

弹性云服务器

# 用户指南

文档版本 01  
发布日期 2024-11-27



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

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

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

# 目录

<b>1 产品介绍</b>	<b>1</b>
1.1 什么是弹性云服务器?	1
1.2 产品优势	2
1.3 应用场景	4
1.4 实例类型和规格	4
1.4.1 实例概述	5
1.4.2 实例生命周期	5
1.4.3 实例类型	6
1.4.4 实例规格 (x86)	7
1.4.4.1 通用型	7
1.4.4.2 通用计算增强型	12
1.4.4.3 内存优化型	22
1.4.4.4 密集存储型	28
1.4.4.5 超高 I/O 型	31
1.4.4.6 高计算型	38
1.4.4.7 GPU 加速型	42
1.4.4.8 AI 加速型	64
1.4.5 实例规格 (鲲鹏)	67
1.4.5.1 鲲鹏通用计算型	67
1.4.5.2 鲲鹏通用计算增强型	69
1.4.5.3 鲲鹏内存优化型	74
1.4.5.4 鲲鹏超高 I/O 型	77
1.4.5.5 鲲鹏 AI 推理加速型	81
1.5 镜像	84
1.5.1 镜像类型	84
1.5.2 Cloud-init	85
1.6 云硬盘	86
1.7 网络	86
1.8 安全	88
1.8.1 故障恢复	88
1.9 约束与限制	88
1.10 与其他服务的关系	90
1.11 权限管理	90

1.12 区域和可用区.....	96
<b>2 快速入门.....</b>	<b>97</b>
2.1 创建弹性云服务器.....	97
2.1.1 概述.....	97
2.1.2 步骤一：基础配置.....	97
2.1.3 步骤二：网络配置.....	99
2.1.4 步骤三：高级配置.....	100
2.1.5 步骤四：确认配置.....	103
2.2 登录弹性云服务器.....	103
2.3 初始化数据盘.....	104
2.3.1 场景及磁盘分区形式介绍.....	104
2.3.2 初始化 Windows 数据盘（Windows 2008）.....	105
2.3.3 初始化 Windows 数据盘（Windows 2019）.....	112
2.3.4 初始化 Linux 数据盘（fdisk）.....	120
2.3.5 初始化 Linux 数据盘（parted）.....	126
2.3.6 初始化容量大于 2TiB 的 Windows 数据盘（Windows 2008）.....	131
2.3.7 初始化容量大于 2TiB 的 Windows 数据盘（Windows 2012）.....	138
2.3.8 初始化容量大于 2TiB 的 Linux 数据盘（parted）.....	147
<b>3 通过 IAM 授予使用 ECS 的权限.....</b>	<b>153</b>
3.1 创建用户并授权使用 ECS.....	153
3.2 ECS 自定义策略.....	154
<b>4 实例.....</b>	<b>157</b>
4.1 登录 Windows ECS.....	157
4.1.1 Windows ECS 登录方式概述.....	157
4.1.2 通过控制台 VNC 登录 Windows ECS.....	158
4.1.3 通过本地 Windows 主机登录 Windows ECS.....	159
4.1.4 通过本地 Linux 主机登录 Windows ECS.....	164
4.1.5 通过本地 macOS 主机登录 Windows ECS.....	166
4.1.6 通过移动设备登录 Windows ECS.....	169
4.2 登录 Linux ECS.....	174
4.2.1 Linux ECS 登录方式概述.....	174
4.2.2 通过 VNC 登录 Linux ECS.....	175
4.2.3 通过 SSH 密码方式登录 Linux ECS.....	177
4.2.4 通过 macOS 主机登录 Linux ECS.....	179
4.2.5 通过移动设备登录 Linux ECS.....	179
4.3 管理 GPU 加速型 ECS 的 GPU 驱动.....	191
4.3.1 GPU 驱动概述.....	191
4.3.2 Tesla 驱动及 CUDA 工具包获取方式.....	192
4.3.3 手动安装 GPU 加速型 ECS 的 GRID 驱动.....	193
4.3.4 手动安装 GPU 加速型 ECS 的 Tesla 驱动.....	203
4.4 管理 ECS 配置信息.....	217



4.4.1 更改时区.....	217
4.4.2 使用实例自定义数据和元数据.....	219
4.4.2.1 通过元数据获取 ECS 基本信息.....	219
4.4.2.2 通过实例自定义数据配置 ECS 实例.....	228
4.4.3 修改 ECS 名称.....	235
4.4.4 管理云服务器组.....	236
4.4.5 自动恢复.....	237
4.4.6 获取控制台日志.....	238
4.4.7 配置同一 VPC 内 ECS 通过主机名进行通信.....	241
4.5 变更 ECS 规格 (vCPU 和内存) .....	242
4.5.1 变更单台 ECS 规格.....	242
4.6 变更 ECS 操作系统.....	243
4.6.1 重装操作系统.....	243
4.6.2 切换操作系统.....	244
4.7 查看弹性云服务器信息.....	246
4.7.1 查看创建状态.....	247
4.7.2 查看失败任务.....	247
4.7.3 查看弹性云服务器详细信息.....	248
4.7.4 导出弹性云服务器列表信息.....	249
<b>5 镜像管理.....</b>	<b>250</b>
5.1 镜像概述.....	250
5.2 创建镜像.....	251
<b>6 磁盘管理.....</b>	<b>253</b>
6.1 为 ECS 新增磁盘.....	253
6.2 为 ECS 挂载磁盘.....	254
6.3 在线卸载磁盘.....	255
6.4 扩容云硬盘.....	257
6.5 开启高级磁盘功能.....	258
<b>7 弹性网卡管理.....</b>	<b>259</b>
7.1 绑定弹性网卡.....	259
7.2 解绑弹性网卡.....	260
7.3 修改私有 IP 地址.....	261
7.4 管理虚拟 IP 地址.....	262
7.5 开启网卡多队列功能.....	262
7.6 动态获取 IPv6 地址.....	267
<b>8 弹性公网 IP 管理.....</b>	<b>285</b>
8.1 绑定弹性公网 IP.....	285
8.2 解绑弹性公网 IP.....	285
8.3 修改弹性公网 IP 的带宽.....	286
8.4 无公网 IP 的弹性云服务器访问 Internet.....	286
<b>9 安全管理.....</b>	<b>290</b>

9.1 安全组.....	290
9.1.1 安全组概述.....	290
9.1.2 默认安全组和规则.....	291
9.1.3 安全组配置示例.....	291
9.1.4 配置安全组规则.....	296
9.1.5 更改安全组.....	298
<b>10 备份管理.....</b>	<b>299</b>
10.1 云备份概述.....	299
10.2 备份弹性云服务器.....	303
<b>11 密码和密钥对管理.....</b>	<b>305</b>
11.1 重置 ECS 密码.....	305
11.1.1 密码使用场景介绍.....	305
11.1.2 在控制台重置 ECS 密码.....	306
11.1.3 在操作系统内部修改 ECS 密码.....	308
11.1.4 重置 Windows 云服务器密码.....	309
11.1.5 重置 Linux 云服务器密码.....	311
11.2 管理 ECS 一键式重置密码插件.....	313
11.2.1 安装单台 ECS 的一键式重置密码插件（单台操作）.....	313
11.3 管理 ECS 密钥对.....	318
11.3.1 密钥对使用场景介绍.....	318
11.3.2 （推荐）通过管理控制台创建密钥对.....	320
11.3.3 通过 PuTTY Key Generator 工具创建密钥对.....	320
11.3.4 导入密钥对.....	324
11.3.5 获取和清除云服务器密码（Windows）.....	325
11.3.5.1 获取 Windows 弹性云服务器的密码.....	325
11.3.5.2 清除 Windows 弹性云服务器的密码.....	326
<b>12 资源.....</b>	<b>327</b>
12.1 配额调整.....	327
<b>13 使用 CES 监控 ECS.....</b>	<b>328</b>
13.1 监控弹性云服务器.....	328
13.2 弹性云服务器支持的基础监控指标.....	328
13.3 弹性云服务器支持的操作系统监控指标（安装 Agent）.....	332
13.4 设置告警规则.....	363
13.5 查看监控指标.....	365
<b>14 常见问题.....</b>	<b>366</b>
14.1 产品咨询.....	366
14.1.1 弹性云服务器使用时有什么限制？.....	366
14.1.2 弹性云服务器可以做什么？.....	366
14.2 ECS 创建.....	366
14.2.1 为什么创建弹性云服务器的任务失败，但是在列表中显示创建成功？.....	366

14.2.2 申请开通弹性云服务器需要多久? .....	367
14.3 ECS 删除与退订.....	367
14.3.1 “删除”按钮是什么意思? .....	367
14.3.2 已删除的弹性云服务器可以再开通吗? .....	367
14.3.3 重启/关机弹性云服务器时, 长时间处于“正在重启”/“正在关机”状态, 怎么办? .....	367
14.4 远程登录.....	368
14.4.1 登录前准备类.....	368
14.4.1.1 启动弹性云服务器时卡在“Waiting for cloudResetPwdAgent” .....	368
14.4.2 远程连接类.....	369
14.4.2.1 MSTSC 方式登录 Windows 2012 的弹性云服务器, 登录失败怎么办? .....	369
14.4.2.2 怎样修改远程登录的端口? .....	370
14.4.2.3 远程登录弹性云服务器时, 对浏览器版本的要求? .....	373
14.4.3 VNC 登录类.....	373
14.4.3.1 VNC 方式登录弹性云服务器后, 较长时间不操作, 界面无响应? .....	373
14.4.3.2 VNC 方式登录弹性云服务器后, 查看数据失败, VNC 无法正常使用? .....	373
14.4.3.3 VNC 方式登录弹性云服务器时, 系统黑屏输入无反应? .....	373
14.4.4 Windows 远程登录报错类.....	373
14.4.4.1 远程连接 Windows 云服务器报错: 出现身份验证错误, 要求的函数不受支持.....	374
14.4.4.2 远程连接 Windows 云服务器报错: 此计算机无法连接到远程计算机.....	375
14.4.4.3 远程连接 Windows 云服务器报错: 没有远程登录的权限.....	380
14.4.4.4 远程连接 Windows 云服务器报错: 没有远程桌面授权服务器可以提供许可证.....	381
14.4.4.5 登录 Windows 云服务器时报错: 0x112f.....	384
14.4.4.6 远程连接 Windows 云服务器报错: 0x1104.....	385
14.4.4.7 远程连接 Windows 云服务器报错: 122.112.....	389
14.4.4.8 使用 Mac 远程连接 Windows 云服务器报错: 证书或相关链无效.....	391
14.4.4.9 远程连接 Windows 云服务器报错: 由于协议错误会话中断.....	395
14.4.4.10 远程连接 Windows 云服务器报错: 无法验证此远程计算机的身份.....	397
14.4.4.11 远程连接 Windows 云服务器报错: 两台计算机无法在分配的时间内连接.....	398
14.4.4.12 远程连接 Windows 云服务器报错: 连接被拒绝未授权此用户.....	398
14.4.4.13 远程连接 Windows 云服务器报错: 您的连接已丢失.....	402
14.4.4.14 RDP 连接已断开, 出现内部错误, 错误代码 4.....	404
14.4.5 Linux 远程登录报错类.....	405
14.4.5.1 远程连接 Linux 云服务器报错: Module is unknown.....	405
14.4.5.2 远程连接 Linux 云服务器报错: Permission denied.....	408
14.4.5.3 远程连接 Linux 云服务器报错: read: Connection reset by peer.....	410
14.4.5.4 远程连接 Linux 云服务器报错: Access denied.....	411
14.4.5.5 远程连接 Linux 云服务器报错: Disconnected: No supported authentication methods available.....	411
14.5 磁盘分区/挂载/扩容.....	412
14.5.1 Windows 操作系统云服务器看不到挂载的数据盘怎么办? .....	412
14.5.2 怎样调整系统盘分区? .....	413
14.5.3 怎样查询 Windows 弹性云服务器磁盘分区与磁盘设备的对应关系? .....	418
14.5.4 怎样查询 Linux 弹性云服务器磁盘分区与磁盘设备的对应关系? .....	421

14.5.5 怎样配置 Windows 弹性云服务器的虚拟内存? .....	423
14.5.6 如何将扩容系统盘的空白分区在线扩容到末尾的 root 分区? .....	425
14.5.7 如何将扩容系统盘的空白分区在线扩容到非末尾的 root 分区? .....	426
14.5.8 一台弹性云服务器可以挂载多块磁盘吗? .....	429
14.5.9 弹性云服务器挂载磁盘时有什么限制? .....	430
14.5.10 哪些弹性云服务器支持挂载 SCSI 类型的云硬盘? .....	430
14.5.11 控制台与弹性云服务器内部之间磁盘挂载点的对应关系.....	430
14.5.12 挂载 SCSI 盘的 Linux 弹性云服务器, 重启失败怎么办? .....	434
14.5.13 磁盘脱机怎么办? .....	435
14.5.14 重启后盘符发生变化怎么办? .....	436
14.5.15 用户误删除 Tools 后, 如何获取数据盘信息? .....	437
14.6 网络配置.....	438
14.6.1 不同账号下跨 VPC 的弹性云服务器内网是否可以互通? .....	438
14.6.2 我购买的云服务器是否在同一子网? .....	438
14.6.3 如何配置端口映射? .....	438
14.6.4 如何查看云服务器的 mac 地址? .....	440
14.6.5 如何查看和修改 Linux 弹性云服务器的内核参数? .....	442
14.6.6 DHCP 无法正常获取内网 IP? .....	446
14.6.7 Linux 云服务器网络性能测试方法.....	448
14.6.8 弹性云服务器添加网卡后是否会自动启动? .....	456
14.6.9 怎样检查 IB 云服务器的网络是否正常? .....	456
14.6.10 怎样给 IB 网卡手动配置 IP 地址? .....	457
14.6.11 如何解决 SR-IOV 场景下 win7 弹性云服务器安装 82599 网卡驱动报错? .....	458
14.7 弹性 IP 类.....	459
14.7.1 一台弹性云服务器是否可以绑定多个弹性公网 IP? .....	459
14.7.2 未绑定弹性公网 IP 的弹性云服务器能否访问公网? .....	460
14.7.3 弹性 IP Ping 不通? .....	460
14.7.4 为什么弹性云服务器可以远程连接, 但是无法 ping 通? .....	465
14.8 密码与密钥对.....	465
14.8.1 怎样设置镜像密码的有效期? .....	465
14.8.2 一键重置密码后无法使用新密码登录弹性云服务器.....	466
14.8.3 业务端口被一键式重置密码插件占用.....	468
14.8.4 一键式重置密码插件占用虚拟内存 VIRT 和常驻内存 SHR 过高怎么办? .....	471
14.8.5 怎样查询弹性云服务器使用的密钥对是哪个? .....	472
14.8.6 无法导入密钥对, 怎么办? .....	472
14.8.7 为什么使用密钥文件无法正常登录 Linux 弹性云服务器? .....	473
14.8.8 通过 puttygen.exe 工具创建的密钥对, 导入管理控制台失败怎么办? .....	473
14.8.9 Windows 弹性云服务器中的 cloudbase-init 账户是什么? .....	475
14.8.10 python 升级导致 Cloud-init 不工作.....	476
14.9 应用搭建及软件安装咨询.....	477
14.9.1 弹性云服务器上是否可以搭建数据库? .....	477
14.9.2 弹性云服务器是否支持 Oracle 数据库? .....	477

14.10 文件上传/数据传输.....	477
14.10.1 上传文件到云服务器方式概览.....	477
14.10.2 本地 Windows 主机通过 MSTSC 上传文件到 Windows 云服务器.....	478
14.10.3 本地 Windows 主机使用 WinSCP 上传文件到 Linux 云服务器.....	481
14.10.4 本地 macOS 系统主机上传文件到 Windows 云服务器.....	482
14.10.5 本地 Linux 主机使用 SCP 上传文件到 Linux 云服务器.....	485
14.10.6 本地 Linux 主机使用 SFTP 上传文件到 Linux 云服务器.....	486
14.10.7 本地 Windows 主机使用 FTP 上传文件到 Windows/Linux 云服务器.....	488
14.10.8 本地 Linux 主机使用 FTP 上传文件到 Linux 云服务器.....	489
14.10.9 使用 FTP 上传文件时客户端连接服务端超时怎么办? .....	490
14.10.10 使用 FTP 上传文件时写入失败, 文件传输失败怎么办? .....	490
14.10.11 打开 FTP 服务器上的文件夹时发生错误, 请检查是否有权限访问该文件夹.....	491
14.10.12 WinSCP 无法连接到 Linux 云服务器怎么办? .....	493
14.11 ECS 故障.....	494
14.11.1 针对管理控制台的异常提示信息, 应该如何处理? .....	495
14.11.2 获取控制台日志时, 系统显示问号是什么原因? .....	496
14.11.3 使用 free 命令查看弹性云服务器的内存, 为什么与实际不符? .....	497
14.11.4 弹性云服务器的主机名带后缀 novalocal, 这正常吗? .....	498
14.11.5 怎样使修改的静态主机名永久生效? .....	499
14.11.6 Linux 操作系统云服务器无法获取元数据怎么办? .....	502
14.12 ECS 卡顿.....	503
14.12.1 Windows 云服务器卡顿怎么办? .....	503
14.12.2 Linux 云服务器卡顿怎么办? .....	507
14.13 规格变更.....	510
14.13.1 Windows 弹性云服务器变更规格后数据盘脱机怎么办? .....	510
14.13.2 Linux 弹性云服务器变更规格后磁盘脱机怎么办? .....	513
14.14 操作系统变更.....	514
14.14.1 我能否自己安装或者升级操作系统? .....	514
14.14.2 弹性云服务器能否更换操作系统? .....	515
14.14.3 切换操作系统需要多长时间? .....	515
14.14.4 重装操作系统时是否可以其他操作系统? .....	515
14.14.5 重装操作系统需要多长时间? .....	515

# 1 产品介绍

## 1.1 什么是弹性云服务器？

弹性云服务器（Elastic Cloud Server，ECS）是由CPU、内存、操作系统、云硬盘组成的基础的计算组件。弹性云服务器创建成功后，您就可以像使用自己的本地PC或物理服务器一样，在云上使用弹性云服务器。

弹性云服务器的开通是自助完成的，您只需要指定CPU、内存、操作系统、规格、登录鉴权方式即可，同时也可以根据您的需求随时调整弹性云服务器的规格，为您打造可靠、安全、灵活、高效的计算环境。

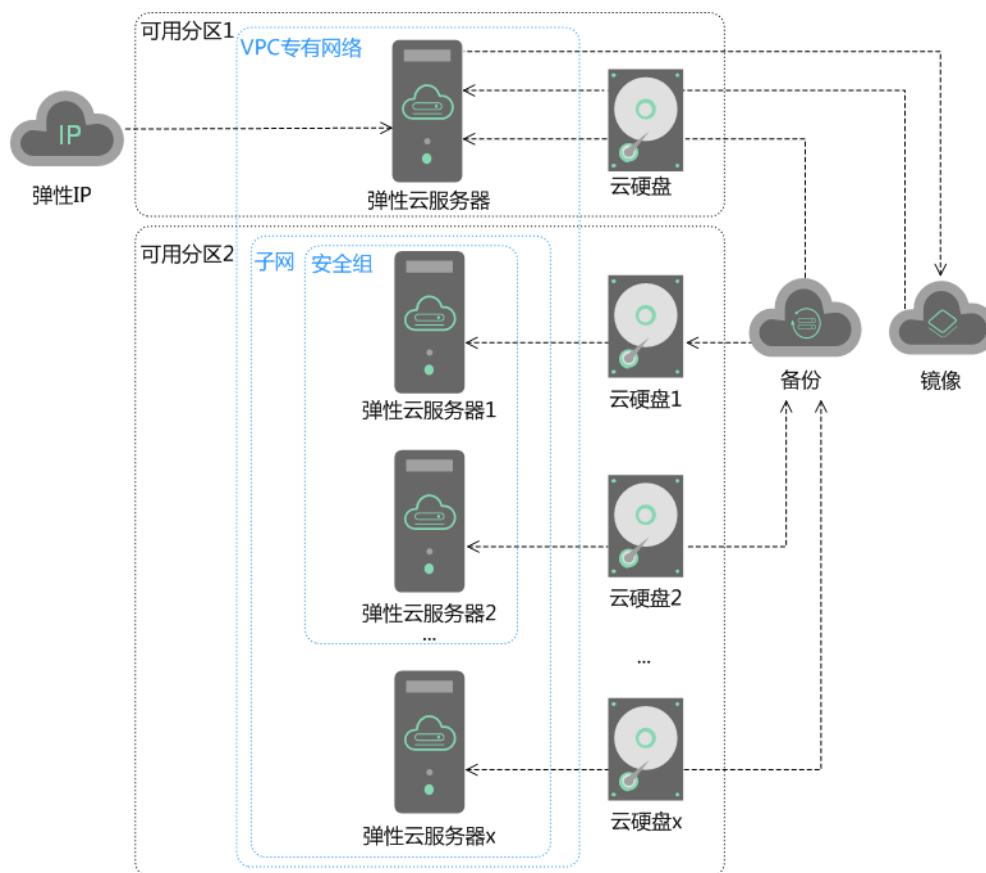
- 操作系统，请参见[镜像类型](#)。
- 登录方式，请参见《弹性云服务器用户指南》的“登录弹性云服务器”。

### 产品架构

通过和其他产品、服务组合，弹性云服务器可以实现计算、存储、网络、镜像安装等功能：

- 弹性云服务器在不同可用区中部署（可用区之间通过内网连接），部分可用区发生故障后不会影响同一区域内的其它可用区。
- 可以通过虚拟私有云建立专属的网络环境，设置子网、安全组，并通过弹性公网IP实现公网链接。
- 通过镜像服务，可以对弹性云服务器安装镜像，也可以通过私有镜像批量创建弹性云服务器，实现快速的业务部署。
- 通过云硬盘服务实现数据存储，并通过云硬盘备份服务实现数据的备份和恢复。
- 云监控是保持弹性云服务器可靠性、可用性和性能的重要部分，通过云监控，用户可以观察弹性云服务器资源。
- 云备份（Cloud Backup and Recovery，CBR）提供对云硬盘和弹性云服务器的备份保护服务，支持利用备份数据恢复服务器和磁盘的数据。

图 1-1 ECS 产品架构



## 访问方式

云服务平台提供了Web化的服务管理平台，即管理控制台和基于HTTPS请求的API（Application programming interface）管理方式。

- API方式  
如果用户需要将云服务平台上的弹性云服务器集成到第三方系统，用于二次开发，请使用API方式访问弹性云服务器，具体操作请参见《弹性云服务器接口参考》。
- 控制台方式  
其他相关操作，请使用管理控制台方式访问弹性云服务器。

## 1.2 产品优势

弹性云服务器可以根据业务需求和伸缩策略，自动调整计算资源。您可以根据自身需要自定义服务器配置，灵活地选择所需的内存、CPU、带宽等配置，帮助您打造可靠、安全、灵活、高效的应用环境。

### 稳定可靠

- 丰富的磁盘种类  
根据IO性能划分云硬盘的磁盘类型，您可根据应用程序要求选择您所需的云硬盘。

更多关于云硬盘规格、性能等信息，请参见《云硬盘用户指南》。

- 高数据可靠性  
基于分布式架构的，可弹性扩展的虚拟块存储服务；具有高数据可靠性，高I/O吞吐能力，能够保证任何一个副本故障时快速进行数据迁移恢复，避免单一硬件故障造成数据丢失。
- 支持云服务器和云硬盘的备份及恢复  
可预先设置好自动备份策略，实现在线自动备份。也可以根据需要随时通过控制台或API，备份云服务器和云硬盘指定时间点的数据。

## 安全保障

- 多种安全服务，多维度防护  
Web应用防火墙、漏洞扫描等多种安全服务提供多维度防护。
- 安全评估  
提供对用户云环境的安全评估，帮助用户快速发现安全弱点和威胁，同时提供安全配置检查，并给出安全实践建议，有效减少或避免由于网络中病毒和恶意攻击带来的损失。
- 智能化进程管理  
提供智能的进程管理服务，基于可定制的白名单机制，自动禁止非法程序的执行，保障弹性云服务器的安全性。
- 漏洞扫描  
支持通用Web漏洞检测、第三方应用漏洞检测、端口检测、指纹识别等多项扫描服务。

## 软硬结合

- 搭载专业的硬件设备  
弹性云服务器搭载在专业的硬件设备上，能够深度进行虚拟化优化技术，用户无需自建机房。
- 随时获取虚拟化资源  
可随时从虚拟资源池中获取并独享资源，并根据业务变化弹性扩展或收缩，像使用本地PC一样在云上使用弹性云服务器，确保应用环境可靠、安全、灵活、高效。

## 弹性伸缩

- 自动调整计算资源  
动态伸缩：基于伸缩组监控数据，随着应用运行状态，动态增加或减少弹性云服务器实例。  
定时伸缩：根据业务预期及运营计划等，制定定时及周期性策略，按时自动增加或减少弹性云服务器实例。
- 灵活调整云服务器配置  
规格、带宽可根据业务需求灵活调整，高效匹配业务要求。



## 1.3 应用场景

### 网站应用

对CPU、内存、硬盘空间和带宽无特殊要求，对安全性、可靠性要求高，服务一般只需要部署在一台或少量的服务器上，一次投入成本少，后期维护成本低的场景。例如网站开发测试环境、小型数据库应用。

推荐使用通用型弹性云服务器，主要提供均衡的计算、内存和网络资源，适用于业务负载压力适中的应用场景，满足企业或个人普通业务搬迁上云需求。

### 企业电商

对内存要求高、数据量大并且数据访问量大、要求快速的数据交换和处理的场景。例如广告精准营销、电商、移动APP。

推荐使用内存优化型弹性云服务器，主要提供高内存实例，同时可以配置超高IO的云硬盘和合适的带宽。

### 图形渲染

对图像视频质量要求高、大内存，大量数据处理，I/O并发能力。可以完成快速的数据处理交换以及大量的GPU计算能力的场景。例如图形渲染、工程制图。

推荐使用GPU加速型弹性云服务器，基于NVIDIA Tesla M60硬件虚拟化技术，提供较为经济的图形加速能力。能够支持DirectX、OpenGL，可以提供最大显存1GiB、分辨率为4096×2160的图形图像处理能力。

### 数据分析

处理大容量数据，需要高I/O能力和快速的数据交换处理能力的场景。例如MapReduce、Hadoop计算密集型。

推荐使用密集存储型弹性云服务器，主要适用于需要对本地存储上的极大型数据集进行高性能顺序读写访问的工作负载，例如：Hadoop分布式计算，大规模的并行数据处理和日志处理应用。主要的数据存储是基于HDD的存储实例，默认配置最高10GE网络能力，提供较高的PPS性能和网络低延迟。最大可支持24个本地磁盘、48个vCPU和384GiB内存。

### 高性能计算

高计算能力、高吞吐量的场景。例如科学计算、基因工程、游戏动画、生物制药计算和存储系统。

推荐使用高性能计算型弹性云服务器，主要使用在受计算限制的高性能处理器的应用程序上，适合要求提供海量并行计算资源、高性能的基础设施服务，需要达到高性能计算和海量存储，对渲染的效率有一定保障的场景。

## 1.4 实例类型和规格

## 1.4.1 实例概述

实例即弹性云服务器，是由CPU、内存、操作系统、云硬盘组成的基础的计算组件。

弹性云服务器创建成功后，您就可以像使用自己的本地PC或物理服务器一样，在云上使用弹性云服务器，打造一个高效、可靠、安全的计算环境。弹性云服务器的开通是自动完成的，您只需要指定CPU、内存、操作系统、规格、登录鉴权方式即可，同时也可以根据您的需求随时调整您的弹性云服务器规格。

云平台提供了多种实例类型供您选择，不同类型的实例可以提供不同的计算能力和存储能力。同一实例类型下可以根据CPU和内存的配置选择不同的实例规格。

- 关于实例类型的信息，请参考[实例类型](#)。
- 了解实例从创建到释放历经的各种状态请参考[实例生命周期](#)。

## 1.4.2 实例生命周期

生命周期是指弹性云服务器从创建到删除（或释放）历经的各种状态。

表 1-1 弹性云服务器状态说明

状态	状态属性	说明
创建中	中间状态	创建弹性云服务器实例后，在弹性云服务器状态进入运行中之前的状态。
正在开机	中间状态	弹性云服务器实例从关机到运行中的中间状态。
运行中	稳定状态	弹性云服务器实例正常运行状态。 在这个状态的实例可以运行您的业务。
正在关机	中间状态	弹性云服务器实例从运行中到关机的中间状态。
关机	稳定状态	弹性云服务器实例被正常停止。 在这个状态下的实例，不能对外提供业务。
重启中	中间状态	弹性云服务器实例正在进行重启操作。
更新规格中	中间状态	弹性云服务器实例接收变更请求，开始进行变更操作。
更新规格校验中	中间状态	弹性云服务器实例正在校验变更完成后的配置。
删除中	中间状态	弹性云服务器实例处于正在被删除的状态。 如果长时间处于该状态，则说明出现异常，需要联系技术支持处理。
已删除	中间状态	弹性云服务器实例已被正常删除。在该状态下的实例，不能对外提供业务，并在短时间内从系统中彻底清除。
故障	稳定状态	弹性云服务器实例处于异常状态。 在这个状态下的实例，不能对外提供业务，需要联系技术支持进行处理。

状态	状态属性	说明
重装操作系统中	中间状态	弹性云服务器实例接收到重装操作系统请求，处于重装操作系统的过程中。
重装操作系统失败	稳定状态	弹性云服务器实例接收到重装操作系统请求，进行重装的过程中发生异常，导致重装失败。 在这个状态下的实例，不能对外提供业务，需要联系技术支持进行处理。
切换操作系统中	中间状态	弹性云服务器实例接收到切换操作系统请求，处于切换操作系统的过程中。
切换操作系统失败	稳定状态	弹性云服务器实例接收到切换操作系统请求，进行切换的过程中发生异常，导致切换失败。 在这个状态下的实例，不能对外提供业务，需要联系技术支持进行处理。
强制重启中	中间状态	弹性云服务器实例正在进行强制重启操作。
更新规格回退中	中间状态	弹性云服务器实例正在回退变更规格的配置。

### 1.4.3 实例类型

云平台提供了几种类型的弹性云服务器供您选择，针对不同的应用场景，可以选择不同规格的弹性云服务器。

- 鲲鹏CPU架构：
  - 鲲鹏通用计算型
  - 鲲鹏通用计算增强型
  - 鲲鹏内存优化型
- x86 CPU架构：
  - 通用型
  - 通用计算增强型
  - 内存优化型
  - 密集存储型
  - 超高I/O型
  - 高计算型
  - GPU加速型
  - AI加速型

### 规格命名规则

AB.C

例如s2.medium

其中，

- A表示系列，例如s表示通用型、c表示计算型、m表示内存型。  
鲲鹏系列规格在系列名称前加k，例如kc表示鲲鹏通用计算增强型。
- B表示系列号，例如s2中2表示计算II型。
- C表示规格，当前系列中的规格大小，例如：medium、large、xlarge。

## vCPU

弹性云服务器的处理器运用超线程HT（Hyper-Threading）技术，允许在CPU的每个物理内核上公开两个执行上下文，即一个物理内核包含两个虚拟的“逻辑内核”，可以处理不同的软件线程。vCPU（virtual CPU）即为虚拟的“逻辑内核”。

