阿里云
专有云Enterprise版

产品简介

产品版本：V3.0.0

文档版本：20171101

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

表 1: 格式约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **禁止:** 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **警告:** 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于警示信息、补充说明等，是用户必须了解的内容。	 **注意:** 导出的数据中包含敏感信息，请妥善保存。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 **说明:** 您也可以通过按**Ctrl** + **A**选中全部文件。
>	多级菜单递进。	**设置** > **网络** > **设置网络类型**
粗体	表示按键、菜单、页面名称等UI元素。	单击 **确定**。
courier字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

1 产品概述

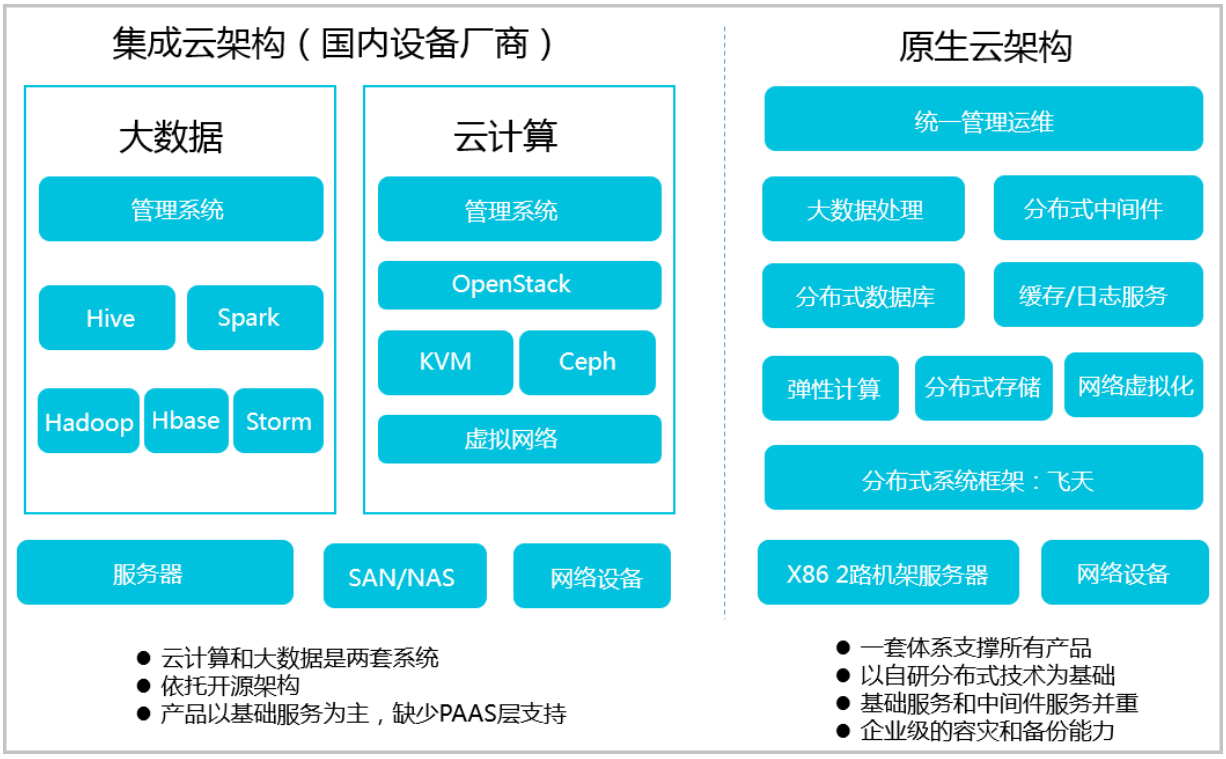
1.1 云计算架构的两个流派

云计算架构可以分为集成云架构和原生云架构。

在集成云架构中，云计算和大数据是两个体系，导致了账号、管理为两套体系，无法融合。云计算以计算虚拟化为主，基于传统的架构做突破，现在被OpenStack开源逐步成为主流。大数据以Hadoop体系为主，逐步加入了Spark，Storm等。开源架构好处是代码共享，容易建立生态；问题是各方利益牵扯进展缓慢，大的厂商往往都是发布一个开源的优化版本，但是优化的越多，反合的难度越大。另外，集成云架构缺乏PaaS层支持，云计算只有基础的计算、存储、网络资源，大数据只有离线计算、流计算、内存计算、列式存储等。

在原生云架构中，云计算和大数据是一套体系，因此账号，管理，运维也为一套融合的体系。原生云架构由互联网的开放架构演进而来，以分布式系统框架为基础，大数据和Web应用先行，然后扩展到各种基础服务。同时，原生云架构不开源，用户对系统掌控力强。另外，原生云架构在PaaS层应用较集成云架构丰富。

可以说，集成云架构是学院派、渐进式，原生云架构是实战派、变革式。



1.2 阿里云专有云

云计算是一种可通过互联网便捷访问，且IT资源（包括网络、服务器、存储、应用、服务）可定制、可共享、可按量付费的模式。这些资源能够快速部署、并且只需要很少的管理工作或很少的与服务提供商的交互。

云计算服务的模式一般认为有三大类：公共云、专有云和混合云。

公共云

公共云是云计算服务提供商为公众提供服务的云计算平台，任何个人或企业都可以通过授权接入该平台。公共云可以充分发挥云计算系统的规模经济效益，能够以低廉的价格，提供有吸引力的服务给最终用户，创造新的业务价值。

专有云

专有云是云计算服务提供商为企业在其内部建设的专有云计算系统。专有云将云基础设施与软硬件资源建立在防火墙内，以供机构或企业内各部门共享数据中心内的资源。管理者可能是组织本身，也可能是第三方；位置可能在组织内部，也可能在组织外部。

在企业IT架构实现向云演进的过程中，越来越多的企业出于安全合规、已有数据中心利旧、本地化体验等自身的建设要求，希望在自己的数据中心内也可以获得大规模云计算带来的服务体验。阿里云专有云解决方案诞生自阿里云公共云，通过帮助企业在自己的数据中心交付完整的可定制的阿里云软件解决方案，让您可以在本地运行同阿里云公共云提供的超大规模云计算和大数据产品相同的特性，为企业提供一致性的混合云体验，从而满足您按需获得IT资源，保持业务持续性的要求。

混合云

混合云是由两个或更多云环境组成云基础设施，这些系统各自独立，并由标准化的方式交换数据。混合云融合了公共云和专有云，既可以尽可能多地发挥云计算系统的规模经济效益，同时又可以保证数据安全性。那些不是很敏感的非关键业务可以由混合云中的公共云实现，而对那些安全性要求较高的应用则可以由专有云实现。

2 系统架构

2.1 架构介绍

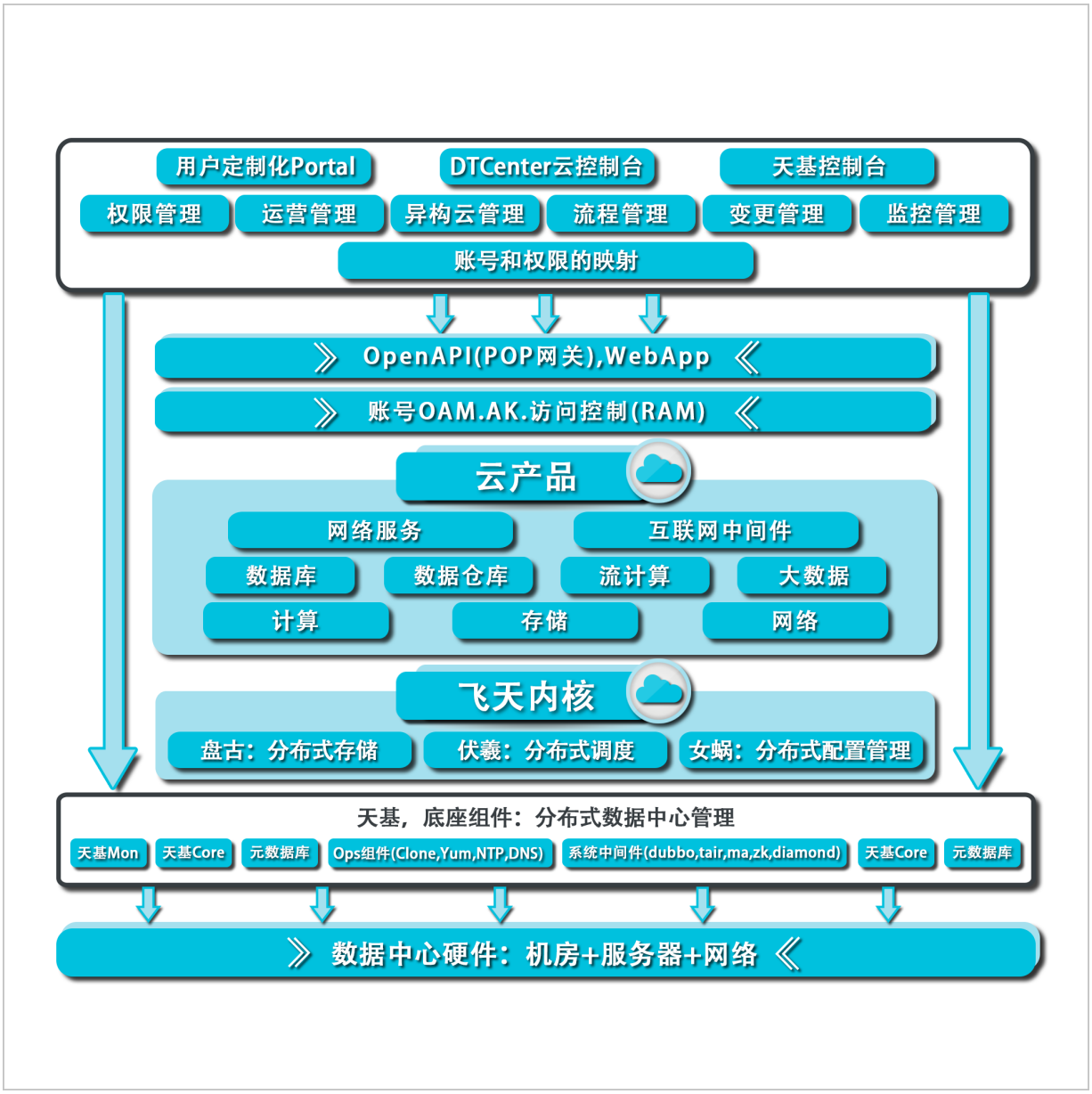
专有云基于多租户与开放API模型，为您提供企业级云安全架构，与统一的管理运维体验。

2.2 逻辑架构

专有云通过将物理服务器的计算和存储能力以及网络设备虚拟化成虚拟计算、分布式存储和软件定义网络，并在此基础上提供云数据库、大数据处理、分布式中间件服务，为您的应用系统提供IT基础服务的支撑能力，同时可以和您现有的账号体系，监控运维系统进行对接。

专有云逻辑架构有如下特点：

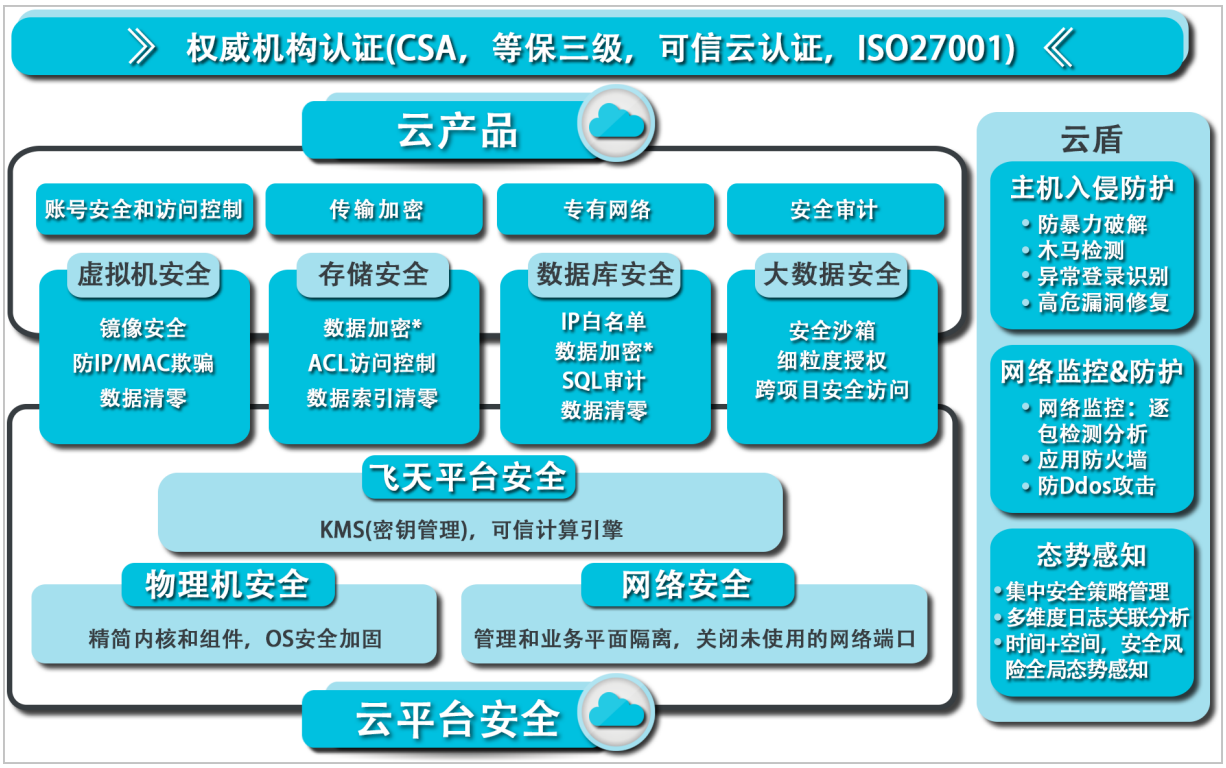
- 硬件基础是IDC+X86服务器+网络设备；
- 飞天内核（分布式引擎），基于飞天提供了各种云产品；
- 所有云产品都要求遵从一个统一的API框架，管理与运维（账号、授权、监控、日志）体系及安全体系；
- 保证所有产品遵从一致性的使用体验。



2.3 安全架构

云产品本身既有前台服务，又有后台系统，所以阿里云专有云安全架构分为两层：平台层和用户层。

平台层安全架构包含底层平台和云产品安全架构，强调对系统的控制力；用户层安全架构强调用户层面的安全策略。

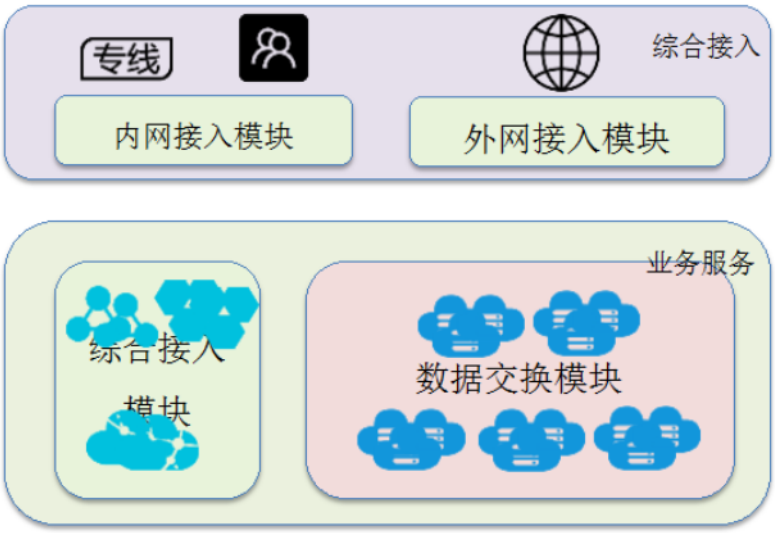

权威机构认证(CSA，等保三级，可信云认证，ISO27001)
云产品
账号安全和访问控制
传输加密
专有网络
安全审计
虚拟机安全
镜像安全
防IP/MAC欺骗
数据清零
存储安全
数据加密*
ACL访问控制
数据索引清零
数据库安全
IP白名单
数据加密*
SQL审计
数据清零
大数据安全
安全沙箱
细粒度授权
跨项目安全访问
飞天平台安全
KMS(密钥管理)，可信计算引擎
物理机安全
精简内核和组件，OS安全加固
网络安全
管理和业务平面隔离，关闭未使用的网络端口
云平台安全
云盾
主机入侵防护
• 防暴力破解
• 木马检测
• 异常登录识别
• 高危漏洞修复
网络监控&防护
• 网络监控：逐包检测分析
• 应用防火墙
• 防Ddos攻击
态势感知
• 集中安全策略管理
• 多维度日志关联分析
• 时间+空间，安全风险全局态势感知

3 网络架构

3.1 架构介绍

专有云网络架构定义了综合接入区和业务服务区两个逻辑区域。业务服务区作为云网络的核心部分为必选区域。综合接入区根据实际需要进行剪裁。

图 1: 网络架构逻辑分区



各层交换机的角色和作用如下表所示。

表 2: 角色定义

角色名称	所属模块	作用
ISW（互联交换机）	外网接入模块	出口交换机，互联ISP或用户网络骨干。
CSR（内网接入交换机）	外网接入模块	出口路由器，对于有大量路由表项要求以及高级三层应用的场景或要求非以太网接口类型可考虑采用CSR。
CSW （内网接入交换机）	内网接入模块	接入用户内网骨干，实现云网络内外部的路由分发交互，包括VPC专线接入。
DSW （分布层交换机）	数据交换模块	核心交换机，用于连接各个ASW接入交换机。
ASW （接入层交换机）	数据交换模块	接入交换机，接入云服务器，上行互联核心交换机DSW。

角色名称	所属模块	作用
LSW （综合接入交换机）	综合接入模块	云产品服务接入交换机，主要提供VPC和SLB等服务。

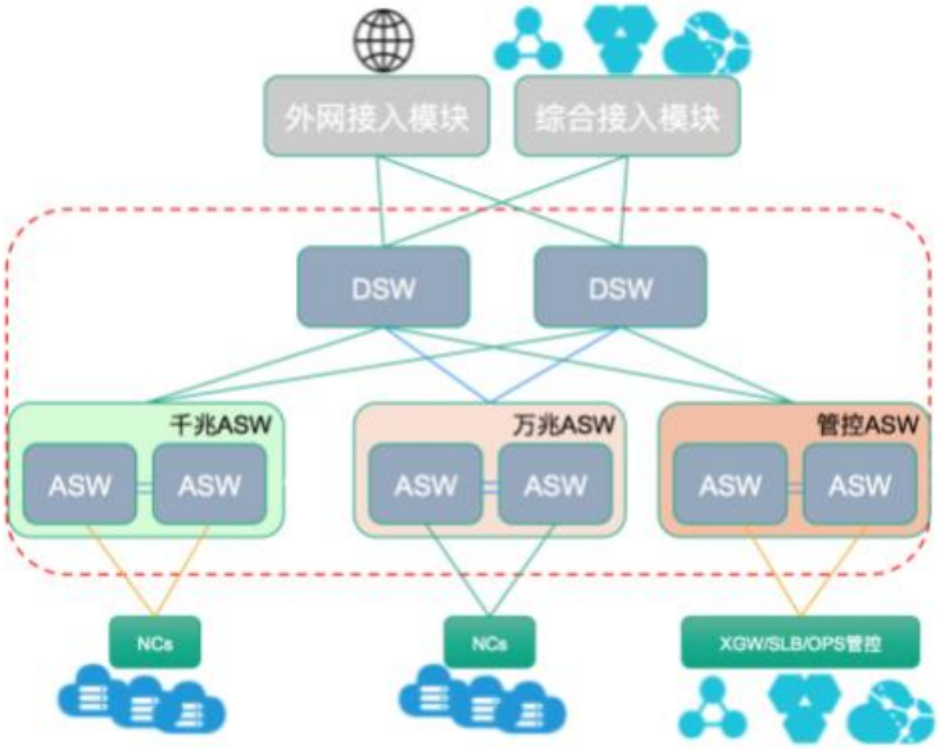
3.2 业务服务区

业务服务区提供所有云业务的网络承载，各个云业务系统的内部流量交互在该区域内完成，包括数据交换模块和综合服务模块。此部分是专有云网络的核心部分，不可裁剪。

数据交换模块

由DSW和ASW组成典型的二层CLOS架构。ASW两两堆叠作为叶子节点，此部分根据网络规模大小可选择不同适用范围的数据交换模型。所有云业务服务器上行至ASW堆叠设备， ASW和DSW之间通过EBGP互联，DSW之间相互没有连接。数据交换模块与其他模块之间通过EBGP互联。数据交换模块接收ISW发布的外网路由，并发布云产品地址网段到ISW。

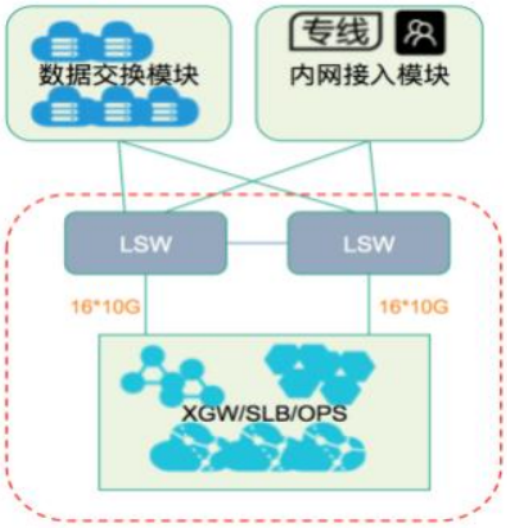
图 2: 数据库交换模块



综合服务模块

各类云产品服务器（XGW/SLB/OPS）分别与两台LSW互联，通过OSPF交换路由信息；两台LSW之间通过IBGP交互路由信息；LSW与DSW、CSW之间通过EBGP交换路由信息。

图 3: 综合服务模块



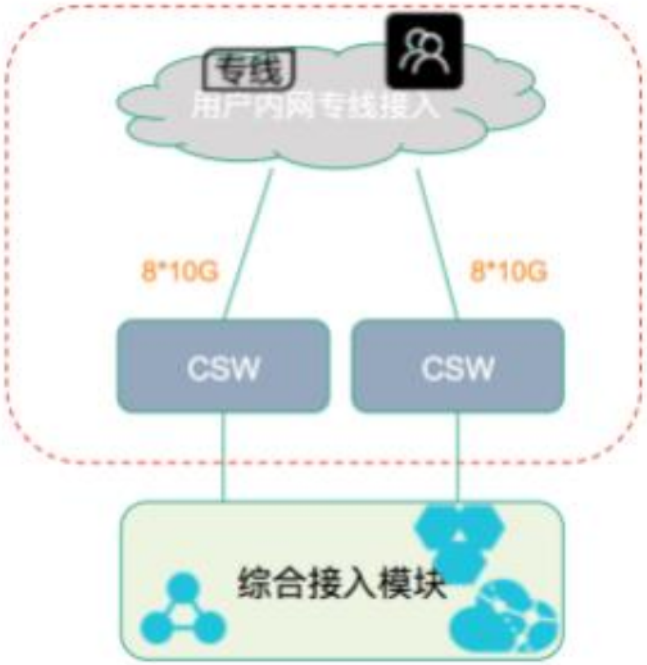
3.3 综合接入区

综合接入区作为业务服务区的外延网络，提供用户管理、用户自有网络、互联网访问专有云网络的通道，由内网接入模块和外网接入模块组成。此部分可根据用户实际部署需求进行裁剪。

内网接入模块

两台CSW为内部用户提供两类接入：VPC接入和普通云服务接入。VPC接入由CSW提供内部用户与VPC的映射关系，将内部用户分别导入各个VPC内，在CSW上，不同用户群保持相互隔离。普通云服务接入，CSW与综合服务模块通过EBGP互联，直接提供到业务服务区的所有资源访问。

图 4: 内网接入模块

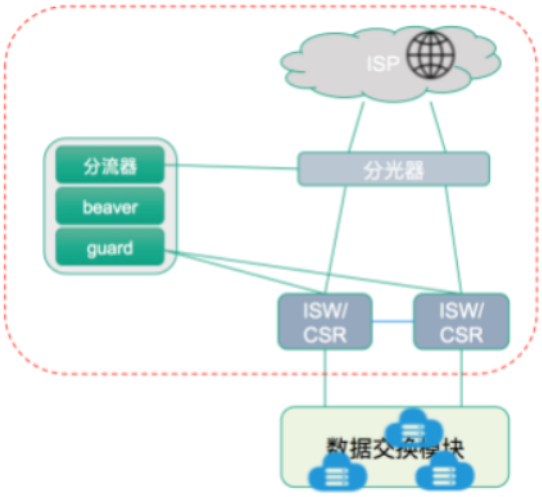


外网接入模块

由两台ISW组成，接入ISP或用户公网骨干，实现内外部的路由分发交互。两台ISW之间运行IBGP协议相互备份路由。上联到ISP或用户公网骨干，互联方式可根据实际情况采用静态路由或者EBGP协议，互联带宽根据用户的阿里云网络规模和用户骨干带宽设计定义。推荐做多运营商接入跑多线BGP，每运营商2*10GE。同时外网接入模块与数据交换模块之间通过EBGP协议交互路由。外网接入模块向数据交换模块发布相关外网路由，接收数据交换模块发出的云服务内部路由，实现云网络内部与外部交互。

此外，在外网接入模块会旁挂阿里云安全防护系统，外网访问云网络的流量通过分光器引流至beaver，beaver监测到攻击流量后通过云盾发布相应的路由将攻击流量引入云盾进行清洗，并将清洗后的流量回注。

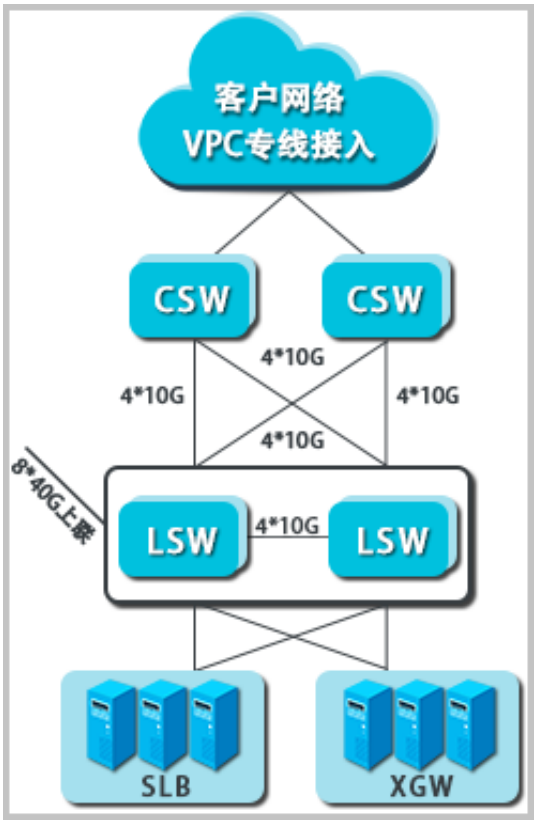
图 5: 外网接入模块



3.4 VPC专线接入方案

VPC专线接入部分的示意图如图 *6: VPC专线接入示意图*所示。

图 6: VPC专线接入示意图



xnet支持配置的CSW设备型号如下：

- H3C 6800-4C
- Huawei CE8860-4C-EI-B

- Huawei CE6855-48S6Q-HI (万兆盒式，48个10GE + 6个40GE）

说明:

选择Huawei CE6855-48S6Q-HI代替CE6850-48S4Q-EI作为CSW，是因为Huawei CE6855-48S6Q-HI有如下优势：

- 包转发率和FIB数量大于ASW和LSW的选型CE6851-48S6Q-HI。
- 支持SDN和VxLAN。新VPC专线方案中，MGW去掉后，CSW可以实现专线侧到XGW的VxLAN封装。
- 更高的包转发率、FIB、MAC，且多两个40GE口。

4 底座组件

专有云底座有三类组件，共同为云平台的部署和运维提供支撑。

表 3: 底座组件

组件		功能说明
Ops组件	Yum	安装软件包 软件源在初始装机阶段已经部署完毕，主要用于物理机上安装操作系统、部署飞天、ECS等专有云的应用软件包及其依赖的组件。
	Clone	机器克隆服务
	NTP	时钟源服务 部署在专有云的物理机，从标准NTP时间源同步，并授时给其他宿主机。
	DNS	域名解析服务 为专有云内部环境提供域名的正解和反解服务，在两个OPS 机器上各运行一个 bind 实例，通过keepalived 提供高可用服务，在一台失效的情况下，另外一台能够主动接管。
底座中间件	dubbo	分布式RPC服务
	tair	缓存服务
	mq	消息队列服务
	ZooKeeper	分布式协同
	Diamond	配置管理服务
	SchedulerX	定时任务服务
底座基础组件	天基	数据中心管理
	天基Mon	数据中心监控
	OTS-inner	表格存储服务
	SLS-inner	云平台日志服务
	元数据库	元数据库
	POP	云平台开放接口Open API
	OAM	账号系统

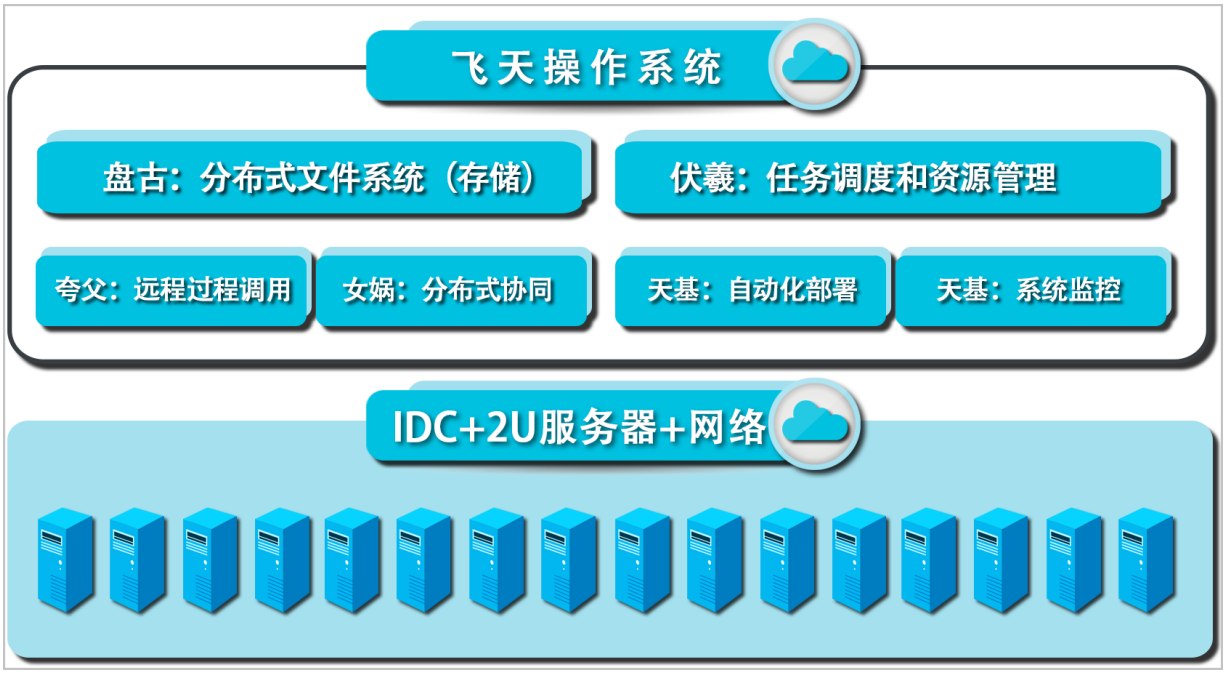
组件		功能说明
	RAM	认证授权系统
	WebApps	运维控制台支撑

5 产品优势

5.1 飞天大规模分布式计算系统内核

飞天大规模分布式计算内核，为上层的服务提供存储、计算和调度等方面的底层支持，其中包括：远程过程调用、安全管理、资源管理、分布式文件系统、任务调度和协调服务。通俗地讲，飞天就是把几千台通用的服务器整合成一台超级计算机。

图 7: 飞天内核系统架构



飞天平台内核包含的模块覆盖了以下主要的功能：

- 分布式系统底层服务

  提供分布式环境下所需要的协调服务、远程过程调用、安全管理和资源管理的服务。这些底层服务为上层的分布式文件系统、任务调度等模块提供支持。

- 分布式文件系统

  提供一个海量的、可靠的、可扩展的数据存储服务，将集群中各个节点的存储能力聚集起来，并能够自动屏蔽软硬件故障，为您提供不间断的数据访问服务；支持增量扩容和数据的自动平衡，提供类似于POSIX的用户空间文件访问API，支持随机读写和追加写的操作。

- 任务调度

为集群系统中的任务提供调度服务，同时支持强调响应速度的在线服务和强调处理数据吞吐量的离线任务；自动检测系统中故障和热点，通过错误重试、针对长尾作业并发备份作业等方式，保证作业稳定可靠地完成。

- 集群监控和部署

  对集群的状态和上层应用服务的运行状态和性能指标进行监控，对异常事件产生警报和记录；为运维人员提供整个飞天平台以及上层应用的部署和配置管理，支持在线集群扩容、缩容和应用服务的在线升级。

5.2 天基部署与管控系统

天基提供了云服务产品的统一部署、验证、授权和管控能力，为云服务提供基础性的支撑。天基框架中包含了部署框架、资源库、元数据库、云盾、认证授权、接口网管、日志服务、管控服务等模块。

- 部署框架为所有的云服务提供了统一的接入平台部署和服务间的依赖关系管理功能。
- 资源库保存了所有云服务和依赖组件的执行文件。
- 云盾为云服务提供Web攻击防护功能。
- 认证授权组件为云服务提供访问控制能力，支持多租户的隔离。
- 接口网关为云服务提供统一的API管理平台。
- 日志服务为云服务提供了日志存储、检索、获取等功能。
- 管控模块监控各云服务的基础健康状态，支撑云平台的运维体系。

5.3 久经考验的阿里云服务

弹性计算包含高可用的云服务器（简称ECS）、负载均衡（简称SLB）、专有网络（简称VPC）等云基础能力。

- ECS是一种简单高效、处理能力可弹性伸缩的计算服务，帮助客户快速构建更稳定、安全的应用。
- SLB是对多台云服务器进行流量分发的负载均衡服务。它可以消除单点故障提升应用系统可用性并提升应用系统服务能力。负载均衡通过设置虚拟服务地址（IP），将位于同一地域（Region）的多台云服务器资源虚拟成一个高性能、高可用的应用服务池。根据指定的方式，将来自客户端的网络请求分发到云服务器池中。
- VPC可帮助您基于阿里云构建出一个隔离的网络环境，使您可以完全掌控自己的虚拟网络，包括选择自有IP地址范围、划分网段、配置路由表和网关等。

存储与数据库包含了对象存储服务（简称OSS）、表格存储服务、关系型数据库（简称RDS）、分布式关系型数据库（简称DRDS）以及其它云数据库等多个服务。

- OSS可给用户提供基于RESTFul API的弹性扩展、高安全和高可靠的云存储服务。
- 表格存储是构建在阿里云飞天分布式系统之上的NoSQL数据存储服务，提供海量结构化数据的存储和实时访问。
- RDS是一种稳定可靠、可弹性伸缩的分布式数据库服务，基于全SSD盘高性能存储，支持MySQL、SQL Server、PostgreSQL和PPAS（高度兼容Oracle）引擎，默认部署主备架构且提供了容灾、备份、恢复、监控、迁移等方面的全套解决方案，彻底解决数据库运维的烦恼。
- DRDS是一种稳定、可靠、容量和服务能力可弹性伸缩的分布式关系型数据库服务。

阿里云专有云提供了大数据服务。

- 大数据计算服务（MaxCompute，原名ODPS）是一种TB/PB级数据仓库解决方案。MaxCompute向您提供了完善的数据导入方案以及多种经典的分布式计算模型，能够更快速的解决您海量数据计算问题，有效降低企业成本，并保障数据安全。
- 分析型数据库（AnalyticDB）是阿里巴巴自主研发的海量数据实时高并发在线分析（Realtime OLAP）云计算服务，使得您可以在毫秒级针对千亿级数据进行即时的多维分析透视和业务探索。
- 阿里云流计算是一套通用的流式计算（Stream Computing）平台，提供实时的流式数据计算服务。使用流计算，可有效降低数据的处理延迟，最大化实时您的业务数据。

阿里云专有云提供了应用中间件服务，支持各类客户应用程序。

- 企业级分布式应用服务（EDAS，Enterprise Distributed Application Service）是企业级互联网架构解决方案的核心产品，充分利用阿里云现有资源管理和服务体系，引入中间件成熟的整套分布式计算框架（包括分布式服务化框架、服务治理、运维管控、链路追踪和稳定性组件等），以应用为中心，帮助企业级客户轻松构建并托管分布式应用服务体系。
- 消息队列（Message Queue，简称MQ）是企业级互联网架构的核心服务，基于高可用分布式集群技术，为客户提供发布订阅、接入、管理、定时（延时）、监控报警等一套完整的高性能的消息服务。
- 容器服务（Container Service）是一种高性能可伸缩的容器管理服务，支持在一组阿里云云服务器上通过Docker容器来运行或编排应用。
- 资源编排（ROS）使用户能够通过简单的方法创建和管理一组有关联的阿里云的资源，对其提供标准化的模板描述方式并提供完整的生命周期管理，包括资源的申请、创建、销毁。

5.4 统一的运维管理系统

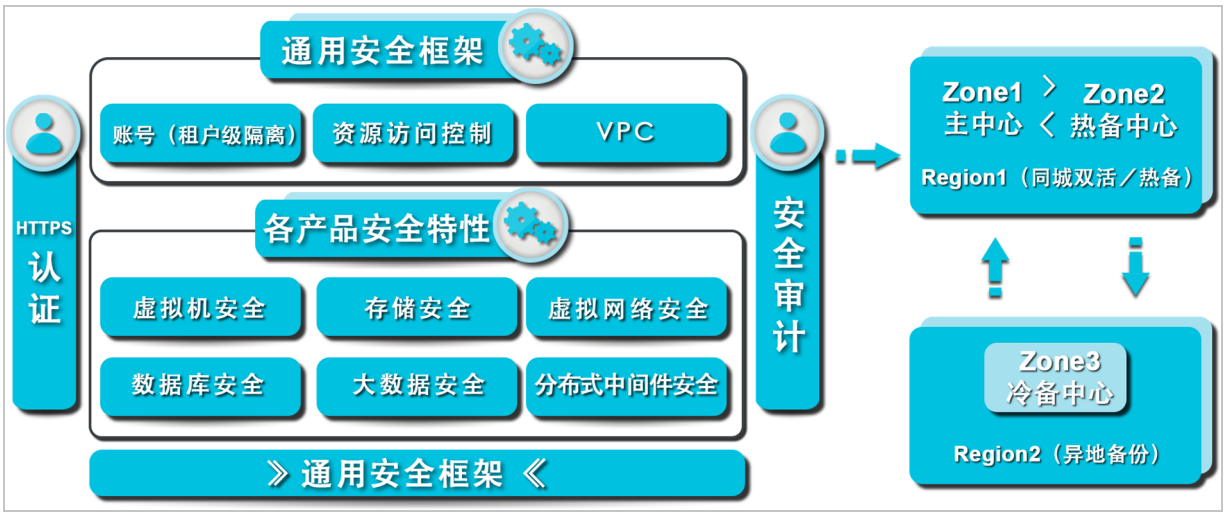
统一的运维管理系统包含云服务控制台和运维监控控制台。您可以通过控制台来进行账号管理、分配云服务资源、处理告警、升级系统、审计管理等操作。

图 8: 统一运维和运营管理Portal

5.5 高级别的安全与容灾能力

阿里云专有云提供了全方位的安全能力，保证您的访问和数据安全。所有控制台均需要通过HTTPS证书的方式访问。专有云提供完善的角色授权机制，保证多租户模式下，资源访问的安全可控。支持不同的安全角色，包括安全管理员、系统管理员、安全审计员。

图 9: 安全与容灾架构



阿里云专有云提供了两地三中心的容灾备份方案。同城（50KM）以内，可通过负载均衡实现无状态服务的双机房双活，以及结构化数据热备方案，实现秒级RTO和RPO。异地（>50KM）可提供数据冷备方案，根据网络情况，可实现分钟级或者小时级RTO和RPO。

5.6 开放的云服务接口

云服务通过OpenAPI平台，提供丰富的SDK包和RESTful API接口。您可以使用开放接口来灵活访问专有云提供的各种云服务。您还可以通过OpenAPI获取云平台的基础管控信息，将专有云平台接入到您统一的管控系统。

6 合规安全解决方案

2017年6月1日，《中华人民共和国网络安全法》正式实施，对等保合规作了明文规定。为了帮助企业用户快速满足等保合规的要求，阿里云整合云盾产品的技术优势，建立“等保合规生态”。联合阿里云在各地的合作测评机构、安全咨询合作厂商，为您提供一站式等保测评。完备的攻击防护、数据审计、加密、安全管理，助您快速省心地通过等保合规。

6.1 重点解读

网络与通信安全

条款解读

- 根据服务器角色和重要性，对网络进行安全域划分。
- 在内外网的安全域边界设置访问控制策略，并要求配置到具体的端口。
- 在网络边界处应当部署入侵防范手段，防御并记录入侵行为。
- 对网络中的用户行为日志和安全事件信息进行记录和审计。

应对策略

- 推荐使用阿里云的VPC和安全组对网络进行安全域划分并进行合理的访问控制。
- Web应用防火墙防范网络入侵。
- 使用日志功能对用户行为日志和安全事件进行记录分析和审计。
- 对于经常面临DDoS威胁系统，还可使用DDoS高防进行异常流量过滤和清洗。

设备与计算安全

条款解读

- 避免账号共享、记录和审计运维操作行为是最基本的安全要求。
- 必要的安全手段保证系统层安全，防范服务器入侵行为。

应对策略

- 对服务器和数据的操作行为进行审计，同时为每个运维人员建立独立的账号，避免账号共享。
- 使用安骑士对服务器进行完整的漏洞管理、基线检查和入侵防御。

应用和数据安全

条款解读

- 应用是具体业务的直接实现，不具有网络和系统相对标准化的特点。大部分应用本身的身份鉴别、访问控制和操作审计等功能，都难以用第三方产品来替代实现。
- 数据的完整性和保密性，除了在其他层面进行安全防护以外，加密是最为有效的方法。
- 数据的异地备份是等保三级区别于二级最重要的要求之一，是实现业务连续最基础的技术保障措施。

应对策略

- 在应用开发之初，就应当考虑应用本身的身份鉴别、访问控制和安全审计等功能。
- 对已经上线的系统，通过增加账号认证、用户权限区分和日志审计等功能设计满足等保要求。
- 数据的安全，使用HTTPS，确保数据在传输的过程中保持处于加密状态。
- 数据备份，推荐使用RDS的异地容灾实例自动实现数据备份，亦可以将数据库备份文件手工同步到阿里云其他地区的服务器。

安全管理策略

条款解读

- 安全策略、制度和管理层人员，是保证持续安全非常重要的基础。策略指导安全方向，制度明确安全流程，人员落实安全责任。
- 等保要求提供了一种方法论和最佳实践，安全可以按照等保的方法论进行持续的建设和管理。

应对策略

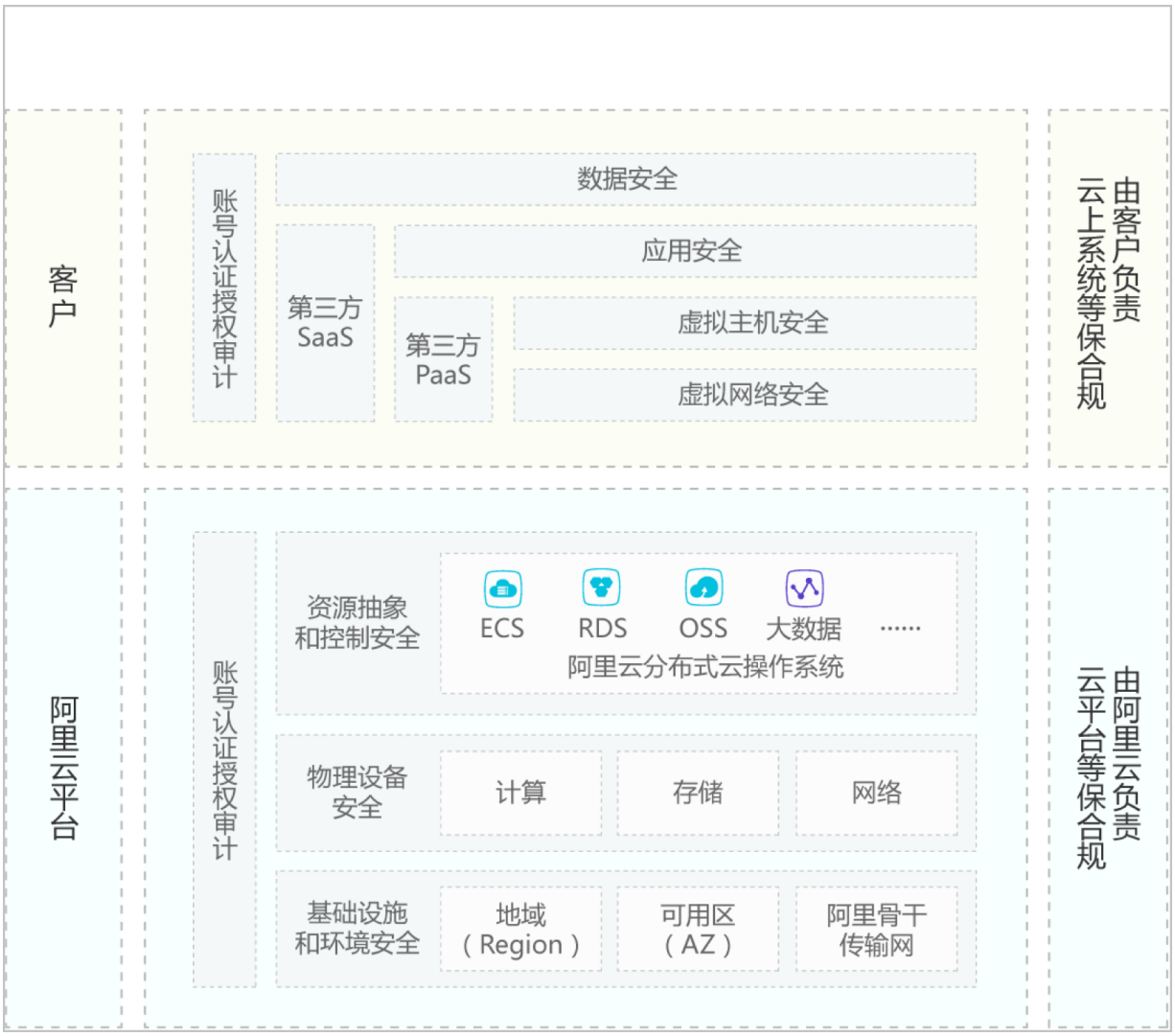
- 安全策略、制度和管理层人员，需要客户管理层根据本企业的实际情况，进行梳理、准备和落实，并形成专门的文件。
- 漏洞管理过程中需要用到的技术手段，推荐使用阿里云的安骑士服务，快速发现云上系统漏洞，及时处理。

6.2 云上等保合规

合规责任共担

阿里云平台与云上租户系统分别定级和测评。阿里云平台测评结论可供租户系统测评时复用。

图 10: 合规责任共担



阿里云可提供如下内容：

- 阿里云平台等保备案证明
- 阿里云测评报告关键页
- 阿里云云盾销售许可证
- 阿里云部分测评项说明

责任分担详解：

- 阿里云是全国唯一一家参与和通过云计算等保标准试点示范的云服务商。公共云、电子政务云通过等级保护三级备案和测评。金融云通过等级保护四级的备案和测评。
- 根据监管部门明确的结论复用原则，阿里云上的租户系统通过等级保护测评时，物理安全、部分网络安全和安全管理的结论可以复用，阿里云可提供说明。

- 阿里云平台完备的安全技术和管理架构，以及阿里云提供的云盾安全防护体系，更有利于租户通过等级保护测评。

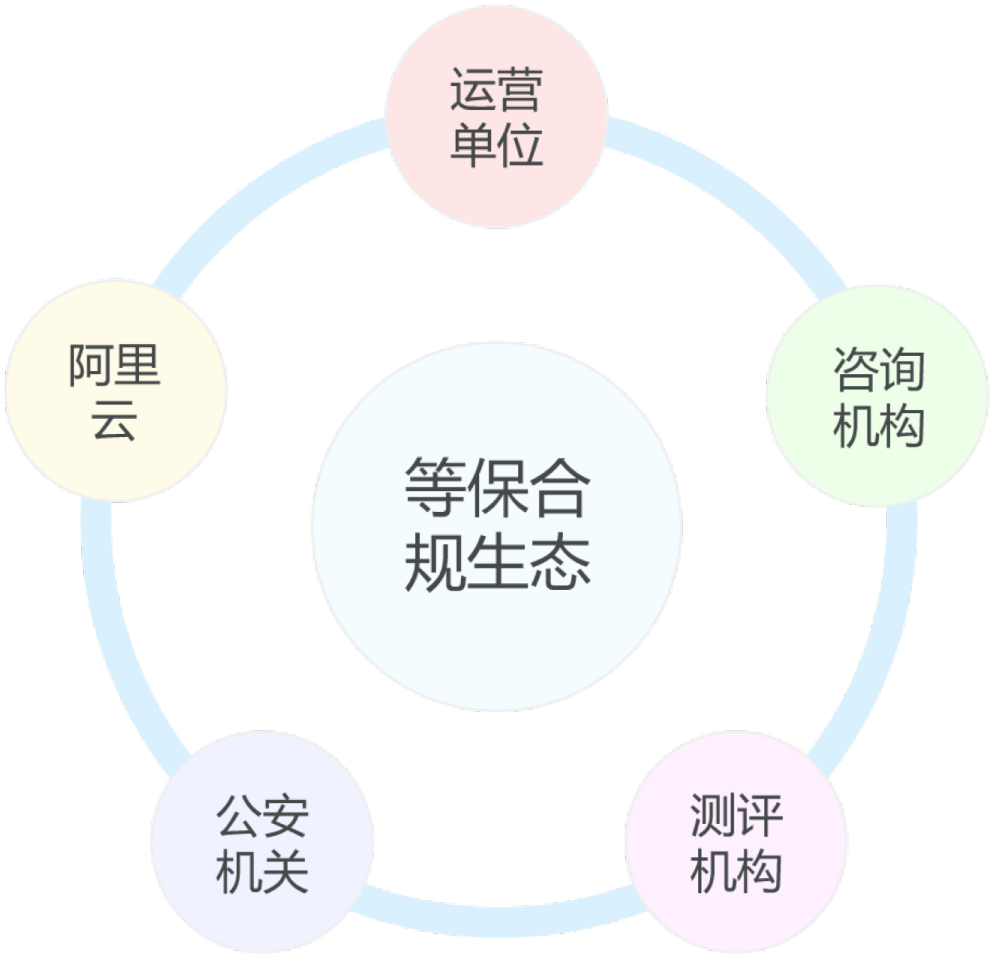
等保合规生态

云上等保现状：

- 大部分租户对等保不了解
- 不知道等保该如何入手
- 不善于与监管部门打交道
- 安全体系落后于业务发展

为了便于云上系统能够快速通过等保测评，阿里云通过建立“等保合规生态”，提供一站式等保合规解决方案。

图 11: 等保合规生态



等保工作分工：

- 阿里云：整合服务机构能力，并提供安全产品
- 咨询厂商：提供全流程技术支撑和咨询服务
- 测评机构：提供测评服务
- 公安机关：负责备案审核和监督检查

6.3 等保实施流程

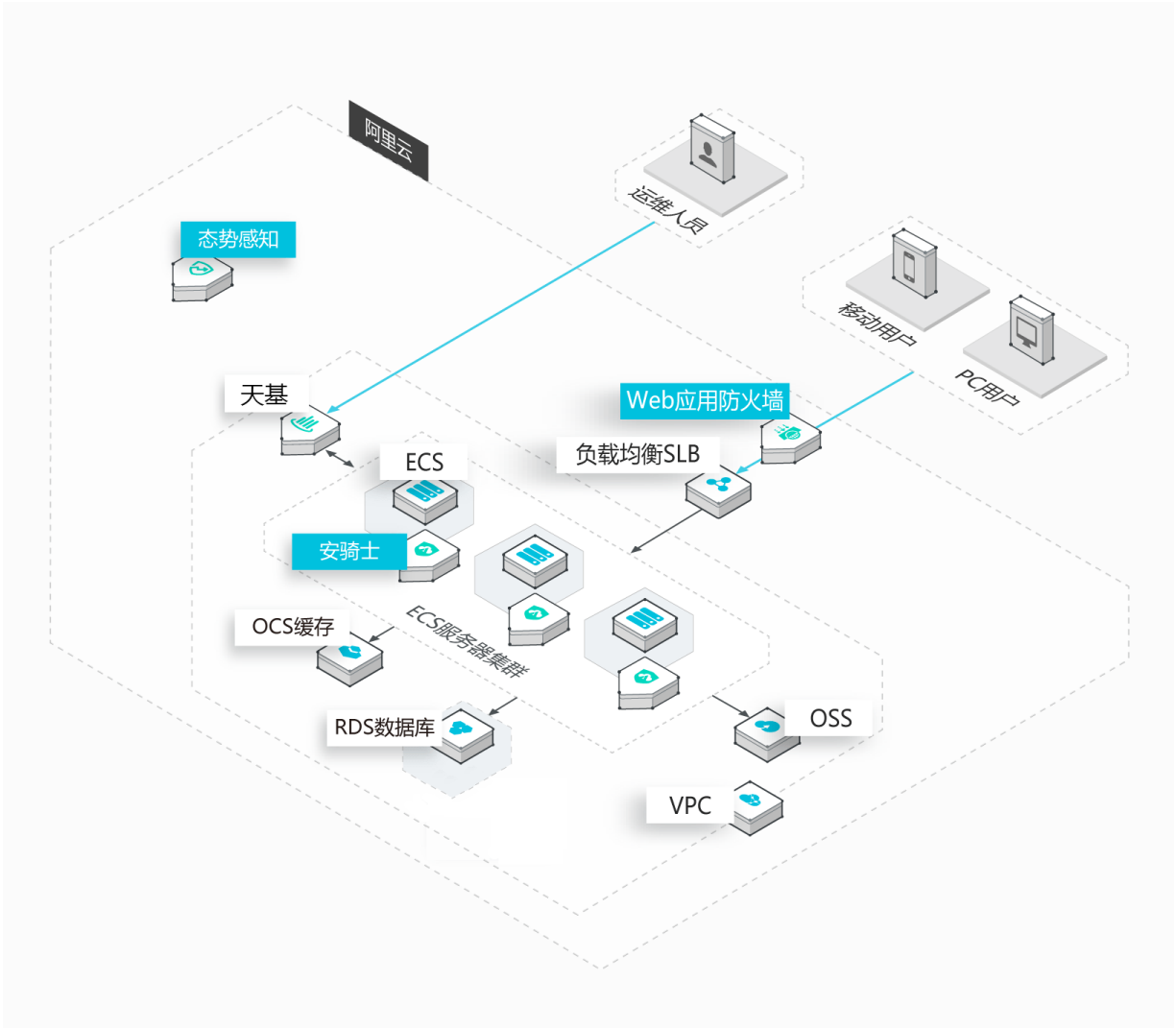
等保实施流程如图 *12:* 等保实施流程所示。

图 12: 等保实施流程

6.4 安全合规架构

快速接入云盾，快速完成安全整改。以最小的安全投入满足等保的基础合规技术要求。

图 13: 安全合规架构



等保基本要求：

- 物理和环境安全：包括机房供电、温湿度控制、防风防雨防雷措施等，可直接复用阿里云的测评结论。
- 网络和通信安全：包括网络架构、边界防护、访问控制、入侵防范、通信加密等。
- 设备和计算安全：包括入侵防范、恶意代码防范、身份鉴别、访问控制、集中管控和安全审计等。
- 应用和数据安全：包括安全审计、数据完整性和保密性。

6.5 方案优势

一站式等保测评服务

甄选并联合各地服务质量优异的咨询和测评机构，提供一站式、全流程合规，大大降低运营单位投入。

- 避免多点沟通和重复工作，减少运营单位投入
- 效率大大提高，最快两周完成测评
- 阿里云提供云上安全和合规最佳实践

完备的安全防护体系

通过完备的云盾安全架构，运营单位可以在阿里云上找到对应的产品，完成对不符合项的整改，全面满足等保要求。

7 云服务器ECS

7.1 产品概述

本节对云服务器的基本概念进行简单的介绍。

云服务器（Elastic Compute Service，简称ECS）是处理能力可弹性伸缩的计算服务，它的管理方式比物理服务器更简单高效。根据业务需要，您可以随时创建实例、扩容磁盘或释放任意多台云服务器实例。

云服务器ECS实例（以下简称ECS实例）是一个虚拟的计算环境，包含CPU、内存等最基础的计算组件，是云服务器呈献给每个用户的实际操作实体。ECS实例是云服务器最为核心的概念。其他的资源，比如磁盘、IP、镜像、快照等，只有与ECS实例结合后才能使用。

云服务器ECS示意图如图 *14: ECS 示意图*所示。

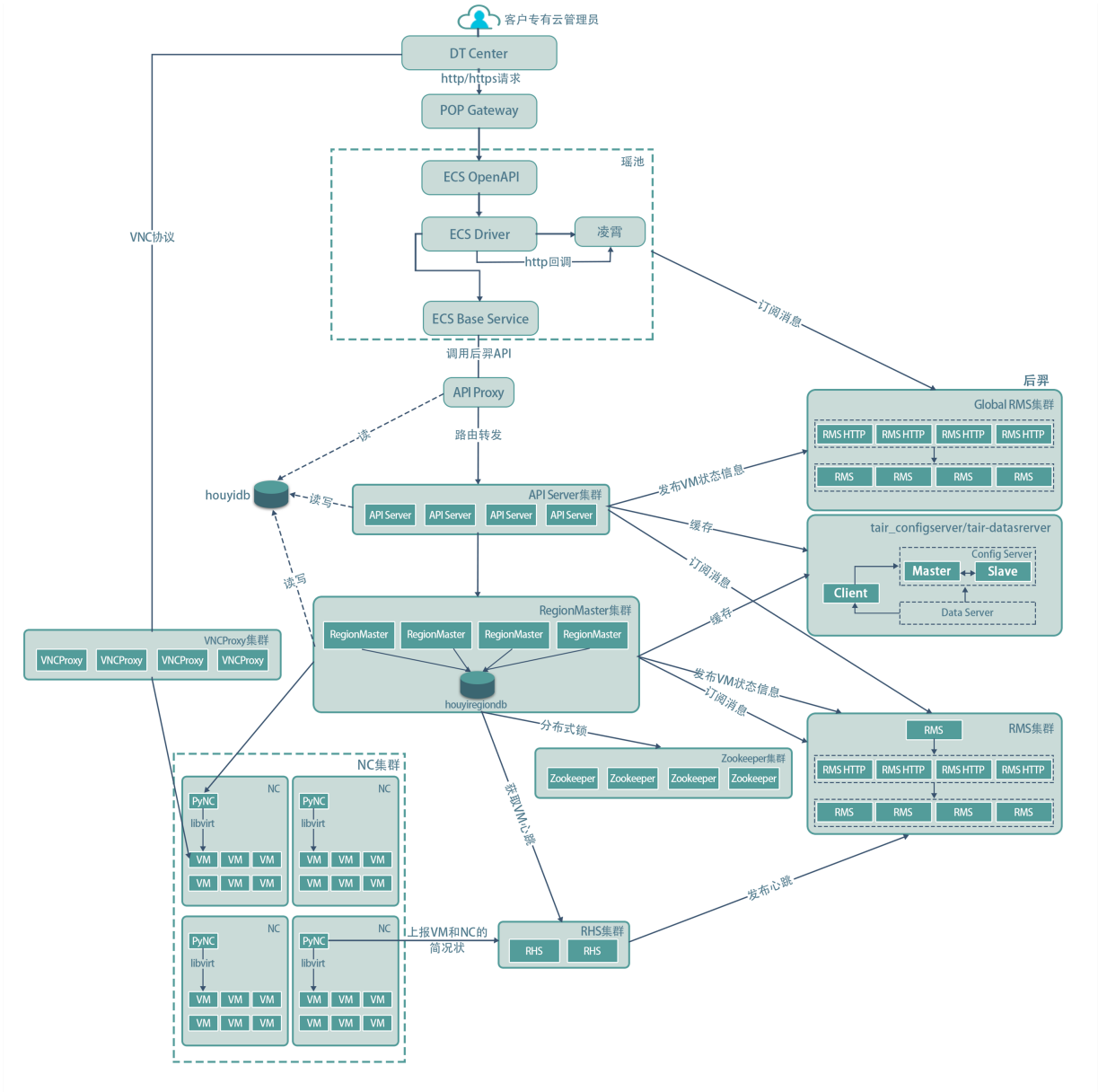
图 14: ECS 示意图

7.2 产品架构

本文通过示意图来展示 ECS 的整体架构图。

ECS 建立在阿里云自主研发的飞天分布式操作系统上。其单机虚拟化通过 Xen/KVM 实现。存储则依赖飞天的盘古（分布式存储）。ECS架构如图 *15: ECS架构图*所示。

图 15: ECS架构图



7.3 功能特性

7.3.1 实例

ECS实例是一个虚拟的计算环境，包含CPU、内存等最基础的计算组件，是云服务器呈献给每个用户的实际操作实体。ECS实例是云服务器最为核心的概念。其他的资源，比如磁盘、IP、镜像、快照等，只有与ECS实例结合后才具有使用意义。

7.3.1.1 实例规格族

本文将对不同的实例规格族进行介绍说明。

根据云服务器ECS实例规格的配置以及应用场景的不同，将实例规格按一定标准划分为若干实例规格族。

XN4

XN4是紧凑共享型实例，面向小型网站Web应用、小型数据库、开发、测试环境、代码存储服务器等场景。

XN4具有如下特点：

- CPU与Memory采用1:1配比。
- 采用2.5GHz主频的Intel Xeon E5-2686 v4 (Broadwell)处理器。
- 最新一代DDR4内存。
- 默认I/O优化。

N4

N4是通用共享型实例，面向中小型Web服务器、批量处理、分布式分析、广告服务和分布式分析等场景。

N4具有如下特点：

- CPU与Memory采用1:2配比。
- 采用2.5GHz主频的Intel Xeon E5-2686 v4 (Broadwell)处理器。
- 最新一代DDR4内存。
- 默认I/O优化。

MN4

MN4是均衡共享型实例，采用更大的CPU与内存配比，面向中型Web服务器、批量处理、分布式分析、广告服务、分布式分析、Hadoop集群等场景。

MN4具有如下特点：

- CPU与Memory采用1:4配比。
- 采用2.5GHz主频的Intel Xeon E5-2686 v4 (Broadwell)处理器。
- 最新一代DDR4内存。
- 默认I/O优化。

SN1

SN1是通用独享型实例，面向中大型Web服务器（高并发）、大型多人在线游戏（MMO）前端/数据分析和计算、利用CPU进行高精度编解码、渲染、基因计算等固定性能计算场景。

SN1具有如下特点：

- 计算性能稳定。
- CPU与Memory采用1:2配比。
- 采用2.5GHz主频的Intel Xeon E5-2686 v4 (Broadwell)或E5-2682 v3 (Haswell)处理器。
- 最新一代DDR4内存。
- 实例网络性能与计算规格同对应（实例计算规格越大则网络性能越强）。
- 默认I/O优化。

SN2

SN2是均衡独享型实例，采用更大的CPU与内存配比，面向中大型Web服务器（高并发）、大型多人在线游戏（MMO）前端、数据分析和计算、利用CPU进行高精度编解码、渲染、基因计算、Hadoop集群等固定性能计算场景。

SN2具有如下特点：

- 计算性能稳定。
- CPU与Memory采用1:4配比。
- 采用2.5GHz主频的Intel Xeon E5-2686 v4 (Broadwell)或E5-2682 v3 (Haswell)处理器。
- 最新一代DDR4内存。
- 实例网络性能与计算规格同对应（实例计算规格越大则网络性能越强）。
- 默认I/O优化。

SE1

SE1是内存独享型实例，采用更大的CPU与内存配比，面向Cache/Redis、搜索类、内存数据库、高I/O的数据库（如Oracle、MongoDB）、Hadoop集群、大量的数据处理加工等固定性能计算场景。

SE1具有如下特点：

- 计算性能稳定。
- CPU与Memory采用1:8配比。
- 采用2.5GHz主频的Intel Xeon E5-2686 v4 (Broadwell)或E5-2682 v3 (Haswell)处理器。
- 最新一代DDR4内存。
- 实例网络性能与计算规格同对应（实例计算规格越大则网络性能越强）。
- 默认I/O优化。

E4

E4是内存型实例，面向需要大量的内存操作、查找和计算的应用，如：做Cache/Redis、搜索类、内存数据库等。

E4具有如下特点：

- CPU与Memory采用1:8配比。
- 采用2.5GHz主频的Intel Xeon E5-2682 v4 (Broadwell)处理器。
- 默认I/O优化。

GN4

GN4是带GPU设备的实例，面向深度学习、渲染、视频处理、视频编解码等。

GN4具有如下特点：

- 采用 NVIDIA M40 GPU 计算卡。
- 多种 CPU 和 Memory 配比。
- 处理器：2.5 GHz 主频的 Intel Xeon E5-2682 v4 (Broadwell)。
- 实例网络性能与计算规格对应（规格越高网络性能强）。

GN5

GN5是带GPU设备的实例，面向深度学习、科学计算如计算流体动力学、计算金融学、基因组学研究、环境分析，高性能计算、渲染、多媒体编解码及其他服务器端 GPU 计算工作负载等。

GN5具有如下特点：

- 采用 NVIDIA P100 GPU 计算卡。
- 多种 CPU 和 Memory 配比。
- 处理器：2.5 GHz 主频的 Intel Xeon E5-2682 v4 (Broadwell)。
- 实例网络性能与计算规格对应（规格越高网络性能强）。

GN5I

GN5I是带GPU设备的实例，面向深度学习、多媒体编解码等服务器端 GPU 计算工作负载等。

GN5I具有如下特点：

- 采用 NVIDIA P4 GPU 计算卡。
- 处理器与内存配比为 1:4。
- 处理器：2.5 GHz 主频的 Intel Xeon E5-2682 v4 (Broadwell)。
- 实例网络性能与计算规格对应（规格越高网络性能强）。

GA1

GA1是带GPU设备的实例，面向渲染、多媒体编解码，机器学习、高性能计算、高性能数据库，其他需要强大并行浮点计算能力的服务器端业务等。

GA1具有如下特点：

- CPU 和 Memory 配比为 1:2.5。
- 实例网络性能与计算规格对应（规格越高网络性能强）。
- 采用 NVIDIA M40 GPU 计算卡。
- 处理器：2.5 GHz 主频的 Intel Xeon E5-2682 v4 (Broadwell)。

规格族之间的变配逻辑

- 三种共享型实例规格族（XN4、N4、MN4）之间及规格族内部可以变配。
- 三种独享型实例规格族（SN1、SN2、SE1）之间及规格族内部可以变配。

7.3.1.2 实例规格

本文针对各实例规格族中的具体规格进行罗列说明。

实例是能够为您的业务提供计算服务的最小单位，它以一定的规格来为您提供相应的计算能力。

ECS实例的规格定义了实例的CPU、内存的配置（包括CPU型号、主频等）这两个基本属性，但是只有同时配合磁盘种类、镜像和网络类型，才能唯一确定这台实例的具体服务形态。

表 4: 实例规格表

实例规格族	规格类型代码	CPU（Core）	内存（GB）
SN1	ecs.sn1.medium	2	4
	ecs.sn1.large	4	8
	ecs.sn1.xlarge	8	16
	ecs.sn1.3xlarge	16	32
	ecs.sn1.7xlarge	32	64
SN2	ecs.sn2.medium	2	8
	ecs.sn2.large	4	16
	ecs.sn2.xlarge	8	32
	ecs.sn2.3xlarge	16	64
	ecs.sn2.7xlarge	32	128
	ecs.sn2.13xlarge	56	224
N4	ecs.n4.small	1	2
	ecs.n4.large	2	4
	ecs.n4.xlarge	4	8
	ecs.n4.2xlarge	8	16
	ecs.n4.4xlarge	16	32
	ecs.n4.8xlarge	32	64
MN4	ecs.mn4.small	1	4
	ecs.mn4.large	2	8
	ecs.mn4.xlarge	4	16
	ecs.mn4.2xlarge	8	32
	ecs.mn4.4xlarge	16	64
	ecs.mn4.8xlarge	32	128
E4	ecs.e4.small	1	8
	ecs.e4.large	2	16
	ecs.e4.xlarge	4	32
	ecs.e4.2xlarge	8	64
	ecs.e4.4xlarge	16	128

实例规格族	规格类型代码	CPU（Core）	内存（GB）
XN4	ecs.xn4.small	1	1
GN4	ecs.gn4-c4g1.xlarge	4	30
	ecs.gn4-c8g1.2xlarge	8	60
	ecs.gn4.8xlarge	32	48
	ecs.gn4-c4g1.2xlarge	8	60
	ecs.gn4-c8g1.4xlarge	16	60
	ecs.gn4.14xlarge	56	96
GN5	ecs.gn5-c4g1.xlarge	4	30
	ecs.gn5-c8g1.2xlarge	8	60
	ecs.gn5-c4g1.2xlarge	8	60
	ecs.gn5-c8g1.4xlarge	16	120
	ecs.gn5-c28g1.7xlarge	28	112
	ecs.gn5-c8g1.8xlarge	32	240
	ecs.gn5-c28g1.14xlarge	56	224
	ecs.gn5-c8g1.14xlarge	56	480
GN5I	ecs.gn5i-c2g1.large	2	8
	ecs.gn5i-c4g1.xlarge	4	16
	ecs.gn5i-c8g1.2xlarge	8	32
	ecs.gn5i-c16g1.4xlarge	16	64
	ecs.gn5i-c28g1.14xlarge	56	224
GA1	ecs.ga1.2xlarge	8	20
	ecs.ga1.4xlarge	16	40
	ecs.ga1.8xlarge	32	80
	ecs.ga1.14xlarge	56	160

7.3.1.3 实例生命周期

本文详细说明了实例在整个生命周期中若干固有的状态。

实例的生命周期是从创建开始到最后释放。

实例固有状态

在这个生命周期中，实例有其固有的几个状态，如表所示。

表 5: 实例生命周期的固有状态

状态	状态属性	解释	API 的对应状态
准备中	中间状态	实例被创建后，在进入运行中之前的状态。如果长时间处于该状态，则说明出现异常。	Pending
已创建	稳定状态	实例已经创建完成，等待启动。如果长时间处于该状态，则说明出现异常。	Stopped
启动中	中间状态	实例在控制台或通过API，被重启、启动等操作后，在进入运行中之前的状态。如果长时间处于该状态，则说明出现异常。	Starting
运行中	稳定状态	实例正常运行状态。在这个状态的实例可以运行您的业务。	Running
停止中	中间状态	实例在控制台或通过API，被停止操作后，在进入已停止之前的状态。如果长时间处于该状态，则说明出现异常。	Stopping
已停止	稳定状态	实例被正常停止。在这个状态下的实例，不能对外提供业务。	Stopped
重新初始化中	中间状态	实例在控制台或通过API，被重新初始化系统盘或/和数据盘后，在进入运行中之前的状态。如果长时间处于该状态，则说明出现异常。	Stopped
更换系统盘中	中间状态	实例在控制台或通过API，被更换操作系统等操作后，在进入运行中之前的状态。如果长时间处于该状态，则说明出现异常。	Stopped
已过期	稳定状态	这个状态下的实例处于停止状态，不能对外提供业务。	Stopped

7.3.2 云盘

本文将对云盘的使用及分类作简要说明。

云服务器ECS的云盘既可以单独使用，又可以组合使用，以满足不同应用场景的需求。您可以根据自己要求，选择使用合适的数据存储选项。

云盘为ECS实例提供数据块级别的数据存储，采用三副本的分布式机制，为ECS实例提供99.9999999%的数据可靠性保证。

根据性能的不同，云盘又可以分为普通云盘、高效云盘和SSD云盘。

- 普通云盘面向低I/O负载的应用场景，为ECS实例提供数百IOPS的I/O性能。
- 高效云盘面向中度I/O负载的应用，为ECS实例提供最高3000随机IOPS的存储性能。
- SSD云盘为I/O密集型应用，提供稳定的高随机IOPS性能。

7.3.2.1 云盘参数对比

本文将对各类云盘的参数进行对比说明。

表 6: 云盘对比

块存储	SSD云盘	高效云盘	普通云盘
最大容量	32768GB	32768GB	2000GB
最大IOPS	20000	3000	数百
最大吞吐量	256MBps	80MBps	20~40MBps
性能计算公式	IOPS=min{30*容量，20000} 吞吐量=min{50+0.5*容量，256}MBps	IOPS=min{1000+6*容量，3000} 吞吐量=min{50+0.1*容量，80}MBps	不适用
访问时延	0.5~2ms	1~3ms	5~10ms
数据可靠性	99.9999999%	99.9999999%	99.9999999%
API 名称	cloud_ssd	cloud_efficiency	cloud
典型应用场景	• I/O密集型应用 • 中大型关系数据库 • NoSQL数据库	• 中小型数据库 • 大型开发测试 • Web服务器日志	不被经常访问或者低I/O负载的应用场景

7.3.2.2 磁盘性能测试方法

在进行下列测试前，请确保磁盘已经4K对齐。

- 测试随机写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

- 测试随机读IOPS：

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

- 测试写吞吐量：

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

- 测试读吞吐量：

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

上述测试时fio相关参数说明：

参数	说明
-direct=1	测试时忽略I/O缓存，数据直写。
-rw=randwrite	测试时的读写策略，可选值randread（随机读）、randwrite（随机写）、read（顺序读）、write（顺序写）、randrw（混合随机读写）。
-ioengine=libaio	测试方式使用libaio（Linux A I/O，异步I/O）。应用使用I/O通常有2种方式：同步和异步。 • 同步的I/O一次只能发出一个I/O请求，等待内核完成才返回。这样对于单个线程iodepth总是小于1，但是可以透过多个线程并发执行来解决。通常会用16-32根线程同时工作把iodepth塞满。 • 异步则通常使用libaio这样的方式一次提交一批I/O请求，然后等待一批的完成，减少交互的次数，会更有效率。

参数	说明
-bs=4k	单次I/O的块文件大小为4k。未指定该参数时的默认大小也是4k。
-size=1G	测试文件大小为1G。
-numjobs=1	测试线程数为1。
-runtime=1000	测试时间为1000 秒。如果未配置则持续将前述-size指定大小的文件，以每次-bs值为分块大小写完。
-group_reporting	测试结果显示模式，group_reporting表示汇总每个进程的统计信息，而非以不同job汇总展示信息。
-filename=iotest	测试时的输出文件路径和名称。测试完成后请记得删除相应文件，以免占用磁盘空间。
-name=Rand_Write_Testing	测试任务名称。

7.3.2.3 云盘的特点及应用场景

本文将分别对几种云盘的特点和应用场景进行介绍。

云盘分为高效云盘、SSD云盘和普通云盘。

高效云盘

产品特点

高效云盘采用固态硬盘与机械硬盘的混合介质作为存储介质，具备如下特点：

- I/O性能：最高提供3000随机读写IOPS、80MBps的吞吐性能。
- 数据可靠性：采用分布式三副本机制，提供99.9999999%的数据可靠性。
- 性能基准：
  - IOPS：起步1000 IOPS、每GB增加6个IOPS，最高3000。
  - 吞吐量：起步50MBps、每GB增加0.1MBps、最高80MBps。

    例如250GB的高效云盘，拥有2500的随机读写IOPS、75MBps的吞吐性能。
- 最大容量：单块高效云盘最大提供32768GB存储空间。
- 单独挂载：高效云盘支持挂载到在相同可用区内的任意ECS实例上。

应用场景

- MySQL、SQL Server、PostgreSQL等中小型关系数据库应用。
- 对数据可靠性要求高、中度性能要求的中大型开发测试应用。

SSD云盘

产品特点

SSD云盘利用分布式三副本机制，能够提供稳定的高随机I/O、高数据可靠性的高性能存储，具备如下特点：

- 高性能：最高提供20000随机读写IOPS、256MB/s的吞吐能力。
- 数据可靠性：采用分布式三副本机制，提供99.9999999%的数据可靠性。
- IOPS：每GB容量提供30个随机IOPS能力，最大提供20000随机读写IOPS性能。比如100GB的SSD云盘提供3000 IOPS性能；334GB的SSD云盘提供10020 IOPS性能。

  说明: SSD云盘只有挂载到I/O优化的实例时，才能获得期望的IOPS性能。挂载到非I/O优化的实例时无法获得期望的IOPS性能。
- 性能基准：
  - 块大小为4KB/8KB时，IOPS可达最大20000；
  - 块大小为16KB 时，IOPS最大16300左右，吞吐量到256MB/s上限；
  - 块大小为32KB时，IOPS最大8150左右，吞吐量到256MB/s上限；
  - 块大小为64KB时，IOPS最大4100左右；
  - 以此类推。
- 吞吐量：SSD云盘的吞吐性能=min{50+0.5xdisk_size，256 }MBps，起步50、每GB增加0.5MBps，上限 256MBps的吞吐性能。
- 最大容量：单块SSD云盘最大提供32768GB存储空间。
- 单独挂载：SSD云盘支持挂载到在相同可用区内的任意ECS实例上。

应用场景

SSD云盘具备稳定的高随机I/O性能及高数据可靠性，适合以下场景：

- PostgreSQL、MySQL、Oracle、SQL Server等中大型关系数据库应用。
- 对数据可靠性要求高的中大型开发测试环境。

普通云盘

产品特点

普通云盘采用机械磁盘作为存储介质，利用分布式三副本机制，提供高数据可靠性，具备如下特点：

- I/O性能：提供数百的随机读写IOPS能力，最大30~40MB/s的吞吐量。
- 数据可靠性：采用分布式三副本机制，提供99.9999999%的数据可靠性。
- 最大容量：单块普通云盘最大提供2000GB存储空间。
- 单独挂载：普通云盘支持挂载到在相同可用区内的任意ECS实例上。

应用场景

- 适合数据不被经常访问或者低I/O负载的应用场景；如果应用需要更高的I/O性能，建议使用SSD云盘。
- 需要低成本并且有随机读写I/O的应用环境。

7.3.2.4 云盘三副本技术介绍

本文将对云盘的三副本技术进行简要的介绍。

阿里云分布式文件系统为ECS提供稳定、高效、可靠的数据随机访问能力。

Chunk

ECS用户对虚拟磁盘的读写最终都会被映射为对阿里云数据存储平台上的文件的读写。阿里云提供一个扁平的线性存储空间，在内部会对线性地址进行切片，一个分片称为一个Chunk；对于每一个Chunk，阿里云会复制出三个副本，并将这些副本按照一定的策略存放在集群中的不同节点上，保证您数据的可靠性。

3份副本的原理

在阿里云数据存储系统中，有3类角色，分别称为Master、Chunk Server以及 Client。ECS用户的一个写操作，经过层层转换，最终会交由Client来执行，执行过程简要说明如下：

1. Client计算出这个写操作对应的Chunk。
2. Client向Master查询该Chunk的3份副本的存放位置。
3. Client根据Master返回的结果，向这3个Chunk Server发出写请求。
4. 如果三份都写成功，Client向用户返回成功；反之，Client向用户返回失败。

Master的分布策略会综合考虑集群中所有Chunk Server的磁盘使用情况，在不同交换机机架下的分布情况、电源供电情况、机器负载情况，尽量保证一个Chunk的所有副本分布在不同机架下的不同Chunk Server上，有效防止由于一个Chunk Server或一个机架的故障导致的数据不可用。

数据保护机制

当有数据节点损坏，或者某个数据节点上的部分硬盘发生故障时，集群中部分Chunk的有效副本数就会小于 3。一旦发生这种情况，Master就会发起复制机制，在Chunk Server之间复制数据，使集群中所有Chunk的有效副本数达到3份。

综上所述，对云盘上的数据而言，所有用户层面的操作都会同步到底层3份副本上，无论是新增、修改还是删除数据。这种模式，能够保障用户数据的可靠性和一致性。

至于ECS实例内由于病毒感染、人为误删除或黑客入侵等软故障原因造成的数据丢失，需要采用备份、快照等技术手段来解决。任何一种技术都不可能解决全部的问题，因地制宜的选择合适的数据保护措施，才能为您宝贵的业务数据筑起一道坚实的防线。

7.3.3 镜像

本文着重介绍了镜像的基本概念以及获取镜像和导入镜像的方法。

镜像是ECS实例运行环境的模板，一般包括操作系统和预装的软件。您可以使用镜像创建新的ECS实例和更换ECS实例的系统盘。

获取镜像的方法

云服务器ECS提供了以下灵活多样的方式让您方便地获取镜像：

- 根据现有的ECS实例创建自定义镜像。
- 选择其他用户共享给您的镜像。

导入镜像和复制镜像

您可以把线下环境的镜像文件导入到ECS的集群中生成一个自定义镜像。

您还可以把自定义镜像复制到其他地域，实现环境和应用的跨地域一致性部署。

7.3.4 快照

本文对快照的概念作了简要的介绍。

所谓快照，就是某一个时间点上某一个磁盘的数据拷贝。

在使用磁盘的过程中，您可能会遇到以下场景：

- 当您在磁盘上进行数据的写入和存储时，希望使某块磁盘上的数据作为其他磁盘的基础数据。
- 云盘（普通云盘、高效云盘或者SSD云盘）虽然提供了安全的存储方式确保您所存储的任何内容都不丢失，但是如果存储在磁盘上的数据本身就是错误的数据，比如由于应用的错误导致的数据

错误，或者黑客利用您的应用的漏洞进行恶意读写，那么就需要其他的机制来保证在您的数据出现问题时，能够恢复到您期望的数据状态。

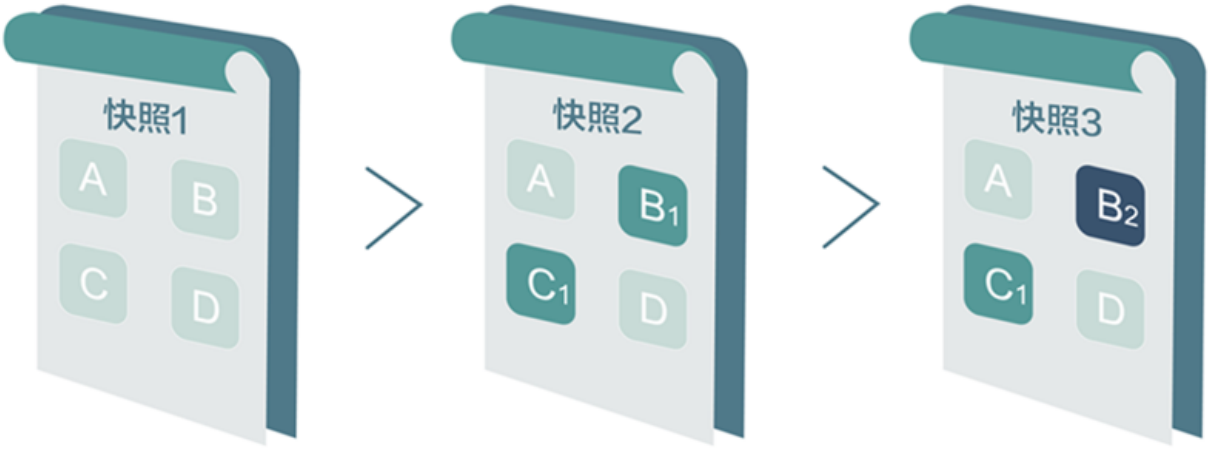
7.3.4.1 原理介绍

本文对快照的基本原理进行了简要介绍。

阿里云提供了快照机制，通过创建快照，您可以保留某一个时间点上一个磁盘的数据拷贝，有计划地对磁盘创建快照，从而保证您的业务可持续运行。

快照使用增量的方式，2个快照之间只有数据变化的部分才会被拷贝，如图所示。

图 16: 快照示意图



图中快照1、快照2和快照3分别是磁盘的第1个、第2个和第3个快照。文件系统对磁盘的数据进行分块检查，当创建快照时，只有数据变化了的块才会被复制到快照中。在该示例中：

- 快照1由于是磁盘的第一个快照，会把这个磁盘上的所有数据都复制一份。
- 而快照2只是复制了有变化的数据块B1和C1，数据块A和D引用了快照1中的A和D。
- 同理快照3则复制了有变化数据块B2，数据块A和D继续引用快照1中的，而数据块C1则引用快照2中的。
- 当磁盘需要恢复到快照3的状态，快照回滚会把数据块A、B2、C1和D复制到磁盘上，从而恢复成快照3的状态。
- 如果快照2被删除，快照中的数据块B1将被删除，但是数据块C1则不会被删除。这样在恢复到快照3时，仍可以恢复数据块C1的状态。

手动创建一个40GB的快照，一般只需要几分钟的时间。

快照存放在对象存储（OSS）上，但是这些存储在OSS上的快照文件对用户不可见，而且也不会计算到OSS的用户的Bucket的占用空间。您只能通过云服务器的控制台或者API进行快照操作。

7.3.4.2 快照2.0产品规格升级

本文对快照2.0产品规格进行对比说明。

ECS快照2.0数据备份服务在原有快照基础功能上，提供了更高的快照额度、更灵活的自动任务策略，并进一步降低了对业务I/O的影响。

表 7: 快照规格对比表

功能点	原快照规格	快照2.0规格	用户价值	示例
快照额度	磁盘数量×6+6	每块磁盘拥有64个快照额度。	更长的保护周期；更细的保护粒度。	• 某块非核心业务数据盘每天零点进行一次快照备份，可以保存超过2个月的备份数据； • 某块核心业务数据盘每隔4小时进行一次快照备份，可以保存超过10天的备份数据。
自动任务策略	系统默认，每天触发一次，无法手工修改。	支持自定义快照时间点、每周重复日期、快照保留时长，可查询自动快照策略关联的磁盘数及详情。	保护策略更灵活	• 您每天有24个快照时间点可供选择，一天之内可以进行多次快照； • 您可以选择周一到周日任意日期作为快照重复日期； • 您可以指定保存时长，或者永久保留（达到自动快照额度上限后会自动删除创建时间最早的那个自动快照）。
实现原理	COW (Copy-On-Write)	ROW (Redirect-On-Write)	减小快照任务对业务I/O写性能影响。	您的业务无感知，随时支持数据快照备份。

7.3.4.3 技术优势对比

本文将对 ECS 快照 2.0 数据服务与传统快照进行对比说明。

阿里云ECS快照2.0数据服务相比于传统存储产品数据快照功能，具备诸多优势。

表 8: 技术优势对比表

对比项	ECS快照2.0数据服务	传统存储快照功能
容量限制	无限容量，满足超大业务规模数据保护需求。	有限容量，受限于初次购买的存储设备容量，只能满足少量核心业务的数据保护需求。
扩展性	弹性伸缩，您可根据业务规模任意扩展，一次点击，秒级生效。	扩展性较差，受限于生产存储性能、可用容量、供应商支持能力等，一次调整周期约为1-2周时间。
成本投入	根据用户业务实际数据变化量，按快照占用容量收费。	前期投入大，涉及软件许可、预留空间、升级维护费用，投资浪费比较严重。
易用性	中文界面，7×24小时线上售后支持。	操作繁琐过程复杂，极大程度上受制于供应商支持能力。

7.3.5 网络和安全

安全是云计算的重要基石。以下是4种常见攻击的应对方案：

- *ARP* 欺骗防御
- 未知协议攻击防御
- *DDoS* 攻击防御
- 口令恶意破解

7.3.5.1 ARP 欺骗防御

ARP 欺骗攻击，是一种攻击成本很低但影响范围很大的攻击手段。为了防御 ARP 欺骗，我们在网络出口设置了 ARP 防火墙，只有使用平台统一分配的 MAC 地址才能够进行正常通讯，此举可以将非法流量阻隔在攻击者的实例之内。

7.3.5.2 未知协议攻击防御

为了防止异常协议通讯包流出实例，对其他实例或者网络设备进行攻击，我们通过专门的防火墙，对协议包进行了过滤，只允许 TCP/IP 协议栈的合法通讯包进出。

7.3.5.3 DDoS 攻击防御

只要您开通了云盾服务，云盾会对实例的网络流量、并发连接数、数据包数量进行监控，来识别和应对 DDoS 攻击，详细内容可以参考云盾产品介绍。

7.3.5.4 口令恶意破解

云盾会对恶意口令破解进行拦截，保证您的登录安全。

7.3.5.5 专有网络的IP

本文介绍了专有网络的使用场景以及收费方式等。

IP类型

专有网络类型的ECS实例有2种外网IP类型：

- 外网Public IP
- 弹性外网IP（EIP）

使用场景

如果您需要一个外网IP，希望在购买VPC类型的ECS实例时自动分配一个外网IP，在释放实例的时候随实例一起释放，不需要保留该外网IP，请选择Public IP。外网Public IP是随实例一起创建分配的，或升级0Mbps带宽时系统自动分配，分配后不能解绑，只能随实例一起释放。

如果您需要可以长期保留某个外网IP，可以绑定和解绑在需要的云服务器上，请选择弹性外网IP（EIP）。弹性外网IP（EIP）是可以单独申请，可以绑定到没有分配外网IP（Public IP或者EIP）的实例，也可以从实例上解绑，绑定到另外一个实例上，还可以单独进行释放。

一个专有网络类型的ECS实例上最多只能分配一个外网IP，要么外网Public IP，要么弹性外网IP。

收费方式

- 按固定带宽：需指定带宽的大小，如10Mbps (单位为bit)，费用合并在实例费用中一起支付。

- 按使用流量：是按实际发生的网络流量进行收费。先使用后付费，按小时计量计费。为了防止突然爆发的流量产生较高的费用，可以指定容许的最大网络带宽进行限制。

专有网络和经典网络的Public IP异同

- **相同点**：都可以通过Public IP进行访问Internet的。在产品上的所有操作都相同，随实例一起购买，可以升级带宽，不能解绑，可以随实例一起释放。
- **区别**：专有网络的Public IP是NAT IP，在机器内部无法通过命令行查询；经典网络是Binding IP，可以在机器中通过命令行查询。

7.3.5.6 内网

本文对阿里云的内网进行简要的介绍说明。

目前阿里云的服务器内网间，非I/O优化的实例为千兆共享的带宽，I/O优化的实例为万兆共享的带宽，没有特殊限制。由于是共享网络，因此无法保证带宽速度是不变的。

如果您在两台同地域的ECS实例传输数据，一般建议使用内网连接。同时，RDS、负载均衡以及OSS相关的内网速度也都是千兆共享的环境。这些产品间也都可以使用内网相互连接使用。

目前只要是相同地域下，负载均衡、RDS 、OSS同ECS之间都是可以直接内网互通连接使用的。

对于内网中**专有网络**下的ECS实例：

- 同一账号、同一地域、同一个VPC网络的实例：如果在同一个安全组，即默认内网互通；如果在不同安全组，可以通过安全组授权实现内网互通。
- 同一账号、同一地域的实例，如果在不同的VPC网络，需通过高速通道实现网络互通。
- 实例的内网IP地址不能进行修改、更换。
- 实例的内网、外网不支持VIP（虚拟IP）配置。
- 实例的网络类型不同，不能内网互通。

7.3.5.7 安全组

7.3.5.7.1 安全组限制

本文将对安全组中的相关限制条件进行说明。

- 单个安全组内的实例个数不能超过1000。如果您有超过1000个实例需要内网互访，可以将他们分配到多个安全组内，并通过互相授权的方式允许互访。
- 每个实例最多可以加入5个安全组。
- 每个用户的安全组最多100个。

- 每个安全组最多有100条安全组规则。
- 对安全组的调整操作，对您的服务连续性没有影响。
- 安全组是有状态的。如果数据包在Outbound方向是被允许的，那么对应的此连接在Inbound方向也是允许的。

7.3.5.7.2 安全组规则

本文将对安全组的相关规则进行具体说明。

安全组规则可以允许或者禁止与安全组相关联的ECS实例的外网和内网的入出方向的访问。

您可以随时授权和取消安全组规则。您的变更安全组规则会自动应用于与安全组相关联的ECS实例上。

在设置安全组规则的时候，安全组的规则务必简洁。如果您给一个实例分配多个安全组，则该实例可能会应用多达数百条规则。访问该实例时，可能会出现网络不通的问题。

7.4 产品优势

本文将介绍云服务器的主要特性和功能。

拥有传统服务器和虚拟主机无法企及的特性

- 稳定性：服务可用性高达99.95%，数据可靠性高达99.9999999%。支持宕机迁移、数据快照备份和回滚、系统性能报警。
- 容灾备份：每份数据多份副本，单份损坏可在短时间内快速恢复。
- 安全性：支持配置安全组规则、云盾防DDos系统、多用户隔离、防密码破解。
- 弹性扩容：10分钟内可启动或释放100台ECS实例；支持在线不停机升级带宽；5分钟内停机升级CPU和内存。
- 可控性：作为云服务器ECS的用户，您拥有超级管理员的权限，能够完全控制ECS实例的操作系统，可以通过管理终端自助解决系统问题，并可以进行部署环境、安装软件等操作。
- 易用性：丰富的操作系统和应用软件，使用镜像可一键简单部署同一镜像；可在多台ECS中快速复制环境，轻松扩展；支持自定义镜像、磁盘快照批量创建ECS实例。
- API接口：使用ECS　API调用管理，通过安全组功能对一台或多台服务器进行访问设置，使开发使用更加方便。

主要功能

- 数十种实例规格，满足各种不同需求。
- 3 种数据存储盘（普通云盘、SSD 云盘、高效云盘），并提供 I/O 优化实例。

- 支持多个版本的操作系统。
- 丰富的镜像资源，让您免安装、快速部署操作系统和应用软件。
- 提供 API 管理方式，给您完善的管理权限。

7.4.1 云计算的高可用性

本文介绍阿里云的重要特性：高可用性。

相较于普通的IDC机房以及服务器厂商，阿里云会使用更严格的IDC标准、服务器准入标准以及运维标准，以保证云计算整个基础框架的高可用性、数据的可靠性以及云服务器的高可用性。

在此基础之上，阿里云所提供的每个地域都存在多可用区。当您需要更高的可用性时，可以利用阿里云的多可用区搭建自己的主备服务或者双活服务。对于面向金融领域的两地三中心的解决方案，您也可以通过多地域和多可用区搭建出更高的可用性服务。其中包括容灾、备份等服务，阿里云都有非常成熟的解决方案。

在阿里云的整个框架下，这些服务可以非常平滑地进行切换。无论是两地三中心，还是电子商务以及视频服务等，都可以在阿里云找到对应的行业解决方案。

此外，阿里云为您提供了如下3项支持：

- 提升可用性的产品和服务，包括云服务器、负载均衡、多备份数据库服务以及数据迁移服务DTS等。
- 行业合作伙伴以及生态合作伙伴，帮助您完成更高、更稳定的架构，并且保证服务的永续性。
- 多种多样的的培训服务，让您从业务端到底层的基础服务端，在整条链路上实现高可用。

7.4.2 云计算的安全性

本文介绍阿里云的重要特性：安全性。

选择了云计算，最关心的问题就是云计算的安全与稳定。阿里云近期通过了很多的国际安全标准认证，包括ISO2007、MTCS等。这些安全合规对于用户数据的私密性、用户信息的私密性以及用户隐私的保护都有非常严格的要求。

- **在阿里云专有网络之上，可以产生更多的业务可能性。**

  您只需进行简单配置，就可在自己的业务环境下，与全球所有机房进行串接，从而提高业务的灵活性、稳定性以及可发展性。
- **对于原来拥有自建的IDC机房，也不会产生问题。**

阿里云专有网络可以拉专线到原有的IDC机房，形成混合云的架构。阿里云可以提供各种混合云的解决方案和非常多的网络产品，形成强大的网络功能，让您的业务更加灵活。结合阿里云的生态，您可以在云上发展出意想不到的业务生态。

- **阿里云专有网络更加稳定和安全。**

  稳定性：业务搭建在专有网络上，而网络的基础设施将会不停进化，使您拥有更新的网络架构以及更新的网络功能，使得您的业务永远保持在一个稳定的状态。专有网络允许您自由地分割、配置和管理自己的网络。

  安全性：面对互联网上不断的攻击流量，专有网络天然就具备流量隔离以及攻击隔离的功能。业务搭建在专有网络上后，专有网络会为业务筑起第一道防线。

总之，专有网络提供了稳定、安全、快速交付、自主可控的网络环境。对于传统行业以及未接触到云计算的行业和企业而言，借助专有网络混合云的能力和混合云的架构，它们将享受云计算所带来的技术红利。

7.4.3 云计算的弹性

本文介绍阿里云的重要特性：弹性。

云计算最大的优势就在于弹性。目前，阿里云已拥有在数分钟内开出一家中型互联网公司所需要的IT资源的能力，这就能够保证大部分企业在云上所构建的业务都能够承受巨大的业务量压力。

计算弹性

- **纵向的弹性**，即单个服务器的配置变更。传统IDC模式下，很难做到对单个服务器进行变更配置。而对于阿里云，当您购买了云服务器或者存储的容量后，可以根据业务量的增长或者减少自由变更自己的配置。
- **横向的弹性**。对于游戏应用或直播平台出现的高峰期，若在传统的IDC模式下，您根本无法立即准备资源；而云计算却可以使用弹性的方式帮助客户度过这样的高峰。当业务高峰消失时，您可以将多余的资源释放掉，以减少业务成本的开支。利用横向的扩展和缩减，配合阿里云的弹性伸缩，完全可以做到定时定量的伸缩，或者按照业务的负载进行伸缩。

存储弹性

阿里云拥有很强的存储弹性。当存储量增多时，对于传统的IDC方案，您只能不断去增加服务器，而这样扩展的服务器数量是有限的。在云计算模式下，将为您提供海量的存储，当您需要时可以直接购买，为存储提供最大保障。

网络弹性

云上的网络也具有非常大的灵活性。只要您购买了阿里云的专有网络，那么所有的网络配置与线下IDC机房配置可以是完全相同的，并且可以拥有更多的可能性。可以实现各个机房之间的互联互通，各个机房之间的安全域隔离，对于专有网络内所有的网络配置和规划都会非常灵活。

总之，对于阿里云的弹性而言，是计算的弹性、存储的弹性、网络的弹性以及您对于业务架构重新规划的弹性。您可以使用任意方式去组合自己的业务，阿里云都能够满足您的需求。

7.4.4 云服务器和传统IDC对比优势

本文主要介绍ECS与传统IDC的优势对比情况。

表 9: 优势对比表

	云服务器	传统 IDC
机房网络	自主研发的直流电服务器，绿色机房设计，PUE低。	传统交流电服务器设计，PUE高。
	骨干机房，出口带宽大，独享带宽。	机房质量参差不齐，用户选择困难，以共享带宽为主。
	BGP多线机房，全国访问流畅均衡。	以单线和双线为主。
	内置主流的操作系统，Windows正版激活。	需您自备操作系统，自行安装。
操作易用	可在线更换操作系统。	无法在线更换操作系统，需要您自己重装。
	Web在线管理，简单方便。	没有在线管理工具，维护困难。
	手机验证密码设置，安全方便。	重置密码麻烦，且被破解的风险大。
容灾备份	每份数据多份副本，单份损坏可在短时间内快速恢复。	您自行搭建，使用传统存储设备，价格高昂。
	用户自定义快照。	数据损坏需用户自己修复。
	快速自动故障恢复。	没有提供快照功能，无法做到自动故障恢复。
安全可靠	有效阻止MAC欺骗和ARP攻击。	很难阻止MAC欺骗和ARP攻击。
	有效防护DDoS攻击，可进行流量清洗和黑洞。	清洗和黑洞设备需要另外购买，价格昂贵。

	云服务器	传统 IDC
	端口入侵扫描、挂马扫描、漏洞扫描等附加服务。	普遍存在漏洞挂马和端口扫描等问题。
灵活扩展	开通云服务器非常灵活，可以在线升级配置。	服务器交付周期长。
	带宽升降自由。	带宽一次性购买，无法自由升降。
	在线使用负载均衡，轻松扩展应用。	硬件负载均衡，价格昂贵，设置也非常麻烦。
节约成本	使用成本门槛低。	使用成本门槛高。
	无需一次性大投入。	一次性投入巨大，闲置浪费情况严重。
	按需购买，弹性付费，灵活应对业务变化。	无法按需购买，必须为业务峰值满配。

8 容器服务

8.1 产品概述

容器服务（Container Service）是一种高性能可伸缩的容器管理服务，支持在一组阿里云云服务器上通过Docker 容器来运行或编排应用。

容器服务免去了您对容器管理集群的搭建，整合了负载均衡、专有网络等云产品，让您可以通过控制台或简单的API（兼容Docker API）进行容器生命周期管理。

8.2 功能特性

集群管理

- 您可以根据自己的需求，选择不同的地域创建和删除集群。
- 多种服务器托管方式。
- 支持将已创建的云服务器添加到指定集群。

一站式容器生命周期管理

- **网络**：支持跨宿主机容器间互联，支持通过container name或hostname定义的域名互访。
- **存储**：支持数据卷管理，支持OSSFS。
- **日志**：支持日志自动采集。
- **监控**：支持容器级别和VM级别的监控。
- **调度**：支持跨可用区高可用和异常节点的reschedule等策略。
- **路由**：支持4层和7层的请求转发和后端绑定。
- **子账号**：支持集群级别的RAM授权管理。

兼容标准Docker API

兼容标准的Docker Swarm和Docker Compose协议，无缝地将已有系统从线下迁移至云上。

阿里云环境特有的增值能力，更好的体验

- 扩展Compose模板定义，增强生命周期管理。
- 整合负载均衡，提供容器的访问能力。
- 高可用调度策略，轻松打通上下游交付流程。
- 支持服务级别的亲和性策略和横向扩展。

- 支持跨可用区高可用和灾难恢复。
- 支持集群和应用管理的OpenAPI，轻松对接持续集成和私有部署系统。

高效可靠

- 支持海量容器秒级启动。
- 支持容器的异常恢复和自动伸缩。
- 支持跨可用区的容器调度。

8.3 产品优势

简单易用

- 一键创建容器集群。
- 全兼容Docker Compose模板编排应用。
- 支持图形化界面和OpenAPI。

安全可控

- 用户拥有并独占云服务器。
- 支持定制安全组安全规则。

协议兼容

- 兼容标准Docker API，支持应用无缝迁云。
- 兼容Compose模板协议。
- 支持通过API对接，实现第三方的调度下发和系统集成。

8.4 基本概念

集群

一个集群指容器运行所需要的云资源组合，关联了若干服务器节点、负载均衡、专有网络等云资源。

节点

一台服务器（可以是虚拟机实例或者物理服务器）已经安装了Docker Engine，可以用于部署和管理容器；容器服务的Agent程序会安装到节点上并注册到一个集群上。集群中的节点数量可以伸缩。

容器

一个通过Docker镜像创建的运行时实例，一个节点可运行多个容器。

镜像

Docker镜像是容器应用打包的标准格式，在部署容器化应用时可以指定镜像，镜像可以来自于Docker Hub、阿里云容器Hub、或者用户的私有Registry。镜像ID 可以由镜像所在仓库URI和镜像Tag（缺省为latest）唯一确认。

编排模板

编排模板包含了一组容器服务的定义和其相互关联，可以用于多容器应用的部署和管理。容器服务支持Docker Compose模板规范并有所扩展。

应用

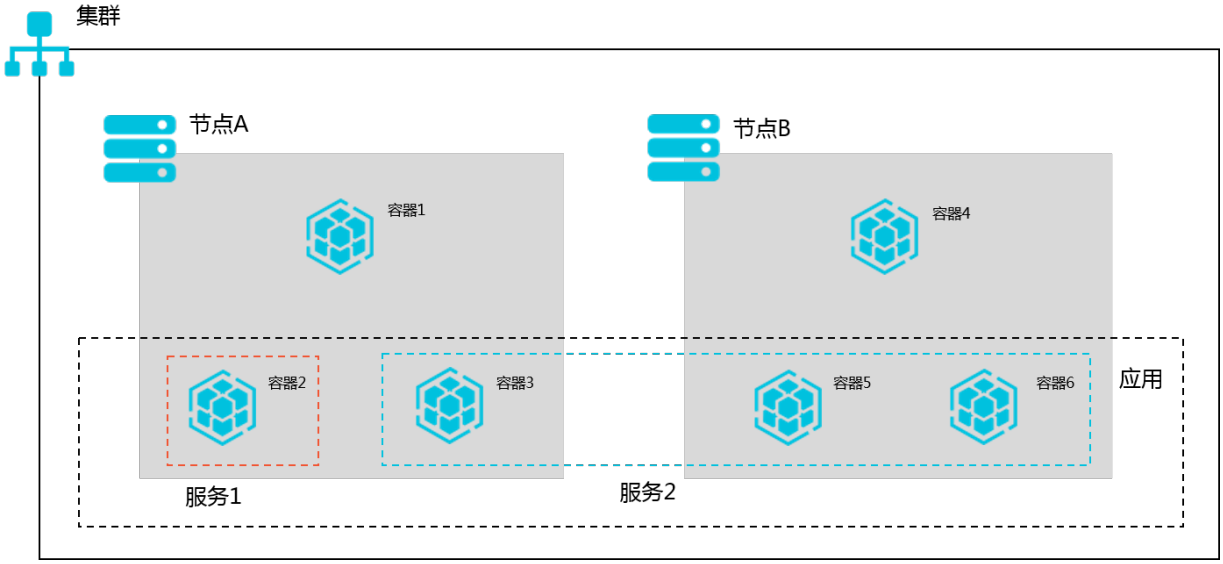
一个应用可通过单个镜像或一个编排模板创建，每个应用可包含1个或多个服务。

服务

一组基于相同镜像和配置定义的容器，作为一个可伸缩的微服务。

关联关系

图 17: 关联关系



9 对象存储OSS

9.1 什么是 OSS

对象存储服务（Object Storage Service，简称 OSS）提供海量、安全、低成本、高可靠的云存储服务。它可以理解为一个即开即用，无限大空间的存储集群。相比传统自建服务器存储，OSS 在可靠性、安全性、成本和数据处理能力方面都有着突出的优势。使用 OSS，您可以通过网络随时存储和调用包括文本、图片、音频和视频等在内的各种非结构化数据文件。

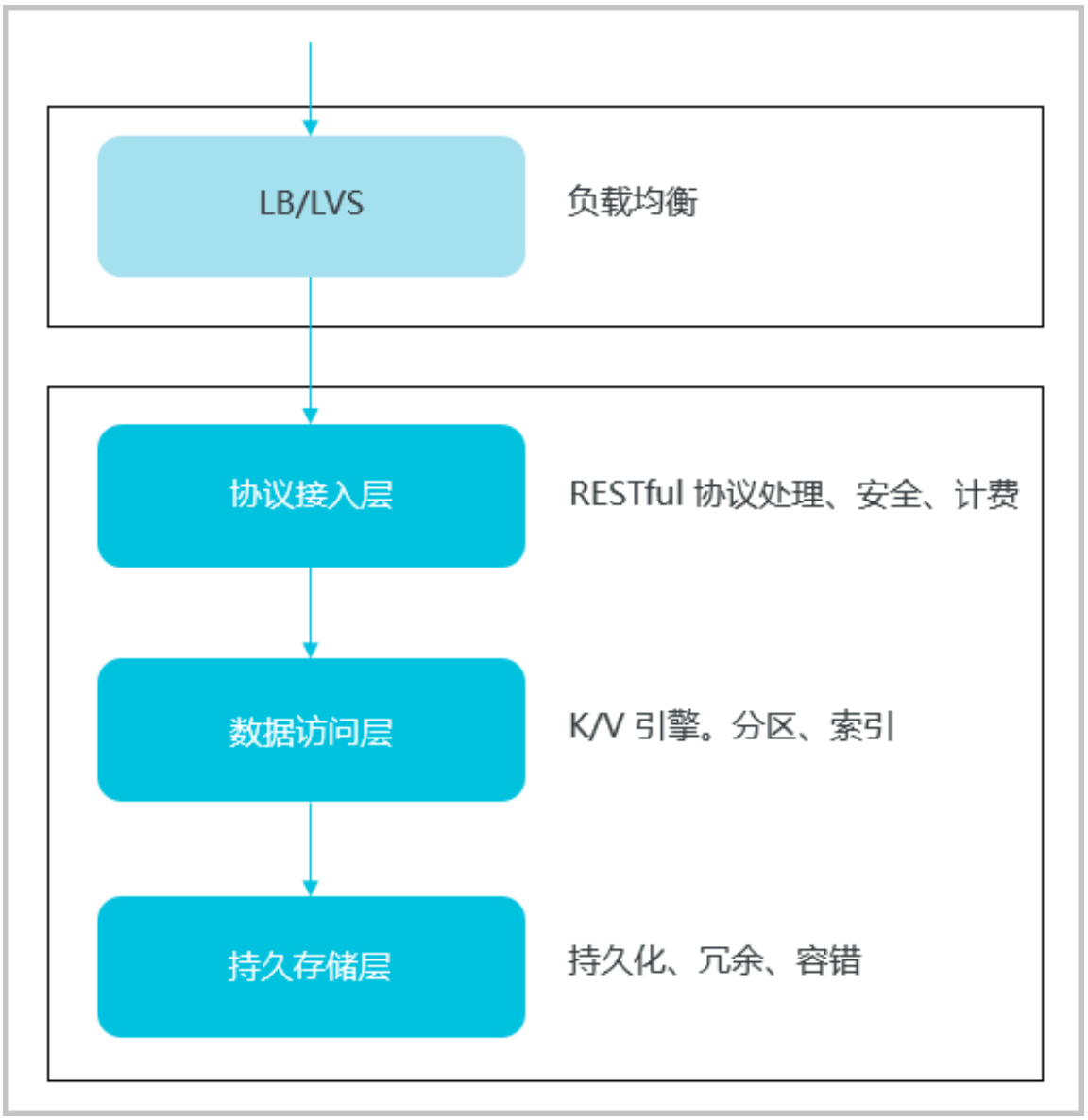
OSS 将数据文件以对象/文件（object）的形式上传到存储空间（bucket）中。 您可以进行以下操作：

- 创建一个或者多个存储空间
- 每个存储空间中添加一个或多个文件
- 通过获取已上传文件的地址进行文件的分享和下载
- 通过修改存储空间或文件的属性或元信息来设置相应的访问权限
- 通过云控制台执行基本和高级 OSS 任务
- 通过开发工具包 SDK 或直接在应用程序中进行 RESTful API 调用执行基本和高级 OSS 任务。

9.2 产品架构

对象存储 OSS 是构建在阿里云飞天平台上的一种存储解决方案。其基础是飞天平台的分布式文件系统，分布式任务调度等基础设施。这些基础设施提供了 OSS 以及其他阿里云服务所需的分布式调度、高速网络、分布式存储等重要特性。OSS 的架构如下图所示：

图 18: OSS 架构图



- 最上层是协议接入层，负责接收用户使用 RESTful 协议发来的请求，进行安全认证。如果认证通过，用户的请求将被转发到 Key-Value 引擎继续处理；如果认证失败，直接返回错误信息给用户。
- 数据访问层负责数据结构化处理，即按照 Key 来查找或存储数据，并支持大规模并发的请求。当协调服务集群变更导致服务被迫改变运行物理位置时，可以快速协调找到接入点。
- 最底层是持久存储层，即大规模分布式文件系统。元数据存储在 Master 上，Master 之间采用分布式消息一致性协议（Paxos）保证元数据的一致性。从而实现高效的文件分布式存储和访问，保证数据在系统中有 3 个备份以及在软硬件错误发生以后的故障恢复。OSS 系统的这一设计提供了不低于 99.9% 可用性和 99.99999999% 数据可靠性。

9.3 功能特性

OSS提供以下主要功能：

表 10: OSS主要功能

功能	描述
创建存储空间	在上传任何文件到 OSS 之前，您需要首先创建存储空间来存储文件。
删除存储空间	如果您不再需要存储空间，请将其删除以免进一步产生费用。
修改存储空间读写权限	OSS 提供权限控制 ACL（Access Control List），您可以在创建存储空间的时候设置相应的 ACL 权限控制，也可以在创建之后修改 ACL。
设置静态网站托管	将存储空间配置成静态网站托管模式，并通过存储空间域名访问该静态网站。
设置防盗链	为了减少您存储于 OSS 的数据被其他人盗链而产生额外费用，OSS 支持设置基于 HTTP header 中表头字段 referer 的防盗链方法。
管理跨域资源共享	OSS 提供 HTML5 协议中的跨域资源共享 CORS 设置，帮助您实现跨域访问。
设置生命周期	定义和管理存储空间内所有对象或对象的某个子集的生命周期。设置生命周期一般用于文件的批量管理和自动碎片删除等操作。
上传文件	您可以上传任意类型文件到存储空间中。
新建文件夹	您可以像管理 Windows 文件夹一样管理 OSS 文件夹。
搜索文件	在存储空间或文件夹中搜索具有相同的名称前缀的文件。
获取文件访问地址	通过获取已上传文件的地址进行文件的分享和下载。
删除文件	删除单个文件或批量删除文件。
删除文件夹	删除单个文件夹或批量删除文件夹。
修改文件读写权限	您可以在上传文件的时候设置相应的 ACL 权限控制，也可以在上传之后修改 ACL。
管理碎片	删除存储空间内的全部或部分碎片文件。
图片服务	对保存在OSS上的图片进行格式转换、剪裁、缩放、旋转、水印、样式封装等各种处理。
API	提供 OSS支持的 RESTful API 操作和相关示例。
SDK	提供主流语言 SDK 的开发操作和相关示例。

9.4 产品优势

OSS与自建存储对比

表 11: OSS与自建存储对比表

对比项	对象存储 OSS	自建服务器存储
可靠性	• 服务可用性不低于 99.9%。 • 规模自动扩展，不影响对外服务。 • 数据持久性不低于 99.99999999%。 • 数据自动多重冗余备份。	• 受限于硬件可靠性，易出问题，一旦出现磁盘坏道，容易出现不可逆转的数据丢失。 • 人工数据恢复困难、耗时、耗力。
安全	• 提供企业级多层次安全防护。 • 多用户资源隔离机制，支持异地容灾机制。（需要选配容灾套件） • 提供多种鉴权和授权机制及白名单、防盗链、主子账号功能。	• 需要另外购买清洗和黑洞设备。 • 需要单独实现安全机制。
数据处理能力	提供图片处理功能。	需要额外采购，单独部署。

方便、快捷的使用方式

提供标准的 RESTful API 接口、丰富的 SDK 包、客户端工具、控制台。您可以像使用文件一样方便地上传、下载、检索、管理用于 Web 网站或者移动应用的海量数据。

- 不限文件数量和大小。您可以根据所需存储量无限扩展存储空间，解决了传统硬件存储扩容问题。
- 支持流式写入和读出。特别适合视频等大文件的边写边读业务场景。
- 支持数据生命周期管理。您可以自定义将到期数据批量删除。

强大、灵活的安全机制

灵活的鉴权，授权机制。提供 STS 和 URL 鉴权和授权机制，以及白名单、防盗链、主子账号功能。

丰富的图片处理服务

支持 jpg、png、bmp、gif、webp、tiff 等多种图片格式的转换，以及缩略图、剪裁、水印、缩放等多种操作。

9.5 基本概念

本部分将向您介绍 OSS 中涉及的几个基本概念，以便于您更好地理解 OSS 产品。

对象/文件（Object）

对象是 OSS 存储数据的基本单元，也被称为 OSS 的文件。对象由元信息（Object Meta），用户数据（Data）和文件名（Key）组成。对象由存储空间内部唯一的 Key 来标识。对象元信息是一个键值对，表示了对象的一些属性，比如最后修改时间、大小等信息，同时用户也可以在元信息中存储一些自定义的信息。

