



# GPU云服务器 用户指南

文档版本: 20220602



## 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。

- 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。
- 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。
- 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

| 格式          | 说明  | 样例   |
|-------------|---|--|
| ⚠ 危险        | 该类警示信息将导致系统重大变更甚至故<br>障,或者导致人身伤害等结果。      | ⚠ 危险 重置操作将丢失用户配置数据。                                      |
| ⚠ 警告        | 该类警示信息可能会导致系统重大变更甚<br>至故障,或者导致人身伤害等结果。    | 會告<br>重启操作将导致业务中断,恢复业务<br>时间约十分钟。                        |
| 〔∫〉 注意      | 用于警示信息、补充说明等,是用户必须<br>了解的内容。              | 大意<br>权重设置为0,该服务器不会再接受新<br>请求。                           |
| ? 说明        | 用于补充说明、最佳实践、窍门等 <i>,</i> 不是<br>用户必须了解的内容。 | ⑦ 说明<br>您也可以通过按Ctrl+A选中全部文件。                             |
| >           | 多级菜单递进。                                   | 单击设置> 网络> 设置网络类型。  |
| 粗体          | 表示按键、菜单、页面名称等UI元素。                        | 在 <b>结果确认</b> 页面,单击 <b>确定</b> 。                          |
| Courier字体   | 命令或代码。                                    | 执行    cd /d C:/window    命令 <i>,</i> 进入<br>Windows系统文件夹。 |
| 斜体          | 表示参数、变量。                                  | bae log listinstanceid                                   |
| [] 或者 [alb] | 表示可选项,至多选择一个。                             | ipconfig [-all -t]                                       |
| {} 或者 {alb} | 表示必选项,至多选择一个。                             | switch {act ive st and}                                  |

## 目录

| 1.NVIDIA驱动安装指引                  | 06 |
|---------------------------------|----|
| 2.常用操作导航                        | 09 |
| 3.创建GPU实例                       | 14 |
| 3.1. 创建配备GPU驱动的GPU实例(Linux)     | 14 |
| 3.2. 创建配备GRID驱动的GPU实例(Windows)  | 26 |
| 3.3. 创建未配备驱动的GPU实例              | 35 |
| 4.连接GPU实例                       | 44 |
| 4.1. 连接方式概述                     | 44 |
| 4.2. 使用Workbench连接实例            | 46 |
| 4.2.1. 通过密码或密钥认证登录Linux实例       | 46 |
| 4.2.2. 通过密码或密钥认证登录Windows实例     | 50 |
| 4.3. 使用VNC连接实例                  | 53 |
| 4.3.1. 通过密码认证登录Linux实例          | 54 |
| 4.3.2. 通过密码认证登录Windows实例        | 56 |
| 5.管理GPU实例                       | 60 |
| 5.1. 停止实例                       | 60 |
| 5.2. 启动实例                       | 62 |
| 5.3. 休眠实例                       | 64 |
| 5.4. 重启实例                       | 67 |
| 5.5. 释放实例                       | 68 |
| 5.6. GPU监控                      | 70 |
| 6.安装NVIDIA驱动                    | 73 |
| 6.1. 在GPU计算型实例中安装GPU驱动(Linux)   | 73 |
| 6.2. 在GPU计算型实例中安装GPU驱动(Windows) | 78 |
| 6.3. 在GPU实例上安装GRID驱动(Windows)   | 80 |
| 6.4. 在GPU虚拟化型实例中安装GRID驱动(Linux) | 83 |

| 7.卸载NVIDIA驱动  | 89 |
|---------------|----|
| 7.1. 卸载GPU驱动  | 89 |
| 7.2. 卸载GRID驱动 | 93 |
| 8.更新NVIDIA驱动  | 96 |

## 1.NVIDIA驱动安装指引

由于GPU实例本身并未配备相关驱动,因此您需要根据实际业务场景选择合适的GPU实例规格,并安装与之 匹配的NVIDIA驱动。本文将为您详细介绍不同场景下购买实例并安装驱动的指引。

## 驱动说明

阿里云GPU实例支持安装以下两种NVIDIA驱动:

- GPU驱动:用于驱动物理GPU。
- GRID驱动:用于获得图形加速能力。

驱动的类型以及安装方式受业务场景、实例规格、操作系统等因素影响,示意图如下所示。



## GPU虚拟化型实例驱动安装指引

GPU虚拟化实例(vgn6i、vgn5i)配备了分片虚拟化后的虚拟GPU,因此仅支持安装GRID驱动。由于GPU实例 配备的NVIDIA GPU计算卡尚未激活NVIDIA GRID License, GPU图形功能受到限制,您可以通过创建GPU实例 直接选购配备了已激活GRID License的镜像,也可以自行准备。具体说明如下:

| 操作系统 | 适用的驱动<br>类型 | 推荐使用场景 | 操作指引 |
|------|-------------|--------|------|
|------|-------------|--------|------|

#### 用户指南·NVIDIA驱动安装指引

| 操作系统   | 适用的驱动<br>类型 | 推荐使用场景                         | 操作指引   |  |  |
|--|-------------|--------------------------------|--|--|--|
| Windows操<br>作系统<br>(Windows<br>Server)                                     |             |                                | <ul> <li>推荐您在创建GPU实例时,直接在镜像市场选用预装GRID驱动的收费镜像,例如Windows<br/>Server 2019 中文版预装GRID 12.3驱动<br/>镜像、Windows Server 2016 中文版预<br/>装GRID 12.3驱动镜像。这些收费镜像带有<br/>已经激活License的GRID驱动,您无需再手动<br/>安装GRID驱动。</li> </ul> |  |  |
|  |             |                                | 具体操作,请参见创建配备GRID驱动的GPU实<br>例(Windows)。   |  |  |
|  | GRID驱动      | 需要支持OpenGL、Direct3D<br>等图形计算场景 | <ul> <li>如果您未在创建GPU实例时,通过镜像市场购买配备了GRID驱动的镜像,或者未找到您需要的操作系统及版本,请提交工单申请GRID<br/>License,然后下载GRID驱动安装包再安装到GPU实例即可。具体操作,请参见在GPU实例上安装GRID驱动(Windows)。</li> </ul>   |  |  |
|  |             |                                | <ul> <li>如果您已经提前购买了GRID License,仅需要<br/>购买未配备驱动的GPU实例。具体操作,请参<br/>见创建未配备驱动的GPU实例。</li> </ul>   |  |  |
|  |             |                                | ⑦ 说明 创建GPU实例后,您需要手<br>动安装GRID驱动。具体操作,请咨询<br>License提供方。  |  |  |
| Linux操作系<br>统(Alibaba<br>Cloud<br>Linux、<br>Cent OS、<br>Ubunt u、<br>SUSE等) | GRID驱动      | 需要进行深度学习、AI等非图<br>形计算的通用计算业务场景 | 请 <mark>提交工单</mark> 申请GRID License,然后手动安装<br>GRID驱动。具体操作,请参见 <mark>在GPU虚拟化型实</mark><br>例中安装GRID驱动(Linux)。  |  |  |

## GPU计算型实例驱动安装指引

GPU计算型实例规格族包含以下实例规格:

- GPU计算型实例规格族: gn7i、gn7、gn6i、gn6e、gn6v、gn5i、gn5
- GPU计算型弹性裸金属服务器实例规格族: ebmgn7i、ebmgn7、ebmgn6e、ebmgn6v、ebmgn6i、 ebmgn5、ebmgn5i

⑦ 说明 其中, gn7和ebmgn7实例规格仅适用于深度学习、AI、科学计算等通用计算场景, 您需要 为该实例规格安装GPU驱动。

不同的业务场景适用的驱动类型有所区别,具体说明如下:

• 深度学习、AI、科学计算等通用计算场景

| 操作系统  | 适用的驱动类型 | 操作指引  |
|---|---------|---|
| Linux操作系统<br>(Alibaba Cloud<br>Linux、CentOS、<br>Ubuntu、SUSE<br>等) | GPU驱动   | <ul> <li>推荐您在创建GPU实例时,在公共镜像选中自动安装GPU驱动,<br/>实例创建成功后首次启动时,将自动执行驱动的安装。具体操作,请参见创建配备GPU驱动的GPU实例(Linux)。</li> <li>如果在公共镜像中没有您需要的操作系统或版本,您也可以先创建未配备驱动的GPU实例,再前往NVIDIA官网下载GPU驱动并手动安装。相关操作指引如下:</li> <li>创建未配备驱动的GPU实例</li> <li>在GPU计算型实例中安装GPU驱动(Linux)</li> </ul> |
| Windows操作系统<br>(Windows<br>Server)                                | GPU驱动   | 目前暂时不支持通过创建GPU实例自动配置Windows操作系统的<br>GPU驱动,您可以先创建未配备驱动的GPU实例,再前往NVIDIA官<br>网下载GPU驱动并手动安装。相关操作指引如下:   |

## ● 需要支持OpenGL、Direct3D等图形计算的场景

| 操作系统  | 适用的驱动类型 | 操作指引  |
|---|---------|---|
|   |         | <ul> <li>推荐您在创建GPU实例时,直接在镜像市场选用预装GRID驱动的收费镜像,例如Windows Server 2019 中文版预装GRID 13 驱动镜像、Windows Server 2016 中文版预装GRID 13驱动镜像。这些收费镜像带有已经激活License的GRID驱动,您无需再手动安装GRID驱动。</li> </ul> |
|   |         | 实例创建完成后,即可自动安装GRID驱动。具体操作,请参见创<br><mark>建配备GRID驱动的GPU实例(Windows)</mark> 。   |
| Windows操作系统<br>(Windows<br>Server)                          | GRID驱动  | <ul> <li>如果您未在创建GPU实例时,通过镜像市场购买配备了GRID驱动<br/>的镜像,或者未找到您需要的操作系统及版本,请提交工单申请<br/>GRID License,然后下载GRID驱动安装包再安装到GPU实例即<br/>可。具体操作,请参见在GPU实例上安装GRID驱动<br/>(Windows)。</li> </ul>      |
|   |         | <ul> <li>如果您已经提前购买了GRID License,仅需要购买未配备驱动的<br/>GPU实例。具体操作,请参见创建未配备驱动的GPU实例。</li> </ul>   |
|   |         | ⑦ 说明 创建GPU实例后,您需要手动安装GRID驱动。<br>具体操作,请咨询License提供方。   |
|   |         |   |
| Linux操作系统<br>(Alibaba Cloud<br>Linux、CentOS、<br>Ubuntu、SUSE | GPU驱动   | <ul> <li>建议您在创建GPU实例时,在公共镜像选中自动安装GPU驱动,</li> <li>实例创建成功后首次启动时,将自动执行驱动的安装。具体操作,请参见创建配备GPU驱动的GPU实例(Linux)。</li> </ul>   |
|   |         | <ul> <li>如果在公共镜像中没有您需要的操作系统或版本,您也可以先创<br/>建未配备驱动的GPU实例,再前往NVIDIA官网下载GPU驱动并手<br/>动安装。相关操作指引如下:</li> </ul>  |
| 寺)  |         | ■ 创建未配备驱动的GPU实例   |
|   |         | ■ 在GPU计算型实例中安装GPU驱动(Linux)  |

## 2.常用操作导航

GPU实例作为云服务器ECS的一类实例规格,保持了与ECS实例相同的操作方式。在使用云服务器ECS时,您可能会遇到各种问题,例如远程连接、更换操作系统、扩容云盘、升高或降低实例配置、使用快照或镜像等。本文介绍了云服务器ECS的常用操作,供您参考。

## 使用限制

- 使用云服务器ECS的注意事项,请参见使用须知。
- 使用云服务器ECS的资源规格限制,请参见使用限制和查看和提升实例配额。
- •

## 创建并管理ECS实例

- 您可以按以下步骤操作ECS实例的生命周期:
  - i. 使用向导创建实例
  - ii. 远程连接ECS实例
  - iii. 停止实例
  - iv. 释放实例
- 如果当前的实例规格或网络配置无法满足业务需求,您可以变更实例规格、IP地址和公网带宽峰值:
  - 。 包年包月实例:
    - 包年包月实例升配规格
    - 续费降配
    - 临时升级带宽(连续时间段)
  - 按量付费实例:
    - 按量付费实例变配规格
    - 按量付费实例修改带宽
  - 。 ECS实例IP地址操作:
    - 更换公网IP地址
    - 专有网络类型ECS公网IP转为弹性公网IP
- 如果当前的操作系统无法满足需求,您可以更换操作系统。具体操作,请参见更换操作系统。
- 您可以使用以下功能精细化控制和管理ECS实例:
  - 实例自定义数据
  - o 实例元数据
  - o 实例标识
  - o 实例RAM角色

## 管理计费

包年包月实例:

您可以使用不同的方式续费包年包月实例:

- o 手动续费实例
- 自动续费实例

- 续费降配
- 续费变配
- 统一包年包月实例的到期日
- 按量付费实例:

您可以为按量付费实例启用节省停机模式,更多信息,请参见按量付费实例节省停机模式。

- 转换实例计费方式:
  - o 按量付费转包年包月
  - 包年包月转按量付费

## 提高计费性价比

- 您可以购买抢占式实例,降低部分场景下的使用成本,搭配弹性供应实现自动化交付。具体操作,请参见创建弹性供应组和创建抢占式实例。
- 您可以购买预留实例券,提高实例抵扣方式灵活性和降低成本。具体操作,请参见<mark>购买预留实例券</mark>。
- 您可以购买存储容量单位包,抵扣同一地域下按量付费云盘的计费账单。具体操作,请参见创建存储容量 单位包。

## 创建并管理云盘



当云盘作数据盘用时,您可以按以下步骤使用云盘:

- 1. 创建云盘。
- 2. 挂载数据盘。
- 3. 分区格式化数据盘(Linux)或分区格式化数据盘(Windows)。
- 4. 创建快照备份数据。具体操作,请参见创建一个云盘快照。
- 5. 如果已有的云盘容量无法满足需求,您可以扩容云盘。具体操作,请参见以下文档:
  - o 在线扩容云盘(Linux系统)
  - o 离线扩容云盘(Linux系统)
  - o 在线扩容云盘 (Windows系统)
  - o 离线扩容云盘 (Windows系统)

6. 如果云盘数据出错,您可以使用某个时刻的云盘快照回滚云盘。具体操作,请参见使用快照回滚云盘。

- 7. 如果要将云盘恢复到初始状态,您可以重新初始化云盘。具体操作,请参见重新初始化数据盘。
- 8. 卸载数据盘。
- 9. 释放云盘。

## 创建和管理快照



#### 您可以按以下步骤使用快照:

- 1. 创建快照, 支持手动创建快照和自动创建快照:
  - o 创建一个云盘快照。
  - 使用自动快照策略,定期自动创建快照。具体操作,请参见<mark>执行或取消自动快照策略</mark>。
- 2. 查看快照容量。
- 3. 为了节省快照存储空间,删除不必要的快照。具体操作,请参见优化快照使用成本。

快照的常见应用场景如下所示:

- 用于拷贝或恢复数据:您可以使用快照创建云盘或者回滚云盘。具体操作,请参见使用快照创建云盘和使用快照回滚云盘。
- 用于快速部署环境:您可以使用系统盘快照创建自定义镜像,并使用自定义镜像创建实例。具体操作,请 参见使用快照创建自定义镜像和使用自定义镜像创建实例。



## 创建并管理自定义镜像

控制台上操作的主要都是自定义镜像。使用自定义镜像,您可以快速部署业务环境。您可以通过以下方式获 取自定义镜像。

- 使用快照创建自定义镜像。
- 使用实例创建自定义镜像。
- 使用Packer创建自定义镜像。
- 不同地域之间复制镜像。具体操作,请参见复制镜像。
- 不同账号之间共享镜像。具体操作,请参见共享镜像。

- 导入自定义镜像。
- 使用Packer创建并导入本地镜像。

您可以导出镜像备份环境。具体操作,请参见导出镜像。

## 创建并管理安全组

您可以按以下步骤使用安全组:

| 1.创建安全组 | <br>2.添加安全组规则 | <br>3.ECS实例加入安全组 | <br>4.管理安全组 | <br>5.管理安全组规则 |  |
|---------|---------------|------------------|-------------|---------------|--|
|         |               |                  |             |               |  |

- 1. 创建安全组。
- 2. 添加安全组规则。
- 3. ECS实例加入安全组。
- 4. 删除安全组规则。
- 5. 删除安全组。

为了方便部署业务,您可以跨地域、跨网络类型克隆安全组。具体操作,请参见克隆安全组。

如果新的安全组规则对线上业务产生了不利影响,您可以全部或部分还原安全组规则。具体操作,请参见还 原安全组规则。

### 创建并授予实例RAM角色

您可以按以下步骤使用密钥对:

- 1. (可选)为RAM用户授予操作实例RAM角色的权限策略。具体操作,请参见授权RAM用户使用实例RAM 角色。
- 2. 创建并授予实例RAM角色。具体操作,请参见授予实例RAM角色。
- 3. 在使用过程中,您可以随时更换实例RAM角色。具体操作,请参见更换实例RAM角色。

## 创建并使用密钥对

您可以按以下步骤使用密钥对:

- 1. 创建SSH密钥对或者导入SSH密钥对。
- 2. 绑定SSH密钥对。
- 3. 通过密钥认证登录Linux实例。
- 4. 解绑SSH密钥对。
- 5. 删除SSH密钥对。

## 创建并使用弹性网卡

您可以按以下步骤使用弹性网卡:

|--|

- 1. 创建弹性网卡。
- 2. 将弹性网卡附加到实例或者在创建实例时附加弹性网卡。
- 3. (可选)配置辅助弹性网卡。

- 4. 分配辅助私网IP地址。
- 5. 解绑弹性网卡。
- 6. 删除弹性网卡。

## 搭建IPv6专有网络

- 运行Windows Server操作系统的ECS实例的具体操作,请参见Windows实例使用IPv6导航。
- 运行Linux操作系统的ECS实例的具体操作,请参见Linux实例使用IPv6导航。

## 使用标签

您可以使用标签管理各种资源,提高效率。您可以按以下步骤使用标签:

- 1. 新建并绑定标签。
- 2. 使用标签检索资源。
- 3. 删除或解绑标签。

## 使用实例启动模板

实例启动模板帮助您快速创建相同配置的ECS实例,您可以按以下步骤使用实例启动模板:

- 1. 创建实例启动模板。
- 2. 创建实例启动模板的新版本。
- 3. 删除实例启动模板和版本。

## 使用部署集

部署集帮助您提供底层应用的高可用性,您可以按以下步骤使用部署集:

- 1. 创建部署集。
- 2. 在部署集内创建ECS实例。
- 3. 调整实例所属部署集。
- 4. 删除部署集。

## 使用云助手

云助手可以发送远程命令,免去了运维过程中的使用跳板机的不便。您可以按以下步骤使用云助手:

- 1. (可选)部分ECS实例需要您手动安装和配置云助手客户端。具体操作,请参见安装云助手客户端。
- 2. 创建命令。
- 3. 执行命令。
- 4. 查看执行结果及修复常见问题。

## 3.创建GPU实例

## 3.1. 创建配备GPU驱动的GPU实例(Linux)

本文适用于仅需要进行深度学习、AI等非图形计算的通用计算业务场景,该场景推荐您使用支持配备GPU驱动的GPU计算型实例。本文为您提供在创建Linux操作系统的GPU计算型实例时,配置自动安装GPU驱动的详细操作指导。

## 背景信息

阿里云目前仅支持在创建GPU计算型实例(如GPU计算型实例、GPU计算型弹性裸金属服务器实例)并选择了Linux操作系统的公共镜像时,配置自动安装GPU驱动,不适用于以下场景:

- 如果您需要购买Windows操作系统的实例及GPU驱动,请购买未配备驱动的GPU实例后再手动安装驱动, 具体操作,请参见创建未配备驱动的GPU实例和在GPU计算型实例中安装GPU驱动(Windows)。
- 如果在公共镜像中没有您需要的Linux操作系统或版本,请您先创建未配备驱动的普通GPU实例,再前往 NVIDIA官网购买GPU驱动并手动安装。具体操作,请参见创建未配备驱动的GPU实例和在GPU计算型实例 中安装GPU驱动(Linux)。
- 如果您需要使用共享镜像和自定义镜像,请自行安装GPU驱动。

不同场景适用的驱动类型和安装方式有所不同,更多信息,请参见NVIDIA驱动安装指引。

## 准备工作

1. 注册阿里云账号,并完善账号信息。

- 完成实名认证。具体操作,请参见阿里云账号注册流程。
- 开通按量付费ECS资源时,您的阿里云账户余额(即现金余额)和代金券的总值不得小于100.00元人 民币。具体充值操作,请参见如何充值。

2. 前往实例创建页。

## 操作流程

- 步骤1: 完成基础配置
- 步骤2: 完成网络和安全组配置
- 步骤3:完成系统配置
- (可选)步骤4:完成分组设置
- 步骤5: 确认订单

#### 步骤1:完成基础配置

基础配置包括购买实例的基础需求(付费模式、地域可用区)以及一台实例所需的基础资源(实例规格、镜像、存储)。完成基础设置后,单击下**一步:网络和安全组设置**。

1. 选择付费模式。

付费模式影响实例的计费和收费规则,不同付费模式的实例遵循的资源状态变化规则也存在差异。

| 付费模式 | 说明               | 相关文档 |
|------|------------------|------|
| 包年包月 | 先付费后使用,最短可以按周购买。 | 包年包月 |

| 付费模式  | 说明   | 相关文档   |
|-------|--|--------|
| 按量付费  | 先使用后付费,计费周期精确到秒,方便您按需购买<br>和释放资源。                            | ◎ 按量付费 |
|       | ⑦ 说明 推荐搭配使用节省计划优化成本。   | ○ 节省计划 |
| 抢占式实例 | 先使用后付费,相对于按量付费实例价格有一定的折<br>扣,但可能因市场价格变化或实例规格库存不足而自<br>动释放实例。 | 抢占式实例  |

#### 2. 选择地域及可用区。

选择距离近的地域可以降低网络时延,实例创建完成后不支持更改地域和可用区。更多信息,请参见<mark>地</mark> 域和可用区。

- 3. 选择实例规格并进行相关配置。
  - i. 定位到**异构计算GPU/FPGA/NPU**架构下的GPU**计算型**,或者**弹性裸金属服务器**架构下的GPU 型,然后选择实例规格。

? 说明

- 可选的实例规格和地域等因素有关,您可以前往ECS实例可购买地域查看实例的可购情况。
- 如果您有特定的配置需求,例如需要挂载多张弹性网卡、使用ESSD云盘、使用本地盘等,请确认实例规格是否支持。关于实例规格的特点、适用场景、指标数据等信息,请参见实例规格族。
- 如果您需要针对特定场景购买实例,可以查看场景化选型页签中的推荐信息,例如适用于AI机器学习等场景的GPU实例规格。
- ii. 在当前选择实例后确认所选实例规格信息。

如果您需要体验阿里云推出的集群极速部署工具FastGPU,帮助您将人工智能计算任务一键部署在 阿里云的IaaS资源上,可以单击FastGPU极速创建GPU集群>,前往阿里云开发者实验室,通过三 个训练场景中的应用实验进行学习与体验。FastGPU的更多信息,请参见什么是集群极速部署工具 FastGPU。

#### iii. 如果选择付费模式为抢占式实例,配置使用时长和上限价格。

使用时长指抢占式实例的保护期,超出保护期后可能因市场价格变化或实例规格库存不足而自动释 放实例。使用时长选项说明如下表所示。

| 抢占式实例使用时长 | 说明                             |
|-----------|--------------------------------|
| 设定使用实例1小时 | 抢占式实例创建后有1小时保护期,在保护期内不会被自动释放。  |
| 无确定使用时长   | 抢占式实例创建后没有保护期,但比有保护期的抢占式实例更优惠。 |

#### 上限价格选项说明如下表所示。

| 单台实例规格上限价 | 说明   |
|-----------|--|
| 使用自动出价    | 始终使用实例规格的实时市场价格,该市场价格不会超过对应按量付费<br>实例的价格。使用自动出价可以避免抢占式实例因实时市场价格超过上<br>限被自动释放,但不能避免因实例规格的库存不足被自动释放。 |
| 设置单台上限价   | 自行输入明确的价格上限,实例规格的实时市场价格超出该上限或者库存不足时,抢占式实例都会被自动释放。  |

iv. 配置购买实例的数量。

使用向导单次最多购买100台实例。此外,您持有的实例数量不能超过配额,具体配额以页面显示为准。更多信息,请参见查看和提升实例配额。

- 4. 选择镜像。
  - i. 在镜像区域的公共镜像中,选择所需的Linux操作系统及版本。
  - ii. 选中**安装GPU驱动**,根据自身业务场景,判断是否需要选中AIACC训练加速和AIACC推理加速。 然后再选择对应的CUDA版本、Driver版本和cuDNN版本。

| 镜像      | 公共镜像                 | 自定义镜像                 | 共享镜像           | 镜像市场                |               |
|---------|----------------------|-----------------------|----------------|---------------------|---------------|
|         | CentOS •             | 7.9 64位               |                | -                   | ✓ 安全加固 ⑦      |
|         | 安装 GPU 驱动 ⑦ [        | AIACC训练加速 ?           | AIACC推理加速 ?    |                     |               |
|         | 请选择 GPU CUDA 版本 / D  | iver 版本 / CUDNN 版本    |                |                     |               |
|         | CUDA 版本 11.2.2 > - 1 | Driver 版本 460.91.03 > | CUDNN 版本 8.1.1 |                     |               |
| 存储      | CUDA 版本 11.0.2 >     |                       |                |                     |               |
| 云盘参数和性能 | CUDA 版本 10.2.89 >    |                       | 主能级另           | リ⑦: PLO (単盘IOPS性能上限 | 1万) 🔹 🔽 随实例释放 |
|         | CUDA版本 10.1.168 >    |                       |                |                     |               |
|         | CUDA 版本 10.0.130 >   |                       |                |                     |               |

