

UniCloud Usphere 云计算管理平台

双机热备配置指导

资料版本：5W100-20220310

© 紫光云技术有限公司 2022 版权所有，保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

对于本手册中出现的其它公司的商标、产品标识及商品名称，由各自权利人拥有。

由于产品版本升级或其他原因，本手册内容有可能变更。紫光云保留在没有任何通知或者提示的情况下对本手册的内容进行修改的权利。本手册仅作为使用指导，紫光云尽全力在本手册中提供准确的信息，但是紫光云并不确保手册内容完全没有错误，本手册中的所有陈述、信息和建议也不构成任何明示或暗示的担保。

目 录

1 简介	1
1.1 原理介绍	1
1.2 组网	1
2 注意事项	2
3 配置前提	3
3.1 搭建双机热备的主机的要求	3
3.2 同步模板数据的方式	3
3.3 热备主机作为 VKS 主机的使用说明	4
4 配置步骤	4
4.1 系统安装要求	4
4.2 双机热备配置	5
4.2.1 搭建场景组合	5
4.2.2 双机热备配置步骤	5
4.3 双机热备产品注册	9
5 常用操作	9
5.1 单机故障后的恢复指导	9
5.1.1 适用范围	9
5.1.2 系统要求	9
5.1.3 故障恢复实现	9
5.1.4 重新注册	11
5.2 双机热备服务启动/关闭控制	11
5.2.1 脚本实现	11
5.2.2 脚本执行举例	12
5.3 常用命令	12
5.3.1 查询双机热备主机状态	12
5.3.2 drbd 脑裂判断	13
5.3.3 删除本地同步分区	13
6 缩略语	13

1 简介

实际生产环境中服务器故障是不可避免的。服务器故障可能由各种原因引起，如设备故障、操作系统故障、软件系统故障等等。通常服务器产生故障正常的恢复的时间可能需要几分钟，甚至几小时。而对于一些重要系统而言，用户是很难忍受这样长时间的服务中断的。因此，就需要通过双机热备，来避免长时间的服务中断，保证系统长期、可靠的服务。

UniCloud Usphere VMS 双机热备使用两台服务器部署 VMS，同时两台服务器互相备份，共同提供服务。当一台服务器出现故障时，可由另一台服务器承担服务，从而在不需要人工干预的情况下，自动保证 VMS 管理平台能持续提供服务。VMS 双机热备由备用的服务器解决了在主服务器故障时管理平台中断的问题。

管理平台可以通过双机热备技术来实现管理平台的高可靠性，与 VMS 配置备份相比，VMS 双机热备是一种实时的高可靠性方案。



注意

- 通过分析系统的重要性以及对服务中断的容忍程度，以此决定是否使用双机热备。即业务能否容忍中断，可容忍多长的时间恢复，或者业务不能恢复会造成多大的影响。
 - 双机热备会有一个短暂的切换过程。在切换过程中，VMS 管理平台服务有可能出现短暂的时间中断。当切换完成后，服务将正常恢复。
-

1.1 原理介绍

双机热备功能是基于数据同步复制的方式来实现的。主备服务器之间数据的同步是采用 DRBD (Distributed Replicated Block Device, 分布式镜像块设备内容的存储复制) 存储复制解决方案来实现，当主服务器数据发生变化时，该数据变化会实时同步到备用服务器，这样就保证了主备服务器之间数据的一致性。

