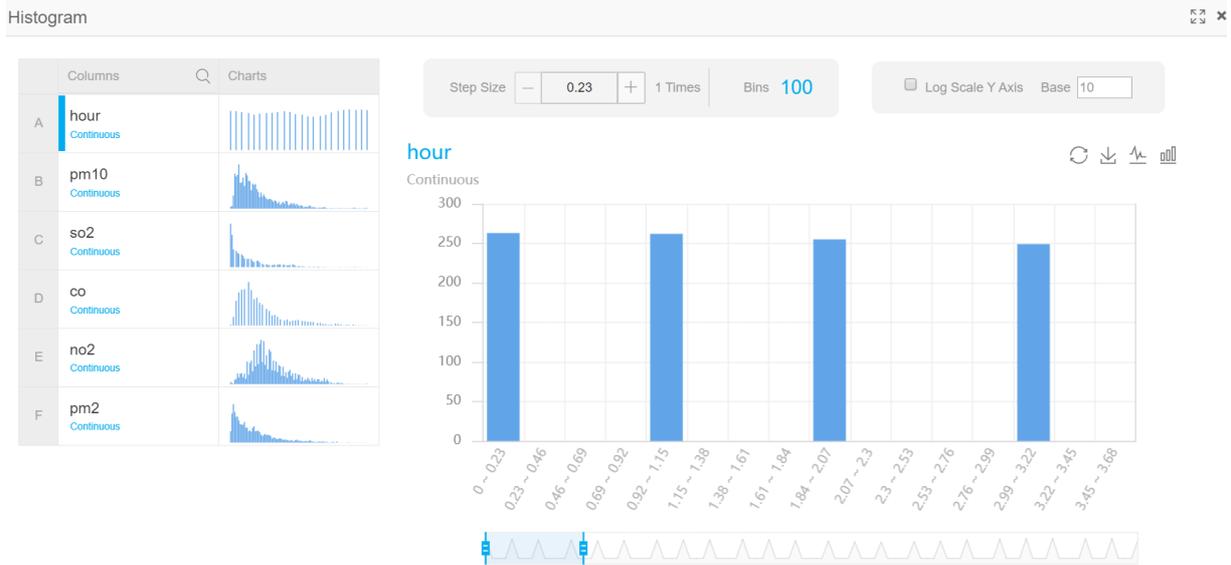
To help data analysts get started with the service, Machine Learning Platform for AI provides a set of experiment templates for scenarios such as product recommendations, text analysis, financial risk management, and weather prediction. These templates contain configurations and data that you can run the experiment with.

You can create experiments from the templates provided on the homepage. You can learn information about how an experiment is configured, how machine learning works, and how data is processed.



31.4.4. Data visualization

You can right-click an output component to view the visual output model. For example, you can view the model evaluation report and data analysis results. Visualized output can be displayed in multiple forms such as line charts, dot charts, and bar charts.



31.4.5. Model management

Visualized model management:

1. In the left-side navigation pane, click **Models**.
2. Expand the **Models** folder to locate the model built from a specified experiment.
3. Right-click the model and then select **Show Model**.

Binary Logistic Regression ⌵ ⌶

If the input is sparse data, feature columns with zero weight are not displayed.

Column Name ▲	Weight	
	1 ▲	0 ▲
pm10	15.7360985405929	-
so2	1.474706135267199	-
co	0.8739231801649705	-
no2	8.652489608506198	-
Constants	-14.47551218333226	0

Save to MaxCompute: pai_lr_coefficient_xlab_m_logisticregres_1165164_v0 Save Close

? **Note** You can also right-click the model to perform other actions. For example, you can export the PMML file or deploy the model.

31.4.6. Multiple algorithm components

The current version of Machine Learning Platform for AI provides up to 84 algorithm components of 10 categories on Apsara Stack, as listed in the following table.

Level 1	Level 2	Level 3	Description
1. Data source/target	1.1 Read MaxCompute table	-	Reads a MaxCompute table.
	1.2 Write MaxCompute table	-	Writes data to a MaxCompute table.
2. Data preprocessing	2.1 Sampling and filtering	2.1.1 Weighted sampling	Collects samples at the specified ratio.
		2.1.2 Random sampling	Collects samples randomly.
		2.1.3 Filtering and mapping	Filters data based on the SQL WHERE conditions.
		2.1.4 Stratified sampling	Collects samples by stratum.
	2.2 Data merge	2.2.1 JOIN	SQL JOIN
		2.2.2 Merge columns	Merges two columns from two tables.
		2.2.3 Merge rows (UNION)	SQL UNION
	2.3 Others	2.3.1 Standardization	Standardizes a column of a table.
		2.3.2 Splitting	Splits data at the specified ratio.
		2.3.3 Normalization	Normalization is a method to simplify computation. Normalization converts a dimensional expression into a dimensionless expression (scalar).
		2.3.4 Missing data imputation	Replaces a null or specified value with the maximum, minimum, average, or custom value.
		2.3.5 KV to Table	Converts the specified KV (key:value) pair to table columns.
		2.3.6 Table to KV	Converts the specified table columns to KV pairs.
		2.3.7 Append ID column	Adds a custom ID column to a table.

Level 1	Level 2	Level 3	Description
3. Feature engineering	3.1 Feature transformation	3.1.1 PCA	A dimensionality reduction algorithm.
	3.2 Feature importance evaluation	3.2.1 Linear model feature importance	Evaluates the features of the corresponding algorithm.
		3.2.2 Random forest feature importance evaluation	
4. Statistical analysis	4.1 Percentile	-	Calculates the percentile of a column.
	4.2 Data pivoting	-	Provides data pivoting functions.
	4.3 Covariance	-	Computes the covariance of the specified data.
	4.4 Empirical probability density chart	-	Obtains nonparametric distribution based on predicted probability distribution.
	4.5 Chi-square goodness of fit test	-	Determines the differences between the observed and the expected frequencies for each classification of a single multiclass classification nominal variable.
	4.6 Chi-square test of independence	-	Verifies whether two factors that each have two or more classes) are mutually independent. The null hypothesis is that the two factors are independent of each other.
	4.7 Two sample T test	-	The two sample T test includes the independent sample T test and the paired sample T test. The two samples independent of each other are called independent samples. An independent sample T test checks whether two samples are significantly different from each other. A paired sample T test checks whether the mean values from two paired populations are significantly different from each other.
	4.8 One sample T test	-	One sample T test tests whether the mean of a normally distributed population differs significantly from the target value.

Level 1	Level 2	Level 3	Description
	4.9 Normality test	-	Determines whether a dataset is well-modeled by a normal distribution.
	4.10 Lorenz curve	-	Compares the total income of a population with the actual distribution of income across the population.
	4.11 Whole table statistics	-	Calculates statistical information of each column in a table, including the default value, maximum value, minimum value, variance, and offset.
	4.12 Pearson coefficient	-	Calculates the Pearson coefficient of two numerical columns.
	4.13 Histogram	-	Queries a histogram based on multiple metrics.
	4.14 Scatter plot	-	A chart where data points are distributed on the Cartesian coordinate plane.
	4.15 Correlation coefficient matrix	-	Calculates a matrix of correlated coefficients for multiple columns.
	5.1 Binary classification	5.1.1 Linear SVM	A type of supervised machine learning algorithm used to identify and classify models, and then perform regression analysis.
		5.1.2 Logistic regression for binary classification	A type of supervised machine learning algorithm used to generate a binary classification model from training data.
		5.1.3 GBDT binary classification	GBDT is an iterative decision tree algorithm that calculates results based on the final conclusions of multiple decision trees.
		5.2.1 Logistic regression for multiclass classification	A linear regression algorithm used for multiclass classification.
		5.2.2 Random forest	A type of classifier that uses multiple trees and sample data to generate models and make predictions.

Level 1	Level 2	Level 3	Description
5. Machine learning	5.2 Multiclass classification	5.2.3 KNN	For each row in the prediction table, this component selects K records nearest to the row from the training table, and takes the class with the maximum number of records among the K records as the class of the row.
		5.2.4 Naive Bayes	A family of classification algorithms based on Bayes' theorem and independent hypothesis of feature conditions.
	5.3 Regression	5.3.1 GBDT regression	An algorithm that uses the GBDT structure for regression.
		5.3.2 PS linear regression	An algorithm that supports a large amount of training data based on parameter servers.
		5.3.3 Linear regression	Analyzes the linear relationship between a dependent variable and multiple independent variables.
	5.4 Clustering	5.4.1 K-means clustering	Clustering similarity is calculated based on a central object obtained by using mean values of objects in different clusters.
	5.5 Evaluation	5.5.1 Binary classification evaluation	Evaluates and predicts the preceding algorithms.
		5.5.2 Multiclass classification evaluation	
		5.5.3 Regression model evaluation	
		5.5.4 Clustering model evaluation	
		5.5.5 Confusion matrix	
	5.6 Prediction	5.6.1 Prediction	
	5.7 Collaborative recommendation	5.7.1 Collaborative filtering (etrec)	Etrec is an item-based collaborative filtering algorithm that uses two input columns and provides the top K items with the highest similarity as the output.

Level 1	Level 2	Level 3	Description
6. Time series	6.1 x13_arma	-	ARIMA is short for Autoregressive Integrated Moving Average Model. x13-arma is an ARIMA algorithm based on the open-source X-13ARIMA-SEATS seasonal adjustment.
	6.2 x13_auto_arma	-	Automatically selects an ARIMA model based on Gomez and Maravall processes.
	7.1 Word splitting	-	An algorithm that is used to split words in the specified text. Currently, only Chinese-language is supported for the Taobao and Internet word splitting models.
	7.2 Word frequency statistics	-	After word splitting, words are listed in the same order of document IDs. The frequency of a word appears in each document is calculated.
	7.3 TF-IDF	-	A statistical method for evaluating the importance of a word within a document in a collection or corpus.
	7.4 PLDA	-	Provides probability distribution of the topic of each document.
	7.5 Word2Vec	-	Converts words to vectors.
	7.6 Convert rows, columns, and values to KV pairs	-	Converts a set of row, column, and value to a KV pair (row, [col_id,value]).
	7.7 Text summarization	-	Algorithmically extracts abstracts from documents.
	7.8 Keyword extraction	-	Extracts the words from a document that are most correlated to the meaning of the document.
	7.9 Sentence splitting	-	Splits sentences by punctuation.
	7.10 Deprecated word filtering	-	A preprocessing method in text analysis. The algorithm is to filter out the noise in word splitting results (such as of, yes, and ah). Custom word libraries are supported.

Level 1	Level 2	Level 3	Description
7. Text analysis	7.11 String similarity		String similarity calculation is a basic operation in machine learning. It is typically used in industries such as information retrieval, natural language processing, and bioinformatics. This algorithm has five methods it can use to calculate similarity: Levenshtein distance, longest common substring, string subsequence kernel, cosine, and simhash_hamming. It also supports two input methods: string-to-string calculation and top N calculation.
	7.12 String similarity-top N	-	The similarity is calculated based on the maximum number of top N data records.
	7.13 Semantic vector distance	-	Calculates the extension words (sentences) of the specified words (sentences) based on the calculated semantic vectors (such as word vectors calculated by the Word2vec component). For example, you can use this component to generate a list of words that are most similar to the word that you have entered based on the word vector results calculated by the Word2vec component.
	7.14 N-gram counting	-	Generates N-grams based on the words and counts the number of the corresponding N-grams within all corpus.
	7.15 PMI	-	Counts the co-occurrence of all words in several documents and calculates the PMI between two documents.
	7.16 Document similarity	-	Calculates the similarity between documents or sentences that are separated with spaces based on the similarity of strings. The document similarity is calculated in same way as string similarity calculation.

Level 1	Level 2	Level 3	Description
8. Deep learning	8.1 TensorFlow 1.4	-	Uses the PAI-Tensorflow framework to implement deep learning.
	8.2 Read OSS buckets	-	Sets the authorization information and domain name used by the machine learning service to read data from and write data to your OSS bucket.
9. Network analysis	9.1 K-Core	-	The KCore of a graph is the subgraph that remains after all nodes with a degree less than or equal to K have been removed.
	9.2 Single-source shortest path	-	Calculates the shortest path between two points.
	9.3 Page ranking	-	Calculates the rank of a Web page.
	9.4 Label propagation clustering	-	A graph-based semi-supervised machine learning algorithm. Nodes are labeled based on the label of its neighboring nodes. The scale at which labels are propagated is determined by the degree of similarity between neighboring nodes. Labels are propagated iteratively and updated over time.
	9.5 Label propagation classification	-	A semi-supervised classification algorithm that uses the label information of labeled nodes to predict that of unlabeled nodes.
	9.6 Modularity	-	A measure of the structure of networks. It measures the closeness of communities divided from a network structure. A value larger than 0.3 represents an obvious community structure.
	9.7 Maximal connected subgraph	-	A maximal connected subgraph is a connected subgraph of an undirected graph that is connected to the vertex connected to the G graph.

