阿里云

ZStack for Alibaba Cloud

多管理节点物理机高可用

产品版本：V3.0.1

文档版本：20180930

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。	 **说明**： 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 安装与部署

1.1 概述

ZStack for Alibaba Cloud以单独的高可用套件形式，提供多管理节点物理机高可用功能。当其中任何一个管理节点失联，秒级触发高可用切换，从而保障管理节点持续提供服务。

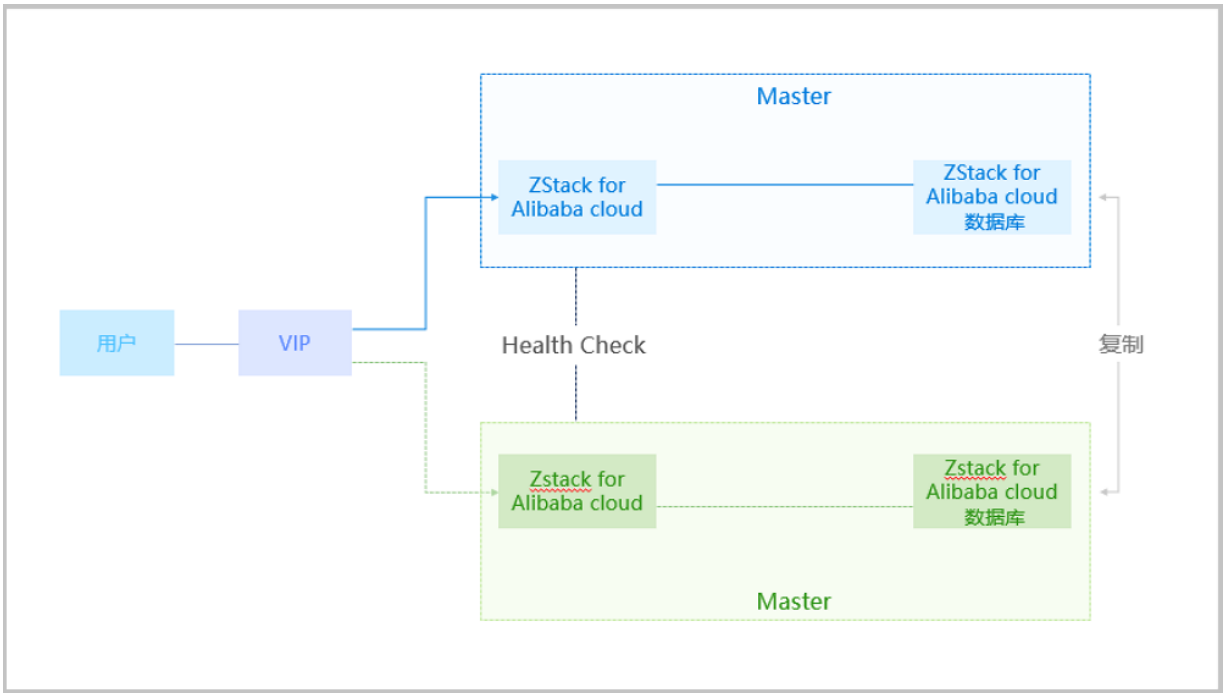
本文档以双管理节点物理机高可用为例进行介绍。

高可用原理

在双管理节点模型下，每个管理节点均运行**zsha2**高可用进程，负责实时监控管理节点上的关键服务（包括：管理节点服务、UI服务、数据库服务），当任何一个关键服务出现宕机，立即通过Keep Alived触发VIP（Virtual IP）迁移，然后尝试恢复宕机服务。

如图 *1-1:* 双管理节点物理机高可用所示：

图 1-1: 双管理节点物理机高可用



1.2 安装与部署

本章节主要介绍双管理节点物理机高可用的安装部署。

1.2.1 准备软件工具

请管理员准备以下必要的软件包，以便安装部署过程顺利执行：

- ZStack for Alibaba Cloud定制版ISO
  - 文件名称：ZStack_Alibaba_Cloud-x86_64-DVD-3.0.1-c74.iso
  - 下载地址：点击这里
- ZStack for Alibaba Cloud安装包
  - 文件名称：ZStack_Alibaba_Cloud-installer-3.0.1.bin
  - 下载地址：点击这里
- 多管理节点高可用套件
  - 文件名称：ZStack-Enterprise-Multinode-HA-Suite-3.0.1.tar.gz
  - 下载地址：点击这里

说明：
软件下载后，需通过MD5校验工具核对校验码，以确保软件完整无损。

1.2.2 核对硬件设备

本场景采用2个x86服务器部署双管理节点物理机高可用，配置信息如表 *1-2:* 服务器配置所示。管理员可根据业务性能需求，合理调配CPU、内存和硬盘的容量配比，以达到合适的平衡状态。

表 1-1: 服务器配置

	配件	型号	数量	总数
服务器	CPU	Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz	2	2个
	内存	DDR4 16GB	8	
	主板	双路服务器标准主板	1	
	阵列卡	阵列卡支持SAS/SATA RAID 0/1/10 支持直通模式	1	
	固态硬盘	Intel SSD DC S3610 480GB	2	
	机械硬盘1	SAS HDD 300GB 3.5"，15k rpm	2	

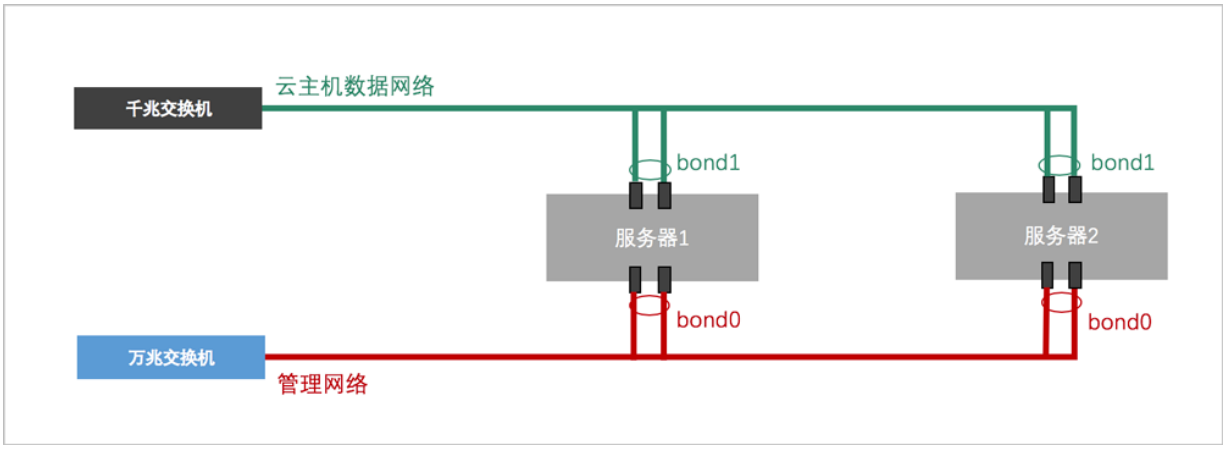
机械硬盘2	NL SAS HDD 2TB 3.5"，7.2k rpm	6
千兆网口	以太网1GbE，RJ45	2
万兆网口	以太网10GbE，SFP+	2
光电模块	-	
光纤HBA卡	-	
远程管理	DELL iDRAC企业版	1
电源	标准电源1100W	2