当前绝大多数规格已经默认开启了超线程，如果在创建弹性云服务器或者变更规格时关闭了超线程，则在弹性云服务器上查看到的CPU核数是规格的Flavor名称中展示的vCPU数量的一半。

例如，2核的物理CPU包含4个vCPU（线程）。

## 网络带宽

弹性云服务器根据不同的规格限制内网带宽和内网收发包能力。

- 内网基准带宽：指弹性云服务器在整机网络带宽存在争抢时，能稳定达到的保证带宽。
- 内网最大带宽：指弹性云服务器在整机网络带宽没有争抢（宿主机上其他虚拟机对网络带宽要求不高）时，可以达到的最大带宽。
- 内网最大收发包能力：指弹性云服务器能达到的最大收发包能力。

### 📖 说明

- 网络收发包测试方法，请参见《弹性云服务器用户指南》的“网络性能测试方法”。
- 开启网卡多队列的方法，请参见《弹性云服务器用户指南》的“开启网卡多队列功能”。
- 如集群采用10GE组网，则实例规格中的性能指标在极端场景下无法保证。
- 最大带宽是实例维度的，即实例如果有多张网卡，所有网卡的带宽之和不超过实例的最大带宽。

## 1.4.4 实例规格（x86）

### 1.4.4.1 通用型

#### 概述

通用型弹性云服务器主要提供基本水平的vCPU性能，同时可根据工作负载的需要实现性能的突增，具有短期发挥更高性能的能力。适用于那些不会经常（或始终）用尽vCPU性能，但会偶尔突然使用的场景，特别适合通用工作负载，如Web服务器、开发人员环境和小型数据库等。

S7n型弹性云服务器搭载第三代英特尔® 至强® 可扩展处理器，配套25GE智能高速网卡，提供较高网络带宽和PPS收发包能力，提供更高性价比。

S6型弹性云服务器和S6nl型弹性云服务器适合平时不会持续高压使用CPU，但偶尔需要提高计算性能完成工作负载的场景，包括但不限于：轻量级的Web服务器、开发、测试环境以及中低性能数据库等。

GS7型弹性云服务器搭载国产海光C86 7380 CPU，基于全新虚拟化平台，多项技术优化，配套25GE智能高速网卡，提供较高网络带宽和PPS收发包能力，提供更高性价比。

GS6型弹性云服务器搭载国产海光C86 7285 CPU，基于全新虚拟化平台，多项技术优化，配套25GE智能高速网卡，提供较高网络带宽和PPS收发包能力，提供更高性价比。

## 使用场景

- 应用：  
Web服务器、开发测试环境以及小型数据库应用等。  
GS7型和GS6型适用轻量级数据库及缓存服务器、中轻载企业应用等。
- 场景特点：  
vCPU、内存、硬盘空间和带宽等无特殊要求，安全性、可靠性要求高，一次投入成本少，后期维护成本低。
- 适用场景：  
部署企业官网、搭建企业办公环境、支撑企业开发测试活动。  
GS7型和GS6型适用政企金融场景，满足政企、金融行业自主可控诉求。

## 规格

表 1-2 S7n 型云主机规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	虚拟化类型
s7n.small.1	1	1	0.8/0.1	10	1	2	4	KVM
s7n.medium.2	1	2	0.8/0.1	10	1	2	4	KVM
s7n.large.2	2	4	1.5/0.2	15	1	2	8	KVM
s7n.xlarge.2	4	8	2/0.35	25	1	2	16	KVM
s7n.2xlarge.2	8	16	3/0.75	50	2	2	32	KVM
s7n.medium.4	1	4	0.8/0.1	10	1	2	4	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	虚拟化类型
s7n.large.4	2	8	1.5/0.2	15	1	2	8	KVM
s7n.xlarge.4	4	16	2/0.35	25	1	2	16	KVM
s7n.2xlarge.4	8	32	3/0.75	50	2	2	32	KVM

表 1-3 S6 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
s6.small.1	1	1	0.8/0.1	10	1	2	KVM
s6.xlarge.1	4	4	2/0.35	25	1	2	KVM
s6.2xlarge.1	8	8	3/0.75	50	2	2	KVM
s6.medium.2	1	2	0.8/0.1	10	1	2	KVM
s6.large.2	2	4	1.5/0.2	15	1	2	KVM
s6.xlarge.2	4	8	2/0.35	25	1	2	KVM
s6.2xlarge.2	8	16	3/0.75	50	2	2	KVM
s6.4xlarge.2	16	32	6/1.5	100	4	2	KVM
s6.medium.4	1	4	0.8/0.1	10	1	2	KVM
s6.large.4	2	8	1.5/0.2	15	1	2	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
s6.xlarge.4	4	16	2/0.35	25	1	2	KVM
s6.2xlarge.4	8	32	3/0.75	50	2	2	KVM
s6.4xlarge.4	16	64	6/1.5	100	4	2	KVM

表 1-4 S6nl 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
s6nl.small.1	1	1	0.8/0.15	10	1	2	KVM
s6nl.medium.2	1	2	0.8/0.15	10	1	2	KVM
s6nl.large.2	2	4	1.5/0.3	15	1	2	KVM
s6nl.xlarge.2	4	8	2/0.5	25	1	2	KVM
s6nl.2xlarge.2	8	16	3/1	50	2	2	KVM
s6nl.4xlarge.2	16	32	6/1.5	100	4	2	KVM
s6nl.medium.4	1	4	0.8/0.15	10	1	2	KVM
s6nl.large.4	2	8	1.5/0.3	15	1	2	KVM
s6nl.xlarge.4	4	16	2/0.5	25	1	2	KVM
s6nl.2xlarge.4	8	32	3/1	50	2	2	KVM
s6nl.4xlarge.4	16	64	6/1.5	100	4	2	KVM

表 1-5 GS7 型弹性服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
gs7.small.1	1	1	0.8/0.1	10	1	2	KVM
gs7.medium.2	1	2	0.8/0.1	10	1	2	KVM
gs7.large.2	2	4	1/0.2	15	1	2	KVM
gs7.xlarge.2	4	8	1.5/0.3	20	1	2	KVM
gs7.2xlarge.2	8	16	2/0.5	35	2	2	KVM
gs7.medium.4	1	4	0.8/0.1	10	1	2	KVM
gs7.large.4	2	8	1/0.2	15	1	2	KVM
gs7.xlarge.4	4	16	1.5/0.3	20	1	2	KVM
gs7.2xlarge.4	8	32	2/0.5	35	2	2	KVM

表 1-6 GS6 型弹性服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
gs6.small.1	1	1	0.8/0.1	10	1	2	KVM
gs6.medium.2	1	2	0.8/0.1	10	1	2	KVM
gs6.large.2	2	4	1/0.2	15	1	2	KVM
gs6.xlarge.2	4	8	1.5/0.3	20	1	2	KVM
gs6.2xlarge.2	8	16	2/0.5	35	2	2	KVM



规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
gs6.medium.4	1	4	0.8/0.1	10	1	2	KVM
gs6.large.4	2	8	1/0.2	15	1	2	KVM
gs6.xlarge.4	4	16	1.5/0.3	20	1	2	KVM
gs6.2xlarge.4	8	32	2/0.5	35	2	2	KVM

#### 1.4.4.2 通用计算增强型

##### 概述

通用计算增强型弹性云服务器是CPU独享型实例，实例间无CPU资源争抢，性能强劲稳定，搭载全新网络加速引擎，以及DPDK(Data Plane Development Kit)快速报文处理机制，提供更高的网络性能，满足不同场景需求。

- C7n型弹性云服务器搭载第三代英特尔® 至强® 可扩展处理器，在性能、安全、稳定性等方面全面升级，最大核数升级至96U，内存频率升级至3200MHz，提供安全可信的云上环境。
- C6s型弹性云服务器搭载第二代英特尔® 至强® 可扩展处理器，兼具高性能、高稳定性、低时延、高性价比的特点，适用于互联网、游戏、渲染等场景，特别是对计算及网络稳定性有较高要求的场景。
- C6和C6nl型弹性云服务器搭载第二代英特尔® 至强® 可扩展处理器，多项技术优化，计算性能强劲稳定，配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力。
- GC7型弹性云服务器搭载国产海光C86 7380 CPU，基于全新虚拟化平台，多项技术优化，计算性能强劲稳定，配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力。
- GC6型弹性云服务器搭载国产海光C86 7285 CPU，基于全新虚拟化平台，多项技术优化，计算性能强劲稳定，配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力。

## 规格

表 1-7 C7n 型云主机规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
c7n.large.2	2	4	4/0.8	40	2	2	16	1.5/6	KVM
c7n.xlarge.2	4	8	8/1.6	80	2	3	32	2/6	KVM
c7n.2xlarge.2	8	16	15/3	150	4	4	64	3/6	KVM
c7n.3xlarge.2	12	24	17/5	200	4	6	96	4/6	KVM
c7n.4xlarge.2	16	32	20/6	280	8	8	128	5/6	KVM
c7n.6xlarge.2	24	48	25/9	400	8	8	192	6/无	KVM
c7n.8xlarge.2	32	64	30/12	550	16	8	256	8/无	KVM
c7n.12xlarge.2	48	96	35/18	750	16	8	256	12/无	KVM
c7n.16xlarge.2	64	128	36/24	800	28	8	256	16/无	KVM
c7n.24xlarge.2	96	192	40/36	850	32	8	256	24/无	KVM
c7n.large.4	2	8	4/0.8	40	2	2	16	1.5/6	KVM
c7n.xlarge.4	4	16	8/1.6	80	2	3	32	2/6	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
c7n.2xlarge.4	8	32	15/3	150	4	4	64	3/6	KVM
c7n.3xlarge.4	12	48	17/5	200	4	6	96	4/6	KVM
c7n.4xlarge.4	16	64	20/6	280	8	8	128	5/6	KVM
c7n.6xlarge.4	24	96	25/9	400	8	8	192	6/无	KVM
c7n.8xlarge.4	32	128	30/12	550	16	8	256	8/无	KVM
c7n.12xlarge.4	48	192	35/18	750	16	8	256	12/无	KVM
c7n.16xlarge.4	64	256	36/24	800	28	8	256	16/无	KVM
c7n.24xlarge.4	96	384	40/36	850	32	8	256	24/无	KVM

表 1-8 C6s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
c6s.large.2	2	4	1/1	30	2	2	KVM
c6s.xlarge.2	4	8	2/2	60	2	3	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
c6s.2xlarge.2	8	16	4/4	120	4	4	KVM
c6s.3xlarge.2	12	24	5.5/5.5	180	4	6	KVM
c6s.4xlarge.2	16	32	7.5/7.5	240	8	8	KVM
c6s.6xlarge.2	24	48	11/11	350	8	8	KVM
c6s.8xlarge.2	32	64	15/15	450	16	8	KVM
c6s.12xlarge.2	48	96	22/22	650	16	8	KVM
c6s.16xlarge.2	64	128	30/30	850	32	8	KVM
c6s.large.4	2	8	1/1	30	2	2	KVM
c6s.xlarge.4	4	16	2/2	60	2	3	KVM
c6s.2xlarge.4	8	32	4/4	120	4	4	KVM
c6s.3xlarge.4	12	48	5.5/5.5	180	4	6	KVM
c6s.4xlarge.4	16	64	7.5/7.5	240	8	8	KVM
c6s.6xlarge.4	24	96	11/11	350	8	8	KVM
c6s.8xlarge.4	32	128	15/15	450	16	8	KVM
c6s.12xlarge.4	48	192	22/22	650	16	8	KVM
c6s.16xlarge.4	64	256	30/30	850	32	8	KVM

表 1-9 C6 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
c6.large.2	2	4	4/1.2	40	2	2	1/5	KVM
c6.xlarge.2	4	8	8/2.4	80	2	3	1.5/5	KVM
c6.2xlarge.2	8	16	15/4.5	150	4	4	2/5	KVM
c6.3xlarge.2	12	24	17/7	200	4	6	2.5/5	KVM
c6.4xlarge.2	16	32	20/9	280	8	8	3.5/5	KVM
c6.6xlarge.2	24	48	25/14	400	8	8	4/5	KVM
c6.8xlarge.2	32	64	30/18	550	16	8	7/10	KVM
c6.12xlarge.2	48	96	35/27	750	16	8	10/15	KVM
c6.16xlarge.2	64	128	40/36	1000	32	8	20/无	KVM
c6.large.4	2	8	4/1.2	40	2	2	1/5	KVM
c6.xlarge.4	4	16	8/2.4	80	2	3	1.5/5	KVM
c6.2xlarge.4	8	32	15/4.5	150	4	4	2/5	KVM
c6.3xlarge.4	12	48	17/7	200	4	6	2.5/5	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
c6.4xlarge.4	16	64	20/9	280	8	8	3.5/5	KVM
c6.6xlarge.4	24	96	25/14	400	8	8	4/5	KVM
c6.8xlarge.4	32	128	30/18	550	16	8	7/10	KVM
c6.12xlarge.4	48	192	35/27	750	16	8	10/15	KVM
c6.16xlarge.4	64	256	40/36	1000	32	8	20/无	KVM

表 1-10 C6nl 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
c6nl.large.2	2	4	4/1	32	2	2	KVM
c6nl.xlarge.2	4	8	8/2	64	2	3	KVM
c6nl.2xlarge.2	8	16	15/4	120	4	4	KVM
c6nl.3xlarge.2	12	24	17/6	160	4	6	KVM
c6nl.4xlarge.2	16	32	20/8	224	8	8	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
c6nl.6xlarge.2	24	48	25/12	320	8	8	KVM
c6nl.8xlarge.2	32	64	30/16	440	16	8	KVM
c6nl.16xlarge.2	64	128	40/32	800	32	8	KVM
c6nl.large.1	2	2	4/1	32	2	2	KVM
c6nl.xlarge.1	4	4	8/2	64	2	3	KVM
c6nl.2xlarge.1	8	8	15/4	120	4	4	KVM
c6nl.3xlarge.1	12	12	17/6	160	4	6	KVM
c6nl.4xlarge.1	16	16	20/8	224	8	8	KVM
c6nl.large.4	2	8	4/1	32	2	2	KVM
c6nl.xlarge.4	4	16	8/2	64	2	3	KVM
c6nl.2xlarge.4	8	32	15/4	120	4	4	KVM
c6nl.3xlarge.4	12	48	17/6	160	4	6	KVM
c6nl.4xlarge.4	16	64	20/8	224	8	8	KVM
c6nl.6xlarge.4	24	96	25/12	320	8	8	KVM
c6nl.8xlarge.4	32	128	30/16	440	16	8	KVM
c6nl.16xlarge.4	64	256	40/32	800	32	8	KVM

表 1-11 GC7 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
gc7.large.2	2	4	2/0.4	30	2	2	KVM
gc7.xlarge.2	4	8	3/0.8	50	2	3	KVM
gc7.2xlarge.2	8	16	4/1.5	80	4	4	KVM
gc7.3xlarge.2	12	24	5/2	110	4	6	KVM
gc7.4xlarge.2	16	32	7/3	140	8	8	KVM
gc7.6xlarge.2	24	48	8/4	200	8	8	KVM
gc7.8xlarge.2	32	64	10/5	260	16	8	KVM
gc7.12xlarge.2	48	96	15/8	350	16	8	KVM
gc7.16xlarge.2	64	128	18/10	400	32	8	KVM
gc7.large.1	2	2	2/0.4	30	2	2	KVM
gc7.xlarge.1	4	4	3/0.8	50	2	3	KVM
gc7.2xlarge.1	8	8	4/1.5	80	4	4	KVM
gc7.3xlarge.1	12	12	5/2	110	4	6	KVM
gc7.4xlarge.1	16	16	7/3	140	8	8	KVM
gc7.large.4	2	8	2/0.4	30	2	2	KVM
gc7.xlarge.4	4	16	3/0.8	50	2	3	KVM



规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
gc7.2xlarge.4	8	32	4/1.5	80	4	4	KVM
gc7.3xlarge.4	12	48	5/2	110	4	6	KVM
gc7.4xlarge.4	16	64	7/3	140	8	8	KVM
gc7.6xlarge.4	24	96	8/4	200	8	8	KVM
gc7.8xlarge.4	32	128	10/5	260	16	8	KVM
gc7.12xlarge.4	48	192	15/8	350	16	8	KVM
gc7.16xlarge.4	64	256	18/10	400	32	8	KVM

表 1-12 GC6 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
gc6.large.2	2	4	2/0.4	30	2	2	KVM
gc6.xlarge.2	4	8	3/0.8	50	2	3	KVM
gc6.2xlarge.2	8	16	4/1.5	80	4	4	KVM
gc6.3xlarge.2	12	24	5/2	110	4	6	KVM
gc6.4xlarge.2	16	32	7/3	140	8	8	KVM
gc6.6xlarge.2	24	48	8/4	200	8	8	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
gc6.8xlarge.2	32	64	10/5	260	16	8	KVM
gc6.12xlarge.2	48	96	15/8	350	16	8	KVM
gc6.16xlarge.2	64	128	18/10	400	32	8	KVM
gc6.large.e.1	2	2	2/0.4	30	2	2	KVM
gc6.xlarge.1	4	4	3/0.8	50	2	3	KVM
gc6.2xlarge.1	8	8	4/1.5	80	4	4	KVM
gc6.3xlarge.1	12	12	5/2	110	4	6	KVM
gc6.4xlarge.1	16	16	7/3	140	8	8	KVM
gc6.large.e.4	2	8	2/0.4	30	2	2	KVM
gc6.xlarge.e.4	4	16	3/0.8	50	2	3	KVM
gc6.2xlarge.e.4	8	32	4/1.5	80	4	4	KVM
gc6.3xlarge.e.4	12	48	5/2	110	4	6	KVM
gc6.4xlarge.e.4	16	64	7/3	140	8	8	KVM
gc6.6xlarge.e.4	24	96	8/4	200	8	8	KVM
gc6.8xlarge.e.4	32	128	10/5	260	16	8	KVM
gc6.12xlarge.4	48	192	15/8	350	16	8	KVM
gc6.16xlarge.4	64	256	18/10	400	32	8	KVM

## 使用场景

- C7n型：  
适用于对计算与网络有更高性能要求的Web应用、电商平台、短视频平台、在线游戏、保险金融等各类中重载企业应用。
- C6s型：  
适用于互联网、游戏、渲染等场景，特别是对计算及网络稳定性有较高要求的场景。
  - 游戏业务场景：满足游戏行业高性能、高稳定性要求。
  - 渲染场景：优质渲染效果下提供极致性价比。
  - 其他场景：游戏加速器、视频弹幕、建站、APP开发等。
- C6型和C6nl型：  
对计算与网络有更高性能要求的网站和Web应用、通用数据库及缓存服务器、中重载企业应用等。
- GC7型和GC6型：  
适用于政企金融场景，满足政企、金融行业自主可控诉求。  
对计算与网络有更高性能要求的网站和Web应用、通用数据库及缓存服务器、中重载企业应用、游戏、渲染等。

### 1.4.4.3 内存优化型

#### 概述

内存优化型弹性云服务器内存要求高，数据量大并且数据访问量大，同时要求快速的数据交换和处理。适用于广告精准营销、电商、车联网等大数据分析场景。

- M7n型弹性云服务器搭载第三代英特尔® 至强® 可扩展处理器，在性能、安全、稳定性等方面全面升级，最大核数升级至96U，内存频率升级至3200MHz；提供安全可信的云上环境，适用于高内存计算应用。
- M6s型弹性云服务器搭载第二代英特尔® 至强® 可扩展处理器，多项技术优化，计算性能强劲稳定。配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力；提供最大512GiB基于DDR4的内存实例，适用于高内存计算应用。
- M6型弹性云服务器和M6nl型弹性云服务器搭载第二代英特尔® 至强® 可扩展处理器，多项技术优化，计算性能强劲稳定。配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力；提供最大512GiB基于DDR4的内存实例，适用于高内存计算应用。
- GM7型弹性云服务器搭载国产海光C86 7380 CPU，基于全新虚拟化平台，多项技术优化，配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力；提供最大512GiB基于DDR4的内存实例，适用于高内存计算应用。
- GM6型弹性云服务器搭载国产海光C86 7285 CPU，基于全新虚拟化平台，多项技术优化，配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力；提供最大512GiB基于DDR4的内存实例，适用于高内存计算应用。

#### 使用场景

- 应用：

关系数据库和NoSQL数据库、内存数据分析等。

GM7型和GM6型适用大规模并行处理 (MPP) 数据仓库、MapReduce和Hadoop 分布式计算、分布式文件系统、网络文件系统、日志或数据处理等应用。

- 场景特点：  
内存要求高，同时要求内存优化。
- 适用场景：  
大数据分析，如广告精准营销、电商、车联网等高性能数据库场景。  
GM7型和GM6型适用政企金融场景，满足政企、金融行业自主可控诉求。

## 规格

表 1-13 M7n 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
m7n.large.8	2	16	4/0.8	40	2	2	16	1.5/6	KVM
m7n.xlarge.8	4	32	8/1.6	80	2	3	32	2/6	KVM
m7n.2xlarge.8	8	64	15/3	150	4	4	64	3/6	KVM
m7n.3xlarge.8	12	96	17/5	200	4	6	96	4/6	KVM
m7n.4xlarge.8	16	128	20/6	280	8	8	128	5/6	KVM
m7n.6xlarge.8	24	192	25/9	400	8	8	192	6/无	KVM
m7n.8xlarge.8	32	256	30/12	550	16	8	256	8/无	KVM
m7n.12xlarge.8	48	384	35/18	750	16	8	256	12/无	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
m7n.16xlarge.8	64	512	36/24	800	28	8	256	16/无	KVM
m7n.24xlarge.8	96	768	40/36	850	32	8	256	24/无	KVM

表 1-14 M6 型弹性服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
m6.large.8	2	16	4/1.2	40	2	2	1/5	KVM
m6.xlarge.8	4	32	8/2.4	80	2	3	1.5/5	KVM
m6.2xlarge.8	8	64	15/4.5	150	4	4	2/5	KVM
m6.3xlarge.8	12	96	17/7	200	4	6	2.5/5	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	云硬盘基础带宽/突发带宽 (Gbps)	虚拟化类型
m6.4xlarge.8	16	128	20/9	280	8	8	3.5/5	KVM
m6.6xlarge.8	24	192	25/14	400	8	8	4/5	KVM
m6.8xlarge.8	32	256	30/18	550	16	8	7/10	KVM
m6.16xlarge.8	64	512	40/36	1000	32	8	20/无	KVM

表 1-15 M6s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
m6s.large.8	2	16	3/1	30	2	2	KVM
m6s.xlarge.8	4	32	6/2	60	2	3	KVM
m6s.2xlarge.8	8	64	12/4	120	4	4	KVM
m6s.3xlarge.8	12	96	14/5.5	160	4	6	KVM
m6s.4xlarge.8	16	128	16/7.5	220	8	8	KVM
m6s.6xlarge.8	24	192	20/11	320	8	8	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发 包能力 (万 PPS)	网卡多 队列数	网卡 个数 上限	虚拟 化类 型
m6s.8xlarge.8	32	256	25/15	450	16	8	KVM
m6s.16xlarge.8	64	512	34/30	850	32	8	KVM

表 1-16 M6nl 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发 包能力 (万 PPS)	网卡多 队列数	网卡 个数 上限	虚拟 化类 型
m6nl.large.8	2	16	4/1	32	2	2	KVM
m6nl.xlarge.8	4	32	8/2	64	2	3	KVM
m6nl.2xlarge.8	8	64	15/4	120	4	4	KVM
m6nl.3xlarge.8	12	96	17/6	160	4	6	KVM
m6nl.4xlarge.8	16	128	20/8	224	8	8	KVM
m6nl.6xlarge.8	24	192	25/12	320	8	8	KVM
m6nl.8xlarge.8	32	256	30/16	440	16	8	KVM
m6nl.16xlarge.8	64	512	40/32	800	32	8	KVM

表 1-17 GM7 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发 包能力 (万 PPS)	网卡多 队列数	网卡 个数上 限	虚拟化 类型
gm7.large.8	2	16	2/0.4	30	2	2	KVM
gm7.xlarge.8	4	32	3/0.8	50	2	3	KVM
gm7.2xlarge.8	8	64	4/1.5	80	4	4	KVM
gm7.3xlarge.8	12	96	5/2	110	4	6	KVM
gm7.4xlarge.8	16	128	7/3	140	8	8	KVM
gm7.6xlarge.8	24	192	8/4	200	8	8	KVM
gm7.8xlarge.8	32	256	10/5	260	16	8	KVM
gm7.16xlarge.8	64	512	18/10	400	32	8	KVM

表 1-18 GM6 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发 包能力 (万 PPS)	网卡多 队列数	网卡 个数上 限	虚拟化 类型
gm6.large.8	2	16	2/0.4	30	2	2	KVM
gm6.xlarge.8	4	32	3/0.8	50	2	3	KVM
gm6.2xlarge.8	8	64	4/1.5	80	4	4	KVM
gm6.3xlarge.8	12	96	5/2	110	4	6	KVM



规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发 包能力 (万 PPS)	网卡多 队列数	网 卡 个 数 上 限	虚拟化 类型
gm6.4xlarge.8	16	128	7/3	140	8	8	KVM
gm6.6xlarge.8	24	192	8/4	200	8	8	KVM
gm6.8xlarge.8	32	256	10/5	260	16	8	KVM
gm6.16xlarge.8	64	512	18/10	400	32	8	KVM

#### 1.4.4.4 密集存储型

##### 概述

密集存储型弹性云服务器自带高存储带宽和IOPS的本地盘，具有高存储IOPS以及读写带宽的优势。同时，本地盘的价格更加低廉，在海量数据存储场景下，具备更高的性价比。密集存储型弹性云服务器具备如下特点：

- 本地磁盘提供更高顺序读写性能和更低时延，提升文件读写性能。
- 提供强大而稳定的计算能力，保障计算作业的高效处理效率。
- 提供更高的内网性能，包括高内网带宽和pps(packet per second)，满足业务高峰期弹性云服务器间数据交互需求。

D7型弹性云服务器搭载第三代英特尔® 至强® 可扩展处理器，在性能、安全、稳定性等方面全面升级，计算性能强劲稳定，内存频率升级至3200MHz，配套25GE智能高速网卡，提供超高网络带宽和PPS收发包能力；配套有本地SATA盘，单盘容量高达3600GiB，最大支持32 × 3600GiB本地盘容量。

D6型弹性云服务器搭载第二代英特尔® 至强® 可扩展处理器，计算性能强劲稳定，提供1:4的vCPU和内存比实例。配套自研25GE智能高速网卡，提供超高网络带宽和PPS收发包能力；配套有本地SATA盘，单盘提升到3600GiB，最大支持36 × 3600GiB本地盘容量。

D3型弹性云服务器搭载英特尔® 至强® 可扩展处理器，计算性能强劲稳定。配套自研25GE智能高速网卡，提供超高网络带宽和PPS收发包能力；配套有本地SAS盘，提供更高带宽的本地存储。

##### 使用场景

- 应用：大规模并行处理 (MPP) 数据仓库，MapReduce和Hadoop分布式计算，大数据计算。
- 场景特点：适合处理海量数据、需要高I/O能力，要求快速数据交换和处理的场景。

- 使用场景：分布式文件系统，网络文件系统、日志或数据处理应用。

## 规格

表 1-19 D7 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	本地盘 (GiB)	虚拟化类型
d7.xlarge.4	4	16	5/1.7	60	2	3	32	2 × 3600	KVM
d7.2xlarge.4	8	32	10/3.5	120	4	4	64	4 × 3600	KVM
d7.4xlarge.4	16	64	20/6.7	240	4	6	96	8 × 3600	KVM
d7.6xlarge.4	24	96	25/10	350	8	8	128	12 × 3600	KVM
d7.8xlarge.4	32	128	30/13.5	450	8	8	192	16 × 3600	KVM
d7.12xlarge.4	48	192	40/20	650	16	8	256	24 × 3600	KVM
d7.16xlarge.4	64	256	42/27	850	16	8	256	32 × 3600	KVM

表 1-20 D6 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	本地盘 (GiB)	虚拟化类型
d6.xlarge.4	4	16	5/2	60	2	3	2 × 3600	KVM
d6.2xlarge.4	8	32	10/4	120	4	4	4 × 3600	KVM
d6.4xlarge.4	16	64	20/7.5	240	8	8	8 × 3600	KVM
d6.6xlarge.4	24	96	25/11	350	8	8	12 × 3600	KVM
d6.8xlarge.4	32	128	30/15	450	16	8	16 × 3600	KVM
d6.12xlarge.4	48	192	40/22	650	16	8	24 × 3600	KVM
d6.16xlarge.4	64	256	42/30	850	32	8	32 × 3600	KVM
d6.18xlarge.4	72	288	44/34	900	32	8	36 × 3600	KVM

表 1-21 D3 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	本地盘 (GiB)	虚拟化类型
d3.xlarge.8	4	32	2.5/2.5	50	2	3	2 × 1675	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	本地盘 (GiB)	虚拟化类型
d3.2xlarge.8	8	64	5/5	100	2	4	4 × 1675	KVM
d3.4xlarge.8	16	128	10/10	120	4	8	8 × 1675	KVM
d3.6xlarge.8	24	192	15/15	160	6	8	12 × 1675	KVM
d3.8xlarge.8	32	256	20/20	200	8	8	16 × 1675	KVM
d3.12xlarge.8	48	384	32/32	220	16	8	24 × 1675	KVM
d3.14xlarge.10	56	560	40/40	500	16	8	28 × 1675	KVM

#### 1.4.4.5 超高 I/O 型

##### 概述

超高I/O型弹性云服务器使用高性能NVMe SSD本地磁盘，提供高存储IOPS以及低读写时延，您可以通过管理控制台创建挂载有高性能NVMe SSD盘的弹性云服务器。

##### 使用场景

- 超高I/O型弹性云服务器适用于高性能关系型数据库
- NoSQL数据库(Cassandra、MongoDB等)以及ElasticSearch搜索等场景

## 规格

表 1-22 Ir7 型弹性服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	本地盘 (GiB)	虚拟化类型
ir7.large.4	2	8	3/0.8	40	2	3	32	2 × 50	KVM
ir7.xlarge.4	4	16	6/1.5	80	2	3	32	2 × 100	KVM
ir7.2xlarge.4	8	32	15/3.1	150	4	4	64	2 × 200	KVM
ir7.4xlarge.4	16	64	20/6.2	300	4	6	96	2 × 400	KVM
ir7.8xlarge.4	32	128	30/12	400	8	8	192	2 × 800	KVM
ir7.16xlarge.4	64	256	40/25	600	16	8	256	2 × 1600	KVM

表 1-23 Ir7n 型弹性服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	本地盘 (GiB)	虚拟化类型
ir7n.large.4	2	8	3/0.9	40	2	3	32	2 × 50	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	本地盘 (GiB)	虚拟化类型
ir7n.xlarge.4	4	16	6/1.8	80	2	3	32	2 × 100	KVM
ir7n.2xlarge.4	8	32	15/3.6	150	4	4	64	2 × 200	KVM
ir7n.4xlarge.4	16	64	20/7.3	300	4	6	96	2 × 400	KVM
ir7n.8xlarge.4	32	128	30/14.5	400	8	8	192	2 × 800	KVM
ir7n.16xlarge.4	64	256	40/29	600	16	8	256	2 × 1600	KVM

