

目录

目录	1
安装操作系统	2
购买时安装操作系统	2
重装操作系统	2
设置RAID级别	2
常用的RAID级别	2
购买时设置RAID级别	2
重装系统时设置RAID级别	3
更改网络配置	3
SSH密钥管理	3
创建密钥对	3
查看和修改密钥对	3
删除密钥对	4
镜像使用说明	4
自定义镜像管理	4
创建自定义镜像	4
查看自定义镜像	4
删除自定义镜像	5
安全组	5
创建安全组	5
编辑安全组进站规则/出站规则	5
复制安全组	6
删除安全组	6
操作前提	6
操作步骤	6
管理安全组中的裸金属服务器	6
远程登录	6
操作步骤	6
监控	7
监控指标	7
操作指南	7
自助安装云监控	7
操作背景	7
操作方法	7
1. 下载安装包	7
2. 不同镜像安装方式	7
CentOS安装nvidia-fabricmanager	7
Ubuntu安装nvidia-fabricmanager	8
3. 启动nvidia-fabricmanager服务	8
4. 查看nvidia-fabricmanager服务状态	8
5. 设置nvidia-fabricmanager服务开机自启动	8

安装操作系统

在购买或者重装GPU裸金属服务器时，用户需指定想要安装的操作系统的。

裸金属服务器控制台中提供的可选操作系统，都已经过服务器厂商和金山云的测试认证，保证裸金属服务器可以在数据中心正常运行。

购买时安装操作系统

1. 登录[裸金属服务器控制台](#)，默认进入裸金属服务器 > 裸金属服务器页面。
2. 单击裸金属服务器，进入新建裸金属服务器页面。
3. 在选择镜像步骤中，选择所需镜像。关于创建裸金属服务器的其他步骤说明，请参见[第一步：创建裸金属服务器](#)。

针对GPU裸金属服务器，标准镜像列表中提供了已预先安装CUDA的镜像，可以从列表中选择使用。



重装操作系统

非标准化租赁机型不支持重装系统操作，标准化租赁机型可在线重装系统。

1. 登录[裸金属服务器控制台](#)，默认进入裸金属服务器 > 裸金属服务器页面。
2. 在目标裸金属服务器对应的操作列中，选择更多 > 重装系统。
3. 选择要使用的镜像、Bond选型和RAID类型等，并单击 下一步。



4. 选择SSH密钥，设置管理员密码，并单击确定。

若要创建新的SSH密钥，请参见[SSH密钥管理](#)。



设置RAID级别

独立冗余磁盘阵列（Redundant Arrays of Independent Disks，简称RAID）以提高读写性能以及容错能力为目标，与单块硬盘相比，具有更快的传输速率，可大幅提高存储系统的数据吞吐量。

用户可以像管理独立硬盘或逻辑存储单元一样来管理磁盘阵列，并根据业务场景，在购买时或者重装系统时设置RAID级别。

常用的RAID级别

RAID级别	介绍	需要磁盘数	磁盘空间利用率
单盘RAID0	最早出现的RAID模式，即数据分条技术（Data Striping），将数据分散存储在所有磁盘中，以独立访问方式实现多块磁盘同时读写操作，是所有RAID级别中性能最高的。但是它并不提供冗余或错误修复能力，适用于对数据安全性要求不高的情况。	$n \geq 1$	100%
RAID1	称为磁盘镜像，它将数据完全一致地分别写到【工作磁盘】和【镜像磁盘】，在不影响性能情况下最大限度的保证系统的可靠性和可修复性。当一块硬盘失效时，系统会忽略该硬盘，转而使用剩余的镜像盘读写数据，具备很好的磁盘冗余能力，多用在保存关键性的重要数据的场合。	$2n$	50%
RAID5	数据以块为单位分布到各个硬盘上。RAID 5不对数据进行备份，而是把数据和与对应的奇偶校验信息存储到各个磁盘上，并且数据和校验信息存储在不同磁盘。当一个磁盘数据损坏后，利用剩下的数据和校验信息去恢复数据。RAID5的读出效率很高，写入效率一般，块式的集体访问效率不错。因为奇偶校验码在不同的磁盘上，所以提高了可靠性。	$n \geq 3$	$(n-1) / n \%$
RAID10	RAID 10 结合了RAID 0数据条带分部与RAID 1镜像冗余两者的特性，既能够提供较高的读写速度，又能够使数据得到镜像保护，储存空间利用率为50%，实际应用比较常用。	$n \geq 4$	50%
RAID50	结合了RAID 5冗余校验和RAID 0条带分布的优势，可以在确保数据安全性的同时保障存储I/O的性能。	$2n, n \geq 3$	$(n-2) / n\%$