Level 1	Level 2	Level 3	Description
	9.8 Vertex clustering coefficient	-	This coefficient is used to calculate the peripheral density of nodes near a node in an undirected graph. The density of a star network is 0. The density of a fully meshed network is 1.
	9.9 Edge clustering coefficient	-	This coefficient is used to calculate the peripheral density of each edge in an undirected graph.
	9.10 Counting triangle	-	Generates all triangles in an undirected graph.
	9.11 Tree depth	-	Generates the depth and tree ID of each node in a network composed of many trees.
10. Tools	10.1 MaxCompute SQL	-	Executes MaxCompute SQL statements.

31.5. Scenarios

Currently, Machine Learning Platform for AI can be applied to the following scenarios:

Marketing

- Use cases: commodity recommendations, user profiling, and precise advertising.
- Example: Machine Learning Platform for AI associates user shopping behavior data with commodities to offer commodity recommendations and evaluate the recommendation results. For more information, see .

Finance

- Use cases: loan delivery prediction, financial risk management, stock trend prediction, and gold price prediction.
- Example 1: Agricultural loan delivery is used in a typical example of data mining. A lender uses machine learning to build an empirical model based on historical data such as the lendee's annual income, cultivated crop type, and debit and credit history. This model is then used to predict the lendee's capacity. For more information, see .
- Example 2: Users' credit card expense records are processed by a machine learning algorithm. After raw data binning and feature engineering transformation, data is used to build a linear model. The final credit score of each user is determined by the model predictions, and can be used in a variety of loan and finance related credit checks. For more information, see [Creditscore evaluation](#).

Text

- Use cases: news classification, keyword extraction, text summarization, and text analysis.
- Example: A simple system for automatic commodity label classification is built using the text analysis function of Machine Learning Platform for AI.

Take online shopping as an example. A commodity typically has labels for multiple dimensions. For example, the commodity description of a pair of shoes may be "Korean Girl Dr. Martens Women's Preppy/British-style Lace-up Dull-polish Ankle High Platform Leather Boots." A bag may be described as "Discount Every Day 2016 Autumn and Winter New Arrival Women's Korean-style Seashell-shaped Tassel Three-way Bag as a Messenger Bag, Hand Carry Bag, and Shoulder Bag."

Each product description contains multiple dimensions such as the time, place of origin, and style. E-commerce platforms face the daunting challenge of how to classify hundreds of thousands of products based on these specified dimensions. The biggest challenge is determining which labels constitute the dimensions of each product. A label classification system can be built by using a machine learning algorithm to automatically learn label terms. For example, it could learn location-related labels such as Japan, Fujian, and Korea. For more information, see .

Unstructured data processing

- Use cases: image classification and image text extraction by using optical character recognition (OCR).
- Example: A prediction model can be quickly built for image recognition by using the TensorFlow deep learning framework. The TensorFlow deep learning framework can begin to recognize images within half an hour. The system returns image classification results. Image recognition by using deep learning can also be used in illicit image filtering, facial recognition, and object detection.

Other prediction cases

- Use cases: rainfall forecast, football match result prediction, microblog leader analysis, and social relationship chain analysis.
- Example: Air quality can be predicted by Machine Learning Platform for AI based on historic air quality index data such as PM 2.5, carbon monoxide concentration, and nitrogen dioxide concentration. The prediction results can then be used to determine which air quality index has the greatest impact on PM 2.5 levels. For more information, see .

31.6. Limits

This topic describes the limits of Machine Learning Platform for AI.

Item	Description
MaxCompute service deployment	The computing service of Machine Learning Platform for AI relies on MaxCompute to store tables and perform some SQL-related computations. Therefore, MaxCompute must be deployed before you can use the machine learning service.

Item	Description
OSS service deployment (deep learning)	The deep learning service of Machine Learning Platform for AI relies on OSS to store, read, and write data. Therefore, OSS must be deployed.
Limit to component use	For more information, see parameter configurations for each component.

31.7. Terms

This topic describes terms used in Machine Learning Platform for AI.

data mining

A broad definition that describes the use of algorithms to explore useful information from large amounts of data. Typically, data mining uses machine learning algorithms.

Alibaba Cloud DTplus

The big data platform of Alibaba Cloud. DTplus provides enterprises with a complete set of end-to-end big data solutions for fields such as enterprise data warehouses, BI, machine learning, and data visualization. These solutions help enterprises become more agile, smarter, and more perceptive in the data technology (DT) era.

table

Data storage units of MaxCompute. Tables used by machine learning are stored in MaxCompute. Logically, a table is a two-dimensional structure that consists of rows and columns. Each row represents a record. Each column represents a field of the same data type. One record can contain one or more columns. The schema of a table consists of column names and column types.

On Machine Learning Platform for AI, you can create a table, add the table to favorites, and import data to the table. The table is automatically stored in MaxCompute. To delete a table, you must log on to MaxCompute.

partition

Certain columns specified in a table when the table is created. In most cases, you can consider a partition as a directory in a file system.

Tables are stored in MaxCompute. MaxCompute uses each value in a partition column as a partition (directory). You can specify multiple hierarchies of partitions by using multiple table columns as table partitions. The relationships between partitions are similar to those between multiple hierarchies of directories.

When using data, if you specify the name of a partition, only the data in the specified partition is read. This removes the need to scan the entire table for data, improves processing efficiency, and minimizes costs.

lifecycle

The period of time that determines how long a table partition is retained since it was last updated. If a table (partition) is not updated within the specified time period, MaxCompute automatically deletes it.

sparse data format

Datasets in which most data entries are null or have a value of 0. Sparse data can be utilized effectively if efficient methods are used to explore the useful information that exists in the relatively incomplete sparse data set.

On Machine Learning Platform for AI, if the data of a feature in a sample is in the sparse format, you must convert the format to the LibSVM format, select **key:value, key:valuesparse data format** on the parameter setting page, and then upload the data.

feature

An attribute that is used to describe an object. For example, a person can be described by age, gender, occupation, and other attributes. Each of these attributes is a feature of the person.

On Machine Learning Platform for AI, a dataset is stored as a table. A column in the table is a feature of this dataset. The features of data are important to machine learning. Data and its features determine the upper limit of machine learning capabilities. Models and algorithms are used to help machine learning reach the upper limit. Therefore, features must be processed before a machine learning experiment can be executed. Typical feature processing methods include data preprocessing, feature selection, and dimension reduction.

dimension reduction

A method that is used to remove the dimensions that have minor impacts and extract key features from a large number of features. A dimension is the way something is observed. In machine learning, dimensions describe the features of a dataset. If a dataset has millions of features, the training model for machine learning will be complex and the training will take a long period of time. In this case, dimension reduction is required. The dimension reduction algorithms on Machine Learning Platform for AI include PCA and LDA.

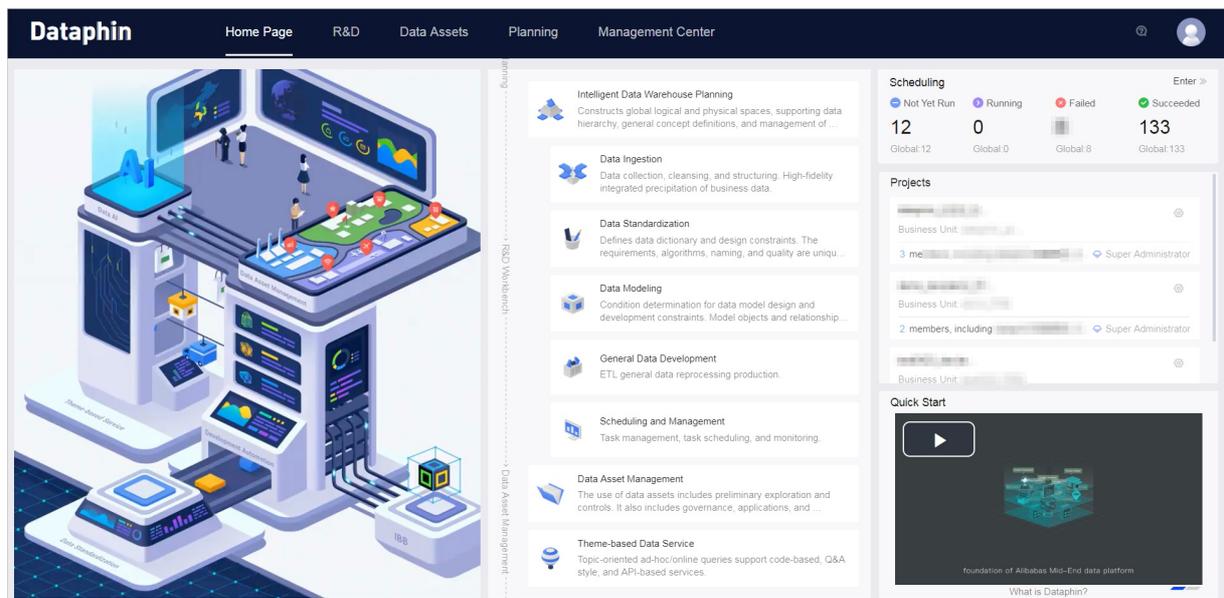
32. Dataphin

32.1. What is Dataphin?

Dataphin is an engine for creating intelligent big data platforms. It is designed to meet the requirements of big data development, management, and utilization across multiple industries. It adopts an OneData, OneEntity, OneService (product, technology, methodology) big data lifecycle management system. The system is developed by Alibaba Cloud and has been proven by years of practice. Dataphin provides an end-to-end intelligent data creation and management solution covering data ingestion, data standardization, data modeling, data development, data distilling, data asset management, and data services. These features help governments and enterprises build an asset-oriented, service-oriented, closed-loop, and self-optimizing intelligent data system with unified standards to stimulate and drive innovation.

Dataphin is integrated with a large amount of compute and storage environments, which enables you to use a single console to process data from various data sources. By using Dataphin, you can import data, produce standard data by data modeling, and create a tag system by extracting tags from entities. This allows you to generate and manage data assets by using your business data knowledge. Dataphin also provides several types of data services including data table search and intelligent voice search.

The following figure shows the Dataphin R&D Workbench.



32.2. Benefits

Dataphin provides the following benefits.

- **Data standardization:** The definitions of dimensions, dimension attributes, business processes, and metrics are standardized based on dimensional modeling. This standardization helps to guarantee the quality of data and accuracy of metrics.
- **Efficient and automatic coding:** You can define atomic metrics, business filters, granularity, and statistical periods. By combining these four types of computing logic components, you can then define derived metrics. You can use these components to create data models. Based on your models, the system will automatically generate code to produce data.
- **Optimal intelligent computation:** You can create logical models from business perspectives.

After you publish your logical models, the system automatically generates the physical representations of the logical models and the code of the logical models. This reduces your dependence on professional data developers.

- **End-to-end development:** Data ingestion, modeling, development, management, data search, and exploration are combined to implement centralized and efficient development.
- **Systematic data catalog:** Based on standardized modeling, efficient and automatic metadata extraction, Dataphin provides a standardized and user-readable data catalog. The data catalog allows you to spend less time finding the data you require.
- **Efficient data search:** An overview of data assets is provided based on your metadata and data from the Dataphin system database to achieve simple, fast, and intelligent search of data tables and data.
- **Visualized data assets:** A business data asset map (data catalog) is built to help represent your business system from different data perspectives, extract business data knowledge, and learn more about key business stages and data.
- **Easy and reliable data utilization:** Data elements can be used for data production after they are created. You can easily search and access logical tables created based on business themes. This simplifies about 80% of query code.
- **High efficiency:** Dataphin provides end-to-end and intelligent data construction and management tools. This reduces data development requirements. Developers can independently run the extract, transform, and load (ETL) procedure to quickly meet the demand for data. The OneData, OneEntity, and OneService methodology (patent pending) enables the abstraction and definition of models and metrics, automatic coding, automatic theme-based data aggregation and output.
- **Low costs:** Dataphin is metadata-based and algorithm intelligence-driven. Automatic data production is independently performed on both the physical platform (backend computing engine) and logical plane (UI). In addition to comprehensive analysis, tracking, and optimization for data assets, Dataphin ensures optimal computation and storage resource allocation. This greatly reduces the cost of data utilization.

32.3. Features

- **Support for compute engines:** Multiple types of compute engines, including MaxCompute and Hadoop.
- **Data import:** You can import and structure data from various data sources, including local and Alibaba Cloud databases, unstructured data storage, and big data storage.
- **Global design:** During the design of the data warehouse architecture, you can define business units, data domains, and projects.
- **Data standardization:** Dataphin allows you to define data standardization elements by configuring parameters in the console. You can define multiple statistical metrics at a time, and then the system processes the metrics to generate aggregate data.
- **Data modeling and development:** You can build logical data models by configuring a graphical user interface. The system generates the code representation of your data models. It also generates tasks to convert your logical data models to physical models. You will not be aware of the code and task generation process. Dataphin also supports custom coding for data development.
- **Scheduling and management:** You can schedule tasks and manage task running.
- **Metadata management:** Dataphin supports standard and automatic metadata extraction to create a unique and centralized metadata warehouse.

- **Asset analysis:** Dataphin visualizes data assets and provides a data catalog. You can gain an overview of your data assets and quickly locate and use the data that you require.
- **Data security:** Dataphin supports access control for projects, tables, and fields.
- **Data service:** Dataphin allows you to perform theme-based queries on logical and physical tables.