表 1-24 I7n 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	本地盘 (GiB)	虚拟化类型
i7n.2xlarge.4	8	32	10/3.4	120	4	4	64	1 × 1600GiB NVMe	KVM
i7n.4xlarge.4	16	64	15/6.7	200	4	6	96	2 × 1600GiB NVMe	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	辅助网卡个数上限	本地盘 (GiB)	虚拟化类型
i7n.8xlarge.4	32	128	25/13.5	400	8	8	192	4 × 1600GiB NVMe	KVM
i7n.12xlarge.4	48	192	30/20	500	16	8	256	6 × 1600GiB NVMe	KVM
i7n.16xlarge.4	64	256	35/27	600	16	8	256	8 × 1600GiB NVMe	KVM
i7n.24xlarge.4	96	420	44/20	800	32	8	256	12 × 1600GiB NVMe	KVM

表 1-25 I3 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	本地盘	虚拟化类型
i3.2xlarge.8	8	64	2.5/2.5	100	4	4	1 × 1600GiB NVMe	KVM
i3.4xlarge.8	16	128	5/5	150	4	8	2 × 1600GiB NVMe	KVM
i3.8xlarge.8	32	256	10/10	200	8	8	4 × 1600GiB NVMe	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	本地盘	虚拟化类型
i3.12xlarge.8	48	384	15/15	240	8	8	6 × 1600GiB NVMe	KVM
i3.15xlarge.8	60	512	25/25	500	16	8	7 × 1600GiB NVMe	KVM
i3.16xlarge.8	64	512	25/25	500	16	8	8 × 1600GiB NVMe	KVM

## 本地盘性能

Ir7型弹性云服务器本地盘的IOPS(Input/Output Operations Per Second)性能如表 1-26所示。

表 1-26 Ir7 型弹性云服务器本地盘 IOPS 性能

规格名称	4KB随机读最大IOPS
ir7.large.4	28125
ir7.xlarge.4	56250
ir7.2xlarge.4	112500
ir7.4xlarge.4	225000
ir7.8xlarge.4	450000
ir7.16xlarge.4	900000

Ir7n型弹性云服务器本地盘的IOPS(Input/Output Operations Per Second)性能如表 1-27所示。

表 1-27 Ir7n 型弹性云服务器本地盘 IOPS 性能

规格名称	4KB随机读最大IOPS
ir7n.large.4	28125



规格名称	4KB随机读最大IOPS
ir7n.xlarge.4	56250
ir7n.2xlarge.4	112500
ir7n.4xlarge.4	225000
ir7n.8xlarge.4	450000
ir7n.16xlarge.4	900000

I7n型弹性云服务器本地盘的IOPS(Input/Output Operations Per Second)性能和单盘指标如表1-28和表1-29所示。

表 1-28 I7n 型弹性云服务器本地盘 IOPS 性能

规格名称	4KB随机读最大IOPS
i7n.2xlarge.4	900000
i7n.4xlarge.4	1800000
i7n.8xlarge.4	3600000
i7n.12xlarge.4	5400000
i7n.16xlarge.4	7200000
i7n.24xlarge.4	10800000

表 1-29 I7n 型弹性云服务器本地盘单盘指标

指标	性能
磁盘容量	1.6T
读IOPS (4KB随机读)	900000
写IOPS (4KB随机写)	250000
读吞吐量	6.2 GiB/s
写吞吐量	2.1 GiB/s
访问时延	微秒级

I3型弹性云服务器本地盘的IOPS(Input/Output Operations Per Second)性能和单盘指标如表1-30和表1-31所示。

表 1-30 I3 型弹性云服务器本地盘 IOPS 性能

规格名称	4KB随机读最大IOPS
i3.2xlarge.8	750000
i3.4xlarge.8	1500000
i3.8xlarge.8	3000000
i3.12xlarge.8	4500000
i3.15xlarge.8	5250000
i3.16xlarge.8	6000000

表 1-31 I3 型弹性云服务器本地盘单盘指标

指标	性能
磁盘容量	1.6T
读IOPS（4KB随机读）	750000
写IOPS（4KB随机写）	200000
读吞吐量	2.9 GiB/s
写吞吐量	1.9 GiB/s
访问时延	微秒级

## 使用须知

- 超高I/O型弹性云服务器当前支持如下版本的操作系统：
  - EulerOS 2.2
  - CentOS 7.2
  - CentOS 7.3
  - Ubuntu Server 16.04
  - SUSE Linux Enterprise Server 12 SP2
  - Fedora 25 64bit
  - OpenSUSE 42.2 64bit

### 说明

建议使用EulerOS 2.2和Ubuntu Server 16.04操作系统。

- 超高I/O型弹性云服务器所在的物理机发生故障时，不支持弹性云服务器的热迁移恢复：
  - 部分宿主机硬件故障或亚健康等场景，需要用户配合关闭ECS完成宿主机硬件维修动作。
  - 因系统维护或硬件故障等，HA重新部署ECS实例后，实例会冷迁移到其他宿主机，本地盘数据不保留。

- 超高I/O型弹性云服务器不支持规格变更。
- 超高I/O型弹性云服务器不支持本地盘的快照和备份。
- 可使用本地盘和云硬盘两类磁盘存储数据，通过挂载云硬盘，可以提供更大的存储空间。关于本地盘和云硬盘的使用，有如下约束与限制：
  - 系统盘只能部署在云硬盘上，不可以部署在本地盘上。
  - 数据盘可以部署在云硬盘和本地盘上。
  - 最多可以挂载60块盘（包括VBD盘+SCSI盘+本地盘）。
- 您可以通过配置fstab文件，设置云服务器系统启动时自动挂载磁盘分区。具体操作请参见《弹性云服务器用户指南》的“设置开机自动挂载磁盘分区”。
- 超高I/O型弹性云服务器的本地磁盘数据有丢失的风险（比如宿主机宕机或本地磁盘损坏时），如果您的应用不能做到数据可靠性的架构，强烈建议您使用云盘搭建您的弹性云服务器。
- 删除超高I/O型弹性云服务器后，本地NVMe SSD盘中的数据会被自动清除，请提前做好数据备份。删除本地盘数据的时间较长，因此，资源释放的时间较之常规云服务器略长。
- 由于本地盘数据的可靠性取决于物理服务器和硬盘的可靠性，存在单点故障风险，建议您在应用层做好数据冗余，以保证数据的可用性，需要长期保存的业务数据建议使用云硬盘存储。
- 超高I/O型弹性云服务器的本地盘设备名为/dev/nvme0n1、/dev/nvme0n2等。
- lr3型弹性云服务器的本地盘为拆分型本地盘，一块本地盘可能被多个弹性云服务器使用。当本地盘损坏时，会影响多个弹性云服务器。  
建议您在创建lr3型弹性云服务器时，将弹性云服务器加入云服务器组，避免出现一块本地盘损坏影响多个弹性服务器的情况。具体操作请参见《弹性云服务器用户指南》的“管理云服务器组”。
- 对于超高I/O型弹性云服务器，关机后其基础资源（包括vCPU、内存、镜像）会继续收费。如需停止计费，需删除弹性云服务器。

#### 1.4.4.6 高计算型

##### 概述

- HC2型（高性能计算型II代）弹性云服务器的处理器与内存配比为1:2或者1:4，每一个vCPU都对应一个Intel Xeon处理器核心的超线程，主要适用于高性能计算业务场景，能够提供海量并行计算资源和高性能的基础设施服务，达到高性能计算和海量存储的要求，保障渲染效率。常用场景包括：
  - 基因工程、游戏动画、生物制药的计算和存储系统。
  - 渲染农场、动漫影视基地等公共渲染平台，以及影视等渲染平台。
  - 高性能前端集群、Web 服务器、高性能科学和工程应用、广告服务、视频编码和分布式分析。
- H2型弹性云服务器主要用于满足高端计算（例如分子建模、计算流体力学）的需要，除了提供强大的CPU能力外，H2型云服务器还提供了多种选择，使用EDR InfiniBand网卡组成的低延迟RDMA网络，支持内存密集的计算需求。
- HL1型弹性云服务器内存容量大、支持云服务器间100Gbps RDMA InfiniBand互联，同时可以挂载56GiB的高I/O共享存储。
- H3型弹性云服务器搭载高性能的英特尔® 至强® 可扩展处理器，每一个vCPU都对应一个Intel Xeon处理器核心的超线程，计算性能稳定，主要适用于高性能计算业

务场景，同时搭载全新网络加速引擎，以及DPDK(Data Plane Development Kit)快速报文处理机制，以高速稳定的网络性能保障应用运行。

## 规格

表 1-32 高计算型弹性云服务器的规格

类型	vCPU	内存 (GiB)	规格名称	虚拟化类型
HC2型	2	4	hc2.large.2	KVM
	4	8	hc2.xlarge.2	KVM
	8	16	hc2.2xlarge.2	KVM
	16	32	hc2.4xlarge.2	KVM
	32	64	hc2.8xlarge.2	KVM
	2	8	hc2.large.4	KVM
	4	16	hc2.xlarge.4	KVM
	8	32	hc2.2xlarge.4	KVM
	16	64	hc2.4xlarge.4	KVM
	32	128	hc2.8xlarge.4	KVM
HL1型	32	256	hl1.8xlarge.8	KVM
H2型	16	128	h2.4xlarge.8	KVM
	16	256	h2.4xlarge.16	KVM
H3型	2	4	h3.large.2	KVM
	4	8	h3.xlarge.2	KVM
	8	16	h3.2large.2	KVM
	12	24	h3.3xlarge.2	KVM
	16	32	h3.4large.2	KVM
	24	48	h3.6xlarge.2	KVM
	32	64	h3.8large.2	KVM
	2	8	h3.large.4	KVM
	4	16	h3.xlarge.4	KVM
	8	32	h3.2large.4	KVM
	12	48	h3.3xlarge.4	KVM
	16	64	h3.4large.4	KVM
	24	96	h3.6xlarge.4	KVM

类型	vCPU	内存 (GiB)	规格名称	虚拟化类型
	32	128	h3.8large.4	KVM

## 使用场景

- 应用：  
H2型、HL1型：高性能计算（High-performance computing, HPC）、大数据、人工智能（Artificial Intelligence, AI）
- 场景特点：  
H2型、HL1型：
  - 高性能硬件：内存、vCPU比为8:1，且具备大量多线程的物理CPU，从而提供高性能的存储I/O和高吞吐量的网络连接。
  - 专为HPC集群设计：通过配置多个HL1型弹性云服务器构建集群，该集群可以用于安装可扩展的集群文件系统（例如Lustre）。在H2型云服务器上运行的HPC应用，可以读取并修改存储在其中的数据。
  - RDMA网络连接：与H2型弹性云服务器一样，HL1型云服务器使用EDR 100Gbps 网卡组成的低延迟RDMA网络，并可以通过RDMA协议与H2型云服务器通信。此外，为了实现高存储I/O，HL1型云服务器可以通过RDMA协议访问云硬盘，存储路径的带宽最大可达56Gbps。
- 适用场景：  
H2型、HL1型：适合内存大、节点与节点间连接性好、且高存储I/O的高集群计算需求，包括HPC、大数据、AI。在HPC解决方案中，HL1型云服务器非常适合于Lustre并行分布式文件系统，常用于大规模集群计算。  
例如在HPC场景，H2型云服务器可以作为计算节点，HL1型云服务器可以作为存储节点使用。

## 功能

高计算型弹性云服务器功能如下：

- 提供大量内存和处理器数
- 最大可支持32个vCPU
- H2型、HL1型弹性云服务器使用带宽为100Gbps的IB网卡
- HL1型弹性云服务器支持挂载如下类型的云硬盘作为系统盘和数据盘：
  - 高IO(性能优化I型)
  - 超高IO(时延优化)
- HL1型弹性云服务器可以挂载56GiB的高I/O共享存储  
HL1型云服务器通过挂载高IO（性能优化I型）、超高IO（时延优化）两种类型的云硬盘，无需做其他配置操作，即可实现挂载56GiB的高I/O共享存储。

## H2 型云服务器使用须知

- H2型云服务器不支持操作系统的重装、切换功能。
- H2型云服务器不支持规格变更。

- H2型云服务器不支持冷迁移、热迁移、HA（High Availability）。
- H2型云服务器当前支持如下版本的操作系统：
  - 对于公共镜像，支持如下版本的操作系统：
    - CentOS 7.3 64bit
    - SUSE Linux Enterprise Server 11 SP4 64bit
    - SUSE Linux Enterprise Server 12 SP2 64bit
  - 对于私有镜像，支持如下版本的操作系统：
    - CentOS 6.5 64bit
    - CentOS 7.2 64bit
    - CentOS 7.3 64bit
    - SUSE Linux Enterprise Server 11 SP4 64bit
    - SUSE Linux Enterprise Server 12 SP2 64bit
    - Red Hat Enterprise Linux 7.2 64bit
    - Red Hat Enterprise Linux 7.3 64bit
- H2型云服务器使用带宽100Gbps的IB网卡。
- H2型云服务器使用一块PCIe 3.2TB SSD卡作为本地临时存储。
- 使用私有镜像创建的H2型云服务器，需在云服务器创建完成后安装InfiniBand网卡驱动。请根据IB类型，在Mellanox官网选择相应版本的InfiniBand网卡驱动下载（建议选择4.2-1.0.0.0的驱动版本下载使用），并根据Mellanox提供的操作指导进行安装。
  - IB网卡类型：“Mellanox Technologies ConnectX-4 Infiniband HBA (MCX455A-ECAT)”
  - Mellanox官网地址：<http://www.mellanox.com/>
  - 网卡驱动下载地址：[https://network.nvidia.com/products/infiniband-drivers/linux/mlnx\\_ofed/](https://network.nvidia.com/products/infiniband-drivers/linux/mlnx_ofed/)
- 使用SUSE操作系统的H2型云服务器，如需使用IPoIB（IP over IB）功能，需在安装InfiniBand网卡驱动后，给IB网卡手动配置一个IP地址。配置操作请参见《弹性云服务器用户指南》的“怎样给IB网卡手动配置IP地址？”。
- 删除H2型云服务器后，SSD磁盘中的数据会被自动清除。因此，在使用H2型云服务器的过程中，请勿将持久性数据保存至SSD磁盘。

## HL1 型云服务器使用须知

- 除高I/O（性能优化I型）、超高I/O（时延优化）两种类型的云硬盘，HL1型云服务器不支持挂载其他类型的云硬盘。

HL1型云服务器通过挂载高I/O（性能优化I型）、超高I/O（时延优化）两种类型的云硬盘，无需做其他配置操作，即可实现挂载56GiB的高I/O共享存储。
- HL1型云服务器不支持规格变更。
- HL1型云服务器使用带宽100Gbps的IB网卡。

- 使用私有镜像创建的HL1型云服务器，需安装InfiniBand网卡驱动。请根据IB类型，在Mellanox官网选择相应版本的InfiniBand网卡驱动下载（建议选择4.2-1.0.0.0的驱动版本下载使用），并根据Mellanox提供的操作指导进行安装。
  - IB网卡类型：“Mellanox Technologies ConnectX-4 Infiniband HBA (MCX455A-ECAT)”
  - Mellanox官网地址：<http://www.mellanox.com/>
- 使用SUSE操作系统的HL1型云服务器，如需使用IPoIB（IP over IB）功能，需在安装InfiniBand网卡驱动后，给IB网卡手动配置一个IP地址。配置操作请参见《弹性云服务器用户指南》的“怎样给IB网卡手动配置IP地址？”。
- HL1型云服务器当前支持如下版本的操作系统：
  - 对于公共镜像，支持如下版本的操作系统：
    - CentOS 7.3 64bit
    - SUSE Linux Enterprise Server 11 SP4 64bit
    - SUSE Linux Enterprise Server 12 SP2 64bit
  - 对于私有镜像，支持如下版本的操作系统：
    - CentOS 6.5 64bit
    - CentOS 7.2 64bit
    - CentOS 7.3 64bit
    - SUSE Linux Enterprise Server 11 SP4 64bit
    - SUSE Linux Enterprise Server 12 SP2 64bit
    - Red Hat Enterprise Linux 7.2 64bit
    - Red Hat Enterprise Linux 7.3 64bit

#### 1.4.4.7 GPU 加速型

GPU加速型云服务器（GPU Accelerated Cloud Server，GACS）能够提供强大的浮点计算能力，从容应对高实时、高并发的大量计算场景。

GPU加速型云服务器包括G系列和P系列两类。其中：

- G系列：图形加速型弹性云服务器，适合于3D动画渲染、CAD等。
- P系列：计算加速型或推理加速型弹性云服务器，适合于深度学习、科学计算、CAE等。

#### GPU 加速实例总览

在售：除主售外的其他GPU机型均为在售机型，如果在售机型售罄，推荐使用主售机型

- 图像加速G系列
  - [图形加速增强型G7r](#)
  - [图形加速增强型G7v](#)

- [图形加速增强型G7](#)
- [图形加速增强型G6](#)
- [图形加速增强型G5r](#)
- [图形加速增强型G5](#)
- 计算加速P系列
  - [计算加速型P3v](#)
  - [计算加速型P2s](#)（主售）
  - [推理加速型Pi3](#)
  - [推理加速型Pi2](#)（主售）
  - [推理加速型Pi2nl](#)

相关操作链接：

- 《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”
- 《弹性云服务器用户指南》的“GPU加速型实例安装Tesla驱动及CUDA工具包”

## GPU 加速型实例支持的镜像

表 1-33 GPU 加速型实例支持的镜像

类别	实例	支持的镜像
图形加速型	G7r	云桌面提供
图形加速型	G7v	<ul style="list-style-type: none"><li>• CentOS 8.2 64bit</li><li>• CentOS 7.6 64bit</li><li>• Ubuntu 20.04 Server 64bit</li><li>• Ubuntu 18.04 Server 64bit</li><li>• Windows Server 2019 Standard 64bit</li><li>• Windows Server 2016 Standard 64bit</li></ul>
图形加速型	G7	<ul style="list-style-type: none"><li>• CentOS 8.2 64bit</li><li>• CentOS 7.6 64bit</li><li>• Ubuntu 20.04 Server 64bit</li><li>• Ubuntu 18.04 Server 64bit</li><li>• Windows Server 2019 Standard 64bit</li><li>• Windows Server 2016 Standard 64bit</li></ul>
图形加速型	G6	<ul style="list-style-type: none"><li>• CentOS 8.2 64bit</li><li>• CentOS 7.6 64bit</li><li>• Ubuntu 20.04 64bit</li><li>• Ubuntu 18.04 64bit</li><li>• Windows Server 2019 Standard 64bit</li><li>• Windows Server 2016 Standard 64bit</li></ul>



类别	实例	支持的镜像
图形加速型	G5r	<ul style="list-style-type: none"><li>Windows Server 2016 Standard 64bit</li><li>Windows Server 2012 R2 Standard 64bit</li></ul>
图形加速型	G5	<ul style="list-style-type: none"><li>CentOS 8.2 64bit</li><li>CentOS 7.6 64bit</li><li>CentOS 7.5 64bit</li><li>Ubuntu 20.04 64bit</li><li>Ubuntu 18.04 64bit</li><li>Windows Server 2019 Standard 64bit</li><li>Windows Server 2016 Standard 64bit</li><li>Windows Server 2019 Datacenter 64bit</li><li>Windows Server 2016 Datacenter 64bit</li></ul>
计算加速型	P3v	<ul style="list-style-type: none"><li>CentOS 8.2 64bit</li><li>CentOS 8.1 64bit</li><li>CentOS 8.0 64bit</li><li>CentOS 7.9 64bit</li><li>CentOS 7.8 64bit</li><li>CentOS 7.7 64bit</li><li>CentOS 7.6 64bit</li><li>Ubuntu 20.04 server 64bit</li><li>Ubuntu 18.04 server 64bit</li></ul>
计算加速型	P2s	<ul style="list-style-type: none"><li>Windows Server 2016 Standard 64bit</li></ul>
推理加速型	Pi3	<ul style="list-style-type: none"><li>CentOS 8.2 64bit</li><li>CentOS 8.1 64bit</li><li>CentOS 8.0 64bit</li><li>CentOS 7.9 64bit</li><li>CentOS 7.8 64bit</li><li>CentOS 7.7 64bit</li><li>CentOS 7.6 64bit</li><li>Ubuntu 20.04 server 64bit</li><li>Ubuntu 18.04 server 64bit</li></ul>
推理加速型	Pi2	<ul style="list-style-type: none"><li>CentOS 7.5 64bit</li><li>Windows Server 2019 Standard 64bit</li><li>Windows Server 2016 Standard 64bit</li></ul>

类别	实例	支持的镜像
推理加速型	Pi2nl	<ul style="list-style-type: none"><li>CentOS 7.5 64bit</li><li>Ubuntu 16.04 Server 64bit</li><li>Windows Server 2016 Standard 64bit</li></ul>

## 图形加速增强型 G7r

### 概述

G7r型弹性云服务器使用NVIDIA Quadro RTX A6000 GPU显卡，能够支持DirectX、Shader Model、OpenGL、Vulkan接口，整卡提供最大48GiB GDDR6显存，理论单精度性能FP32：38.7TFLOPS，Tensor性能（开启稀疏性）：309.7TFLOPS。更多的Tensor核心带来的强大的性能满足专业级的图形处理需求。

请按需选择您所需要的类型和规格。

### 规格

表 1-34 G7r 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g7r.3xlarge.4	12	48	17/5	200	4	6	1 × NVIDIA RTX A6000-6Q	6	KVM
g7r.4xlarge.4	16	64	20/6	280	8	8	1 × NVIDIA RTX A6000-8Q	8	KVM
g7r.6xlarge.4	24	96	25/9	400	8	8	1 × NVIDIA RTX A6000-12Q	12	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g7r.8xlarge.4	32	128	30/12	550	16	8	1 × NVIDIA RTX A6000-16Q	16	KVM
g7r.12xlarge.4	48	192	35/18	750	16	8	1 × NVIDIA RTX A6000-24Q	24	KVM
g7r.24xlarge.4	96	384	40/36	1100	32	8	1 × NVIDIA RTX A6000	48	KVM

#### G7r型弹性云服务器功能如下：

- 处理器：第三代英特尔® 至强® 可扩展处理器 8378A，主频3.0GHz，睿频3.5GHz。
- 支持图形加速接口：
  - DirectX 12.0、Direct2D、DirectX Video Acceleration (DXVA)
  - Shader Model 5.1
  - OpenGL 4.6
  - Vulkan 1.1
- 支持CUDA、DirectCompute和OpenCL。
- 单卡搭载10752个CUDA核心，84个二代RT核心以及576个三代Tensor核心。
- 支持图形加速应用。
- 支持CPU重载推理应用。
- 提供和弹性云服务器相同的申请流程。
- 自动化的调度G7r型弹性云服务器到装有NVIDIA Quadro RTX A6000 GPU卡的可用区。
- 内置1个NVENC和2个NVDEC编解码引擎（含AV1解码引擎）。

#### 常规支持软件列表

G7r型弹性云服务器主要用于图形加速场景，例如图像渲染、云桌面、3D可视化。应用软件如果依赖GPU的DirectX、OpenGL硬件加速能力可以使用G7r型云服务器。常用的图形处理软件支持列表如下：

- AutoCAD
- 3ds Max
- MAYA
- Agisoft PhotoScan
- ContextCapture

### 使用须知

- G7r型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 说明

G7r型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- 使用公共镜像创建的G7r型弹性云服务器默认已安装特定版本的GRID驱动，但GRID License需自行购买和配置使用，请提前确认云GRID驱动版本是否符合需求，并配置GRID License。  
配置GRID License步骤请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。
- 使用私有镜像创建的G7r型弹性云服务器，请确认在制作私有镜像时安装GRID驱动。如果未安装，请在创建完成后安装GRID驱动，以实现图形加速功能。  
详细安装操作请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 图形加速增强型 G7v

### 概述

G7v型弹性云服务器使用NVIDIA A40 GPU显卡，能够支持DirectX、Shader Model、OpenGL、Vulkan接口，提供单卡48GiB显存，理论浮点性能峰值FP32：37.4TFLOPS，理论Tensor性能峰值TF32：74.8TFLOPS | 149.6TFLOPS（开启稀疏性）。可提供理论2倍于RTX6000显卡的渲染性能和1.4倍于RTX6000的图形处理性能，满足专业级的图形处理需求。

请按需选择您所需要的类型和规格。

### 规格

表 1-35 G7v 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g7v.2xlarge.8	8	64	15/3	100	4	4	1 × NVIDIA A40-8Q	8	KVM
g7v.4xlarge.8	16	128	20/6	150	8	8	1 × NVIDIA A40-16Q	16	KVM
g7v.6xlarge.8	24	192	25/9	200	8	8	1 × NVIDIA A40-24Q	24	KVM

**G7v型弹性云服务器功能如下：**

- 处理器：第三代英特尔® 至强® 可扩展处理器 8378A，主频3.0GHz，睿频3.5GHz。
- 支持图形加速接口：
  - DirectX 12.07、Direct2D、DirectX Video Acceleration (DXVA)
  - Shader Model 5.17
  - OpenGL 4.68
  - Vulkan 1.18
- 支持CUDA、DirectCompute、OpenACC和OpenCL。
- 单卡带有10752个CUDA核心，84个二代RT核心以及336个三代Tensor核心。
- 支持图形加速应用。
- 支持CPU重载推理应用。
- 提供和弹性云服务器相同的申请流程。
- 自动化的调度G7v型弹性云服务器到装有NVIDIA A40 GPU卡的可用区。
- 内置1个NVENC和2个NVDEC(含AV1解码)。

**常规支持软件列表**

G7v型弹性云服务器主要用于图形加速场景，例如图像渲染、云桌面、3D可视化。应用软件如果依赖GPU的DirectX、OpenGL硬件加速能力可以使用G7v型云服务器。常用的图形处理软件支持列表如下：

- AutoCAD
- 3ds Max
- MAYA
- Agisoft PhotoScan
- ContextCapture
- Adobe Premiere Pro
- Solidworks
- Unreal Engine
- Blender
- Vray

### 使用须知

- G7v型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 说明

G7v型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- 使用公共镜像创建的G7v型弹性云服务器默认已安装特定版本的GRID驱动，但GRID License需自行购买和配置使用，请提前确认云GRID驱动版本是否符合需求，并配置GRID License。  
配置GRID License步骤请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。
- 使用私有镜像创建的G7v型弹性云服务器，请确认在制作私有镜像时安装GRID驱动。如果未安装，请在创建完成后安装GRID驱动，以实现图形加速功能。  
详细安装操作请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 图形加速增强型 G7

### 概述

G7型弹性云服务器使用NVIDIA A40 GPU显卡，能够支持DirectX、Shader Model、OpenGL、Vulkan接口，单卡提供48GiB显存，理论浮点性能峰值FP32：37.4TFLOPS，理论Tensor性能峰值TF32：74.8TFLOPS | 149.6TFLOPS（开启稀疏性）。可提供理论2倍于RTX6000显卡的渲染性能和1.4倍于RTX6000的图形处理性能，满足专业级的图形处理需求。

请按需选择您所需要的类型和规格。

### 规格

表 1-36 G7 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g7.1 2xlarge.8	48	384	35/18	750	16	8	1 × NVIDIA A-A40	1 × 48	KVM
g7.2 4xlarge.8	96	768	40/36	850	16	8	2 × NVIDIA A-A40	2 × 48	KVM

#### G7型弹性云服务器功能如下：

- 处理器：第三代英特尔® 至强® 可扩展处理器 8378A，主频3.0GHz，睿频3.5GHz。
- 支持图形加速接口：
  - DirectX 12.07、Direct2D、DirectX Video Acceleration (DXVA)
  - Shader Model 5.17
  - OpenGL 4.68
  - Vulkan 1.18
- 支持CUDA、DirectCompute、OpenACC和OpenCL。
- 单卡带有10752个CUDA核心，84个二代RT核心以及336个三代Tensor核心。
- 支持图形加速应用。
- 支持CPU重载推理应用。
- 提供和弹性云服务器相同的申请流程。
- 自动化的调度G7型弹性云服务器到装有NVIDIA A40 GPU卡的可用区。
- 内置1个NVENC和2个NVDEC(含AV1解码)。

#### 常规支持软件列表

G7型弹性云服务器主要用于图形加速场景，例如图像渲染、云桌面、3D可视化。应用软件如果依赖GPU的DirectX、OpenGL硬件加速能力可以使用G7型云服务器。常用的图形处理软件支持列表如下：

- AutoCAD
- 3ds Max
- MAYA
- Agisoft PhotoScan
- ContextCapture

- Adobe Premiere Pro
- Solidworks
- Unreal Engine
- Blender
- Vray

### 使用须知

- G7型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 说明

G7型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- 使用公共镜像创建的G7型弹性云服务器默认已安装特定版本的GRID驱动，但GRID License需自行购买和配置使用，请提前确认云GRID驱动版本是否符合需求，并配置GRID License。  
配置GRID License步骤请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。
- 使用私有镜像创建的G7型弹性云服务器，请确认在制作私有镜像时安装GRID驱动。如果未安装，请在创建完成后安装GRID驱动，以实现图形加速功能。  
详细安装操作请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 图形加速增强型 G6

### 概述

G6型弹性云服务器使用NVIDIA Tesla T4 GPU显卡，能够支持DirectX、OpenGL、Vulkan接口，提供16GiB显存，理论性能Pixel Rate: 101.8GPixel/s，Texture Rate: 254.4GTexel/s，满足专业级的图形处理需求。

请按需选择您所需要的类型和规格。

### 规格

表 1-37 G6 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g6.xlarge.4	4	16	6/2	200	8	8	1 × T4	16	KVM



规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g6.4xlarge.4	16	64	15/8	200	8	8	1 × T4	16	KVM
g6.6xlarge.4	24	96	25/15	200	8	8	1 × T4	16	KVM
g6.9xlarge.7	36	252	25/15	200	16	8	1 × T4	16	KVM
g6.10xlarge.7	40	280	25/15	200	16	8	1 × T4	16	KVM
g6.18xlarge.7	72	504	30/30	400	32	16	2 × T4	32	KVM
g6.20xlarge.7	80	560	30/30	400	32	16	2 × T4	32	KVM

#### G6型弹性云服务器功能如下：

- 处理器：第二代英特尔® 至强® 可扩展处理器 6266，主频3.0GHz，睿频3.4GHz。
- 支持图形加速接口：
  - DirectX 12、Direct2D、DirectX Video Acceleration (DXVA)
  - OpenGL 4.5
  - Vulkan 1.0
- 支持CUDA和OpenCL。
- 支持NVIDIA T4 GPU卡。
- 支持图形加速应用。
- 支持CPU重载推理应用。
- 自动化的调度G6型弹性云服务器到装有NVIDIA T4 GPU卡的可用区。
- 内置1个NVENC和2个NVDEC。

#### 常规支持软件列表

G6型弹性云服务器主要用于图形加速场景，例如图像渲染、云桌面、3D可视化。应用软件如果依赖GPU的DirectX、OpenGL硬件加速能力可以使用G6型云服务器。常用的图形处理软件支持列表如下：

- AutoCAD
- 3ds Max
- MAYA
- Agisoft PhotoScan
- ContextCapture

### 使用须知

- G6型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 说明

G6型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- 使用公共镜像创建的G6型弹性云服务器默认已安装特定版本的GRID驱动，但GRID License需自行购买和配置使用，请提前确认云GRID驱动版本是否符合需求，并配置GRID License。

配置GRID License步骤请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。

- 使用私有镜像创建的G6型弹性云服务器，请确认在制作私有镜像时安装GRID驱动。如果未安装，请在创建完成后安装GRID驱动，以实现图形加速功能。

详细安装操作请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。

- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 图形加速增强型 G5r

### 概述

G5r型弹性云服务器基于PCI直通技术，独享整张GPU卡，提供专业级的图形加速能力。同时，G5r型弹性云服务器使用NVIDIA Quadro RTX5000 GPU卡，能够支持DirectX、OpenGL接口，可以提供最大显存16GiB，支持渲染，云游戏，图形工作站等场景。