根据不同的上传方式，对象的大小限制是不一样的。分片上传最大支持 48.8TB 的对象大小，其他的上传方式最大支持 5GB。

对象的生命周期是从上传成功到被删除为止。在整个生命周期内，对象信息不可变更。重复上传同名的对象会覆盖之前的对象，因此，OSS 不支持修改文件的部分内容等操作。

OSS 提供了追加上传功能，用户可以使用该功能不断地在 Object 尾部追加写入数据。

说明: 如无特殊说明，本文档中的对象、文件称谓等同于 Object。

存储空间（Bucket）

存储空间是您用于存储对象（Object）的容器，所有的对象都必须隶属于某个存储空间。您可以设置和修改存储空间属性用来控制访问权限、生命周期等，这些属性设置直接作用于该存储空间内所有对象，因此您可以通过灵活创建不同的存储空间来完成不同的管理功能。

- 同一个存储空间的内部是扁平的，没有文件系统的目录等概念，所有的对象都直接隶属于其对应的存储空间。
- 每个用户可以拥有多个存储空间。
- 存储空间的名称在 OSS 范围内必须是全局唯一的，一旦创建之后无法修改名称。
- 存储空间内部的对象数目没有限制。

强一致性

Object 操作在 OSS 上具有原子性，操作要么成功要么失败，不会存在有中间状态的Object。OSS 保证用户一旦上传完成之后读到的 Object 是完整的，OSS 不会返回给用户一个部分上传成功的 Object。

Object 操作在 OSS 上同样具有强一致性，用户一旦收到了一个上传（PUT）成功的响应，该上传的 Object 就已经立即可读，并且数据的三份副本已经写成功。不存在一种上传的中间状态，即

read-after-write 却无法读取到数据。对于删除操作也是一样的，用户删除指定的 Object 成功之后，该 Object 立即变为不存在。

强一致性方便了用户架构设计，可以使用跟传统存储设备同样的逻辑来使用OSS，修改立即可见，无需考虑最终一致性带来的各种问题。

OSS与文件系统的对比

OSS 是一个分布式的对象存储服务，提供的是一个 Key-Value 对形式的对象存储服务。用户可以根据 Object 的名称（Key）唯一地获取该Object的内容。虽然用户可以使用类似 test1/test.jpg 的名字，但是这并不表示用户的 Object 是保存在test1 目录下面的。对于 OSS 来说，test1/test.jpg 仅仅只是一个字符串，和 a.jpg 这种并没有本质的区别。因此不同名称的 Object 之间的访问消耗的资源是类似的。

文件系统是一种典型的树状索引结构，一个名为 test1/test.jpg 的文件，访问过程需要先访问到 test1 这个目录，然后再在该目录下查找名为 test.jpg 的文件。因此文件系统可以很轻易的支持文件夹的操作，比如重命名目录、删除目录、移动目录等，因为这些操作仅仅只是针对目录节点的操作。这种组织结构也决定了文件系统访问越深的目录消耗的资源也越大，操作拥有很多文件的目录也会非常慢。

对于 OSS 来说，可以通过一些操作来模拟类似的功能，但是代价非常昂贵。比如重命名目录，希望将 test1 目录重命名成 test2，那么 OSS 的实际操作是将所有以 test1/ 开头的 Object 都重新复制成以 test2/ 开头的 Object，这是一个非常消耗资源的操作。因此在使用 OSS 的时候要尽量避免类似的操作。

OSS 保存的 Object 不支持修改（追加写 Object 需要调用特定的接口，生成的 Object 也和正常上传的 Object 类型上有差别）。用户哪怕是仅仅需要修改一个字节也需要重新上传整个 Object。而文件系统的文件支持修改，比如修改指定偏移位置的内容、截断文件尾部等，这些特点也使得文件系统拥有广泛的适用性。但另外一方面，OSS 能支持海量的用户并发访问，而文件系统会受限于单个设备的性能。

因此，将 OSS 映射为文件系统是非常低效的，也是不建议的做法。如果一定要挂载成文件系统的话，建议尽量只做写新文件、删除文件、读取文件这几种操作。使用 OSS 应该充分发挥其优点，即海量数据处理能力，优先用来存储海量的非结构化数据，比如图片、视频、文档等。

10 消息服务

10.1 产品概述

阿里云消息服务（Message Service，简称 MNS）是一种高效、可靠、安全、便捷、可弹性扩展的分布式消息服务。MNS 能够帮助应用开发者在他们应用的分布式组件上自由地传递数据、通知消息，构建松耦合系统。

消息服务提供了队列模型，旨在提供高可靠高并发的一对一消费模型，即队列中的每一条消息都只能够被某一个消费者进行消费。队列模型就如同一家旋转寿司店，寿司店中有多个寿司师傅（生产者）在制作精美的寿司（消息），每一份寿司都是独特的，每位顾客（消费者）同时从传送带上拿取中意的寿司进行食用（消费）。

10.2 功能特性

丰富的队列操作和队列属性配置

阿里云消息服务提供了丰富的队列属性配置选项，您可以进行队列属性的个性化配置来满足不同的应用场景，支持普通队列、延迟队列、优先级队列等多种队列模式。

支持的队列功能包括创建消息队列、修改队列属性、获取队列属性、删除队列、获取队列列表等。

支持的消息功能包括发送消息到指定队列、消费队列的消息、删除被消费过的消息、查看队列的消息、修改消息下次可消费时间等。

支持并发访问

阿里云消息服务支持多个生产者和消费者并发访问同一个队列，无需特殊设置即可自由调整并发度，并能确保某条消息在取出之后的特定时间段内，无法被其他消费者获得。可以根据业务需求自由伸缩并发访问数。

消息投递保障

在消息有效期内，确保消息至少能被成功消费一次。用户间资源隔离，确保您队列中的消息不会被非法获取。

消息服务支持多种消息投递方式，包括推送到 HTTP(S) Server，推动到队列，推动到邮箱，推送到短信等。

分布式事务消息

提供完善的分布式环境下事务消息解决方案。

支持日志管理

可以通过日志管理的方式，查看每一条消息发送、接收和删除的完整生命周期。用户可以通过日志管理，方便地进行问题调查。

支持云监控

可以通过云监控查看队列情况，并且可以自定义报警项，当队列情况不符合期望时，能够及时知晓。

10.3 产品优势

消息服务 MNS 与自建队列集群的区别

核心优势	使用消息服务 MNS	自建队列集群
简单易用	• 您无需自行搭建消息服务，免运维； • 标准 HTTP RESTful 接口，接入方便； • 多种语言 SDK 支持且不断丰富。	• 需要采购服务器，初期成本投入高； • 需安装和运维，后期成本不断增加； • 非HTTP RESTful 接口，私有协议，通用性以及安全性偏低。
稳定可靠	• 消息三份拷贝，可靠性高达 99.99999999%； • 服务可用性高达 99.9%； • 特有机制确保 Always Writable。	• 数据单机或主从存储，可靠性差； • 单机或小集群服务，可用性无保证； • 无法确保 Always Writable。
安全防护	• 多层次安全防护和防 DDoS攻击； • 多用户隔离机制，每个用户配备独立命名空间； • 支持主子账号，支持鉴权和授权； • 支持 HTTPS，VPC 访问。	• 安全防护能力有限。
高性价比	• 队列数量以及队列存储容量可扩展性强； • 系统规模自动扩展，对用户完全透明； • 全球多地域提供服务； • 零启动成本，长期综合成本相比物理机降低 50%。	• 队列数量和消息堆积能力有限； • 不具备多地域服务能力； • 前期较高服务器成本和日益增长的运维成本。

10.4 典型应用

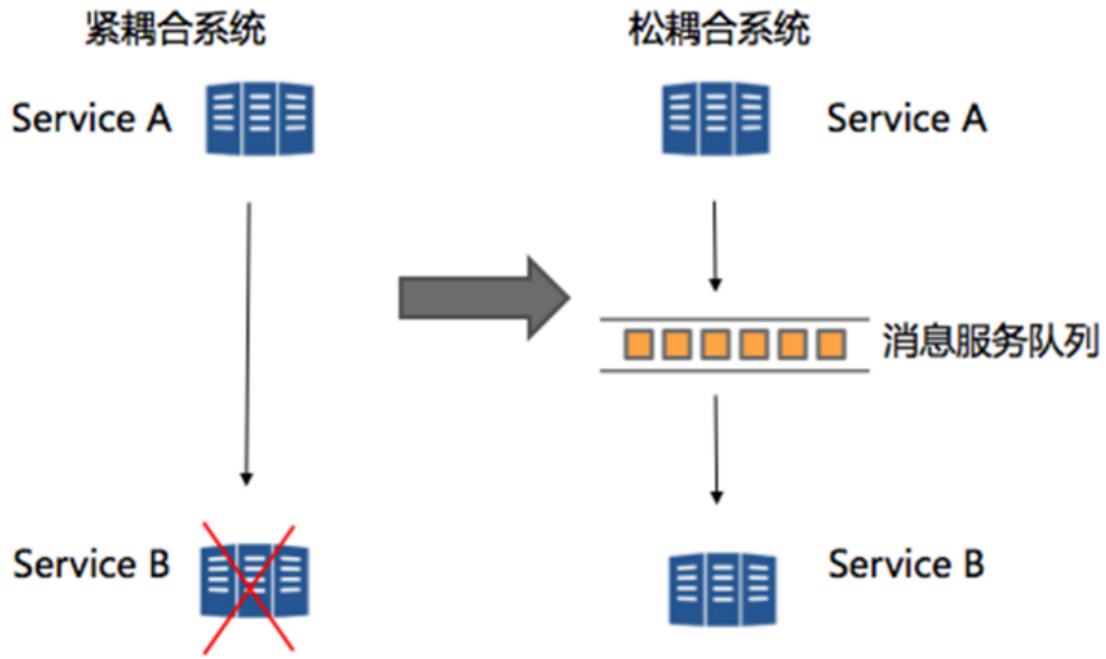
消息服务典型的使用场景包括：

- 将消息服务与其他阿里云产品集成，让应用程序更加可靠、灵活。
- 将消息服务用作工作队列，其中每条消息代表一项任务，需要通过一个流程来完成。一台或多台 ECS 可以从队列中读取并执行任务。
- 将业务流程中重要事件的通知保存在消息服务中，每个事件在队列中都有一条对应的消息，需要知晓该事件的应用程序可以读取和处理对应的消息。

系统解耦

消息服务可应用于系统解耦，如下图所示。在紧耦合系统中，当服务 B 出现问题或升级，都会影响服务 A。而在应用了消息服务的松耦合系统中，服务 B 出现问题或升级则不会影响服务 A。

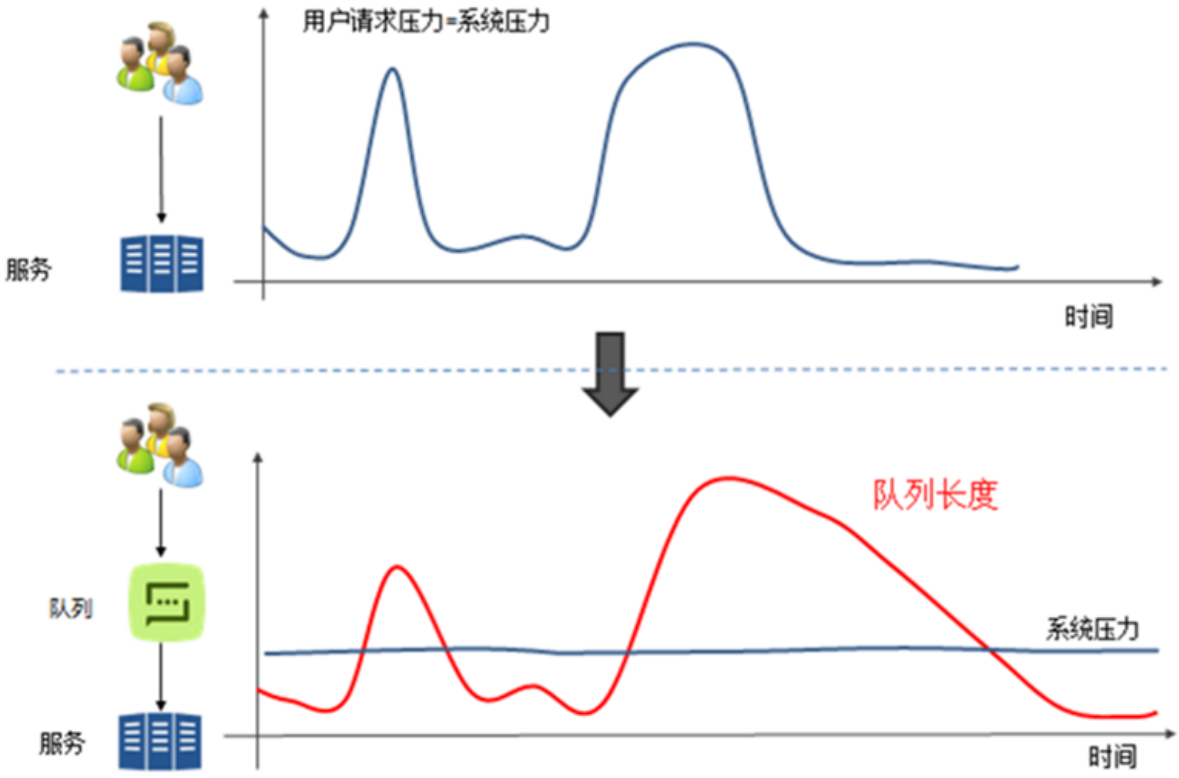
图 19: 系统解耦



削峰填谷

当流量洪流突然来袭时，消息服务可以缓冲突发流量，避免系统因突发流量崩溃。如下图所示，可以通过队列对用户请求压力来实现削峰填谷，从而降低系统峰值压力。

图 20: 削峰填谷



数据交换

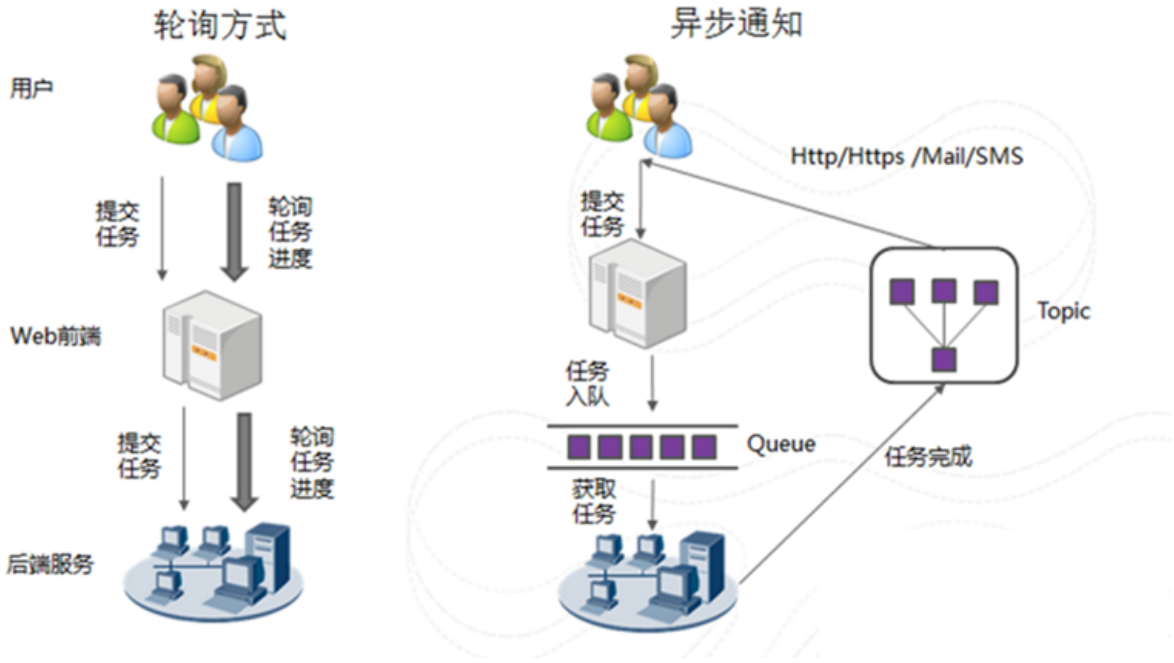
如下图所示，通过使用阿里云消息服务，您无需打通企业 A 和 B 的内网，也无需暴露企业 A 内网服务，就可以实现企业 A 向企业 B 数据同步和交互。同时消息服务已经支持 RAM 访问控制，可以灵活设置公网队列的访问策略。

图 21: 数据交换

异步通知

消息服务的通知功能，可以在后端服务处理完成任务时，回调通知用户，从而减少用户、Web 前端和后端服务之间大量不必要的轮询请求。

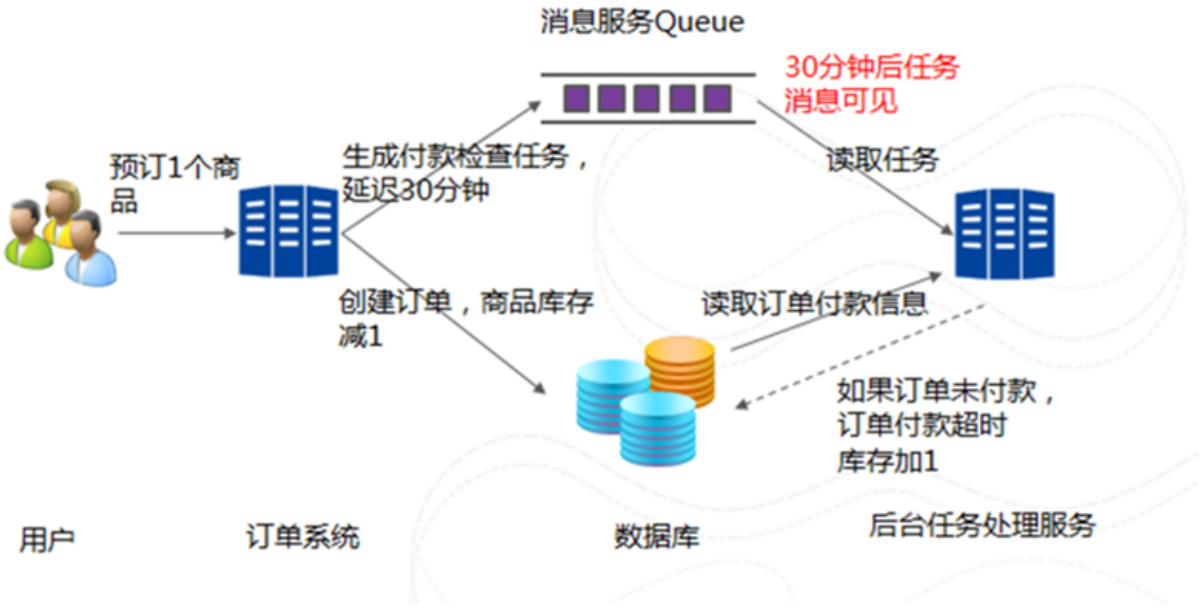
图 22: 异步通知



定时任务

消息服务提供延迟任务消息的功能，通过队列延迟发送任务消息，从而帮助您轻松实现定时任务场景。

图 23: 定时任务


消息服务Queue
30分钟后任务
消息可见
预订1个商品
生成付款检查任务，
延迟30分钟
读取任务
创建订单，商品库存
减1
读取订单付款信息
如果订单未付款，
订单付款超时
库存加1
用户
订单系统
数据库
后台任务处理服务

11 表格存储TableStore

11.1 什么是表格存储

表格存储（Table Store）是构建在阿里云飞天分布式系统之上的 NoSQL 数据存储服务，提供海量结构化数据的存储和实时访问。

- 表格存储以实例和表的形式组织数据，通过数据分片和负载均衡技术，达到规模的无缝扩展。
- 表格存储向应用程序屏蔽底层硬件平台的故障和错误，能自动从各类错误中快速恢复，提供非常高的服务可用性。
- 表格存储管理的数据全部存储在 SSD 中并具有多个备份，提供了快速的访问性能和极高的数据可靠性。

11.2 产品优势

扩展性

- 动态调整预留读/写吞吐量

  在创建表的时候，应用程序可以根据业务访问的情况来配置预留读/写吞吐量。表格存储根据表的预留读/写吞吐量进行资源的调度和预留。在使用过程中，还可以根据应用情况动态修改预留读/写吞吐量。

- 无限容量

  表格存储中表的数据量没有上限，随着表数据量的不断增大，表格存储会进行数据分区的调整从而为该表配置更多的存储。

数据可靠性

表格存储将数据的多个备份存储在不同机架的不同机器上，并会在备份失效时进行快速恢复，提供了极高的数据可靠性。

高可用性

通过自动的故障检测和数据迁移，表格存储对应用屏蔽了机器和网络的硬件故障，提供了高可用性。

管理便捷

应用程序无需关心数据分区的管理、软硬件升级、配置更新、集群扩容等繁琐的运维任务。

访问安全性

表格存储对应用程序的每一次请求都进行身份认证和鉴权，以防止未授权的数据访问，确保数据访问的安全性。

强一致性

表格存储保证数据写入强一致，写操作一旦返回成功，应用就能立即读到最新的数据。

灵活的数据模型

表格存储的表无固定格式要求，每行的列数及不同行同名列的类型可以不相同，支持多种数据类型，如 Integer、Boolean、Double、String 和 Binary。

监控集成

用户可以从表格存储控制台实时获取每秒请求数、平均响应延时等监控信息。

11.3 功能特性

针对表的功能和操作

- ListTable：列出实例下的所有表。
- CreateTable：创建表。
- DeleteTable：删除表。
- DescribeTable：获取表的属性信息。
- UpdateTable：更新表的预留读/写吞吐量配置。

针对数据的功能和操作

- 单行操作
  - GetRow：读取单行数据。
  - PutRow：新插入一行。如果该行内容已经存在，先删除旧行，再写入新行。
  - UpdateRow：更新一行。应用可以增加、删除一行中的属性列，或者更新已经存在的属性列的值。如果该行不存在，则新增一行。
  - DeleteRow：删除一行。
- 批量操作
  - BatchGetRow：批量读取一张或者多张表的多行数据。
  - BatchWriteRow：批量插入、更新、删除一张表或者多张表的多行数据。

- 范围读取
  - GetRange：读取表中某个范围内的数据。

针对写的功能

- 原子性

  PutRow、UpdateRow、DeleteRow 操作的结果保证原子性，即：要么全部成功，要么全部失败，不会存在中间状态。
- 强一致性

  应用程序获得写操作成功的响应后，本次操作的修改会立即生效，应用程序可以读取到该行最新的修改。

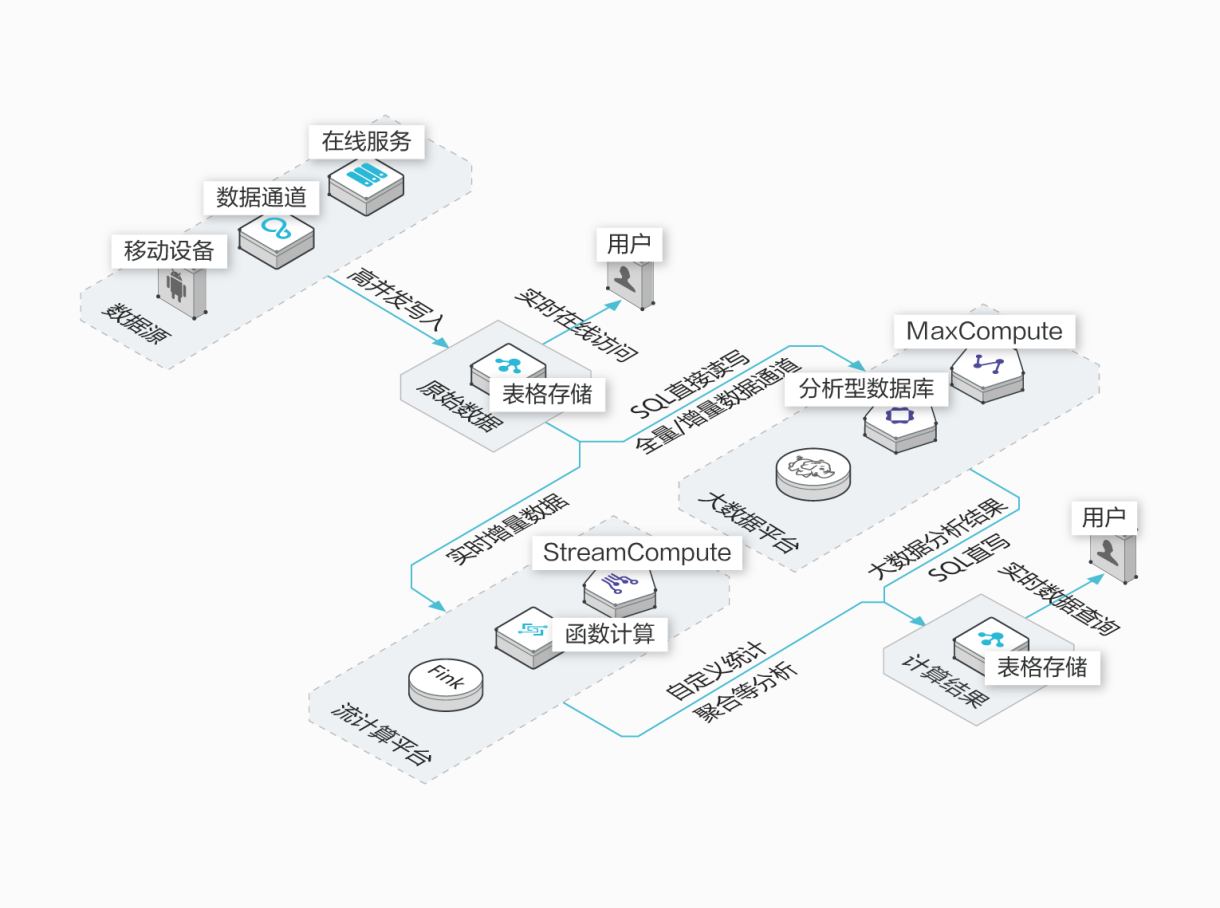
另外，表格存储提供 BatchWriteRow 操作，对多个单行写操作进行聚集，应用程序可以将多个 PutRow、UpdateRow、DeleteRow 操作放到一个 BatchWriteRow 操作中。需要特别注意的是，BatchWriteRow 操作是多个单行写操作的聚集，本身不保证原子性，可能会出现部分行操作执行成功，部分行操作执行失败的情况，但是 BatchWriteRow 的子操作具有原子性。

11.4 应用场景

大数据存储与分析

表格存储提供高并发、低延时的海量数据存储与在线访问，提供增量以及全量数据通道，并支持 MaxCompute 等大数据分析平台的 SQL 直读直写。高效的增量流式读接口让数据轻松完成实时流计算。

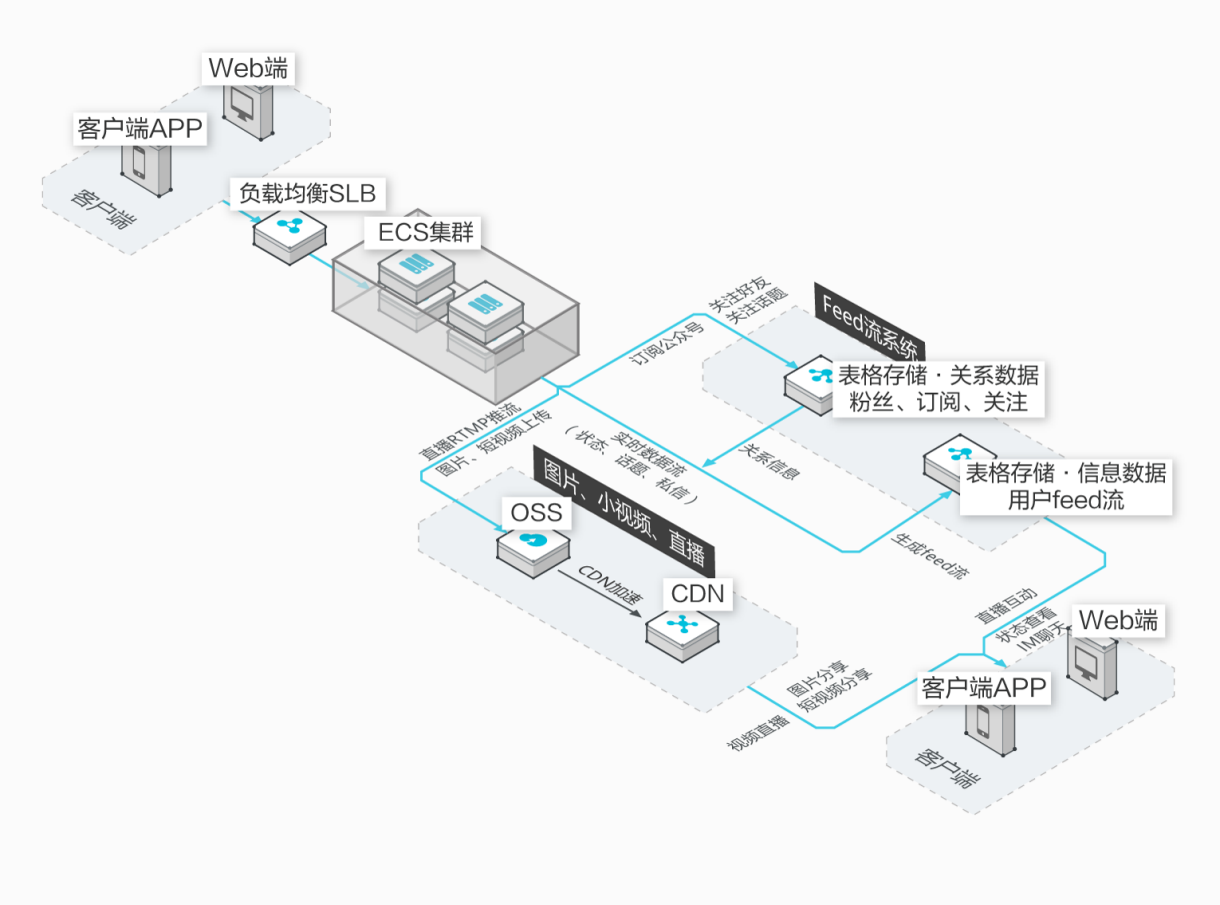
图 24: 大数据存储与分析



社交 Feed 流

表格存储能够存储人与人之间产生的大量社交信息，包括 IM 聊天，以及评论、跟帖和点赞等 Feed 流信息，满足访问波动明显、高并发、低延时的需要。图片与视频存储在 OSS 上通过 CDN 加速带来最优的用户体验。

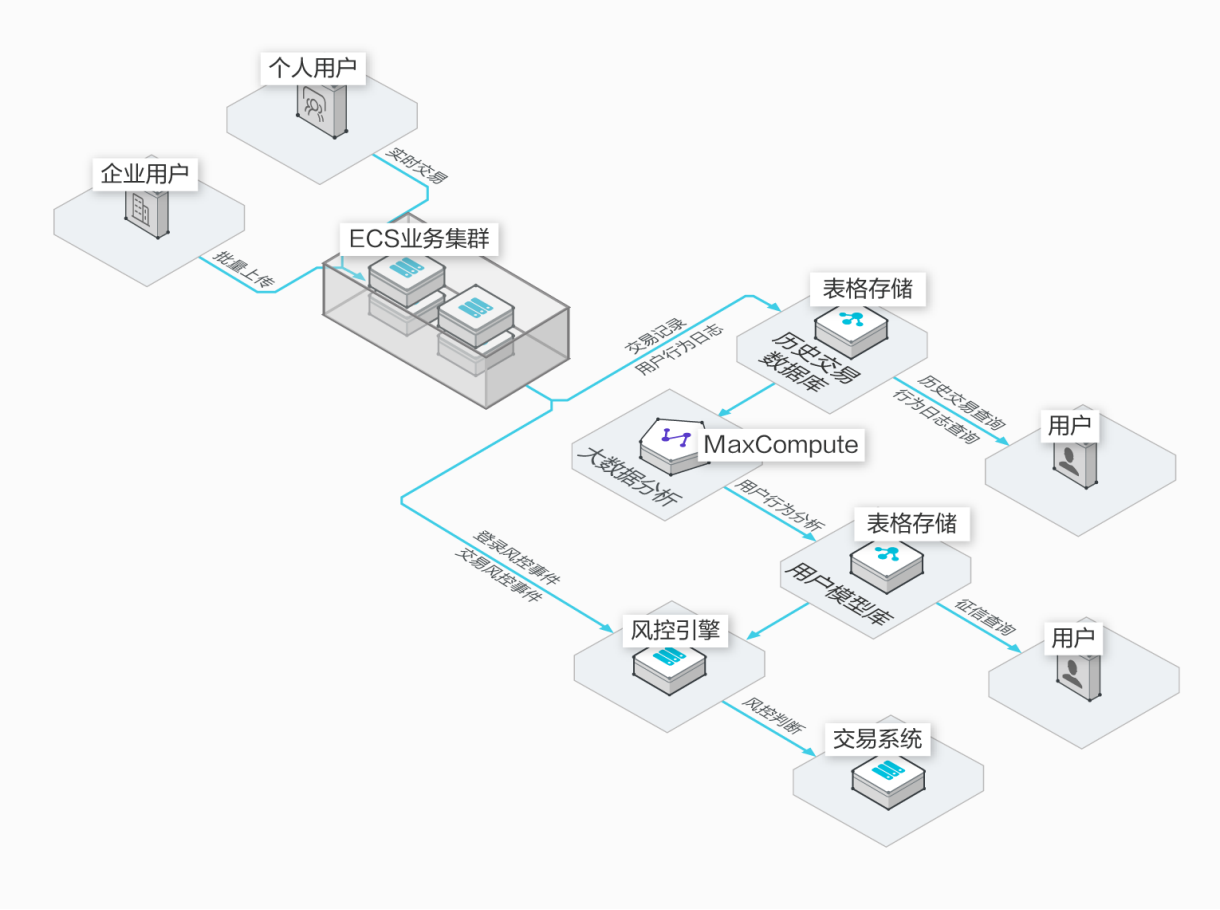
图 25: 社交 Feed 流场景



金融风控

低延时、高并发，弹性资源让您的风控系统永远工作在最佳状态，牢牢控制交易风险。灵活的数据结构能够让业务模式跟随市场需求快速迭代。

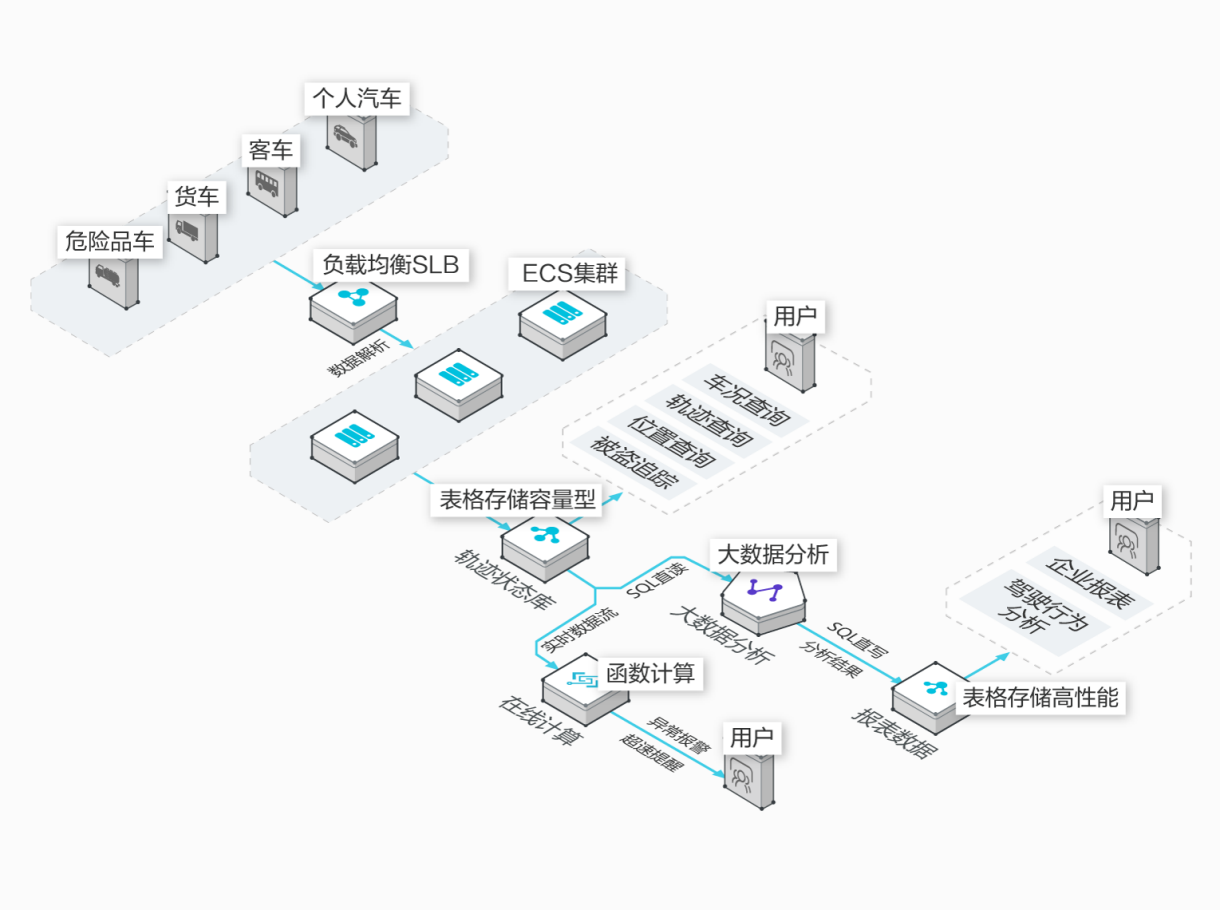
图 26: 金融风控场景



车联网数据

单表支撑 PB 级数据量，无需分库分表，简化业务逻辑，schemafree 数据模型轻松接入不同车辆设备的监控数据。与多种大数据分析平台、实时计算服务等无缝结合，轻松完成实时在线查询以及业务报表分析。

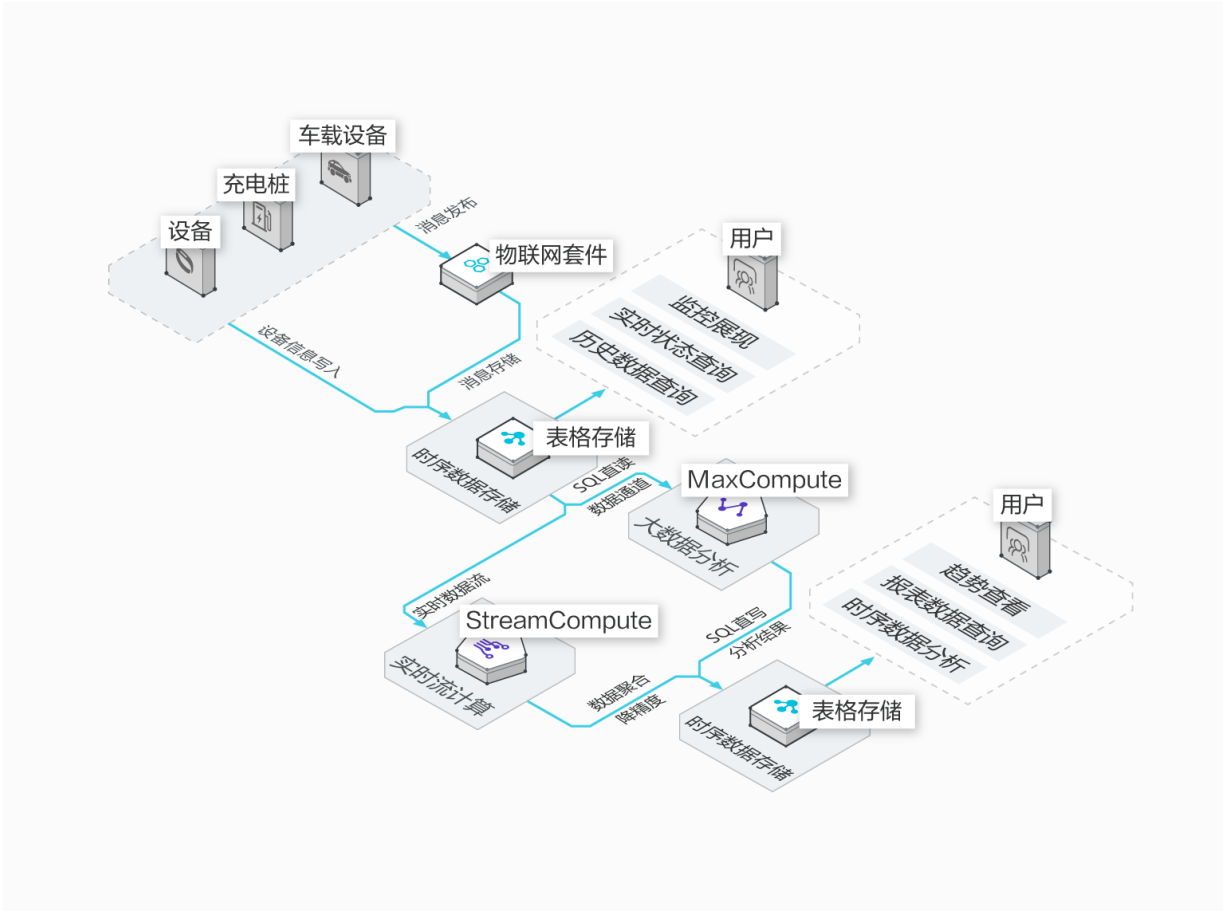
图 27: 车联网数据场景



物联网时序数据

单表 PB 级数据规模及千万级 QPS 让表格存储轻松满足 IoT 设备、监控系统等时序数据的存储需求，大数据分析 SQL 直读以及高效的增量流式读接口让数据轻松完成离线分析与实时流计算。

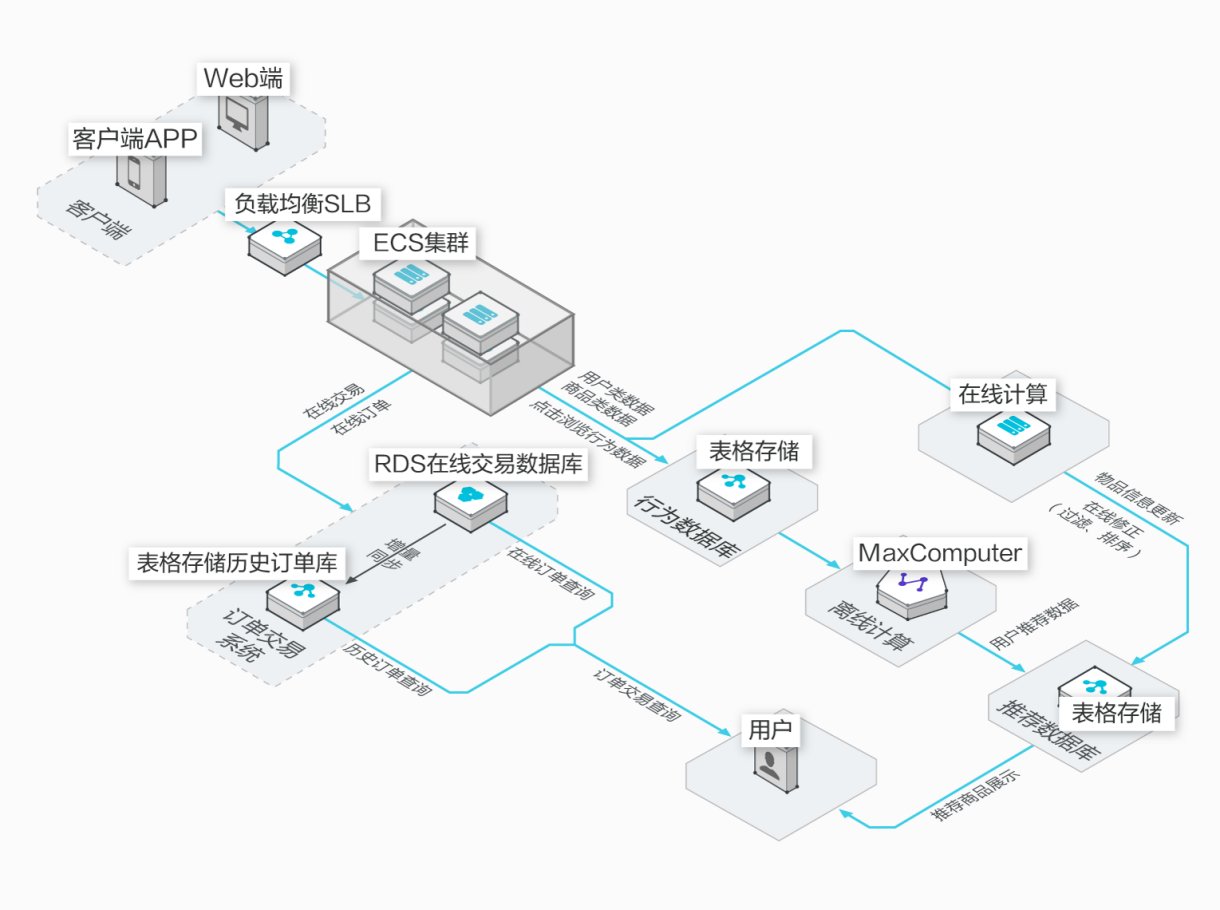
图 28: 物联网时序数据场景



电商推荐

使用表格存储让您无需担心大量历史交易订单的数据规模与访问性能，配合大数据计算服务，轻松实现精准营销，让您从容应对所有用户在线高峰时刻。

图 29: 电商推荐场景



11.5 使用限制

下表为表格存储的使用限制汇总，部分限制范围表明用户能够使用的最大值，而不是建议值。为保证更好的性能，请合理设计表结构和单行数据大小。

表 12: 使用限制

限制项	限制范围	说明
一个阿里云用户账号下可以保有实例数	不超过 10	如有需求提高上限，请联系管理员。
一个实例中表的个数	不超过 64	如有需求提高上限，请联系管理员。
实例名字长度	3-16 Bytes	字符集为 [a-z、A-Z、0-9] 和连词符（-），首字符必须是字母且末尾字符不能为连词符（-）。
表名长度	1-255 Bytes	字符集为 [a-z、A-Z、0-9] 和下划线（_），首字符必须是字母或下划线（_）。

限制项	限制范围	说明
列名长度限制	1-255 Bytes	字符集为 [a-z、A-Z、0-9] 和下划线（_），首字符必须是字母或下划线（_）。
主键包含的列数	1-4	至少 1 列，至多 4列。
String 类型主键列列值大小	不超过 1 KB	单一主键列 String 类型的列列值大小上限 1 KB。
String 类型属性列列值大小	不超过 2 MB	单一属性列 String 类型的列列值大小上限 2 MB。
Binary 类型主键列列值大小	不超过 1 KB	单一主键列 Binary 类型的列列值大小上限 1 KB。
Binary 类型属性列列值大小	不超过 2 MB	单一属性列 Binary 类型的列列值大小上限 2 MB。
一行中属性列的个数	不限制	不限制单一行拥有的属性列个数。
单行数据大小	不限制	不限制单一行中所有列名与列值总和大小。
读请求中 columns_to_get 参数的列的个数	0-128	读请求一行数据中获取的列的最大个数。
单表 UpdateTable 的次数	上调：无限制 下调：无限制	需要遵循单表的调整频率限制。
单表 UpdateTable 的频率	每 2 分钟 1 次	单表在 2 分钟之内，最多允许调整 1 次预留读/写能力值。
BatchGetRow 一次操作请求读取的行数	不超过 100	N/A
BatchWriteRow 一次操作请求写入行数	不超过 200	N/A
BatchWriteRow 一次操作的数据大小	不超过 4 MB	N/A
GetRange 一次返回的数据	5000 行或者 4 MB	一次返回数据的行数超过 5000 行，或者返回数据的数据大小大于 4 MB。以上任一条件满足时，超出上限的数据将会按行级别被截掉并返回下一行数据主键信息。
一次 HTTP 请求 Request Body 的数据大小	不超过 5 MB	N/A

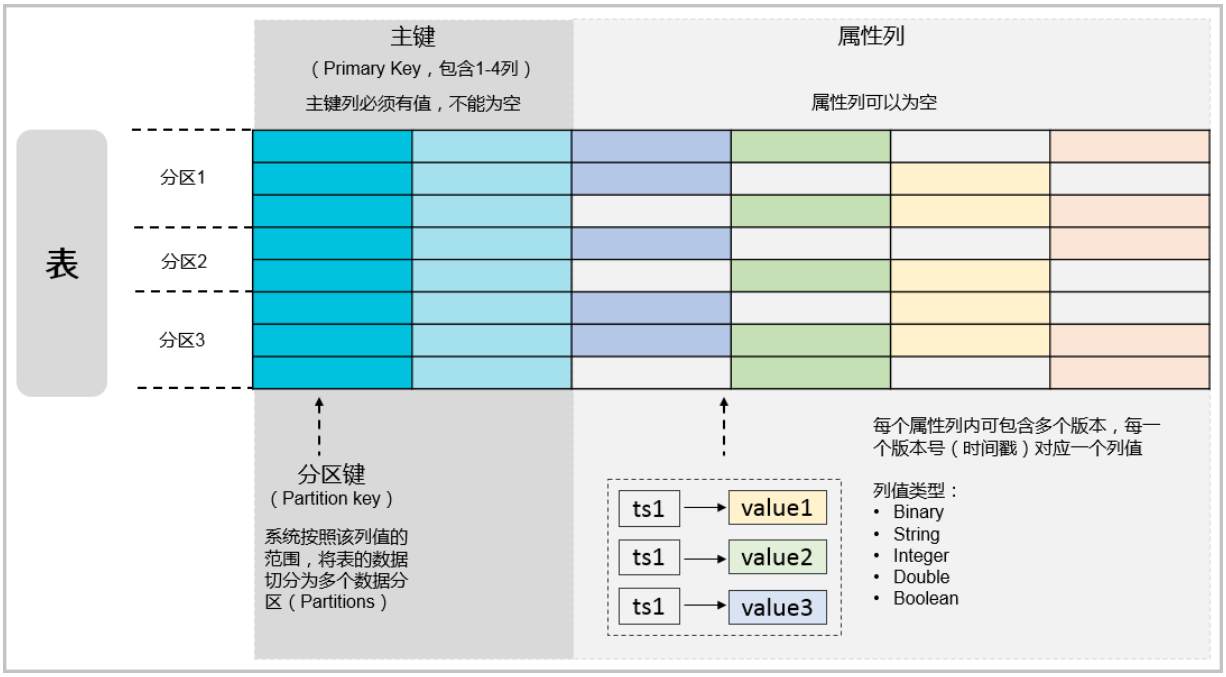
11.6 基本概念

11.6.1 数据模型

表格存储的数据模型概念包括表、行、主键和属性。

数据模型图如图 *30:* 数据模型图所示：

图 30: 数据模型图



- 表是行的集合，行由主键和属性组成。
- 主键列和属性列均由名称和值组成。
- 所有行都必须包含主键列，其主键列的数目和名称必须相同。
- 每行包含的属性列的数目可以不固定，名字和数据类型也可以不同。
- 与主键列不同，每个属性列可以包含多个版本，每个版本号（时间戳）对应一个列值。

说明: 时间戳是从 1970-01-01 00:00:00 UTC 时间到当前写入时间的毫秒数。

如下例子展示了同一张表中的两行：

ID	Type	ISBN	PageCount	Length
‘4776’	timestamp = 1466676354000, value = ‘Book’	timestamp = 1466676354000, value = ‘123*45678912345’	timestamp = 1466676354000, value = 666	空

ID	Type	ISBN	PageCount	Length
‘6555’	timestamp = 1466676354000, value = ‘Music’	空	空	timestamp = 1466676354000, value = 400; timestamp = 1466762754000, value = 500

上述表格的含义为：

- ID 是表的主键，ID 为 ‘4776’ 和 ‘6555’ 的行拥有不同的属性，它们可以被存在一张表中。
- ID 为 ‘4776’ 行的 Type 属性列只有一个版本数据，版本号为 1466676354000 的数据为 ‘Book’ 。
- ID 为 ‘6555’ 行的 Length 属性列有两个版本数据，版本号为 1466676354000 的数据为 400，版本号为 1466762754000 的数据为 500。

11.6.2 最大版本数

最大版本数（Max Versions）是数据表的一个属性，表示该数据表中的属性列能够保留多少个版本的数据。当一个属性列的版本个数超过 Max Versions 时，最早的版本将被异步删除。

建表后，您可以通过 UpdateTable 接口动态更改数据表的 Max Versions。

注意:

- 超过 Max Versions 的数据版本为无效数据，即使数据还没有被真正删除，该数据对用户已经不可见，无法读出。
- 当调小 Max Versions 时，如果数据版本个数超过新设的 Max Versions，最早的版本会被系统异步删除。
- 当调大 Max Versions 时，如果以前版本个数超过旧的 Max Versions 但还没有被系统删除的，数据会被重新读出来。

11.6.3 数据生命周期

数据生命周期（Time to live，简称 TTL）是数据表的一个属性，即数据的存活时间，单位为秒。表格存储会在后台对超过存活时间的数据进行清理，以减少用户的数据存储空间，降低存储成本。

- TTL 由用户在建表时进行设置，如果希望数据永不过期，将其设置为 -1。
- 建表后，可以通过 UpdateTable 接口动态更改 TTL。

- TTL 的单位为秒，例如期望过期时间为 30 天，TTL 应设置为 2592000（即 30 * 24 * 3600）。

假设数据表的 TTL 设置为 86400（一天），在 2016-07-21 00:00:00 UTC 时，该数据表上所有版本号小于 1468944000000（除以 1000 换算成秒之后即 2016-07-20 00:00:00 UTC）的属性列都将过期，系统会自动清理这些过期的数据。

注意:

- 超过 TTL 的过期数据为无效数据，即使数据还没有被真正删除，该数据对用户已经不可见，无法读出。
- 当调小 TTL 时，可能会有数据因为 TTL 变小而过期，这部分数据会被系统异步删除。
- 当调大 TTL 时，如果有版本号在上个 TTL 之外的数据还没有被系统删除，数据会被重新读出。

11.6.4 有效版本偏差

有效版本偏差（Max Version Offset）是数据表的一个属性，单位为秒。

为了防止非期望的写入，服务端在处理写请求时会对属性列的版本号进行检查。当版本号小于当前写入时间减去 Max Version Offset，或者大于等于当前写入时间加上 Max Version Offset 的值时，该行数据写入失败。

属性列的有效版本范围为：[数据写入时间 - 有效版本偏差，数据写入时间 + 有效版本偏差)。数据写入时间为 1970-01-01 00:00:00 UTC 时间到当前写入时间的秒数。属性列版本号为毫秒，其除以 1000 换算成秒之后必须属于这个范围。

例如，当数据表的有效版本范围为 86400（一天），在 2016-07-21 00:00:00 UTC 时，只能写入版本号大于 1468944000000（换算成秒之后即 2016-07-20 00:00:00 UTC）并且小于 1469116800000（换算成秒之后即 2016-07-22 00:00:00 UTC）的数据。当某一行的某个属性列版本号为 1468943999000（换算成秒之后即 2016-07-19 23:59:59 UTC，小于一天）时，该行数据写入失败。

- 建数据表时，用户若不设置有效版本偏差，将使用默认值 86400。
- 建表后，可以通过 UpdateTable 接口动态更改有效版本偏差。
- 有效版本偏差为非 0 值，可以大于 1970-01-01 00:00:00 UTC 时间到当前时间的秒数。

11.6.5 主键和属性

主键

- 主键是表中每一行的唯一标识。主键由 1 到 4 个主键列组成。
- 创建表的时候，必须明确指定主键的组成、每一个主键列的名字和数据类型以及它们的顺序。
- 主键列的数据类型只能是 String、Integer 和 Binary。如果为 String 或者 Binary 类型，长度不超过 1 KB。

分区键

组成主键的第一个主键列又称为分区键。表格存储会根据表中每一行分区键的值所属的范围自动将这一行数据分配到对应的分区和机器上，以达到负载均衡的目的。

具有相同分区键的行属于同一个数据分区，一个分区可能包含多个分区键。分区键是最小的分区单位，一个分区键下的数据无法再做切分。为了防止分区过大无法切分，单个分区键下所有行的大小总和建议不超过 1 GB。

表格存储服务会根据特定的规则对分区进行分裂和合并，以达到更好的负载均衡。这个过程是自动的，应用程序无需关心。

属性

属性存放行的数据。每一行包含的属性列个数没有限制。

版本号

在一个属性列上，当写入的版本数超过数据表的最大版本数时，较早版本的数据会被删除，只保留最新的 Max Versions 的版本数。

在写入数据时可以指定属性列的版本号，如果不指定版本号，服务端会将当前时间的毫秒单位时间戳（从 1970-01-01 00:00:00 UTC 计算起的毫秒数）作为属性列生成版本号。比如属性列版本号为 1468944000000（即 2016-07-20 00:00:00 UTC），当数据表的 TTL 设置为 86400（一天）时，该版本的数据将会在 2016-07-21 00:00:00 UTC 过期，随后会被后台系统自动删除。

读取一行数据时，可以指定每列最多读取多少版本或者读取的版本号范围。

注意:

- 版本号的单位为毫秒，在进行 TTL 比较和有效版本偏差计算时，需要除以 1000 换算成秒。

- 当数据的版本号（即时间戳）完全由服务端决定时，写入的数据在写入后经过 TTL 秒后会被系统清理。
- 为了防止无效的写入，写入过期数据将会直接失败。

  例如在 2016-07-21 00:00:00 向 TTL 为 86400 的数据表中写入版本号小于 1468944000000（即 2016-07-20 00:00:00 UTC）的数据将会直接失败。
- 为了防止错误的写入，写入的属性列的版本号换算成秒后，需要在 [数据写入时间-有效版本偏差，数据写入时间+有效版本偏差) 的范围内。

列名的命名规范

主键列和属性列遵循如下命名规范：

- 必须由英文字母、数字或下划线（_）组成
- 首字符必须为英文字母或下划线（_）
- 大小写敏感
- 长度在 1~255 个字符之间

列值类型

表格存储支持 5 种类型的列值。

表 13: 列值类型

数据类型	定义	是否可为主键	大小限制
String	UTF-8，可为空	是	为主键列时最大为 1 KB，为属性列时请参考限制项汇总。
Integer	64 bit，整型	是	8 Bytes
Double	64 bit，Double 类型	否	8 Bytes
Boolean	True/False，布尔类型	否	1 Byte
Binary	二进制数据，可为空	是	为主键列时最大为 1 KB，为属性列时请参考限制项汇总。

11.6.6 读/写吞吐量

读/写吞吐量的单位为读服务能力单元和写服务能力单元，简称 CU （Capacity Unit），是数据读写操作的最小计费单位。

- 1 单位读能力表示从数据表中读一条 4 KB 数据。
- 1 单位写能力表示向数据表写一条 4 KB 数据。
- 操作数据大小不足 4 KB 的部分向上取整，如写入 7.6 KB 数据消耗 2 单位写能力，读出 0.1 KB 数据消耗 1 单位读能力。

应用程序通过 API 进行表格存储读写操作时，会消耗对应的写服务能力单元和读服务能力单元。

预留读/写吞吐量

预留读/写吞吐量是表的一个属性。应用程序在创建表的时候，可以为该表指定预留读/写吞吐量。预留读写吞吐量的配置不影响该数据表的访问性能和服务能力。

- 预留读/写吞吐量可以设置为 0。
- 当预留读/写吞吐量大于 0 时，表格存储根据该配置为表分配和预留相应的资源，从而获得更低的资源使用成本。
- 当预留读/写吞吐量大于 0 时，即使没有读写请求也会进行计费，所以表格存储限制用户能够自行设置的单表预留读写吞吐量最大为 5000（读和写分别不超过 5000）。如果用户有单表预留读写吞吐量需要超出 5000 的需求，可以联系管理员提高预留读写吞吐量。
- 不存在的表将被视作预留读和预留写吞吐量均为 0，访问不存在的表将根据操作类型消耗 1 个按量读 CU 或者 1 个按量写 CU。

应用程序可以通过 UpdateTable 操作动态修改表的预留读/写吞吐量配置。

按量读/写吞吐量

按量读/写吞吐量是数据表在每一秒钟实际消耗的读/写吞吐量中超出预留读/写吞吐量的部分，统计周期为 1 秒。

假如某数据表设置的预留读吞吐量为 100，某 1 秒内读操作实际消耗 120 读吞吐量，则这 1 秒内消耗的按量读吞吐量为 20。如果数据表设置的预留读吞吐量为 0，则该数据表上所有的读访问消耗的读吞吐量均为按量读吞吐量。

由于按量读/写吞吐量的模式无法预估需要为数据表预留的计算资源，表格存储需要提供足够的服务能力以应对突发的访问高峰，所以按量吞吐量的单价高于预留吞吐量的单价。合理设置数据表的预留吞吐量能够有效地降低使用成本。

注意： 由于按量读/写吞吐量无法准确估计需要预留的资源，在某些极端访问情况下，若单个分片键每秒钟的访问需要消耗 10000 CU，表格存储可能会返回

OTSCapacityUnitExhausted 错误给应用程序。此时，应用程序需要使用退避重试等策略来减少访问该表的频率。

11.6.7 实例

实例是用户使用和管理表格存储服务的实体，用户在开通表格存储服务之后，需要通过云控制台来创建实例，然后在实例内进行表的创建和管理。实例是表格存储资源管理的基础单元，表格存储对应用程序的访问控制和资源计量都在实例级别完成。

用户可以为不同的业务创建不同的实例来管理相关的表，也可以为同一个业务的开发测试和生产环境创建不同的实例。

表格存储允许一个账号最多创建 10 个实例，每个实例内最多创建 64 张表。如果您需要增加限额，请联系管理员。

实例的名字在单个节点内必须唯一，用户可以在不同的节点内创建名字相同的实例。实例命名规范如下：

- 必须由英文字母、数字或连字符（-）组成
- 首字符必须为英文字母
- 末尾字符不能为连字符（-）
- 大小写不敏感
- 长度在 3 Byte ~ 16 Byte 之间

目前表格存储支持两种实例规格：高性能实例和容量型实例。

注意: 创建实例时请谨慎选择规格类型，创建后无法修改。

两种实例规格在功能上保持一致，都能够支持单表 PB 级别的数据量，主要区别在于使用成本及适用场景上。

- 高性能实例

  高性能实例能够提供百万级读写 TPS，单行的读写操作的平均延时为单个毫秒，适用于对读写性能和并发都要求非常高的场景，例如游戏、金融风控、社交应用、 推荐系统、舆情监控等。

- 容量型实例

  容量型实例能够提供极高的写吞吐量，能够提供接近于高性能实例的写性能以及更低的使用成本，但读性能和并发能力均低于高性能实例，适用于写多读少，对读性能不敏感但对成本较为敏感的业务，比如日志监控数据、车联网数据、设备数据、时序数据以及物流数据。

注意: 容量型实例不支持预留读/写吞吐量，所有的读写访问均按照按量读/写吞吐量进行计费。

12 云数据库RDS版

12.1 产品概述

阿里云关系型数据库（Relational Database Service，简称RDS）是一种稳定可靠、可弹性伸缩的在线数据库服务。基于阿里云分布式文件系统和高性能存储，云数据库提供了容灾、备份、恢复、监控、迁移等方面的全套解决方案，彻底解决数据库运维的烦恼。

云数据库MySQL版基于Alibaba的MySQL源码分支，经过双11高并发、大数据量的考验，拥有优良的性能和吞吐量。除此之外，云数据库MySQL版还拥有经过优化的读写分离、数据压缩、智能调优等高级功能。

MySQL是全球最受欢迎的开源数据库，作为开源软件组合LAMP（Linux + Apache + MySQL + Perl/PHP/Python）中的重要一环，广泛应用于各类应用。

Web2.0时代，风靡全网的社区论坛软件系统Discuz和博客平台Wordpress均基于MySQL实现底层架构。Web3.0时代，阿里巴巴、Facebook、Google等大型互联网公司都采用更为灵活的MySQL构建了成熟的大规模数据库集群。

12.2 功能特性

12.2.1 数据链路服务

阿里云数据库提供全数据链路服务，包括DNS、负载均衡、Proxy等。因为RDS使用原生的DB Engine，对数据库的操作高度类似，基本没有学习成本。另外，阿里云数据库提供DMS服务，极大的方便了用户访问使用数据库。

DNS

DNS模块提供域名到IP的动态解析功能，以便规避RDS实例IP地址改变带来的影响。在连接池中设置域名后，即使对应的IP地址发生了变化，仍然可以正常访问RDS实例。

例如，某RDS实例的域名为**test.rds.aliyun.com**，对应的IP地址为**10.10.10.1**。某程序连接池中设置为**test.rds.aliyun.com** 或 **10.10.10.1** 都可以正常访问RDS实例。

一旦该RDS实例发生了可用区迁移或者版本升级后，IP地址可能变为**10.10.10.2**。如果程序连接池中设置的是域名**test.rds.aliyun.com**，则仍然可以正常访问RDS实例。但是如果程序连接池中设置的是IP地址**10.10.10.1**，就无法访问RDS实例了。

负载均衡

负载均衡模块提供实例IP地址（包括内网IP 和外网IP），以便屏蔽物理服务器变化带来的影响。

例如，某RDS实例的内网IP地址为**10.1.1.1**，对应的Proxy或者DB Engine运行在**192.168.0.1**上。在正常情况下，负载均衡模块会将访问**10.1.1.1**的流量重定向到**192.168.0.1**上。当**192.168.0.1**发生了故障，处于热备状态的**192.168.0.2**接替了**192.168.0.1**的工作。此时负载均衡模块会将访问**10.1.1.1**的流量重定向到 **192.168.0.2** 上，RDS实例仍旧正常提供服务。

Proxy

Proxy模块提供数据路由、流量探测和会话保持等功能。

- 数据路由功能：支持大数据场景下的分布式复杂查询聚合和相应的容量管理。
- 流量探测功能：降低SQL注入的风险，在必要情况下支持SQL日志的回溯。
- 会话保持功能：解决故障场景下的数据库连接中断问题。

DMS

DMS（Data Management Service）是一个访问管理云端数据的Web服务，提供了数据管理、对象管理、数据流转和实例管理等功能。

DMS支持 MySQL、SQL Server、PostgreSQL和PPAS等数据源。

12.2.2 高可用服务

高可用服务由Detection、Repair、Notice等模块组成，主要保障数据链路服务的可用性，除此之外还负责处理数据库内部的异常。

另外，RDS还通过迁移到支持多可用区的地域和采用适当的高可用策略，提升RDS的高可用服务。

Detection

Detection模块负责检测DB Engine的主节点和备节点是否提供了正常的服务。通过间隔为8~10秒的心跳信息，HA节点可以轻易获得主节点的健康情况，结合备节点的健康情况和其它HA节点的心跳信息，Detection模块可以排除网络抖动等异常引入的误判风险，在30秒内完成异常切换操作。

Repair

Repair模块负责维护DB Engine的主节点和备节点之间的复制关系，还会修复主节点或者备节点在日常运行中出现的错误。

例如：

- 主备复制异常断开的自动修复
- 主备节点表级别损坏的自动修复
- 主备节点Crash的现场保存和自动修复

Notice

Notice模块负责将主备节点的状态变动通知到负载均衡或者Proxy，保证用户访问正确的节点。

例如：Detection模块发现主节点异常，并通知Repair模块进行修复。Repair模块进行了尝试后无法修复主节点，通知Notice进行流量切换。Notice模块将切换请求转发至负载均衡或者Proxy，此时用户流量全部指向备节点。与此同时，Repair在别的物理服务器上重建了新的备节点，并将变动同步给Detection模块。Detection模块开始重新检测实例的健康状态。

多可用区

多可用区是在单可用区的级别上，将同一地域的多个单可用区组合成的物理区域。相对于单可用区RDS实例，多可用区RDS实例可以承受更高级别的灾难。

例如，单可用区RDS实例可以承受服务器和机架级别的故障，而多可用区RDS实例可以承受机房级别的故障。

目前多可用区RDS不额外收取任何费用，在已开通多可用区地域的用户可以直接购买多可用区RDS实例，也可以通过跨可用区迁移将单可用区RDS实例转化成多可用区RDS实例。

注意： 因为多可用区之间存在一定的网络延迟，因此多可用区RDS实例在采用半同步数据复制方案的时候，对于单个更新的响应时间会比单可用区实例长。这种情况最好通过提高并发量的方式来实现整体吞吐量的提高。

高可用策略

高可用策略是根据用户自身业务的特点，采用服务优先级和数据复制方式之间的不同组合，以组合出适合自身业务特点的高可用策略。

服务优先级有以下两个级别：

- RTO（Recovery Time Objective）优先：数据库应该尽快恢复服务，即可用时间最长。对于数据库在线时间要求比较高的用户应该使用 TO优先策略。
- RPO（Recovery Point Objective）优先：数据库应该尽可能保障数据的可靠性，即数据丢失量最少。对于数据一致性要求比较高的用户应该使用RPO优先策略。

数据复制方式有以下三种方式：

- 异步复制（Async）：应用发起更新（含增加、删除、修改操作）请求，Master完成相应操作后立即响应应用，Master向Slave异步复制数据。因此异步复制方式下，Slave不可用不影响主库上的操作，而Master不可用有较小概率会引起数据不一致。
- 强同步复制（Sync）：应用发起更新（含增加、删除、修改操作）请求，Master完成操作后向Slave复制数据，Slave接收到数据后向Master返回成功信息，Master接到Slave的反馈后再响应应用。Master向Slave复制数据是同步进行的，因此Slave不可用会影响Master上的操作，而Master不可用不会引起数据不一致。
- 半同步复制（Semi-sync）：正常情况下数据复制方式采用强同步复制方式，当Master向Slave复制数据出现异常的时候（Slave不可用或者双节点间的网络异常），Master会暂停对应用的响应，直到复制方式超时退化成异步复制。如果允许应用在此时更新数据，则Master不可用会引起数据不一致。当双节点间的数据复制恢复正常（Slave恢复或者网络恢复），异步复制会恢复成强同步复制。

用户可以根据自身业务特点，选择服务优先级和数据复制方式的不同组合方式，提高可用性，不同组合方式的特点如表 *14:* 不同组合的特点所示。

表 14: 不同组合的特点

云数据引擎	服务优先级	数据复制方式	组合特点
MySQL 5.6	RPO	Async	在Master发生故障的情况下，切换会发生在Slave应用完所有的Relay Log之后。 在Slave发生故障的情况下，应用操作Master不受影响。在Slave恢复之后再同步Master上面的数据。
MySQL 5.6	RTO	Semi-sync	在Master发生故障且数据复制未退化的情况下，因为数据一致性已经得到保障，RDS将立即触发切换操作把流量导向Slave。 在Slave发生故障的情况下，应用操作Master将会出现超时，而后数据复制方式退化为异步复制方式；在Slave恢复并同步完Master上的数据之后，数据复制方式恢复为强同步。 在双节点数据不一致且数据复制方式已经退化为异步复制方式的情况下，如果Master发生了故障，则切换会发生在Slave应用完所有的Relay Log之后。
MySQL 5.6	RPO	Semi-sync	在Master发生故障且数据复制未退化的情况下，因为数据一致性已经得到保障，RDS将立即触发切换操作把流量导向Slave。

云数据引擎	服务优先级	数据复制方式	组合特点
			在Slave发生故障的情况下，应用操作Master将会出现超时，而后数据复制方式退化为异步复制方式；在Slave重新获取到Master信息时（Slave恢复或者网络故障恢复），数据复制方式恢复为强同步方式。 在双节点数据不一致且Slave上的数据差异无法补全的情况下，如果Master发生了故障，则用户可以通过API获取Slave的时间点并决定何时切换以及补全数据的方法。

12.2.3 备份恢复服务

备份恢复服务主要提供数据的离线备份、转储和恢复。

Backup

Backup模块负责将主备节点上面的数据和日志压缩和上传。RDS默认将备份上传到OSS中，在特定场景下还支持将备份文件转储到更加廉价和持久的归档存储上。在备节点正常运作的情况下，备份总是在备节点上面发起，以避免对主节点提供的服务带来冲击；在备节点不可用或者损坏的情况下，Backup模块会通过主节点创建备份。

Recovery

Recovery模块负责将OSS上面的备份文件恢复到目标节点上。

- 回滚主节点功能：客户发起数据相关的误操作后可以通过回滚功能按时间点恢复数据。
- 修复备节点功能：在备节点出现不可修复的故障时自动新建备节点来降低风险。
- 创建只读实例功能：通过备份来创建只读实例。

Storage

Storage模块负责备份文件的上传、转储和下载。目前备份数据全部上传至OSS进行存储，您可以根据需要获取临时链接来下载。在某些特定场景下，Storage模块支持将OSS上面的备份文件转储至归档存储来提供更长时间和更低费用的离线存储。

12.2.4 监控服务

云数据库提供物理层、网络层、应用层等多方位的监控服务，保证业务可用性。

Service

Service 模块负责服务级别的状态跟踪，监控负载均衡、OSS、归档存储和日志服务等RDS依赖的其他云产品是否正常，包括功能和响应时间等。另外对RDS内部的服务，Service也会通过日志来判定是否正常运作。

Network

Network模块负责网络层面的状态跟踪，包括ECS与RDS之间的连通性监控，RDS物理机之间的连通性监控，路由器和交换机的丢包率监控。

OS

OS模块负责硬件和OS内核层面的状态跟踪，包括：

- 硬件检修：不断检测CPU、内存、主板、存储等设备的工作状态，预判是否会发生故障，并提前进行自动报修。
- OS内核监控：跟踪数据库的所有调用，并从内核态分析调用缓慢或者出错的原因。

Instance

Instance模块负责RDS实例级别的信息采集，包括：

- 实例的可用信息
- 实例的容量和性能指标
- 实例的SQL执行记录

12.2.5 调度服务

调度服务由Resource模块和Version模块组成，主要提供资源调配和实例版本管理。

Resource

Resource模块主要负责RDS底层资源的分配和整合，对您而言就是实例的开通和迁移。例如，您通过RDS控制台或者API创建实例，Resource模块会计算出最适合的物理服务器来承载流量。RDS实例跨可用区迁移所需的底层资源也由Resource负责分配和整合。在经过长时间的实例创建、删除和迁移后，Resource模块会计算可用区内的资源碎片化程度，并定期发起资源整合以提高可用区的服务承载量。

Version

Version模块主要负责RDS实例的版本升级。例如：

- MySQL大版本升级：MySQL 5.1升级至MySQL 5.5，MySQL 5.5升级至MySQL 5.6等。

- MySQL小版本升级：MySQL源码存在的bug修复。

12.2.6 迁移服务

迁移服务主要帮助用户把数据从本地数据库迁移到阿里云数据库，或者把阿里云数据库的一个实例迁移到另一实例中。云数据库提供了DTS（Data Transmission Service）工具，方便用户快速的迁移数据库。

DTS

DTS是一个云上的数据传输服务，能快速的将本地数据库或者RDS中的实例迁移到另一个RDS实例中。目前DTS支持MySQL、SQL Server和PostgreSQL三种数据库。

DTS还提供了三种迁移模式，分别为结构迁移、全量迁移和增量迁移。

- 结构迁移

  DTS会将迁移对象的结构定义迁移到目标实例，目前支持结构迁移的对象有表、视图、触发器、存储过程和存储函数。

- 全量迁移

  DTS会将源数据库迁移对象已有数据全部迁移到目标实例中。

  注意: 在全量迁移过程中，为了保证数据一致性，无主键的非事务表会被锁定。锁定期间这些表无法写入，锁定时长依赖于这些表的数据量大小。在这些无主键非事务表迁移完成后，锁才会释放。

- 增量迁移

  DTS会将迁移过程中数据变更同步到目标实例。

  注意: 如果迁移期间进行了DDL操作，这些结构变更不会同步到目标实例。

12.3 产品优势

12.3.1 易于使用

即开即用

您可以通过API进行RDS规格定制，创建后RDS实时生产出目标实例。

按需升级

随着数据库压力和数据存储量的变化，您可以灵活调整实例规格，且升级期间RDS不会中断数据链路服务。

透明兼容

RDS与原生数据库引擎的使用方法一致，您无需二次学习，上手即用。另外RDS兼容您现有的程序和工具。使用通用的数据导入导出工具即可将数据迁移到RDS，迁移过程中的人力开销非常低。

管理便捷

阿里云负责RDS的日常维护和管理，包括但不限于软硬件故障处理、数据库补丁更新等工作，保障RDS运转正常。您也可自行通过阿里云控制台完成数据库的增加、删除、重启、备份、恢复等管理操作。

12.3.2 高性能

参数优化

阿里云聚集国内顶尖的数据库专家，所有RDS实例的参数都是经过多年的生产实践优化而得。在RDS实例的生命周期内，DBA持续对其进行优化，确保RDS一直基于最佳实践在运行。

SQL 优化建议

针对您的应用场景特点，RDS会锁定效率低下的SQL语句并提出优化建议，以便您优化业务代码。

高端硬件投入

RDS使用的所有服务器硬件都经过多方评测，保证在性能和稳定性上都遥遥领先。

12.3.3 高安全性

防DDoS攻击

说明: 此功能需开通阿里云的安全产品。

当您使用外网连接和访问RDS实例时，可能会遭受DDoS攻击。当RDS安全体系认为您的实例正在遭受DDoS攻击时，会首先启动流量清洗的功能，如果流量清洗无法抵御攻击或者攻击达到黑洞阈值时，将会进行黑洞处理。

流量清洗和黑洞处理的方法及触发条件如下：

- 流量清洗

只针对外网流入流量进行清洗，处于流量清洗状态的RDS实例可正常访问。

流量清洗的触发和结束由系统自动完成，单个RDS实例满足以下任一条件即触发流量清洗：

- PPS（Package Per Second）达到3万；
- BPS（Bits Per Second）达到180Mb；
- 每秒新建并发连接达到1万；
- 激活并发连接数达到1万；
- 非激活并发连接数达到10万。

- 黑洞处理

  只针对外网流入流量进行黑洞处理，处于黑洞状态的RDS实例不可被外网访问，此时应用程序通常也处于不可用状态。黑洞处理是保证RDS整体服务可用性的一种手段。

  黑洞触发条件如下：

  - BPS（Bits Per Second）达到2GB；
  - 流量清洗无效。

  黑洞结束条件为：黑洞在2.5小时后自动解除。

  说明: 建议您通过内网访问RDS实例，可以使RDS实例免受DDoS攻击的风险。

访问控制策略

用户可定义允许访问RDS的IP地址，指定之外的IP地址将被拒绝访问。

每个账号只能看到、操作自己的数据库。

系统安全

RDS处于多层防火墙的保护之下，可以有力地抗击各种恶意攻击，保证数据的安全。

RDS服务器不允许直接登录，只开放特定的数据库服务所需要的端口。

RDS服务器不允许主动向外发起连接，只能接受被动访问。

12.3.4 高可靠性

双机热备

RDS采用热备架构，物理服务器出现故障后服务秒级完成切换。整个切换过程对应用透明。

多副本冗余

RDS 服务器中的数据构建于RAID之上，数据备份存储在OSS上。

数据备份

RDS提供自动备份的机制。您可以自行选择备份周期，也可以根据自身业务特点随时发起临时备份。

数据恢复

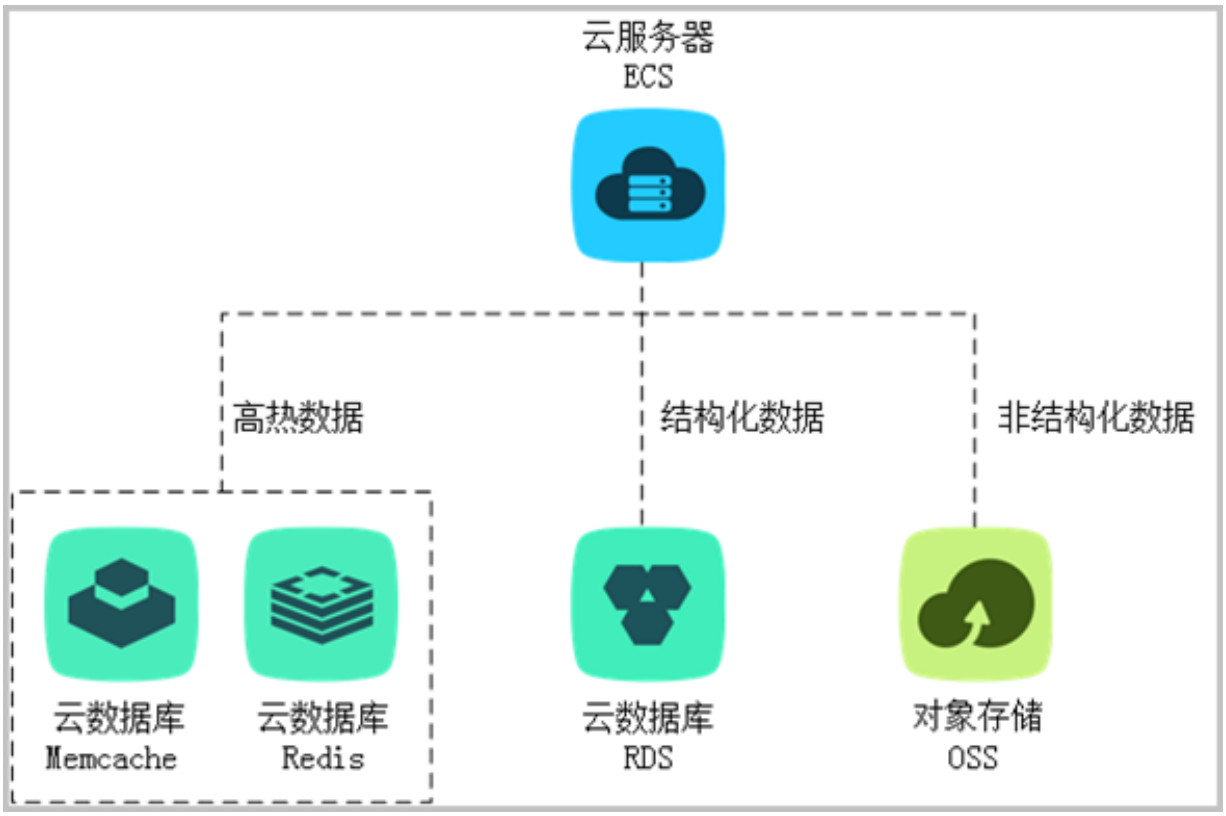
支持按备份集和指定时间点的恢复。在大多数场景下，您可以将7天内任意一个时间点的数据恢复到RDS临时实例上，数据验证无误后即可将数据迁回RDS主实例，从而完成数据回溯。

12.4 典型应用

12.4.1 数据多样化存储

RDS支持搭配云数据库Memcache版、云数据库Redis版和对象存储OSS等存储产品使用，实现多样化存储扩展，如图 *31:* 数据多样化存储所示。

图 31: 数据多样化存储



缓存数据持久化

RDS可以和云数据库Memcache版和云数据库Redis版搭配使用，组成高吞吐、低延迟的存储解决方案。与RDS相比，云数据库缓存产品有两个特性：

- 响应速度快，云数据库Memcache版和云数据库Redis版请求的时延通常在几毫秒以内。
- 缓存区能够支持比 RDS更高的QPS（每秒处理请求数）。

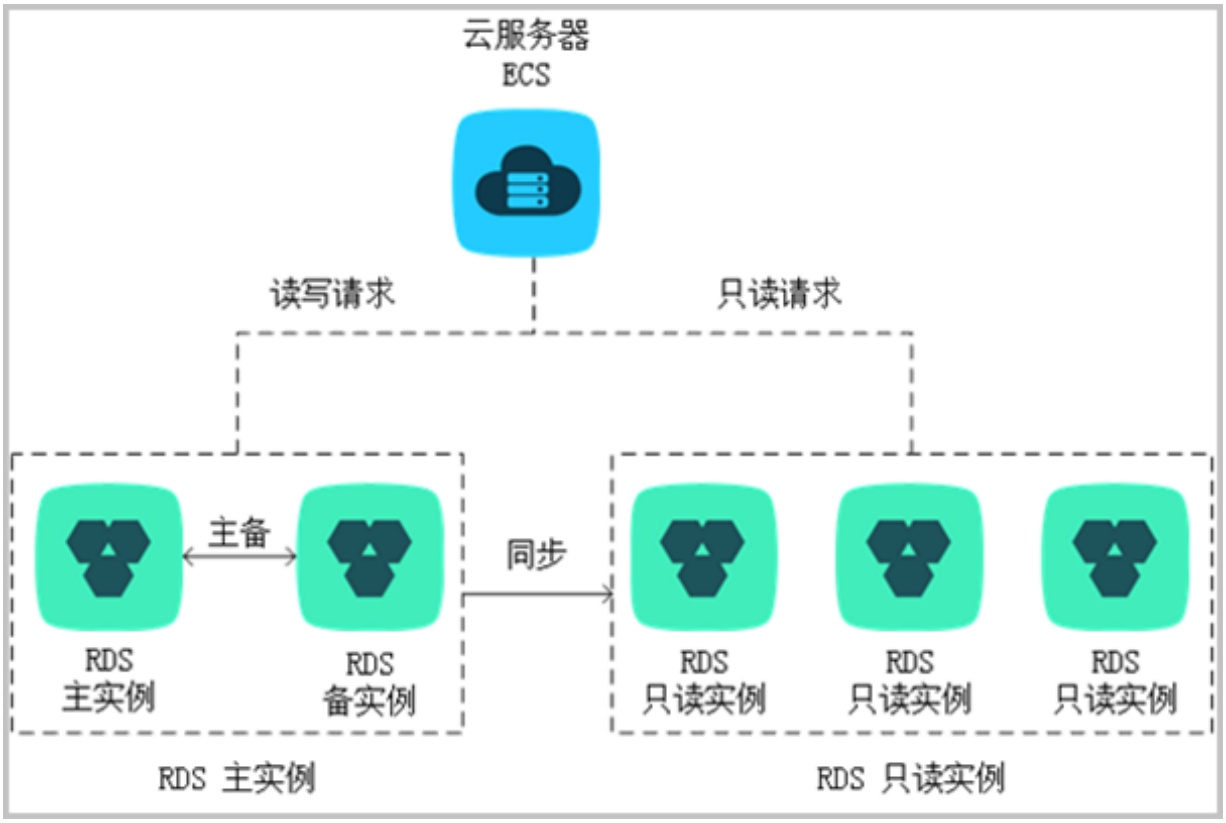
多结构数据存储

OSS是阿里云对外提供的海量、安全、低成本、高可靠的云存储服务。RDS可以和OSS搭配使用，组成多类型数据存储解决方案。例如，当业务应用为论坛时，RDS搭配OSS使用，注册用户的图像、帖子内容的图像等资源可以存储在OSS中，以减少RDS的存储压力。

12.4.2 读写分离

云数据库MySQL版支持直接挂载只读实例，分担主实例读取的压力。每个只读实例有独立的连接串，可由应用端自动分配读取压力，如图 *32:* 读写分离所示。

图 32: 读写分离

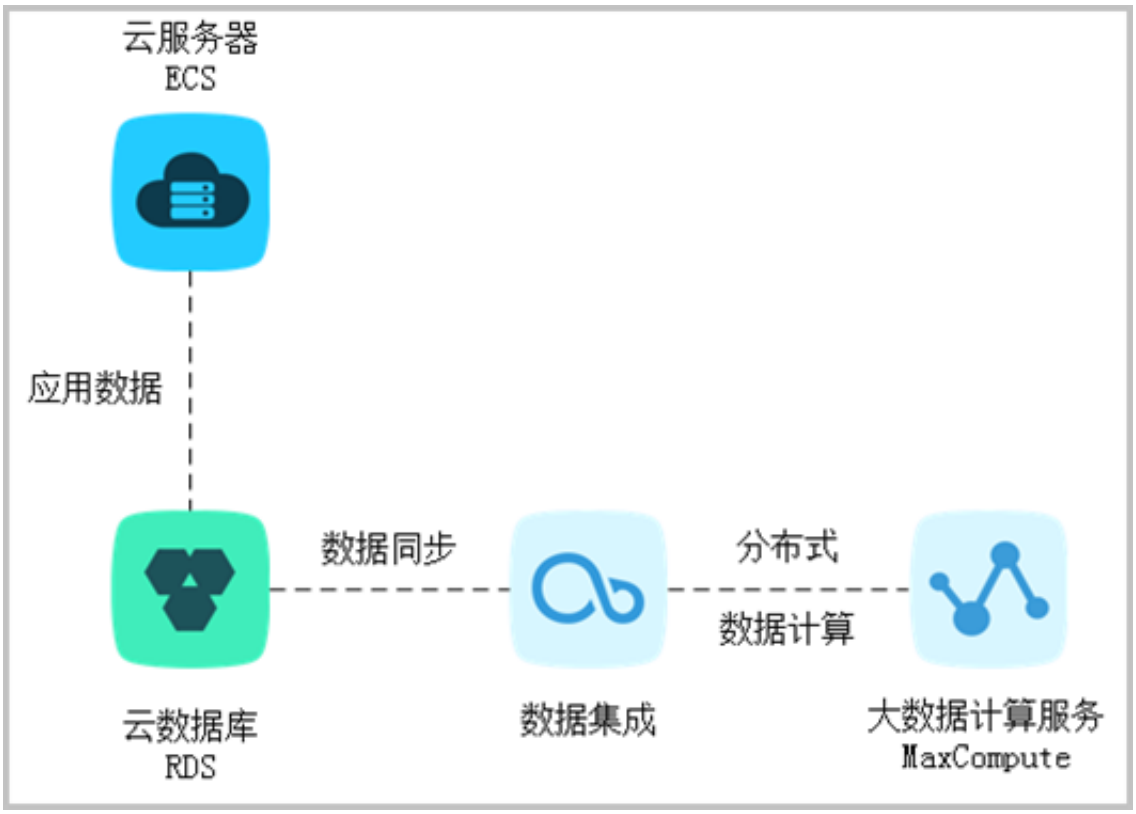


12.4.3 大数据分析

开放数据处理服务又称大数据计算服务（MaxCompute，原名ODPS），可服务于批量结构化数据的存储和计算，提供海量数据仓库的解决方案以及针对大数据的分析建模服务。

通过数据集成服务，可将RDS数据导入MaxCompute，实现大规模的数据计算，如图 *33:* 大数据分析所示。

图 33: 大数据分析



13 云数据库Redis版

13.1 产品概述

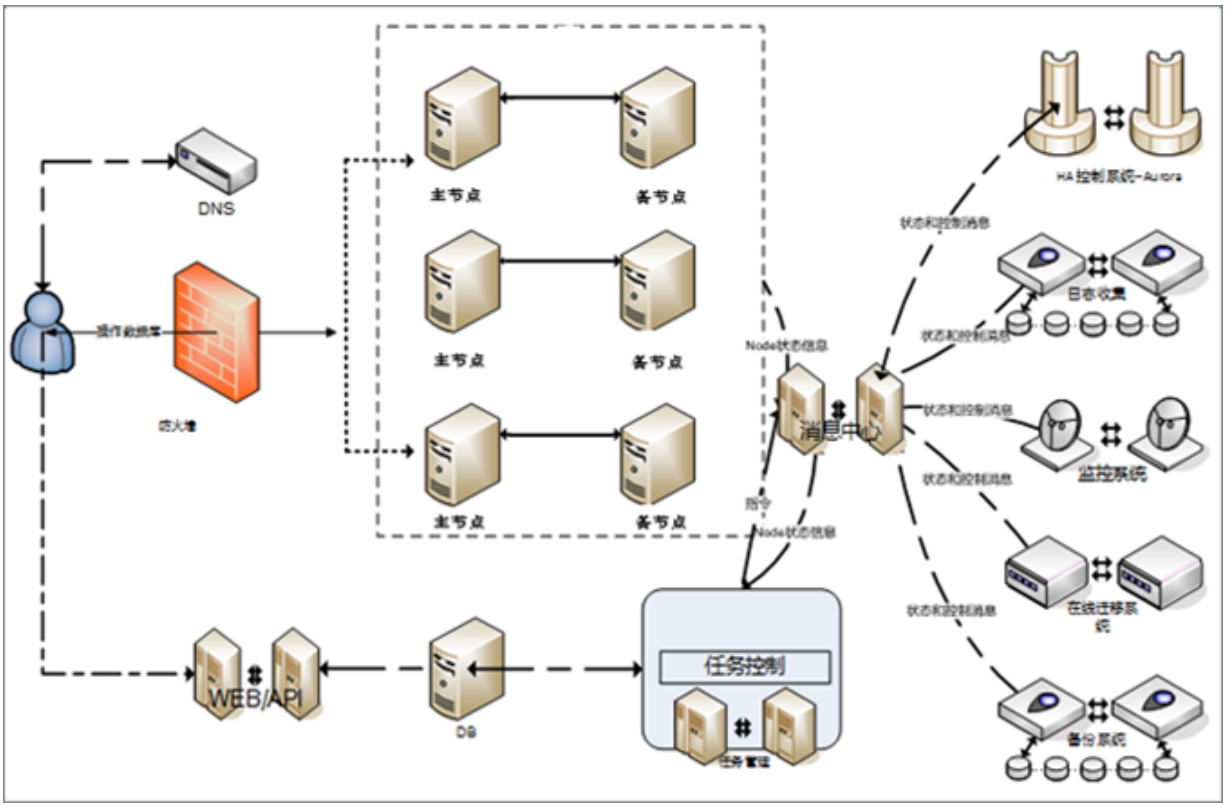
阿里云数据库 Redis 版（ApsaraDB for Redis）是兼容开源 Redis 协议的 Key-Value 类型在线存储服务。它支持字符串（String）、链表（List）、集合（Set）、有序集合（SortedSet）、哈希表（Hash）等多种数据类型，及事务（Transactions）、消息订阅与发布（Pub/Sub）等高级功能。通过“内存+硬盘”的存储方式，云数据库Redis版在提供高速数据读写能力的同时满足数据持久化需求。

除此之外，云数据库 Redis 版作为云计算服务，其硬件和数据部署在云端，有完善的基础设施规划、网络安全保障、系统维护服务。所有这些都无需用户考虑，确保用户专心致力于自身业务创新。

13.2 产品架构

云数据库 Redis 版的基础架构图如图 *34:* 架构图所示。

图 34: 架构图



云数据库Redis版自动搭建好主备双节点结构供用户使用。

- **HA 控制系统**

  实例高可用探测模块，用于探测监听Redis实例运行状况。如果判断为主节点实例不可用，进行主备节点的切换操作，保证Redis实例的高可用。

- **日志收集**

  进行Redis运行情况的日志收集，包括实例慢查询日志，访问控制日志等。

- **监控系统**

  进行Redis实例性能监控信息的收集工作，目前包括基本信息组监控，keys组信息监控，String信息组监控等核心信息。

- **在线迁移系统**

  当实例所运行的物理机出现故障，在线迁移系统会根据备份系统中的备份文件进行实例重新搭建，保证业务不受影响。

- **备份系统**

  针对Redis实例进行备份处理，并且将生成的备份文件存储至OSS系统上进行保存。目前Redis备份系统支持用户自定义备份设置，临时备份并且保存7天内的备份文件。

- **任务控制**

  云数据库Redis版实例支持多种管理控制任务，如创建实例，变更配置，备份实例等，任务系统会根据用户下发的操作指令，进行灵活控制并且进行任务跟踪及出错管理。

13.3 规格说明

说明: 带宽上限是出入流量的总和。

表 15: 标准套餐

规格	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
1 GB 主从版	10000	10	单核	主-从双节点实例
2 GB 主从版	10000	16	单核	主-从双节点实例
4 GB 主从版	10000	24	单核	主-从双节点实例
8 GB 主从版	10000	24	单核	主-从双节点实例

规格	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
16 GB 主从版	10000	32	单核	主-从双节点实例
32 GB 主从版	10000	32	单核	主-从双节点实例
64 GB 主从版	20000	48	单核	主-从双节点实例

13.4 规格性能

说明: 带宽上限是出入流量的总和。

标准版-双副本

表 16: 标准套餐

规格	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
1 GB 主从版	10000	10	单核	主-从双节点实例
2 GB 主从版	10000	16	单核	主-从双节点实例
4 GB 主从版	10000	24	单核	主-从双节点实例
8 GB 主从版	10000	24	单核	主-从双节点实例
16 GB 主从版	10000	32	单核	主-从双节点实例
32 GB 主从版	10000	32	单核	主-从双节点实例
64 GB 主从版	20000	48	单核	主-从双节点实例

表 17: 定制套餐

规格	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
1 GB 主从高配版	20000	48	单核	主-从双节点实例
2 GB 主从高配版	20000	48	单核	主-从双节点实例
4 GB 主从高配版	20000	48	单核	主-从双节点实例
8 GB 主从高配版	20000	48	单核	主-从双节点实例
16 GB 主从高配版	20000	48	单核	主-从双节点实例

规格	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
32 GB 主从高配版	20000	48	单核	主-从双节点实例

标准版-单副本

表 18: 标准套餐

规格	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
1 GB 单机版	10000	10	单核	单节点实例
2 GB 单机版	10000	16	单核	单节点实例
4 GB 单机版	10000	24	单核	单节点实例
8 GB 单机版	10000	24	单核	单节点实例
16 GB 单机版	10000	32	单核	单节点实例
32 GB 单机版	10000	32	单核	单节点实例
64 GB 单机版	20000	48	单核	单节点实例

表 19: 定制套餐

规格	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
1 GB 单机高配版	20000	48	单核	单节点实例
2 GB 单机高配版	20000	48	单核	单节点实例
4 GB 单机高配版	20000	48	单核	单节点实例
8 GB 单机高配版	20000	48	单核	单节点实例
16 GB 单机高配版	20000	48	单核	单节点实例
32 GB 单机高配版	20000	48	单核	单节点实例

集群版

规格（GB）	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	说明
16 GB 集群版	80000	384	8核	高性能集群实例
32 GB 集群版	80000	384	8核	高性能集群实例
64 GB 集群版	80000	384	8核	高性能集群实例
128 GB 集群版	160000	768	16核	高性能集群实例
256 GB 集群版	160000	768	16核	高性能集群实例

QPS 能力参考

表 20: QPS 能力参考

规格（GB）	连接数上限（个）	内网带宽上限（MByte）	CPU 处理能力	QPS 参考值
8	10000	24	单核	80000

说明: 非集群版实例的 QPS 能力参考范围为8-10万， 集群实例的 QPS 参考值为节点数目乘以8-10万。

测试场景说明

图 35: 网络拓扑图

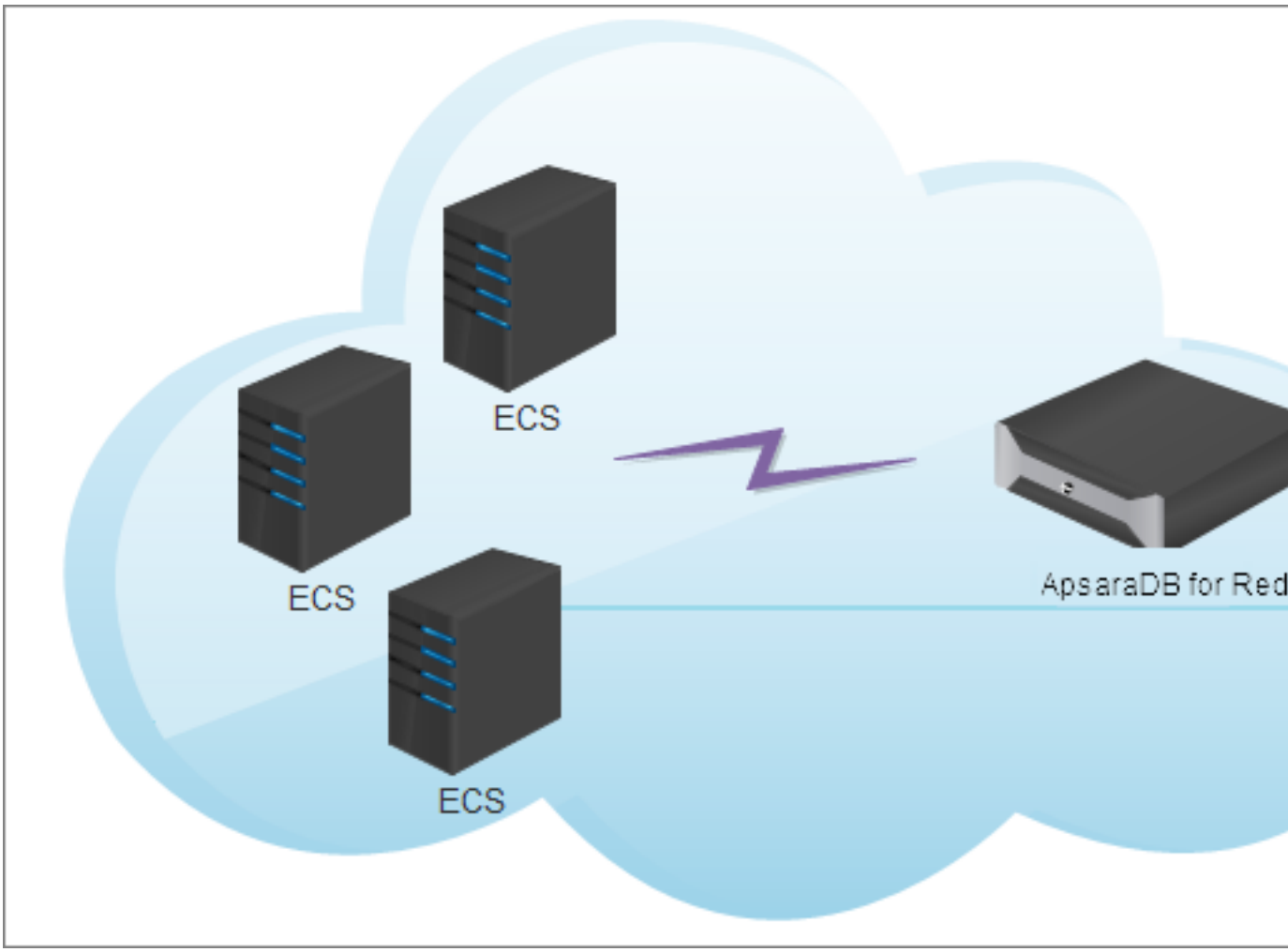


表 21: 云主机规格

操作系统	CPU (个数)	内存	区域	个数
Ubuntu 14.04 64位	1	2048 MB	华南1	3

1. 在3台 ECS 上下载 并安装redis-2.8.19 源码包。

```
$ wget http://download.redis.io/releases/redis-2.8.19.tar.gz
$ tar xzf redis-2.8.19.tar.gz
$ cd redis-2.8.19
$ make
$ make install
```

2. 在3台 ECS 上同时执行以下命令。

```
redis-benchmark -h **********.m.cnsza.kvstore.aliyuncs.com -p 6379 -a password -t set -c 50 -d 128 -n 25000000 -r 5000000
```

3. 汇总3台 ECS 上的测试数据，QPS 为3台 ECS 总和。

13.5 功能特性

支持丰富的数据类型

兼容开源 Redis 协议中定义的所有数据类型，如String、Hash、List、Set、SortedSet等，支持多种数据操作，充分满足业务需求。

持久化存储

内存+硬盘的存储方式，数据存储到物理磁盘，满足用户数据持久化需求。

支持消息通知机制

基于事件通知机制解耦消息发布者和消息订阅者之间的耦合，实现消息发布及订阅（PUB/SUB）功能，满足多个客户端使用者之间的互联互通。

支持事务操作

支持Redis协议中定义的事务（Transaction）处理，实现单个客户端发送的多个命令组成的原子性操作。

13.6 功能特性

架构灵活

- 单节点架构

  单节点架构适用于纯缓存场景，支持单节点集群弹性变配，满足高 QPS 场景，提供超高性价比。

- 双机热备架构

  系统工作时主节点（Master）和备节点（Slave）数据实时同步，主节点故障时系统自动秒级切换，备节点接管业务，全程自动且对业务无影响，主备架构保障系统服务具有高可用性。

- 集群架构

  集群（cluster）实例采用分布式架构，每个节点都采用一主一从的高可用架构，自动容灾切换，故障迁移，多种集群规格可适配不同的业务压力，无限扩展数据库性能。

数据安全

- 备份及一键恢复

  每天自动备份数据，数据容灾能力强，免费支持数据一键恢复，有效防范数据误操作，业务损失降到最低。

- 多层网络安全防护

  VPC 私有网络在 TCP 层直接进行网络隔离保护；DDOS 防护实时监测并清除大流量攻击；支持1000个以上 IP 白名单配置

- 深度内核优化

  阿里云专家团队对源码 Redis 进行深度内核优化，有效防止内存溢出，修复安全漏洞，为您保驾护航。

弹性扩展

- 数据容量扩展

  云数据库 Redis 版支持多种内存规格的产品配置，可根据业务量大小进行自由升级内存规格。

- 性能扩展

  支持集群架构下弹性无限扩展数据库系统的存储空间及吞吐性能，突破海量数据高 QPS 性能瓶颈，轻松应对每秒百万次的读写需求。

- 业务形态扩展

  支持单节点缓存架构和双节点存储架构，适配不同业务场景，支持标准版和双节点版之间的灵活变配。

智能运维

- 监控平台

  提供 CPU 利用率、连接数、磁盘空间利用率等实例信息实时监控及报警，随时随地了解实例动态。

- 可视化管理平台

  管理控制平台对实例克隆、备份、数据恢复等高频高危操作可便捷的进行一键式操作。

- 可视化 DMS 平台

  专业的 DMS 数据管理平台，提供可视化的数据管理，全面提升研发、运维效率。

- 数据库内核版本管理

  主动升级，快速修复缺陷，免去日常版本管理苦恼；优化 Redis 参数配置，最大化利用系统资源。

13.7 产品优势

集群功能

- 可支持超大容量，超高性能。支持集群功能，提供128 G及以上集群实例规格，可满足大容量和高性能需求。
- 提供64 G及以下的主-从双节点实例，满足一般用户的容量和性能需求。

弹性扩容

- 存储容量一键扩容：您可根据业务需求通过控制台对实例存储容量进行调整。
- 在线扩容不中断服务：调整实例存储容量可在线进行，无需停止服务，不影响您自身业务。

资源隔离

针对实例级别的资源隔离，可以更好地保障单个用户服务的稳定性。

安全可靠

- 数据持久化存储：内存+硬盘的存储方式，在提供高速数据读写能力的同时满足数据持久化需求。
- 数据主从双备份：所有数据在主从节点上进行双备份。
- 支持密码认证方式以确保访问安全可靠。

高可用

- 每个实例均有主从双节点：避免单点故障引起的服务中断。
- 硬件故障自动检测与恢复：自动侦测硬件故障并在数秒内切换，恢复服务。

秒级别监控

- 提供秒级别实时监控，分钟级别历史监控。
- 提供各数据结构各接口的监控信息，访问情况一目了然，便于您对云数据库 Redis 版的使用情况有充分的了解。

简单易用

- 服务开箱即用：支持即开即用的方式，购买之后即可使用，方便业务快速部署。
- 兼容开源 Redis：兼容 Redis 命令，任何 Redis 客户端都可以轻松与云数据库 Redis 版建立连接进行数据操作。
- 可视化的管理监控面板：控制台提供多项监控统计信息，方便您对 Redis 实例进行管理。

13.8 产品优势

性能卓越

- 集群功能可支持超大容量，超高性能。支持集群功能，提供128 GB 及以上集群实例规格，可满足大容量和高性能需求。
- 提供 64 GB 及以下的主-从双节点实例，满足一般用户的容量和性能需求。

弹性扩容

- 存储容量一键扩容：用户可根据业务需求通过控制台对实例存储容量进行调整。
- 在线扩容不中断服务：调整实例存储容量可在线进行，无需停止服务，不影响用户自身业务。

数据安全

- 数据持久化存储：内存+硬盘的存储方式，在提供高速数据读写能力的同时满足数据持久化需求。
- 数据主从双备份：所有数据在主从节点上进行双备份。
- 支持密码认证方式以确保访问安全可靠。

高可用

- 双副本与集群版实例均有主从双节点，避免单点故障引起的服务中断。
- 硬件故障自动检测与恢复：自动侦测硬件故障并在数秒内切换，恢复服务。
- 实例级别的资源隔离可以更好地保障单个用户服务的稳定性。

秒级别监控

- 提供秒级别实时监控，分钟级别历史监控。
- 提供各数据结构和接口的监控信息，访问情况一目了然，便于用户对云数据库 Redis 版的使用情况有充分的了解。

简单易用

- 服务开箱即用：支持即开即用的方式，购买之后即可使用，方便业务快速部署。
- 兼容开源 Redis：兼容 Redis 命令，任何 Redis 客户端都可以轻松与云数据库 Redis 版建立连接进行数据操作。
- 可视化的管理监控面板：控制台提供多项监控统计信息，方便用户对 Redis 实例进行管理。

13.9 典型应用

游戏行业应用

游戏行业可以选择云数据库 Redis 版作为重要的部署架构组件。

场景一：**Redis 作为存储数据库使用**

游戏部署架构相对简单，主程序部署在ECS上，所有业务数据存储在 Redis 中，作为持久化数据库。云数据库Redis版支持持久化功能，主备双机冗余数据存储。

场景二：**Redis 作为缓存加速应用访问**

Redis 作为缓存层，加速应用访问。数据存储在后端的数据库中（RDS）。

Redis 的服务可靠性至关重要，一旦 Redis 服务不可用，将导致后端数据库无法承载业务访问压力。云数据库Redis版提供双机热备的高可用架构，保障极高的服务可靠性。主节点对外提供服务，当主节点出现故障，系统自动切换备用节点接管服务，整个切换过程对用户全部透明。

视频直播类应用

视频直播类业务往往会重度依赖 Redis 业务。存储用户数据及好友互动关系。

双机热备保障高可用

云数据库 Redis 版提供双机热备的方式，可以极大的提供服务可用性保障。

集群版解决性能瓶颈

云数据库 Redis 提供集群版实例，破除 Redis 单线程机制的性能瓶颈，可以有效的应对视频直播类流量突起，对于高性能的需求可以有效的支撑。

轻松扩容应对业务高峰

云数据库 Redis 版可支持一键扩容，整个升级过程对用户全透明，可以从容应对流量突发对业务产生的影响。

电商行业应用

电商行业中对于 Redis 大量使用，多数在商品展示、购物推荐等模块。

场景一：**秒杀类购物系统**

大型促销秒杀系统，系统整体访问压力非常大，一般的数据库根本无法承载这样的读取压力。

云数据库 Redis 版支持持久化功能，可以直接选择 Redis 作为数据库系统使用。

场景二：**带有计数系统的库存系统**

底层用 RDS 存储具体数据信息，数据库字段中存储具体计数信息。数据库 Redis 来进行计数的读取，RDS 存储计数信息。云数据库 Redis 版部署在物理机上，底层基于 SSD 高性能存储，可以提供极高的数据存储能力。

13.10 基本概念

Redis

Redis 是一款依据 BSD 开源协议发行的高性能 Key-Value 存储系统（cache and store）。

实例 ID

实例对应一个用户空间，是使用 Redis 的基本单位。

Redis 对单个实例根据不同的容量规格有不同的连接数、带宽、CPU 处理能力等限制。用户可在控制台中看到自己购买的实例 ID 列表。Redis 实例分为主-从双节点实例和高性能集群实例两种。

主-从双节点实例

指具备主-从架构的云数据库 Redis 版实例。主-从双节点能扩展的容量和性能有限。

高性能集群实例

指具有集群扩展性的云数据库 Redis 版实例。集群实例有更好的扩展性和性能，但是在功能上也有一定的限制。

连接地址

用于连接云数据库 Redis 版的 Host 地址。以域名方式展示，可在**实例信息** > **连接信息**中查询到。

连接密码

用于连接云数据库 Redis 版的密码。密码拼接方法为：实例 ID：自定义密码。比如，在购买时设置的密码为1234，分配的实例 ID 为xxxx，那么密码即为xxxx:1234。

逐出策略

与 Redis 的逐出策略保持一致。

DB

即 Redis 中的 Database。云数据库 Redis 版支持256个 DB，默认写入到第0个 DB 中。

13.11 Redis 小版本最新特性介绍

本文会介绍云数据库 Redis 版的最新版本内核提供的产品特性及功能。

云数据库 Redis 版由资深阿里云专家对内核进行深度优化，修复安全漏洞，提升服务稳定性。同时由于客户需求的不断演进，云数据库 Redis 版也会通过对内核版本的优化逐步开放一些产品功能及 Redis 的原生命令的支持。

您可以根据相应特性在控制台上一键操作将内核版本升级至最新版本。升级内核版本会出现30s内的连接闪断，请您在业务低峰期运行，并确保应用程序具备重连机制。

说明: 用户在控制台自主升级内核版本功能预计在2017年6月底支持。

白名单

- 标准版-双节点、标准版-单节点配置支持用户自定义白名单。
- 集群版本暂时不支持。

GEO 功能

云数据库 Redis 版目前的版本为2.8，为了跟随 Redis 开源社区的发展脚步，云数据库 Redis 版目前已经全面支持 Redis 社区3.2版本的 GEO 功能。

Config Get 命令

Config Get 命令放开限制。

LUA 支持

LUA 脚本放开限制，标准版-双节点、标准版-单节点支持用户直接调用。

集群版本条件性支持：

- 所有 key 都应该由 KEYS 数组来传递，redis.call/pcall 里面调用的 redis命令，key 的位置，必须是 KEYS array, 否则直接返回 error。

  ```
  `"-ERR bad lua script for redis cluster, all the keys that the script uses should be passed using the KEYS array\r\n"`
  ```

- 所有 key，必须在1个 slot 上，否则直接返回 error。

  ```
  "-ERR eval/evalsha command keys must in same slot\r\n"
  ```

Client list 命令

Client list 命令放开限制，标准版-双节点，标准版-单节点支持用户调用，集群版本暂时不支持。

14 云数据库Memcache版

14.1 产品概述

云数据库 Memcache 版是基于内存的缓存服务，支持海量小数据的高速访问。云数据库 Memcache 版可以极大缓解对后端存储的压力，提高网站或应用的响应速度。

云数据库 Memcache 版支持 Key-Value 的数据结构，兼容 Memcached 协议的客户端都可与云数据库 Memcache 版进行通信。

云数据库 Memcache 版支持即开即用的方式快速部署。对于动态 Web、APP 应用，可通过缓存服务减轻对数据库的压力，从而提高网站整体的响应速度。

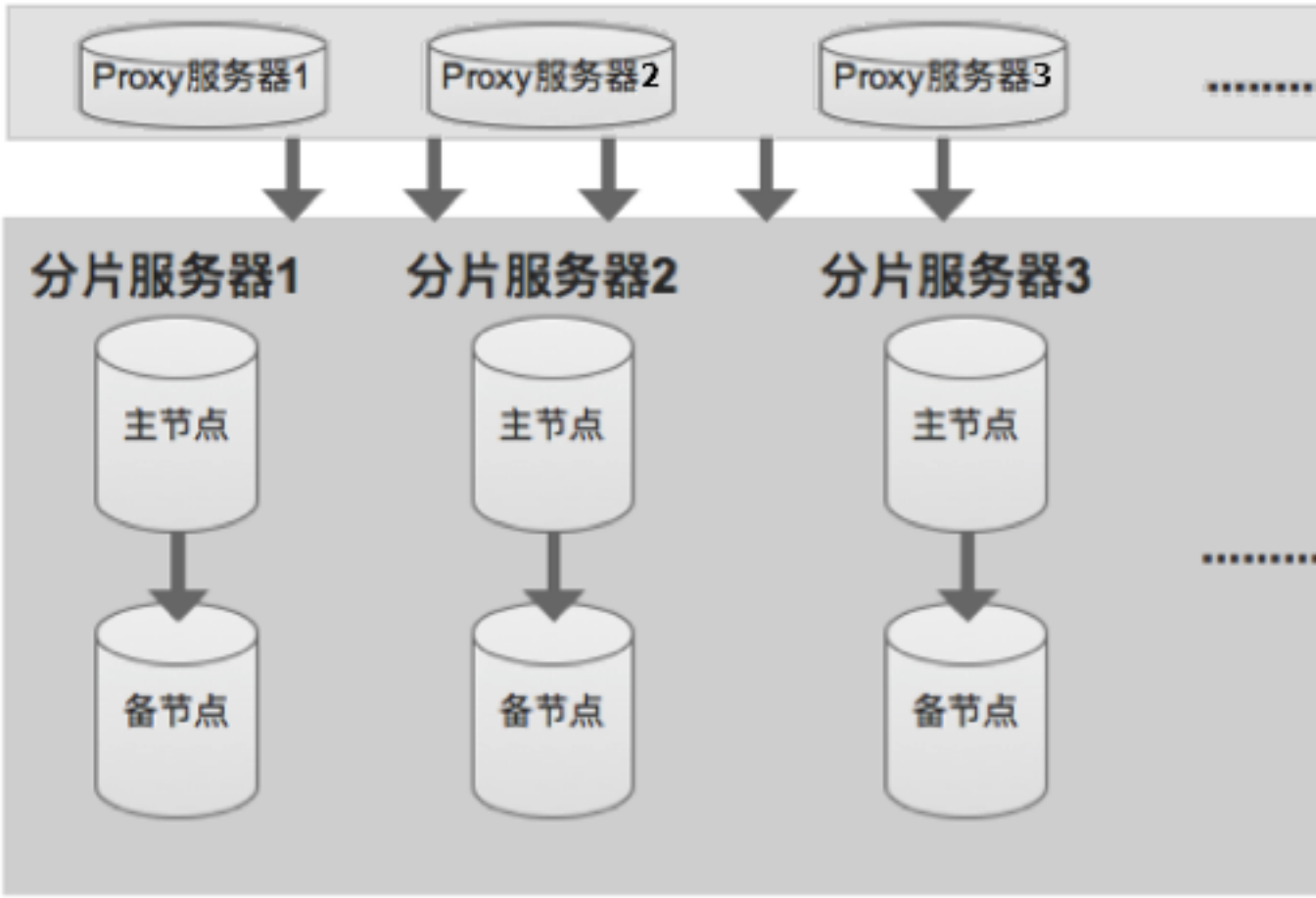
与本地自建 Memcached 相同之处在于云数据库 Memcache 版同样兼容 Memcached 协议，与用户环境兼容，可直接使用。不同之处在于硬件和数据部署在云端，有完善的基础设施、网络安全保障、系统维护服务。所有的这些服务，都不需要投资，只需根据使用量进行付费即可。