32.4. Functions

32.4.1. Overview

Dataphin has the following modules:

Platform

This module helps you learn more about the entire product system and global settings, and understand the product functions to quickly get started. It also implements system management and control to ensure that all the other modules are running as expected.

Global design

Based on a global view of your business and data, you can design an architecture for your data warehouse. During the design, you need to define namespaces (business units), theme domains (data domains), and terms (global objects). You also need to create projects as management units and add data sources.

Data ingestion

Based on the projects and physical data sources defined during global design, the data ingestion module supports data extraction. This involves extracting all kinds of data from all business systems and loading the data into the target databases. This process achieves data synchronization and integration, which facilitates the building of the source data layer by using data cleansing strategies.

Data standardization

Based on the architecture defined in global design and the source data layer built by data ingestion, you can create data elements such as statistical metrics. You can use these data elements to ensure that clear and standardized data will be produced.

Modeling

You can use the data elements created for data standardization to design data models. After the data models are submitted and published, Dataphin automatically generates code for the models and recurring data production tasks. This is a full suite of services that provides complete management of data production on the common dimensional model layer.

Coding

Dataphin provides a code editor for you to configure and submit code tasks.

Resource and function management

Dataphin allows you to manage resource packages (such as JAR type and other file types) to meet data processing requirements. Dataphin supports searching for and using built-in functions. You can also create user-defined functions to meet the specific requirements for functional processing.

Data distilling

Based on the source data layer and the common dimensional model layer, Dataphin can identify the links between target entities and IDs, extract behaviors of the entities, and define tags. This achieves data integration and data mining. Dataphin can then generate and schedule tasks to tag target entities. This is a full suite of services that provides complete management of the data distilling process.

Scheduling and management

Dataphin supports policy-based scheduling and management of tasks generated by modeling, coding, and data distilling. The scheduling and management involves data production task deployment, task implementation, dependency checking, and task management. This ensures that all tasks can run as expected and without interruption.

Metadata warehouse

Dataphin allows you to collect, parse, and manage metadata of the source data layer, common dimensional model layer, and distilled data center.

Data asset management

Based on the metadata warehouse, this module supports deep metadata analysis and data asset management. It shows asset distribution and metadata details. This makes it easy for you to search for data assets and learn about data assets in more detail.

Security management

Dataphin supports managing data quality and security. It allows defining data standardization elements, presenting data permission details, managing data permission approval processes, and monitoring data production tasks and alerts. It also supports end-to-end tracing from data sources to applications. This helps you discover data asset optimization problems and provide solutions.

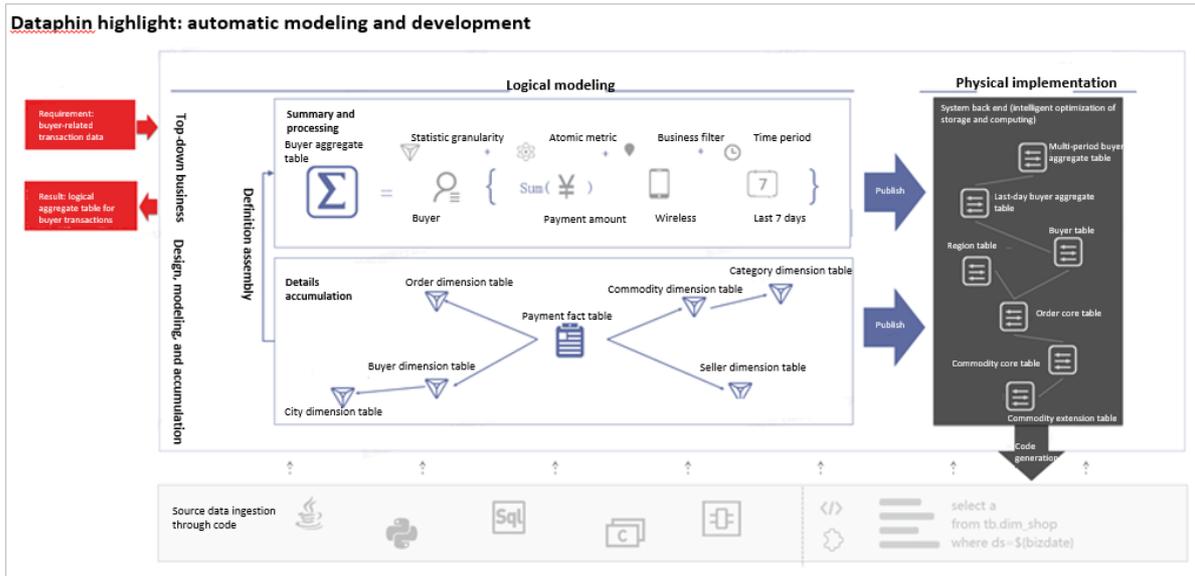
Ad hoc query

This module supports asset data searches through custom SQL queries. You can use the search and analysis engine to quickly search for data in physical tables and theme-based logical tables. Theme-based logical tables are also known as data models or logical models.

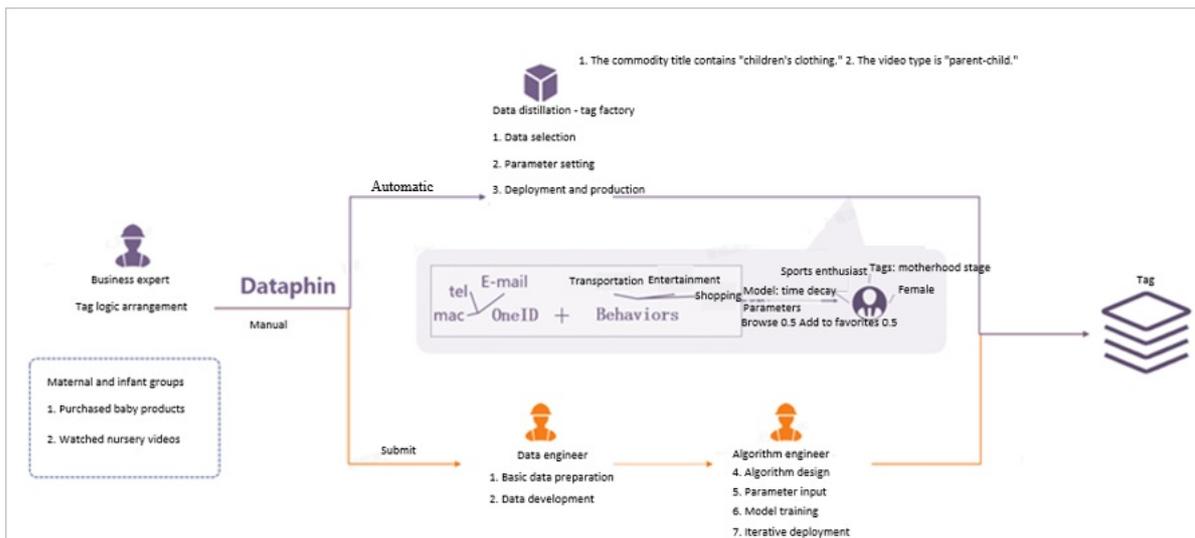
32.4.2. Resolved issues

With Dataphin, you can resolve the following issues:

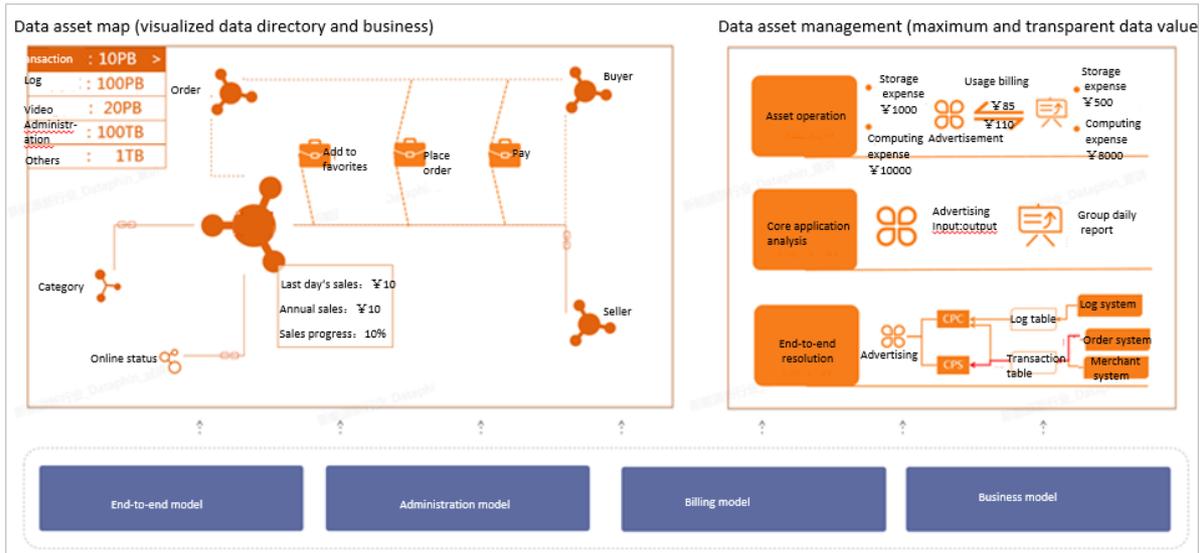
- **Modeling:** You can build data models by using a graphical user interface rather than writing SQL code. The system then automatically publishes the models and generates tasks to produce data. All metrics and standards are clearly defined.



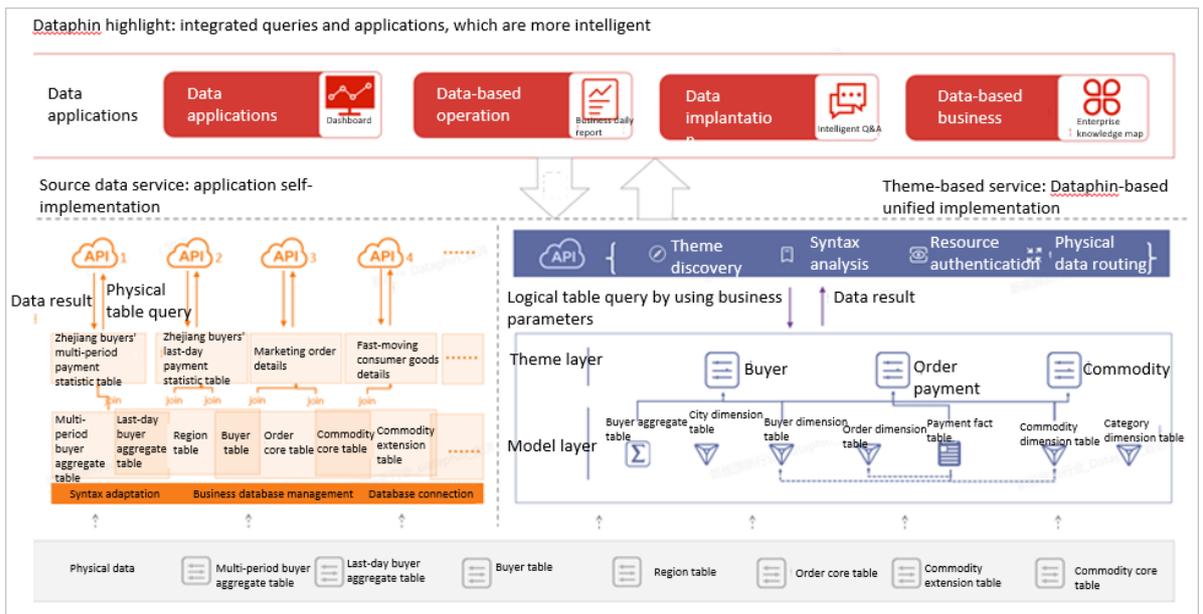
- **Data distilling (coming soon):** You can extract master business data and build a data management platform (DMP) based on entities. This will include three steps and involve customizing parameters, ID recognition, and automatic tag creation following a standard process.



- **Data asset management:** You can create and manage data assets, gain a deep understanding of data assets from a unique perspective, and get more value from your business data.



- Ad hoc queries: Dataphin supports theme-based queries for logical tables. This ensures quick data query and locating, and greatly simplifies SQL query statements. This also ensures that data is produced in a standard, regular, and clear manner. The standardized output data can be used by several business applications.



32.4.3. Platform management

Platform management is a basic function provided by Dataphin. It ensures that all users carry out controlled, orderly, and smooth data development. Project management refers to configuring global settings. Dataphin allows the super administrator to manage the entire platform, including managing buyers accounts and computation resources. On the Dataphin homepage, you can view the introduction of service modules provided by Dataphin and gain access to Dataphin documentation. You can also quickly access target modules.

Account management

User accounts are managed for secure use of the Dataphin system. The existing account system of your enterprise can be connected to Dataphin. Then, the users who need to use Dataphin can be added to Dataphin as members. Users with the highest privileges can manage accounts and permissions of other users.