### 规格

表 1-38 G5r 型弹性服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g5r.8xlarge.2	32	64	10/4	100	4	8	1 × RTX 5000	16	KVM

### 说明

NVIDIA Quadro RTX5000显卡采用新一代的图灵架构，且搭载Nvidia新一代RTX平台。

### G5r型弹性云服务器功能如下：

- 处理器：第二代英特尔® 至强® 可扩展处理器 6278，主频2.6GHz，睿频3.5GHz，或英特尔® 至强® 可扩展处理器 6151，主频3.0GHz，睿频3.4GHz。
- 支持光栅化处理加速，含48个RT core。
- 支持NVIDIA RTX5000 GPU卡。
- 支持渲染类图形加速。
- 支持深度学习应用，含3072个CUDA core和384个Tensor core。
- 提供GPU直通的能力。
- 提供最大显存16GiB。

### 使用须知

- G5r型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

### 说明

G5r型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- G5r型弹性云服务器不支持规格变更。
- G5r型弹性云服务器处于公测阶段，请联系客服申请公测。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 图形加速增强型 G5

### 概述

G5型弹性云服务器基于NVIDIA GRID虚拟化GPU技术，能够提供全面的专业级的图形加速能力。G5型弹性云服务器使用NVIDIA Tesla V100 GPU显卡，能够支持DirectX、

OpenGL、Vulkan接口，提供1/2/4/8/16 GiB显存规格，支持最大4096 × 2160分辨率，满足从入门级到专业级的图形处理需求。

请选择您所需要的类型和规格。

## 规格

表 1-39 G5 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	虚拟化类型
g5.xlarge.4	32	128	25/15	200	16	8	1 × V100	16	KVM

### 说明

V100-xQ，表示云服务器使用GRID技术将V100 GPU虚拟为不同规格和型号的vGPU，x对应于vGPU的显存大小，Q表示该类型的vGPU主要适用于工作站和桌面场景。更多关于GRID虚拟GPU的介绍，请参见[NVIDIA官方文档](#)。

### G5型弹性云服务器功能如下：

- 处理器：第二代英特尔® 至强® 可扩展处理器 6278，主频2.6GHz，睿频3.5GHz，或英特尔® 至强® 可扩展处理器 6151，主频3.0GHz，睿频3.4GHz。
- 支持图形加速接口：
  - DirectX 12、Direct2D、DirectX Video Acceleration (DXVA)
  - OpenGL 4.5
  - Vulkan 1.0
- 支持CUDA和OpenCL。
- 支持Quadro vDWS特性，为专业级图形应用提供加速。
- 支持NVIDIA V100 GPU卡。
- 支持图形加速应用。
- 提供GPU硬件虚拟化（vGPU）。
- 自动化的调度G5型弹性云服务器到装有NVIDIA V100 GPU卡的可用区。
- 可以提供最大显存16GiB，分辨率为4096×2160的图形图像处理能力。

### 常规支持软件列表

G5型弹性云服务器主要用于图形加速场景，例如图像渲染、云桌面、3D可视化。应用软件如果依赖GPU的DirectX、OpenGL硬件加速能力可以使用G5型云服务器。常用的图形处理软件支持列表如下：

- AutoCAD
- 3ds Max

- MAYA
- Agisoft PhotoScan
- ContextCapture

### 使用须知

- G5型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 📖 说明

G5型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- G5型Windows操作系统云服务器启动时默认加载了GRID驱动，使用vGPU显卡作为默认视频输出，暂不支持使用管理控制台提供的“远程登录”功能。请使用RDP协议（如Windows远程桌面MSTSC）访问G5型实例，之后安装远程访问工具，如VNC工具等第三方VDI工具。
- G5型弹性云服务器，需在弹性云服务器创建完成后配置GRID License使用。
- 使用公共镜像创建的G5型弹性云服务器默认已安装特定版本的GRID驱动，但GRID License需自行购买和配置使用，请提前确认云GRID驱动版本是否符合需求，并配置GRID License。

配置GRID License步骤请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。

- 使用私有镜像创建的G5型弹性云服务器，请确认在制作私有镜像时安装GRID驱动。如果未安装，请在创建完成后安装GRID驱动，以实现图形加速功能。

详细安装操作请参考《弹性云服务器用户指南》的“GPU加速型实例安装GRID驱动”。

- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 计算加速型 P3v

### 概述

P3v型弹性云服务器采用NVIDIA A800 GPU，在提供云服务器灵活性的同时，提供超高性能计算能力。适用于AI深度学习、科学计算，在深度学习训练、科学计算、计算流体力学、计算金融、地震分析、分子建模、基因组学等领域都能表现出巨大的计算优势。理论单精度浮点性能：FP32：19.5TFLOPS。Tensor核心浮点性能：TF32：156TFLOPS，BFLOAT16：312TFLOPS。

### 规格

表 1-40 P3v 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	GPU 连接技术	显存 (GiB)	虚拟化类型
p3v.3xlarge.8	12	96	17/5	200	4	4	1 × NVIDIA A800 80GB	N/A	80	KVM
p3v.24xlarge.8	96	768	40/36	850	32	8	8 × NVIDIA A800 80GB	NVLink	640	KVM

**P3v型弹性云服务器功能如下：**

- 处理器：第三代英特尔® 至强® 可扩展处理器 6348，主频2.6GHz，睿频3.5GHz。
- 支持NVIDIA A800 GPU卡，每台云服务器支持最大8张A800显卡。
- 支持NVIDIA CUDA并行计算，支持常见的深度学习框架Tensorflow、Caffe、PyTorch、MXNet等。
- 单精度能力19.5 TFLOPS，双精度能力9.7 TFLOPS。
- 支持NVIDIA Tensor Core能力，深度学习混合精度运算能力达到156 TFLOPS。
- 单实例最大网络带宽40Gb/s。
- 单卡 80GB HBM2显存，显存带宽2039Gb/s，支持多卡NVLINK互联技术。
- 完整的基础能力：
  - 网络自定义，自由划分子网、设置网络访问策略。
  - 海量存储，弹性扩容，支持备份与恢复，让数据更加安全。
  - 弹性伸缩，快速增加或减少云服务器数量。
- 灵活选择：  
与普通云服务器一样，P3v型云服务器可以做到分钟级快速发放。
- 优秀的超算生态：  
拥有完善的超算生态环境，用户可以构建灵活弹性、高性能、高性价比的计算平台。大量的HPC应用程序和深度学习框架已经可以运行在P3v实例上。

**常规软件支持列表**

P3v型云服务器主要用于计算加速场景，例如深度学习训练、推理、科学计算、分子建模、地震分析等场景。应用软件如果使用到GPU的CUDA并行计算能力，可以使用P3v型云服务器。常用的软件支持列表如下：

- Tensorflow、Spark、PyTorch、MXNet、Caffe等常用深度学习框架

- RedShift for Autodesk 3ds Max、V-Ray for 3ds Max等支持CUDA的GPU渲染
- Agisoft PhotoScan
- MapD
- Amber、NAMD、VASP等2000余支持GPU加速的应用

### 使用须知

- P3v型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 📖 说明

P3v型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- 使用私有镜像创建的P3v型云服务器，请确认在制作私有镜像时已安装驱动。如果未安装，请在创建完成后安装驱动，以实现计算加速功能。详细操作请参考《弹性云服务器用户指南》“GPU加速型实例安装Tesla驱动及CUDA工具包”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 计算加速型 P2s

### 概述

P2s型弹性云服务器采用NVIDIA Tesla V100 GPU，在提供云服务器灵活性的同时，提供高性能计算能力和优秀的性价比。P2s型弹性云服务器能够提供超高的通用计算能力，适用于AI深度学习、科学计算，在深度学习训练、科学计算、计算流体力学、计算金融、地震分析、分子建模、基因组学等领域都能表现出巨大的计算优势。

### 规格

表 1-41 P2s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	GPU连接技术	显存 (GiB)	虚拟化类型
p2s.2xlarge.8	8	64	10/4	50	4	4	1 × V100	PCIe Gen 3	1 × 32GiB	KVM
p2s.4xlarge.8	16	128	15/8	100	8	8	2 × V100	PCIe Gen 3	2 × 32GiB	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	GPU连接技术	显存 (GiB)	虚拟化类型
p2s.8xlarge.8	32	256	25/15	200	16	8	4 × V100	PCIe Gen 3	4 × 32GiB	KVM
p2s.16xlarge.8	64	512	30/30	400	32	8	8 × V100	PCIe Gen 3	8 × 32GiB	KVM

#### P2s型弹性云服务器功能如下：

- 处理器：第二代英特尔® 至强® 可扩展处理器 6278，主频2.6GHz，睿频3.5GHz，或英特尔® 至强® 可扩展处理器 6151，主频3.0GHz，睿频3.4GHz。
- 支持NVIDIA Tesla V100 GPU卡，每台云服务器支持最大8张Tesla V100显卡。
- 支持NVIDIA CUDA并行计算，支持常见的深度学习框架Tensorflow、Caffe、PyTorch、MXNet等。
- 单精度能力14 TFLOPS，双精度能力7 TFLOPS。
- 支持NVIDIA Tensor Core能力，深度学习混合精度运算能力达到112 TFLOPS。
- 单实例最大网络带宽30Gb/s。
- 使用32GiB HBM2显存，显存带宽900Gb/s。
- 完整的基础能力：
  - 网络自定义，自由划分子网、设置网络访问策略。
  - 海量存储，弹性扩容，支持备份与恢复，让数据更加安全。
  - 弹性伸缩，快速增加或减少云服务器数量。
- 灵活选择：  
与普通云服务器一样，P2s型云服务器可以做到分钟级快速发放。
- 优秀的超算生态：  
拥有完善的超算生态环境，用户可以构建灵活弹性、高性能、高性价比的计算平台。大量的HPC应用程序和深度学习框架已经可以运行在P2s实例上。

#### 常规软件支持列表

P2s型云服务器主要用于计算加速场景，例如深度学习训练、推理、科学计算、分子建模、地震分析等场景。应用软件如果使用到GPU的CUDA并行计算能力，可以使用P2s型云服务器。常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet等常用深度学习框架
- RedShift for Autodesk 3ds Max、V-Ray for 3ds Max等支持CUDA的GPU渲染
- Agisoft PhotoScan
- MapD



## 使用须知

- P2s型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

### 📖 说明

P2s型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- 使用私有镜像创建的P2s型云服务器，请确认在制作私有镜像时已安装Tesla驱动。如果未安装，请在创建完成后安装驱动，以实现计算加速功能。详细操作请参考《弹性云服务器用户指南》“GPU加速型实例安装Tesla驱动及CUDA工具包”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 推理加速型 Pi3

### 概述

Pi3型弹性云服务器采用专为AI推理打造的NVIDIA A30 GPU，能够提供超强的实时推理能力。单卡24GB显存，933GB/s带宽使得Pi3也同时可以支持一定的训练场景。其理论AI训练吞吐量三倍于NVIDIA V100显卡，六倍于前代Pi2型服务器的T4显卡。Pi3型弹性云服务器借助NVIDIA A30的运算器，能够提供最大330TOPS的INT8算力（开启稀疏性）。理论峰值单精度浮点性能：TF32：10.3TFLOPS，Tensor核心性能：TF32：82TFLOPS | 165TFLOPS（开启稀疏性）。

### 规格

表 1-42 Pi3 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发能力 (万PPS)	网卡多队列数	网卡个数/个数上限	GPU	GPU连接技术	显存 (GiB)	虚拟化类型
pi3.6xlarge.4	24	96	25/9	400	8	8	1 × NVIDIA A30	24	-	KVM
pi3.12xlarge.4	48	192	35/18	750	8	8	2 × NVIDIA A30	48	-	KVM

### Pi3型弹性云服务器功能如下：

- 处理器：第三代英特尔® 至强® 可扩展处理器 6348，主频2.6GHz，睿频3.5GHz。

- 支持NVIDIA A30 GPU卡NVLINK互联技术，单实例最大支持2张A30 GPU卡。
- 单GPU单精度计算能力最高10.3 TFLOPS。
- 单GPU INT8计算能力最高330 TOPS。
- 单GPU提供24GiB HBM2显存，带宽933GiB/s。
- 内置1个OFA和1个NVJPEG和4个NVDEC。

### 常规支持软件列表

Pi3实例主要用于GPU推理计算场景，例如图片识别、语音识别、自然语言处理等场景。也可以支持一定的训练场景。

常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet、Spark等深度学习框架。
- AMBER、NAMD、OPENFOAM等2000+支持GPU加速的软件应用。

### 使用须知

- Pi3型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性IP、带宽等，按各自产品的计费方法进行收费。

### 说明

Pi3型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- Pi3型云服务器，所在物理机发生故障时，云服务器支持自动恢复。
- 创建Pi3型云服务器时，请确认在制作私有镜像时已安装Tesla驱动。如果未安装，请在云服务器创建完成后安装驱动，以实现计算加速功能。详细操作请参考《弹性云服务器用户指南》“GPU加速型实例安装Tesla驱动及CUDA工具包”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 推理加速型 Pi2

### 概述

Pi2型弹性云服务器采用专为AI推理打造的NVIDIA Tesla T4 GPU，能够提供超强的实时推理能力。Pi2型弹性云服务器借助T4的INT8运算器，能够提供最大130TOPS的INT8算力。Pi2也可以支持轻量级训练场景。

### 规格

表 1-43 Pi2 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	本地盘	虚拟化类型
pi2.xlarge.4	4	16	8/2	25	2	2	1 × T4	1 × 16	-	KVM
pi2.2xlarge.4	8	32	10/4	50	4	4	1 × T4	1 × 16	-	KVM
pi2.3xlarge.4	12	48	12/6	80	6	6	1 × T4	1 × 16	-	KVM
pi2.4xlarge.4	16	64	15/8	100	8	8	2 × T4	2 × 16	-	KVM
pi2.8xlarge.4	32	128	25/15	200	16	8	4 × T4	4 × 16	-	KVM

#### Pi2型弹性云服务器功能如下：

- 处理器：第二代英特尔® 至强® 可扩展处理器 6278，主频2.6GHz，睿频3.5GHz，或英特尔® 至强® 可扩展处理器 6151，主频3.0GHz，睿频3.4GHz。
- 支持NVIDIA Tesla T4 GPU卡，单实例最大支持4张T4 GPU卡。
- 提供GPU硬件直通能力。
- 单GPU单精度计算能力最高8.1 TFLOPS。
- 单GPU INT8计算能力最高130 TOPS。
- 单GPU提供16GiB GDDR6显存，带宽320GiB/s。
- 内置1个NVENC和2个NVDEC。

#### 常规支持软件列表

Pi2实例主要用于GPU推理计算场景，例如图片识别、语音识别、自然语言处理等场景。也可以支持轻量级训练场景。

常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet等深度学习框架。

#### 使用须知

- Pi2型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 📖 说明

Pi2型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- Pi2型云服务器，所在物理机发生故障时，云服务器支持自动恢复。
- 使用公共镜像创建的Pi2型云服务器，默认已安装Tesla驱动。
- 使用私有镜像创建的Pi2型云服务器，请确认在制作私有镜像时已安装Tesla驱动。如果未安装，请在云服务器创建完成后安装驱动，以实现计算加速功能。详细操作请参考《弹性云服务器用户指南》“GPU加速型实例安装Tesla驱动及CUDA工具包”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

## 推理加速型 Pi2nl

### 概述

Pi2nl型弹性云服务器采用专为AI推理打造的NVIDIA Tesla T4 GPU，能够提供超强的实时推理能力。Pi2nl型弹性云服务器借助T4的INT8运算器，能够提供最大130TOPS的INT8算力。Pi2nl也可以支持轻量级训练场景。

### 规格

表 1-44 Pi2nl 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	GPU	显存 (GiB)	本地盘	虚拟化类型
pi2nl.2xlarge.4	8	32	10/4	50	4	4	1 × T4	1 × 16	-	KVM
pi2nl.4xlarge.4	16	64	15/8	100	8	8	2 × T4	2 × 16	-	KVM
pi2nl.8xlarge.4	32	128	25/15	200	16	8	4 × T4	4 × 16	-	KVM

Pi2nl型弹性云服务器功能如下：

- 处理器：第二代英特尔® 至强® 可扩展处理器 6278，主频2.6GHz，睿频3.5GHz，或英特尔® 至强® 可扩展处理器 6151，主频3.0GHz，睿频3.4GHz。
- 支持NVIDIA Tesla T4 GPU卡，单实例最大支持4张T4 GPU卡。
- 提供GPU硬件直通能力。
- 单GPU单精度计算能力最高8.1 TFLOPS。
- 单GPU INT8计算能力最高130 TOPS。
- 单GPU提供16GiB GDDR6显存，带宽320GiB/s。
- 内置1个NVENC和2个NVDEC。

#### 常规支持软件列表

Pi2nl实例主要用于GPU推理计算场景，例如图片识别、语音识别、自然语言处理等场景。也可以支持轻量级训练场景。

常用的软件支持列表如下：

- Tensorflow、Caffe、PyTorch、MXNet等深度学习框架。

#### 使用须知

- Pi2nl型云服务器，关机后基础资源（包括vCPU、内存、镜像、GPU）不计费，但系统盘仍会收取容量对应的费用。如有其他绑定的产品，如云硬盘、弹性公网IP、带宽等，按各自产品的计费方法进行收费。

#### 说明

Pi2nl型云服务器，关机后资源会被释放，下次开机时如果后台资源不足，可能会导致云服务器开机失败。如果您需要长期使用该云服务器，建议保持开机状态。

- Pi2nl型云服务器，所在物理机发生故障时，云服务器支持自动恢复。
- 使用公共镜像创建的Pi2nl型云服务器，默认已安装Tesla驱动。
- 使用私有镜像创建的Pi2nl型云服务器，请确认在制作私有镜像时已安装Tesla驱动。如果未安装，请在云服务器创建完成后安装驱动，以实现计算加速功能。详细操作请参考《弹性云服务器用户指南》“GPU加速型实例安装Tesla驱动及CUDA工具包”。
- GPU型弹性云服务器因通用算力和异构算力差异大，仅支持变更规格至同类型规格内的细分规格。
- GPU型弹性云服务器不支持热迁移。

### 1.4.4.8 AI 加速型

AI加速型云服务器（AI Accelerated Cloud Server, AIACS）是专门为AI业务提供加速服务的云服务器。搭载昇腾系列芯片及软件栈。

AI推理加速型系列：搭载昇腾310芯片，为AI推理业务加速。

#### AI 加速型实例总览

AI推理加速型：[AI推理加速增强I型Ai1s](#)

## 适用于 AI 加速型实例的公共镜像列表

表 1-45 适用于 AI 加速型实例的公共镜像列表

类别	实例	支持的公共镜像
AI推理加速增强I型	Ai1s	Ubuntu Server 18.04 64bit CentOS 7.6 64bit

## AI 推理加速增强 I 型 Ai1s

### 概述

AI推理加速增强I型实例Ai1s是以昇腾310（Ascend 310）芯片为加速核心的AI加速型弹性云服务器。基于Ascend 310芯片低功耗、高算力特性，实现了能效比的大幅提升，助力AI推理业务的快速普及。通过AI推理加速型实例Ai1s将Ascend 310芯片的计算加速能力在云平台上开放出来，方便用户快速简捷地使用Ascend 310芯片强大的处理能力。

AI推理加速型实例Ai1s基于Atlas 300I加速卡设计，更多详细信息请参考[昇腾社区](#)。

AI加速型云服务器可用于机器视觉，安防监控，智慧园区，智慧城市，智慧交通，智慧零售，互联网音视频，视频编解码等业务场景。

### 规格

表 1-46 Ai1s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽	最大收发包能力 (万/PPS)	Ascend 310	Ascend RAM (GiB)	网卡多队列数	网卡个数上限	虚拟化类型
ai1s.large.4	2	8	4/1.3	20	1	8	2	2	KVM
ai1s.xlarge.4	4	16	6/2	35	2	16	2	3	KVM
ai1s.2xlarge.4	8	32	10/4	50	4	32	4	4	KVM
ai1s.4xlarge.4	16	64	15/8	100	8	64	8	8	KVM
ai1s.8xlarge.4	32	128	25/15	200	16	128	8	8	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽	最大收发包能力 (万/PPS)	Ascend 310	Ascend RAM (GiB)	网卡多队列数	网卡个数上限	虚拟化类型
ai1s.6xlarge.4	24	96	25/14	400	4	96	8	8	KVM
ai1s.9xlarge.4	36	144	30/18	550	4	144	16	8	KVM
ai1s.18xlarge.4	72	288	40/36	1000	8	288	32	8	KVM

## 功能

Ai1s加速型弹性云服务器功能如下：

- 处理器与内存配比为1:4。
- 处理器：第二代英特尔® 至强® 可扩展处理器 6278，主频2.6GHz，睿频3.5GHz，或英特尔® 至强® 可扩展处理器 6151，主频3.0GHz，睿频3.4GHz。
- 支持Ascend 310芯片，每张Atlas 300I加速卡包含4个Ascend 310芯片。
- 单芯片整数精度（INT8）16TeraOPS。
- 单芯片提供8GiB显存，内存带宽50GiB/s。
- 内置硬件视频编解码引擎，支持5路全高清视频解码器（H.264/265）。

## 使用须知

1. Ai1s实例支持如下版本的公共镜像：
  - Ubuntu Server 18.04 64bit
  - CentOS 7.6 64bit
2. Ai1s型弹性云服务器不支持规格变更。
3. Ai1s型云服务器所在物理机发生故障时，云服务器支持自动恢复。

## 如何使用 AI 加速型云服务器

使用AI加速型云服务器的流程如下：

1. 创建云服务器，详细步骤，请参考详细操作请参考《弹性云服务器用户指南》“步骤一：基础配置”。
  - 在“规格”设置时，选择AI加速型相关规格。
  - 在“镜像”设置时，可以选择“公共镜像”和“私有镜像”。
    - 公共镜像：已经默认安装了CANN 3.1.0的开发套件包，并配置了环境变量，请验证环境的可用性。

- 私有镜像：需要安装驱动、固件和开发套件，并配置环境变量。详细步骤，请参考[昇腾文档中心](#)对应版本的《CANN软件安装指南》。
2. 远程登录云服务器。  
对于Linux操作系统的云服务器，登录时请以SSH密码方式登录。
  3. 验证环境可用性。  
请参考[昇腾文档中心](#)中对应CANN版本《应用开发指南》中的“样例使用指导”，选择一个样例进行编译运行，验证环境的正确性。  
例如，选择“基于Caffe ResNet-50网络实现图片分类（图片解码+缩放+同步推理）”样例进行验证。

## 1.4.5 实例规格（鲲鹏）

### 1.4.5.1 鲲鹏通用计算型

#### 概述

鲲鹏通用计算型云服务器搭载鲲鹏920处理器及25GE智能高速网卡，以较低的成本提供了基础的计算能力，可以满足基础业务上云的诉求。同时也提供了突发能力，可以根据工作负载的需要进行突发，满足部分业务高峰，是一款性价比极高的产品。

#### 规格

表 1-47 ks1 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队列数	网卡 个数 上限	虚拟 化类 型
ks1.large.1	2	2	1.5/0.3	15	1	1	KVM
ks1.xlarge.1	4	4	2/0.5	20	1	2	KVM
ks1.2xlarge.1	8	8	3.5/1	35	2	2	KVM
ks1.3xlarge.1	12	12	5/1.5	45	2	2	KVM
ks1.4xlarge.1	16	16	7/2	60	2	2	KVM
ks1.large.2	2	4	1.5/0.3	15	1	1	KVM
ks1.xlarge.2	4	8	2/0.5	20	1	2	KVM
ks1.2xlarge.2	8	16	3.5/1	35	2	2	KVM



规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
ks1.3xlarge.2	12	24	5/1.5	45	2	2	KVM
ks1.4xlarge.2	16	32	7/2	60	2	2	KVM
ks1.large.4	2	8	1.5/0.3	15	1	1	KVM
ks1.xlarge.4	4	16	2/0.5	20	1	2	KVM
ks1.2xlarge.4	8	32	3.5/1	35	2	2	KVM
ks1.3xlarge.4	12	48	5/1.5	45	2	2	KVM
ks1.4xlarge.4	16	64	7/2	60	2	2	KVM

表 1-48 ks1s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
ks1s.large.1	2	2	1.5/0.3	15	1	1	KVM
ks1s.xlarge.1	4	4	2/0.5	20	1	2	KVM
ks1s.2xlarge.1	8	8	3.5/1	35	2	2	KVM
ks1s.3xlarge.1	12	12	5/1.5	45	2	2	KVM
ks1s.4xlarge.1	16	16	7/2	60	2	2	KVM
ks1s.large.2	2	4	1.5/0.3	15	1	1	KVM
ks1s.xlarge.2	4	8	2/0.5	20	1	2	KVM
ks1s.2xlarge.2	8	16	3.5/1	35	2	2	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
ks1s.3xlarge.2	12	24	5/1.5	45	2	2	KVM
ks1s.4xlarge.2	16	32	7/2	60	2	2	KVM
ks1s.large.4	2	8	1.5/0.3	15	1	1	KVM
ks1s.xlarge.4	4	16	2/0.5	20	1	2	KVM
ks1s.2xlarge.4	8	32	3.5/1	35	2	2	KVM
ks1s.3xlarge.4	12	48	5/1.5	45	2	2	KVM
ks1s.4xlarge.4	16	64	7/2	60	2	2	KVM

## 使用场景

鲲鹏通用计算型云服务器适合平时不会持续高压使用CPU，但偶尔需要提高计算性能完成工作负载的场景，包括但不限于轻量级的Web服务器、开发、测试环境以及中低性能数据库等。

### 1.4.5.2 鲲鹏通用计算增强型

#### 鲲鹏通用计算增强型实例类型总览

鲲鹏通用计算增强型云服务器搭载鲲鹏处理器，提供强劲的鲲鹏算力和高性能网络，更好地满足企业对云上业务性价比高、安全可靠等诉求。

在售：kC1、kC1s、kC1m

表 1-49 鲲鹏通用计算增强型实例特点

规格名称	计算	磁盘类型	网络
鲲鹏通用计算增强型 kC1	<ul style="list-style-type: none"><li>• CPU/内存配比：1:1/1:2/1:4</li><li>• vCPU数量范围：1-60</li><li>• 处理器：鲲鹏920处理器</li><li>• 基频：2.6GHz</li></ul>	<ul style="list-style-type: none"><li>• 高IO</li><li>• 超高IO</li></ul>	<ul style="list-style-type: none"><li>• 超高网络收发包能力</li><li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li><li>• 最大网络收发包：400万PPS</li><li>• 最大内网带宽：30Gbps</li></ul>

规格名称	计算	磁盘类型	网络
鲲鹏通用计算增强型 kC1s	<ul style="list-style-type: none"><li>• CPU/内存配比：1:2/1:4</li><li>• vCPU数量范围：2-60</li><li>• 处理器：鲲鹏920处理器</li><li>• 基频：2.6GHz</li></ul>	<ul style="list-style-type: none"><li>• 超高IO</li><li>• 高IO</li></ul>	<ul style="list-style-type: none"><li>• 超高网络收发包能力</li><li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li><li>• 最大网络收发包：400万PPS</li><li>• 最大内网带宽：18Gbps</li></ul>
鲲鹏通用计算增强型 kC1m	<ul style="list-style-type: none"><li>• CPU/内存配比：1:2/1:4</li><li>• vCPU数量范围：2-48</li><li>• 处理器：鲲鹏920处理器</li><li>• 基频：2.6GHz</li></ul>	<ul style="list-style-type: none"><li>• 超高IO</li><li>• 高IO</li></ul>	<ul style="list-style-type: none"><li>• 超高网络收发包能力</li><li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li><li>• 最大网络收发包：350万PPS</li><li>• 最大内网带宽：25Gbps</li></ul>

## 规格

表 1-50 kC1 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
kc1.small .1	1	1	2/0.5	20	1	2	KVM
kc1.large .2	2	4	3/0.8	30	2	2	KVM
kc1.xlarge.2	4	8	5/1.5	50	2	3	KVM
kc1.2xlarge.2	8	16	7/3	80	4	4	KVM
kc1.3xlarge.2	12	24	9/4.5	110	4	5	KVM
kc1.4xlarge.2	16	32	12/6	140	4	6	KVM
kc1.6xlarge.2	24	48	15/8.5	200	8	6	KVM
kc1.8xlarge.2	32	64	18/10	260	8	6	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
kc1.12xlarge.2	48	96	25/16	350	16	6	KVM
kc1.15xlarge.2	60	120	30/20	400	16	6	KVM
kc1.large.4	2	8	3/0.8	30	2	2	KVM
kc1.xlarge.4	4	16	5/1.5	50	2	3	KVM
kc1.2xlarge.4	8	32	7/3	80	4	4	KVM
kc1.3xlarge.4	12	48	9/4.5	110	4	5	KVM
kc1.4xlarge.4	16	64	12/6	140	4	6	KVM
kc1.6xlarge.4	24	96	15/8.5	200	8	6	KVM
kc1.8xlarge.4	32	128	18/10	260	8	6	KVM
kc1.12xlarge.4	48	192	25/16	350	16	6	KVM

表 1-51 kc1s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
kc1s.large.2	2	4	2/0.4	25	2	2	KVM
kc1s.xlarge.2	4	8	3/0.8	40	2	3	KVM
kc1s.2xlarge.2	8	16	4/1.5	65	4	4	KVM
kc1s.3xlarge.2	12	24	5/2	90	4	5	KVM
kc1s.4xlarge.2	16	32	7/3	120	4	6	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
kc1s.6xlarge.2	24	48	8/4	160	8	6	KVM
kc1s.8xlarge.2	32	64	10/5	200	8	6	KVM
kc1s.12xlarge.2	48	96	15/8	280	16	6	KVM
kc1s.15xlarge.2	60	120	18/10	320	16	6	KVM
kc1s.large.4	2	8	2/0.4	25	2	2	KVM
kc1s.xlarge.4	4	16	3/0.8	40	2	3	KVM
kc1s.2xlarge.4	8	32	4/1.5	65	4	4	KVM
kc1s.3xlarge.4	12	48	5/2	90	4	5	KVM
kc1s.4xlarge.4	16	64	7/3	120	4	6	KVM
kc1s.6xlarge.4	24	96	8/4	160	8	6	KVM
kc1s.8xlarge.4	32	128	10/5	200	8	6	KVM
kc1s.12xlarge.4	48	192	15/8	280	16	6	KVM

表 1-52 kc1m 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
kc1m.large.2	2	4	3/0.8	30	2	2	KVM
kc1m.xlarge.2	4	8	5/1.5	50	2	3	KVM
kc1m.2xlarge.2	8	16	7/3	80	4	4	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
kc1m.3xlarge.2	12	24	9/4.5	110	4	5	KVM
kc1m.4xlarge.2	16	32	12/6	140	4	6	KVM
kc1m.6xlarge.2	24	48	15/8.5	200	8	6	KVM
kc1m.8xlarge.2	32	64	18/10	260	8	6	KVM
kc1m.12xlarge.2	48	96	25/16	350	16	6	KVM
kc1m.large.4	2	8	3/0.8	30	2	2	KVM
kc1m.xlarge.4	4	16	5/1.5	50	2	3	KVM
kc1m.2xlarge.4	8	32	7/3	80	4	4	KVM
kc1m.3xlarge.4	12	48	9/4.5	110	4	5	KVM
kc1m.4xlarge.4	16	64	12/6	140	4	6	KVM
kc1m.6xlarge.4	24	96	15/8.5	200	8	6	KVM
kc1m.8xlarge.4	32	128	18/10	260	8	6	KVM
kc1m.12xlarge.4	48	192	25/16	350	16	6	KVM