14.2 系统架构

云数据库 Memcache 版采取集群版架构。 云数据库 Memcache 版内置数据分片及读取算法，整体过程对用户透明，免去用户开发及运维烦恼。每个分片节点采取主备架构保证服务高可用。

云数据库 Memcache 版由 Proxy 服务器（服务代理）、分片服务器和配置服务器三个组件组成。

图 36: Memcache 架构



Proxy 服务器

单节点配置，集群版结构中会有多个 Proxy 组成，系统会自动对其实现负载均衡及故障转移。

分片服务器

每个分片服务器均是双副本高可用架构，主节点故障之后，系统会自动进行主备切换保证服务高可用。

配置服务器

用于存储集群配置信息及分区策略，目前采用双副本高可用架构，保证高可用。

注意:

- 三个组件的个数和配置，在购买集群版相应规格时由系统固定指定，用户暂时不能灵活选择。规格详情如下：

规格	proxy 个数	分片服务器个数	单个分片服务器内存大小
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

- Memcache 集群统一暴露一个访问域名，用户访问该域名进行正常的 Memcache 访问及数据操作，proxy 服务器、分片服务器和配置服务器均不提供域名访问，用户不可以直接连接访问对其进行操作。

14.3 规格说明

云数据库 Memcache 版采用集群版架构，规格定义如下。

规格	CPU 处理能力	节点个数	最大连接数	最大内网带宽
1 GB	单核	1	10000	10
2 GB	单核	1	10000	16
4 GB	单核	1	10000	24
8 GB	单核	1	10000	24
16 GB	双核	2	10000	96
32 GB	4核	4	40000	192
64 GB	8核	8	80000	384
128 GB	16核	16	160000	768
256 GB	16核	16	160000	768
512 GB	32核	32	320000	1536

注意: 目前不支持直接购买512 GB 规格的实例，需要您提交工单去申请开通。

14.4 产品功能

分布式架构，单节点故障业务不受影响

- 云数据库 Memcached 版采用分布式集群架构，每个节点均由双机热备架构组成，具备自动容灾及故障迁移能力。
- 多种规格可适配不同的业务压力，数据库性能支持无限扩展。
- 支持数据持久化及备份恢复策略，有效的保证数据可靠性，可避免物理节点故障缓存失效对后端数据库造成的巨大压力冲击。

多层安全防护体系，为您抵御90%以上的网络攻击

- DDOS 防护 ：在网络入口实时监测，当发现超大流量攻击时，对源 IP 进行清洗，清洗无效情况下可以直接拉进黑洞。
- IP 白名单配置：最多支持配置100个允许连接实例的服务器IP地址，从访问源进行直接的风险控制。
- VPC 虚拟网络：云数据库 Memcache 版全面接入VPC，可基于阿里云构建出一个隔离的网络环境。
- SASL 鉴权：采用 SASL 进行用户身份认证鉴权，保障数据访问安全性。

完善的工具为您分担缓存数据库的运维工作

- 监控报警：提供 CPU 利用率、IOPS、连接数、磁盘空间等实例信息实时监控及报警，随时随地了解实例动态。
- 数据管理：提供可视化数据管理工具，轻松搞定数据操作。
- 源码、分布式维护：专业的数据库内核专家维护，免除 Memcache 源码及分布式算法的维护工作。

14.5 产品优势

简单易用

- 服务开箱即用：支持即开即用的方式，购买之后即可使用，方便业务快速部署。
- 兼容开源 Memcache：兼容 Memcache binary protocol，符合该协议的客户端（binary SASL）都可连接云数据库 Memcache 版。

- 可视化的管理监控面板：控制台提供多项监控统计信息，方便用户对 Memcache 实例进行管理。

集群功能

可支持超大容量，超高性能。默认采用集群功能输出，提供超大集群实例规格，可满足大容量和高性能需求。

弹性扩容

- 存储容量一键扩容：用户可根据业务需求通过控制台对实例存储容量进行调整。
- 在线扩容不中断服务：调整实例存储容量可在线进行，无需停止服务，不影响用户自身业务。

资源隔离

针对实例级别的资源隔离，可以更好地保障单个用户服务的稳定性。

安全可靠

- 支持密码认证方式以确保访问安全可靠。
- 数据持久化存储：内存+硬盘的存储方式，在提供高速数据读写能力的同时满足数据持久化需求。

秒级别监控

- 提供基于引擎和资源的分钟级别历史监控。
- 提供各数据结构和接口的监控信息，访问情况一目了然，便于用户对云数据库 Memcache 版的使用情况有充分的了解。

高可用

- 每个实例均有主从双节点：避免单点故障引起的服务中断。
- 硬件故障自动检测与恢复：自动侦测硬件故障并在数秒内切换，恢复服务。

14.6 应用场景

访问频度极高业务

如社交网络、电子商务、游戏、广告等。可以将访问频度非常高的数据存储在云数据库 Memcache 版中，底层数据存储在 RDS 中。

大型促销类业务

大型促销秒杀系统，系统整体访问压力非常大。一般的数据库根本无法承载这样的读取压力，可选用云数据库 Memcache 版存储。

带有计数器的库存系统

云数据库 RDS 与云数据库 Memcache 版搭配使用。RDS 存储具体数据信息，数据库字段中存储具体计数信息。云数据库 Memcache 版来进行计数的读取，RDS 存储计数信息。

数据分析业务

云数据库 Memcache 版搭配开放数据处理服务 MaxCompute。实现对大数据的分布式分析处理，适用于商业分析、挖掘等大数据处理场景。通过数据集成服务可自助实现数据在云数据库 Memcache 版与 MaxCompute 间的同步，简化数据操作流程。

14.7 使用限制

- 云数据库 Memcache 版仅支持 Key/Value 格式的数据，不支持 array、map、list 等复杂类型，复杂类型的数据不适合使用。
- 云数据库 Memcache 版的数据存储在内存中，服务并不保证缓存数据不会丢失，有强一致性要求的数据不适合存储。
- 云数据库 Memcache 版支持的单条缓存数据的Key最大不超过1 KB，Value 最大不超过1 MB，过大的数据不适合存储。
- 云数据库 Memcache 版不支持事务，有事务性要求的数据不适合写入，而应该直接写入数据库。
- 当数据访问分布比较均匀，数据没有明显的冷热分别时，大量的访问请求在云数据库 Memcache 版无法命中，使用云数据库 Memcache 版作为数据库缓存的效果不明显。在选择缓存时，需要充分考虑到业务模式对数据访问的要求。

14.8 名词解释

表 22: Memcache 相关术语列表

术语	说明
Memcached	Memcached 是一个高性能的分布式内存对象缓存系统。 Memcached 官方介绍可参见这里。云数据库 Memcache 版兼容 Memcached 二进制协议和文本协议两种方式。

术语	说明
实例 ID	实例对应一个用户空间，是使用云数据库 Memcache 版的基本单位。云数据库 Memcache 对单个实例根据不同的容量规格有不同的 QPS 和流量限制。用户可在控制台中看到自己购买的实例 ID 列表。
连接地址	用于连接云数据库 Memcache 的 Host 地址，以域名方式展示，可在**实例信息>基本信息>实例详情>内网地址**中查询到。
连接密码	用于连接云数据库 Memcache 的密码。可在购买时设置，或者在购买后重置密码。
命中率	用户读取成功的成功次数/用户读取次数。
免用户名密码访问	指用户可以在已获授权的 ECS 上无需用户名密码即可访问对应的云数据库 Memcache。更多详情请参见用户指南中的免密码访问。
SASL	SASL 全称 SimpleAuthentication and Security Layer，是一种用来扩充 C/S 模式验证能力的机制。Memcached 从1.4.3版本开始，支持 SASL 认证。由于云数据库 Memcache 版的多租户共享特性，也采用 SASL 作为鉴权机制。SASL 本质上是使用密码保证的缓存数据安全，建议采用强密码和定期修改密码的策略。云数据库 Memcache 将每60秒自动进行一次鉴权。

Memcached

Memcached 是一个高性能的分布式内存对象缓存系统。Memcached 官方介绍可参见这里。云数据库 Memcache 版兼容 Memcached 二进制协议和文本协议两种方式。

实例 ID

实例对应一个用户空间，是使用云数据库 Memcache 版的基本单位。云数据库 Memcache 版对单个实例根据不同的容量规格有不同的 QPS 和流量限制。您可以在控制台中看到自己购买的实例 ID 列表。

连接地址

用于连接云数据库 Memcache 版的 Host 地址，以域名方式展示，可在**实例信息>基本信息>实例详情>内网地址**中查询到。

连接密码

用于连接云数据库 Memcache 版的密码。可在购买时设置，或者在购买后重置密码。

命中率

用户读取成功的成功次数/用户读取次数。

免用户名密码访问

指您可以在已获授权的 ECS 上无需用户名密码即可访问对应的云数据库 Memcache版。更多详情请参见操作指南中的免密码访问。

SASL

SASL 全称 SimpleAuthentication and Security Layer，是一种用来扩充 C/S 模式验证能力的机制。Memcached 从1.4.3版本开始，支持 SASL 认证。由于云数据库 Memcache 版的多租户共享特性，也采用 SASL 作为鉴权机制。SASL 本质上是使用密码保证的缓存数据安全，建议采用强密码和定期修改密码的策略。云数据库 Memcache 版将每60秒自动进行一次鉴权。

14.9 云数据库 Memcache 重磅升级

背景

云数据库 Memcache 版（原版本），采用分布式缓存架构，不提供数据可靠性保障，当服务节点故障，虽然服务可靠性得到保障，但是由于不提供数据持久化策略，数据丢失后，需要用户自行预热 Memcache 系统，造成了极大的不便。

为了更好的服务客户，阿里云云数据库团队将云数据库 Memcache 版本产品重磅升级（2017年5月10日发布），在保障服务可靠性的同时，提供了双机热备、数据持久化、备份恢复等高级功能，为用户提供容灾、恢复、监控、迁移等方面的全套数据库解决方案。

产品形态对比

模块细分	新版 Memcache	旧版 Memcache
分布式架构	支持	支持
数据持久化保障	支持	不支持

模块细分	新版 Memcache	旧版 Memcache
双机热备架构	支持	不支持
Memcache 协议兼容	完全兼容	完全兼容

售卖模式对比

售卖模式更加灵活，支持包月和按量付费的售卖形态。

模块细分	新版 Memcache	旧版 Memcache
包年包月新购	支持	不支持
包年包月升级	支持	不支持
包年包月续费	支持	不支持
包年包月续费变配	支持	不支持
包年包月自动续费	支持	不支持
按量付费新购	支持	支持
按量付费变配	支持	支持
按量付费释放	支持	支持
按量付费转包年包月	支持	不支持

售卖地域支持

售卖地域支持更广泛，国际地域全方位支持，国内地域全部对齐。

表 23: 国际地域列表

地域	可用区	新版 Memcache	旧版 Memcache
亚太（新加坡）	亚太1可用区A	支持	不支持
日本	日本可用区A	支持	不支持
德国（法兰克福）	德国可用区A	支持	不支持
亚太东南2(悉尼)	澳洲可用区A	支持	不支持
香港	香港可用区C	支持	不支持
美东	美东1可用区A	支持	不支持
美国硅谷	美西1可用区B	支持	支持

表 24: 国内地域列表

地域	可用区	新版 Memcache	旧版 Memcache
华东1	华东1可用区B	支持	支持
	华东1可用区D	支持	支持
	华东1可用区E	支持	支持
华东2	华东2可用区A	支持	支持
	华东2可用区B	支持	支持
华南1	华南1可用区A	支持	支持
	华南1可用区B	支持	支持
华北1	华北1可用区B	支持	支持
华北2	华北2可用区A	支持	支持
	华北2可用区B	支持	支持
	华北2可用区C	支持	支持

功能模块对比

功能覆盖度全面提升，更加注重数据库的高级功能。

功能大类	功能点	新版 Memcache	旧版 Memcache
备份恢复	全量备份	支持	不支持
	备份恢复	支持	不支持
	克隆实例	支持	不支持
	数据流操作	预计6月底 DMS 支持图形化	简要命令行
监控报警	资源监控	支持	支持
	资源报警	5月底支持	支持
数据安全	白名单	支持	支持
	VPC 支持	支持	支持
	白名单免密	支持	支持
	多个 Memcache 实例支持免密	支持	不支持

相关 FAQ

- 旧版本的云数据库 Memcache 版如何管理和变配？

  答：旧版的云数据库 Memcache 将继续可以在阿里云控制台进行管理，旧版本已经开通的实例可以正常管理，变配及释放。

- 旧版本的云数据库 Memcache 版如何新购？

  答：旧版本的云数据库 Memcache 将不再支持新购实例，新购实例只能购买新版 Memcache 产品。

  预计2017年8月，阿里云将推出单节点类型的云数据库 Memcache 版本，届时单节点类型的产品定价将于旧版本 Memcache 定价相同。客户可以灵活选择单节点或双节点的产品形态满足不同的业务需求。

- 旧版本的云数据库 Memcache 如何升级至新版 Memcache？

  答：目前还不支持一键将旧版 Memcache 升级至新版的产品化功能。如果需要升级至新版，需要购买一个新版的 Memcache 实例，随后手工预热数据，将应用程序指向新的 Memcache 域名，随后释放老版本的 Memcache 实例。

15 数据传输服务DTS

15.1 产品概述

数据传输（Data Transmission，简称DTS）是阿里云提供的一种支持RDBMS（关系型数据库）、NoSQL、OLAP等多种数据源之间数据交互的数据服务。它提供了数据迁移、实时数据订阅及数据实时同步等多种数据传输能力。通过数据传输可实现不停服数据迁移、数据异地灾备、跨境数据同步、缓存更新策略等多种业务应用场景，助您构建安全、可扩展、高可用的数据架构。

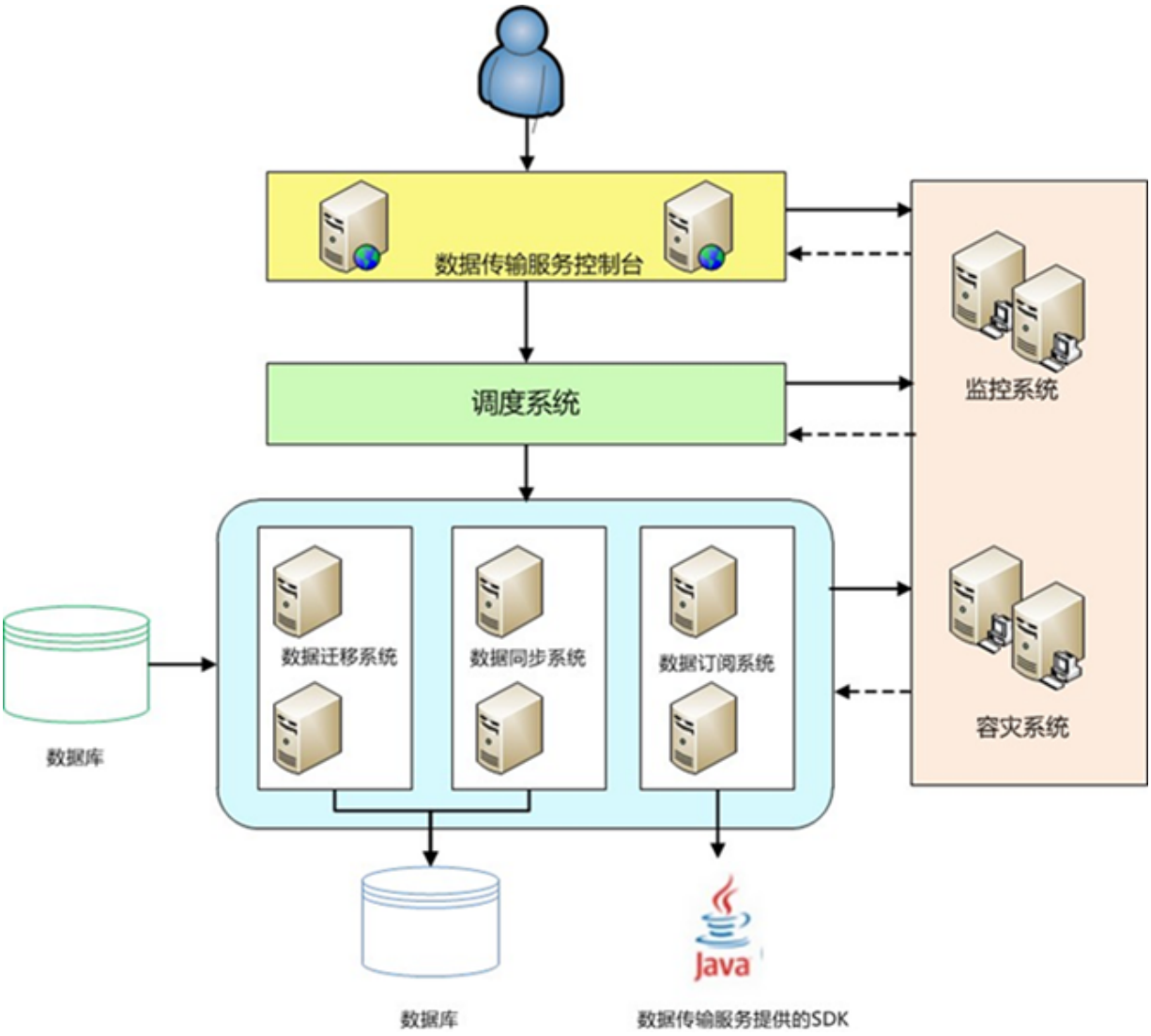
- 数据传输服务DTS的目标是帮您将复杂的数据交互工作承担下来，让您可以专注于上层的业务开发，数据传输服务承诺99.5%的链路稳定性及99.999%的数据可靠性。
- 数据传输服务DTS支持多种数据源类型，例如：
  - 关系型数据库：Oracle、MySQL、SQLServer、PostgreSQL、DRDS、PetaData、OceanBase。
  - NoSQL：MongoDB、Redis。
  - OLAP：MaxCompute、AnalyticDB、流计算。
- 数据传输服务DTS支持RAM主子账号体系，您可以使用子账号创建并管理DTS实例，极大程度提升企业安全性。RAM主子账号相关授权方法请参考《用户指南》中**DTS 支持RAM 主子账号**章节内容。

15.2 产品架构

系统架构

如图 *37:* 系统架构图所示为数据传输服务的系统架构图。

图 37: 系统架构图



- **系统高可用**

  数据传输服务内部每个模块都有主备架构，保证系统高可用。容灾系统实时检测每个节点的健康状况，一旦发现某个节点异常，会将链路秒级切换到其他节点。

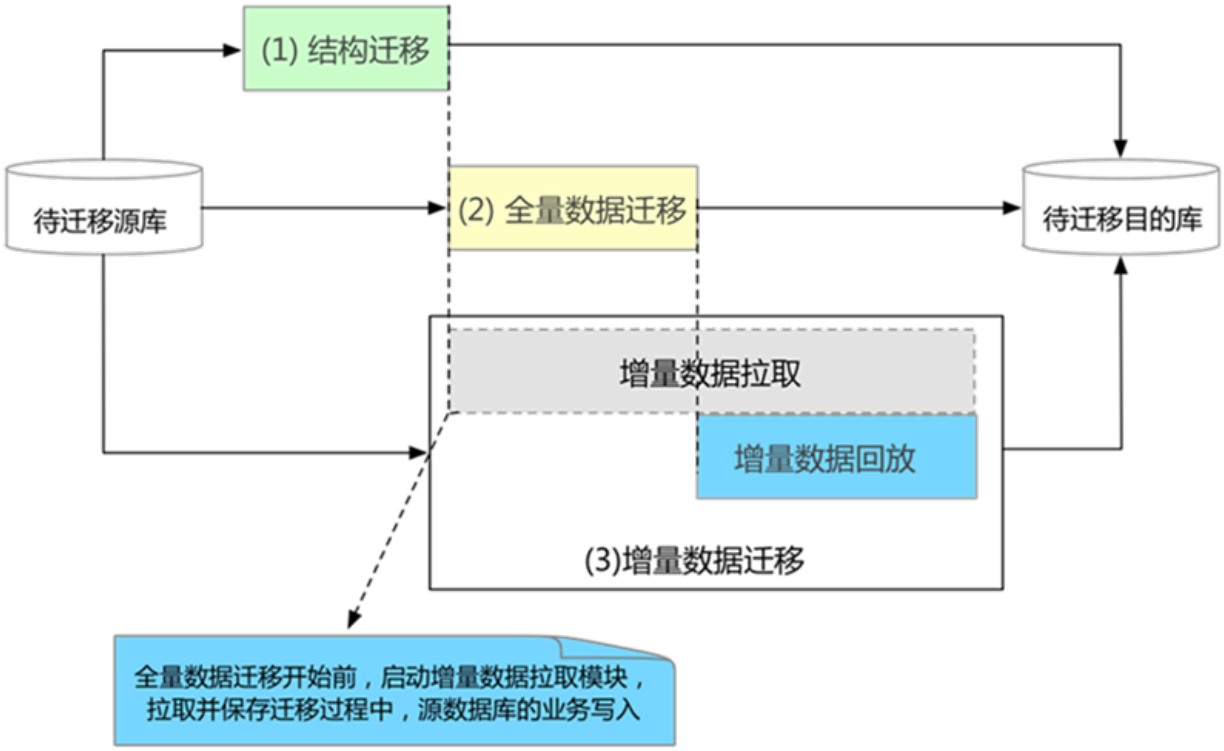
- **数据源地址变更**

  对于数据订阅及同步链路，容灾系统还会监测数据源的连接地址切换等变更操作，一旦发现数据源发生连接地址变更，它会动态适配数据源新的连接方式，在数据源变更的情况下，保证链路的稳定性。

数据迁移基本原理

数据迁移基本原理如图 *38:* 数据迁移基本原理图所示。

图 38: 数据迁移基本原理图



数据迁移任务提供多种迁移类型：结构对象迁移、全量数据迁移以及增量数据迁移。如果需要实现不停服迁移，那么迁移过程需要经历：

1. 结构对象迁移
2. 全量迁移
3. 增量数据迁移

对于异构数据库之间的迁移，进行结构迁移时，DTS会从源库读取结构定义语法后，再根据目标数据库的语法定义，组装成目标数据库的语法定义格式，然后导入到目标实例中。

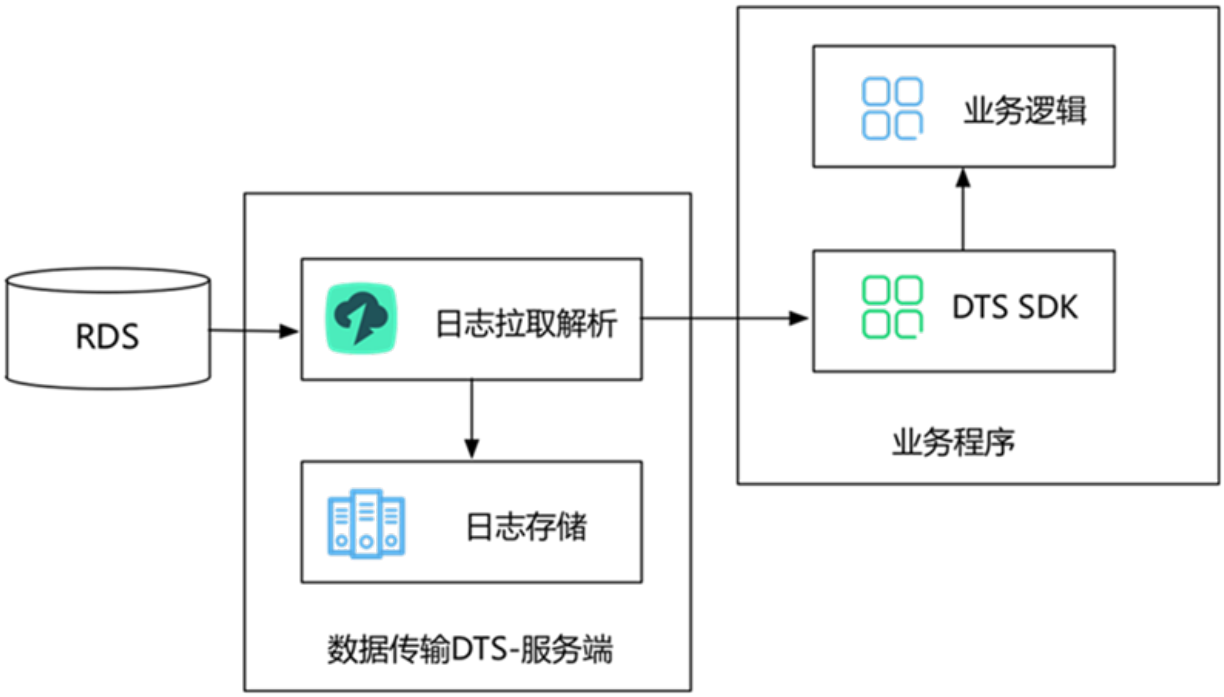
全量数据迁移过程持续较久，在这过程中，源实例不断有业务写入，为保证迁移数据的一致性，在全量数据迁移之前会启动增量数据拉取模块，增量数据拉取模块会拉取源实例的增量更新数据，并解析、封装、存储在本地存储中。

当全量数据迁移完成后，DTS会启动增量数据回放模块，增量数据回放模块会从增量拉取模块中获取增量数据，经过反解析、过滤、封装后同步到目标实例，从而实现源实例、目标实例数据实时同步。

数据订阅基本原理

数据订阅基本原理图如图 *39:* 数据订阅基本原理图所示。

图 39: 数据订阅基本原理图



如上图所示，数据订阅支持实时拉取RDS实例的增量日志，用户可以通过DTS SDK来数据订阅服务端订阅增量日志，根据业务需求，实现数据定制化消费。

DTS服务端的日志拉取模块主要实现从数据源抓取原始数据，并通过解析、过滤、标准格式化等流程，最终将增量数据在本地持久化。

日志抓取模块通过数据库协议连接并实时拉取源实例的增量日志。例如源实例为RDS for MySQL，那么数据抓取模块通过Binlog dump协议连接源实例。

DTS实现了日志拉取模块及下游消费SDK的高可用。

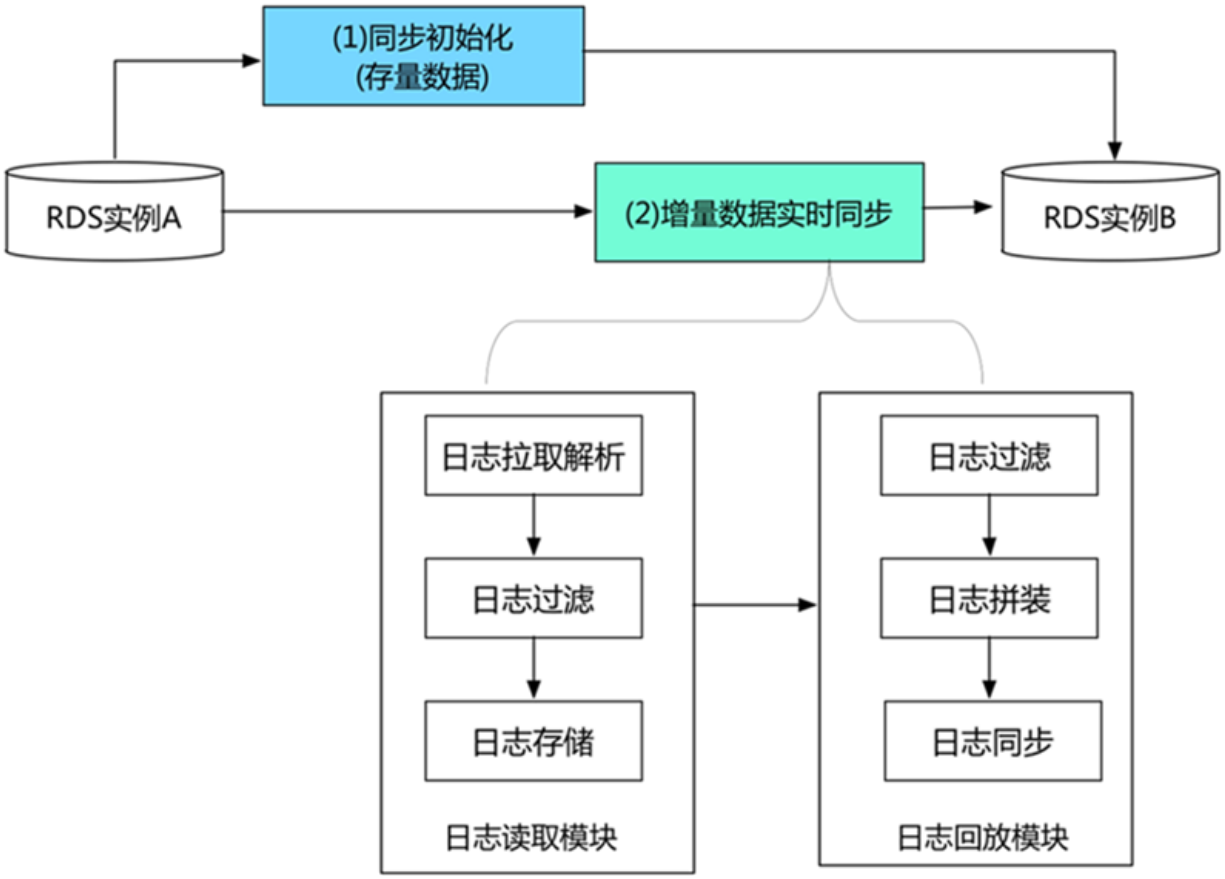
DTS容灾系统一旦检测到日志拉取模块出现异常，就会在健康服务节点上断点重启日志拉取模块，保证日志拉取模块的高可用。

DTS支持在服务端实现下游SDK消费进程的高可用。用户同时对一个数据订阅链路，启动多个下游SDK消费进程，服务端同时只向一个下游消费推送增量数据，当这个消费进程异常后，服务端会从其他健康下游中选择一个消费进程，向这个消费进程推送数据，从而实现下游消费的高可用。

实时同步基础原理

实时同步基础原理如图 *40:* 实时同步基础原理所示。

图 40: 实时同步基础原理



数据传输服务的实时同步功能能够实现任何两个RDS实例之间的增量数据实时同步。

同步链路的创建过程包括：

- 同步初始化：主要将源实例的历史存量数据在目标实例初始化一份。
- 增量数据实时同步：当初始化完成后进入两边增量数据实时同步阶段，在这个阶段DTS会实现源实例跟目标实例之间数据动态同步过程。

增量数据实时同步过程，DTS的底层实现模块主要包括：

- 日志读取模块

  日志读取模块从源实例读取原始数据，经过解析、过滤及标准格式化，最终将数据在本地持久化。日志读取模块通过数据库协议连接并读取源实例的增量日志。如果源数据库为RDS for MySQL，那么数据抓取模块通过Binlog dump协议连接源库。

- 日志回放模块

  日志回放模块从日志读取模块中请求增量数据，并根据用户配置的同步对象进行数据过滤，然后在保证事务时序性及事务一致性的前提下，将日志记录同步到目标实例。

DTS实现了日志读取模块、日志回放模块的高可用，DTS容灾系统一旦检测到链路异常，就会在健康服务节点上断点重启链路，从而有效保证同步链路的高可用。

15.3 功能特性

15.3.1 数据迁移

数据迁移功能旨在帮助用户方便、快速地实现各种数据源之间的数据迁移。实现数据迁移上云、阿里云内部跨实例数据迁移、数据库拆分扩容等业务场景。数据传输服务提供的数据迁移功能能够支持同异构数据源之间的数据迁移，同时提供了库表列三级映射、数据过滤多种ETL特性。

多种数据源类型

数据迁移支持多种数据源之间的数据迁移，不同数据源的支持情况如表 *25:* 不同数据源的支持情况所示。

表 25: 不同数据源的支持情况

数据源	结构迁移	全量迁移	增量迁移
Oracle->RDS for MySQL	支持	支持	不支持
Oracle->RDS for PPAS	支持	支持	支持
MySQL->RDS for MySQL	支持	支持	支持
SQLServer-> RDS for SQL Server	支持	支持	支持
PostgreSQL->RDS for PostgreSQL	支持	支持	支持
MongoDB->MongoDB	支持	支持	支持
Redis->Redis	支持	支持	支持
MySQL->DRDS	不支持	支持	支持
MySQL-> PetaData	不支持	支持	不支持
MySQL-> OceanBase	支持	支持	支持

数据迁移支持的源实例类型包括：

- RDS实例
- 本地自建数据库
- ECS自建数据库
- MongoDB实例

- Redis实例

数据迁移支持的目标实例包括：

- RDS实例
- ECS自建数据库
- MongoDB实例
- Redis实例
- DRDS实例
- PetaData实例
- OceanBase实例

多种迁移方式

数据传输服务提供的数据迁移方式包括：

- 在线迁移

  默认使用在线迁移，在线迁移只要用户配置迁移的源、目标实例及迁移对象即可，DTS会自动完成整个数据迁移过程。在线迁移支持数据不停服迁移，然而在线迁移要求DTS服务器能够同时跟源实例、目标实例连通。

- 离线文件迁移

  离线文件迁移要求用户先使用DTS客户端将源实例的数据导出成文件后，再将文件导入到目标实例。 离线迁移不支持数据不停服迁移，它主要用于解决DTS服务器跟源数据库网络不通的情况。

如果DTS服务器同源实例网络连通的情况下，建议使用在线迁移，降低数据迁移成本。

多种迁移步骤

数据迁移支持结构迁移、全量数据迁移及增量数据迁移多种迁移步骤。其中：

- 结构迁移，帮助用户将源实例中的结构对象定义一键迁移至目标实例。
- 全量数据迁移，帮助用户将源实例中的历史存量数据迁移至目标实例。
- 增量数据迁移，帮助用户将迁移过程中，源实例提供服务产生的增量数据实时同步到目标实例。结构迁移 + 全 量数据迁移 + 增量数据迁移可以简单实现业务不停服迁移。

多种ETL特性

数据迁移支持多种ETL特性，主要包括：

- 支持库表列三级对象名映射。库表列三级对象名映射是指可以实现源实例和目标实例的库名或表名，甚至列名不同的两个对象之间的数据迁移。
- 支持迁移数据过滤，用户可以对要迁移的表配置某种SQL条件过滤要迁移的数据。例如用户可以配置时间条件，只迁移最新的数据。

报警机制

数据迁移提供迁移异常报警，一旦迁移任务出现异常，即会向任务的owner发送报警短信，让用户第一时间了解并处理异常任务。

迁移任务

迁移任务是数据传输服务进行数据迁移的基本单元。如果需要进行数据迁移，必须在数据传输服务控制台创建一个迁移任务。当创建迁移任务时，需要配置待迁移源实例的连接方式、目标实例的连接方式、迁移对象及迁移类型等信息。用户可以在数据传输服务控制台进行迁移任务的创建、管理、停止及删除等操作。

迁移任务在创建及运行过程中，不同阶段会处于不同的状态，具体如表 *26:* 迁移状态及说明所示。

表 26: 迁移状态及说明

迁移状态	状态说明	可进行操作
未启动	迁移任务已经完成任务配置，但是还没有进行迁移前的预检查的任务。	• 预检查 • 删除
预检中	迁移任务正在进行前期的预检查阶段。	删除
预检通过	迁移任务已经通过迁移之前的预检查，但是还没有启动迁移。	• 启动 • 删除
迁移中	迁移任务正在进行正常的数据迁移。	• 暂停 • 结束 • 删除
迁移失败	迁移任务异常，可以根据任务的进度确认具体是哪个阶段失败。	删除
暂停中	这个迁移任务已经被暂停迁移。	• 启动 • 删除
完成	迁移任务已经完成数据迁移，或者用户单击**结束**停止数据迁移。	删除

15.3.2 数据同步

数据实时同步功能旨在帮助用户实现两个数据源之间的数据实时同步。通过数据实时同步功能可实现数据异地灾备、本地数据灾备、跨境数据同步及在线离线数据打通（OLTP到OLAP）数据同步等多种业务场景。

功能列表

- 支持任何两个RDS For MySQL实例间的数据实时同步。
- 支持RDS for MySQL实例和分析型数据库AnalyticDB实例间的数据实时同步。
- 支持RDS for MySQL实例和MaxCompute实例间的数据实时同步。
- 支持RDS for MySQL实例和DataHub实例间的数据实时同步。

同步对象

- 数据同步的同步对象的选择粒度可以为：库、表、列。用户可以根据需要同步某几个表的数据。
- 数据同步支持库、表、列名映射，即用户可以进行两个不同库名的数据库之间的同步，或两个不同表名的表之间的数据同步。
- 数据同步支持列选择，即用户可以根据业务需求，只同步表中的某几列数据。

同步作业

同步作业是数据实时同步的基本单元。如果要进行两个实例间的数据同步，必须在数据传输控制台创建同步作业。

同步作业在创建及运行过程中，不同阶段会处于不同的状态，具体如表 *27:* 作业状态及说明所示。

表 27: 作业状态及说明

作业状态	状态说明	可进行操作
预检中	同步作业正在进行启动前的预检查。	• 查看同步配置 • 删除同步 • 复制同步配置 • 配置监控报警
预检查失败	同步作业预检查没有通过。	• 预检查 • 查看同步配置 • 修改同步对象 • 修改同步速度 • 删除同步

作业状态	状态说明	可进行操作
		• 复制同步配置 • 配置监控报警
未启动	迁移任务已经通过迁移之前的预检查，但是还没有启动。	• 预检查 • 开始同步 • 修改同步对象 • 修改同步速度 • 删除同步 • 复制同步配置 • 配置监控报警
同步初始化中	同步作业正在进行同步初始化。	• 查看同步配置 • 删除同步 • 复制同步配置 • 配置监控报警
同步初始化失败	同步作业在初始化过程中，迁移失败。	• 查看同步配置 • 修改同步对象 • 修改同步速度 • 删除同步 • 复制同步配置 • 配置监控报警
同步中	同步作业正常同步中。	• 查看同步配置 • 修改同步对象 • 修改同步速度 • 暂停同步 • 删除同步 • 复制同步配置 • 配置监控报警
同步失败	同步作业同步异常。	• 查看同步配置 • 修改同步对象 • 修改同步速度 • 启动同步 • 删除同步 • 复制同步配置

作业状态	状态说明	可进行操作
		• 配置监控报警
暂停中	同步作业执行了暂停，处于暂停状态。	• 查看同步配置 • 修改同步对象 • 修改同步速度 • 启动同步 • 删除同步 • 复制同步配置 • 配置监控报警

高级特性

数据订阅支持多种特性，有效降低用户使用门槛，主要包括：

- 动态增减同步对象

  在数据同步过程中，用户可以随时增加或减少需要同步的对象。
- 完善性能查询体系

  数据同步提供同步延迟、同步性能（RPS、流量）趋势图，用户可以方便查看同步链路的性能趋势。
- 完善监控体系

  数据同步提供同步作业状态、同步延迟的报警监控功能。用户可以根据业务敏感度，自定义同步延迟报警阈值。

15.3.3 数据订阅

实时数据订阅功能旨在帮助用户获取RDS/DRDS的实时增量数据，用户能够根据自身业务需求自由消费增量数据，例如实现缓存更新策略、业务异步解耦、异构数据源数据实时同步及含复杂ETL的数据实时同步等多种业务场景。

功能列表

- 支持经典网络、VPC网络下RDS For MySQL实例的数据订阅。

数据源类型

实时数据订阅支持的数据源类型包括：

- RDS For MySQL

- DRDS

其中，DRDS不记录事务日志，如果需要订阅DRDS的实时增量数据，则需要通过订阅DRDS底层挂载的RDS实例的增量日志来实现。

订阅对象

数据订阅的订阅对象可以为：库和表。用户可以根据需要订阅某几个表的增量数据。

数据订阅将增量数据细分为数据变更（Data Manipulation Language 简称DML）和结构变更（Data Definition Language，简称DDL），配置数据订阅时，可以选择需要订阅的具体数据变更类型。

订阅通道

订阅通道是进行增量数据订阅与消费的基本单元。如果要订阅RDS的增量数据，必须在数据传输控制台创建一个针对这个RDS实例的订阅通道。订阅通道会实时拉取RDS的增量数据，并将最新一段时间的增量数据保存在订阅通道中，用户可以使用数据传输提供SDK从这个订阅通道中订阅增量数据，并进行相应的消费。同时，用户可以在数据传输控制台进行订阅通道的创建、管理及删除等操作。

一个订阅通道同时只能被一个下游SDK订阅消费，如果用户有多个下游需要订阅同一个RDS实例时，需要创建多个订阅通道。这些订阅通道订阅的RDS实例均为同一个实例ID。

订阅通道在创建及运行过程中，不同阶段会处于不同的状态，具体如表 *28:* 通道状态及说明所示。

表 28: 通道状态及说明

通道状态	状态说明	可进行操作
预检中	订阅通道已经完成任务配置，正在进行启动之前的简单预检查。	删除订阅
未启动	迁移任务已经通过迁移之前的预检查，但是还没有启动订阅。	• 开始订阅 • 删除订阅
初始化	订阅通道正在进行启动初始化，一般需要1分钟左右。	删除订阅
正常	订阅通道正在正常拉取RDS实例的增量数据。	• 查看示例代码 • 查看订阅数据 • 删除订阅
异常	订阅通道拉取RDS实例增量数据异常。	• 查看示例代码 • 删除订阅

高级特性

数据订阅支持多种特性，有效降低用户使用门槛，主要包括：

- 动态增减订阅对象

  在数据订阅过程中，用户可以随时增加或减少需要订阅的对象。

- 在线查看订阅数据

  数据传输DTS控制台支持在线查看订阅通道中的增量数据。

- 修改消费时间点

  数据订阅支持用户随时修改需要消费数据对应的时间点。

- 完善监控体系

  数据订阅提供订阅通道状态、下游消费延迟的报警监控功能。用户可以根据业务敏感度，自定义消费延迟报警阈值。

15.4 产品优势

数据传输（Data Transmission）服务支持RDBMS、NoSQL、OLAP等数据源间的数据传输。它提供了数据迁移、实时数据订阅及数据实时同步等多种数据传输方式。相对于第三方迁移同步工具，数据传输服务提供更丰富多样、高性能、高安全可靠的传输链路，同时它提供了诸多便利功能，极大地方便了传输链路的创建及管理。

丰富多样

数据传输服务能够支持多种同异构数据源之间的迁移同步，例如Oracle->MySQL、Oracle->Postgres Plus Advanced Server。对于异构数据源之间的迁移，数据传输服务支持结构对象定义的转化，例如将Oralce中的同义词转换为Postgres Plus Advanced Server中对应的同义词定义。

数据传输服务支持多种传输方式，数据迁移、实时数据订阅及数据实时同步。其中实时数据订阅及数据实时同步均为实时数据传输方式。

为了降低数据迁移对应用的影响，数据迁移功能支持不停服迁移方式。不停服迁移，可实现在数据迁移过程中，应用停机时间降低到分钟级别。

高性能

数据传输服务使用高规格服务器来保证每条迁移同步链路都能拥有良好的传输性能。

对于数据迁移，数据传输服务底层使用了多种性能优化措施，全量数据迁移高峰期时性能可以达到70MB/s，20W TPS。

相对于传统的数据同步工具，数据传输服务的实时同步功能能够将并发粒度缩小到事务级别，能够并发同步同张表的更新数据，从而极大得提升同步性能，高峰期时，同步性能可以达到30000W/s。

安全可靠

数据传输服务底层为服务集群，集群内任何一个节点宕机或发生故障，控制中心都能够将这个节点上的所有任务秒级切换到其他节点上，链路稳定性高达99.95%。

数据传输服务内部对部分传输链路提供7×24小时的数据准确性校验，快速发现并纠正传输数据，保证传输数据可靠性。

数据传输服务各模块间采用安全传输协议及安全token认证，有效得保证数据传输可靠性。

简单易用

数据传输服务提供可视化管理界面，提供向导式的链路创建流程，用户可以在其控制台简单轻松地创建自己的传输链路。

数据传输服务控制台展示了链路的传输状态及进度，传输性能等信息，用户可以方便管理自己的传输链路。

为了解决网络或系统异常等导致的链路中断问题，数据传输服务提供链路断点续传功能，且定期监测所有链路的状态，一旦发现链路异常，先尝试自动修复重启，如果链路需要用户介入修复，那么用户可以直接在控制台修复后触发链路重启。

15.5 典型应用

数据传输服务支持数据迁移、数据实时订阅及数据实时同步等多种功能，通过数据传输可以帮助您满足下面多种典型应用场景。

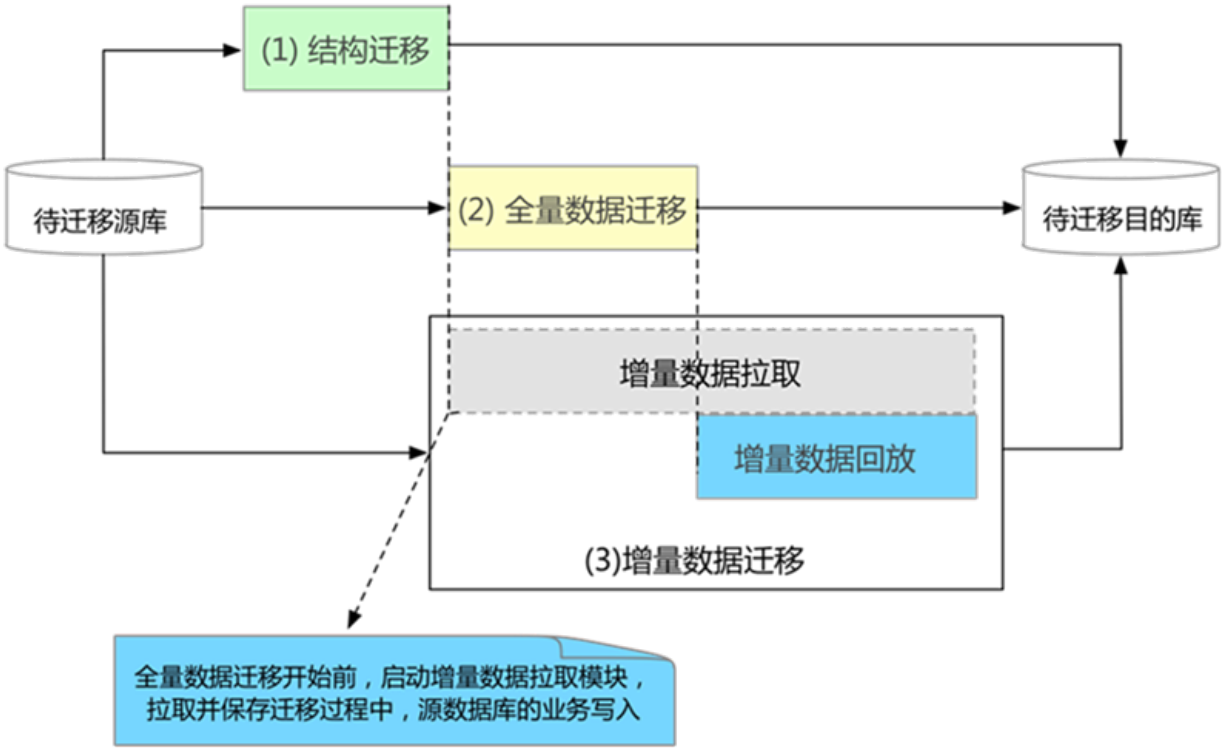
不停服迁移，系统迁移时业务停服时间降低到分钟级别

很多用户希望系统迁移时，尽可能不影响业务提供服务。然而在系统迁移过程中，如果业务不停服，那么迁移数据就会发生变化，无法保证迁移数据的一致性。为了保证迁移数据一致性，很多第三方迁移工具，要求在数据迁移期间，应用停止服务。整个迁移过程，业务可能需要停服数小时甚至上天，这对业务伤害极大。

为了降低数据库迁移门槛，数据传输提供不停服迁移解决方案，让数据迁移过程中，业务停服时间降低到分钟级别。

不停服迁移的实现原理如图 *41:* 不停服迁移所示。

图 41: 不停服迁移



不停服迁移的迁移类型需包含结构迁移、全量数据迁移及增量数据迁移三个阶段。当进入增量数据迁移阶段时，目标实例会保持跟源数据库之间的数据实时同步，用户可以在目标数据库进行业务验证，当验证通过后，直接将业务切换到目标数据库，从而实现整个系统迁移。

由此可见，在整个迁移过程中，只有当业务从源实例切换到目标实例期间，会产生业务闪断，其他时间业务均能正常服务。

数据异地灾备，地区故障对服务无影响

由于地区断电、断网等客观原因，产品可用性并不能达到100%。当出现这些故障时，如果用户业务部署在单个地区，那么就会因为地区故障导致服务不可用，且不可用时间完全依赖故障恢复时间。

为了解决地区故障导致的服务不可用，提高服务可用性，可以在构建异地灾备中心。当业务中心发生地区故障时，直接将业务流量切换到灾备中心，立刻恢复服务。数据灾备架构如图 42: 异地灾备所示。

图 42: 异地灾备

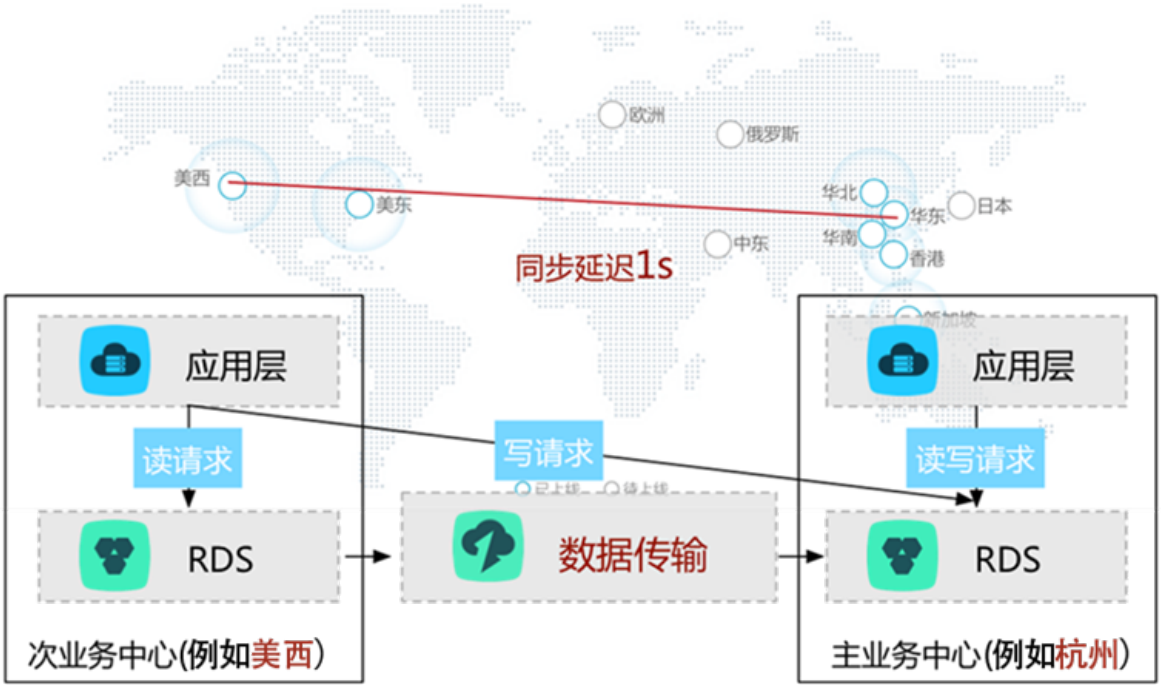
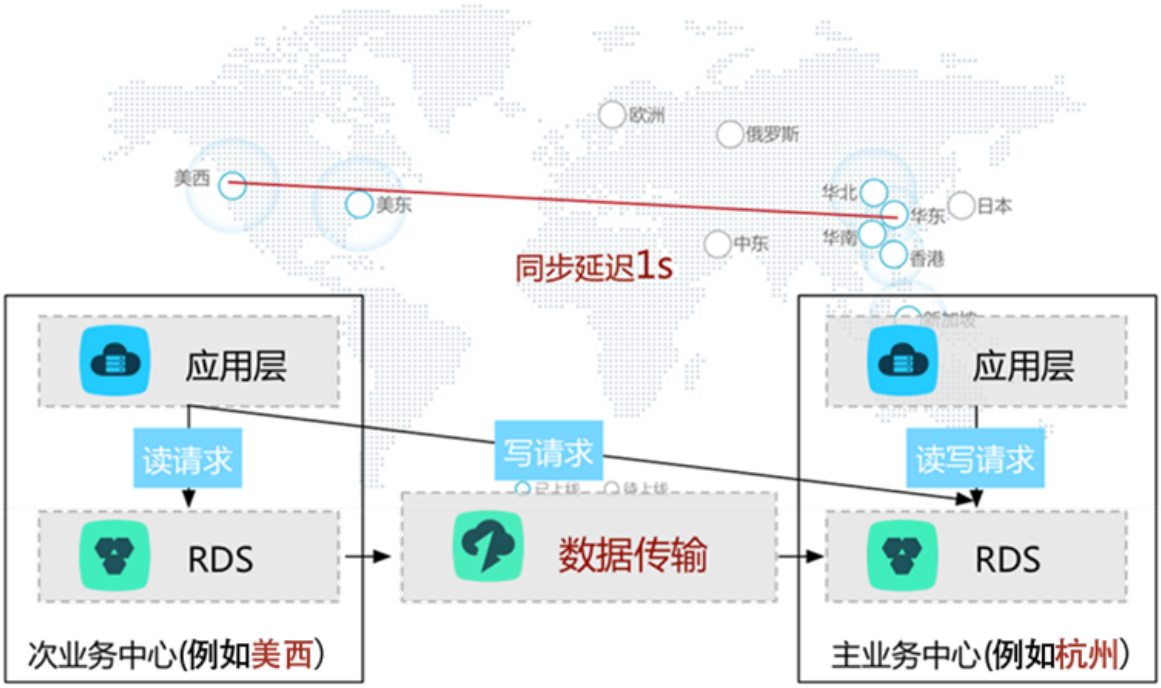


图 *42:* 异地灾备中，当业务部署在杭州时，在异地(例如北京)构建灾备中心。灾备中心同业务中心的数据库通过数据传输进行数据实时同步，当业务中心故障时，可以保证数据灾备的数据完整性。

加速全球化业务访问速度，为跨境业务赋能

对于用户分布比较广的业务，例如全球化业务，如果按照传统架构，只在单地区部署服务，那么其他地区的用户需要跨地区远距离访问服务，导致访问延迟大、用户体验差的问题。为了加速全球化业务访问速度，优化访问体验，可以将架构调整为如图 *43:* 降低跨地区访问延迟所示。

图 43: 降低跨地区访问延迟



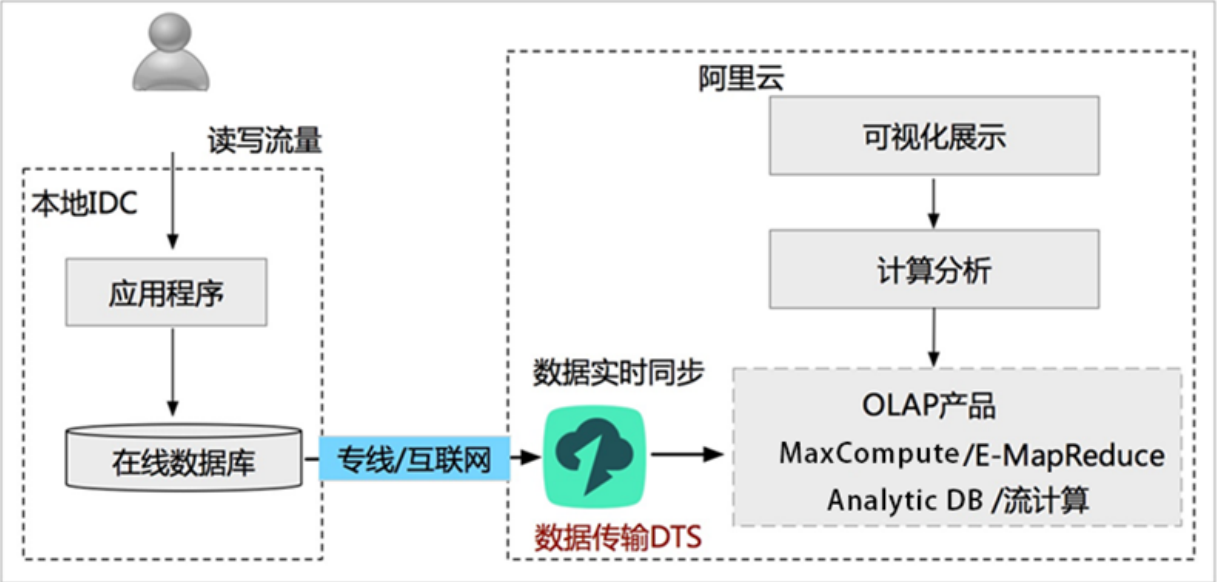
在这个架构中，我们定义了中心和单元的概念，所有地区用户的写请求全部路由回中心。通过数据传输将中心的数据实时同步到各个单元，各个地区的用户的读请求，可以路由到就近的单元，从而避免远距离访问，降低访问延迟，加速全球化访问速度。

云BI，快速搭建定制化BI系统

自建BI系统无法满足越来越高的实时性要求，且操作复杂。而阿里云提供的BI体系，可以在不影响现有架构的情况下快速搭建BI系统，所以越来越多的用户选择在阿里云上搭建满足自身业务定制化要求的BI系统。

数据传输DTS可以帮助用户将本地自建数据库的数据实时同步到阿里云的BI存储系统（例如MaxCompute、Analytic数据库或流计算），然后，用户使用各种计算引擎进行后续的数据分析，同时可以通过可视化工具进行计算结果的实时展示，或通过迁移工具将计算结果同步回本地IDC，具体实现架构如图 *44: 云BI*所示。

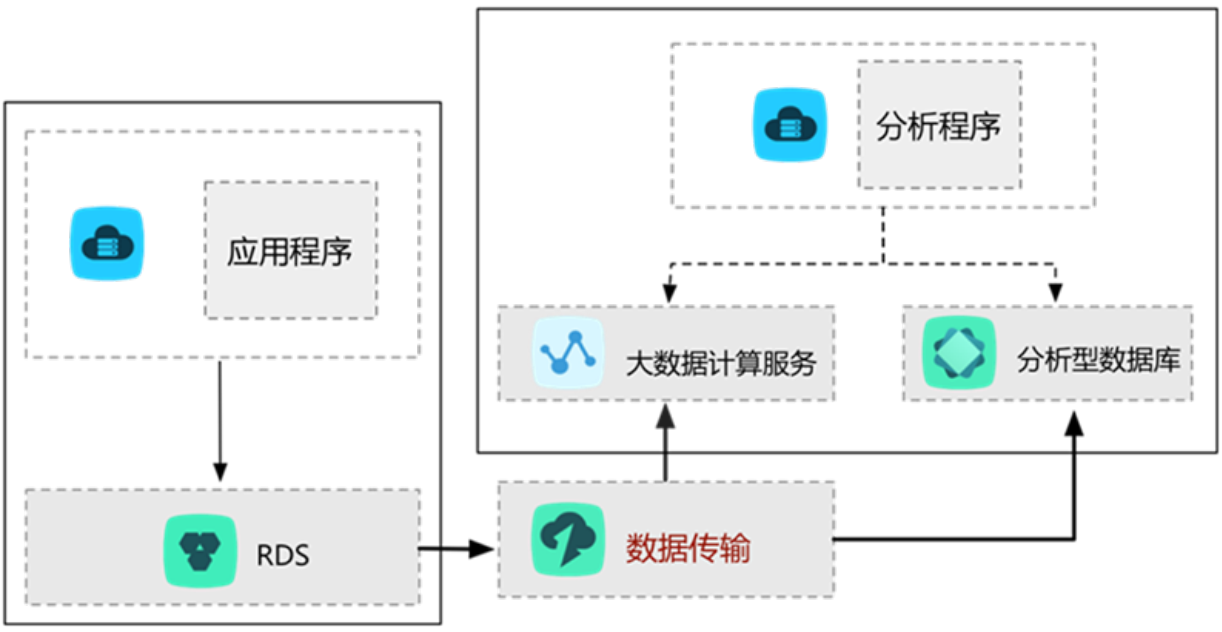
图 44: 云BI



数据实时分析，快速抢占商业先机

数据分析在提高企业洞察力和用户体验方面发挥着举足轻重的作用，且实时数据分析能够让企业更快速、灵活地调整市场策略，适应快速变化的市场方向及消费者体验。为了在不影响线上业务的情况下实现实时数据分析，需要将业务数据实时同步到分析系统中，由此可见，实时获取业务数据必不可少。数据传输提供的数据订阅功能，可以在不影响线上业务的情况下，帮助您获取业务的实时增量数据，通过SDK可将其同步至分析系统中进行实时数据分析，如图 *45:* 数据实时分析所示。

图 45: 数据实时分析

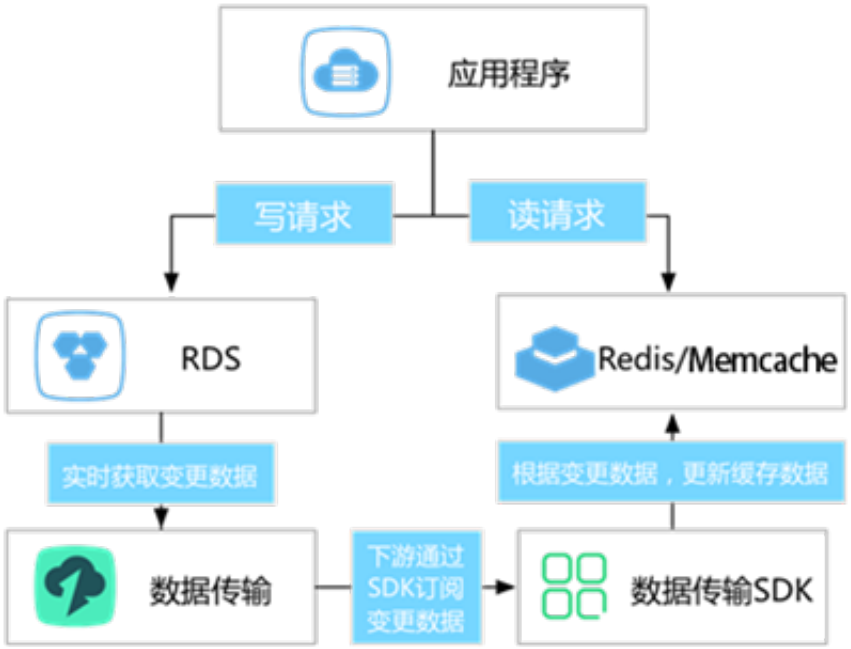


轻量级缓存更新策略，让核心业务更简单可靠

为了提高业务访问速度，提升业务读并发，很多用户都会在业务架构中引入缓存层。业务所有读请求全部路由到缓存层，通过缓存的内存读取机制大大提升业务读取性能。缓存中的数据不能持久化，一旦缓存异常退出，那么内存中的数据就会丢失，所以为了保证数据完整，业务的更新数据会落地到持久化存储中，例如数据库。

如上所述，业务会遇到缓存跟持久化数据库数据同步的问题。数据传输DTS提供的数据订阅功能，可以通过异步订阅数据库的增量数据，并更新缓存的数据，实现轻量级的缓存更新策略。这种策略的架构如图 *46:* 缓存更新策略所示。

图 46: 缓存更新策略



这种缓存更新策略的优势在于：

- 更新路径短，延迟低

  缓存失效为异步流程，业务更新数据库完成后直接返回，不需要关心缓存失效流程，整个更新路径短，更新延迟低。

- 应用简单可靠

  应用无需实现复杂双写逻辑，只需启动异步线程监听增量数据，更新缓存数据即可。

- 应用更新无额外性能消耗

因为数据订阅是通过解析数据库的增量日志来获取增量数据，获取数据的过程对业务、数据库性能无损。

业务异步解耦，让核心业务更简单可靠

通过数据订阅，可以将深耦合业务优化为通过实时消息通知实现的异步耦合，让核心业务逻辑更简单可靠。这个应用场景在阿里巴巴内部得到了广泛的应用，目前淘宝订单系统每天有上万个下游业务，通过数据订阅获取订单系统的实时数据更新，触发自身的变更逻辑。

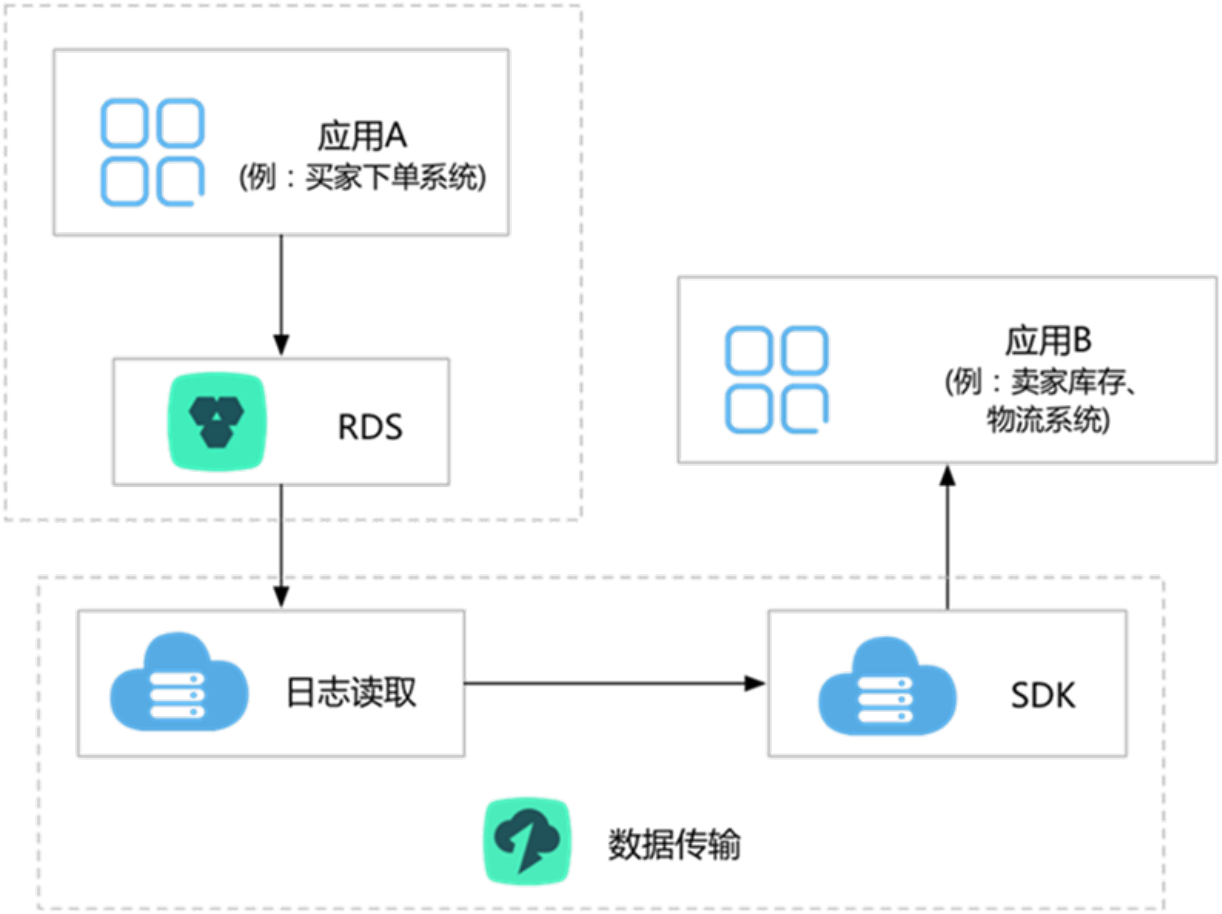
下面举个简单的逻辑，描述下整个应用场景的优势。

例如电商行业，涉及下单系统、卖家库存、物流发货等多个业务逻辑。如果将这些逻辑全部在下单流程中，那么下单流程为：用户下单，系统通知卖家库存，物流发货等下游业务进行逻辑变更，当全部变更完成后，返回下单结果。这种下单逻辑存在如下问题：

- 下单流程长、时间、用户体验差。
- 系统稳定性差，任何一个下游发生故障，直接影响下单系统的可用性。

为了提升核心应用用户体验，提高稳定性，可以将核心应用和依赖的下游业务异步解耦。让核心应用更稳定可靠。具体调整如图 *47:* 业务异步解耦所示。

图 47: 业务异步解耦



下单系统只要下完单就直接返回，底层通过数据传输实时获取订单系统的变更数据，业务通过SDK订阅这个变更数据，并触发库存、物流等下游业务逻辑。由此，保证下单系统的简单可靠。

读能力横向扩展，快速适应业务发展

对于有大量读请求的应用场景，单个RDS实例可能无法承担全部的读取压力， 甚至可能对主流程业务产生影响。为了实现读取能力的弹性扩展，分担数据库压力，可以使用数据传输服务的实时同步功能构建只读实例，利用这些只读实例承担大量的数据库读取工作负载，从而方便地扩展了应用的吞吐量。

15.6 基本概念

预检查

预检查是迁移任务启动之前的必经阶段，主要是对影响迁移成功的前置条件进行检查。例如源目标实例的连通性，迁移账号的权限等的检查。如果预检查失败了，那么可以根据修复方法修复后，重新进行预检查。

结构迁移

结构迁移是迁移任务中的其中一种迁移类型。在数据库迁移中，它是指进行结构对象定义语法的迁移，包括表、视图、触发器、存储过程、存储函数、同义词等结构对象的语法迁移。对于异构数据库之间的迁移，在结构迁移阶段进行数据类型的映射，并根据源实例和目标实例的语法定义，对对象定义语法进行调整。例如Oracle到MySQL的迁移时，会将Oracle中的number映射为MySQL中的decimal类型。

全量数据迁移

全量数据迁移是迁移任务的一种迁移类型。它是指将源实例数据库中的所有数据，不包括结构语法定义，迁移到目标实例。如果创建迁移任务时，只选择全量数据迁移，而不选增量数据迁移，那么在迁移过程中，如果源实例有数据写入，那么对于迁移过程中源实例的新增数据，不会迁移到目标实例。

增量数据迁移

增量数据迁移是迁移任务的一种迁移类型。它是指将迁移过程中，将源实例写入的增量数据同步到目标实例。如果创建迁移任务时，选择了全量数据迁移及增量数据迁移，那么数据传输服务会先在源实例实现静态快照，先将快照数据迁移到目标实例之后，再将迁移过程中源实例写入的增量数据同步到目标实例中。增量数据迁移是一个保持目标实例跟源实例数据实时同步的过程，不会自动结束，如果需要结束迁移，那么需要在控制台手动结束任务。

同步初始化

同步初始化是指在同步链路增量数据同步前，将同步对象的历史数据初始化到目标实例。

同步初始化分为结构初始化和全量数据初始化。结构初始化是进行同步对象的结构定义的初始化。全量数据初始化是进行同步对象数据的初始化。

同步性能

同步性能是指每秒同步到目标实例的记录数，单位为RPS。同步性能为数据同步售卖的规格定义。不同的规格，每秒同步的记录数不同。

同步延迟

同步延迟是指同步到目标实例的最新数据在源数据库执行的时间戳，和源实例当前时间戳的差值。同步延迟反映了目标实例和源实例的数据时间差。当同步延迟为0时，表示源实例和目标实例数据完全一致。

订阅通道ID

订阅通道ID是订阅通道的唯一标识，购买完订阅通道，数据传输会自动生成订阅通道ID。用户使用SDK消费增量数据时，需要配置对应的订阅通道ID。在数据传输控制台的订阅列表中，显示每个订阅通道对应的订阅通道ID。

数据更新

数据传输服务将数据库中的更新数据类型分为：数据更新和结构更新。数据更新是指只修改数据，但是不修改结构对象定义。例如insert、update、delete等。

结构更新

数据传输服务将数据库中的更新数据类型分为：数据更新和结构更新。结构更新是指修改了结构对象定义的语法。例如create table、alter table、drop view 等。用户可以在创建订阅通道时，选择是否订阅结构更新。

数据范围

数据范围是指订阅通道中存储的增量数据时间戳的范围，增量数据对应的时间戳是这条增量数据在RDS实例中应用完并写入事务日志的时间戳。默认情况下订阅通道中只保留最新一天的增量数据。数据传输服务会定期清理过期的增量数据，同时更新订阅通道的数据范围。

消费时间点

消费时间点是指下游SDK订阅数据且已经被消费掉的最新一条增量数据对应的时间戳。SDK每消费一条数据都向数据传输服务服务端汇报ACK，服务端会更新并保存这个SDK对应的消费时间点，当SDK异常重启时，服务端会自动从最后的消费位点推送订阅数据。

16 数据管理

16.1 产品概述

数据管理（Data Management，简称DMS）支持MySQL、SQL Server、PostgreSQL、MongoDB、Redis等关系型数据库和NoSQL的数据库管理。

它是一种集数据管理、结构管理、访问安全、BI图表、数据趋势、数据轨迹、性能与优化和服务器管理于一体的数据管理服务。

16.2 产品架构

阿里云数据管理提供的数据库管理服务包含三层结构：业务层、调度层、连接层，用于对RDBMS、NoSQL的实时数据访问和后台数据任务的调度。

业务层

- DMS业务层为您提供数据库实时点击和SQL访问，业务层为无状态节点，可线性扩展，确保DMS整体服务能力的提升。
- 宕机无状态切换，确保7×24小时服务。

调度层

- DMS调度层为您提供的调度主要包含：导出、导入、表结构对比、数据趋势。其后台主要通过线程池进行调度，分为实时调度和后台定时调度两类。
- 实时调度提供前端点击后立即调度并一次性任务处理，用户提交任务后无需等待结果，数据管理后台自动完成所有工作，结束后用户下载或查看结果。
- 后台定时调度任务为用于定时获取用户指定的数据（如数据趋势），数据管理后台将统一定义调度各项任务对各项业务数据进行采集，提供查阅和分析。

连接层

连接层为数据管理的访问数据的核心部件，主要包含如下几点：

- 兼容MySQL、SQLServer、PostgreSQL、PPAS、Redis、MongoDB的请求。
- 前端操作上会话隔离及保持，即通过数据管理打开多个SQL窗口，各SQL窗口间的会话相互隔离，并尽可能保持各SQL窗口内的会话状态，接近客户端的体验。
- 实例会话数量控制，防止对单个实例建立大量的连接数。
- 按功能分级回收连接策略，尽可能确保不同功能体验的基础上减少对数据库连接数。

16.3 功能特性

本章节主要介绍关系型数据库和NoSQL的功能特性。

关系型数据库

- 数据管理：提供SQL窗口、SQL命令行、表数据管理、SQL智能提示、SQL格式化、自定义SQL、SQL模板、SQL执行计划、导入导出等管理功能。
- 结构管理：提供库、表、视图、函数、存储过程、触发器、事件、序列、同义词等对象管理以及表结构对比功能。
- 性能与优化：提供实时性能展示、实时SQL索引建议、图形化锁管理、实例会话管理、诊断报告等功能。
- BI图表：提供基于SQL结果集直接绘制图表，支持折线图、饼图、柱状图等常规图表以及地图、动态图、环比等高级图表。
- 数据趋势：提供直观的业务表读取/插入/删除/更新行数的实时监控和历史趋势。
- 访问安全：提供4层认证体系，提供登录/操作审计，提供云账号授权、访问地址授权、功能开关等细粒度授权功能。

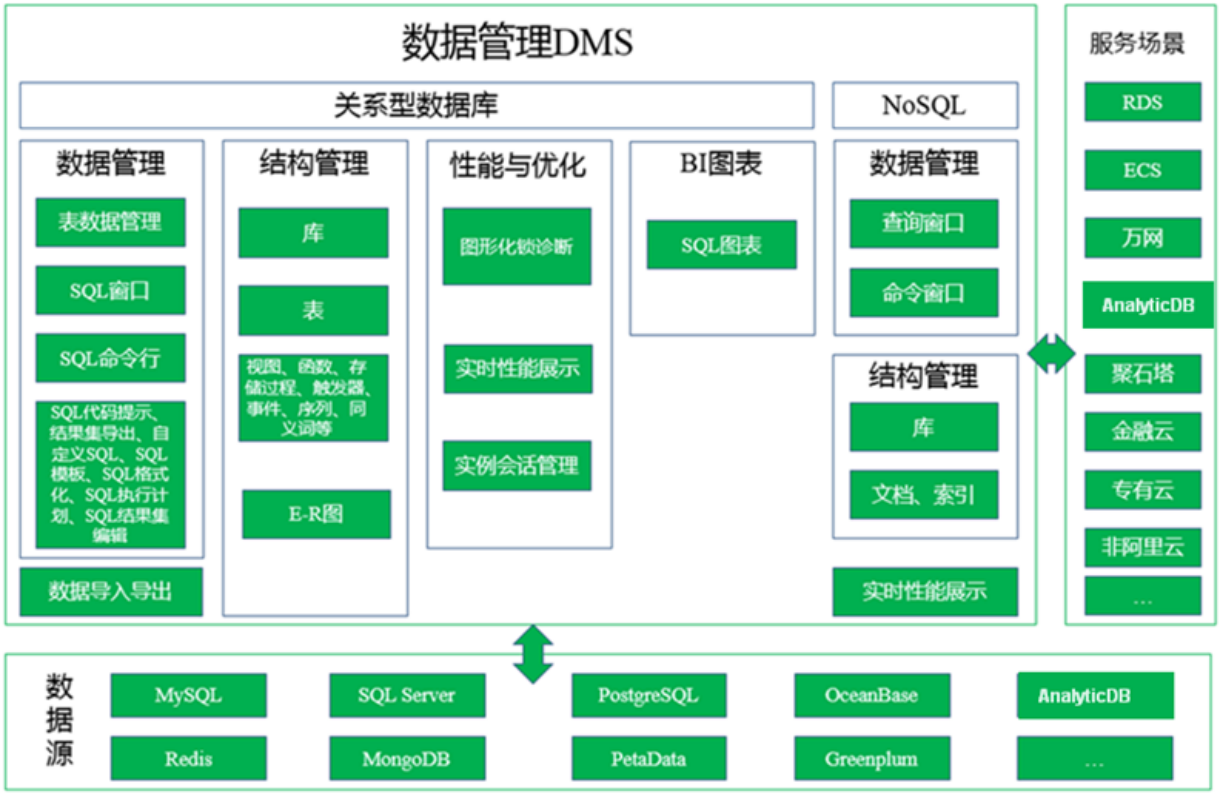
NoSQL

- 数据管理：提供查询窗口和命令窗口功能。
- 结构管理：提供库、文档和索引等对象管理功能。
- 实时性能展示：提供核心性能指标的实时展示。

功能特性图

DMS功能特性图如图 *48:* 功能特性图 所示。

图 48: 功能特性图



16.4 产品优势

本章节主要介绍DMS的产品优势。

轻松拥有数据分析能力

- 基于SQL结果集直接绘制图表。
- 灵活的行/列变化追踪，精准的库/表/行数据回滚。
- 直观的业务表读取/插入/删除/更新行数分析。

极大提升研发效率

- 表结构对比。
- SQL智能提示。
- 自定义SQL/SQL模板的复用。
- 工作环境自动恢复。
- 字典文档导出。

实时优化数据库性能

- 实用的会话管理。
- 核心指标的秒级监控。

- 图形化锁管理。
- SQL索引的实时建议。
- 整体性能的诊断报告。

丰富的数据源支持

- 关系型数据库：MySQL、SQL Server、PostgreSQL、PPAS、OceanBase、PetaData等。
- NoSQL：Redis、MongoDB等。
- OLAP：AnalyticDB等。

16.5 典型应用

16.5.1 便捷的数据操作

本章节主要介绍DMS在数据操作上的应用情况。

现存问题

用户需要通过一款便捷而功能全面的产品来完成SQL操作，保存常用的操作数据，并应用到具体的业务中。

解决方案

- 通过DMS打开表功能，用户可以类似EXECL方式操作表数据，不懂SQL也能对表数据增删改查以及统计分析。
- 通过DMS自定义SQL，用户可以保存常用业务SQL，并在管理其他数据库/实例中直接应用这些SQL。

16.5.2 实时优化数据库性能

本章节主要介绍DMS在优化数据库性能时的应用情况。

现存问题

数据库的性能优化需要以长时间详细的监控记录为基础，并进行细致分析和异常定位，才能更好地有效优化并提升其性能。

解决方案

DMS提供了数据库性能秒级监控，包含select/insert/update/delete、活跃连接数以及网络流量等指标，让用户不错过任何性能波动。

DMS提供了数据库会话中运行SQL展示，还支持结束会话，而会话分类统计能帮助用户快速定位异常SQL来源。

16.5.3 禁止数据导出

本章节主要介绍DMS在企业合作中禁止数据导出中的相关应用。

现存问题

企业合作中，通常由一方来掌控数据，另外一方负责功能开发。另一方可以访问数据，但不能将数据导出，否则会造成数据泄露的情况。

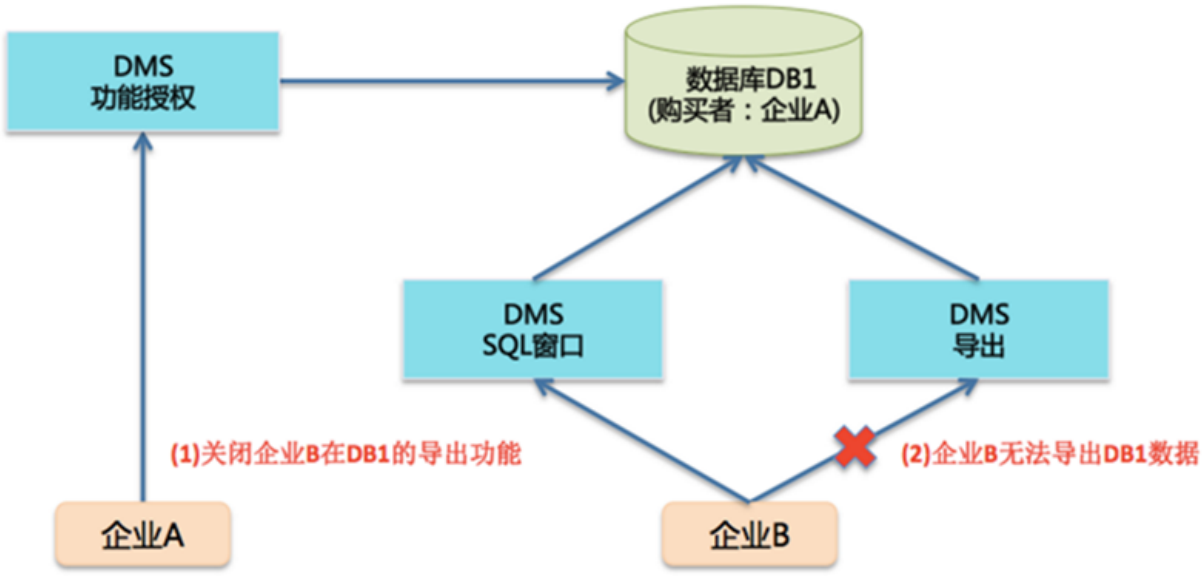
解决方案

现在，企业用户只需登录DMS，通过功能授权将数据库实例授权给合作方，但禁用数据导出功能。

合作方可以查询数据，但无法导出数据，消除了数据泄露的风险。

禁止合作方导出数据，用户如何使用**功能授权**功能来禁止合作方导出数据，如图 *49:* 【数据管理*DMS*-功能授权】用于禁止合作方导出数据所示。

图 49: 【数据管理DMS-功能授权】用于禁止合作方导出数据



16.5.4 绘制SQL结果集的图表

本章节主要介绍DMS在绘制SQL绘制结果集图表时的应用情况。

现存问题

过去，用户通常通过SQL来查出数据。然后，再将数据导入到 EXECL中，制作折线图、饼图等静态图表，过程相对繁琐。

解决方案

用户可以通过DMS直接基于SQL结果集来绘制图表，还可以制作很多高级图表，如动态图表、环比、个性化Tooltip等，让用户高质量完成工作。

16.5.5 SQL复用

本章节主要介绍DMS在SQL复用中的应用。

现存问题

用户访问数据库，几乎都会执行SQL，简单查询SQL比较容易上手，但对于复杂分析或带有业务逻辑的查询，每次重新编写成本太高，保存到文本中，需要持续维护对应关系且无法随时随地使用。

解决方案

用户通过DMS的**我的SQL**功能可以将常用SQL保存到DMS，SQL复用不受本地保存的限制，复用范围灵活，无论当前数据库、当前实例还是全部实例。

16.6 数据源支持

本章节主要介绍数据管理对各数据源的功能支持情况。

关系型数据库

关系型数据库数据源如表 *29:* 关系型数据库数据源所示。

表 29: 关系型数据库数据源

模块	功能	MySQL	SQL Server	PostgreSQL	PPAS	OceanBase	PetaData
数据管理	表数据管理	√	√	√	√	√	√
	SQL窗口	√	√	√	√	√	√
	SQL命令行	√		√	√	√	√

模块	功能	MySQL	SQL Server	PostgreSQL	PPAS	OceanBase	PetaData
	SQL模板	√				√	
	SQL格式化	√	√	√	√	√	√
	自定义SQL	√	√			√	√
	SQL智能提示	√	√			√	√
	SQL执行计划	√	√	√	√		
结构管理	库管理	√	√	√	√	√	
	表管理	√	√	√	√	√	
	索引、视图、存储过程、函数、触发器、事件等对象管理	√	√	√	√		
	E-R图展示	√	√				
	数据字典	√					
BI图表	常规SQL图表	√（RDS）					
性能与优化	实例会话	√	√			√	
	图形化锁诊断	√					
	实时性能	√		√	√		
导入导出	基础导入导出	√	√	√	√	√	√
	超大数据量导出	√	√	√	√		

NoSQL

NoSQL数据源如表 *30: NoSQL数据源*所示。

表 30: NoSQL数据源

模块	功能	Redis	MongoDB
数据管理	查询窗口	√	√
	命令窗口	√	
结构管理	库管理		√

模块	功能	Redis	MongoDB
	文档管理		√
	索引管理		√
性能与优化	实时性能	√	√

16.7 普通版和高级版区别

本章节主要介绍数据管理中普通版和高级版的功能参数差异，如表 *31:* 普通版和高级版差异所示。

表 31: 普通版和高级版差异

模块	功能	免费版	高级版
数据管理	表数据管理	√	√
	SQL窗口	√	√
	SQL命令行	√	√
	SQL模板	√	√
	SQL格式化	√	√
	自定义SQL		√
	SQL智能提示	√	√
	SQL执行计划	√	√
结构管理	库管理	√	√
	表管理	√	√
	索引、视图、存储过程、函数、触发器、事件等对象管理	√	√
	E-R图展示	√	√
	数据字典		√
	表结构对比		√
BI图表	常规SQL图表	√	√
性能与优化	诊断报告	√	√
	实例会话	√	√
	图形化锁诊断	√	√

模块	功能	免费版	高级版
	实时性能	√	√
访问安全	云账号授权	√	√
	访问地址授权	√	√
	访问日志		√
	功能授权		√
数据趋势	表数据变化		√
	库数据变化		√
导入导出	基础导入导出	√	√
	超大数据量导出		√
Linux管理	文件管理	√	√
	命令终端	√	√
	多屏终端	√	√
	实时监控	√	√
	系统管理	√	√

17 负载均衡SLB

17.1 产品概述

负载均衡（Server Load Balancer）是将访问流量根据转发策略分发到后端多台云服务器（Elastic Compute Service，简称 ECS） 的流量分发控制服务。通过流量分发扩展应用系统对外的服务能力，通过消除单点故障提升应用系统的可用性。

负载均衡服务通过设置虚拟服务地址，将添加的ECS虚拟成一个高性能、高可用的应用服务池。根据应用指定的方式，将来自客户端的网络请求分发到云服务器池中。

该服务地址都是每个负载均衡实例独占的，更改转发策略不会导致负载均衡服务地址的变更。您可以将域名解析到负载均衡的服务地址，对外提供服务。除非必要，不建议您删除负载均衡服务。删除了负载均衡服务以后，相应的服务配置和服务地址将会被释放掉，数据一旦删除，不可恢复。

负载均衡服务由以下三个部分组成：

- **负载均衡实例**（Load Balancer）

  如果您想使用负载均衡服务，必须先创建一个负载均衡实例。一个负载均衡实例可以添加多个监听和后端服务器。

- **监听器**（Listener）

  在使用负载均衡服务前，您必须为负载均衡实例添加一个监听，指定监听规则和转发策略，并配置健康检查。

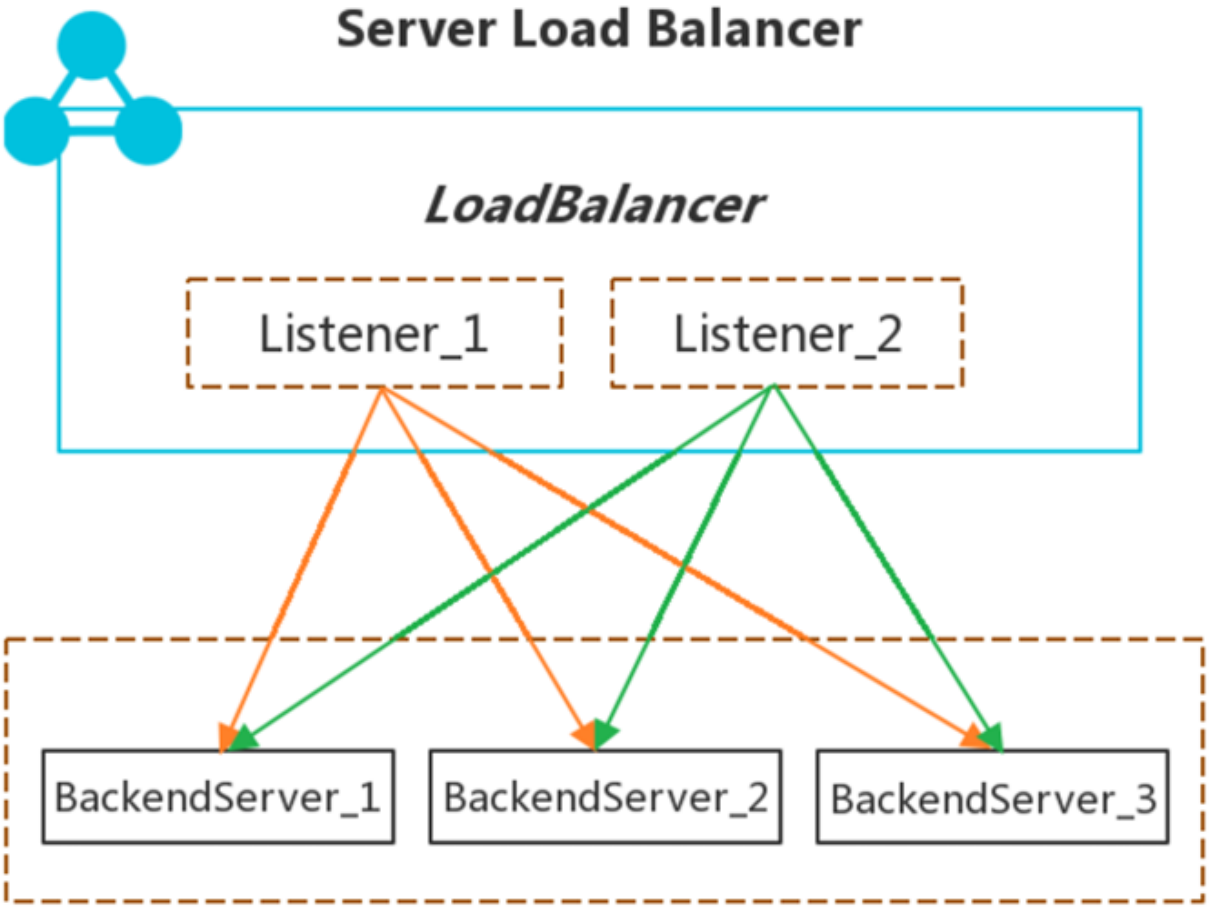
  针对不同的需求，您可以配置四层（TCP/UDP）或七层（HTTP/HTTPS）监听。

- **后端服务器**（Backend Server）

  一组接收前端请求的ECS实例。

如下图所示，来自客户端的请求经过负载均衡实例后，系统根据配置的监听规则，将请求转发到对应的后端ECS实例上。

图 50: 负载均衡构成

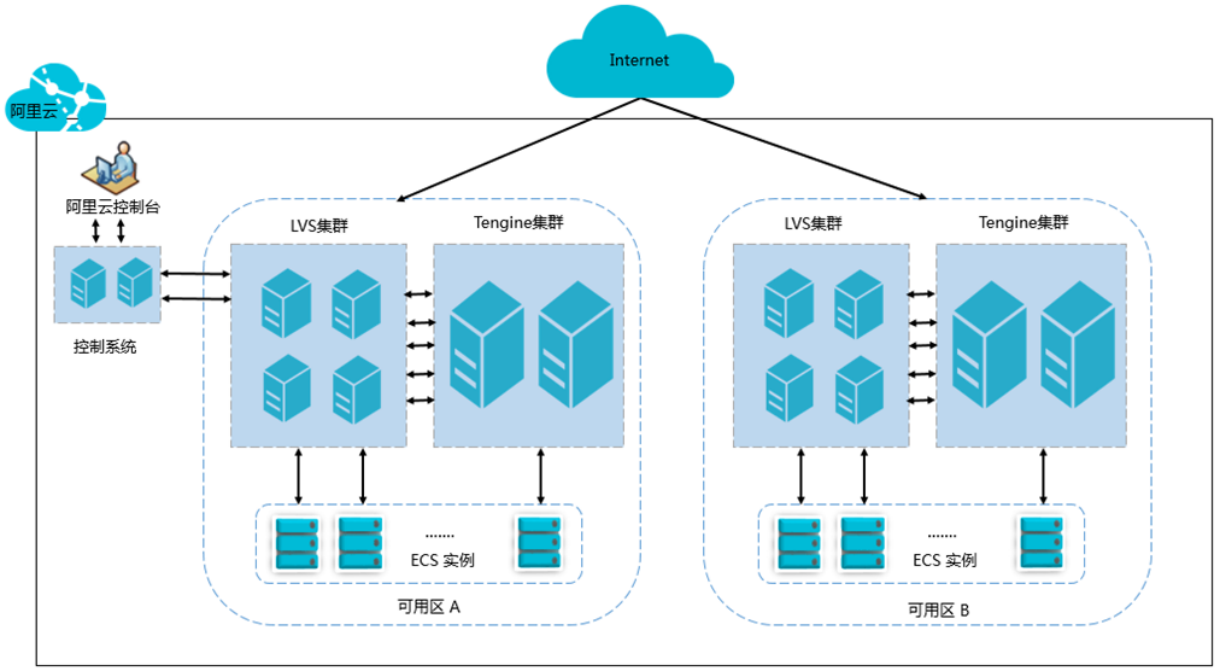


17.2 产品架构

负载均衡采用集群部署，可实现会话同步，以消除服务器单点，提升冗余，保证服务的稳定性。专有云当前提供四层（TCP协议和UDP协议）和七层（HTTP和HTTPS协议）的负载均衡服务。

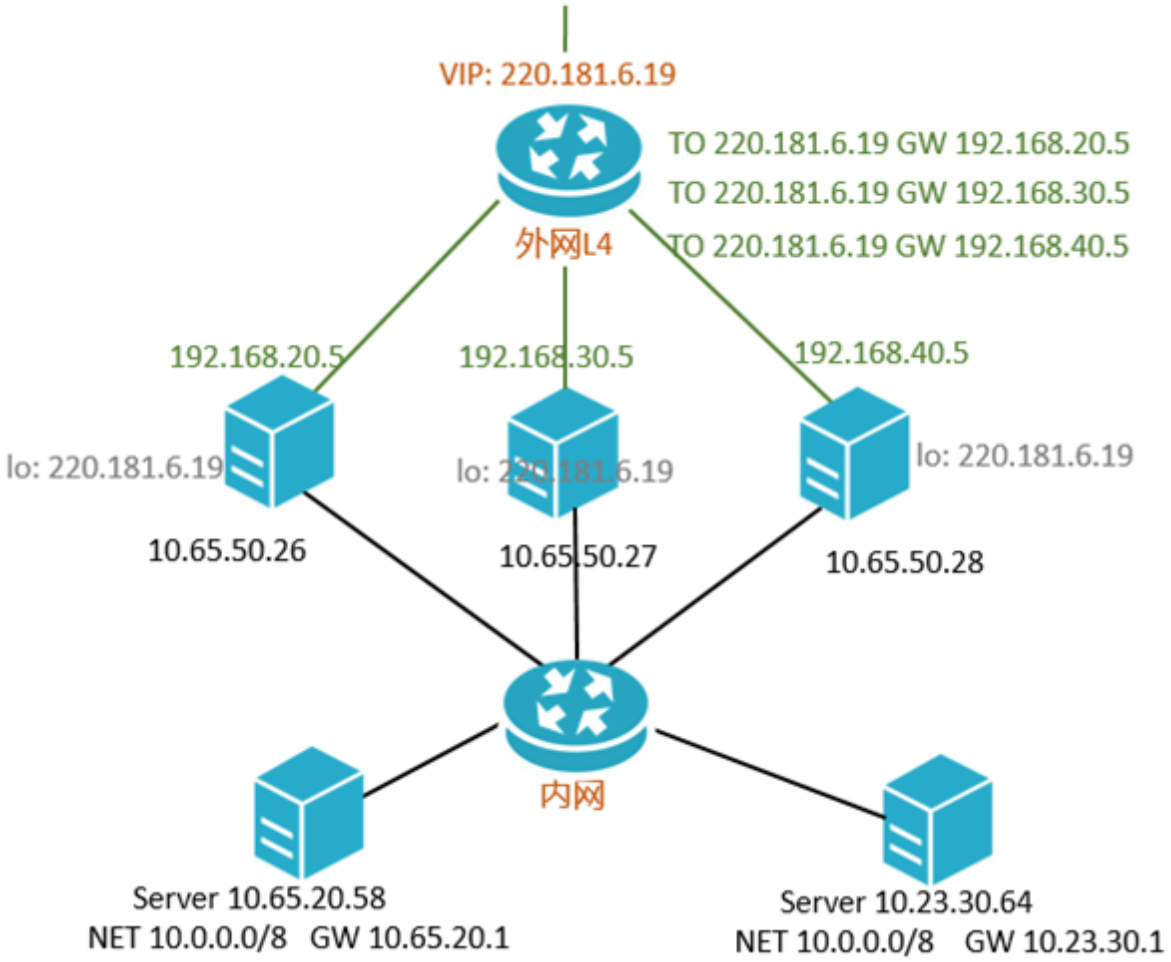
- 四层采用开源软件 LVS（Linux Virtual Server）+ keepalived的方式实现负载均衡，并根据云计算需求对其进行了定制化。
- 七层采用Tengine实现负载均衡。Tengine是由淘宝网发起的Web服务器项目，它在Nginx的基础上，针对大访问量网站的需求，添加了很多高级功能和特性。

图 51: 负载均衡架构



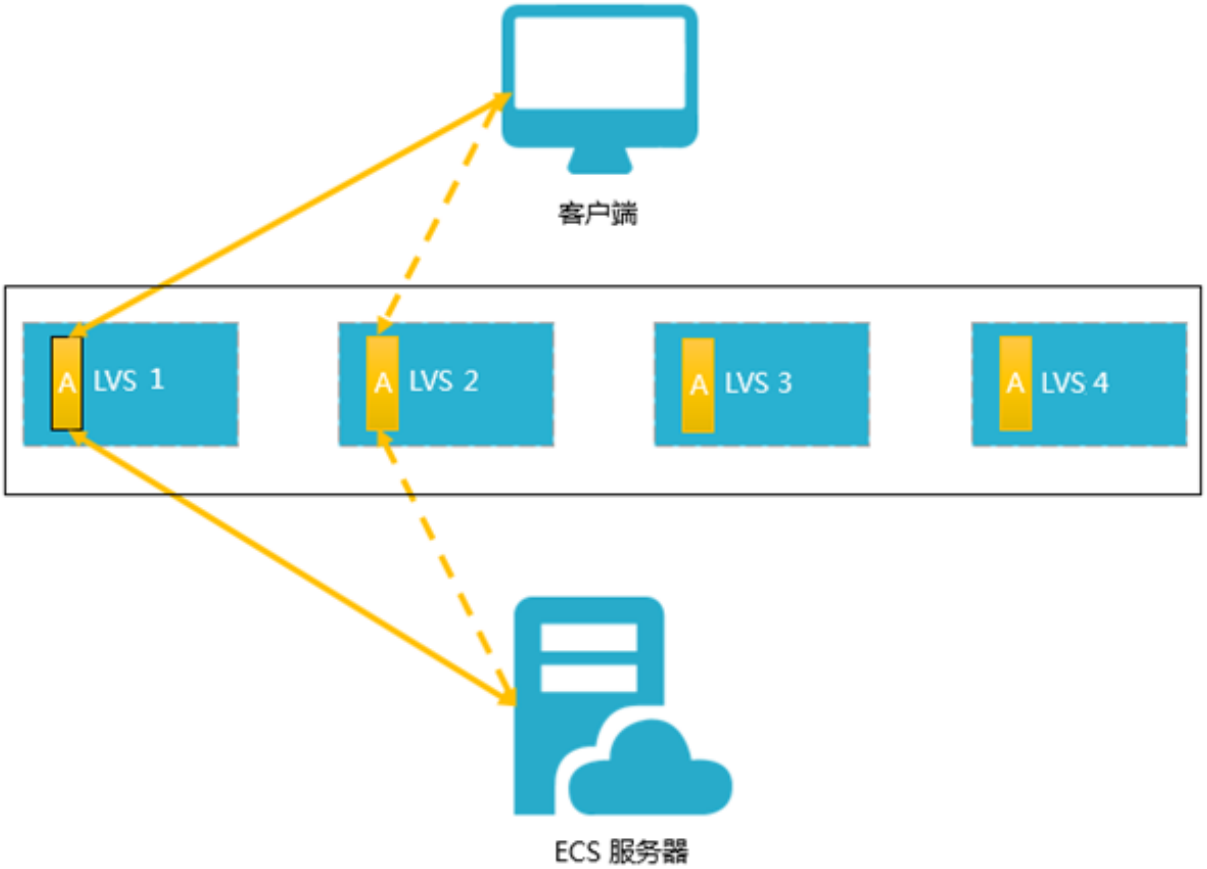
如下图所示，四层负载均衡实际上是由多台LVS机器部署成一个LVS集群来运行的。采用集群部署模式极大地保证了异常情况下负载均衡服务的可用性、稳定性与可扩展性。

图 52: 集群部署



LVS集群内的每台LVS都会进行会话，通过组播报文同步到该集群内的其它LVS机器上，从而实现LVS集群内各台机器间的会话同步。如下图所示，当客户端向服务端传输三个数据包后，在LVS1上建立的会话A开始同步到其它LVS机器上。图中实线表示现有的连接，图中虚线表示当LVS1出现故障或进行维护时，这部分流量会走到一台可以正常运行的机器LVS2上。因而负载均衡集群支持热升级，并且在机器故障和集群维护时最大程度对用户透明，不影响用户业务。

图 53: 会话同步



17.3 功能特性

协议支持

当前提供四层（TCP协议和UDP协议）和七层（HTTP和HTTPS协议）的负载均衡服务。

健康检查

支持对后端ECS进行健康检查。负载均衡服务会自动屏蔽异常状态的ECS，待该ECS恢复正常后自动解除屏蔽。

会话保持

提供会话保持功能。在会话的生命周期内，可以将同一客户端的请求转发到同一台后端ECS上。

调度算法

支持轮询、最小连接数两种调度算法。

- 轮询：按照访问次数依次将外部请求依序分发到后端ECS上。
- 最小连接数：连接数越小的后端服务器被轮询到的次数（概率）也越高。

访问控制

支持白名单访问控制。通过添加负载均衡监听的访问白名单，仅允许特定IP访问负载均衡服务。

证书管理

针对HTTPS协议，提供统一的证书管理服务。证书无需上传到后端ECS，解密处理在负载均衡上进行，降低后端ECS CPU开销。

实例类型

提供内网和外网两种类型的负载均衡服务。您可以根据业务场景来选择配置对外公开或对内私有的负载均衡服务。

管理方式

提供控制台、API、SDK多种管理方式。

17.4 产品优势

高可用

采用全冗余设计，无单点，支持同城容灾，可用性高达99.95%。根据应用负载进行弹性扩容，在流量波动情况下不中断对外服务。

低成本

与传统硬件负载均衡系统高投入相比，成本可下降60%。内网类型实例免费使用，无需一次性采购昂贵的负载均衡设备，无需运维投入。

安全

结合云盾提供防DDoS攻击能力，包括CC、SYN flood等DDoS攻击方式。

17.5 典型应用

负载均衡主要应用于以下场景中：

场景一：**应用于高访问量的业务**

如果您的应用访问量很高，您可以通过配置监听规则将流量分发到不同的ECS实例上。此外，您可以使用会话保持功能将同一客户端的请求转发到同一台后端ECS上，提高访问效率。

场景二：**横向扩张系统的服务能力**

您可以根据业务发展的需要，通过随时添加和移除ECS实例来扩展应用系统的服务能力，适用于各种Web服务器和App服务器。

场景三：消除单点故障

您可以在负载均衡实例下添加多台ECS实例。当其中一部分ECS发生故障后，负载均衡会自动屏蔽故障的ECS实例，将请求分发给正常运行的ECS实例，保证应用系统仍能正常工作。

17.6 使用限制

- 在四层（TCP协议）服务中，不支持后端ECS既作为Real Server，又作为客户端向所在的负载均衡实例发送请求。因为，返回的数据包只在云服务器内部转发，不经过负载均衡，所以使用负载均衡内的后端ECS去访问负载均衡的服务地址是不通的。
- 在通过负载均衡对外提供服务前，首先要确保已经完成并正确配置了所有负载均衡后端ECS上的应用服务，且能通过ECS的服务地址访问该服务。
- 负载均衡不提供ECS间的数据同步服务。如果部署在负载均衡后端ECS上的应用服务是无状态的，那么可以通过独立的ECS或RDS服务来存储数据；如果部署在负载均衡后端ECS上的应用服务是有状态的，那么需要确保这些ECS上的数据是同步的。
- 当负载均衡实例的服务地址已经进行了域名解析，对外服务时，请不要随意删除该负载均衡实例。删除负载均衡实例会将实例的服务地址一同释放掉，从而导致服务中断。

17.7 基本概念

负载均衡

阿里云计算提供的一种网络负载均衡服务。结合ECS ，提供四层和七层负载均衡服务。

负载均衡实例

负载均衡实例是一个运行的负载均衡服务。要使用负载均衡服务，必须先创建一个负载均衡实例。LoadBalancerId 是识别负载均衡实例的唯一标识。

负载均衡服务地址

负载均衡实例的IP地址。根据创建的负载均衡实例的类型，服务地址可能是外网IP也可能是内网IP。

监听

负载均衡服务监听规定了如何将请求转发给后端服务器。监听配置包括监听端口、负载均衡策略和健康检查配置等，每个监听对应后端的一个应用服务。

后端服务器

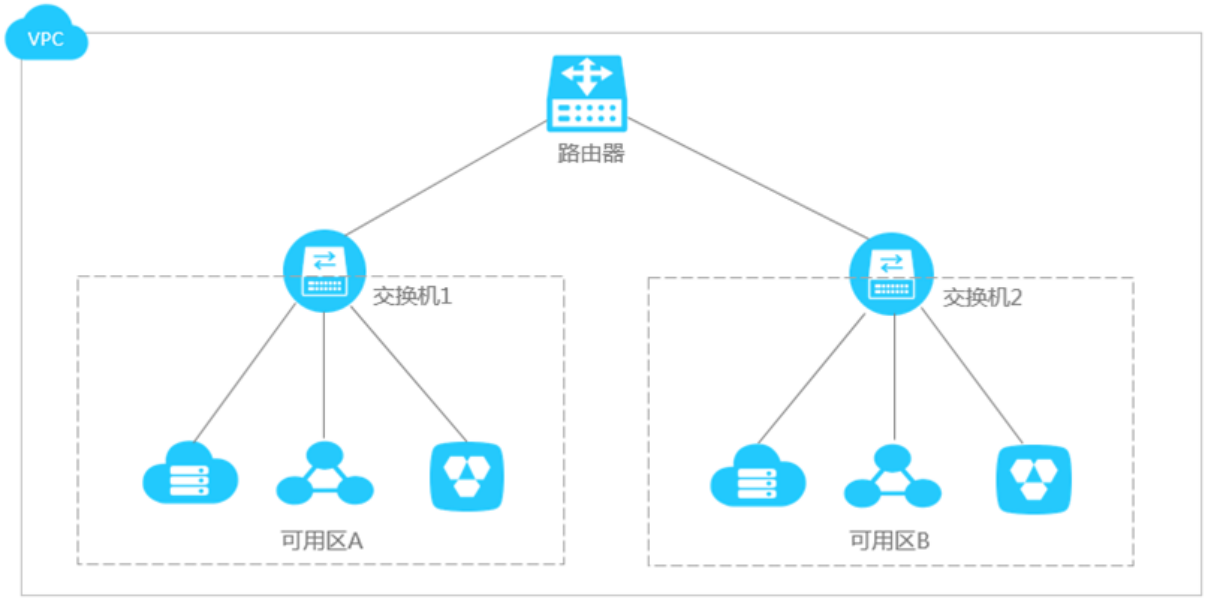
接收负载均衡分发请求的ECS实例。负载均衡服务将访问请求按照设定的规则转发到添加的后端ECS上进行处理。

18 专有网络VPC

18.1 产品概述

专有网络VPC（Virtual Private Cloud），帮助您基于阿里云构建出一个隔离的网络环境。您可以完全掌控自己的虚拟网络，包括选择自有IP地址范围、划分网段、配置路由表和网关等。此外您也可以通过专线 、VPN等连接方式将VPC与传统数据中心组成一个按需定制的网络环境，实现应用的平滑迁移上云。

图 54: 专有网络



专有网络和经典网络

阿里云提供如下两种网络类型：

- 经典网络

  经典网络类型的云产品，统一部署在阿里公共基础内，规划和管理由阿里云负责，更适合对网络易用性要求比较高的客户。

- 专有网络

  专有网络是一个可以自定义隔离专有网络，您可以自定义这个专有网络的拓扑和IP地址，适用于对网络安全性要求较高和有一定的网络管理能力的客户。

经典网络和专有网络的功能差异如下表所示。

表 32: 经典网络和专有网络功能对比

功能	经典网络	专有网络
二层逻辑隔离	不支持	支持
自定义私网网段	不支持	支持
私网IP规划	经典网络内唯一	专有网络内唯一，专有网络间可重复
自建VPN	不支持	支持
私网互通	账号内相同地域内互通	专有网络内互通，专有网络间隔离
路由表	不支持	支持
交换机	不支持	支持
自定义路由器	不支持	支持
SDN	不支持	支持
隧道技术	不支持	支持
自建NAT网关	不支持	支持

18.2 产品架构

背景信息

随着云计算的不断发展，对虚拟化网络的要求越来越高，比如弹性（scalability）、安全性（security）、可靠性（reliability）、私密性（privacy），并且还有极高的互联性能（performance）需求，因此催生了多种多样的网络虚拟化技术。
比较早的解决方案，是将虚拟机的网络和物理网络融合在一起，形成一个扁平的网络架构，例如大二层网络。随着虚拟化网络规模的扩大，这种方案中的ARP欺骗、广播风暴、主机扫描等问题会越来越严重。为了解决这些问题，出现了各种网络隔离技术，把物理网络和虚拟网络彻底隔开。其中一种技术是用户之间用VLAN进行隔离，但是VLAN的数量最大只能支持到4096个，无法支撑公共云的巨大用户量。

原理描述

基于目前主流的隧道技术，专有网络（Virtual Private Cloud，简称VPC）隔离了虚拟网络。每个VPC都有一个独立的隧道号，一个隧道号对应着一张虚拟化网络。一个VPC内的ECS之间的传输数据包都会加上隧道封装，带有唯一的隧道ID标识，然后送到物理网络上进行传输。不同VPC内

的ECS因为所在的隧道ID不同，本身处于两个不同的路由平面，所以不同VPC内的ECS无法进行通信，天然地进行了隔离。

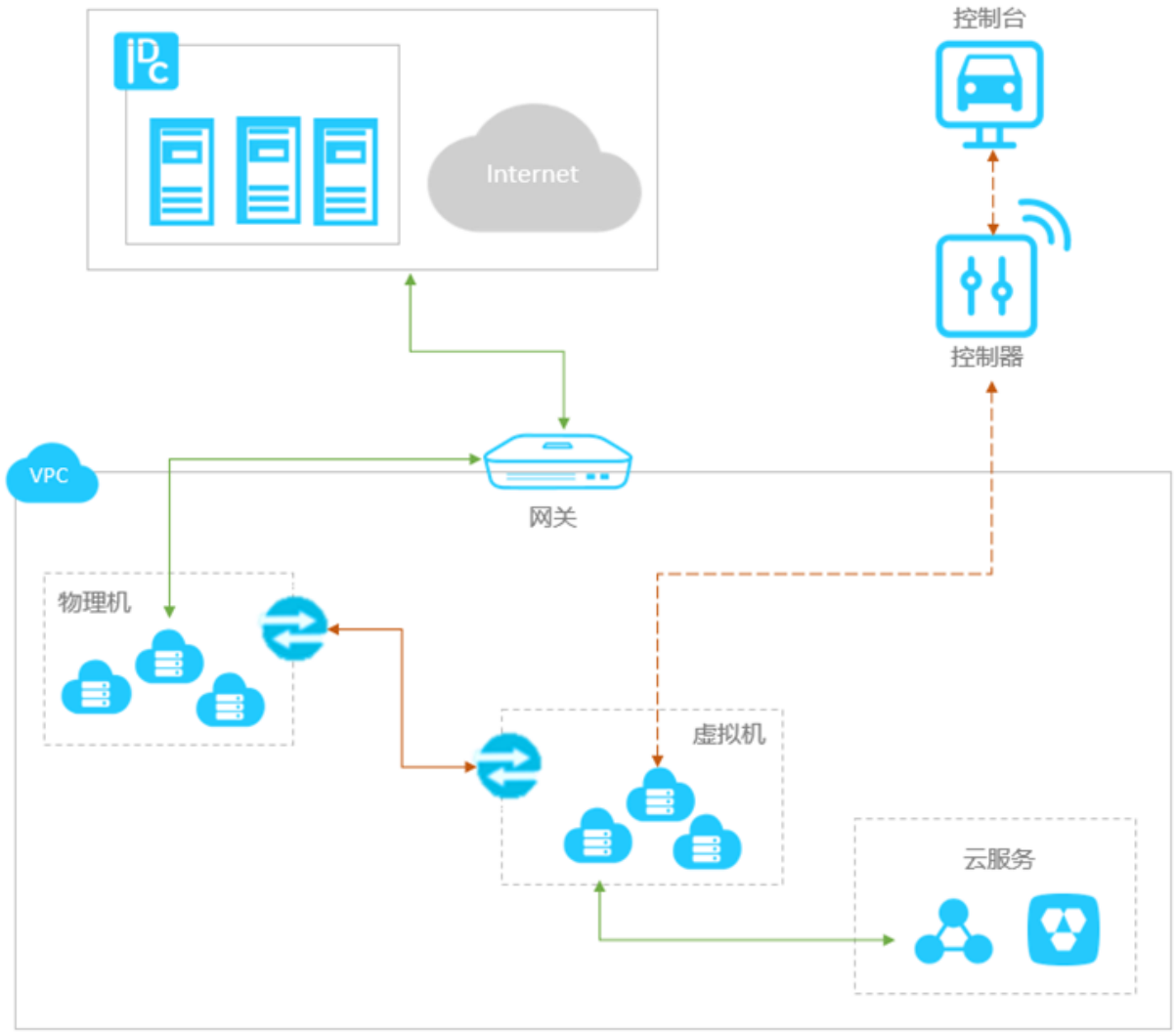
基于隧道技术，阿里云的研发团队自研了交换机，软件自定义网络（Software Defined Network，简称SDN）技术和硬件网关，在此基础上实现了VPC产品。