Computation resource management

Dataphin is a Platform as a Service (PaaS) product. To achieve unified and stable computation, you must select one computing engine type and configure the connection settings for your computing engine sources. This makes Dataphin compatible with the computing engine sources at the Infrastructure as a Service (IaaS) layer. You can select a computing engine type that is suitable for data development in your environment. Dataphin supports two major types of computing engines: MaxCompute and Hadoop. Dataphin can automatically collect and parse the metadata of these two types of engines. For more information about how to collect and initialize metadata, see Metadata warehouse.

Homepage

The homepage provides access to data development and management. It also provides product introduction, and allows you to manage projects and go to the workbench for specific projects. The homepage provides access to the entire process of data production, data management, and data servicing. This allows you to learn about product functions before you get started, and ensure that you can quickly access specific function modules.

Internationalization: language support

The supported languages are Chinese or English. This allows users from different countries and regions to use Dataphin. Dataphin identifies the language that your system uses and automatically selects a default language.

32.4.4. Global design

Based on a global view of your business and data, you can design an architecture for your data warehouse. This is a fundamental step in data development. The architectural design ensures that data is manageable and controllable. The data systems defined and designed during data development, distilling, and management meet mid- and long-term business requirements. The produced business data is service-oriented, theme-based, and easy to use.

The global design involves the following:

- Data warehouse architecture definition based on business characteristics includes business unit management and access control, data domain management and access control, and management of the defined global objects.
- Project definition based on requirements for independent data management and collaborative development includes member management and the management of basic project information and computing resources.
- Data source configuration based on computing resources for projects and requirements for business data includes data source management.

Data warehouse architecture

The data warehouse architecture defines logical namespaces (business units), theme domains (data domains), and terms (global objects) based on business characteristics. This standardizes data definitions during architectural design management and data development control.

Projects

A project is a physical namespace used to isolate users from resources. Projects are created to meet the requirements for independent management of data development projects and efficient management of data resource quality. Data development constraints can be configured for each project.

Physical data sources

Dataphin supports data source creation, modification, and other features that allow you to register and cancel the registration of databases. The data source types supported by Dataphin include MaxCompute, MySQL, SQL Server, and PostgreSQL. Data sources can be used as the source storage or target storage for data synchronization. Some special types of data sources (such as MaxCompute) can serve as the computing engine for projects to function as the computation and storage base.

32.4.5. Data ingestion

The source data layer is built through data ingestion. Before ingesting data, you need to select a business data storage system as the data source. Then, you need to formulate data synchronization, cleansing, and structuring policies to satisfy your data requirements in terms of storage, accuracy (up-to-date), and quality.

Data ingestion is an important initial stage in data development. The data synchronization suite of Dataphin is developed based on several years of industry practice. In the past, Alibaba has overseen the synchronization and exchange of many types of data including business and log data. This helps achieve efficient ingestion of raw business data. The data transmission channel can collect and analyze metadata to check the amount and content of data that has been transmitted. The flexible management of custom error tolerance mechanisms is also supported. This helps achieve high-quality data synchronization.

Data source configuration

You can import and manage multiple data sources. The data source list allows you to manage imported data sources and add various different types of additional data sources. Currently, data sources that can be used for data synchronization include MaxCompute, MySQL, SQL Server, PostgreSQL, and Hive.

Data synchronization

You can select source data and target data, configure parameters for incremental or full synchronization, and identify mappings between source data fields and target data fields. You can also configure the data transfer rate and the number of concurrent sync tasks. With these configurations, synchronization tasks can be generated and scheduled.

32.4.6. Data standardization

In most cases that involve traditional development, specific and important data creation and development (such as data modeling and metric definition), depend on the developer's professional capabilities. Without a uniform naming convention, standards for development and designs are transferred based on individual and changing documents. This may cause a series of problems such as metric name conflicts or repeated calculation.

Based on the OneData methodology, Dataphin standardizes the definition of important data elements such as dimensions, business processes, and metrics. This ensures unique computing logic and names, and eliminates metric ambiguities during the initial stages of architectural design. In addition, Dataphin provides form-based interfaces for you to create multiple metrics at a time. This lowers the requirements of data development and increases overall development efficiency. This also allows business users with limited data analysis expertise to carry out development work by using Dataphin.

Data standardization involves defining five types of data elements: dimensions, business processes, atomic metrics, business filters, and derived metrics. Dataphin helps you design a data architecture by creating business units and data domains. You can extract standard data elements and reuse data elements based on the data architecture. Standard data elements include data warehouse themes (such as granularity that is composed of dimensions) and metric creation elements (such as atomic metrics and business filters).

32.4.6.1. Dimensions

- A dimension is unique within a business unit and it exclusively belongs to a data domain. This standardizes naming and theme classification.
- You can create dimensions by adding additional attributes to an existing dimension, which is used as a parent dimension.
- Dataphin supports the creation of various types of dimensions, including common, common (hierarchy), enumeration, and virtual dimensions.
- Dataphin allows you to view and manage the list of dimensions created in a specific business unit or a specific project. You can also view and modify each dimension.

View and manage the dimension list

Dataphin allows you to view the list of dimensions created in a specific project. You can view the name, creator, and publishing status of each dimension. You can search for a specific dimension in the list, and then modify, unpublish, or delete the dimension.

View and manage dimensions

Dataphin provides form-based interfaces for you to view, create (using a standard template), and modify dimensions. A dimension is a key concept of business. You need to specify the following information when creating a dimension:

- **Basic information:** the data domain (to which the dimension belongs), name, display name, and description. The name is prefixed with `dim_` by default to distinguish the name from other names.
- **Logic information:** The logic information is used to describe and define the scope of the dimension. This is to ensure that the dimension is accurate and unique when you later need to add dimension attributes. The required configurations vary by dimension type.

Quick view of dimensions: You can click a dimension in the left-side navigation pane to view basic information of the dimension and then perform supported operations on the dimension. This does not affect your previous operations.

32.4.6.2. Business processes

A business process is a collection of the smallest unit of behaviors or events that occur in a business activity. For example, the smallest unit of behavior can be to create an order or browse a web page. The behaviors occurring in a business process, such as paying for an order and browsing a web page, are recorded in a fact table. The fact table models a particular business process.

Similar to dimension, business process is a key concept in the OneData methodology used for designing the data architecture. It works with dimensions to define the data architecture. Dataphin supports standard definition for business processes. This allows you to check the overall business data of your organization and easily categorize fact tables by business process.

To ensure that a fact-based model is built in a unified and standard manner, a business process is unique within a business unit and it exclusively belongs to a data domain. This standardizes naming and theme classification.

Dataphin allows you to view and manage the list of business processes created in a specific business unit or a specific project. You can also view and modify each business process.

View and manage the business process list

Dataphin allows you to view the list of business processes created in a specific project. You can view the name, creator, and publishing status of each business process. You can search for a specific business process in the list, and then modify or delete the business process.

View and manage business processes

Dataphin provides form-based interfaces for you to view, create (using a standard template), and edit business processes. A business process is a key concept of business. You need to specify the following information when creating a business process: data domain (to which the business process belongs), business process name, display name, and description.

32.4.6.3. Atomic metrics

An atomic metric is an abstraction of computing logic. To eliminate definition and development inconsistency, Dataphin introduces the concept of "Design to Code". When a metric is defined, the statistical criteria (computing logic) is also defined. Re-engineering of the ETL process is not required, which increases development efficiency and ensures the consistency of statistical results. Based on the complexity of computing logic, Dataphin categorizes atomic metrics into native atomic metrics and composite metrics. An example of a native atomic metric is payment amount. A composite metric is created based on the combination of atomic metrics. For example, the average sales per customer is calculated by dividing the total sales by the number of customers.

An atomic metric is unique within a business unit and has only one source logical table. The computing logic of an atomic metric is defined based on the fields of the source logical table model. This ensures that all statistical metrics are created in a unified and standard manner. The data domain of each logical table linked to the source logical table is retrieved to trace the data domains to which the atomic metric belongs. For example, an atomic metric may belong to multiple data domains. This ensures that names and logic are normalized and themes are classified in a standard manner.

View and manage the atomic metric list

Dataphin allows you to view the list of atomic metrics created in a specific project. You can view the name, creator, and publishing status of each atomic metric. You can search for a specific atomic metric in the list, and then modify, unpublish, or delete the atomic metric.

View and manage atomic metrics

- Native atomic metrics

To ensure standard creation of atomic metrics, Dataphin allows you to define an atomic metric that is only based on a logical table and its model. You can select a source table. Select a field from the snowflake or star schema that contains the source table, and define the computing logic for the atomic metric based on the field.

- Composite metrics

Composite metrics are calculated based on multiple atomic metrics. For example, you can obtain the payment conversion rate metric based on several atomic metrics. You can first define two atomic metrics: the number of customers who pay for orders and the number of customers who place orders. The payment conversion rate metric is expressed as the number of customers who place orders divided by the number of customers who pay for orders.

32.4.6.4. Business filters

An atomic metric is the standardized definition of computing logic, and a business filter is the standardized definition of a query condition. Similar to an atomic metric, a business filter is unique within a business unit and has only one source logical table. The computing logic of a business filter is defined based on the fields of the source logical table model. This ensures that all statistical metrics are created in a unified and standard manner. The data domain of each logical table linked to the source logical table is retrieved to trace the data domains to which the business filter belongs. For example, a business filter may belong to multiple data domains. This ensures that names and logic are normalized and themes are classified in a standard manner.

View and manage the business filter list

Dataphin allows you to view the list of business filters created in a specific project. You can view the name, creator, and publishing status of each business filter. You can search for a specific business filter in the list, and then modify, unpublish, or delete the business filter.

View and manage business filters

To ensure the standard creation of business filters, you can only define a business filter based on the source logical table and the models associated with the table. You can select a source table. Select fields from the snowflake or star schema that contains the source table, and define the computing logic for the business filter based on the fields.

32.4.6.5. Derived metrics

Derived metrics are commonly used statistical metrics. To create derived metrics in a standard, regular, and clear manner, each derived metric is a calculation based on the following criteria:

- Atomic metric: statistical criteria, that is, the computing logic.
- Business filter: the scope of business to be measured. It is used to filter the records that comply to specific business rules.
- Statistical period: a period during which statistics are collected, for example, the last 1 or 30

days.

- **Granularity:** a statistical object or perspective that defines the level of data aggregation. It can be considered as a grouping condition for aggregation, that is, GROUP BY clauses in SQL statements. Granularity is a combination of dimensions. For example, if a derived metric is a seller's turnover in a province, the granularity is the combination of the seller and the region dimensions.

By combining the preceding parts, multiple derived metrics can be quickly created at a time while ensuring that the definitions and computing logic are clear without any duplication. This metric creation method is simple, available to all users, and does not require a high level of technical expertise. For example, business users can also complete metric creation. A derived metric is a concept that is based on the same level as a field. Each derived metric is unique and defined at the specified granularity level.

View and manage the derived metric list

Dataphin allows you to view the list of derived metrics created in a specific project. You can view the name, creator, and publishing status of each derived metric. You can search for a specific derived metric in the list, and then modify, unpublish, or delete the derived metric.

View and manage derived metrics

To standardize the creation of derived metrics, the scope and objects to be measured must be determined based on the statistical computing logic. Therefore, you must select an atomic metric and the granularity, statistical period, and business filter related to the atomic metric. Then, you can follow a standard process to create multiple derived metrics at the same time.

- **Select statistic granularity**

Granularity is a combination of dimensions. The dimensions in the selection box are all the dimensions linked to the logical table model where the atomic metric resides. This provides a strong basis for useful and practical calculation at the specified granularity.

- **Select a statistical period**

Dataphin provides default statistical periods and also allows you to add custom statistical periods on the Planning page.

- **Select a business filter**

A business filter is a constraint or filter condition defined for a logical table. You may need to obtain a group or a type of business data. For example, you want to define metrics for the same statistical scope and computing logic for different statistical periods, such as the last one day, seven days, and 30 days. Dataphin allows you to define multiple levels of granularity, statistical periods, and business filters. These elements can be combined to create multiple derived metrics. This ensures standard metric creation and improves development efficiency.

32.4.7. Modeling

32.4.7.1. Overview

Dataphin provides systematic modeling and development functions to deeply implement the data warehouse theory. You can create business dimensions and business processes by using a top-down approach, and then enrich dimension tables, fact tables, aggregate tables, and the application data store layer. This process allows you to produce standardized data assets, which provides you with layered business data. The data standardization process can also optimize computation and storage.

32.4.7.2. Logical dimension tables

A logical dimension table contains details about a dimension. Dataphin allows you to view and manage the list of created logical dimension tables, and to view and modify a specific logical dimension table.

View and manage the logical dimension table list

Dataphin allows you to view the list of logical dimension tables created in a specific project. You can view the name, creator, creation time, and publishing status of each table. You can search for a specific logical dimension table in the list, and then modify or delete the table.