## 使用场景

鲲鹏通用计算增强型弹性云服务器适用于对自主研发、安全隐私要求较高的政企金融场景，对网络性能要求较高的互联网场景，对核数要求较多的大数据、HPC场景，对成本比较敏感的建站、电商等场景等。

### 1.4.5.3 鲲鹏内存优化型

#### 鲲鹏内存优化型实例类型总览

鲲鹏内存优化型弹性云服务器搭载鲲鹏920处理器及25GE智能高速网卡，提供最大480GiB基于DDR4的内存实例和高性能网络，擅长处理大型内存数据集和高网络场景。

在售：kM1、kM1s、kM1m

表 1-53 鲲鹏内存优化型实例特点

规格名称	计算	磁盘类型	网络
鲲鹏内存优化型kM1	<ul style="list-style-type: none"> <li>• CPU/内存配比：1:8</li> <li>• vCPU数量范围：2-60</li> <li>• 处理器：鲲鹏920处理器</li> <li>• 基频：2.6GHz</li> </ul>	<ul style="list-style-type: none"> <li>• 高IO</li> <li>• 通用型SSD</li> <li>• 超高IO</li> </ul>	<ul style="list-style-type: none"> <li>• 超高网络收发包能力</li> <li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li> <li>• 最大网络收发包：400万PPS</li> <li>• 最大内网带宽：30Gbps</li> </ul>
鲲鹏内存优化型kM1s	<ul style="list-style-type: none"> <li>• CPU/内存配比：1:8</li> <li>• vCPU数量范围：2-60</li> <li>• 处理器：鲲鹏920处理器</li> <li>• 基频：2.6GHz</li> </ul>	<ul style="list-style-type: none"> <li>• 超高IO</li> <li>• 高IO</li> </ul>	<ul style="list-style-type: none"> <li>• 超高网络收发包能力</li> <li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li> <li>• 最大网络收发包：400万PPS</li> <li>• 最大内网带宽：18Gbps</li> </ul>
鲲鹏内存优化型kM1m	<ul style="list-style-type: none"> <li>• CPU/内存配比：1:8</li> <li>• vCPU数量范围：2-48</li> <li>• 处理器：鲲鹏920处理器</li> <li>• 基频：2.6GHz</li> </ul>	<ul style="list-style-type: none"> <li>• 超高IO</li> <li>• 高IO</li> </ul>	<ul style="list-style-type: none"> <li>• 超高网络收发包能力</li> <li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li> <li>• 最大网络收发包：350万PPS</li> <li>• 最大内网带宽：25Gbps</li> </ul>

## 规格

表 1-54 kM1 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡多队列数	网卡个数上限	虚拟化类型
km1.large.8	2	16	3/0.8	30	2	2	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
km1.xlarge.8	4	32	5/1.5	50	2	3	KVM
km1.2xlarge.8	8	64	7/3	80	4	4	KVM
km1.3xlarge.8	12	96	9/4.5	110	4	5	KVM
km1.4xlarge.8	16	128	12/6	140	4	6	KVM
km1.6xlarge.8	24	192	15/8	200	8	6	KVM
km1.8xlarge.8	32	256	18/10	260	8	6	KVM
km1.12xlarge.8	48	384	25/16	350	16	8	KVM
km1.15xlarge.8	60	480	30/20	400	16	8	KVM

表 1-55 km1s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
km1s.large.8	2	16	2/0.4	30	2	2	KVM
km1s.xlarge.8	4	32	3/0.8	50	2	3	KVM
km1s.2xlarge.8	8	64	4/1.5	80	4	4	KVM
km1s.3xlarge.8	12	96	5/2	110	4	5	KVM
km1s.4xlarge.8	16	128	7/3	140	4	6	KVM
km1s.6xlarge.8	24	192	8/4	200	8	6	KVM
km1s.8xlarge.8	32	256	10/5	260	8	6	KVM



规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
km1s.12x large.8	48	384	15/8	350	16	8	KVM
km1s.15x large.8	60	480	18/10	400	16	8	KVM

表 1-56 km1m 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/ 基准带宽 (Gbps)	最大收发包 能力 (万PPS)	网卡 多队 列数	网卡 个数 上限	虚拟 化类 型
km1m.la rge.8	2	16	3/0.8	30	2	2	KVM
km1m.xl arge.8	4	32	5/1.5	50	2	3	KVM
km1m.2x large.8	8	64	7/3	80	4	4	KVM
km1m.3x large.8	12	96	9/4.5	110	4	5	KVM
km1m.4x large.8	16	128	12/6	140	4	6	KVM
km1m.6x large.8	24	192	15/8.5	200	8	6	KVM
km1m.8x large.8	32	256	18/10	260	8	6	KVM
km1m.12 xlarge.8	48	384	25/16	350	16	8	KVM

## 使用场景

大数据分析，如广告精准营销、电商、车联网等大数据分析场景。内存存储系统，如 Memcache 等。

### 1.4.5.4 鲲鹏超高 I/O 型

#### 鲲鹏超高 I/O 型实例类型总览

鲲鹏超高I/O型弹性云服务器搭载鲲鹏920处理器及25GE智能高速网卡，提供最大480GiB基于DDR4的内存实例和高性能网络，擅长处理大型内存数据集和高网络场景。

在售：ki1s、ki1

表 1-57 鲲鹏超高 I/O 型实例特点

规格名称	计算	磁盘类型	网络
鲲鹏超高 I/O型ki1s	<ul style="list-style-type: none"><li>• CPU/内存配比：1:4</li><li>• vCPU数量范围：8-64</li><li>• 处理器：鲲鹏920处理器</li><li>• 基频：2.6GHz</li></ul>	<ul style="list-style-type: none"><li>• 超高IO</li><li>• 高IO</li></ul>	<ul style="list-style-type: none"><li>• 超高网络收发包能力</li><li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li><li>• 最大网络收发包：320万PPS</li><li>• 最大内网带宽：18Gbps</li></ul>
鲲鹏超高 I/O型ki1	<ul style="list-style-type: none"><li>• CPU/内存配比：1:4</li><li>• vCPU数量范围：8-64</li><li>• 处理器：鲲鹏920处理器</li><li>• 基频：2.6GHz</li></ul>	<ul style="list-style-type: none"><li>• 高IO</li><li>• 超高IO</li></ul>	<ul style="list-style-type: none"><li>• 超高网络收发包能力</li><li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li><li>• 最大网络收发包：400万PPS</li><li>• 最大内网带宽：30Gbps</li></ul>

## 规格

表 1-58 ki1s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡个数上限	网卡多队列数	本地盘	虚拟化类型
ki1s.2xlarge.4	8	32	4/1.5	65	4	4	1 × 3200GiB	KVM
ki1s.4xlarge.4	16	64	7/3	120	6	4	2 × 3200GiB	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡个数上限	网卡多队列数	本地盘	虚拟化类型
ki1s.6xlarge.4	24	96	8/4	160	6	8	3 × 3200GiB	KVM
ki1s.8xlarge.4	32	128	10/5	200	6	8	4 × 3200GiB	KVM
ki1s.12xlarge.4	48	192	15/8	280	6	16	6 × 3200GiB	KVM
ki1s.16xlarge.4	64	228	18/10	320	6	16	8 × 3200GiB	KVM

表 1-59 ki1 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡个数上限	网卡多队列数	本地盘	虚拟化类型
ki1.2xlarge.4	8	32	7/3	80	4	4	1 × 3200GiB	KVM
ki1.4xlarge.4	16	64	12/6	140	6	4	2 × 3200GiB	KVM
ki1.6xlarge.4	24	96	15/8.5	200	6	8	3 × 3200GiB	KVM
ki1.8xlarge.4	32	128	18/10	260	6	8	4 × 3200GiB	KVM
ki1.12xlarge.4	48	192	25/16	350	6	16	6 × 3200GiB	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽 (Gbps)	最大收发包能力 (万PPS)	网卡个数上限	网卡多队列数	本地盘	虚拟化类型
ki1.16xlarge.4	64	228	30/20	400	6	16	8 × 3200GiB	KVM

## 功能

ki1s型弹性服务器的IOPS(Input/Output Operations Per Second)性能和单盘指标如表1-60和表1-61所示。

表 1-60 ki1s 型弹性云服务器 IOPS 性能

规格名称	4KB随机读最大IOPS
ki1s.2xlarge.4	750000
ki1s.4xlarge.4	1500000
ki1s.6xlarge.4	2250000
ki1s.8xlarge.4	3000000
ki1s.12xlarge.4	4500000
ki1s.16xlarge.4	6000000

表 1-61 ki1s 型弹性云服务器 NVMe 单盘指标

指标	性能
磁盘容量	3.2T
读IOPS (4KB随机读)	750000
写IOPS (4KB随机写)	200000
读吞吐量	2.9 GiB/s
写吞吐量	1.9 GiB/s
访问时延	微秒级

ki1型弹性服务器的IOPS(Input/Output Operations Per Second)性能和单盘指标如表1-62和表1-63所示。

表 1-62 ki1 型弹性云服务器 IOPS 性能

规格名称	4KB随机读最大IOPS
ki1.2xlarge.4	750000
ki1.4xlarge.4	1500000
ki1.6xlarge.4	2250000
ki1.8xlarge.4	3000000
ki1.12xlarge.4	4500000
ki1.16xlarge.4	6000000

表 1-63 ki1 型弹性云服务器 NVMe 单盘指标

指标	性能
磁盘容量	3.2T
读IOPS（4KB随机读）	750000
写IOPS（4KB随机写）	200000
读吞吐量	2.9 GiB/s
写吞吐量	1.9 GiB/s
访问时延	微秒级

## 使用须知

- 鲲鹏超高I/O型弹性云服务器不支持规格变更。
- 鲲鹏超高I/O型弹性云服务器不支持本地盘的快照和备份。
- 可使用本地盘和云硬盘两类磁盘存储数据，通过挂载云硬盘，可以提供更大的存储空间。关于本地盘和云硬盘的使用，有如下约束与限制：
  - 系统盘只能部署在云硬盘上，不可以部署在本地盘上。
  - 数据盘可以部署在云硬盘和本地盘上。
  - 最多可以挂载60块盘（包括VBD盘+SCSI盘+本地盘）。其中，SCSI盘最多只能挂载30块，VBD盘最多只能挂载22块（含系统盘）。
  - 建议您在应用中使用wwn号进行本地盘的相关操作，不要直接使用盘符，因为Linux操作系统会有低概率盘符漂移的可能。以挂载本地盘为例：  
假设本地盘的wwn号为wwn-0x50014ee2b14249f6，则执行的命令为：  
**mount /dev/disk/by-id/wwn-0x50014ee2b14249f6**

## 📖 说明

如何查看本地盘wwn号？

1. 登录弹性云服务器操作系统。
2. 执行以下命令，查看wwn号。

**ll /dev/disk/by-id**

- 鲲鹏超高I/O型弹性云服务器的本地磁盘数据有丢失的风险（比如宿主机宕机或本地磁盘损坏时），如果您的应用不能做到数据可靠性的架构，强烈建议您使用云盘搭建您的弹性云服务器。
- 删除鲲鹏超高I/O型弹性云服务器后，本地NVMe SSD盘中的数据会被自动清除，请提前做好数据备份。删除本地盘数据的时间较长，因此，资源释放的时间较之常规云服务器略长。
- 由于本地盘数据的可靠性取决于物理服务器和硬盘的可靠性，存在单点故障风险，建议您在应用层做好数据冗余，以保证数据的可用性，需要长期保存的业务数据建议使用云硬盘存储。
- 鲲鹏超高I/O型弹性云服务器的本地盘设备名为/dev/nvme0n1、/dev/nvme0n2等。
- 对于鲲鹏超高I/O型弹性云服务器，关机后其基础资源（包括vCPU、内存、镜像）会继续收费。如需停止计费，需删除弹性云服务器。

## 使用场景

鲲鹏超高I/O型弹性云服务器适用于高性能关系型数据库，NoSQL数据库(Cassandra、MongoDB等)以及ElasticSearch搜索等场景。

### 1.4.5.5 鲲鹏 AI 推理加速型

鲲鹏AI推理加速型云服务器是专门为AI业务提供加速服务的云服务器。搭载昇腾系列芯片及软件栈。

鲲鹏AI推理加速型系列：搭载昇腾310芯片，为AI推理业务加速。

表 1-64 鲲鹏 AI 推理加速型实例特点

规格名称	计算	磁盘类型	网络
鲲鹏AI推理加速增强I型 kAi1s	<ul style="list-style-type: none"><li>• CPU/内存配比：1:1/1:2</li><li>• vCPU数量范围：4-48</li><li>• 处理器：鲲鹏920处理器</li><li>• 基频：2.6GHz</li></ul>	<ul style="list-style-type: none"><li>• 高IO</li><li>• 通用型SSD</li><li>• 超高IO</li><li>• 极速型SSD</li><li>• 通用型SSD V2</li></ul>	<ul style="list-style-type: none"><li>• 超高网络收发包能力</li><li>• 实例网络性能与计算规格对应，规格越高网络性能越强</li><li>• 最大网络收发包：200万PPS</li><li>• 最大内网带宽：12Gbps</li></ul>

## 📖 说明

kAi1s使用的驱动和CANN仅支持21.0.2（3.0.1）版本，不支持驱动和CANN升级。

## 鲲鹏 AI 推理加速增强 I 型 kAi1s

### 概述

鲲鹏AI推理加速型实例kAi1s是以昇腾310（Ascend 310）芯片为加速核心的AI加速型弹性云服务器。基于Ascend 310芯片低功耗、高算力特性，实现了能效比的大幅提升，助力AI推理业务的快速普及。通过AI推理加速型实例kAi1s将Ascend 310芯片的计算加速能力在云平台上开放出来，方便用户快速简捷地使用Ascend 310芯片强大的处理能力。

鲲鹏AI推理加速型实例kAi1s基于Atlas 300I加速卡设计，更多详细信息请参考[昇腾社区](#)。

鲲鹏AI推理加速型云服务器可用于机器视觉、语音识别、自然语言处理通用技术，支撑智能零售、智能园区、机器人云大脑、平安城市等场景。

### 规格

表 1-65 kAi1s 型弹性云服务器的规格

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽	最大收发包能力 (万/PPS)	网卡多队列数	网卡个数上限	Ascend 310 个数	虚拟化类型
kai1s.xlarge.1	4	4	3/0.8	20	2	2	1	KVM
kai1s.2xlarge.1	8	8	4/1.5	40	2	3	2	KVM
kai1s.4xlarge.1	16	16	6/3	80	4	4	4	KVM
kai1s.3xlarge.2	12	24	8/4	100	4	4	4	KVM
kai1s.4xlarge.2	16	32	10/6	140	4	5	6	KVM
kai1s.6xlarge.2	24	48	12/8	200	8	6	8	KVM
kai1s.9xlarge.2	36	72	12/8	200	8	6	12	KVM

规格名称	vCPU	内存 (GiB)	最大带宽/基准带宽	最大收发包能力 (万/PPS)	网卡多队列数	网卡个数上限	Ascend 310 个数	虚拟化类型
kai1s.12xlarge.2	48	96	12/8	200	16	6	12	KVM

## 功能

kAi1s加速型弹性云服务器功能如下：

- 处理器与内存配比为1:1/1:2。
- 处理器：Kunpeng 920 2.6GHz。
- 支持Ascend310芯片，每张Atlas 300I加速卡包含4个Ascend310芯片。
- 单芯片半精度（FP16）8TeraFLOPS。
- 单芯片整数精度（INT8）16TeraOPS。
- 单芯片提供8GiB显存，内存带宽50GiB/s。
- 内置硬件视频编解码引擎，支持H.264/265。

## 使用须知

- kAi1s实例支持如下版本的操作系统：
  - Ubuntu Server 18.04 64bit
  - CentOS 7.6 64bit
- kAi1s型弹性云服务器不支持规格变更。
- kAi1s型云服务器所在物理机发生故障时，云服务器支持自动恢复。

## 如何使用 kAi 系列加速型云服务器

购买与使用kAi系列加速型云服务器的流程如下：

1. 创建云服务器，详细步骤，请参考详细操作请参考《弹性云服务器用户指南》“步骤一：基础配置”。
  - 在“规格”设置时，选择kAi系列加速型相关规格。
  - 在“镜像”设置时，可以选择“公共镜像”和“私有镜像”。
    - 公共镜像：已经默认安装了CANN 3.1.0的开发套件包，并配置了环境变量，请验证环境的可用性。
    - 私有镜像：需要安装驱动、固件和开发套件，并配置环境变量。详细步骤，请参考[昇腾文档中心](#)对应版本的《CANN软件安装指南》。
2. 远程登录云服务器。  
对于Linux操作系统的云服务器，登录时请以SSH密码方式登录。
3. 验证环境可用性。



请参考[昇腾文档中心](#)中对应CANN版本《应用开发指南》中的“样例使用指导”，选择一个样例进行编译运行，验证环境的正确性。

例如，选择“基于Caffe ResNet-50网络实现图片分类（图片解码+缩放+同步推理）”样例进行验证。

## 1.5 镜像

### 1.5.1 镜像类型

#### 什么是镜像

镜像是一个包含了软件及必要配置的弹性云服务器模板，至少包含操作系统，还可以包含应用软件（例如，数据库软件）和私有软件。通过镜像，您可以创建弹性云服务器。

镜像分为公共镜像和私有镜像，公共镜像为系统默认提供的镜像，私有镜像为用户自己创建的镜像。用户可以灵活便捷地使用公共镜像或者私有镜像申请弹性云服务器。同时，用户还能通过已有的弹性云服务器或者外部镜像文件创建私有镜像，这样能快速轻松地启动能满足您一切需求的新弹性云服务器。例如，如果您的应用程序是网站或Web服务，您的镜像可能会包含Web服务器、相关静态内容和动态页面代码。因此，您通过这个镜像创建弹性云服务器之后，您的Web服务器将启动。

#### 镜像类型

镜像类型	说明
公共镜像	常见的标准操作系统镜像，所有用户可见，包括操作系统以及预装的公共应用。公共镜像具有高度稳定性，皆为正版授权，请放心使用，您也可以根据实际需求自助配置应用环境或相关软件。
私有镜像	包含操作系统或业务数据、预装的公共应用以及用户的私有应用的镜像，仅用户个人可见。 私有镜像包括系统盘镜像、数据盘镜像、ISO 镜像和整机镜像，其中： <ul style="list-style-type: none"><li>● 系统盘镜像：包含用户运行业务所需的操作系统、应用软件的镜像。系统盘镜像可以用于创建云服务器，迁移用户业务到云。</li><li>● 数据盘镜像：只包含用户业务数据的镜像。数据盘镜像可以用于创建云硬盘，将用户的业务数据迁移到云上。</li><li>● ISO 镜像：将外部镜像的ISO文件注册到云平台的私有镜像。ISO 镜像是特殊的镜像，只能发放用作临时过渡的云服务器。</li><li>● 整机镜像：也叫全镜像，包含用户运行业务所需的操作系统、应用软件和业务数据的镜像。整机镜像基于增量备份制作，相比同样磁盘容量的系统盘镜像和数据盘镜像，创建效率更高。</li></ul>
共享镜像	由其他用户共享而来的私有镜像。 更多关于共享镜像的使用，请参见《镜像服务用户指南》的“共享镜像”章节。

## 1.5.2 Cloud-init

Cloud-init是开源的云初始化程序，能够对新创建弹性云服务器中指定的自定义信息（主机名、实例自定义数据等）进行初始化配置。

通过Cloud-init进行弹性云服务器的初始化配置，将对您使用弹性云服务器、镜像服务和弹性伸缩产生影响。

### 对镜像服务的影响

为了保证使用私有镜像新创建的弹性云服务器可以自定义配置，您需要在创建私有镜像前先安装Cloud-init/Cloudbase-init。

- 如果是Windows操作系统，需下载并安装Cloudbase-init。
- 如果是Linux操作系统，需下载并安装Cloud-init。

在镜像上安装Cloud-init/Cloudbase-init后，即可在创建弹性云服务器时，按照用户的需要自动设置弹性云服务器的初始属性。

更多关于安装的信息，请参见《镜像服务用户指南》。

### 对弹性云服务器的影响

- 在创建弹性云服务器时，如果选择的镜像支持Cloud-init特性，此时，您可以通过系统提供的“实例自定义数据注入”功能，注入初始化自定义信息（例如为弹性云服务器设置登录密码），完成弹性云服务器的初始化配置。
- 支持Cloud-init特性后，弹性云服务器的登录方式会产生影响。
- 对于运行中的弹性云服务器，支持Cloud-init特性后，用户可以通过查询、使用元数据，对正在运行的弹性云服务器进行配置和管理。

### 对弹性伸缩的影响

- 创建伸缩配置时，您可以使用“实例自定义数据注入”功能，指定弹性云服务器的初始化自定义信息。如果伸缩组使用了该伸缩配置，则伸缩组新创建的弹性云服务器会自动完成初始化配置。
- 对于已有的伸缩配置，如果其私有镜像没有安装Cloud-init/Cloudbase-init，则使用该伸缩配置的伸缩组创建的弹性云服务器在登录时会受到影响。

解决方法请参见《弹性伸缩用户指南》的“支持Cloud-init特性后，对使用弹性伸缩有哪些影响？”。

### 使用须知

- 使用Cloud-init特性时，需开启弹性云服务器所在VPC中子网的DHCP。
- 使用Cloud-init特性时，安全组出方向规则需满足如下要求：
  - 协议：TCP
  - 端口范围：80
  - 远端地址：169.254.0.0/16

#### 说明

如果您使用的是默认安全组出方向规则，则已经包括了如上要求，可以正常访问元数据。默认安全组出方向规则请参见[安全组](#)。

## 1.6 云硬盘

### 什么是云硬盘

云硬盘（Elastic Volume Service, EVS）可以为弹性云服务器提供高可靠、高性能、规格丰富并且可弹性扩展的块存储服务，满足不同场景的业务需求，适用于分布式文件系统、开发测试、数据仓库以及高性能计算等场景。

### 云硬盘的类型

不同类型云硬盘的性能各不相同，您可根据应用程序要求选择您所需的云硬盘。

更多关于云硬盘规格、性能等信息，请参见《云硬盘用户指南》。

### 云硬盘的磁盘模式

云硬盘的磁盘模式分为VBD（虚拟块存储设备，Virtual Block Device）类型和SCSI（小型计算机系统接口，Small Computer System Interface）类型。

- VBD类型：  
当您通过管理控制台创建云硬盘时，云硬盘的磁盘模式默认为VBD类型。VBD类型的云硬盘只支持简单的SCSI读写命令。
- SCSI类型：  
您可以通过管理控制台创建SCSI类型的云硬盘，该类型的云硬盘支持SCSI指令透传，允许弹性云服务器操作系统直接访问底层存储介质。除了简单的SCSI读写命令，SCSI类型的云硬盘还可以支持更高级的SCSI命令。

#### 说明

更多关于SCSI类型云硬盘的使用（如驱动安装），请参见《云硬盘用户指南》的“磁盘模式及使用方法”章节。

## 1.7 网络

### 虚拟私有云

虚拟私有云（Virtual Private Cloud，以下简称VPC）为弹性云服务器构建了一个逻辑上完全隔离的专有区域，您可以在自己的逻辑隔离区域中定义虚拟网络，为弹性云服务器构建一个逻辑上完全隔离的专有区域。您还可以在VPC中定义安全组、VPN、IP地址段、带宽等网络特性，方便管理、配置内部网络，进行安全、快捷的网络变更。同时，您可以自定义安全组内与组间弹性云服务器的访问规则，加强弹性云服务器的安全保护。

虚拟私有云更多信息，请参见《虚拟私有云用户指南》。

### 子网

子网是用来管理弹性云服务器网络平面的一个网络，可以提供IP地址管理、DNS服务，子网内的弹性云服务器IP地址都属于该子网。

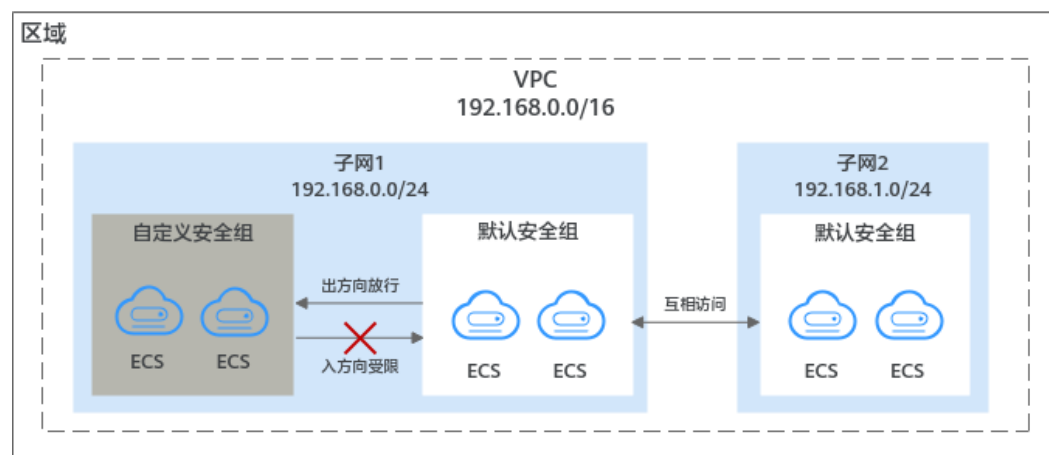
默认情况下，同一个VPC的所有子网内的弹性云服务器均可以进行通信，不同VPC的弹性云服务器不能进行通信。

## 安全组

安全组是一个逻辑上的分组，为具有相同安全保护需求并相互信任的弹性云服务器提供访问策略。安全组创建后，用户可以在安全组中定义各种访问规则，当弹性云服务器加入该安全组后，即受到这些访问规则的保护。

系统会为每个用户默认创建一个默认安全组，默认安全组的规则是在出方向上的数据报文全部放行，入方向访问受限，安全组内的弹性云服务器无需添加规则即可互相访问。

图 1-2 默认安全组



默认安全组规则如表1-66所示：

表 1-66 默认安全组规则

方向	协议	端口范围	目的地址/源地址	说明
出方向	全部	全部	目的地址： 0.0.0.0/0	允许所有出站流量的数据报文通过。
入方向	全部	全部	源地址：当前安全组(例如：sg-xxxxx)	仅允许安全组内的云服务器彼此通信，丢弃其他入站流量的全部数据报文。

## 弹性公网 IP

弹性公网IP (Elastic IP, 简称EIP) 提供独立的公网IP资源，包括公网IP地址与公网出口带宽服务。可以与弹性云服务器、裸金属服务器、虚拟IP、弹性负载均衡等资源灵活地绑定及解绑。

一个弹性公网IP只能绑定一个云资源使用。

## 1.8 安全

### 1.8.1 故障恢复

弹性云服务器通过云备份实现故障恢复，当云服务器或磁盘出现故障或者人为错误导致数据误删时，可以自助快速恢复数据。

#### 什么是云备份

云备份（Cloud Backup and Recovery, CBR）可以为云服务器、云硬盘提供简单易用的备份服务，当发生病毒入侵、人为误删除、软硬件故障等事件时，可将数据恢复到任意备份点。

云备份保障用户数据的安全性和正确性，确保业务安全。

#### 云服务器备份与云硬盘备份

目前弹性云服务器备份可以通过“云服务器备份”和“云硬盘备份”功能实现：

- 云服务器备份（推荐）：如果是对弹性云服务器中的所有云硬盘（系统盘和数据盘）进行备份，推荐使用云服务器备份功能，同时对所有云硬盘进行备份，避免因备份创建时间差带来的数据不一致问题。
- 云硬盘备份：如果对指定的单个或多个云硬盘（系统盘或数据盘）进行备份，推荐使用云硬盘备份功能，在保证数据安全的同时降低备份成本。

表 1-67 云服务器备份与云硬盘备份

对比维度	云服务器备份	云硬盘备份
备份/恢复对象	服务器中的所有云硬盘（系统盘和数据盘）。	指定的单个或多个磁盘（系统盘或数据盘）。
推荐场景	需要对整个云服务器进行保护。	系统盘没有个人数据，因而只需要对部分的数据盘进行备份。
优势	备份的同一个服务器下的所有磁盘数据具有一致性，即同时对所有云硬盘进行备份，不存在因备份创建时间差带来的数据不一致问题。	保证数据安全的同时降低备份成本。

## 1.9 约束与限制

本文介绍弹性云服务器（ECS）在使用过程中的约束和限制。

#### ECS 使用须知

- 禁止使用ECS搭建赌博、违规私服、违法跨境VPN等违法违规业务。
- 禁止使用ECS对电商网站开展刷单、刷广告等虚假交易操作。

- 禁止利用ECS对外部系统发起网络攻击，例如：暴力破解，传播病毒、木马等，包括但不限于DDoS攻击、CC攻击，Web攻击。
- 禁止使用ECS提供流量穿透服务。
- 禁止利用ECS搭建爬虫环境，对外部系统发起爬虫搜索。
- 未经外部系统主体授权，禁止利用ECS对外部系统发起扫描、渗透等探测行为。
- 禁止在ECS上部署任何违法违规网站和应用。
- 禁止利用ECS发送垃圾、骚扰邮件或从事其他侵犯个人隐私的行为。

## ECS 使用限制

- 请勿卸载云服务器硬件的驱动程序。
- 弹性云服务器不支持加载外接硬件设备，例如：硬件加密狗、U盘、外接硬盘、银行Ukey等。
- 请勿修改网卡的MAC地址。
- 弹性云服务器不支持二次虚拟化，即不支持安装虚拟化软件。
- 部分软件的鉴权模式可能会导致license与物理机的硬件信息绑定，云服务器的迁移操作可能会引起物理信息变更进而导致license失效。
- 由于物理机故障导致云服务器发生迁移，迁移时可能会出现重启或关机现象，建议您在系统业务进程配置自动拉起和开机自启动，或者通过业务集群部署、主备部署等方式实现业务的高可用。
- 建议您为部署核心关键业务的云服务器做好数据备份。
- 建议您为云服务器上的应用业务指标做好监控配置。
- 不建议您修改默认的DNS，如您有公网DNS配置需求，可以在云服务器上配置公网DNS和内网DNS。
- 弹性云服务器系统盘的引导模式（Legacy）为BIOS（Basic Input Output System）或者UEFI（Unified Extensible Firmware Interface）两者中的一种，由镜像的启动模式决定。
  - 通过切换操作系统操作，可以选择支持BIOS或UEFI启动模式的镜像为已有的弹性云服务器变更系统盘引导模式。
  - 通过“导入私有镜像”方式创建私有镜像时，可以创建UEFI或BIOS启动方式的私有镜像。

## Windows 操作系统使用须知

- 不要结束系统进程，结束系统进程可能会导致服务器发生蓝屏、重启。
- 建议您预留2G以上的内存，否则在使用过程中可能会出现蓝屏、卡顿、服务无法正常运行等情况。
- 不建议修改注册表，修改注册表可能会导致系统启动失败。如果一定要修改，请在修改前备份注册表。
- 不建议修改服务器时钟，修改服务器时钟可能会使DHCP租约失效进而导致IP丢失。
- 不建议删除重置密码进程CloudResetPwdAgent和CloudResetPwdUpdateAgent，否则会导致一键重置密码功能不可用。
- 不建议关闭虚拟内存，关闭虚拟内存可能会使系统性能降低或运行异常。
- 不要删除VMTool程序，会导致云服务器运行异常。