逻辑架构

如下图所示，VPC包含交换机、网关和控制器三个重要的组件。

- 交换机和网关组成了数据通路的关键路径，控制器使用自研的协议下发转发表到网关和交换机，完成了配置通路的关键路径，整体架构里面，配置通路和数据通路互相分离。
- 交换机是分布式的结点，网关和控制器都有集群部署并且是多机房互备的，所有链路上都有冗余容灾，提升了VPC产品的整体可用性。
- 交换机和网关性能在业界都是领先的，自研的SDN协议和控制器，能轻松管控成千上万张虚拟网络。

图 55: 专有网络架构



在产品上，除了给您一张独立的虚拟化网络，阿里云还为每个VPC提供了独立的路由器、交换机组件，让您可以更加丰富地进行组网。

针对内网安全需求，您可以使用安全组功能在一个VPC内进行更加细粒度的访问控制和隔离。缺省情况下，VPC内的ECS只能和本VPC内的ECS或云服务进行通信。此外，您可以使用阿里云提供的和VPC相关的产品，比如弹性外网IP和路由器接口等使VPC可以和外网、其它VPC进行通信。

18.3 功能特性

私网网段

在创建VPC和交换机时您需要指定专有网络和其子网的网段。每个专有网络只能指定一个网段，网段范围如下表所示。

表 33: 专有网络网段

网段	可用主机数	备注
192.168.0.0/16	65532	去除系统占用地址
172.16.0.0/12	1048572	去除系统占用地址
10.0.0.0/8	16777212	去除系统占用地址

交换机

交换机是组成专有网络的基础网络设备，它可以连接不同的云产品实例。创建专有网络之后，您可以通过添加交换机为专有网络划分一个或多个子网。在创建交换机时，您也要指定交换机的网段，交换机的网段可以和它所属的VPC网段一样或者是其VPC网段的子集，子网掩码必须在16到29之间。

路由器

路由器是一个专有网络的枢纽。作为专有网络中重要的功能组件，它可以连接VPC内的各个交换机，同时也是连接VPC与其它网络的网关设备。

每个路由器中维护一张路由表，它会根据具体的路由条目的设置来转发网络流量。您创建VPC时，系统会自动为VPC创建一个路由器。删除VPC时，系统也会自动删除对应的路由器。目前不支持直接创建和删除路由器。

路由表

路由表是指路由器上管理路由条目的列表。新建VPC时，系统会自动创建一个路由表。删除VPC时，系统也会自动删除对应的路由表。不支持直接创建和删除路由表。

说明: 每个路由器只能有一个路由表。路由表中的路由条目会影响VPC中的所有云产品实例。

路由条目

路由表中的每一项是一条路由条目，路由条目定义了通向指定目标网段的网络流量的下一跳地址，路由条目包括系统路由和自定义路由两种类型的路由条目。

专有网络创建时，系统会自动创建1条系统路由条目，用于专有网络内的云产品实例访问专有网络外的云服务。创建交换机，系统也会创建1条对应的系统路由条目，目的地址为所创建交换机的网段。您可以创建和删除自定义路由条目。

选路规则

路由表中采用最长前缀匹配作为流量的路由选路规则。最长前缀匹配是指IP网络中当路由表中有多条条目可以匹配目的IP时，采用掩码最长（最精确）的一条路由作为匹配项并确定下一跳。

例如，某专有网络中路由表中路由条目如下表所示。

表 34: 路由表示例

目标网段	下一跳类型	下一跳地址	类型
100.64.0.0/10	-	-	System
192.168.0.0/24	-	-	System
0.0.0.0/0	Instance	i-12345678	Custom
10.0.0.0/24	Instance	i-87654321	Custom

目的地址为 100.64.0.0/10和192.168.0.0/24的两条路由均为系统路由，前者为系统保留的地址段，后者为专有网络中为交换机配置的系统路由。

目的地址为0.0.0.0/0和10.0.0.0/24的两条路由为自定义路由，表示将访问0.0.0.0/0地址段的流量转发至ID为i-12345678的ECS实例，将访问10.0.0.0/24地址段的流量转发至ID为 i-87654321的ECS实例。根据最长前缀匹配规则，在该专有网络中，访问10.0.0.1的流量会转发至 i-87654321，而访问10.0.1.1的流量会转发至i-12345678。

18.4 产品优势

安全隔离

- 使用隧道技术，达到与传统VLAN方式相同的隔离效果。
- 广播域隔离在网卡级别。
- VLAN级别的隔离，彻底阻断网络通讯。
- 划分不同的安全域，进行访问控制。

访问控制

- 灵活的访问控制规则。
- 满足政务，金融的安全隔离规范。

软件定义网络

- 按需配置网络设置，软件定义网络。

- 管理操作实时生效。

丰富的网络连接方式

- 支持软件VPN。
- 支持专线连接。

易用

- 软件定义网络，自由组网。
- 专线VPN接入，实现传统架构平滑迁移。
- 网络规模弹性伸缩，突破物理设备限制。

18.5 典型应用

专有网络可用于以下经典场景：

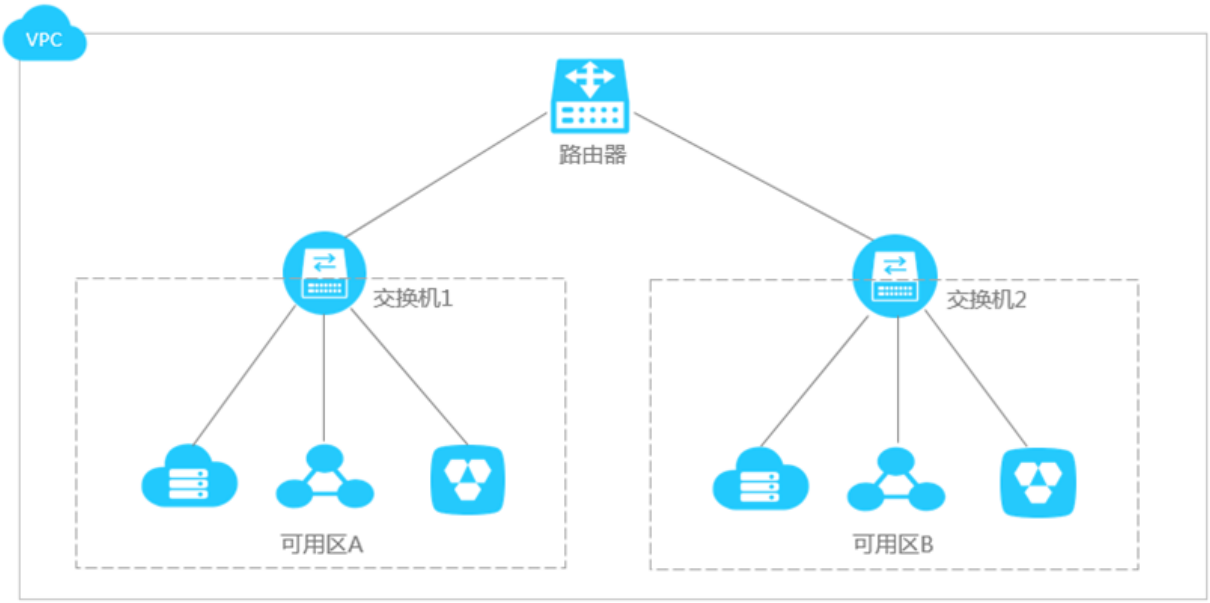
安全隔离

希望在云上构建一个完全隔离的网络环境，您可以自主定义私有网络配置。

多可用区容灾

您可以通过将资源部署在不同可用区的交换机中，实现跨可用区容灾。

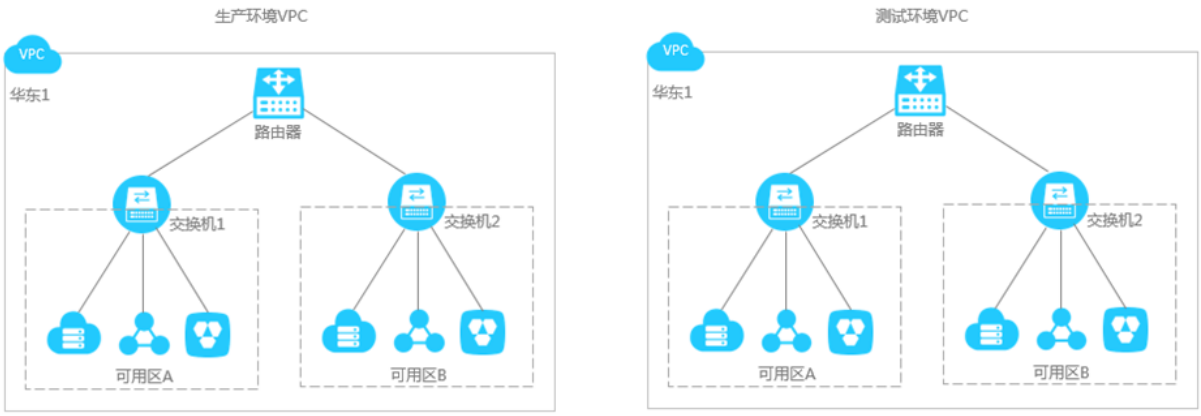
图 56: 多可用区部署



业务系统隔离

如果您有多个业务系统需要通过VPC进行严格隔离，比如生产环境和测试环境要严格进行隔离，那么可以使用多个VPC。

图 57: 业务隔离架构



18.6 基本概念

专有网络

专有网络是基于阿里云创建的自定义私有网络，不同的专有网络之间彻底逻辑隔离，您可以在自己创建的专有网络内创建和管理云产品实例，例如ECS、负载均衡、RDS等。

路由器

路由器是VPC网络的枢纽，它可以连接VPC内的各个交换机，同时也是连接VPC与其他网络的网关设备。它会根据具体的路由条目的设置来转发网络流量。

交换机

交换机是组成VPC网络的基础网络设备。它可以连接不同的云产品实例。在VPC网络内创建云产品实例时，必须指定云产品实例所在的交换机。

路由表

路由表是指路由器上管理路由条目的列表。

路由条目

路由表中的每一项是一条路由条目，路由条目定义了通向指定目标网段的网络流量的下一跳地址。路由条目包括系统路由和自定义路由两种类型。

18.7 VPC通信

VPC是隔离的私有网络，默认VPC与VPC之间、VPC与外网无法通信。但您可使用弹性外网IP、NAT网关和路由器接口等功能实现VPC互通或与外网通信。

18.7.1 弹性外网IP

弹性外网IP（Elastic IP Address，简称EIP），是可以独立持有的外网IP地址资源，能动态绑定到不同的ECS实例上，绑定和解绑时无需停机。

弹性外网IP是一种NAT IP。它实际位于阿里云的公网网关上，通过NAT方式映射到了被绑定在ECS实例位于私网的网卡上。因此，绑定了弹性外网IP的ECS实例可以直接使用这个IP进行公网通信，但是在网卡上并不能看到这个IP地址。

EIP具有以下特性：

- 弹性绑定

  您可以在需要时将该弹性外网IP绑定到ECS实例上，使绑定的ECS实例具备使用该IP地址进行公网通信的能力；在不需要时，可以将之解绑。
- 可配置的网络能力

  您可以根据需要来调整弹性外网IP的带宽值，带宽的修改即时生效。

18.7.2 NAT网关

NAT网关（NAT Gateway）是一款企业级的VPC公网网关，提供NAT代理（SNAT、DNAT）、10Gbps 级别的转发能力、以及跨可用区的容灾能力。

NAT网关提供以下功能：

- SNAT

  NAT网关提供SNAT功能，为VPC内无公网IP的ECS实例提供访问互联网的代理服务。
- DNAT

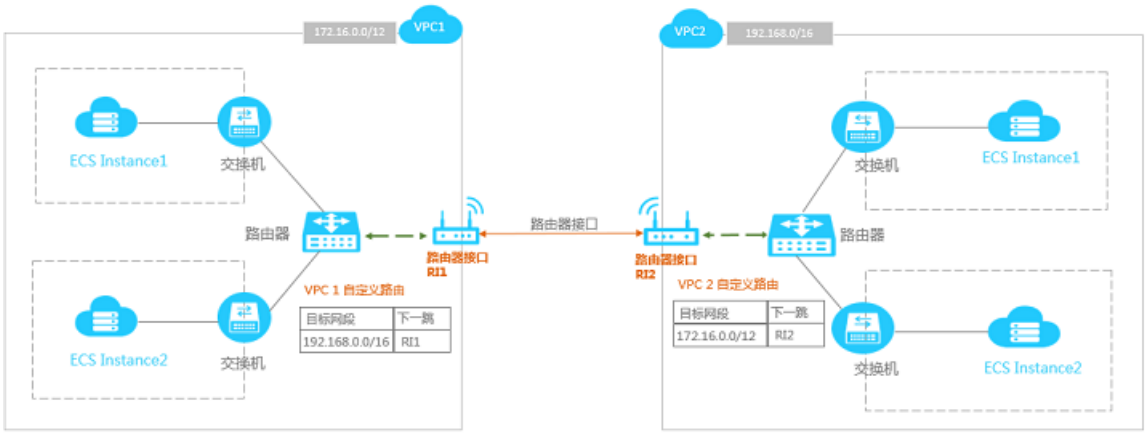
  NAT网关支持DNAT功能，将NAT网关上的公网IP映射给ECS实例使用，使得ECS可以面向互联网提供服务。DNAT支持端口映射和IP映射。

18.7.3 路由器接口

路由器接口是一种虚拟设备，具备搭建通信通道并控制其工作状态的功能。通过在VPC的路由器上分别创建路由器接口，并进行互连，从而使这两个VPC可以通过这个通道向对方转发消息。

路由器接口支持同账号和跨账号VPC互通。

图 58: VPC互通



19 日志服务

19.1 产品概述

日志服务（Log Service，简称 Log）是针对日志类数据的一站式服务，在阿里巴巴集团经历大量大数据场景锤炼而成。

日志服务（Log Service，简称 Log）是针对日志类数据的一站式服务，在阿里巴巴集团经历大量大数据场景锤炼而成。您无需开发就能快捷完成日志数据采集、消费、投递以及查询分析等功能，提升运维、运营效率，建立 DT 时代海量日志处理能力。

核心功能如下：

实时采集与消费（LogHub）

功能：

- 通过ECS、容器、移动端，开源软件，JS等接入实时日志数据（例如Metric、Event、BinLog、TextLog、Click等）
- 提供实时消费接口，与实时计算及服务对接

用途：数据清洗（ETL），流计算（Stream Compute），监控与报警， 机器学习与迭代计算

投递数仓（LogShipper）

稳定可靠的日志投递。将日志中枢数据投递至存储类服务进行存储与大数据分析。支持压缩、自定义Partition、以及行列等各种存储方式

用途：数据仓库 + 数据分析, 审计, 推荐系统与用户画像

查询与实时分析（Search/Analytics）

实时索引、查询分析数据数据。

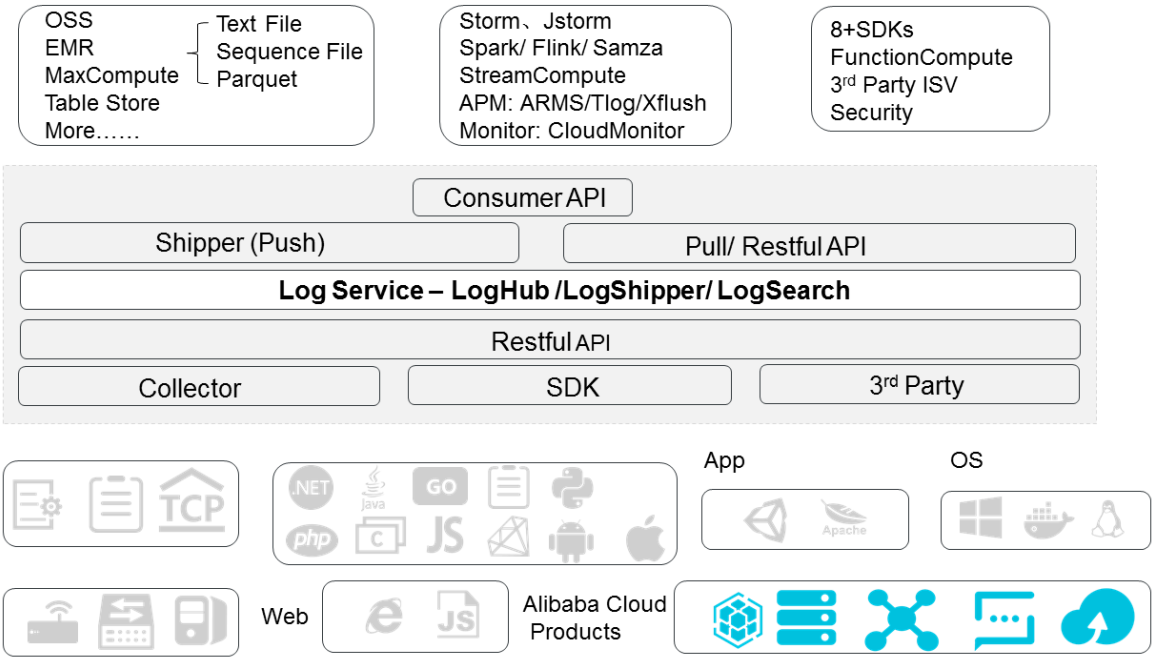
- 查询：关键词、模糊、上下文、范围
- 统计：SQL聚合等丰富查询手段
- 可视化：Dashboard + 报表功能
- 对接：Grafana，JDBC/SQL92

用途：DevOps/线上运维，日志实时数据分析，安全诊断与分析，运营与客服系统

19.2 产品架构

日志服务的架构如下图所示。

图 59: 产品架构



Logtail

帮助您快速收集日志的Agent。其特点如下所示：

- 基于日志文件、无侵入式的收集日志
  - 只读取文件。
- 安全、可靠
  - 支持文件轮转不丢失数据。
  - 支持本地缓存。
  - 网络异常重试。
- 方便管理
  - Web端可视化配置
- 完善的自我保护
  - 实时监控进程CPU、内存消耗，限制使用上限。

前端服务器

采用LVS + Nginx构建的前端机器。其特点如下所示：

- HTTP、REST协议
- 水平扩展
  - 流量上涨时可快速通过增加前端机来提高处理能力。
- 高吞吐、低延时
  - 纯异步处理，单个请求异常不会影响其他请求。
  - 内部采用专门针对日志的Lz4压缩，提高单机处理能力，降低网络带宽。

后端服务器

后端是分布式的进程，部署在多个机器上，完成实时对Logstore数据的持久化、索引、查询以及投递至MaxCompute。整体后端服务的特点如下所示：

- 数据高安全性：
  - 您写入的每条日志，都会被保存3份。
  - 任意磁盘损坏、机器宕机情况下，数据自动复制修复。
- 稳定服务：
  - 进程崩溃和机器宕机时，Logstore会自动迁移。
  - 自动负载均衡，确保无单机热点。
  - 严格的Quota限制，防止单个用户行为异常对其他用户产生影响。
- 水平扩展：
  - 以分区（Shard）为单位进行水平扩展，用户可以按需动态增加分区来增加吞吐量。

19.3 产品优势

全托管服务

- 用性强，5分钟即可接入服务进行使用，Agent支持任意网络下数据采集
- LogHub覆盖Kafka 100%功能，并提供完整监控、报警等功能数据，弹性伸缩等（可支持PB/Day规模），使用成本为自建50%以下
- LogSearch/Analytics 提供保存查询、仪表盘和报警功能、使用成本为自建 20%以下
- 0+接入方式，与云产品 （OSS/E-MapReduce/MaxCompute/Table Store/MNS/CDN/ARMS等）、开源软件（Storm、Spark）无缝对接

生态丰富

- LogHub 支持30+采集端，包括Logstash、Fluent等，无论是从嵌入式设备，网页，服务器，程序等都能轻松接入。在消费端，支持与Spark Streaming、Storm、Spark Streaming、云监控、ARMS等对接
- LogShipper 支持丰富数据格式（TextFile、SequenceFile、Parquet等），支持自定义Partition，数据可以直接被Presto、Hive、Spark、Hadoop、E-MapReduce、MaxCompute、HybridDB等存储引擎
- LogSearch/Analytics 查询分析语法完整，兼容SQL92，即将提供：JDBC协议与Grafana对接

实时性强

- LogHub：写入即可消费；Logtail（采集Agent）实时采集传输，1秒内到服务端（99.9%情况）
- LogSearch/Analytics：写入即可查询分析，在多个查询条件下1秒可查询10亿级数据，多个聚合条件下1秒可分析1亿级数据

完整API/SDK

- 轻松支持自定义管理及二次开发
- 所有功能均可通过API/SDK实现，提供多种语言SDK，可轻松管理服务和百万级设备
- 查询分析语法简单便捷（兼容SQL92），接口友好适合与生态软件对接（即将提供Grafana对接方案）

19.4 应用场景

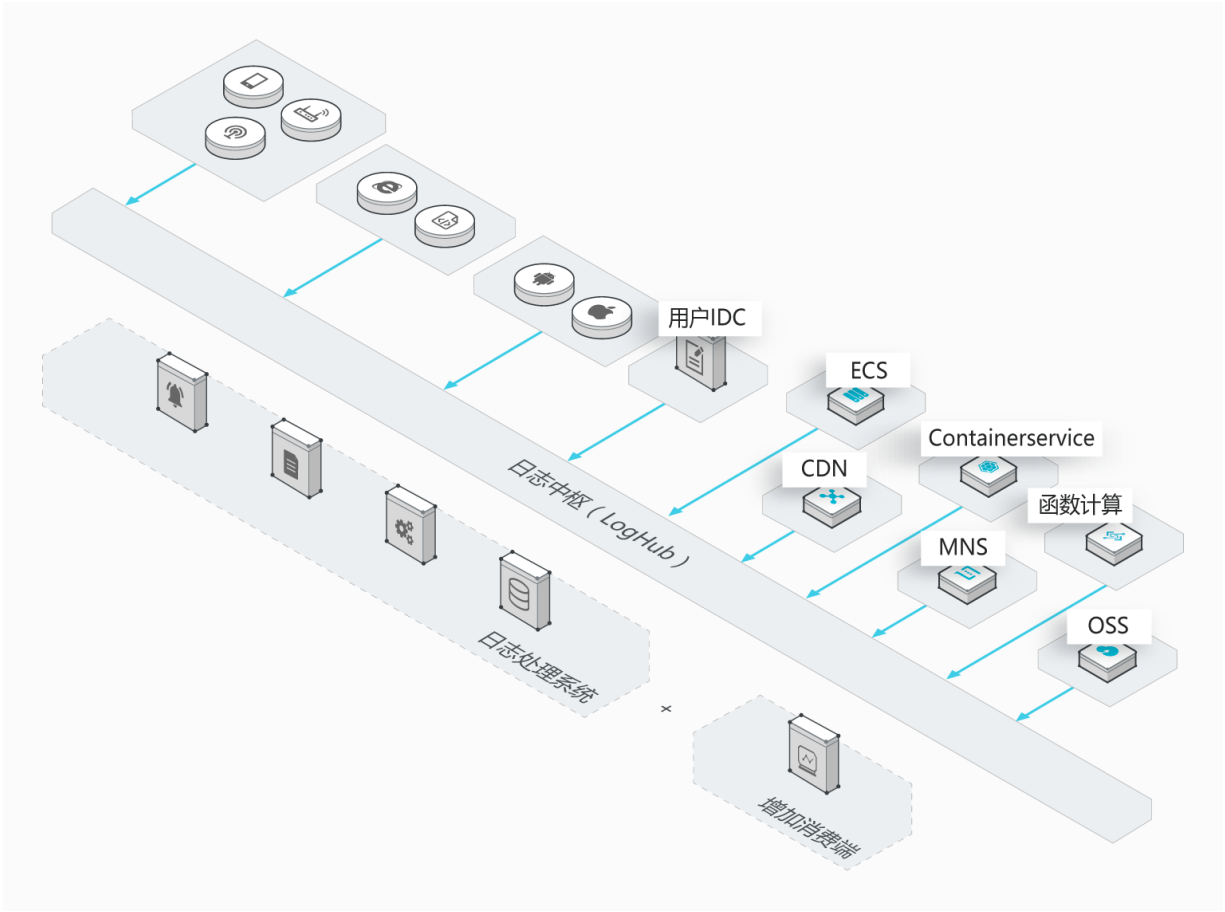
日志服务的典型应用场景包括：数据采集、实时计算、数仓与离线分析、产品运营与分析、运维与管理等场合。典型应用场景如下。

数据采集与消费

通过日志服务LogHub功能，可以大规模低成本接入各种实时日志数据（包括Metric、Event、BinLog、TextLog、Click等）。

方案优势：

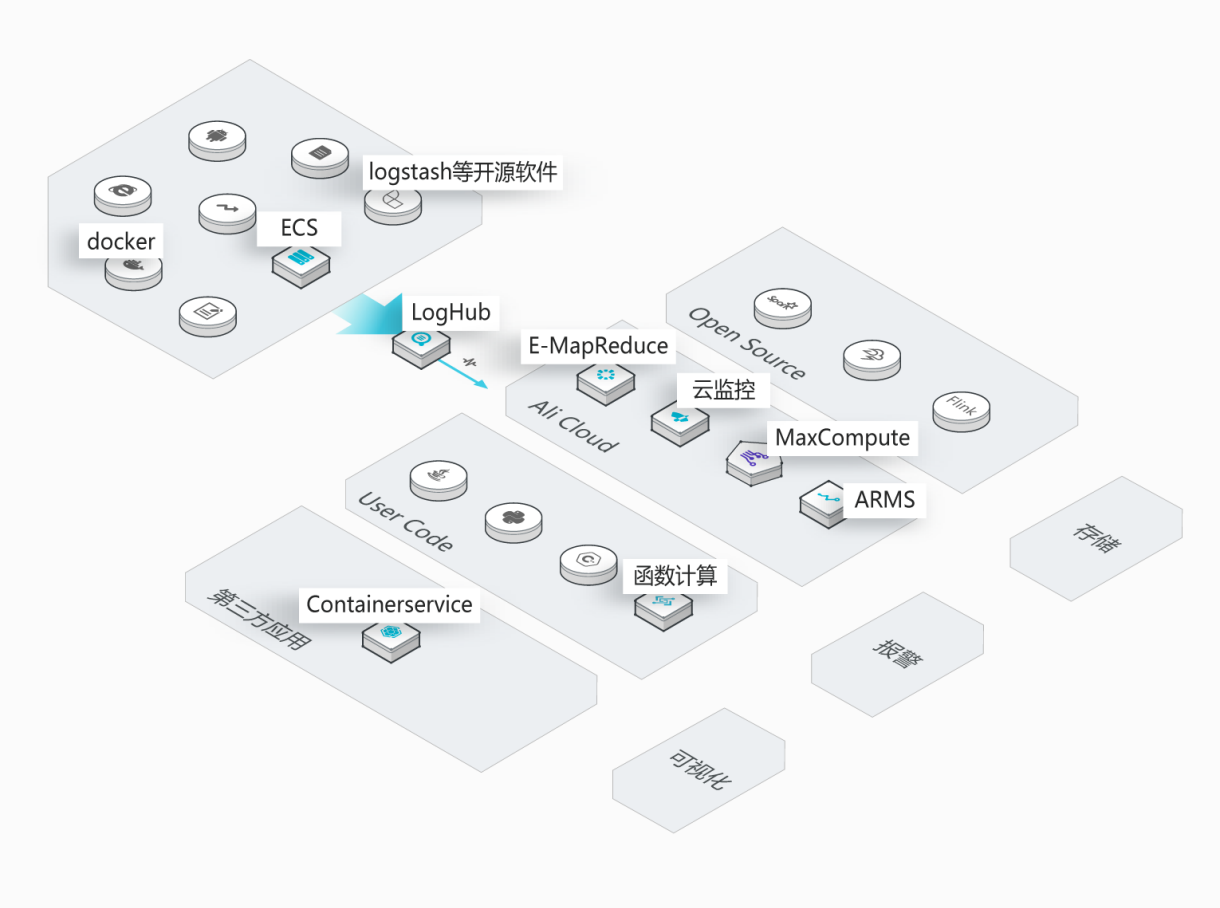
- 使用便捷：提供30+实时数据采集方式，让您快速搭建平台；强大配置管理能力，减轻运维负担；节点遍布全国与全球
- 弹性伸缩：无论是流量高峰还是业务增长都能轻松应对



数据清洗与流计算 (ETL/Stream Processing)

日志中枢（LogHub）支持与各种实时计算及服务对接，并提供完整的进度监控，报警等功能，并可以根据SDK/API实现自定义消费

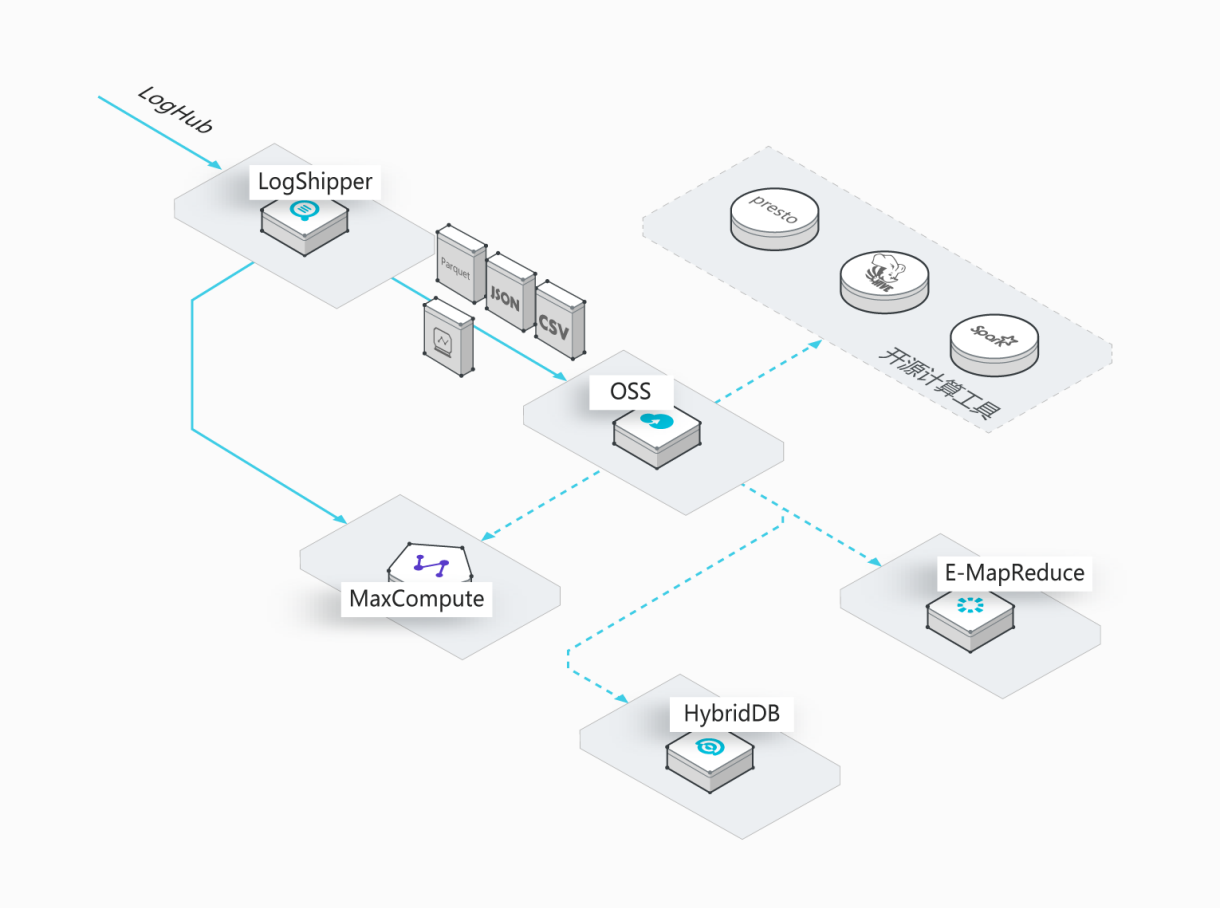
- 操作便捷：提供丰富SDK以及编程框架，与各流计算引擎无缝对接
- 功能完善：提供丰富监控数据，以及延迟报警机制
- 弹性伸缩：PB级弹性能力，0延迟



数据仓库对接(Data Warehouse)

日志投递（LogShipper）功能可以将日志中枢（LogHub）中数据投递至存储类服务，过程支持压缩、自定义Partition、以及行列等各种存储格式

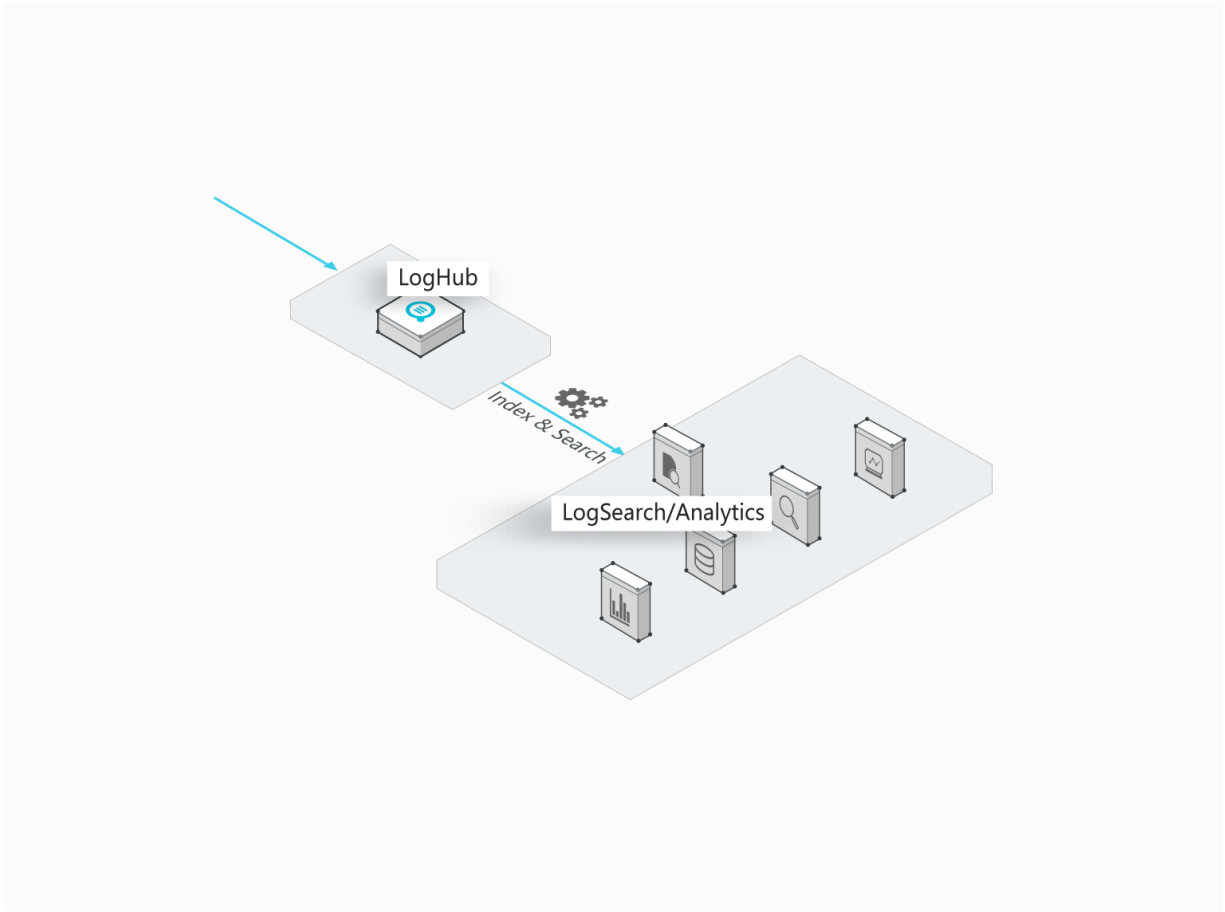
- 海量数据：对数据量不设上限
- 种类丰富：支持行、列、TextFile等各种存储格式
- 配置灵活：支持用户自定义Partition等配置



日志实时查询与分析

实时查询分析（LogAnalytics）可以实时索引LogHub中数据，提供关键词、模糊、上下文、范围、SQL聚合等丰富查询手段

- 实时性强：写入后即可查询
- 海量低成本：支持PB/Day索引能力，成本为自建方案15%
- 分析能力强：支持多种查询手段，及SQL进行聚合分析，并提供可视化及报警功能



19.5 基本概念

日志

日志（Log）是系统在运行过程中变化的一种抽象，其内容为指定对象的某些操作和其操作结果按时间的有序集合。文件日志（LogFile）、事件（Event）、数据库日志（BinLog）、度量（Metric）数据都是日志的不同载体。在文件日志中，每个日志文件由一条或多条日志组成，每条日志描述了一次单独的系统事件，是日志服务中处理的最小数据单元。

日志组

一组日志的集合，写入与读取的基本单位。

日志主题

一个日志库内的日志可以通过日志主题（Topic）来划分。用户可以在写入时指定日志主题，并在查询时指定查询的日志主题。

项目

项目（Project）是日志服务中的资源管理单元，用于资源隔离和控制。您可以通过项目来管理某一个应用的所有日志及相关的日志源。它管理着用户的所有日志库（Logstore），采集日志的机器配置等信息，同时它也是用户访问日志服务资源的入口。

日志库

日志库是日志服务中日志数据的收集、存储和查询单元。每个日志库隶属于一个项目，且每个项目可以创建多个日志库。

分区

每个日志库分若干个分区（Shard），每个分区由MD5左闭右开区间组成，每个区间范围不会相互覆盖，并且所有的区间的范围是MD5整个取值范围。

19.5.1 日志

半世纪前说起日志，想到的是船长、操作员手里厚厚的笔记。如今计算机诞生使得日志产生与消费无处不在：服务器、路由器、传感器、GPS、订单、及各种IoT设备通过不同角度描述着我们生活的世界。借助于计算力量，通过采集、处理、使用日志，我们不断更新对整个世界以及体系的认知。

日志是什么?

从船长日志中我们可以发现，日志除了带一个记录的时间戳外，可以包含几乎任意的内容，例如：一段记录文字、一张图片、天气状况、船行方向等。几个世纪过去了，“船长日志”的方式已经扩展到一笔订单、一项付款记录、一次用户访问、一次数据库操作等多样的领域。

日志这种广泛使用模式之所以经久不衰，在于“日志是一种简单的不能再简单的存储抽象”。它是一个只能增加的，完全按照时间排序的一系列记录。

我们可以给日志末尾添加记录，并且可以从左到右读取日志记录。每一条记录都指定了一个唯一的有一定顺序的日志记录编号。

日志顺序由“时间”来确定，从图上可以看到日志从右到左的时间顺序，新产生的事件被记录，过去的事件渐渐远去，但它记录了什么时间发生了什么事情，这无论对于计算机、人类、还是整个世界而言，是认知与推理的基础。

日志服务中的日志（Log）

日志（Log）是系统在运行过程中变化的一种抽象，其内容为指定对象的某些操作和其操作结果按时间的有序集合。文件日志（LogFile）、事件（Event）、数据库日志（BinLog）、度量（Metric）数据都是日志的不同载体。在文件日志中，每个日志文件由一条或多条日志组成，每条日志描述了一次单独的系统事件，是日志服务中处理的最小数据单元。

日志服务采用半结构数据模式定义一条日志。该模式中包含主题（Topic）、时间（Time）、内容（Content）和来源（Source）四个数据域。

与此同时，日志服务对日志各字段的格式有不同要求，具体如下表所示：

数据域	含义	格式
主题（Topic）	用户自定义字段，用以标记一批日志。例如访问日志可根据不同站点进行标记。	包括空字符串在内的任意字符串，长度不超过128字节。默认情况下，该字段为空字符串。
时间（Time）	日志中的保留字段，用以表示日志产生的时间，一般由日志中的时间信息直接提取生成。	整型，Unix标准时间格式。单位为秒，表示从1970-1-1 00:00:00 UTC计算起的秒数。
内容（Content）	用以记录日志的具体内容。内容部分由一个或多个内容项组成，每一个内容项为一个Key-Value对。	Key为UTF-8编码字符串，包含字母、下划线和数字，且不以数字开头。长度不超过128字节。不可以使用如下关键字：__time__、__source__、__topic__、__和__extract_others__。Value为任意字符串，长度不超过1024*1024字节。
来源（Source）	日志的来源地，例如产生该日志机器的IP地址。	任意字符串，长度不超过128字节。默认情况下该字段为空。

实际使用场景中，日志的格式多样。为了帮助理解，以下以一条nginx原始访问日志如何映射到日志服务日志数据模型为例说明。假设用户nginx服务器的IP地址为10.249.201.117，以下为该服务器的一条原始日志：

```
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"
```

把该条原始日志映射到日志服务日志数据模型，如下：

数据域	内容	说明
Topic	“”	沿用默认值，即空字符串。
Time	1330589527	日志产生的精确时间,表示从1970-1-1 00:00:00 UTC计算起的秒数。从原始日志中的时间戳转换而来。
Content	Key-Value对	日志具体内容。
Source	“10.249.201.117”	使用服务器IP地址作为日志源。

用户可以自己决定如何提取日志原始内容并组合成Key-Value对，例如下表：

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”

19.5.2 日志组

一组日志的集合，写入与读取的基本单位。

日志组的限制为：最大 4096 条日志，或 10MB 空间。

19.5.3 日志主题

一个日志库内的日志可以通过日志主题（Topic）来划分。您可以在写入时指定日志主题，并在查询时指定查询的日志主题。例如，一个平台用户可以使用用户编号作为日志主题写入日志。这样在查询时可利用日志主题让不同用户仅看到自己的日志。如果不需要划分一个日志库内的日志，让所有日志使用相同的日志主题即可。

说明: 空字符串是一个有效的日志主题（Topic），且无论是写入还是查询日志时，默认的日志主题都是空字符串。所以，如果不需要使用日志主题，最简单的方式就是在写入和查询日志时都使用默认日志主题，即空字符串。

19.5.4 项目

项目（Project）是日志服务中的资源管理单元，用于资源隔离和控制。您可以通过项目来管理某一个应用的所有日志及相关的日志源。它管理着用户的所有日志库（Logstore），采集日志的机器配置等信息，同时它也是用户访问日志服务资源的入口。

具体来说，项目可以提供如下功能：

- 帮助您组织、管理不同的Logstore。在实际使用中，您可能需要使用日志服务集中收集、存储不同项目、产品或者环境的日志。您可以把不同项目、产品或者环境的日志分类管理在不同的项目中，方便后续的日志消费、导出或者索引。同时，项目还是日志访问权限管理的载体。
- 提供您日志服务资源的访问入口。每创建一个项目，日志服务会为该项目分配一个独有的访问入口。该访问入口支持通过网络写入、读取及管理日志。

您可以通过日志服务管理控制台进行以下项目操作：

- 创建项目
- 查看项目列表
- 管理项目
- 删除项目

19.5.5 日志库

日志库（Logstore）是日志服务中日志数据的采集、存储和查询单元。每个日志库隶属于一个项目，且每个项目可以创建多个日志库。您可以根据实际需求为某一个项目生成多个日志库，其中常见的做法是为一个应用中的每类日志创建一个独立的日志库。例如，用户有一个“big-game”游戏应用，服务器上有三种日志：操作日志（operation_log）、应用程序日志（application_log）以及访问日志（access_log），用户可以首先创建名为“big-game”的项目，然后在该项目下面为这三种日志创建三个日志库，分别用于它们的采集、存储和查询。

无论是写入或者查询日志，您都需要指定操作的 Logstore。如果您希望投递日志数据到 MaxCompute 做离线分析，其数据投递也是以 Logstore 为单元进行数据同步，即一个 Logstore 内的日志数据投递到一张 MaxCompute 的 Table。

具体来说，日志库提供如下功能：

- 采集日志，支持实时日志写入
- 存储日志，支持实时消费
- 建立索引，支持日志实时查询

- 提供投递到 MaxCompute 的数据通道

您可以通过日志服务管理控制台进行以下日志库操作：

- 创建日志库
- 查看日志库列表
- 修改日志库配置
- 删除日志库

19.5.6 分区

Logstore读写日志必定保存在某一个分区（Shard）上。每个日志库（Logstore）分若干个分区，每个分区由MD5左闭右开区间组成，每个区间范围不会相互覆盖，并且所有的区间的范围是MD5整个取值范围。

分区范围

创建Logstore时，指定分区个数，会自动平均划分整个MD5的范围。每个分区均有范围，可用MD5方式来表示，且必定包含于以下范围中：[00000000000000000000000000000000, ffffffffffffffffffffffffffffffff）。

分区的范围均为左闭右开区间，由以下Key组成：

- BeginKey：分区起始的Key值，分区范围中包含该Key值
- EndKey：分区结束的Key值，分区范围中不包含该Key值

分区的范围用于支持指定Hash Key的模式写入，以及分区的分裂和合并操作。在向分区读写数据过程中，读必须指定对应的分区，而写的过程中可以使用负载均衡模式或者指定Hash Key的模式。负载模式下，每个数据包随机写入某一个当前可用的分区中，在指定Hash Key模式下，数据写入分区范围包含指定Key值的分区。

例如，某Logstore共有4个分区，且该Logstore的MD5取值范围是[00,FF)。各个分区范围如下表所示。

分区号	范围
Shard0	[00,40)
Shard1	[40,80)
Shard2	[80,C0)
Shard3	[C0,FF)

当写入日志时，通过指定Hash Key模式指定一个MD5的Key值是5F，日志数据会写入包含5F的Shard1分区上；如果指定一个MD5的Key值是8C，日志数据会写入包含8C的Shard2分区上。

分区的读写能力

每个分区可提供一定的服务能力：

- 写入：5MB/s，2000次/s
- 读取：10MB/s，100次/s

建议您根据实际数据流量规划分区个数，流量超出读写能力时，及时分裂分区以增加分区个数，从而达到更大的读写能力；如您的流量远远达不到分区的最大读写能力时，建议您合并分区以减少分区个数，从而节约分区租赁费用。

例如，如果您的有两个readwrite状态的分区，最大可以提供10MB/s的数据写入服务，但如果您实时写入数据流量达到14MB，建议分裂其中一个分区，使readwrite分区数量达到3个。如果您实施写入数据流量仅为3MB/s，那么一个分区即可满足需要，建议您合并两个分区。

- 当写入的API持续报告403或者500错误时，通过Logstore云监控查看流量和状态码判断是否需要增加分区。
- 对超过分区服务能力的读写，系统会尽可能服务，但不保证服务质量。

说明:

- 当写入的API持续报告403或者500错误时，通过Logstore云监控查看流量和状态码判断是否需要增加分区。
- 对超过分区服务能力的读写，系统会尽可能服务，但不保证服务质量。

分区的状态

分区的状态包括：

- readwrite：可以读写
- readonly：只读数据

- readwrite：可以读写
- readonly：只读数据

创建分区时，所有分区状态均为readwrite状态，分裂或合并操作会改变分区状态为readonly，并生成新的readwrite分区。分区状态不影响其数据读取的性能，同时，readwrite分区保持正常的数据写入性能，readonly状态分区不提供数据写入服务。

在分裂分区时，需要指定一个处于readwrite状态的ShardId和一个MD5。MD5要求必须大于分区的BeginKey并且小于EndKey。分裂操作可以从一个分区中分裂出另外两个分区，即分裂后分区数量增加2。在分裂完成后，被指定分裂的原分区状态由readwrite变为readonly，数据仍然可以被消费，但不可写入新数据。两个新生成的分区状态为readwrite，排列在原有分区之后，且两个分区的MD5范围覆盖了原来分区的范围。

在合并操作时，必须指定一个处于readwrite状态的分区，指定的分区不能是最后一个readwrite分区。服务端会自动找到所指定分区的右侧相邻分区，并将两个分区范围合并。在合并完成后，所指定的分区和其右侧相邻分区变成只读（readonly）状态，数据仍然可以被消费，但不能写入新数据。同时新生成一个 readwrite 状态的分区，新分区的MD5范围覆盖了原来两个分区的范围。

通过日志服务管理控制台您可以进行以下分区操作：

- 扩容
- 缩容
- 删除

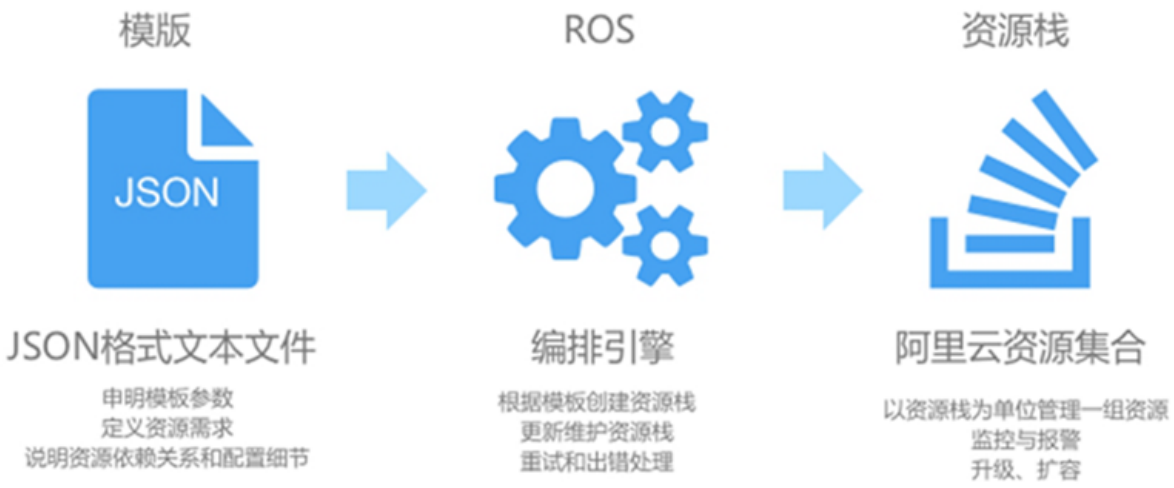
20 资源编排

20.1 产品概述

阿里云资源编排服务（ROS）是一款帮助阿里云用户简化云计算资源管理和自动化运维的服务。您遵循ROS定义的模板规范，编写模板文件，在模板中定义所需云计算资源的集合及资源间的依赖关系、资源配置细节等，ROS通过编排引擎自动完成所有资源的创建和配置，以达到自动化部署、运维的目的。编排模板是一种用户可读、易于编写的文本文件，用户可以通过SVN、Git等版本控制工具来控制模板的版本，以达到控制基础设施版本的目的，您可以通过API、SDK等方式把ROS的编排能力与自己的应用整合，做到基础设施即代码（Infrastructure is Code）。

编排模板同时也是一种标准化的资源和应用交付方式，您可以通过编排模板交付包含云资源和应用的整体系统和解决方案。ISV可以通过这种交付能力，轻松的整合阿里云的资源和ISV的软件系统，达到统一交付的目的。

编排服务是通过资源栈（Stack）这样的逻辑集合来统一管理一组云资源，所以，对于云资源的创建、删除、克隆等操作都可以以资源组为单位来完成。在DevOps实践中，可以很轻松的克隆开发、测试、线上环境。同时，也可以更容易做到应用的整体迁移和扩容。



20.2 使用限制

资源编排对用户有以下限制：

- 每个堆栈允许创建的最大资源数为200个。
- 每个用户允许创建的堆栈数最大为50个。

- 每个模板文件的大小不超过512kb。

21 API网关

21.1 产品概述

API网关为您提供完整的API托管服务，辅助您将能力、服务、数据以API的形式开放给合作伙伴，也可以发布到API市场供更多的开发者采购使用。

- 提供防攻击、防重放、请求加密、身份认证、权限管理、流量控制等多重手段保证API安全，降低API开放风险。
- 提供API定义、测试、发布、下线等全生命周期管理，并生成SDK、API说明文档，提升API管理、迭代的效率。
- 提供便捷的监控、报警、分析、API市场等运维、运营工具，降低API运营、维护成本。

API网关将能力的复用率最大化，企业间能够互相借力，企业发展能够专注自身业务，实现共赢。

图 60: API网关

21.2 功能特性

API生命周期管理

- 支持包括API发布、API测试、API下线等生命周期管理功能。
- 支持API日常管理、API版本管理、API快速回滚等维护功能。

全面的安全防护

- 支持多种认证方式，支持HMAC（SHA-1、SHA-256）算法签名。
- 支持HTTPS协议，支持SSL加密。
- 防攻击、防注入、请求防重放、请求防篡改。

灵活的权限控制

- 您以APP作为请求API的身份，网关支持针对APP的权限控制。
- 只有已经获得授权的APP才能请求相应的API。
- API提供者可以主动授权某个APP调用某个API的权限。
- API若上架到API市场，则创建者可以将已创建的API授权给自己的APP。

精准的流量控制

- 流量控制可以用于管控API的被访问频率、APP的请求频率、用户的请求频率。
- 流量控制的时间单位可以是分钟、小时、天。
- 同时支持流控例外，允许设置特殊的APP或者用户。

请求校验

支持参数类型、参数值（范围、枚举、正则、Json Schema）校验，无效校验直接会被API网关拒绝，减少无效请求对后端造成的资源浪费，大大降低后端服务的处理成本。

数据转换

通过配置映射规则，实现前、后端数据翻译。

- 支持前端请求的数据转换。
- 支持返回结果的数据转换。

监控报警

- 提供可视化的API实时监控，包括：调用量、流量大小、响应时间、错误率，在陆续增加维度。
- 支持历史情况查询，以便统筹分析。
- 可配置预警方式（短信、Email），订阅预警信息，以便实时掌握API运行情况。

自动工具

- 自动生成API文档，可供在线查看。
- API网关提供多种语言SDK的示例。降低API的运维成本。
- 提供可视化的界面调试工具，快速测试，快速上线。

API市场

- 可将API上架到API市场，供更多开发者采购和使用。

21.3 产品优势

解放生产力

完成API录入后，即可告别API管理的一切繁杂，API网关为您解决API文档维护、SDK维护、API版本管理等繁琐事务，大大降低您的日常维护成本。

只为实际服务付费

随时开通，API日常管理、生成文档、生成SDK、流量控制、权限控制。

大规模且高性能

API网关采用分布式部署，自动扩展，能够承载大规模的API访问；同时还能保证较低的延时，为您的后端服务提供高保障高效率的网关功能。

安全稳定

您的服务只需在内网对API网关开放，不必顾虑安全问题。API网关还提供严格的权限管理功能、精准的流量控制功能、全面的监控报警功能，让您的服务安全、稳定、可控。

21.4 基本概念

使用API网关，您需要对以下基本概念有所了解。

应用

APP

您需要创建APP作为调用API时的身份。

AppKey、AppSecret

每个APP都有这样一对密钥对，加密计算后放入请求中作为签名信息。

加密签名

API请求中携带签名信息，用于网关对请求做身份验证。

授权

授予某个APP调用某个API的权限，由API服务方完成。 APP被授权后才能调用API。

API生命周期

API服务方分阶段的管理API ，包括API的创建、测试、发布、下线、版本切换等。

API定义

API服务方创建API时，设置的API的后端服务、请求格式、接收格式、返回格式等规则内容。

参数映射

您实际请求的参数与API服务后端参数不一致时，支持API服务方配置参数映射。

参数校验

API服务方对入参设置校验规则，由网关根据规则对无效请求进行过滤。

常量参数

不需要用户传入的，但是后端服务需要始终接收的常量参数。

系统参数

您可以设置网关向您后端抛请求时，附带一些系统参数，如CaClientIp，即请求IP等。

API分组

API服务方管理API的单元。创建API需要先创建分组。

二级域名

创建分组时，系统给分组绑定的域名，用于测试API调用。

独立域名

开放API服务，需要为分组绑定独立域名。您通过访问该独立域名调用API。

签名密钥

API服务方可以创建签名密钥并绑定到API上，网关到服务方后端的请求就会带上签名信息，用于后端的安全验证。

流量控制策略

用于API服务方对API 、用户、 APP按天、小时、分钟进行流量限制。

22 云盾（基础版）

22.1 产品概述

阿里云提供的云产品及服务来自于阿里巴巴集团在电子商务和互联网金融行业多年的技术沉淀，云盾则是在阿里云上为客户的业务系统安全护航的首要和关键组件。

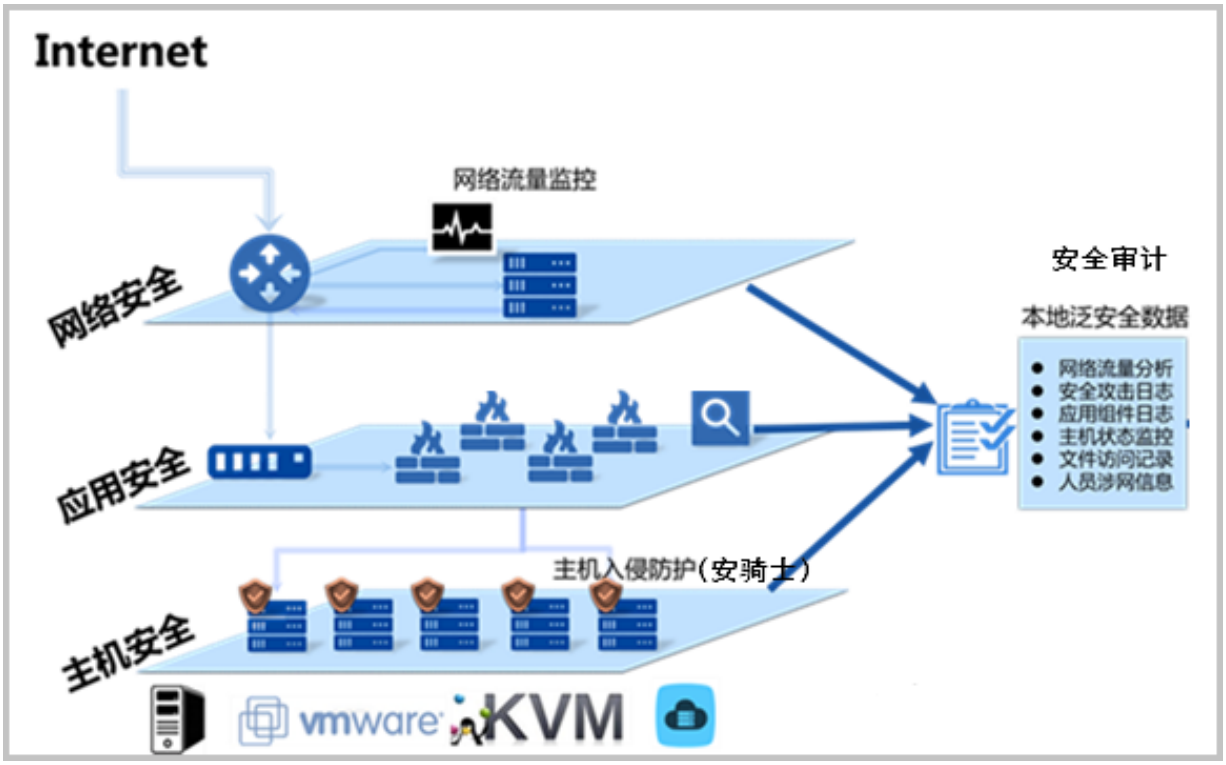
专有云云盾是阿里巴巴集团多年来安全技术研究积累的成果，结合阿里云云计算平台强大的数据分析能力以及阿里云专业的安全运营团队，为云数据中心提供多层面一体化的云安全防护解决方案。

专有云V3云盾分为云盾基础版和高级版。云盾基础版为专有云V3内置，不需要额外购买。

22.2 产品架构

云盾基础版在专有云中的结构如下图所示。

图 61: 云盾基础版部署逻辑拓扑结构



22.3 功能特性

专有云云盾默认提供云平台安全和云产品安全功能，同时面向云租户提供云盾基础版和高级版安全功能。

云盾基础版由网络流量监控系统、主机入侵防御系统、安全审计、和集中管控系统四大功能模块组成。

表 35: 云盾基础版主要功能

模块	功能
网络流量监控	流量采集，简单攻击检测。
主机入侵防御（安骑士）	防暴力破解、网站后门查杀、异地登录告警。
安全审计	云运维审计。

- 在专有云的网络边界，通过流量镜像的方式将出入专有云的所有网络流量输入到云盾网络流量监控模块进行逐包检测分析，分析结果将作为云盾其他防护模块的参考依据。
- 主机入侵防御模块部署在云服务器上，实现对云服务器上的Web木马进行查杀、对密码暴力破解进行拦截、对异常的登录行为进行告警。
- 集中管控系统部署在云服务器集群中，负责对云盾所有安全模块的集中策略管理以及进行统一的日志分析。

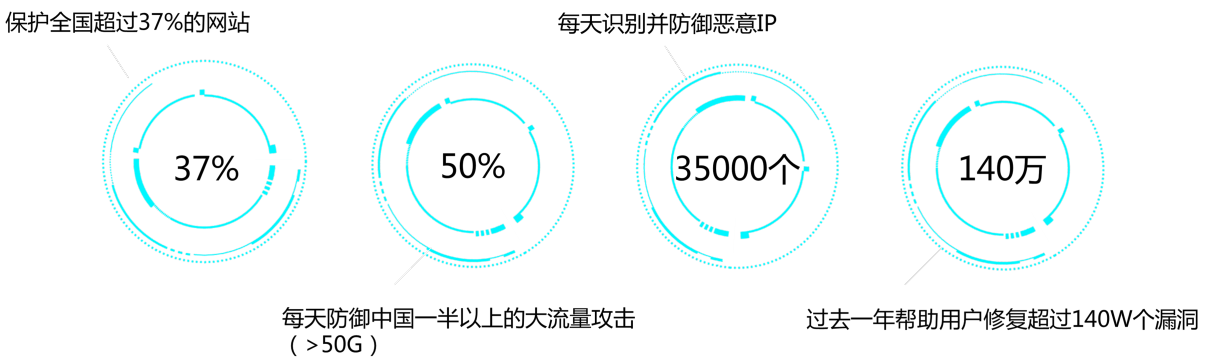
22.4 产品优势

云上安全先行者

阿里安全团队从2005年起护航阿里巴巴集团内部所有业务系统的信息安全，不断积累安全经验。自2011年首次推出云盾产品，全方位保证阿里云的安全体系，成为云上安全先行者。

阿里云保护全国超过37%的网站，每天防御中国一半以上的大流量攻击，每天识别并防御恶意IP 35,000个，过去一年帮助用户修复超过140万个漏洞。

图 62: 阿里云处理海量互联网数据



权威认证，安全可靠

依靠阿里云平台自身的安全特性以及云盾为云上客户提供的攻击防御特性，阿里云先后取得了国内外多项云安全认证。

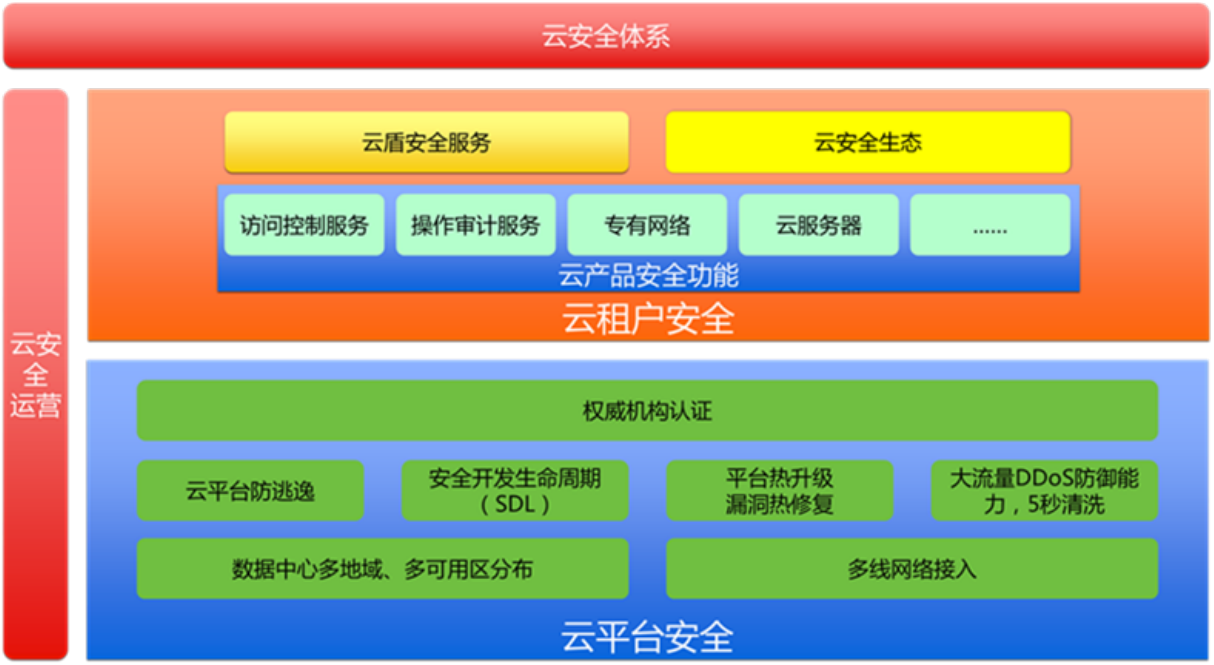
图 63: 阿里云获得的安全认证

- 全球首家获得云安全国际认证金牌（CSA STAR Certification）的云服务供应商
- 全国首家获得ISO27001信息安全管理体系国际认证的云安全服务供应商
- 全国首个通过公安部等级保护测评（DJCP）的云计算系统
- 阿里云电子政务云平台首批通过党政部门云服务网络安全审查（增强级）
- 全国首家云等保试点示范平台
- 金融云高分通过等保四级测评，成为全国首个四级云平台

体系完整，技术领先

十年攻防，一朝成盾。在经历了为阿里巴巴集团自身业务十年来的安全护航后，阿里巴巴积累了大量的安全研究成果、安全数据、安全运营和安全管理方法，形成了一支专业的云安全专家团队。 云盾是集合这些安全专家多年攻防经验开发出来的面向云计算平台安全最佳实现的成熟体系，可有效的保护专有云上客户的云平台、云网络环境和云业务系统的安全。

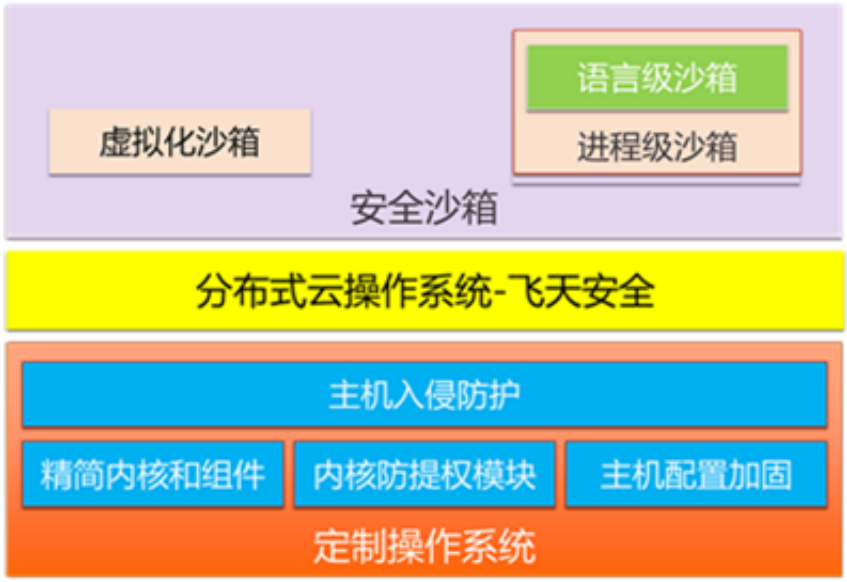
图 64: 阿里云云安全体系



22.4.1 云平台安全

22.4.1.1 纵深防御

图 65: 纵深防御结构



物理网络安全

- 管理平面和业务平面网络隔离。
- 关闭未使用的网络端口防止非法接入。

- 回收服务器默认路由防止主动外联。

宿主机安全

- 操作系统内核和组件都经过精简。
- 符合业界安全规范的配置加固。
- 内核防提权模块。
- 主机入侵防护软件。
- 租户程序运行在安全沙箱中。

飞天安全

22.4.1.2 多租户隔离

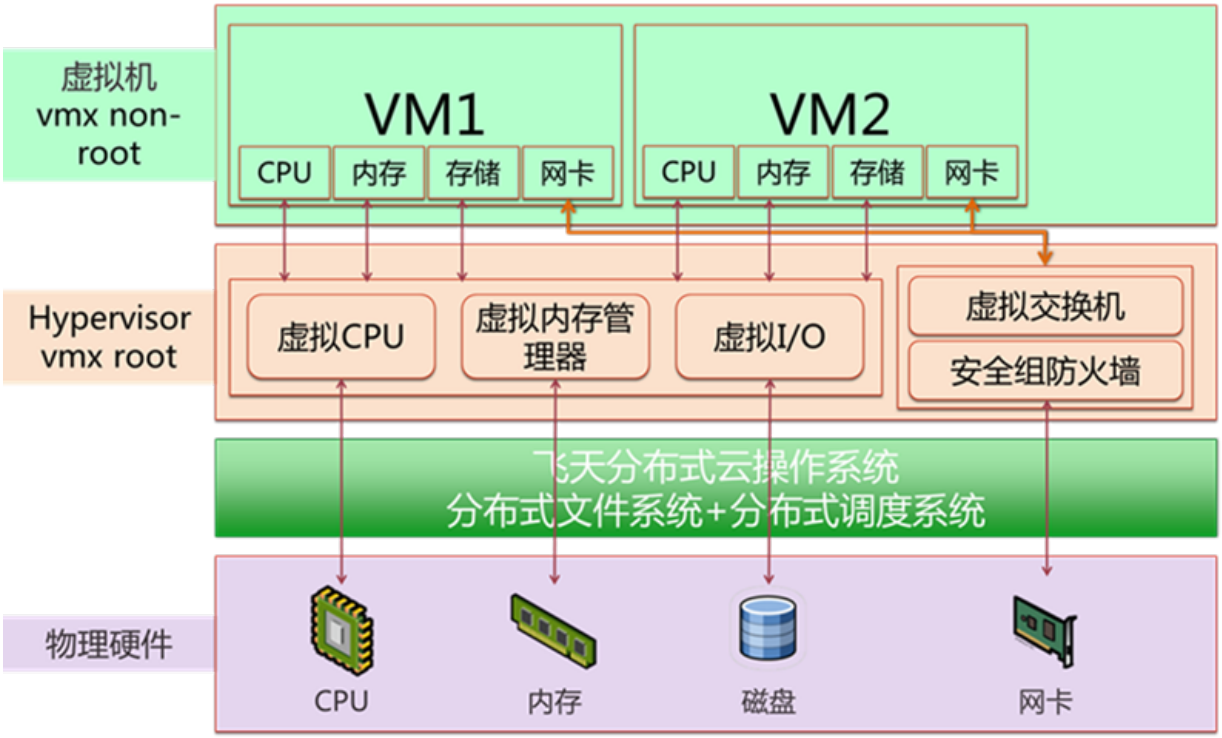
云服务器租户隔离

- 应用硬件协助的虚拟化技术模式（Xen Full Virtualization，简称HVM模式），基于VT-x技术隔离CPU。
- 硬件辅助扩展页表技术（Extended Page Table，简称EPT）EPT技术隔离内存。
- 分离设备驱动I/O模型隔离存储。
- 通过交换型VSwitch，不同虚拟机的数据包被转发到对应的虚拟端口。
- 虚拟机的IP和MAC地址绑定，防地址欺骗及网络嗅探。
- VPC、安全组、防火墙隔离租户网络。
- 物理内存、物理存储重分配前清零。

其他云产品租户隔离

- 用户数据打标签隔离存储。
- 基于身份验证进行访问控制。

图 66: 多租户隔离架构



22.4.1.3 数据安全

数据安全承诺

2015年7月，阿里云发起中国云计算服务商首个“数据保护倡议”：运行在云计算平台上的开发者、公司、政府、社会机构的数据，所有权绝对属于客户；云计算平台不得将这些数据移作它用。平台方有责任和义务，帮助客户保障其数据的私密性、完整性和可用性。

多副本分布式存储

专有云使用分布式存储，文件被分割成许多数据片段分散存储在不同的设备上，并且每个数据片段存储多个副本。分布式存储不但提高了数据的可靠性，也提高了数据的安全性。

加密存储

专有云加密服务使用经国家密码管理局检测认证的硬件密码机，帮助您满足数据安全方面的监管合规要求，保护云上业务数据的机密性。借助加密服务，您可以实现对加密密钥的完全控制并进行解密操作。

加密传输

专有云平台提供标准的加密传输协议，便于云平台与外界以及系统间传输敏感数据的需求。云平台支持标准的TLS协议，可提供高达256位密钥的加密强度，完全满足敏感加密传输需求。

残留信息

对于曾经存储过客户数据的内存和磁盘，一旦释放和回收，其上的残留信息将被自动清零。同时，任何更换和淘汰的存储设备，都将统一执行消磁处理并物理折弯之后，才能运出数据中心。

数据审计

通过平台级的访问审计，以及产品级的SQL审计、上传下载审计，确保数据的生成、变更、删除、传播有迹可循，使违规的数据操作无所遁形。

22.4.1.4 开发发全

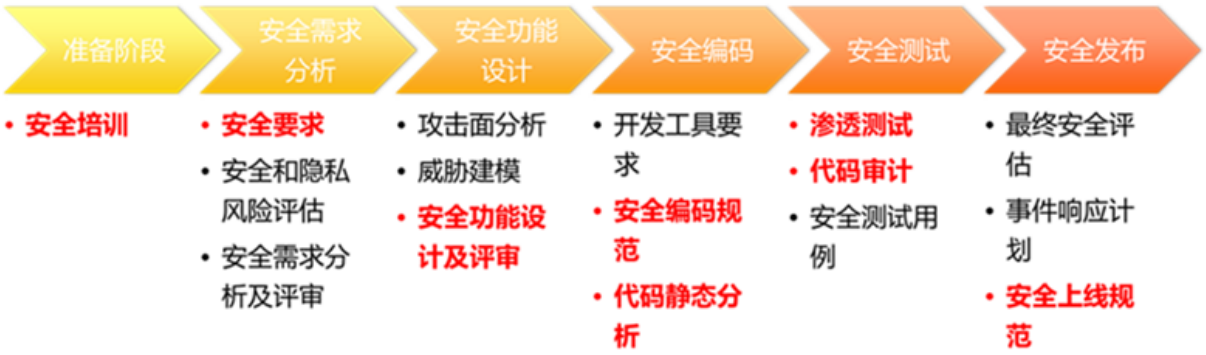
源代码是组成云计算平台的最小单元，大量信息安全问题根源是由开发设计缺陷引起的。

专有云平台开发遵循软件安全开发生命周期（Security Development Lifecycle，简称SDL），解决绝大部分因源代码安全缺陷而引发的安全问题。SDL覆盖飞天操作系统、云产品、大数据产品、OpenAPI等，从产品开发初期就基于云端安全威胁构建云服务的安全功能或属性，极大提高了云平台安全的健壮性。

软件安全开发生命周期包含：

- 需求分析：识别云服务安全需求和风控需求。
- 产品设计：攻击面分析、威胁建模、安全架构/功能设计。
- 编码阶段：采用安全开发框架，遵循安全编码规范。
- 测试阶段：渗透测试结合代码审计。未经安全测试的产品禁止上线。
- 发布阶段：按照安全规范实施整体加固。

图 67: 安全开发生命周期管理



22.4.1.5 漏洞热修复

漏洞热修复是指在不影响系统正常运行的情况下进行系统漏洞修复，防止由于冷补丁修复导致的业务中断、业务崩溃和宕机等问题。

漏洞热修复基于阿里云六年的云安全运营经验，支持以下系统和平台：

- Linux内核
- 飞天分布式云操作系统及各云产品
- ECS Xen Hypervisor
- RDS MySQL

案例

2015年3月，Xen曝出安全漏洞：XSA-123，该漏洞造成客户机指令提权，任意一台云主机可读取到其他另一台云主机的重要数据，从而导致客户数据泄密。Rackspace 服务器大规模重启，影响客户业务超过10分钟。 亚马逊99.9%的服务器热修复，近0.1%的服务器重启。阿里云100%的服务器热修复，对客户业务无任何影响。阿里云在漏洞曝出前就拿到了相关漏洞细节，并提前发布了热补丁修复该漏洞，并在没有影响任何用户主机的情况下完成了整个系统的修复。

22.4.2 云产品安全

22.4.2.1 云服务器安全

专有云云服务器提供以下安全功能：

镜像安全

- 定期修复高危漏洞。
- 内置主机入侵防护软件。

热升级

- 宿主机Linux内核热升级。
- Hypervisor热升级。

租户隔离

- Hypervisor隔离不同虚拟机的CPU、内存、存储。
- 通过专有网络VPC和安全组隔离不同租户网络。
- 内存、存储释放后数据清零。

可靠性

- 分布式冗余存储保障数据可靠性。
- 基于磁盘快照的快捷备份和回滚。
- 基于故障迁移的即时恢复。
- 基于在线迁移的智能资源调度。
- 可用性高达99.95%。

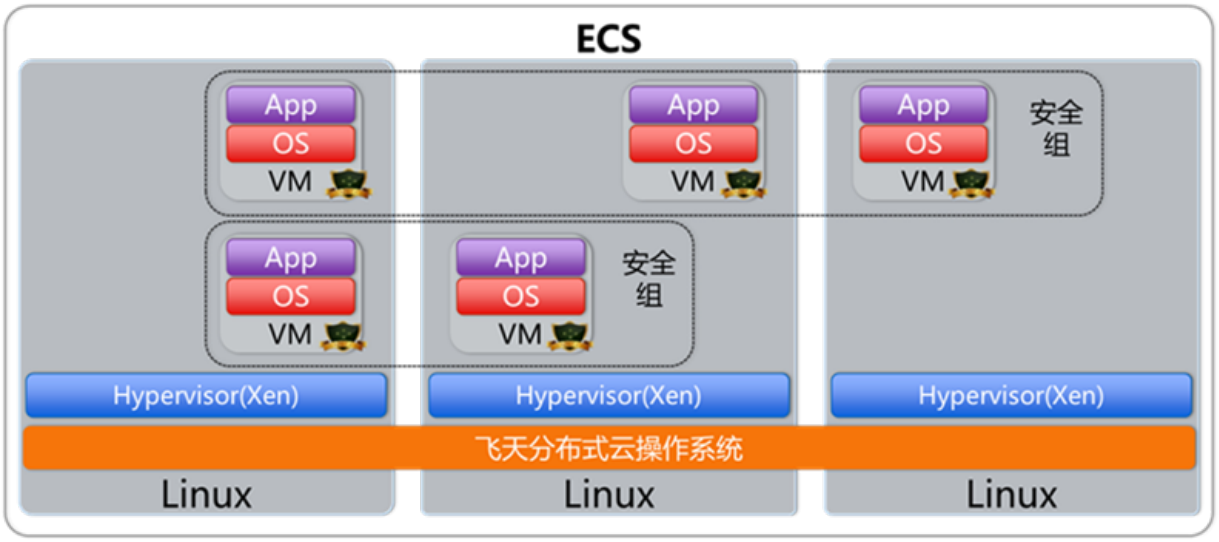
网络安全

- 专有网络VPC（基于VxLAN）隔离。
- 状态检测虚拟防火墙，划分安全域。
- 防IP、MAC伪造和地址解析协议（Address Resolution Protocol，简称ARP）欺骗。
- 防网络嗅探。

主机安全

- 租户拥有最高权限，阿里云没有登录权限。
- Linux支持SSH Key认证，防暴力破解。

图 68: 云服务器安全架构



22.4.2.2 云数据库安全

专有云提供以下云数据库安全功能：

租户隔离

- 数据库实例隔离。

可靠性

- RAID5磁盘阵列存储，保障数据可靠性。
- 每个RDS实例拥有两个物理节点进行主从热备份。
- 支持秒级主备切换。
- 数据库定时备份，能够根据备份文件将数据库恢复至七日内任意时间点。
- 可用性高达99.95% 。

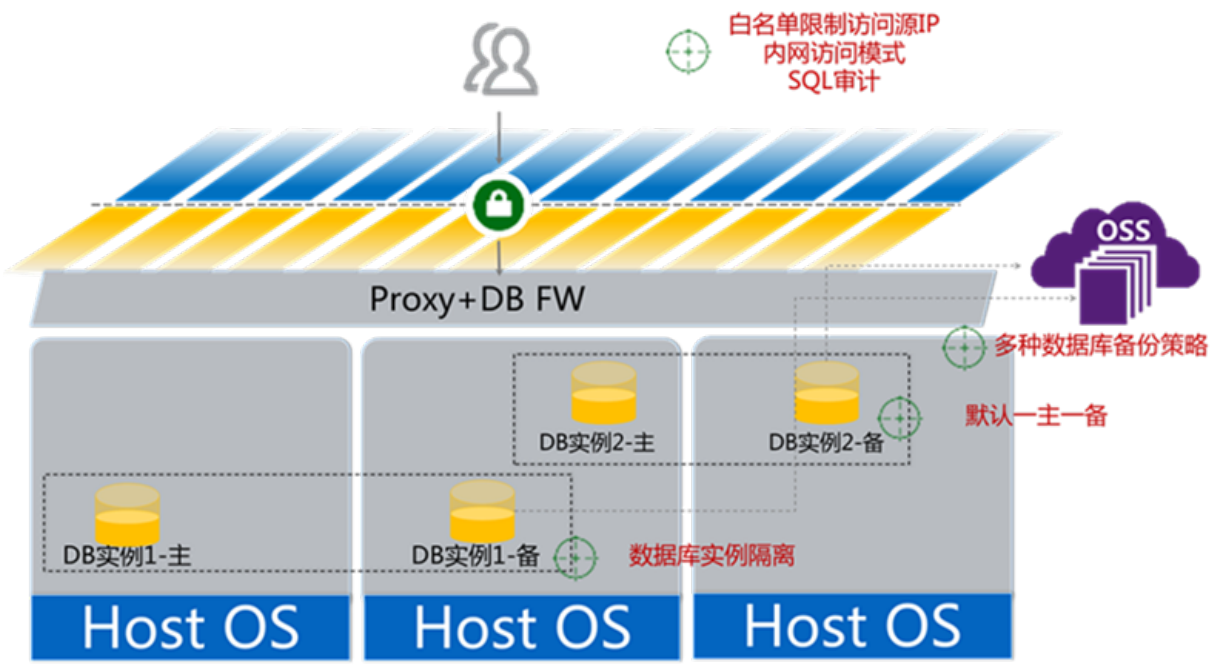
网络安全

- 通过IP白名单限定允许访问RDS服务的源IP。

热升级

- RDS For MySQL实例热升级，客户业务无感知。

图 69: 云数据库安全



22.4.2.3 云存储安全

专有云提供以下云存储安全功能:

租户隔离

- 租户数据打标签区分。
- 服务接入层对称密钥认证技术鉴别用户。

可靠性

- 分布式冗余存储保障数据可靠性。
- 可用性高达99.9%。

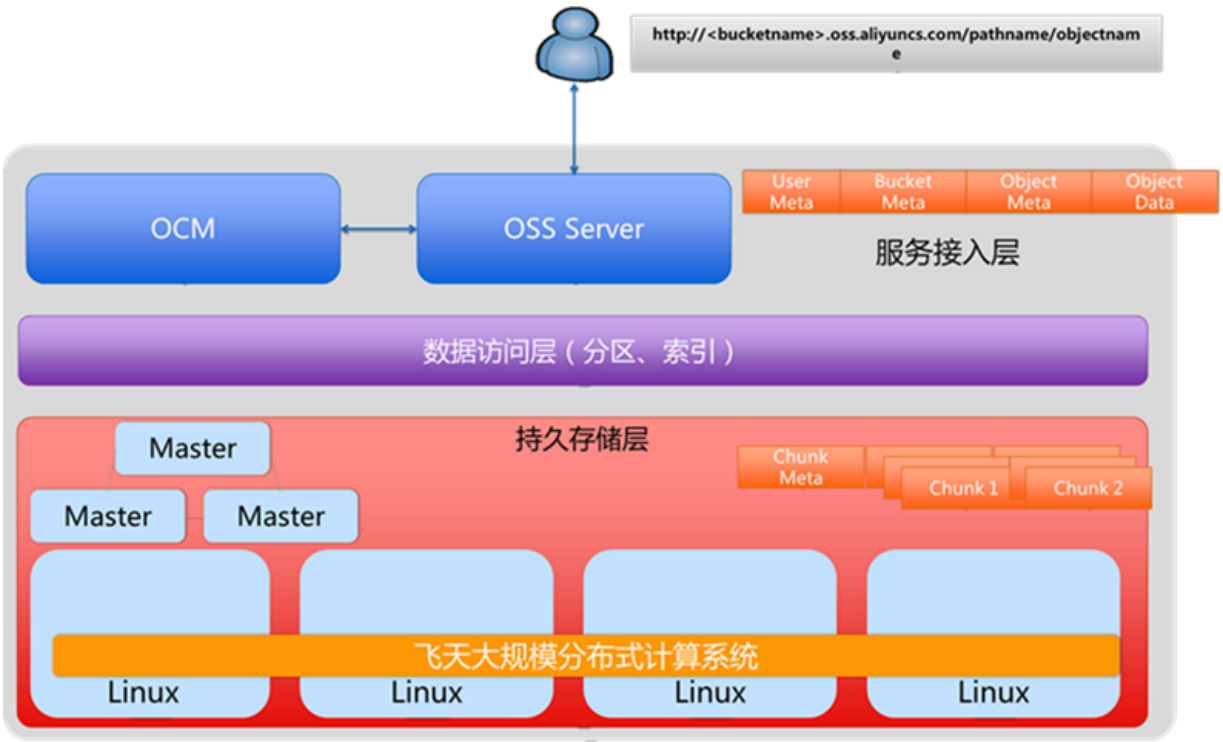
访问控制

- 通过访问控制列表（ACL）进行访问控制。
- 基于访问控制管理(Resource Access Management,简称 RAM)授权策略的访问控制。

加密传输

- 支持SSL传输加密。
- 支持服务器端加密存储。

图 70: 云存储安全架构



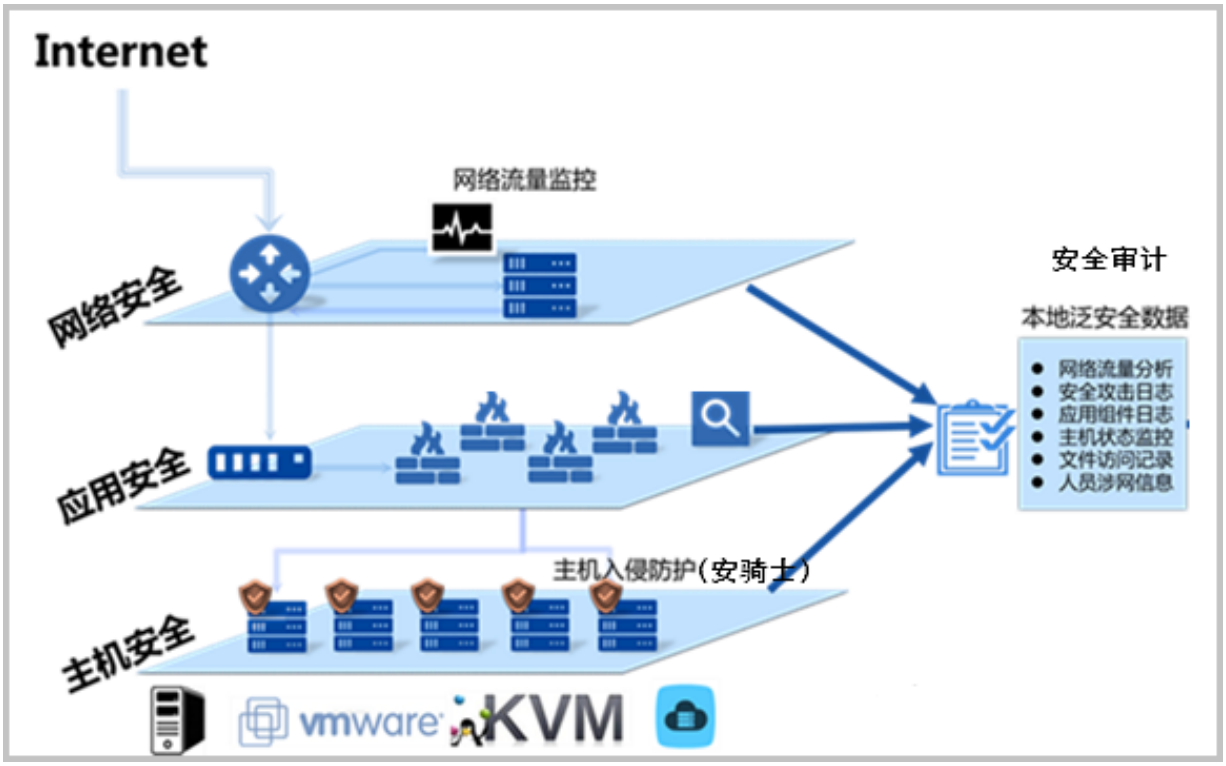
22.5 典型应用

云盾不同于传统的软硬件安全产品，它采用纵深防御、多点联动的云安全架构，完全基于阿里云的云计算环境研发，从网络层、应用层、主机层等多个层面为您提供全面的、一体化的云安全防护能力。

单数据中心部署

专有云云盾基础版包含网络流量监控模块、主机入侵防御模块、安全审计模块，从网络层、应用层、主机层多个层面为您的单数据中心提供全面的、一体化的云安全防护能力。

图 71: 云盾基础版单数据中心部署结构



23 云盾（高级版）

23.1 产品概述

阿里云提供的云产品及服务来自于阿里巴巴集团在电子商务和互联网金融行业多年的技术沉淀，云盾则是在阿里云上为客户的业务系统安全护航的首要和关键组件。

专有云云盾是阿里巴巴集团多年来安全技术研究积累的成果，结合阿里云云计算平台强大的数据分析能力以及阿里云专业的安全运营团队，为云数据中心提供多层面一体化的云安全防护解决方案。

专有云V3云盾分为云盾基础版和高级版，高级版包含了基础版的功能。

云服务商保障云平台自身安全：

- 确保云平台底层安全，防止逃逸。
- 云产品开发遵循安全开发生命周期（SDL）流程，上线前必须经过安全测试。
- 云数据中心多地域、多可用区分布。
- 云数据中心支持多线BGP接入。
- 具备平台热升级、漏洞热修复能力。
- 具备大流量DDoS防御能力，流量清洗应在10秒内完成。
- 云平台通过国际、国家、行业的安全合规认证。

云服务商为云租户提供保护云端系统及数据的技术手段：

- 云账号支持主子账号、双因素认证、分组授权、细粒度授权、临时授权。
- 为租户提供安全审计手段。
- 为租户提供数据加密手段。
- 为租户提供云安全SaaS服务。
- 引入第三方安全厂商，为租户提供个性化的行业安全解决方案。

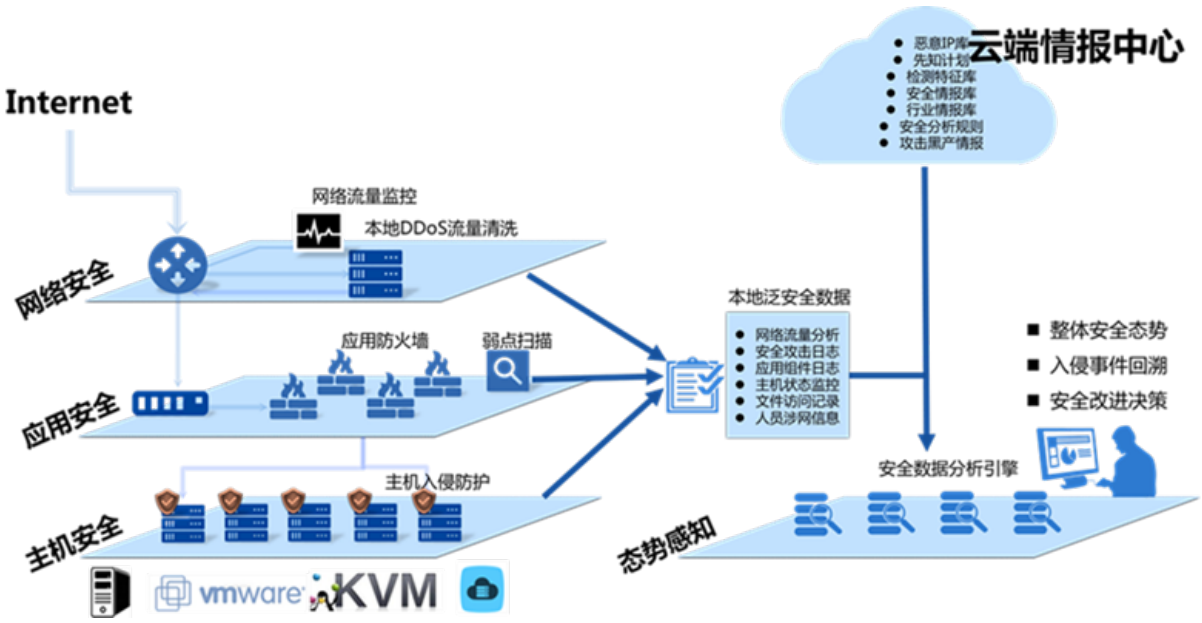
云租户保障自身云产品安全：

- 安全管理责任不变，数据归属关系不变，安全管理标准不变。
- 建立安全管理体系，七分管理，三分技术。
- 合理使用云服务商提供的技术手段，包括云产品安全功能、云安全服务、云安全生态中第三方安全厂商的产品。

23.2 产品架构

云盾高级版在专有云V3中的结构如下图所示。

图 72: 云盾部署逻辑拓扑结构



- 在专有云的网络边界，通过流量镜像的方式将出入专有云的所有网络流量输入到云盾网络流量监控模块进行逐包检测分析，分析结果将作为云盾其他防护模块的参考依据。
- Web应用防火墙（WAF）模块可以对常见的Web攻击、CC攻击进行阻断。
- 主机入侵防御（安骑士）模块部署在云服务器上，实现对云服务器上的Web木马进行查杀、对密码暴力破解进行拦截、对异常的登录行为进行告警，同时还能对高危漏洞进行修复。
- 态势感知系统通过汇集网络流量、主机端信息，结合从云端下发的威胁情报和大数据分析模型，在本地部署的大数据集群中进行威胁态势分析。
- 集中管控系统部署在云服务器集群中，负责对云盾所有安全模块的集中策略管理以及进行统一的日志分析。

23.3 功能特性

专有云V3云盾默认提供云平台安全和云产品安全功能，面向租户安全的功能分为云盾基础版和高级版。

云盾基础版由网络流量监控系统、主机入侵防御系统、安全审计三大功能模块组成；云盾高级版在基础版的基础上增加了DDoS清洗、云防火墙、WAF和态势感知等功能。结合阿里云专业的安全运营服务为云用户提供了入侵防御、安全审计、态势感知和集中管控等一站式安全保障。

图 73: 云盾高级版主要功能

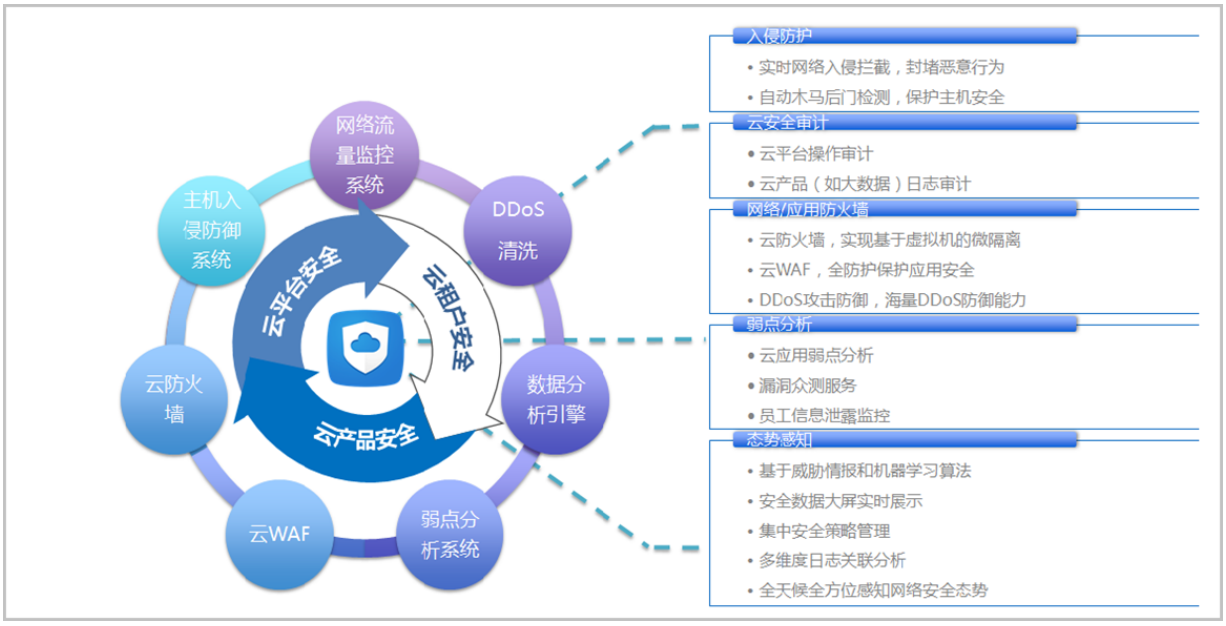


表 36: 云盾详细功能列表

产品	模块	功能
态势感知	网络流量监控	流量采集，简单攻击检测。
		基于模型的攻击检测。
	安全审计	云日志审计。
	安骑士	防暴力破解、网站后门查杀、异地登录告警。
		补丁管理。
		安全运维。
		基线检查。
	弱点扫描	弱点管理、安全监测预警、安全大屏。
云防火墙	云防火墙	云主机访问控制。
DDoS清洗	DDoS清洗	DDoS攻击防御。
Web应用防火墙	Web应用防火墙	应用入侵防护。

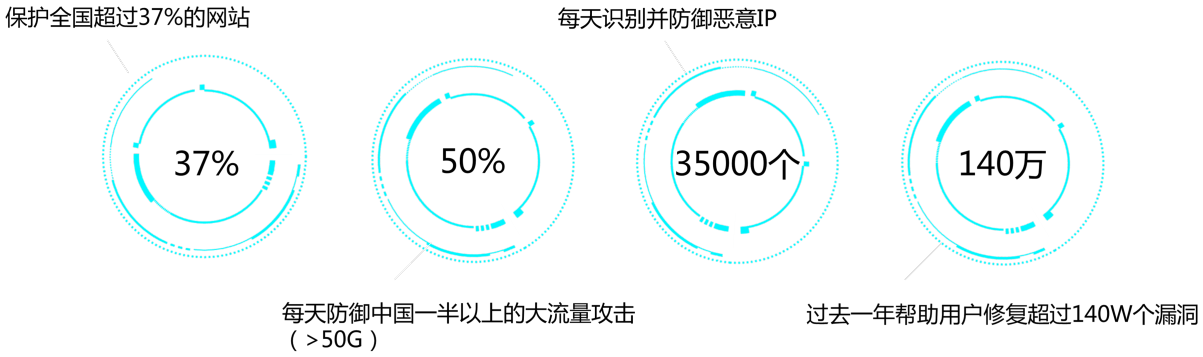
23.4 产品优势

云上安全先行者

阿里安全团队从2005年起护航阿里巴巴集团内部所有业务系统的信息安全，不断积累安全经验。自2011年首次推出云盾产品，全方位保证阿里云的安全体系，成为云上安全先行者。

阿里云保护全国超过37%的网站，每天防御中国一半以上的大流量攻击，每天识别并防御恶意IP 35,000个，过去一年帮助用户修复超过140万个漏洞。

图 74: 阿里云处理海量互联网数据



权威认证，安全可靠

依靠阿里云平台自身的安全特性以及云盾为云上客户提供的攻击防御特性，阿里云先后取得了国内外多项云安全认证。

图 75: 阿里云获得的安全认证

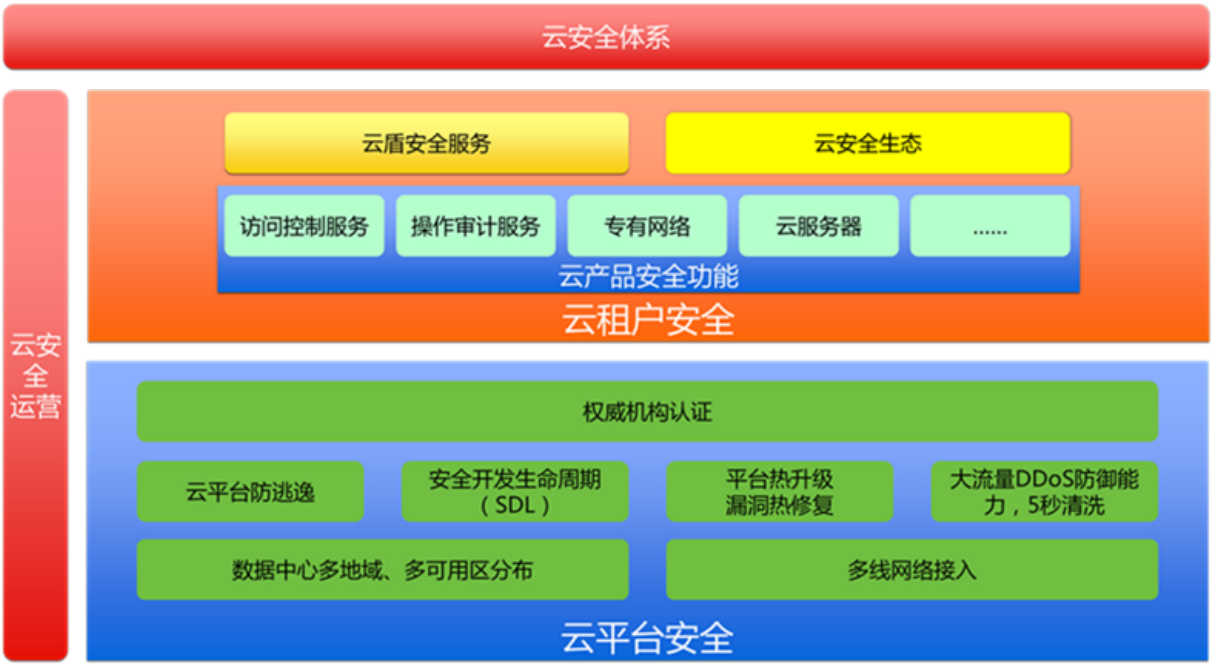
- 全球首家获得云安全国际认证金牌（CSA STAR Certification）的云服务供应商
- 全国首家获得ISO27001信息安全管理体系国际认证的云安全服务供应商

- 全国首个通过公安部等级保护测评（DJCP）的云计算系统
- 阿里云电子政务云平台首批通过党政部门云服务网络安全审查（增强级）
- 全国首家云等保试点示范平台
- 金融云高分通过等保四级测评，成为全国首个四级云平台

体系完整，技术领先

十年攻防，一朝成盾。在经历了为阿里巴巴集团自身业务十年来的安全护航后，阿里巴巴积累了大量的安全研究成果、安全数据、安全运营和安全管理方法，形成了一支专业的云安全专家团队。 云盾是集合这些安全专家多年攻防经验开发出来的面向云计算平台安全最佳实现的成熟体系，可有效的保护专有云上客户的云平台、云网络环境和云业务系统的安全。

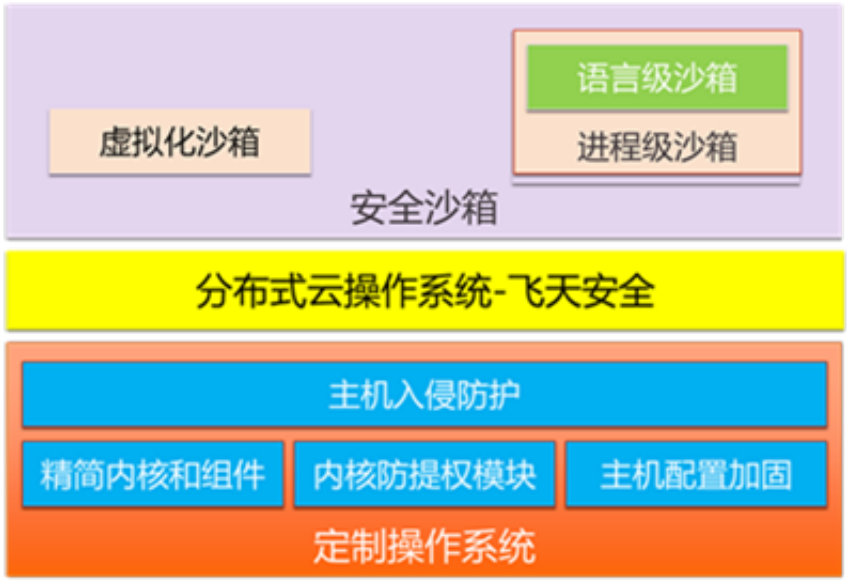
图 76: 阿里云云安全体系



23.4.1 云平台安全

23.4.1.1 纵深防御

图 77: 纵深防御结构



物理网络安全

- 管理平面和业务平面网络隔离。
- 关闭未使用的网络端口防止非法接入。
- 回收服务器默认路由防止主动外联。

宿主机安全

- 操作系统内核和组件都经过精简。
- 符合业界安全规范的配置加固。
- 内核防提权模块。
- 主机入侵防护软件。
- 租户程序运行在安全沙箱中。

飞天安全

23.4.1.2 多租户隔离

云服务器租户隔离

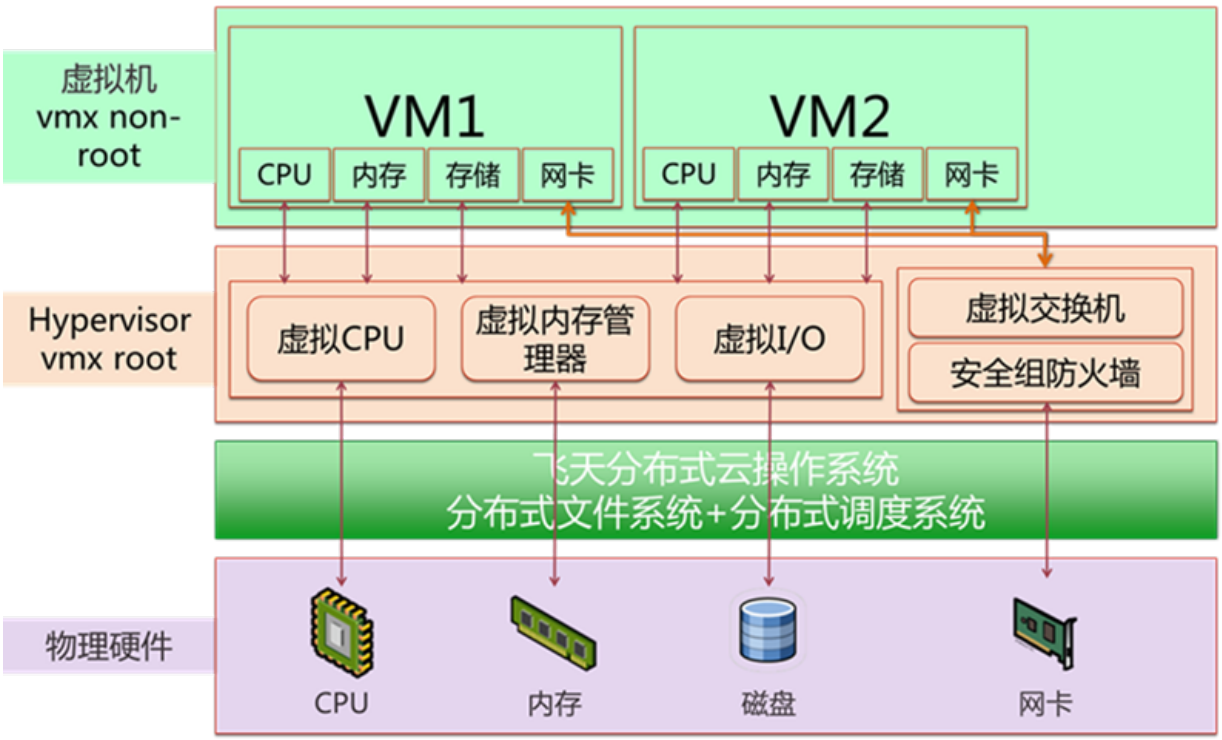
- 应用硬件协助的虚拟化技术模式（Xen Full Virtualization，简称HVM模式），基于VT-x技术隔离CPU。
- 硬件辅助扩展页表技术（Extended Page Table，简称EPT）EPT技术隔离内存。
- 分离设备驱动I/O模型隔离存储。
- 通过交换型VSwitch，不同虚拟机的数据包被转发到对应的虚拟端口。

- 虚拟机的IP和MAC地址绑定，防地址欺骗及网络嗅探。
- VPC、安全组、防火墙隔离租户网络。
- 物理内存、物理存储重分配前清零。

其他云产品租户隔离

- 用户数据打标签隔离存储。
- 基于身份验证进行访问控制。

图 78: 多租户隔离架构



23.4.1.3 数据安全

数据安全承诺

2015年7月，阿里云发起中国云计算服务商首个“数据保护倡议”：运行在云计算平台上的开发者、公司、政府、社会机构的数据，所有权绝对属于客户；云计算平台不得将这些数据移作它用。平台方有责任和义务，帮助客户保障其数据的私密性、完整性和可用性。

多副本分布式存储

专有云使用分布式存储，文件被分割成许多数据片段分散存储在不同的设备上，并且每个数据片段存储多个副本。分布式存储不但提高了数据的可靠性，也提高了数据的安全性。

加密存储

专有云加密服务使用经国家密码管理局检测认证的硬件密码机，帮助您满足数据安全方面的监管合规要求，保护云上业务数据的机密性。借助加密服务，您可以实现对加密密钥的完全控制并进行解密操作。

加密传输

专有云平台提供标准的加密传输协议，便于云平台与外界以及系统间传输敏感数据的需求。云平台支持标准的TLS协议，可提供高达256位密钥的加密强度，完全满足敏感加密传输需求。

残留信息

对于曾经存储过客户数据的内存和磁盘，一旦释放和回收，其上的残留信息将被自动清零。同时，任何更换和淘汰的存储设备，都将统一执行消磁处理并物理折弯之后，才能运出数据中心。

数据审计

通过平台级的访问审计，以及产品级的SQL审计、上传下载审计，确保数据的生成、变更、删除、传播有迹可循，使违规的数据操作无所遁形。

23.4.1.4 开发发全

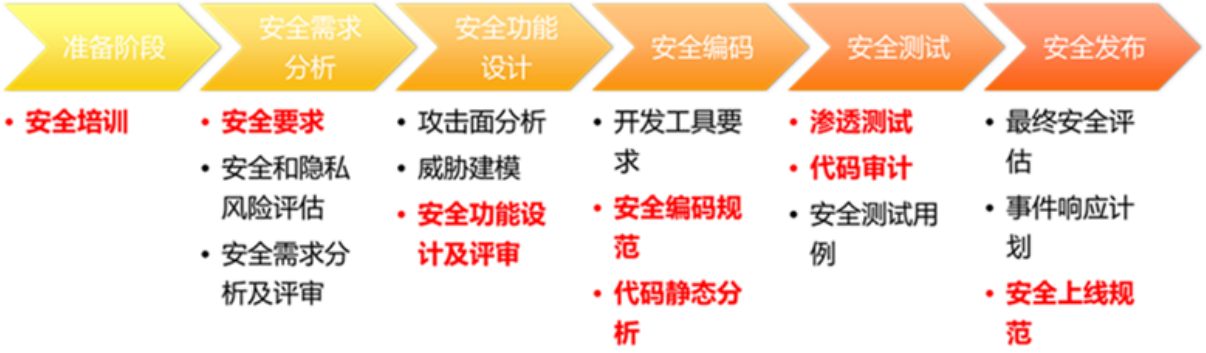
源代码是组成云计算平台的最小单元，大量信息安全问题根源是由开发设计缺陷引起的。

专有云平台开发遵循软件安全开发生命周期（Security Development Lifecycle，简称SDL），解决绝大部分因源代码安全缺陷而引发的安全问题。SDL覆盖飞天操作系统、云产品、大数据产品、OpenAPI等，从产品开发初期就基于云端安全威胁构建云服务的安全功能或属性，极大提高了云平台安全的健壮性。

软件安全开发生命周期包含：

- 需求分析：识别云服务安全需求和风控需求。
- 产品设计：攻击面分析、威胁建模、安全架构/功能设计。
- 编码阶段：采用安全开发框架，遵循安全编码规范。
- 测试阶段：渗透测试结合代码审计。未经安全测试的产品禁止上线。
- 发布阶段：按照安全规范实施整体加固。

图 79: 安全开发生命周期管理



23.4.1.5 漏洞热修复

漏洞热修复是指在不影响系统正常运行的情况下进行系统漏洞修复，防止由于冷补丁修复导致的业务中断、业务崩溃和宕机等问题。

漏洞热修复基于阿里云六年的云安全运营经验，支持以下系统和平台：

- Linux内核
- 飞天分布式云操作系统及各云产品
- ECS Xen Hypervisor
- RDS MySQL

案例

2015年3月，Xen曝出安全漏洞：XSA-123，该漏洞造成客户机指令提权，任意一台云主机可读取到其他另一台云主机的重要数据，从而导致客户数据泄密。Rackspace 服务器大规模重启，影响客户业务超过10分钟。 亚马逊99.9%的服务器热修复，近0.1%的服务器重启。阿里云100%的服务器热修复，对客户业务无任何影响。阿里云在漏洞曝出前就拿到了相关漏洞细节，并提前发布了热补丁修复该漏洞，并在没有影响任何用户主机的情况下完成了整个系统的修复。

23.4.2 云产品安全

23.4.2.1 云服务器安全

专有云云服务器提供以下安全功能：

镜像安全

- 定期修复高危漏洞。
- 内置主机入侵防护软件。

热升级

- 宿主机Linux内核热升级。
- Hypervisor热升级。

租户隔离

- Hypervisor隔离不同虚拟机的CPU、内存、存储。
- 通过专有网络VPC和安全组隔离不同租户网络。
- 内存、存储释放后数据清零。

可靠性

- 分布式冗余存储保障数据可靠性。
- 基于磁盘快照的快捷备份和回滚。
- 基于故障迁移的即时恢复。
- 基于在线迁移的智能资源调度。
- 可用性高达99.95%。