双机热备的主备切换是通过 CMSD (热备服务管理程序) 来实现的。

CMSD 主程序完成以下功能：

- 热备服务的启动、关闭、切换等功能。
- 网络故障或主机故障时，热备服务能够自动进行切换。

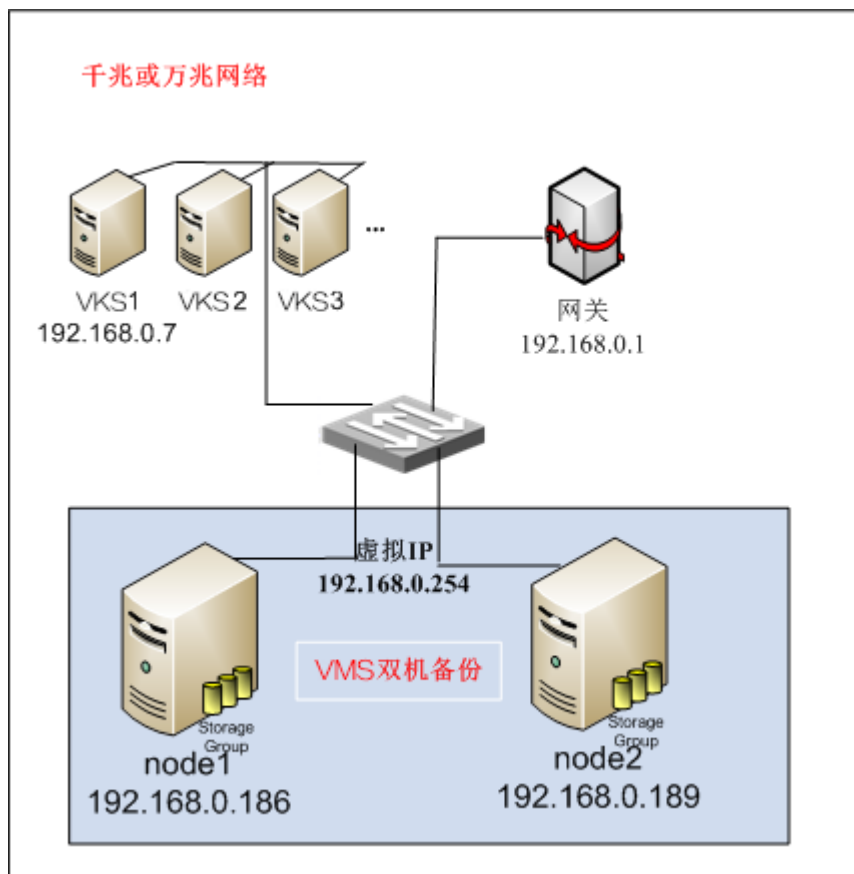
1.2 组网

双机热备系统由三个实体组成：主节点、备节点和仲裁节点。

- 主备节点是安装了 CMSD 主程序的两个节点。
- 仲裁节点分为高级仲裁节点和简易 ping 仲裁节点。
 - 高级仲裁节点可以是同管理平台 VMS 安装了 CMSD 程序的节点。
 - 简易 ping 仲裁节点可以是主备节点可 ping 通的交换机、路由器。

双机热备的组网拓扑图如下所示。

图1-1 组网拓扑图



- node1 和 node2 是构成双机热备的两台 VMS 主机。
- 多台 VKS 主机和 node1、node2 主备 VMS 管理平台组成 Usphere 云服务平台。

 注意

- VKS 主机与主备 VMS 管理平台之间的交换网络要求至少是千兆网络。
- 双机热备的两台主机 (node1、node2) 可以再作为 VKS 主机加入到热备主机自己管理的主机池中使用，但不能再分离成单独的 VMS 管理平台使用。

2 注意事项

- 仅支持新部署的 Usphere 云平台，VMS 管理平台采用双机热备配置。不支持在原有在线使用的 Usphere 云平台中的 VMS 管理平台升级成双机热备，因为原有数据无法保留。
- 搭建双机热备的两台主机名称建议使用字母、数字等简单的命名方式，不建议使用点“.”等特殊字符，否则可能会导致搭建双机热备失败。
- 服务器安装完 Usphere 后，搭建双机热备之前，若需要修改主机名称，可以在服务器的控制台进行修改；双机热备搭建完成后，请勿修改 VMS 主机名。
- Usphere 仅支持在 root 模式下搭建双机热备。

- 双机热备主机因为网络异常出现脑裂，等网络恢复正常后，会自动协商出主服务器，若脑裂时 VKS 主机的性能数据发送的是现在的备服务器，脑裂恢复后，VKS 主机的性能数据仍向现在的备服务器发送性能数据，导致主管理平台性能数据丢失，此时在主管理平台上需执行连接主机操作，保证 VKS 主机性能数据发送正确。
- 双机热备部署或增加本地同步分区盘的过程中，要保证服务器健壮性，不能重启和关闭服务器。若出现故障服务器重启或断电，需要重新部署搭建。
- 如果服务器使用了软 raid 构建的硬盘、NVMe 硬盘或 VROC 构建的硬盘作为系统盘，由于其盘符不是 sdX 的格式，因此不支持搭建双机热备。