购买时设置RAID级别

在购买GPU裸金属服务器（GPU Elastic Physical Compute，简称GEPC）时，在“选择镜像”步骤中设置数据盘RAID参数。关于其他步骤说明，请参见[创建并配置裸金属服务器](#)。

重装系统时设置RAID级别

在为GEPC重装系统时，可以设置数据盘RAID参数。但注意此操作会格式化磁盘，需要提前做好数据备份。

有关重装系统的详细步骤，请参见[安装操作系统](#)。

更改网络配置

GPU裸金属服务器使用内网IP与处于同一VPC的其他裸金属服务器或云服务器实现内网互访，产生的流量不收取费用。

内网IP由系统统一分配，用户可以在控制台进行更换网络配置的操作。不支持在操作系统内部自行变更内网IP，否则会导致内网通信失败。

更换网络配置的具体操作如下：

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
 2. 找到目标裸金属服务器，在对应的操作列中，选择“更多 > 更换网络”。
- 注意：** 仅支持为未绑定弹性IP的GPU裸金属服务器更换网络。
3. 选择虚拟私有网络、子网与IP地址，单击“确定”。

SSH密钥管理

GPU裸金属服务器（GPU Elastic Physical Compute，简称GEPC）支持使用SSH密钥对认证登录，用户在创建GEPC时需要指定SSH密钥对，所有已创建的密钥对在“SSH密钥”页面统一管理。

创建密钥对

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 在左侧导航栏中，选择“SSH密钥”。
3. 在SSH密钥对列表页面，单击“创建密钥”。
4. 选取一种方式创建密钥对：
 - “创建新密钥对”：输入密钥对名称，单击“确定”后在弹出框中单击“下载”，保存私钥到本地。私钥只有一次下载机会，后续无法找回，只能创建其他密钥对。

- “使用已有公钥”：输入密钥对名称和公钥，单击“确定”。

查看和修改密钥对

密钥对通过ID/指纹作为唯一标识，用户可以查看密钥对应的公钥内容，或批量修改密钥对名称。

1. 进入“[裸金属服务器 > SSH密钥](#)”页面。
2. 选择想要查看的密钥对，可以看到创建时间、密钥ID和公钥内容。
3. 选择一个或多个想要修改名称的密钥对，单击“修改”。
4. 输入新的密钥对名称，单击“确定”。

密钥对将修改为新名称。如果是批量修改，命名方式为“新名称 + 后缀”。

删除密钥对

用户可根据需要删除密钥对，但需要注意删除后无法恢复。

1. 进入“[裸金属服务器 > SSH密钥](#)”页面。
2. 选择一个或多个想要删除的密钥对，单击列表上方“删除”。
3. 在弹出页面确认信息，单击“确定”。