## Linux 操作系统使用须知

- 不要修改/etc/issue文件内容，否则可能导致系统发行版本无法被识别。
- 不要删除系统目录或文件，否则可能导致系统无法正常运行或启动。
- 不要修改系统目录的权限或名称，否则可能导致系统无法正常运行或启动。
- 请勿随意升级Linux操作系统的内核。
- 不建议删除重置密码进程CloudResetPwdAgent和CloudResetPwdUpdateAgent，否则会导致一键重置密码功能不可用。
- 不建议修改系统默认DNS SERVER（/etc/resolv.conf），否则可能会导致软件源、NTP等无法使用。
- 不建议修改主机默认的内网网络配置信息（IP、子网掩码及网关地址），否则可能会导致网络异常。
- 使用手动配置网络的场景，以静态地址为主，不需要系统的网络管理工具时（例如部署k8s的场景），不建议开启NetWorkManager服务，NetWorkManager服务会跟系统内部网络服务出现冲突，导致网络异常。

## 1.10 与其他服务的关系

### 弹性云服务器与其他服务的关系

- **弹性伸缩：**  
利用弹性伸缩，可以根据您定义的伸缩策略进行弹性云服务器的伸缩，以节省您的资源使用成本和提高资源利用率。
- **负载均衡：**  
将访问流量自动分发到多台弹性云服务器上，提高应用系统对外的服务能力，提高应用程序容错能力。
- **云硬盘：**  
可以将云硬盘挂载到弹性云服务器，并可以随时扩容云硬盘容量。
- **虚拟私有云：**  
为弹性云服务器提供一个逻辑上完全隔离的专有网络，您还可以在VPC中定义安全组、VPN、IP地址段、带宽等网络特性。用户可以通过VPC方便地管理、配置内部网络，进行安全、快捷的网络变更。同时，用户可以自定义安全组内与组间弹性云服务器的访问规则，加强弹性云服务器的安全保护。
- **镜像：**  
您可以通过镜像创建弹性云服务器，提高弹性云服务器的部署效率。
- **云监控：**  
当用户开通了弹性云服务器后，无需额外安装其他插件，即可在云监控查看对应服务的实例状态。
- **云备份：**  
提供对弹性云服务器的备份保护服务。支持对弹性云服务器中的所有云硬盘（系统盘和数据盘）进行备份，并利用备份数据恢复弹性云服务器数据。

## 1.11 权限管理

如果您需要对创建的ECS资源，给企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以使用统一身份认证服务（Identity and Access



Management, 简称IAM) 进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能, 可以帮助您安全地控制资源的访问。

通过IAM, 您可以在账号中给员工创建IAM用户, 并授权控制他们对资源的访问范围。例如您的员工中有负责软件开发的人员, 您希望他们拥有ECS的使用权限, 但是不希望他们拥有删除ECS等高危操作的权限, 那么您可以使用IAM为开发人员创建用户, 通过授予仅能使用ECS, 但是不允许删除ECS的权限策略, 控制他们对ECS资源的使用范围。

如果账号已经能满足您的要求, 不需要创建独立的IAM用户进行权限管理, 您可以跳过本章节, 不影响您使用ECS服务的其他功能。

IAM是提供权限管理的基础服务, 无需付费即可使用, 您只需要为您账号中的资源进行付费。关于IAM的详细介绍, 请参见《IAM产品介绍》。

## ECS 权限

默认情况下, 新建的IAM用户没有任何权限, 您需要将其加入用户组, 并给用户组授予策略或角色, 才能使得用户组中的用户获得对应的权限, 这一过程称为授权。授权后, 用户就可以基于被授予的权限对云服务进行操作。

ECS部署时通过物理区域划分, 为项目级服务。授权时, “作用范围” 需要选择“区域级项目”, 然后在指定区域对应的项目中设置相关权限, 并且该权限仅对此项目生效; 如果在“所有项目”中设置权限, 则该权限在所有区域项目中都生效。访问ECS时, 需要先切换至授权区域。

权限根据授权精细程度分为角色和策略。

- **角色:** IAM最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度, 提供有限的服务相关角色用于授权。由于各服务之间存在业务依赖关系, 因此给用户授予角色时, 可能需要一并授予依赖的其他角色, 才能正确完成业务。角色并不能满足用户对精细化授权的要求, 无法完全达到企业对权限最小化的安全管控要求。
- **策略:** IAM最新提供的一种细粒度授权的能力, 可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式, 能够满足企业对权限最小化的安全管控要求。例如: 针对ECS服务, 管理员能够控制IAM用户仅能对某一类云服务器资源进行指定的管理操作。

多数细粒度策略以API接口为粒度进行权限拆分, ECS支持的API授权项请参见《弹性云服务器API参考》中“ECS策略及授权项说明”章节。

如表1-68和表1-69所示, 包括了ECS的所有系统策略和系统角色。

表 1-68 ECS 系统策略 (推荐)

策略名称	描述	策略内容
ECS FullAccess	弹性云服务器管理员权限, 拥有该权限的用户可以拥有ECS支持的全部权限, 包括创建、删除、查询、变更规格等操作。	<a href="#">ECS FullAccess策略内容</a>
ECS CommonOperations	弹性云服务器普通用户权限, 拥有该权限的用户可以执行开机、关机、重启、查询弹性云服务器的操作。	<a href="#">ECS CommonOperations策略内容</a>



策略名称	描述	策略内容
ECS ReadOnlyAccess	弹性云服务器只读权限，拥有该权限的用户仅能查看弹性云服务数据。	<a href="#">ECS ReadOnlyAccess 策略内容</a>

表 1-69 ECS 系统角色

角色名称	描述	角色内容
Server Administrator	弹性云服务器的所有执行权限，该角色有依赖，需要在同项目中勾选依赖的角色：Tenant Guest。 如果在操作过程中涉及其他服务资源的创建、删除、变更等，则还需要在同项目中勾选对应服务的 Administrator 权限。 例如：在控制台创建 ECS 时如需创建新的 VPC，则需额外授予创建 VPC 的 VPC Administrator 权限。	<a href="#">Server Administrator 角色内容</a>

表 1-70 列出了 ECS 常用操作与系统策略的授权关系，您可以参照该表选择合适的系统策略。

表 1-70 ECS 操作与系统策略关系

操作	ECS FullAccess	ECS CommonOperations	ECS ReadOnlyAccess
创建弹性云服务器	√	x	x
控制台远程登录云服务器	√	√	x（不支持 VNC 方式远程登录）
查看弹性云服务器列表	√	√	√
查询弹性云服务器详情	√	√	√
修改弹性云服务器	√	x	x
启动弹性云服务器	√	√	x
关闭弹性云服务器	√	√	x
重启弹性云服务器	√	√	x
删除弹性云服务器	√	x	x
重装操作系统	√	x	x
切换操作系统	√	x	x

操作	ECS FullAccess	ECS CommonOperations	ECS ReadOnlyAccess
弹性云服务器挂载磁盘	√	x	x
弹性云服务器卸载磁盘	√	x	x
查询弹性云服务器磁盘列表	√	√	√
弹性云服务器挂载网卡	√	x	x
弹性云服务器卸载网卡	√	x	x
查询弹性云服务器网卡列表	√	√	√
变更弹性云服务器规格	√	x	x
查询弹性云服务器规格列表	√	√	√
查询云服务器组	√	√	√

## ECS FullAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:*",
        "evs:*get",
        "evs:*list",
        "evs:volumes:create",
        "evs:volumes:delete",
        "evs:volumes:attach",
        "evs:volumes:detach",
        "evs:volumes:manage",
        "evs:volumes:update",
        "evs:volumes:use",
        "evs:volumes:uploadImage",
        "evs:snapshots:create",
        "vpc:*get",
        "vpc:*list",
        "vpc:networks:create",
        "vpc:networks:update",
        "vpc:subnets:update",
        "vpc:subnets:create",
        "vpc:ports:*",
        "vpc:routers:get",
        "vpc:routers:update",
        "vpc:securityGroups:*",
        "vpc:securityGroupRules:*",
        "vpc:floatingIps:*",
        "vpc:publicIps:*",
        "ims:images:create",
        "ims:images:delete",
        "ims:images:get",
        "ims:images:list",
        "ims:images:update",

```

```
        "ims:images:upload"  
      ]  
    }  
  ]  
}
```

## ECS CommonOperations 策略内容

```
{  
  "Version": "1.1",  
  "Statement": [  
    {  
      "Effect": "Allow",  
      "Action": [  
        "ecs:*:get*",  
        "ecs:*:list*",  
        "ecs:*:start",  
        "ecs:*:stop",  
        "ecs:*:reboot",  
        "ecs:blockDevice:use",  
        "ecs:cloudServerFpgaImages:relate",  
        "ecs:cloudServerFpgaImages:register",  
        "ecs:cloudServerFpgaImages:delete",  
        "ecs:cloudServerFpgaImages:unrelate",  
        "ecs:cloudServers:setAutoRecovery",  
        "ecs:cloudServerPasswords:reset",  
        "ecs:cloudServerPorts:modify",  
        "ecs:cloudServers:vnc",  
        "ecs:diskConfigs:use",  
        "ecs:securityGroups:use",  
        "ecs:serverGroups:manage",  
        "ecs:serverFloatingIps:use",  
        "ecs:serverKeyPairs:*",  
        "ecs:serverPasswords:manage",  
        "ecs:servers:createConsole",  
        "ecs:servers:createImage",  
        "ecs:servers:setMetadata",  
        "ecs:servers:setTags",  
        "ecs:serverVolumes:use",  
        "evs:*:get*",  
        "evs:*:list*",  
        "evs:snapshots:create",  
        "evs:volumes:uploadImage",  
        "evs:volumes:delete",  
        "evs:volumes:update",  
        "evs:volumes:attach",  
        "evs:volumes:detach",  
        "evs:volumes:manage",  
        "evs:volumes:use",  
        "vpc:*:get*",  
        "vpc:*:list*",  
        "vpc:floatingIps:create",  
        "vpc:floatingIps:update",  
        "vpc:floatingIps:delete",  
        "vpc:publicIps:update",  
        "vpc:publicIps:delete",  
        "ims:images:create",  
        "ims:images:delete",  
        "ims:images:get",  
        "ims:images:list",  
        "ims:images:update",  
        "ims:images:upload"  
      ]  
    }  
  ]  
}
```

## ECS ReadOnlyAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:*get*",
        "ecs:*list*",
        "ecs:serverGroups:manage",
        "ecs:serverVolumes:use",
        "evs:*get*",
        "evs:*list*",
        "vpc:*get*",
        "vpc:*list*",
        "ims:*get*",
        "ims:*list*"
      ]
    }
  ]
}
```