3 配置前提

3.1 搭建双机热备的主机的要求

搭建双机热备的两台 VMS 主机需同时满足以下条件：

- 由于双机热备的数据库分区创建在/vms 分区下，并且/vms 分区自身需要一定量的空间，因此安装 VMS 进行手动分区时，至少要为/vms 分区分配 30G 的空间。/vms 分区最小空间估算方法为：数据库分区（预估主机个数*10M+预估虚拟机个数*15M）*15/1024M +10 G。建议在保证其他分区空间充足的前提下，为/vms 分区分配尽量大的空间。
- 为了缩短双机热备的搭建时间，当设置数据库分区大小超过 200G 时，建议两台 VMS 之间的管理网使用万兆光纤。
- 全新安装的系统，没有作为 VKS 主机使用过，管理平台中无纳管的 VKS 主机。
- 主机上的存储池必须保持系统默认配置，即只有 isopool 和 defaultpool 两个存储池。
- 安装的 Usphere 软件版本号必须相同，且安装的 Usphere 组件必须相同。
- 安装系统时，主机的系统引导方式必须相同。
- 两台主机之间的管理网络必须相通。

搭建双机热备的仲裁节点需满足以下条件：

- 高级仲裁节点的系统需要和搭建双机的两台主机一致。
- 推荐用同一管理平台中的 VKS 主机作为高级仲裁。

3.2 同步模板数据的方式

VMS 双机热备推荐两种方式同步虚拟机模板数据：

- FC 或 iSCSI 共享模板存储方式。建议使用 FC 或 iSCSI 共享模板存储方式同步虚拟机模板数据。
- 本地同步分区方式。



说明

- 若有 FC 或 iSCSI 类型的共享存储，优先推荐使用共享存储。
 - 若既要使用同步分区又要使用 FC/iSCSI 类型的共享存储，请先增加本地同步分区类型的模板存储，再挂载 FC/iSCSI 存储设备上的 LUN。
-

当双机热备的两台主机无法使用共享存储 LUN 时，可以使用本地同步分区，用于同步主备管理平台之间的虚拟机模板数据。比如使用虚拟机作为 VMS 双机节点，主备管理平台无法使用 FC 存储设备上的 LUN，又无法添加 iSCSI 存储设备上的 LUN 时，可以配置本地同步分区。

双机热备搭建成功后，在 VMS 管理平台中，进入[云资源/虚拟机模板/模板存储]页面单击<增加模板存储>按钮或者进入[系统管理/双机热备配置/模板存储]页面单击<增加>按钮，可增加一个本地同步分区类型的模板存储。



只能增加一个本地同步分区类型的模板存储。

由于涉及到格式化，须使用本地磁盘作为新的同步分区。创建本地同步分区类型的模板存储时，主备节点存在如下限制条件：

- 主、备节点不能作为 VKS 主机加入到管理平台中。
- 主、备节点上不能添加除默认存储池外的其他存储池。
- 主、备节点上不能发现和挂载 FC 和 iSCSI 存储设备上的 LUN（此种情况下建议拔除 cable 线后再重新操作）。
- 主、备节点的本地磁盘要求大小一致，其识别的设备名称相同(即都为/dev/sd*)。

只有当上述所有条件都满足要求时，才能成功创建本地同步分区类型的模板存储。

3.3 热备主机作为VKS主机的使用说明

- 双机热备（CMSD）的主备两台主机可以作为 VKS 主机，允许加入到热备主机自己管理的主机池中使用。
- 双机热备的主备两台主机不能作为业务主机加入到另外一个管理平台中。
- 双机热备（CMSD）创建本地同步分区后，双机热备主机不能再作为 VKS 主机使用。
- 双机热备主机作为 VKS 主机使用后不能再增加本地同步分区，需要在管理平台中删除作为 VKS 主机的热备主机后，才可以添加。

4 配置步骤

4.1 系统安装要求

- 两台服务器安装的 Usphere 软件版本号必须相同，且安装的 Usphere 组件（VMS、ISC、SSP）必须相同。
- 两台服务器采用相同的引导方式进行系统安装：即都为 Legacy 方式或者都为 UEFI 方式，二者不可混用。
- 两台服务器的分区方式及分区大小要求，请参考《UniCloud Usphere 安装指导》。
- 要求主备两台服务器的时间保持一致。如果不一致请手动同步服务器时间。比如可以使用 `date -s '2013-6-14 8:30:00'` 命令手动配置时间，也可以配置 NTP 服务器来同步时间。