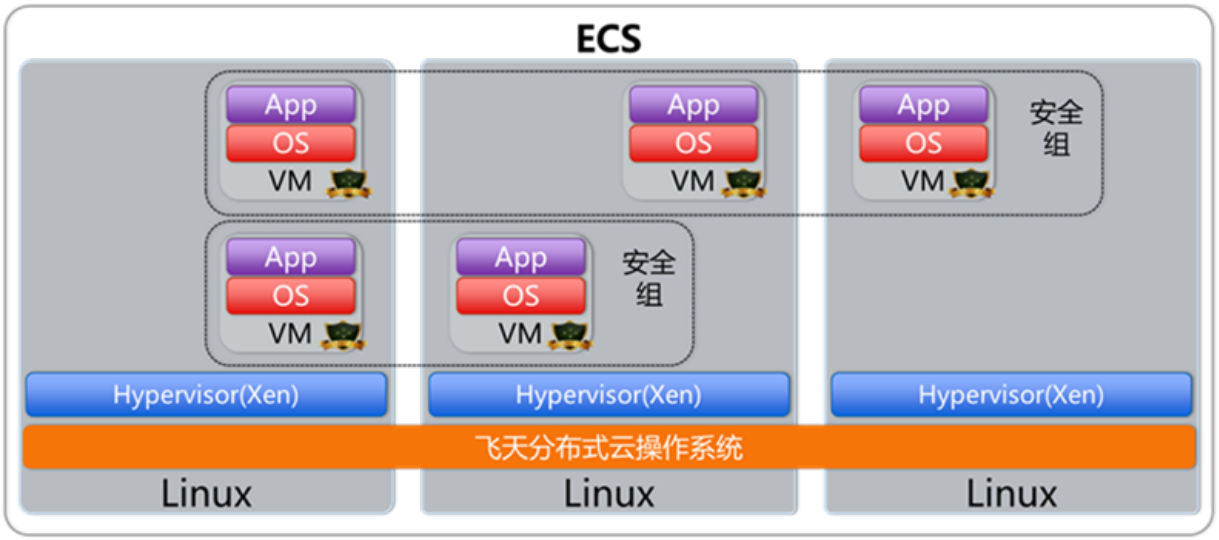
网络安全

- 专有网络VPC（基于VxLAN）隔离。
- 状态检测虚拟防火墙，划分安全域。
- 防IP、MAC伪造和地址解析协议（Address Resolution Protocol，简称ARP）欺骗。
- 防网络嗅探。

主机安全

- 租户拥有最高权限，阿里云没有登录权限。
- Linux支持SSH Key认证，防暴力破解。

图 80: 云服务器安全架构



23.4.2.2 云数据库安全

专有云提供以下云数据库安全功能：

租户隔离

- 数据库实例隔离。

可靠性

- RAID5磁盘阵列存储，保障数据可靠性。
- 每个RDS实例拥有两个物理节点进行主从热备份。
- 支持秒级主备切换。
- 数据库定时备份，能够根据备份文件将数据库恢复至七日内任意时间点。
- 可用性高达99.95% 。

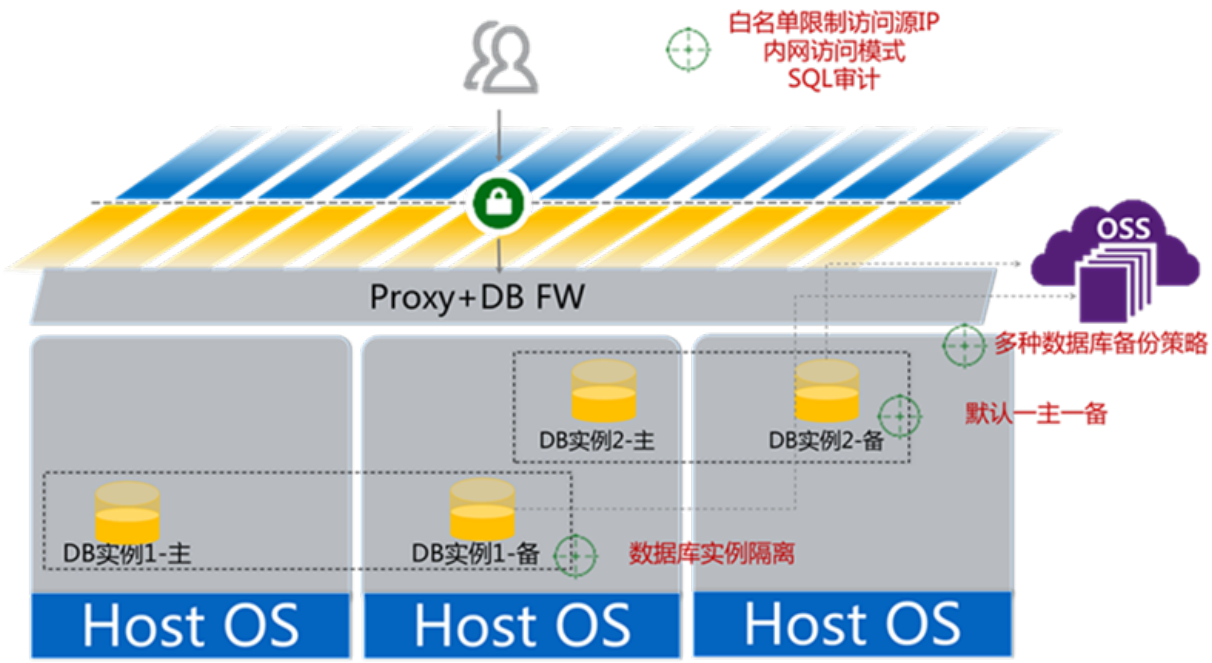
网络安全

- 通过IP白名单限定允许访问RDS服务的源IP。

热升级

- RDS For MySQL实例热升级，客户业务无感知。

图 81: 云数据库安全



23.4.2.3 云存储安全

专有云提供以下云存储安全功能:

租户隔离

- 租户数据打标签区分。
- 服务接入层对称密钥认证技术鉴别用户。

可靠性

- 分布式冗余存储保障数据可靠性。
- 可用性高达99.9%。

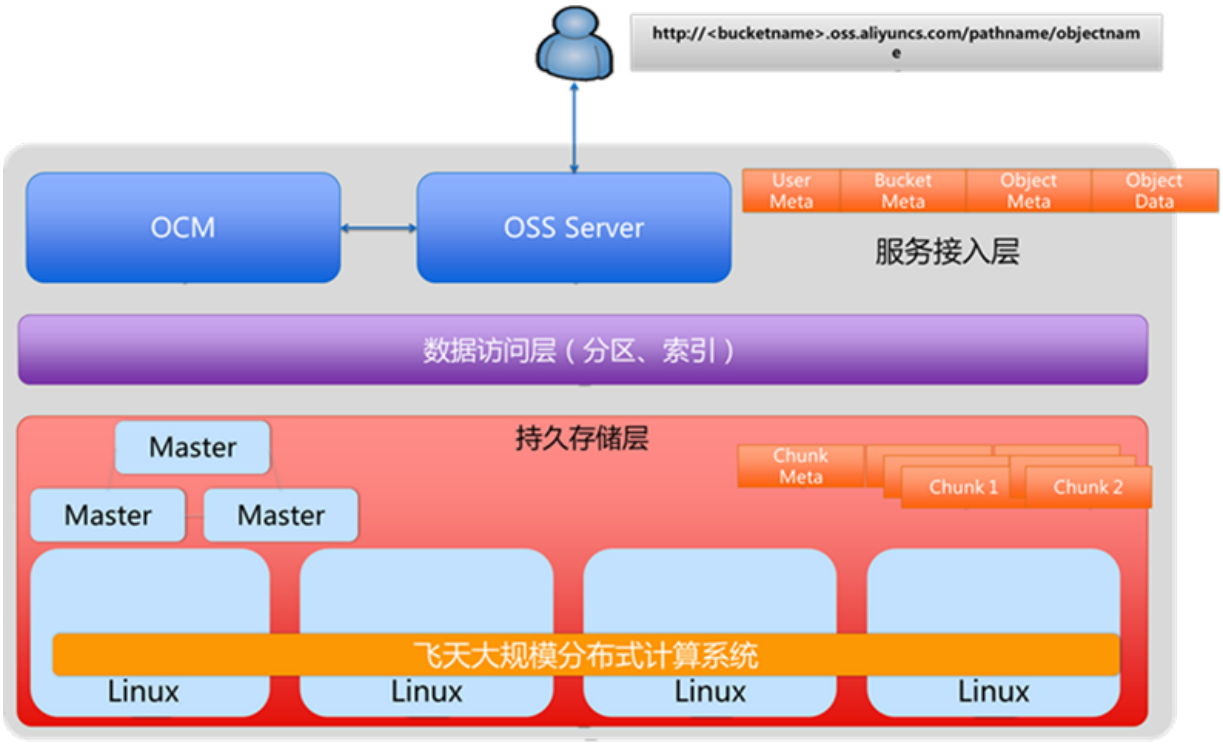
访问控制

- 通过访问控制列表（ACL）进行访问控制。
- 基于访问控制管理(Resource Access Management,简称 RAM)授权策略的访问控制。

加密传输

- 支持SSL传输加密。
- 支持服务器端加密存储。

图 82: 云存储安全架构



23.5 典型应用

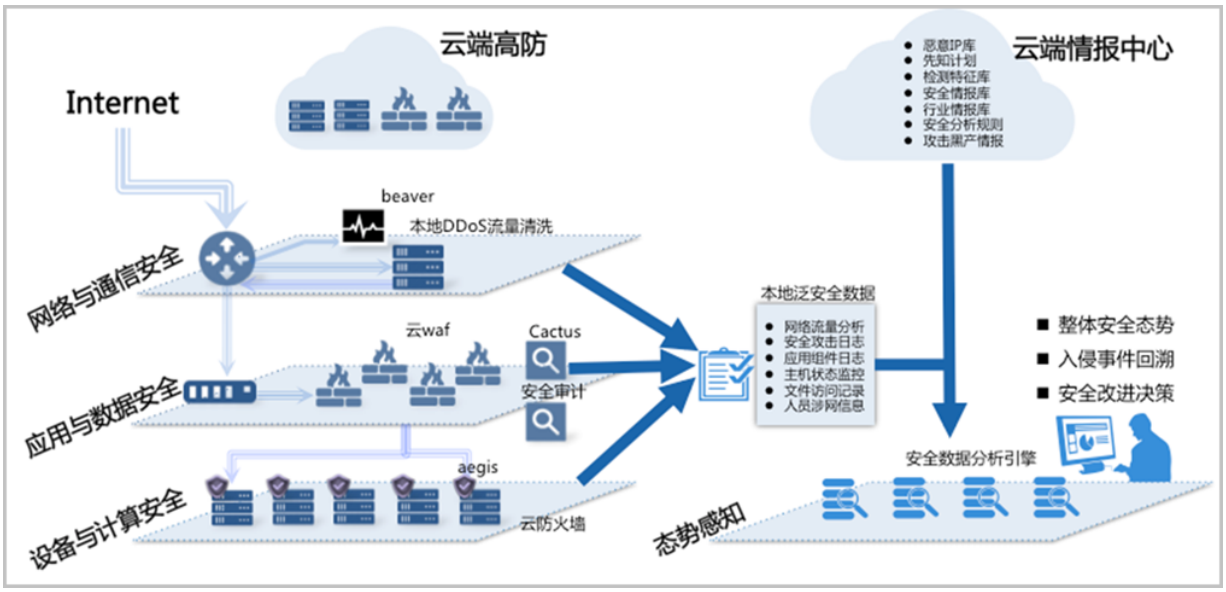
云盾不同于传统的软硬件安全产品，它采用纵深防御，多点联动的云安全架构，完全基于阿里云的云计算环境研发，从网络层、应用层、主机层等多个层面为用户提供全面的、一体化的云安全防护能力。

单数据中心部署

专有云V3云盾高级版包含网络流量监控模块、安全审计模块、主机入侵防御（安骑士）模块、弱点扫描模块、云防火墙模块、DDoS清洗模块、Web应用防护模块、和态势感知模块等。

专有云云盾在单数据中心的场景下的部署架构如下图所示：

图 83: 云盾高级版典型部署架构图

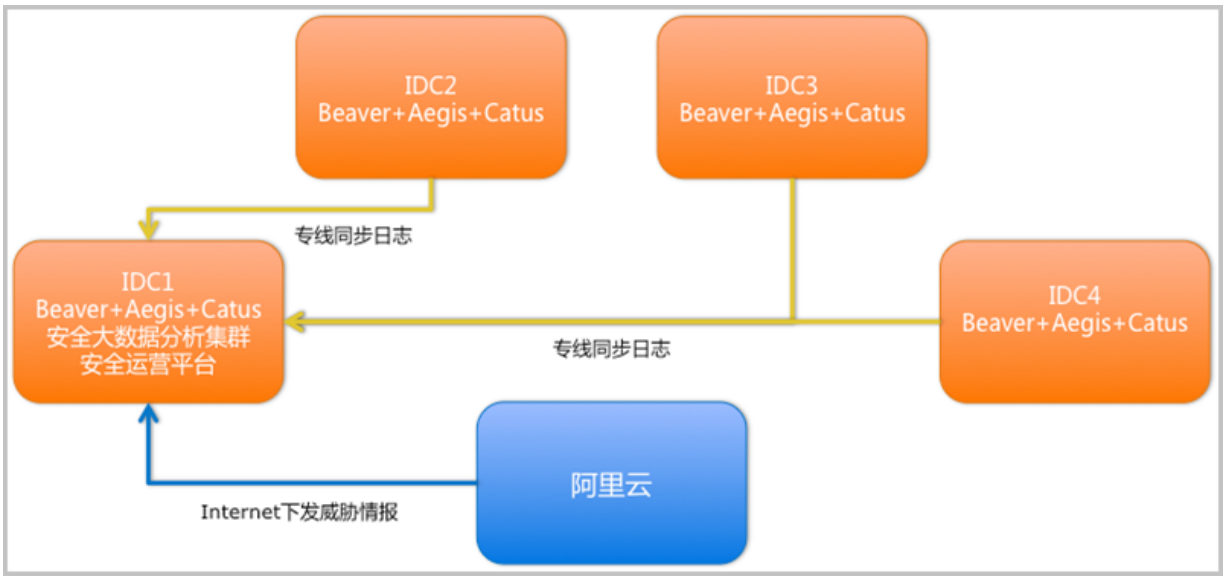


多数据中心部署

专有云云盾支持多数据中心、多链路部署，各数据中心日志通过专线同步，威胁情报统一下发，实现跨数据中心的云安全一体化管理。

专有云云盾在多数据中心场景下的部署架构如下图所示：

图 84: 多机房多链路部署架构图



23.6 基本概念

攻击可见

把过去那些看不见的攻击痕迹呈现在用户面前，并结合当下的安全情况，给出一份可持续的安全策略方案。

安全可控

通过对主机、网络、应用和数据进行全面的弱点分析，对网络行为实时监控。

DDoS攻击

Distributed Denial of Service攻击，即分布式拒绝服务。攻击指借助于客户/服务器技术，将多个计算机联合起来作为攻击平台，对一个或多个目标发动DDoS攻击，从而成倍地提高拒绝服务攻击的威力。

畸形报文

包括Frag flood、Smurf、Stream flood、Land flood攻击、IP畸形包、TCP畸形包、及UDP畸形包。

传输层DDoS攻击

包括Syn flood、Ack flood、UDP flood、ICMP flood、Rstflood等攻击。

Web应用DDoS攻击

包括HTTP get flood、HTTP post flood、 cc等攻击。

DNS DDoS攻击

包括DNS request flood、DNS response flood、虚假源+真实源DNS query flood、权威服务器和本地服务器攻击。

连接型DDoS攻击

包括TCP慢速连接攻击、连接耗尽攻击、Loic、Hoic、Slowloris、Pyloris、Xoic等慢速攻击。

WAF

Web应用防护系统（又名：网站应用级入侵防御系统)。Web应用防火墙是通过执行一系列针对HTTP/HTTPS的安全策略来专门为Web应用提供保护。

针对性攻击

通过Spark模型的规则计算出的一些特定攻击。

24 云监控

24.1 产品概述

云监控（CloudMonitor）是一项针对阿里云资源进行监控的服务。云监控服务可用于收集获取阿里云资源的监控指标，以及针对指标设置警报。

云监控服务能够监控云服务器 ECS、云数据库 RDS 和负载均衡等各种阿里云服务资源，同时也能够通过 HTTP，ICMP 等通用网络协议监控互联网应用的可用性。借助云监控服务，您可以全面了解您在阿里云上的资源使用情况、性能和运行状况。借助报警服务，您可以及时做出反应，保证应用程序顺畅运行。

24.2 产品优势

云监控是阿里巴巴集团多年来服务器监控技术研究积累的成果，结合阿里云云计算平台强大的数据分析能力，为您提供云服务监控、站点监控和自定义监控，为您的产品、业务保驾护航。

天然集成

云监控服务无需特意购买和开通，您注册好阿里云账号后，便自动为您开通了云监控服务，方便您在购买和使用阿里云产品后直接到云监控查看产品运行状态并设置报警规则。

数据可视化

云监控通过 Dashboard 为您提供丰富的图表展现形式，并支持全屏展示和数据自动刷新。满足各种场景下的监控数据可视化需求。

监控数据处理

云监控支持您通过 Dashboard 对监控数据进行时间维度和空间维度的聚合处理。

灵活报警

云监控还为您提供了监控项的报警服务。您在为监控项设置好合理的报警规则和通知方式后，一旦发生异常便会立刻为您发出报警通知，让您及时知晓服务异常并处理异常，从而提高用户产品的可用性。

24.3 典型应用

云监控为您提供了非常丰富的使用场景，下面按服务为您举例说明。

云服务监控

您购买和使用了云监控支持的阿里云服务后，即可方便地在云监控对应的产品页面查看您的产品运行状态、各个指标的使用情况并对监控项设置报警规则。

系统监控

监控实例性能指标数据，及时获取实例使用情况。

通过监控 ECS 的 CPU 使用率、内存使用率、公网流出流速（带宽）等基础指标，确保实例的正常使用，避免因为对资源的过度使用造成您的业务无法正常运转。

及时处理异常场景

云监控会根据您设置的报警规则，在监控数据达到报警阈值时发送报警信息，让您及时获取异常通知、查询异常原因。

及时扩容场景

对带宽、连接数、磁盘使用率等监控项设置报警规则后，可以让您方便地了解云服务现状，在业务量变大后及时收到报警通知进行服务扩容。

24.4 基本概念

以下名词是云监控的关键概念。

云服务监控

为阿里云服务用户提供各个产品的性能指标查看，当前支持 ECS、RDS、负载均衡、OSS 等主要云产品的监控指标。

报警服务

支持用户对上述三种监控服务的指标设置报警规则。当监控数据满足报警规则的设置时，会发出报警通知。

监控项

用户设置或者系统默认的监控数据类型，例如站点监控的 HTTP 监控默认有两个监控项响应时间和状态码。ECS 的监控项有 CPU 利用率、内存利用率等等。

维度

定位监控项数据位置的维度，例如磁盘 IO 这个监控项，通过实例和磁盘名称两个维度可以定位到唯一的监控数据。在自定义监控中，目前维度用字段信息表示。

报警规则

报警规则是一个条件。例如内存使用率统计周期为 5 分钟，连续 3 次大于等于 50% 是一个规则。

通道沉默

当某一条报警发出后，如果这个指标 24 小时之内持续超过报警阈值，则 24 小时内不会再次触发报警。

报警联系人

报警通知的接收人。

报警组

一组报警联系人，可以包含一个或多个报警联系人。在报警设置中，均通过报警组发送报警通知。报警系统根据预先设定的报警方式，在到达报警阈值时向报警组成员发送报警通知。

通知方式

给用户发送报警通知的方式。包含短信、旺旺（淘宝）、邮件、MNS 消息队列推送。

25 访问控制

25.1 产品概述

RAM（Resource Access Management）是阿里云为客户提供的用户身份管理与访问控制服务。使用 RAM，您可以创建、管理用户账号（比如员工、系统或应用程序），并可以分配这些用户账号对您名下资源具有的操作权限。当您的企业存在多用户协同操作资源时，使用 RAM 可以让您避免与其他用户共享云账号密码或访问密钥，按需为用户分配最小权限，从而降低您的企业信息安全风险。

RAM 设计思路

RAM 允许在一个云账号下创建并管理多个用户身份，并允许给单个身份或一组身份分配不同的授权策略（ Policy），从而实现不同用户拥有不同的云资源访问权限。

RAM 用户身份是指任意的通过控制台或 OpenAPI 操作专有云资源的人、系统或应用程序。为了支持多种应用场景的身份管理，RAM 支持两种不同的用户身份类型：RAM-User 和 RAM-Role。RAM-User 是一种实体身份，有确定的身份 ID 和身份认证密钥，它通常与某个确定的人或应用程序一一对应。 RAM-Role 是一种虚拟身份，有确定的身份 ID，但没有确定的身份认证密钥。RAM-Role 需要与某个实体身份进行关联之后才能被使用。一个 RAM-Role 可以与多种实体身份关联，比如可以与当前云账号下的 RAM-User 关联，与其他云账号下的 RAM-User 关联，与专有云服务（EMR / MTS / ...）关联，与外部实体身份（如企业本地账号）关联。

RAM 允许在云账号下创建并管理多个授权策略，每个授权策略本质上是一组权限的集合。管理员可以将一个或多个授权策略分配给 RAM 用户（包括 RAM-User 和 RAM-Role）。RAM 授权策略语言可以表达精细的授权语义，可以指定对某个 API-Action 和 Resource-ID 授权，也可以支持多种限制条件（源 IP、访问时间、多因素认证等）。

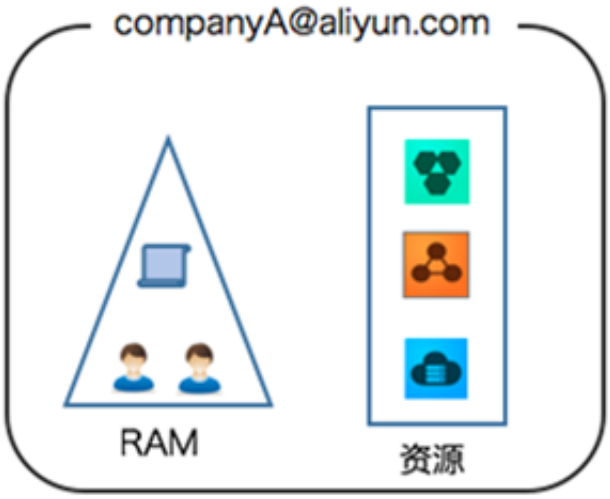
云账户 VS RAM 用户

- 从归属关系上看，云账户与 RAM 用户是一种主从关系。云账户是专有云资源归属、资源使用计量的基本主体。RAM 用户只能存在于某个云账户下的 RAM 实例中。RAM 用户不拥有资源，在被授权操作时所创建的资源归属于主账户。
- 从权限角度看，云账户与 RAM 用户是一种 root 与 user 的关系（类比 Linux 系统）。Root 对资源拥有一切操作控制权限，而 user 只能拥有被 root 所授予的某些权限，而且 root 在任何时刻都可以撤销 user 身上的权限。

RAM 与企业级云资源管理

需求说明（如下图所示）：

图 85: 需求示例图



- 您的企业只需使用一个云账号（比如 companyA@aliyun.com）。
- 所有资源都归属于该云账号的名下，云账号是资源的 Owner（掌握完全控制权的人），也是账单的支付者。
- 通过 RAM 为您名下的操作员（对资源进行运维管控操作）创建独立的用户账号并进行授权管理。
- 用户账号不拥有资源（对其所创建的资源默认没有访问权限），只能操作被授权的资源。

适用具有如下特点的企业场景：希望能够直观地管理每个操作人员（或应用）的账号及权限。

25.2 基本概念

云账户（主账户）

云账户是阿里云资源归属、资源使用计量的基本主体。当您开始使用专有云服务时，首先需要注册一个云账户。云账户对其名下所有资源拥有完全权限。默认情况下，资源只能被属主（ResourceOwner）所访问，任何其他用户访问都需要获得属主的显式授权。所以从权限管理的角度来看，云账户就是操作系统的 root 或 Administrator，所以我们有时称它为**根账户**或**主账户**。

云账户别名

每个云账户可以在 RAM 中为自己设置一个全局唯一的别名。别名主要用于 RAM 用户登录以及成功登录后的显示名。比如，云账号 admin@abc.com 为自己设置一个别名为 abc.com，那么其名下的 RAM 用户 alice 成功登录后，显示名就是 alice@abc.com。

RAM 用户

RAM 允许在一个云账户下创建多个 RAM 用户（可以对应企业内的员工、系统或应用程序）。RAM 用户不拥有资源，没有独立的计量，这些用户由所属云账户统一控制。RAM 用户是归属于云账户，只能在所属云账户的空间下可见，而不是独立的云账户。RAM 用户必须在获得云账户的授权后才能登录控制台或使用 API 操作云账户下的资源。

RAM 用户有两种身份类型：**RAM-User** 和 **RAM-Role**。RAM-User 类型是一种实体身份类型，有确定的身份 ID 和身份凭证，它通常与某个确定的人或应用程序一一对应。RAM-Role 类型是一种虚拟身份类型，它没有确定的身份凭证，它必须关联到某个实体身份上才能使用。

RAM-Role 与 Textbook-Role（教科书式角色）的差异：

- （相同点）RAM-Role 和 Textbook-Role 都可以绑定一组权限集。
- （不同点）RAM-Role 是一种虚拟身份或影子账号，它有独立的身份 ID，除了绑定权限之外，还需要指定角色列表（Roleplayers），它主要用于解决与身份联盟（Identity Federation）相关的问题。Textbook-Role 通常只表示一组权限的集合，它不是身份，主要用于简化授权管理。

RAM-Role 的扮演与切换：

- 从登录身份切换到角色身份（SwitchRole）：一个实体用户（比如 RAM-User）登录到控制台后，可以选择**切换到某个角色**，前提是这个实体用户已经被关联了角色。每次只能切换进入某一种角色。当用户从**登录身份**进入**角色身份**时，用户只能使用**角色身份**上所绑定的权限，而**登录身份**上绑定的权限会被屏蔽。如果需要使用**登录身份**的权限，那么需要从**角色身份**切换回到**登录身份**。
- 从实体身份通过程序调用方式扮演角色（AssumeRole）：如果一个实体用户（比如 RAM-User）关联了某个 RAM-Role，那么该用户可以使用访问密钥（AccessKey）来调用 STS 服务的 AssumeRole 接口来获得这个 RAM-Role 的一个临时访问密钥。临时访问密钥有过期时间和受限制的访问权限（不会超过该角色所绑定的权限集），通常用于解决临时授权问题。

身份凭证（Credential）

身份凭证是用于证明用户真实身份的凭据，它通常是指登录密码或访问密钥（Access Key）。身份凭证是秘密信息，用户必须保护好身份凭证的安全。

- 登录名/密码（Password）

  您可以使用登录名和密码登入专有云控制台，查看订单或购买资源，并通过控制台进行资源操作。

- 访问密钥（AccessKey）

  您可以使用访问密钥构造一个 API 请求（或者使用云服务 SDK）来操作资源。

- 多因素认证

  多因素认证（Multi-Factor Authentication， MFA）是一种简单有效的最佳安全实践方法，它能够在用户名和密码之外再额外增加一层安全保护。启用 MFA 后，用户登录专有云网站时，系统将要求输入用户名和密码（第一安全要素），然后要求输入来自其 MFA 设备的可变验证码（第二安全要素）。这些多重要素结合起来将为您的账户提供更高的安全保护。

资源（Resource）

资源是云服务呈现给用户与之交互的对象实体的一种抽象，如对象存储 OSS 的存储空间，ECS 实例等。

我们为每个资源定义了一个全局的专有云资源名称（Alibaba Cloud Resource Name， ARN）。格式如下：

`acs:<service-name>:<region>:<account-id>:<resource-relative-id>`

格式说明：

- **acs**：它是 Alibaba Cloud Service 的首字母缩写，表示阿里云平台。
- **service-name**：阿里云提供的 Open Service 的名字，如 *ecs, oss, odps* 等。
- **region**：地区信息。如果不支持该项，可以使用通配符“*”号来代替。
- **account-id**：账号 ID，比如 1234567890123456。
- **resource-relative-id**：与 service 相关的资源描述部分，其语义由具体 service 指定。以 OSS 为例，acs:oss:*:1234567890123456:sample_bucket/file1.txt 表示云平台 OSS 资源，OSS 对象名称是*sample_bucket/file1.txt*，对象的 Owner 是*1234567890123456*。

权限（Permission）

权限是允许（Allow）或拒绝（Deny）一个用户对某种资源执行某种操作。

操作可以分为两大类：**资源管控操作**和**资源使用操作**。资源管控操作是指云资源的生命周期管理及运维管理操作，比如 ECS 的实例创建、停止、重启等，OSS 存储空间的 创建、修改、删除等。资源使用操作是指使用资源的核心功能，比如 ECS 实例操作系统中的用户操作，OSS 存储空间的数据上传 / 下载。资源管控所面向的用户一般是资源购买者或您组织内的运维员工，资源使用所面向的用户则是您组织内的研发员工或应用系统。

对于弹性计算和数据库产品，资源管控操作可以通过 RAM 来管理，而资源使用操作是在每个产品的实例内进行管理，比如 ECS 实例操作系统的权限控制，MySQL 数据库提供的权限控制。单对于存储类产品，如 OSS， Table Store 等，资源管控操作和资源使用操作都可以通过 RAM 来管理。

授权策略（Policy）

授权策略是描述权限集的一种简单语言规范。

RAM 支持两种类型的授权策略：专有云平台管理的**系统访问策略**和客户管理的**自定义访问策略**。对于专有云云平台管理的系统访问策略，用户只能使用，不能修改，云平台会自动完成系统访问策略的版本更新。对于客户管理的自定义访问策略，用户可以自主创建和删除，策略版本由客户自己维护。

25.3 典型应用

企业子账号管理与分权

企业 A 购买了多种云资源（如 ECS 实例/RDS 实例/负载均衡实例/OSS 存储空间/ ...），A 的员工需要操作这些云资源，比如有的负责购买，有的负责运维，还有的负责线上应用。由于每个员工的工作职责不一样，需要的权限也不一样。出于安全或信任的考虑，A 不希望将云账号密钥直接透露给员工，而希望能给员工创建相应的用户账号。用户账号只能在授权的前提下操作资源，不需要对用户账号进行独立的计量。当然，A 随时可以撤销用户账号身上的权限，也可以随时删除其创建的用户账号。

不同企业之间的资源操作与授权管理

A 和 B 代表不同的企业。A 购买了多种云资源（如 ECS 实例/RDS 实例/负载均衡实例/OSS 存储空间/ ...）来开展业务。A 希望能专注于业务系统，而将云资源运维监控管理等任务委托或授权给企业 B。当然，企业 B 可以进一步将代运维任务分配给 B 的员工。B 可以精细控制其员工对 A 的云资源操作权限。如果 A 和 B 的这种代运维合同终止，A 随时可以撤销对 B 的授权。

针对不可信客户端 APP 的临时授权管理

企业 A 开发了一款移动 APP ，并购买了 OSS 服务。移动 APP 需要上传数据到 OSS（或从 OSS 下载数据），A 不希望所有 APP 都通过 APP Server 来进行数据中转，而希望让 APP 能直连 OSS 上传/下载数据。由于移动 APP 运行在用户自己的终端设备上，这些设备并不受 A 的控制。出于安全考虑，A 不能将访问密钥保存到移动 APP 中。A 希望将安全风险控制到最小，比如，每个移动 APP 直连 OSS 时都必须使用最小权限的访问令牌，而且访问时效也要很短（比如30分钟）。

25.4 支持的云服务列表

所有阿里云服务都会与 RAM 集成。目前支持产品服务如下：

表 37: 支持的云服务

产品类别	产品名称
弹性计算	• 云服务器 • 高性能计算 • 容器服务
存储	• 对象存储 OSS • 消息服务 • 表格存储
数据库	• 云数据库RDS版 • 云数据库Redis版 • 云数据库MongoDB版 • 云数据库Memcache版 • 云数据库OceanBase版 • 数据传输服务DTS • 数据管理
网络	• 负载均衡 • 高速通道 • NAT网关 • 专有网络VPC
应用服务	• 日志服务 • API网关

产品类别	产品名称
云盾	• 云盾
互联网中间件	• 企业级分布式应用服务EDAS • 分布式关系型数据库DRDS • 消息队列MQ • 业务实时监控服务ARMS • 全局事务服务GTS • 云服务总线CSB
数据分析展现	• 大数据解决方案 • 关系网络分析 • 画像分析
大数据基础服务	• 大数据计算分析 • 大数据开发套件 • 分析型数据库 • 流计算 • 大数据管家
人工智能	• 机器学习
管理与监控	• 云监控 • 访问控制 • 资源编排 • OMS

26 计量服务OMS

26.1 产品概述

阿里云计量服务（OpenMetering Service，简称OMS），是阿里云对外提供用户在阿里云平台上使用各个服务（如OSS、Table Store、MaxCompute、RDS、VM、ACE）产生的计量数据存储和查询服务。

用户可以通过OpenAPI，获取格式化的自己消耗的计量数据。

26.2 基本概念

Domain

Domain（计量实体）是用户可查询的数据Object所归属的服务的类别。比如用户使用阿里云的OSS服务，希望能够看到自己使用OSS的计量数据，可以查询OSS相关的Domain，比如/OSS。用户可以用传统数据中的表来类比开放计量服务中的Domain。用户开通阿里云计量服务后，可以获得自己使用的其他云服务的计量信息。用户也可以创建自己的Domain用来对自己的业务的子用户进行计量服务。与数据库中的Table概念类似。

Object

Object（计量数据对象）是指用户某一个Domain里面的计量数据记录的集合，用户可以指定时间范围，需要的数据类别等参数，从中筛选出自己需要的数据，Object是数据查询的最小单位，用户可以使用OpenAPI从相应的Domain里面获取数据Object。Object可以类比为传统数据库中的查询记录。

Field

Field（计量项），如Memory、CPU等，类似于Table的Column。目前OMS支持的Field数据类型包括：Integer、String、Double和Timestamp。Integer是64位带符号位的整数。Timestamp是一种特殊的数据类型，是秒值，表示自1970年1月1日00:00:00 GMT以来经过的秒数，并且是由OMS自动生成。

Key

某个Object的唯一标识，Key（计量数据标识）可以由多个Field组成，类似于Table的Primary Key。

Dimension

Dimension（可查询维度），可以由多个Field组成，类似于Table中的Index。查询目标Domain采用的查询条件需要有对应的dimension定义才会有效。

27 企业级分布式应用服务EDAS

27.1 产品概述

企业级分布式应用服务（Enterprise Distributed Application Service，简称 EDAS）是企业级互联网架构解决方案的核心产品。EDAS 充分利用阿里云现有资源管理和服务体系，引入中间件成熟的整套分布式计算框架（包括HSF 或 Dubbo 分布式服务化框架、服务治理、运维管控、链路追踪和稳定性组件等），以应用为中心，帮助企业级客户轻松构建并托管分布式应用服务体系。

图 86: EDAS产品示意图

27.2 功能特性

EDAS 作为阿里巴巴分布式服务架构的核心产品，涵盖了应用生命周期管理、运维管控等众多功能。

27.2.1 容器

作为 EDAS 平台应用运行的基础容器，EDAS Container 集成了阿里巴巴中间件技术栈，在容器启动、监控、稳定性及性能上得到极大的提升。同时，EDAS Container 全面兼容 Apache Tomcat 。

27.2.2 以应用为中心的中间件 PaaS 平台

应用基本管理和运维

在 EDAS 控制台上，您可以一站式完成应用生命周期的管控，包括创建、部署、启动、停止、扩容、缩容和应用下线等，依托阿里巴巴平台超大规模集群运维管理经验，轻松运维上千个实例的应用。

弹性伸缩

EDAS 支持手动和自动两种方式来实现应用的扩容与缩容，可以通过对 CPU、内存和负载的实时监控来实现对应用的秒级扩容和缩容。

主/子账户体系

EDAS 独创主子体系，您可以根据自己企业的部门划分、团队划分和项目划分在 EDAS 平台上建立对应的主子账号关系；同时，ECS 资源也以主子账号关系进行划分，便于您进行资源的分配。

角色与权限控制

应用的运维通常涉及应用研发负责人、应用运维负责人和机器资源负责人。不同的角色对于一个应用的管理操作各不一致，因此 EDAS 提供了角色和权限控制机制，可以为不同的账号定义各自的角色，并分配相应的权限。

27.2.3 丰富的分布式服务

分布式服务框架

自2007年，伴随着阿里巴巴电商平台大规模分布式改造的持续进行，自主研发的分布式服务框架 HSF（High-Speed Service Framework）和 Dubbo 应运而生。HSF 是一款面向企业级互联网架构的分布式服务框架，以高性能网络通信框架为基础，提供了诸如服务发布与注册、服务调用、服务路由、服务鉴权、服务限流、服务降级和服务调用链路跟踪等一系列久经考验的功能特性。

分布式配置管理

集中式系统变成分布式系统后，如何有效地对分布式系统中每一个机器上的配置信息进行有效的实时管理成了一个难题。EDAS 提供高效的分布式配置管理，能够将分布式系统的配置信息在 EDAS 控制台上集中管理起来，做到一处配置，处处使用。更重要的是，EDAS 允许您在控制台上对配置信息进行修改，在秒级时间内就能够实时通知到所有的机器。

分布式任务调度

任务调度服务允许您创建任意周期性的单机或者分布式调度任务，并能对任务运行周期进行管理，同时提供对任务的历史执行记录进行查询。适用于诸如每天凌晨 2 点定时迁移历史数据，每隔 5 分钟进行任务触发，每个月的第一天发送系统月报等任务调度场景。

分布式事务

分布式事务（GTS）是一款高性能、高可靠、接入简单的分布式事务中间件，用于解决分布式环境下的数据一致性问题。EDAS配合GTS使用，能够轻松实现分布式数据库事务、多库事务、消息事务、服务链路级事务及其各种组合，策略丰富，易用性和性能兼顾。

27.2.4 运维管控与服务治理

服务鉴权

HSF 服务框架致力于保证每一次分布式调用的稳定与安全。在服务注册、服务订阅以及服务调用等每一个环节，都进行严格的服务鉴权。

服务限流

EDAS 可以对每一个应用提供的众多服务配置限流规则，以实现对服务的流控，确保服务能够稳定运行。EDAS 提供了从 QPS 和线程两个维度提供对限流规则的配置，保证系统在应对流量高峰时能以最大的支撑能力平稳运行。

服务降级

每一个应用会调用许多外部服务，对于这些服务配置降级规则可以实现对劣质服务的精准屏蔽，确保应用自身能够稳定运行，防止劣质的服务依赖影响应用自身的服务能力。EDAS 从响应时间维度对降级规则进行配置，可以在应对流量高峰时合理地屏蔽劣质依赖。

自动化压测

自动化压测工具模型，能够帮助客户将性能压测融入到日常生活中。所有自动化压测的流量，全是生产环境真实流量，通过对服务权重的控制，在保证稳定的前提下，真正意义上做到线上压测。

容量规划

将性能压测工作日常化之后，应用的负责人能够非常方便的看到应用的性能指标，并根据这些性能指标，结合当前系统运行水位，实现对应用精准的容量规划。

27.2.5 立体化监控与数字化运营

分布式链路跟踪

EDAS 鹰眼监控系统能够分析分布式系统的每一次系统调用、消息发送和数据库访问，从而精准发现系统的瓶颈和隐患。

服务调用监控

EDAS 能够针对应用的服务调用情况，对服务的 QPS 、响应时间和出错率进行全方面的监控。

IaaS 基础监控

EDAS 能够针对应用的运行状态，对机器的 CPU 、内存、负载、网络和磁盘等基础指标进行详细的监控。

27.3 产品优势

EDAS 支撑了整个阿里巴巴 99% 以上的大规模应用系统，其中涵盖了包括会员、交易、商品、店铺、物流和评价在内的所有在线核心系统，在稳定性、可靠性等多个维度具有独特的优势。

更可靠

- 阿里巴巴近 10 年使用与沉淀的核心技术产品。
- 支持全集团所有核心应用稳定运行。
- 历次双十一大促考验。
- 完善的鉴权体系保证每一次服务调用的安全可靠。

更全面

- 完善的 PaaS 平台支持应用生命周期的管理。
- 完整的服务治理解决方案管理分布式服务。
- 全面的应用诊断排查系统轻松定位故障根源。
- 线上压测，容量规划轻松获取线上机器运行性能指标和实时运行水位。
- 自动弹性伸缩从容应对突发流量高峰。

更深入

- 深入业务指标，实现全盘报表。
- 立体化多维度监控，实现全息排查。
- 链路跟踪洞察每一次分布式调用。
- 依赖分析剖析每一处系统瓶颈。

更开放

- 多款互联网中间件已经开源。
- 捐献 Apache 顶级项目，极佳的业界口碑。
- 无捆绑，可以轻松使用开源软件替换。

27.4 典型应用场景

应用发布与管理

在复杂的云环境，应用发布与管理会变得十分复杂。本地开发完成的应用需要逐个部署到服务器，然后登录每一台服务器终端进行应用的发布和部署；后续还会有应用的重启，扩容等。服务器的不断增加对于运维人员将是一个极大的挑战。

针对这个场景，EDAS 提供了一个可视化的应用发布与管理平台，无论集群规模多大，都可以在 Web 控制台上轻松地进行应用生命周期管理。

构建一个分布式系统

当集中式应用转变成分布式系统的时候，系统之间的相互可靠调用一直以来都是分布式架构的难题，比如网络通信，序列化协议设计等很多技术细节需要确定。

EDAS 提供了一个高性能的 RPC 框架，能够构建高可用的分布式系统，系统地考虑到了各个应用之间的分布式服务发现、服务路由、服务调用以及服务安全等细节。

透过数据来剖析系统运行状态

应用开发完毕部署到生产环境之后，通常需要对应用运行时状态进行一些监控，比如 CPU 使用率、机器负载、内存使用率和网路流量等。但此类基础监控通常满足不了业务需求，比如系统运行变慢却无法定位瓶颈所在，或者页面打开出错但是无法排查具体调用错误等。

对此， EDAS 提供了一系列系统数据化运营组件，针对分布式系统的每一个组件和每一个服务进行精细化的监控和跟踪，建立数字化剖析系统，帮助精准地找到系统瓶颈所在。

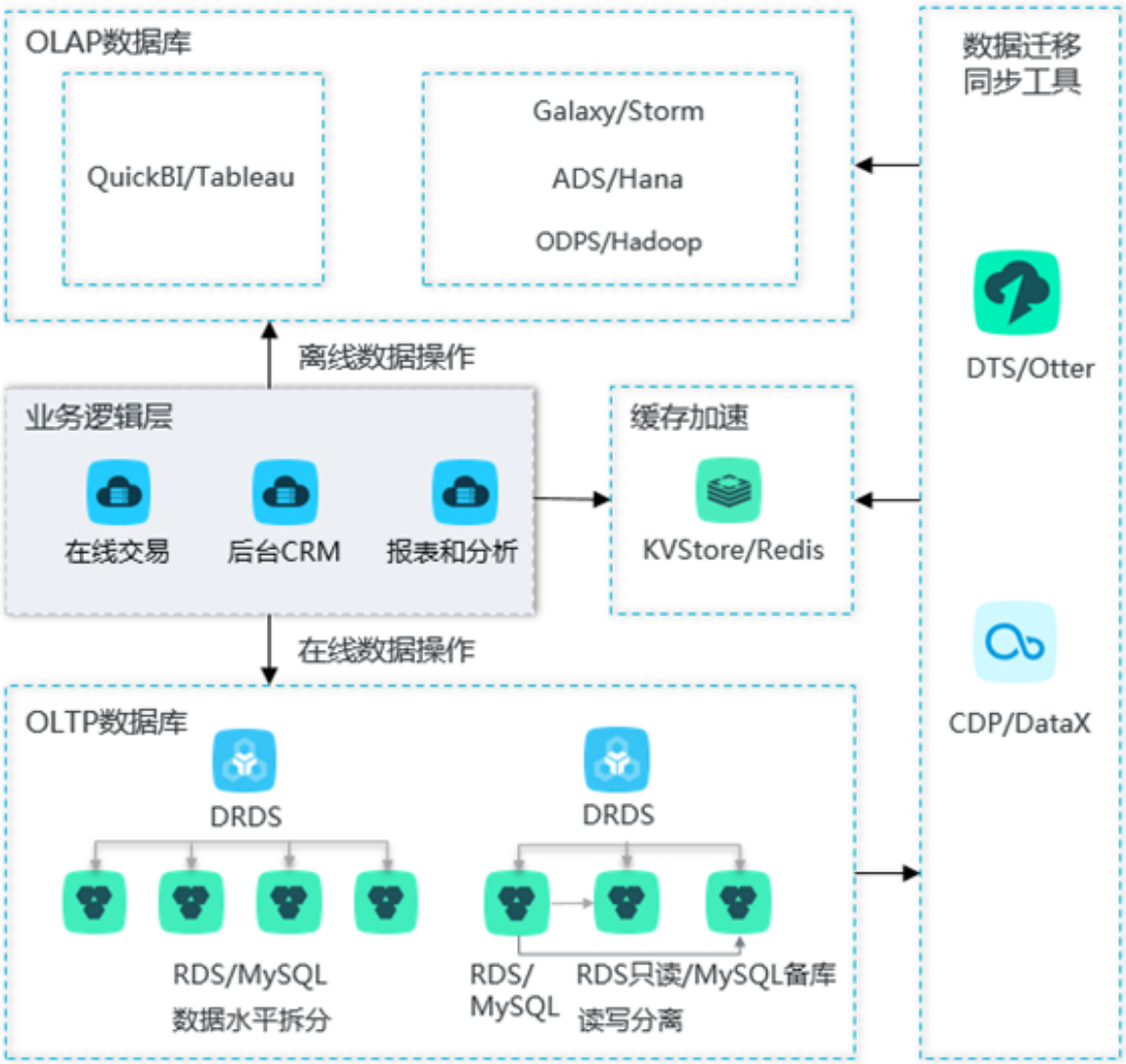
28 分布式关系型数据库DRDS

28.1 产品概述

分布式关系型数据库服务（Distributed Relational Database Service ，简称DRDS）是阿里巴巴集团自主研发的中间件产品，专注于解决单机关系型数据库扩展性问题。DRDS兼容MySQL协议和语法，支持分库分表、平滑扩容、服务升降配、透明读写分离和分布式事务等特性。产品具备轻量(无状态)、灵活、稳定、高效等特点，可以为您提供分布式数据库全生命周期的运维管控能力。

DRDS主要应用场景在大规模在线数据操作上。通过贴合业务的拆分方式，将操作效率提升到极致，有效满足在线业务对关系性数据库要求。

图 87: DRDS产品结构示意图



解决的问题

- **单机数据库容量瓶颈**： 随着数据量和访问量的增长，传统单机数据库会遇到很大的挑战，依赖硬件升级并不能完全解决问题。
- **单机数据库扩展困难**：传统单机数据库容量扩展往往意味着服务中断，很难做到业务无感知或者少感知。
- **单机数据库使用成本高**： 当业务数据和访问量增加到一定量时，单机数据库需要依赖特定的高端存储和小型机设备，成本曲线快速上升。

28.2 功能特性

本文介绍 DRDS 的主要功能和特点。

分库分表

支持RDS/MySQL的分库分表。在创建分布式数据库后，您只需选择拆分键，DRDS 就可以按照拆分键生成拆分规则，实现数据水平拆分。

透明读写分离

通过使用RDS只读实例或者MySQL备机实现读写分离，帮助应用解决事务、只读实例或者备机挂掉、指定主备访问等细节问题。对应用无侵入，在 DRDS 控制台即可完成读写分离相关操作。

数据存储平滑扩容

当出现数据存储容量和访问量瓶颈时，DRDS 支持在线存储容量扩展。扩容无需应用改造，进度支持可视化跟踪。

服务升降配

DRDS 实例可以通过改变资源数量实现服务能力的弹性扩展。

分布式运维指令集

DRDS 提供独有分布式数据库运维指令集，如SHOW SLOW、TRACE、SHOW NODE 等指令，帮助您快速发现和定位问题。

全局唯一数字序列

DRDS 支持分布式全局唯一且有序递增的数字序列。满足业务在使用分布式数据库时对主键或者唯一键以及特定场景的需求。

数据账号权限体系

DRDS支持类似MySQL的账号和权限体系，确保不同角色使用的账号操作安全。

分布式事务

DRDS 结合分布式事务套件 GTS，可以支持分布式事务，保证分布式数据库数据一致性。

监控报警

DRDS 支持对核心资源指标和数据库实例指标的实时监控和报警，如实例 CPU、网络IO、活跃线程等。帮助您实时发现资源和性能瓶颈。

28.3 产品优势

分布式

数据读写存储集群化，不受单机限制，业务使用无连接数限制。

弹性

数据服务可升降配，数据存储白屏化scale-up和scale-out，读写分离线性提升读能力。

高性能

分库分表经典方案让操作聚焦少量数据。多种拆分方式适应数据特点，并具备特定SQL并行执行能力，进一步提升执行效率。

安全

完整的类似MySQL的账号体系，提供具备授权鉴权的Open API，方便集成能力到业务管控中，产品服务支持体系化。

简单易用

兼容 MySQL 协议和大部分MySQL SQL语法，无业务侵入式使用读写分离，全面的运维和监控能力。

成熟度高

参与阿里巴巴全部双十一活动，是阿里巴巴集团接入关系型数据库的标准。

28.4 应用场景

本文重点介绍 DRDS 的典型应用场景。

高并发实时交易场景

面向客户端的电商、金融、O2O、零售等行业普遍存在用户基数大、营销活动频繁的特点，导致核心交易系统数据库响应日益变慢的问题，制约业务发展。 DRDS 提供线性水平扩展能力，能够实时提升数据库处理能力，提高访问效率，峰值 TPS 达150万+，轻松应对高并发的实时交易场景。

海量数据存储访问场景

企业客户随着业务的快速发展，业务数据增长迅猛，会产生超过单机数据库存储能力极限的数据，造成数据库容量瓶颈，限制业务发展。DRDS 可以线性扩展存储空间，目前可以支持 200+ MySQL 实例的单数据库集群，提供 PB 级存储能力。可广泛应用于工业制造、智能家居、车联网等超大规模数据存储访问场景。

高性价比数据库解决方案

对于政务机构、大型企业、银行等行业，为了支持大规模数据存储和高并发数据库访问，传统方案需要强依赖小型机和高端存储等高成本的商业解决方案，以达到服务能力扩展的目的。DRDS 能够利用普通服务器提供阿里巴巴双十一同等处理能力的高性价比国产化数据库解决方案。

低运维成本数据库

初创型企业初期发展阶段技术积累相对比较薄弱，资金投入有限，业务发展快，数据库的稳定性风险高。DRDS 继承了阿里巴巴多年的分布式数据库技术积累，能够提供简单易用的数据库运维系统，降低企业的技术运维成本，赋予企业强大的数据库支撑能力，为企业发展保驾护航。

29 消息队列MQ

29.1 产品概述

消息队列（Message Queue，简称 MQ）是阿里巴巴集团中间件技术部自主研发的专业消息中间件。产品基于高可用分布式集群技术，提供消息发布订阅、消息轨迹查询、定时（延时）消息、资源统计、监控报警等一系列消息云服务，是企业级互联网架构的核心产品。MQ 历史超过 9 年，为分布式应用系统提供异步解耦、削峰填谷的能力，同时具备海量消息堆积、高吞吐、可靠重试等互联网应用所需的特性，是阿里巴巴双11使用的核心产品，每年天猫双十一全天提供 99.99% 可用性。

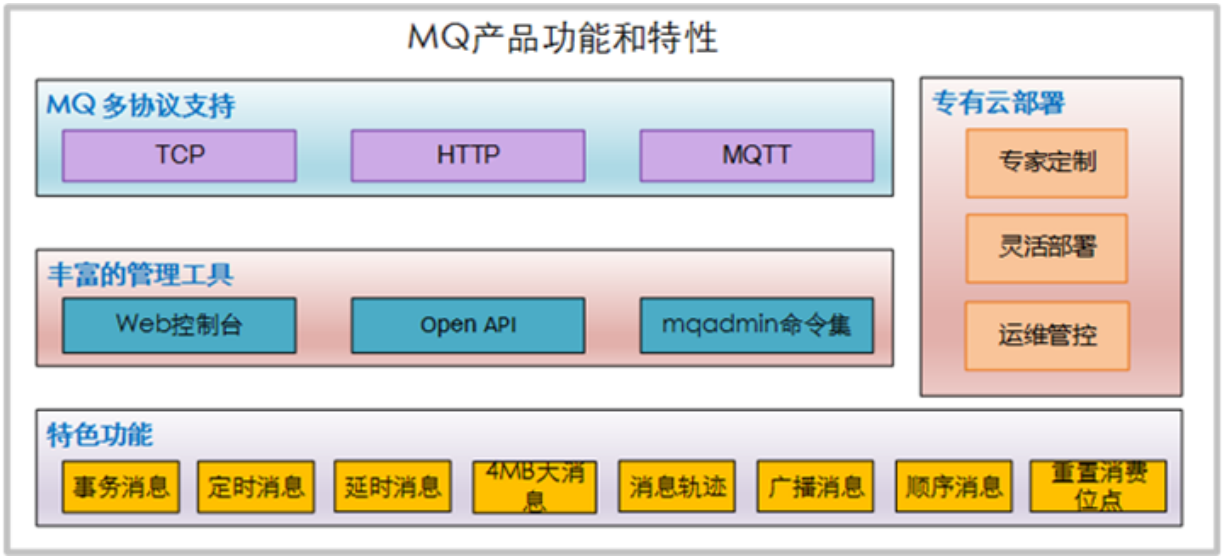
MQ 目前提供 TCP、HTTP、MQTT 等协议层面的接入方式，支持 Java、C++ 以及 .NET 不同语言，方便不同编程语言开发的应用快速接入 MQ 消息云服务。您可以将应用部署在阿里云 ECS、企业自建云，或者嵌入到移动端、物联网设备中与 MQ 建立连接进行消息收发，同时本地开发者也可以通过公网接入 MQ 服务进行消息收发。

图 88: MQ产品示意图

29.2 功能特性

MQ 提供了多种协议和开发语言的接入方式以及多维度的管理工具，同时针对不同的应用场景提供了一系列的特色功能。

功能和特性概览



多协议接入

- 支持 HTTP 协议：支持 RESTful 风格 HTTP 协议完成收发消息，可以解决跨语言使用 MQ 问题。
- 支持 MQTT 协议：支持主动推送模型，多级 Topic 模型支持一次触达 1000万+ 终端，可广泛应用于物联网和社交即时通信场景。
- 支持 TCP 协议：区别于 HTTP 简单的接入方式，提供更为专业、可靠、稳定的 TCP 协议的 SDK 接入。

管理工具

- Web 控制台：支持 Topic 管理、发布管理、订阅管理、消息查询、消息轨迹、资源报表以及监控报警管理。
- Open API：提供 API 允许您将 MQ 管理工具集成到自己的控制台。
- mqadmin 命令集：专有云输出提供一套丰富的管理命令集，以命令方式对 MQ 服务进行管理。

特色功能

- 事务消息：实现类似 X/Open XA 的分布事务功能，以达到事务最终一致性状态。
- 定时（延时）消息：允许消息生产者指定消息进行定时（延时）投递，最长支持 40 天。
- 大消息：目前默认支持最大 256KB 消息，华北 2 地域支持最大 4MB 消息。

- 消息轨迹：通过消息轨迹能清晰定位消息从发布者发出，经由 MQ 服务端，投递给消息订阅者的完整链路，方便定位排查问题。
- 广播消息：允许一个 Consumer ID 所标识的所有 Consumer 都会各自消费某条消息一次。
- 顺序消息：允许消息消费者按照消息发送的顺序对消息进行消费。
- 重置消费进度：根据时间重置消费进度，进行消息回溯或者丢弃堆积消息。

专有云部署

- 专家定制：提供技术方案设计；专家现场技术支持与培训。
- 灵活部署：支持专有云独立部署，同时支持混合云架构。
- 运维管控：专有云支持 mqadmin 命令集、Open API 运维管理工具，方便管控平台集成以及统一运维。

29.3 产品优势

MQ 相比其他消息中间件具备以下优势。

专业

- 消息领域业内专业的消息中间件，产品历史超过 9 年，消息保证不丢，技术体系丰富成熟。
- 阿里巴巴内部产品名 MetaQ、Notify；开源社区产品名为 RocketMQ；产品多次在国内外获奖。
- 阿里内部 1000+ 核心应用使用，每天流转几千亿条消息，经过双11交易、商品等核心链路真实场景的验证，稳定可靠。

高可靠

- 一份消息多份落盘存储，经过严格的断电测试，消息依然保证不丢失。数据可靠性SLA：99.99999999999%。
- 允许海量消息堆积，单个 Topic 即使堆积 100亿+ 条消息，系统的可用性仍然保持不变。
- 默认消息持久化存储 3 天，支持重置消费位点消费 3 天之内任何时间点的消息。

高性能

- 同一网络内，消息传输网络时延在 10 毫秒之内，性能测试下，网卡可被打满。
- 默认单 Topic 发送消息上限为每秒 5000 条，最高可申请扩展至 10W 以上。
- 默认单条消息大小最大支持 256KB，华北2 地域支持 4MB 大消息。

多协议接入

- 支持 HTTP 协议：支持 RESTful 风格 HTTP 协议完成收发消息，可以解决跨语言使用 MQ 问题。

- 支持 MQTT 协议：支持主动推送模型，多级 Topic 模型支持一次触达1000万+ 终端，可广泛应用于物联网和社交即时通信场景。
- 支持 TCP 协议：区别于 HTTP 简单的接入方式，提供更为专业、可靠、稳定的 TCP 协议的 SDK 接入。

独立部署

- 支持专有云独立输出，支持物理机和虚拟机，仅几台机器便可搭建完整消息云服务。
- 专有云配套 mqadmin 命令集和管理类 Open API ，方便运维人员实时监控系统状态。
- 支持混合云架构，允许通过专线的方式接入服务。

29.4 典型应用

MQ 可应用在多个领域，包括异步通信解耦、企业解决方案、金融支付、电信、电子商务、快递物流、广告营销、社交、即时通信、手游、视频、物联网、车联网等。

MQ 可以应用但不局限于以下业务场景：

- 一对多，多对多异步解耦：基于发布订阅模型，对分布式应用进行异步解耦，增加应用的水平扩展能力。
- 削峰填谷：大促等流量洪流突然来袭时，MQ 可以缓冲突发流量，避免下游订阅系统因突发流量崩溃。
- 日志监控：作为重要日志的监控通信管道，将应用日志监控对系统性能影响降到最低。
- 消息推送：为社交应用和物联网应用提供点对点推送，一对多广播式推送的能力。
- 金融报文：发送金融报文，实现金融准实时的报文传输，可靠安全。
- 电信信令：将电信信令封装成消息，传递到各个控制终端，实现准实时控制和信息传递。

30 企业实时监控服务ARMS

30.1 产品概述

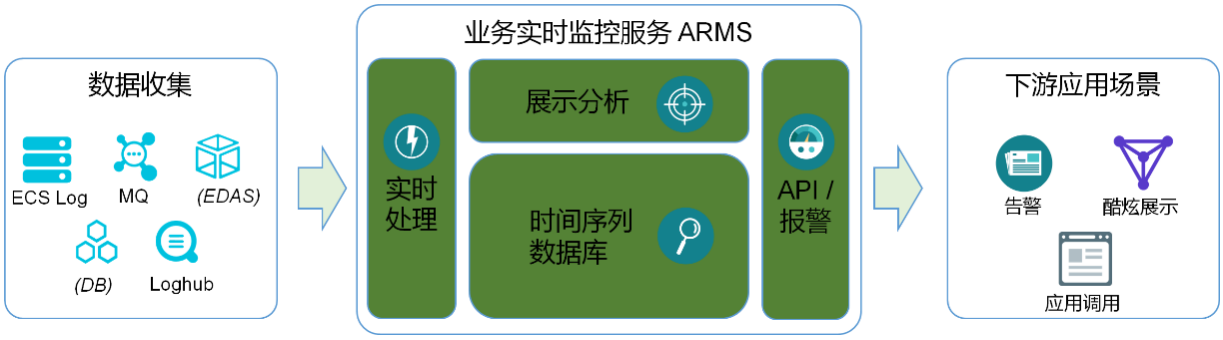
业务实时监控服务（Application Real-Time Monitoring Service，简称 ARMS ）是一款端到端一体化实时监控解决方案的 PaaS 级阿里云产品。

基于ARMS，您可以利用海量数据迅速定制秒级响应的业务监控能力。在技术架构上，ARMS 整合和封装了数据收集、消息通道、实时计算、列式存储，以及在线报表等多种先进互联网技术组件。在用户体验方面，ARMS 为您屏蔽了复杂的监控计算逻辑，针对不同行业提供了快速搭建基于系统监控、商品销售、网站分析等各种场景监控方案的能力。

通过 ARMS，您只需要拖拽式操作三个步骤，即可搭建出一套监控服务：

- 日志和数据接入
- 实时计算任务编排
- 告警和数据集定制

图 89: ARMS示意图



30.2 功能特性

数据接口

提供各类云上产品或企业定制的数据接口。

- **云服务器 ECS 数据源**：应用直接将数据在 ECS 上输出到相应的日志文件上，再通过 ECS 上的 ARMS Agent（也就是 SLS Agent）将日志数据传输到 ARMS 计算节点中。通过 ECS 方式获取数据的应用需要在相应的 ECS 上安装和配置 ARMS Agent，并在某些情况下要求应用就相关数据日志进行改造。

- **LogHub 数据源**：也叫做 SLS 日志源。ARMS 可以通过重用用户在阿里云上的 SLS-日志服务来直接拉取 SLS 中的数据。
- **SDK 数据源**：通过提供 SDK 进行数据收集。可以将 SDK 集成到各类不方便收集日志的终端进行数据收集，典型场景如各类移动终端的数据收集。
- MQ（**敬请期待**）：通过配置 MQ 相应 Topic 的接收端，将指定的数据以消息方式传输到 ARMS 计算节点。

可视化任务编排接口

基于实时计算引擎提供清洗和聚合计算的各类编程 GUI 接口。

- 清洗计算：单/多/顺序分隔符，KV 清洗，JSON 清洗，以及其他各类定制化(如异常堆栈)清洗逻辑。
- 聚合计算：基于各类时间粒度的所有常规聚合计算，TopN，SUM， COUNT (DISTINCT)，MAX，MIN， IDENTITY 等。
- 实时计算的静态或动态 JOIN。
- 其他各类脚本语言（如 Groovy 等）的自定义的清洗和聚合逻辑定制（敬请期待）。
- 报表和报警定制。

拖拽式报表定制

方便快捷地进行各类报警和报表的拖拽式定制开发。

- **业务报警设置**：支持各类内容指标定义、等级区分定义、各类联系人通知方式定义。
- **展示图表定制**：提供时间序列或其他类似维度的全套解决方案，集成柱状、折线、饼图、翻牌器、表格等常见展现形式及大盘配置，提供数据下钻、上钻能力。
- **大盘定制**：支持基于已定义的报警和图标的大盘或报表拖拽式开发。大盘支持在线浏览或离线订阅、下载。

快速业务监控定制能力

提供丰富的各类业务监控模板，提高业务监控定制效率。

- 行业销售监控模板：如基于地域、种类等的销售额统计。
- 其他各类业务或系统监控如关键字监控，系统性能等场景（敬请期待）。

30.3 产品优势

久经考验

经过阿里巴巴集团内部历时 4 年打磨，目前内部支撑超过 1000 个线上任务，涵盖了商品中心、交易、菜鸟、卖家中心、航旅等业务真实监控场景验证。

海量吞吐

在阿里内部的最大集群数据处理量超过每秒 15GB/s ，后端数据结果写入超过 5GB/s 。

秒级延迟

实时聚合统计计算达到秒级延迟，快速响应业务事件。

持续计算能力

自我修复能力，自动修复故障节点，解决流乱序、流延迟造成的数据异常问题。

低学习成本

全图形化配置流程，不需要编写代码即可完成大部分的监控配置任务。

一站式集成

整合和封装了数据收集、消息通道、实时计算、列式存储以及在线报表等多种先进互联网技术组件，为您提供实时监控的一站式解决方案。

30.4 典型应用

ARMS 适合各类监控场景，如基础架构监控、电商监控、移动终端监控等。以下为应用场景参考。

30.4.1 零售行业实时监控方案

以某服装行业龙头公司为例，其IT系统采用基于ARMS的混合云解决方案，搭建零售业务实时监控方案。

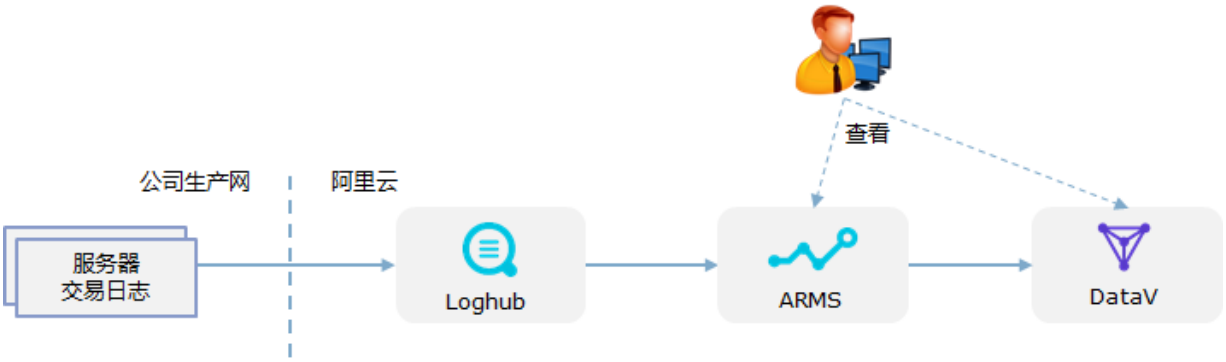
最初业务痛点

- 监控平台使用传统的商业OLAP数据库，License费用昂贵。
- 监控平台在横向扩展以及实时性方面难以满足业务的要求。

基于ARMS的监控方案

总体架构图如下所示。

图 90: 零售行业监控方案



其中：

- 交易日志通过logtail Agent实时上传到阿里云Loghub日志服务。
- ARMS业务实时监控服务对接Loghub，进行计算和存储，并通过自带交互大盘进行销售业务的实时分析查看。
  - 计算编排和存储：从日志中抽取每条交易的详细数据，包括总价、件数，按照交易发生的地点、销售公司的名称，以及客户会员的信息等多个维度进行聚合。
  - 交互展示：基于地域、门店、会员、类目等各个维度的销售状态展示，以及各类下钻场景的分析。
- ARMS数据输出到下游DataV数据可视化组件作大屏展示。

业务价值

基于 ARMS 的 IT 运维监控系统带来的业务价值如下：

- TCO分析成本成百倍下降的同时，满足高时效性的多维度分析。
- 前线销售实时掌控，让销售策略和库存配置策略及时调整成为可能。
- 监控展示方案满足多场合需求：DataV的酷炫大盘用于监控室总体展示，ARMS的交互大盘用于问题的深度排查。

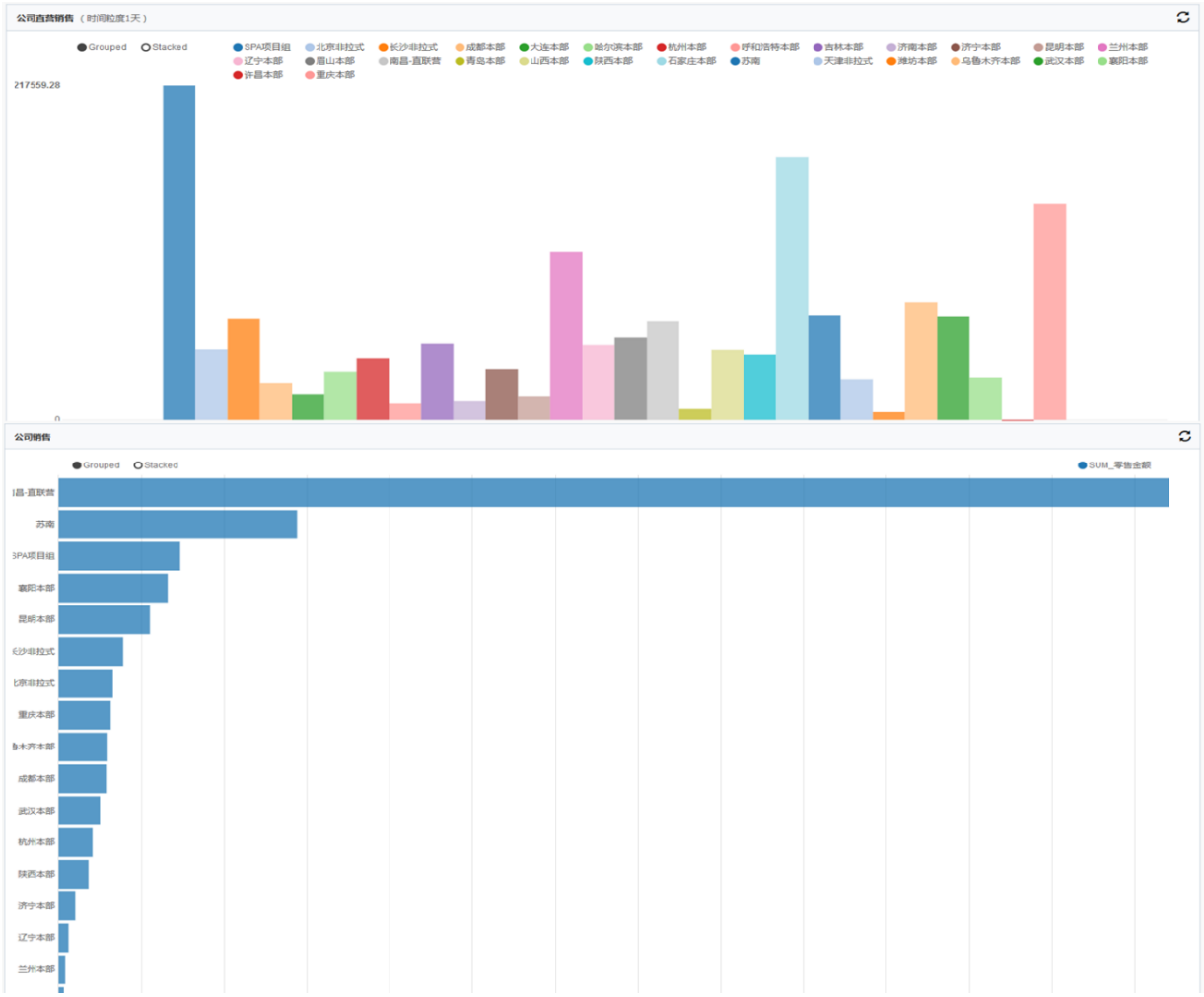
样例展示

监控大盘：

图 91: 监控大盘样例

监控报表：

图 92: 监控报表样例



30.4.2 车联网实时监控方案

以上海某车联网行业方案提供商为例，其IT系统采用基于ARMS的方案进行车辆的在线情况统计。

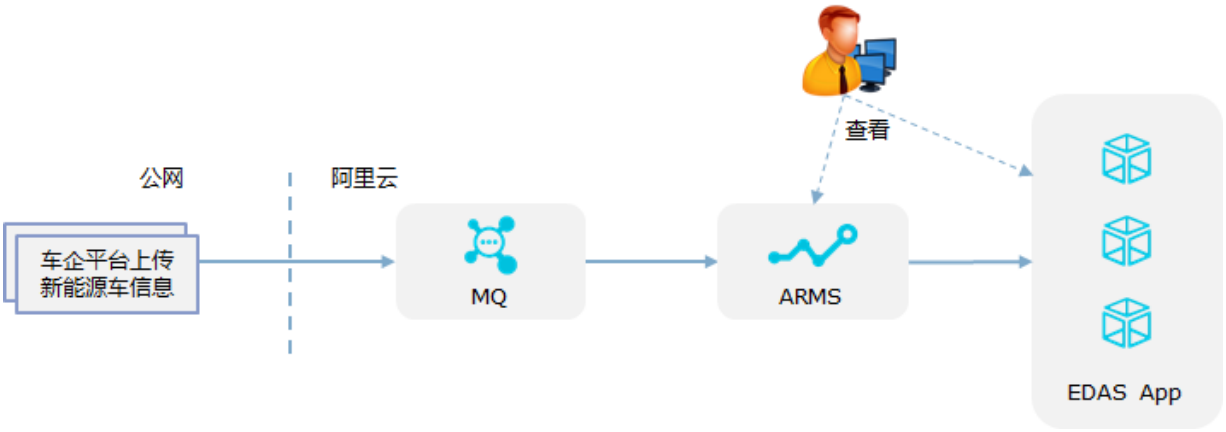
最初业务痛点

由于数据量巨大 (每秒10万级的车辆信息)，无法基于数据库对原始数据进行多维度统计。

基于ARMS的实时监控方案

总体架构如下图所示：

图 93: 车联网监控方案



其中：

- 车企平台把新能源车的实时信息通过 MQ 消息队列上传到阿里云。
- ARMS业务实时监控服务对接MQ，获取所有车辆的在线信息，并进行实时统计和存储。
  - 计算编排和存储：基于车辆上报信息、区域、车辆类型、企业等维度进行在线率、故障率的实时统计，并基于自定义聚合维度将结果进行列式格式的存储。
  - 数据透出：通过数据API对数据进行下游输出。
- 下游EDAS应用通过API进行数据调用，通过用户自身应用进行数据对外展示和分析。

业务价值

基于ARMS的监控方案带来的业务价值如下：

- 实时掌握车辆的运行状态，针对不同车型进行实时故障统计和反馈，质量改进效率大大提升。
- 通过新能源车行驶状态监控，第一时间排查骗补等违规行径。

样例展示

监控展示样例如下图：

图 94: 车联网监控


最近5分钟车辆在线数
在线车数量 12292
最近1一小时车辆在线数
在线车数量 23995
实时大于30s （时间粒度1分钟）
Grouped
Stacked
sum_left30s
sum_over30s
77
0
17:51:00
18:17:00
18:43:00
19:09:00
实时间隔小于等于30s （时间粒度1分钟）
Grouped
Stacked
sum_left30s
77
0
17:51:00
18:17:00
18:43:00
19:09:00

31 全局事务服务GTS

31.1 GTS简介

全局事务服务（Global Transaction Service，简称 GTS）是一款高性能、高可靠、接入简单的分布式事务中间件，用于解决分布式环境下的数据一致性问题。

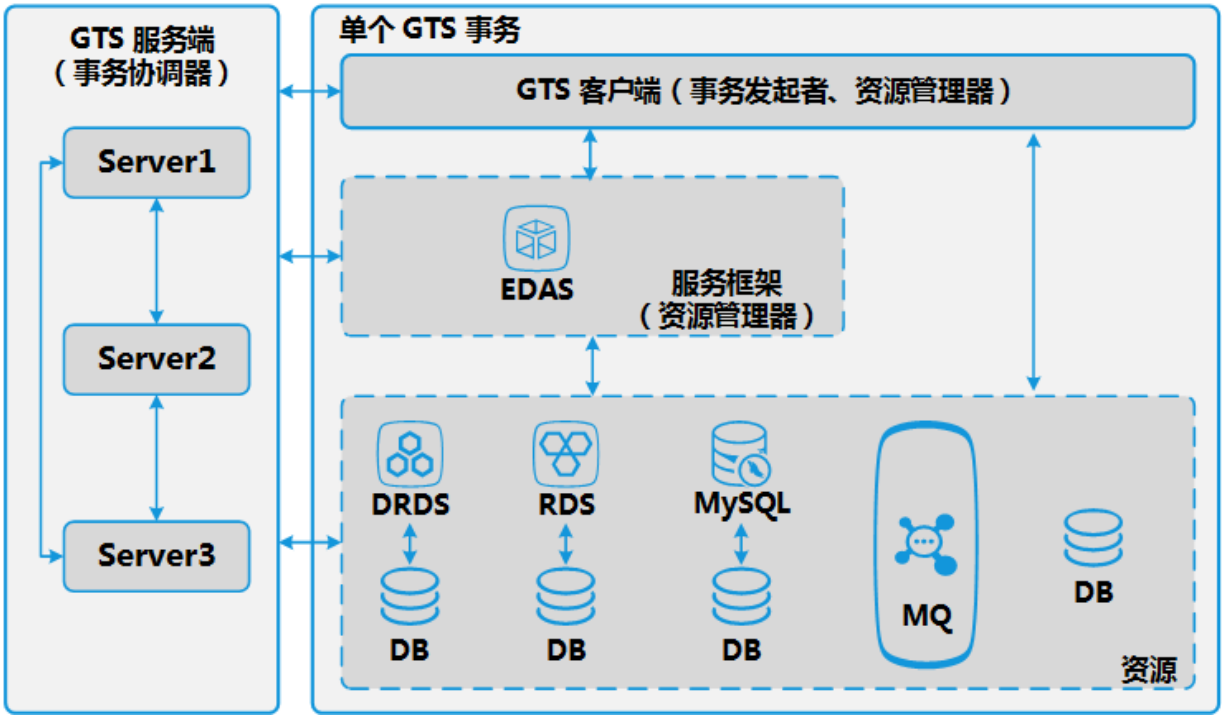
一个完整的业务往往需要调用多个子业务或服务，随着业务的不断增多，涉及的服务及数据也越来越多，越来越复杂。传统的系统难以支撑，出现了应用和数据库等的分布式系统。分布式系统又带来了数据一致性的问题，从而产生了分布式事务。

在单机数据库下很容易维持事务的 ACID（Atomicity、Consistency、Isolation、Durability）特性。在分布式事务中，由于事务发起者、资源管理器、事务协调者及资源分别位于不同的分布式系统的不同节点之上，则很难保持事务的一致性。GTS 产品的优势就在于可以保证分布式系统中的分布式事务的 ACID 特性。

GTS 支持 DRDS、RDS、MySQL 等多种数据源，可以配合 EDAS 和 Dubbo 等微服务框架使用，兼容 MQ 实现事务消息。通过各种组合，您可以轻松地在一个跨多种服务节点的复杂环境里实现分布式事务服务。

GTS 的架构如下图所示：

图 95: GTS 架构



- **GTS 服务端**：即事务协调器。负责分布式事务的推进，管理事务生命周期。服务端的 GTS 相关配置通过内部组件进行下发或同步。
- **GTS 客户端**：即事务发起者。通过事务协调器，开启、提交、回滚分布式事务。同时还包含部分资源管理器组件，负责管理和控制资源，与 GTS 服务器进行交互。
- **服务框架**：GTS 可以和 EDAS 等服务框架配合使用，管理服务框架中的事务。服务框架可以集成资源管理器组件，管理和控制资源。
- **资源**： 包括RDS、DRDS、MySQL以及其它数据库事务，还包括 MQ 消息事务。

31.2 应用场景

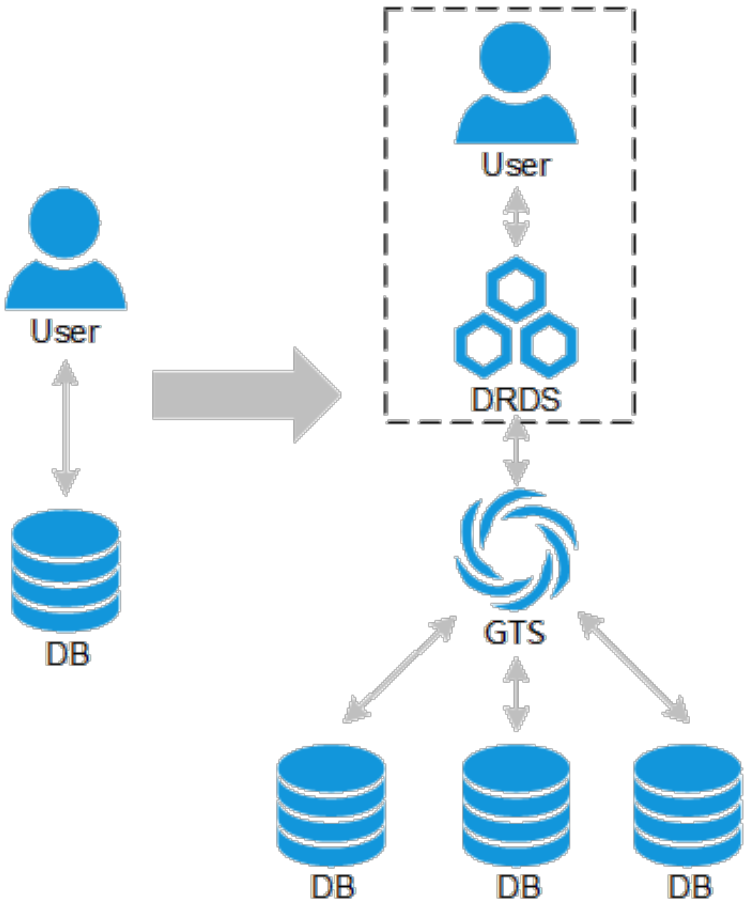
GTS 可应用在涉及数据库操作的多个领域，包括但不限于金融支付、电信、电子商务、快递物流、广告营销、社交、即时通信、手游、视频、物联网、车联网等，典型的应用场景如下：

解决使用DRDS分库分表后产生的跨分库事务问题

DRDS 通过分库分表实现数据水平拆分，来解决单机关系型数据库扩展性问题. 但是原有单库单表进行分库分表后, 单表的数据被分散到多个库的表中, 原来对单表多行数据进行的变更, 可能会变为对多库多表的数据变更，即单机本地事务变成了分布式事务。

DRDS 本身不支持分布式事务，上述场景下再采用原来的单库事务进行操作会导致失败。在 DRDS 中加入 GTS 能够实现这种多个库交易操作的原子性，解决分布式数据库跨库事务的问题

图 96: 跨分库事务场景



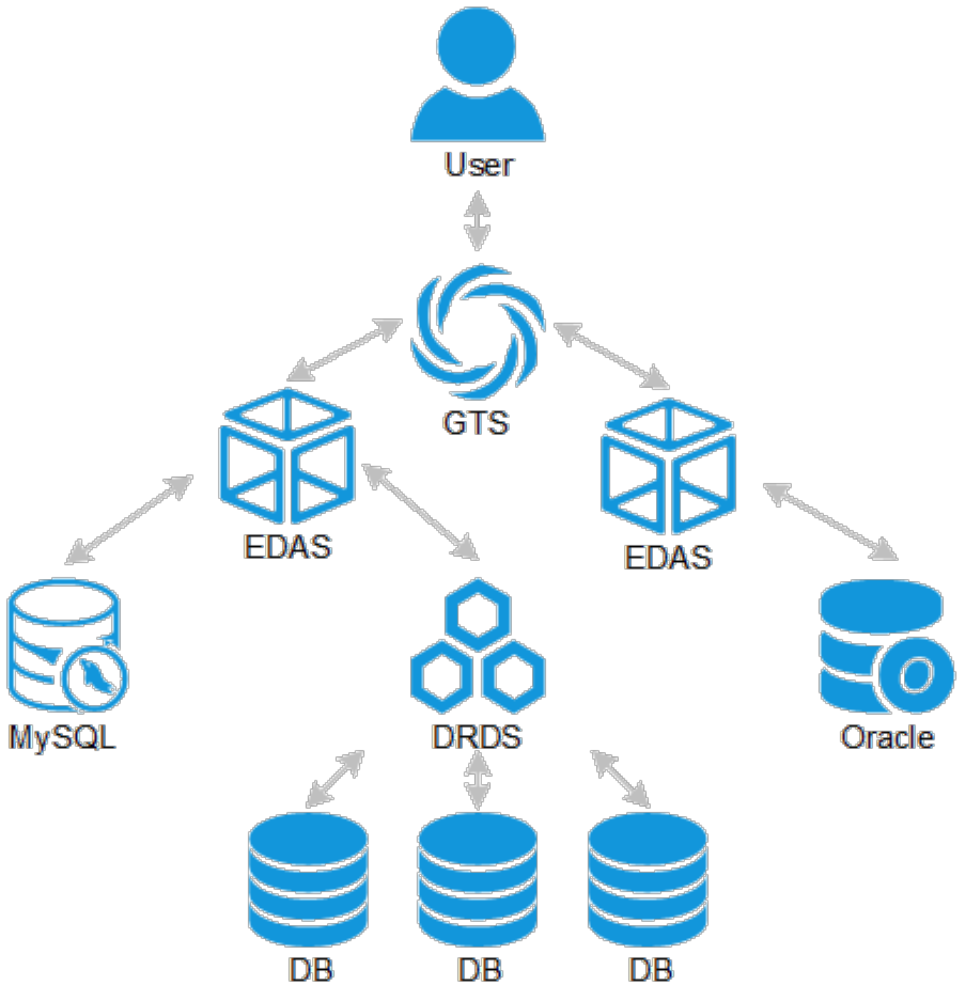
解决跨服务的事务问题

现代 IT 应用中，服务化的模式得到了广泛使用。比如大型电商应用中，为了系统解耦，常常将整个应用划分为多个系统，如商品系统、商家系统、用户系统、账务系统、物流系统等，各个系统会提供各自的服务。

一个简单的商品加入购物车的操作，会调用商品系统的服务来减掉库存，调用购物车系统的服务增加记录，调用结算系统的服务变更待结算金额等等操作。

使用 GTS 可以将调用这些服务的操作加入到一个全局事务中，让他们要么同时成功，要么同时失败，保证了各个系统的数据一致性。

图 97: 跨服务事务场景



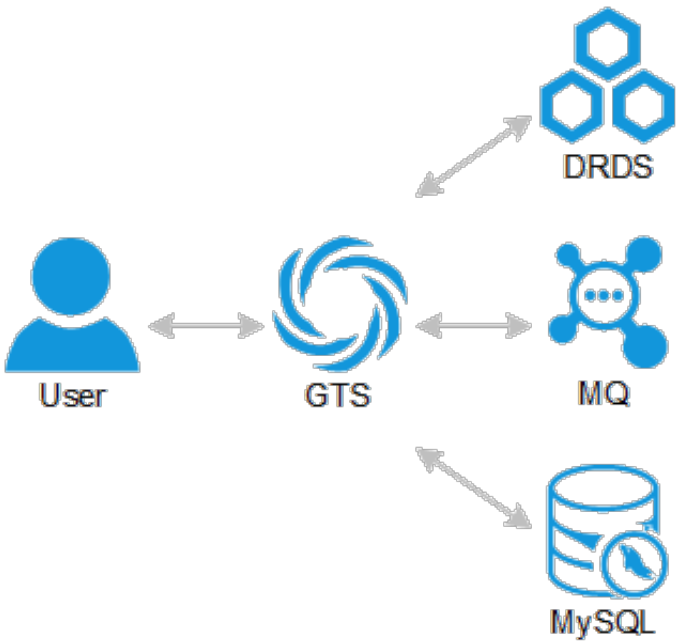
GTS配置MQ可以快速解决事务消息问题

有些系统在使用数据库保证系统内数据一致的同时，也会使用消息队列（MQ）作为和其他系统间的消息传递，完成不同系统间的数据一致。

一个典型的场景，A 系统成功将本地数据 1 保存到数据库后， 通过 MQ 向 B 系统发送一条通知消息， B 系统收到消息后保存与数据 1 关联的数据 2， A、B 两个系统保持数据一致。但是当 A 系统成功保存数据但是未能成功调用消息系统发送通知时，会导致 A 系统中有数据 1 而 B 系统中没有相应的数据 2，即 A、B 两个系统出现数据不一致， 造成系统故障。

对于上述类似场景，GTS 能够将 A 系统向数据库写入数据 1 的本地事务，和通过 MQ 向 B 系统发送通知放到一个全局事务中，保证数据写入则消息一定发出，数据未写入则消息一定不会发出。

图 98: 事务消息场景



31.3 产品功能

GTS 主要能够解决以下三方面的问题。

跨数据库的分布式事务

业务初始阶段往往规模比较小，大多情况下，单库就可以满足需求。经过一段时间后，业务规模也会随之变得大而复杂，会出现分库的情况，这时原有的单机事务往往会变成分布式事务。使用 GTS 可以让您像使用传统单机数据库事务一样，轻松接入分布式事务，且代码改动成本极低。

跨消息和数据库的分布式事务

在某些业务场景中，需要进行多个数据库操作的同时，还会调用消息系统。数据库操作成功、消息发送失败或者反过来都会造成业务的不完整。GTS 可以帮助您轻松统筹消息系统和数据库的事务，将各个资源加入事务范畴。

跨服务的分布式事务

业务完成服务化后，资源与客户端调用解耦，同时又要保证多个服务调用间资源的变化保持强一致，否则会造成业务数据的不完整。GTS 支持跨服务的事务，无论您使用的是 Dubbo 服务框架还是 EDAS 服务框架，都可以接入事务，让您的服务操作没有事务的后顾之忧。

31.4 产品优势

简单易用

应用开发者不再需要考虑复杂的事务问题，仅需简单配置及一句GTS注解（@TxcTransaction）就能轻松实现分布式事务，对已有业务代码无侵入。

节约成本

节省运维成本，避免了分布式场景下产生的数据异常；节省开发成本，像使用单机事务一样使用分布式事务。

高性能

在某些业务场景下，可以达到传统分布式事务性能 10 倍左右；热点数据可以高效处理，无惧数据冲突。

高可靠

中间状态多份落盘存储，经过严格断电测试，严格保证数据一致性。

高可用

GTS 具有同 region 高可用特性，即使突发事件造成集群中某一台机器挂掉，GTS 仍然能够提供原本一半的服务能力。

支持广泛

支持 DRDS、RDS、MySQL、PostgreSQL、H2 等多种数据源，可以配合使用 EDAS、Dubbo 及多种私有 RPC 框架，同时还兼容 MQ 等中间件产品。

31.5 名词解释

GTS包含一些事务相关的术语，通过本文档，可以有一个初步的了解。

事务（Transaction）	事务，是指作为单个逻辑工作单元执行的一系列操作，要么完全执行，要么完全不执行。
分布式事务	事务的发起者、资源及资源管理器和事务协调者分别位于不同的分布式系统的不同节点之上。
事务分支	一个分布式事务可能包含多个分支，只有当所有的分支全部成功时，分布式事务才能成功，一个分支的失败将导致分布式事务的回滚。在 GTS 框架下，分支可能是一个分库上执行的SQL语句，或是一个手动模式分支。

事务边界	分布式事务需要进行开启，在执行结束后需要进行结束（提交或回滚），事务开启和关闭即划定了一个事务边界。
事务模式	GTS 提供的预先定义好的事务模式，不同的事务模式提供了不同的易用性和性能，不同的事务模式组合可解决极度复杂的场景。
事务分组	每个 GTS 应用都需要申请一个事务分组名称，这个唯一名称由客户指定的参数部分以及系统数据组成。
事务分组	每个 GTS 应用都需要申请一个事务分组名称，这个唯一名称由客户指定的参数部分以及系统数据组成。
事务实例名	事务实例名为客户应用中开启事务的代码块的标识，可以帮助用户了解应用的哪部分代码开启了全局事务，此名称可以在控制台上看到。
事务发起者	即 GTS 客户端，通过事务协调协调者开启 / 提交分布式事务。
资源管理器（Resource Manager，简称**RM** ）	事务中的资源管理器抽象，定义了资源参与到事务中的行为，不同事务模式对应不同的资源管理器。
事务管理器	即 GTS Server，负责分布式事务的推进，为客户端发起的分布式事务请求分配全局唯一的事务ID，并记录资源管理器提交的事务分支的状态，最终负责全局事务的提交或回滚。
ACID	数据库事务正确执行的四个特性的缩写。包含：原子性（Atomicity）、一致性（Consistency）、隔离性（Isolation）、持久性（Durability）。一个支持事务的数据库，必需要具有这四种特性，否则在事务过程当中无法保证数据的正确性，交易过程极可能达不到交易方的要求。
两阶段提交	两阶段提交（Two-Phase Commit protocol，2PC）协议是分布式事务的处理协议。
XID	即 GTS 分布式事务的全局事务 ID，GTS 服务会为每一个分布式事务生成一个全局唯一的分布式事务 ID。由于其全局唯一性，我们可以通过 GTS 日志中的 XID 帮助排查问题。
BranchId	即 GTS 分布式事务的分支事务 ID，它是事务分支的唯一标识。XID 和 BranchId 是一对多的包含关系，即一个全局事务可能包含多个事务分支。通过在 GTS 日志中跟踪某个 BranchId，可以帮助排查问题，观察事务分支提交和回滚的原因。

GlobalCommit 全局事务提交，GTS 中用于表示一个全局事务中所有操作都提交了。

GlobalRollback 全局事务回滚，GTS 中用于表示一个全局事务中所有操作都回滚了。

BranchCommit 分支事务提交，GTS 中用于表示一个全局事务的某个分支操作提交了。

BranchRollback 分支事务回滚，GTS 中用于表示一个全局事务的某个分支操作回滚了。

32 云服务总线CSB

32.1 产品概述

云服务总线（Cloud Service Bus，简称 CSB）是一个基于高可用分布式技术构建的服务 API 开放平台。主要针对需要进行管理和控制，包括安全授权、流量限制的内部系统间服务访问和对外服务发布和订阅。可以应用于专有云和混合云，帮助企业打通内外系统，实现跨技术平台、应用系统、企业组织的服务能力互通。各个系统以发布、订阅服务 API 的形式相互开放，围绕 API 互动。CSB 对这些服务 API 进行统一的组织和管控，从而构建企业内部以及与上下游、第三方企业之间，灵活、开放、安全、稳定、高效的业务能力融合、重塑、创新的合作平台。

图 99: CSB产品示意图

云服务总线（CSB）可以应用于以下场景：

- 企业内部不同架构平台的服务之间通过 CSB 实现互通
- 企业内部服务开放给外部，允许通过 CSB 以不同的协议同时访问
- 企业内部通过 CSB 访问外部以不同协议提供的服务
- 不同企业内部的服务之间通过 CSB 实现互通

32.2 功能特性

CSB的功能主要包括以下三个方面。

API服务总线

提供高可用、稳定高效、可线性扩容的服务能力以及丰富全面的访问控制。

- **协议转换**

  支持常用协议服务的接入和开放 (HTTP/HSF/Dubbo/Web Service)，可扩展支持定制化参数以及各种复杂架构的映射。

- **认证鉴权**

  提供灵活的访问鉴权，并可对接企业账号体系。

- **服务控制**

  提供服务的流量控制、黑白名单、端点路由和响应过滤。

API管理组织

提供可灵活定制的 API 全环节管理和组织。

- **服务发布**

  提供服务分组、版本管理、发布审批、导入和导出，以及 API 生命周期管理和组织，支持跨 CSB 实例联动发布已适应多环境的复杂互联场景。

- **服务授权**

  提供完整灵活的服务订阅审批授权机制。

- **服务消费**

  提供服务调用 SDK ，以及服务的消费计量。

API运维监控

提供多样的运维管控工具，用以获取及时详尽的系统状态信息，系统维护方便快捷。

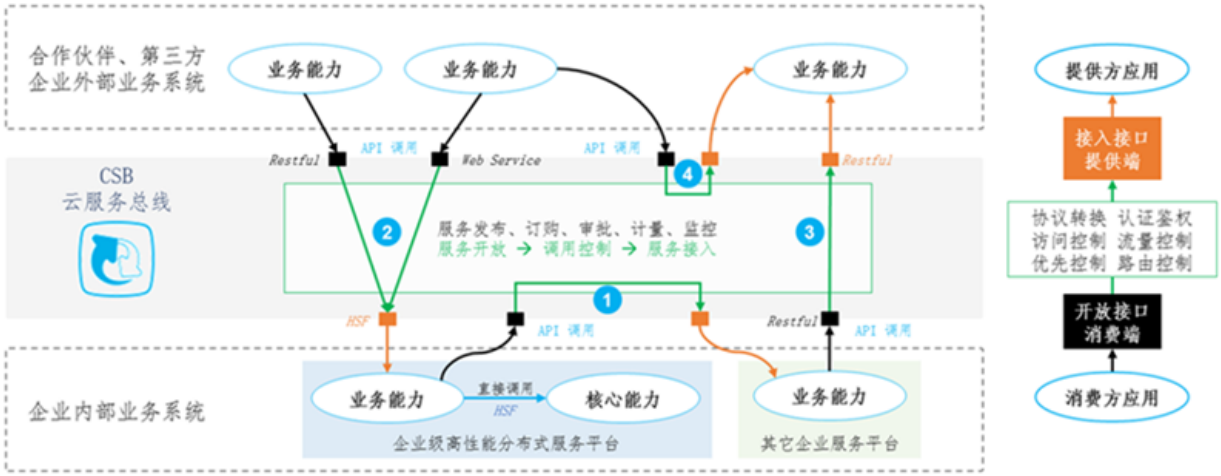
- **日志监控**

  提供服务消费日志、系统监控日志、管理审批日志、服务路由分析、以及矩阵式监测的自动化告警。

- **系统管理**

  提供实例管理、用户管理、角色和权限的灵活定制。

图 100: CSB功能示意图



32.3 产品优势

高性能稳定可靠

基于阿里巴巴内部长期使用与沉淀的高可用高性能分布式集群技术产品构建。稳定可靠地支持大规模请求。

跨协议开放互联

支持不同系统、协议服务安全可控的互联，对接企业原有系统和账号体系；支持参数自定义和各种类型复杂结构的映射，以及多种常用协议的服务API发布和消费。灵活应对各种系统互通需求。

一站式服务管理

完整的API生命周期管理，服务目录，服务授权，用户管理；及时详细的服务质量监控和消费报告。实现能力开放的整体把握和统一管理、组织和互动。

轻量级简便易用

服务发布、配置变更便捷灵活，服务调用简单方便，处理能力按需简便扩容。实现企业业务能力的高效数字化输出和迅捷变更。

32.4 典型应用

云服务总线CSB可以应用于专有云、公共云，以及混合云场景，实现跨系统跨协议的服务互通。

主要针对需要进行管理和控制的系统间服务访问和对外开放场景，包括服务接入授权、流量限制等。

- **异构系统打通**

和 ESB 产品类似，采用不同技术构建的系统，要相互进行服务调用，需要解决服务协议差异的问题。例如用 EDAS 构建的去中心化服务新系统，与 SOAP Web Service 系统的互通。

- **服务组织管控**

  更进一步，实现对服务的规范化和一致化地管理和访问，例如服务的组织分类、搜索查询、审批授权、监控统计等，常见于企业内部业务能力的数字化整合管理。

- **服务开放平台**

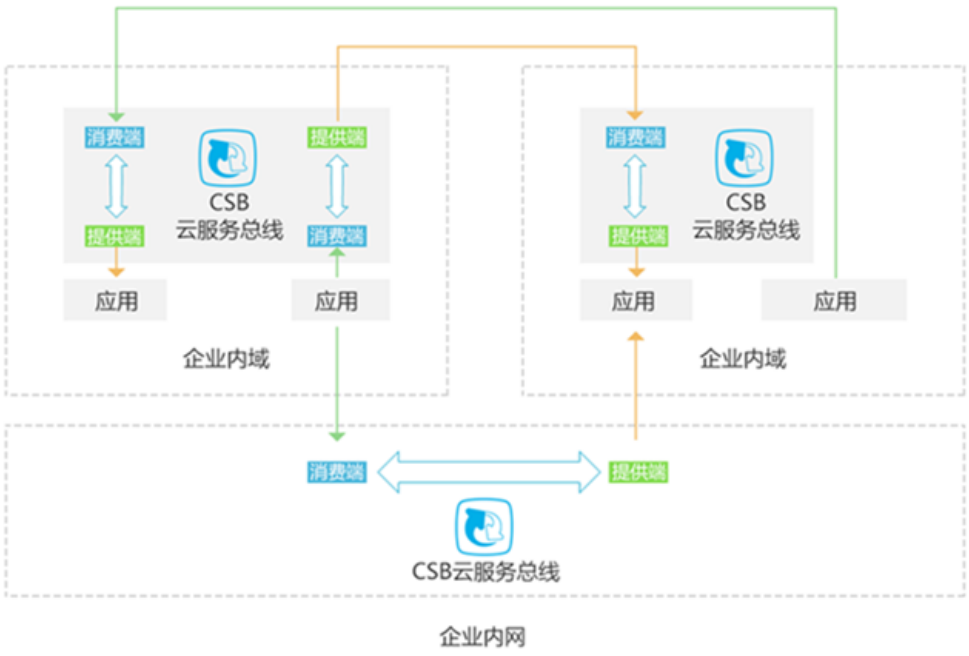
  是组织管控的延伸，但更着眼在不同组织系统能力的融合、转化、创新上，强调的是开放性，把企业的核心业务能力以 API 的方式开放出去，互相借力，推动、开拓业务。

内部互通

企业内部服务能力通过 CSB 可管可控地开放互通。

- 一个域的应用通过另一个域的 CSB 实例访问其内部应用服务。
- 两个域的 CSB 实例构成的桥接通道实现互通。
- 各个域之间通过企业统一的 CSB 实例实现互通。

图 101: 内部互通

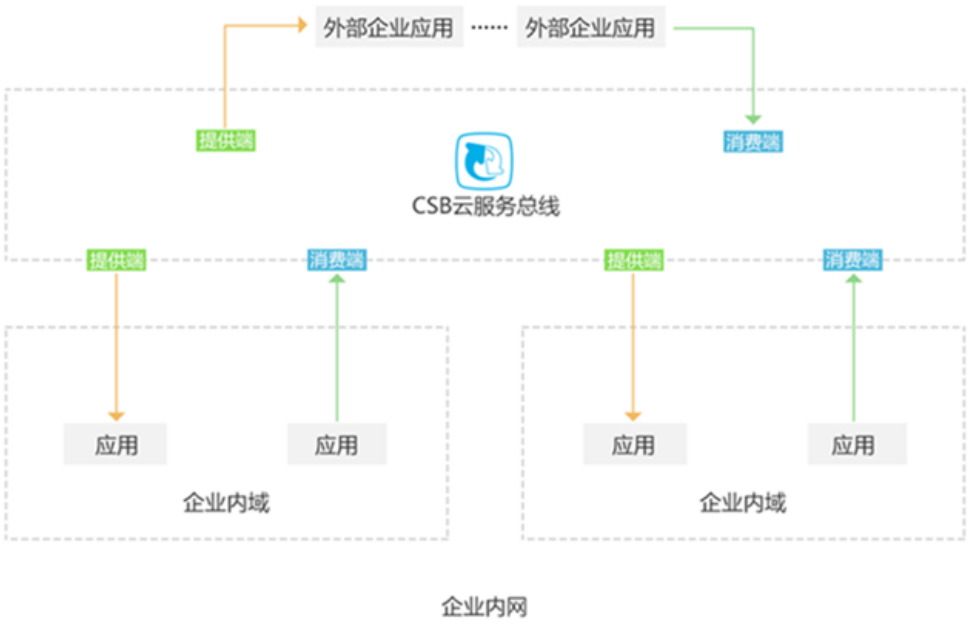


内外互通

企业内部以及与合作伙伴和第三方的系统通过 CSB 可管可控地开放互通。

- 企业内部以及合作伙伴和第三方都可以在 CSB 上发布、订购服务。
- 各自的服务通过授权做访问控制（包括流量控制等）。
- 所有开放的服务在 CSB 上形成由企业自主管理的综合服务集市。

图 102: 内外互通

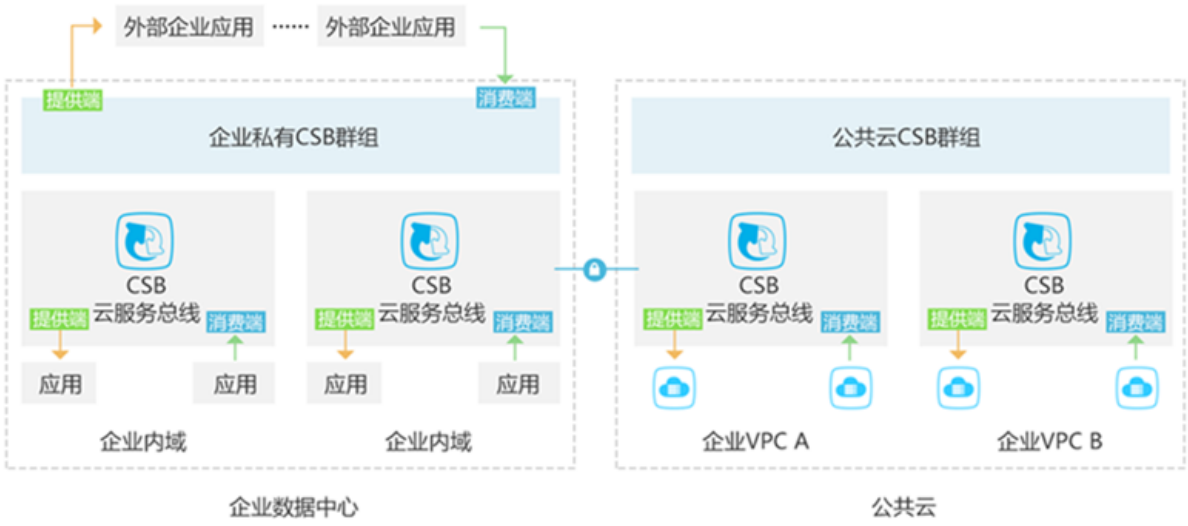


混合云、多群组

企业内部以及企业在公共云上的专有域（VPC 虚拟专有云）之间互通。

- 企业内部某个 CSB 与其在公共云上的某个 VPC 内的 CSB 通过专线互通。
- 公共云上的多个企业 VPC 之间通过各自的 CSB 互通。
- 每个 CSB（实例）上的用户和服务由 CSB 实例拥有者自主管理。

图 103: 混合云、多群组


外部企业应用
外部企业应用
提供端
消费端
企业私有CSB群组
公共云CSB群组
CSB
云服务总线
提供端
消费端
应用
应用
应用
应用
企业内域
企业内域
企业VPC A
企业VPC B
企业数据中心
公共云

33 MaxCompute

33.1 产品概述

大数据计算服务（MaxCompute）： 是阿里巴巴内部发展的一个高效能、低成本，高可用的**EB级**大数据计算服务，在集团内部每天处理超过EB级的数据量。

MaxCompute是面向大数据处理的分布式系统，主要提供结构化数据的存储和计算，是阿里巴巴云计算整体解决方案中最核心的主力产品之一。

多租户、数据安全、水平扩展等特性是MaxCompute的核心设计目标，采用抽象的作业处理框架为不同用户对各种数据处理任务提供统一的编程接口和界面。

MaxCompute产品特点如下：

- 采用分布式架构，规模可以根据需要平行扩展。
- 自动存储容错机制，保障数据高可靠性。
- 所有计算在沙箱中运行，保障数据高安全性。
- 以RESTful API的方式提供服务。
- 支持高并发、高吞吐量的数据上传下载。
- 支持离线计算、机器学习两类模型及计算服务。
- 支持基于SQL、Mapreduce、Graph、MPI等多种编程模型的数据处理方式。
- 支持多租户，多个用户可以协同分析数据。
- 支持基于ACL和policy的用户权限管理，可以配置灵活的数据访问控制策略，防止数据越权访问。

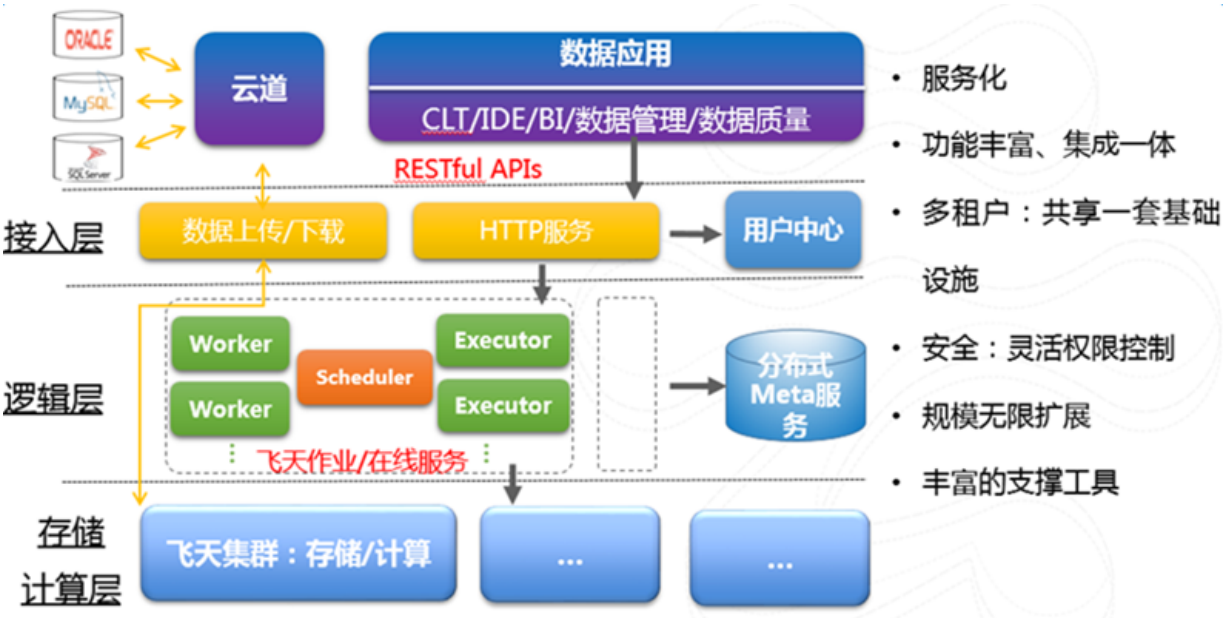
33.2 产品架构

MaxCompute主要服务于批量结构化数据的存储和计算，可以提供海量数据仓库的解决方案以及针对大数据的分析建模服务。

随着社会数据收集手段的不断丰富及完善，越来越多的行业数据被积累下来 。数据规模已经增长到了传统软件行业无法承载的海量数据（百GB、TB、乃至PB）级别。在分析海量数据场景下，由于单台服务器的处理能力限制，数据分析者通常采用分布式计算模式。但分布式的计算模型对数据分析人员提出了较高的要求，且不易维护 。使用分布式模型，数据分析人员不仅需要了解业务需求，同时还需要熟悉底层计算模型。

MaxCompute 的目的是为用户提供一种便捷的分析处理海量数据的手段 。用户可以不必关心分布式计算细节，从而达到分析大数据的目的。

图 104: MaxCompute架构



33.3 功能特性

33.3.1 Tunnel

33.3.1.1 Tunnel特点

- 数据进出MaxCompute的通道。
- 高并发上传下载。
- 服务能力水平扩展。
- 可支持每天1P吞吐量。
- 分为批量及实时两种模式。
- 实时模式支持pub/sub（发布/订阅）模型。
- 基于MaxCompute Tunnel的工具有TT, CDP、Flume、Fluentd等。
- 支持对表的读写，不支持视图。
- 写表是追加（Append）模式。
- 并发以提高总体吞吐量。
- 避免频繁提交。
- 上传数据时，目标分区必须存在。

- 实时上传模式。

33.3.1.2 Tunnel数据上传下载

Tunnel命令

```
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;
```

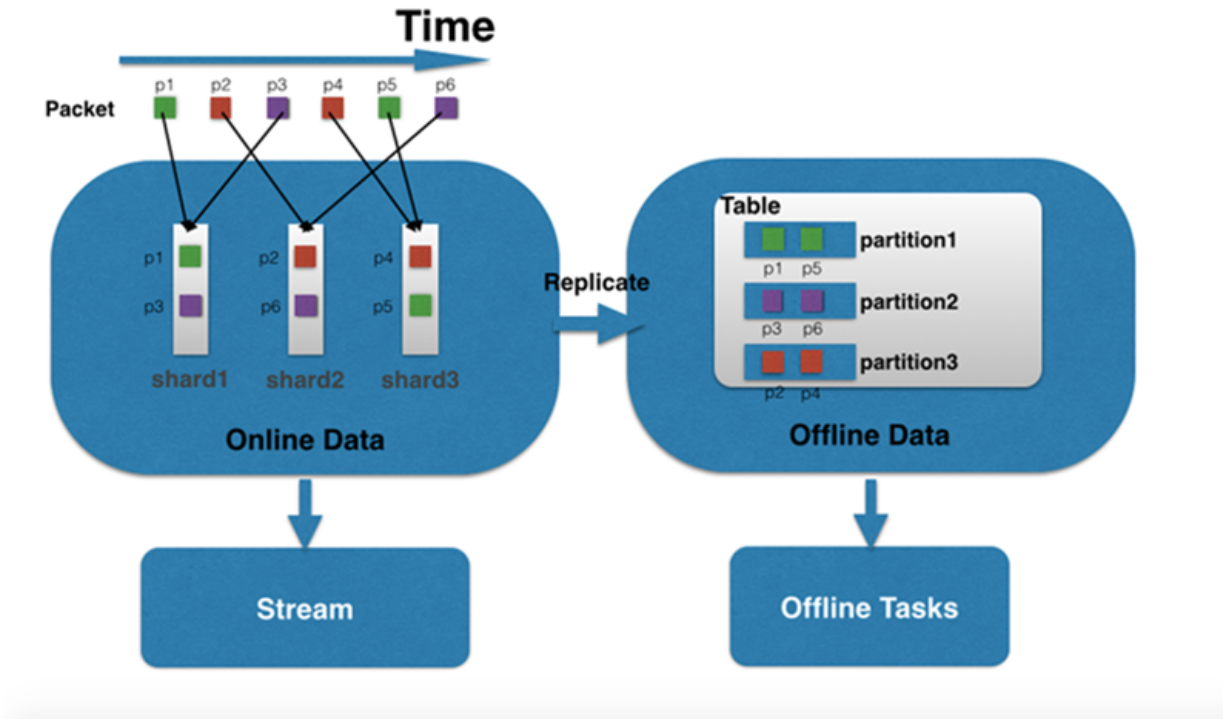
使用说明

- 基于Tunnel SDK的一个命令行工具，可用于本地文本文件上传到MaxCompute或下载表数据到本地。
- 表的分区要先建好。
- DataX，CDP，TT等已基于Tunnel实现了更为完善的工具，可用于支持MaxCompute与关系数据库的数据交互。
- 日志数据可以使用Flume、Fluentd工具导入。
- 特殊的场景用户可以基于Tunnel实现自定义的工具。

实时上传

- 小batch上传。
- 高QPS。
- 毫秒级latency。
- 可订阅。

图 105: 实时上传



33.3.2 SQL

33.3.2.1 SQL特点

- 适用于大数据量处理（T级别到P级别）。
- 延迟比较高，每个SQL的运行时间在几十秒到小时级别。
- 语法类似Hive的HQL，是在标准SQL的基础上有所扩展。
- 没有事务，没有主键。
- 不支持UPDATE，DELETE。

33.3.2.2 与开源对比

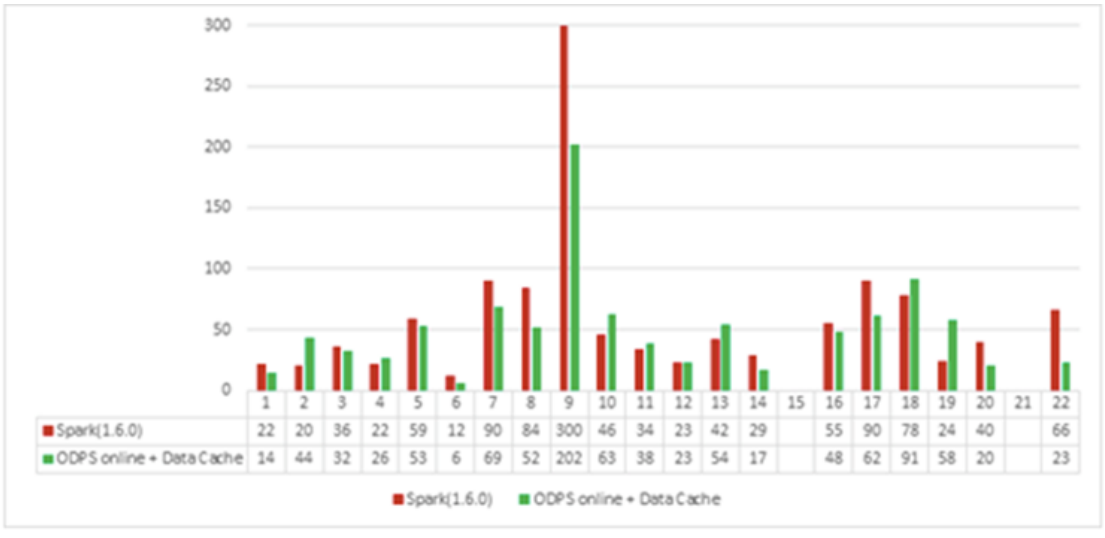
- TPCH 1 TB 数据，MaxCompute与Hive（Apache-hive-1.2.1-bin +TEZ-ui-0.70. with CBO）相比， MaxCompute提升95.6%。