此外，本场景还配备了1台万兆交换机、1台千兆交换机以及若干五类跳线。

1.2.3 检查网络连接

管理员根据如图 *1-2:* 网络拓扑图所示的网络拓扑图，对上述服务器和网络设备进行上架并连线。

图 1-2: 网络拓扑图



1.2.4 安装操作系统

操作步骤

1. 准备

   管理员对上架的网络设备和服务器加载电源，手动启动服务器进入BIOS，检查以下内容：

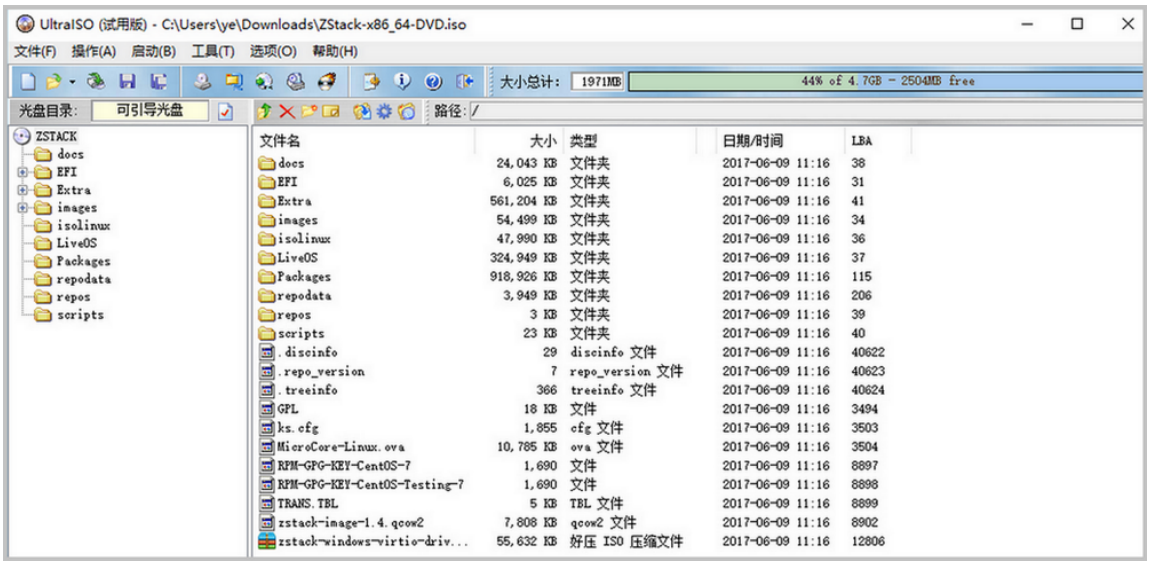
   - 激活所有CPU核心和启用超线程功能，设定系统性能为最高性能状态；
   - 打开硬件虚拟化VT功能，支持硬件虚拟化技术加速优化功能；
   - 进入阵列卡设定，对两块系统硬盘配置RAID1（Mirror），其余硬盘设定直通模式。

2. 在UltraISO打开ZStack for Alibaba Cloud DVD镜像

   - ZStack for Alibaba Cloud操作系统ISO镜像可通过DVD-RW设备刻录成安装光盘，也可通过UltraISO工具将把ISO文件刻录到U盘。
   - 打开UltraISO，点击**文件**按钮，选择打开已下载好的ISO文件。