注意

请注意在系统时间修改后，需要执行命令 `hwclock -w` 命令将时间信息写入 BIOS 保存，以免主机重启后时间发生变化。

- 系统安装完成后，可以通过主机 Xconsole 控制台来修改主机 `hostname` 和 IP 地址。
-



注意

启用双机热备功能后，请不要再修改主机名称和 IP 地址，否则会出现严重问题。

- 两台服务器系统盘分区编号和挂载点需保持一致，可通过 `lsblk` 命令查看系统盘分区信息。

4.2 双机热备配置

只需在一个 VMS 管理平台中进行部署，即可实现双机备份功能。搭建过程中，因为需要初始化本地及对端硬盘分区，花费时间会比较长，请耐心等待，具体花费时间长短与实际环境中的硬盘大小及网络速度有关。一般来讲，千兆网络环境下，每 100G 硬盘大约花费 20 分钟，而在万兆网络环境中，花费时间会适当减少。

4.2.1 搭建场景组合

- 服务器+高级仲裁
 - 服务器+简易 Ping 方式
 - 虚拟机+高级仲裁
 - 虚拟机+简易 Ping 方式
-



说明

按照推荐优先级从高到低排列。

当主备两节点网络出现故障，但两节点对外通信正常，仲裁机制介入。

- 高级仲裁主机：安装了 CMSD 程序且担任仲裁角色的节点。通常为指定一台已安装 VKS 组件的主机作为仲裁主机，当主备 VMS 管理平台之间通信异常时，主备节点同时向仲裁主机发送本机状态信息，仲裁主机根据接收到的信息进行仲裁，将重新确定的主备角色信息回送给两节点。
- 简易 ping 方式：两个正在使用且比较稳定可达的 IP 地址，一般为网关或其他能够常态可达的 IP 地址，用于 ping 测试。

4.2.2 双机热备配置步骤

- (1) 单击顶部“系统”页签，选择左侧导航树[双机热备配置]菜单项，进入双机热备配置页面。
- (2) 单击<热备搭建>按钮，弹出热备搭建对话框，配置各项参数。

图4-1 热备搭建对话框

热备搭建

双机热备搭建完成后,如需创建模板存储,建议使用共享模板存储,若无共享模板存储,可创建本地同步分区,将本地同步分区作为模板存储。

* 虚IP地址: 172.16.60.55

* 子网掩码/前缀: 255.255.0.0

* 备机IP地址: 172.16.60.52

* 备机root密码:

仲裁方式: **高级仲裁** | 简易Ping方式

* 仲裁主机IP地址: 172.16.60.53

* 仲裁主机root密码:

* 预估主机个数: 4

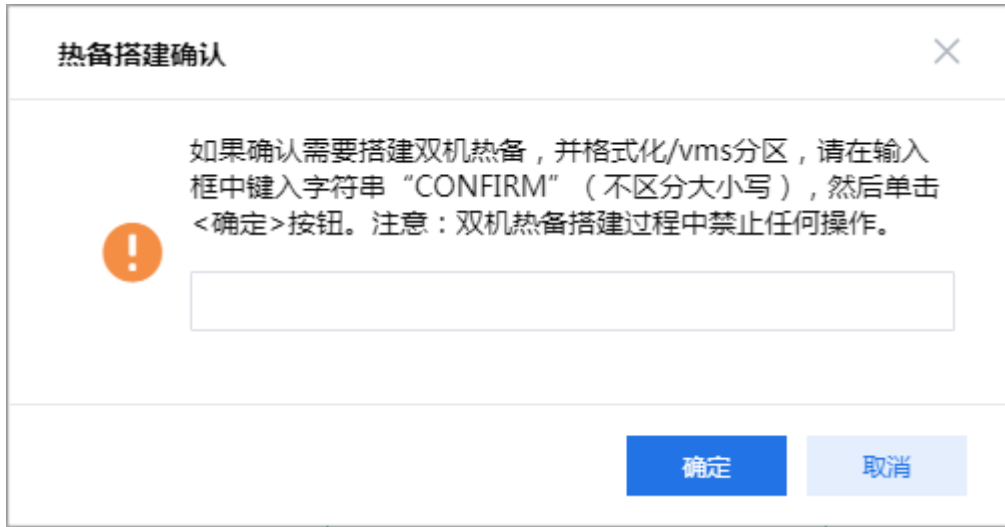
* 预估虚拟机个数: 20

* 数据库分区大小(GB): 20

确定 取消

- (3) 输入虚 IP 地址、子网掩码、备机 IP 地址、备机 root 密码、选择仲裁方式并根据选择输入仲裁主机 IP 地址、仲裁主机 root 密码或者仲裁 IP 地址 1、仲裁 IP 地址 2、输入预估主机个数、预估虚拟机个数、数据库分区大小, 单击<确定>按钮。
- (4) 在弹出的热备搭建确认对话框中输入参数 CONFIRM, 单击<确定>按钮。

图4-2 热备搭建确认对话框



若预检查出错，会提示相应的错误信息，修改配置重新搭建即可。详细的错误信息将保存在/var/log/cvm_master_slave.log 日志文件中。

- (5) 开始搭建双机热备环境时，管理平台将跳转到双机热备配置页面，展示搭建进度。

图4-3 双机热备配置



- (6) 双机热备搭建成功，跳转到登录页面。

如果双机热备搭建失败，搭建页面中途会中止，并且页面下方提示“配置失败，请尝试重新登录并配置”，详细的错误信息将保存在/var/log/cvm_master.log 和 /var/log/cvm_master_slave.log 日志文件中。根据失败所在的阶段和日志可以找到对应的原因，修改配置，重新搭建。

(7) 增加模板存储。

热备搭建成功后，第一次登录 VMS 管理平台时，会自动跳转到模板存储配置界面。如果要正常使用 VMS 双机热备功能，需要配置模板存储。

- 单击<现在创建>按钮，创建模板存储。推荐使用 FC 共享目录或 iSCSI 共享目录类型的模板存储。

图4-4 增加模板存储

- 若单击“暂不创建”，后续需要增加模板存储，可以在[系统管理/双机热备配置/模板存储]中增加。

(8) 配置 VNC 代理服务器（可选）。

如果虚拟机登录时需要使用 VNC 代理，则需要登录管理平台，在[系统管理]中配置 VNC 代理服务器的信息。

图4-5 VNC 代理服务器参数配置

 注意

- 如果提供 VMS 服务的主机本身配置了 VNC 代理服务器，则要求主备两台 VMS 主机必须都配置 VNC 代理服务器，且其 IP 参数必须设置为 VMS 管理平台的虚拟 IP 地址。
- 主备 VMS 后台 root 密码需要一致，避免切换备节点后虚拟机无法正常连接 VNC 控制台。