You can view details about a specific logical dimension table. You can view the primary key, dimension-associated fields, and attributes in the logical table. You can also view the star schema and snowflake schema containing this logical dimension table. If an inheritance relationship is defined, you can view settings of the parent and child dimension tables. You can also publish a logical dimension table after unlocking and modifying the table, zoom in or zoom out from the canvas, and view the published version. Dataphin provides a graphical user interface for you to configure a specific logical dimension table. You can define dimension attributes, associate dimensions with the table, and add child dimensions. Other supported operations include configuring the logical table conversion settings, viewing table details, and customizing the scheduling policy for the logical table conversion task.

32.4.7.3. Logical fact tables

Dataphin supports using logical fact tables to model a specific business process (such as placing an order and paying for a commodity) or a state measure (such as account balance and inventory). A logical fact table is created in an optimized schema that is similar to a snowflake schema. Apart from measures and dimension-associated fields, this type of schema allows a fact table to also contain fact attributes. This reduces the complexity of the model design and makes it more user-friendly.

View and manage the logical fact table list

Dataphin allows you to view the list of logical fact tables created in a specific project. You can view the name, creator, and publishing status of each table. You can search for a specific logical fact table in the list, and then modify, unpublish, or delete the table.

View and modify logical fact tables

Dataphin allows you to view details about a specific logical fact table model on a form-based interface. You can view the dimension-associated fields, measures, fact attributes in the logical fact table, and the logical dimension tables associated with the table. You can also publish a logical fact table after unlocking and modifying the table, zoom in or zoom out from the canvas, and view the published version. Dataphin provides a graphical user interface for you to configure a specific logical fact table model. The configurations include defining basic information, primary key, and fields, configuring the logical table conversion settings, and customizing the scheduling policy for the logical table conversion task.

32.4.7.4. Logical aggregate tables

The logical aggregate table model is an important data warehouse model. It contains two types of elements. The first type of element refers to various statistical values used to describe statistic granularity. The statistical values form a derived metric, for example, the sales in the last seven days. Granularity is a combination of several dimensions, such as the province and the product line dimensions. The second type of element refers to the attributes of the dimensions that constitute granularity. Examples of attributes are province name, product line name, product line level.

View and manage the logical aggregate table list

Dataphin allows you to view the list of logical aggregate tables created in a specific project. You can view the name and creation time of each table. You can search for a specific logical aggregate table in the list, and then modify, unpublish, or delete the table.

View and modify logical aggregate tables

A logical aggregate table can be created by aggregating the derived metrics defined following a standard process. You can also associate the logical aggregate table with fields of physical tables generated by code tasks.

32.4.7.5. Coding automation

After a logical dimension table, logical fact table, or logical aggregate table is published, Dataphin automatically designs the corresponding physical model, generates code and tasks to produce required data. Multiple tasks are usually generated to convert a logical table to a physical model. If you want to view the task running logic, go to the Scheduling page.

32.4.8. Coding

32.4.8.1. Overview

Coding is an important data development method. This method can be used to achieve the same goal as building data models on graphical user interfaces. Dataphin allows you to edit scripts by using the coding method supported by your computing engine. You can submit the scripts to the scheduling system, which schedules the code tasks to produce data. You can also view historical versions of each code task. Multiple types of scripts are supported, such as SQL, Shell, and MapReduce scripts. The requirements for coding and configuration vary by script type. The requirements include syntax requirements and requirements for scheduling configuration. After a script is submitted and published, Dataphin creates a code task to run and produce data. In a directed acyclic graph (DAG), a task is also called a node. Dataphin supports the following operations for code task management: create, view, modify, and delete code tasks, edit scripts, configure task scheduling policies, publish tasks, and manage task versions.

32.4.8.2. Code editor

The code editor provides an online code editing interface to complete data development tasks. It supports SQL, MapReduce, Spark, and Shell programming.

32.4.8.3. Task scheduling configuration and publishing

Scheduling configuration

You can configure the scheduling policy for one-time and recurring tasks. Tasks with a scheduling policy configured can be published. The system can check the integrity of task scheduling configurations. Only tasks with a complete scheduling configuration can be published. All published tasks are recurring tasks. You can choose **Scheduling > Recurring Tasks** and view the published recurring tasks in the left-side navigation pane.

Publish

Members of a project can publish tasks if they have required permissions. Only a scheduling configuration with complete parameter settings, valid dependencies, and no circular dependencies can be published to create tasks. This guarantees that stable and orderly data production can be completed on schedule.

32.4.8.4. Code management

Dataphin supports various code operations to facilitate code file management and use. You can create, delete, update, rename, and view code files, and place code files in specific folders to categorize the code files.

Manage files

Dataphin allows you to edit, delete, unpublish, and rename each code file. You can also view the publishing status, creator, and creation time of each code file. This facilitates easy creation, clear display, and systematic management of code files.

Manage folders

When there are many code files, sort them in different folders to save and display these files in an orderly manner. You can create, rename, and delete folders, and move historical and new code files to specified folders for better management. Dataphin also supports hierarchical folder structures.

32.4.8.5. Collaborative programming

Manage node versions

Dataphin allows you to view historical task node versions. You can view the version number, submitter, submission time, and description. You can also view the code of each version to identify differences in code. Dataphin supports multiple node types, including MaxCompute_SQL, MaxCompute MR, and Shell.

Collaborative development

To achieve more efficient development by allowing collaboration between multiple developers, Dataphin provides a script locking mechanism, which prevents conflicts during collaborative development. This mechanism ensures that a line of code can only be edited by one user at a time. A user can steal the lock of another user to obtain the script editing permission. The user whose lock is stolen can obtain editing permission again by stealing the lock.

32.4.9. Resource and function management

32.4.9.1. Overview

Resource and function management assists code development. Data developers can upload local resources and configure task nodes for calling these resources to meet specific data processing requirements. These developers can also complete common data processing by using the built-in functions in the programming language supported by the computing engine. If a data logic (such as data conversion in compliance with a business logic) requires frequent processing and this cannot be achieved with the built-in functions, developers can define custom functions based on self-uploaded resources.

32.4.9.2. Resource management

Dataphin allows the data developers of a project to add, edit, and perform other operations on resources in the project. You can name and upload resource files, and then copy the resource file name to reference the resource file in the code. You can also delete unnecessary resource files.

Create and upload resource files

By default, the following types of local resource files can be uploaded: XLS, DOC, TXT, CSV, JAR, Python, and other types (such as ZIP packages). New file types that are different from these types can be quickly added in three days by using the standard interface. Each resource file name is unique within a project. The file name and resource package cannot be changed after a resource file is submitted. Only one resource file can be uploaded each time, and the type of the uploaded file must be the same as the selected file type.

Reference resources

You can copy and paste a resource file name to a specific position in the code editor, and write a statement to call this resource.

Update resources

You can update the description of managed resources and delete existing resources to save storage space.

32.4.9.3. Function management

You can search, use, and manage functions. Functions are classified into two types: built-in functions of the system and user defined functions based on uploaded resources such as JAR and Python packages. You can extend user defined functions by referencing standard functions.

Create user defined functions

Each user defined function must have a unique name within its project and cannot be renamed after being registered.

Reference functions

You can click Copy to copy the name of a built-in function or a user defined function, and then paste the name to a specific position in the code editor. Then, write a statement in the format of the sample command to process data.

Update functions

You can update user defined functions by editing related information (except name) and delete unnecessary user defined functions.

32.4.10. Scheduling and management

The scheduling center allows you to perform management work during the later stages of data development. The scheduling center provides the list of all data processing tasks and task instances. Data processing tasks include recurring and one-time tasks. Task instances include instances of the data processing tasks and retroactive data generation tasks. The scheduling center also provides the directed acyclic graphs (DAGs) showing task dependencies, task instance dependencies, and instance status. You can set the task running sequence, schedule specific nodes in a DAG, achieve optimal allocation of resources, and discover abnormal tasks. This ensures that all the tasks are run on schedule. The scheduling center also reports alerts during task running to ensure that errors can be handled in time. The scheduling center allows you to view and manage tasks.

Task list

You can view the lists of recurring and one-time tasks created in a specific project and the DAGs showing task dependencies.

Recurring tasks

You can view the recurring task list, search for specific tasks, and view the dependencies of each task. You can switch between different projects to view and search for tasks in a specific project. You can search for tasks by task node name or task node ID. You can also filter the task nodes that you own and nodes published the current day. This helps narrow down the scope of tasks or find specific tasks that you want to manage.

One-time tasks

You can view the one-time task list, search for specific tasks, and view details of each task. You can switch between different projects to view and search for tasks in a specific project. You can search for tasks by task node name or task node ID. You can also filter the task nodes that you own and nodes published the current day. This helps narrow down the scope of tasks or find specific tasks that you want to manage.

Task instance management

You can view the lists of recurring, one-time, and retroactive data generation task instances created in a specific project while viewing details of each task instance.

Recurring task instances

You can view the instance list, search for specific instances, and view details of each instance. You can view the running status and details of each recurring task instance. The details include task node ID, node name, task owner, task start time, end time, and run duration. You can switch between different projects to view and search for task instances in a specific project. You can search for task instances by task node name or task node ID. You can also filter the task instances that you own, instances with errors, and incomplete instances. This helps narrow down the scope of instances or find specific instances that you want to manage.

One-time task instances

You can view the instance list, search for specific instances, and view details of each instance. You can view the running status and details of each one-time task instance. The details include task node ID, node name, task owner, task start time, end time, and run duration. You can switch between different projects to view and search for task instances in a specific project. You can search for task instances by task node name or task node ID. You can also filter the task instances that you own and instances that run the current day. This helps narrow down the scope of instances or find specific instances that you want to manage.

Retroactive data generation instances

You can view the list of created retroactive data generation task instances and details of each instance. The details include the data timestamp, status, and run duration. You can also view the node ID, node name, and owner of the task for which you generate retroactive data. Dataphin also supports search and filter for retroactive data generation task instances.

Logical tables

You can search for and view logical tables and their conversion tasks. You can also view the fields of each logical table. You can switch between a logical table task and a logical table task instance to view details. By default, the DAG on the right of the logical table task list shows all conversion task nodes of the current logical table and the dependencies between the nodes, including indirect dependencies. By default, the DAG on the right of the logical table task instance list shows all conversion task instances of the current logical table and their status. The status may be running, success or failed.

32.4.11. Metadata warehouse

Dataphin provides powerful metadata management capabilities. It can collect and extract metadata from MaxCompute, Hadoop, Hive, MySQL, PostgreSQL, and Oracle data sources. It supports real-time tracing of metadata in the preceding computing and storage engines, and builds a unified metadata model by extracting metadata from different types of storage engines. Dataphin supports the rapid enrichment of multiple types of metadata and provides diverse metadata that complies with unified standards. This provides a rich source of stable metadata to catalog and handle data.

The metadata warehouse is the core foundation of data asset management. We recommend that you ensure that the following items are available or guaranteed when building the metadata warehouse:

- **Metadata collection standard:** A unified data development standard is required to ensure the consistency of metadata for modeling, data table creation, and data lineage. This improves the availability of metadata for data retrieval and data services.
- **Metadata accuracy (up-to-date) and quality:** The metadata output time and quality must be guaranteed to improve the accuracy of the data in the data asset module and the efficiency of data retrieval performed by developers.
- **Metadata model system:** A unified public metadata model is used to ensure compatibility with various types of data and deliver a comprehensive data map service.

32.4.12. Data asset management

After data acquisition, integration, processing are complete, you can systematically manage data assets. Based on OneData and data assets methodologies, Dataphin designs the data use principle and provides core technologies, including metadata acquisition, extraction, and processing technologies. You can classify and manage data in the form of assets, monitor data quality, and optimize resources. This allows you to minimize the cost of data, obtain the maximum value from data, and use this value to benefit your business.

Data asset management is implemented by using a series of core technologies. The real-time event subscription service provides real-time metadata update for tables and tasks. The rules engine ensures efficient and accurate judgment of data governance rules and the creation of health scoring models. Dynamic log analysis supports analyzing numerous daily operational logs for production tasks and daily machine management logs. Graph computing supports the analysis and creation of data lineage. The Onelog data tracking technology ensures the consistency of metadata between the data production, service, and consumption phases. You can access metadata during each of the three phases. The metadata import and processing architecture (in the form of a plug-in) supports management for data from different computing and storage engines. This architecture provides a set of services including data collection, analysis, governance, application, and operation. It is developed by Alibaba and based on the extensive experience with mass data management. It covers the entire data lifecycle, including data creation, management, application, and destruction.

Based on the data catalog established through an analysis of enterprise data assets, the data map module provides a search engine and data profiling (both derived from user behavior data). This allows you to efficiently retrieve an enterprise's data assets.

Asset overview

Dataphin can display the structure of the enterprise data assets that are created based on OneData. Components in different shapes represent business entities, whereas lines of different styles represent business links between these entities. This helps to visualize the structure of the data for a business unit.

Asset map

An asset map summarizes the relationships between dimensions and business processes in a data domain of a business unit to show the composition of your enterprise data. In addition, the asset map provides efficient, fast, and accurate data search and exploration based on your self-initiated behaviors, such as searches, access history, and favorites.