图 106: MaxCompute 2.0 VS Hive

TPCH 1 TB 数据，ODPS 2.0 VS Hive(Apache-hive-1.2.1-bin +TEZ-ui-0.70. with CBO)
ODPS 提升95.6%

- 标准TPCH 450GB数据，MaxCompute与Spark SQL（1.6.0 latest release）相比，MaxCompute提升 17.8%。

图 107: MaxCompute 2.0 VS Spark SQL



标准 TPCH 450GB 数据，ODPS 2.0 VS Spark SQL (1.6.0 latest release),
ODPS 提升 17.8%

33.3.3 MapReduce

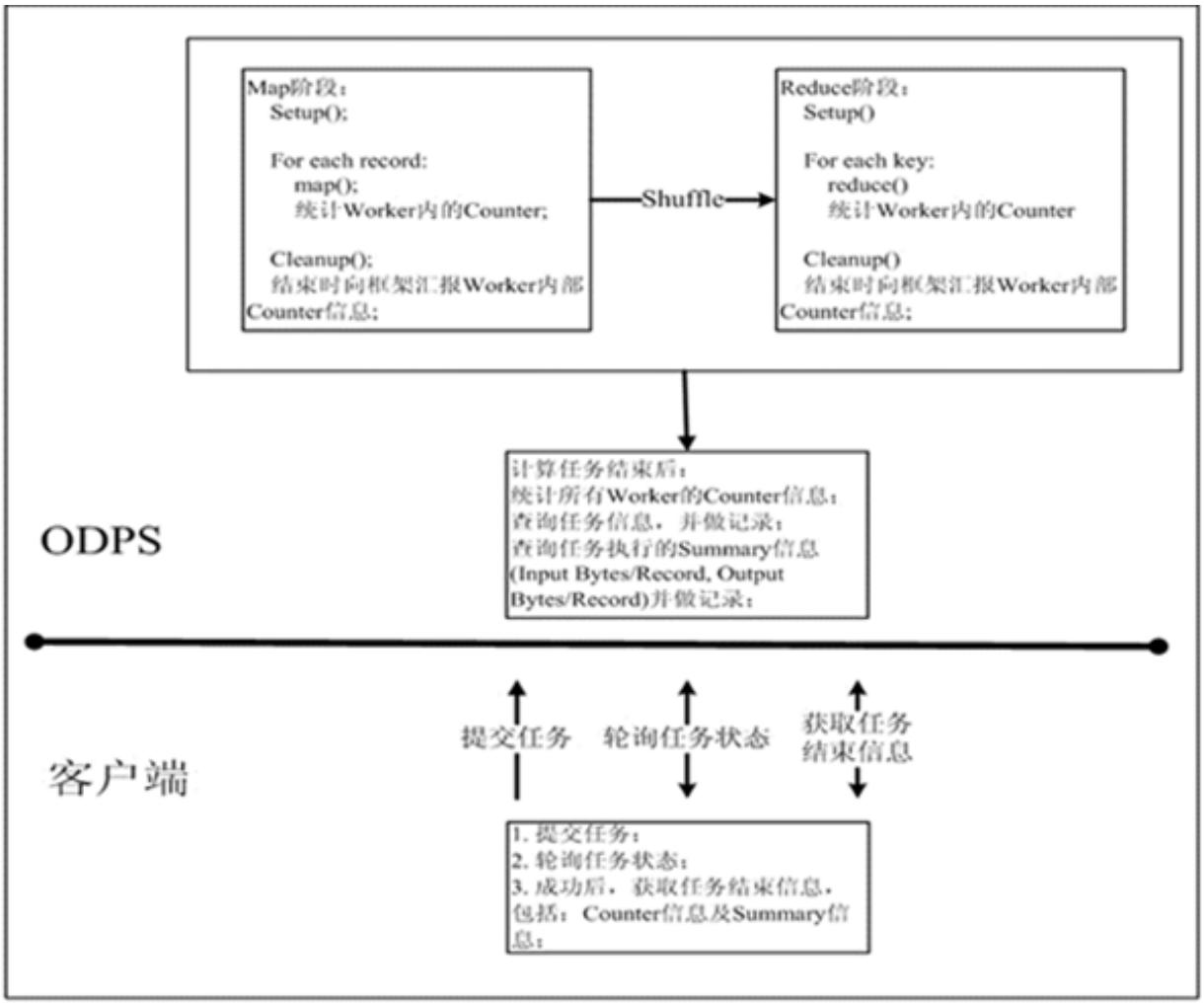
33.3.3.1 MapReduce特点

- 输入输出仅支持MaxCompute内置类型。

- 可以输入多表，输出多表或到不同分区。
- 可以读资源（Resource）。
- 不支持输入view。
- 编译MR程序需要JDK 1.6环境。
- 受限的沙箱安全环境。

33.3.3.2 MaxCompute MR过程

图 108: MR过程



33.3.3.3 Hadoop MR VS MaxCompute MR

表 38: 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）

图 109: MR

```
@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);
  }
```

33.3.4 Graph

33.3.4.1 Graph特点

- 图计算编程模型（类似Google Pregel）。
- 数据装载到内存，在迭代次数较多时优势明显。
- 可用于开发机器学习算法。
- 可以支持100亿顶点和1500亿边的规模。
- 典型应用。
  - Pagerank。
  - K-Means聚类。
  - 一度、二度关系，最短路径等。
- Graph作业处理数据是一个图。
- 原始数据存储在Table中，用户自定义的GraphLoader将Table中的数据加载为点和边。
- 迭代计算。

33.3.4.2 Graph关系网络模型

关系网络引擎为关系型数据的挖掘提供了多种针对业务优化的关系网络模型，帮助用户快速实现对关系网络数据的复杂挖掘。

社区发现

- 引擎输入（关系数据）。
- 引擎输出（ID，社区ID）。
- 计算逻辑：利用全局网络连接最优找到N个社区，社区内部足够紧密，社区间足够稀疏。

半监督分类

- 引擎输入（问题ID）。
- 引擎输出（疑似问题ID, 权重）。
- 计算逻辑：利用已有问题ID（某一类或多类），根据整个网络连接关系判断全局中疑似此类（或多类）的ID以及权重。

孤立点检测

- 引擎输入（关系数据）。
- 引擎输出（孤立点ID, 权重）。
- 计算逻辑：利用关系网络中连接关系判断是否有相对孤立的节点并输出。

关键点挖掘

- 引擎输入（关系数据）。
- 引擎输出（关键点ID, 类别）。
- 计算逻辑：利用关系网络中连接关系计算网络中关键类型节点（中间度、影响力、媒介能力）。

N度关系

- 引擎输入（关系数据）。
- 引擎输出（可检索的关系网络）。
- 计算逻辑：利用关系网络中连接关系整理多维关系，并建立索引方便查询某ID的具体关联情况。

33.3.5 系统安全

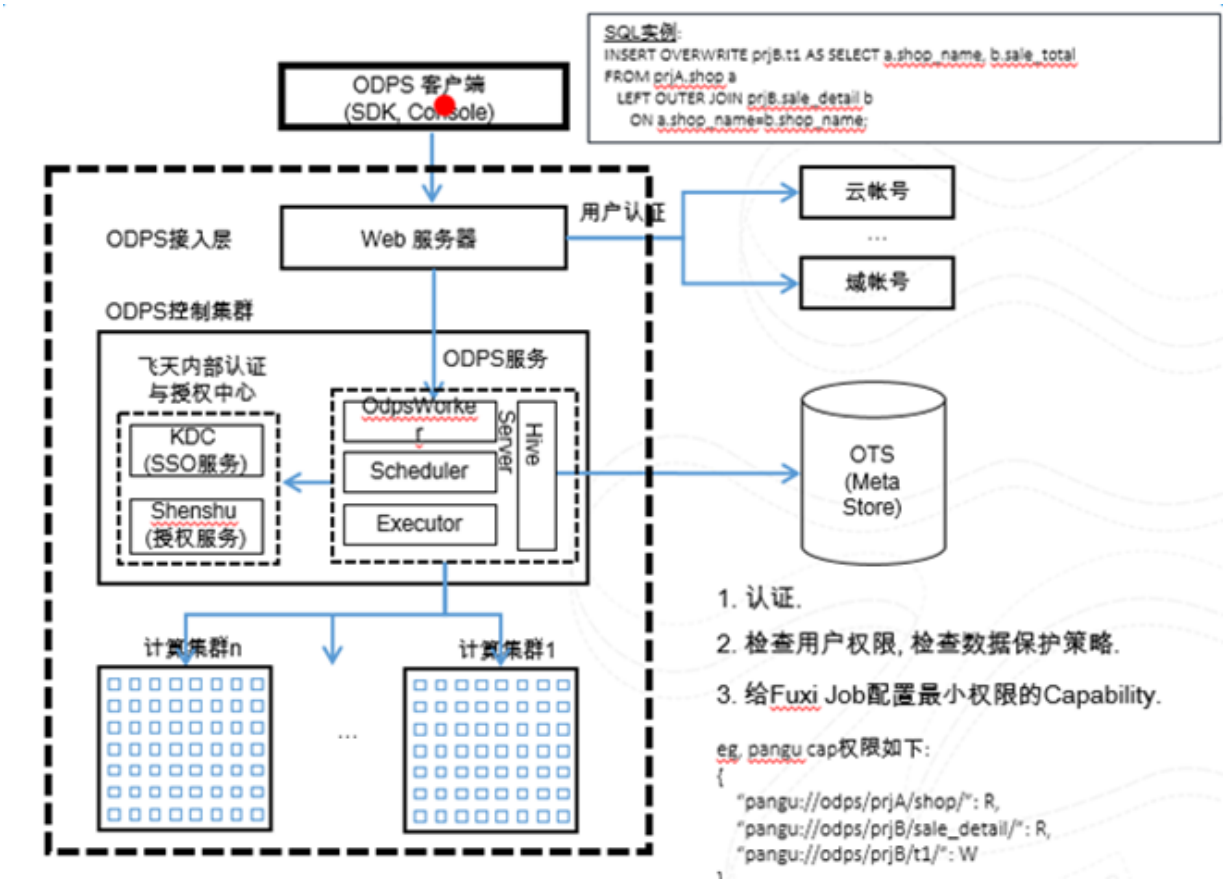
33.3.5.1 安全特点

- 支持多租户的使用场景，同时满足多用户协同、数据共享、数据保密和安全的需要。
- 用户访问需要认证，用户操作需要鉴权，所有操作记录审计日志。

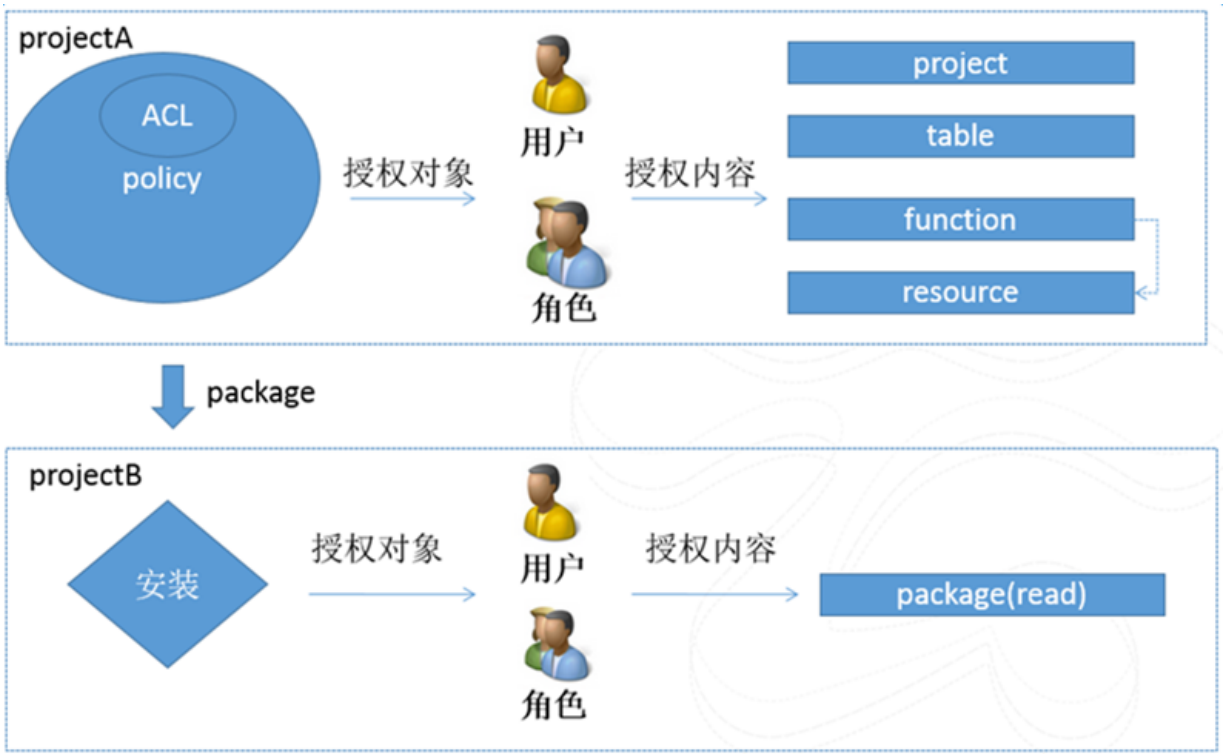
- 支持ACL授权、policy授权、角色授权、跨project授权多种权限管理方法，满足多种场景的需求。
- 同时提供DAC和MAC的安全管理方案，满足对于部分敏感数据的管理需求，可以提供精确到列级别的数据管理。
- 对于安全等级较高的数据，提供项目保护模式，防止数据泄露。
- 所有计算在受限的沙箱中运行，多层次的应用沙箱、系统沙箱配合请求鉴权管理机制，保证数据的安全。

33.3.5.2 安全架构

图 110: 安全架构



33.3.5.3 权限管理模型



33.3.5.4 ACL授权

- 对已存在的对象和用户进行授权管理。
- 用户必须存在于project中。
- 角色（role）是一组权限的集合。
- 系统预设admin角色。
- 对象删除时，所有相关的ACL也被删除。

```
grant CreateTable, CreateInstance, List on project myprj to user Alice;
-- 增加用户权限。
```

```
grant worker to aliyun$abc@aliyun.com;
-- 加用户进角色。
```

```
revoke CreateTable on project myprj from user Alice;
```

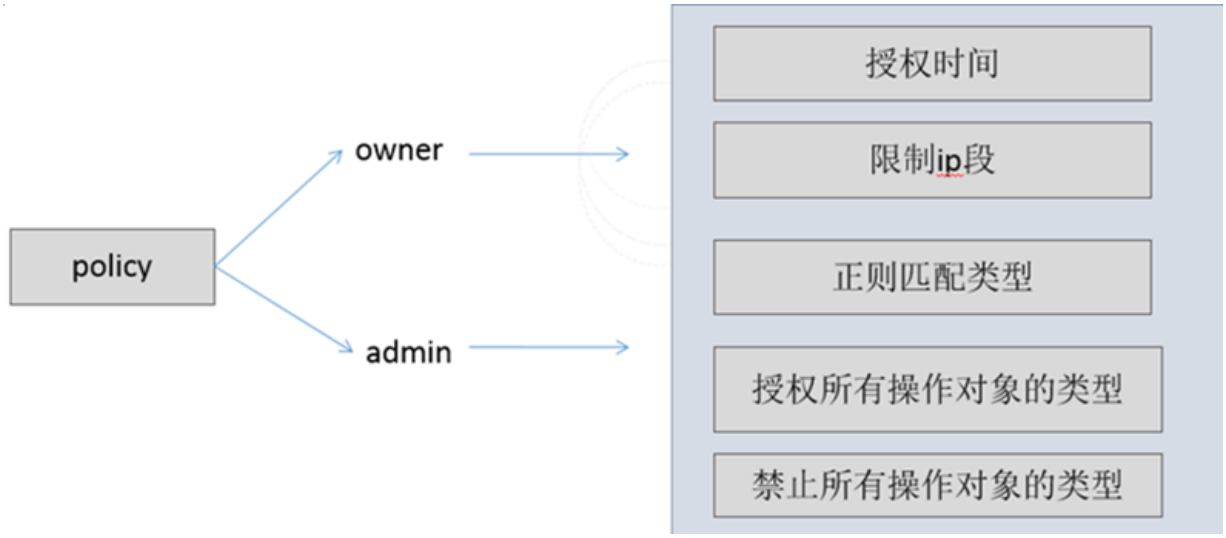
```
-- 删除用户权限。
```

33.3.5.5 Policy授权

Policy授权策略描述了授权的具体内容，它是一种规则，比如：允许所有用户读所有开发环境的表、允许A用户从某个IP访问数据；当管理的数据对象和用户数非常多时，可以极大的减轻管理负担；存在于project和角色上，分为project Policy与role Policy，不会随表的删除而消失。

```
GET POLICY
PUT POLICY <policyFile>;
GET POLICY ON ROLE <roleName>;
PUT POLICY <policyFile> ON ROLE <roleName>;
```

图 111: policy授权




```
{
"Version": "1",
"Statement":
 [{
 "Effect":"Allow",
                    "Principal":"ALIYUN$alice@aliyun.com",                                    "Action":
["odps:CreateTable","odps:CreateInstance","odps:List"],          "Resource":"acs:odps:*:projects/
test_project", "Condition":{    "DateLessThan": { "acs:CurrentTime":"2013-11-11T23:59:59Z" },
  "IpAddress": { "acs:SourceIp":"10.32.180.0/23" }
 }
 },
{
 "Effect":"Deny",
 "Principal":"ALIYUN$alice@aliyun.com",
 "Action":"odps:Drop", "Resource":"acs:odps:*:projects/test_project/tables/*"
}
```

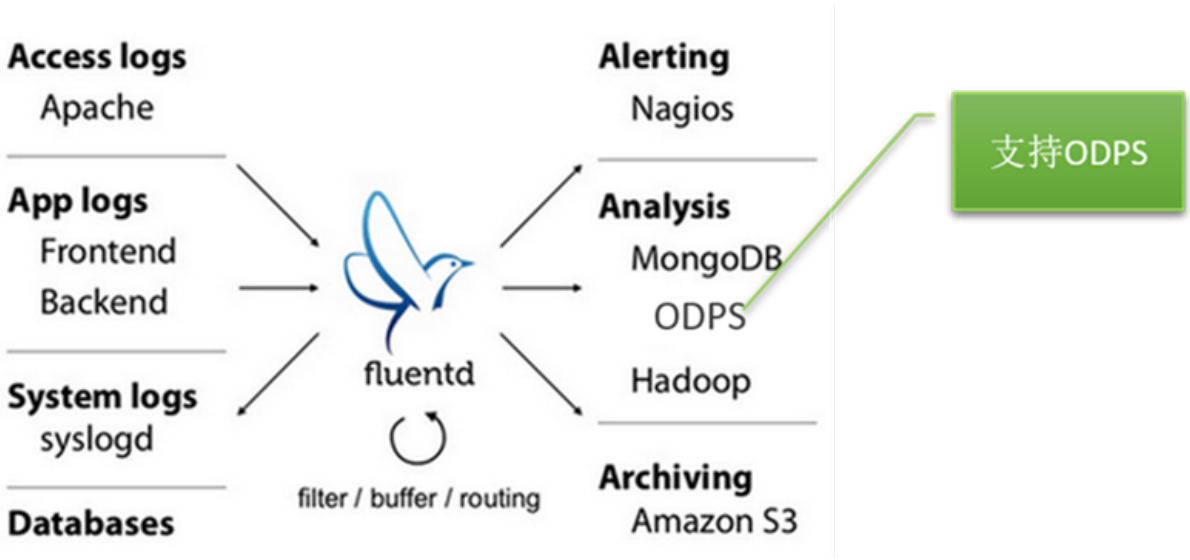
```
]}
```

33.3.6 开源生态

33.3.6.1 日志导入工具-Fluentd

官网：*http://www.fluentd.org*

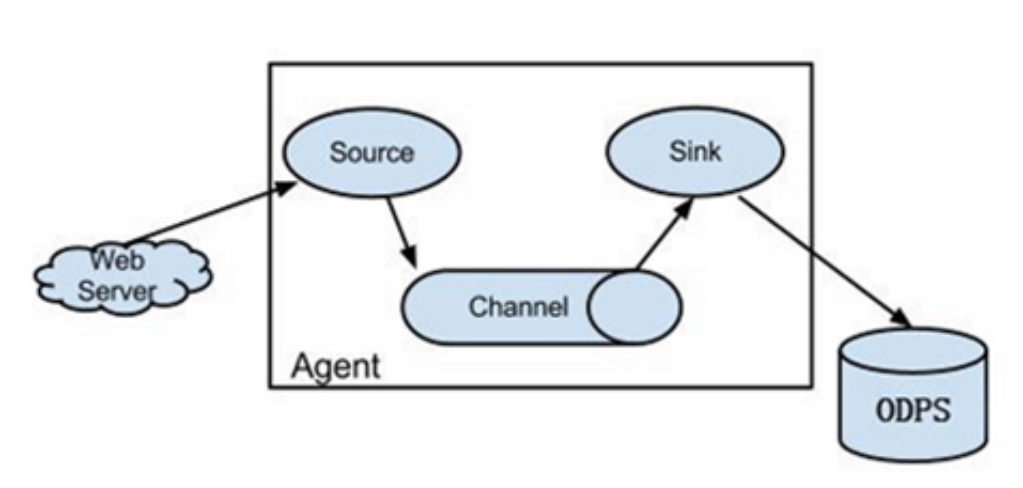
图 112: fluented



33.3.6.2 日志导入工具-Flume

官网：*http://flume.apache.org/*

图 113: flume



33.3.6.3 开源代码

- MaxCompute R插件： *https://github.com/aliyun/aliyun-odps-r-plugin*
- MaxCompute Sqoop: *https://github.com/aliyun/aliyun-odps-sqoop*
- MaxCompute ogg: *https://github.com/aliyun/aliyun-odps-ogg-plugin*
- MaxCompute eclipse插件： *https://github.com/aliyun/aliyun-odps-eclipse-plugin*
- MaxCompute JDBC Driver: *https://github.com/aliyun/aliyun-odps-jdbc*
- MaxCompute Python SDK: *https://github.com/aliyun/aliyun-odps-python-sdk*
- MaxCompute Java SDK: *https://github.com/aliyun/aliyun-odps-java-sdk*

33.4 产品优势

国内唯一的大数据云服务平台，真正的数据分享平台

- 数据仓库、数据挖掘、数据分析、数据分享。
- 阿里集团内部使用的统一数据处理平台，支持阿里贷款、数据魔方、DMP（阿里巴巴广告联盟）、余额宝等多款产品。

MaxCompute规模

- 单一集群规模可以达到10000+服务器（保持80%线性扩展）。
- 单个MaxCompute部署可以支持100万服务器以上，无限制（线性扩展略差），支持同城、异地多数据中心模式。
- 10000+用户数，1000+项目应用、100+部门（多租户）。
- 100万以上作业（目前单日平均提交任务），20000以上并发作业。

33.4.1 存储

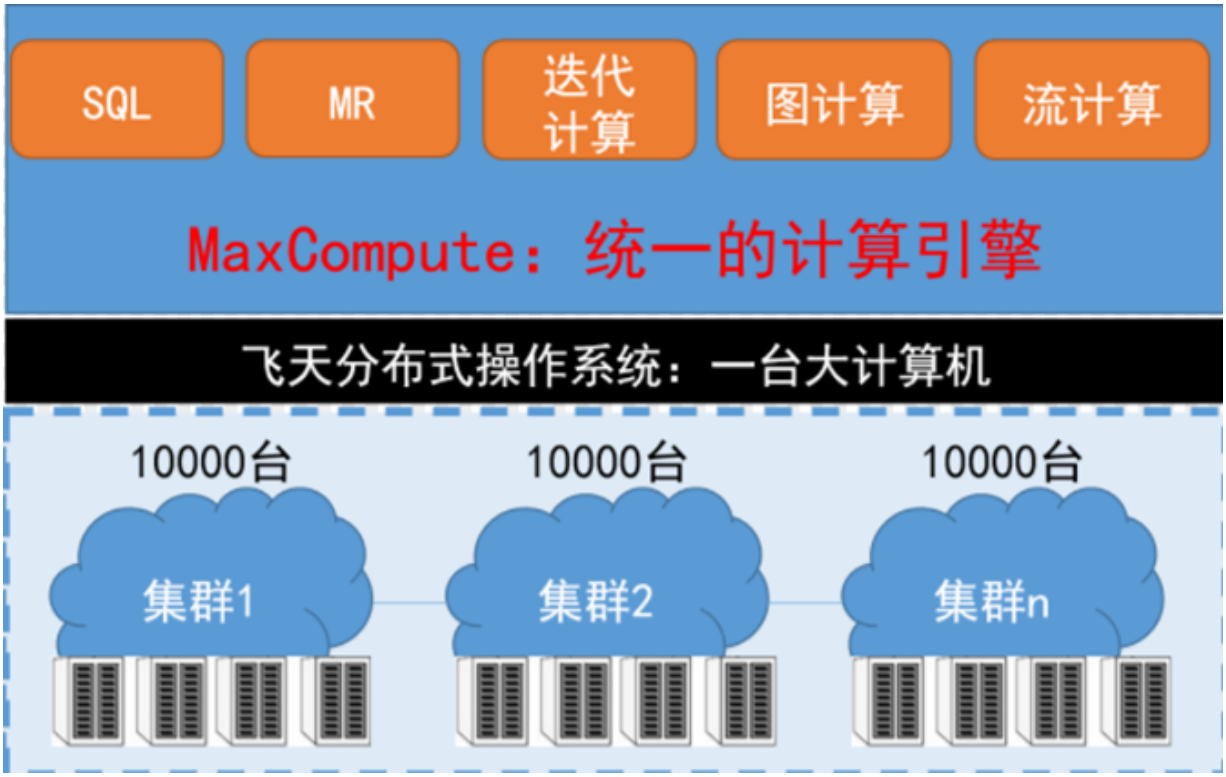
- MaxCompute处理数据大都存储在结构化的二维表中（Table）。
  - 表隶属于project。
  - 表可以进行分区。
  - 表中的数据类型

    有bigint，double，boolean，datetime，decimal，string，tinyint，smallint，int，float，varchar，binary，及struct。
- 数据由盘古存储系统管理，自动化的多副本存储策略提高数据可用性，屏蔽底层硬件故障。
- 列存储结构，压缩存储。

- 内置的数据生命周期管理策略。
- 基于存储配额的（Quota）多租户管理机制。

33.4.2 计算引擎ALL IN ONE BOX

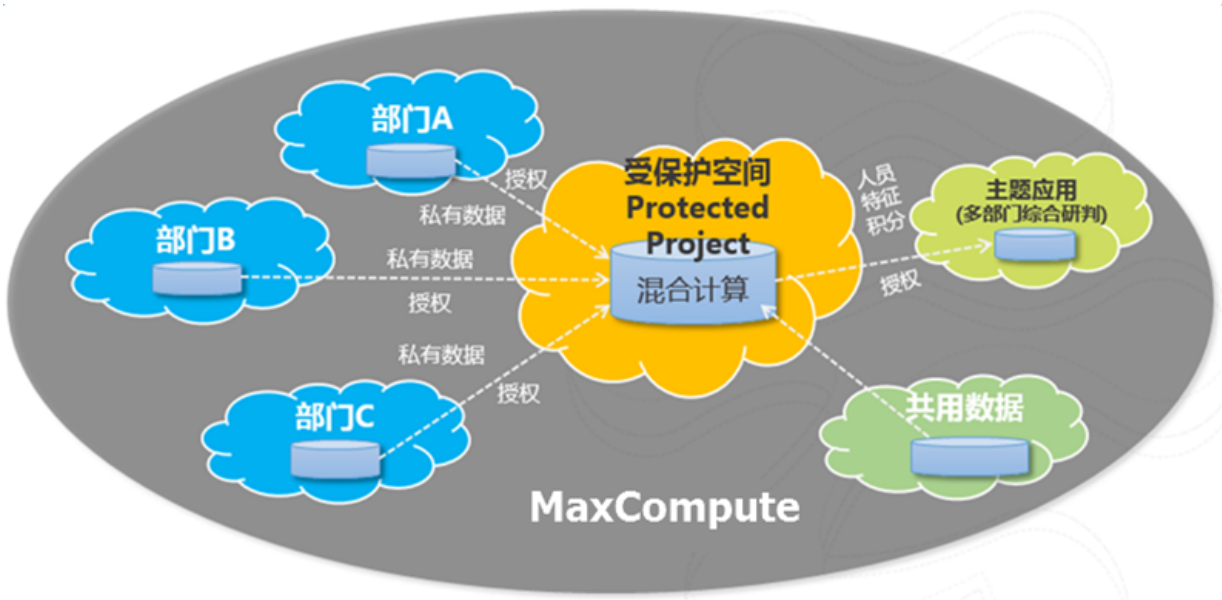
图 114: 操作系统



33.4.3 MaxCompute多租户机制

敏感数据共享式服务：数据不搬家，可用不可见。

图 115: 多租户机制

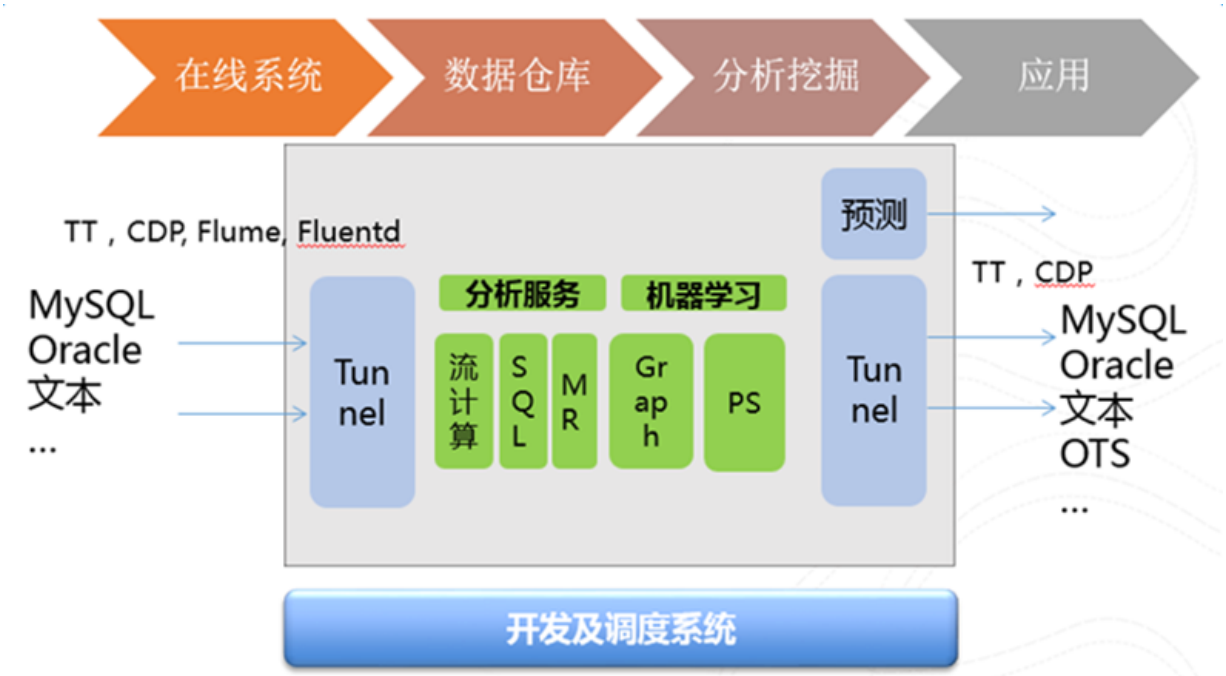


33.4.4 MaxCompute多集群支持

图 116: 多集群双活/灾备

33.4.5 数据处理流程

图 117: 数据处理流程



33.5 典型应用

海量数据的存储

只有大量的数据是不够的，还要将数据集中起来才能更好的发挥作用。

海量数据的计算

通过SQL，MR， Graph等方式，可以在一个任务中轻松处理TB级别的数据，并且有专门的分布式矩阵运算、数据挖掘算法（PAI）等对数据进行流式处理引擎及实时数据分析。

多组织间的数据交换

不仅能满足一个组织内部的数据管控，也能用于多个组织间的数据隔离和交换。

开箱即用的服务

您不用关注基础设施管理而是关注自己业务内容，MaxCompute本身提供各服务的一致性与连续性。

说明: MaxCompute并非在线OLTP关系数据库，它不支持事务，没有主键约束。

33.6 基本概念

项目空间

项目空间是 MaxCompute 的基本组织单元，它类似于传统数据库的Database或Schema的概念，是进行多用户隔离和访问控制的主要边界 。一个用户可以同时拥有多个项目空间的权限。通过安全授权，可以在一个项目空间中访问另一个项目空间中的对象，例如：表(Table) ，资源(Resource) ，函数(Function)，实例(Instance)

用户可以通过Use Project 命令进入一个项目空间，例如：

```
use my_project
-- 进入一个名为my_project的项目空间。
```

运行此命令后，用户会进入一个名为my_project的项目空间，从而可以操作该项目空间下的对象，例如：表(Table) ，资源(Resource) ，函数(Function) ，实例(Instance)等，而不需要关心操作对象所在的项目空间。Use Project是MaxCompute客户端提供的命令，相关命令的具体说明请参考： MaxCompute 常用命令 。

表

表是MaxCompute的数据存储单元，它在逻辑上也是由行和列组成的二维结构，每行代表一条记录，每列表示相同数据类型的一个字段，一条记录可以包含一个或多个列，各个列的名称和类型构成这张表的Schema 。阿里云数加平台的数据管理模块可以对MaxCompute表进行新建、收藏、修改数据生命周期管理、修改表结构和数据表/资源/函数权限管理审批等操作，详情请参考：基本介绍 以对数据管理模块功能进行深入了解。

MaxCompute的表格分两种类型：外部表及内部表。

对于内部表，所有的数据都被存储在MaxCompute中，表中的列可以是MaxCompute支持的任意种数据类型。MaxCompute中的各种不同类型计算任务的操作对象（输入、输出）都是表 。用户可以创建表，删除表以及向表中导入数据。

对于外部表，MaxCompute并不真正持有数据，表格的数据可以存放在 OSS 中。MaxCompute仅会记录表格的Meta信息。用户可以通过MaxCompute的外部表机制处理OSS上的非结构化数据，例如：视频、音频、基因、气象、地理信息等。处理流程包括：

1. 将数据上传至OSS。
2. 在 RAM产品中授予MaxCompute服务读取OSS数据权限。
3. 自定义Extractor，用于读取 OSS 上的特殊格式数据。目前，MaxCompute默认提供CSV格式的Extractor，并提供视频格式数据读取的代码样例。

4. 创建外部表。
5. 执行SQL作业分析数据。

分区

分区表指的是在创建表时指定分区空间，即指定表内的某几个字段作为分区列。在大多数情况下，用户可以将分区类比为文件系统下的目录。MaxCompute将分区列的每个值作为一个分区（目录）。用户可以指定多级分区，即将表的多个字段作为表的分区，分区之间类似多级目录的关系。在使用数据时如果指定了需要访问的分区名称，则只会读取相应的分区，避免全表扫描，提高处理效率，降低费用。

示例：

```
create table src (key string, value bigint) partitioned by (pt string);
-- 目前，MaxCompute仅承诺String类型分区。
```

```
select * from src where pt='20151201';
-- 正确使用方式。MaxCompute在生成查询计划时只会将'20151201'分区的数据纳入输入中。
```

```
select * from src where pt = 20151201;
--  错误的使用方式。在这样的使用方式下，MaxCompute并不能保障分区过滤机制的有效性。pt是String类型，当String类型与Bigint(20151201)比较时，MaxCompute会将二者转换为Double类型，此时有可能会有精度损失。
```

数据类型

MaxCompute 表中的列必须是下列描述的任意一种类型，各种类型的描述及取值范围包括：

类型	常量定义	描述	取值范围
Tinyint	1Y，-127Y	8位有符号整形。	-128 ~ 127
Smallint	32767S， -100S	16位有符号整形。	-32768 ~ 32767
Int	1000，-15645787	32位有符号整形。	-2^{31} ~ $2^{31}-1$
Bigint	100000000000L， -1L	64位有符号整形。	$-2^{63}+1$ ~ $2^{63}-1$
String	“abc”，’bcd’，”alibaba”，’inc’	字符串，支持UTF-8编码。其他编码的字符行为未定义。	单个String列最长允许8MB。
Float	无	32位二进制浮点型。	/
Boolean	True，False	布尔型。	True或False
Double	3.1415926 1E+7	64位二进制浮点型。	-1.0 10 308 ~ 1.0 10308

Datetime	Datetime '2017-11-11 00:00:00'	日期时间类型。使用东八区时间作为系统标准时间。	0001-01-01 00:00:00 000 ~ 9999- 12-31 23:59:59 999
Decimal	3.5BD，99999999999.9999999BD	10进制精确数字类型。	整形部分：$-10^{36}+1$ ~ $10^{36}-1$ 小数部分：精确到10^{-18}
Varchar	无	变长字符类型，n为长度。	1 ~ 65535
Binary	无	二进制数据类型。	单个Binary列最长允许8MB。
Timestamp	Timestamp '2017-11-11 00:00:00.123456789'	与时区无关的时间戳类型。	0001-01-01 00:00:00 000000000 ~ 9999-12-31 23:59:59 999999999

注意:

- 上述的各种数据类型均可为NULL。
- 对于Int类型常量，如果超过Int类型取值范围，会转换为Bigint类型；如果超过Bigint类型取值范围，会转换为Double类型。MaxCompute在未设定**odps.sql.type.system.odps2**为**true**的情况下，保留此转换，但会报告一个警告，提示Int类型被当作Bigint类型处理了。如果用户的脚本有此种情况，建议全部改写为Bigint类型，避免混淆。
- Varchar类型常量可通过String类型常量的隐式转换表示。
- String类型常量支持连接，例如 'abc' 'xyz' 会解析为 'abcxyz' ，不同部分可以写在不同行上。

MaxCompute支持的复杂数据类型的定义及构造方法，如下表所示：

类型	定义方法	构造方法

Array	array< int >;
	array< struct< a:int, b:string >>
		array(1, 2, 3);
		array(array(1, 2);
		array(3, 4))

Map	map< string, string >; map< smallint, array< string>>	map("k1", "v1", "k2", "v2");

		map(1S, array('a', 'b'), 2S, array('x', 'y'))
Struct	struct< x:int, y:int>; struct< field1:bigint, field2:array< int>, field3:map< int, int>>	named_struct('x', 1, 'y', 2); named_struct('field1', 100L, 'field2', array(1, 2), 'field3', map(1, 100, 2, 200)

资源

资源（Resource）是MaxCompute的特有概念。用户如果想使用MaxCompute的自定义函数#UDF#或*MapReduce*功能需要依赖资源来完成，例如：

- SQL UDF：用户在编写UDF后，需要将编译好的jar包以资源的形式上传到MaxCompute。运行这个UDF时，MaxCompute 会自动下载这个jar包，获取用户代码，运行UDF，无需用户干预。上传jar包的过程就是在 MaxCompute上创建资源的过程，这个jar包是MaxCompute资源的一种。
- MapReduce：用户编写MapReduce程序后，将编译好的jar包作为一种资源上传到MaxCompute。运行MapReduce作业时，MapReduce框架会自动下载这个jar资源，获取用户代码。用户同样可以将文本文件以及MaxCompute中的表作为不同类型的资源上传到MaxCompute。用户可以在UDF及MapReduce的运行过程中读取、使用这些资源。MaxCompute 提供了读取、使用资源的接口。详细示例请参考资源使用示例及*UDTF使用说明*中的描述。需要注意的是，MaxCompute的 自定义函数*(UDF)*或*MapReduce*对资源的读取有一定的限制，请参考应用限制 。

MaxCompute资源的类型包括：

- File类型。
- Table类型：MaxCompute 中的表。
- Jar类型：编译好的Java Jar包。
- Archive类型：通过资源名称中的后缀识别压缩类型，支持的压缩文件类型包括：.zip/.tgz/.tar.gz/.tar/jar。

函数

MaxCompute为用户提供了SQL计算功能，用户可以在MaxCompute SQL中使用系统的内建函数完成一定的计算和计数功能。但当内建函数无法满足要求时，用户可以使用MaxCompute提供的Java编程接口开发自定义函数（User Defined Function，以下简称UDF）。自定义函数#UDF#又可以进一步分为标量值函数（UDF），自定义聚合函数（UDAF）和自定义表值函数（UDTF）三种。

用户在开发完成UDF代码后，需要将代码编译成jar包，并将此jar包以jar资源的形式上传到MaxCompute，最后在MaxCompute中注册此UDF。在使用UDF时，只需要在SQL中指明UDF的函数名及输入参数即可，使用方式与MaxCompute提供的内建函数相同。

任务

任务(Task)是MaxCompute的基本计算单元。SQL及MapReduce功能都是通过任务(Task)完成的。

对于用户提交的大多数任务，特别是计算型任务，例如：*SQL DML语句*，*MapReduce*等，MaxCompute会对其进行解析，得出任务的执行计划。执行计划是由具有依赖关系的多个执行阶段(Stage)构成的。目前，执行计划逻辑上可以被看做一个有向图，图中的点是执行阶段，各个执行阶段的依赖关系是图的边。MaxCompute会依照图（执行计划）中的依赖关系执行各个阶段。在同一个执行阶段内，会有多个进程，也称之为Worker，共同完成该执行阶段的计算工作。同一个执行阶段的不同Worker只是处理的数据不同，执行逻辑完全相同。计算型任务在执行时，会被实例化，用户可以操作这个实例（Instance）的信息，例如：获取实例状态*(Status Instance)*，终止实例运行*(Kill Instance)*等。

另一方面，部分MaxCompute任务并不是计算型的任务，例如：SQL中的DDL语句，这些任务本质上仅需要读取、修改MaxCompute中的元数据信息。因此，这些任务无法被解析出执行计划。

任务实例

在MaxCompute中，部分任务*(Task)* 在执行时会被实例化，以MaxCompute实例（下文简称实例或Instance）的形式存在。实例会经历运行（Running）及结束（Terminated）两个阶段。运行阶段的状态为Running（运行中），而结束阶段的状态将会是Success（成功），Failed（失败）或Canceled（被取消）。用户可以根据运行任务时MaxCompute给出的实例ID查询、改变任务的状态，例如：

```
status <instance_id>;
-- 查看某实例的状态。
```

```
kill <instance_id>;
```

```
-- 停止某实例，将其状态设置为Canceled。
```

资源配额

配额（Quota）分为存储和计算两种。对于存储，在MaxCompute中可以设置一个project中允许使用的存储上限，在接近上限到一定程度时会触发报警。对于计算资源的限制，有内存和CPU两方面，即在project 中同时运行的进程所占用的内存和CPU资源不可以超过指定的上限。

34 大数据开发套件

34.1 产品概述

大数据开发套件（Data IDE）是阿里云数加重要的Paas平台产品，是"DataWorks"中最重要的核心组件。提供全面托管的工作流服务，一站式开发管理的界面，帮助企业专注于数据价值的挖掘和探索。

使用大数据开发套件（Data IDE），可对数据进行数据传输、数据转换等相关操作，从不同的数据存储引入数据，对数据进行转化处理，最后将数据提取到其他数据系统。

图 118: 产品优势

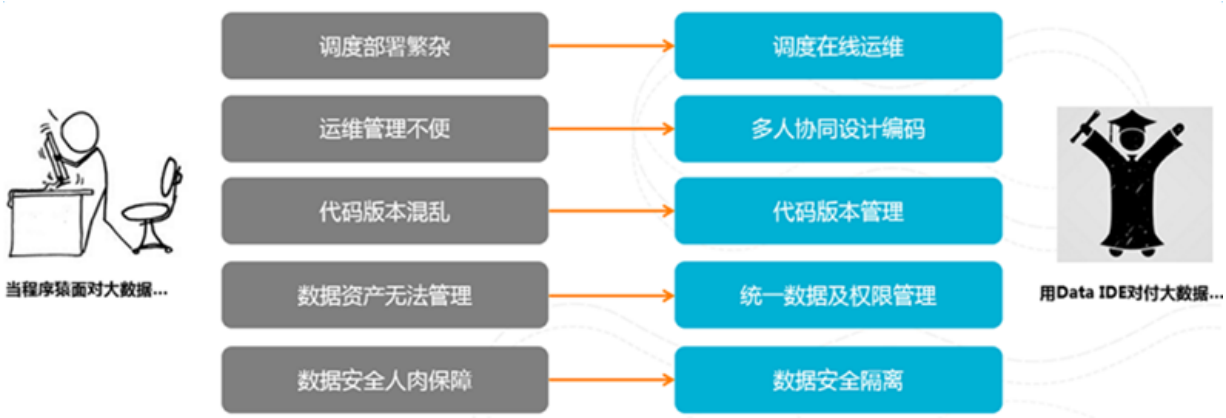
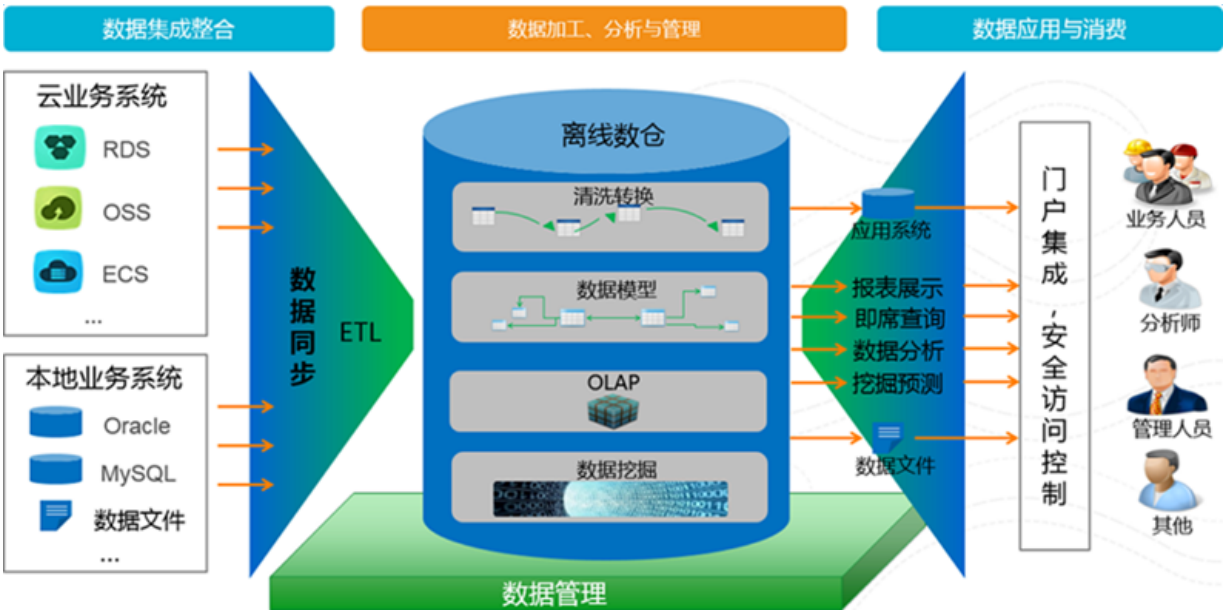


图 119: 数据开发流程



34.2 功能特性

大数据开发套件，为您提供一站式大数据集成、处理、分析的平台。为您提供稳定高效的数据上云、可视化在线数据开发工具、海量离线调度系统和海量数据管理的功能。

数据上云

图 120: 数据上云

在线数据加工

图 121: 数据加工

Job离线调度

图 122: 离线调度

全局数据管理

图 123: 数据管理

34.3 产品优势

多人协同操作

支持100人以上协同设计、开发、运维。通过操作日志审计、版本diff/回滚、互斥锁机制等提供面向企业级的线上协同开发的数仓开发平台。

良好的扩展性

可以对单个工作项目快速扩展，平滑迁移数据及离线任务，并且对接基于大数据开发平台的BI分析、机器学习、数据可视化等功能，为用户提供一个强大的数据价值最大化的产品平台。

开放的生态基础

提供各个产品功能模块的OpenAPI，提供良好的二次开发的能力。

数据安全保障

提供多个数据实例（MaxCompute Project）之间的数据授权机制，真实的线上数据只能使用却不可见，让用户的数据做到真正的安全。

强大的调度运维能力

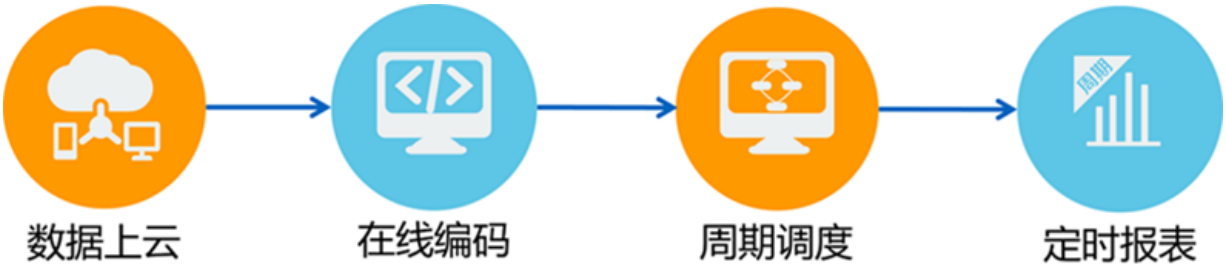
提供白屏化的运维能力，以及任务执行状态监控、自定义执行机器等功能，让用户更好的掌控自己的任务。

34.4 典型应用

34.4.1 BI应用

如何用Data IDE做报表。

图 124: BI应用流程



基于网络日志，完成如下分析需求：

- 统计网站的PV（浏览次数）、UV（独立访客），按用户的终端类型（如Android、iPad、iPhone、PC等）分别统计，并给出这一天的统计报表。
- 网站的访问来源，了解网站的流量从哪里来。

截取一条真实的数据如下：

```
18.111.79.172 - - [12/Feb/2014:03:15:52 +0800] "GET /articles/4914.html HTTP/1.1" 200 37666 "http://coolshell.cn/articles/6043.html" "Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.107 Safari/537.36" -
```

【新建表】：在导入数据之前，需要先创建一张MaxCompute目标表，我们把表名命名为ods_log_tracker。

图 125: 建表

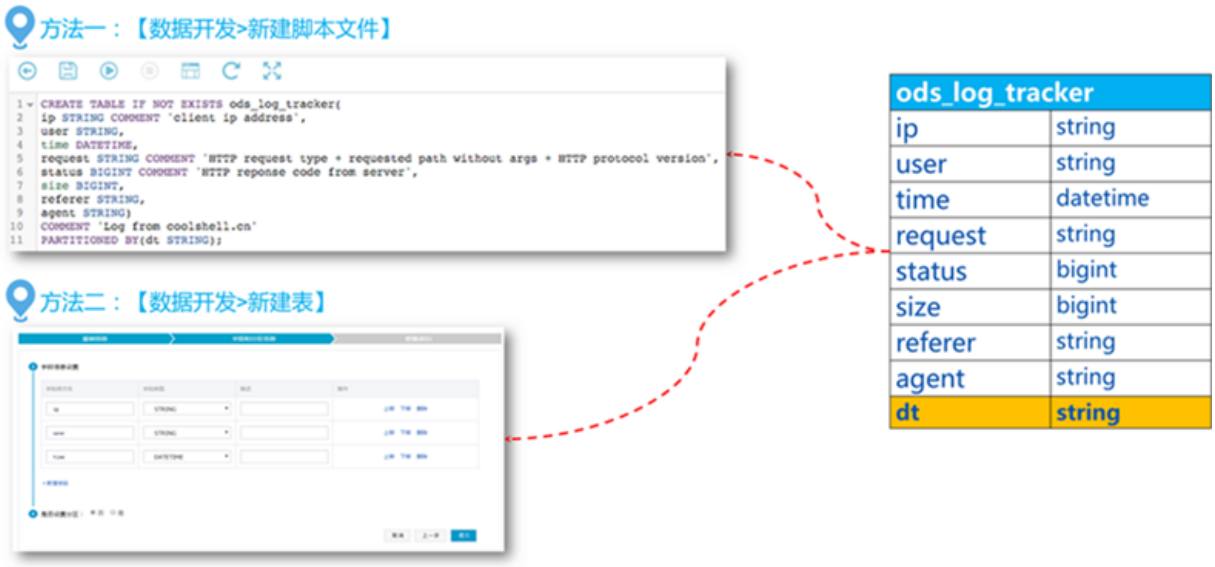


图 126: 各表依赖

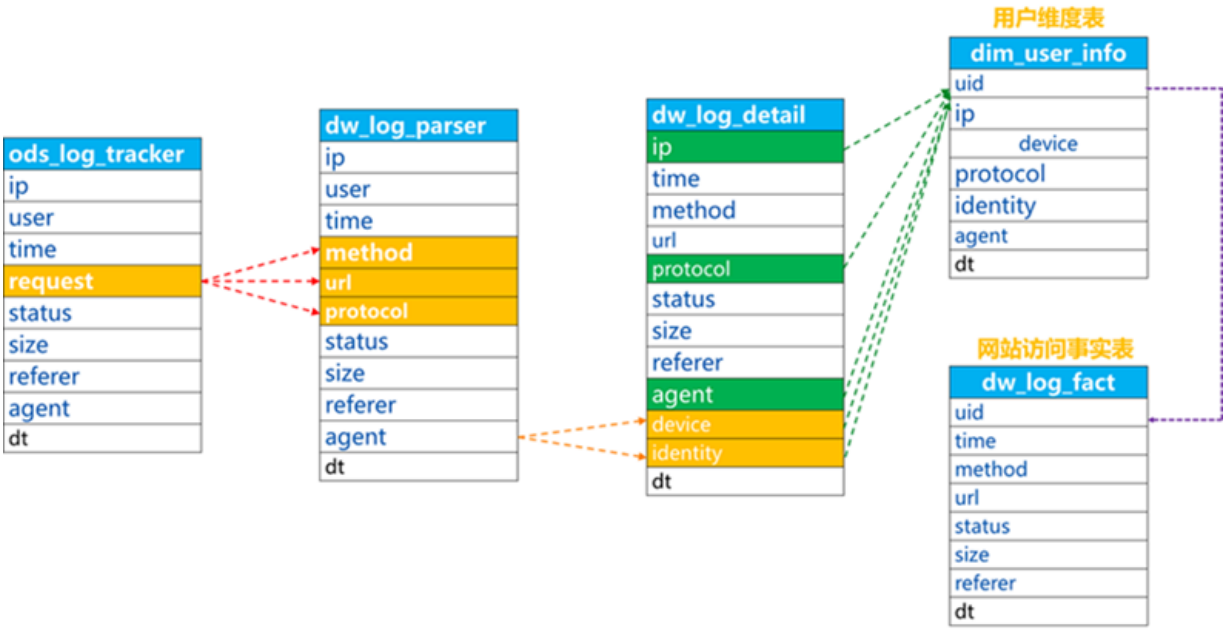


图 127: 新建任务

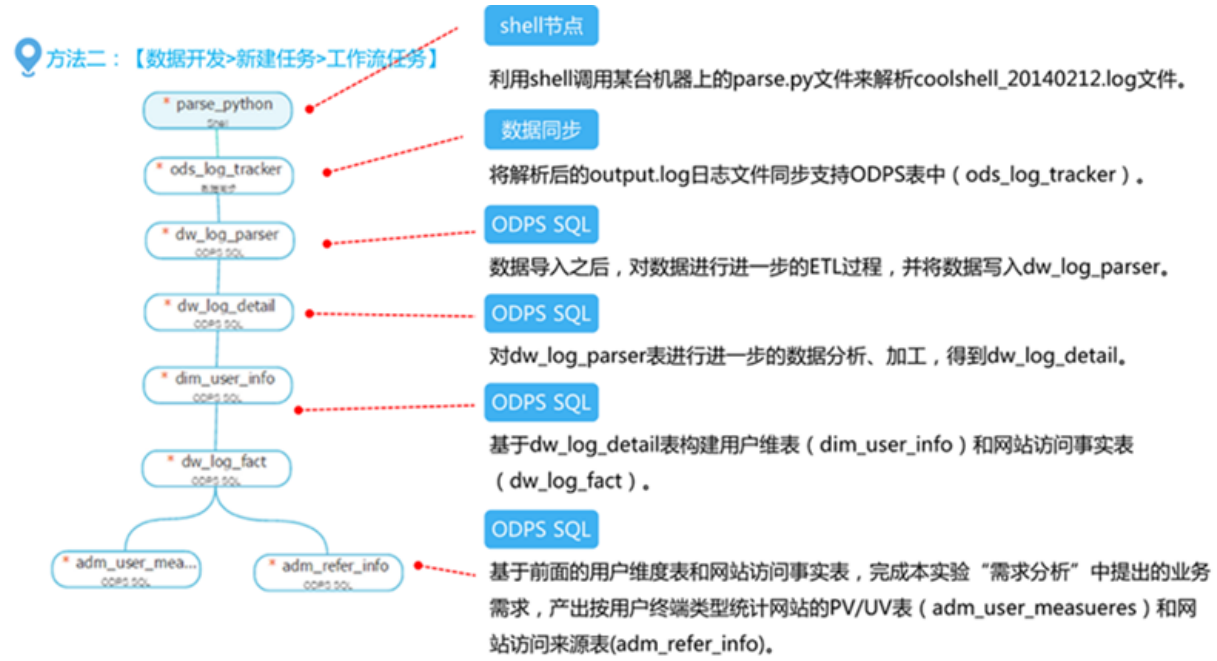


图 128: 开发过程

34.4.2 云上数仓

图 129: 数据采集到应用

图 130: 数据同步

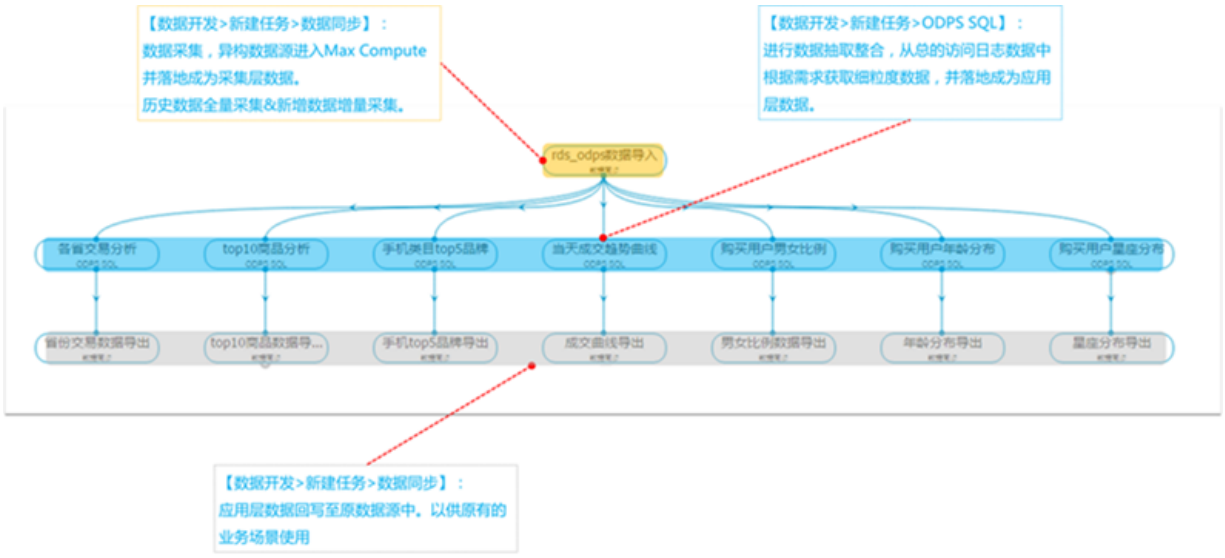


图 131: 任务管理

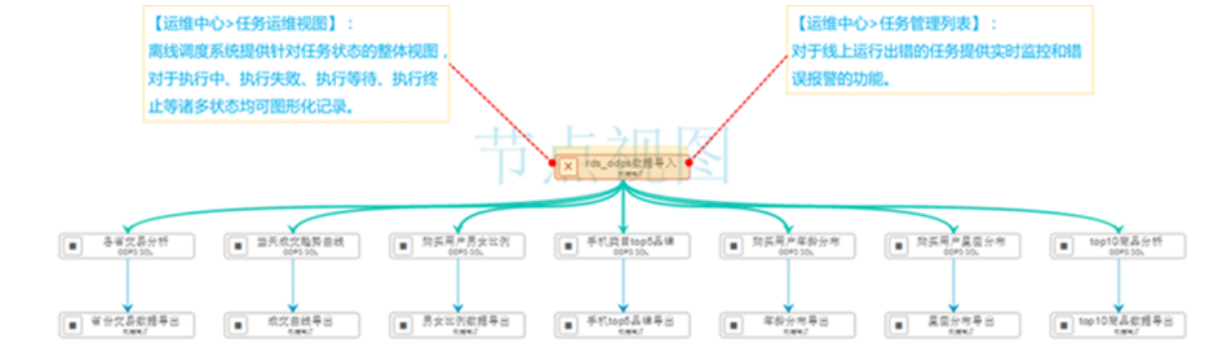
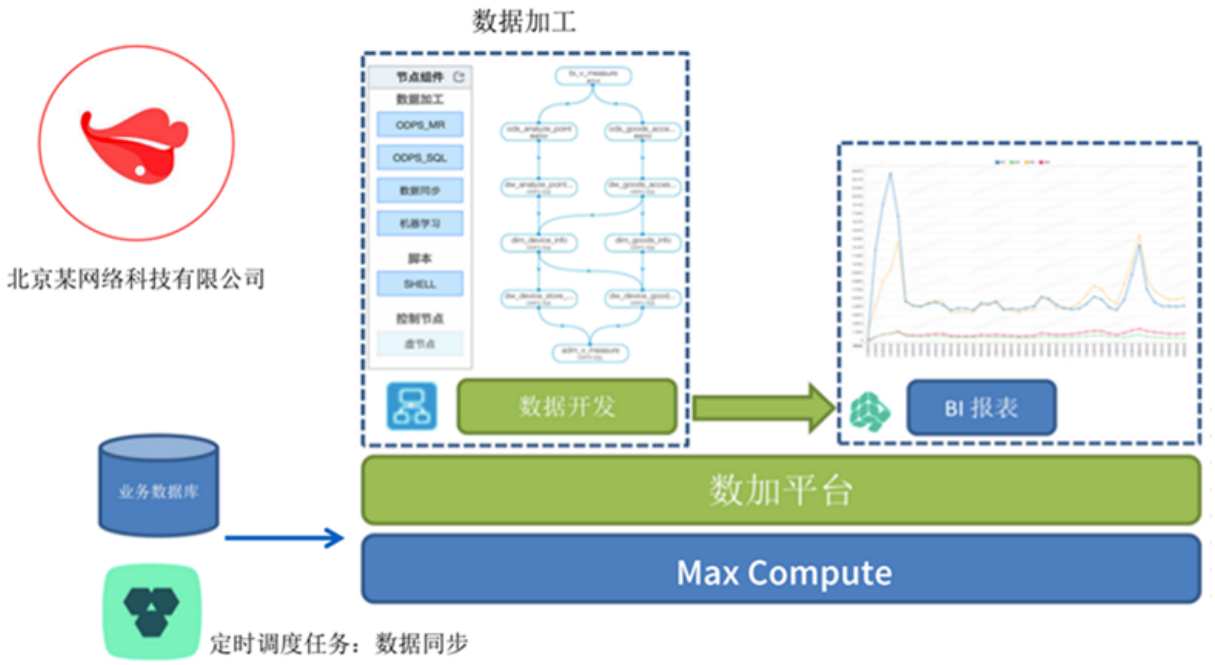


图 132: 数据查询

34.4.3 实践案例

图 133: 实践案例



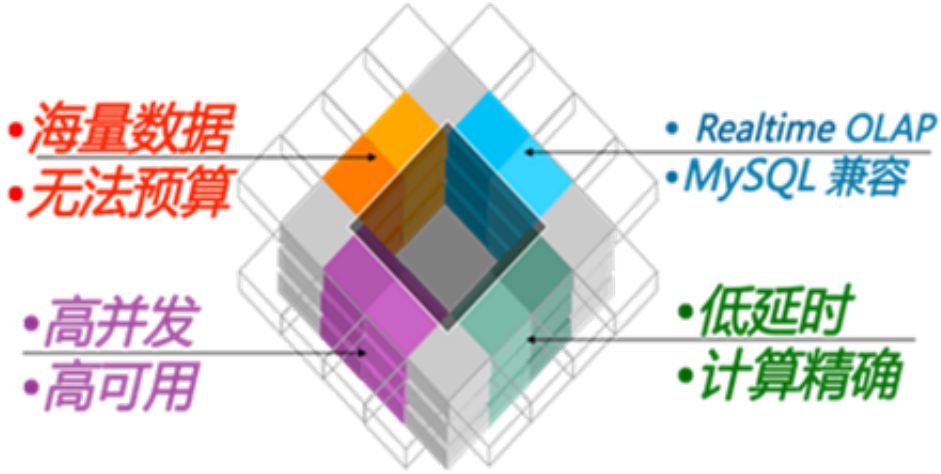
35 分析型数据库

35.1 产品概述

根据IDC 2013年发布的数字宇宙研究报告（Digital Universe）显示，在接下来的8年中，我们所产生的数据量将超过40ZB（泽字节）。作为大数据特征中最重要的 Volume（容量）、Velocity（数据生产速度）的两个原始特征都在发生急剧变化，使得数据处理从业务系统的一部分演变得愈发独立，企业待加速数据分析和挖掘过程，并由报表展现为主到强调数据洞察转型，让数据直接快速产生价值（Value）。在业务系统中，我们通常使用的是OLTP （OnLine Transaction Processing，联机事务处理）系统，如MySQL、Microsoft SQL Server 等关系数据库系统。这些关系数据库系统擅长事务处理，在数据操作中保持着严格的一致性和原子性，能够很好支持频繁的数据插入和修改，但是，一旦待进行查询或计算的数据量过大，达到数千万甚至数十亿条，或待进行的计算非常复杂的情况下，OLTP类数据库系统便力不从心了。

图 134: 数据库对比

图 135: 优势对比



分析型数据库（AnalyticDB，原ADS）和主流数据系统进行对比时，我们等待OLAP（On-Line Analytical Processing，联机分析处理）系统来进行处理。从广义上，OLAP系统是针对OLTP系统而言，并不特别关心对数据进行输入、修改等事务性处理，而是关心对已有大量数据进行多维度的、复杂的分析的一类数据处理系统。而在具体的产品中，我们通常将OLAP系统分为MOLAP、ROLAP和HOLAP三类。

多维OLAP（Multi-Dimensional OLAP，简称MOLAP），是预先根据数据待分析的维度进行建模，在数据的物理存储层面以“立方体”（Cube）的结构进行存储，具有查询速度快等优点，但是数据必须预先建模，无法依据使用者的意愿进行即时灵活的修改。而关系型OLAP（Relational OLAP，简称 ROLAP），则使用类似关系数据库的模型进行数据存储，通过类似SQL等语言进行查询和计算，优点是数据查询计算自由，可以灵活地根据使用者的要求进行分析，但是缺点是在海量数据的情况下分析计算缓慢。至于HOLAP，则是MOLAP和ROLAP的混合模式。

而分析型数据库，则是一套RT-OLAP（Realtime OLAP，实时OLAP）系统。在数据存储模型上，采用自由灵活的关系模型存储，可以使用SQL进行自由灵活的计算分析，无需预先建模，而利用分布式计算技术，分析型数据库可以在处理百亿条甚至更多量级的数据上达到甚至超越MOLAP类系统的处理性能，真正实现百亿数据毫秒级计算。

分析型数据库让海量数据和实时与自由的计算可以兼得，实现了速度驱动的大数据商业变革。一方面，分析型数据库拥有快速处理迁移级别海量数据的能力，使得数据分析中使用的数据可以不再是抽样的，而是业务系统中产生的全量数据，使得数据分析的结果具有最大的代表性。

另外，AnalyticDB采用分布式计算技术，拥有强大的实时计算能力，通常可以在数百毫秒内完成百亿级的数据计算，使得使用者可以根据自己的想法在海量数据中自由的进行探索，而不是根据预先设定好的逻辑查看已有的数据报表。

更重要的是，由于分析型数据库能够支撑较高并发查询量，并且通过动态的多副本数据存储计算技术来保证较高的系统可用性，所以能够直接作为面向最终用户（End User）的产品（包括互联网产品和企业内部的分析产品）的后端系统。如淘宝数据魔方、淘宝指数、快的打车、阿里妈妈达摩盘（DMP）、淘宝美食频道等拥有数十万至上千万最终用户的互联网业务系统中，都使用了分析型数据库。

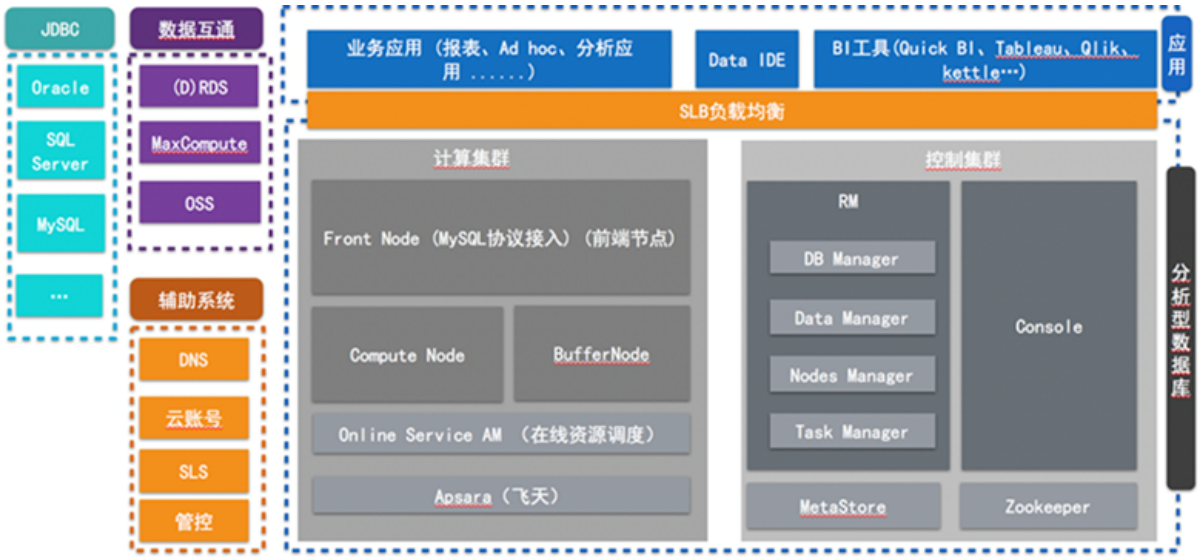
大数据领域，愈发向“4M”趋势发展：

图 136: 4M趋势

分析型数据库作为海量数据下的实时计算系统，给使用者带来极速自由的大数据在线分析计算体验，最终期望为大数据行业带来巨大的变革。

35.2 产品架构

图 137: AnalyticDB架构



AnalyticDB（原ADS）是构建在阿里云分布式操作系统飞天之上的基于MPP架构并融合了分布式检索技术的分布式实时计算系统。如图所示，Analytic DB的主体部分主要由四个部分组成：

- 底层依赖

  包括用于进行资源虚拟化隔离、数据持久化存储、构建数据结构和索引而使用的阿里云飞天分布式操作系统套件，用于存储分析型数据库的各类元数据（注意并不是实际参与计算用的数据）的阿里云分析型数据库或阿里云表格存储，以及用于各个组件间进行分布式协调的开源Apache ZooKeeper模块。

- 计算集群

  是计算资源实际包括的内容，均可进行横向扩展。包括用于处理用户连接接入认证、鉴权、查询路由与分发以及提供元数据查询管理服务的FrontNode、用于进行实际的数据存储与计算的ComputeNode、用于处理数据实时更新数据缓冲和实时数据写入版本控制的BufferNode。计算集群运行在阿里云超大规模分布式操作系统飞天上，并通过在线资源调度模块来使用飞天调度计算资源。

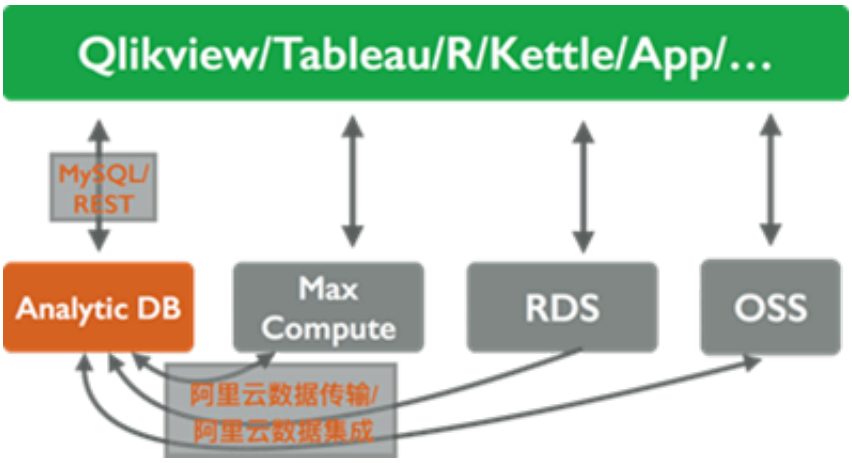
- 控制集群

  暨资源管理器RM，用于控制计算集群中数据库资源分配、数据库内数据和计算资源的分布、管理飞天集群上的计算节点、管理数据库后台运行的任务等。控制集群实际上由多个模块组成，一个控制集群可以同时管理位于不同机房部署的多套计算集群。

- 外围模块

如用于管理Front Node的分组和负载均衡的阿里云负载均衡、用于发布数据库域名的阿里云DNS系统、阿里云账号系统、Analytic DB的控制台（Admin Console）和用户控制台（DMS for Analytic DB）等。

图 138: 外部系统交互



在对外部系统的交互上，分析型数据库能够从MaxCompute批量导入数据，并且可以快速批量导出海量数据到MaxCompute；可以实时的将(D)RDS的数据同步到分析型数据库中（须借助外部同步工具）。

对于前端业务，分析型数据库允许任何遵循MySQL 5.1/5.5/5.6系列协议的客户端和驱动进行连接。例如：MySQL 5.1.x jdbc driver、MySQL 5.3.x odbc connector(driver)、MySQL 5.1.x/5.5.x/5.6.x 客户端、java、python、C/C++、Node.js、PHP、R(RMySQL)、Websphere Application Server 8.5、Apache Tomcat、JBoss等。

35.3 产品优势

海量数据计算能力

最高支持单表万亿记录、PB数据级别。

自由灵活的查询能力

通过SQL对海量数据灵活的进行多维分析、数据透视、数据筛选。

极速的查询响应时间

毫秒级的百亿级数据多维透视。

多通道并行数据导入

离线通道、在线通道双模式并行数据导入，导入性能随集群规模线性扩展。

精细化的安全机制

提供精确到列级别的权限管理和超细粒度用户操作审计，利用公私钥机制保护数据安全。

良好的兼容性

全面兼容MySQL协议（包括数据元信息），天生具备与商业分析工具、应用的兼容性，内置支持多种数据源数据快速接入，大幅度降低业务系统和商业软件的接入成本。

35.4 典型应用

电商行业

A-CRM、爆款选品、自动化运营、SKU组合分析等。

O2O

数据分析和CRM系统、地理围栏系统。

广告行业

数字营销，M-DMP系统。

金融行业

实时多维数据分析、交易流水查询系统、报表系统等。

大安全

人群透视分析，潜在关键元素挖掘，关系网络分析，明细查询等。

交通、交警

车辆卡口数据分析和研判。

物流和物联网业

车联网数据分析、企业安监数据分析、传感器数据存储和检索、物流实时数仓。

35.4.1 某银行

图 139: 银行应用

业务问题

通过分析型数据库，将原有的需要运行20分钟的报表加速到1秒钟，并且能够支撑冠字号查询等业务系统

已有方案

Oracle

数据规模

700G 30亿

优势

实时计算处理时间从20分钟提升到1秒

35.4.2 某交警

图 140: 交警应用

业务系统特点：

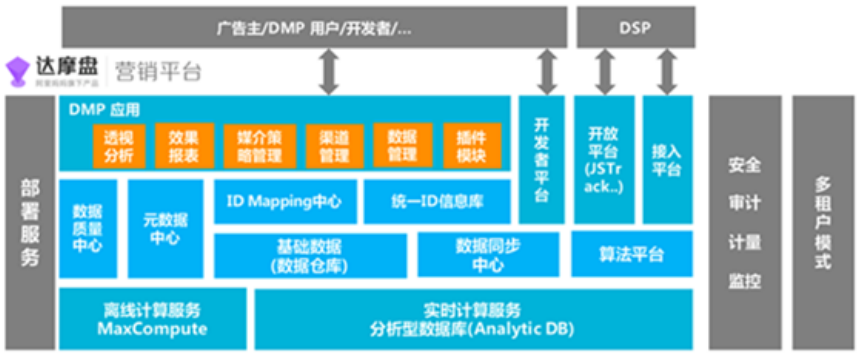
海量数据：一个市仅交通卡口过车纪录表达到300亿 - 500亿级别（保存半年），折合数据20-30T

飞速增长：市级系统每天数据增量1000万条

复杂查询：多个部门需要，多种插叙方式，业务应用查询复杂，包括单表查询、多表查询（join）、模糊查询（like）、轨迹分析（in）、区域碰撞（intersect）、短时过车（having count）、多用户等多种应用场景，对表结构设计、内存使用效率、CPU使用率等要求较高，对查询的并发性有要求

35.4.3 阿里妈妈DMP

图 141: DMP应用



广告DMP应用架构

在DMP系统中，大数据处于整个系统的核心位置。MaxCompute进行用户数据清洗，标签的挖掘。AnalyticDB承接了广告主对大数据的透视和人群管理的计算工作。Analytic DB极速的海量数据DUMP功能可以将用户圈定完成的人群DUMP到查询更快速的KV存储系统中。定向引擎根据KV存储系统的数据，服务于DSP系统。

36 流计算

36.1 产品概述

目前对信息高时效性、可操作性的需求不断增长，这要求软件系统在更少的时间内能处理更多的数据。传统的大数据处理模型将在线事务处理和离线分析从时序上将两者完全分割开来，但显然该架构目前已经越来越落后于人们对于大数据实时处理的需求。

流计算的产生即来源于对于上述数据加工时效性的严苛需求： **数据的业务价值随着时间的流失而迅速降低**，**因此在数据发生后必须尽快对其进行计算和处理。**而传统的大数据处理模式对于数据加工均遵循传统日清日毕模式，即以小时甚至以天为计算周期对当前数据进行累计并处理，显然这类处理方式无法满足数据实时计算的需求。在诸如实时大数据分析、风控预警、实时预测、金融交易等诸多业务场景领域，批量（或者说离线）处理对于上述对于数据处理时延要求苛刻的应用领域而言是完全无法胜任其业务需求的。而流计算作为一类针对流数据的实时计算模型，可有效地缩短全链路数据流时延、实时化计算逻辑、平摊计算成本，最终有效满足实时处理大数据的业务需求。

什么是流数据？

从广义上说，所有大数据的生成均可以看作是一连串发生的离散事件。这些离散的事件以时间轴为维度进行观看就形成了一条条事件流/数据流。不同于传统的离线数据，流数据是指由数千个数据源**持续生成的数据**，流数据通常也以数据记录的形式发送，但相较于离线数据，流数据普遍的规模较小。流数据产生源头来自于源源不断的事件流，例如客户使用您的移动或 Web 应用程序生成的日志文件、网购数据、游戏内玩家活动、社交网站信息、金融交易大厅或地理空间服务，以及来自数据中心内所连接设备或仪器的遥测数据。

通常而言，流计算具备三大类特点：

- 实时（realtime）且无界（unbounded）的数据流

  流计算面对计算的数据源是实时且流式的，流数据是按照时间发生顺序地被流计算**订阅**和**消费**。且由于数据发生的持续性，数据流将长久且持续地集成进入流计算系统。例如，对于网站的访问点击日志流，只要网站不关闭其点击日志流将一直不停产生并进入流计算系统。因此，对于流系统而言，数据是实时且不终止（无界）的。

- 持续（continuos）且高效的计算

  流计算是一种“事件触发”的计算模式，触发源就是上述的无界流式数据。一旦有新的流数据进入流计算，流计算立刻发起并进行一次计算任务，因此整个流计算是持续进行的计算

- 流式（streaming）且实时的数据集成。

  流数据触发一次流计算的计算结果，可以被直接写入目的数据存储，例如将计算后的报表数据直接写入RDS进行报表展示。因此流数据的计算结果可以类似流式数据源一样持续写入目的数据存储。

36.1.1 产品历程

阿里云流计算脱胎于阿里集团内部双十一实时大屏业务，在阿里集团内部从最开始支持双十一大屏展现和部分实时报表业务的实时数据业务团队，历经4、5年的长期摸索和发展，到最终成长一个独立稳定的云计算产品团队。阿里云流计算期望将阿里集团本身沉淀多年的流计算产品、架构、业务能够以云产品的方式对外提供服务，助力更多中小企业实时化自身大数据业务。

最初阿里集团支撑双十一大屏等业务同样采用的是开源的 Storm 作为基础系统支持，并在上面开发相关 Storm 代码。这个时期的实时业务处于萌芽阶段，规模尚小。数据开发人员使用 Storm 原生 API 开发流式作业，开发门槛高，系统调试难，存在大量重复的人肉工作。

阿里集团的工程师针对这类大量重复工作，开始考虑进行业务封装和抽象。工程师们基于 Storm 的 API 开发出大量可复用的数据统计组件 ，例如实现了简单过滤、聚合、窗口等等作为基础的编程组件，并基于这类组件提供了一套 XML 语义的业务描述语言。基于这套设计，流式计算用户可以使用 XML 语言将不同的组件进行拼装描述，最终完成一整套完整的流计算处理流程。基于 XML +Storm 组件的编程方式，从底层上避免了用户大量的重复开发工作，同时亦降低了部分使用门槛。但我们的数据分析人员仍然需要熟悉整套编程组件和 XML 描述语法，这套编程方式离分析人员最熟悉的 SQL 方式仍然差距甚远。

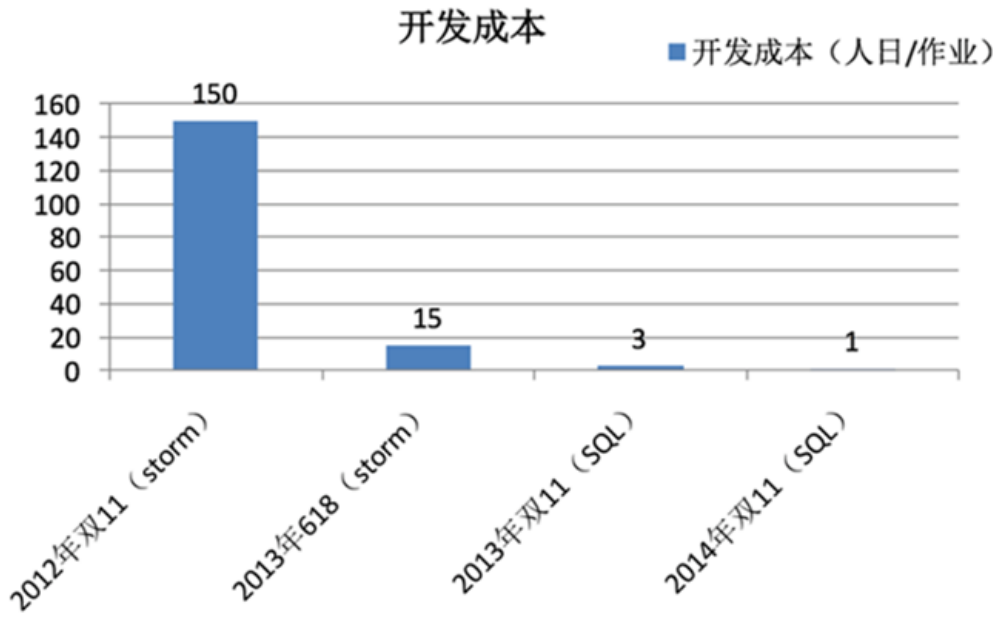
任何技术的发展一定遵循 **小众/创新** 到 **大众/普及** 的成长轨迹，而从小众到大众，从创新到普及的转折点一定在于技术的功能成熟和成本降低。阿里工程师开始思考如何更大程度降低数据分析产品门槛从而普及到更多的用户。得益于关系型数据库几十年沉淀的用户群体，使用经典的 SQL 模式去计算和处理数据一则可以对标 SQL 功能从而提炼我们的技术成熟度，二则可以利用用户熟悉的 SQL 模型可极大降低用户上手使用流计算的门槛。因此，阿里工程师最终开发一套 StreamSQL 替换了原有的 XML+ 组件的编程方式，这套系统成为今天阿里云流计算的核心计算引擎（Galaxy）。当前这套系统以单机群数千台机器规模，在阿里集团内部服务20+BU，日均消息处理数千亿，流量近 PB 级别，成为阿里集团最核心的流式计算集群。

当前阿里云流计算在原有 Galaxy 系统基础上，更加丰富和提升了用户的使用体验，包括提供一整套的开发平台，完整的流式数据处理业务流程。使用阿里云流计算，受益于阿里大数据多年的技术和业务沉淀，用户可以完全享受到阿里集团最新最前沿的计算引擎能力，业务上可规避阿里集团多年

在流式大数据的试错和教训，让用户自身可以更快、更轻松地实时化大数据处理流程，助力业务发展。

阿里集团工程师使用阿里云流计算开发工期对比如图 *142: 开发成本*所示。

图 142: 开发成本



36.1.2 产品定位

阿里云流计算提供类标准的 StreamSQL 语义协助用户简单轻松完成流式计算逻辑的处理。同时，受限于 SQL 代码功能有限无法满足某些特定场景的业务需求，阿里云流计算同时为部分授信用户提供全功能的 UDF 函数，帮助用户完成业务定制化的数据处理逻辑。在流数据分析领域用户直接使用 StreamSQL+UDF 即可完成大部分流式数据分析处理逻辑，目前的流计算**更擅长于做流式数据分析、统计、处理**，对于非 SQL 能够解决的领域，例如复杂的迭代数据处理、复杂的规则引擎告警则不适合现有的流计算产品去解决。

目前流计算擅长解决的几个领域的应用场景：

- 实时的网络点击 PV、UV 统计。
- 统计交通卡口的平均五分钟通过车流量。
- 水利大坝的压力数据统计和展现。
- 网络支付涉及金融盗窃固定行为规则的告警。

曾经阿里云流计算对接，但发现无法满足的情况：

- Oracle 存储过程使用阿里云流计算替换：流计算无法从功能上完全替换掉 Oracle 存储过程，两者面向问题领域不一致。
- 现有的 Spark 作业无缝迁移到流计算： Spark 部分涉及流计算可以考虑改造并迁移到流计算，用户可以完全省去运维 Spark 和开发 Spark 的各类成本，但无法做到 Spark 作业无缝迁移到流计算。
- 多种复杂规则引擎告警：针对单一一条数据存在多条复杂规则告警，且该规则在系统运行时变化。这类应该有专门的规则引擎系统解决，当前流计算面对不是该问题域。

当前流计算对外接口定义为 StreamSQL/UDF，提供服务于流式数据分析、统计、处理的一站式开发工具，面向客户更多是数仓开发人员、数据分析师，这类客户不希望更多参与底层代码开发，而希望**简单编写流计算 SQL 即可完成自身流式数据分析业务。**

36.1.3 产品特性

相较于其他流计算产品，阿里云流计算提供一些极具竞争力的产品优势，用户可以充分利用阿里云流计算提供的产品优势，方便快捷的解决自身业务实时化大数据分析的问题。

强大的实时处理能力

不同于其他开源流计算中间件只提供粗陋的计算框架，大量的流计算细节需要业务人员造轮子重新实现。阿里云流计算集成诸多全链路功能，方便用户进行全链路流计算开发，包括：

- 强大的流计算引擎，阿里云流计算提供提供标准的 StreamSQL，支持各类Fail场景的自动恢复，保证故障情况下数据处理的准确性；支持多种内建的字符串处理、时间、统计等类型函数；精确的计算资源控制，彻底保证用户作业的隔离性。
- 关键性能指标超越 Storm 的六到八倍，数据计算延迟优化到秒级乃至毫秒级，单个作业吞吐量可做到百万级别，单集群规模在数千台。
- 深度整合各类云数据存储，包括 DataHub、日志服务（SLS）、RDS、OTS、ADS、IOTHub 等各类数据存储系统，无需额外的数据集成工作，阿里云流计算可以直接读写上述产品数据。

托管的实时计算服务

不同于开源或者自建的流式处理服务，阿里云流计算是完全托管的流式计算引擎，阿里云可针对流数据运行查询，无需预置或管理任何基础设施。在阿里云流计算，您可以享受一键启用的流式数据服务能力。阿里云流计算天然集成数据开发、数据运维、监控预警等服务，方便您最小成本试用和迁移流式计算。

提供完全租户隔离的托管运行服务，从最上层工作空间到最底层执行机器，提供最有效的隔离和全面防护，让用户放心使用流计算。

良好的流式开发体验

支持标准 SQL（产品名称为：StreamSQL），提供内建的字符串处理、时间、统计等各类计算函数，替换业界低效且复杂的 Storm 开发，让更多的BI人员、运营人员通过简单的 StreamSQL 可以完成实时化大数据分析和处理，让实时大数据处理普适化、平民化。

提供全流程的流式数据处理方案，针对全链路流计算提供包括数据开发、数据运维、监控预警等不同阶段辅助套件，让数据开发仅需三步，即可完成流式计算任务上线。

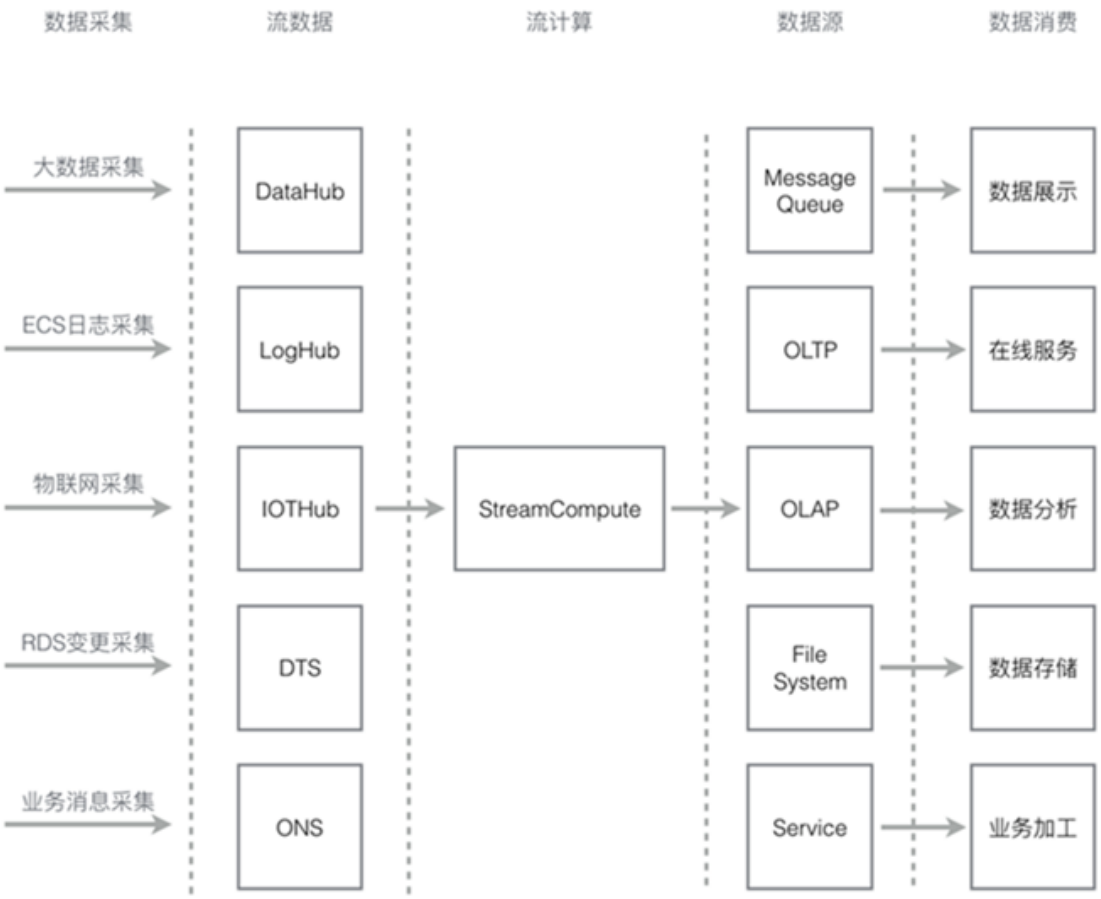
低廉的人力和集群成本

大量优化的 SQL 执行引擎，会产生比手写原生Storm任务更高效且更廉价的计算任务，无论开发成本和运行成本，阿里云流计算均要远低于开源流式框架。考虑下，编写一个复杂业务逻辑下 Storm 任务 Java 代码行数，考虑下针对这个任务的调试、测试、调优、上线工作成本，考虑下后续长期对于 Storm、Zookeeper 等开源软件的运维成本。如果使用阿里云流计算，上述问题完全交由阿里云平台承担，用户完全聚焦业务，快速实现市场目标。

36.1.4 业务流程

在阿里云流计算使用前，对流式数据处理整体全链路有个简单认识可以极大方便用户梳理业务流程，制定相应的系统设计方案。阿里云流计算全流程系统架构如图 *143: 系统架构*所示。

图 143: 系统架构



部分数据源并不是通过流计算直接打通，需要通过中转进入流计算，请参看本节最后的说明。

1. **数据采集**

   广义的实时数据采集指： 用户使用流式数据采集工具将数据**流式且实时地采集并传输到大数据 Pub/Sub 系统**，该系统将为下游流计算提供源源不断的事件源去触发流式计算任务的运行。阿里云大数据生态中提供了诸多针对不同场景领域的流式数据 Pub/Sub 系统，阿里云流计算天然集成上图中诸多 Pub/Sub 系统，以方便用户可以轻松集成各类流式数据存储系统。例如用户可以直接使用流计算对接 SLS 的 LogHub 系统，以做到快速集成并使用 ECS 日志。

2. **流式计算**

   流数据作为流计算的触发源驱动流计算运行。因此，**一个流计算任务必须至少使用一个流数据作为数据源**。同时，对于一些业务较为复杂的场景，流计算还支持和静态数据存储进行关联查询。例如针对每条 DataHub 流式数据，流计算将根据流式数据的主键和 RDS 中数据进行关联查询（即 join 查询）；同时，阿里云流计算还支持针对多条数据流进行关联操作，StreamSQL 支持阿里集团量级的复杂业务也不在话下。

3. **实时数据集成**

   为尽可能减少数据处理时延，同时减少数据链路复杂度。阿里云流计算将计算的结果数据可不经其他过程直接写入目的数据源，从而最大程度降低全链路数据时延，保证数据加工的新鲜度。为了打通阿里云生态，阿里云流计算天然集成了OLTP（RDS产品线等）、NoSQL（OTS等）、OLAP（ADS等）、MessageQueue（DataHub、ONS等）、MassiveStorage（

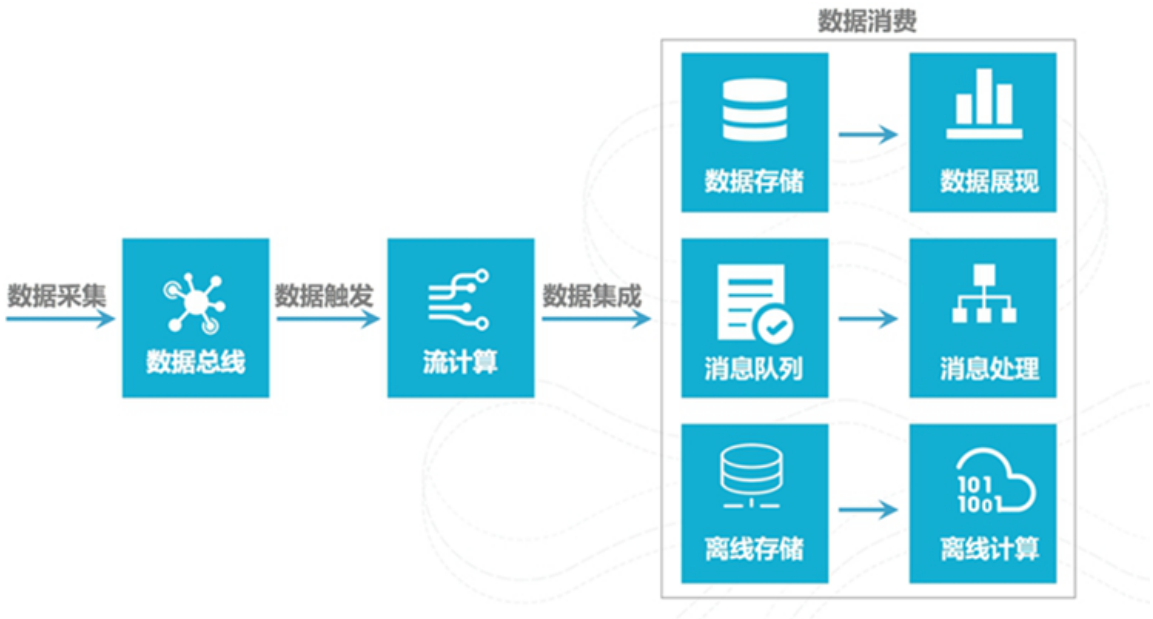
4. **数据消费**

   流式计算的结果数据进入各类数据源后，用户可以使用各类个性化的应用消费结果数据： 用户可以使用数据存储系统访问数据，使用消息投递系统进行信息接收，或者直接使用告警系统进行告警。

36.1.5 流计算全链路

不同于现有的离线/批量计算模型（和批量计算差异性在下一小节细述），流计算全链路整体上更加强调数据的实时性，包括**数据实时采集**、**数据实时计算**、**数据实时集成**。三大类数据的实时处理逻辑在全链路上保证了流式计算的低时延。全链路流计算示意图如图 *144:* 全链路流计算所示。

图 144: 全链路流计算



1. **数据采集**

   用户使用流式数据采集工具将数据**流式且实时**地采集并传输到大数据消息Pub/Sub系统，该系统将为下游流计算提供源源不断的事件源去触发流式计算任务的运行。

2. **流式计算**

流数据作为流计算的触发源驱动流计算运行。因此，**一个流计算任务必须至少使用一个流数据作为数据源**。一批进入的数据流将直接触发下游流计算的一次流式计算处理，并针对但批次流式数据得出计算结果。

3. **数据集成**

   流计算将计算的结果数据直接写入目的数据源，这其中包括多种数据源，包括数据存储系统、消息投递系统，甚至直接对接业务规则告警系统发出告警信息。不同于批量计算（例如阿里云MaxCompute或者开源Hadoop），**流计算天生自带数据集成模块**，**可以将结果数据直接写入到目的数据源。**

4. **数据消费**

   流计算一旦将结果数据投递到目的数据源后，后续的数据消费从系统划分来说，和流计算已经完全解耦。用户可以使用数据存储系统访问数据，使用消息投递系统进行信息接收，或者直接使用告警系统进行告警。

36.1.6 流计算和批量计算区别

相比于批量大数据计算，流（式）计算整体上还属于较为新颖的计算概念，下面我们从用户/产品层面来理解下两类计算方式的区别。需要注意的是，这里的说明并非严谨的科学/理论解释，更详细的理论解析请参看 Wikipedia 的相关章节*Stream Processing*。

36.1.6.1 批量计算

目前绝大部分传统数据计算和数据分析服务均是基于批量数据处理模型： 使用 ETL 系统或者 OLTP 系统进行构造数据源，在线的数据服务（包括 Ad-Hoc 查询、DashBoard 等服务）通过构造 SQL 语言访问上述数据源并取得分析结果。这套数据处理的方法论伴随着关系型数据库在工业界的演进而被广泛采用。但在大数据时代下，伴随着越来越多的人类活动被信息化、进而数据化，越来越多的数据处理要求实时化、流式化，当前这类处理模型开始面临实时化的巨大挑战。传统的批量数据处理通常基于如下处理模型：

1. 使用 ETL 系统或者 OLTP 系统构造原始的数据源，以提供给后续的数据服务进行数据分析和数据计算。即下图，用户装载数据，系统将根据自己的存储和计算情况，对于装载的数据进行索引构建等一系列查询优化工作。因此，对于批量计算，**数据一定需要预先加载到计算系统**，**后续计算系统才在数据加载完成后方能进行计算。**
2. 用户/系统**主动**发起一个计算任务（例如 MaxCompute 的 SQL 任务，或者 Hive 的 SQL 任务）并向上述数据系统进行请求。此时计算系统开始调度（启动）计算节点进行大量数据计算，该过程的计算量可能巨大，耗时长达数分钟乃至于数小时。同时，由于数据累计的不可

及时性，上述计算过程的数据一定是历史数据，无法保证数据的“新鲜”。用户可以根据自己的需要随时调整计算SQL，甚至于使用 AdHoc 查询，可以做到即时修改即时查询。

3. 计算结果返回，计算任务完成后将数据以结果集形式返回用户，或者可能由于计算结果数据量巨大保存着数据计算系统中，用户进行再次数据集成到其他系统。一旦数据结果巨大，整体的数据集成过程漫长，耗时可能长达数分钟乃至于数小时。

批量计算是一种批量、高时延、主动发起的计算任务。 用户使用批量计算的顺序是：

1. 预先加载数据。
2. 提交计算任务，并且可以根据业务需要修改计算任务，再次提交任务。
3. 计算结果返回。

36.1.6.2 流式计算

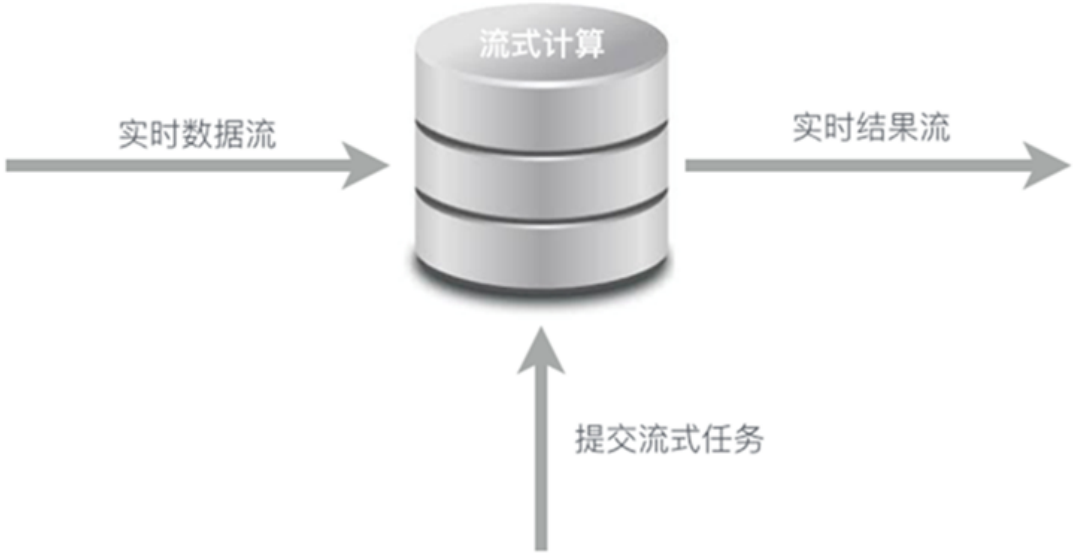
不同于批量计算模型，流式计算更加强调计算数据流和低时延，流式计算数据处理模型如下：

1. 使用实时数据集成工具，将数据实时变化传输到流式数据存储（即消息队列，如 DataHub）；此时数据的传输变成实时化，将长时间**累积**大量的数据平摊到每个时间点不停地小批量实时传输，因此数据集成的时延得以保证。

   此时数据将源源不断写入流数据存储，不需要预先加载的过程。同时流计算对于流式数据不提供存储服务，数据是持续流动，在计算完成后就立刻丢弃。
2. 数据计算环节在流式和批量处理模型差距更大，由于数据集成从**累积**变为**实时**，不同于批量计算**等待数据集成全部就绪后才启动计算任务**，流式计算任务是一种常驻计算**服务**，一旦启动将一直处于**等待事件触发**的状态，一旦有小批量数据进入流式数据存储，流计算立刻计算并迅速得到结果。同时，阿里云流计算还使用了增量计算模型，将大批量数据分批进行增量计算，进一步减少单次运算规模并有效降低整体运算时延。

从用户角度，对于流式任务，必须预先定义计算逻辑，并提交到流式计算系统中。在整个运行期间，流计算作业逻辑不可更改，用户通过停止当前作业运行后再次提交作业，此时之前已经计算完成的数据是无法重新再次计算。

3. 不同于批量计算结果数据需等待数据计算结果完成后，批量将数据传输到在线系统；流式计算任务在每次小批量数据计算后可以立刻将数据写入在线/批量系统，无需等待整体数据的计算结果，可以立刻将数据结果投递到在线系统，进一步做到实时计算结果的实时化展现。



流计算是一种持续、低时延、事件触发的计算任务，用户使用流计算的顺序是：

1. 提交流计算任务。
2. 等待流式数据触发流计算任务。
3. 计算结果持续不断对外写出。

36.1.6.3 模型对比

下表给出了流计算与批量计算两类计算模型的差别。

表 39: 模型对比

对比指标	批量计算	流式计算
数据集成方式	预先加载数据	实时加载数据实时计算

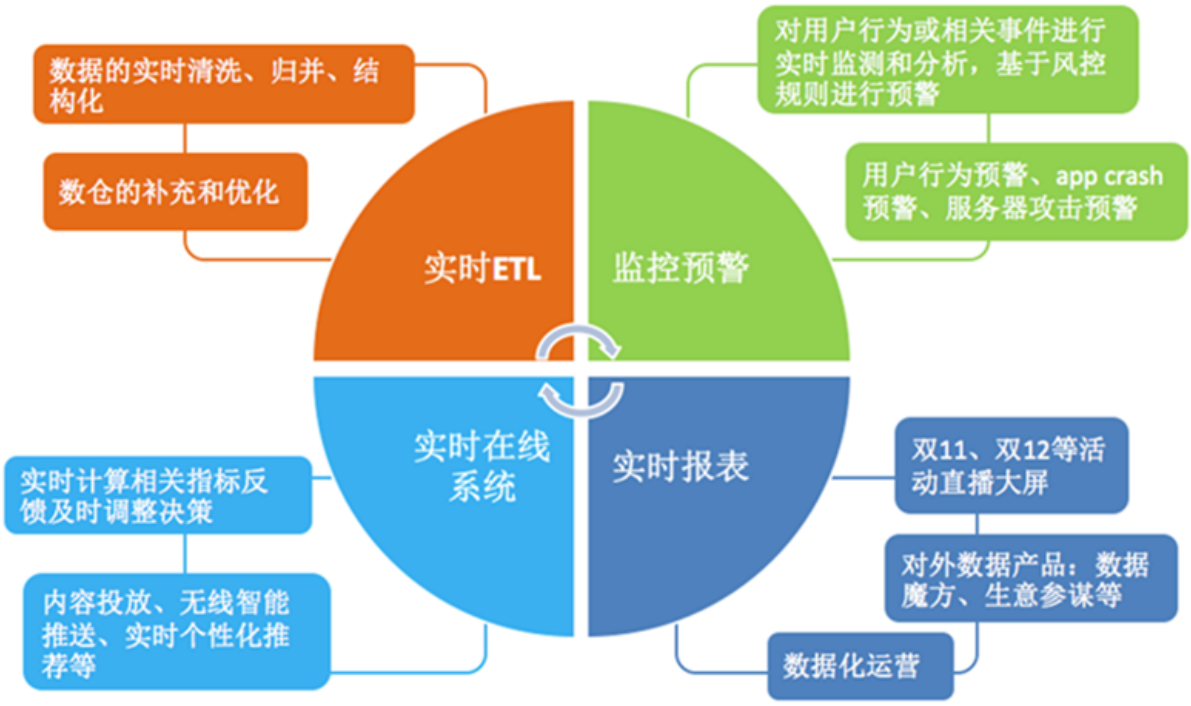
对比指标	批量计算	流式计算
使用方式	业务逻辑可以修改，数据可重新计算	业务逻辑一旦修改，之前的数据不可重新计算（流数据易逝性）。
数据范围	对数据集中的所有或大部分数据进行查询或处理。	对滚动时间窗口内的数据或仅对最近的数据记录进行查询或处理。
数据大小	大批量数据。	单条记录或包含几条记录的微批量数据。
性能	几分钟至几小时的延迟。	只需大约几秒或几毫秒的延迟。
分析	复杂分析。	简单的响应函数、聚合和滚动指标。

在大部分大数据处理场景下，受限于当前流计算的整个计算模型较为简单，流计算是批量计算的**有效增强**，特别在于对于事件流处理时效性上，**流计算对于大数据计算是一个不可或缺的增值服务。**

36.2 典型应用

流计算使用 StreamSQL 主打流式数据分析场景，如图 *145:* 使用场景所示。

图 145: 使用场景



- **实时ETL**

集成流计算现有的诸多数据通道和 SQL 灵活的加工能力，对流式数据进行实时清洗、归并、结构化。作为离线数仓有效的补充和优化，作为数据实时传输的可计算通道。

- **实时报表**

  实时化采集、加工流式数据源，实时监控和展现业务、客户各类指标，让数据化运营实时化。

- **监控预警**

  对系统和用户行为进行实时检测和分析，实时监测和发现危险行为。

- **在线系统**

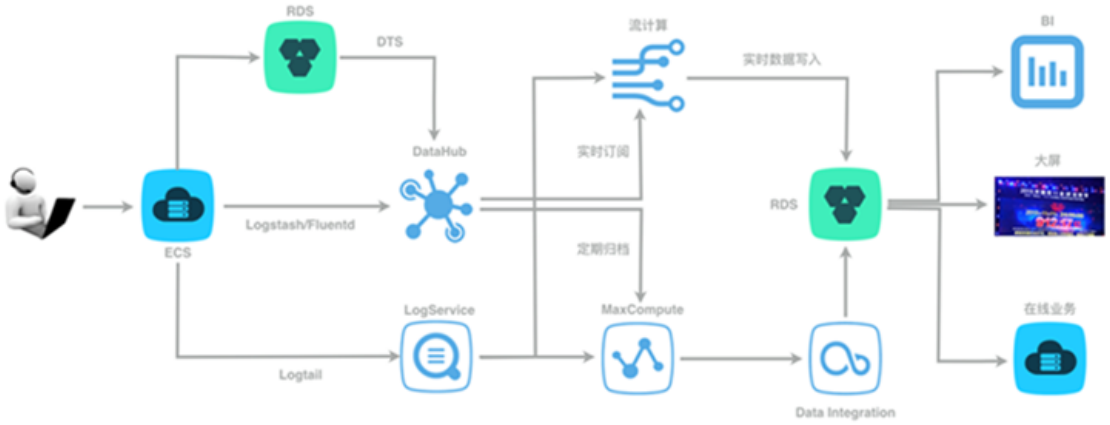
  实时计算各类数据指标，并利用实时结果及时调整在线系统相关策略。在各类内容投放、无线智能推送领域有大量场景。

36.2.1 电商案例

阿里云流计算脱胎于阿里集团电商行业大数据架构，可以说天生适合电商行业各类流式数据分析和报表支持。从电商行业来看，对于流式数据实时处理需求主要集中在：

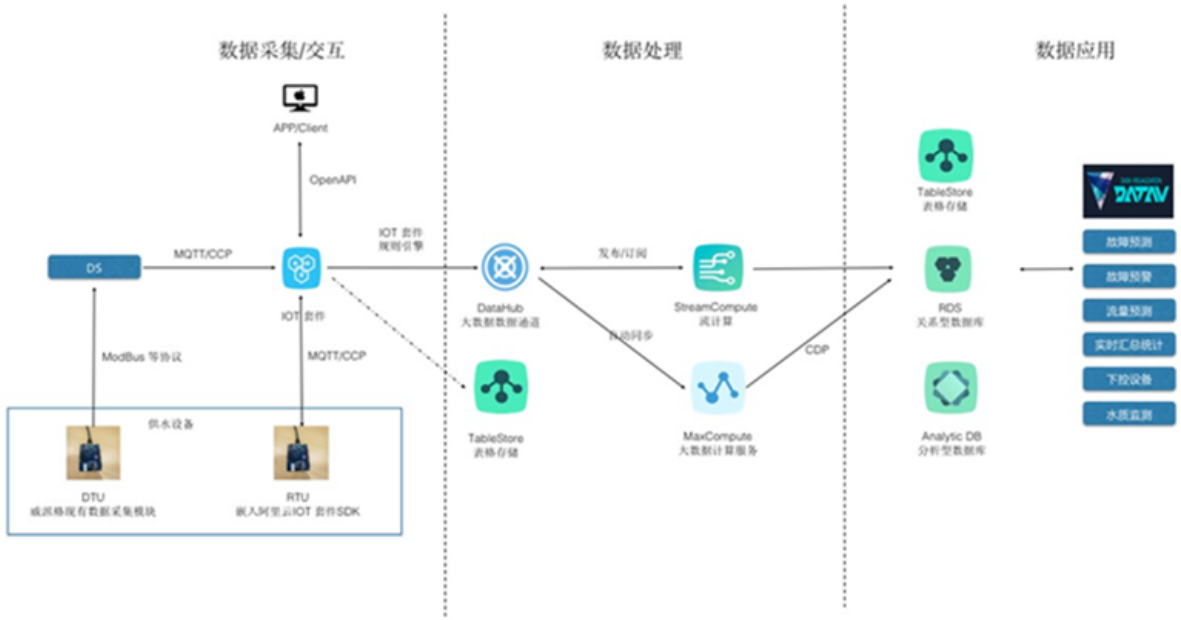
- 对于用户行为的实时分析，例如交易大屏、用户大屏等一系列用户行为大屏展示。使用传统的离线分析不仅整体速度慢时延高，而且可能对于在线系统存在一定压力，从而系统稳定性得不到完全保证。
- 对于用户、业务、系统的实时监控，例如全站交易时段曲线图可以协助运营或者技术人员了解当前时间点全站交易情况，如果交易出现不正常波动（例如突然下跌），应该立刻触发报警机制以方便相关人员介入排查问题，以减少交易波动对于公司业务的巨大影响。
- 对于一些大促活动的实时监控，例如双十一、618等电商大促，运营人员需要实时获取到各类指标信息，用以迅速决定是否需要更换大促运营方案。

使用阿里云流计算，结合阿里云各类计算、存储系统，可以方便支持上述各类个性化的流式数据分析需求。不同于其他数据分析系统，阿里云流计算既满足业务灵活性，同时又采用 SQL 保证业务开发的低门槛。



36.2.2 物联网应用

在 IoT 领域，由于存在大量的实时数据流（物联网场景下传感器产生的数据流更多），同时核心关键业务对于数据监控的时效性要求非常高，在这类领域流计算、实时计算的应用场景将非常广泛。阿里云流计算提供全链路的流式数据加工、生产服务，极大方便 IoT 开发者能够迅速搭建一整套完整的实时流数据监控、分析系统，助力 IoT 的实时大数据业务发展。