     如图 *1-3: 在UltraISO打开DVD镜像*所示：

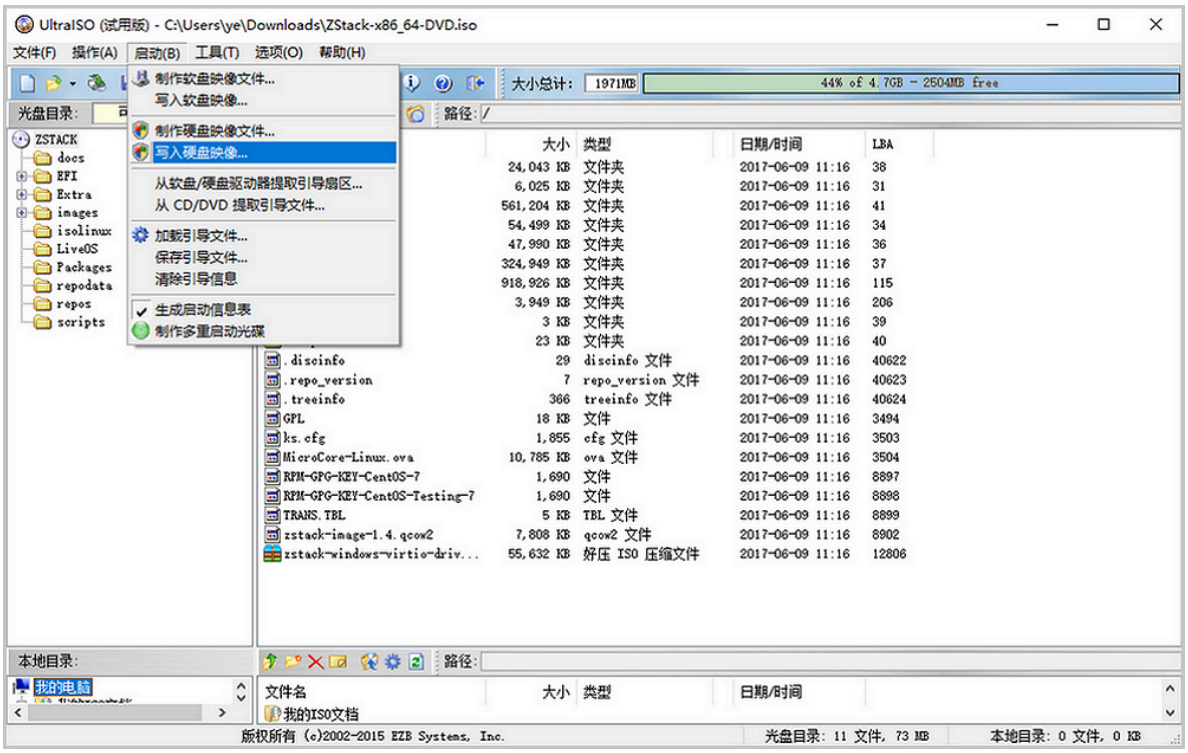
     图 1-3: 在UltraISO打开DVD镜像



3. 写入硬盘镜像

   在UltraISO点击**启动**按钮，选择**写入硬盘镜像**，如图 *1-4: 在UltraISO写入DVD镜像*所示：

图 1-4: 在UltraISO写入DVD镜像



4. 在UltraISO确认写入ZStack for Alibaba CloudDVD镜像

   - 如果系统只插了一个U盘，则默认以此U盘进行刻录和写入，在刻录前，**注意备份U盘之前的内容。**
   - 其他选项，按照默认设置，无须额外配置，点击**写入。**

   如图 *1-5: 在UltraISO确认写入ISO镜像*所示：

图 1-5: 在UltraISO确认写入ISO镜像

5. 进入安装导航

ISO镜像已经刻录到U盘。此时U盘可用来作为启动盘，支持Legacy模式和UEFI模式引导。管理员通过安装介质，引导节点启动，并进入安装导航，如图 *1-6: U*盘引导界面所示：

图 1-6: U盘引导界面

6. 安装操作系统

   默认选择**Install ZStack for Alibaba Cloud**开始安装操作系统。

   在进入安装界面后，已经预先配置默认选项：

   - **时区**：亚洲东八区
   - **语言**：English(United States)
   - **键盘**：English(US)

   一般情况下管理员无需更改配置。管理员需自行执行硬盘的分区，推荐分区如下（UEFI 模式）：

   - */boot/efi*：创建分区500MB
   - */boot*：创建分区1GB
   - *swap*（交换分区）：创建分区32GB
   - */*（根分区）：配置剩下容量

   分区配置完后，选择**Software Selection**进入服务器安装角色候选，选择**ZStack for Alibaba Cloud Management Node**管理节点模式，确定后回到主界面。

   点击**Begin Installation**进行安装。安装过程将会自动进行，管理员需要设定root账户密码。

   安装结束后，重新引导服务器并拔掉U盘。如安装成功，则服务器重启后进入操作系统登录提示符，使用root和设置的密码登录到操作系统。

说明：
管理员可根据自身需要更改密码。

1.2.5 配置网络

管理员对两台服务器均安装操作系统后，可进行网络配置。在目录*/usr/local/bin/*，ZStack for Alibaba Cloud提供便利的网络配置脚本，管理员可通过此脚本快速配置接口（Interface）和网桥（Bridge）信息。

本场景对2个管理节点的网络信息、以及用于Keep Alived通信的VIP设定如下：

表 1-2: 管理网络

服务器	网卡 1	网卡 2	聚合接口	网桥	IP地址	掩码	网关
管理节点1	eth0	eth1	bond0	br_bond0	*192.168.195.200*	*255.255.0.0*	*192.168.0.1*
管理节点2	eth0	eth1	bond0	br_bond0	*192.168.196.125*	*255.255.0.0*	*192.168.0.1*

表 1-3: 云主机数据网络

节点	网卡 1	网卡 2	聚合接口	网桥	IP地址	掩码	网关
管理节点1	em1	em2	bond1	-	-	-	-
管理节点2	em1	em2	bond1	-	-	-	-

表 1-4: VIP

-	IP地址	掩码
VIP	*192.168.199.151*	*255.255.0.0*

- 以上均为示例数据，管理员可根据实际部署环境自行更改；
- 网关需由物理网络设备提供，同时作为**网络状态仲裁检测**。

以下分别介绍管理网络和云主机数据网络的配置。

1.2.5.1 配置管理网络

本场景对管理网络设定如下：

表 1-5: 管理网络

服务器	网卡 1	网卡 2	聚合接口	网桥	IP地址	掩码	网关
管理节点1	eth0	eth1	bond0	br_bond0	*192.168.195.200*	*255.255.0.0*	*192.168.0.1*
管理节点2	eth0	eth1	bond0	br_bond0	*192.168.196.125*	*255.255.0.0*	*192.168.0.1*

对**管理节点1**执行以下配置命令：

```
# 创建聚合网卡bond1
[root@localhost ~]# zs-bond-lacp -c bond0

# 将网卡eth0与eth1均添加到bond0
[root@localhost ~]# zs-nic-to-bond -a bond0 eth0
[root@localhost ~]# zs-nic-to-bond -a bond0 eth1

# 配置上述链路聚合后,请管理员在对应的交换机网口配置LACP聚合

# 创建网桥br_bond0,指定网络IP、掩码和网关
[root@localhost ~]# zs-network-setting -b bond0 192.168.195.200 255.255.0.0 192.168.0.1

# 查看聚合端口bond0是否创建成功
[root@localhost ~]# zs-show-network
...
--------------------------------------------------------------------------------
| Bond Name  | SLAVE(s)       | BONDING_OPTS                                  |
--------------------------------------------------------------------------------
| bond0      | eth0           | miimon=100 mode=4 xmit_hash_policy=layer2+3   |
|            | eth1           |                                               |
--------------------------------------------------------------------------------
```

对**管理节点2**执行类似的配置命令。

说明：

- eth0和eth1加载到bond0后，对应交换机的端口需要配置LACP聚合，否则网络通信将异常；如果交换机不支持LACP聚合，请联系网络设备厂商更换设备。
- 通过bond0创建网桥后，网桥命名为br_bond0，将提供管理网络服务。
- 关于网桥的IP地址、子网掩码和网关参数，用户需按照实际情况填写。
- 管理网络配置完成后，可通过**ping**命令进行检测；若配置正确，则两管理节点的管理网络对应的IP地址可互**ping**。
- 管理网络建议采用万兆以上带宽，若独立部署，允许千兆带宽。

管理网络配置完成后，随之可配置云主机数据网络。

1.2.5.2 配置云主机数据网络

本场景对云主机数据网络设定如下：

表 1-6: 云主机数据网络

节点	网卡 1	网卡 2	聚合接口	网桥	IP地址	掩码	网关
管理节点1	em1	em2	bond1	-	-	-	-
管理节点2	em1	em2	bond1	-	-	-	-