32.4.13. Security management

32.4.13.1. Overview

The wide use of big data services makes data security an important issue. In China, the Cyber Security Law of the People's Republic of China was implemented on June 1, 2017. The Cyber Security Law encourages the development of network data security precautions and utilization technologies. EU General Data Protection Regulation (GDPR) was enacted on May 25, 2018. It aims to enhance the protection of data such as personal information. Dataphin focuses on intelligent development and management of data and places great importance on data security management. It provides comprehensive data security protection throughout the entire lifecycle (from data production to destruction). The protection is implemented by data access control, data isolation, and data security level classification. Other data protection methods include privacy compliance, data masking, and auditing of data security.

Data access control and data isolation require the highest priority in data security management. Dataphin provides management of data access permission requests, approvals, and lifecycle. It supports data isolation for multi-tenancy and field level access control, and offers a data access authorization model based on access control lists (ACLs).

Dataphin establishes a comprehensive data security guarantee system covering the entire lifecycle of data. This system provides technologies and management measures to protect data from the perspectives of data access behaviors, data content, and data environment. During big data development and management, Dataphin works with the Alibaba Cloud data security management system to provide an "available but invisible" environment for secure big data exchange. Dataphin also supports field level access control, control of permission request approval processes, and tracing and auditing of data use behaviors. All these combined methods help to guarantee data security during the storage, transfer, and use of big data.

Dataphin offers a hierarchical permission control system and a full range of management, covering the request, approval, assignment, handover, and authentication of data access permissions.

32.4.13.2. Permission types

Dataphin provides data access control based on user roles and resources. This allows you to use Dataphin and access data in a secure and controllable manner.

Role privileges

Dataphin provides account management mechanisms to obtain the super administrator and system members for centralized management of user operations. This controls the access methods of users at the platform level. Dataphin also allows you to control resource access at the organizational level by using project management. This access control method is role-based access control. It assigns specific roles a set of data resource permissions. Users acquire permissions through the roles to which the users are assigned.

Resource permissions

Dataphin provides a data access control mechanism to centrally manage user operations on project data resources. When each project is independently managed, and system members are isolated from resources, cross-project resource access can be controlled. This helps achieve data sharing by allowing users to use data of a specific project in another project without data migration.

32.4.13.3. Permission management

Permission requests

Data developers can find the required data table on the [Data Map](#) page and view the metadata details of this table. However, if they want to query data in the table, they must apply for permissions.

In a permission request process, Dataphin displays information about the requested data table by default, including the table type and the business unit to which the table belongs. Field metadata of the table is also displayed. Dataphin supports permission requests that follow the principle of least privilege. Specifically, requests for field-level permissions are supported. Multiple options of permission validity period are provided. You can customize a date range or select 30 days, 90 days, 180 days, or 1 year as the validity period. You can describe the purposes for which you intend to use the requested permissions. The approver can determine whether to grant you the permissions based on the description.

Request management

Dataphin allows you to view your requests and the status of the requests. You can click **Details** to view details of a request and click **Cancel** to cancel a request. After your request is approved, you can view your permission details, including the accessible fields.

Permission approval

After a permission request is submitted, the system randomly assigns the ticket to an administrator of the project to which the requested data table belongs. The administrator needs to approve the request. Approvers can view details about the submitted requests on the **My Approvals** tab and decide whether to approve or reject the request.

Permission handover

Users must hand over their permissions before shifting to another position or leaving the company. This ensures that related data and data production tasks can be handed over to appropriate staff. On the **My Permissions** page, you can click **Revoke** to hand over your permissions to the project administrator. Then, Dataphin reclaims the permission.

32.4.14. Ad hoc query

Dataphin supports high-performance ad hoc queries based on the OneService engine. Dataphin supports both traditional simple query and theme-based query methods, and enables code simplicity and fast query.

Syntax

- Dataphin supports offline queries on all logical tables. The intelligent engine selects the optimal physical table based on factors such as the output time and query performance.
- Dataphin supports join queries based on snowflake schemas. This makes it simpler to write SQL queries.
- Dataphin supports queries on physical tables, logical tables, and combinations of physical tables and logical tables.
- Dataphin supports multiple computing engines (each with unique syntax), such as MaxCompute SQL and Hive SQL.
- Dataphin provides intelligent code completion, precompilation, and beautification for SQL statements.
- Dataphin can manage permissions and authenticate users for access to fields in a logical or physical table.

Query implementation

You can enter any query statements in a query script. The script editor provides intelligent prompts based on the input content, quickly locates the required data table or field, and verifies the validity of the script syntax.

32.5. Scenarios

A retail group plans to launch a marketing program for members on New Year's Eve and wants to invite a celebrity for a promotional event. For this purpose, its business team needs to analyze the members' reaction to promotional offers for each quarter to determine the total amount of coupons to issue. In addition, the team also studies the members' celebrity preferences to determine whom to invite and the key commodities to promote.

The group has imported all transaction data and commercial-related music and video data into a MaxCompute database. Dataphin needs to calculate the promotion-based sales amount for each member and the celebrities each member follows. The group will then determine the activity plan.

32.6. Limits

None.

32.7. Concepts

Business unit

A business unit is used to define the name and business space of a data warehouse. If your business only involves retail, and the systems in the business are less isolated, you only need to build one business unit: retail.

Global object

A global object is a global concept. By defining global objects, you can universally reference the definitions of global concepts and ensure consistency throughout the entire system.

Project management

A project is a physical space division that allows users to isolate developers from resources. After setting a name for a project, you can start data modeling and development in the project.

Physical data source

You can register your physical databases to Dataphin. Physical databases serve as the underlying data sources for projects and data synchronization.

Dimension

A dimension is a statistical object. It is an entity that actually exists. By creating a dimension, you can standardize your business entities (or master data) during architectural design to ensure that they are unique.

Business process

A business process is a collection of all events in a business activity. By creating a business process, you can standardize a type of transaction event in business to ensure that it is unique.

Logical dimension table

One logical dimension table corresponds to one dimension. A logical dimension table stores dimension attributes that describe facts. Logical dimension tables are used to extract details of common objects from business data.

Logical fact table

A logical fact table models a specific business process and provides detailed information of transactions in the business process. Logical fact tables are used to extract details of common transactions from business data.

Atomic metric and business filter

An atomic metric and business filter are the computing logic and attributive limitation commonly used in business. An atomic metric and business filter are expressions formulated based on fields in a logical table. These are reusable common data elements extracted to calculate aggregate data.

Derived metric

A derived metric is a commonly used statistical metric. It is used to aggregate the data of an object group in a specific range during a time period. Therefore, a derived metric is defined by the time period (statistical period), statistical object (statistic granularity), range (business filter), and calculation method (atomic metric). After specifying the preceding elements, you need to set a name and a display name for the derived metric to complete metric creation. For example, you can define the promotion-based sales amount for each member in a quarter (Q1, Q2, Q3, and Q4) as a derived metric. You can also add other conditions as required.

33.Elasticsearch

33.1. What is Elasticsearch?

Elasticsearch is a distributed search and data analysis tool based on Lucene. It provides a distributed multi-tenant search engine which supports full-text search and is based on RESTful Web APIs. This Java-based enterprise-class search engine complies with the Apache license terms and conditions. Elasticsearch is designed for real-time search and is stable, reliable, fast, and easy to install and use.

Elasticsearch provides the open-source Elasticsearch V5.5.3 engine. It is designed for scenarios such as data search and analysis. Elasticsearch provides enterprise-class permission control, security monitoring and alarming, and automatic reporting based on the open-source Elasticsearch engine.

The default plug-ins provided by Elasticsearch include but are not limited to:

- **IK Analyzer:** an open-source Java-based lightweight Chinese language analysis kit. It is a popular plug-in for language analysis in the open-source community, and a distributed real-time analysis and research engine.
- **Smart Chinese Analysis plug-in:** the default Chinese language analysis tool in Lucene. It can be automatically scaled to hundreds of servers to process PBs of structured or unstructured data.
- **ICU Analysis plug-in:** a Lucene ICU analyzer. ICU is a set of stable, mature, powerful, and easy-to-use libraries, providing Unicode and globalization support for software applications.
- **Japanese (Kuromoji) Analysis plug-in:** a Japanese language analysis tool.
- **Stempel (Polish) Analysis plug-in:** a French language analysis tool.
- **Mapper Attachments Type plug-in:** an attachment-type plug-in which can parse files of different types into strings through the Tika library.

33.2. Benefits

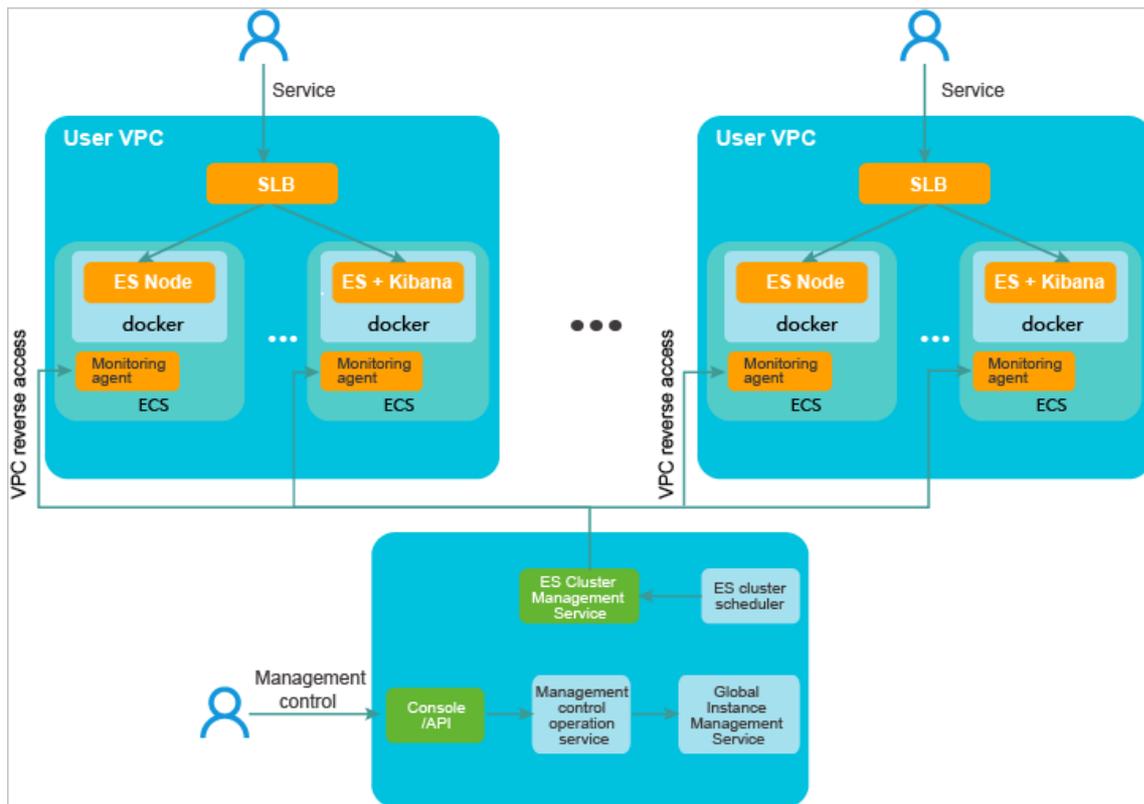
Elasticsearch has the following features and advantages.

- **Real-time retrieval and analysis**
Searches and analyzes PBs of data in real time, and responds to search and analysis in milliseconds.
- **Stability and reliability**
Uses Alibaba Cloud IaaS to support disaster recovery and fault tolerance mechanisms to provide stable and reliable data storage.
- **Easy deployment and maintenance**
Supports automatic deployment, and provides a perfect system monitoring module and zero-cost O&M.
- **Visual analysis**
Integrates with the Kibana module to provide visual data analysis and background management.
- **Chinese word segmentation**
Integrates with mainstream plug-ins by default, including the third-party IK Chinese word segmentation plug-ins.

- Auto scaling
 - Supports auto scaling to hundreds of servers and scaling of server hardware configurations.
- Technical support
 - Alibaba Cloud provides 24/7 technical support, as well as product documentation and training services.

33.3. Architecture

The following figure shows the Elasticsearch architecture when an Elasticsearch instance is created.



You can use Apsara Stack Management Console or API to submit configurations to create an Elasticsearch instance.

1. Select ECS as the Elasticsearch node and specify the disk capacity.
2. The control service obtains the relevant instance and disk configurations from the ECS instance, records the configurations to the database, and submits the data to the global instance management service.
3. The instance management service converts the configurations into the configuration file of the Elasticsearch cluster based on the request type, and submits the configuration file to the Elasticsearch cluster management service.
4. The Elasticsearch cluster management service, an offline processing system, executes the corresponding task state machine based on the request type, until the task reaches the final state.

Take instance creation for example. The Elasticsearch cluster management service tags the ECS instance, attaches the instance to your VPC, configures the SLB instance, and hosts it to the cluster scheduler. The cluster scheduler pulls up the Elasticsearch and Kibana processes on the ECS instance.

Elasticsearch and Kibana processes run in the ECS instance through a container. The monitoring agent (an independent process) collects the metrics and sends the metrics through Log Service to CloudMonitor for monitoring and alarming. Your instances are isolated through VPC. The control service implements VPC reverse access through port mapping to manage your Elasticsearch instances.