4.3 双机热备产品注册

双机热备产品的注册，需采取 License Server 授权方式，申请授权时，只需要在主 VMS 上申请即可。授权的具体操作，请联系技术支持。

5 常用操作

5.1 单机故障后的恢复指导

5.1.1 适用范围

已经部署双机热备的 Usphere 云平台，其中一台热备服务器发生硬件故障，需重新安装系统或更换服务器时，必须手动进行双机热备功能的恢复。

其他诸如网络不可达、断电等未重装系统的故障，不需要手动进行恢复，故障解决后双机热备功能将自动恢复。

5.1.2 系统要求

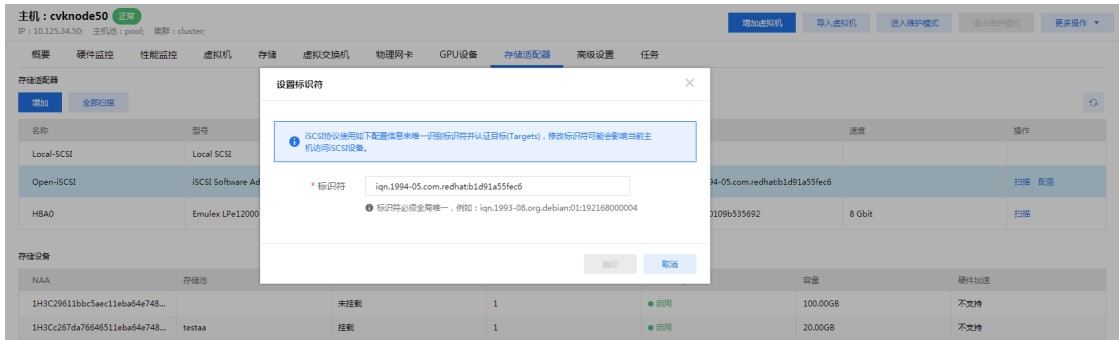
- 热备服务器单机故障时，用于恢复的服务器安装的 Usphere 软件的版本必须与原版本一致。
- 用于恢复的服务器如果是新更换的服务器或原热备服务器更换硬盘，则要求新服务器的硬盘或更换后的硬盘与原服务器的硬盘大小相同。如果为原服务器，只需重新安装原 Usphere 软件。
- 进行单机故障恢复时，主服务器（原双机备份的两台服务器中只剩下主服务器状态正常）必须对外提供 VMS 服务，这样恢复时才能进行原始数据的同步。
- 进行单机故障恢复时，要求新装服务器启用 root 权限。如果主服务器禁用了 root，请先执行命令 `sudo su root`，获取 root 权限后，再进行单机故障恢复。

5.1.3 故障恢复实现

1. 注意事项

如果双机环境增加了 iSCSI 共享存储类型的模板存储，在执行故障恢复前，需要先登录备机的 VMS 界面，修改存储适配器的标识符与之前的标识符一致（若采用新的备机，请修改存储适配器的标识符与故障主机存储适配器的标识符一致）。

若之前没有记录存储适配器的标识符，请咨询技术支持进行找回。



2. 操作步骤

(1) 使用 root 用户登录主服务器，执行 ms_recovery.sh 脚本，根据提示输入 yes。

运行脚本 ms_recovery.sh 示例：

```
root@cvknode1:~# ms_recovery.sh 172.16.60.51 123456 123456
Please confirm to recovery 172.16.60.51 and format peer's /vms partition [yes/no]:
yes
```

-----Ip address Check-----

该脚本共 3 个参数，说明如下。

表5-1 脚本参数说明表

参数序号	参数名称	说明
\$1	peer-node-ip	对端主机管理网络的IP地址，即新装服务器中作为Slave主机的IP地址。
\$2	peer-node-root-password	对端主机root用户密码。
\$3	controller-root-password	仲裁主机root用户密码，仲裁方式是简易Ping方式时，不需要设置该参数。



注意

如果密码有&、#、@等特殊字符，需要加单引号，否则无法正确识别密码，不能正常执行脚本。

例如：若对端主机 root 密码为 123&456，则正确命令为 ms_recovery.sh 172.16.60.51 123'&'456 123456。

运行脚本的过程中，因为需要初始化对端硬盘分区以及同步对端数据，花费时间会比较长，请耐心等待，具体花费时间长短与实际环境中的硬盘大小及网络速度有关。一般来讲，千兆网络环境下，每 100G 硬盘大约花费 20 分钟，而在万兆网络环境中，花费时间会适当减少。



说明

在脚本执行过程中，不能中断程序执行。

(2) 执行结果说明。

○ 执行成功

脚本执行完毕后，出现以下字样，则表示双机热备单机故障恢复成功。

“-----CVM Master/Slave **Restore** Success-----”

○ 执行失败

如果执行失败，则在执行脚本过程中，会出现“**ERROR....**”失败提示信息，且程序将会自动中止执行。

详细的错误信息保存在/var/log/cvm_master.log 和/var/log/cvm_master_slave.log 日志文件中。



注意

恢复完成后再次登录 VMS 管理平台，如果热备管理的非双机热备主机的 VKS 主机状态显示为不正常，重新连接 VKS 主机即可。

5.1.4 重新注册

双机热备单机故障恢复后，产品可能需要重新注册。

- 如果是硬盘故障，更换硬盘恢复成功后，采用原先申请的 License 注册文件在已恢复主机中重新注册即可。即故障恢复成功后，已恢复主机是作为 Slave 主机运行，此时需要进行主备切换，切换后，已恢复主机作为 Master 主机提供 VMS 服务，此时通过虚拟 IP 地址登录 VMS 管理平台进行注册即可。
- 如果是新更换服务器，需要重新申请 License 注册文件，并重新注册，具体请参考 [4.3 双机热备产品注册](#)。
- 其他诸如断电、网络等故障，不必重新进行产品注册。