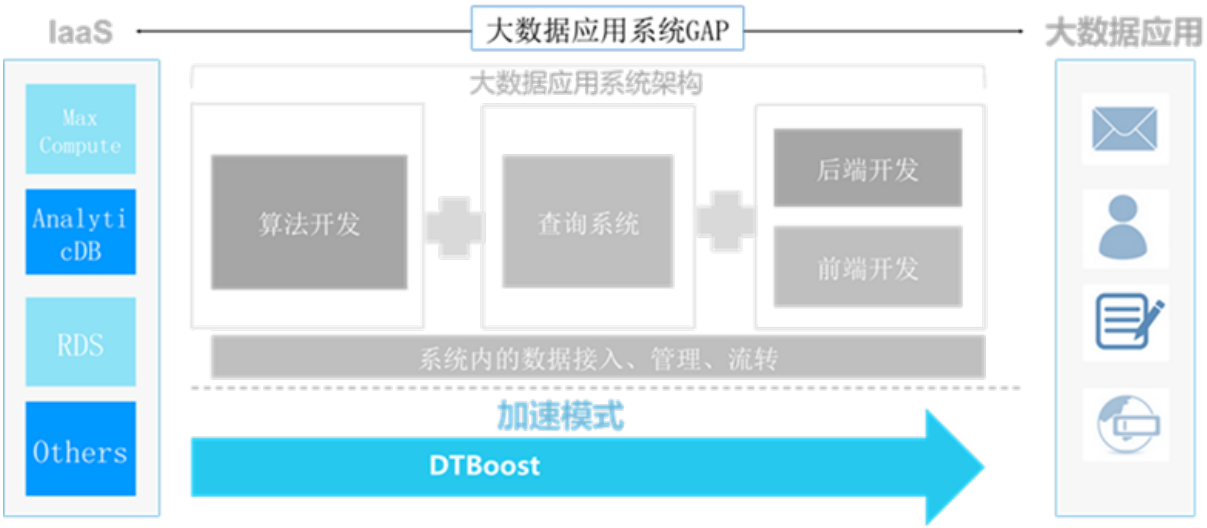
在 IoT 领域，由于存在大量的实时数据流（物联网场景下传感器产生的数据流更多），同时核心关键业务对于数据监控的时效性要求非常高，在这类领域流计算、实时计算的应用场景将非常广泛。阿里云流计算提供全链路的流式数据加工、生产服务，极大方便 IoT 开发者能够迅速搭建一整套完整的实时流数据监控、分析系统，助力 IoT 的实时大数据业务发展。

37 大数据应用加速器

37.1 产品概述

阿里云 DTBoost 数据加速器产品从大数据应用落地点出发，提供了一套大数据应用开发套件，能够帮助开发者从业务需求的角度有效的整合阿里云各个大数据产品，大大降低搭建大数据应用系统当中绝大部分的系统工程工作，在相应行业应用解决方案的结合下，能够让不是很熟悉大数据应用系统开发的程序员也能够快速为企业搭建大数据应用，从而实现大数据价值的快速落地。

图 146: 大数据应用系统

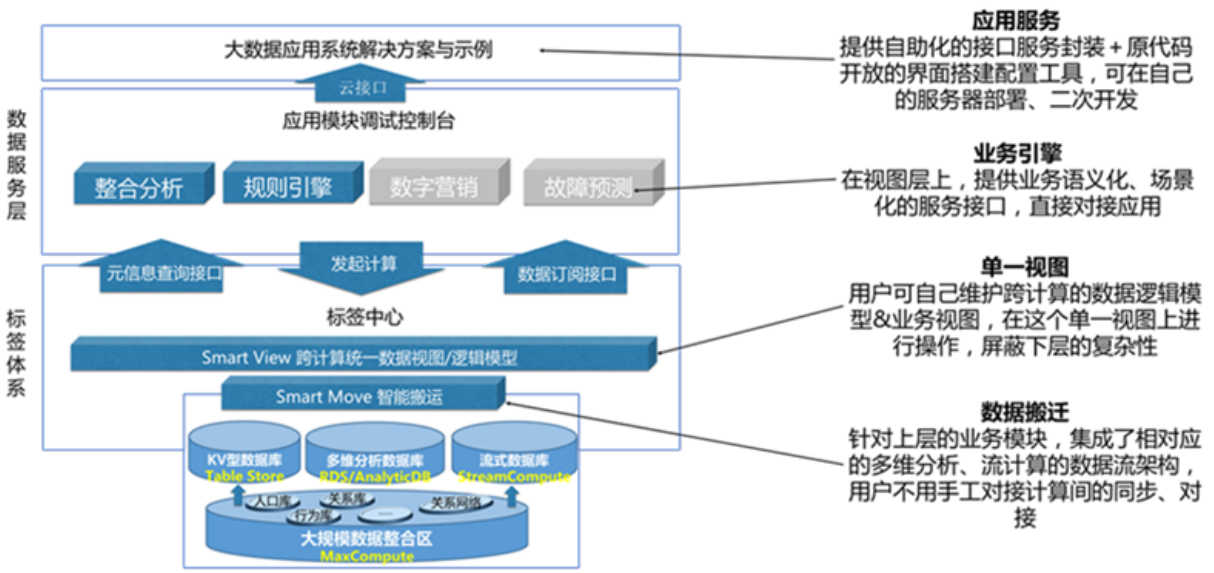


概括来讲，DTBoost 是以标签中心为基础，建立跨多个云计算资源之上的统一逻辑模型，开发者可以在“标签”这种逻辑模型视图上结合画像分析、规则预警、文本挖掘、个性化推荐、关系网络等多个业务场景的数据服务模块，通过接口的方式进行快速的应用搭建。这种方式的好处一在于屏蔽掉应用开发人员对于下层多个计算存储资源的深入理解与复杂的系统对接工作，二在于通过数据服务的形式透出也有助于 IT 部门对数据使用的管理，避免资源的重复和冗余。

简单来说，因为大数据计算能力的增强，开发者只需要把需要使用的数据在模型当中进行管理后，即可通过 API 方式进行相应的计算对接到产品界面端上，或通过提供的界面配置功能直接生成可以独立部署的代码快速搭建相应的大数据产品。

37.2 产品架构

图 147: DTBoost架构图



37.3 功能特性

37.3.1 标签中心

以逻辑模型的建立提到耗时耗力的传统数据仓库物理模型，为分布在多个计算存储资源上的明细数据建立统一的“实体-关系-标签”模型，并将数据一键加速整合入阿里云分析型数据库等在线分析库当中，为分析应用做好数据准备。

功能简介

- 数据资源整合

  标签中心提供一种业务视角的数据发现、模型探索的工具，便于业务人员、开发人员、数据管理人员透视企业的数据资产。

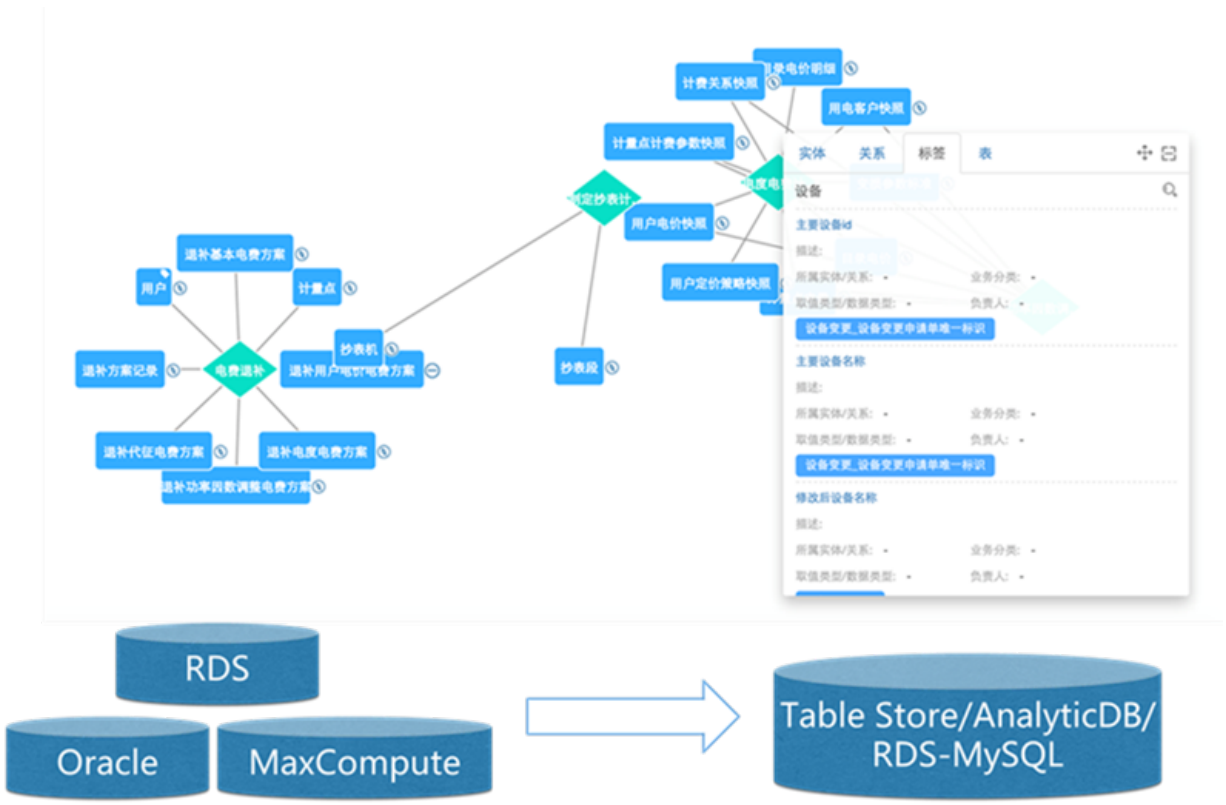
- 为数据服务提供视图支撑

  为多个计算引擎上的数据提供一个统一的数据视图，结合数据服务能够屏蔽下层多个计算资源的对接与数据同步。

- 数据访问管理

  可以通过逻辑层对数据访问权限进行有效控制，比物理表的访问管理更加安全有效。

图 148: DTBoost功能



产品特性

- 业务视角管理

  围绕实体-关系-标签这三个元素进行建模，是从业务的角度出发对数据进行组织管理，而不是从表的概念出发进行建模，便于应用层对数据运用和管理的理解、操作，以近似于概念模型的形态透出，让人人都能看得懂。

- 跨计算的统一逻辑模型

  传统建模的数据来源和模型的使用一般在同一数据库当中，而大数据环境下因为数据采集类型的多样性，和数据计算的多样性使得来源和使用分散在不同的计算存储资源当中，数据产生与加工首先就可能分布在不同的数据库当中，其次同一份数据需要进行跨流式、Adhoc 类多维分析、离线算法加工等多种方式的计算，数据需要能在多个存储和计算资源当中自由流转。

- 灵活拓展性

  表/标签之间的逻辑关系的建立也是在逻辑层上完成的，这就使得模型的维护是可以动态建设的，便于模型的维护和管理，而无需在物理层将数据进行归并后再使用。每一个标签之间可以独立使用，这种离散的列化操作方式也使的数据的使用上更为灵活。

37.3.2 整合分析

在逻辑模型上像分析“大宽表”一样进行灵活自由的分析表达，通过快速自主封装可实时计算的分析接口头透出到前端，或是在线可视化界面配置生成可以独立部署交互式分析应用代码，实现敏捷的交互式分析应用搭建。

功能简介

- 能够提供可视化的界面配置工具，在标签中心所管理的模型基础上，生成交互式分析应用；
- 能够支持交互式筛选、钻取、计数、列表、统计分布、地理分布等分析操作。内置多种可视化图表，包括常规条形图、饼状图、折线图、地图、表格等；
- 拖拽查询配置支持多种查询模式，支持子查询的配置。支持筛选数据导出配置，方便与其他系统打通；
- 所生成应用以源代码方式提供，能够自由修改界面样式，独立部署。

图 149: 整合分析功能

图 150: 分析结果

产品特性

表 40: 一般 BI 工具和整合分析对比表

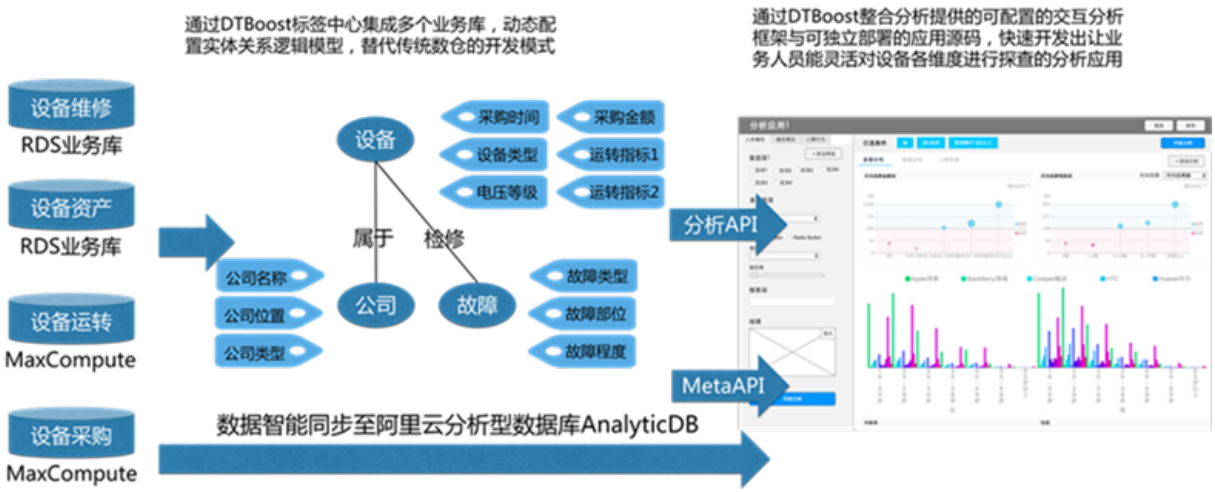
-	一般 BI 工具	整合分析
查询灵活度	一般来说以一层关联较多	与标签中心模型配合，能够自定义拖拽配置多层子查询
分析交互	以固定报表查看为主	强化交互式 SQL 模型不定由用户自己选择的分析
可拓展性	拓展性较低，以固定产品	开放应用层的源代码，可以自行进行集成拓展
系统集成性	一般作为独立分析工具，不与其他系统集成	能够方便的进行对下游系统的对接（如广告投放系统）

37.4 典型应用

- 结构化数据全量数据200G以上；
- 希望把多个业务系统/小数仓的数据进行整合打通；
- 分析场景围绕一个主体展开（人、设备、车辆.....）；
- 数据维度（包括衍生出来的维度）超过10个以上；

- 业务人员希望自己来进行分析，而不是只看固定报表。

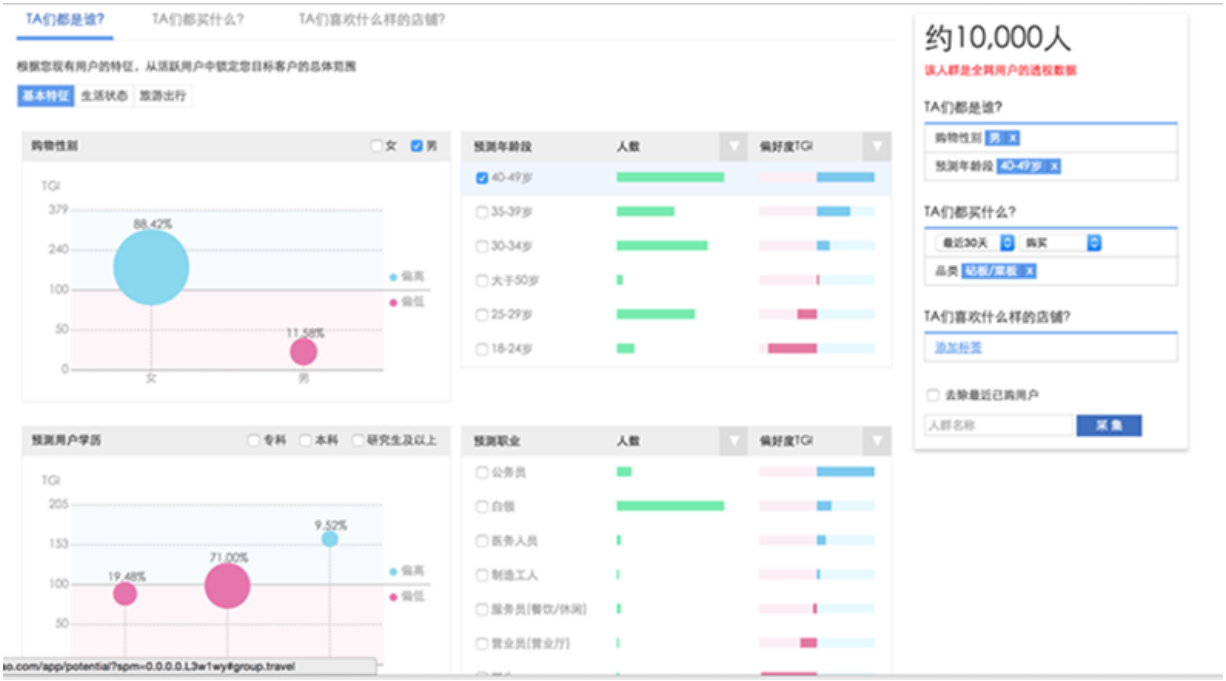
图 151: DTBoost应用



37.4.1 画像分析

整合用户收藏、成交、点击、注册信息、定位、以及衍生加工的标签等多份数据，全方位分析用户各种行为之间的关联性，从而更有效的设计交叉销售、营销内容、人群定向等运营策略。

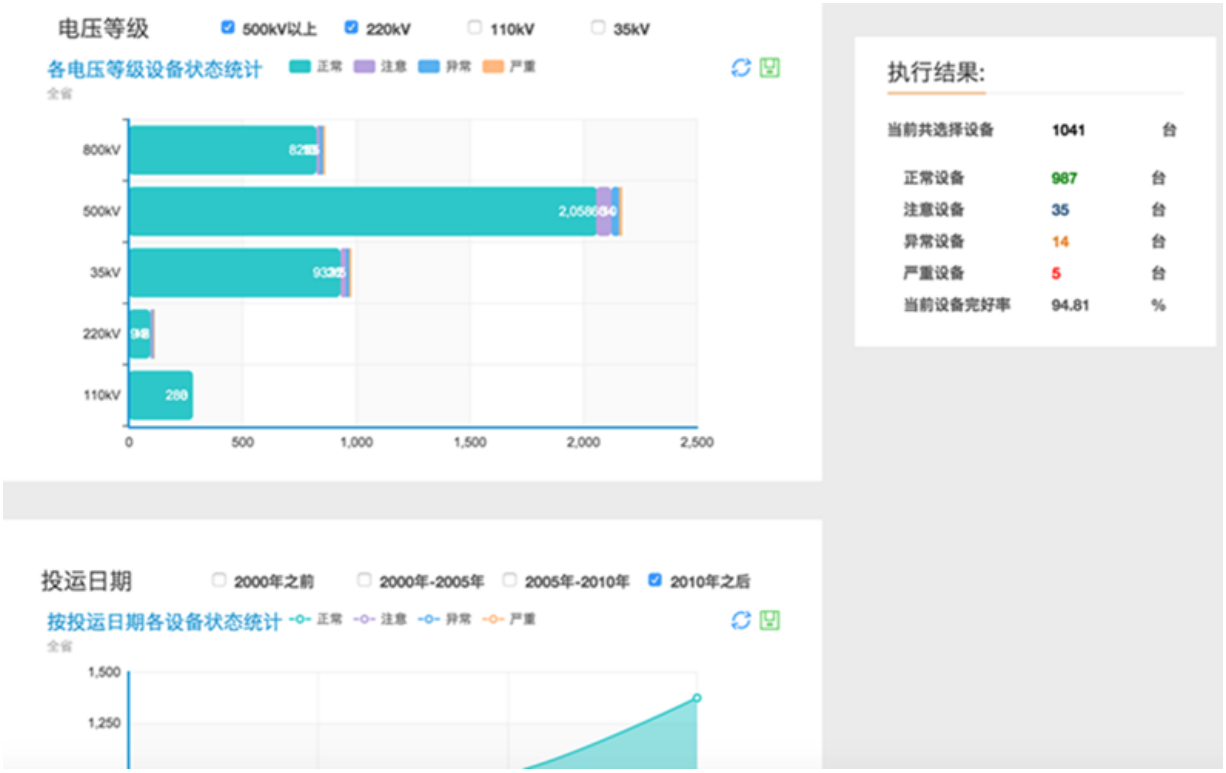
图 152: 画像分析



37.4.2 设备履历

打通设备在采购、运维、检修、报废、技改等多个环节的数据，能全方位的对设备资产情况进行分析梳理，以及剖析各种外界的数据对设备状况的影响，大大提升设备资产管理水平。

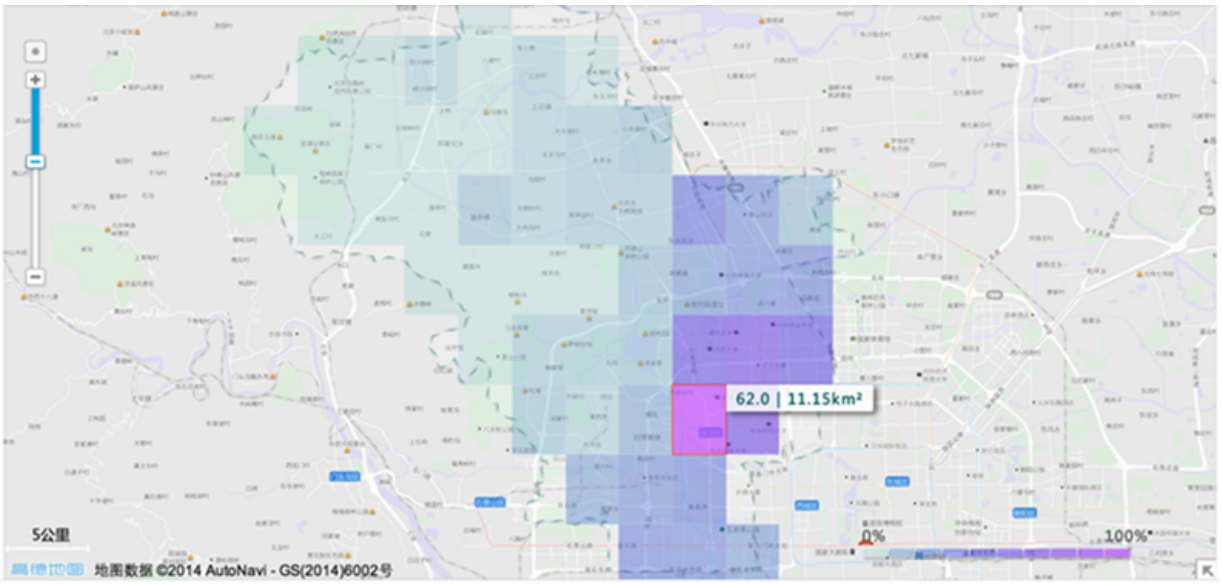
图 153: 设备履历



37.4.3 地理分析

结合地理位置信息，无需在每个地理网格上的分布提前进行预计算，直接在 POI 明细数据之上进行多种多样的筛选、汇总分析，大大提高了时空数据分析的灵活度。

图 154: 地理分析


62.0 | 11.15km²
5公里
0%
100%
地图数据 ©2014 AutoNavi - GS(2014)6002号

38 大数据管家

38.1 产品概述

大数据管家（Bigdata Cloudconsle，以下简称BCC），是为阿里巴巴专有云大数据产品定制的运维产品。通过大数据管家目前可以管理MaxCompute、AnalyticDB、StreamCompute、BaseIDE等产品。

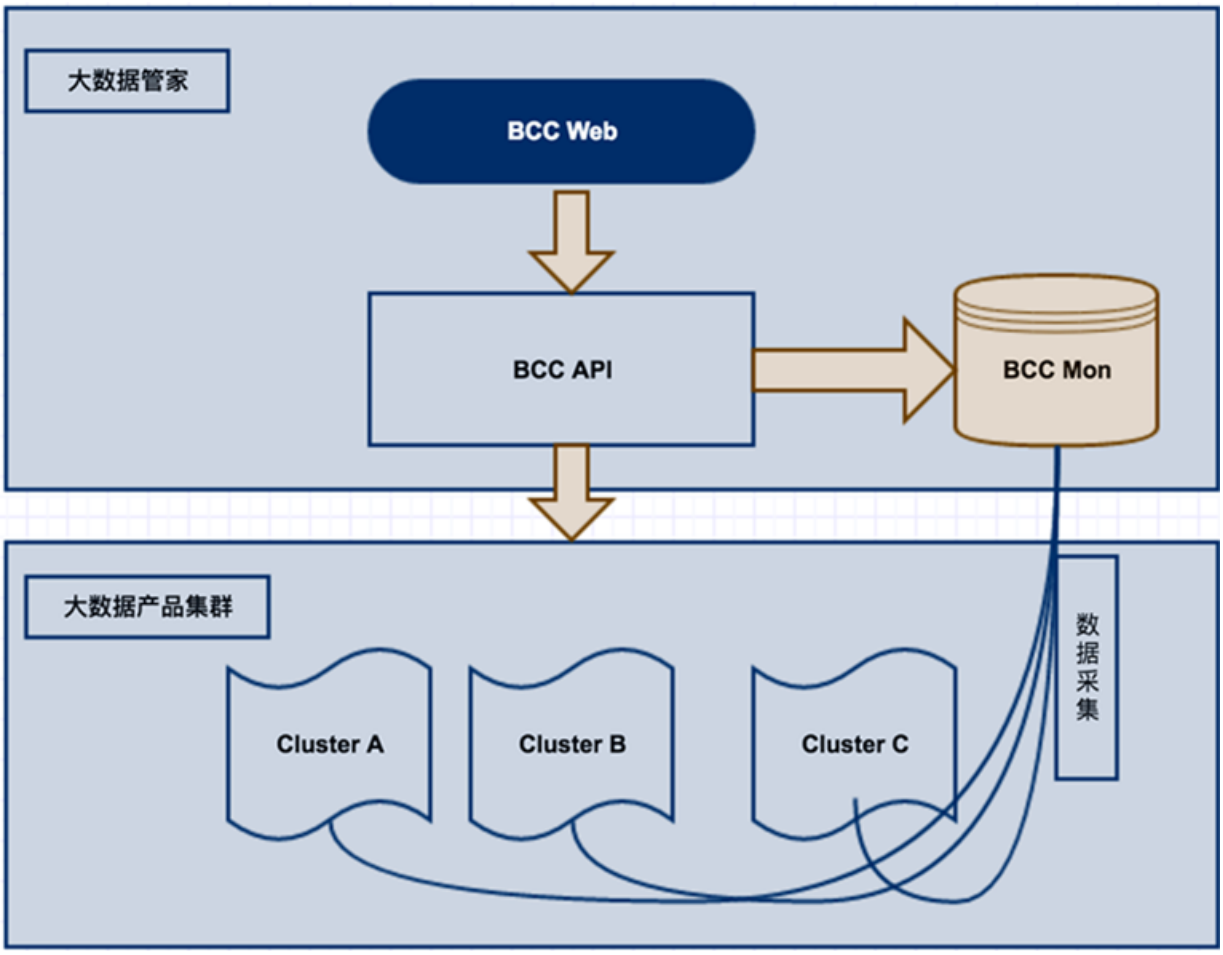
通过大数据管家的赋能，专有云驻场人员可以轻松地管理大数据产品，如查看大数据产品的运行指标，修改大数据产品运行配置，对大数据产品进行自检，搜索日志等功能。

38.2 产品架构

大数据管家组成部分：

- BCC WEB（前端）
- BCC API（管控API）
- BCC Mon（数据采集）

图 155: BCC组成部分



38.3 功能特性

下面对大数据管家（BCC）的主要功能做一些介绍，详细的功能使用可以参考《运维指南》中的**大数据管家**章节。

38.3.1 业务概览

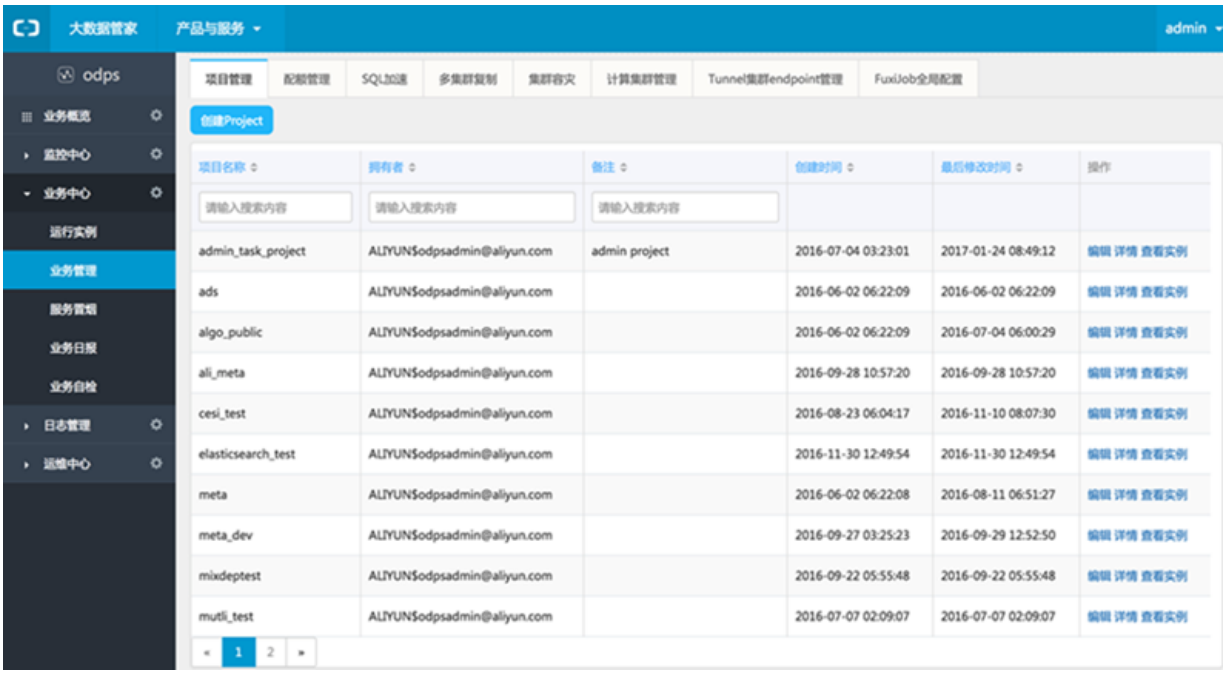
概览页面显示对应集群的几项重要的信息，包括计算配额的CPU利用率、操作系统平均负载、磁盘利用率、网络流量、集群内机器数量、盘古物理容量（已使用和总容量）、系统任务数量、盘古已创建文件数量及部分数据的一周内变化比例。概览能够直观得展示当前集群的状况，集群的切换通过集群选择下列表中选择。

图 156: 业务概览图

38.3.2 业务管理

业务管理是项目管理、配合管理、SQL加速、多集群复制、集群容灾、计算集群管理、Tunnel集群endpoint管理、FuxiJob全局配置的总入口。

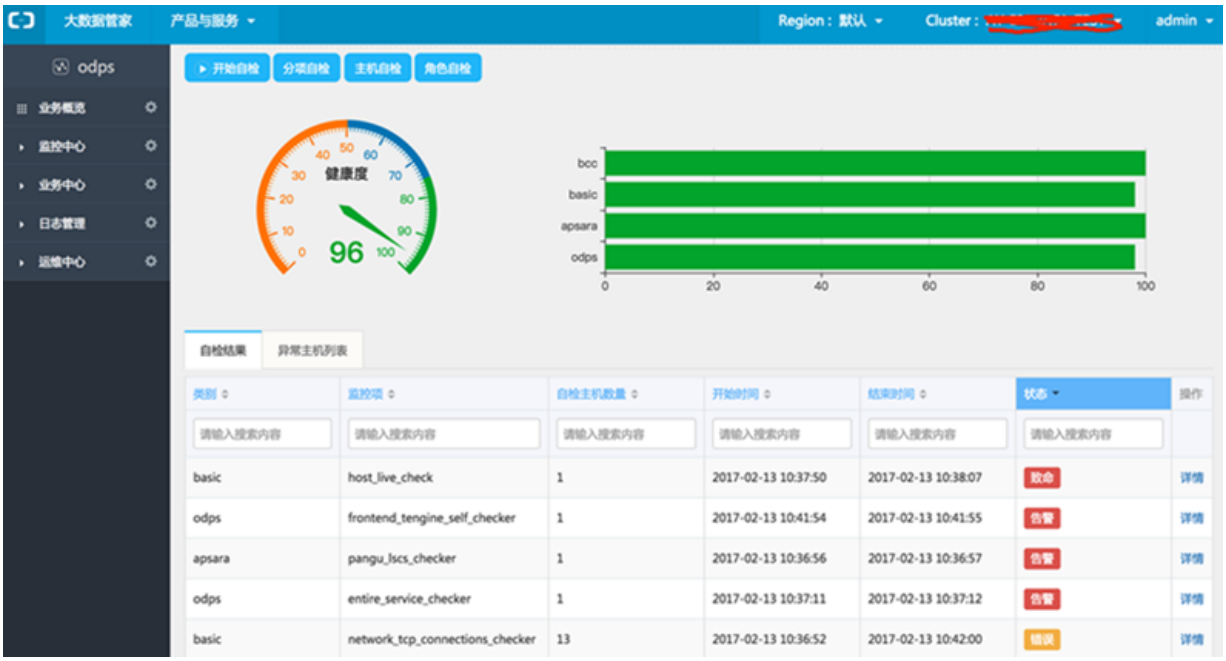
图 157: 业务管理图



38.3.3 业务自检

业务自检是指可以对飞天、AnalyticDB、MaxCompute、Galaxy等业务进行健康检查，并按分值（满分100）作出健康度的判断。

图 158: 业务自检图



38.3.4 系统监控

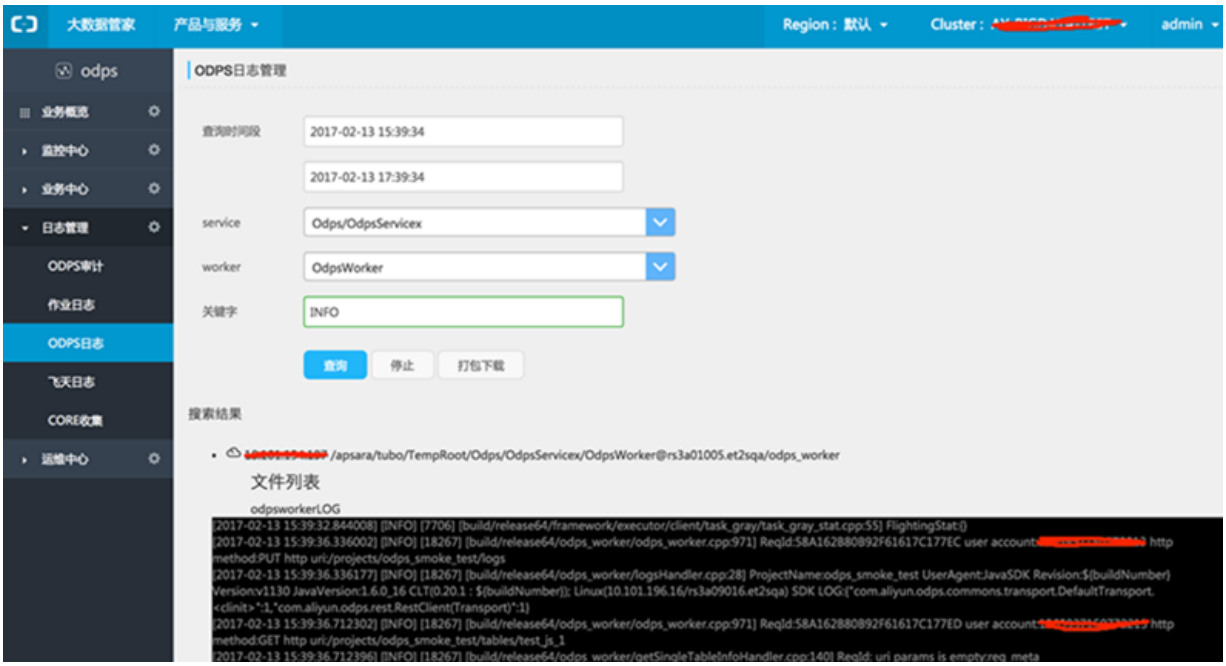
系统监控包括了集群CPU、内存、磁盘IO、网络等的使用量。可选择日期查看集群使用量趋势。

图 159: 系统监控图

38.3.5 日志管理

日志管理可以拿到各种service中的各个worker的日志，可以根据关键字筛选打包下载。

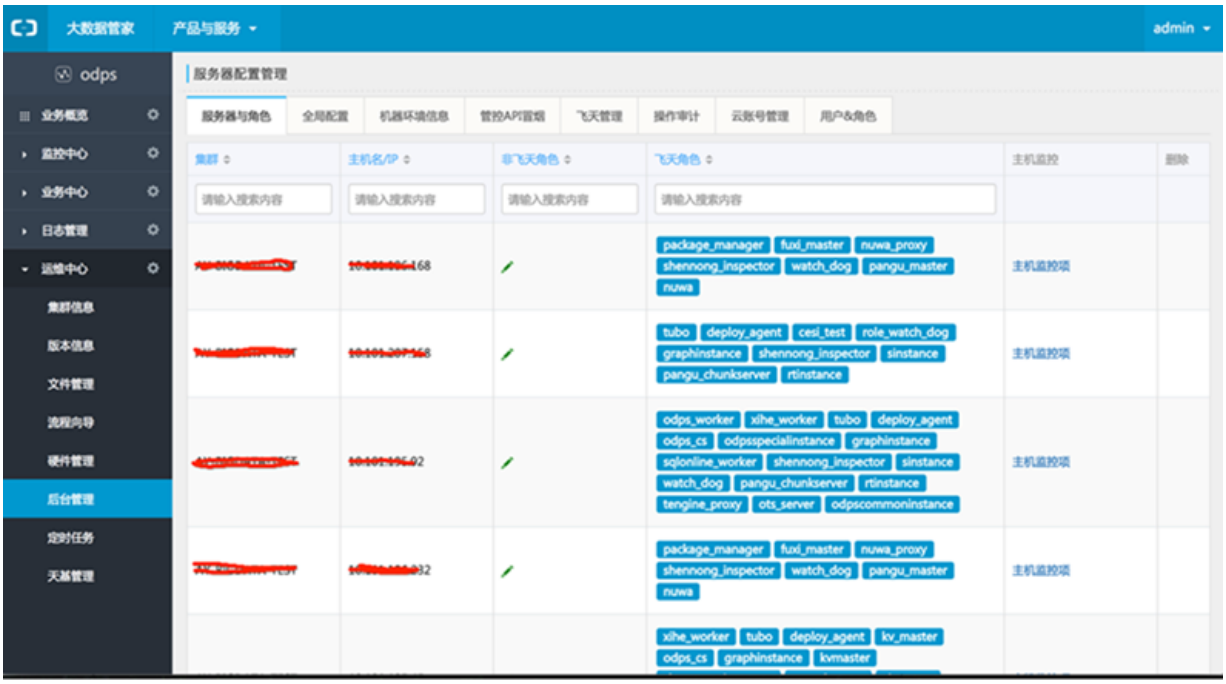
图 160: 日志管理图



38.3.6 机器管理

后台管理是服务器与角色、全局配置、机器环境信息、管控API冒烟、飞天管理、操作审计、云账号管理、用户&角色的总入口。

图 161: 机器管理图



39 Quick BI

39.1 产品概述

Quick BI是一个基于云计算的灵活的轻量级的自助BI工具服务平台。

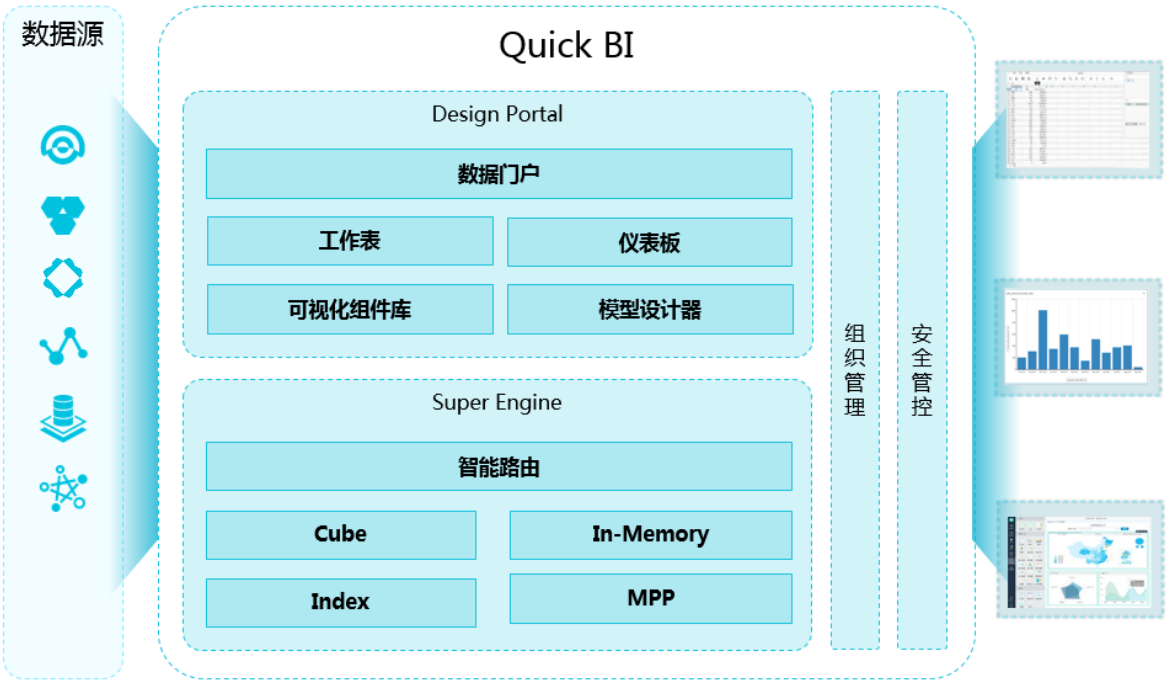
Quick BI支持众多种类的数据源，既可以连接MaxCompute（ODPS）、RDS、AnalyticDB、HybridDB（Greenplum）等云数据源，也支持连接ECS上您自有的MySQL数据库。Quick BI可以为您提供海量数据实时在线分析服务，通过提供智能化的数据建模工具，极大降低了数据的获取成本和使用门槛，通过支持拖拽式操作和提供丰富的可视化图表控件，帮助您轻松自如地完成数据透视分析、自助取数、业务数据探查、报表制作和搭建数据门户等工作。

Quick BI不止是业务人员看数据的工具，更能让每个人都成为数据分析师，帮助企业实现数据化运营。

39.2 产品架构

Quick BI产品架构如下图所示。

图 162: Quick BI架构



Quick BI的主要模块和相关功能。

- **数据连接模块**

  负责适配各种云数据源，包括但不限于MaxCompute、RDS（MySQL、PostgreSQL、SQL Server）、AnalyticDB、HybridDB（MySQL、PostgreSQL）等，封装数据源的元数据/数据的标准查询接口。

- **数据预处理模块**

  负责针对数据源的轻量级 ETL 处理，目前主要是支持MaxCompute的自定义SQL功能，未来会扩展到其他数据源。

- **数据建模**

  负责数据源的OLAP建模过程，将数据源转化为多维分析模型，支持维度（包括日期型维度、地理位置型维度）、度量、星型拓扑模型等标准语义，并支持计算字段功能，允许用户使用当前数据源的SQL语法对维度和度量进行二次加工。

- **工作表**

  负责在线电子表格（webexcel）的相关操作功能，涵盖行列筛选、普通/高级过滤、分类汇总、自动求和、条件格式等数据分析功能，并支持数据导出，以及文本处理、表格处理等丰富功能。

- **仪表板**

  负责将可视化图表控件拖拽式组装为仪表板，支持线图、饼图、柱状图、漏斗图、树图、气泡地图、色彩地图、指标看板等17种图表，支持查询条件、TAB、IFRAME和文本框4种基本控件，支持图表间数据联动效果。

- **数据门户**

  负责将仪表板拖拽式组装为数据门户，支持内嵌链接（仪表板）和外嵌链接（第三方URL），支持模板和菜单栏的基本设置。

- **QUERY引擎**

  负责针对数据源的查询过程。

- **组织权限管理**

  负责 <组织-工作空间> 的两级权限架构体系管控，以及工作空间下的用户角色体系管控，实现基本的权限管理，实现不同人看不同报表。

- **行级权限管理**

  负责数据的行级粒度权限管控，实现不同人看同一张报表展现不同数据。

- **转让/分享/公开**

支持将工作表、仪表板、数据门户转让或分享给其他登录用户访问，支持将仪表板公开到互联网供非登录用户访问。

39.3 功能特性

Quick BI提供以下功能：

无缝集成云上数据库

支持阿里云多种数据源，包括但不限于MaxCompute、RDS（MySQL、PostgreSQL、SQL Server）、AnalyticDB、HybridDB（MySQL、PostgreSQL）等。

图表

丰富的数据可视化效果。系统内置柱状图、线图、饼图、雷达图、散点图等17种可视化图表，满足不同场景的数据展现需求，同时自动识别数据特征，智能推荐合适可视化方案。

分析

多维数据分析。基于Web页面的工作环境，拖拽式、类似于Excel的操作方式，一键导入、实时分析，可以灵活切换数据分析的视角，无需重新建模。

快速搭建数据门户

拖拽式操作、强大的数据建模、丰富的可视化图表，帮助您快速搭建数据门户。

实时

支持海量数据的在线分析，您无需提前进行大量的数据预处理，大大提高分析效率。

安全管控数据权限

内置组织成员管理，支持行级数据权限，满足不同人看不同的报表，以及同一份报表不同人看到不同数据的需求。

39.4 产品优势

Quick BI的总体优势可总结为多，块，强大和易用。

多

支持RDS、MaxCompute、AnalyticDB等多种数据源。

快

亿级数据秒级响应。

强大

内置完整的电子表格工具，可以让您轻松完成复杂的中国式报表的制作。

易用

丰富的数据可视化功能，自动识别数据特征，自动智能为您生成最合适的图表。

39.5 应用场景

39.5.1 编辑数据集

背景信息

数据集创建好以后，您可以根据图表的实际需求来简单编辑一下数据集，比如，切换字段类型或者新增计算字段等等。

如果您想了解如何创建数据集，请参阅《*Quick BI*用户指南》。

以下示例均以company_sales_record为例。

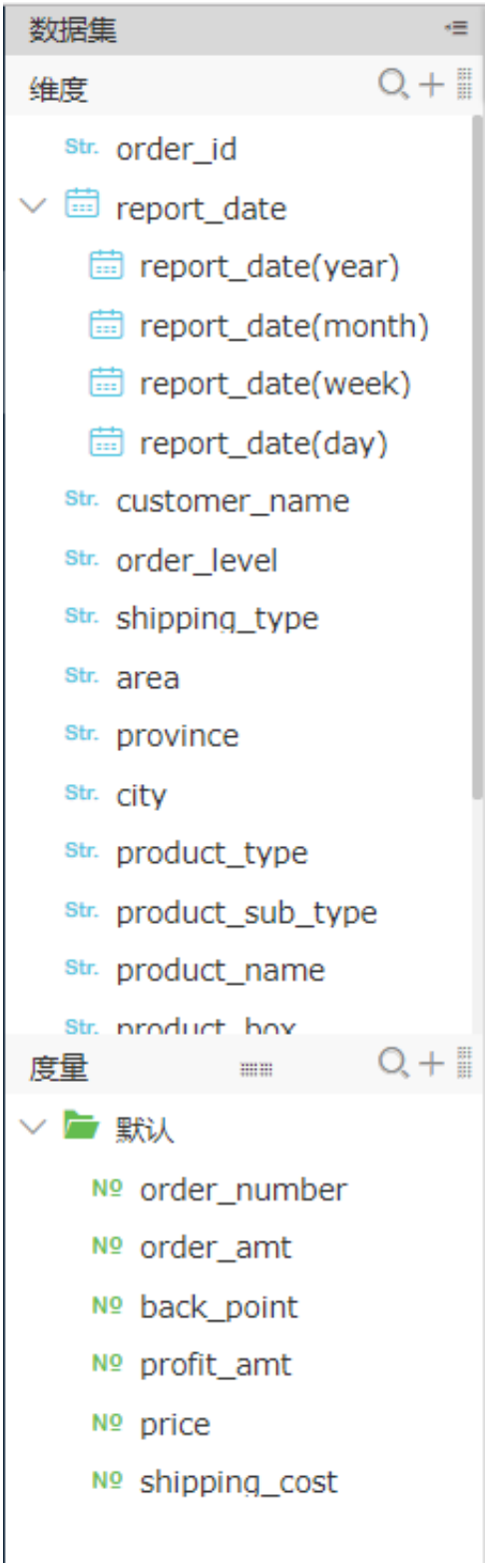
操作步骤

1. 在数据集列表中，找到需要编辑的数据集，比如company_sales_record，单击后面的**编辑**，进入到数据集的编辑页面。

   图 163: 编辑数据集

2. 系统会按照预设将数据集中的字段分别列在维度列表和度量列表中。

   图 164: 维度/度量字段列表

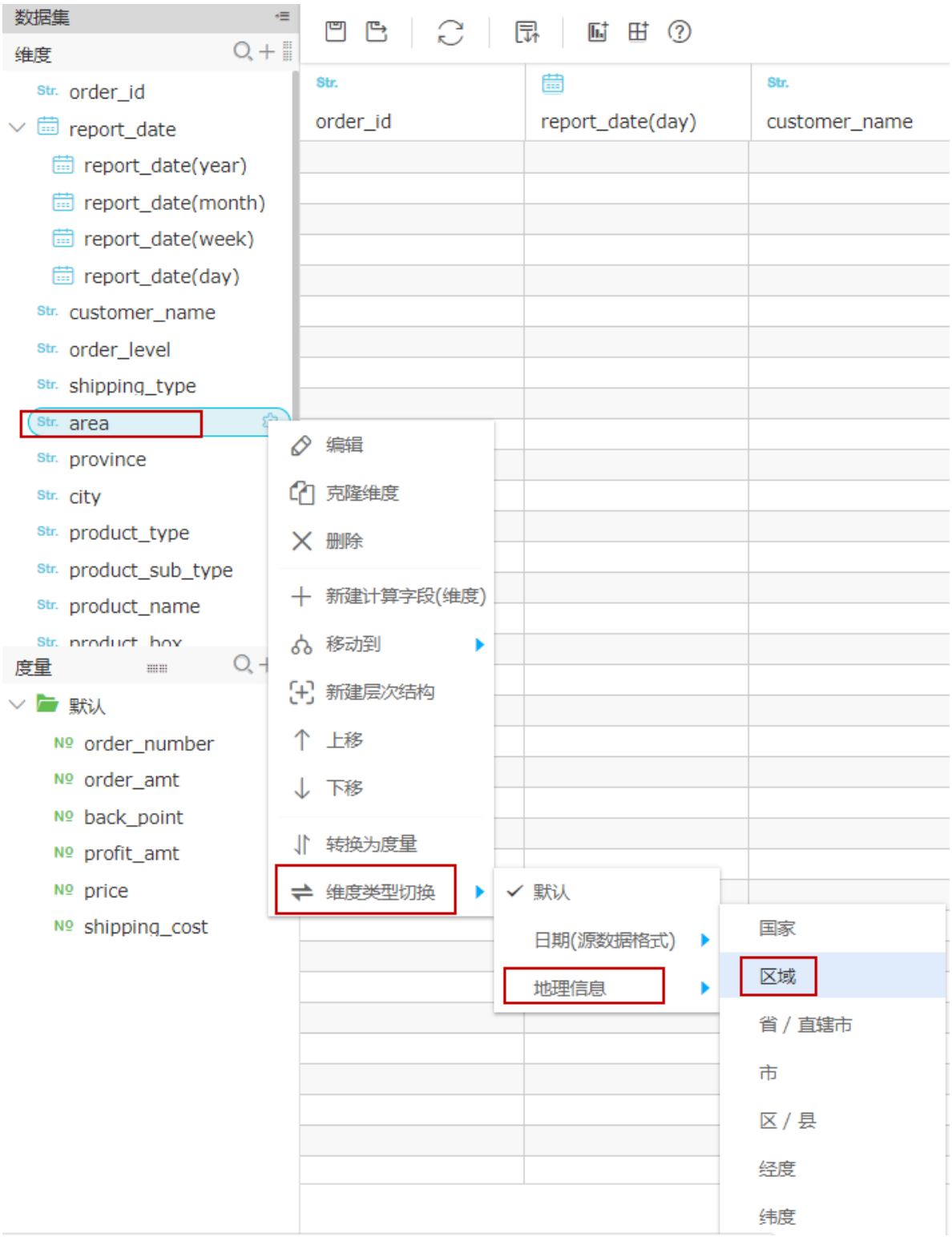


制作图表时，如果您需要用到地图类的图表，比如气泡地图和色彩地图，那么您需要找到含有地理信息的维度字段，并将它们的字段类型从字符串切换为地理信息，否则，地图将无法展示。

3. 在维度列表中，找到area （区域），单击鼠标右键，在下拉菜单中选择**维度类型切换 > 地理信息 > 区域。**

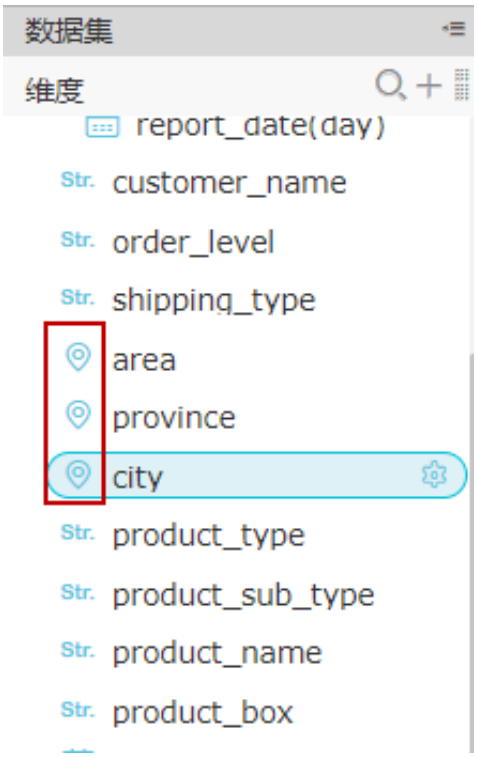
   说明: 切换地理信息时，地理信息的选项一定要与字段完全匹配。比如，字段为area （区域），那么在地理信息列表中，也一定要选择区域，这样字段类型才能被切换。

图 165: 切换维度字段类型



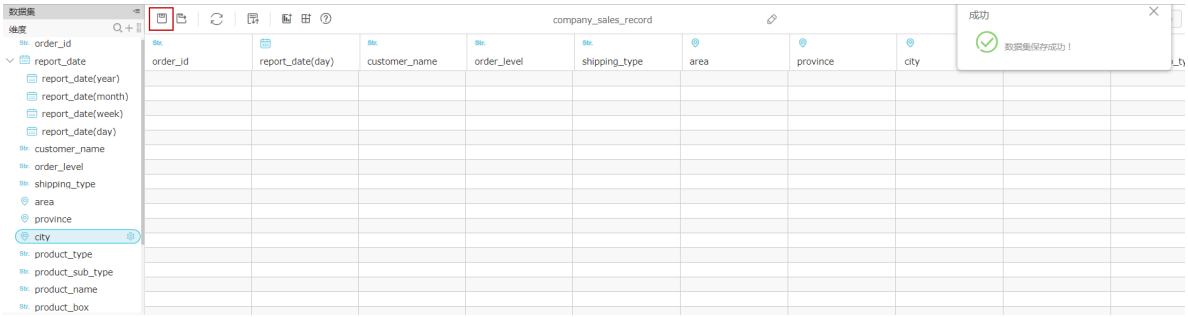
4. 同理，将province （省份）和city（城市）也用同样的方式切换。切换完成后，三个维度字段前面会出现地理信息的标识。

图 166: 切换维度字段类型



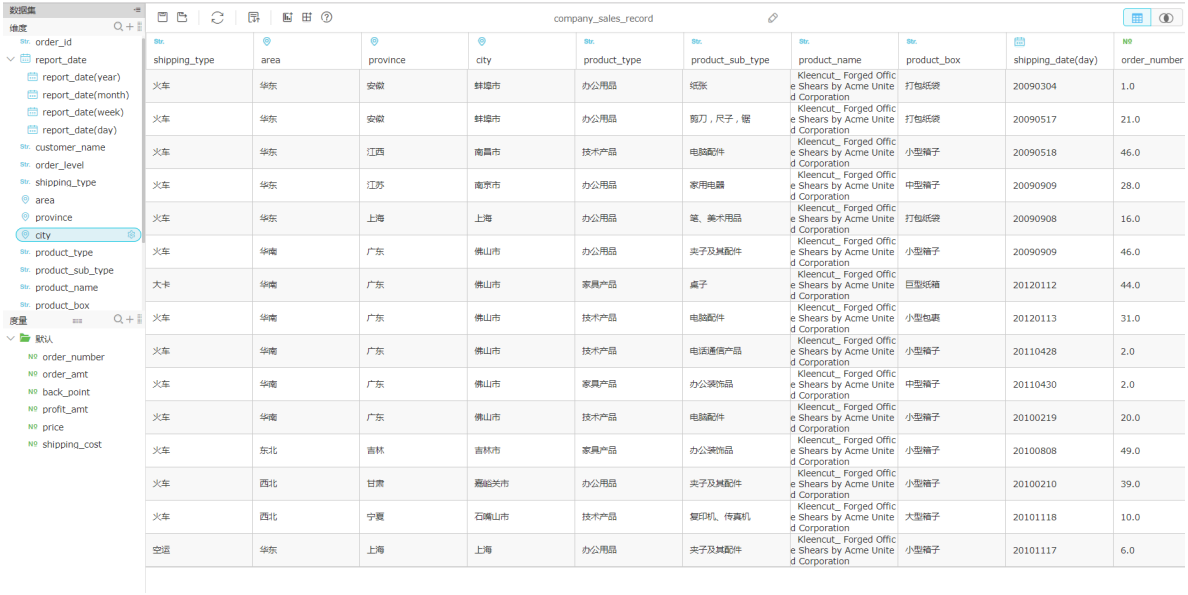
5. 数据集编辑完成后，单击**保存**。

图 167: 保存数据集



6. 单击**刷新**，系统会自动将数据显示在表格中，方便您查看编辑后的效果。

图 168: 刷新数据集



39.5.2 制作数据图表

背景信息

本章节只展示柱图的制作流程，如果您想了解其它图表的制作流程，请参阅《*Quick BI*用户指南》。

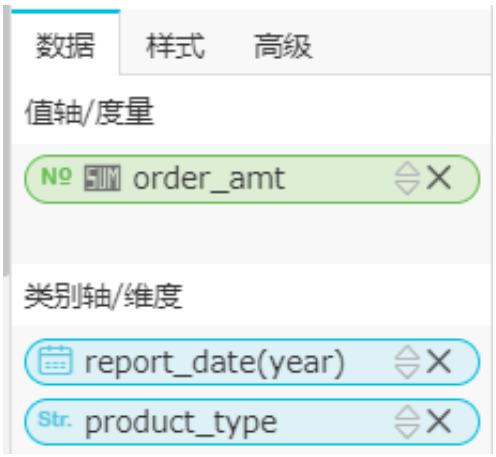
操作步骤

1. 在仪表板配置区，选择一张需要的图表，双击其图标，选中的图表会自动显示在仪表板展示区。

图 169: 选择图表

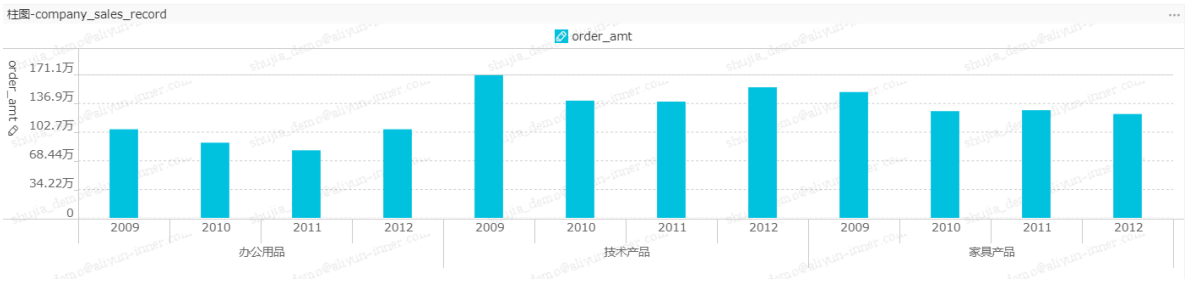
2. 在数据标签页中，根据图表的需求，从维度和度量列表中选择需要的数据，双击数据名称，数据会自动填充到指定区域。

图 170: 选择字段



3. 数据选择完成后，单击**更新**，更新后的图表展示如下。

图 171: 更新后的柱图

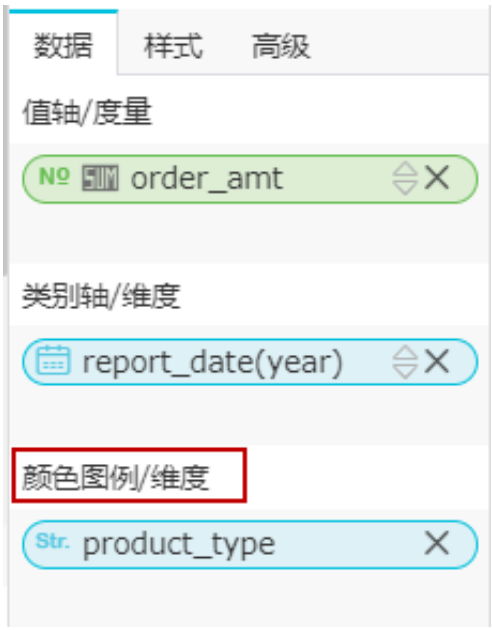


启用颜色图例

4. 如果展示的数据较多，您还可以启用颜色图例功能。将一个维度字段拖拽到颜色图例区域，该字段的内容将会以不同的颜色展示到图表中。

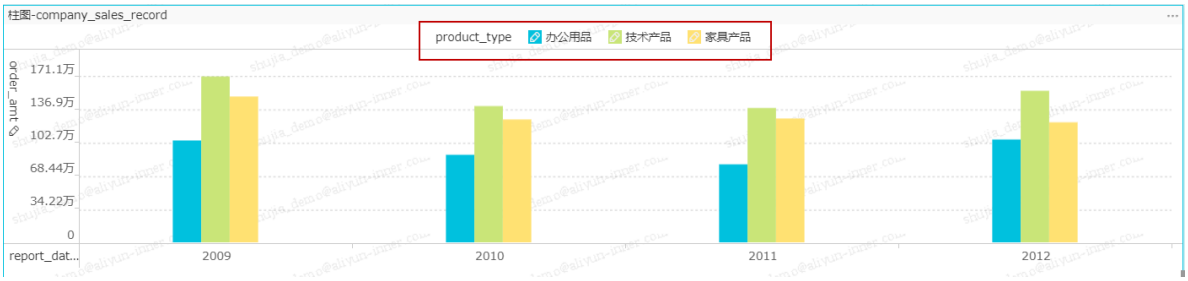
> **说明:** 当值轴区域只有一个度量字段时，颜色图例功能才可用，否则，此功能禁用。当颜色图例功能不能使用时，系统会自动给出提示和原因，您可以根据系统提示，手动调整度量字段和维度字段。

图 172: 启用颜色图例



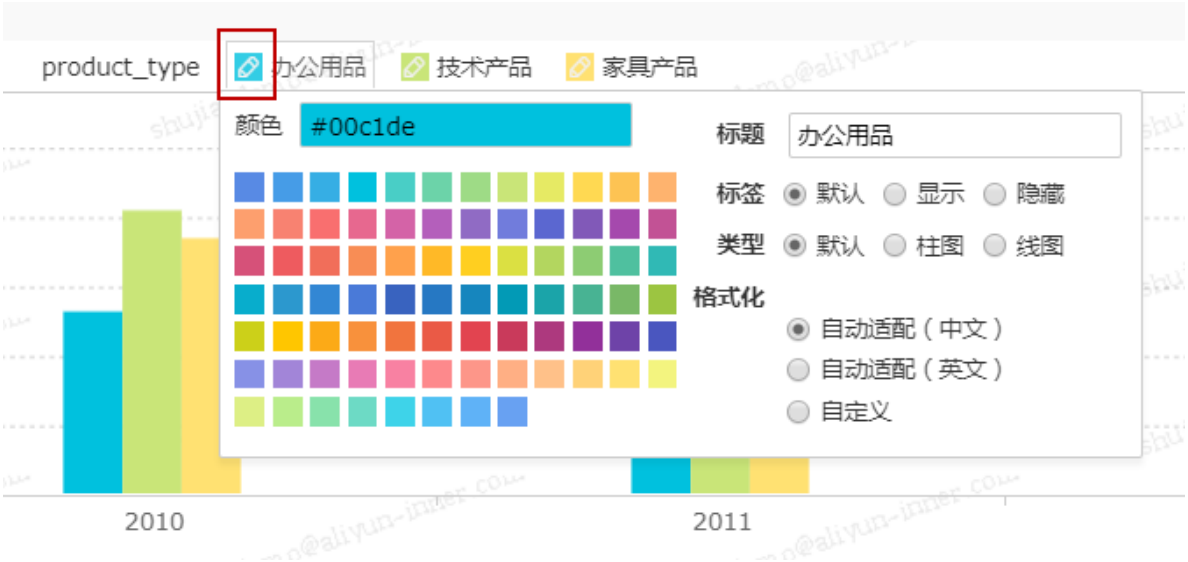
5. 字段调整完毕后，单击**更新**，更新后的图表展示如下。

图 173: 更新后的柱图



6. 单击图例前面的色块，您可以编辑图例的配色方案。

图 174: 编辑图例的配色方案



7. 在**样式**标签页中，您可以编辑图表的显示标题，布局和显示图例。

图 175: 样式标签页


数据
样式
高级
标题
柱图-company_sales_recorc
显示标题
布局
横向
堆积
百分比堆积
双Y轴
显示Y轴
显示X轴
轴标题
设计
显示图例
上
图例标题
产品类型
显示ToolTip

40 关系网络分析

40.1 产品概述

关系网络分析软件（Graph Analytics），又名Alibaba Cloud I+（以下简称I+）是基于关系网络的大数据可视化分析平台，在阿里巴巴、蚂蚁金服集团内广泛应用于反欺诈、反作弊、反洗钱等风控业务，面向公安、税务、海关、银行、保险、互联网等提供行业解决方案。

产品围绕“大数据多源融合、计算应用、可视分析、业务智能”设计实现，结合关系网络、时空数据、地理制图学建立可视化表征，揭示对象间的关联和对象时空相关的模式及规律。产品提供关联网络、时空网络、即问网络、信息立方、智能研判、协作共享、动态建模等功能，以可视化的方式有效融合机器的计算能力和人的认知能力，获得对于海量数据的洞察力，帮助您更为直观、高效地获取信息和知识。

40.2 产品定位

I+= Information to Image to Intelligence to ...

解决的问题：**如何从海量数据中发现有价值的情报。**

基于阿里云的大数据平台，处理PB级别数据。

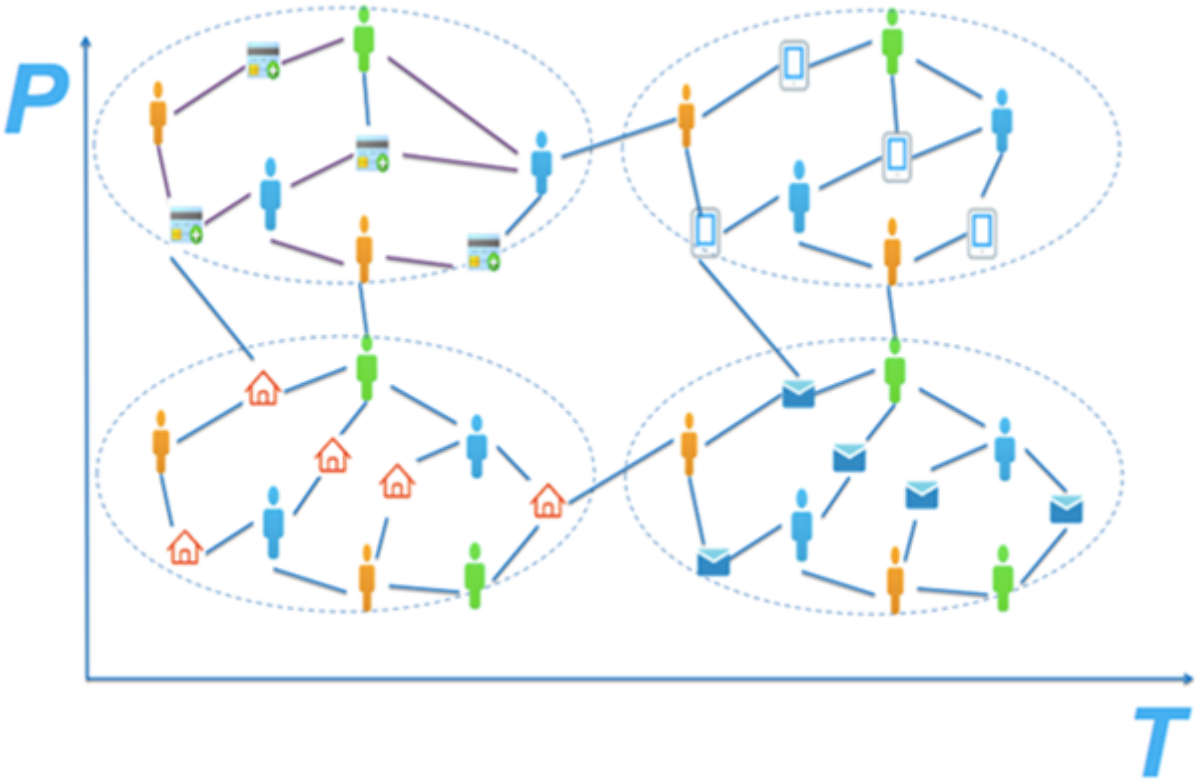
一个可视化的情报解决方案，不只是一个 BI 工具，也不只是可视化工具。

- 融合多张网，用户关系网络、资金关系网络、媒介关系网络、位置关系网络。
- 超过百亿对象，千亿关系。支付宝内部几亿用户，千亿关系对。
- 发现人和人，人和物，物和物等对象之间的关系。

40.2.1 风险防控价值

传统的风险管理中，基于单个设备、账户名单的账户黑白名单体系，存在被绕过的风险，所以需要融合风险事件中的位置、用户行为、关系等大数据元素，构建风险**可疑网络、可信网络、可视化分析体系**，并在此基础上预测潜在的风险、识别并监控团伙网络、还原风险事件现场、识别反洗钱团伙资金链路、预防营销活动反作弊。

图 176: 风险防控



40.2.2 公共安全价值

911 是发生在美国本土最为严重的**恐怖袭击**，遇难者总数高达**2996**人，对美经济损失达**2000**亿美元，相当于当年生产总值的**2%**。同时，此次事件对全球经济所造成的损害甚至达到**1万亿**美元左右。

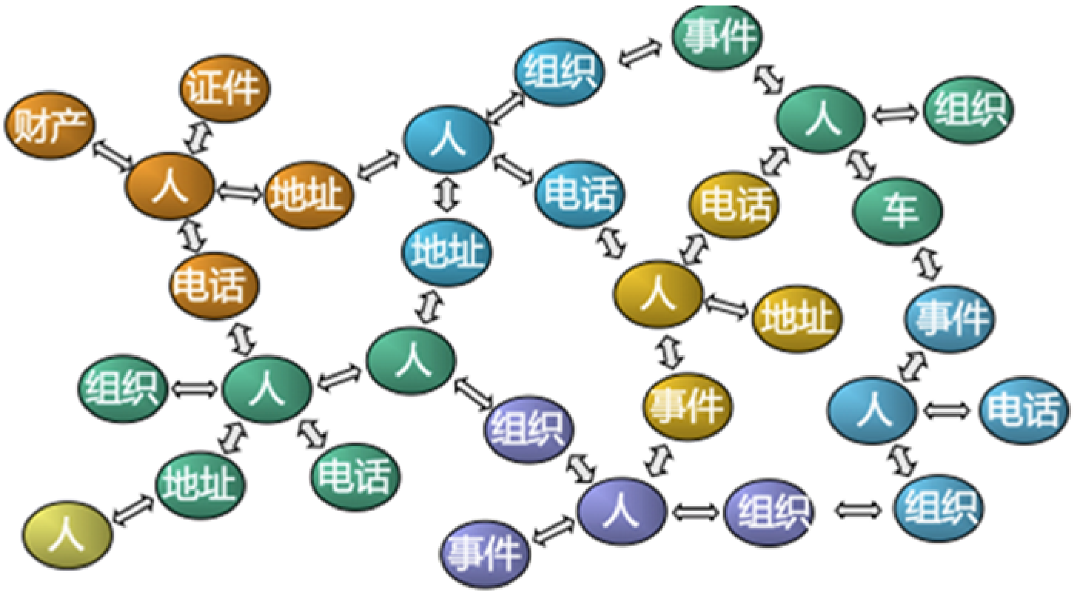
图 177: 恐怖袭击

图 178: 难民

40.2.3 世界万物相连

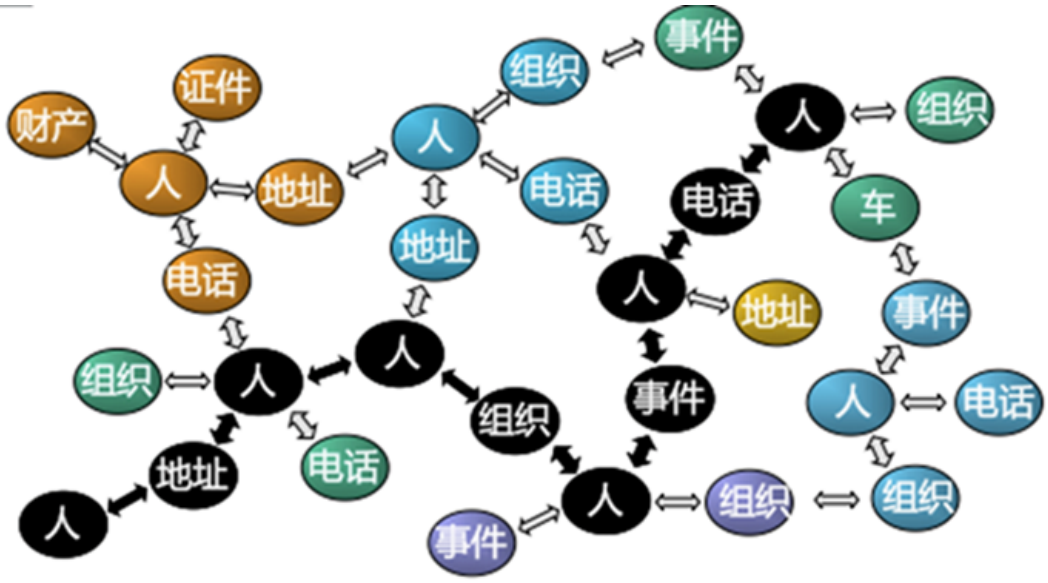
世界本就一张网，如下：

图 179: 世界本就一张网



任何两物之间都可以寻找关联，留下线索：

图 180: 两物关联

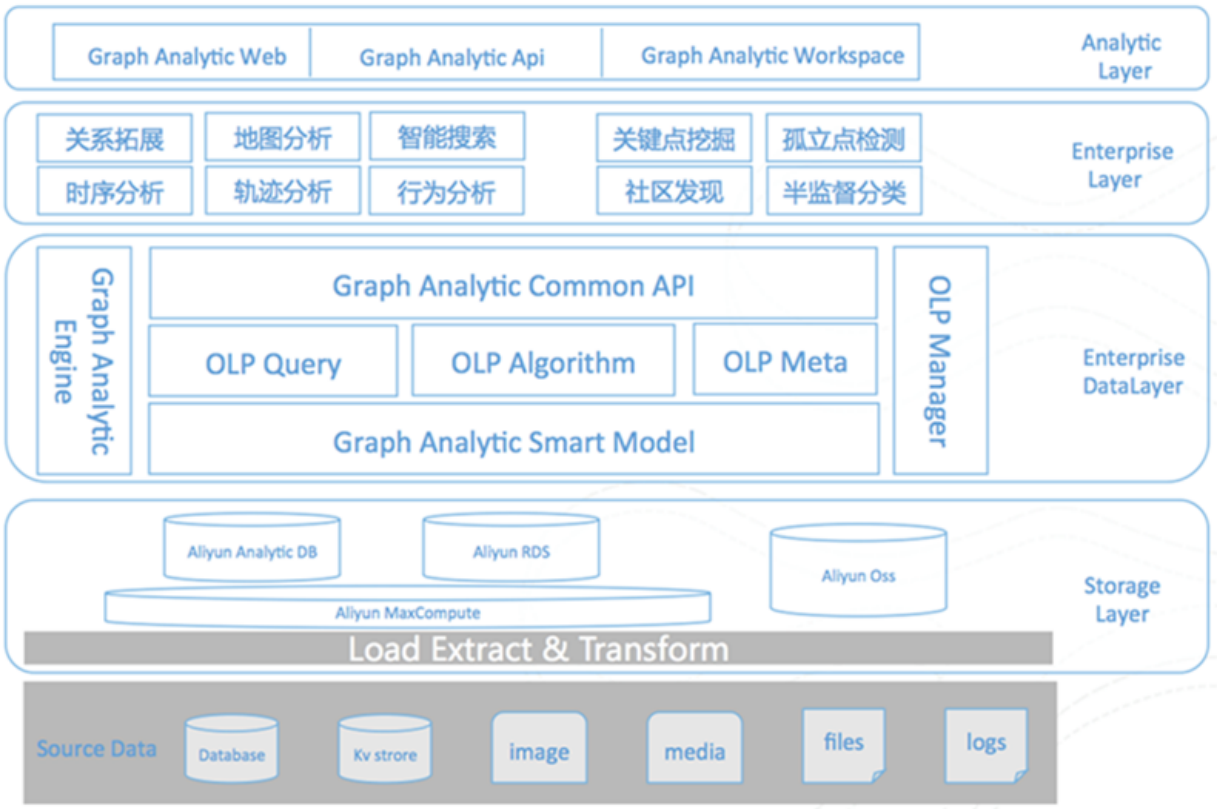


40.3 产品架构

40.3.1 系统架构

关系网络分析软件（Graph Analytics）采用组件化、服务化设计理念，多层次体系架构。数据存储计算平台建立在阿里云自主研发的大数据基础服务平台（数加平台）上，支持PB / EB级别数据的存储和计算，具有强大的数据整合、处理、分析、计算能力。

图 181: 系统架构



整个系统分为存储计算层、数据服务层、业务应用层、分析展现层。

- **存储计算层**

  基于阿里云大数据平台，支持多种开放数据源。计算平台分为离线和在线，离线计算为MaxCompute，实现数据的整合、处理，在线计算实现数据的实时计算，包括分析型数据库AnalyticDB、图数据库（BigGraph）、流计算（StreamCompute）。

- **数据服务层**

  按照关系域、关系类型、关系事项抽象出的“实体 - 属性 - 关系”模型，提取自然对象关系、社会对象关系、空间对象关系，进行业务关系逻辑建模，通过逻辑业务定义整合异域多源的数据，支持逻辑模型的灵活管理和维护。数据服务引擎为业务应用层提供统一的业务逻辑查询语言，执行各种复杂的关系网络查询、算法分析。

- **业务应用层**

  将关联网络、时空网络、搜索网络、信息立方、智能研判、协作共享、动态建模等多业务业务应用封装成API接口，提供给分析层调用。

- **应用分析层**

  提供多元智能可视化交互分析界面，支持多种终端。提供外部API接口及可视化组件服务，支持第三方系统的接入。

40.3.2 OLP模型

图 182: OLP模型

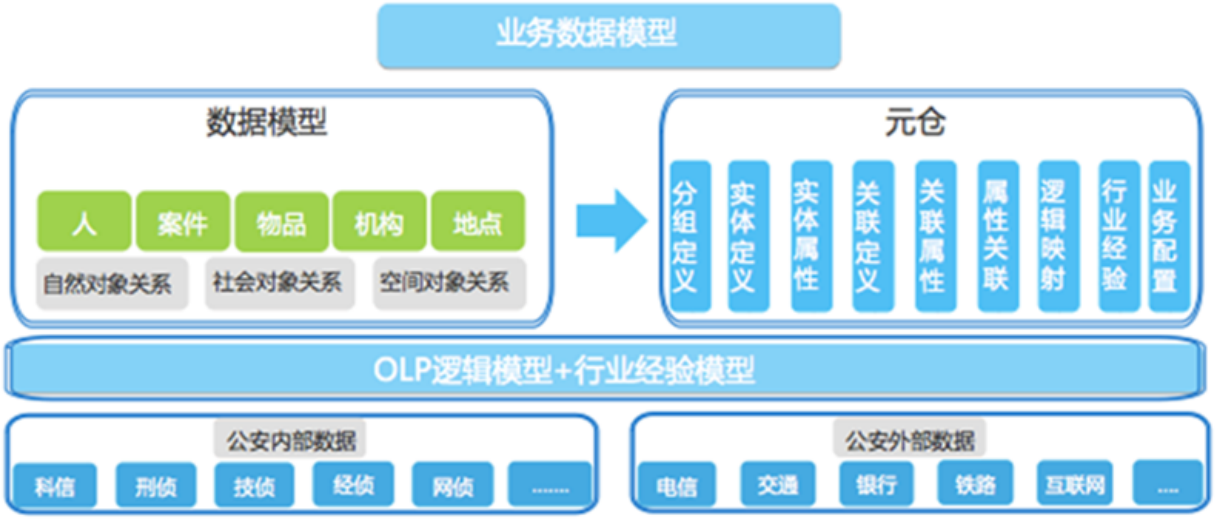


图 183: 以高铁为例

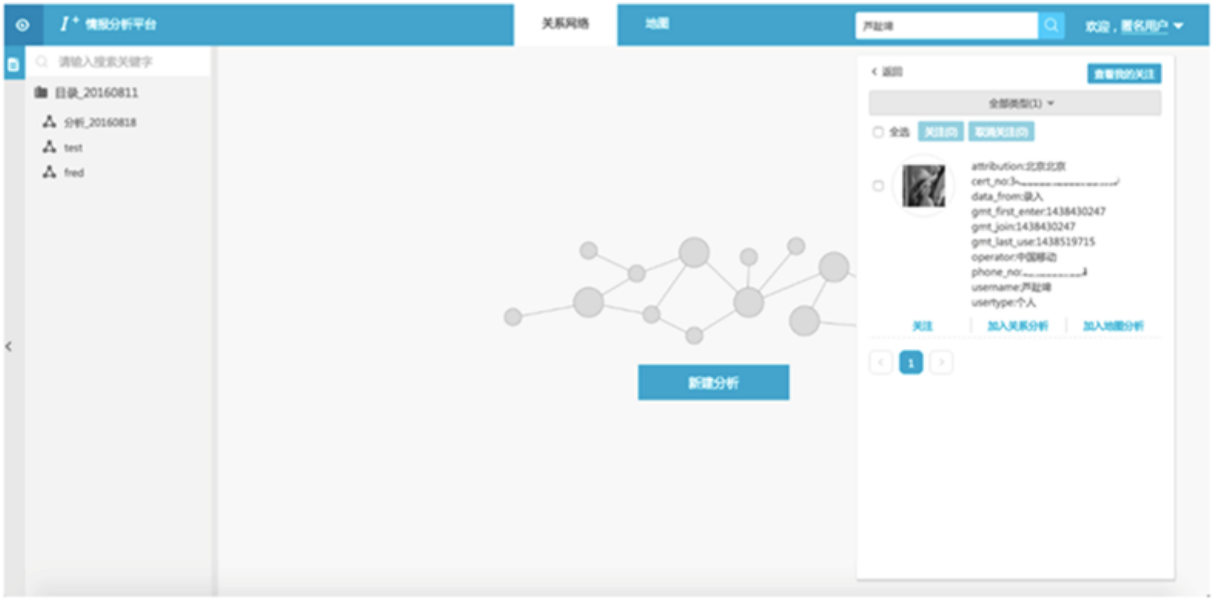
40.4 功能特性

三大功能模块

40.4.1 搜索

搜索作为I+三大独立模块之一，可以作为研判人员查找对象的工具，该对象包含手机，身份证等不同对象实体，方便研判人员快速查看对象信息。同时搜索出的对象也可以作为单独对象节点添加到关系分析和GIS分析中去，可以作为研判起点。

图 184: 搜索



40.4.2 关系网络

关联反查

以任意单个或一批对象为起点进行关系的无限拓展分析。关联反查，可帮助实现信息的无限关联。情报分析工作的核心是从大量的、没有关联的信息中发现少量的关联性线索和情报，即将信息转换为可操作情报的过程。关联反查提供两种方式：简单和高级。

群体分析

分析一批相同或不同类型的对象内部之间的关联关系，包含直接关系和间接关系。

共同邻居

分析一批相同或不同类型的对象共同联系对象。

路径分析

分析两个对象之间的关系路径。

骨干分析

针对团伙网络，通过智能业务算法，探索关系网络中核心骨干节点。

血缘分析

以家族户号为血缘脉络，展示所有人之间的血缘关联。

时序分析

在时间维度上详细展示每个事件发生细节。

信息立方

- 行为分析

  展示事件在时间维度上发生的分布频率情况。

- 行为明细

  将事件的明细信息（原数据记录按规则筛选）展示出来。

- 对象信息

  汇总关系网络中的实体，并按实体类型分类。

- 统计信息

  统计关系网络中的关系和实体，包含总体分布、对象熟悉和关系属性。

群体统计

群集统计是统计关系网络分析中的群集分布。这里面群集是指一群体对象节点，任意两两对象节点在拓扑上拥有联通路径，其中合并节点作为一个联通桥，在拓扑上认为其内部全部联通。

标签统计

标签统计是为了统计关系网络中的对象节点的标签信息。在I+系统中，标签分为两类：系统标签和用户标签。系统标签是指业务系统对某些节点定义的标签，如红黑名单等。而用户标签是指每个I+用户对某些节点通过I+添加的标签。

图区布局

关系网络分析图区支持矩阵布局、圆环布局、横直线布局、竖直线布局、力导向布局、层次布局。

右键操作

关系网络分析图区中的信息包括对象、关系内容，映射的网络结构中的图。其中点和边是核心元素，所有的分析都是基于图对其中的点和边做操作。“右键操作”是围绕着对关系网络编辑分析设计的功能点。

目录管理

目录管理可以对目录和分析管理操作。对于目录提供“新建分析”，“删除目录”，“重命名”，“新建目录”操作；对于分析提供“打开”，“删除”，“重命名”的功能操作。同时目录管理提供“搜索”对于内容检索定位和“分享”功能将个人的分析内容分享给他人延续分析。

协作共享

“协作共享”是产品提供的一种新分析模式，可以将个人的分析共享给其他人，将个人的思路和经验传递给其他分析用户，同时可以将其他分析用户的智慧和经验统一起来，形成多人协作，团体共同推进的局面，能够将团队融合为一个整体。

“协作共享”的角色分为发起者、协同者。整体流程为：发起者对个人的分析内容分享，分享过程中会设定分享范围，指定分享对象。协同者在接收到分析内容后，可以延续分析，同时将结果保存，形成新的分析版本。

同时产品支持对协同分析的管理，包括“删除”、“重命名”，“历史版本”等管理。

40.4.3 地图分析

GIS模块为I+情报分析平台三大主要模块之一。GIS模块的前端界面如下所示：

图 185: GIS模块的前端界面

GIS模块的功能如下所示。

空间关系网络

以对象的坐标位置布局关系。辅以多维统计分析和不同层次的关系探索，构建时空网络，支持自定义业务坐标，建立不同的地图模式，方便业务差异化展示。

动态轨迹

自定义不同对象实体的动态轨迹任务，支持点或线轨迹模式切换和视觉可视差异化表现。

圈地选人

支持用户自定义地理位置和空间范围，搜寻时空交错中显现的对象。

伴随分析

构建降纬hash空间，极速计算潜在伴随对象，支持业务时空自定义。

共现分析

以自定义时空为基础，海量数据中探查较高概率共同出现的对象。

40.5 产品优势

海量数据实时挖掘

I+能够在百亿节点、千亿边、万亿记录的PB量级数据中，按照用户的业务指令进行关系挖掘和时空计算，并且实时交互响应。

模型认知万物相连

I+用OLP模型认知世界万物相连，以实体（Object）和关联（Link）显像表征，以属性（Property）实现异构数据的理解和整合。

业务场景灵活赋能

I+自带以OLP为核心的中枢控制，通过业务配置和感知实现人机交互学习，赋能公安、欺诈、金融、税务等不同场景业务研判。

可视分析高效体验

I+全面分析潜在用户体验要素和业务痛点，沉淀出数据、交互、结果的分阶可视化体验和协同共享，使得有证可查，有据可说。

智能深入以人为本

I+精准智能化帮助业务员思考学习，解决常见的业务难题，目前已有涉恐指数、伴随分析、亲密度、涉毒指数等深度训练模型。

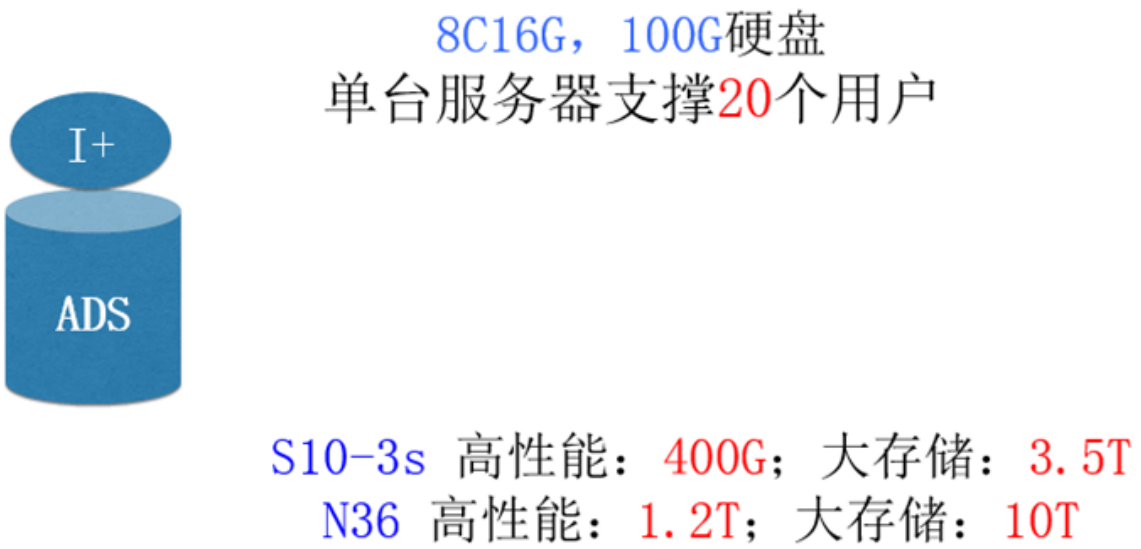
重大项目深度考验

I+已经作为亮点应用参与多个国家级重点项目，在安保、反恐、关税等领域让客户耳目一新，业务价值得到深度考验和认可。

40.6 性能指标

40.6.1 规格参考

图 186: 规格参考图



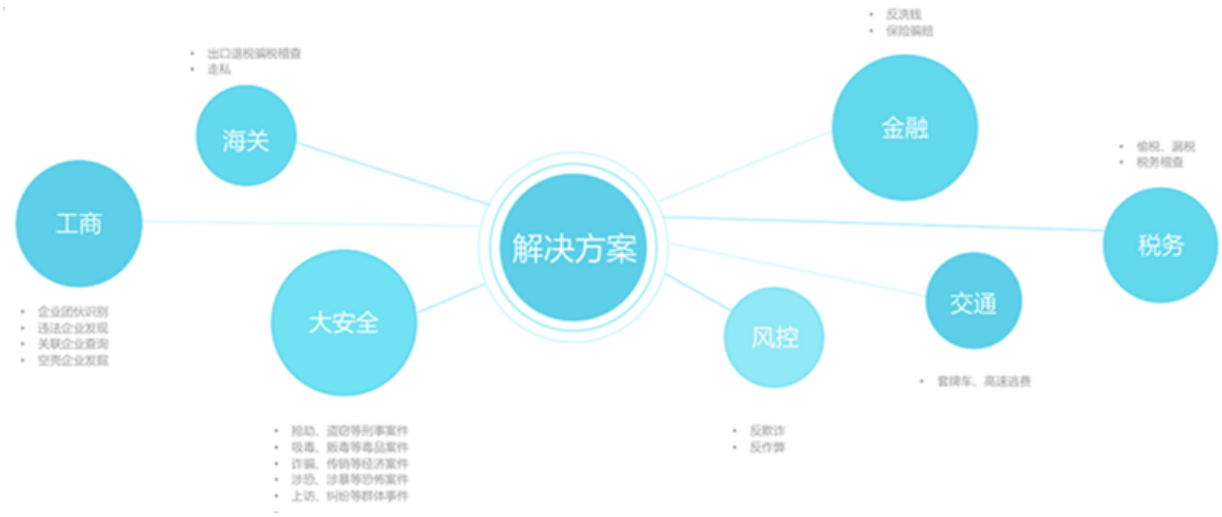
40.6.2 性能参考

压测数据#仅供参考。

图 187: 性能参考图

40.7 典型应用

图 188: 典型应用



40.7.1 典型案例

案例说明

本案例讲述如何通过I+平台侦破复杂BK案。

数据环境

本案例数据资源如下：

- 通讯工具、即时通讯通联
- 人员基本信息
- 手机轨迹、车轨迹
- 租车信息、住宿信息、火车信息
- 护照信息、出入境记录

案件背景

2015年底，某部门监测到部分小众通讯工具异常活跃，使用该工具的账号与境外某国家频繁联系，形迹可疑，按照工作要求，分析人员通过I+快速核实情况，甄别风险。

注意: 本故事不代表真实案件，仅为 I+平台功能测试杜撰！

分析过程

分析人员筛选2015年12月份活跃用户。

图 189: 筛选活跃用户

结果显示，有一批账号在该时段频繁使用小众工具。

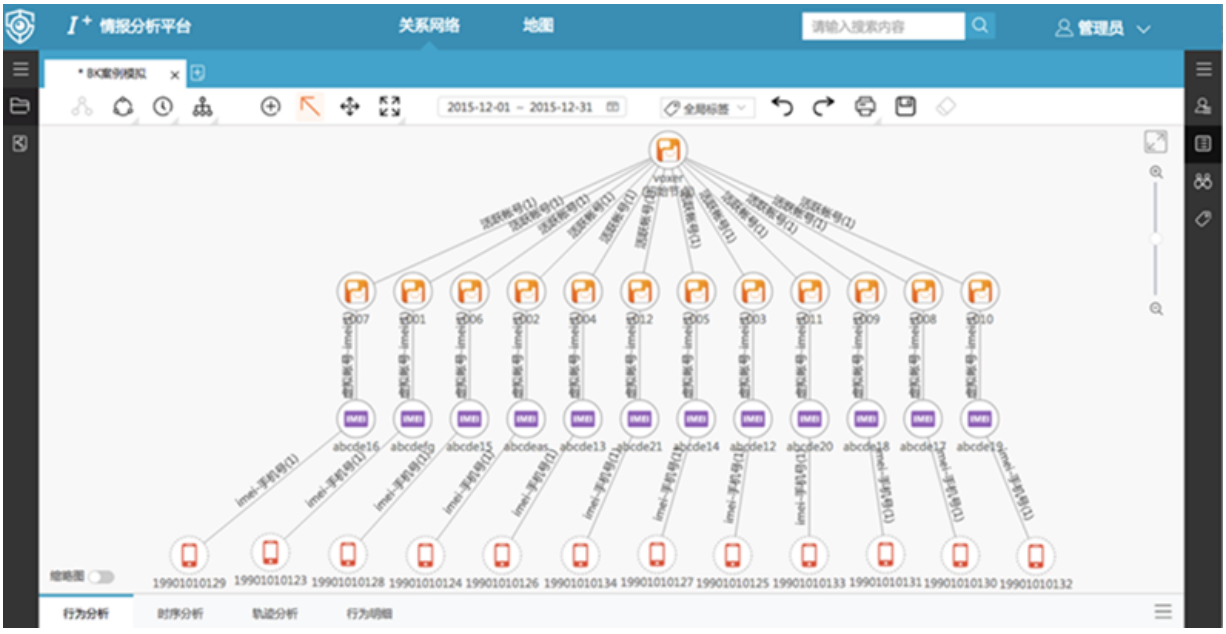
图 190: 关系网络图

进一步分析这批账号在这段时间的通联关系，结果显示，这批账号与境外某国的另外一批账号存在单线联系，分析人员尝试获取境外账号的情况，没有查到相关信息。

图 191: 单线联系

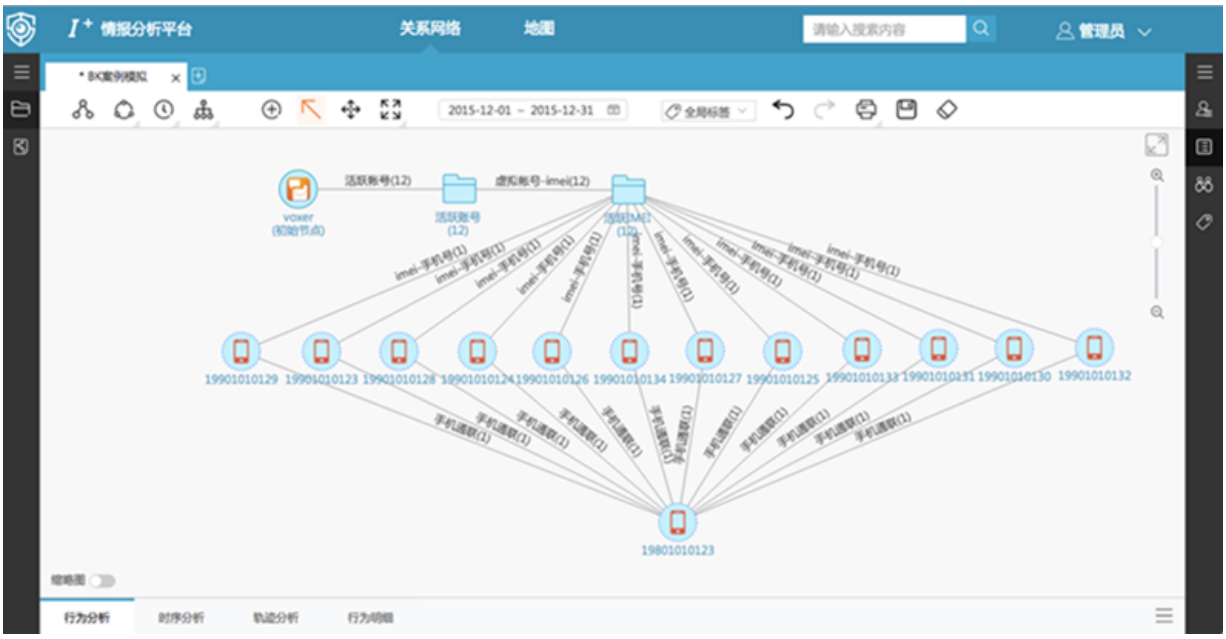
分析人员尝试获取到这批账号的实际使用人，进一步拓展分析发现，这批账号分别在不同的手机上使用，关联反查得到对应的手机号，结果显示，这批手机号只在单独的手机上使用过。

图 192: 单独手机使用



对这批手机进行分析，未获取到其开户信息，但发现与同一个手机号码存在联系。

图 193: 同一手机联系



进一步分析发现，此号码在多个地点出现。

图 194: 多地点出现

接下来，分析人员对该号码展开分析，尝试获取与其同行的其他号码。

图 195: 其他号码

结果显示，19701010123与该号码存在6次伴行关系。

图 196: 手机通联一

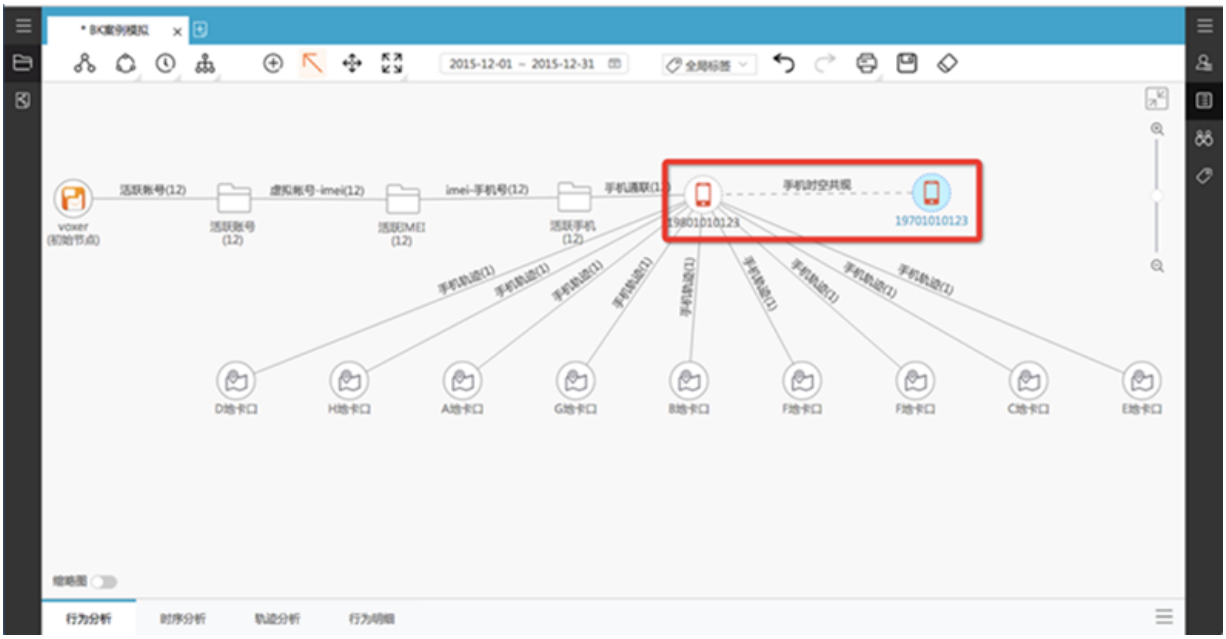
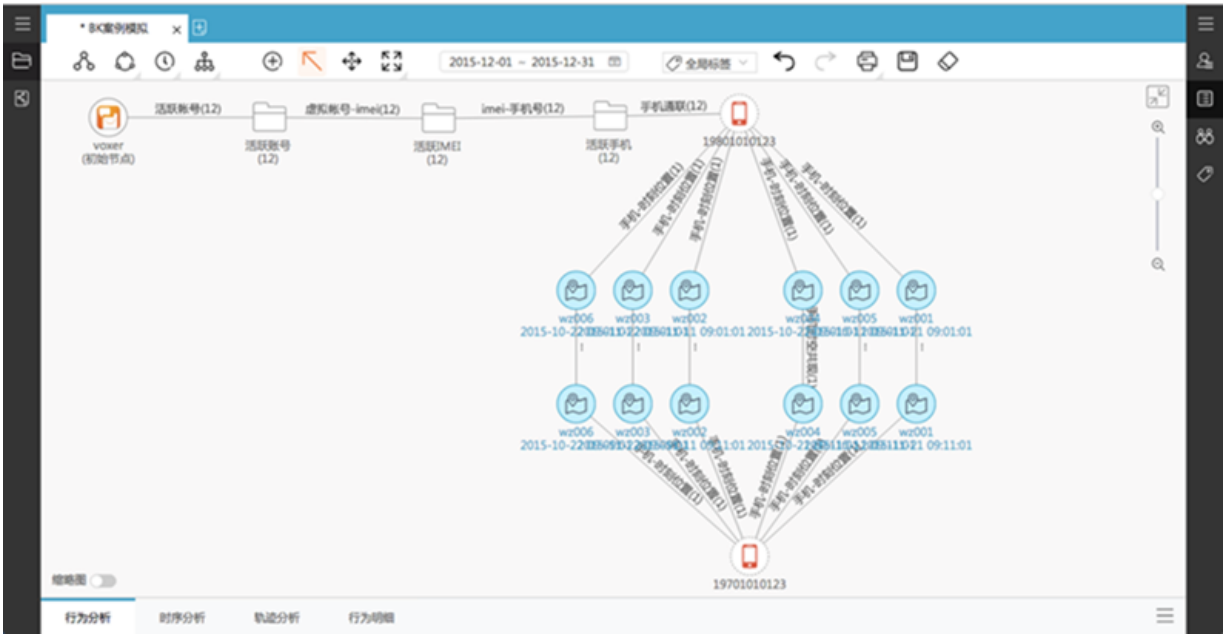
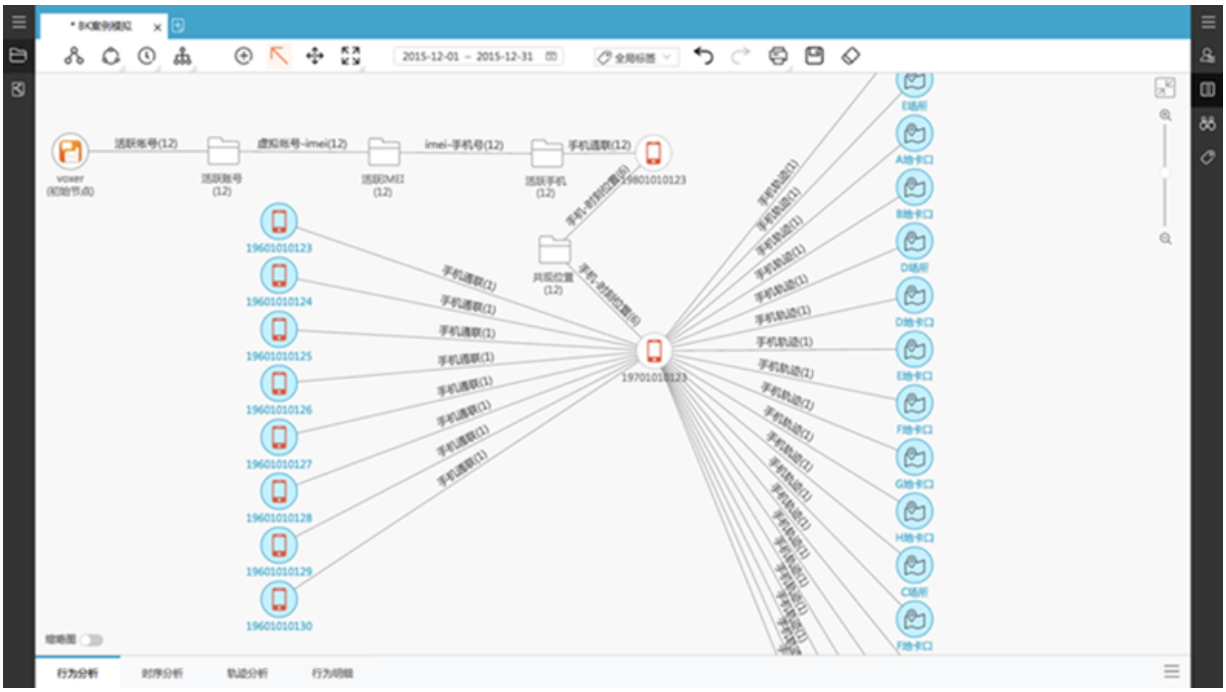


图 197: 手机通联二



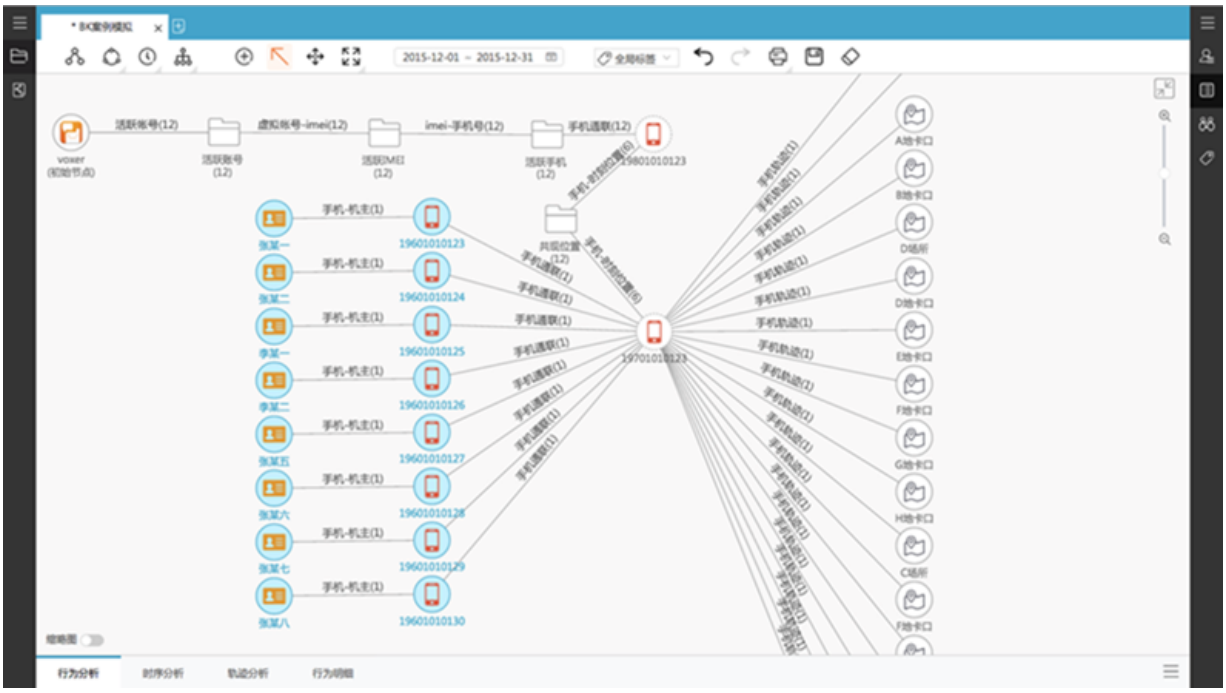
根据经验判断，19701010123很可能为19801010123持有的另一部手机。分析人员对19701010123开展分析，发现该部手机存在大量的通联信息、位置信息。

图 198: 手机信息



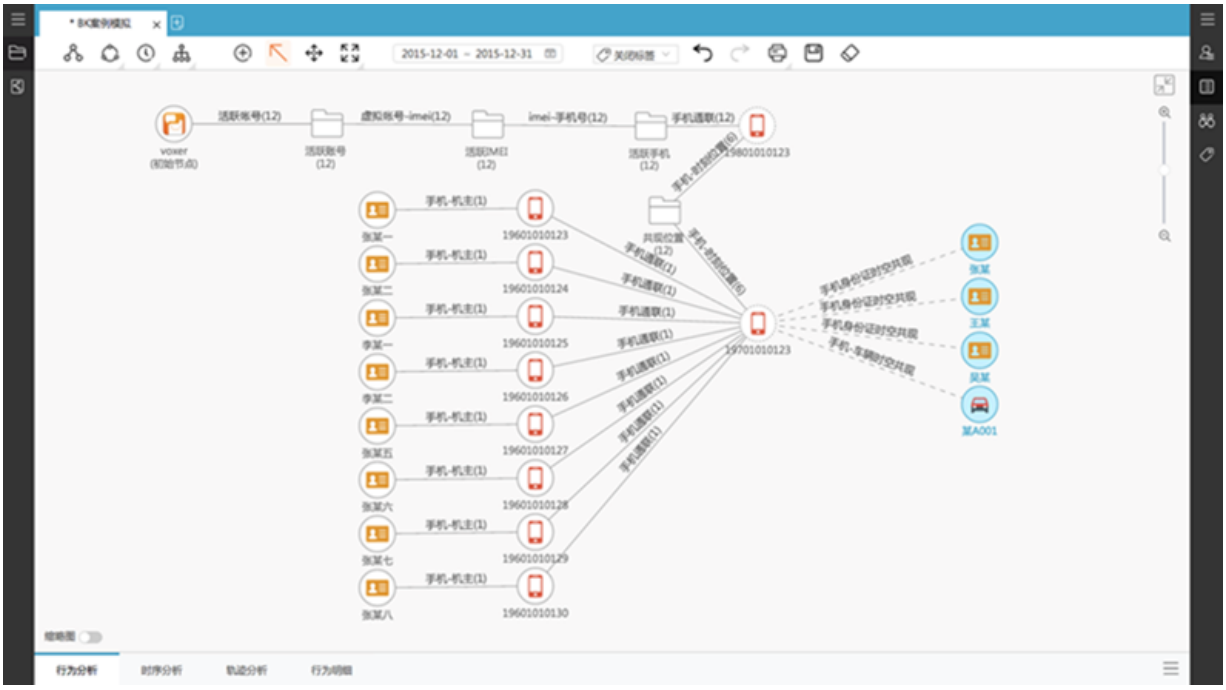
对该手机的通联对象展开分析，发现这批通联号码存在机主登记信息。

图 199: 机主登记信息



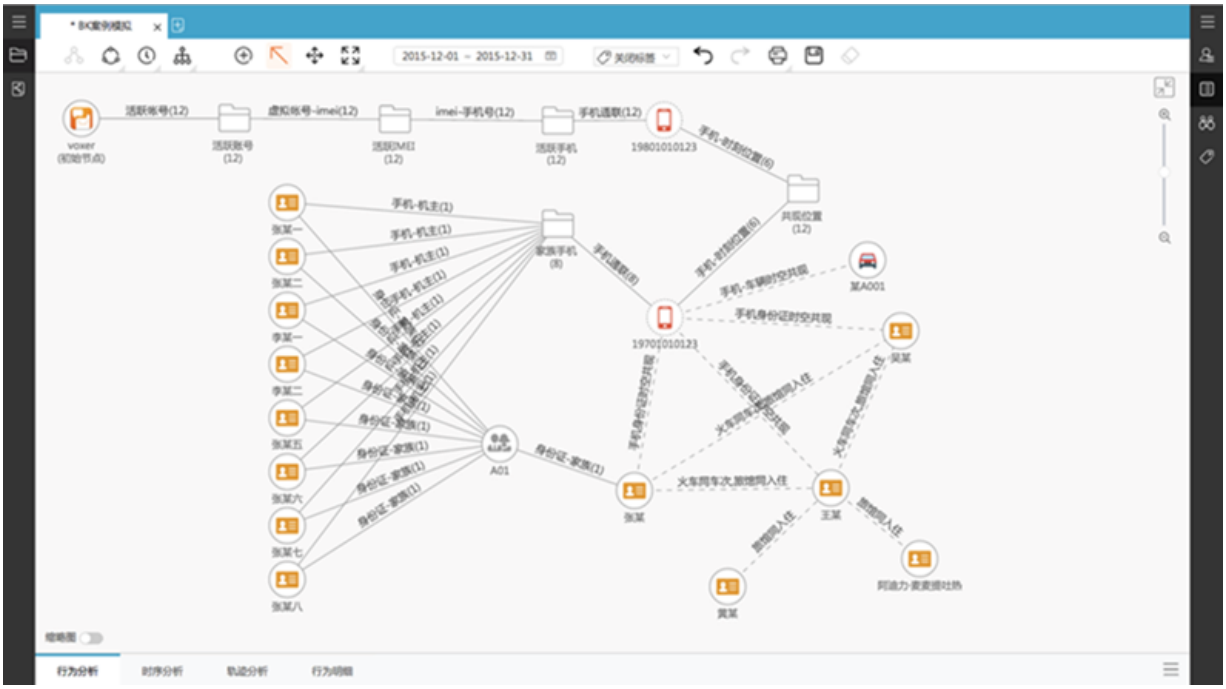
通过对19701010123的轨迹进行分析，发现该号码与张某等三人及某A001车存在伴行关系，此手机号很有可能为此三人所有。

图 200: 手机号持有人



进一步对上述发现的身份证进行分析，发现其中9人为同一家族成员，且张某与另外两人存在多次同行同住关系，同时，还挖掘到黄某、尚某与b12345123存在密切往来，三人曾经多次入住相同酒店。

图 201: 密切往来人员



对黄某、尚某等四人进行重点分析，发现此四人为SK前科人员，2011年由于涉嫌某BK案件被逮捕，于2015年释放出狱，结合上述异常行为，分析人员判断此四人很可能预谋作案。