33.4. Features

- **Distributed search engine and data analysis engine**

Search: Elasticsearch provides Internet search engine services such as Google and retrieval in IT systems.

Data analysis: Elasticsearch carries out analysis on top 10 toothpaste sellers in the last seven days on an e-commerce website, and analysis on top 3 news blocks with the largest access volume in the last month on a news website.

- **Full-text retrieval, structured retrieval, and data analysis**

Full-text retrieval: Elasticsearch can search for commodities whose names contain the word toothpaste.

Structured retrieval: Elasticsearch can search for commodities classified as daily chemical products.

Data analysis: Elasticsearch carries out analysis on how many commodities each commodity category contains.

- **Quasi-real-time data processing**

Data distribution: Elasticsearch automatically stores a large amount of data on multiple servers.

Data processing: Elasticsearch can process and retrieve a large amount of data from multiple servers.

Quasi-real-time processing: Elasticsearch can perform data search and analysis in seconds.

33.5. Scenarios

Elasticsearch can be used in the following scenarios.

- To provide powerful search functions. It supports general search modes, similar to those on Google.
- Search for logs or transaction data. It helps analyze business trends, system bottlenecks, or running status and development, and collect logs.
- To provide alerting functions. The system can continuously query and analyze data. If the specified value is exceeded, an alert is issued.
- To perform business information analysis. It can easily locate key information from millions of data records.

Examples of Elasticsearch scenarios in and outside Mainland China:

- Outside Mainland China

- GitHub (open-source code management platform): Elasticsearch can be used to search through hundreds of billions of rows of code.
 - The Guardian: Elasticsearch can be used to analyze logs of user behaviors (clicking, browsing, favoriting, and commenting) and social network data (comments on a piece of news). It notifies the author of analysis results and reader opinions about the news.
 - Stack Overflow (an IT discussion forum): Users can submit IT-related issues, such as program errors, on the forum. Other forum members may discuss with you or provide a solution. You can paste the program error messages and search for possible answers through full-text retrieval.
 - Log data analysis: Logstash collects logs, and Elasticsearch conducts complex data analysis (This combination of Elasticsearch, Logstash, and Kibana is also known as ELK).
 - Commodity price monitoring website: You can set a price threshold for a commodity. When the commodity price is lower than the threshold, the system sends a message to you. For example, if you have subscribed to the toothpaste price monitoring service, when the price of Colgate toothpaste family package is less than RMB 50, the system sends a message to you.
- Mainland China
 - Site search (such as e-commerce, recruiting, and portal websites)
 - IT system search (such as OA, CRM, and ERP)
 - Data analysis

33.6. Limits

Limits on the disk capacity

- There must be sufficient space available to store a minimum of one replica.
- The overhead of indexes is generally 10% more than that of the source data (the index overhead of `_all` type is not calculated).
- Linux reserves 5% of the total disk space for critical process handling, system recovery, and disk fragment storage by default.
- 20% of the total disk space is reserved for internal overhead, segment consolidation, and log operations in Elasticsearch.
- At least 15% of the total disk space should be reserved for data security reasons.



Note

- Minimum total disk size = Source data × 3.4.

```
Total disk size = Source data × (1 + Number of replicas) × (1 + Index overhead) / (1 - Linux reserved space) / (1 - Elasticsearch overhead) / (1 - Security threshold)
= Source data × (1 + Number of replicas) × 1.7
= Sources data × 3.4
```

- We recommend that you do not enable the `_all` field unless it is required in your business.
- Indexes with this field enabled have a larger overhead. Alibaba Cloud test results and user feedback indicate that the preceding estimation should be doubled for use in this case.

```
Total disk size = Source data × (1 + Number of replicas) × 1.7 × (1 + 0.5)
= Source data × 5.1
```

Limits on cluster types

The cluster capacity is limited by the specifications of each node in Alibaba Cloud Elasticsearch. Based on Alibaba Cloud test results and user feedback, we recommend that you use the following guidelines to determine the appropriate specifications for your clusters.

Maximum number of nodes in a cluster = Number of CPUs per node × 5

The maximum workload of an Elasticsearch node varies depending on the following scenarios:

- Data acceleration and query aggregation

Maximum workload per node = Memory size per node (GB) × 10

- Logging and offline analysis

Maximum workload per node = Memory size per node (GB) × 50

- General scenarios

Maximum workload per node = Memory size per node (GB) × 30

Cluster types

Type	Maximum cluster nodes	Maximum memory size per node (query)	Maximum memory size per node (logging)	Maximum memory size per node (general)
2-core 4 GB	10	40 GB	200 GB	100 GB
2-core 8 GB	10	80 GB	400 GB	200 GB
4-core 16 GB	20	160 GB	800 GB	512 GB
8-core 32 GB	40	320 GB	1.5 TB	1 TB
16-core 64 GB	50	640 GB	2 TB	2 TB

Limits on the shard size and quantity

An appropriate shard plan is required for each index in the Elasticsearch cluster. Generally, a better policy can be found to replace the default 5-shard policy in Elasticsearch.

- For small Elasticsearch nodes, the size of each shard must be smaller than or equal to 30 GB. For large Elasticsearch nodes, the size of each shard must be smaller than or equal to 50 GB.
- For log analysis or extremely large indexes, the size of each shard must be smaller than or equal to 100 GB.
- The number of shards, including replicas, must be equal to the number of nodes or an integer multiple of the number of nodes.
- We recommend that you create a maximum of five shards for each index on a node.

Limits on resources

- Number of nodes: 2 to 50
- Disk size: 160 GB to 2048 GB
- Specifications:
 - `elasticsearch.sn2ne.xlarge` (4-core 16 GB)
 - `elasticsearch.sn2ne.2xlarge` (8-core 32 GB)
 - `elasticsearch.sn2ne.4xlarge` (16-core 64 GB)

33.7. Terms

Cluster

The Elasticsearch cluster. A cluster consists of one primary node and one or more secondary nodes. A node can be elected as the primary node of an Elasticsearch cluster. The concept of primary and secondary nodes exist only within the clusters. Elasticsearch uses a decentralized architecture. It does not have a central node. To external nodes, an Elasticsearch cluster is considered as a whole. Communicating with any node within an Elasticsearch cluster is the same as communicating with the Elasticsearch cluster.

Shard

A part of an index. Elasticsearch divides an index into multiple shards and distributes these shards among nodes to implement distributed search capabilities. The number of shards for an index must be specified when the index is created. After an index is created, you cannot change the number of shards for the index.

Replica

A copy of an index. You can create multiple index replicas to enhance the fault tolerance of the system. If a shard is damaged or lost, it can restore itself from a replica. Replicas also improve the search efficiency. Elasticsearch balances search requests through these replicas.

Recovery

The process of redistributing shards for a node, also called data redistribution. It guarantees the integrity of data when the node joins or leaves a cluster, or when the node recovers from a failure.

River

A data source in Elasticsearch. It can transfer data from external storage (such as databases) to Elasticsearch. It is an Elasticsearch service in the form of a plug-in. It reads data from the river and builds indexes in Elasticsearch. The officially supported river types include couchDB, RabbitMQ, Twitter, and Wikipedia.

Gateway

The device used to store snapshots of indexes. An Elasticsearch node stores all the indexes in its memory by default. When the node memory is full, the node stores persistent indexes in local disks. Index snapshots stored in a gateway can be restored upon a cluster reboot for fault recovery, which is faster than reading indexes from local disks. Elasticsearch supports multiple types of gateways, including local file system (default), distributed file system, Hadoop HDFS, and Amazon S3.

discovery.zen

The automatic node discovery mechanism. Elasticsearch is a peer to peer (P2P) system that uses broadcasting to discover nodes. Nodes communicate with each other through a multicast protocol. P2P interaction is also supported.

Transport

The process used for communication between nodes within a cluster, or between clusters and clients. Nodes communicate with each other over TCP by default. Elasticsearch also supports other transmission protocols, including JSON over HTTP, Thrift, Servlet, Memcached, and ZeroMQ.

34.DataHub

34.1. What is DataHub?

DataHub is a real-time data distribution platform designed to process streaming data. You can publish and subscribe to applications for streaming data in DataHub and distribute the data to other platforms. DataHub allows you to analyze streaming data and build applications based on the streaming data.

DataHub collects, stores, and processes streaming data from mobile devices, applications, website services, and sensors. You can use your own applications or Alibaba Cloud Realtime Compute to process streaming data in DataHub, such as real-time website access logs, application logs, and events. The processing results such as alerts and statistics presented in graphs and tables are updated in real time.

Based on the Apsara system of Alibaba Cloud, DataHub features high availability, low latency, high scalability, and high throughput. DataHub is seamlessly integrated with Realtime Compute, allowing you to use SQL to analyze streaming data.

DataHub also supports synchronizing streaming data to various Alibaba Cloud services such as MaxCompute and OSS.

34.2. Benefits

High throughput

You can write terabytes (TB) of data into a topic and up to 80 million records into a shard every day.

Real-time processing

DataHub makes it easy to collect and process various types of streaming data in real time so you can react quickly to new information.

Ease of use

- DataHub provides a variety of SDKs for C++, Java, Python, Ruby, and Go.
- In addition to SDKs, DataHub provides RESTful APIs so that you can manage DataHub by using existing protocols.
- You can use collection tools such as Fluentd, Logstash, and Oracle GoldenGate to write streaming data into DataHub.
- DataHub supports structured and unstructured data. You can write unstructured data to DataHub, or create a schema for the data before it is written into the system.

High availability

- The processing capacity of DataHub is automatically scaled out without affecting your services.
- DataHub automatically stores multiple copies of data.

Scalability

You can dynamically adjust the throughput of each topic. The maximum throughput of a topic is 256,000 records per second.

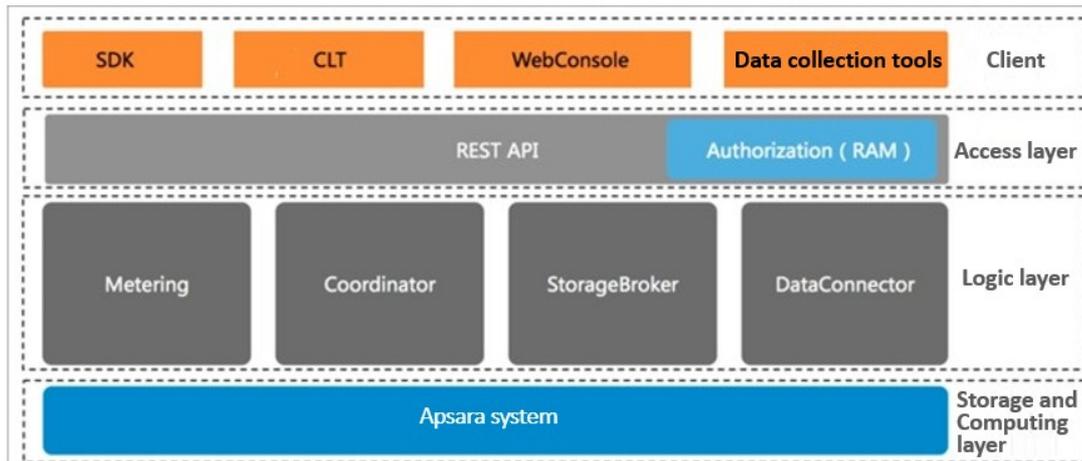
Data security

- DataHub provides enterprise-level security measures and isolates resources between users.
- It also provides several authentication and authorization methods, including whitelist configuration and RAM user management.

34.3. Architecture

Architecture shows the DataHub architecture.

Architecture



The architecture of DataHub consists of four layers: client, access layer, logical layer, and storage and scheduling layer.

Client

DataHub supports the following types of clients:

- SDKs: DataHub provides a variety of SDKs for C++, Java, Python, Ruby, and Go.
- Command line tool (CLT): You can run commands in Windows, Linux, or Mac operating systems to manage projects and topics.
- Console: In the console, you can manage projects and topics, create subscriptions, view shard details, monitor topic performance, and manage DataConnector.
- Data collection tools: Logstash, Fluentd, and Oracle GoldenGate (OGG).

Access layer

DataHub can be accessed through HTTP and HTTPS. DataHub supports RAM authorization and horizontal scaling of topic performance.

Logical layer

The logical layer handles the key features of DataHub, including project and topic management, data read and write, checkpoint-based data restoration, traffic statistics, and data archives.

Based on these key features, the logical layer is composed of the following modules:

StorageBroker, Metering, Coordinator, and DataConnector.

- **StorageBroker:** Enables the reading and writing of data in DataHub. Adopts the log file storage model of the Apsara Distributed File System, halving the read/write volume compared with the transfer of write-ahead logs. Stores three copies of data to ensure that no data is lost if a server fault occurs. Supports disaster recovery between data centers. Supports data write caching to ensure efficient consumption of real-time data. Supports independent read caching of historical data to enable concurrent consumption of the same data.
- **Metering:** Supports shard-level billing based on the consumption period.
- **Coordinator:** Supports checkpoint-based data restoration. Provides 150,000 QPS per node. Supports horizontal scaling of the processing capacity.
- **DataConnector:** Supports automatic data synchronization from DataHub to other Alibaba Cloud services, including MaxCompute, Object Storage Service (OSS), AnalyticDB, ApsaraDB RDS for MySQL, Table Store, and Elasticsearch.