5.2 双机热备服务启动/关闭控制

ms_resource_stop.sh 和 ms_resource_start.sh 两个脚本的主要作用是临时中止业务，进行主机维护。



注意

通过 ms_resource_stop.sh 命令关闭双机热备服务后，两台热备主机并未分离，不能单独把其中一台主机再作为 VMS 管理平台或 VKS 主机使用。

5.2.1 脚本实现

- ms_resource_stop.sh

关闭双机热备功能。命令执行过程中，命令不能中断（Ctrl +C 中断）；执行命令时，主机不能关机或重启。该脚本没有调用参数。

- **ms_resource_start.sh**

开启双机热备功能。当使用 **ms_resource_stop.sh** 关闭双机热备功能后，可以使用 **ms_resource_start.sh** 重新启用双机热备功能。命令执行过程中，命令不能中断（Ctrl +C 中断），执行命令时，主机不能关机或重启。该脚本有一个可选参数。

5.2.2 脚本执行举例

在两台主机中任意一台执行即可。

以 root 用户登录主机，执行 **ms_resource_stop.sh** 脚本或 **ms_resource_start.sh** 脚本。

```
root@cvknode01:~#ms_resource_stop.sh
或者
root@cvknode01:~#ms_resource_start.sh
```

5.3 常用命令

5.3.1 查询双机热备主机状态

双机热备功能安装后，可以通过 **crm status** 命令查询双机热备主机中各服务的启动状态。

- 双机热备启用成功，双机都正常的状态，如下所示。

```
root@cvknode1:~# crm status
master : cvknode1 (Online)
slave  : cvknode2 (Online)
quorum : [controller] cvknode3 (Online)
-----
webip  : [172.16.60.55] Started cvknode1
Master/Slave Set: [/var/lib/mysql-share]
  Master: [cvknode1]
  Slave  : [cvknode2]
mysql  : Started cvknode1
tomcat : Started cvknode1
-----
```

- 双机热备启用成功，**cvknode1** 主机故障，主备切换后的状态，如下所示。

```
root@cvknode2:~# crm status
master : cvknode2 (Online)
slave  : cvknode1 (Offline)
quorum : [controller] cvknode3 (Online)
-----
webip  : [172.16.60.55] Started cvknode2
Master/Slave Set: [/var/lib/mysql-share]
  Master: [cvknode2]
  Slave  : [Unknown]
mysql  : Started cvknode2
tomcat : Started cvknode2
-----
```

- 双机热备安装或启用失败，主机状态，如下所示。

```
root@cvknode1:~# crm status
cmsadm: failed to connect to cmsd
```

5.3.2 drbd 脑裂判断

双机热备环境中，可以通过 `drbd-overview` 命令查看是否出现 drbd 脑裂。

若有 `StandAlone`，表示已出现 drbd 脑裂。

已出现 drbd 脑裂：

```
root@cvknode1:~# drbd-overview
0:r0/0 StandAlone Secondary/Unknow UpToDate/DUnknown r-----
0:r1/0 Connected Secondary/Primary UpToDate/UpToDate C      r-----
```

5.3.3 删除本地同步分区

同步分区不支持界面删除，也不建议后台删除。但若增加的同步分区因为某些原因必须要删除时，可以在后台主服务器上执行 `ms_rm_sync_partition.sh r1` 命令，确定要删除时输入 `yes`。请慎用。

(1) 删除本地同步分区。

```
root@cvknode1:~# ms_rm_sync_partition.sh r1
Please confirm to delete r1 sync partition [yes/no]:
yes
Success to delete sync r1.
```

(2) 后台查看 `drbd-overview` 输出只有 `r0/0` 行，说明删除成功。

```
root@cvknode2:~# drbd-overview
0:r0/0 Connected Primary/Secondary UpToDate/UpToDate C r----- /var/lib/mysql-
share ext4 19G 91M 18G 1%
```

6 缩略语

缩略语 Abbreviations	中文解释 Chinese explanation	英文全名 Full spelling
VMS	虚拟化管理系统	Virtualization Management System
VKS	虚拟化内核平台	Virtualization Kernel System
DRBD	分布式镜像块设备内容的存储复制解决方案	Distributed Replicated Block Device
CMSD	热备服务管理程序	CVM Master Slave Daemon