图 202: 嫌疑人一

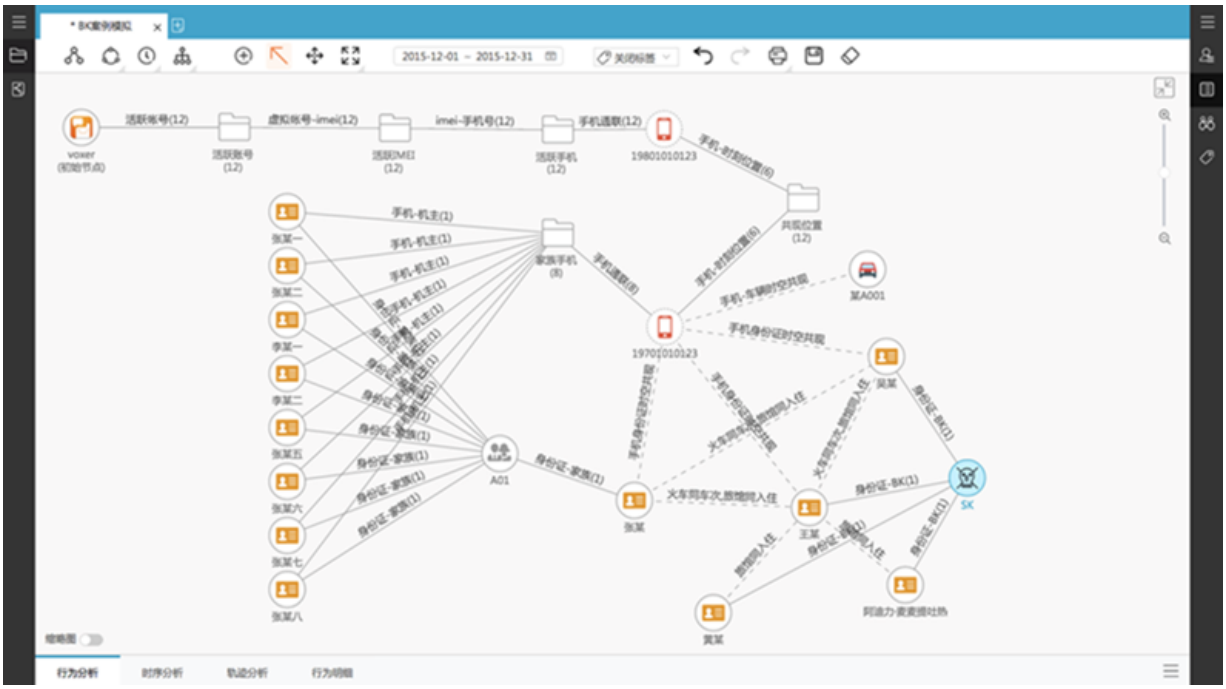
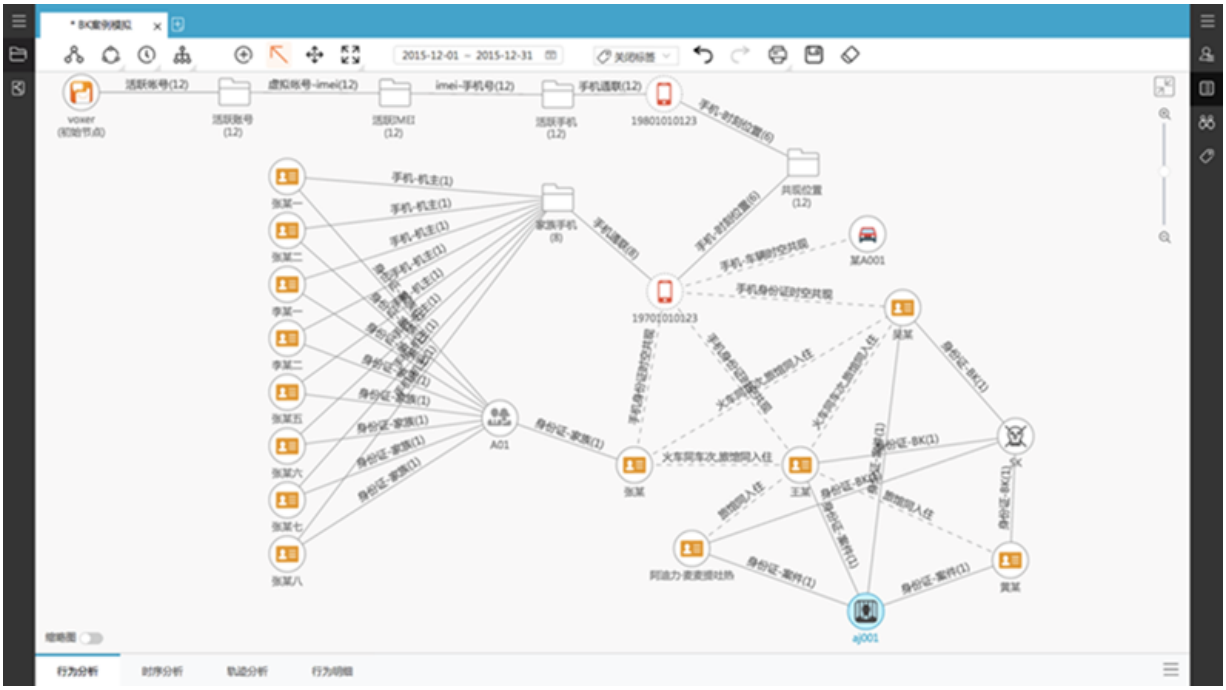


图 203: 嫌疑人二



再次拓展分析发现，黄某、尚某以及张某家族共有9人持有护照，并于2009年、2014年、2015年相继往返B国。

图 204: 拓展分析

B国为境外SK敏感国家，上述异常监测中所发现的小众通讯工具通联对象所在国家恰好也在B国，分析人员初步认为，张某、黄某、尚某等人很有可能预谋作案。

以上情报立即上报相关部分，经过民警实地排查，在某出租屋内发现张某、黄某、尚某等人正预谋在2016年元旦期间实施BK袭击，相关部门当即调派精干警力，摧毁了该团伙，避免了恶性案件的发生。

案情回顾如下：

图 205: 案情回顾

40.7.2 智能关系网络

图 206: 智能关系网络一

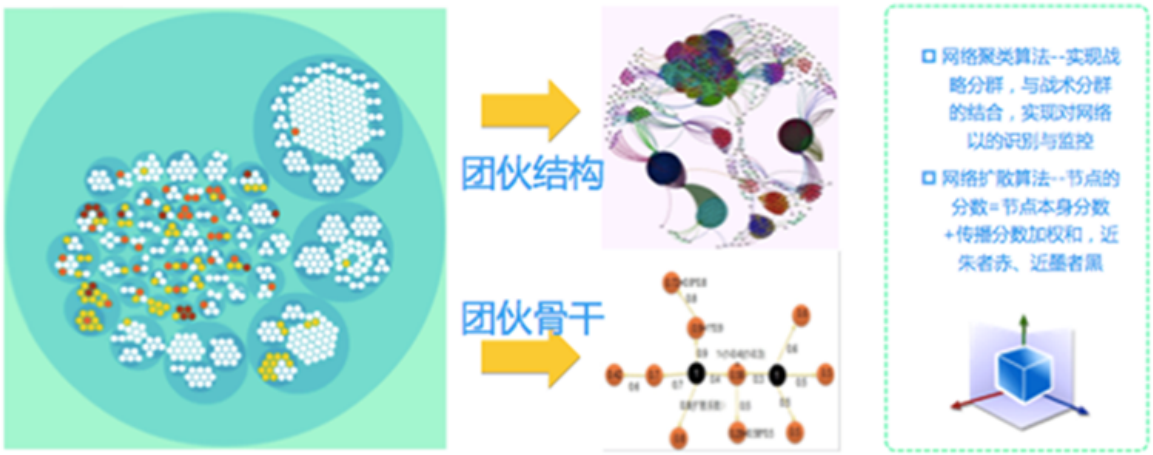


图 207: 智能关系网络二

40.7.3 行业风控

图 208: 行业风控图

40.7.4 公安安防

图 209: 公安安防图


信息系统建设
信息检索、查询
可视分析

41 采云间（DPC）

41.1 数据集成平台

41.1.1 产品概述

阿里云数据集成平台（Data Integration Platform，简称DIP）是蚂蚁大数据平台解决方案的一个分支产品，DIP基于MaxCompute，提供强大的产品功能，支持复杂的企业级数据集成，涵盖数据的集成、ETL开发、任务调度、发布部署和运维等功能，完成数据的价值重构和可信赖交付，帮助企业从大数据中获得更多价值。

数据集成平台主要包括两个产品模块：ETL开发框架和任务管理中心。

41.1.2 产品架构

大数据集成平台在整个蚂蚁大数据解决方案中的位置如图 *210:* 大数据解决方案所示。

图 210: 大数据解决方案

大数据集成平台DIP主要包括两个功能模块： ETL开发框架和任务管理中心，下面将对这两个功能模块进行分别的介绍。

41.1.3 功能特性

41.1.3.1 ETL开发框架

ETL开发服务提供基于大数据平台的ETL开发的编程环境。

ETL设计与开发是大数据的DW和BI建设的关键环节之一，负责系统数据的生成，将数据在EDW数据架构的层次之间进行加工传输。

ETL开发主要包含的功能模块如下：

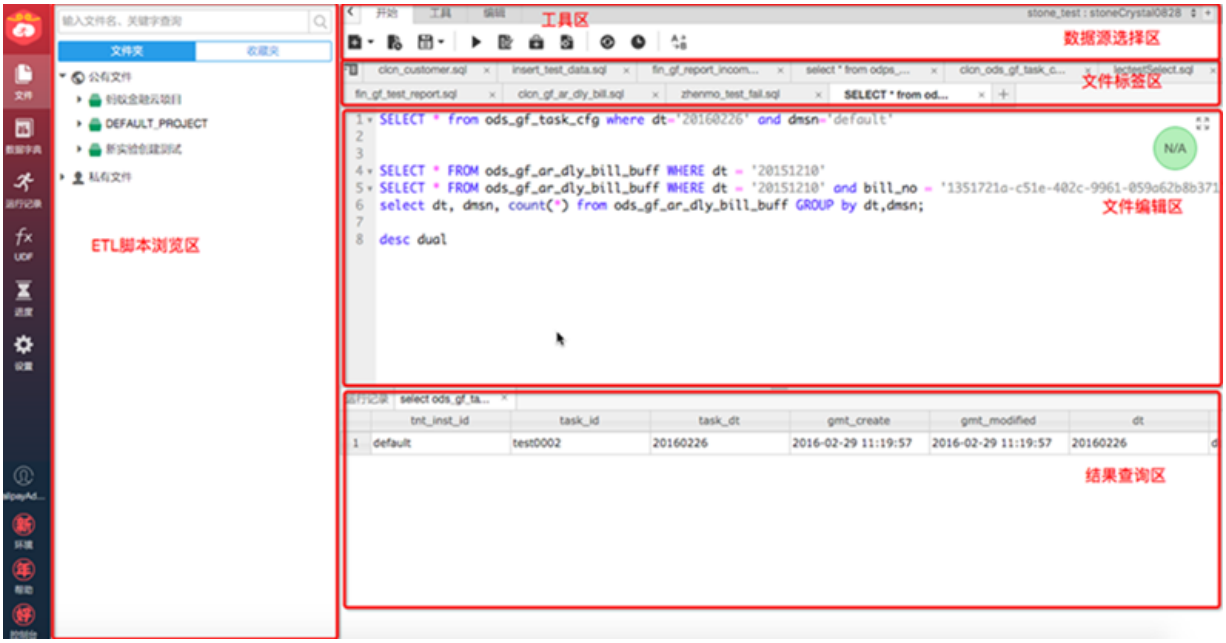
- 脚本开发
- 数据字典
- 数据管道

41.1.3.1.1 脚本开发

脚本开发模块提供脚本编辑和调试的开发环境，可以进行数据清洗和加工的脚本开发，也可以作为灵活查询的工具，执行查询命令。

如图 *211:* 脚本开发界面所示为脚本开发界面，脚本开发具备如下功能特性：

图 211: 脚本开发界面



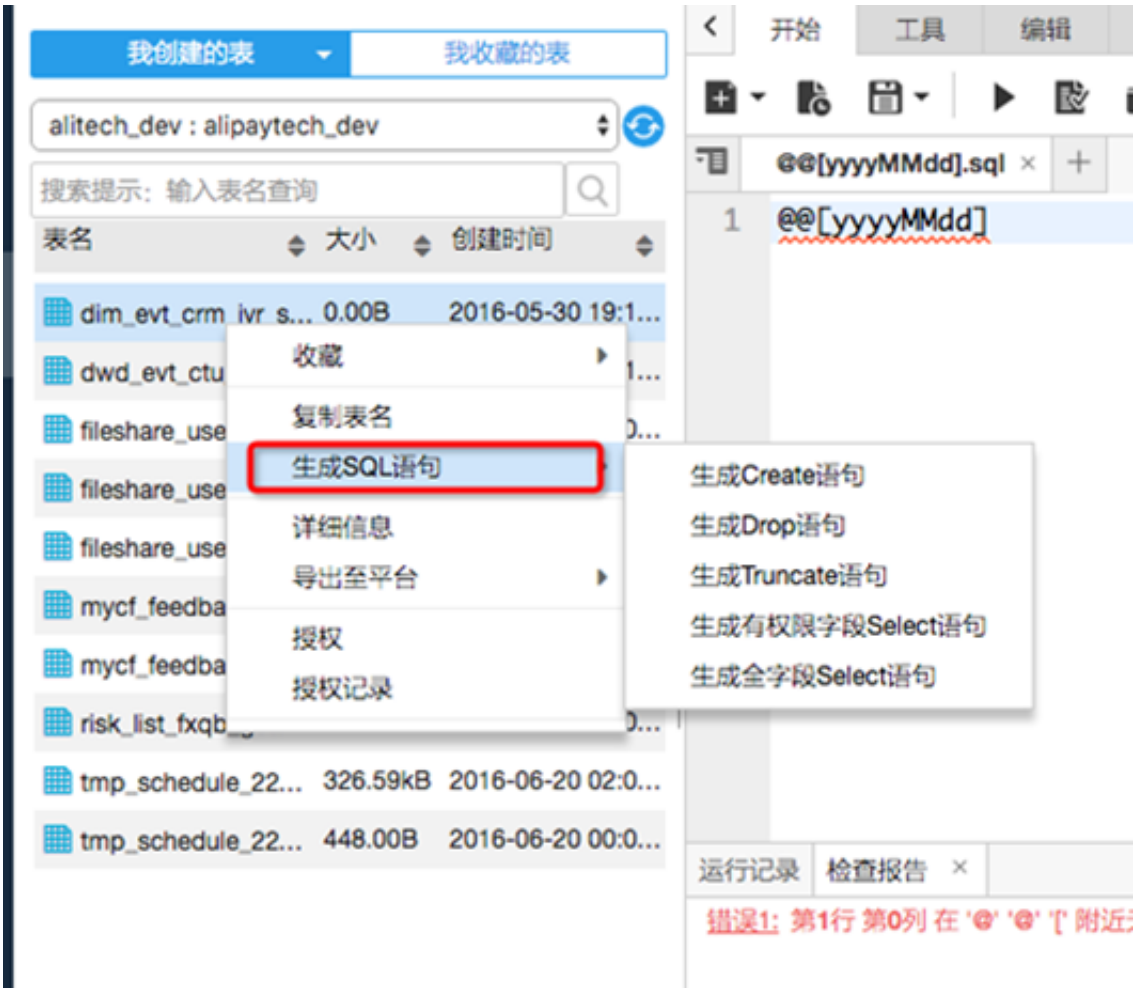
- 提供个人和团队进行ETL脚本协同开发。
- 方便的SQL执行和调试功能，可以对执行结果集进行各类操作，满足灵活查询的需要。
- 以文件树的方式记录、编辑和管理代码脚本，支持文件加锁和版本管理。
- 支持UDF函数上传，支持数据分析师上传自己的数据进行关联查询。

41.1.3.1.2 数据字典

数据字典功能可以查看表清单和字段信息，也可以收藏个人关心的表。

如图 *212:* 数据字典所示，数据字典具备如下功能特性：

图 212: 数据字典



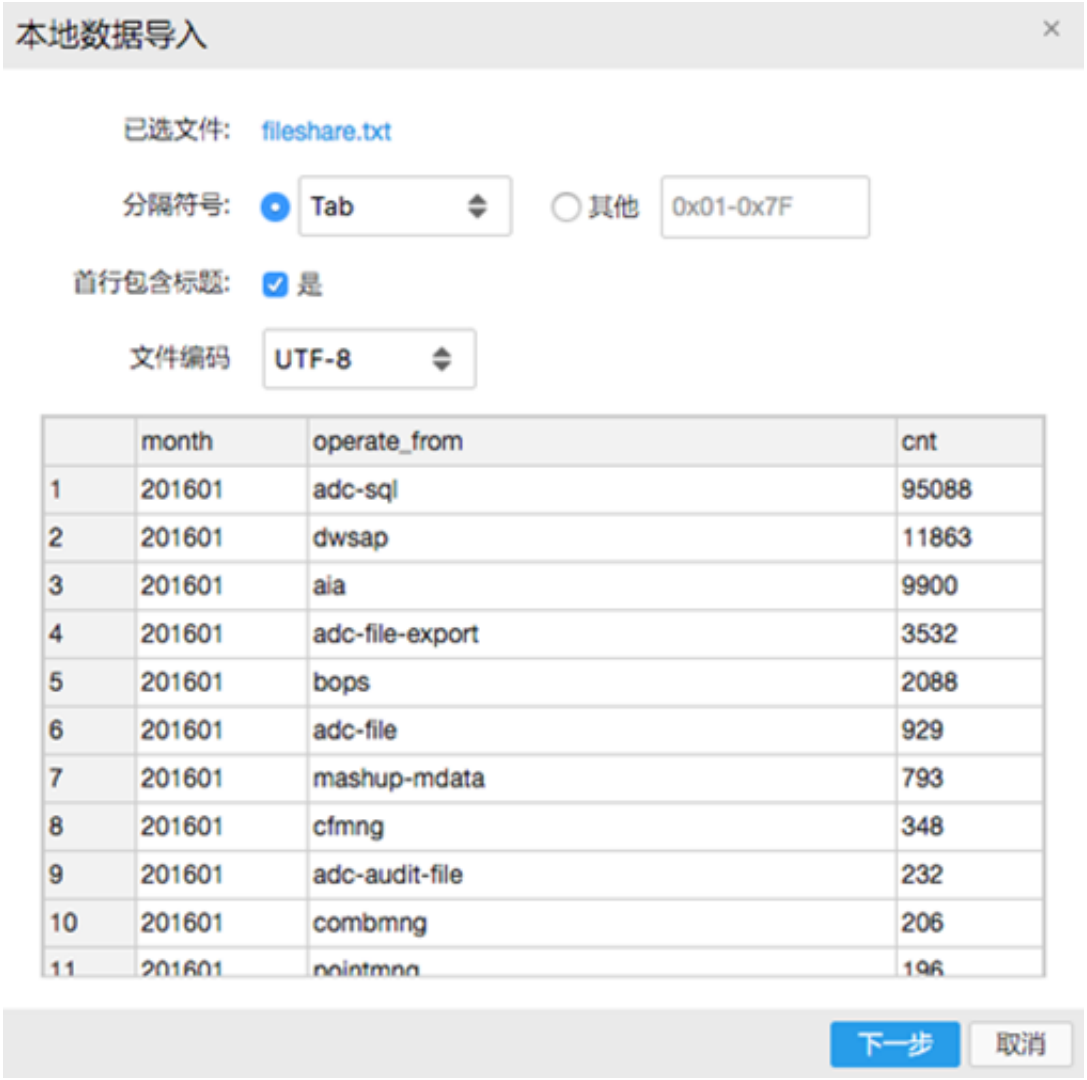
- 可以查询指定数据源的所有的表和我创建的表。
- 显示查看表的详情，包括创建人、存储和字段清单。
- 分组收藏关心的表，方便使用。
- 自动生成常见SQL。
- 可以把表授权给他人，并查看授权记录。

41.1.3.1.3 数据管道

数据管道支持将数据导入到MaxCompute，也支持将数据从MaxCompute中导出。

如图 *213:* 数据管道所示，数据管道具备如下功能特性：

图 213: 数据管道



- 支持最大500M文件的上传，支持csv和txt格式。
- 支持查询结果导出（默认不开通，需要设置）。

41.1.3.2 任务管理中心

任务管理服务提供数据任务的全生命周期管理功能，可以周期性的运行数据采集、加工等任务，保证数据的日常开发。

任务管理主要包含的功能模块如下：

- 任务配置
- 任务监控
- 报警配置

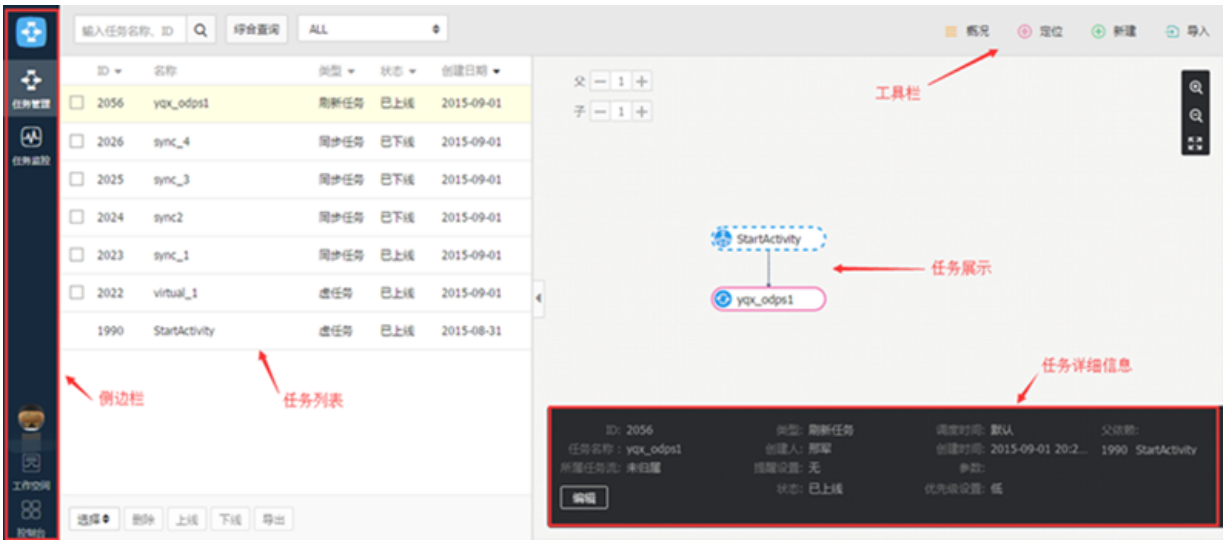
41.1.3.2.1 任务配置

任务配置支持任务的创建、编辑、上线和删除等操作，可以灵活设置任务依赖。

如图 *214: 任务配置*所示，任务配置具备如下功能特性：

- 云端数据同步和加载，通过简单易用的配置，支持RDS、Table Store、MaxCompute等异构数据库之间的双向同步。
- 图形化任务调度的设置，支持复杂的任务调度规则，包括任务定义、任务上下线、任务挂起、补数据、任务监控和异常报警等。
- 支持自动解析依赖，省却手动配置工作。

图 214: 任务配置



41.1.3.2.2 任务监控

任务监控提供任务运行的实时监控大图，方便查看任务执行进度，也可快速发现和解决出错任务。

如图 *215: 任务监控*所示，任务监控具备如下功能特性：

- 实时数据刷新，第一时间发现问题。
- 可视化的任务流显示，方便查找异常任务。
- 支持任务的挂起、跳过和重试等功能。

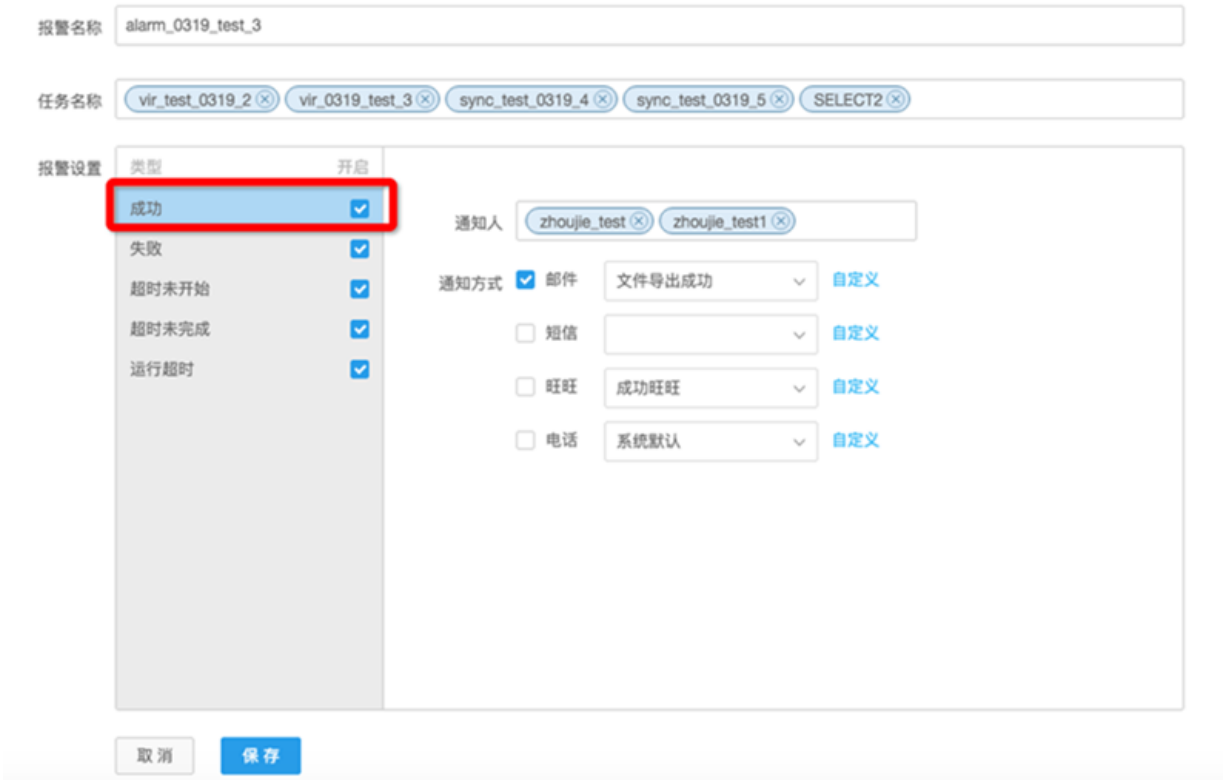
图 215: 任务监控

41.1.3.2.3 报警配置

报警配置功能支持设置报警规则，方便在第一时间在任务执行出现异常情况时，发出报警。

如图 *216:* 报警配置所示，报警配置具备如下功能特性：

图 216: 报警配置



- 支持多种渠道进行提醒。
- 支持多种预警规则类型：成功、失败、超时未开始、超时未结束和运行时长超时。

- 支持任务的挂起、跳过和重试等功能。

41.1.3.2.4 发布部署

发布部署支持把已经开发好的任务和脚本，发布到其他环境。

发布部署支持同集群发布和跨集群发布。

在发布部署模块，您可以：

- 创建发布包

  选择待发布的任务，设置发布目标环境，如图 *217:* 创建发布包所示。

  图 217: 创建发布包

- 发布监测

系统会自动检测发布包中的任务，在目标环境的依赖关系是否完整，数据源是否存在，如图 *218: 发布监测*所示。

图 218: 发布监测

- 发布审核

  发布管理员可以审阅待发布的发布包，以决定是否需要发布通过，如图 *219: 发布审核*所示。

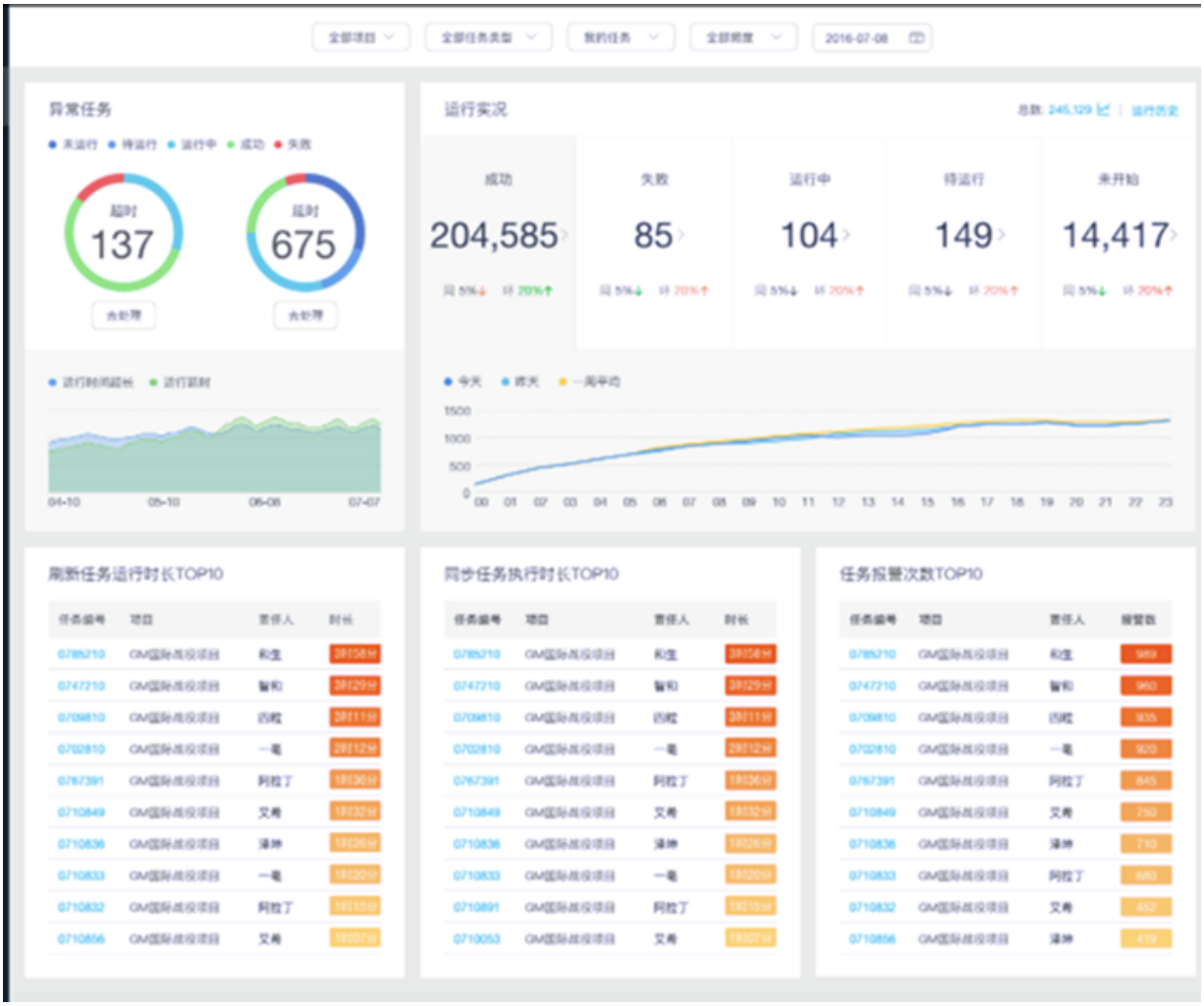
图 219: 发布审核

41.1.3.2.5 智能运维

如图 *220: 智能运维*所示，智能运维提供运维监控大屏，实时显示当前任务执行的情况：

- 异常任务：包括任务运行超时和启动延迟。
- 任务实况：显示任务的实时运行情况。
- 刷新任务运行时长Top10：显示运行时长Top10的刷新任务。
- 同步任务运行时长Top10：显示运行时长Top10的同步任务。
- 任务报警Top0：显示报警次数Top10的任务。

图 220: 智能运维



41.2 数据分析平台

41.2.1 产品概述

数据分析平台的主要作用是帮助用户在海量数据下，进行多维分析，例如，在交易网站中，一天有数百万比的记录，可以通过数据分析平台的托拉拽等功能，快速的筛选出产品或者运营的数据，以帮助产品或者运营进行日常的决策。

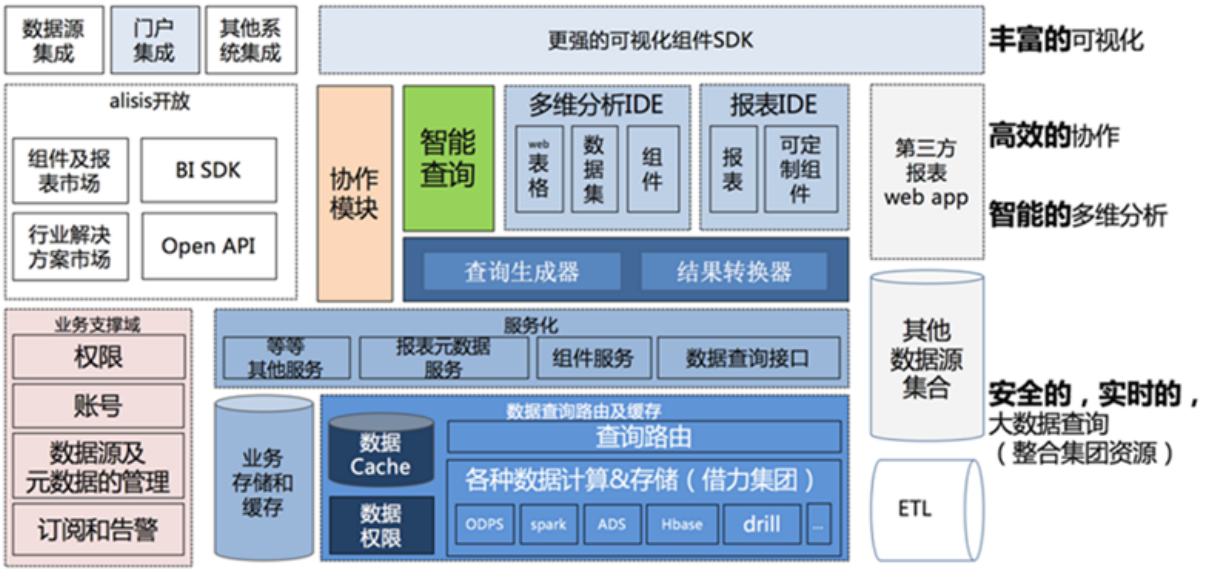
假设运营想要各省、各城市、各区域，男女的平均交易金额。如果您懂SQL，您可以写 select avg(money) from xxx group by province, city, area, gender这样的语句从而获得对应的表格式结果；也可以在alisis中通过鼠标的点击来完成类似的功能，而且还可以将结果以柱状图的形式表达出来，使数据更加容易被理解。

上述的场景是一个独立的场景，如果将数据分析和整个工作体系结合起来，比如说和growth hacking结合起来，那么在growth hacking中的每一个步骤都可以通过数据分析平台来分析对应的数据。从而增强growth hacking的效率，帮助企业获取竞争优势。

41.2.2 产品架构

数据分析平台整体技术架构如图 *221: 数据平台整体技术架构*所示。

图 221: 数据平台整体技术架构



其中主要模块为：多维分析 IDE、报表制作 IDE、数据查询、数据加速及开放集成。

41.2.3 功能特性

41.2.3.1 数据集

数据集是数据的集合，它是一个逻辑概念，它背后的实现可能来自多个物理表，最终汇总成一个数据的集合。针对这个数据集合，我们可以做一系列的操作：

- 数据集和一张物理表或者多张物理表进行映射。
- 预览数据集，查看数据集中的内容。
- 对数据集进行行级权限的设置。

- 对数据集进行加速，以提升数据集访问速度。

数据集的基本结构如图 222: 数据集的逻辑结构所示。

图 222: 数据集的逻辑结构

其中，维度一般是字符类型；度量一般是计算类型。

在前文举的例子中：各省、各城市、各区域、男女的平均交易金额，**省**、**市**、**区域**及**性别**是维度；**交易金额**是度量。

41.2.3.2 工作表

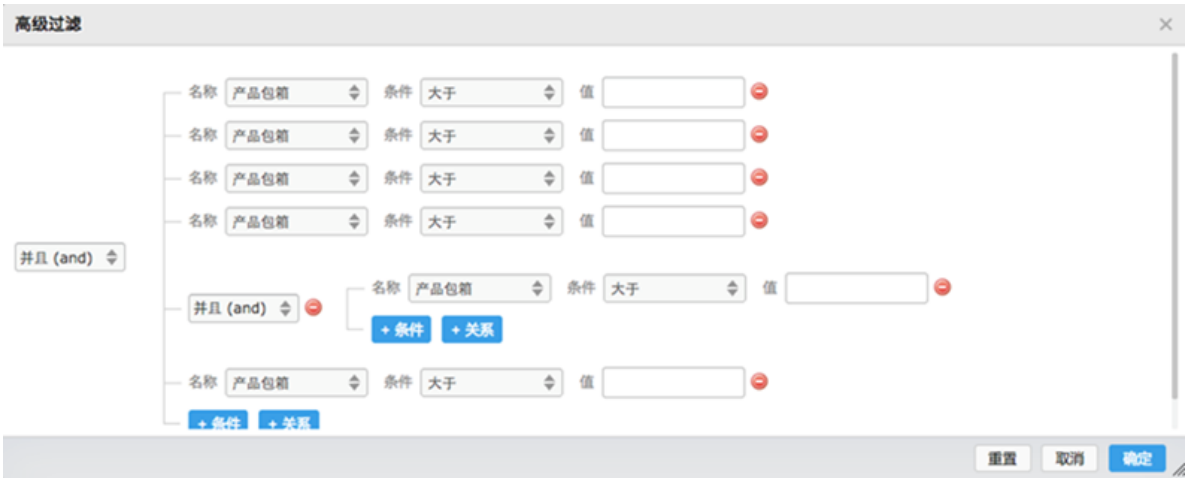
您需要对数据集进行各种操作，从而产出工作表。

操作方法有很多种，举例如下：

- 合并单元格
- 分类汇总和自动求和
- 排序
- 交换x、y轴
- 导出：将数据导出到本地的excel文件
- 图表：直方图、折线图和漏斗图等

- 函数：avg、sum count、distinct 等
- 条件格式
- 行级权限
- 高级过滤，如图 *223:* 高级过滤所示。

图 223: 高级过滤



工作表是大数据分析平台的核心功能之一。在工作表中，我们可以对数据进行各种计算，常见的计算如平均值、总数和加总等。

41.2.3.3 SQL 查询生成

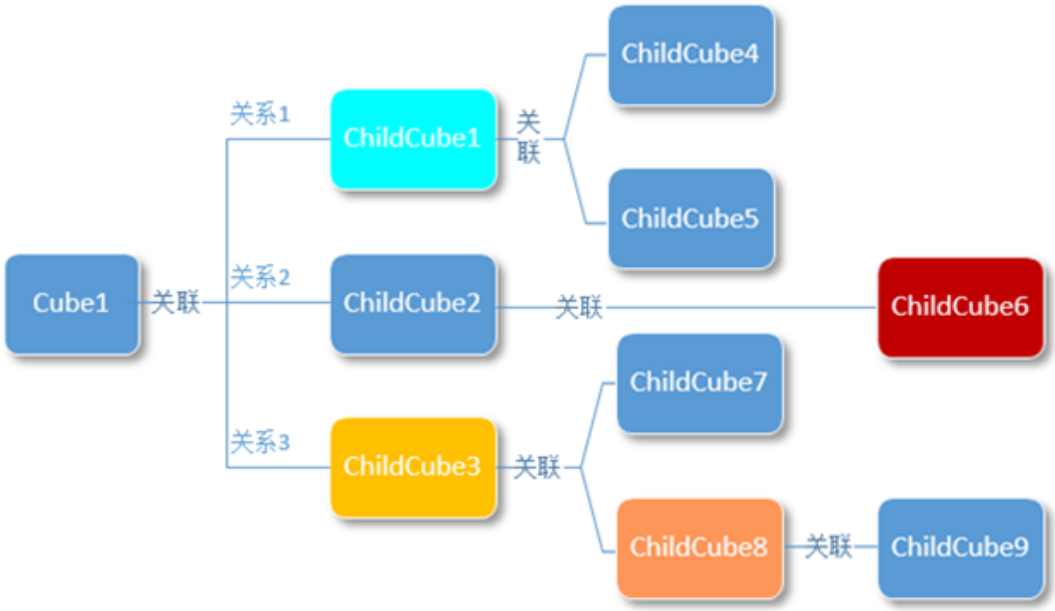
当您编辑表格之后，系统会根据页面上您的设置生成一个数据集的抽象结构，然后再根据这个抽象的结构来生成SQL语句，以便到大数据分析引擎中执行。

41.2.3.3.1 数据集抽象

维度/度量的血缘分析，就是通过逻辑模型中的维度/度量去找到物理模型中对应的表和字段以及查询方式，是生成查询结构的第一步。无论数据集内部的结构如何复杂，对外展现的是其逻辑结构。

一个数据集包含了多个维度或度量，维度和度量之间可以进行组合分析查询。如果一个数据集没有子数据集，那么我们可以认为维度和度量都会映射到事实表中列，这种情况比较简单，从主数据集就能找到所有信息，查询结构也简单。如果主数据集关联了多个子数据集，而子数据集又是多个数据集关联而成，情况就比较复杂了。图 *224:* 数据集抽象表现了这种关系。

图 224: 数据集抽象



41.2.3.3.2 查询路径分析

元数据分析结束后，下一步是分析查询路径。

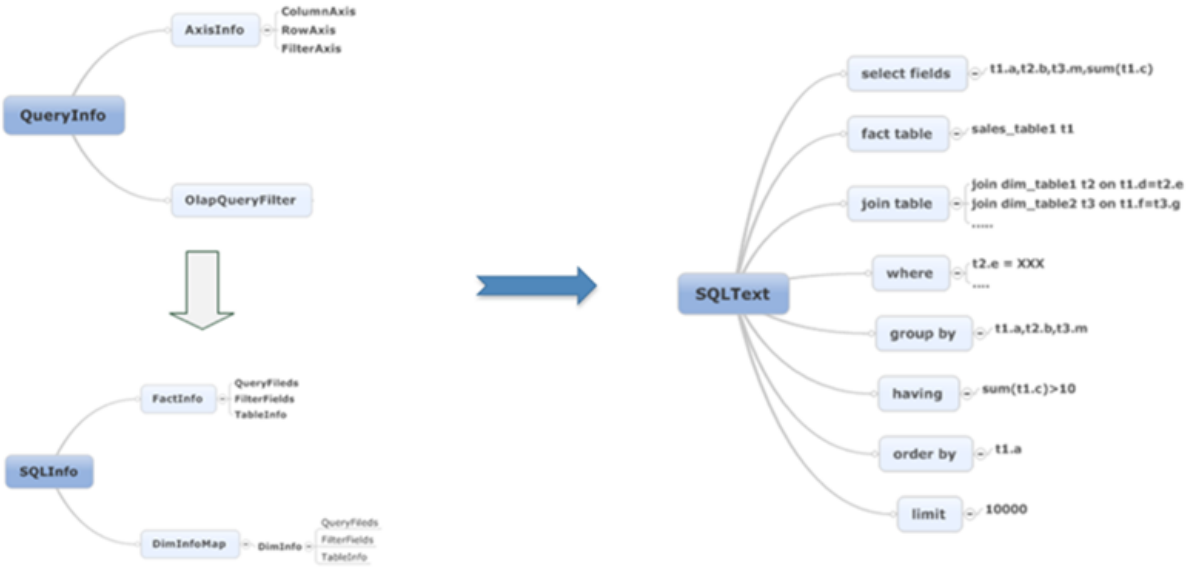
当您选择维度时，维度的来源childCube可能是数据集关系树上任意一个节点。而节点必须与主cube建立关系才能连接起多个维度/度量的查询，否则查询是松散而没有意义的。为了建立起这种关系，本发明的方案是从叶子节点（即维度/度量节点）开始，从下往上追溯关系路径。扫描出所有路径之后，再对路径做分段、聚合，得到合理的查询路径。

41.2.3.3.3 生成查询语句

查询路径分析完毕之后，就可以开始最后一步，生成查询语句。

作为一种工具性的 SQL 生成引擎，必须遵循一定的规则来生成 SQL。表现这种树型关联，要使分析复杂度不随着层次深度而线性增加，因此必须是可递归的处理模型。因此本方案使用 SQL 的子查询作为关联关系的逻辑处理单元。每一个childCube都会映射为一个子查询，向上层屏蔽内部处理细节。而每个逻辑处理单元的 SQL 生成逻辑，与普通多张事实表之间的关联查询类似，可以重用该场景下的逻辑，如图 225: 生成查询语句所示。

图 225: 生成查询语句



41.2.3.4 查询执行

在真正的执行查询时，需要有一个查询引擎进行分布式计算，其大概的逻辑如下：

1. 解析SQL。
2. 生成逻辑执行计划。
3. 生成物理执行计划。
4. 分布式执行。
5. 结果汇总。

相关举例如图 *226:* 举例、图 *227: join*举例和图 *228:* 执行计划所示。

图 226: 举例

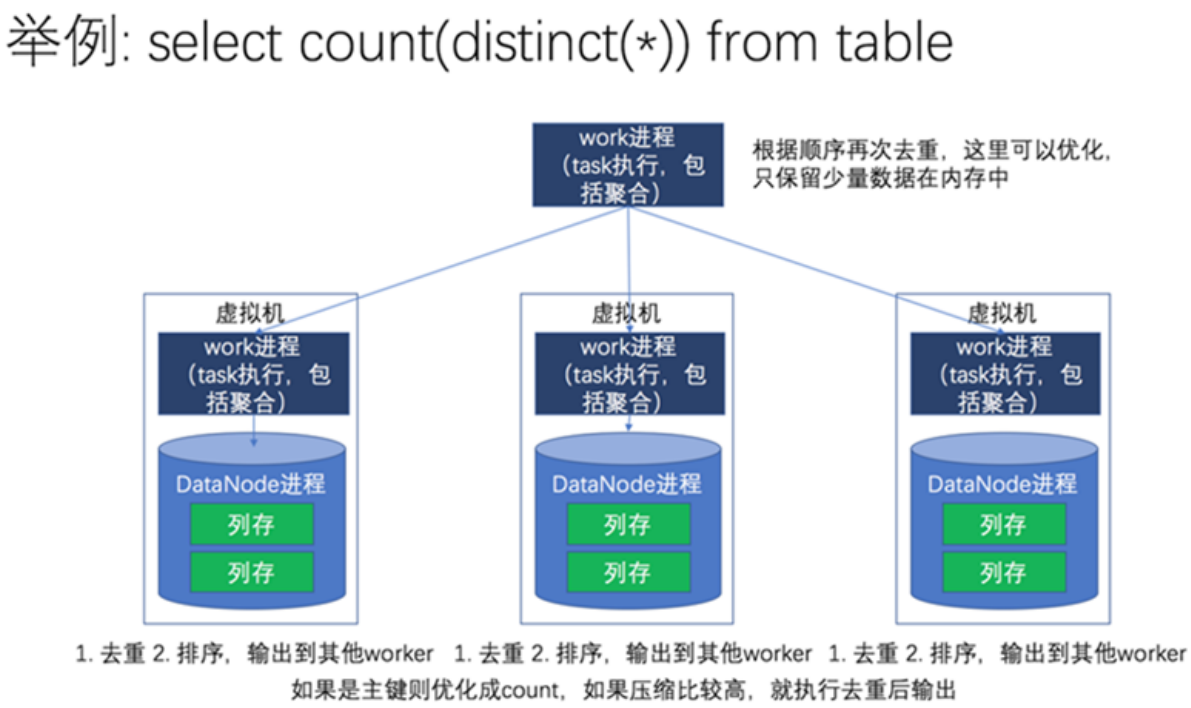


图 227: join举例

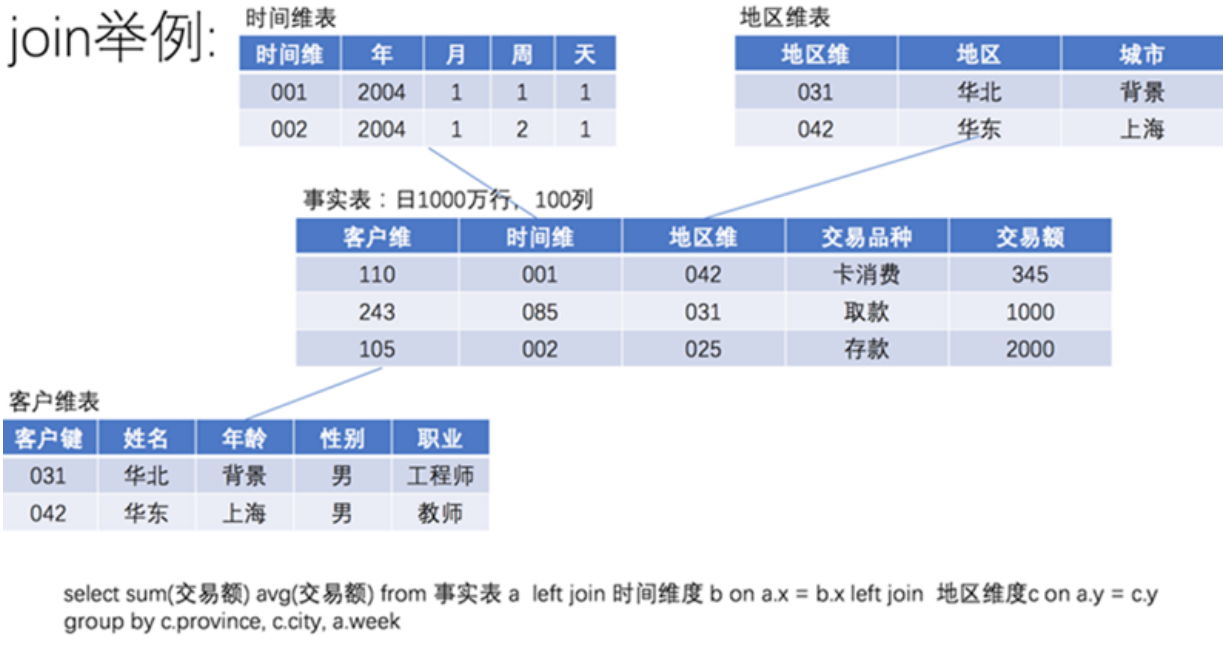
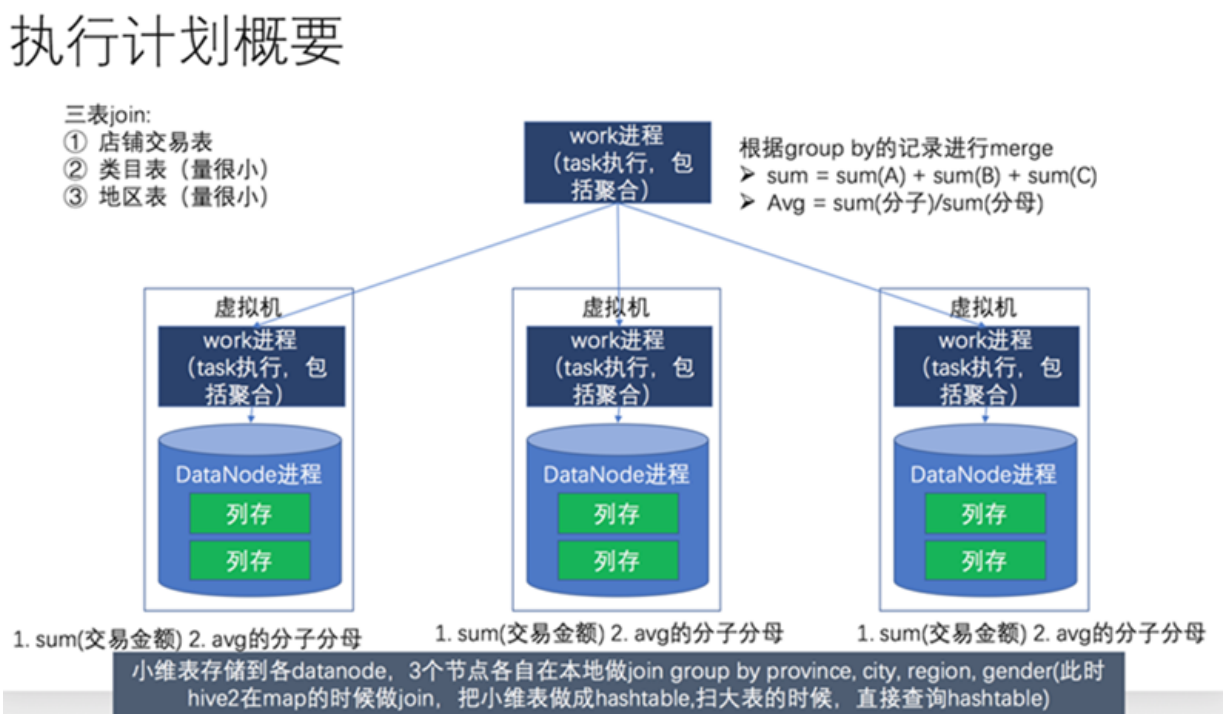


图 228: 执行计划

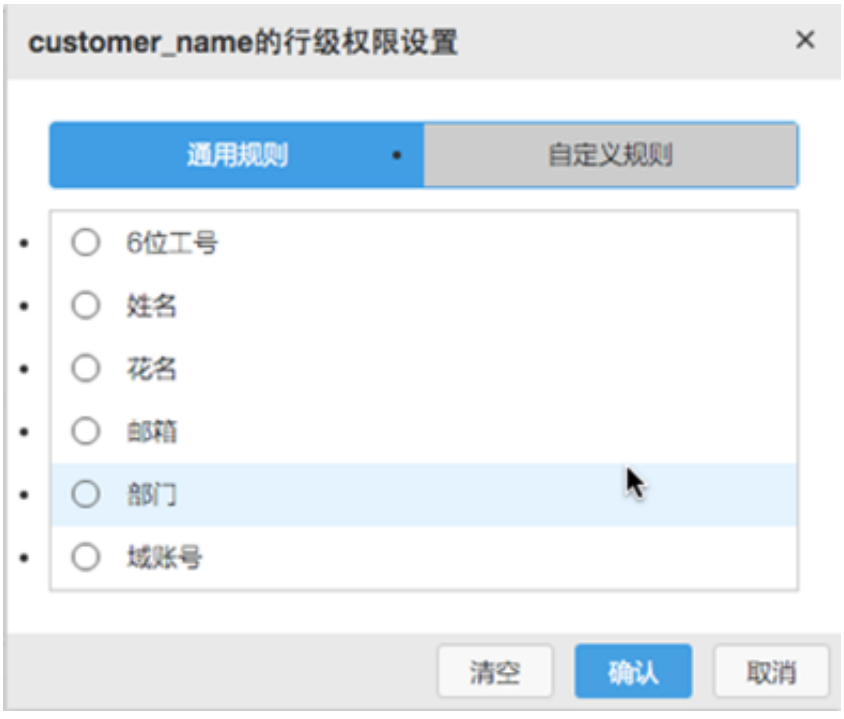


41.2.3.5 行级权限

行级权限是大数据分析平台的重要功能，通过行级权限的设置，我们可以实现对一张报表，不同的人看到不同的内容，大大地提升了数据展示的效率及数据安全性。

在使用金融云账户登录情况下的行级权限设置，配置如图 *229:* 行级权限设置所示。

图 229: 行级权限设置



其中以下四个参数的设置是有效的：

- 姓名：对应到金融云的登录账号中的**姓名**
- 花名：对应到金融云的登录账号中的**昵称**
- 邮箱：对应到金融云的登录账号中的**邮箱**
- 域账号：对应到金融云的登录账号中的**账户名**，即登录系统的账号名称。

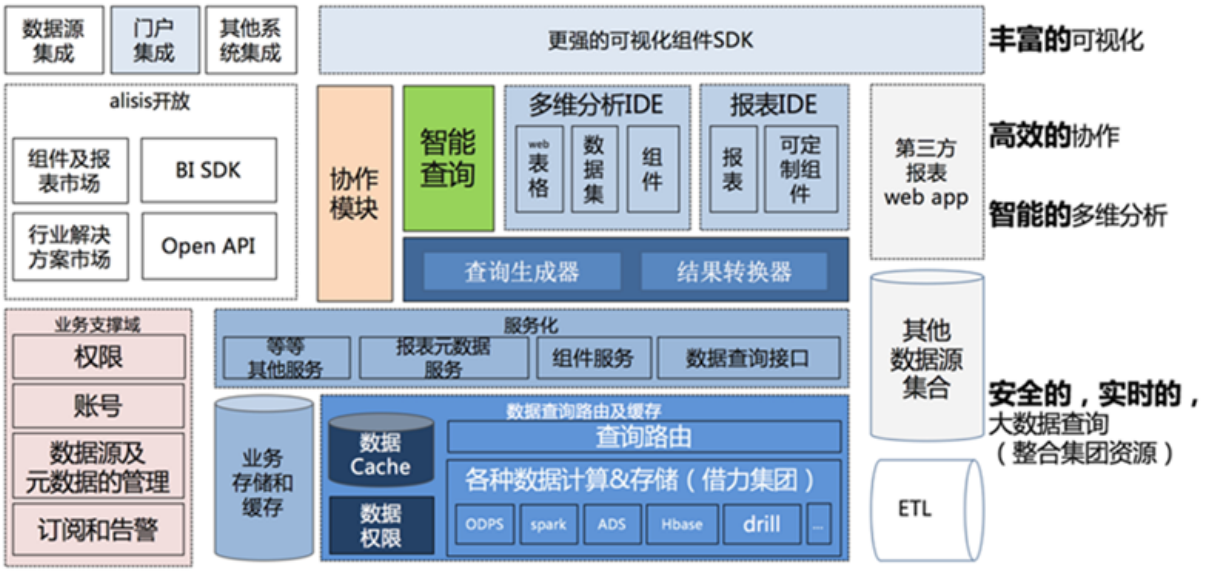
在行级权限设置的情况下，系统会根据报表阅读者的上述信息，对数据进行额外的过滤，并将满足条件的数据返回给报表页面进行呈现。

41.2.4 产品优势

基于 MaxCompute 的存储

如图 *230: 架构图*所示，数据的存储是存储在阿里集团自研的高性能、高扩展、高稳定性的分布式存储系统上，这套分布式存储可以构建在廉价的服务器上，从而在较低的成本情况下构建高稳定性，高扩展性的存储。

图 230: 架构图



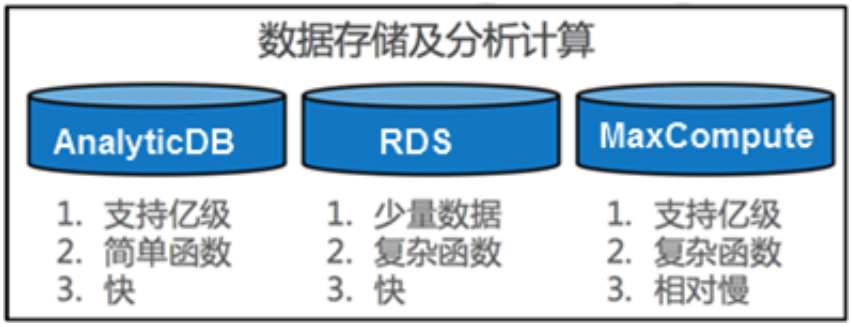
基于 MaxCompute 的ETL

MaxCompute 的计算模型是 MRR 模型，其和经典的MR的模型的区别在于 MRR 模型中，Reduce 之后会将数据重新写回磁盘，增加了 IO 的次数，加大了总的处理时间。

基于 RDS/AnalyticDB/MaxCompute 的多维分析

根据不同数据量的情况，我们会将数据加载到不同的加速器上，如图 *231:* 数据存储及分析计算所示。

图 231: 数据存储及分析计算



尤其是在使用 AnalyticDB 的情况下，行式结构被转换成列式结构，会大大的加快数据加载及分析的速度，有着相当大的性能优势，而且压缩后占用的空间可以缩小1/10到1/40。

丰富的可视化效果

支持各类图表，如图 *232:* 各类图表所示。

图 232: 各类图表

41.3 机器学习平台

41.3.1 产品概述

机器学习平台（Machine Learning Platform，简称MLP）是蚂蚁大数据解决方案（DPC）的一个分支产品，机器学习平台构建于MaxCompute计算集群之上，汇集了阿里集团大量优质分布式算法，包括数据处理、特征工程、机器学习算法、文本算法等等，可高效的完成海量、亿级维度数据的复杂计算，给业务带来更为精准的洞察力；同时，该平台提供了一套极易操作的可视化编辑页面，大大降低了数据挖掘的门槛，提高建模效率，最终帮您快速得到大数据后背隐藏的秘密。

41.3.2 产品架构

机器学习平台在整个蚂蚁大数据解决方案的位置如图 *233:* 大数据解决方案所示。

图 233: 大数据解决方案

机器学习平台PAI主要包括三个功能模块：算法组件管理、实验管理和模型管理。

下面将对这三个功能模块进行分别的介绍。

41.3.3 功能特性

41.3.3.1 算法组件

算法组件提供了一套完整的支持数据建模的算法组件库，从数据预处理到模型评估。

目前提供的算法组件近100个，包括常用的算法，如逻辑回归、随机森林、kmeans等；也包括最近流程的算法GBDT、word2vec、PLDA等。

41.3.3.1.1 数据IO组件

平台与 MaxCompute 云计算集群是直接绑定的，因为可以无感知地读写 MaxCompute 的表；平台的数据和计算任务都在 MaxCompute 集群中进行，因此保证大数据存储和分析能力。

数据 IO 组件具备如下功能特征：

- 读数据表，可查看数据字段和前100条数据。
- 写数据库，运行的结果、报告都可以保存到 MaxCompute 中。
- 支持本地数据上传。

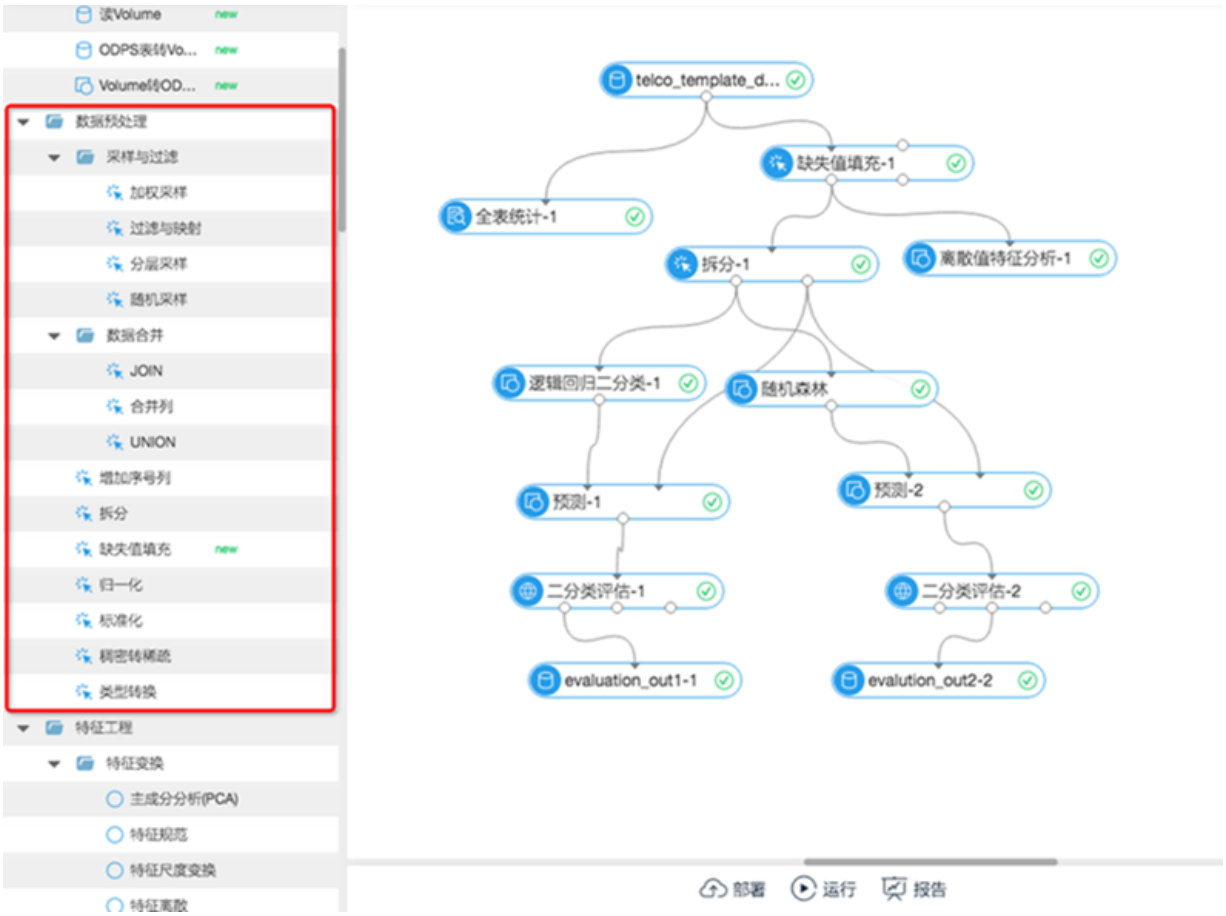
41.3.3.1.2 预处理组件

系统提供了24个数据预处理的组件，可以快速进行数据清理、数据加工等操作，同时提供各种统计分析组件，可以进行数据探查、数据质量的检测，为数据建模提供必须的数据准备。

如图 *234:* 预处理组件所示，预处理组件具备如下功能特征：

- 数据采样：包含加权采样、分层采样、随机采样。
- 数据过滤：提供 SQL 脚本级别的数据过滤功能。
- 数据合并和拆分：数据融合。
- 类型转换：5种数据库类型转换。
- 缺省值填充：采用min、max、mean等统计方法的填充。
- 统计分析：包含直方图、百分位、相关系数矩阵、特征各种统计指标等。

图 234: 预处理组件



41.3.3.1.3 特征工程组件

特征工程组件提供了13个算法组件，可以从多个维度进行特征选择和特征变换。

如图 *235:* 特征工程组件所示，特征工程组件具备如下功能特征：

- 特征变换：包含PCA、特征线性变换、特征离散化等功能。
- 特征选择：提供基于过滤式和算法模型两种方式的特征选择。
- 特征生成：将非线性特征通过GBDT编码成线性特征。

图 235: 特征工程组件

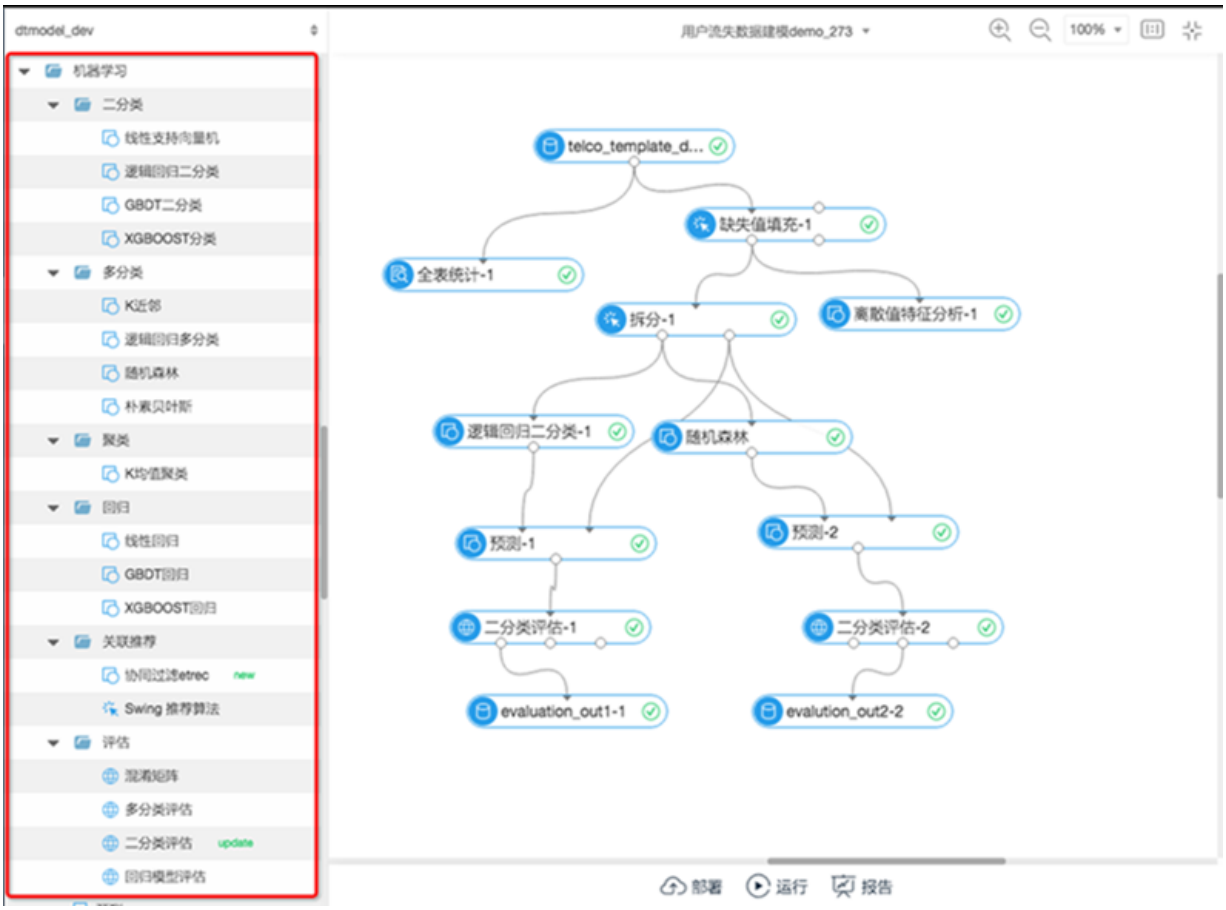
41.3.3.1.4 机器学习组件

机器学习组件提供了20多个算法组件，包括常用的机器学习算法以及强大的模型评估功能。

如图 *236:* 机器学习组件所示，机器学习组件具备如下功能特征：

- 提供了分类、回归、聚类、关联等类别的机器学习算法。
- 支持模型预测功能。
- 支持ROC、KS、混淆矩阵、PR等模型评估指标。

图 236: 机器学习组件



41.3.3.1.5 垂直领域组件

垂直领域组件提供了文本分析、网络分析等相关17种组件，可以支持在文本领域、社交领域的业务。

如图 *237:* 垂直领域套件所示，垂直领域组件具备如下功能特征：

- 支持分词、TF-IDF、PLDA、word2Vec等文本分析算法。
- 支持K-core、pagerank、标签传播聚类、标签传播分类等网络分析算法。

图 237: 垂直领域套件

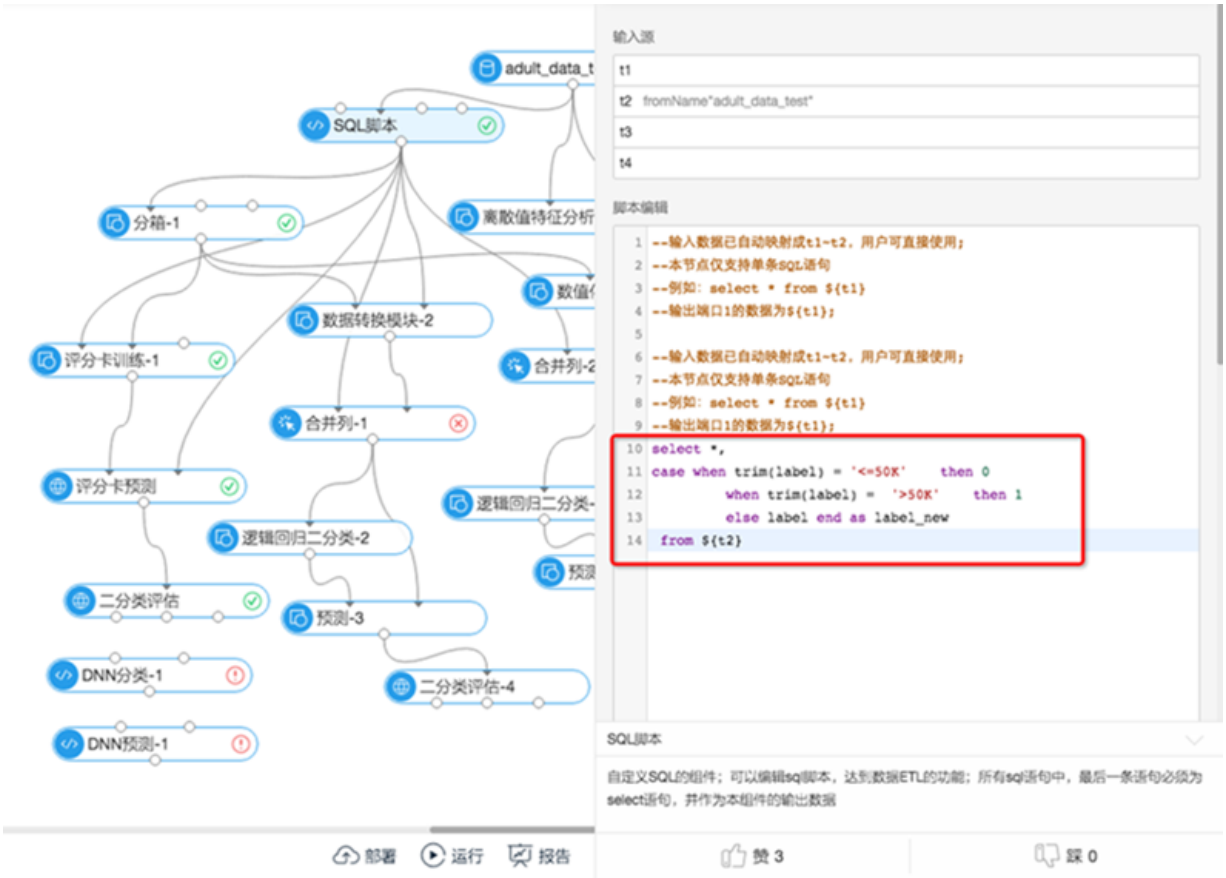
41.3.3.1.6 自定义组件

为了提供更灵活的建模流程，满足个性化需求，平台提供了自定义 SQL 组件、MR 组件等，可直接编写自己的数据处理逻辑。

如图 *238:* 自定义组件所示，自定义组件具备如下功能特征：

- SQL 组件可支持任意 MaxCompute SQL 的脚本语法。
- MR 组件支持 MaxCompute MR 的任务，支持设置 MR 的各种运行参数。

图 238: 自定义组件



41.3.3.2 建模流程管理

通过一系列的算法组件组合以及组件的关联关系，构建一个具有一定数据分析或数据挖掘功能的实验流程。

平台通过提供更好地实验编辑功能和实验管理功能，为您沉淀建模知识和业务解决能力。

41.3.3.2.1 实验模板

如图 *239: 创建实验*所示，创建一个空白的实验或从官方提供的实验模板中创建一个自己的实验。

图 239: 创建实验

在实验中，您可以使用官方提供的近100种算法组件，来完成自己建模流程，实现业务价值。

41.3.3.2.2 实验分享

将实验流程分享给指定的人，并具有该实验的所有权限，如图 *240:* 分享实验所示。

图 240: 分享实验

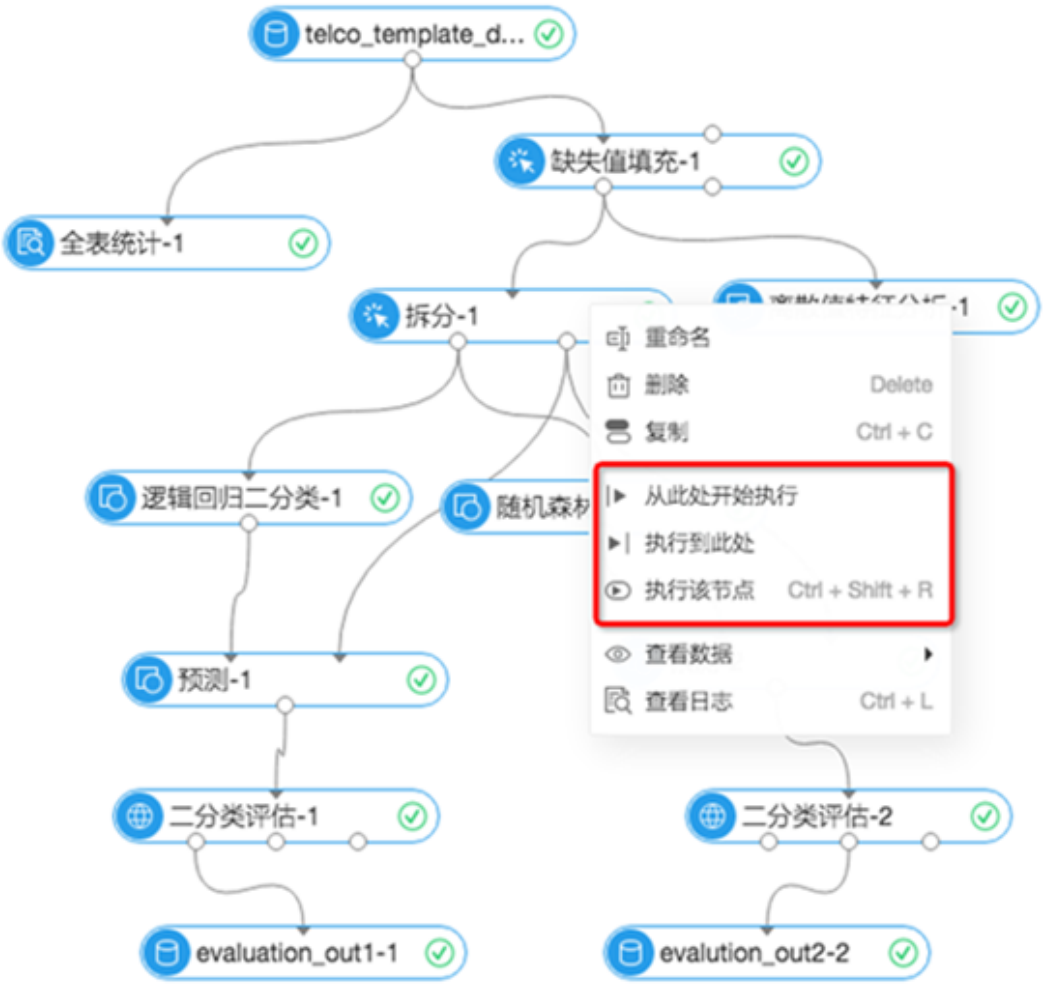
41.3.3.2.3 实验运行策略

为了更有效率的完成数据挖掘的工作，平台提供了多种DAG流程运行策略。

如图 *241:* 实验运行策略所示，实验运行策略具备如下功能特征：

- 支持全局运行，依次运行实验中每一个组件。
- 支持单点运行，运行指定的某个组件。
- 支持并行运行，同时可以运行多个分支上的组件。
- 支持起点运行，依次运行指定组件以及后续组件。

图 241: 实验运行策略



41.3.3.2.4 画布操作

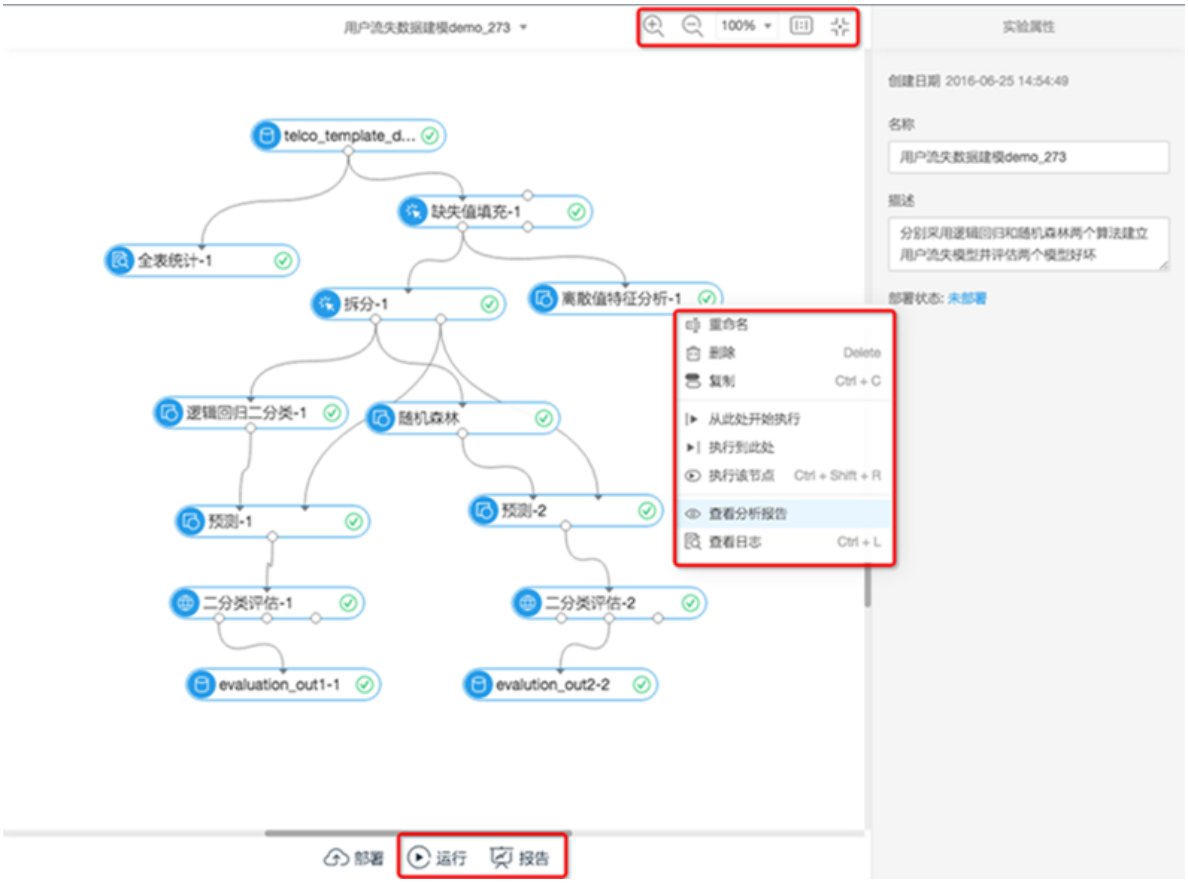
画布操作提供强大的画布编辑功能以及对于算法组件的编辑操作。

画布操作具备如下功能特征：

- 支持组件拖拽到画布。
- 支持画布缩放、居中定位等操作。
- 在组件连线时，支持智能数据类型提示。
- 实时监控组件运行的状态和时间。
- 支持组件复制、组件参数编辑。
- 支持查看组件运行日志。

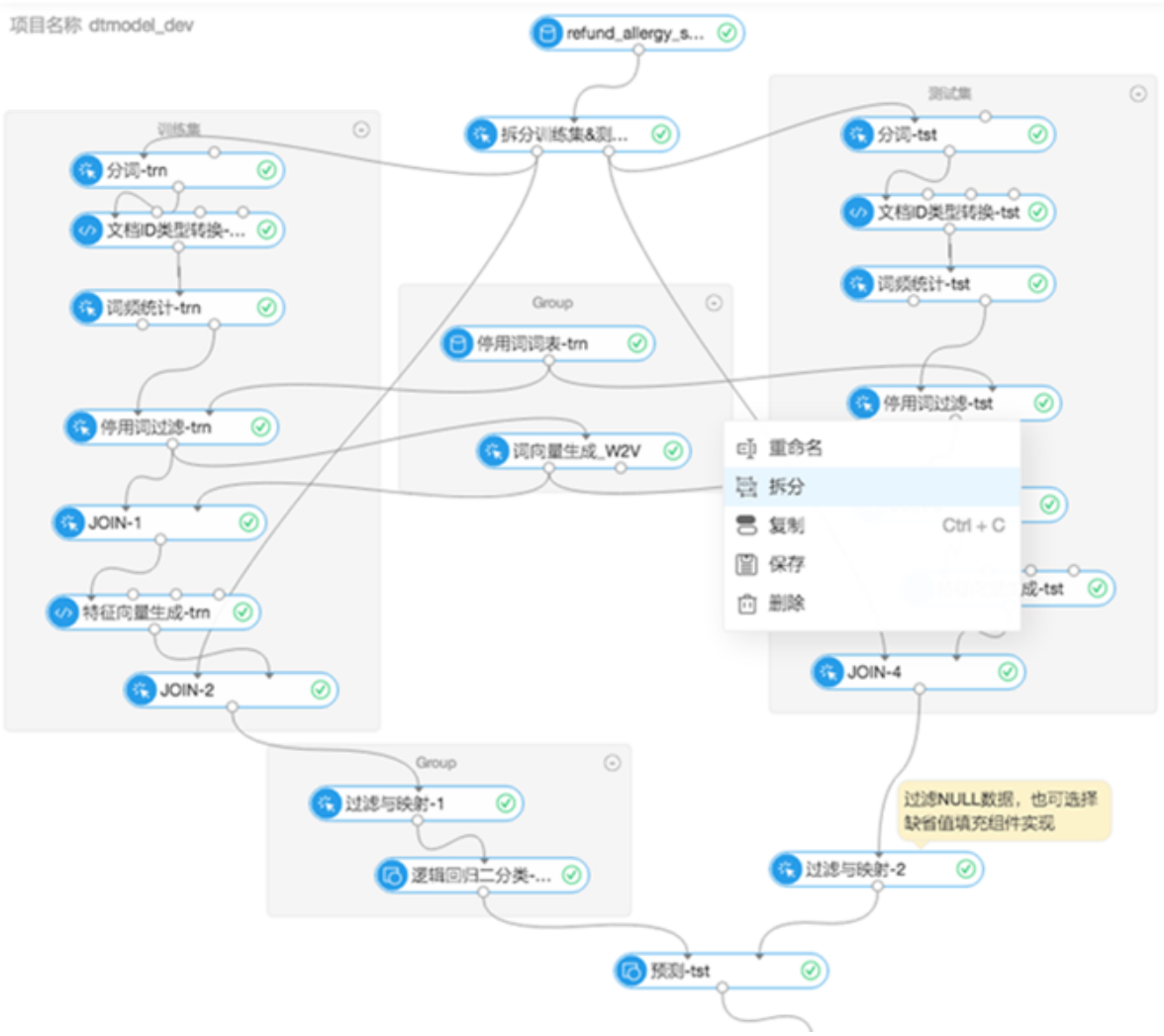
- 支持查看组件运行的结果数据或分析报告，如图 *242:* 分析报告所示。

图 242: 分析报告



- 支持组件合并功能和文本注释，如图 *243:* 组件合并及文本注释所示。

图 243: 组件合并及文本注释



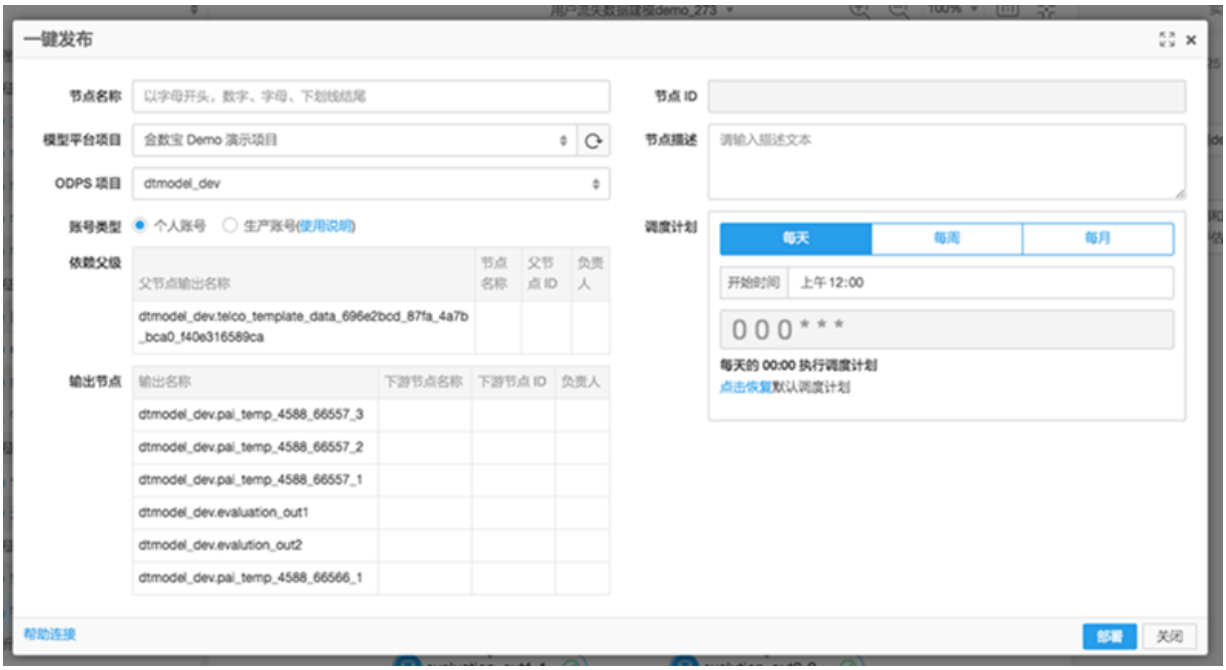
41.3.3.2.5 实验部署

配置调度时间、依赖上下游的数据，一键发布到金数宝平台，实现实验流程定时调度和依赖调度。

如图 *244: 一键发布*所示，实验部署具备如下功能特征：

- 支持个人账号和生成账号。
- 支持添加上下游表的依赖关系。
- 支持按月、按周和按天调度。

图 244: 一键发布



41.3.3.3 模型管理

平台提供了模型生命周期的管理功能，支持模型生成、模型使用、模型上线、模型实时更新等操作。

模型管理具备如下功能特征：

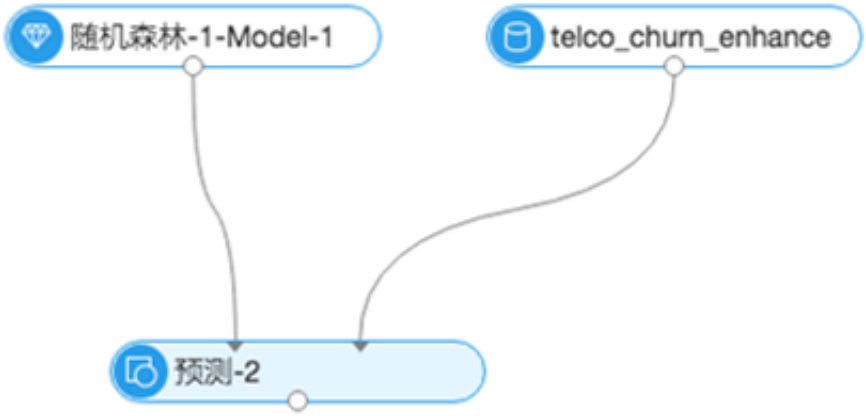
- 支持离线数据的模型预测。
- 支持线上数据模型预测，且可根据预测效果，实时更新模型。
- 支持模型生成 PMML 文件导出。
- 支持丰富的模型可视化。

41.3.3.3.1 模型预测

平台训练出来的模型会自动保存在模型列表中，可查看模型的相关信息。

模型可以先直接拖拽到画布中，用于离散数据的预测，如图 *245:* 模型预测所示。

图 245: 模型预测



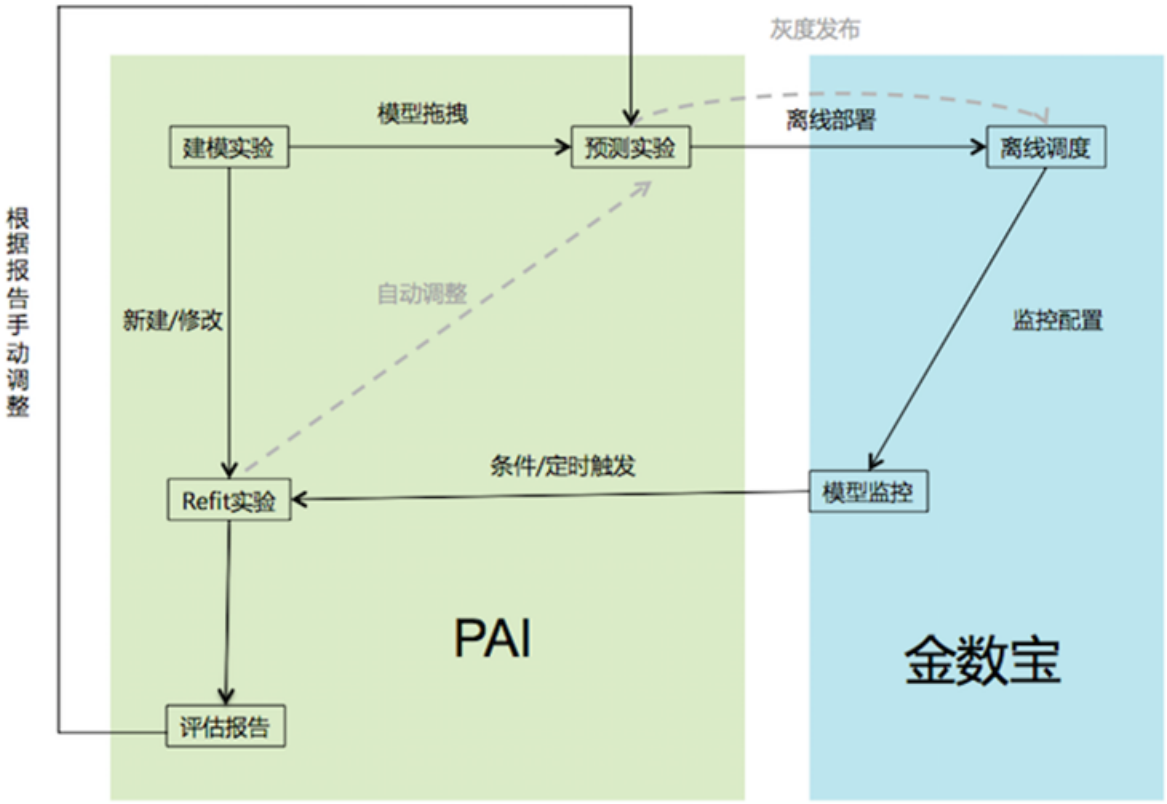
41.3.3.3.2 模型输出

模型可生成标准的 PMML 模型文件导出。

41.3.3.3.3 模型回溯

模型平台金数宝实时监控模型预测结果，并返回算法平台，触发实验训练过程，采用最新的数据训练模型，一键部署最新的数据模型，完整线上模型的refit功能，如图 *246:* 模型回溯所示。

图 246: 模型回溯



41.3.3.4 可视化分析

平台提供丰富的数据可视化功能，支持更方便的数据探查、数据分布、特征分析等操作；支持数据分箱交互功能，效果优于model builder。

41.3.3.4.1 数据可视化

产品支持多种维度的数据可视化。

如图 *247: 数据可视化*所示，数据可视化具备如下功能特征：

- 支持原始数据的查看。
- 支持数据分布查看，可调整是分箱数，支持动态调整数据展示范围。
- 支持分布统计值的可视化。
- 支持柱状图与线图的切换。
- 支持保存到本地。

图 247: 数据可视化

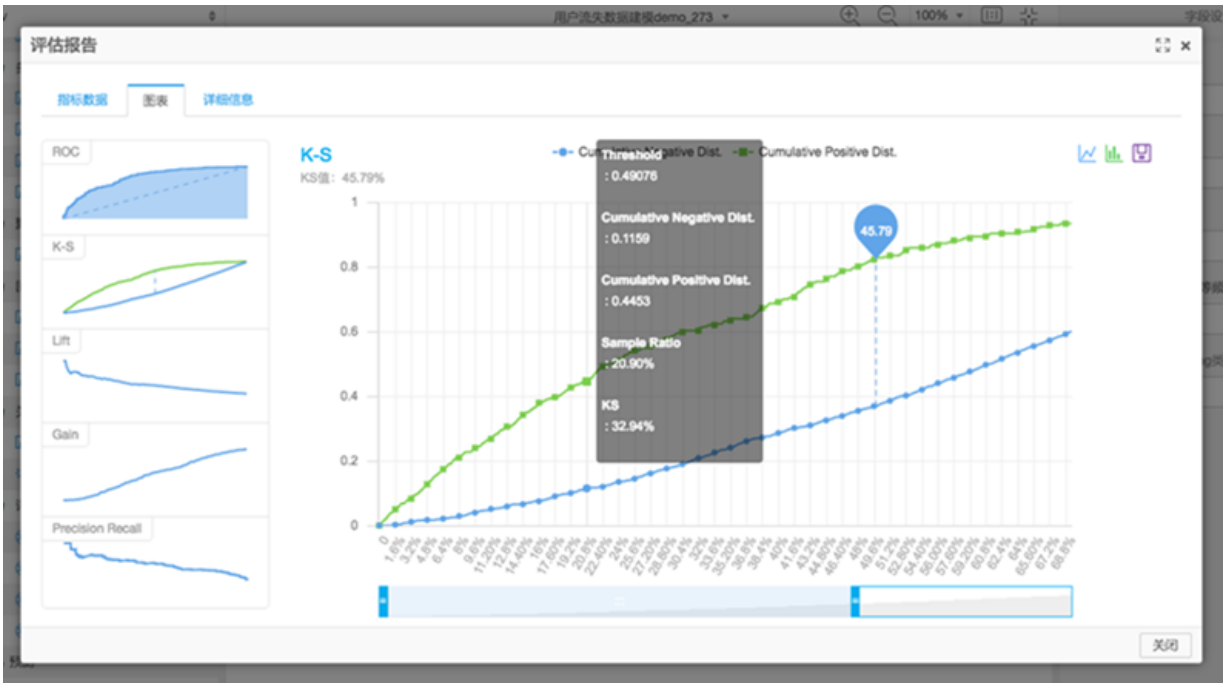
41.3.3.4.2 评估可视化

产品支持常用模型评估指标的可视化。

如图 *248:* 评估可视化所示，评估可视化具备如下功能特征：

- 支持ROC、KS、PR、Lift、Gain等指标的可视化。
- 支持指标图像的缩放功能。
- 支持保存到本地。

图 248: 评估可视化



41.3.3.4.3 分箱可视化

产品支持等频、等距操作，也支持自定义的分箱操作。

如图 *249: 列表显示分箱详情*和图 *250: 图表显示分箱详情*所示，分箱可视化具备如下功能特征：

- 支持等频、等距、按照 WOE 排序自动分箱。
- 支持先分箱的合并和拆分。
- 支持设置分箱的约束条件。
- 支持计算分箱的WOE和IV值。
- 支持分箱的数据分布可视化。

图 249: 列表显示分箱详情

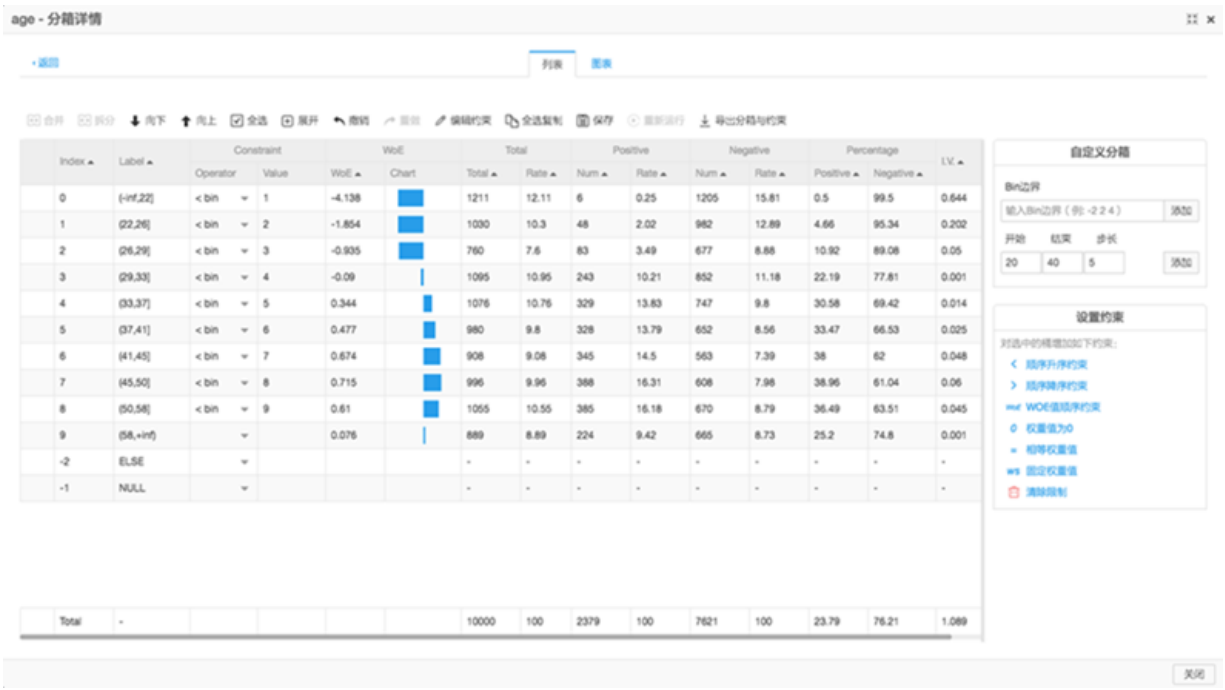
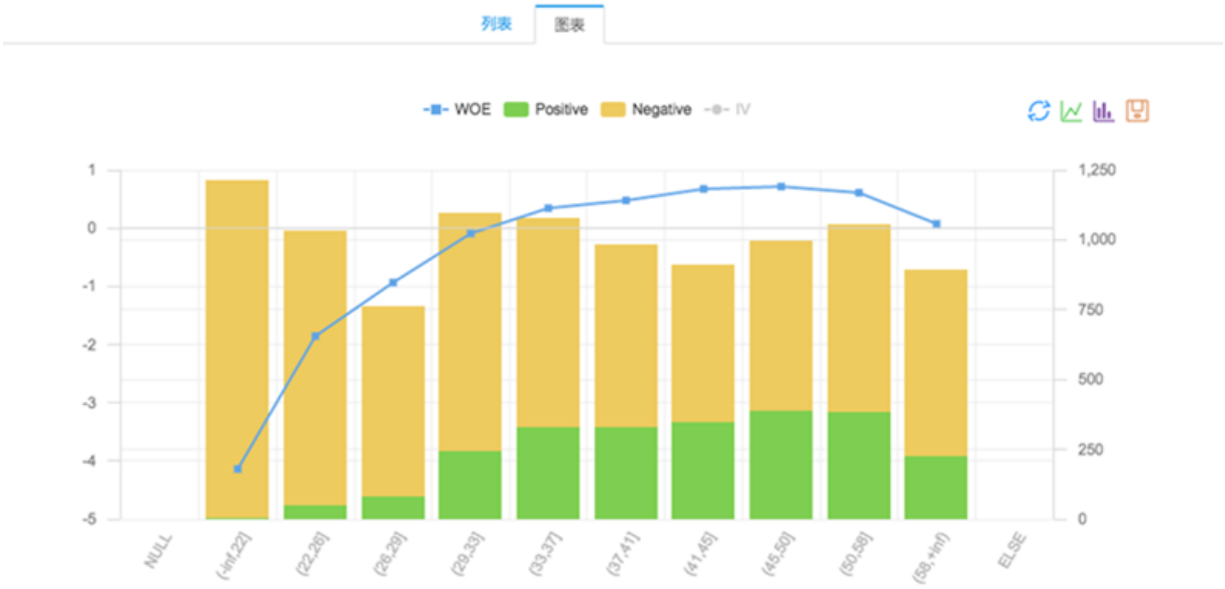


图 250: 图表显示分箱详情



41.3.3.4.4 模型可视化

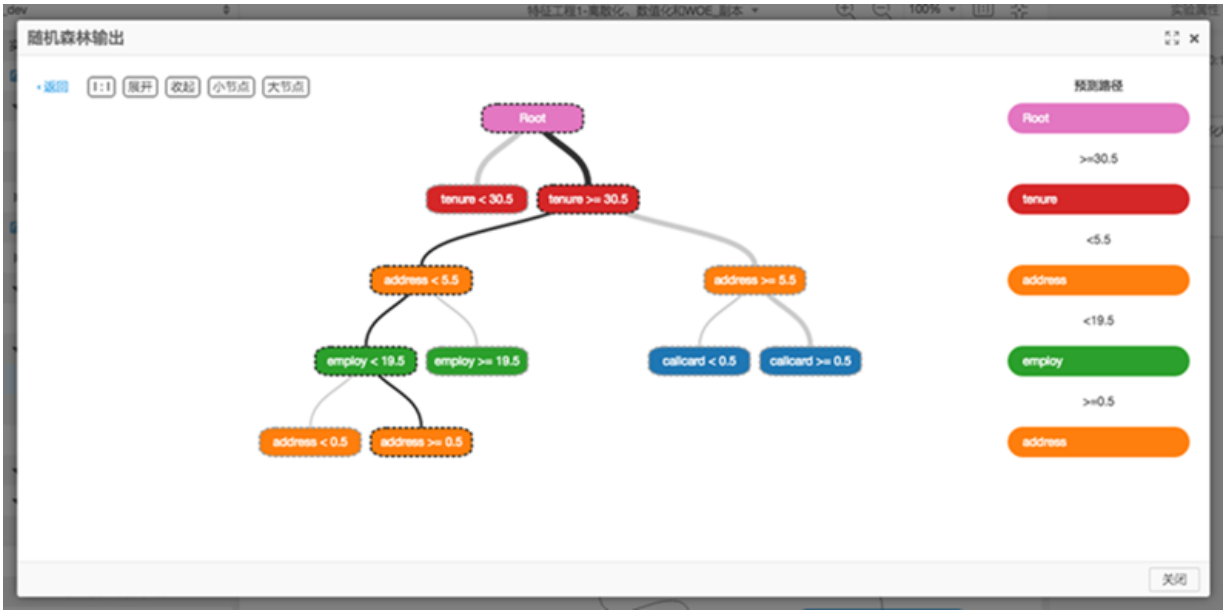
产品支持决策树结构模型的可视化功能。

如图 *251:* 模型可视化所示，模型可视化具备如下功能特征：

- 支持展示每个节点的规则。

- 支持展示决策规则链路。

图 251: 模型可视化



42 机器学习PAI

42.1 产品概述

机器学习简单来说就是人教机器在我们积累的数据当中发现规律，然后能够辅助我们来做一些预测和决策。

机器学习笼统地讲可以分为三类：

- **有监督学习**（supervised learning）是指每个样本都有对应的期望值，然后通过搭建模型，完成从输入的特征向量到目标值映射，典型的例子是回归和分类问题；
- **无监督学习**（unsupervised learning）是指在所有的样本中没有任何目标值，我们期望从数据本身发现一些潜在的规律，比如说做一些简单的聚类；
- **增强学习**（Reinforcement learning）相对来说比较复杂，是指一个系统和外界环境不断地交互，获得外界反馈，然后决定自身的行为，达到长期目标的最优化，其中典型的案例就是阿法狗下围棋，或者无人驾驶。

机器学习兴起的因素

最近几年，机器学习比以前更火了，主要是我们在深度学习技术上取得了一定的进展，总结起来应该是三大因素：

- 第一个因素是数据的因素。互联网上每天生成海量的数据，有图像、语音、视频、还有各类传感器产生的数据，例如各种定位信息、穿戴设备；非结构化的文本数据也是重要的组成部分。数据越多，深度学习越容易得到表现好的模型。
- 第二个因素是大规模分布式高性能计算能力的提升。这些年来，GPU高性能计算、分布式云计算等计算平台迅猛发展，让大规模的数据挖掘和数据建模成为可能，也为深度学习的飞跃创造了物质基础；阿里云的愿景之一就是成为和水电煤一样的基础设施。
- 第三个因素是指算法上的创新。随着数据和计算能力的提升，算法本身也有了很大的进展，尤其在深度学习方面，譬如从脑神经学上得到的灵感，在激活函数上进行了稀疏性的处理，等等。

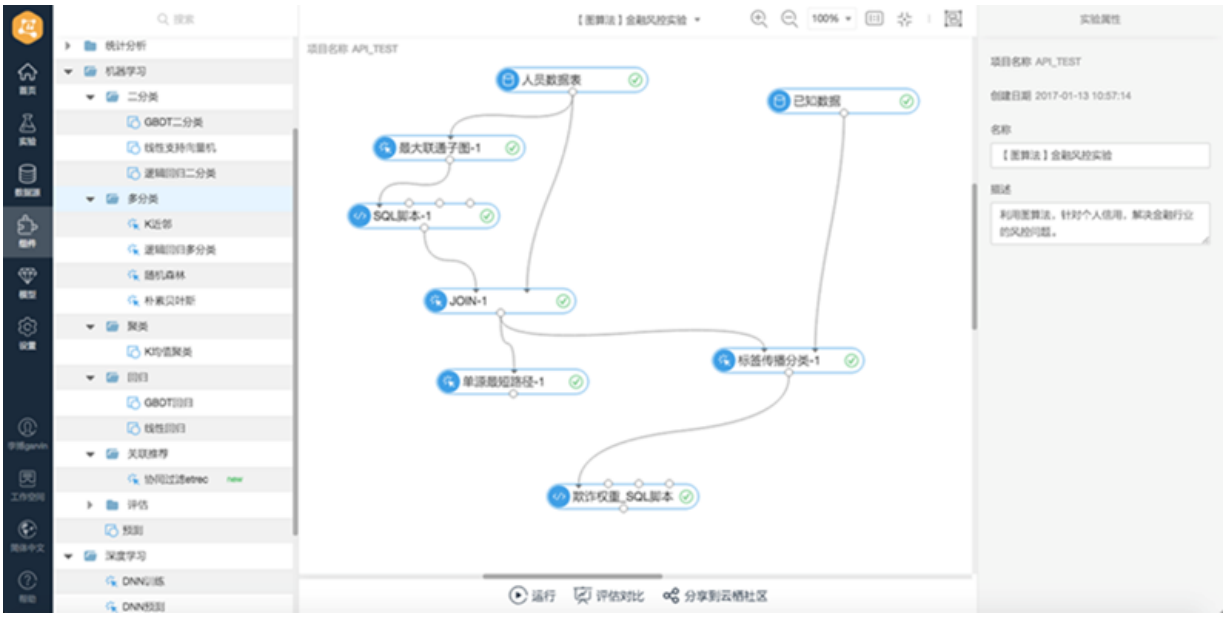
基于上述三点，人工智能又迎来了它的第二个春天。人工智能将以更快的速度进入我们的生产和生活中来，成为我们的眼睛，我们的耳朵，帮助我们更快捷地获取信息，辅助我们做出决策。机器学习平台产品也因此而产生，加速迭代过程，助力技术的发展。

42.2 功能特性

主流机器学习平台

阿里去年发布了自己的智能平台，其目的是为了加速整个创新过程，提高工作效率。该平台是基于阿里云的云计算平台，具有处理超大规模数据的能力和分布式的存储能力，同时整个模型支持超大规模的建模以及计算。该智能平台主要分为三层，第一层是 Web UI 界面，第二层是 IDST 算法层，最后一层是 MAXCOMPUTE 平台层。下文更加详细地介绍。

PAI 平台界面



上图所示的是 PAI 平台内部的界面。左边是主要功能区，中间是一个画布。使用者可以用鼠标将相应的组件拖拽到画布上，形成一个有向的工作流，完成从元数据到数据处理再到建模等一系列的数据挖掘工作。右边主要是用于设置组件内参数。

下面来介绍下它主要功能：

- 搜索功能：当我们有很多数据、表、实验时，可以通过搜索功能快速查找到您需要的资料；
- 实验列表：通过双击实验名称，在画布上显示原来的有向实验流图，可以继续之前没有完成的实验；
- 数据表，它类似于文件管理器，可以查看您所有的数据表；
- 算法和工具列表，包含了常用的机器学习算法组件等核心功能；
- 模型列表，通过该功能，使用者可以管理所有的模型。

建立一个模型训练实验

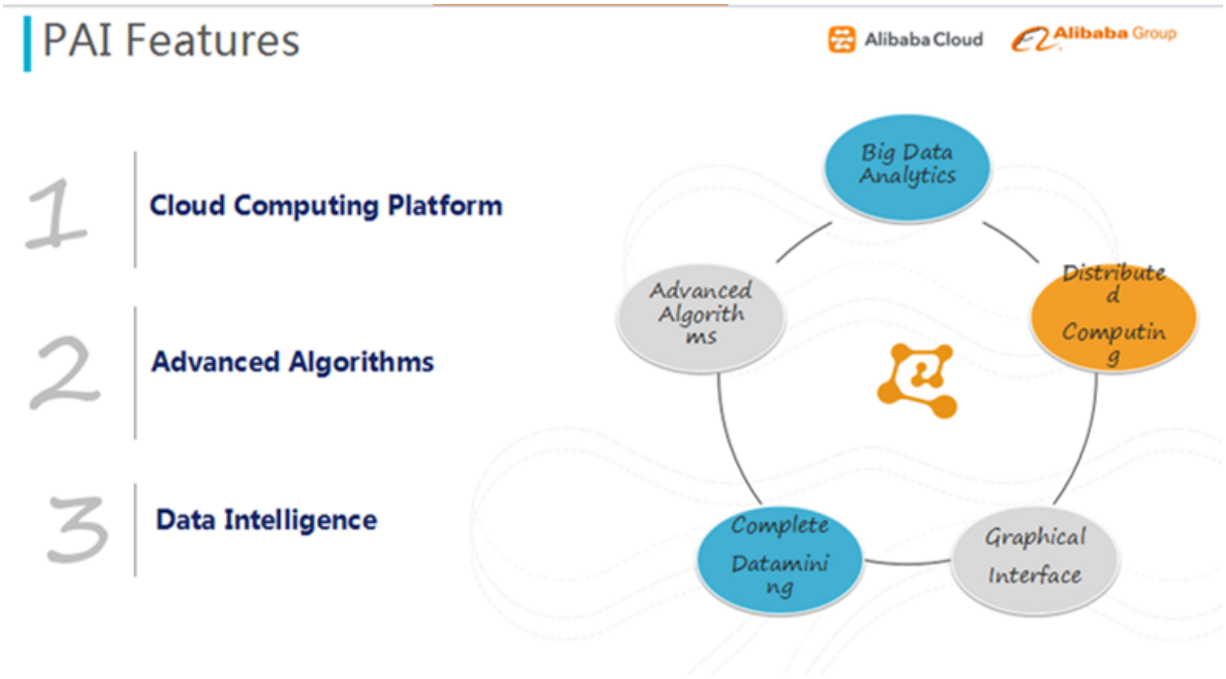
建立一个模型训练实验的步骤包括：首先，选择有运行权限的 Project，新建算法实验；然后在组件列表中拖拉数据源、特征工程、机器学习算法等组件搭建挖掘流；最后单击运行即可。

数据探索分析

在整个过程当中，需要做进行一些数据的探索分析，所有的算法都会提供一个重要性的分析，提供特征排序。这里面最为重要的是评估分析，我们希望整个评估是全面而准确的。

图形工作流的优势在于，使用者可以很容易地进行循环实验。因为在模型优化时，需要多次循环迭代优化，通过该模板，每次修改相应的参数再进行重复实验，大大提高了工作效率。

PAI 平台特点



该平台通过交互的界面降低了技术门槛，使用者可以轻松实现数据挖掘的工作，而无需太多经验；其次，其内嵌的算法，都是经过阿里内部多年的淬炼，在性能和准确率上都有较大的提升；最后是数据智能，该平台提供了从元数据到模型部署整套流程，通过提供基本的组件，使用者可以搭建各个垂直场景下的解决方案。

我们的客户主要包括以下几部分：一类是传统的大型企业和政府部门；另一类是中小企业，主要是公共云上的初创用户；以及一些个人用户，如数据科学家、研究人员等。

PAI 平台架构图

上图是 PAI 整体框架图，最底层是基础设施层，包括 CPU 和 GPU 集群；其上一层是阿里提供的计算框架，包括 MapReduce/SQL/MPI 等计算方式；中间一层是模型算法层，包含数据预处理，特征工程，机器学习算法等基本组件，帮助使用者完成简单的工作；平台化产品主要是项目管理、算法模型分享，以及一些特定的需求；最上层是应用层，阿里内部的搜索、推荐、蚂蚁金服等项目在进行数据挖掘工作时，都是依赖 PAI 平台产品。

目前依托于 PAI 机器学习平台，提供基本组件，通过拖拽的操作，完成工作流程的整体布局；同时在垂直场景下，提供一些更为专业的组件，例如在文本分析方面，我们提供了一套完整的文本分析的算法组件。

阿里云机器学习平台既支持丰富的机器学习算法。接下来发展的方向，是数据智能的方向，一个是提供智能化的组件，例如将参数调优的工作也替用户完成；另一个就是开发垂直应用领域需要的算法组件，逐渐形成行业解决方案。