对**管理节点1**执行以下配置命令：

```
# 创建聚合网卡bond1
[root@localhost ~]# zs-bond-lacp -c bond1

# 将网卡em1与em2均添加到bond1
[root@localhost ~]# zs-nic-to-bond -a bond1 em1
[root@localhost ~]# zs-nic-to-bond -a bond1 em2

# 配置上述链路聚合后,请管理员在对应的交换机网口配置LACP聚合

# 云主机数据网络,无需创建网桥

# 查看聚合端口bond1是否创建成功
[root@localhost ~]# zs-show-network
...
--------------------------------------------------------------------------------
| Bond Name  | SLAVE(s)       | BONDING_OPTS                                  |
--------------------------------------------------------------------------------
| bond1      | em1            | miimon=100 mode=4 xmit_hash_policy=layer2+3       |
|            | em2            |                                               |
--------------------------------------------------------------------------------
```

对**管理节点2**执行类似的配置命令。

说明：

em1和em2加载到bond1后，对应交换机的端口需要配置LACP聚合，否则网络通信将异常；如果交换机不支持LACP聚合，请联系网络设备厂商更换设备。

1.2.6 安装许可证

本场景下，两个管理节点安装的许可证类型要求完全一致。

CLI方式

管理员可通过CLI方式分别向两个管理节点中导入许可证。更多详情可参考*ZStack*官网教程《许可（license）更新说明》。

1.2.7 安装高可用套件

背景信息

本场景下，管理员已安装两个最新版ZStack for Alibaba Cloud管理节点，并对两个管理节点安装许可证完毕，现在对其中一个管理节点安装**多管理节点高可用套件**，即可实现双管理节点高可用。

- 管理节点1（*192.168.195.200*）
- 管理节点2（*192.168.196.125*）

假定对管理节点1安装高可用套件，则管理节点1为主管理节点，管理节点2为备管理节点。

操作步骤

1. 导入高可用套件。

   管理员已获得高可用套件，可将其导入管理节点1并解压，执行以下命令：

   ```
   # 通过scp工具将高可用套件传输到管理节点1
   [root@localhost ~]# ls
   ZStack-Enterprise-Multinode-HA-Suite-3.0.1.tar.gz

   # 将高可用套件解压，生成两个可执行文件：zsha2和zstack-hamon
   [root@localhost ~]# tar zxvf ZStack-Enterprise-Multinode-HA-Suite-3.0.1.tar.gz
   zsha2 //多管理节点高可用的安装和管理程序
   zstack-hamon //多管理节点高可用的监控程序
   ```

2. HA初始化。

   在管理节点1中安装高可用套件，执行以下命令：

   ```
   [root@localhost ~]# ./zsha2 install-ha -nic br_bond0 -gateway 192.168.0.1 -slave "root:
   password@192.168.196.125" \
   -vip 192.168.199.151 -db-root-pw zstack.mysql.password -yes
   ```

   说明：

   - 安装高可用套件，需将**zsha2**和**zstack-hamon**放在一个目录，安装过程中，**zsha2**会自动部署**zstack-hamon**以及相关配置文件。
   - 安装命令中，相关参数说明：
     - -nic：物理设备名，用于配置VIP，生产环境一般是一个管理网络的网桥，例如-nic br_bond0
     - -gateway：主备管理节点的冲裁网关，例如-gateway 192.168.0.1
     - -slave：指定备管理节点，例如-slave "root:password@192.168.196.125"

说明：

安装过程中，备管理节点的数据库会被主管理节点的数据库覆盖，请谨慎配置。

- -vip：指定Keep Alived通信的VIP，例如 -vip 192.168.199.151
- -db-root-pw：主备管理节点的数据库root密码（必须相同），例如-db-root-pw zstack.mysql.password
- -time-server：可选参数，指定时间同步服务器，用于统一时间同步，例如./zsha2 install-ha -time-server 192.168.196.125
- -cidr：可选参数，指定网络段，需覆盖主备管理节点IP、VIP和网关，例如./zsha2 install-ha -cidr 192.168.0.0/16

说明：

如果不指定，系统会自动计算出一个最小网络段，可能无法满足需求，推荐指定网络段。

- -force：可选参数，当主备管理节点的数据库始终无法完成自动同步，对主管理节点强制执行**zsha2**安装命令，例如./zsha2 install-ha -force

说明：

执行强制安装前，建议对两个数据库进行备份。

- -repo：可选参数，指定Yum源，默认为本地源，例如./zsha2 install-ha -repo zstack-local
- -timeout：可选参数，主备管理节点的数据库初始化复制超时时间，默认值为600，单位为秒，例如./zsha2 install-ha -timeout 600
- -yes：可选参数，所有设置均允许

高可用套件初始化完成后，可执行以下命令查看管理节点的状态：

```
# 查看管理节点1的状态
[root@localhost ~]# zsha2 status
Status report from 192.168.195.200
==================================
Owns virtual address:          yes //管理节点1已获取VIP，同一时刻只允许一个管理节点获取VIP
Self 192.168.195.200 reachable:   yes //管理节点1可达
Gateway 192.168.0.1 reachable:    yes //当前网关可达
VIP 192.168.199.151 reachable:    yes //VIP可达
Peer 192.168.196.125 reachable:   yes //管理节点2可达
Keepalived status:             active //Keep Alived服务处于工作状态
ZStack HA Monitor:             active //高可用监控服务处于工作状态
MySQL status:                  mysqld is alive //数据库正常工作
MN status: Running [PID:6500] //管理节点正常工作
```

```
UI status: Running [PID:9785] http://192.168.195.200:5000 //UI正常工作

Slave Status:
------------
            Slave_IO_Running: Yes //Slave IO正常运行
           Slave_SQL_Running: Yes //Slave SQL正常运行
                 Last_Error:
       Seconds_Behind_Master: 0
             Last_IO_Error:
            Last_SQL_Error:

pass '-peer user:pass@host[:port]' to show peer status
or, '-peer host' if SSH pubkey-login has been enabled. //查看管理节点2状态的说明
```

说明：

为确保双管理节点间的监控数据实时同步，建议在这两个管理节点之间做SSH免密登录。

```
# 登录管理节点1，对管理节点2做SSH免密登录
[root@localhost ~]# ssh-keygen
[root@localhost ~]# ssh-copy-id 192.168.196.125

# 登录管理节点2，对管理节点1做SSH免密登录
[root@localhost ~]# ssh-keygen
[root@localhost ~]# ssh-copy-id 192.168.195.200
```