镜像使用说明

GPU裸金属服务器提供标准镜像，也支持自定义镜像。每个镜像都是GPU裸金属服务器运行环境的模板，包含了操作系统和预装软件以及配置。

目前GPU裸金属服务器支持以下标准镜像：

操作系统	内核版本
CentOS-6.5 64位	2.6.32-431.el6.x86_64
CentOS-7.2 64位	3.10.0-327.el7.x86_64
CentOS-6.7 64位	2.6.32-573.el6.x86_64
CentOS-7.0 64位	3.10.0-123.el7.x86_64
CentOS-7.1 64位	3.10.0-229.el7.x86_64
CentOS-7.3 64位	3.10.0-514.el7.x86_64
centos-7.2-x86_64-cuda8	3.10.0-327.el7.x86_64
centos-7.3-x86_64-cuda8	3.10.0-514.el7.x86_64
Ubuntu-12.04 64位	3.13.0-32-generic
Ubuntu-14.04 64位	4.4.0-31-generic
Ubuntu-16.04 64位	4.4.0-62-generic
ubuntu-16.04-x86_64-cuda8	4.4.0-62-generic
ubuntu-14.04-x86_64-cuda8	4.4.0-31-generic
Windows Server 2008 R2 英文 数据中心版 NT 6.1	
Windows Server 2008 R2 中文 数据中心版 NT 6.1	
Windows Server 2012 R2 英文 数据中心版 NT 6.3	
Windows Server 2012 R2 中文 数据中心版 NT 6.3	

自定义镜像管理

用户可以制作自定义镜像，并对自定义镜像进行集中管理。

创建自定义镜像

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 选择目标裸金属服务器，并单击列表上方的“关机”。
3. 在目标裸金属服务器对应的“操作”列中，选择“更多 > 制作镜像”。

4. 填写镜像名称，单击“确定”。

查看自定义镜像

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 在左侧导航栏中，选择“自定义镜像”。

在列表中可以查看自定义镜像的相关属性，包括ID、操作系统和创建时间。

删除自定义镜像

3. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
4. 在左侧导航栏中，选择“自定义镜像”。
5. 选择一个或多个想要删除的自定义镜像，并单击列表上方的“删除”。
6. 在弹出页面中确认信息，单击“删除”。

安全组

安全组是一种有状态的服务器虚拟防火墙，它用于设置单台或多台裸金属服务器的网络访问控制。有关安全组的详细介绍，请参见[安全组介绍](#)。

用户在创建GPU裸金属服务器实例时，可以将一个安全组与该实例相关联，为安全组添加规则，规定流入或流出其关联实例的流量。安全组规则按照从上到下的顺序与数据包进行逐条匹配。

创建安全组

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 在左侧导航栏中，选择“安全组”。
3. 单击“新建安全组”。
4. 在新建安全组页面，输入安全组名称，选择虚拟私有网络，并配置规则：
 1. 在“进站规则”或“出站规则”页签上，单击“新增一行”。
 2. 选择“协议”，输入“起始端口”、“结束端口”、“源IP”和“备注”。
 3. 重复上述步骤添加更多进站或者出站规则。

注意： 安全组中必须包含至少一条进站规则或出站规则。

5. 单击“确定”。

系统开始创建安全组并弹出“创建成功”提示。

编辑安全组进站规则/出站规则

编辑出站规则与进站规则步骤基本一致，本节以进站规则的编辑作为示例进行介绍。

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 在左侧导航栏中，选择“安全组”。
3. 选择需要编辑的某个安全组，单击列表上方或安全组进站规则中的“编辑进站规则”。
4. 在规则列表中编辑已有进站规则，并使用页面其他功能辅助规则的编辑：
 - 新增一行：单击“新增一行”为列表增加新的规则。
 - 导出规则：编辑安全组进站规则中，单击“导出规则”将现有进站规则列表另存为.xlsx文件，并在本地编辑保存。
 - 批量导入：单击“批量导入”并选择文件，将当前进站规则列表替换为所选文件中的规则列表。

批量导入的详细步骤为：

- a. 准备一个.xlsx格式的文件，内容的具体格式参考下图。

- b. 在进站规则页面，单击“批量导入”。
- c. 如果想备份原有规则列表，在弹出页面中单击“导出现有规则”。

d. 单击“选择文件”，并选择已准备的文件。系统将自动解析并列出规则。

e. 检查规则正确无误后，单击“开始导入”。

操作成功后，可以在进站规则页面查看新导入的条目。



5. 单击“确定”。

复制安全组

用户可以将已有安全组快速地复制到其他地域或VPC。

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 在左侧导航栏中，选择“安全组”。
3. 选择需要复制的安全组，单击列表上方的“复制安全组”。
4. 在“复制安全组”页面，完成下列配置：
 - 选择目标地域和目标VPC。
 - 输入新的安全组名称。
 - 编辑进站规则和出站规则。
5. 单击“确定”。