⑦ 说明 如果您选择了超级计算集群sccgn7ex实例规格族,那么还需要根据自身业务,选择 是否自动安装RDMA软件栈。

GPU驱动、RDMA软件栈、AIACC训练加速以及AIACC推理加速的说明如下:

GPU驱动用于驱动物理GPU,配合CUDA、cuDNN库可以高效地使用GPU。安装GPU驱动会同时安装GPU驱动、CUDA和cuDNN库。并且仅部分Linux公共镜像支持自动安装GPU驱动。可选版本如下表所示:

⑦ 说明 如果是新业务系统,建议选择最新版本的GPU驱动、CUDA、cuDNN库。

## 用户指南·创建GPU实例

| CUDA版本   | Driver版本   | cuDNN版<br>本   | 支持的公共镜像版本(仅<br>支持自营镜像)  | 支持的实例规格  |
|----------|--|---|---|--|
| 11.4.1   | 470.82.01  | 8.2.4   | <ul> <li>Alibaba Cloud Linux<br/>2、Alibaba Cloud<br/>Linux 3</li> <li>Ubuntu 20.04、<br/>18.04、16.04</li> <li>Centos 8.x、7.x</li> <li>Debian 10.10</li> <li>⑦ 说明<br/>Debian 10.10版本<br/>仅支持超级计算集<br/>群sccgn7ex。</li> </ul> | <ul> <li>gn7i、gn7e、gn6v、<br/>gn6i、gn6e、gn5、<br/>gn5i</li> <li>ebmgn7、ebmgn7i、<br/>ebmgn7e、<br/>ebmgn6v、ebmgn6i、<br/>ebmgn6e、ebmgn5i</li> <li>sccgn7ex</li> </ul> |
| 11.2.2   | 460.91.03  | 8.1.1   | <ul> <li>Alibaba Cloud Linux<br/>2、Alibaba Cloud<br/>Linux 3</li> <li>Ubuntu 20.04、<br/>18.04、16.04</li> <li>Centos 8.x、7.x</li> </ul>  | <ul> <li>gn7、gn7i、gn7e、<br/>gn6v、gn6i、gn6e、<br/>gn5、gn5i</li> <li>ebmgn7、ebmgn7i、<br/>ebmgn7e、<br/>ebmgn6v、ebmgn6i、<br/>ebmgn6e、ebmgn5i</li> </ul>               |
| 11.0.2   | 460.91.03  | <ul><li>8.1.1</li><li>8.0.4</li></ul>               | <ul> <li>Alibaba Cloud Linux 2</li> <li>Ubuntu 20.04,<br/>18.04, 16.04</li> <li>Centos 8.x, 7.x</li> </ul>  | <ul> <li>gn7、gn7e、gn6v、<br/>gn6i、gn6e、gn5、<br/>gn5i</li> <li>ebmgn7、ebmgn7e、<br/>ebmgn6v、ebmgn6i、<br/>ebmgn6e、ebmgn5i</li> </ul>                                 |
| 10.2.89  | 460.91.03  | <ul><li>8.1.1</li><li>8.0.4</li><li>7.6.5</li></ul> | <ul> <li>Alibaba Cloud Linux 2</li> <li>Ubuntu 18.04、16.04</li> <li>Centos 8.x、7.x</li> </ul>   | <ul> <li>gn6v、gn6i、gn6e、<br/>gn5、gn5i</li> <li>ebmgn6v、ebmgn6i、<br/>ebmgn6e、ebmgn5i</li> </ul>   |
| 10.1.168 | <ul> <li>450.80.</li> <li>02</li> <li>440.64.</li> <li>00</li> </ul> | <ul><li>8.0.4</li><li>7.6.5</li><li>7.5.0</li></ul> | <ul><li>Ubuntu 18.04、16.04</li><li>Centos 7.x</li></ul>   | <ul> <li>gn6v、gn6i、gn6e、<br/>gn5、gn5i</li> <li>ebmgn6v、ebmgn6i、<br/>ebmgn6e、ebmgn5i</li> </ul>   |

#### GPU云服务器

#### 用户指南·创建GPU实例

| CUDA版本   | Driver版本  | cuDNN版<br>本  | 支持的公共镜像版本(仅<br>支持自营镜像)   | 支持的实例规格  |
|----------|---|--|--|--|
| 10.0.130 | <ul> <li>450.80.</li> <li>02</li> <li>440.64.</li> <li>00</li> </ul>                  | <ul> <li>7.6.5</li> <li>7.5.0</li> <li>7.4.2</li> <li>7.3.1</li> </ul>                               | <ul> <li>Ubuntu 18.04、16.04</li> <li>Centos 7.x</li> </ul>           | <ul> <li>gn6v、gn6i、gn6e、<br/>gn5、gn5i</li> <li>ebmgn6v、ebmgn6i、<br/>ebmgn6e、ebmgn5i</li> </ul> |
| 9.2.148  | <ul> <li>450.80.</li> <li>02</li> <li>440.64.</li> <li>00</li> <li>390.116</li> </ul> | <ul> <li>7.6.5</li> <li>7.5.0</li> <li>7.4.2</li> <li>7.3.1</li> <li>7.1.4</li> </ul>                | <ul><li>Ubuntu 16.04</li><li>Centos 7.x</li></ul>                    | <ul> <li>gn6v、gn6e、gn5、<br/>gn5i</li> <li>ebmgn6v、<br/>ebmgn6e、ebmgn5i</li> </ul>              |
| 9.0.176  | <ul> <li>450.80.</li> <li>02</li> <li>440.64.</li> <li>00</li> <li>390.116</li> </ul> | <ul> <li>7.6.5</li> <li>7.5.0</li> <li>7.4.2</li> <li>7.3.1</li> <li>7.1.4</li> <li>7.0.5</li> </ul> | <ul><li>Ubuntu 16.04</li><li>Centos 7.x</li><li>SUSE 12sp2</li></ul> | <ul> <li>gn6v、gn6e、gn5、<br/>gn5i</li> <li>ebmgn6v、<br/>ebmgn6e、ebmgn5i</li> </ul>              |
| 8.0.61   | <ul> <li>450.80.</li> <li>02</li> <li>440.64.</li> <li>00</li> <li>390.116</li> </ul> | <ul><li>7.1.3</li><li>7.0.5</li></ul>  | <ul><li>Ubuntu 16.04</li><li>Centos 7.x</li></ul>                    | ■ gn5、gn5i<br>■ ebmgn5i  |

⑦ 说明 如果您在实例创建完成后需要更换操作系统,请确保使用支持自动安装GPU驱动的镜像,避免导致自动安装失败。

#### ■ RDMA软件栈

为进一步优化神龙架构GPU服务器的网络性能,阿里云推出了GPU计算型超级计算集群实例规格 族,即sccgn系列实例。该系列机型具备了超强的计算能力和网络通信能力,其中sccgn7ex实例 规格族支持RDMA软件栈的自动安装,方便您的业务使用GPUDirect RDMA功能。更多信息,请参 见sccgn系列实例使用说明。

 AIACC-Training是阿里云自研的AI加速器,支持统一加速AI主流计算框架TensorFlow、 PyTorch、MxNet和Caffe,可以显著提升训练性能。更多信息,请参见神龙AI训练加速引擎 AIACC-Training(AIACC训练加速)。

⑦ 说明 Cent OS 6、SUSE Linux、Alibaba Cloud Linux暂时不支持安装AIACC-Training。

 AIACC-Inference是阿里云自研的AI加速器,支持统一加速AI主流计算框架TensorFlow和可导出 ONNX格式的框架,可以显著提升推理性能。更多信息,请参见神龙AI推理加速引擎AIACC-Inference(AIACC推理加速)。

⑦ 说明 Cent OS 6、SUSE Linux、Alibaba Cloud Linux暂时不支持安装AIACC-Inference。

5. 选择存储并进行相关配置。

实例通过添加系统盘、数据盘和共享盘NAS获得存储能力, 云服务器ECS提供了云盘和本地盘满足不同场景的需求。

云盘可以用作系统盘和数据盘,包括ESSD云盘、SSD云盘、高效云盘等类型。更多信息,请参见云盘概述。

⑦ 说明 随实例一起创建的云盘和实例的付费模式相同。

本地盘只能用作数据盘,如果实例规格配备了本地盘(例如本地SSD型、大数据型等),页面中会显示 本地盘的信息。更多信息,请参见本地盘。

? 说明 不支持自行为实例挂载本地盘。

i. 选择系统盘。

系统盘用于安装操作系统,默认容量为40 GiB,但实际可设置的最低容量和镜像类型有关,如下表 所示。

| 镜像                           | 系统盘容量范围(GiB)           |
|------------------------------|------------------------|
| Linux(不包括CoreOS和Red<br>Hat ) | [max{20, 镜像文件大小}, 500] |
| FreeBSD                      | [max{30, 镜像文件大小}, 500] |
| CoreOS                       | [max{30, 镜像文件大小}, 500] |
| Red Hat                      | [max{40, 镜像文件大小}, 500] |
| Windows                      | [max{40, 镜像文件大小}, 500] |

ii. (可选)选择数据盘。

支持创建空云盘或者用快照创建云盘。快照是云盘在某一时间点数据状态的备份文件,用快照创建 云盘便于快速导入数据。选择数据盘时,您还可以加密云盘满足数据安全或法规合规等场景的要 求。关于数据加密的介绍,请参见加密概述。

⑦ 说明 单台实例支持挂载的数据盘存在数量限制。更多信息,请参见块存储使用限制。

iii. (可选)选择共享盘NAS。

如果您有较多数据需要供多台实例共享访问,推荐使用NAS文件系统,可以节约大量拷贝与同步成本。

选择已有的NAS文件系统,或者单击创建新NAS文件系统前往NAS文件系统控制台即时创建NAS 文件系统。具体操作,请参见通过控制台创建通用型NAS文件系统。创建完成后,返回ECS实例创建向 导并单击 • 按钮,查看NAS文件系统列表。关于挂载NAS文件系统时的注意事项,请参见新购ECS时 挂载NAS文件系统。

6. (可选)配置快照服务。

创建实例时即可为云盘开启自动备份,有效应对数据误删等风险。

选择已有的自动快照策略,或者单击创建自动快照策略前往快照页面即时创建自动快照策略。具体操 作,请参见创建自动快照策略。创建完成后,返回ECS实例创建向导并单击。按钮,查看自动快照策略列 表。

## 步骤2:完成网络和安全组配置

网络和安全组配置提供了公网以及与其他阿里云资源通信的能力,并保障实例在网络中的安全。完成网络和 安全组配置后,单击下**一步:系统配置**。

1. 选择网络及公网IP。

| 配置项 | 说明   | 相关文档   |
|-----|--|--|
| 网络  | <ul> <li>选择专有网络。</li> <li>专有网络是您在阿里云自己定义的一个隔离网络环境,您可以完全掌控自己的专有网络,例如选择IP地址范围、配置路由表和网关等。</li> <li>如果在创建实例时不需要自定义专有网络配置,您可以跳过本步骤,系统会自动创建默认专有网络和交换机。</li> <li>② 说明 仅当实例所属地域不存在已创建的自定义专有网络时,才支持跳过本步骤。</li> <li>选择已有的专有网络和交换机,或者单击前往控制台的建前往专有网络控制台即时创建专有网络和交换机。</li> <li>动建完成后,返回ECS实例创建向导并单击 &lt;&gt;</li></ul> | <ul> <li>什么是专有网络</li> <li>创建专有网络</li> <li>创建交换机</li> <li>为已有VPC开通IPv6网段</li> <li>为已有交换机开通IPv6网段</li> </ul> |

| 配置项  | 说明   | 相关文档      |
|------|--|-----------|
|      | 如果您在基础配置页面中选择了Windows 2008 R2及<br>以下版本的镜像,您必须在此处选中分配公网IPv4地<br>址复选框,或者在创建实例后绑定弹性公网IP,以便通<br>过其他协议连接实例,例如RDP(Windows自带的远<br>程连接)、PCOIP、XenDesktop HDX 3D等。否则,<br>在GPU驱动安装完成后,您将无法通过管理终端连接实<br>例,连接实例时会始终显示黑屏或停留在启动界面。 |           |
|      | ⑦ 说明 RDP不支持DirectX、OpenGL等应<br>用,您需要自行安装VNC服务和客户端。   |           |
|      | 具体操作如下:  |           |
| 公网IP | i. 选中 <b>分配公网IPv4地址</b> 。  | 什么是弹性公网IP |
|      | ii. 选择带宽计费模式。  |           |
|      | <ul> <li>按固定带宽:按您选择的带宽值计费,适用于<br/>对网络带宽有稳定要求的场景。</li> </ul>  |           |
|      | 按使用流量:按实际使用的流量计费,此时选择的带宽峰值用于防止突然爆发的流量产生较高费用。适用于对网络带宽要求变化大的场景,例如大部分时间流量较低,但间歇性出现流量高峰。   |           |
|      | iii.选择按固定带宽的带宽值或按使用流量的带宽峰<br>值。  |           |

2. 选择安全组。

安全组是一种虚拟防火墙,用于控制安全组内实例的入流量和出流量。更多信息,请参见安全组概述。

如果在创建实例不需要自定义安全组配置,您可以跳过本步骤,系统会自动创建默认安全组。默认安全 组入方向放行22端口、3389端口及ICMP协议,在创建完成后您可以修改安全组配置。

i. 如果您需要创建一个新的安全组, 单击新建安全组。

关于安全组各项配置的详细说明,请参见创建安全组。

- ii. 单击重新选择安全组。
- iii. 在**请选择安全组**对话框,选择一个或多个安全组,然后单击**选择**。
- 3. 配置弹性网卡。

弹性网卡分为主网卡和辅助网卡。主网卡不支持从实例解绑,只能随实例一起创建和释放。辅助网卡支持自由绑定至实例和从实例解绑,方便您在实例之间切换网络流量。如需随实例一起创建辅助网卡,请 单击+图标,然后选择辅助网卡所属的交换机。

⑦ 说明 创建实例时只能添加1块辅助网卡,您也可以在实例创建完成后单独创建辅助网卡并绑定至实例。关于各实例规格支持绑定的弹性网卡的数量,请参见实例规格族。

## 步骤3:完成系统配置

系统配置包括登录凭证、主机名、实例自定义数据等,用于定制实例在控制台和操作系统内显示的信息或使 用方式。完成系统配置后,单击**下一步:分组设置**。 1. 配置登录凭证。

建议选择**密钥对**或**自定义密码**。如果您选择**创建后设置**,在通过管理终端连接实例前必须先绑定SSH 密钥对或者重置密码,然后重启实例使修改生效。如果此时GPU驱动尚未安装完成,重启操作会导致安 装失败。

2. 配置控制台中显示的实例名称和操作系统内部显示的主机名称。

创建多台实例时,设置有序的实例名称和主机名称便于从名称了解实例的批次等信息。关于设置有序名 称的规则,请参见<mark>批量设置有序的实例名称或主机名称</mark>。

- 3. 配置高级选项。
  - i. 选择实例RAM角色。

实例通过实例RAM角色获得该角色拥有的权限,可以基于临时安全令牌STS(Security Token Service)访问指定云服务的API和操作指定的云资源,安全性更高。

选择已有的实例RAM角色,或者单击创建实例RAM角色前往RAM控制台即时创建实例RAM角色。

创建完成后,返回ECS实例创建向导并单击 • 按钮,查看实例RAM角色列表。具体操作,请参见创

建实例RAM角色并为角色授予权限。

ii. 选择实例元数据访问模式。

实例元数据(metadata)包含了实例在阿里云系统中的信息,您可以在运行中的实例内方便地查 看实例元数据,并基于实例元数据配置或管理实例。关于如何通过普通模式和加固模式查看实例元 数据,请参见查看实例元数据。

| 实例元数据访问模式    | 说明                             |
|--------------|--------------------------------|
| 普通模式(兼容加固模式) | 实例创建完成后,支持通过普通模式或者加固模式查看实例元数据。 |
| 仅加固模式        | 实例创建完成后,仅支持通过加固模式查看实例元数据。      |

 警告 选择仅加固模式会导致cloud-init组件初始化配置失败,影响实例的元数据、自定 义数据等配置,请慎重选择。

#### iii. 配置实例自定义数据。

实例自定义数据可以作为实例自定义脚本在启动实例时执行,实现自动化配置实例,或者仅作为普通数据传入实例。更多信息,请参见使用实例自定义数据(Linux实例)和使用实例自定义数据(Windows实例)。

由于您在基本配置页面中选择了安装GPU驱动、RDMA软件栈(仅超级计算集群sccgn7ex涉及)、 AIACC训练加速、AIACC推理加速,在该区域会显示自动安装脚本的内容。实例创建成功后首次启 动时,cloud-init会自动执行该自动安装脚本。

| 实例RAM角色 ③ | 法正将实例AAA条合 で 0 査者详慎 1 包建实例AAA条合  |
|-----------|--|
| 实例元数据访问模式 | <b>普通電式 (展音加图電式)</b> 行加图模式 ①   |
| 实例自定义数据③  | 输入已采用 Base64 编码  |
|           | (健康ならない) 組織的なMARH者, MACCIMELIE, MACCIMELIE, MACCIMELIE, SERIES, SFREENDER, STREENDER,                              |
|           | Terretere<br>Menses Spal version to install<br>RE_MENIA_LAGE_MENTERC*<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_MENDERVIEW<br>RE_MENIA_AGEC_ |

⑦ 说明 您也可以自行准备自动安装脚本并录入,同样能够实现自动安装GPU驱动、RDMA 软件栈(仅超级计算集群sccgn7ex涉及)、AIACC训练加速、AIACC推理加速的配置。具体操 作,请参见通过自定义安装脚本配置。

## (可选)步骤4:完成分组设置

分组设置提供标签、资源组等批量管理实例的方式。完成分组设置后,单击下一步:确认订单。

1. 配置标签。

标签由一对键值(Key-Value)组成。使用标签标识具有相同特征的资源后,例如所属组织或用途相同 的资源,您可以基于标签方便地检索和管理资源。更多信息,请参见<mark>标签概述</mark>。

选择已有的标签,或者填写标签键和标签值即时创建标签。

2. 选择资源组。

资源组供您从业务角度管理跨地域、跨产品的资源,并支持针对资源组管理权限。更多信息,请参见<mark>资</mark> <mark>源组</mark>。

选择已有的资源组,或者单击**去创建**前往资源管理控制台即时创建资源组。创建完成后,返回ECS实例 创建向导并单击 ③ 按钮,查看资源组列表。具体操作,请参见创建资源组。

3. 选择部署集。

部署集支持高可用策略,部署集内实例会严格分散在不同的物理服务器上,保证业务的高可用性和底层 容灾能力。

4. 选择专有宿主机。

专有宿主机是一台由单租户独享物理资源的云主机,具有满足严格的安全合规要求、允许自带许可证 (BYOL)上云等优势。

选择已有的专有宿主机,或者单击**创建专有宿主机**即时创建专有宿主机。创建完成后,返回ECS实例创 建向导并单击 ③ 按钮,查看专有宿主机列表。具体操作,请参见创建DDH。

#### 5. 选择私有池。

创建弹性保障或容量预定后,系统会自动生成私有池,预留特定属性特定数量的实例。从关联的私有池 中创建这一类实例,可以提供资源确定性保障。更多信息,请参见资源保障概述。

⑦ 说明 弹性保障和容量预定仅支持为按量付费实例保障资源供应确定性。

| 私有池容量 | 说明  |
|-------|---|
| 开放    | 优先使用开放类型私有池的容量,如果开放类型私有池无可用容量,则尝<br>试使用公共池的容量。        |
| 不使用   | 不使用任何私有池的容量。  |
| 指定    | 继续指定一个专用或开放类型私有池的ID,使用其容量创建实例。如果该<br>私有池没有可用容量,则创建失败。 |

## 步骤5: 确认订单

在最终创建实例前,请检查实例的整体配置并配置使用时长等选项,确保各项配置符合您的要求。

1. 检查所选配置。

如需修改配置,单击 

❷标前往对应页面。您可以基于所选配置生成模板,方便后续快捷创建配置类似的实例,如下表所示。

| 快捷操作                 | 说明   | 相关文档   |
|----------------------|--|--|
| 保存为启动模板              | 将所选配置保存为启动模板,使用启动模板创建实例<br>可以免去重复配置的操作。      | 使用实例启动模板创建实<br>例   |
| 生成Open API最佳实<br>践脚本 | 自动生成API工作流和SDK示例供您参考。                        | <ul> <li>RunInstances</li> <li>批量创建ECS实例<br/>(Java)</li> <li>批量创建ECS实例<br/>(Python)</li> </ul> |
| 保存当前购买配置为<br>ROS模板   | 将所选配置保存为ROS模板,方便继续使用ROS模板创<br>建资源栈,实现一键交付资源。 | 创建资源栈  |

- 2. 配置实例的使用时长。
  - 按量付费实例:设置自动释放实例的时间。您也可以在创建实例后手动释放实例,或者配置自动释放。具体操作,请参见释放实例。
  - 包年包月实例:选择购买时长以及是否启用自动续费。您也可以在创建实例后手动续费实例,或者配置自动续费。具体操作,请参见续费概述。
- 3. 阅读*云服务器ECS服务条款*,如无疑问,选中《云服务器ECS服务条款》。
- 4. 在页面底部查看实例的总费用,如无疑问,确认创建实例并按提示完成支付。

如果您为实例设置了自动安装GPU驱动,实例创建成功后,将自动执行GPU驱动的安装任务。自动安装 耗时受不同实例规格的内网带宽和CPU核数的影响,约10~20分钟。您可以远程连接实例查看安装过 程,也可以在安装完成后查看安装日志/root/auto\_install/auto\_install.log。安装过程中的显示效果如 下表所示:

| 安装过程 | 显示效果                  |
|------|-----------------------|
| 正在安装 | 显示安装进度条。              |
| 安装成功 | 安装结果提示ALL INSTALL OK。 |
| 安装失败 | 安装结果提示INSTALL FAIL。   |

注意 在安装过程中无法使用GPU,请勿对实例进行任何操作,也不要安装其它GPU相关软件,避免自动安装失败导致实例不可用。

## 通过自定义安装脚本配置

自动安装脚本可以应用于以下两种场景:

- 如果您没有在基本配置页面中选择安装GPU驱动、RDMA软件栈(仅超级计算集群sccgn7ex涉及)、
   AIACC训练加速和AIACC推理加速,并且想要在系统配置页面手动输入自动安装脚本。
- 如果您通过调用RunInstances接口来创建GPU实例,那么,您只能通过UserData参数上传自动安装脚本。

以下操作指导您准备自动安装脚本,并在创建实例阶段输入该脚本完成自动安装GPU驱动的配置。

1. 准备自动安装脚本。

#### 自动安装脚本的内容如下:

```
#!/bin/sh
#Please input version to install
IS INSTALL RDMA=""
IS INSTALL AIACC TRAIN=""
IS INSTALL AIACC INFERENCE=""
DRIVER VERSION=""
CUDA VERSION=""
CUDNN VERSION=""
IS INSTALL RAPIDS="FALSE"
INSTALL DIR="/root/auto install"
#using .run to install driver and cuda
auto install script="auto install.sh"
script download url=$(curl http://100.100.100.200/latest/meta-data/source-address | hea
d -1)"/opsx/ecs/linux/binary/script/${auto install script}"
echo $script download url
mkdir $INSTALL DIR && cd $INSTALL DIR
wget -t 10 --timeout=10 $script download url && sh ${INSTALL DIR}/${auto install script
} $DRIVER VERSION $CUDA VERSION $CUDNN VERSION $IS INSTALL AIACC TRAIN $IS INSTALL AIAC
C INFERENCE $IS INSTALL RDMA $IS INSTALL RAPIDS
```

⑦ 说明 自动安装脚本使用.run安装包方式安装GPU驱动等模块。

使用自动安装脚本时,您需要根据实际业务场景,补充如下参数。

根据您所创建的实例规格和镜像,设置对应GPU驱动、CUDA和cuDNN库的版本。详细的版本支持情况,请参见驱动支持的镜像版本和实例规格。示例如下:

```
DRIVER_VERSION="470.82.01"
CUDA_VERSION="11.4.1"
CUDNN VERSION="8.2.4"
```

◦ 设置是否安装RDMA软件栈。

⑦ 说明 仅超级计算集群sccgn7ex的实例规格族支持安装RDMA软件栈。

如果安装RDMA软件栈,则需要将IS\_INSTALL\_RDMA的值设置为*TRUE*,否则设置为*FALSE*。示例如下:

IS INSTALL RDMA="TRUE"

- 设置是否安装AIACC训练加速和AIACC推理加速。
  - 如果安装AIACC训练加速,则将IS\_INSTALL\_AIACC\_TRAIN的值设置为TRUE,否则设置为FALSE。
  - 如果安装AIACC推理加速,则将IS\_INSTALL\_AIACC\_INFERENCE的值设置为*TRUE*,否则设置 为*FALSE*。

示例如下:

IS\_INSTALL\_AIACC\_TRAIN="TRUE" IS INSTALL AIACC INFERENCE="FALSE"

2. 脚本准备完成后,您可以在系统配置阶段,在高级选项的实例自定义数据区域内,手动输入该脚本。

| 实例自定义数据⑦ | □ 输入已采用 Base64 编码   |   |
|----------|---|---|
|          | #//bin/sh<br>#PRaea proversion to install<br>IS_INSTAL_AACC_TRANK-''<br>IS_INSTAL_AACC_INFRENCE-''                  | • |
|          | Windows 支持 bat 和 powershell 两种指式,在 Base64 编码帅,第一行为 [bat] 或者 [powershell], Linux 支持 shell 脚本, 更多招格式参考 doud-init   書看 |   |
|          | 评例  |   |