Storage and scheduling layer

- **Storage:** Based on the log file storage model of the Apsara Distributed File System, DataHub supports append operations and solid state drive (SSD) storage. Data in each shard is stored in a separate file based on the recording time of the data.
- **Scheduling:** Based on the scheduling module of Job Scheduler, DataHub assigns shards to nodes based on the traffic that occurs on each shard. This ensures that the shards do not occupy the CPU or memory of Job Scheduler. The number of partitions on a single node has no upper limit. DataHub supports failovers within milliseconds and hot upgrades.

34.4. Features

34.4.1. Data queue

DataHub automatically generates a cursor for each record in a shard. The cursor is a unique sequence of numbers. You can improve the performance of a topic by increasing the number shards in the topic.

34.4.2. Checkpoint-based data restoration

DataHub supports saving checkpoints for subscribed applications in the system. You can restore data from any checkpoint you saved if your subscribed application fails.

34.4.3. Data synchronization

Data in DataHub is automatically synchronized to other Alibaba Cloud services.

DataConnector

You can create a DataConnector to synchronize DataHub data in real time or near real time to other Alibaba Cloud services, including MaxCompute, OSS, Elasticsearch, ApsaraDB RDS for MySQL, AnalyticDB, and Table Store.

You can configure the DataConnector so that the data you write to DataHub can be used in other cloud platforms. At-least-once semantics is applied in data synchronization. This ensures that no data is lost, but may result in duplicated records in the destination platform if an error occurs during the synchronization process.

Destination platforms

The following table describes the platforms to which DataHub records can be synchronized.

Destination platforms

Destination platform	Timeliness	Description
MaxCompute	Near real-time. Latency: 5 minutes.	The column names and data types in the source topic must be the same as those in MaxCompute. The MaxCompute table must have one or more corresponding partition columns.
OSS	Real-time	Records are synchronized to the specified bucket in OSS and are saved as CSV files.
Elasticsearch	Real-time	Records are synchronized to the specified index in Elasticsearch. Records may not be synchronized in the order of the recording time. If you want to synchronize data in the order of the recording time, you must write the records with the same partition key into the same shard.
ApsaraDB RDS for MySQL	Real-time	Records are synchronized to the specified table in ApsaraDB RDS for MySQL.
AnalyticDB	Real-time	Records are synchronized to the specified table in AnalyticDB.
Table Store	Real-time	Records are synchronized to the specified table in Table Store.

34.4.4. Scalability

The throughput of each topic can be scaled by splitting or merging shards.

You can adjust the number of shards in a topic according to the service load.

For example, if the topic throughput cannot handle a surge in the service load during Double 11, you can split existing shards to up to 256 to increase the throughput to 256 MB/s.

As the service load decreases after Double 11, you can reduce the number of shards as needed by performing the merge operation.

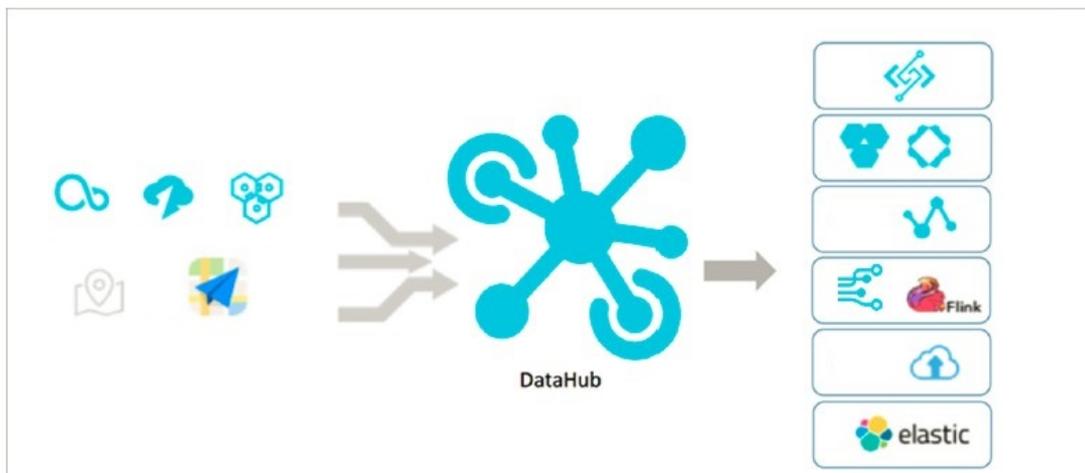
34.5. Scenarios

34.5.1. Overview

As a streaming data processing platform, DataHub can be used with various Alibaba Cloud products to provide one-stop data processing services.

34.5.2. Data uploading

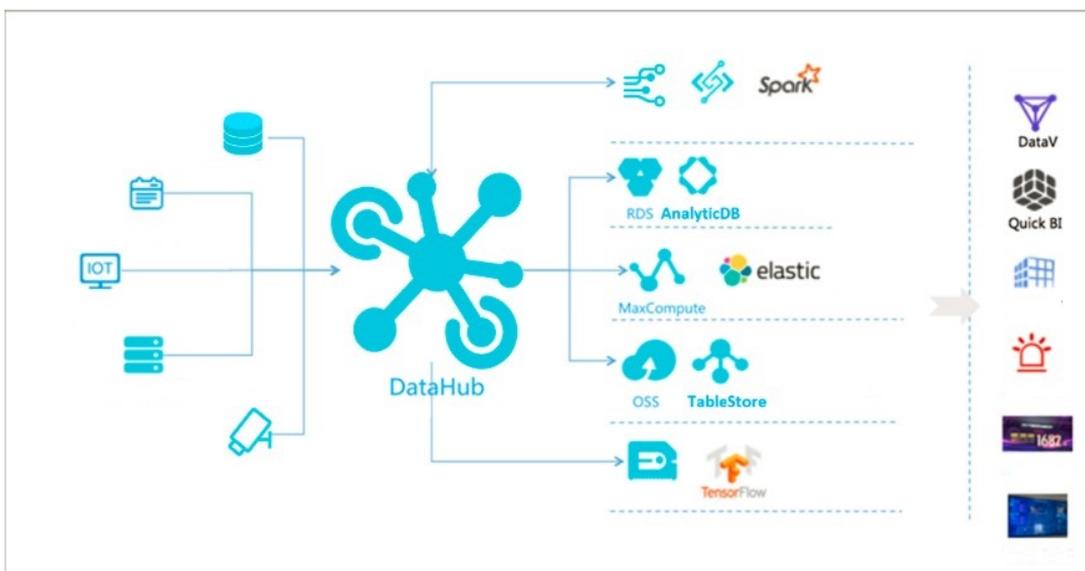
Data uploading



DataHub is connected to other Alibaba Cloud services, saving you the trouble of uploading the same data to different platforms.

34.5.3. Data collection

Data collection

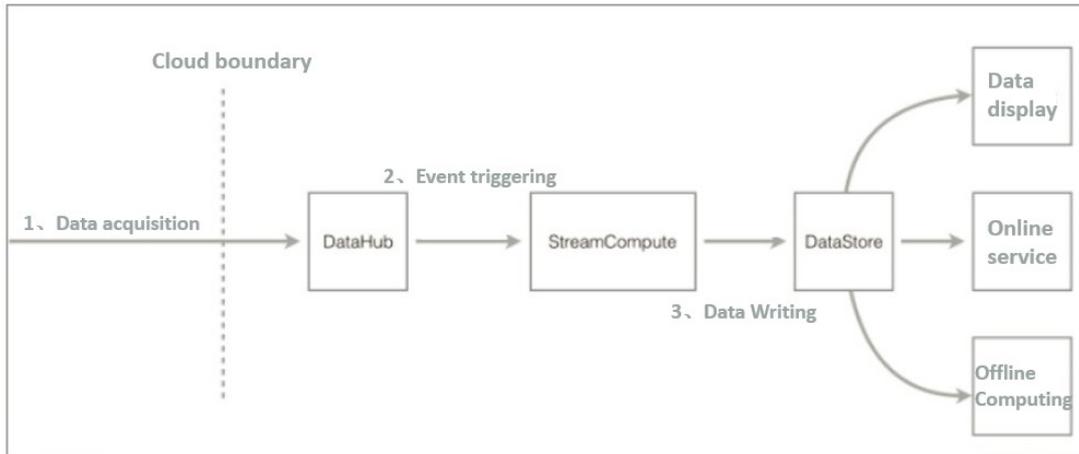


DataHub provides several types of data collection tools for you to write your data into DataHub. DataHub supports log collection from Logstash and Fluentd, and binary log collection from Data Transmission Service (DTS) and Oracle GoldenGate (OGG). DataHub also supports the collection of surveillance videos through GB28181.

34.5.4. Realtime Compute

Realtime Compute is a real-time computing engine of Alibaba Cloud, which allows you to use a language similar to SQL to analyze streaming data. Data can be transferred from DataHub to Realtime Compute or from Realtime Compute to DataHub.

DataHub and Realtime Compute



34.5.5. Data utilization

You can build an application to consume the data in DataHub, process the data in real time, and output the process results.

You can also use another application to process the streaming data output from the previous application to form a directed acyclic graph (DAG)-based data processing procedure.

34.5.6. Data archiving

You can create a DataConnector to periodically archive data in DataHub to MaxCompute.

34.6. Limits

Limits

Item	Range	Description
Active shards	(0,10]	Each topic can contain up to 10 active shards.
Shards	(0,512]	You can create up to 512 shards in each topic.
HTTP body size	≤ 4 MB	The HTTP request body size cannot exceed 4 MB.
String size	≤ 1 MB	The size of a string cannot exceed 1 MB.

Item	Range	Description
Merge and split operations on new shards	$\geq 5s$	You cannot merge a shard with another shard or split the shard in less than 5 seconds after it is created.
Queries per second (QPS)	$\leq 1,000$	The write QPS limit for each shard is 1,000. Multiple queries in one batch are considered one query.
Throughput	$\leq 1 \text{ MB/s}$	Each shard provides a throughput of up to 1 MB/s.
Projects	≤ 5	You can create up to 5 projects with each account.
Topics	≤ 20	You can create up to 20 topics in each project. Contact the administrator if you need to create more topics.
Time-to-live of records	[1,7]	The time-to-live of each record in the topic is from 1 to 7 days.

34.7. Terms

project

A project is an organizational unit in DataHub and contains one or more topics. DataHub projects and MaxCompute projects are independent of each other. Projects that you create in MaxCompute cannot be used in DataHub.

topic

The smallest unit for data subscription and publishing. You can use topics to distinguish different types of streaming data. For more information about projects and topics, see Limits in Product Introduction.

time-to-live of records

The period that each record can be retained in the topic. Unit: day. Minimum value: 1. Maximum value: 7.

shard

A shard in a topic. Shards ensure the concurrent data transmission of a topic. Each shard has a unique ID. A shard can be in a different status. For more information about shard status, see the following table. Each active shard consumes server resources. We recommended that you create shards as needed.

 Note

Shard status

Status	Description
Activating	All shards in a topic are in the Activating state when the topic is created. You cannot perform read or write operations on shards because they are being activated.
Active	Read and write operations are enabled when a shard is in the Active state.
Deactivating	A shard is in the Deactivating state when it is being split or merged with another shard. You cannot perform read or write operations on the shard because it is being deactivated.
Deactivated	A shard is in the Deactivated state when the split or merge operation is completed. The shard is read-only when it is in the Deactivated state.

hash key range

The range of hash key values for a shard, which is in [Starting hash key,Ending hash key) format. The hashing mechanism ensures that all records with the same partition key are written to the same shard.

merge

The operation that merges two adjacent shards. Two shards are considered adjacent if the hash key ranges for the two shards form a contiguous set with no gaps.

split

The operation that splits one shard into two adjacent shards.

record

A unit of data that is written into DataHub.

record type

The data type of records in a topic. Tuple and blob are supported. A tuple is a sequence of immutable objects. A blob is a chunk of binary data stored as a single entity.

Note

- The following data types are supported in a tuple topic.

Tuple data types

Type	Description	Value range
Bigint	An 8-byte signed integer. <div style="border: 1px solid #add8e6; padding: 5px; margin-top: 5px;">  Note Do not use the minimum value (-9223372036854775808) because this is a system reserved value. </div>	-9223372036854775807 to 9223372036854775807
String	A string. Only UTF-8 encoding is supported.	The size of a string cannot exceed 1 MB.
Boolean	One of two possible values.	Valid values: True and False, true and false, or 0 and 1.
Double	A double-precision floating-point number. It is 8 bytes in length.	$-1.0 \cdot 10^{308}$ to $1.0 \cdot 10^{308}$
Timestamp	A timestamp.	It is accurate to microseconds.

- In a blob topic, a chunk of binary data is stored as a record. Records written into DataHub are Base64 encoded.