当出现“复制成功”的提示时，表示安全组复制成功，用户可以在安全组页面查看复制的安全组信息。

删除安全组

用户可以删除不再使用的安全组，但是VPC中的默认安全组不能删除。

操作前提

安全组中不包含任何裸金属服务器和云服务器，否则安全组不能删除。取消关联裸金属服务器和取消关联云服务器的操作步骤类似，请参见[管理安全组中的裸金属服务器](#)。

操作步骤

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 在左侧导航栏中，选择“安全组”。
3. 选择一个或多个需要删除的安全组，并单击“删除”。
4. 在删除确认页面，单击“删除”。当出现“删除成功”的提示时，表示安全组已成功删除。

管理安全组中的裸金属服务器

1. 登录[裸金属服务器控制台](#)，默认进入“裸金属服务器 > 裸金属服务器”页面。
2. 在左侧导航栏中，选择“安全组”。
3. 选择一个安全组，并单击列表上方的“管理裸金属服务器”。
4. 在管理裸金属服务器页面，单击“添加”或“移除”来管理此安全组下的裸金属服务器。左侧列表展示未加入此安全组的网卡，右侧列表展示已加入此安全组的网卡。
5. 单击“确定”。

远程登录

绑定公网IP地址的GPU裸金属服务器，支持远程登录。

本节以Xshell 6软件为例进行介绍。

操作步骤

1. 在Xshell左上角选择“文件 > 新建”。
2. 在新建会话窗口中，输入GPU裸金属服务器的公网IP地址，单击“确定”。
3. 在左侧导航栏中，选择“连接 > 用户身份验证”，并根据认证类型配置验证信息。本例采用默认的root用户。

- 如果采用密码登录，选择“Password”方法并输入密码。

- 如果采用SSH密钥登录，选择“Public Key”方法，并单击“浏览”导入SSH密钥。

4. 单击“确定”。

监控

监控指标

GPU裸金属服务器支持Linux、Windows系统监控。监控项具体参考[各产品监控项](#)。

操作指南

点击您的裸金属服务器进入详情页，在详情页可以对您的机器状况进行监控，包括机器详情、流量统计、云监控、进站规则、出站规则和硬件监控。

如果您在创建裸金属服务器时选择了免费开通云监控功能，点击您的裸金属服务器详情页中的“云监控”按钮，就可以监控到30天内的CPU利用率、磁盘读写带宽、内存使用情况、内存使用率、GPU利用率、GPU温度、GPU可用显存信息。

自助安装云监控

如果您在创建时未选择免费开通云监控功能，也可在需要时手动安装云监控服务。

安装步骤参考云监控[安装代理>裸金属服务器](#)。

操作背景

当所选GPU实例搭载的GPU卡支持NvLink&NvSwitch，需额外安装与驱动版本对应的nvidia-fabricmanager服务使GPU卡间能够互联。

操作方法

1. 下载安装包

从nvidia官网选择对应镜像、系统架构、驱动版本的安装包进行下载安装即可，包含rpm和deb格式。下载地址：<https://developer.download.nvidia.cn/compute/cuda/repos/> 比如rpm的nvidia-fabric-manager-xxxx，nvidia-fabric-manager-devel-xxxx

2. 不同镜像安装方式

CentOS安装nvidia-fabricmanager

```
version=470.103.01 #已经安装的驱动版本
yum -y install yum-utils
```

```
yum-config-manager --add-repo https://developer.download.nvidia.com/compute/cuda/repos/rhel7/x86_64/cuda-rhel7.repo  
yum install -y nvidia-fabric-manager-${version}-1
```

Ubuntu安装nvidia-fabricmanager

```
version=470.103.01  
main_version=$(echo $version | awk -F '.' '{print $1}')  
apt-get update  
apt-get -y install nvidia-fabricmanager-${main_version}=${version}-*
```

3. 启动nvidia-fabricmanager服务

```
systemctl start nvidia-fabricmanager
```

4. 查看nvidia-fabricmanager服务状态

```
systemctl status nvidia-fabricmanager
```

5. 设置nvidia-fabricmanager服务开机自启动

```
systemctl enable nvidia-fabricmanager
```