在实例启动后会自动安装GPU驱动、CUDA和cuDNN库,并根据您配置的脚本判断是否自动安装RDMA软 件栈(仅超级计算集群sccgn7ex涉及)、AIACC训练加速以及AIACC推理加速。安装完成后实例会自动 重启,然后GPU驱动才能正常工作 。

⑦ 说明 GPU驱动在Persistence Mode下工作更稳定。自动安装脚本会自动开启GPU驱动的 Persistence Mode,并将该设置添加到Linux系统服务中,在开机时自动启动服务,从而确保实例重 启后还能默认开启Persistence Mode。

## 3.2. 创建配备GRID驱动的GPU实例 (Windows)

本文适用于需要支持OpenGL、Direct3D等图形计算的业务场景,该场景推荐您使用能够获取图形能力的GRID 驱动。您可以在创建GPU实例时,前往镜像市场直接购买配备了GRID驱动的Windows镜像并完成驱动的自动 安装,本文为您提供详细操作指导。

#### 背景信息

由于GPU实例配备的NVIDIA GPU计算卡尚未激活NVIDIA GRID License, GPU图形功能受到限制,因此,必须获取相关License后才能使用GRID驱动。阿里云镜像市场提供配备了GRID驱动的Windows收费镜像,这些收费镜像带有已经激活License的GRID驱动,您无需再手动安装GRID驱动。

如果您需要购买Linux操作系统的GPU实例及GRID驱动,请创建未配备驱动的Linux操作系统的GPU实例后, 再<mark>提交工单</mark>申请GRID驱动的License。创建GPU实例的具体操作,请参见创建未配备驱动的GPU实例。

不同场景适用的驱动类型和安装方式有所不同,更多信息,请参见NVIDIA驱动安装指引。

## 准备工作

- 1. 注册阿里云账号,并完善账号信息。
  - 完成实名认证。具体操作,请参见阿里云账号注册流程。
  - 开通按量付费ECS资源时,您的阿里云账户余额(即现金余额)和代金券的总值不得小于100.00元人 民币。具体充值操作,请参见如何充值。
- 2. 前往实例创建页。

### 操作流程

- 步骤1: 完成基础配置
- 步骤2: 完成网络和安全组配置
- (可选)步骤3:完成系统配置
- (可选)步骤4:完成分组设置
- 步骤5: 确认订单

## 步骤1:完成基础配置

基础配置包括购买实例的基础需求(付费模式、地域可用区)以及一台实例所需的基础资源(实例规格、镜像、存储)。完成基础设置后,单击下**一步:网络和安全组设置**。

1. 选择付费模式。

付费模式影响实例的计费和收费规则,不同付费模式的实例遵循的资源状态变化规则也存在差异。

| 付费模式  | 说明   | 相关文档                                    |
|-------|--|---|
| 包年包月  | 先付费后使用,最短可以按周购买。   | 包年包月                                    |
| 按量付费  | 先使用后付费, 计费周期精确到秒, 方便您按需购买<br>和释放资源。<br>⑦ 说明 推荐搭配使用节省计划优化成本。  | <ul><li>● 按量付费</li><li>● 节省计划</li></ul> |
| 抢占式实例 | 先使用后付费,相对于按量付费实例价格有一定的折<br>扣,但可能因市场价格变化或实例规格库存不足而自<br>动释放实例。 | 抢占式实例                                   |

#### 2. 选择地域及可用区。

选择距离近的地域可以降低网络时延,实例创建完成后不支持更改地域和可用区。更多信息,请参见<mark>地</mark> 域和可用区。

- 3. 选择实例规格并进行相关配置。
  - i. 定位到异构计算GPU/FPGA/NPU > GPU计算型或者异构计算GPU/FPGA/NPU > GPU虚拟化型,然后选择实例规格。

? 说明

- 可选的实例规格和地域等因素有关,您可以前往ECS实例可购买地域查看实例的可购情况。
- 如果您有特定的配置需求,例如需要挂载多张弹性网卡、使用ESSD云盘、使用本地盘等,请确认实例规格是否支持。关于实例规格的特点、适用场景、指标数据等信息,请参见实例规格族。
- 如果您需要针对特定场景购买实例,可以查看场景化选型页签中的推荐信息,例如适用于AI机器学习等场景的GPU实例规格。
- ii. 在当前选择实例后确认所选实例规格信息。

如果您需要体验阿里云推出的集群极速部署工具FastGPU,帮助您将人工智能计算任务一键部署在 阿里云的laaS资源上,可以单击FastGPU极速创建GPU集群>,前往阿里云开发者实验室,通过三 个训练场景中的应用实验进行学习与体验。FastGPU的更多信息,请参见什么是集群极速部署工具 FastGPU。

iii. 如果选择付费模式为抢占式实例, 配置使用时长和上限价格。

使用时长指抢占式实例的保护期,超出保护期后可能因市场价格变化或实例规格库存不足而自动释 放实例。使用时长选项说明如下表所示。

| 抢占式实例使用时长 | 说明                             |
|-----------|--------------------------------|
| 设定使用实例1小时 | 抢占式实例创建后有1小时保护期,在保护期内不会被自动释放。  |
| 无确定使用时长   | 抢占式实例创建后没有保护期,但比有保护期的抢占式实例更优惠。 |

上限价格选项说明如下表所示。

| 单台实例规格上限价 | 说明   |
|-----------|--|
| 使用自动出价    | 始终使用实例规格的实时市场价格,该市场价格不会超过对应按量付费<br>实例的价格。使用自动出价可以避免抢占式实例因实时市场价格超过上<br>限被自动释放,但不能避免因实例规格的库存不足被自动释放。 |
| 设置单台上限价   | 自行输入明确的价格上限,实例规格的实时市场价格超出该上限或者库<br>存不足时,抢占式实例都会被自动释放。  |

iv. 配置购买实例的数量。

使用向导单次最多购买100台实例。此外,您持有的实例数量不能超过配额,具体配额以页面显示 为准。更多信息,请参见<mark>查看和提升实例配额</mark>。

4. 选择镜像。

i. 在镜像区域单击镜像市场,然后单击从镜像市场获取更多选择(含操作系统)。

| 镜像 | 公共镜像        | 自定义镜像   | 共享镜像 | 镜像市场 | 0 |
|----|-------------|---------|------|------|---|
|    | 从镜像市场获取更多选择 | (含操作系统) |      |      |   |

- ii. 在镜像市场对话框的搜索框中, 输入关键字GRID, 并单击搜索。
- iii. 选择需要购买的Windows收费镜像,单击使用。

预装了GRID驱动的收费镜像如Windows Server 2019 中文版预装GRID驱动镜像、Windows Server 2016 中文版预装GRID驱动,这些收费镜像带有已经激活License的GRID驱动,您无需再手动安装GRID驱动。本文以搜索并选用Windows Server 2019 中文版预装GRID驱动镜像镜像为例。

| 镜像市场[华北 2 (北]  | 页) ]  |        |                  | ×            |
|--|---|--------|------------------|--------------|
|  | GRID  |        | 搜索               |              |
| 精选镜像   | ▲ 全部操作系统 ∨ 全部架构 ∨   |        |                  |              |
| <ul> <li>镜像分类 ∧</li> <li>✓ 全部</li> <li>操作系统</li> </ul> | BatchCompute SGE<br>基础系统: linux 采构: 64位<br>在批量计算中使用该模像,可以自动化插键Sun Grid Engine (SGE                  | v1.3 v | ★★★★★<br>9人已使用   | ¥0/月<br>使用   |
| 运行环境<br>管理与监控  | Windows Server 2019 中文版预装GRID 9起討機像<br>基础系统: windows 原物: 64位<br>Windows Server 2019 中文版预绘GRID 9驱动機像 | V1.0 V | ★★★★★<br>502人已使用 | ¥220/月<br>使用 |
| 建站系统<br>应用开发<br>数据库                                    | Windows Server 2016 中文版预读GRID 9驱动機像<br>基础系统: windows 架构: 64位<br>Windows Server 2016 中文版预装GRID 9驱动機像 | V1.0 V | ★★★★★<br>105人已使用 | ¥220/月<br>使用 |
| 服务器软件<br>企业应用<br>云安全市场                                 |   |        |                  |              |

↓ 注意 如果您购买了美国(弗吉尼亚)地域的vgn6i实例,请务必选择GRID 11版本的驱动 镜像,否则将导致实例中的vGPU不可用。

返回创建实例页面,显示类似如下信息,表示镜像选择成功。

| 镜像 | 公共鏡像   | 自定义镜像                                    | 共享現像  | 镜像市场                             |                                |
|----|--|--|---|----------------------------------|--------------------------------|
|    | 当前选择的确象 Windows Server 2019 中文版预编GRID 9驱动摘象 V1.0 ⑦ |  |   |                                  |                                |
|    | 重新选择镜像   |  |   |                                  |                                |
|    | Windows 系统执行 sysprep 制<br>该实例如果用于图形图像计算            | 刀始化时间较长(3~5分钟),<br>I (如:有 OpenGL 或者 Wind | 如无法登陆请稍后重试。初始<br>lows DirectX)需要 <mark>购买 GRID</mark> | 化过程中重启实例可能导致系<br>软件授权 或到 镜像市场 购买 | 统异常,请耐心等待。<br>包含 GRID 软件授权的镜像。 |

#### 5. 选择存储并进行相关配置。

实例通过添加系统盘、数据盘和共享盘NAS获得存储能力,云服务器ECS提供了云盘和本地盘满足不同场景的需求。

云盘可以用作系统盘和数据盘,包括ESSD云盘、SSD云盘、高效云盘等类型。更多信息,请参见云盘概述。

⑦ 说明 随实例一起创建的云盘和实例的付费模式相同。

本地盘只能用作数据盘,如果实例规格配备了本地盘(例如本地SSD型、大数据型等),页面中会显示本地盘的信息。更多信息,请参见本地盘。

⑦ 说明 不支持自行为实例挂载本地盘。

#### i. 选择系统盘。

系统盘用于安装操作系统,默认容量为40 GiB,但实际可设置的最低容量和镜像类型有关,如下表 所示。

| 镜像                           | 系统盘容量范围(GiB)           |
|------------------------------|------------------------|
| Linux(不包括CoreOS和Red<br>Hat ) | [max{20, 镜像文件大小}, 500] |
| FreeBSD                      | [max{30, 镜像文件大小}, 500] |
| CoreOS                       | [max{30, 镜像文件大小}, 500] |
| Red Hat                      | [max{40, 镜像文件大小}, 500] |
| Windows                      | [max{40, 镜像文件大小}, 500] |

ii. (可选)选择数据盘。

支持创建空云盘或者用快照创建云盘。快照是云盘在某一时间点数据状态的备份文件,用快照创建 云盘便于快速导入数据。选择数据盘时,您还可以加密云盘满足数据安全或法规合规等场景的要 求。关于数据加密的介绍,请参见加密概述。

⑦ 说明 单台实例支持挂载的数据盘存在数量限制。更多信息,请参见块存储使用限制。

iii. (可选)选择共享盘NAS。

如果您有较多数据需要供多台实例共享访问,推荐使用NAS文件系统,可以节约大量拷贝与同步成本。

选择已有的NAS文件系统,或者单击创建新NAS文件系统前往NAS文件系统控制台即时创建NAS 文件系统。具体操作,请参见通过控制台创建通用型NAS文件系统。创建完成后,返回ECS实例创建向 导并单击 • 按钮,查看NAS文件系统列表。关于挂载NAS文件系统时的注意事项,请参见新购ECS时 挂载NAS文件系统。

6. (可选)配置快照服务。

创建实例时即可为云盘开启自动备份,有效应对数据误删等风险。

选择已有的自动快照策略,或者单击创建自动快照策略前往快照页面即时创建自动快照策略。具体操 作,请参见创建自动快照策略。创建完成后,返回ECS实例创建向导并单击 • 按钮,查看自动快照策略列 表。

#### 步骤2:完成网络和安全组配置

网络和安全组配置提供了公网以及与其他阿里云资源通信的能力,并保障实例在网络中的安全。完成网络和 安全组配置后,单击下**一步:系统配置**。

1. 选择网络及公网IP。

| 配置项 | 说明 | 相关文档 |
|-----|----|------|
|     |    |      |

| 配置项  | 说明   | 相关文档  |  |
|------|--|---|--|
|      | 选择 <b>专有网络</b> 。<br>专有网络是您在阿里云自己定义的一个隔离网络环<br>境,您可以完全掌控自己的专有网络,例如选择IP地址<br>范围、配置路由表和网关等。<br>如果在创建实例时不需要自定义专有网络配置,您可<br>以跳过本步骤,系统会自动创建默认专有网络和交换<br>机。   | 0. 4.//目去方网络  |  |
| 网络   | ⑦ 说明 仅当实例所属地域不存在已创建的自定义专有网络时,才支持跳过本步骤。   | <ul> <li>创建专有网络</li> <li>创建交换机</li> </ul>                   |  |
|      | 选择已有的专有网络和交换机,或者单击前往控制台<br>创建前往专有网络控制台即时创建专有网络和交换<br>机。创建完成后,返回ECS实例创建向导并单击 <>>> 按<br>钮,查看最新的专有网络和交换机列表。   | <ul><li>● 为已有VPC开通IPv6网段</li><li>● 为已有交换机开通IPv6网段</li></ul> |  |
|      | ⑦ 说明 如果您需要为实例分配IPv6地址,请选择已开通IPv6网段的专有网络和交换机。   |   |  |
|      | 如果您在基础配置页面中选择了Windows 2008 R2及<br>以下版本的镜像,您必须在此处选中分配公网IPv4地<br>址复选框,或者在创建实例后绑定弹性公网IP,以便通<br>过其他协议连接实例,例如RDP(Windows自带的远<br>程连接)、PCOIP、XenDesktop HDX 3D等。否则,<br>在GPU驱动安装完成后,您将无法通过管理终端连接实<br>例,连接实例时会始终显示黑屏或停留在启动界面。   |   |  |
|      | <ul><li>⑦ 说明 RDP不支持DirectX、OpenGL等应</li><li>用,您需要自行安装VNC服务和客户端。</li></ul>  |   |  |
| 公网IP | <ul> <li>具体操作如下:</li> <li>i. 选中分配公网IPv4地址。</li> <li>ii. 选择带宽计费模式。</li> <li>按固定带宽:按您选择的带宽值计费,适用于对网络带宽有稳定要求的场景。</li> <li>按使用流量:按实际使用的流量计费,此时选择的带宽峰值用于防止突然爆发的流量产生较高费用。适用于对网络带宽要求变化大的场景,例如大部分时间流量较低,但间歇性出现流量高峰。</li> <li>iii. 选择按固定带宽的带宽值或按使用流量的带宽峰值。</li> </ul> | 什么是弹性公网IP   |  |

2. 选择安全组。

安全组是一种虚拟防火墙,用于控制安全组内实例的入流量和出流量。更多信息,请参见安全组概述。

如果在创建实例不需要自定义安全组配置,您可以跳过本步骤,系统会自动创建默认安全组。默认安全 组入方向放行22端口、3389端口及ICMP协议,在创建完成后您可以修改安全组配置。

- i. 如果您需要创建一个新的安全组, 单击**新建安全组**。
  - 关于安全组各项配置的详细说明,请参见创建安全组。
- ii. 单击重新选择安全组。
- iii. 在**请选择安全组**对话框,选择一个或多个安全组,然后单击选择。
- 3. 配置弹性网卡。

弹性网卡分为主网卡和辅助网卡。主网卡不支持从实例解绑,只能随实例一起创建和释放。辅助网卡支持自由绑定至实例和从实例解绑,方便您在实例之间切换网络流量。如需随实例一起创建辅助网卡,请 单击+图标,然后选择辅助网卡所属的交换机。

⑦ 说明 创建实例时只能添加1块辅助网卡,您也可以在实例创建完成后单独创建辅助网卡并绑定至实例。关于各实例规格支持绑定的弹性网卡的数量,请参见实例规格族。

### (可选)步骤3:完成系统配置

系统配置包括登录凭证、主机名、实例自定义数据等,用于定制实例在控制台和操作系统内显示的信息或使 用方式。完成系统配置后,单击**下一步:分组设置**。

1. 配置登录凭证。

登录凭证用于安全地登录实例,关于实例连接方式的介绍,请参见连接方式概述ECS远程连接操作指 <mark>南</mark>。

| 登录凭证  | 说明   |
|-------|--|
| 密钥对   | 选择已有的密钥对,或者单击创建密钥对即时创建密钥对。创建完成后,<br>返回ECS实例创建向导并单击 • 按钮,查看密钥对列表。具体操作,请参<br>见创建SSH密钥对。<br>⑦ 说明 仅Linux实例支持使用密钥对登录。 |
| 自定义密码 | 输入并确认密码。使用用户名和密码登录实例时,Linux实例默认用户名为 root ,Windows实例默认用户名为 administrator 。  |
| 创建后设置 | 在实例创建完成后,自行绑定密钥对或者重置实例密码。具体操作,请参<br>见 <mark>绑定SSH密钥对和重置实例登录密码</mark> 。   |

2. 配置控制台中显示的实例名称和操作系统内部显示的主机名称。

创建多台实例时,设置有序的实例名称和主机名称便于从名称了解实例的批次等信息。关于设置有序名 称的规则,请参见<mark>批量设置有序的实例名称或主机名称</mark>。

3. 配置高级选项。

i. 选择实例RAM角色。

实例通过实例RAM角色获得该角色拥有的权限,可以基于临时安全令牌STS(Security Token Service)访问指定云服务的API和操作指定的云资源,安全性更高。

选择已有的实例RAM角色,或者单击创建实例RAM角色前往RAM控制台即时创建实例RAM角色。 创建完成后,返回ECS实例创建向导并单击 <>>>>>按钮,查看实例RAM角色列表。具体操作,请参见创

建实例RAM角色并为角色授予权限。

ii. 选择实例元数据访问模式。

实例元数据(metadata)包含了实例在阿里云系统中的信息,您可以在运行中的实例内方便地查 看实例元数据,并基于实例元数据配置或管理实例。关于如何通过普通模式和加固模式查看实例元 数据,请参见查看实例元数据。

| 实例元数据访问模式    | 说明                             |
|--------------|--------------------------------|
| 普通模式(兼容加固模式) | 实例创建完成后,支持通过普通模式或者加固模式查看实例元数据。 |
| 仅加固模式        | 实例创建完成后, 仅支持通过加固模式查看实例元数据。     |

 警告 选择仅加固模式会导致cloud-init组件初始化配置失败,影响实例的元数据、自定 义数据等配置,请慎重选择。

iii. 配置实例自定义数据。

实例自定义数据可以作为实例自定义脚本在启动实例时执行,实现自动化配置实例,或者仅作为普通数据传入实例。更多信息,请参见使用实例自定义数据(Linux实例)和使用实例自定义数据(Windows实例)。

在输入框输入您准备的实例自定义数据。如果实例自定义数据已进行Base64编码,请选中**输入已** 采用Base64编码。

### (可选)步骤4:完成分组设置

分组设置提供标签、资源组等批量管理实例的方式。完成分组设置后,单击下一步:确认订单。

1. 配置标签。

标签由一对键值(Key-Value)组成。使用标签标识具有相同特征的资源后,例如所属组织或用途相同 的资源,您可以基于标签方便地检索和管理资源。更多信息,请参见<mark>标签概述</mark>。

选择已有的标签,或者填写标签键和标签值即时创建标签。

2. 选择资源组。

资源组供您从业务角度管理跨地域、跨产品的资源,并支持针对资源组管理权限。更多信息,请参见<mark>资</mark> <mark>源组</mark>。

3. 选择部署集。

部署集支持高可用策略,部署集内实例会严格分散在不同的物理服务器上,保证业务的高可用性和底层 容灾能力。 4. 选择专有宿主机。

专有宿主机是一台由单租户独享物理资源的云主机,具有满足严格的安全合规要求、允许自带许可证 (BYOL)上云等优势。

选择已有的专有宿主机,或者单击**创建专有宿主机**即时创建专有宿主机。创建完成后,返回ECS实例创 建向导并单击 ③ 按钮,查看专有宿主机列表。具体操作,请参见创建DDH。

5. 选择私有池。

创建弹性保障或容量预定后,系统会自动生成私有池,预留特定属性特定数量的实例。从关联的私有池 中创建这一类实例,可以提供资源确定性保障。更多信息,请参见资源保障概述。

⑦ 说明 弹性保障和容量预定仅支持为按量付费实例保障资源供应确定性。

| 私有池容量 | 说明  |
|-------|---|
| 开放    | 优先使用开放类型私有池的容量,如果开放类型私有池无可用容量,则尝<br>试使用公共池的容量。        |
| 不使用   | 不使用任何私有池的容量。  |
| 指定    | 继续指定一个专用或开放类型私有池的ID,使用其容量创建实例。如果该<br>私有池没有可用容量,则创建失败。 |

## 步骤5: 确认订单

在最终创建实例前,请检查实例的整体配置并配置使用时长等选项,确保各项配置符合您的要求。

1. 检查所选配置。

如需修改配置,单击 
图标前往对应页面。您可以基于所选配置生成模板,方便后续快捷创建配置类似的实例,如下表所示。

| 快捷操作                 | 说明   | 相关文档   |
|----------------------|--|--|
| 保存为启动模板              | 将所选配置保存为启动模板,使用启动模板创建实例<br>可以免去重复配置的操作。      | 使用实例启动模板创建实<br>例   |
| 生成Open API最佳实<br>践脚本 | 自动生成API工作流和SDK示例供您参考。                        | <ul> <li>RunInstances</li> <li>批量创建ECS实例<br/>(Java)</li> <li>批量创建ECS实例<br/>(Python)</li> </ul> |
| 保存当前购买配置为<br>ROS模板   | 将所选配置保存为ROS模板,方便继续使用ROS模板创<br>建资源栈,实现一键交付资源。 | 创建资源栈  |

2. 配置实例的使用时长。

- 按量付费实例:设置自动释放实例的时间。您也可以在创建实例后手动释放实例,或者配置自动释放。具体操作,请参见释放实例。
- 包年包月实例:选择购买时长以及是否启用自动续费。您也可以在创建实例后手动续费实例,或者配置自动续费。具体操作,请参见续费概述。
- 3. 阅读*云服务器ECS服务条款*,如无疑问,选中《云服务器ECS服务条款》。
- 4. 在页面底部查看实例的总费用,如无疑问,确认创建实例并按提示完成支付。

## 3.3. 创建未配备驱动的GPU实例

如果您仅需要购买未配备驱动的普通GPU实例,可以参考本文操作指导创建GPU实例,但为了不影响您的使用,请务必在创建完成后手动安装相关驱动。

## 背景信息

本文适用于以下场景:

- 已经通过NVIDIA官网或其他渠道获取了GRID License及驱动。
- 想要通过阿里云购买未配备驱动的GPU实例,再前往NVIDIA购买GPU驱动并手动安装。

不同场景适用的驱动类型和安装方式有所不同,更多信息,请参见NVIDIA驱动安装指引。

### 准备工作

- 1. 注册阿里云账号,并完善账号信息。
  - 完成实名认证。具体操作,请参见阿里云账号注册流程。
  - 开通按量付费ECS资源时,您的阿里云账户余额(即现金余额)和代金券的总值不得小于100.00元人 民币。具体充值操作,请参见如何充值。
- 2. 前往实例创建页。

### 操作流程

- 步骤1: 完成基础配置
- 步骤2: 完成网络和安全组配置
- (可选)步骤3:完成系统配置
- (可选)步骤4:完成分组设置
- 步骤5: 确认订单

## 步骤1:完成基础配置

基础配置包括购买实例的基础需求(付费模式、地域可用区)以及一台实例所需的基础资源(实例规格、镜像、存储)。完成基础设置后,单击下一步:网络和安全组设置。

1. 选择付费模式。

付费模式影响实例的计费和收费规则,不同付费模式的实例遵循的资源状态变化规则也存在差异。

| 付费模式 | 说明               | 相关文档 |
|------|------------------|------|
| 包年包月 | 先付费后使用,最短可以按周购买。 | 包年包月 |

| 付费模式  | 说明                                | 相关文档                                  |
|-------|-----------------------------------|---------------------------------------|
| 按量付费  | 先使用后付费,计费周期精确到秒,方便您按需购买<br>和释放资源。 | <ul><li> 按量付费</li><li> 节省计划</li></ul> |
|       | ⑦ 说明 推荐搭配使用节省计划优化成本。              |                                       |
|       | 先使用后付费,相对于按量付费实例价格有一定的折           |                                       |
| 抢占式实例 | 扣,但可能因市场价格变化或实例规格库存不足而自<br>动释放实例。 | 抢占式实例                                 |

#### 2. 选择地域及可用区。

选择距离近的地域可以降低网络时延,实例创建完成后不支持更改地域和可用区。更多信息,请参见<mark>地</mark> 域和可用区。

- 3. 选择实例规格并进行相关配置。
  - i. 定位到异构计算GPU/FPGA/NPU架构下的GPU计算型或者GPU虚拟化型,然后选择实例规格。

? 说明

- 可选的实例规格和地域等因素有关,您可以前往ECS实例可购买地域查看实例的可购情况。
- 如果您有特定的配置需求,例如需要挂载多张弹性网卡、使用ESSD云盘、使用本地盘等,请确认实例规格是否支持。关于实例规格的特点、适用场景、指标数据等信息,请参见实例规格族。
- 如果您需要针对特定场景购买实例,可以查看场景化选型页签中的推荐信息,例如适用于AI机器学习等场景的GPU实例规格。
- ii. 在当前选择实例后确认所选实例规格信息。

如果您需要体验阿里云推出的集群极速部署工具FastGPU,帮助您将人工智能计算任务一键部署在 阿里云的IaaS资源上,可以单击Fast GPU极速创建GPU集群>,前往阿里云开发者实验室,通过三 个训练场景中的应用实验进行学习与体验。Fast GPU的更多信息,请参见什么是集群极速部署工具 Fast GPU。
#### iii. 如果选择付费模式为抢占式实例, 配置使用时长和上限价格。