3. 云平台初始化。

   管理员可通过VIP（*192.168.199.151*）访问管理节点1的UI界面（*http://192.168.199.151:5000*），并完成云平台初始化操作。如登录界面所示：

图 1-7: 登录界面

在管理节点1中执行以下命令，管理节点1在线切换为备管理节点，管理节点2获取VIP（*192.168.199.151*），成为主管理节点。

```
[root@localhost ~]# zsha2 demote
```

管理员可通过该VIP刷新访问管理节点2的UI界面（*http://192.168.199.151:5000*），并完成云平台初始化操作。

1.2.8 集群升级

高可用套件升级

管理员获得新版高可用套件后，可用于升级当前的**zsha2**服务。

当主备管理节点的数据库完成自动同步后，请将新版高可用套件导入主管理节点并解压，在主管理节点中执行以下命令，就可完成高可用套件升级：

```
[root@localhost ~]# ./zsha2 upgrade-ha
```

管理节点升级

在双管理节点高可用场景下，管理员只需在一个管理节点中执行以下命令，就可对两个管理节点进行升级：

```
[root@localhost ~]# ./zsha2 upgrade-mn -peerpass password ZStack_Alibaba_Cloud-installer-3.0.1.bin
```

1.3 其他操作

1.3.1 监控报警

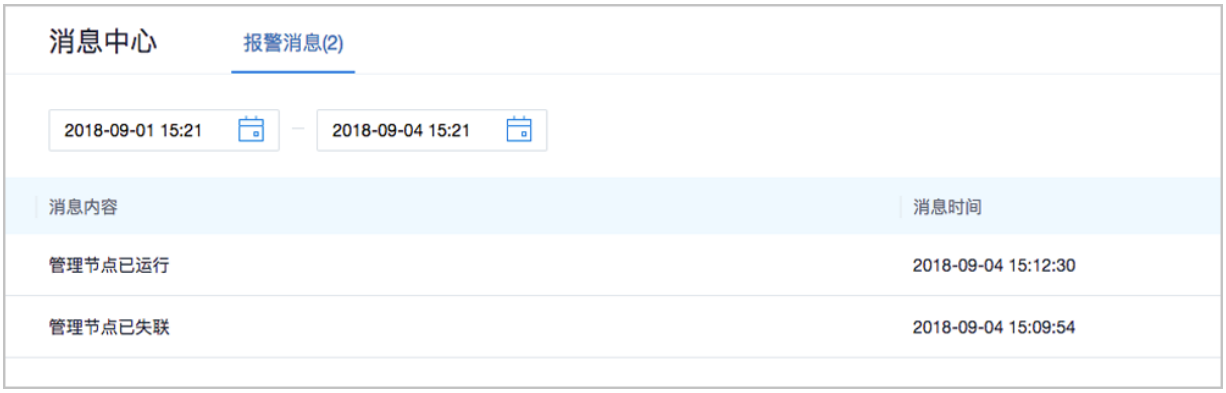
双管理节点高可用场景下，若主管理节点失联，管理员可在ZWatch中创建事件报警器，并添加相关报警条目，指定接收端，系统将以邮件/钉钉/HTTP POST方式发送报警信息，如图 *1-8: ZWatch*监控报警 主管理节点失联所示：

图 1-8: ZWatch监控报警 主管理节点失联

更多详情可参考*ZStack*官网教程《ZWatch监控系统 使用教程》。

若备管理节点失联，管理员可直接在消息中心接收到相关通知，如图 *1-9:* 消息中心 备管理节点失联所示：

图 1-9: 消息中心 备管理节点失联



1.3.2 日志输出

双管理节点高可用场景下，管理员可执行以下命令，收集**zsha2**服务相关日志。

```
[root@localhost ~]# zsha2 collect-log
Collecting logs ...
Collected log: zsha2-log-2018-09-17T154358+0800.tgz

# 将日志压缩包解压
[root@localhost ~]# tar zxvf zsha2-log-2018-09-17T154358+0800.tgz
tmp/zsha2-log588815976/
tmp/zsha2-log588815976/zsha2-status.log
tmp/zsha2-log588815976/zstack-ha.log
tmp/zsha2-log588815976/keepalived.data
tmp/zsha2-log588815976/zs-vip-192.168.199.151.log
tmp/zsha2-log588815976/keepalived_status.log
```

2 高可用测试与恢复

2.1 计划运维

2.1.1 单管理节点需要维护

主管理节点需要维护

双管理节点高可用场景下，假定管理节点1为主管理节点，管理节点2为备管理节点。

若管理员需要临时关闭管理节点1进行维护。

1. 将管理节点1主动切换为备管理节点。

   在管理节点1中执行zsha2 demote命令，管理节点1在线切换为备管理节点，管理节点2获取VIP，成为主管理节点。

2. 关闭管理节点1。

   - 若管理节点1没有被复用为计算节点添加到ZStack for Alibaba Cloud：

     1. 对管理节点1执行zsha2 stop-node，关闭**zsha2**相关服务；
     2. 对管理节点1进行shutdown关机操作；
     3. 对管理节点1下电后进行维护。

   - 若管理节点1被复用为计算节点，并已添加到ZStack for Alibaba Cloud：

     1. 对管理节点1执行zsha2 stop-node，关闭**zsha2**相关服务；
     2. 将管理节点1进入维护模式；
     3. 对管理节点1进行shutdown关机操作；
     4. 对管理节点1下电后进行维护。

3. 启动管理节点1。

   a. 对管理节点1通电后，通过手动或IPMI启动服务器；

   b. 等待管理节点1启动，成功引导操作系统；

   c. 对管理节点1执行zsha2 start-node，启动**zsha2**相关服务；

   d. 对管理节点1执行zsha2 status，查看**zsha2**服务是否正常运行；

   e. 对管理节点1执行zstack-ctl status，查看管理节点服务是否正常运行，UI服务是否正常运行。

备管理节点需要维护

双管理节点高可用场景下，假定管理节点1为主管理节点，管理节点2为备管理节点。

若管理员需要临时关闭管理节点2进行维护。

1. 关闭管理节点2。
   - 若管理节点2没有被复用为计算节点添加到ZStack for Alibaba Cloud：
     1. 对管理节点2执行`zsha2 stop-node`，关闭**zsha2**相关服务；
     2. 对管理节点2进行shutdown关机操作；
     3. 对管理节点2下电后进行维护。
   - 若管理节点2被复用为计算节点，并已添加到ZStack for Alibaba Cloud：
     1. 对管理节点2执行`zsha2 stop-node`，关闭**zsha2**相关服务；
     2. 将管理节点2进入维护模式；
     3. 对管理节点2进行shutdown关机操作；
     4. 对管理节点2下电后进行维护。
2. 启动管理节点2。
   a. 对管理节点2通电后，通过手动或IPMI启动服务器；
   b. 等待管理节点2启动，成功引导操作系统；
   c. 对管理节点2执行`zsha2 start-node`，启动**zsha2**相关服务；
   d. 对管理节点2执行`zsha2 status`，查看**zsha2**服务是否正常运行；
   e. 对管理节点2执行`zstack-ctl status`，查看管理节点服务是否正常运行，UI服务是否正常运行。