## Server Administrator 角色内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "ecs:*",
        "evs:*get",
        "evs:*list",
        "evs:volumes:create",
        "evs:volumes:delete",
        "evs:volumes:attach",
        "evs:volumes:detach",
        "evs:volumes:manage",
        "evs:volumes:update",
        "evs:volumes:uploadImage",
        "evs:snapshots:create",
        "vpc:*get",
        "vpc:*list",
        "vpc:networks:create",
        "vpc:networks:update",
        "vpc:subnets:update",
        "vpc:subnets:create",
        "vpc:routers:get",
        "vpc:routers:update",
        "vpc:ports:*",
        "vpc:privateIps:*",
        "vpc:securityGroups:*",
        "vpc:securityGroupRules:*",
        "vpc:floatingIps:*",
        "vpc:publicIps:*",
        "vpc:bandwidths:*",
        "vpc:firewalls:*",
        "ims:images:create",
        "ims:images:delete",
        "ims:images:get",
        "ims:images:list",
        "ims:images:update",
        "ims:images:upload"
      ],
      "Effect": "Allow"
    }
  ]
}
```

## 1.12 区域和可用区

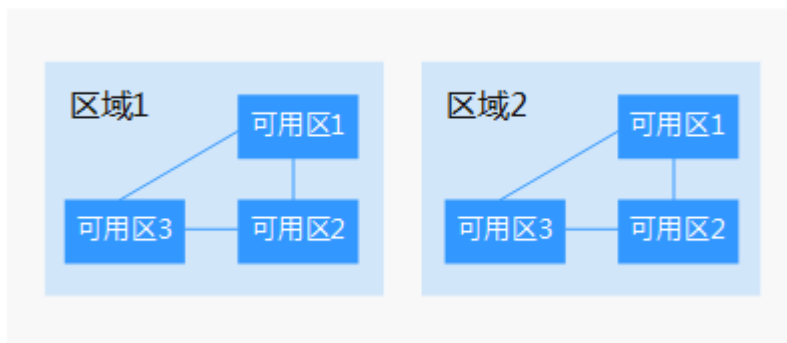
### 什么是区域、可用区？

区域和可用区用来描述数据中心的位置，您可以在特定的区域、可用区创建资源。

- 区域（Region）指物理的数据中心。每个区域完全独立，这样可以实现最大程度的容错能力和稳定性。资源创建成功后不能更换区域。
- 可用区（AZ，Availability Zone）是同一区域内，电力和网络互相隔离的物理区域，一个可用区不受其他可用区故障的影响。一个区域内可以有多个可用区，不同可用区之间物理隔离，但内网互通，既保障了可用区的独立性，又提供了低价、低时延的网络连接。

图1-3阐明了区域和可用区之间的关系。

图 1-3 区域和可用区



### 如何选择区域？

建议就近选择靠近您或者您的目标用户的区域，这样可以减少网络时延，提高访问速度。

### 如何选择可用区？

是否将资源放在同一可用区内，主要取决于您对容灾能力和网络时延的要求。

- 如果您的应用需要较高的容灾能力，建议您将资源部署在同一区域的不同可用区内。
- 如果您的应用要求实例之间的网络延时较低，则建议您将资源创建在同一可用区内。

### 区域和终端节点

当您通过API使用资源时，您必须指定其区域终端节点。有关的区域和终端节点的更多信息，请参阅[地区和终端节点](#)。

# 2 快速入门

## 2.1 创建弹性云服务器

### 2.1.1 概述

#### 操作场景

如果您需要在服务器上部署相关业务，较之物理服务器，弹性云服务器的创建成本较低，并且可以在几分钟之内快速获得基于云服务平台的弹性云服务器设施，并且这些基础设施是弹性的，可以根据需求伸缩。下面介绍如何在管理控制台创建弹性云服务器。

#### 创建流程：

- [步骤一：基础配置](#)
- [步骤二：网络配置](#)
- [步骤三：高级配置](#)
- [步骤四：确认订单](#)


### 2.1.2 步骤一：基础配置

#### 进入创建页面

1. 登录管理控制台。
  2. 选择“计算 > 弹性云服务器”。
  3. 单击“创建弹性云服务器”。
- 系统进入创建页。

#### 基础配置

1. 确认当前“区域”。

如果所在区域不正确，请单击页面左上角的  进行切换。

## 2. 选择“可用区”。

可用区指在同一区域下，电力、网络隔离的物理区域，可用区之间内网互通，不同可用区之间物理隔离。

- 如果您需要提高应用的高可用性，建议您将弹性云服务器创建在不同的可用区。
- 如果您需要较低的网络时延，建议您将弹性云服务器创建在相同的可用区。

### 📖 说明

- 创建云服务器时可以选择随机分配可用区，系统使用哈希算法基于用户的UUID（Universally Unique Identifier）选择一个默认的AZ。
- 不同可用区上线的云服务器类型、规格不同。如需查看当前平台支持的全部云服务器类型、规格，请选择“随机分配”可用区，系统会根据您选择的规格，分配弹性云服务器所在的可用区。

例如：S3实例仅在AZ1上线；S2实例在AZ2、AZ3有货，在AZ1已售罄。选择“随机分配”可用区，可以同时查看当前平台支持的S3实例和S2实例。如果选择创建S3实例，那么系统分配至AZ1可用区；如果选择创建S2实例，系统随机分配至AZ2或AZ3可用区。

## 3. 选择“规格”。

云服务平台提供了多种类型的弹性云服务器供您选择，针对不同的应用场景，可以选择不同规格的弹性云服务器。您可以在列表中查看已上线的弹性云服务器类型、规格，或输入规格名称，或根据vCPU、内存大小搜索目标规格。

### 📖 说明

- 选择云服务器类型前，请认真阅读各种类型云服务器的介绍和注意事项，具体内容请参见[实例类型](#)。

## 4. 选择“镜像”。

### - 公共镜像

常见的标准操作系统镜像，所有用户可见，包括操作系统以及预装的公共应用。请根据您的实际情况自助配置应用环境或相关软件。

### - 私有镜像

用户基于弹性云服务器创建的个人镜像，仅用户自己可见。包含操作系统、预装的公共应用以及用户的私有应用。选择私有镜像创建弹性云服务器，可以节省您重复配置弹性云服务器的时间。

### 📖 说明

- 当整机镜像为可用区内可用时（状态栏为“正常”，且提示“AZx可用”），那么该整机镜像只能在该可用区中创建云服务器，不支持在创建时修改系统盘和数据盘的SCSI模式、共享属性。
- 当整机镜像为区域内可用时（状态栏为“正常”，且未提示“AZx可用”），那么该整机镜像可以在整个区域内创建云服务器，支持在创建时修改数据盘的SCSI模式、共享属性。
- 为了确保使用私有镜像创建的弹性云服务器开启网卡多队列功能，建议您在创建私有镜像时完成网卡多队列的配置。弹性云服务器开启网卡多队列功能可以将弹性云服务器中的网卡中断分散给不同的CPU处理，以满足网卡的需求，从而提升网络PPS和带宽性能。

详细操作请参考镜像服务常见问题：如何设置镜像的网卡多队列属性？

### - 共享镜像

您将接受其他用户共享的私有镜像，作为自己的镜像进行使用。


## 5. 设置磁盘，包括系统盘和数据盘。

- 系统盘

弹性云服务器支持使用的磁盘类型，请参见[云硬盘](#)。

- 数据盘

您可以为弹性云服务器添加多块数据盘，并设置数据盘的相关功能。创建弹性云服务器时，您最多可以为弹性云服务器添加23块数据盘，磁盘大小可以根据需要自定义。

请单击“展开  ”后，设置如下功能。

- SCSI：勾选后，数据盘的磁盘模式为SCSI。更多关于SCSI盘、支持挂载SCSI盘的弹性云服务器等信息，请参见[云硬盘](#)。
- 共享盘：勾选后，数据盘为共享云硬盘。该共享盘可以同时挂载给多台弹性云服务器使用。

6. 单击“下一步：网络配置”。

## 2.1.3 步骤二：网络配置

### 网络配置

1. 设置“网络”：在下拉列表中选择可用的虚拟私有云、子网，并设置私有IP地址的分配方式。

弹性云服务器网络使用虚拟私有云（VPC）提供的网络，包括子网、安全组等。

您可以选择使用已有的虚拟私有云网络，或者创建新的虚拟私有云。

更多关于虚拟私有云的信息，请参见《虚拟私有云用户指南》。

#### 说明

- 弹性云服务器使用的VPC网络DHCP不能禁用。
  - 第一次使用云服务平台时，系统将自动为您创建一个虚拟私有云，包括安全组、网卡。
2. 添加“扩展网卡”：可选配置。您可以添加多张扩展网卡，并指定网卡（包括主网卡）的IP地址。

#### 说明

指定IP地址时，如果是批量创建多台弹性云服务器：

- 此时，该IP地址为起始IP地址。
  - 请确保IP地址在子网范围内且连续可用。
  - 其他子网不能与指定IP的子网相同。
- 系统默认分配IPv4地址，勾选“自动分配IPv6地址”后，网卡支持双栈类型，分配IPv4/IPv6双栈地址。在同一VPC内，弹性云服务器通过IPv6地址在双栈ECS之间进行内网访问。如需访问公网，您需要开启“IPv6带宽”并选择共享带宽，此时弹性云服务器可以通过IPv6地址与互联网上的IPv6网络进行访问。
- 弹性云服务器创建成功后，需手动配置弹性云服务器，动态获取IPv6地址，启用IPv6功能。具体操作请参见“[动态获取IPv6地址](#)”。



### 说明

- 当前仅支持在创建弹性云服务器时开启IPv6功能，开启成功后，不能修改。如果创建弹性云服务器时未开启“IPv6带宽”功能，系统支持在弹性云服务器创建成功后开启“IPv6带宽”功能。
  - 暂不支持独享带宽。
3. 设置“安全组”：在下拉列表中选择可用的安全组，或新建安全组使用。
- 安全组用来实现安全组内和安全组间弹性云服务器的访问控制，加强弹性云服务器的安全保护。用户可以在安全组中定义各种访问规则，当弹性云服务器加入该安全组后，即受到这些访问规则的保护。
- 创建弹性云服务器时，可支持选择多个安全组（建议不超过5个）。此时，弹性云服务器的访问规则遵循几个安全组规则的并集。

### 说明

弹性云服务器初始化需要确保安全组出方向规则满足如下要求：

- 协议：TCP
- 端口范围：80
- 远端地址：169.254.0.0/16

如果您使用的是默认安全组出方向规则，则已经包括了如上要求，可以正常初始化。默认安全组出方向规则为：

- 协议：ANY
- 端口范围：ANY
- 远端地址：0.0.0.0/16

4. 设置“弹性公网IP”。

弹性公网IP是指将公网IP地址和路由网络中关联的弹性云服务器绑定，以实现虚拟私有云内的弹性云服务器通过固定的公网IP地址对外提供访问服务。

您可以根据实际情况选择以下三种方式：

- 自动分配：自动为每台弹性云服务器分配独享带宽的弹性公网IP，带宽值可以由您设定。
- 使用已有：为弹性云服务器分配已有弹性公网IP。使用已有弹性公网IP时，不能批量创建弹性云服务器。
- 不使用：弹性云服务器不能与互联网互通，仅可作为私有网络中部署业务或者集群所需弹性云服务器进行使用。

5. 设置“带宽大小”。

根据业务需要，选择所需的带宽大小，单位Mbit/s。

## 2.1.4 步骤三：高级配置

### 高级配置

1. 设置“云服务器名称”。

名称可自定义，但需符合命名规则：只能由中文字符、英文字母、数字及“\_”、“-”、“.”组成。

如果同时创建多台弹性云服务器，系统会自动按序增加后缀。

一次创建多台弹性云服务器时，系统自动增加后缀，例如：输入ecs，云服务器名称为ecs-0001，ecs-0002，……。再次创建多台云服务器时，命名从上次最大值连续增加，例如：输入ecs，已有云服务器ecs-0010，新创云服务器名称为ecs-0011、ecs-0012、……，命名达到9999时，从0001开始。

允许重名：允许创建的云服务器名称相同。如果是批量创建，勾选“允许重名”后，批量创建的多台云服务器名称全部相同，不再按序增加。

本步骤中的“云服务器名称”，会被设置为虚拟机操作系统中初始的主机名。

### 说明

弹性云服务器操作系统中的主机名(hostname)的命名规则遵循 [RFC 952](#)和[RFC 1123](#)规范。

在将“云服务器名称”设置为虚拟机操作系统中的主机名时，为避免出现未知问题，建议使用a-z或0-9以及中划线“-”组成的名称命名。

- 以“.”、“-”、“\_”和中文字符开头将在弹性云服务器内部被忽略，从第一个非以上字符开始。
- 不作为开头的“.”和中文字符及后面内容将在弹性云服务器内部被忽略。

## 2. 设置“登录方式”。

### - 密码

指使用设置初始密码方式作为弹性云服务器的鉴权方式，此时，您可以通过用户名密码方式登录弹性云服务器。

Linux操作系统时为root用户的初始密码，Windows操作系统时为Administrator用户的初始密码。

密码复杂度需满足[表2-1](#)要求。

表 2-1 密码设置规则

参数	规则
密码	<ul style="list-style-type: none"><li>密码长度范围为8到26位。具体以界面显示为准。</li><li>密码至少包含以下4种字符中的3种：<ul style="list-style-type: none"><li>大写字母</li><li>小写字母</li><li>数字</li><li>Windows操作系统云服务器特殊字符：包括“\$”、“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“,”和“?”</li><li>Linux操作系统云服务器特殊字符：包括“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“^”、“,”、“{”、“}”和“?”</li></ul></li><li>密码不能包含用户名或用户名的逆序。</li><li>Windows操作系统的云服务器，不能包含用户名中超过两个连续字符的部分。</li></ul>

### 说明

系统不会定期自动修改弹性云服务器密码。为安全起见，建议您定期修改密码。

### - 使用镜像密码

如果使用Linux操作系统私有镜像创建云服务器，支持使用私有镜像的密码。请确保所选择的私有镜像已经设置了密码。

- 创建后设置

暂不设置弹性云服务器的密码。待弹性云服务器创建成功后，您需要单击“操作”列下的“重置密码”，根据界面提示，为弹性云服务器设置密码，然后登录弹性云服务器。

3. 设置“云备份”。

云备份提供对云硬盘和弹性云服务器的备份保护，并支持利用备份数据恢复云服务器和云硬盘的数据。云备份设置完成后，系统会将弹性云服务器绑定至云备份存储库并绑定所选备份策略，定期备份弹性云服务器。

您可以根据实际情况选择以下三种方式。

- 自动分配：

- i. 输入云备份存储库的名称：只能由中文字符、英文字母、数字、下划线、中划线组成，且长度小于等于64个字符。例如：vault-f61e。默认的命名规则为“vault\_xxxx”。
- ii. 输入存储库的容量：此容量为备份云服务器所需的容量。存储库的空间不能小于云服务器的空间。取值范围为[云服务器总容量，10485760]GB。
- iii. 设置备份策略：在下拉列表中选择备份策略，或进入云备份控制台查看或编辑备份策略。

- 使用已有：

- i. 选择云备份存储库的名称：在下拉列表中选择已有的云备份存储库。
- ii. 设置备份策略：在下拉列表中选择备份策略，或进入云备份控制台查看或编辑备份策略。

- 不使用：跳过云备份的配置步骤。如云服务器创建成功后仍需设置备份保护，请进入云备份控制台找到目标存储库，绑定服务器。

4. 设置“云服务器组（可选）”

可选配置，云服务器组内的弹性云服务器将遵循反亲和策略，尽量分散地创建在不同主机上。创建云服务器组的详细操作，请参见[管理云服务器组](#)。

#### 说明

包含本地盘的弹性云服务器无法在创建后加入弹性云服务器组，如需使用弹性云服务器组功能，请在创建时选择弹性云服务器组。

5. 设置“自动恢复”。

开启自动恢复功能后，当弹性云服务器所在的硬件出现故障时，系统会自动将弹性云服务器迁移至正常的物理机，保障您受到的影响最小，该过程会导致云服务器重启。如果未开启自动恢复功能，当云服务器所在的硬件出现故障时，需等待系统管理员进行问题修复。更多关于自动恢复功能的介绍，请参见[自动恢复](#)。

#### 说明

对于包含如下特殊资源的云服务器，暂不支持自动恢复功能：

- 本地盘
- 直通FPGA卡
- 直通IB网卡

6. 配置“高级选项”：如需使用“高级选项”中的功能，请勾选“现在配置”。否则，请勿勾选。

- 实例自定义数据注入  
可选配置，主要用于创建云服务器时向云服务器注入实例自定义数据。配置后，云服务器首次启动时会自行注入数据信息。  
例如：您可以通过注入一段脚本，激活待创建云服务器的root用户权限，注入成功后，您可以使用root用户登录弹性云服务器。  
实例自定义数据注入的详细操作，请参见[通过实例自定义数据配置ECS实例](#)。
  - 委托  
可选配置。当需要与其他账号共享云服务器资源，或者委托更专业的人或团队来代为管理时，租户管理员可以在IAM创建委托并授予云服务器资源的管理权限。被委托方使用自己的用户登录系统后，切换到您的账号下管理资源，避免您将自己的安全凭证（密码）共享给其他账号，确保了您的账号安全。  
如果您在IAM上创建了委托，可以通过单击下拉列表选择委托名称，获取相应权限。如何创建委托，请参见《统一身份认证服务用户指南》。
7. 单击“下一步：确认配置”。

## 2.1.5 步骤四：确认配置

### 确认订单

1. 在“确认配置”页面，查看云服务器配置详情。
2. 设置您创建弹性云服务器的数量。
3. 如果您确认配置无误，单击“立即申请”。

### 后续处理

- 如果您是使用外部镜像文件制作的私有镜像创建的云服务器，且未安装一键式重置密码插件，那么建议您成功创建、登录云服务器后，安装密码重置插件。当云服务器的密码丢失或过期时，则可以应用一键式重置密码功能，给云服务器设置新密码。  
具体操作请参考安装一键式重置密码插件。
- 对于设置了启动自动恢复功能的云服务器，创建成功后，可以通过查看“失败任务”，检查自动恢复功能是否开启成功，具体操作请参见[查看失败任务](#)。

## 2.2 登录弹性云服务器

### 登录 Windows 弹性云服务器（中国区）

有2种方法可以登录Windows弹性云服务器：管理控制台提供的“VNC远程登录方式”和“MSTSC方式”。

- 管理控制台远程登录（VNC方式）  
具体操作，请参见[通过控制台VNC登录Windows ECS](#)。
- 远程桌面连接（MSTSC方式）  
具体操作，请参见[通过本地Windows主机登录Windows ECS](#)。

## 登录 Windows 弹性云服务器（非中国区）

有2种方法可以登录Windows弹性云服务器：管理控制台提供的“VNC远程登录方式”和“MSTSC方式”。

- 管理控制台远程登录（VNC方式）  
具体操作，请参见[通过控制台VNC登录Windows ECS](#)。
- 远程桌面连接（MSTSC方式）  
具体操作，请参见[通过本地Windows主机登录Windows ECS](#)。

## 登录 Linux 弹性云服务器（中国区）

弹性云服务器设置的登录鉴权方式不同，登录弹性云服务器的方法也存在差异。

- 首次登录密码方式鉴权的弹性云服务器时，有2种方法可以登录：
  - 管理控制台远程登录（VNC方式）。登录用户名为“root”  
VNC方式，具体操作请参见[通过VNC登录Linux ECS](#)。
  - SSH密码方式，登录用户名为“root”，且弹性云服务器必须绑定弹性公网IP。  
SSH密码方式，具体操作请参见[通过SSH密码方式登录Linux ECS](#)。

## 登录 Linux 弹性云服务器（非中国区）

根据需要选择登录方式，登录弹性云服务器。

- VNC方式  
具体操作，请参见[通过VNC登录Linux ECS](#)。

## 后续处理

- 创建弹性云服务器时如果添加了数据盘，待登录成功后，需要初始化数据盘。  
操作方法请参见[场景及磁盘分区形式介绍](#)。
- 对于部分类型的弹性云服务器，待登录成功后需安装驱动。[实例类型](#)章节列举了当前提供的云服务器类型，并对每种云服务器类型的功能、使用进行介绍，具体请以各类型云服务器的“使用须知”为准。

## 2.3 初始化数据盘

### 2.3.1 场景及磁盘分区形式介绍

如果您在创建弹性云服务器时添加了数据盘，待登录成功后，需要执行初始化数据盘操作，即格式化云硬盘，之后云硬盘才可以正常使用。

## 操作场景

磁盘挂载至云服务器后，需要登录云服务器初始化磁盘，即格式化磁盘，之后磁盘才可以正常使用。

- 系统盘

系统盘不需要初始化，创建云服务器时会自带系统盘并且自动初始化，默认磁盘分区形式为主启动记录分区（MBR, Master boot record）。

- 数据盘
  - 创建云服务器时直接创建数据盘，数据盘会自动挂载至云服务器。
  - 单独创建数据盘，然后将该数据盘挂载至云服务器。

以上两种情况创建的数据盘挂载至云服务器后，均需要初始化后才可以使⽤，请您根据业务的实际规划选择合适的分区方式。

## 磁盘分区操作指导

常用的磁盘分区形式如表2-2所示，并且针对Linux操作系统，不同的磁盘分区形式需要选择不同的分区工具。

表 2-2 磁盘分区形式

磁盘分区形式	支持最大磁盘容量	支持分区数量	Linux分区工具
主启动记录分区（MBR）	2 TiB	<ul style="list-style-type: none"><li>• 4个主分区</li><li>• 3个主分区和1个扩展分区</li></ul> MBR分区包含主分区和扩展分区，其中扩展分区里面可以包含若干个逻辑分区。扩展分区不可以直接使用，需要划分成若干个逻辑分区才可以使用。以创建6个分区为例，以下两种分区情况供参考： <ul style="list-style-type: none"><li>• 3个主分区，1个扩展分区，其中扩展分区中包含3个逻辑分区。</li><li>• 1个主分区，1个扩展分区，其中扩展分区中包含5个逻辑分区。</li></ul>	以下两种工具均可以使用： <ul style="list-style-type: none"><li>• fdisk工具</li><li>• parted工具</li></ul>
全局分区表（GPT, Guid Partition Table）	18 EiB 1 EiB = 1048576 TiB	不限制分区数量 GPT格式下没有主分区、扩展分区以及逻辑分区之分。	parted工具

### 2.3.2 初始化 Windows 数据盘（Windows 2008）

#### 操作场景

本文以云服务器的操作系统为“Windows Server 2008 R2 Enterprise 64bit”为例，提供磁盘的初始化操作指导。

MBR格式分区支持的磁盘最大容量为2 TiB，GPT分区表最大支持的磁盘容量为18 EiB，因此当为容量大于2 TiB的磁盘分区时，请采用GPT分区方式。具体操作请参见[初始化容量大于2TiB的Windows数据盘（Windows 2008）](#)。关于磁盘分区形式的更多介绍，请参见[场景及磁盘分区形式介绍](#)。



不同云服务器的操作系统的格式化操作可能不同，本文仅供参考，具体操作步骤和差异请参考对应的云服务器操作系统的产品文档。

### 须知

首次使用磁盘时，如果您未参考本章节对磁盘执行初始化操作，主要包括创建分区和文件系统等操作，那么当后续扩容磁盘时，新增容量部分的磁盘可能无法正常使用。

## 前提条件

- 已挂载数据盘至云服务器，且该数据盘未初始化。
- 已登录。
  - 弹性云服务器请参见《弹性云服务器用户指南》。
  - 裸金属服务器请参见《裸金属服务器用户指南》。

## 操作指导

**步骤1** 在云服务器桌面，选择“开始”，右键单击后在菜单列表中选择“计算机”，选择“管理”。

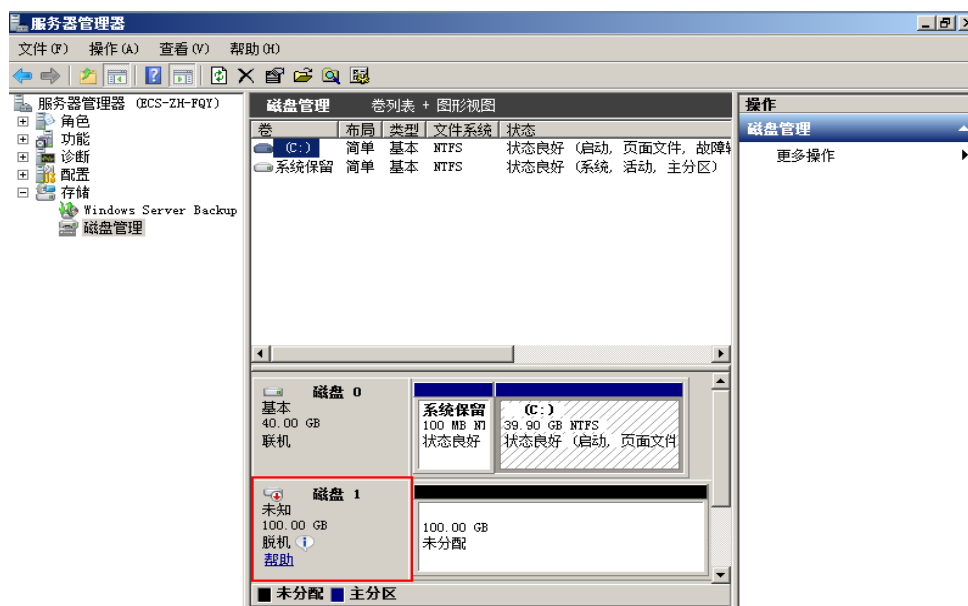
弹出“服务器管理”窗口。

**步骤2** 在左侧导航树中，选择“存储 > 磁盘管理”。

进入“磁盘管理”页面。

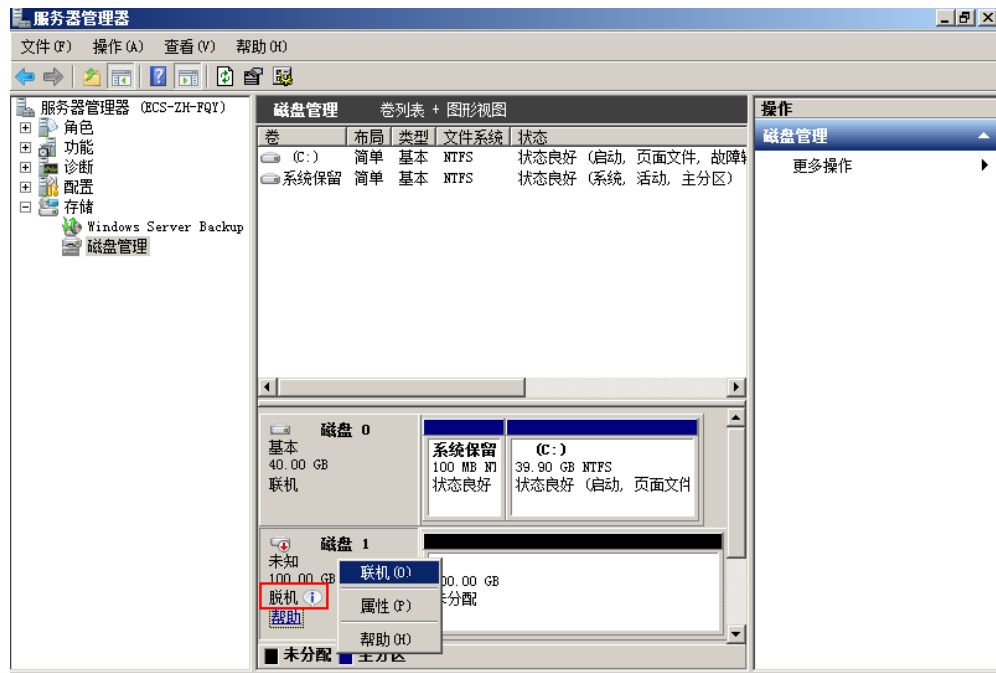
- 如果如图2-1，新挂载磁盘为“脱机”状态，请执行**步骤3**。
- 如果如图2-4，直接弹出“初始化磁盘”对话框，执行**步骤5**。

图 2-1 磁盘管理



**步骤3** 在右侧窗格中出现磁盘列表，在磁盘1区域，右键单击后在菜单列表中选择“联机”，进行联机。

图 2-2 联机

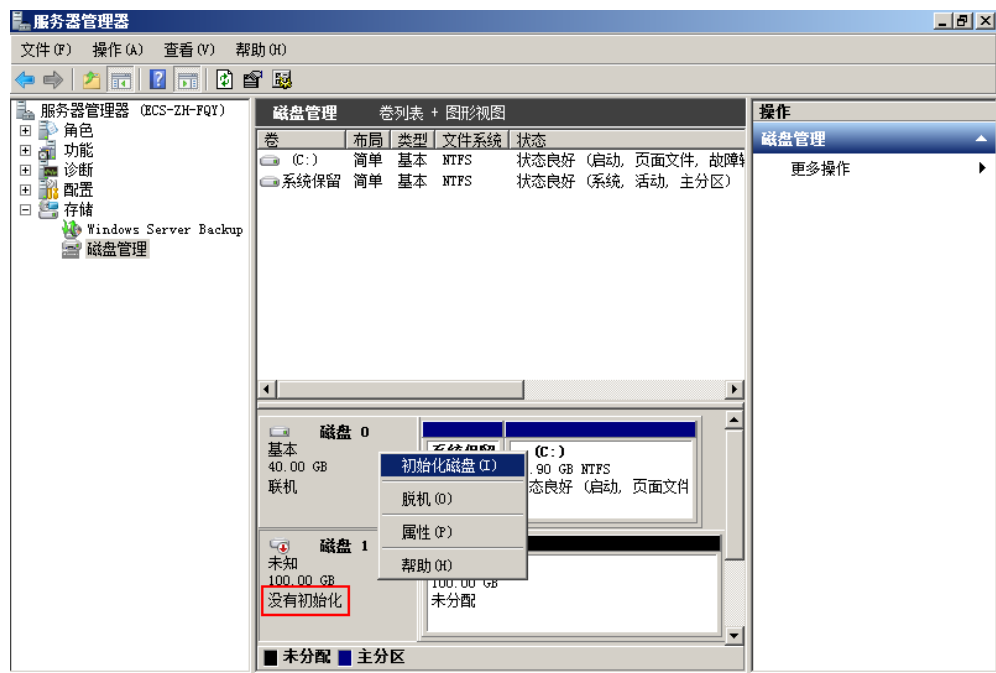


### 说明

如果新增磁盘处于脱机状态，需要先联机然后进行初始化。

**步骤4** 联机后，磁盘1由“脱机”状态变为“没有初始化”，右键单击在菜单列表中选择“初始化磁盘”。如图2-3所示。

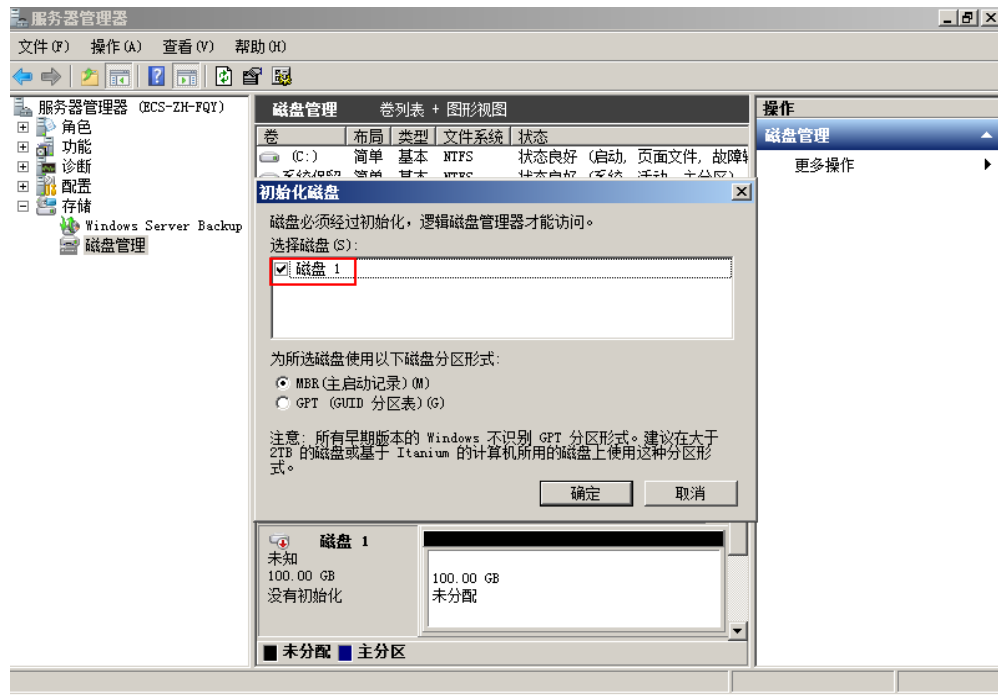
图 2-3 初始化磁盘



**步骤5** 在“初始化磁盘”对话框中显示需要初始化的磁盘，选中“MBR (主启动记录)”或者“GPT (GUID 分区表)”，单击“确定”，如图2-4所示。



图 2-4 未分配磁盘



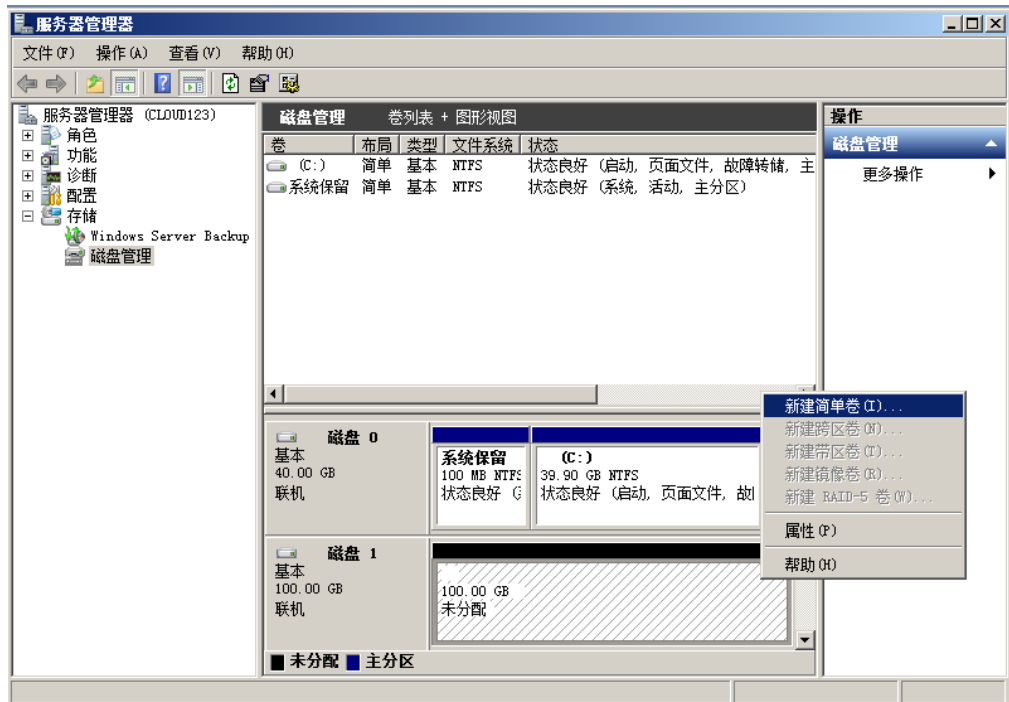
### 须知

MBR支持的云硬盘最大容量为2 TiB，GPT最大支持的云硬盘容量为18 EiB，当前数据盘支持的最大容量为32 TiB，如果您需要使用大于2 TiB的云硬盘容量，分区形式请采用GPT。

当云硬盘已经投入使用后，此时切换云硬盘分区形式时，云硬盘上的原有数据将会清除，因此请在云硬盘初始化时谨慎选择云硬盘分区形式。切换GPT分区形式前，请对云硬盘数据备份后，再格式化硬盘。

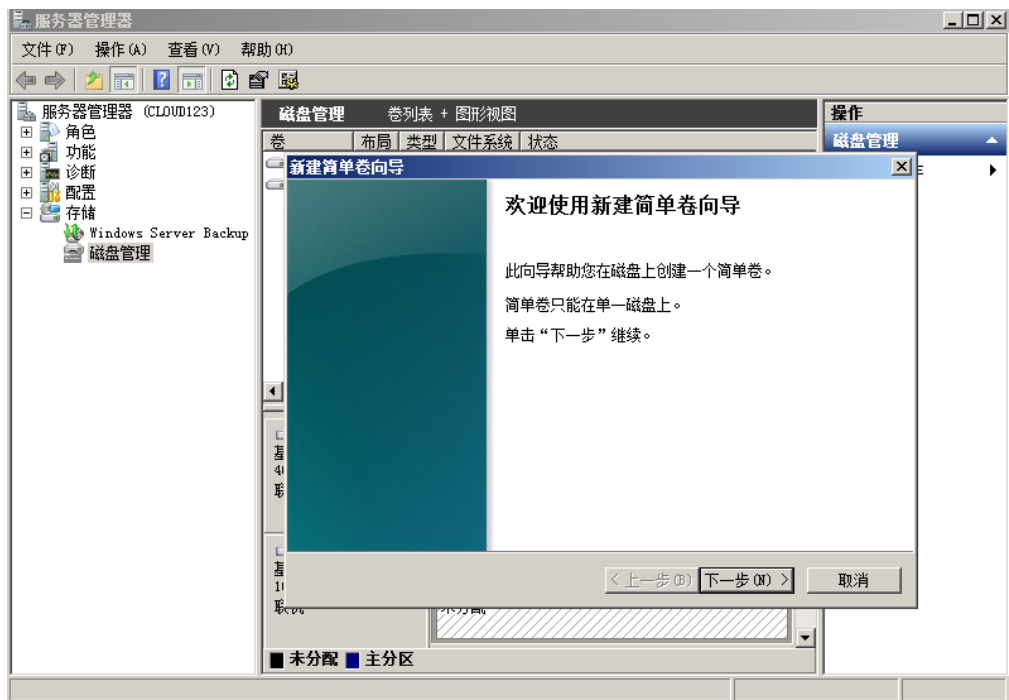
**步骤6** 右键单击磁盘上未分配的区域，选择“新建简单卷”，如图2-5所示。

图 2-5 新建简单卷



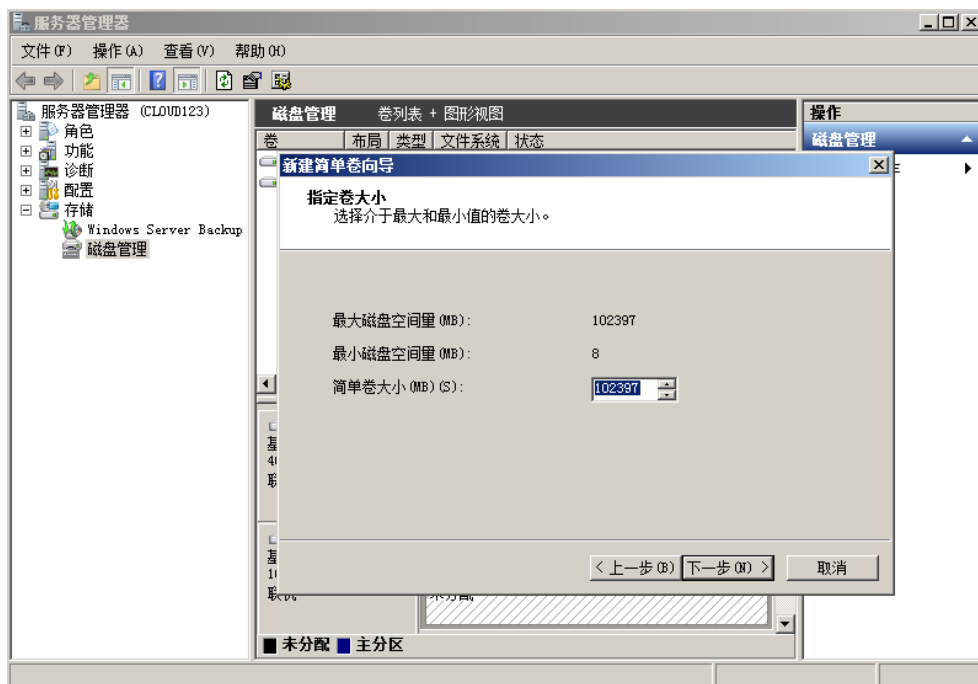
步骤7 弹出“新建简单卷向导”对话框，根据界面提示，单击“下一步”。

图 2-6 新建简单卷向导



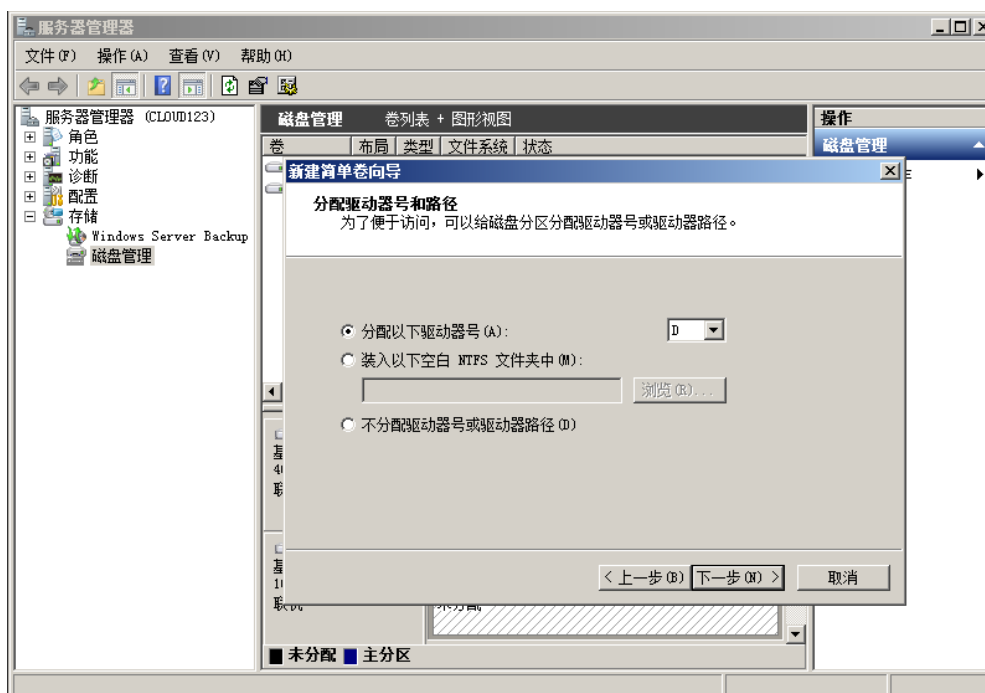
步骤8 根据需要指定卷大小，默认为最大值，单击“下一步”。

图 2-7 指定卷大小



步骤9 分配驱动器号，单击“下一步”。

图 2-8 分配驱动器号和路径



步骤10 勾选“按下列设置格式化这个卷”，并根据实际情况设置参数，格式化新分区，单击“下一步”完成分区创建。

图 2-9 格式化分区

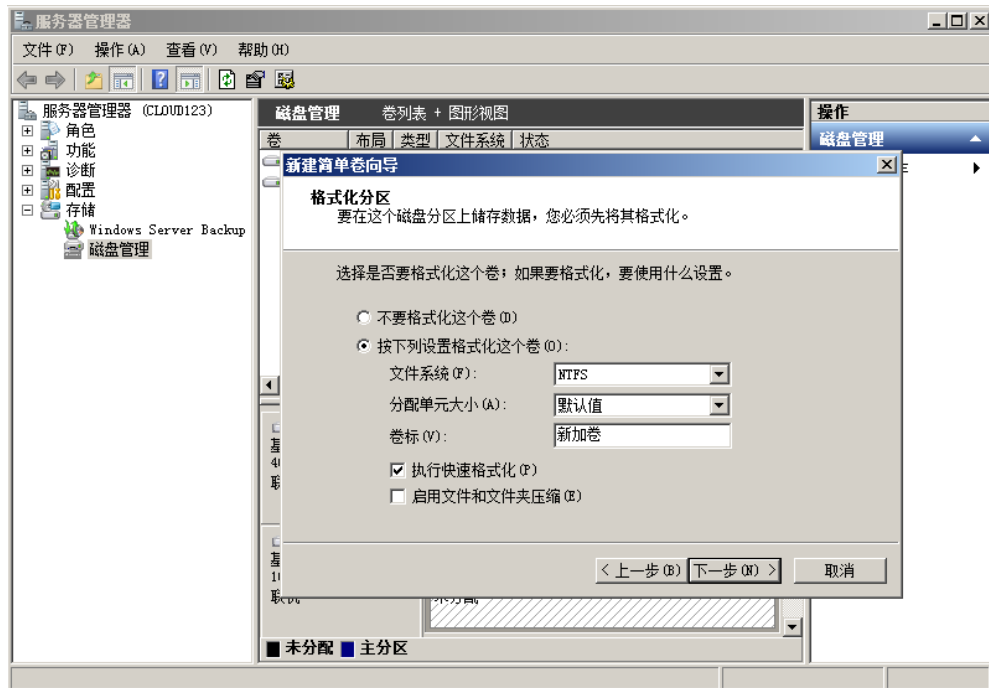
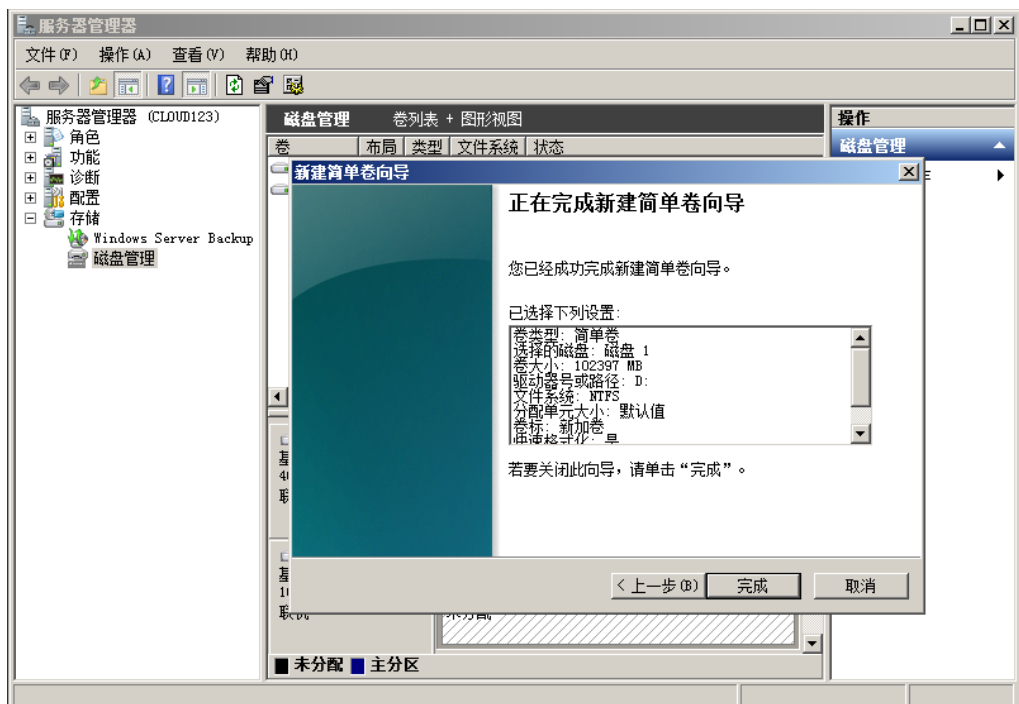


图 2-10 完成分区创建

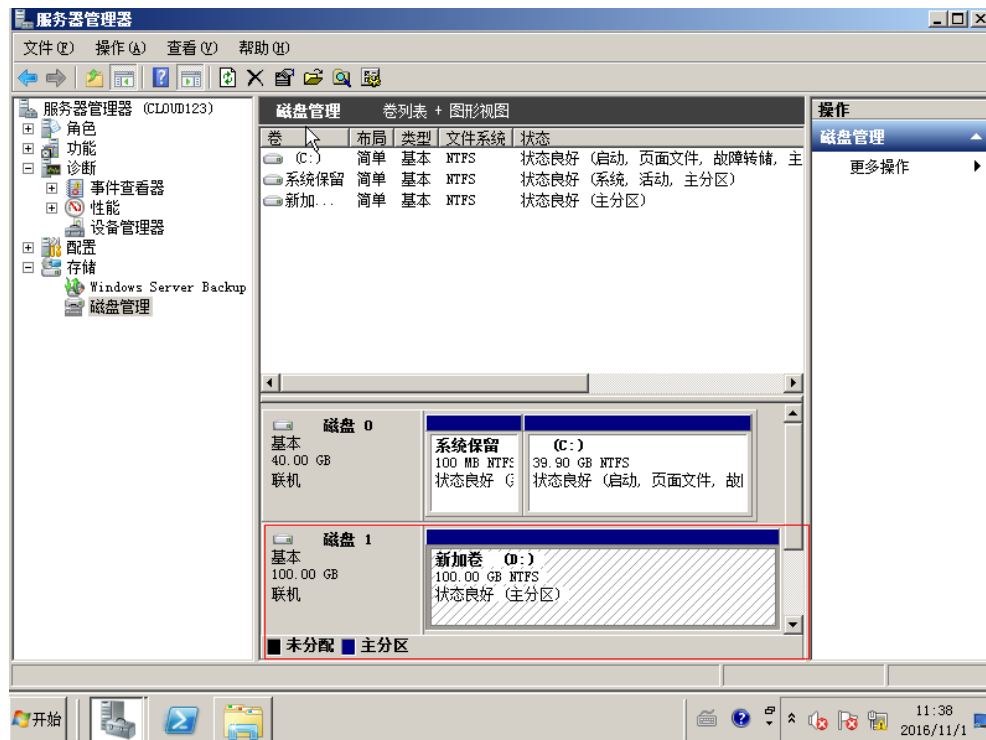


**须知**

不同文件系统支持的分区大小不同，请根据您的业务需求选择合适的文件系统。

**步骤11** 单击“完成”完成向导。需要等待片刻让系统完成初始化操作，当卷状态为“状态良好”时，表示初始化磁盘成功，如图2-11所示。

图 2-11 初始化磁盘成功



----结束

## 2.3.3 初始化 Windows 数据盘（Windows 2019）

### 操作场景

本文以云服务器的操作系统为“Windows Server 2019 Standard 64bit”为例，提供磁盘的初始化操作指导。

MBR格式分区支持的磁盘最大容量为2 TiB，GPT分区表最大支持的磁盘容量为18 EiB，因此当为容量大于2 TiB的磁盘分区时，请采用GPT分区方式。具体操作请参见[初始化容量大于2TiB的Windows数据盘（Windows 2008）](#)。关于磁盘分区形式的更多介绍，请参见[场景及磁盘分区形式介绍](#)。

不同云服务器的操作系统的格式化操作可能不同，本文仅供参考，具体操作步骤和差异请参考对应的云服务器操作系统的产品文档。

#### 须知

首次使用磁盘时，如果您未参考本章节对磁盘执行初始化操作，主要包括创建分区和文件系统等操作，那么当后续扩容磁盘时，新增容量部分的磁盘可能无法正常使用。

## 前提条件

- 已挂载数据盘至云服务器，且该数据盘未初始化。
- 已登录。
  - 弹性云服务器请参见《弹性云服务器用户指南》。
  - 裸金属服务器请参见《裸金属服务器用户指南》。

## 操作指导

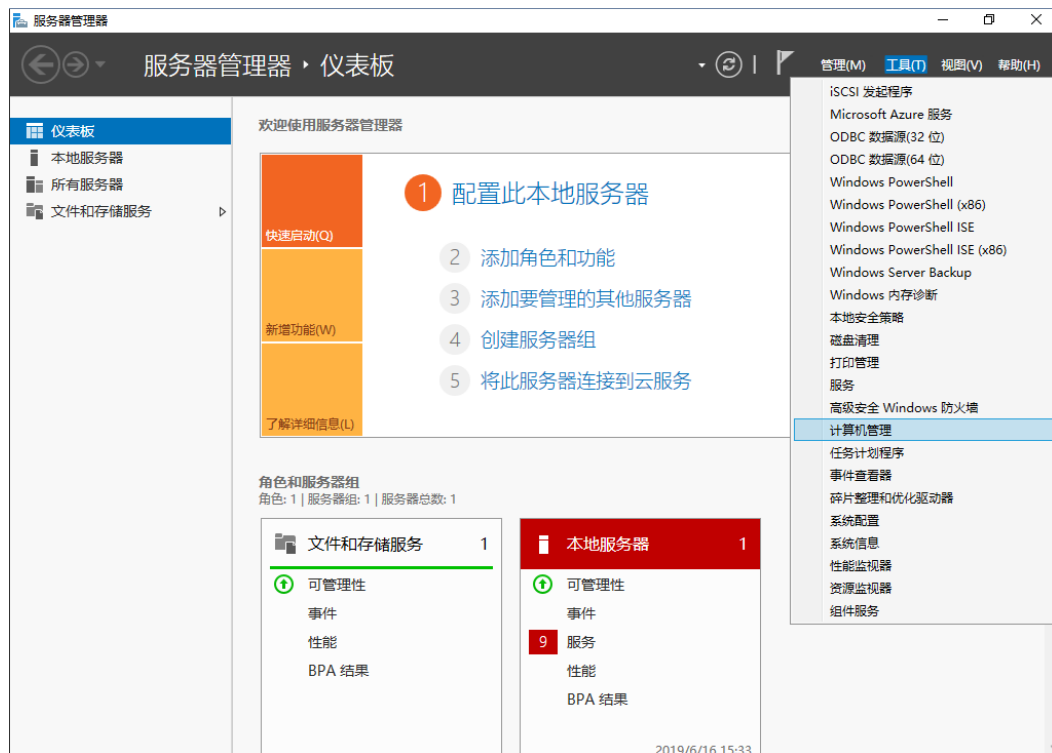
**步骤1** 在云服务器桌面，单击左下方开始图标。

弹出Windows Server窗口。

**步骤2** 单击“服务器管理器”。

弹出“服务器管理器”窗口，如图2-12所示。

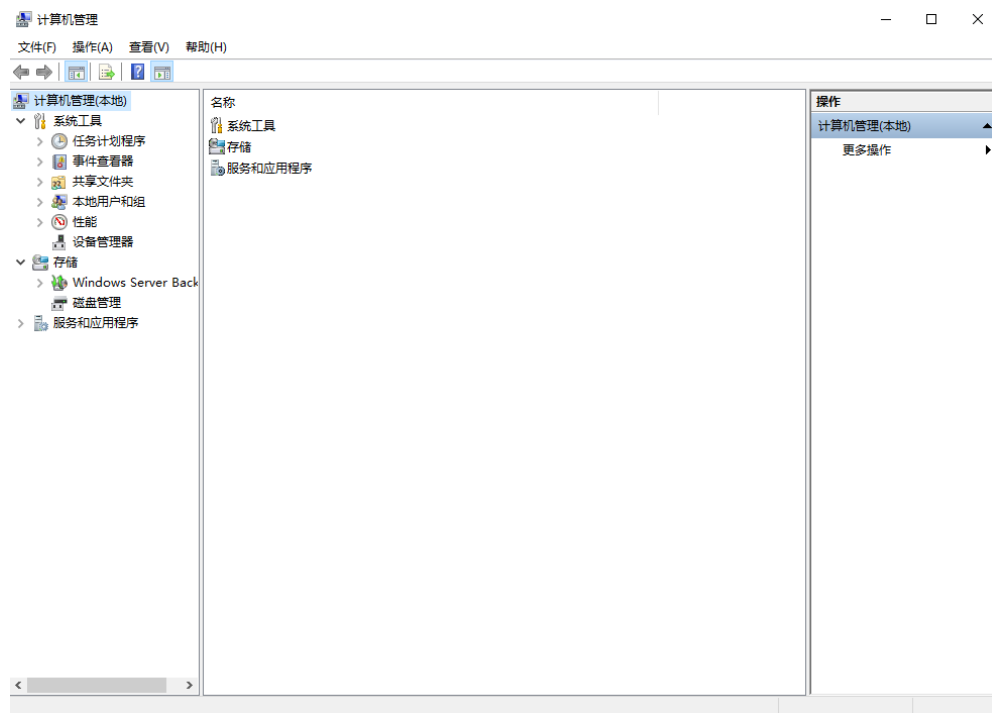
图 2-12 服务器管理器



**步骤3** “服务器管理器”页面右上方选择“工具 > 计算机管理”。

弹出“计算机管理”窗口，如图2-13所示。

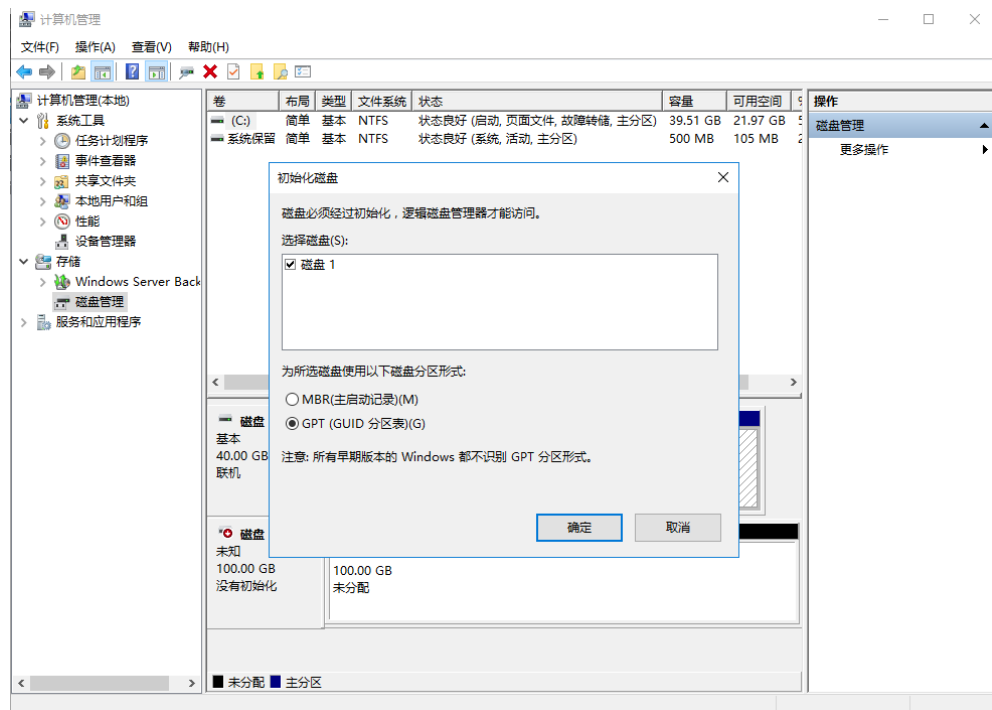
图 2-13 计算机管理



步骤4 选择“存储 > 磁盘管理”。

进入磁盘列表页面，存在未初始化的磁盘时，系统会自动弹出“初始化磁盘”对话框，如图2-14所示。

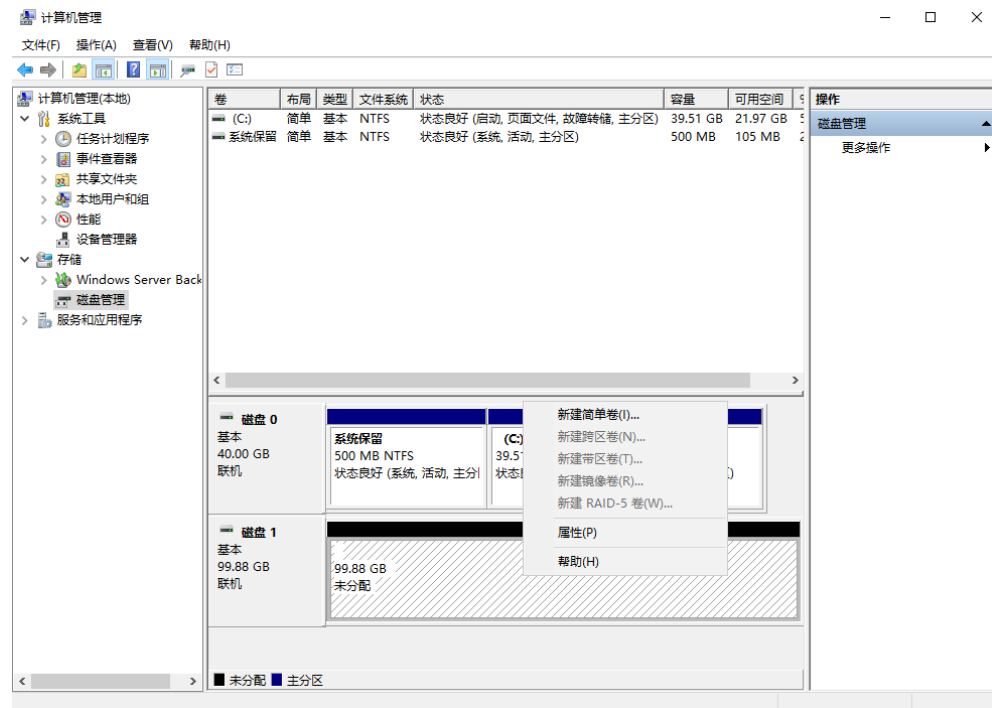
图 2-14 磁盘列表



步骤5 在“初始化磁盘”对话框中显示需要初始化的磁盘，此处以选择“GPT（GUID分区表）”为例，单击“确定”。

返回“计算机管理”窗口，如图2-15所示。

图 2-15 计算机管理



### 须知

MBR支持的云硬盘最大容量为2 TiB，GPT最大支持的云硬盘容量为18 EiB，当前数据盘支持的最大容量为32 TiB，如果您需要使用大于2 TiB的云硬盘容量，分区形式请采用GPT。

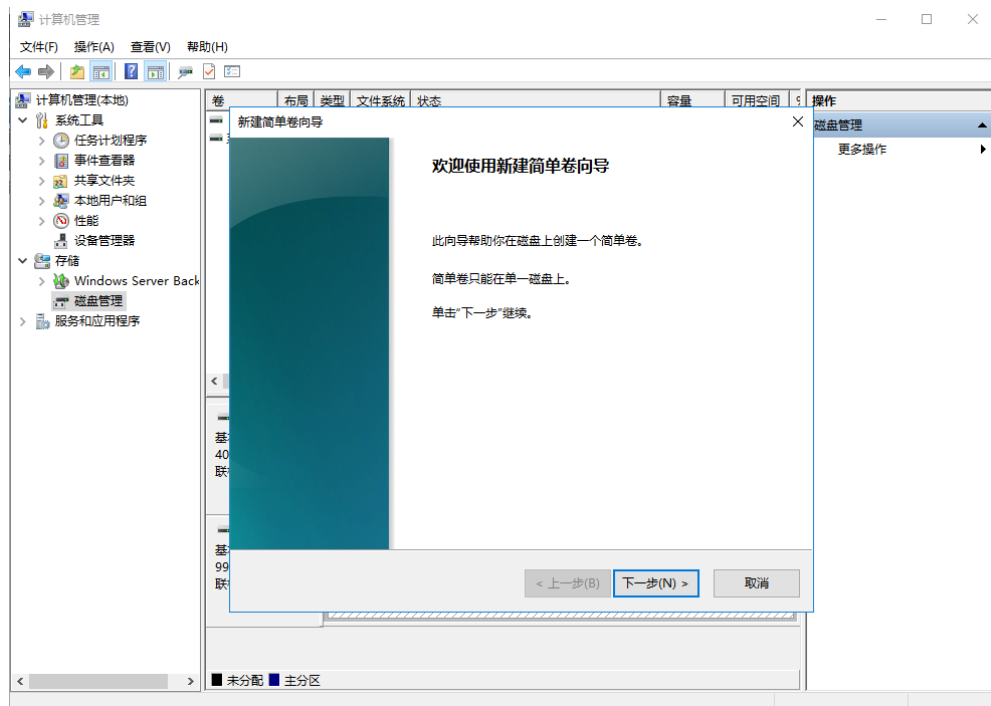
当云硬盘已经投入使用后，此时切换云硬盘分区形式时，云硬盘上的原有数据将会清除，因此请在云硬盘初始化时谨慎选择云硬盘分区形式。切换GPT分区形式前，请对云硬盘数据备份后，再格式化硬盘。

**步骤6** 在磁盘1右侧的未分配的区域，右键单击选择“新建简单卷”。

弹出“新建简单卷向导”窗口，如图2-16所示。

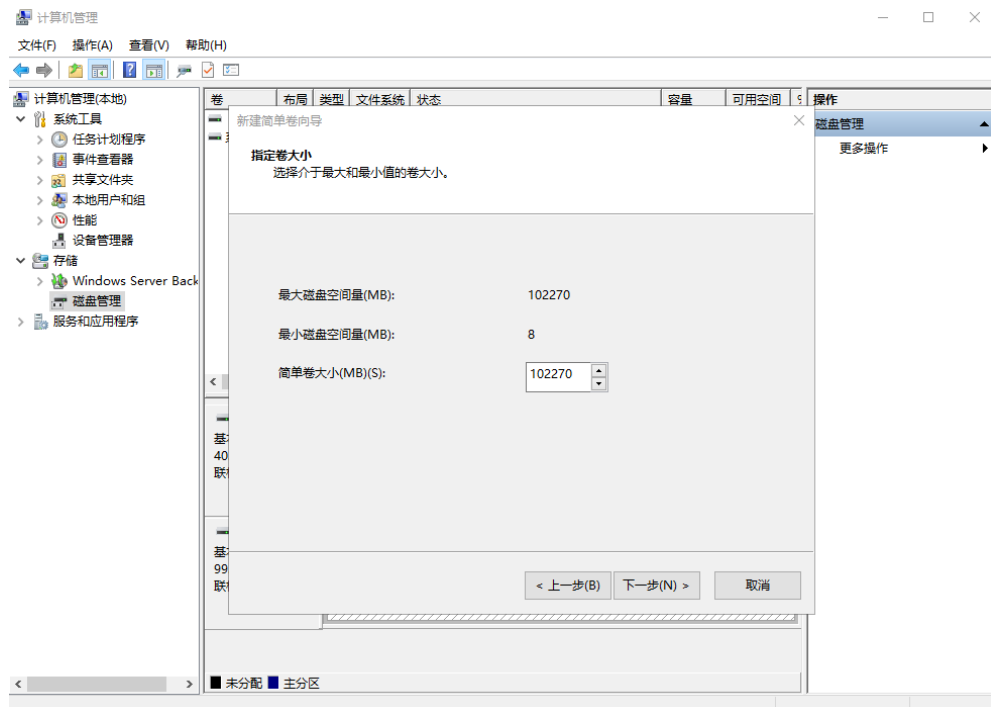


图 2-16 新建简单卷向导



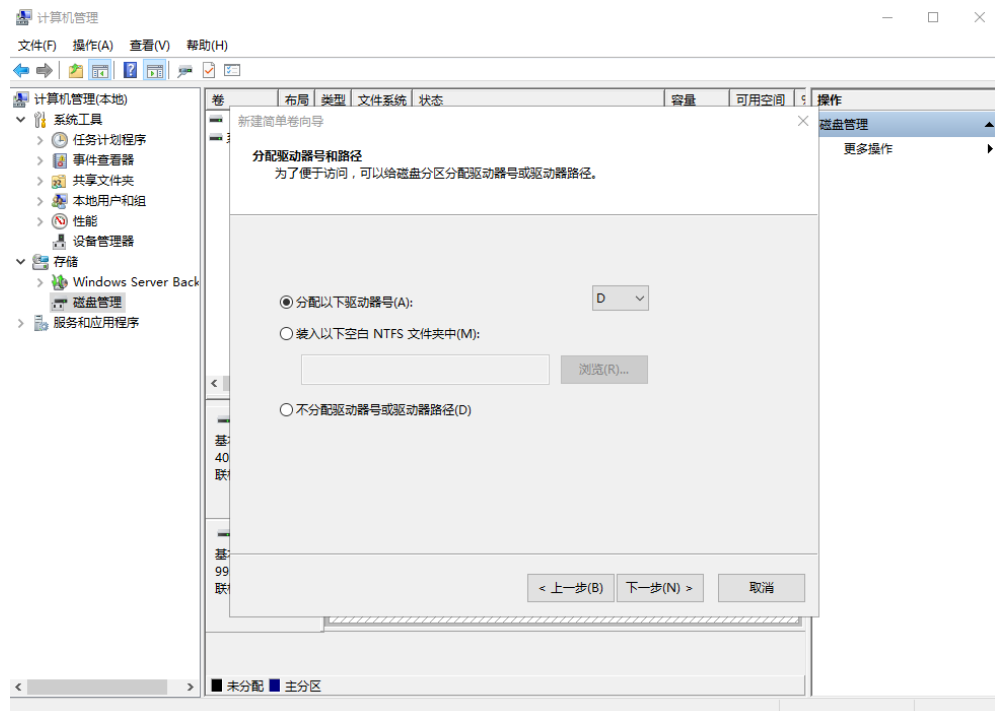
**步骤7** 根据界面提示，单击“下一步”。  
进入“指定卷大小”页面，如图2-17所示。

图 2-17 指定卷大小



**步骤8** 指定卷大小，系统默认卷大小为最大值，您还可以根据实际需求指定卷大小，此处以保持系统默认配置为例，单击“下一步”。  
进入“分配驱动器号和路径”页面，如图2-18所示。

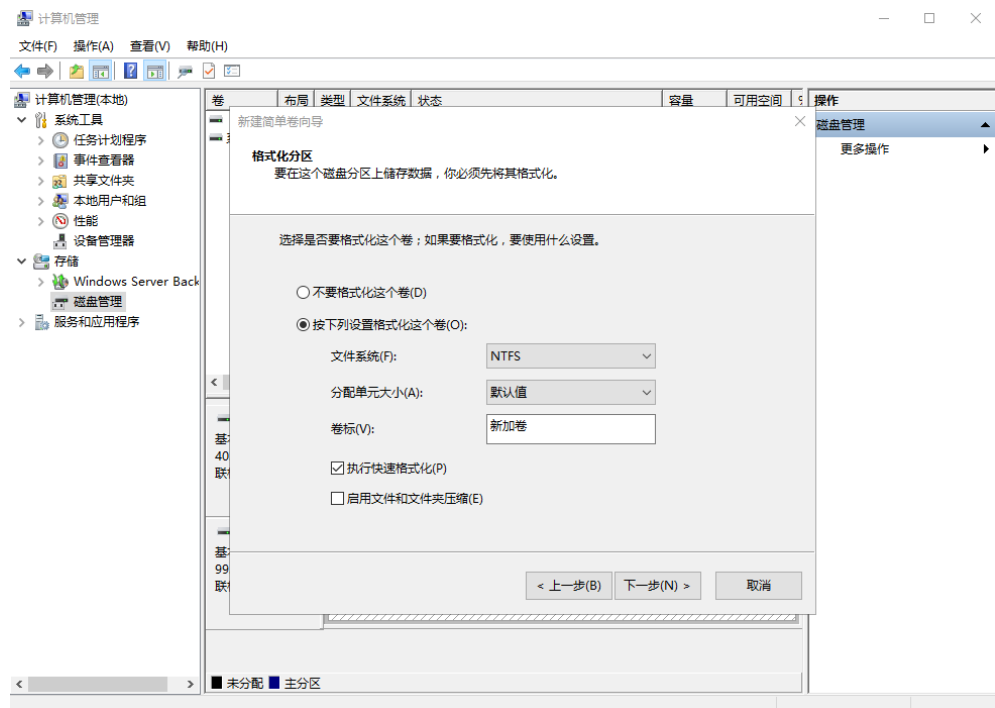
图 2-18 分配驱动器号和路径



**步骤9** 分配到驱动器号和路径，系统默认为磁盘分配驱动器号，驱动器号默认为“D”，此处以保持系统默认配置为例，单击“下一步”。

进入“格式化分区”页面，如图2-19所示。

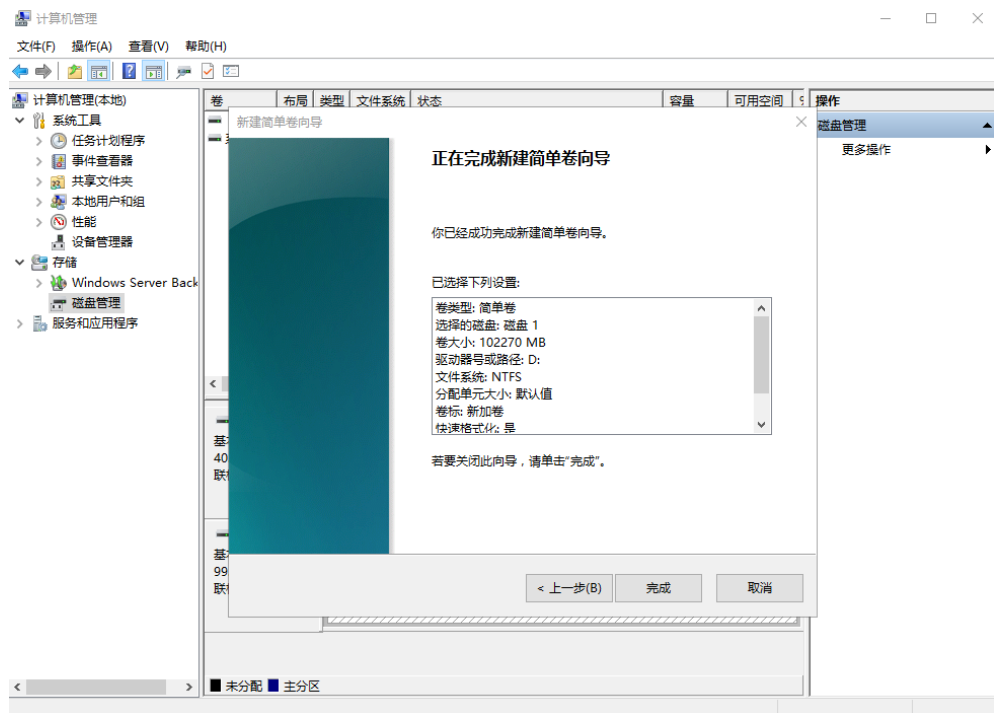
图 2-19 格式化分区



**步骤10** 格式化分区，系统默认的文件系统为NTFS，并根据实际情况设置其他参数，此处以保持系统默认设置为例，单击“下一步”。

进入“完成新建卷”页面，如图2-20所示。

图 2-20 完成新建卷



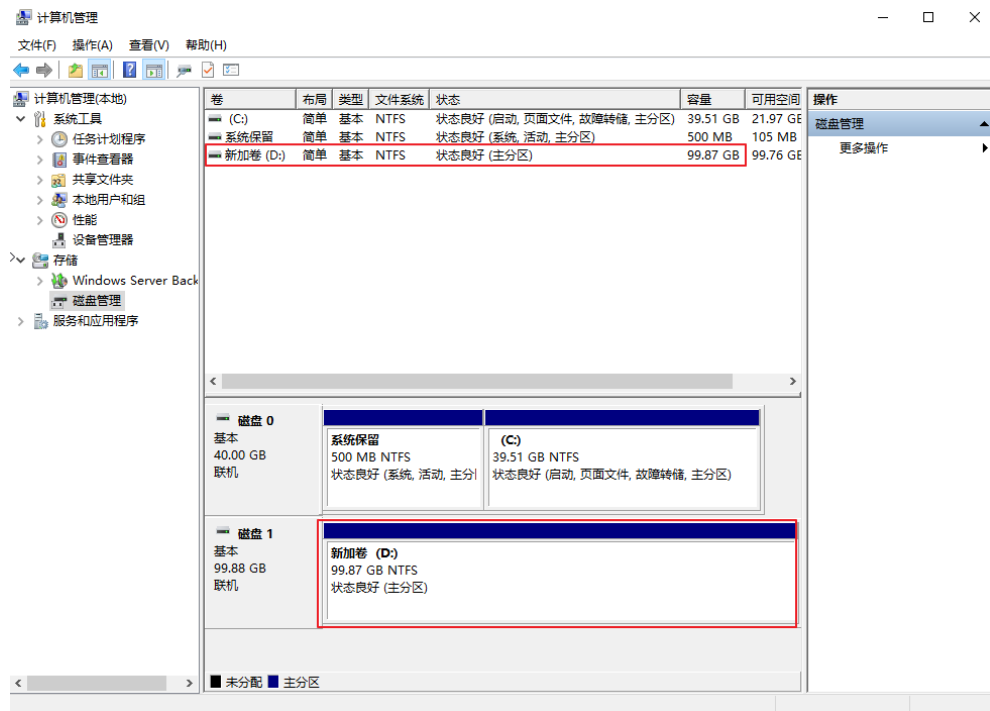
### 须知


不同文件系统支持的分区大小不同，请根据您的业务需求选择合适的文件系统。

**步骤11** 单击“完成”。

需要等待片刻让系统完成初始化操作，当卷状态为“状态良好”时，表示初始化磁盘成功，如图2-21所示。

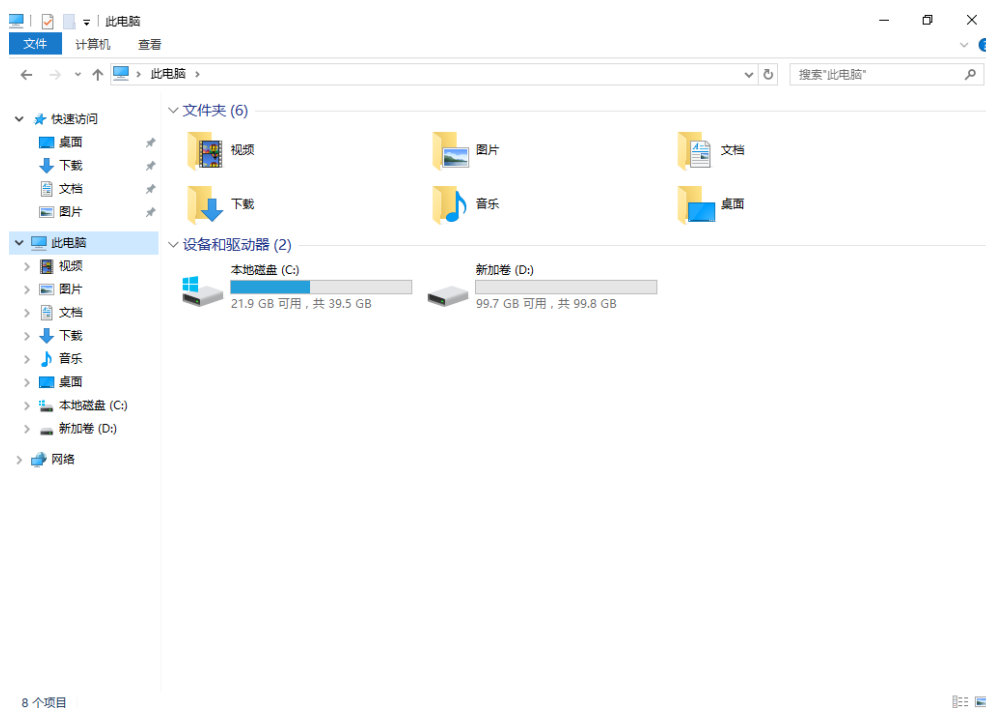
图 2-21 初始化磁盘成功



**步骤12** 新建卷完成后，单击下方任务栏中，在文件资源管理器中查看是否有新建卷，此处以“新建卷 (D:)”为例。

单击“此电脑”，如果如图2-22所示，可以看到“新建卷 (D:)”，表示磁盘初始化成功，任务结束。

图 2-22 文件资源管理器



----结束

## 2.3.4 初始化 Linux 数据盘（fdisk）

### 操作场景

本文以云服务器的操作系统为“CentOS 7.4 64位”为例，采用fdisk分区工具为数据盘设置分区。

MBR支持的磁盘最大容量为2 TiB，GPT最大支持的磁盘容量为18 EiB，因此当您初始化容量大于2 TiB的磁盘时，分区形式请采用GPT。

fdisk分区工具只适用于MBR分区，parted工具适用于MBR分区和GPT分区。关于磁盘分区形式的更多介绍，请参见[场景及磁盘分区形式介绍](#)。

不同云服务器的操作系统的格式化操作可能不同，本文仅供参考，具体操作步骤和差异请参考对应的云服务器操作系统的产品文档。

#### 须知

首次使用磁盘时，如果您未参考本章节对磁盘执行初始化操作，主要包括创建分区和文件系统等操作，那么当后续扩容磁盘时，新增容量部分的磁盘可能无法正常使用。

### 前提条件

- 已挂载数据盘至云服务器，且该数据盘未初始化。
- 已登录。
  - 弹性云服务器请参见《弹性云服务器用户指南》。

- 裸金属服务器请参见《裸金属服务器用户指南》。

## 划分分区并挂载磁盘

本操作以该场景为例，当云服务器挂载了一块新的数据盘时，使用fdisk分区工具将该数据盘设为主分区，分区形式默认设置为MBR，文件系统设为格式，挂载在“/mnt/sdc”下，并设置开机启动自动挂载。

**步骤1** 执行以下命令，查看新增磁盘情况。

**fdisk -l**

回显类似如下信息：