使用时长指抢占式实例的保护期,超出保护期后可能因市场价格变化或实例规格库存不足而自动释 放实例。使用时长选项说明如下表所示。

| 抢占式实例使用时长 | 说明                             |
|-----------|--------------------------------|
| 设定使用实例1小时 | 抢占式实例创建后有1小时保护期,在保护期内不会被自动释放。  |
| 无确定使用时长   | 抢占式实例创建后没有保护期,但比有保护期的抢占式实例更优惠。 |

#### 上限价格选项说明如下表所示。

| 单台实例规格上限价 | 说明   |
|-----------|--|
| 使用自动出价    | 始终使用实例规格的实时市场价格,该市场价格不会超过对应按量付费<br>实例的价格。使用自动出价可以避免抢占式实例因实时市场价格超过上<br>限被自动释放,但不能避免因实例规格的库存不足被自动释放。 |
| 设置单台上限价   | 自行输入明确的价格上限,实例规格的实时市场价格超出该上限或者库存不足时,抢占式实例都会被自动释放。  |

iv. 配置购买实例的数量。

使用向导单次最多购买100台实例。此外,您持有的实例数量不能超过配额,具体配额以页面显示 为准。更多信息,请参见<mark>查看和提升实例配额</mark>。

4. 选择镜像。

镜像提供了运行实例所需的信息,阿里云提供多种镜像来源供您方便地获取镜像,如下表所示。

| 镜像来源  | 说明   | 相关文档      |
|-------|--|-----------|
| 公共镜像  | 阿里云官方提供的基础镜像,均已获得正版授权,涵<br>盖Windows Server系统镜像和主流的Linux系统镜像。 | 公共镜像概述    |
| 自定义镜像 | 您自行创建或导入的镜像,包含了初始系统环境、应<br>用环境、软件配置等信息,可以节省重复配置的时<br>间。      | 自定义镜像概述   |
| 共享镜像  | 其他阿里云账号共享的自定义镜像,方便跨账号使用<br>同一镜像创建实例。                         | 共享或取消共享镜像 |
| 镜像市场  | 镜像市场中的镜像均经过严格审核,种类丰富,方便<br>您一键部署用于建站、应用开发等场景的云服务器。           | 镜像市场      |

#### 5. 选择存储并进行相关配置。

实例通过添加系统盘、数据盘和共享盘NAS获得存储能力, 云服务器ECS提供了云盘和本地盘满足不同场景的需求。

云盘可以用作系统盘和数据盘,包括ESSD云盘、SSD云盘、高效云盘等类型。更多信息,请参见<mark>云盘概</mark> 述。

⑦ 说明 随实例一起创建的云盘和实例的付费模式相同。

本地盘只能用作数据盘,如果实例规格配备了本地盘(例如本地SSD型、大数据型等),页面中会显示 本地盘的信息。更多信息,请参见本地盘。

⑦ 说明 不支持自行为实例挂载本地盘。

i. 选择系统盘。

系统盘用于安装操作系统,默认容量为40 GiB,但实际可设置的最低容量和镜像类型有关,如下表 所示。

| 镜像                           | 系统盘容量范围(GiB)           |
|------------------------------|------------------------|
| Linux(不包括CoreOS和Red<br>Hat ) | [max{20, 镜像文件大小}, 500] |
| FreeBSD                      | [max{30, 镜像文件大小}, 500] |
| CoreOS                       | [max{30, 镜像文件大小}, 500] |
| Red Hat                      | [max{40, 镜像文件大小}, 500] |
| Windows                      | [max{40, 镜像文件大小}, 500] |

ii. (可选)选择数据盘。

支持创建空云盘或者用快照创建云盘。快照是云盘在某一时间点数据状态的备份文件,用快照创建 云盘便于快速导入数据。选择数据盘时,您还可以加密云盘满足数据安全或法规合规等场景的要 求。关于数据加密的介绍,请参见加密概述。

⑦ 说明 单台实例支持挂载的数据盘存在数量限制。更多信息,请参见块存储使用限制。

iii. (可选)选择共享盘NAS。

如果您有较多数据需要供多台实例共享访问,推荐使用NAS文件系统,可以节约大量拷贝与同步成本。

选择已有的NAS文件系统,或者单击创建新NAS文件系统前往NAS文件系统控制台即时创建NAS 文件系统。具体操作,请参见通过控制台创建通用型NAS文件系统。创建完成后,返回ECS实例创建向 导并单击 • 按钮,查看NAS文件系统列表。关于挂载NAS文件系统时的注意事项,请参见新购ECS时 挂载NAS文件系统。

6. (可选)配置快照服务。

创建实例时即可为云盘开启自动备份,有效应对数据误删等风险。

选择已有的自动快照策略,或者单击创建自动快照策略前往快照页面即时创建自动快照策略。具体操 作,请参见创建自动快照策略。创建完成后,返回ECS实例创建向导并单击。按钮,查看自动快照策略列 表。

### 步骤2:完成网络和安全组配置

网络和安全组配置提供了公网以及与其他阿里云资源通信的能力,并保障实例在网络中的安全。完成网络和 安全组配置后,单击下**一步:系统配置**。

1. 选择网络及公网IP。

| 配置项  | 说明   | 相关文档  |
|--|--|---|
|  | 选择 <b>专有网络</b> 。<br>专有网络是您在阿里云自己定义的一个隔离网络环<br>境,您可以完全掌控自己的专有网络,例如选择IP地址<br>范围、配置路由表和网关等。<br>如果在创建实例时不需要自定义专有网络配置,您可<br>以跳过本步骤,系统会自动创建默认专有网络和交换<br>机。   | o 什么是专有网络   |
| 网络   | ⑦ 说明 仅当实例所属地域不存在已创建的自定义专有网络时,才支持跳过本步骤。   | <ul> <li>创建专有网络</li> <li>创建交换机</li> </ul>                   |
| 送<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1 | 选择已有的专有网络和交换机,或者单击 <b>前往控制台<br/>创建</b> 前往专有网络控制台即时创建专有网络和交换<br>机。创建完成后,返回ECS实例创建向导并单击 <>>> 按<br>钮,查看最新的专有网络和交换机列表。<br>⑦ 说明 如果您需要为实例分配IPv6地址,请<br>选择已开通IPv6网段的专有网络和交换机。   | <ul><li>◎ 为已有VPC开通IPv6网段</li><li>◎ 为已有交换机开通IPv6网段</li></ul> |
|  | 如果您在基础配置页面中选择了Windows 2008 R2及<br>以下版本的镜像,您必须在此处选中分配公网IPv4地<br>址复选框,或者在创建实例后绑定弹性公网IP,以便通<br>过其他协议连接实例,例如RDP(Windows自带的远<br>程连接)、PCOIP、XenDesktop HDX 3D等。否则,<br>在GPU驱动安装完成后,您将无法通过管理终端连接实<br>例,连接实例时会始终显示黑屏或停留在启动界面。   |   |
|  | ⑦ 说明 RDP不支持DirectX、OpenGL等应用,您需要自行安装VNC服务和客户端。   |   |
| 公网IP   | <ul> <li>具体操作如下:</li> <li>i. 选中分配公网IPv4地址。</li> <li>ii. 选择带宽计费模式。</li> <li>按固定带宽:按您选择的带宽值计费,适用于对网络带宽有稳定要求的场景。</li> <li>按使用流量:按实际使用的流量计费,此时选择的带宽峰值用于防止突然爆发的流量产生较高费用。适用于对网络带宽要求变化大的场景,例如大部分时间流量较低,但间歇性出现流量高峰。</li> <li>iii. 选择按固定带宽的带宽值或按使用流量的带宽峰值。</li> </ul> | 什么是弹性公网IP   |

2. 选择安全组。

安全组是一种虚拟防火墙,用于控制安全组内实例的入流量和出流量。更多信息,请参见安全组概述。

如果在创建实例不需要自定义安全组配置,您可以跳过本步骤,系统会自动创建默认安全组。默认安全 组入方向放行22端口、3389端口及ICMP协议,在创建完成后您可以修改安全组配置。

- i. 如果您需要创建一个新的安全组, 单击**新建安全组**。
  - 关于安全组各项配置的详细说明,请参见创建安全组。
- ii. 单击重新选择安全组。
- iii. 在**请选择安全组**对话框,选择一个或多个安全组,然后单击选择。
- 3. 配置弹性网卡。

弹性网卡分为主网卡和辅助网卡。主网卡不支持从实例解绑,只能随实例一起创建和释放。辅助网卡支持自由绑定至实例和从实例解绑,方便您在实例之间切换网络流量。如需随实例一起创建辅助网卡,请 单击+图标,然后选择辅助网卡所属的交换机。

⑦ 说明 创建实例时只能添加1块辅助网卡,您也可以在实例创建完成后单独创建辅助网卡并绑定至实例。关于各实例规格支持绑定的弹性网卡的数量,请参见实例规格族。

### (可选)步骤3:完成系统配置

系统配置包括登录凭证、主机名、实例自定义数据等,用于定制实例在控制台和操作系统内显示的信息或使 用方式。完成系统配置后,单击**下一步:分组设置**。

1. 配置登录凭证。

登录凭证用于安全地登录实例,关于实例连接方式的介绍,请参见<mark>连接方式概述ECS远程连接操作指</mark> <mark>南</mark>。

| 登录凭证  | 说明   |  |  |
|-------|--|--|--|
| 密钥对   | 选择已有的密钥对,或者单击创建密钥对即时创建密钥对。创建完成后,<br>返回ECS实例创建向导并单击 • 按钮,查看密钥对列表。具体操作,请参<br>见创建SSH密钥对。<br>⑦ 说明 仅Linux实例支持使用密钥对登录。 |  |  |
| 自定义密码 | 输入并确认密码。使用用户名和密码登录实例时,Linux实例默认用户名为 root ,Windows实例默认用户名为 administrator 。  |  |  |
| 创建后设置 | 在实例创建完成后,自行绑定密钥对或者重置实例密码。具体操作,请参<br>见 <mark>绑定SSH密钥对和重置实例登录密码</mark> 。   |  |  |

2. 配置控制台中显示的实例名称和操作系统内部显示的主机名称。

创建多台实例时,设置有序的实例名称和主机名称便于从名称了解实例的批次等信息。关于设置有序名 称的规则,请参见<mark>批量设置有序的实例名称或主机名称</mark>。

3. 配置高级选项。

i. 选择实例RAM角色。

实例通过实例RAM角色获得该角色拥有的权限,可以基于临时安全令牌STS(Security Token Service)访问指定云服务的API和操作指定的云资源,安全性更高。

选择已有的实例RAM角色,或者单击创建实例RAM角色前往RAM控制台即时创建实例RAM角色。 创建完成后,返回ECS实例创建向导并单击 <>>>>>按钮,查看实例RAM角色列表。具体操作,请参见创

建实例RAM角色并为角色授予权限。

ii. 选择实例元数据访问模式。

实例元数据(metadata)包含了实例在阿里云系统中的信息,您可以在运行中的实例内方便地查 看实例元数据,并基于实例元数据配置或管理实例。关于如何通过普通模式和加固模式查看实例元 数据,请参见查看实例元数据。

| 实例元数据访问模式    | 说明                             |
|--------------|--------------------------------|
| 普通模式(兼容加固模式) | 实例创建完成后,支持通过普通模式或者加固模式查看实例元数据。 |
| 仅加固模式        | 实例创建完成后, 仅支持通过加固模式查看实例元数据。     |

 警告 选择仅加固模式会导致cloud-init组件初始化配置失败,影响实例的元数据、自定 义数据等配置,请慎重选择。

iii. 配置实例自定义数据。

实例自定义数据可以作为实例自定义脚本在启动实例时执行,实现自动化配置实例,或者仅作为普通数据传入实例。更多信息,请参见使用实例自定义数据(Linux实例)和使用实例自定义数据(Windows实例)。

在输入框输入您准备的实例自定义数据。如果实例自定义数据已进行Base64编码,请选中**输入已** 采用Base64编码。

### (可选)步骤4:完成分组设置

分组设置提供标签、资源组等批量管理实例的方式。完成分组设置后,单击下一步:确认订单。

1. 配置标签。

标签由一对键值(Key-Value)组成。使用标签标识具有相同特征的资源后,例如所属组织或用途相同 的资源,您可以基于标签方便地检索和管理资源。更多信息,请参见<mark>标签概述</mark>。

选择已有的标签,或者填写标签键和标签值即时创建标签。

2. 选择资源组。

资源组供您从业务角度管理跨地域、跨产品的资源,并支持针对资源组管理权限。更多信息,请参见<mark>资</mark> <mark>源组</mark>。

3. 选择部署集。

部署集支持高可用策略,部署集内实例会严格分散在不同的物理服务器上,保证业务的高可用性和底层 容灾能力。 4. 选择专有宿主机。

专有宿主机是一台由单租户独享物理资源的云主机,具有满足严格的安全合规要求、允许自带许可证 (BYOL)上云等优势。

选择已有的专有宿主机,或者单击**创建专有宿主机**即时创建专有宿主机。创建完成后,返回ECS实例创 建向导并单击 ③ 按钮,查看专有宿主机列表。具体操作,请参见创建DDH。

5. 选择私有池。

创建弹性保障或容量预定后,系统会自动生成私有池,预留特定属性特定数量的实例。从关联的私有池 中创建这一类实例,可以提供资源确定性保障。更多信息,请参见资源保障概述。

⑦ 说明 弹性保障和容量预定仅支持为按量付费实例保障资源供应确定性。

| 私有池容量 | 说明  |
|-------|---|
| 开放    | 优先使用开放类型私有池的容量,如果开放类型私有池无可用容量,则尝<br>试使用公共池的容量。        |
| 不使用   | 不使用任何私有池的容量。  |
| 指定    | 继续指定一个专用或开放类型私有池的ID,使用其容量创建实例。如果该<br>私有池没有可用容量,则创建失败。 |

# 步骤5: 确认订单

在最终创建实例前,请检查实例的整体配置并配置使用时长等选项,确保各项配置符合您的要求。

1. 检查所选配置。

如需修改配置,单击 
图标前往对应页面。您可以基于所选配置生成模板,方便后续快捷创建配置类似的实例,如下表所示。

| 快捷操作                 | 说明   | 相关文档   |
|----------------------|--|--|
| 保存为启动模板              | 将所选配置保存为启动模板,使用启动模板创建实例<br>可以免去重复配置的操作。      | 使用实例启动模板创建实<br>例   |
| 生成Open API最佳实<br>践脚本 | 自动生成API工作流和SDK示例供您参考。                        | <ul> <li>RunInstances</li> <li>批量创建ECS实例<br/>(Java)</li> <li>批量创建ECS实例<br/>(Python)</li> </ul> |
| 保存当前购买配置为<br>ROS模板   | 将所选配置保存为ROS模板,方便继续使用ROS模板创<br>建资源栈,实现一键交付资源。 | 创建资源栈  |

2. 配置实例的使用时长。

- 按量付费实例:设置自动释放实例的时间。您也可以在创建实例后手动释放实例,或者配置自动释放。具体操作,请参见释放实例。
- 包年包月实例:选择购买时长以及是否启用自动续费。您也可以在创建实例后手动续费实例,或者配置自动续费。具体操作,请参见续费概述。
- 3. 阅读*云服务器ECS服务条款*,如无疑问,选中《**云服务器ECS服务条款》**。
- 4. 在页面底部查看实例的总费用,如无疑问,确认创建实例并按提示完成支付。

# 相关文档

# 相关文档

- RunInstances
- 在GPU计算型实例中安装GPU驱动(Linux)
- 在GPU计算型实例中安装GPU驱动(Windows)

# 4.连接GPU实例

# 4.1. 连接方式概述

GPU实例作为云服务器ECS的一类实例规格,保持了与连接ECS实例相同的相关操作。ECS支持通过多种方式 连接实例,包括Workbench、VNC和第三方客户端工具。您可以综合考虑目标实例的操作系统、本地设备的 操作系统、需要执行的操作等因素,选择合适的连接方式。

# 连接方式介绍

| 实例操作系统 | 本地设备操作系统                 | 连接方式   |  |
|--------|--------------------------|--|--|
|        | Windows                  | <ul> <li>Workbench</li> <li>登录凭证为用户名密码或密钥,请参见通过密码或密钥认证登录<br/>Linux实例。</li> <li>登录凭证为PAM,请参见通过PAM凭据认证登录实例。</li> <li>VNC<br/>具体操作,请参见通过密码认证登录Linux实例。</li> <li>PuTTY等客户端工具</li> <li>登录凭证为SSH密钥对,请参见在Windows环境中使用密钥<br/>对。</li> <li>登录凭证为用户名密码,请参见在Windows环境中使用用户名<br/>密码验证。</li> </ul>               |  |
| Linux  | Linux、macOS等类<br>Unix系统  | <ul> <li>Workbench</li> <li>登录凭证为用户名密码或密钥,请参见通过密码或密钥认证登录<br/>Linux实例。</li> <li>登录凭证为PAM,请参见通过PAM凭据认证登录实例。</li> <li>VNC<br/>具体操作,请参见通过密码认证登录Linux实例。</li> <li>SSH命令连接</li> <li>登录凭证为SSH密钥对,请参见在支持SSH命令的环境中使用密<br/>钥对(通过命令配置信息)。</li> <li>登录凭证为用户名密码,请参见在Linux或Mac OS X环境中使用<br/>用户名密码验证。</li> </ul> |  |
|        | iO5、Android等移<br>动设备操作系统 | SSH Control Lite、JuiceSSH等App<br>具体操作,请参见 <mark>在移动设备上连接Linux实例</mark> 。   |  |
|        |                          |  |  |

| 实例操作系统                            | 本地设备操作系统  | 连接方式   |  |
|-----------------------------------|---|--|--|
| Windows .<br>Windows .<br>Macos . | Windows   | <ul> <li>Workbench</li> <li>登录凭证为用户名密码或密钥,请参见通过密码或密钥认证登录<br/>Windows实例。</li> <li>登录凭证为PAM,请参见通过PAM凭据认证登录实例。</li> <li>VNC<br/>具体操作,请参见通过密码认证登录Windows实例。</li> <li>远程桌面连接(MSTSC)等客户端工具<br/>具体操作,请参见本地设备使用Windows操作系统。</li> </ul> |  |
|                                   | Linux   | <ul> <li>Workbench</li> <li>登录凭证为用户名密码或密钥,请参见通过密码或密钥认证<br/>Windows实例。</li> <li>登录凭证为PAM,请参见通过PAM凭据认证登录实例。</li> <li>VNC<br/>具体操作,请参见通过密码认证登录Windows实例。</li> <li>rdesktop等客户端工具<br/>具体操作,请参见本地设备使用Linux操作系统。</li> </ul>          |  |
|                                   | <ul> <li>Workbench</li> <li>登录凭证为用户名密码或密钥,请参见通过密码或密钥认证登录<br/>Windows实例。</li> <li>登录凭证为PAM,请参见通过PAM凭据认证登录实例。</li> <li>VNC<br/>具体操作,请参见通过密码认证登录Windows实例。</li> <li>Microsoft Remote Desktop Connection for Mac等客户端工具<br/>具体操作,请参见微软官网文档。</li> </ul> |  |  |
|                                   | iO5、Android等移<br>动设备操作系统  | Microsoft Remote Desktop等App<br>具体操作,请参见 <mark>在移动设备上连接Windows实例</mark> 。  |  |

### ? 说明

- 除Workbench和VNC外,其他连接方式均要求待连接实例分配了固定公网IP或EIP。
- Windows实例创建成功后,实例的操作系统内部需要进行初始化,一般需要2~3分钟,初始化期间请不要重启实例。如果您创建了非I/O优化的Windows实例,需要10分钟完成初始化,请等待初始化完成后再连接实例。

# 连接方式对比

#### Workbench、VNC和其他第三方客户端工具均有各自的优势,对比如下。

| 对比项目                 | Workbench                                      | VNC  | 第三方客户端工具                                  |
|----------------------|--|--|---|
| 实例分配固定公网IP或EIP       | <b>可选</b><br>⑦ 说明 不能用<br>于排查误开启防火墙<br>等网络配置问题。 | 可选,因此可以用于排查<br>网络配置错误等异常,例<br>如误开启了防火墙。    | 必需  |
| 实例开启SSH等服务           | 必需   | 可选,因此可以用于排查<br>SSH服务异常等异常,例<br>如SSHD服务被关闭。 | 必需  |
| 控制台一键登录              | 支持   | 支持   | 不支持 <i>,</i> 需要安装本地客<br>户端。               |
| 目标实例操作系统无关性          | 支持连接Linux实例和<br>Windows实例。                     | 支持连接Linux实例和<br>Windows实例。                 | 一般支持连接Linux实例或<br>Windows实例,视客户端<br>工具而定。 |
| 多操作系统用户登录同一<br>台实例   | 支持   | 不支持  | 视客户端工具而定                                  |
| 交互便利性                | 交互便利 <i>,</i> 支持复制粘贴<br>文本。                    | 不支持直接复制粘贴文<br>本,需要借助复制长命令<br>功能。           | 视客户端工具而定                                  |
| 可视化查看Linux系统文件<br>资源 | 支持   | 不支持  | 视客户端工具而定                                  |
| 硬件控制和修改权限            | 不支持  | 支持,可以用于操作BIOS<br>等资源,排查操作系统无<br>法启动等异常。    | 不支持                                       |
| 终端可配置                | 支持,但依赖Workbench<br>开放的能力。                      | 不支持  | 支持 <i>,</i> 具体能力视客户端<br>工具而定。             |

# 4.2. 使用Workbench连接实例

# 4.2.1. 通过密码或密钥认证登录Linux实例

相比VNC,Workbench支持多用户远程连接同一台实例,并且支持可视化管理Linux实例中的文件,更加高效 便捷。

### 前提条件

- 实例已设置登录密码或者绑定密钥对。
- 实例处于**运行中**状态。
- 实例所在安全组已允许Workbench服务相关的IP访问实例,安全组规则详情和网络类型有关。更多信息, 请参见Workbench相关安全组规则详情。

# 背景信息

Workbench的远程连接会话默认维持6个小时,如果您超过6小时没有任何操作,连接会自动断开,您需要重新连接实例。

使用Workbench连接实例时,支持以下协议:

- 终端连接(SSH)协议:连接Linux实例时默认使用的协议,也支持连接安装了仿GNU系统(例如Cygwin)的Windows实例。关于如何连接Linux实例,请参见基于终端连接(SSH)协议连接Linux实例。
- 远程桌面(RDP)协议:连接Windows实例时默认使用的协议,也支持连接开启了远程桌面服务的Linux实例。关于如何连接Linux实例,请参见基于远程桌面(RDP)协议连接Linux实例。

⑦ 说明 如果通过远程桌面(RDP)协议连接实例,请保证公网带宽大于等于5 Mbit /s, 否则远程 桌面会卡顿。

Workbench还支持可视化管理Linux实例中的文件。具体操作,请参见使用Workbench管理Linux实例文件。

# 基于终端连接(SSH)协议连接Linux实例

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像>实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 在实例列表页面,找到需要连接的实例,单击该实例对应操作列下的远程连接。
- 5. 在弹出的远程连接与命令对话框中,单击Workbench远程连接对应的立即登录。
- 6. 在弹出的登录实例对话框中, 输入信息。

一般情况下按精简选项输入信息即可,如下表所示。

| 配置项       | 说明  |
|-----------|---|
| 实例        | 自动填充当前实例的信息,您也可以手动输入其他实例的IP或名称。   |
| 网络连接      | <ul><li>专有网络实例支持选择公网IP或私网IP连接。</li><li>经典网络实例支持选择公网IP或内网IP连接。</li></ul>                             |
| 用户名、密码、密钥 | 输入用户名(例如root),并选择认证方式。支持的认证方式如下:<br>• 密码认证:需要继续输入登录密码。<br>• 证书认证:需要继续输入或上传证书。如果证书已加密,还需要输入密<br>钥口令。 |

#### 在对话框底部单击完整选项可以展开更多配置项,如下表所示。

| 配置项  | 说明                              |
|------|---------------------------------|
| 资源组  | 默认为 <b>全部</b> ,即可以手动选择任一资源组的资源。 |
| 区域   | 默认为 <b>全部</b> ,即可以手动选择任一地域的资源。  |
| 连接协议 | 默认为 <b>终端连接(SSH)</b> 。          |

| 配置项  | 说明  |
|------|---|
| 端口   | 连接协议为 <b>终端连接(SSH)</b> 时,默认端口为22。                                     |
| 语言环境 | 偏好语言影响输出的内容,选择 <b>默认</b> 时,Workbench自动探测您远程主机<br>的语言设置并进行合适的配置。       |
| 字符集  | 偏好字符集影响输出内容的显示结果,选择 <b>默认</b> 时,Workbench自动探测<br>您远程主机的字符集设置并进行合适的配置。 |

7. 单击确定。

如果确定已满足本文的前提条件,连接实例时仍失败,请检查实例内部配置是否满足要求:

- 开启SSHD的远程服务,例如Linux系统中的SSHD服务。
- 开放终端连接端口,通常为22端口。
- 如果使用root用户登录Linux实例,需要保证在/etc/ssh/sshd\_config文件中配置 PermitRootLogin yes
   和 PasswordAuthentication yes ,具体操作请参见为Linux实例开启root用户远程登录。