2.1.2 双管理节点需要维护

双管理节点高可用场景下，假定管理节点1为主管理节点，管理节点2为备管理节点。

若管理员需要临时关闭两个管理节点进行维护。

1. 对两个管理节点执行`zsha2 stop-node`，关闭**zsha2**相关服务；
2. 对两个管理节点进行shutdown关机操作；
3. 对两个管理节点下电后进行维护；
4. 对两个管理节点通电后，通过手动或IPMI启动服务器；
5. 等待两个管理节点启动，成功引导操作系统；
6. 对两个管理节点执行`zsha2 start-node`，启动**zsha2**相关服务；

7. 对两个管理节点执行zsha2 status，查看**zsha2**服务是否正常运行；
8. 对两个管理节点执行zstack-ctl status，查看管理节点服务是否正常运行，UI服务是否正常运行。

2.2 节点修复

2.2.1 单管理节点故障修复

双管理节点高可用场景下，若其中某个管理节点损坏后需要执行修复。

1. 对故障节点执行zsha2 stop-node，关闭**zsha2**相关服务；
2. 尝试恢复故障节点， 如果不能恢复，需使用相同版本的ZStack for Alibaba Cloud定制版ISO修复原节点或安装新节点。
3. 以安装新节点为例：
   a. 调配备用服务器，使得硬件规格与故障节点相近；
   b. 安装基础操作系统，安装完成后，配置root的密码和网络信息与故障节点一致，详情可参考安装与部署章节；
   c. 对置换节点安装高可用套件，详情可参考安装与部署章节；
   d. 对置换节点执行zsha2 status，查看**zsha2**服务是否正常运行；
   e. 对置换节点执行zstack-ctl status，查看管理节点服务是否正常运行，UI服务是否正常运行。

2.2.2 双管理节点故障无法修复

双管理节点高可用场景下，若两个管理节点均损坏无法修复。

3 命令行使用手册

3.1 简介

zsha2是ZStack for Alibaba Cloud针对多管理节点物理机高可用场景设计的命令，帮助用户快速完成该场景下的多种操作。

zsha2下有多条子命令，本手册将对**zsha2**每条子命令的作用和使用方法进行说明。

3.2 -h 帮助内容

描述

显示帮助，可查看**zsha2**全部子命令。

使用方法

```
[root@localhost ~]# zsha2 -h
usage:
  zsha2 [ global options ] command [ command options ]

Global options:
 -h,--help           Display this message

Commands:
 install-ha      install two-node HA environement
 stop-node       stop zstack service in HA environement
 start-node      start zstack service in HA environement
 upgrade-mn      upgrade the MN in HA environment
 upgrade-ha      upgrade the HA suites
 demote          demote current node as backup
 status          show HA status
 show-config     show HA configuration
 collect-log     collect HA related log files
 help            show this help message
```

3.3 version 版本信息

描述

查看版本信息，包括版本号和Commit ID。

使用方法

```
[root@localhost ~]# zsha2 version
```

```
version 3.0.1.0, commit 5ecf6c4a6d6ddca22d9c652494e6a74d46920737
```

3.4 install-ha 安装命令

描述

安装命令。假定用户已安装两个ZStack for Alibaba Cloud管理节点，对主管理节点执行**zsha2**安装命令，即可切换到双管理节点高可用模式。

使用方法

参数	介绍	示例
-nic	物理设备名，用于配置VIP，生产环境一般是一个管理网络的网桥	`./zsha2 install-ha -nic br_bond0`
-gateway	主备管理节点的仲裁网关	`./zsha2 install-ha -gateway 192.168.0.1`
-slave	指定备管理节点 **说明**： 安装过程中，备管理节点的数据库会被主管理节点的数据库覆盖，请谨慎配置。	`./zsha2 install-ha -slave "root:password@192.168.196.125"`
-vip	指定Keep Alived通信的VIP	`./zsha2 install-ha -vip 192.168.199.151`
-db-root-pw	主备管理节点的数据库root密码（必须相同）	`./zsha2 install-ha -db-root-pw zstack.mysql.password`
-time-server	可选参数，指定时间同步服务器，用于统一时间同步	`./zsha2 install-ha -time-server 192.168.196.125`
-cidr	可选参数，指定网络段，需覆盖主备管理节点IP、VIP和网关。 **说明**： 如果不指定，系统会自动计算出一个最小网络段，可能无法满足需求，推荐指定网络段。	`./zsha2 install-ha -cidr 192.168.0.0/16`