### 基于远程桌面(RDP)协议连接Linux实例

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择**实例与镜像 > 实例**。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 在实例列表页面,找到需要连接的实例,单击该实例对应操作列下的远程连接。
- 5. 在弹出的远程连接与命令对话框中,单击Workbench远程连接对应的立即登录。
- 6. 在弹出的登录实例对话框, 输入信息。
  - i. 在对话框底部单击完整选项。
  - ii. 在连接协议区域,选择远程桌面(RDP)。
  - iii. 在弹出的对话框中, 单击确定。
  - iv. 输入其他信息。

| 配置项    | 说明  |
|--------|---|
| 资源组    | 默认为 <b>全部</b> ,即可以手动选择任一资源组的资源。   |
| 区域     | 默认为 <b>全部</b> ,即可以手动选择任一地域的资源。  |
| 实例     | 自动填充当前实例的信息,您也可以手动输入其他实例的IP或名称。   |
| 网络连接   | <ul><li>专有网络实例支持选择公网IP或私网IP连接。</li><li>经典网络实例支持选择公网IP或内网IP连接。</li></ul> |
| 端口     | 连接协议为 <b>远程桌面(RDP)</b> 时,默认端口为3389。                                     |
| 用户名、密码 | 输入用户名(例如Administrator)和密码。  |

7. 单击确定。

如果确定已满足本文的前提条件,连接实例时仍失败,请检查实例内部配置是否满足要求:

- 开启远程桌面服务,例如Linux系统中自行安装的XFreeRDP服务。
- 开放远程桌面端口,通常为3389端口。
- 如果使用root用户登录Linux实例,需要保证在/etc/ssh/sshd\_config文件中配置 PermitRootLogin yes
   和 PasswordAuthentication yes ,具体操作请参见为Linux实例开启root用户远程登录。

## 为Linux实例开启root用户远程登录

部分Linux系统中,SSHD服务默认禁用root用户远程登录,导致登录时提示用户名或密码错误。您可以按照 以下步骤开启root用户远程登录。

- 1. 通过密码认证登录Linux实例(VNC)。
- 2. 打开SSH配置文件。

vi /etc/ssh/sshd\_config

- 3. 修改如下参数。
  - PermitRootLogin no 修改为 PermitRootLogin yes 。
  - PasswordAuthentication no 修改为 PasswordAuthentication yes 。
- 4. 按Esc键, 输入:wq保存修改。
- 5. 重启SSHD服务。

service sshd restart

### Workbench相关安全组规则详情

Workbench相关安全组规则详情和网络类型有关:

如需连接专有网络实例,请在安全组规则页面的入方向页签添加安全组规则,详情如下表所示。

| 授权策略 | 优先级 | 协议类<br>型   | 端口范围   | 授权对象  |
|------|-----|------------|--|---|
| 允许   | 1   | 自定义<br>TCP | <ul> <li>○ Linux实例默认开放22端口:选择SSH(22)。</li> <li>○ 加思你天动业的之其他端口,天</li> </ul> | <ul> <li>如果通过实例的公网IP(包括固定公网IP和EIP)连接:添加47.96.60.0/24和118.31.243.0/24。</li> <li>如果通过实例的专有网络私网IP连接:添加100.104.0.0/16。</li> </ul> |
|      | ТСР | 动输入端口范围。   | <ul> <li>⑦ 说明 您也可以添加</li> <li>0.0.0.0/0,但存在安全风险,<br/>请谨慎使用。</li> </ul>     |   |

 如需通过公网连接经典网络实例,请在安全组规则页面的公网入方向页签添加安全组规则,详情如下表 所示。

| 授权策略 | 优先级 | 协议类<br>型    | 端口范围   | 授权对象  |
|------|-----|-------------|--|---|
| 允许   | 1   | 自定义<br>T CP | <ul> <li>Linux实例默认开放22端口:选择SSH(22)。</li> <li>如果您手动开放了其他端口:手动输入端口范围。</li> </ul> | 通过实例的公网IP(包括固定公网IP<br>和EIP)连接:添加47.96.60.0/24<br>和118.31.243.0/24。<br>⑦ 说明 您也可以添加<br>0.0.0.0/0,但存在安全风险,<br>请谨慎使用。 |

如需通过内网连接经典网络实例,请在安全组规则页面的入方向页签添加安全组规则,详情如下表所示。

| 授权策略 | 优先级 | 协议类<br>型   | 端口范围   | 授权对象  |
|------|-----|------------|--|---|
| 允许   | 1   | 自定义<br>TCP | <ul> <li>Linux实例默认开放22端口:选择SSH(22)。</li> <li>如果您手动开放了其他端口:手动输入端口范围。</li> </ul> | 通过实例的经典网络内网IP连接:<br>添加11.195.184.0/24和<br>11.246.55.0/24。 |

### 常见问题

无法连接实例时,您可以自行排查,详情请参见远程连接FAQ和GuestOS常见问题与修复方案。

# 4.2.2. 通过密码或密钥认证登录Windows实例

相比VNC, Workbench支持多用户远程连接同一台实例,更加高效便捷。

# 前提条件

• 实例已设置登录密码或者配置密钥。

⑦ 说明 控制台中不支持为Windows实例绑定密钥对,如需通过密钥认证登录Windows实例,您可以自行在实例内部开启SSHD的远程服务并配置密钥,例如Windows系统中的Cygwin SSHD服务或WinSSHD服务。关于Windows如何开启SSHD服务的具体操作,请参见微软官方文档。

- 实例处于运行中状态。
- 实例所在安全组已允许Workbench服务相关的IP访问实例,安全组规则详情和网络类型有关。更多信息, 请参见使用Workbench连接Windows实例时相关安全组规则的详情。

# 背景信息

Workbench的远程连接会话默认维持6个小时,如果您超过6小时没有任何操作,连接会自动断开,您需要重新连接实例。

使用Workbench连接实例时,支持以下协议:

远程桌面(RDP)协议:连接Windows实例时默认使用的协议,也支持连接开启了远程桌面服务的Linux实例。关于如何连接Windows实例,请参见基于远程桌面(RDP)协议连接Windows实例。

⑦ 说明 如果通过远程桌面(RDP)协议连接实例,请保证公网带宽大于等于5 Mbit/s,否则远程 桌面会卡顿。

终端连接(SSH)协议:连接Linux实例时默认使用的协议,也支持连接安装了仿GNU系统(例如Cygwin)的Windows实例。关于如何连接Windows实例,请参见基于终端连接(SSH)协议连接Windows实例。

基于远程桌面(RDP)协议连接Windows实例

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像>实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 在实例列表页面,找到需要连接的实例,单击该实例对应操作列下的远程连接。
- 5. 在弹出的远程连接与命令对话框中,单击Workbench远程连接对应的立即登录。
- 6. 在弹出的登录实例对话框中, 输入信息。

一般情况下按精简选项输入信息即可,如下表所示。

| 配置项    | 说明  |
|--------|---|
| 实例     | 自动填充当前实例的信息,您也可以手动输入其他实例的IP或名称。   |
| 网络连接   | <ul><li>专有网络实例支持选择公网IP或私网IP连接。</li><li>经典网络实例支持选择公网IP或内网IP连接。</li></ul> |
| 用户名、密码 | 输入用户名(例如Administrator)和登录密码。  |

#### 在对话框底部单击完整选项可以展开更多配置项,如下表所示。

| 配置项  | 说明                                  |
|------|-------------------------------------|
| 资源组  | 默认为 <b>全部</b> ,即可以手动选择任一资源组的资源。     |
| 区域   | 默认为 <b>全部</b> ,即可以手动选择任一地域的资源。      |
| 连接协议 | 默认为 <b>远程桌面(RDP)</b> 。              |
| 端口   | 连接协议为 <b>远程桌面(RDP)</b> 时,默认端口为3389。 |

7. 单击确定。

如果确定已满足本文的前提条件,连接实例时仍失败,请检查实例内部配置是否满足要求:

- 开启远程桌面服务,例如Windows系统中的远程桌面服务。
- 开放远程桌面端口,通常为3389端口。
- 如果使用非管理员用户登录Windows实例,该用户需要属于Remote Desktop Users组。

### 基于终端连接(SSH)协议连接Windows实例

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像>实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 在实例列表页面,找到需要连接的实例,单击该实例对应操作列下的远程连接。
- 5. 在弹出的远程连接与命令对话框中,单击Workbench远程连接对应的立即登录。
- 6. 在弹出的登录实例对话框, 输入信息。
  - i. 在对话框底部单击完整选项。
  - ii. 在连接协议区域,选择终端连接(SSH)。
  - iii. 在弹出的对话框中, 单击确定。
  - iv. 输入其他信息。

| 配置项       | 说明   |
|-----------|--|
| 资源组       | 默认为 <b>全部</b> ,即可以手动选择任一资源组的资源。  |
| 区域        | 默认为 <b>全部</b> ,即可以手动选择任一地域的资源。   |
| 实例        | 自动填充当前实例的信息,您也可以手动输入其他实例的IP或名称。  |
| 网络连接      | <ul><li>专有网络实例支持选择公网IP或私网IP连接。</li><li>经典网络实例支持选择公网IP或内网IP连接。</li></ul>  |
| 端口        | 连接协议为 <b>终端连接(SSH)</b> 时,默认端口为22。  |
| 用户名、密码、密钥 | 输入用户名(例如root),并选择认证方式。支持的认证方式如下: <ul> <li>密码认证:需要继续输入登录密码。</li> <li>证书认证:需要继续输入或上传证书。如果证书已加密,还需要输入密钥口令。</li> </ul> |
| 语言环境      | 偏好语言影响输出的内容,选择 <b>默认</b> 时,Workbench自动探测您远程<br>主机的语言设置并进行合适的配置。  |
| 字符集       | 偏好字符集影响输出内容的显示结果,选择 <b>默认</b> 时,Workbench自动<br>探测您远程主机的字符集设置并进行合适的配置。  |

### 7. 单击确定。

如果确定已满足本文的前提条件,连接实例时仍失败,请检查实例内部配置是否满足要求:

- 开启SSHD的远程服务,例如Windows系统中的Cygwin SSHD服务或WinSSHD服务。
- 开放终端连接端口,通常为22端口。
- 如果使用非管理员用户登录Windows实例,该用户需要属于Remote Desktop Users组。

# 使用Workbench连接Windows实例时相关安全组规则的详情

安全组规则详情和网络类型有关:

### • 如需连接专有网络实例,请在**安全组规则**页面的入方向页签添加安全组规则,如下表所示。

| 授权策略 | 优先级          | 协议类<br>型 | 端口范围   | 授权对象  |
|------|--------------|----------|--|---|
| 允许   | 1 自定义<br>TCP | 自定义      | <ul> <li>○ Windows实例默认开放3389端</li> <li>□:选择RDP(3389)。</li> <li>○ 加思你表动立立之其他端口:表</li> </ul> | <ul> <li>如果通过实例的公网IP(包括固定公网IP和EIP)连接:添加47.96.60.0/24和118.31.243.0/24。</li> <li>如果通过实例的专有网络私网IP连接:添加100.104.0.0/16。</li> </ul> |
|      |              | 动输入端口范围。 | <ul> <li>⑦ 说明 您也可以添加</li> <li>0.0.0.0/0,但存在安全风险,</li> <li>请谨慎使用。</li> </ul>                |   |

# 如需通过公网连接经典网络实例,请在安全组规则页面的公网入方向页签添加安全组规则,如下表所示。

| 授权策略 | 优先级 | 协议类<br>型    | 端口范围   | 授权对象  |
|------|-----|-------------|--|---|
| 允许   | 1   | 自定义<br>T CP | <ul> <li>Windows实例默认开放3389端<br/>口:选择RDP(3389)。</li> <li>如果您手动开放了其他端口:手<br/>动输入端口范围。</li> </ul> | 通过实例的公网IP(包括固定公网IP<br>和EIP)连接:添加47.96.60.0/24<br>和118.31.243.0/24。<br>⑦ 说明 您也可以添加<br>0.0.0.0/0,但存在安全风险,<br>请谨慎使用。 |

• 如需通过内网连接经典网络实例,请在**安全组规则**页面的入方向页签添加安全组规则,如下表所示。

| 授权策略 | 优先级 | 协议类<br>型 | 端口范围  | 授权对象  |
|------|-----|----------|---|---|
|      |     | 白豆♡      | ○ Windows实例默认开放3389端<br>□:选择 <b>RDP(3389)</b> 。 | 通过实例的经典网络内网IP连接:<br>添加11.195.184.0/24和<br>11.246.55.0/24。             |
| 允许   | 1   | TCP      | <ul> <li>如果您手动开放了其他端口:手<br/>动输入端口范围。</li> </ul> | <ul> <li>✓ 注意 为内网入方向规则添加0.0.0.0/0存在高安全风</li> <li>№, 不建议使用。</li> </ul> |

# 常见问题

无法连接实例时,您可以自行排查,详情请参见远程连接FAQ和GuestOS常见问题与修复方案。

# 4.3. 使用VNC连接实例

# 4.3.1. 通过密码认证登录Linux实例

无法使用Workbench和远程连接软件(例如PuTTY、Xshell、SecureCRT等)连接Linux实例时,您可以通过 控制台的VNC远程连接登录实例,查看云服务器操作界面的实时状态。

## 前提条件

实例已设置登录密码。

⑦ 说明 如果您未设置或者忘记了登录密码,可以为实例重置密码。具体操作,请参见重置实例登录 密码。

### 背景信息

使用VNC连接实例时涉及以下密码:

- 远程连接密码(VNC密码):用于连接ECS管理控制台的管理终端。
- 实例登录密码:用于登录实例操作系统。

VNC的远程连接会话默认维持300秒左右,如果您超过300秒左右没有任何操作,连接会自动断开,您需要重新连接实例。

在无法使用Workbench和远程连接软件登录实例的场景中,您可以通过VNC远程连接登录实例,了解实例 情况并解决问题,如下表所示。

| 场景               | 连接实例后的操作  |
|------------------|-----------|
| 实例启动自检等原因导致启动速度慢 | 查看自检进度    |
| 实例误开启操作系统防火墙     | 关闭防火墙     |
| 云服务器失陷,CPU和带宽消耗高 | 排查并结束异常进程 |

## 操作步骤

操作流程图如下图所示。



- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像 > 实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 在实例列表页面,找到需要连接的实例,在操作列中,单击远程连接。
- 5. 在弹出的远程连接与命令对话框中,单击VNC远程连接对应的立即登录。

6. 连接VNC管理终端。

⑦ 说明 此步骤中使用VNC密码。

- 首次连接VNC管理终端:
  - a. 修改VNC密码。
  - b. 在输入VNC密码对话框中, 输入新修改的密码。
  - c. 单击确定。
- 非首次连接VNC管理终端:
  - a. 在输入VNC密码对话框中, 输入密码。
  - b. 单击确定。
- 7. 登录实例操作系统。

? 说明 此步骤中使用实例登录密码。

- i. 输入用户名root, 然后按回车键。
- ii. 输入实例登录密码, 然后按回车键。

⑦ 说明 输入完成后按回车键即可,登录Linux实例时界面不会显示密码的输入过程。

支持切换不同的VNC管理终端连接Linux实例,默认为CTRL+ALT+F1,最多支持10个。例如,单击 界面左上角的**发送远程命令 > CTRL+ALT+F2**,切换至CTRL+ALT+F2。黑屏可能是因为Linux实 例处于休眠状态,单击键盘上任意键即可唤醒实例。

以下示例演示如何通过VNC连接Linux实例。

|                     | 部资源 ▼ 华东1 (杭州) ▼                            |                          |                                      | Q                            | 建重文档、控制台、API、解决方面和资源                                | 奠用 | 工準             | <b>新家</b> 企业 | 支持    | 宣网      |      | д. д | 1           | R(X 📀              |
|---------------------|---|--------------------------|--------------------------------------|------------------------------|---|----|----------------|--------------|-------|---------|------|------|-------------|--------------------|
| 云服勞闢 ECS            | 子账号设置全局过滤标签,可以更加方便的查看和管                     | 理已被授权的云资源。设置             |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 獨克                  | 实例列表  |                          |                                      |                              |   |    |                |              | đ     | ) 实例使用( | 90 C | 9523 | 691 - E     | 北京編作               |
|                     |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
|                     |   | 2 O                      |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 自动化部署 ROS 🖸 🚥       | ▼ 检索项: 实例名称 jsy × 清除                        |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 容離服务 Kubernetes 🖸 🎫 | 实例D/名称                                      | 标签 监控 可用区 ▼              | IP地址                                 | 状志 🖌 🛛 网络美型 🗸                |   |    | 付费方式           | *            |       |         |      |      |             | 操作                 |
|                     | i-hg1-kQuebefj1ytosnii9<br>jsyleni i intOS6 | ※                        | 121-11166.158(公)<br>19011-06.031(私街) | ●运行中 专有网络                    | 2 vCPU 4 GiB (I/O优化)<br>ecs.c6e.large 1Mbps (細道)    |    | 包年包月<br>2020年8 | 月1日 23:59 到  | 網     |         |      | 83   | 過程注格<br>  換 | 及  升時配<br>器   更多 ▼ |
| <b>买例与現象</b> へ      | i-tup (* 111) Izhihbold78ps                 | ● <sup>○</sup> △ 华东1可用区8 | 1921162.22.94 (私有)                   | <ul> <li>送行中 专有网络</li> </ul> | 1 vCPU 1 GiB (I/O优化)<br>ers yo4 small (Miltons (時間) |    | 包年包月<br>2020年8 | B1円 23-50 副  | 10    |         |      | 15   | 近程注意<br>(数) | ●  升降配             |
| \$\$M (-            | 1974000 0                                   | •                        |                                      |                              | (max)   |    |                |              | ~     |         |      |      | _           |                    |
| 弹性容器实例 ECI II       | 自动 停止 重启 重置实例结                              | 两 续费 按量付费转包年包月 释         | 放设置 更多▲                              |                              |   |    |                |              | 共有2条。 | 每页显示    | 20 🗸 | e e  | c 1.        | > >                |
| 专有宿主机 DDH           |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 超级计算集群              |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 预留实例券               |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 現像                  |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 部署与弹性 >             |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 存储与快照 🚥 🗸 🗸         |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             |                    |
| 网络与安全 🗸             |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             | 9                  |
| 运维与监控 🗸             |   |                          |                                      |                              |   |    |                |              |       |         |      |      |             | 88                 |
| 常用服务推荐              |   | 1、选持                     | 锋 <b>实例,通</b> 过                      | VNC进行过                       | 远程连接  |    |                |              |       |         |      |      |             |                    |

# 修改VNC密码

第一次连接VNC管理终端时,您需要修改VNC密码。您也可以在忘记或者需要更新VNC密码时使用修改功 能。 ↓ 注意 修改非I/O优化实例的VNC密码后,您必须在ECS管理控制台重启实例使新密码生效。重启操 作需要停止您的实例,导致业务中断,请谨慎执行。

- 1. 在实例列表页面,找到需要连接的实例,在操作列中,单击远程连接。
- 2. 在弹出的远程连接与命令对话框中,单击VNC远程连接对应的立即登录。
- 3. 在输入VNC密码对话框中,单击修改VNC密码。
- 4. 在修改VNC密码对话框中, 输入并确认新密码, 然后单击确定。
- (可选)如果实例为非I/O优化实例,重启实例。
   具体操作,请参见重启实例。

### 复制长命令

如果需要从本地设备向实例拷贝较长的文本,例如文件下载地址,您可以使用复制命令输入功能。

- 1. 在实例列表页面,找到需要连接的实例,在操作列中,单击远程连接。
- 2. 在弹出的远程连接与命令对话框中,单击VNC远程连接对应的立即登录。
- 3. 连接VNC管理终端。
- 4. 在界面左上角, 单击复制命令输入。
- 5. 在复制粘贴命令对话框中,输入待拷贝的内容, 然后单击确定。

### 常见问题

无法连接实例时,您可以自行排查,详情请参见远程连接FAQ和GuestOS常见问题与修复方案。

# 4.3.2. 通过密码认证登录Windows实例

无法使用Workbench和远程连接软件(例如RDP、rdesktop等)连接Windows实例时,您可以通过控制台的VNC远程连接登录实例,查看云服务器操作界面的实时状态。

### 前提条件

实例已设置登录密码。

⑦ 说明 如果您未设置或者忘记了登录密码,可以为实例重置密码。具体操作,请参见重置实例登录 密码。

# 背景信息

使用VNC连接实例时涉及以下密码:

- 远程连接密码(VNC密码):用于连接ECS管理控制台的管理终端。
- 实例登录密码:用于登录实例操作系统。

VNC的远程连接会话默认维持300秒左右,如果您超过300秒左右没有任何操作,连接会自动断开,您需要重新连接实例。

在无法使用Workbench和远程连接软件登录实例的场景中,您可以通过**VNC远程连接**登录实例,了解实例 情况并解决问题,如下表所示。

| 场景               | 连接实例后的操作  |
|------------------|-----------|
| 实例启动自检等原因导致启动速度慢 | 查看自检进度    |
| 实例误开启操作系统防火墙     | 关闭防火墙     |
| 云服务器失陷,CPU和带宽消耗高 | 排查并结束异常进程 |

# 操作步骤

操作流程图如下图所示。



### 1. 登录ECS管理控制台。

2. 在左侧导航栏,选择实例与镜像 > 实例。

- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 在实例列表页面,找到需要连接的实例,在操作列中,单击远程连接。
- 5. 在弹出的远程连接与命令对话框中,单击VNC远程连接对应的立即登录。
- 6. 连接VNC管理终端。

⑦ 说明 此步骤中使用VNC密码。

- 首次连接VNC管理终端:
  - a. 修改VNC密码。
  - b. 在输入VNC密码对话框中, 输入新修改的密码。
  - c. 单击**确定**。
- 非首次连接VNC管理终端:
  - a. 在输入VNC密码对话框中, 输入密码。
  - b. 单击**确定**。
- 7. 在管理终端页面的左上角, 单击发送远程命令 > CTRL+ALT+DELETE。

| 连接VNC           |  |
|-----------------|--|
|                 |  |
| 断开VNC           |  |
| CTRL+ALT+DELETE |  |
| CTRL+ALT+F1     |  |
| CTRL+ALT+F2     |  |
| CTRL+ALT+F3     |  |
| CTRL+ALT+F4     |  |
| CTRL+ALT+F5     |  |
| CTRL+ALT+F6     |  |
| CTRL+ALT+F7     |  |
| CTRL+ALT+F8     |  |
| CTRL+ALT+F9     |  |
| CTRL+ALT+F10    |  |

 选择用户账户,输入实例登录密码,然后单击登录图标。 默认有Administrator账户。

### 修改VNC密码

第一次连接VNC管理终端时,您需要修改VNC密码。您也可以在忘记或者需要更新VNC密码时使用修改功能。

↓ 注意 修改非I/O优化实例的VNC密码后,您必须在ECS管理控制台重启实例使新密码生效。重启操 作需要停止您的实例,导致业务中断,请谨慎执行。

- 1. 在实例列表页面,找到需要连接的实例,在操作列中,单击远程连接。
- 2. 在弹出的远程连接与命令对话框中,单击VNC远程连接对应的立即登录。
- 3. 在输入VNC密码对话框中,单击修改VNC密码。
- 4. 在修改VNC密码对话框中,输入并确认新密码,然后单击确定。
- (可选)如果实例为非I/O优化实例,重启实例。
   具体操作,请参见重启实例。

# 复制长命令

如果需要从本地设备向实例拷贝较长的文本,例如文件下载地址,您可以使用复制命令输入功能。

- 1. 在实例列表页面,找到需要连接的实例,在操作列中,单击远程连接。
- 2. 在弹出的远程连接与命令对话框中,单击VNC远程连接对应的立即登录。
- 3. 连接VNC管理终端。
- 4. 在界面左上角, 单击复制命令输入。
- 5. 在复制粘贴命令对话框中, 输入待拷贝的内容, 然后单击确定。

# 常见问题

- 无法连接实例时,您可以自行排查,详情请参见远程连接FAQ和GuestOS常见问题与修复方案。
- 如果您需要调整Windows桌面的分辨率,请参见如何调整Windows桌面的分辨率?。

# 5.管理GPU实例

# 5.1. 停止实例

GPU实例作为云服务器ECS的一类实例规格,保持了与管理ECS实例相同的相关操作。本文介绍停止ECS实例,以及VPC内实例节省停机模式相关的操作。

### 前提条件

实例必须处于运行中状态。

⑦ 说明 停止实例会中断您的业务,请谨慎执行。

# 背景信息

停止包年包月实例不会影响实例的计费。

停止按量付费实例可能会影响实例的计费,具体动作由是否启用了节省停机模式决定。

- 经典网络内按量付费实例不支持节省停机模式,实例停止后继续收费。只有释放实例后,实例才会停止计费。具体操作,请参见释放实例。
- VPC内按量付费实例支持节省停机模式:
  - 如果未启用节省停机模式,实例停止后继续收费。
  - 如果启用了节省停机模式,实例停止后,vCPU、内存和公网IP将不再计费,但其他资源将继续计费。更多信息,请参见按量付费实例节省停机模式。

### 停止包年包月实例

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像>实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 请根据实际情况停止实例。
  - 停止单台实例:找到需要停止的包年包月实例,在操作列,单击更多 > 实例状态 > 停止。
  - 批量停止实例:选中一台或多台需要停止的包年包月实例,并在实例列表底部,单击停止。
- 5. 设置停止方式。
  - **停止**:正常关机流程。
  - **强制停止**:等同于断电处理,可能丢失实例操作系统中未写入磁盘的数据。
- 6. 单击确定。

### 停止按量付费实例

抢占式实例的停止过程和按量付费实例一致,但停止后能否顺利再启动会受到更多因素的影响。更多信息, 请参见停止抢占式实例。

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像 > 实例。
- 3. 在顶部菜单栏左上角处,选择地域。

- 4. 请根据实际情况停止实例。
  - 停止单台实例:找到需要停止的按量付费实例,在操作列,单击更多 > 实例状态 > 停止。
  - 批量停止实例:选中一台或多台需要停止的按量付费实例,并在实例列表底部,单击停止。
- 5. 设置停止方式和停止模式。
  - 如果实例的网络类型为经典网络:
    - a. 设置停止方式。
      - 停止:正常关机流程。
      - 强制停止:等同于断电处理,可能丢失实例操作系统中未写入磁盘的数据。
    - b. 单击确定。
  - 如果实例的网络类型为专有网络:
    - a. 设置停止方式。
      - 停止:正常关机流程。
      - 强制停止:等同于断电处理,可能丢失实例操作系统中未写入磁盘的数据。
    - b. 设置停止模式。
      - 普通停机模式:停止实例后保留实例的资源等并收费。
      - 节省停机模式:停止实例后实例计算资源将不再计费(实例的vCPU和内存将不再保留);云盘(包括系统盘、数据盘)、弹性公网IP和带宽仍旧收费;固定公网IP将被回收,私网IP将保留。
    - c. 单击确定。

| 停止实例 🛛  | X   |
|---|---|
| <ul> <li>包年包月实行如果您停止部<br/>如果您停止部<br/>"停止后仍旧<br/>Windows 实<br/>待。</li> </ul> | 列停止后不会改变到期时间<br>实例是为了更换操作系统、重新初始化云盘、更改实例规格、修改私网IP等操作,建议您勾选<br>l保留实例并继续收费" 选项,避免启动失败。<br>例创建时的初始化时间较长(3~5分钟),初始化过程中请勿重启或停止实例,请耐心等  |
| 您所选的 1个实  | 例▼将执行停止操作, 您是否确定操作?   |
| 停止方式:   |   |
| 停止模式:   | 會止后仍旧保留实例并继续收费  |
|   | <ul> <li>节省停机模式(原停机不收费)</li> <li>1. 停机节省模式下,计算资源(vCPU和内存)、固定公网IP和带宽不再收费。</li> <li>2. 仍旧收费的资源有:系统盘、数据盘、弹性公网IP和带宽(固定带宽模式)、收费镜像。</li> <li>3. 由于计算资源被回收,重启后可能因为库存不足导致启动失败,您可以稍后尝试再次启动,或者尝试换用其它实例规格。</li> <li>4. 由于固定公网IP地址被回收,重启后固定公网IP地址可能变更,但不会导致私网IP地址变更。若实例停机前绑定的是弹性公网IP地址,则重启后IP地址保持不变。</li> </ul> |
|   | 确定取消  |

# 执行结果

实例停止后,进入已**停止**状态。

# 相关文档

• StopInstance

# 5.2. 启动实例

GPU实例作为云服务器ECS的一类实例规格,保持了与管理ECS实例相同的相关操作。本文介绍如何在ECS管理控制台上启动实例。

## 前提条件

实例满足以下条件之一:

- 实例处于已停止状态。
- 按量付费实例处于已过期状态,您已经结清欠费账单但自动重开机失败。

⑦ 说明 您的阿里云账户余额(即现金余额)和代金券的总值不得小于100.00元人民币。

按量付费实例处于欠费回收中状态,您已经结清欠费账单但自动重开机失败。

⑦ 说明 您的阿里云账户余额(即现金余额)和代金券的总值不得小于100.00元人民币。

### 背景信息

按量付费实例欠费停机后,如果您已结清欠费账单但未能重开机,实例仍然会被释放,请及时手动启动实例,避免实例被释放影响业务。更多信息,请参见按量付费。

按量付费实例结清欠费账单后,系统会自动重开机。自动重开机偶尔会出现失败的情况,请您关注自动重开机是否成功,实例在约10分钟内是否进入了运行中状态,如果自动重开机失败则及时手动重开机(即手动启动实例)。

⑦ 说明 欠费停机的按量付费实例启动成功后,可以正常使用实例并持续产生按量账单,请及时释放 不再使用的实例,避免花费不必要的成本。

如果手动启动实例失败,可能因为当前实例规格库存不足,请尝试修改实例规格,具体操作,请参见按量付 费实例变配规格。如果问题仍然无法解决,请提交工单。

### 操作步骤

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像 > 实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 请根据实际情况启动实例。
  - 启动单台实例:找到待启动的实例,在操作列中,选择更多 > 实例状态 > 启动。



• 批量启动实例:选中多台待启动的实例,在页面底部,单击启动。

| 实例             |                        |       |             |                |      |        |  |
|----------------|------------------------|-------|-------------|----------------|------|--------|--|
| 检查到安全组中包含允许对特定 | 2端口进行不受限制              | 削访问的规 | 观则,存在       | 潜在高危风险。查看详情    |      |        |  |
| 创建实例 选择实例属性环   | ·搜索, 或者输入÷             | (键字识别 | 搜索          | ② Q 标签         |      |        |  |
| □ 实例ID/名称      | 标签                     | 监控    | 可用区<br>₽    | IP地址           | 状态 🖌 | 网络类型 🎖 | 配置   |
| i-b            | <b>⊳ 0</b> #           |       | 杭州 可<br>用区I | 192            | ●已停止 | 专有网络   | 1 vCPU 1 GiB (I/O优化)<br>ecs.s6-c1m1.small<br>10Mbps (峰值) |
| ES:            | <ul><li>♥ ♥.</li></ul> |       | 杭州 可<br>用区I | 47.<br>17:     | ✔运行中 | 专有网络   | 2 vCPU 8 GiB (l/O优化)<br>ecs.gбr.large<br>5Mbps (峰值)      |
| ✓ i-b<br>Win   |                        |       | 杭州 可<br>用区I | 47.<br>192     | ●已停止 | 专有网络   | 1 vCPU 1 GiB (I/O优化)<br>ecs.s6-c1m1.small<br>10Mbps (峰值) |
| bw.            | <b>∿ ≎</b> \$          |       | 杭州 可<br>用区I | 172            | ●已停止 | 专有网络   | 4 vCPU 16 GiB (I/O优化)<br>ecs.g7se.xlarge<br>5Mbps (峰值)   |
| bw             | <b>∿ °</b> ‡           |       | 杭州 可<br>用区I | 121 <b>121</b> | ●已停止 | 专有网络   | 1 vCPU 1 GiB (I/O优化)<br>ecs.s6-c1m1.small 1Mbps          |
| 日 启动 停止 重      |                        | 利密码   | 续费          | 按量付费转包年包月      | 释放设置 | 更多▲    |  |

5. 在启动实例对话框中确认实例信息,并单击确定。

### 执行结果

实例启动完成后进入运行中状态。

### 相关文档

• StartInstance

# 5.3. 休眠实例

GPU实例作为云服务器ECS的一类实例规格,保持了与管理ECS实例相同的相关操作。如果您在某个时间段内 无需运行ECS实例,也不会执行某些操作(如升降配等),建议您休眠ECS实例。相比停止实例,休眠实例后 再次启动实例,ECS实例会自动恢复至休眠前的应用程序状态,便于您快速进入休眠前ECS实例的运行状态。

### 背景信息

休眠实例时,操作系统会将内存中的数据(操作系统内正在运行的各种应用程序及使用状态)保留至实例的 系统盘。当您下次启动实例时,操作系统会读取上次保留的内存数据,自动恢复至休眠前ECS实例的应用程 序状态,快速进入休眠前ECS实例的运行状态。而停止实例后再次启动实例时,操作系统会重新启动后台服 务和应用程序。

⑦ 说明 休眠失败时,会自动正常关机,即内存数据不会保存至系统盘。等实例关机后启动实例,操 作系统会重新启动后台服务和应用程序,无法恢复休眠前ECS实例的应用程序状态。

#### 在休眠时,不同购买方式的ECS实例计费有如下差异:

- 包年包月实例:休眠后不影响实例的到期时间,计费不受影响。
- 按量付费实例:取决于休眠时是否选择了休眠不收费选项,具体的区别如下表所示。

休眠实例的计费区别

| 资源项           | 休眠不收费  | 休眠后仍旧保留实例并继续收费 |
|---------------|--------|----------------|
| 计算资源(vCPU和内存) | 释放,不收费 | 保留并继续收费        |

| 资源项 | 休眠不收费 | 休眠后仍旧保留实例并继续收费 |
|-----|-------|----------------|
|-----|-------|----------------|

| 云盘(系统盘和数据盘) | 保留并继续收费                  | 保留并继续收费 |
|-------------|--------------------------|---------|
| 内网IP        | 保留,不收费                   | 保留,不收费  |
| 公网IP        | 释放,启动ECS实例后会获取新的公网IP,不收费 | 保留,不收费  |
| EIP         | 保留并继续收费                  | 保留并继续收费 |
| 带宽          | 继续收费                     | 继续收费    |

# 使用限制

- 休眠功能目前仅对美国(硅谷)、德国(法兰克福)地域开放,其他地域逐步开放中。
- 当ECS实例满足以下条件时,您才能在需要时使实例进入休眠状态:
  - 实例在创建时开启了休眠功能。

⑦ 说明 开启休眠功能后不支持关闭该功能。创建时未开启该功能的实例无法设置该功能。

- 实例安装了休眠Agent。
- 创建ECS实例时, 您只能使用加密的自定义镜像才能开启休眠功能。并且镜像版本支持以下几种:
  - 。 Windows Server 2016及以上
  - 。 Ubunt u 18及以上
  - 。 Cent OS 7及以上
- 如果ECS实例在创建时开启了休眠功能,则ECS实例不支持以下功能:
  - 。 不支持创建自定义镜像。
  - 不支持创建快照。
  - 不支持变更实例规格。
  - 不支持变更操作系统或系统盘。
  - 。 不支持变更包年包月ECS实例的带宽。
- 如果抢占式实例开启了休眠功能,在休眠时仅支持休眠不收费功能。
- 弹性伸缩组内的ECS实例不支持休眠功能。

### 步骤一:开启休眠功能

您必须在创建ECS实例时,选择开启休眠功能,否则无法使实例进入休眠状态。创建时,必须使用加密镜 像。

1. 获取加密的自定义镜像。

有两种方式获取:

- 您已准备了满足休眠要求的加密自定义镜像。
- 通过复制镜像获取,在复制过程中将镜像加密。具体步骤,请参见复制镜像。

⑦ 说明 镜像的限制,请参见使用限制。

### 2. 创建ECS实例,并在创建实例时开启休眠功能。

#### 具体操作步骤,请参见使用向导创建实例。请注意以下配置信息:

| 配置项 | 配置说明  | 举例   |
|-----|---|--|
| 实例  | <ul> <li>。 实例规格: 除裸金属服务器外的任一I/O优化 实例。</li> <li>。 实例内存大小:</li> <li>Windows: 必须小于16 GiB。</li> <li>Linux: 必须小于150 GiB。</li> </ul>  | ecs.g6e.large  |
| 镜像  | <ul> <li>选择第一步创建完成的加密自定义镜像,或者已有的满足休眠要求的加密自定义镜像。</li> <li>选中开启实例休眠状态,开启休眠功能。</li> </ul>  | <ul> <li>encrypted.windows2016</li> <li>选中开启实例休眠状态</li> </ul>  |
| 云盘  | <ul> <li>系统盘:必选项,且需要满足以下要求:</li> <li>类型:高效云盘、SSD云盘或者ESSD云盘。</li> <li>容量:必须足够大,建议至少是内存的2倍。因为,在开启休眠能力时,系统盘会预留部分空间来存放内存数据。因此,系统盘的容量需要确保在存放内存数据时,还能使操作系统和内部应用程序正常运行。</li> <li>加密:使用加密镜像后,系统盘会默认加密。</li> <li>数据盘:可选项。如果在此时创建云盘作为数据盘,必须选择云盘类型、容量、数量,并设置是否加密。</li> </ul> | <ul> <li>系统盘: ESSD云盘, 60 GiB,<br/>加密且密钥为Default<br/>Service CMK。</li> <li>数据盘: ESSD云盘, 40 GiB,<br/>未加密。</li> </ul> |
| 网络  | 请选择专有网络。  | [默认]vpc-bp1opxu1zkhn00g****  |

# 步骤二:安装休眠Agent

在ECS实例开启休眠功能后,您必须为该实例安装休眠Agent。安装完成后,您才能使该实例进入休眠状态。

- 1. 创建并执行如下命令,为单台ECS实例安装休眠Agent。具体方式,请参见立即执行命令。
  - Windows实例:

acs-plugin-manager.exe --exec --plugin ecs-hibernate-win --params "install"

○ Linux实例:

```
acs-plugin-manager --exec --plugin ecs-hibernate-linux --params "install"
```

2. 重启ECS实例,令休眠Agent生效。具体操作,请参见重启实例。

## 步骤三: 使实例进入休眠状态

当ECS实例已开启了休眠功能,并安装了休眠Agent,您可以使运行中的该实例进入休眠状态。在实例休眠 期间,您无法连接该实例。

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像>实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 找到需要休眠的实例, 在操作列, 单击更多 > 实例状态 > 停止。
- 5. 在停止实例对话框中,完成休眠相关设置。
  - i. 选择停止方式为休眠。
  - ii. 选择停止模式,您可以根据ECS实例的计费方式选择休眠后仍旧保留实例并继续收费或者休眠不收费:
    - 包年包月的ECS实例休眠后会继续收费,请选择休眠后仍旧保留实例并继续收费。
    - 按量付费ECS实例支持休眠后仍旧保留实例并继续收费或者休眠不收费,您可以自由选择。

⑦ 说明 抢占式ECS实例只支持休眠不收费,请选择休眠不收费。

休眠后仍旧保留实例并继续收费或者休眠不收费的区别,请参见休眠实例的计费区别。

#### iii. 单击确定。

⑦ 说明 实例停止,进入已停止状态。如需启动实例,具体操作,请参见启动实例。

# 相关文档

您也可以利用OOS定时休眠和启动ECS实例,自动化管理大批量实例的休眠和启动时间,配合休眠不收费功 能节省成本。具体内容,请参见定时开关机。

# 5.4. 重启实例

GPU实例作为云服务器ECS的一类实例规格,保持了与管理ECS实例相同的相关操作。本文介绍如何通过控制 台重启实例。

### 前提条件

待重启的实例必须处于运行中状态。

# 背景信息

重启操作会造成您的实例停止工作,造成业务中断,请谨慎执行。

### 操作步骤

1. 登录ECS管理控制台。

- 2. 在左侧导航栏,选择实例与镜像>实例。
- 3. 选择地域。
- 4. 找到需要重启的实例。
  - 如果重启一个实例,在操作列中,单击更多 > 实例状态 > 重启。
  - 如果重启多个实例,选择多个实例(所选实例的状态必须一致),并在页面底部单击重启。
- 5. 在重启实例对话框中,选择重启方式,并单击确定。

### 相关文档

RebootInstance

# 5.5. 释放实例

GPU实例作为云服务器ECS的一类实例规格,保持了与管理ECS实例相同的相关操作。因此,GPU实例仅支持 释放按量付费实例(包括抢占式实例)和已到期的包年包月实例。包年包月实例到期前,需申请退款或转为 按量付费实例后再释放。本文介绍手动及自动释放按量付费实例的相关操作。

### 前提条件

实例释放后数据无法恢复,建议您在释放之前先创建快照备份数据。具体操作,请参见创建快照。

⑦ 说明 释放实例不影响已经手动创建好的快照和镜像。

# 背景信息

- 包年包月实例到期前,不支持释放实例。您需要将实例转为按量付费实例,再手动释放实例。更多信息, 请参见包年包月转按量付费。如需申请退款,请参见退款规则及退款流程。
- 包年包月实例到期后,您可以手动释放。如果一直未续费,实例也会自动释放。
- 按量付费实例如果未启用节省停机模式,停止后所有资源会继续计费,只有释放后才停止计费。
- 为按量付费实例开启实例释放保护功能,可避免误操作释放实例造成不可挽回的后果。具体操作,请参见开启实例释放保护功能。
- 为ECS实例上挂载的云盘关闭随实例释放的功能,云盘会被自动转换成按量付费数据盘并被保留下来。更 多详情,请参见释放云盘。

#### 手动释放实例

您可以通过控制台立即释放按量付费实例。

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像 > 实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 请根据实际情况停止实例。
  - 释放一台实例:找到需要释放的实例,在操作列中,单击更多 > 实例状态 > 释放设置。

| 实例列表           |                     |  |               | () iji                             | E全局标签管    | 理云资源 设置  | ⑦ ECS控                       | 制台操作指南    | C   | 创建实例               | 批量操作                    |
|----------------|---------------------|--|---------------|------------------------------------|-----------|--|------------------------------|-----------|-----|--------------------|-------------------------|
| 检查到安全组中包含允许对特  | 定端口进行不受限            | 灵制访问的规则,4                                  | 存在潜在高危风险。查看详情 |                                    |           |  |                              |           |     |                    |                         |
| ▼ 选择实例属性项搜索,或者 | 「輸入关键字识别            | 搜索   | 0             | Q                                  | 标签        |  |                              |           |     | 高级搜索               | <u>a</u> o              |
| ■ 实例ID/名称      | 标签                  | 监<br>控 可用区 <del>•</del>                    | IP地址          | 状态 🔻                               | 网络类型<br>▼ | 配置   | 付费方式 ▼                       | 续费方式<br>▼ | 停止楔 | 钍                  | 操(                      |
| Ballinger B.   | 9 💊 🌞               | 坐<br>G     G     G     G     G     G     G | 用区            | ●已停<br>止                           | 专有网络      | 2 vCPU 8 GiB (I/O优化)<br>ecs.g5.large 5Mbps (峰值)    | 按量<br>2019年1月2日 17:19 億      | 収建        | 停机不 | <sup>收费</sup> 更改实例 | 管]<br>规格   更多           |
| -              | ۰ م                 | 坐 <sup> 华东1可</sup><br>G                    |               | <ul> <li>●运行</li> <li>中</li> </ul> | 专有网络      | 2 vCPU 8 GiB (I/O优化)<br>ecs.g5.large 5Mbps (峰值)    | 按量<br>2018年12月20日 14:20<br>建 | ) 创       | -   | ■<br>更改实例          | 里   远程连<br>规格 <b>更多</b> |
| Tanal allow 1. | ۰ 🖗                 | ▶ 华东1可                                     | 刑区 F          | <ul> <li>●运行</li> <li>中</li> </ul> | 专有网络      | 2 vCPU 8 GiB (I/O优化)<br>ecs.sn2ne.large 5Mbps (峰值) | 按量<br>2018年12月20日 1(<br>建    | 启动        |     | 购买相同配置<br>实例状态     | t                       |
|                | ♥ <sup>♥</sup><br>■ | 坐 华东1可 G                                   | 用区            | <ul> <li>)运行</li> <li>中</li> </ul> | 专有网络      | 2 vCPU 8 GiB (I/O优化)<br>ecs.g5.large 10Mbps (峰值)   | 按量<br>2018年12月4日 16:         | 停止<br>重启  |     | 实例设置<br>密码/密钥      | )                       |
| -              | <u>ې</u>            | ▶ 华东1可                                     | <b>∃</b> 区 B  | ●运行<br>中                           | 专有网络      | 1 vCPU 1 GiB (I/O优化)<br>ecs p1 tiny 5Mbps (峰值)     | 按量<br>2018年11月22日 1          | 释放设置      |     | 资源变配               | )                       |

 释放一台或多台实例:根据付费方式筛选按量付费实例,选中需要释放的实例,在列表下方,单 击释放设置。

|   | Text States - Los Str. | ۲  | 0        | ⊵   | 华东 1 可用区<br>G | 1 (10.75,200.0)<br>(10.000.0)/(10.000 | <ul> <li>●运行</li> <li>中</li> </ul> | 专有网络 | 2 vCPU 8 GiB (I/O优化)<br>ecs.g5.large 10Mbps (峰值) |
|---|------------------------|----|----------|-----|---------------|---------------------------------------|------------------------------------|------|--|
|   | Next Address 1         | ۲  | $\Delta$ | Ł   | 华东1可用区 B      | 4140,830,000<br>18.486,000,00         | <ul> <li>●运行</li> <li>中</li> </ul> | 专有网络 | 1 vCPU 1 GiB (I/O优化)<br>ecs.n1.tiny 5Mbps (峰值)   |
|   | And Statements         | ۲  | ♥<br>♦   | Ł   | 华东 1 可用区<br>G | 11.00.00111                           | ●已停<br>止                           | 专有网络 | 1 vCPU 1 GiB (I/O优化)<br>ecs.xn4.small 5Mbps (峰值) |
| 1 | 7                      | ۲  | ♥<br>∰   | ⊾   | 华东 1 可用区<br>G | 11,01,01,04,04                        | ●已停<br>止                           | 专有网络 | 2 vCPU 8 GiB (I/O优化)<br>ecs.g5.large 10Mbps (峰值) |
| • | manager P              | ۲  | •        | ⊾   | 华东1可用区<br>G   | 11.00.004010                          | <ul> <li>●已停</li> <li>止</li> </ul> | 专有网络 | 2 vCPU 8 GiB (I/O优化)<br>ecs.g5.large 0Mbps (峰值)  |
|   | 启动停止                   | 重启 | 重置       | 实例图 | 四 续费          | 按量付费转包年包月                             | 释放                                 | 没置 更 | 19×  |

- 5. 在弹出的对话框中,选择立即释放。
- 6. 单击下一步,并单击确定。

## 开启自动释放

您也可以开启自动释放功能,设置自动释放实例的时间。如果重复设置自动释放时间,以最新的一次设置为 准。

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像 > 实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 设置释放。
  - 释放一台实例:找到需要设置自动释放的实例,在操作列中,单击更多 > 实例状态 > 释放设置。
  - 释放一台或多台实例:根据付费方式筛选按量付费实例,选中需要设置自动释放的实例,在列表下方,单击释放设置。
- 5. 在弹出的对话框中,选择定时释放。
- 6. 打开设置自动释放开关,并设置释放的日期和时间。

⑦ 说明 最早只能设置在30分钟后自动释放实例,时间精确到分钟。

| 释放设置  |               | ×  |  |  |  |  |  |  |
|---|---------------|----|--|--|--|--|--|--|
| *释放行为:  | ○ 立即释放 ● 定时释放 |    |  |  |  |  |  |  |
| 设置自动释放:   |               |    |  |  |  |  |  |  |
| *释放日期:  | 2018-01-29    |    |  |  |  |  |  |  |
| *释放时间:  | 17 : 59       |    |  |  |  |  |  |  |
| 温馨提示: <ul> <li>系统执行释放的定时任务间隔5分钟,系统会按定时任务的释放时间停止计费。</li> </ul> |               |    |  |  |  |  |  |  |
|   | 下一步           | 取消 |  |  |  |  |  |  |

7. 单击下一步, 然后单击确定。

### 关闭自动释放

- 1. 登录ECS管理控制台。
- 2. 在左侧导航栏,选择实例与镜像 > 实例。
- 3. 在顶部菜单栏左上角处,选择地域。
- 4. 释放设置。
  - 为一台实例关闭自动释放:找到需要关闭自动释放的实例,在操作列中,单击更多 > 实例状态 > 释 放设置。
  - 为一台或多台实例关闭自动释放:根据付费方式筛选按量付费实例,选中需要关闭自动释放的实例, 在列表下方,单击释放设置。
- 5. 在弹出的对话框中,选择定时释放。
- 6. 关闭设置自动释放开关。
- 7. 单击下一步, 然后单击确定。

# 相关文档

# 相关文档

- DeleteInstance
- ModifyInstanceAutoReleaseTime

# 5.6. GPU监控

本文为您介绍如何通过云监控控制台和API查询GPU监控数据。

### 前提条件

- 请确保您已在云服务器ECS上创建GPU计算型实例,且已安装GPU驱动。具体操作,请参见创建未配备驱动的GPU实例。
- 请确保您已在云服务器ECS上安装云监控插件。具体操作,请参见安装和卸载C++版本插件。
- 请确保您已添加监控图表,且包含GPU的监控项。具体操作,请参见添加监控图表。

## GPU监控项说明

您可以从GPU、实例和应用分组维度查看GPU相关监控项。GPU的监控项如下表所示。

| 监控项                    | 单位   | MetricName              | Dimensions                  |
|------------------------|------|-------------------------|-----------------------------|
| (Agent)GPU维度解码<br>器使用率 | %    | gpu_decoder_utilization | userld、instanceld、<br>gpuld |
| (Agent)GPU维度编码<br>器使用率 | %    | gpu_encoder_utilization | userld、instanceld、<br>gpuld |
| (Agent )GPU维度GPU温<br>度 | °C   | gpu_gpu_temperature     | userld、instanceld、<br>gpuld |
| (Agent )GPU维度GPU使      | %    | gpu_gpu_usedutilizatio  | userld、instanceld、          |
| 用率                     |      | n                       | gpuld                       |
| (Agent )GPU维度显存<br>空闲量 | Byte | gpu_memory_freespace    | userld、instanceld、<br>gpuld |
| (Agent )GPU维度显存        | %    | gpu_memory_freeutiliza  | userld、instanceld、          |
| 空闲率                    |      | tion                    | gpuld                       |
| (Agent )GPU维度显存        | Byte | gpu_memory_userdspa     | userld、instanceld、          |
| 使用量                    |      | ce                      | gpuld                       |
| (Agent)GPU维度显存         | %    | gpu_memory_usedutiliz   | userld、instanceld、          |
| 使用率                    |      | ation                   | gpuld                       |
| (Agent )GPU维度GPU功      | W    | gpu_power_readings_p    | userld、instanceld、          |
| 率                      |      | ower_draw               | gpuld                       |

# 通过云监控控制台查看GPU的监控数据

- 1. 登录云监控控制台。
- 2. 在左侧导航栏,单击主机监控。
- 3. 在主机监控页面,单击目标主机的实例名称链接,或单击目标主机对应操作列的一图标。
- 4. 单击GPU监控页签。

在GPU监控页签,查看GPU相关监控图表。

### 通过API查看GPU的监控数据

通过DescribeMetricList接口查询GPU监控数据。更多信息, DescribeMetricList。

 ⑦ 说明
 Namespace
 取值为
 acs\_ecs\_dashboard
 MetricName
 和
 Dimensions
 的取值,请

 参见GPU监控项说明。
# 6.安装NVIDIA驱动

# 6.1. 在GPU计算型实例中安装GPU驱动 (Linux)

如果您在创建GPU计算型实例时没有配置自动安装GPU驱动,或者在公共镜像中没有您需要的操作系统或版本,为保证您能正常使用您创建的GPU实例,请在创建后手动安装驱动。本文为您介绍如何为Linux操作系统的GPU实例手动安装GPU驱动。