参数	介绍	示例
-force	可选参数，当主备管理节点的数据库始终无法完成自动同步，对主管理节点强制执行**zsha2**安装命令 **说明**： 执行强制安装前，建议对两个数据库进行备份。	./zsha2 install-ha -force
-repo	可选参数，指定Yum源，默认为本地源	./zsha2 install-ha -repo zstack-local
-timeout	可选参数，主备管理节点的数据库初始化复制超时时间，默认值为600，单位为秒	./zsha2 install-ha -timeout 600
-yes	可选参数，所有设置均允许	./zsha2 install-ha -yes

```
[root@localhost ~]# ./zsha2 install-ha -nic br_bond0 -gateway 192.168.0.1 -slave "root:
password@192.168.196.125" \
-vip 192.168.199.151 -db-root-pw zstack.mysql.password -yes
Master IPv4 address:  192.168.195.200
ZStack version @ 192.168.195.200: 2.6.0
ZStack version @ 192.168.196.125: 2.6.0
Calculated CIDR: 192.168.0.0/16

Start installation ...


x checking network interface and gateway ...
✓ Task 1: checking network interface and gateway ... completed.
x prepare HA-services ...
✓ Task 2: prepare HA-services ... completed.
+ setting up DB config before replication ...
✓ Task 3: setting up DB config before replication ... completed.
x creating DB user for replication ...
✓ Task 4: creating DB user for replication ... completed.
+ update iptables rules ...
✓ Task 5: update iptables rules ... completed.
+ starting the initial replication ...
*************************** 1. row ***************************
            File: mysql-bin.000002
        Position: 1844
    Binlog_Do_DB:
Binlog_Ignore_DB:

+ starting the initial replication ...
✓ Task 6: starting the initial replication ... completed.
x wait peer slave sync status ...

 Slave_IO_Running: Yes
Slave_SQL_Running: Yes
```

```
    Last_IO_Error:
   Last_SQL_Error:
      Last_Error:
      Last_Errno: 0

✓ Task 7: wait peer slave sync status ... completed.
+ wait local DB sync status ...
*************************** 1. row ***************************
           File: mysql-bin.000002
       Position: 245
    Binlog_Do_DB:
Binlog_Ignore_DB:

x wait local DB sync status ...

 Slave_IO_Running: Yes
Slave_SQL_Running: Yes
    Last_IO_Error:
   Last_SQL_Error:
      Last_Error:
      Last_Errno: 0

✓ Task 8: wait local DB sync status ... completed.
+ setting up keepalived ...
✓ Task 9: setting up keepalived ... completed.
x check slave virtual IP settings ...
✓ Task 10: check slave virtual IP settings ... completed.
x configuring ZStack servers ...
✓ Task 11: configuring ZStack servers ... completed.
x installing HA scripts ...
✓ Task 12: installing HA scripts ... completed.
x starting ZStack HA service ...
✓ Task 13: starting ZStack HA service ... completed.
x waiting management node up and running ...
✓ Task 14: waiting management node up and running ... completed.

OK, installation completed.

Hints:
- Stop server with:   zsha2 stop-node,
- Start server with:  zsha2 start-node,
- Get HA status with: zsha2 status -peer 192.168.196.125

Please also setup SSH pubkey-login between 192.168.195.200 and 192.168.196.125
```

3.5 stop-node 关闭管理节点

描述

在双管理节点高可用场景下，关闭其中一个管理节点，同时关闭所有**zsha2**服务。

使用方法

```
[root@localhost ~]# zsha2 stop-node
stopping zstack-ha service ...
stopping zstack management node ...
```

```
stopping keepalived ...
```

3.6 start-node 启动管理节点

描述

在双管理节点高可用场景下，将处于停止状态的管理节点启动，同时启动所有**zsha2**服务。

使用方法

```
[root@localhost ~]# zsha2 start-node
starting keepalived ...
starting zstack-ha service ...
starting zstack management node ...
```

3.7 upgrade-mn 升级管理节点

描述

在双管理节点高可用场景下，仅升级两个管理节点。

使用方法

参数	介绍	示例
-force	可选参数，强制升级管理节点	./zsha2 upgrade-mn -force ZStack_Alibaba_Cloud-installer-3.0.1.bin
-peerpass	可选参数，输入Peer管理节点SSH登录密码	./zsha2 upgrade-mn -peerpass password ZStack_Alibaba_Cloud-installer-3.0.1.bin
-yes	可选参数，所有设置均允许	./zsha2 upgrade-mn -yes

```
[root@localhost ~]# ./zsha2 upgrade-mn -peerpass password ZStack_Alibaba_Cloud-installer-3
.0.1.bin
```

3.8 upgrade-ha 升级高可用套件

描述

在双管理节点高可用场景下，升级当前的**zsha2**服务。

使用方法

```
[root@localhost ~]# ./zsha2 upgrade-ha

Start upgrading ...
```

```
+ Stopping HA-services ...
✓ Task 1: Stopping HA-services ... completed.
+ Upgrading HA suites ...
✓ Task 2: Upgrading HA suites ... completed.
x starting ZStack HA service ...
✓ Task 3: starting ZStack HA service ... completed.

OK, upgrade HA completed.

Hints:
- Stop server with: zsha2 stop-node,
- Start server with: zsha2 start-node,
- Get HA status with: zsha2 status -peer 192.168.196.125
```

3.9 demote 主备切换

描述

在双管理节点高可用场景下，将主管理节点在线切换为备管理节点。

使用方法

```
[root@localhost ~]# zsha2 demote
```

3.10 status 状态信息

描述

在双管理节点高可用场景下，显示当前管理节点的状态，包括是否已获取VIP、自身可达性、网关可达性、VIP可达性、Peer管理节点可达性、Keep Alived服务状态、高可用监控服务状态、数据库状态、管理节点状态、UI状态、Slave状态，以及如何查看Peer管理节点状态的说明。

使用方法

```
[root@localhost ~]# zsha2 status
Status report from 192.168.195.200
================================
Owns virtual address:           yes
Self 192.168.195.200 reachable:   yes
Gateway 192.168.0.1 reachable:    yes
VIP 192.168.199.151 reachable:    yes
Peer 192.168.196.125 reachable:   yes
Keepalived status:              active
ZStack HA Monitor:              active
MySQL status:                   mysqld is alive
MN status: Running [PID:6500]
UI status: Running [PID:9785] http://192.168.195.200:5000

Slave Status:
-------------
            Slave_IO_Running: Yes
           Slave_SQL_Running: Yes
                 Last_Error:
```

```
        Seconds_Behind_Master: 0
               Last_IO_Error:
              Last_SQL_Error:

pass '-peer user:pass@host[:port]' to show peer status
or, '-peer host' if SSH pubkey-login has been enabled.
```

3.11 show-config 显示配置

描述

在双管理节点高可用场景下，显示当前环境的配置信息。

使用方法

```
[root@localhost ~]# zsha2 show-config
{
  "nodeip": "192.168.195.200",
  "peerip": "192.168.196.125",
  "dbvip": "192.168.199.151",
  "nic": "br_bond0",
  "gw": "192.168.0.1",
  "dbnetwork": "192.168.0.0/16",
  "repo": "zstack-local",
  "version": 0
}
```

3.12 collect-log 收集日志

描述

在双管理节点高可用场景下，收集**zsha2**服务相关日志。

使用方法

```
[root@localhost ~]# zsha2 collect-log
Collecting logs ...
Collected log: zsha2-log-2018-09-17T154358+0800.tgz

# 将日志压缩包解压
[root@localhost ~]# tar zxvf zsha2-log-2018-09-17T154358+0800.tgz
tmp/zsha2-log588815976/
tmp/zsha2-log588815976/zsha2-status.log
tmp/zsha2-log588815976/zstack-ha.log
tmp/zsha2-log588815976/keepalived.data
tmp/zsha2-log588815976/zs-vip-192.168.199.151.log
tmp/zsha2-log588815976/keepalived_status.log
```