## 背景信息

GPU实例仅支持安装与其操作系统一致的GPU驱动。本文重点为您介绍手动安装Linux操作系统的GPU驱动的 相关操作。如果您创建的GPU实例为Windows操作系统,请参见在GPU计算型实例中安装GPU驱动 (Windows),根据指导安装Windows操作系统的GPU驱动。

## 操作步骤

- 1. 访问NVIDIA驱动下载。
- 2. 设置搜索条件后,单击搜索查找适用的驱动程序。

| 高级搜索                |   |                      |        |      |             |  |  |
|---------------------|---|----------------------|--------|------|-------------|--|--|
| 产品类型:               |   | 操作系统                 | :      |      |             |  |  |
| Data Center / Tesla | ~ | Linux 64             | l-bit  | ~    |             |  |  |
| 产品系列:               |   | CUDA To              | olkit: |      |             |  |  |
| P-Series            | ~ | 11.2 🗸               |        |      |             |  |  |
| 产品家族:               |   | 语言:                  |        |      |             |  |  |
| Tesla P100          | ~ | Chinese (Simplified) |        |      |             |  |  |
|                     |   | 最新:                  |        |      |             |  |  |
|                     |   | 全部                   |        | ~ ?  |             |  |  |
| 搜索                  |   |                      |        |      |             |  |  |
| 名稱                  |   |                      | 版本     | 發行日期 | CUDA Toolki |  |  |
|                     |   |                      |        |      |             |  |  |

设置项说明如下所示:

| 设置项  | 说明   | 示例  |
|--|--|---|
| <ul><li>○ 产品类型</li><li>○ 产品系列</li><li>○ 产品家族</li></ul> | 根据实例规格配备的GPU选择对应的产品类型、<br>产品系列和产品家族。查看实例规格信息的具体<br>操作,请参见 <mark>查看实例信息</mark> 。<br>各GPU计算型规格的GPU信息的如下表所示。                              | <ul> <li>Data Center / Tesla</li> <li>P-Series</li> <li>Tesla P100</li> </ul> |
| 操作系统   | 根据实例使用的镜像选择对应的Linux操作系统版本。<br>如果下拉列表中没有显示所需的操作系统,请单<br>击下拉列表底部的 <b>选择所有操作系统</b> 。如果您<br>未能找到与实例镜像完全匹配的Linux操作系统<br>版本,请选择Linux 64-bit。 | Linux 64-bit  |

| 设置项          | 说明                        | 示例                   |
|--------------|---------------------------|----------------------|
| CUDA Toolkit | 选择CUDA Toolkit版本。         | 11.2                 |
| 语言           | 选择驱动对应的语言。                | Chinese (Simplified) |
| 最新           | 默认项为 <b>全部</b> ,保持默认选项即可。 | 全部                   |

#### 各GPU计算型规格的GPU信息如下所示:

| 信息项      | gn4                          | gn5                          | gn5i                                 | gn6v                         | gn6i                         | gn6e                         | gn7                          | gn7i                         | gn7e                         |
|----------|------------------------------|------------------------------|--------------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| 产品类<br>型 | Data<br>Center<br>/<br>Tesla | Data<br>Center<br>/<br>Tesla | Dat <i>a</i><br>Center<br>/<br>Tesla | Data<br>Center<br>/<br>Tesla | Data<br>Center<br>/<br>Tesla | Data<br>Center<br>/<br>Tesla | Data<br>Center<br>/<br>Tesla | Data<br>Center<br>/<br>Tesla | Data<br>Center<br>/<br>Tesla |
| 产品系      | M-                           | P-                           | P-                                   | V-                           | T -                          | V-                           | A-                           | A-                           | A-                           |
| 列        | Class                        | Series                       | Series                               | Series                       | Series                       | Series                       | Series                       | Series                       | Series                       |
| 产品家      | M40                          | Tesla                        | Tesla                                | Tesla                        | Tesla                        | Tesla                        | NVIDIA                       | NVIDIA                       | NVIDIA                       |
| 族        |                              | P100                         | P4                                   | V100                         | T4                           | V100                         | A100                         | A10                          | A100                         |

- 3. 在搜索到的驱动程序列表下,选择需要下载的驱动版本,单击对应的驱动名称。
- 在打开驱动程序下载页面,单击下载,然后在NVIDIA驱动程序下载页面,右键单击下载并选择复制 链接地址。

| 驱动程序 ·   支持   关于 NVIDIA ,               | 链接另存为<br>复制链接地址                       |         |
|---|---------------------------------------|---------|
| <sup>w载</sup><br>NVIDIA 驱动程序下载          | 在新标签页中打开图片<br>图片另存为<br>复制图片<br>复制图片地址 |         |
| 该下载内容包括 NVIDIA 图形驱动程序和一个额外安装 GeForce Ex | 从 Google 获取图片说明                       | ▶ Force |
|   | 检查 Ctrl+Sh                            | nift+I  |

5. 远程连接实例。

选择以下任一种方式远程连接GPU实例。

| 连接方式      | 操作指引               |
|-----------|--------------------|
| Workbench | 通过密码或密钥认证登录Linux实例 |
| VNC       | 通过密码认证登录Linux实例    |

6. 使用 wget 命令,并粘贴您在步骤3中复制的驱动下载链接,执行命令下载安装包。命令示例如下所示:

wget https://cn.download.nvidia.com/tesla/460.73.01/NVIDIA-Linux-x86\_64-460.73.01.run

7. 安装GPU驱动。

i. 如果您的操作系统为CentOS,请执行以下命令,查询实例中是否安装kernel-devel和kernelheaders包。对于ubuntu等其他操作系统,已经在其中预装了kernel-devel和kernel-headers包,请 跳过此步骤。

rpm -qa | grep \$(uname -r)

■ 如果回显类似如下信息,即包含了kernel-devel和kernel-headers包的版本信息,表示已安装。

```
kernel-3.10.0-1062.18.1.el7.x86_64
kernel-devel-3.10.0-1062.18.1.el7.x86_64
kernel-headers-3.10.0-1062.18.1.el7.x86_64
```

■ 如果在回显信息中, 您没有找到 *kernel-devel-\**和 *kernel-headers-\**内容, 您需要自行下载并安装 kernel对应版本的kernel-devel和kernel-headers包。

⑦ 说明 kernel-devel和kernel版本不一致会导致在安装driver rpm过程中driver编译出
 错。因此,请您确认回显信息中*kernel-\**的版本号后,再下载对应版本的kernel-devel。在示例回显信息中,kernel的版本号为3.10.0-1062.18.1.el7.x86\_64。

ii. 安装GPU驱动。

以操作系统是Linux 64-bit的驱动为例,您下载的GPU驱动为.run格式,例如:NVIDIA-Linux-x86\_64-xxxx.run。分别执行以下命令,授权并安装GPU驱动。

chmod +x NVIDIA-Linux-x86 64-xxxx.run

sh NVIDIA-Linux-x86 64-xxxx.run

iii. 安装完成后,执行如下命令,查看是否安装成功。

nvidia-smi

回显信息类似如下所示,表示GPU驱动安装成功。

| root@i                | Z                  |          |                    | :~# r                 | ividia-smi                      |                                  |                      |                                     |                   |
|-----------------------|--------------------|----------|--------------------|-----------------------|---------------------------------|----------------------------------|----------------------|-------------------------------------|-------------------|
| Tue Ju                | n 29 1             | 7:58:0   | 4 2021             |                       |                                 |                                  |                      |                                     |                   |
| +<br>  NVID           | IA-SMI             | 460.3    | 2.03               | Driver                | Version: 460.32.0               | I3 (                             | CUDA Versio          | on: 11.2                            | +-                |
| GPU<br>  Fan<br>      | Name<br>Temp       | Perf     | Persist<br>Pwr:Usa | ence-Ml<br>ge/Capl    | Bus-Id Di<br>Memory-U           | sp.A  <br> sage  <br>            | Volatile<br>GPU-Util | Uncorr. ECC<br>Compute M.<br>MIG M. | -+-<br> <br> <br> |
| 0<br>  N/A<br>        | Tesla<br>32C       | т4<br>Р8 | 9W /               | ======<br>0n  <br>70W | 00000000:00:08.0<br>0MiB / 1612 | =====+<br>0 Off  <br>27MiB  <br> |                      | Off<br>Default<br>N/A               | = <br> <br>       |
| +<br>  1<br>  N/A<br> | Tesla<br>31C       | T4<br>P8 | 9W /               | On  <br>70W           | 00000000:00:09.0<br>0MiB / 1612 | 0ff  <br>7MiB                    | 0%                   | Off<br>Default<br>N/A               | -+<br> <br> <br>  |
| 2<br>  N/A<br>        | Tesla<br>31C       | T4<br>P8 | 9W /               | On  <br>70W           | 00000000:00:0A.0<br>0MiB / 1612 | 0ff  <br>7MiB                    | 0%                   | Off<br>Default<br>N/A               | <br> <br>         |
| 3<br>  N/A<br>        | Tesla<br>33C       | T4<br>P8 | 9W /               | 0n  <br>70W           | 00000000:00:0B.0<br>0MiB / 1612 | 0ff  <br>7MiB                    | 0%                   | Off<br>Default<br>N/A               | <br> <br>         |
| +                     |                    |          |                    |                       |                                 |                                  |                      |                                     | -+                |
| Proc<br>  GPU<br>     | esses:<br>GI<br>ID | CI<br>ID | PI                 | D Typ                 | e Process name                  |                                  |                      | GPU Memory<br>Usage                 |                   |
| =====<br>  No<br>+=   | runnin             | g proc   | esses fo           | und                   |                                 |                                  |                      |                                     |                   |

- 8. 如果您创建的GPU实例的规格族为ebmgn7、ebmgn7e或sccgn7ex,您需要继续执行以下操作,安装与 驱动版本对应的nvidia-fabricmanager服务,否则,将无法正常使用GPU实例。
  - i. 安装nvidia-fabricmanager服务。

您可以通过源或者下载安装包两种方式安装nvidia-fabricmanager服务,并且不同操作系统对应的 安装命令有所区别,以下操作均以驱动版本为460.91.03为例,您需要根据实际情况替换 driver\_v ersion 后的驱动版本。命令示例如下:

- 通过源安装nvidia-fabricmanager服务。
  - CentOS 7.x

```
driver_version=460.91.03
yum -y install yum-utils
yum-config-manager --add-repo https://developer.download.nvidia.com/compute/cud
a/repos/rhel7/x86_64/cuda-rhel7.repo
yum install -y nvidia-fabric-manager-${driver version}-1
```

Cent OS 8.x

```
driver_version=460.91.03
driver_version_main=$(echo $driver_version | awk -F '.' '{print $1}')
distribution=rhel8
ARCH=$( /bin/arch )
dnf config-manager --add-repo http://developer.download.nvidia.com/compute/cuda
/repos/$distribution/${ARCH}/cuda-$distribution.repo
dnf module enable -y nvidia-driver:${driver_version_main}
dnf install -y nvidia-fabric-manager-0:${driver_version}-1
```

#### Ubuntu 16.04、Ubuntu 18.04或Ubuntu 20.04

```
driver_version=460.91.03
driver_version_main=$(echo $driver_version | awk -F '.' '{print $1}')
distribution=$(. /etc/os-release;echo $ID$VERSION_ID | sed -e 's/\.//g')
wget https://developer.download.nvidia.com/compute/cuda/repos/$distribution/x86
_64/cuda-$distribution.pin
mv cuda-$distribution.pin /etc/apt/preferences.d/cuda-repository-pin-600
wget https://developer.download.nvidia.com/compute/cuda/repos/$distribution/x86
_64/7fa2af80.pub
apt-key add 7fa2af80.pub
rm 7fa2af80.pub
echo "deb http://developer.download.nvidia.com/compute/cuda/repos/$distribution
/x86_64 /" | tee /etc/apt/sources.list.d/cuda.list
apt-get update
apt-get update
apt-get -y install nvidia-fabricmanager-${driver_version_main}=${driver_version
}-*
```

- 通过下载安装包安装nvidia-fabricmanager服务。
  - CentOS 7.x

```
driver_version=460.91.03
wget http://mirrors.cloud.aliyuncs.com/nvidia-cuda/rhel7/x86_64/nvidia-fabric-m
anager-${driver_version}-1.x86_64.rpm
rpm -ivh nvidia-fabric-manager-${driver version}-1.x86_64.rpm
```

CentOS 8.x

```
driver_version=460.91.03
wget http://mirrors.cloud.aliyuncs.com/nvidia-cuda/rhel8/x86_64/nvidia-fabric-m
anager-${driver_version}-1.x86_64.rpm
rpm -ivh nvidia-fabric-manager-${driver_version}-1.x86_64.rpm
```

#### Ubuntu 16.04

```
driver_version=460.91.03
driver_version_main=$(echo $driver_version | awk -F '.' '{print $1}')
wget http://mirrors.cloud.aliyuncs.com/nvidia-cuda/ubuntul604/x86_64/nvidia-fab
ricmanager-${driver_version_main}_${driver_version}-1_amd64.deb
dpkg -i nvidia-fabricmanager-${driver_version_main}_${driver_version}-1_amd64.deb
eb
```

#### Ubuntu 18.04

```
driver_version=460.91.03
```

```
driver_version_main=$(echo $driver_version | awk -F '.' '{print $1}')
wget http://mirrors.cloud.aliyuncs.com/nvidia-cuda/ubuntul804/x86_64/nvidia-fab
ricmanager-${driver_version_main}_${driver_version}-1_amd64.deb
dpkg -i nvidia-fabricmanager-${driver_version_main}_${driver_version}-1_amd64.d
eb
```

#### Ubuntu 20.04

```
driver_version=460.91.03
driver_version_main=$(echo $driver_version | awk -F '.' '{print $1}')
wget http://mirrors.cloud.aliyuncs.com/nvidia-cuda/ubuntu2004/x86_64/nvidia-fab
ricmanager-${driver_version_main}_${driver_version}-1_amd64.deb
dpkg -i nvidia-fabricmanager-${driver_version_main}_${driver_version}-1_amd64.d
eb
```

#### ii. 执行如下命令,启动nvidia-fabricmanager服务。

```
systemctl enable nvidia-fabricmanager
systemctl start nvidia-fabricmanager
```

iii. 执行如下命令, 查看nvidia-fabricmanager服务状态。

systemctl status nvidia-fabricmanager

如果回显信息如下所示,则表示nvidia-fabricmanager服务安装成功。

| root@      | # systematl status nvidia-fabricmanager  |
|------------|--|
| nvidia-fi  | abricmanager.service - NVIDIA fabric manager service   |
| Loaded:    | loaded (/lib/systemd/system/nvidia-fabricmanager.service; enabled; vendor preset: enabled)   |
| Active:    | active (running) since Mon 2021-09-13 19:14:45 CST: 1 weeks 1 days ago   |
| Process:   | 1928 ExecStart=/usr/bin/ny-fabricmanager -c /usr/share/nyidia/nyswitch/fabricmanager.cfg (code=exited, status=0/SUCCESS)   |
| Main PID:  | 2140 (nv-fabricmanage)   |
| Tasks:     | 18 (limit: 19660)  |
| CGroup:    | /system.slice/nyidia-fabricmanager.service   |
|            | -2140 /usr/bin/nv-fabricmanager -c /usr/share/nvidia/nvswitch/fabricmanager.cfg  |
| Sep 13 19: | 14:26  |
| Sep 13 19: | 14:45 nv-fabricmanager[2140]: Successfully configured all the available GPUs and NVSwitches.   |
| Son 12 101 | added and a second all constant and an and a second a |

 如果您创建的GPU实例的规格族为gn7,目前仅支持安装Driver 510.47.03(不含510.47.03)以下版本的 驱动。如果您需要使用NVIDIA A100安装CUDA 11.6、Driver 510.47.03及以上版本的驱动,请选择GPU实 例的规格族为gn7e。

GPU实例创建或选择完成后,请您手动安装GPU驱动。具体操作,请参见步骤6。

# 6.2. 在GPU计算型实例中安装GPU驱动 (Windows)

由于目前暂不支持通过创建GPU实例自动配置Windows操作系统的GPU驱动,为保证您能正常使用所创建的 GPU实例,需要在创建实例后前往NVIDIA官网获取GPU驱动并手动安装。本文为您提供详细操作指导。

## 背景信息

GPU实例仅支持安装与其操作系统一致的GPU驱动。本文重点为您介绍手动安装Windows操作系统的GPU驱动的相关操作。如果您创建的GPU实例为Linux操作系统,请参见在GPU计算型实例中安装GPU驱动(Linux),根据指导安装Linux操作系统的GPU驱动。

## 操作步骤

1. 远程连接实例。

选择以下任一种方式远程连接GPU实例。

| 连接方式      | 操作指引                 |
|-----------|----------------------|
| Workbench | 通过密码或密钥认证登录Windows实例 |
| VNC       | 通过密码认证登录Windows实例    |

- 2. 在远程桌面中,通过浏览器访问NVIDIA驱动下载地址。
- 3. 设置搜索条件后,单击搜索查找适用的驱动程序。

| 高级搜索                    |                       |
|-------------------------|-----------------------|
| 产品类型:                   | 操作系统:                 |
| Data Center / Tesla 🔹 🥆 | Windows 10 64-bit 🗸 🗸 |
| 产品系列:                   | Windows 驱动程序类型:       |
| P-Series                | 标准・・                  |
| 产品家族:                   | CUDA Toolkit:         |
| Tesla P100 💊            | 11.2 ~                |
|                         | 语言:                   |
|                         | English (US) 🗸        |
|                         | 最新:                   |
|                         | 全部 🗸 🗸                |

#### 设置项说明如下所示:

| 设置项  | 说明  | 示例  |
|--|---|---|
| <ul><li>○ 产品类型</li><li>○ 产品系列</li><li>○ 产品家族</li></ul> | 根据实例规格配备的GPU选择对应的产品类型、<br>产品系列和产品家族。查看实例规格信息的具体<br>操作,请参见 <mark>查看实例信息</mark> 。<br>各GPU计算型规格的GPU信息的如下表所示。 | <ul> <li>Data Center / Tesla</li> <li>P-Series</li> <li>Tesla P100</li> </ul> |
| 操作系统   | 根据实例使用的镜像选择对应的Windows操作<br>系统版本。  | Windows 10 64-bit   |
| CUDA Toolkit   | 选择CUDA Toolkit版本。   | 11.2  |
| 语言   | 选择驱动对应的语言。  | English (US)  |
| 最新   | 默认项为 <b>全部</b> ,保持默认选项即可。   | 全部  |

### 各GPU计算型规格的GPU信息如下所示:

| 信息项      | gn4                          | gn5                          | gn5i                         | gn6v                         | gn6i                         | gn6e                         | gn7                          | gn7i                         | gn7e                         |
|----------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| 产品类<br>型 | Data<br>Center<br>/<br>Tesla |
| 产品系      | M-                           | P-                           | P-                           | V-                           | T -                          | V-                           | A-                           | A-                           | A-                           |
| 列        | Class                        | Series                       |
| 产品家      | M40                          | Tesla                        | Tesla                        | Tesla                        | Tesla                        | Tesla                        | NVIDIA                       | NVIDIA                       | NVIDIA                       |
| 族        |                              | P100                         | P4                           | V100                         | T4                           | V100                         | A100                         | A10                          | A100                         |

### 4. 在搜索到的驱动程序列表下,选择需要下载的驱动版本,单击对应的驱动名称。

| 版本     | 發行日期                             | CUDA Toolkit  |
|--------|----------------------------------|---|
| 462.31 | 19.4.2021                        | 11.2  |
| 461.33 | 8.2.2021                         | 11.2  |
| 461.09 | 7.1.2021                         | 11.2  |
|        | 版本<br>462.31<br>461.33<br>461.09 | 版本 發行日期<br>462.31 19.4.2021<br>461.33 8.2.2021<br>461.09 7.1.2021 |

5. 在打开驱动程序下载页面,单击下载,然后在NVIDIA驱动程序下载页面再次单击下载。

| 本:<br>布日期:<br>作系统:<br>JDA Toolkit:<br>言:<br>件大小:<br>下载 | 462.31 WHQL<br>2021.4.19<br>Windows 10 64-bit<br>11.2<br>Chinese (Simplified<br>445.1 MB | )      |      |
|--|--|--------|------|
|  | ちままし   | 产品支持列表 | 其他信息 |

6. 下载完成后,打开下载驱动程序所在的文件夹。然后双击安装文件,按提示完成安装。

⑦ 说明 对于Windows系统,在GPU驱动安装生效后,由于Windows自带的远程连接(RDP)协议不支持DirectX、OpenGL等相关应用,您需要自行安装VNC服务和客户端,或使用其他支持此类应用的协议,例如PCOIP、XenDesktop HDX 3D等。

完成安装后,您可以前往设备管理器,展开显示适配器,查看是否存在安装的GPU驱动信息。如果存在,表示安装成功。



## 6.3. 在GPU实例上安装GRID驱动(Windows)

当您需要使用Windows操作系统的GPU计算型实例或GPU虚拟化型实例应用于OpenGL、Direct 3D等图形计算场景时,请必须安装GRID驱动。如果您未在创建GPU实例时,通过镜像市场购买配备了GRID驱动的镜像,或者未找到您需要的操作系统及版本,那么,您可以参考本文操作安装GRID驱动并激活License。

## 前提条件

• 创建一台GPU实例,并确保实例可以访问公网。具体操作,请参见创建未配备驱动的GPU实例。

⑦ 说明 建议您选择公共镜像中的Windows镜像。

- 在本地机器上安装了远程连接工具。
- 已经通过提交工单获取了GRID License及对应的GRID驱动安装包。

⑦ 说明 由于GRID驱动的版本必须与您的GPU实例规格及实例所在地域相匹配,因此,请在提交工单时,提供您的GPU实例规格及实例所在地域。

## 背景信息

对于Windows操作系统的GPU计算型实例和GPU虚拟化型实例,安装GRID驱动的操作基本一致。本文以操作 系统为Windows Server 2019的GPU虚拟化型实例vgn6i为例,介绍如何安装NVIDIA GRID驱动并激活GRID License。

### 操作步骤

- 1. 远程连接实例。
- 2. 安装已获取的驱动。
  - i. 双击安装程序,在弹出的对话框,单击OK。



安装程序将自动进行解压。

ii. 解压完成后,在弹出的NVIDIA 安装程序对话框中,单击同意并继续(A)。



iii. 保持默认安装选项,单击下一步(N)。



iv. 安装完成后,单击马上重新启动(R)。



- 3. GPU实例重启完成后,连接实例并查看GRID驱动是否安装成功。
  - i. 使用Workbench远程连接实例。

↓ 注意 请勿通过云服务器管理控制台使用VNC远程连接GPU实例,否则会由于VNC无法获取集成显卡画面导致黑屏,影响您的使用。

- ii. 在Windows桌面的一图标上,单击鼠标右键,然后单击设备管理器(M)。
- iii. 在设备管理器对话框中,查看显示适配器目录下是否存在NVIDIA GRID T4-8Q设备。

如下图所示,表示GRID驱动已安装成功。

| 8.4 | 過管理      | 88 (L)            |  |  |      |      | - | ×    |
|-----|----------|-------------------|--|--|------|------|---|------|
| 文件  | 戸撮       | 作(A) 查看(V) 報助(H)  |  |  |      |      |   |      |
|     |          | 2 🖬 😕             |  |  |      |      |   |      |
| 1.8 | iZbyn    | lowtfsacvZ        |  |  | <br> | <br> |   | <br> |
| 5   | ID 📼     | E ATA/ATAPI 控制器   |  |  |      |      |   |      |
| >   | 重处       | 1988              |  |  |      |      |   |      |
| >   | - 83     | 盘影动器              |  |  |      |      |   |      |
| >   | 公存       | <b>前注空制器</b>      |  |  |      |      |   |      |
| >   | 言打       | 印队列               |  |  |      |      |   |      |
| >   | 前来       | COM HELPT)        |  |  |      |      |   |      |
| >   | - 田      | 30.61             |  |  |      |      |   |      |
| >   | - at     | 1998              |  |  |      |      |   |      |
| >   | = 12     | #                 |  |  |      |      |   |      |
| ~   | E? 30    | 他设备               |  |  |      |      |   |      |
|     | - 52     | 多媒体音须控制器          |  |  |      |      |   |      |
| >   | 晶人       | 体学输入设备            |  |  |      |      |   |      |
| >   | 晶软       | 金驱动器控制器           |  |  |      |      |   |      |
| >   | () R     | 标和其他指针设备          |  |  |      |      |   |      |
| >   | ų ž      | 用串行总线控制器          |  |  |      |      |   |      |
| >   | 💇 P0     | 络近面器              |  |  |      |      |   |      |
| >   | 1 系      | 统设备               |  |  |      |      |   |      |
| ~   | <b>1</b> | 示话情况情             |  |  |      |      |   |      |
|     |          | Microsoft 基本显示透影器 |  |  |      |      |   |      |
|     |          | NVIDIA GRID T4-8Q |  |  |      |      |   |      |