专有云术语表

区域（Zone）

ZStack中最大的一个资源定义，包括集群、二层网络、主存储等资源。

集群（Cluster）

一个集群是类似物理主机（Host）组成的逻辑组。在同一个集群中的物理主机必须安装相同的操作系统（虚拟机管理程序，Hypervisor），拥有相同的二层网络连接，可以访问相同的主存储。在实际的数据中心，一个集群通常对应一个机架（Rack）。

管理节点（Management Node）

安装系统的物理主机，提供UI管理、云平台部署功能。

计算节点（Compute Node）

也称之为物理主机（或物理机），为云主机实例提供计算、网络、存储等资源的物理主机。

主存储（Primary Storage）

用于存储云主机磁盘文件的存储服务器。支持本地存储、NFS、 Ceph、Shared Mount Point、Shared Block等类型。

镜像服务器（Backup Storage）

也称之为备份存储服务器，主要用于保存镜像模板文件。建议单独部署镜像服务器。

镜像仓库（Image Store）

镜像服务器的一种类型，可以为正在运行的云主机快速创建镜像，高效管理云主机镜像的版本变迁以及发布，实现快速上传、下载镜像，镜像快照，以及导出镜像的操作。

云主机（VM Instance）

运行在物理机上的虚拟机实例，具有独立的IP地址，可以访问公共网络，运行应用服务。

镜像（Image）

云主机或云盘使用的镜像模板文件，镜像模板包括系统云盘镜像和数据云盘镜像。

云盘（Volume）

云主机的数据盘，给云主机提供额外的存储空间，共享云盘可挂载到一个或多个云主机共同使用。

计算规格（Instance Offering）

启动云主机涉及到的CPU数量、内存、网络设置等规格定义。

云盘规格（Disk Offering）

创建云盘容量大小的规格定义。

二层网络（L2 Network）

二层网络对应于一个二层广播域，进行二层相关的隔离。一般用物理网络的设备名称标识。

三层网络（L3 Network）

云主机使用的网络配置，包括IP地址范围、网关、DNS等。

公有网络（Public Network）

由因特网信息中心分配的公有IP地址或者可以连接到外部互联网的IP地址。

私有网络（Private Network）

云主机连接和使用的内部网络。

L2NoVlanNetwork

物理主机的网络连接不采用Vlan设置。

L2VlanNetwork

物理主机节点的网络连接采用Vlan设置，Vlan需要在交换机端提前进行设置。

VXLAN网络池（VXLAN Network Pool）

VXLAN网络中的 Underlay 网络，一个 VXLAN 网络池可以创建多个 VXLAN Overlay 网络（即 VXLAN 网络），这些 Overlay 网络运行在同一组 Underlay 网络设施上。

VXLAN网络（VXLAN）

使用 VXLAN 协议封装的二层网络，单个 VXLAN 网络需从属于一个大的 VXLAN 网络池，不同 VXLAN 网络间相互二层隔离。

云路由（vRouter）

云路由通过定制的Linux云主机来实现的多种网络服务。

安全组（Security Group）

针对云主机进行第三层网络的防火墙控制，对IP地址、网络包类型或网络包流向等可以设置不同的安全规则。

弹性IP（EIP）

公有网络接入到私有网络的IP地址。

快照（Snapshot）

某一个时间点上某一个磁盘的数据备份。包括自动快照和手动快照两种类型。

混合云术语表

访问密钥（AccessKey）

用于调用阿里云API或大河云联API的唯一凭证，AccessKey包括AccessKeyID（用于标识用户）和AccessKeySecret（用于验证用户密钥）。

数据中心（Data Center）

包含阿里云的地域和可用区等地域资源，用于匹配阿里云资源的地域属性。

地域（Region）

物理的数据中心，划分地区的基本单位，ZStack混合云的地域对应了阿里云端的地域。

可用区（Identity Zone）

在同一地域内，电力和网络互相独立的物理区域，ZStack混合云的可用区对应了阿里云端的可用区（Zone）。

存储空间（Bucket）

用于存储对象（Object）的容器，ZStack使用对象存储（OSS）里的Bucket来上传镜像文件。

ECS云主机（Elastic Compute Service）

阿里云端创建的ECS实例，可在ZStack混合云界面进行ECS云主机生命周期的管理。

专有网络VPC（Virtual Private Cloud）

用户基于阿里云构建的一个隔离的网络环境，不同的专有网络之间逻辑上彻底隔离。

虚拟交换机（VSwitch）

组成专有网络VPC的基础网络设备，可以连接不同的云产品实例。ZStack混合云的虚拟交换机对应了阿里云VPC下的虚拟交换机。

虚拟路由器（VRouter）

专有网络VPC的枢纽，可以连接专有网络的各个虚拟交换机，同时也是连接专有网络与其它网络的网关设备。ZStack支持查看VPC下的虚拟路由器。

路由表（Route Table）

虚拟路由器上管理路由条目的列表。

路由条目（Route Entry）

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。ZStack支持自定义类型的路由条目。

安全组（Security Group）

针对云主机进行第三层网络的防火墙控制。ZStack混合云的安全组对应了阿里云端ECS云主机三层隔离的防火墙约束。

镜像（Image）

云主机使用的镜像模板文件，一般包括操作系统和预装的软件。ZStack支持上传本地镜像到阿里云，以及使用阿里云端镜像。

弹性公网IP（EIP）

阿里云端公有网络池中的IP地址，绑定弹性公网IP的ECS实例可以直接使用该IP进行公网通信。

VPN连接（VPN Connection）

通过建立点对点的IPsec VPN通道，实现企业本地数据中心的私有网络与阿里云端VPN网络进行通信。

VPN网关（VPN Gateway）

一款基于Internet，通过加密通道将本地数据中心和阿里云专有网络VPC安全可靠连接起来的服务。用户在阿里云VPC创建的IPsec VPN网关，与本地数据中心的用户网关配合使用。

VPN用户网关（Customer Gateway）

本地数据中心的VPN服务网关。可通过ZStack混合云创建VPN用户网关，并将VPN用户网关与VPN网关连接起来。

高速通道（Express Connect）

通过物理专线（即租用运营商的专线：电缆或光纤），连通本地数据中心到阿里云专线接入点，与阿里云VPC环境打通，实现云上云下不同网络间高速，稳定，安全的私网通信。

边界路由器（VBR）

用户申请的物理专线接入交换机的产品映射。用户在物理专线上可以创建边界路由器，边界路由器负责专线上的数据在阿里云上进行转发。通过边界路由器，用户数据可以直达阿里云VPC网络。

路由器接口（Router Interface）

一种虚拟的网络设备，可以挂载在路由器并与其他路由器接口进行高速通道互联，实现不同网络间的内网互通。