- 4. 添加License服务器并激活License。
  - i. 在Windows桌面单击鼠标右键,然后单击NVIDIA 控制面板。

| 查看(V)                   | > |
|-------------------------|---|
| 排序方式(O)                 | > |
| 刷新(E)                   |   |
| 粘贴(P)                   |   |
| 粘贴快捷方式(S)               |   |
| NVIDIA 控制面板             |   |
| 新建(W)                   | > |
| 🕺 nView Desktop Manager |   |
| 🔲 显示设置(D)               |   |
| 🗹 个性化(R)                |   |

ii. 在NVIDIA 控制面板对话框,选择许可>管理许可证。

| 🛃 NVIDIA 控制面板  | - 0   | ) × [ |
|--|---|-------|
| 文件(F) 编辑(E) 桌面(K) 帮助(H)  |   |       |
| 🔇 📾 🔹 🕥 🐇  |   |       |
| 38-965<br>9 3 32<br>- 2010 002<br>- 2010<br>- 2010 | 管理作可能     Gold 出生用作可证未会用的加加体。     Acrouge 出生用作可证未会用的加加体。     Honge 是書     Gold 出生用作可证都符書:     Gold 出生用作可证和行用作可:     Gold 出生用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和目前的证明计正确行     Gold 出生用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作可证和行用作用作用作用作用作用作用作用作用作用作用作用作用作用作用作用作用作用作用作 | ~     |
|  |   | >     |
| 0 系统信息   | 应用(A) 取   | 5jij  |

iii. 在管理许可证区域,输入您提前获取到的License服务器的IP和端口,然后单击应用。 重新打开NVIDIA 控制面板,如果提示如下信息,表示License激活成功。

| NVIDIA 控制面板                         |   | - | × |
|-------------------------------------|---|---|---|
| 文件(F) 编码(E) 成面(K) 和助(H) 〇 后思 - 〇 〇  |   |   |   |
| 5择一项任务<br>30 通过预防调整图像设置<br>管理 30 设置 | 管理许可证<br>Graulestockeenseenseenseense                       |   |   |
|                                     | 许可证类型:<br>정加系统已获 Quadro Virtual Data Center Verkatation 许可。 |   |   |
|                                     | 一级许可证服务器:   |   |   |
|                                     | 端口号:  |   |   |
|                                     | 二级许可证服务器:<br>   |   |   |
|                                     | 说明:<br>必頃: 接收许可请求的主要许可证服务器的 17 地址或全纲定主机名。                   |   |   |
|                                     | 典型的使用循形:  |   |   |
| ① <u>系统信息</u>                       |   |   |   |

## 6.4. 在GPU虚拟化型实例中安装GRID驱动 (Linux)

如果您创建的是GPU虚拟化实例,则需要安装GRID驱动。由于GPU实例配备的NVIDIA GPU计算卡尚未激活 NVIDIA GRID License,GPU图形功能受到限制,您需要购买并激活GRID License。本文以Ubunt u 16.04 64-bit 操作系统的GPU虚拟化型实例vgn6i和vgn5i为例,介绍如何安装GRID驱动和激活GRID License,并验证图形化 加速效果。

## 前提条件

• 创建一台vgn6i或vgn5i实例,确保实例可以访问公网,建议您选择公共镜像中的镜像。

⑦ 说明 本文介绍如何为运行Linux操作系统的实例安装GRID驱动,对需要运行Windows操作系统的实例,在创建实例时选用预装GRID驱动的付费镜像即可。具体操作,请参见创建配备GRID驱动的GPU实例(Windows)。

- 在本地机器上安装了远程连接工具,例如VNC Viewer。
- 已经通过提交工单获取了GRID License及对应的GRID驱动安装包。

⑦ 说明 由于GRID驱动的版本必须与您的GPU实例规格及实例所在地域相匹配,因此,请在提交工单时,提供您的GPU实例规格及实例所在地域。

## 背景信息

本文介绍如何为配备vGPU的GPU虚拟化型实例安装GRID驱动,如果您使用的是Linux操作系统的GPU计算型实例,建议您安装GPU驱动即可。具体操作,请参见在GPU计算型实例中安装GPU驱动(Linux)。

## 安装GRID驱动

1. 禁用nouveau。

nouveau是部分Linux系统中预装的一款开源驱动,必须禁用nouveau才能成功安装新的驱动。

i. 远程连接实例。

选择以下任一种方式远程连接GPU实例。

| 连接方式      | 操作指引               |
|-----------|--------------------|
| Workbench | 通过密码或密钥认证登录Linux实例 |
| VNC       | 通过密码认证登录Linux实例    |

#### ii. 查看是否存在*blacklist-nouveau.conf*文件。

ls /etc/modprobe.d/blacklist-nouveau.conf

#### 如果不存在,请执行以下命令创建该文件。

vim /etc/modprobe.d/blacklist-nouveau.conf

iii. 在*blacklist-nouveau.conf*文件中,添加以下内容禁用nouveau。

⑦ 说明 如果已配置以下内容,则跳过本步骤。

blacklist nouveau blacklist lbm-nouveau options nouveau modeset=0

#### iv. 生成kernel init ramfs。

rmmod nouveau update-initramfs -u

v. 重启实例。

reboot

#### 2. 安装已获取的GRID驱动。

#### ∘ vgn5i

```
chmod +x NVIDIA-Linux-x86_64-418.226.00-grid.run
./NVIDIA-Linux-x86 64-418.226.00-grid.run
```

#### ∘ vgn6i

```
chmod +x NVIDIA-Linux-x86_64-430.63-grid.run
./NVIDIA-Linux-x86_64-430.63-grid.run
```

#### vgn6i-vws、vgn7i-vws

chmod +x NVIDIA-Linux-x86\_64-460.91.03-grid.run ./NVIDIA-Linux-x86\_64-460.91.03-grid.run

#### 3. 测试GRID驱动是否安装成功。

nvidia-smi

如果返回以下GRID驱动信息,说明驱动安装成功。

| root@i<br>Thu Aug 13 11:12:5          | :∼# nvidia-smi<br>8 2020                            |   |
|---------------------------------------|---|---|
| NVIDIA-SMI 430.6                      | 3 Driver Version: 430.63                            | CUDA Version: 10.1  |
| GPU Name<br>  Fan Temp Perf           | Persistence-M Bus-Id Dis<br>Pwr:Usage/Cap Memory-Us | sp.A   Volatile Uncorr. ECC  <br>sage   GPU-Util Compute M. |
| ===================================== | On   00000000:00:07.0<br>N/A / N/A   528MiB / 8128  | Off   N/A  <br>MiB   0% Default                             |
| +                                     |   | +   |
| Processes:<br>GPU PID                 | Type Process name                                   | GPU Memory  <br>Usage                                       |
| No running proc                       | esses found   | <br> <br> +   |

其中, Driver Version 代表当前GRID驱动版本, CUDA Version 代表驱动支持的最高CUDA版本。 其他驱动版本及其支持的CUDA版本情况,如下表所示:

| GRID驱动版本  | 支持的CUDA最高版本 |
|-----------|-------------|
| 430.99    | 10.1        |
| 460.91.03 | 11.2        |
| 470.82.01 | 11.4        |

#### 4. 添加License服务器。

### i. 切换到/etc/nvidia目录。

cd /etc/nvidia

#### ii. 创建gridd.conf文件。

cp gridd.conf.template gridd.conf

iii. 在gridd.conf文件中添加License服务器的信息。

```
ServerAddress=<License服务器的IP>
ServerPort=<License服务器的端口(默认为7070)>
FeatureType=1
```

5. 重启实例使License服务器配置生效。

reboot

- 6. 查看License是否激活成功。
  - i. 远程连接实例。
  - ii. 查看License状态。

systemctl status nvidia-gridd

返回结果显示License acquired successfully, 表示License激活成功。

| rootĝi<br>• nvidia-gridd.service - NVIDIA Grid Daemon<br>Loaded: Loaded (/tet/systemd/system/nvidia-gridd.service; enabled; vendor preset: enabled)<br>Active: active (running) since Thu 2020-08-13 13:26:34 CST; 105 ago<br>Process: 79 ExeCstart-/usr/bin/nvidia-gridd (code=exited, status=0/SUCCESS)<br>Main PID: 853 (nvidia-gridd)<br>Tasks: 3 (limit: 4015)<br>CGroup: /system.slice/nvidia-gridd.service |
|---|
| Aug 13 13:26:34 i   |
| Aug 13 13:26:36 i / nvidia-gridd[853]: Calling load byte_array(tra)<br>Aug 13 13:26:38 i / nvidia-gridd[853]: License acquired successfully. (Info: http:///request; Quadro-Virtual-DWS,5.0)  |

## 测试图形化界面

本文以Ubunt u 16.04 64-bit操作系统为例,使用GLX测试程序验证安装GRID驱动后的图形化加速效果。

- 1. 准备图形化测试环境。
  - i. 运行命令安装x11vnc。

```
apt-get install x11vnc
```

ii. 运行命令 lspci | grep NVIDIA 查询GPU BusID。

本示例中,查询到的GPU BusID为 00:07.0 。

- iii. 配置X Server环境并重启。
  - a. 运行命令 nvidia-xconfig --enable-all-gpus --separate-x-screens 。
  - b. 编辑/etc/X11/xorg.conf,在 Section "Device" 段添加GPU BusID,如本示例中为 BusID "PCI:0:7:0" 。

| Section "Device" |                      |
|------------------|----------------------|
| Identifier       | "Device0"            |
| Driver           | "nvidia"             |
| VendorName       | "NVIDIA Corporation" |
| BoardName        | "Tesla P4"           |
| BusID            | "PCI:0:7:0"          |
| EndSection       |                      |

c. 运行 reboot 命令重启系统。

2. 运行命令安装GLX测试程序。

```
apt-get install mesa-utils
```

- 3. 运行命令 startx 启动X Server。
  - 如果没有 startx 命令,执行 apt-get install xinit 命令安装。
  - startx 启动时可能会提示 hostname: Name or service not known 。这个提示不会影响X
     Server启动。您可以运行命令 hostname 查得主机名后,再修改/etc/hosts文件,
     将 127.0.0.1 后的 hostname 改为本机的hostname。

root@iZ Z:~# startx hostname: Name or service not known xauth: (stdin):1: bad display name "iZl Z:1" in "add" command

4. 开启一个新的SSH客户端终端,运行命令启动x11vnc。

x11vnc -display :1

如果看到如下图所示的信息,表示x11vnc已经成功启动。此时,您能通过VNC Viewer等VNC远程连接软件连接实例。

The VNC desktop is: localhost:0 PORT=5900

- 5. 登录ECS管理控制台,在实例所在安全组中添加安全组规则,允许TCP 5900端口的入方向访问。具体操 作,请参见添加安全组规则。
- 6. 在本地主机上,使用VNC Viewer等VNC远程连接软件,通过 实例公网IP地址:5900 连接实例,进入KDE 桌面。
- 7. 按以下步骤使用 glxinfo 命令查看当前GRID驱动支持的配置。
  - i. 开启一个新的SSH客户端终端。
  - ii. 运行命令 export DISPLAY=:1 。
  - iii. 运行命令 glxinfo -t 列出当前GRID驱动支持的配置。
- 8. 按以下步骤使用 glxgears 命令测试GRID驱动。
  - i. 在KDE桌面上,右键单击桌面,并选择Run Command。

ii. 运行 glxgears 启动齿轮图形测试程序。如果出现如下图所示的窗口,表明GRID驱动正常工作。



## 7.卸载NVIDIA驱动

## 7.1. 卸载GPU驱动

您可以手动卸载GPU驱动,卸载命令需要根据GPU驱动的安装方式和操作系统类型而定。

## 背景信息

GPU实例必须配备了相关驱动才可以正常使用,如果您因某种原因需要卸载当前驱动,请务必再安装与您实例规格及操作系统相匹配的正确驱动,以免影响您的使用。更多信息,请参见更新NVIDIA驱动。

如果您是通过创建GPU实例自动安装了GPU驱动,那么GPU驱动的卸载方式需要根据操作系统类型而定:

- Ubuntu操作系统使用deb方式。
- CentOS和SUSE操作系统使用run方式。

本文为您提供不同操作系统中卸载GPU驱动的具体操作。如下所示:

• Windows操作系统:

在Windows操作系统中卸载GPU驱动

- Linux操作系统:
  - 。 在Ubuntu操作系统中卸载GPU驱动
  - 。 在Cent OS操作系统中卸载GPU驱动
  - 。 在SUSE操作系统中卸载GPU驱动

? 说明

本文Linux操作系统的相关步骤均需要以 root 用户登录并进行操作。如果您是普通用户,请使用 s udo 命令获取 root 权限后再操作。

## 在Windows操作系统中卸载GPU驱动

以下操作将以操作系统为Windows Server 2019的GPU计算型实例gn6i为例。

- 1. 远程连接实例。
- 2. 单击Windows桌面左下角 💶 图标, 然后单击控制面板。



3. 在控制面板中,单击卸载程序。

| ← → · ↑ | 🔝 > 控制面板 |                                |           |                                    |
|---------|----------|--------------------------------|-----------|------------------------------------|
|         |          | 调整计算机的设置                       |           |                                    |
|         |          | 系统和安全<br>查看你的计算机状态<br>學 查看哪件目志 | <u>88</u> | 用户帐户                               |
|         |          | 网络和 Internet<br>重要网络状态和任务      |           | 外观和个性化                             |
|         |          | 硬件<br>查看设备和时的机                 | ٢         | 时钟和区域<br>设置时间和日期<br>更改日期、时间或数字格式   |
|         |          | 10.000mm<br>程序<br>10.000月78    | G         | 轻松使用<br>使用 Windows 建议的设置<br>优化视觉显示 |

4. 右键单击需要卸载的GPU驱动,然后单击卸载/更改(U)。

|                  | 101 > 1914 > 191442346                               |                       |            | V D 1991 | 9.49100 atr   | 2 |
|------------------|--|-----------------------|------------|----------|---------------|---|
|                  |  |                       |            |          |               |   |
| 控制国転工具           | 卸载或肥改程序  |                       |            |          |               |   |
| 全面已安装的英格         | 被要的数据中、进入到费中的利益中、然后参考"的数"、"要求                        | 111日月月1日              |            |          |               |   |
| 自己成大市 Windows 取録 |  |                       |            |          |               |   |
|                  | 他们 - 印献/英政   |                       |            |          | (SE -         |   |
|                  | 48   | 258                   | 0.0010     | 大小       | 85            |   |
|                  | Cocole Chrome  | Google Inc.           | 2021/11/26 |          | 05.0.3359.159 |   |
|                  | Microsoft Visual C++ 2015-2019 Redistributable (x64) | Microsoft Corporation | 2021/11/26 | 23.1 MB  | 14.22.27821.0 |   |
|                  | Microsoft Visual C++ 2015-2019 Redistributable (x86) | Microsoft Corporation | 2021/11/26 | 20.1 MB  | 14.22.27821.0 |   |
|                  | NVIDIA RTX Desktop Manager 202.21                    | NVIDIA Corporation    | 2021/11/26 |          | 202.21        |   |
|                  | NVIDIA WMI 2,96.0                                    | NVIDIA Corporation    | 2021/11/26 |          | 2,36.0        |   |
|                  | NVIDIA BEREIRER 472.50                               |                       | 2021/11/26 |          | 472.50        |   |
|                  | 2046   | /@at(U)               |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |
|                  |  |                       |            |          |               |   |

5. 在弹出的NVIDIA 卸载程序对话框,单击卸载(U)。



6. 卸载完成后,单击马上重新启动(R)。

| NVIDIA 卸载程序                    |           | _   |       |
|--------------------------------|-----------|-----|-------|
| 图形驱动程序<br><sup>版本 472.50</sup> |           |     |       |
| NVIDIA 卸载程序已完成                 |           |     |       |
| • 要完成删除,请重新启动计算机。<br>是否马上重新启动? |           |     |       |
|                                |           |     |       |
|                                | 马上重新启动(R) | 稍后重 | 重启(L) |

重启完成后,则GPU驱动已卸载成功。

## 在Ubuntu操作系统中卸载GPU驱动

如果您使用deb安装包安装了GPU驱动,建议您按照以下步骤卸载GPU驱动,以下操作以Driver 410.104、 CUDA 10.0.130、cuDNN 7.5.0为例。

1. 卸载GPU驱动。

apt-get remove --purge nvidia-\*

#### 2. 卸载CUDA和cuDNN库。

```
apt autoremove --purge cuda-10-0
rm -rf /usr/local/cuda-10.0
```

#### 3. 重启实例。

reboot

如果您使用run安装包安装了GPU驱动,建议您按照以下步骤卸载GPU驱动。

#### 1. 卸载GPU驱动。

/usr/bin/nvidia-uninstall

#### 2. 卸载CUDA和和cuDNN库。

/usr/local/cuda/bin/cuda-uninstaller
rm -rf /usr/local/cuda-10.0

⑦ 说明 不同CUDA版本,卸载命令可能存在差别,如果未找到cuda-uninstaller文件,请到/usr/ local/cuda/bin/目录下查看是否存在uninstall\_cuda开头的文件。如果有,则将命令中的cuda-unin staller替换为该文件名。

3. 重启实例。

reboot

## 在CentOS操作系统中卸载GPU驱动

如果您使用run安装包安装了GPU驱动,建议您按照以下步骤卸载GPU驱动,以下步骤以Driver 410.104、 CUDA 10.0.130、cuDNN 7.5.0为例。

1. 卸载GPU驱动。

/usr/bin/nvidia-uninstall

2. 卸载CUDA和cuDNN库。

/usr/local/cuda/bin/cuda-uninstaller
rm -rf /usr/local/cuda-10.0

⑦ 说明 不同CUDA版本,卸载命令可能存在差别,如果未找到cuda-uninstaller文件,请到/usr/ local/cuda/bin/目录下查看是否存在uninstall\_cuda开头的文件。如果有,则将命令中的cuda-unin staller替换为uninstall\_cuda开头的文件名。

3. 重启实例。

reboot

如果您在Cent OS 7中使用rpm安装包安装了GPU驱动,建议您按照以下步骤卸载GPU驱动。

1. 卸载GPU驱动。

```
yum remove xorg-x11-drv-nvidia nvidia-kmod cuda-drivers
yum remove nvidia-diag-driver-local-repo-rhel7-410.104
```

#### 2. 卸载CUDA和cuDNN库。

```
yum remove /usr/local/cuda-10.0
rm -rf /usr/local/cuda-10.0
```

#### 3. 重启实例。

reboot

如果您在Cent OS 6中使用rpm安装包安装了GPU驱动,建议您按照以下步骤卸载GPU驱动。

1. 卸载GPU驱动。

```
yum remove xorg-x11-drv-nvidia nvidia-kmod cuda-drivers
yum remove nvidia-diag-driver-local-repo-rhel6-410.104
```

2. 卸载CUDA。

```
yum remove /usr/local/cuda-10.0
```

3. 重启实例。

reboot

## 在SUSE操作系统中卸载GPU驱动

如果您使用run安装包安装GPU驱动,建议您按照以下步骤卸载GPU驱动,以下步骤以CUDA 9.0.176为例。

1. 卸载GPU驱动。

/usr/bin/nvidia-uninstall

2. 卸载CUDA和cuDNN库。

```
/usr/local/cuda/bin/uninstall_cuda_9.0.pl
rm -rf /usr/local/cuda-9.0
```

3. 重启实例。

reboot

## 7.2. 卸载GRID驱动

如果您需要卸载GRID驱动,可以根据您已购GPU实例中安装的操作系统类型,参考本文对应的操作进行驱动 卸载。

## 背景信息

GPU实例必须配备了相关驱动才可以正常使用,如果您因某种原因需要卸载当前驱动,请务必再安装与您实例规格及操作系统相匹配的正确驱动,以免影响您的使用。更多信息,请参见更新NVIDIA驱动。

### 在Windows操作系统中卸载GRID驱动

以下操作将以操作系统为Windows Server 2019的GPU虚拟化型实例vgn6i为例。

- 1. 远程连接实例。
- 2. 单击Windows桌面左下角 🛃 图标, 然后单击控制面板。

| ≡ | 最近 | ទីវាព                      |   | Windows Server |                       |                           |
|---|----|----------------------------|---|----------------|-----------------------|---------------------------|
|   | 9  | nView Desktop Manager      |   |                |                       |                           |
|   |    |                            |   |                | Σ                     | 2                         |
|   | ł  | 服务器管理器                     |   | 服务器管理器         | Windows<br>PowerShell | Windows<br>PowerShell ISE |
|   |    |                            |   |                |                       |                           |
|   |    | NVIDIA Corporation<br>最近添加 | ř | Windows 1978   | - <b>-</b>            | 22                        |
|   |    |                            |   | T具             | 任务管理器                 | 控制面板                      |
|   | ۵  | 设置                         |   |                | _                     |                           |
|   | Q  | 搜卖                         |   | -              | Ö,                    |                           |
|   |    | 1                          |   | 远程桌面连接         | 事件查看器                 | 文件资源管理器                   |
| 8 |    | Windows PowerShell         | ~ |                |                       |                           |
| ۵ | #  | Windows 安全中心               |   |                |                       |                           |
|   |    | Windows 附件                 | ~ |                |                       |                           |
|   |    | Windows 管理工具               | ~ |                |                       |                           |
| 0 |    | Windows 轻松使用               | ~ |                |                       |                           |
| Ф |    | Windows 系统                 |   |                |                       |                           |
|   | Q  | H 健 🛤                      |   |                |                       |                           |

3. 在控制面板中,单击卸载程序。

| ← → * ↑ | :::::::::::::::::::::::::::::::::::: |      |                                |           |                                   |
|---------|--------------------------------------|------|--------------------------------|-----------|-----------------------------------|
|         |                                      | 调整计算 | 算机的设置                          |           |                                   |
|         |                                      |      | 系统和安全<br>查看你的计算机状态<br>學 查看事件日志 | <u>88</u> | 用户帐户                              |
|         |                                      |      | 网络和 Internet<br>查看网络状态和任务      |           | 外观和个性化                            |
|         |                                      | -0   | 硬件<br>查看设备和打印机<br>运动设备         | ٢         | 时钟和区域<br>设置时间和日期<br>更改日期、时间或数字格3  |
|         |                                      | ñ    | 程序                             | G         | 轻松使用<br>使用 Windows 建议的创<br>优化规范显示 |

4. 右键单击需要卸载的GRID驱动,然后单击**卸载/更改(U)**。

| 1 相序和功能              |  |   |                        |                   | -           |   | × |
|----------------------|--|---|------------------------|-------------------|-------------|---|---|
| ← → + ↑ ○ + HNER     | ( ) 程序 ) 程序(功能)                        |   |                        | ~ 心 推測1           | 很多纪功的"      |   | p |
| 1240回任主页<br>直面已会纳的更新 | 卸载或更改程序<br>需要如数据来,请从月春中将第四中,然后单非"如果"。  | "更欢"或"排算"。  |                        |                   |             |   |   |
| BREETS Windows 2010  | (8)(7 * 30戦/美政                         |   |                        |                   |             |   | 0 |
|                      | SR A                                   | 发布者<br>MATCIA Comparation                               | 安務时间<br>2020年0月        | 大小                | 版本<br>13042 |   |   |
|                      | NVIDIA nivšew 149.77                   | NVIDIA Corporation                                      | 2020/5/16              |                   | 149.77      |   |   |
| 1                    | NVIDIA 数形描的程序 432.44                   | MIDIA Corporation                                       | 2020/6/16              |                   | 432,44      | 5 |   |
| `                    | Python 3.0.3 (64-bit)  Python Launcher | ython Sattware Foundation<br>Python Sattware Foundation | 2020/5/15<br>2020/5/15 | 100 MB<br>1.28 MB | 3.8.3150.0  |   |   |
|                      |  |   |                        |                   |             |   |   |

5. 在弹出的NVIDIA 卸载程序对话框,单击卸载(U)。



6. 卸载完成后,单击马上重新启动(R)。



重启完成后,则GRID驱动已卸载成功。

## 在Linux操作系统中卸载GRID驱动

以下操作将以操作系统为Ubuntu 18.04的GPU虚拟化型实例vgn6i为例。

1. 远程连接实例。

选择以下任一种方式远程连接GPU实例。

| 连接方式      | 操作指引               |
|-----------|--------------------|
| Workbench | 通过密码或密钥认证登录Linux实例 |
| VNC       | 通过密码认证登录Linux实例    |

- 2. 执行如下命令卸载GRID驱动,并根据提示完成操作。
  - i. 执行如下命令, 卸载驱动。

nvidia-uninstall

ii. 当系统显示如下信息,询问您是否需要备份X screen的配置文件时,建议您保持默认选项 №, 然后按回车键。

⑦ 说明 不同操作系统的回显内容可能存在部分差别,您只需根据自身业务场景选择即可。

ОК

iii. 当系统显示如下信息时,表示卸载已完成。选择OK,然后按回车键。

Jninstallation of existing driver: NVIDIA Accelerated Graphics Driver for Linux-x86\_64 (430.99) is complete.

# 8.更新NVIDIA驱动

如果您的驱动版本已不适用于当前场景,或者您安装了错误的驱动类型或版本导致GPU实例无法使用,您可 以通过卸载当前驱动再安装所需驱动的方式,完成NVDIA驱动的更新。本文将为您提供卸载及安装不同类型 和操作系统的驱动的操作指引。

## 卸载NVIDIA驱动

根据您的驱动类型,选择对应操作指引:

- 卸载GPU驱动
- 卸载GRID驱动

## 安装NVIDIA驱动

请根据您需要安装的驱动类型,选择对应操作指引。

- 如果您需要安装GPU驱动,不同操作系统对应的操作指引如下:
  - 在GPU计算型实例中安装GPU驱动(Windows)
  - 。在GPU计算型实例中安装GPU驱动(Linux)
- 如果您需要安装GRID驱动,不同操作系统对应的操作指引如下:
  - o 在GPU虚拟化型实例中安装GRID驱动(Linux)
  - o 在GPU实例上安装GRID驱动(Windows)

↓ 注意 GPU虚拟化型实例仅支持安装与其规格族相匹配的版本的GRID驱动,如果安装了错误版本的GRID驱动,会导致GPU虚拟化型实例无法正常使用。因此,如果您需要为GPU虚拟化型实例重新安装GRID驱动,请根据您的实例规格安装对应版本的GRID驱动。