```
[root@ecs-test-0001 ~]# fdisk -l

Disk /dev/vda: 42.9 GiB, 42949672960 bytes, 83886080 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x000bcb4e

   Device Boot      Start         End      Blocks   Id  System
/dev/vda1 *         2048     83886079     41942016   83  Linux

Disk /dev/vdb: 107.4 GiB, 107374182400 bytes, 209715200 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
```

表示当前的云服务器有两块磁盘，“/dev/vda”是系统盘，“/dev/vdb”是新增数据盘。

**步骤2** 执行以下命令，进入fdisk分区工具，开始对新增数据盘执行分区操作。

**fdisk 新增数据盘**

以新挂载的数据盘“/dev/vdb”为例：

**fdisk /dev/vdb**

回显类似如下信息：

```
[root@ecs-test-0001 ~]# fdisk /dev/vdb
Welcome to fdisk (util-linux 2.23.2).

Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.

Device does not contain a recognized partition table
Building a new DOS disklabel with disk identifier 0x38717fc1.

Command (m for help):
```

**步骤3** 输入“n”，按“Enter”，开始新建分区。

回显类似如下信息：

```
Command (m for help): n
Partition type:
   p   primary (0 primary, 0 extended, 4 free)
   e   extended
```

表示磁盘有两种分区类型：

- “p”表示主分区。
- “e”表示扩展分区。

## 📖 说明

磁盘使用MBR分区形式，最多可以创建4个主分区，或者3个主分区加1个扩展分区，扩展分区不可以直接使用，需要划分成若干个逻辑分区才可以使用。

磁盘使用GPT分区形式时，没有主分区、扩展分区以及逻辑分区之分。

**步骤4** 以创建一个主要分区为例，输入“p”，按“Enter”，开始创建一个主分区。

回显类似如下信息：

```
Select (default p): p
Partition number (1-4, default 1):
```

“Partition number”表示主分区编号，可以选择1-4。

**步骤5** 以分区编号选择“1”为例，输入主分区编号“1”，按“Enter”。

回显类似如下信息：

```
Partition number (1-4, default 1): 1
First sector (2048-209715199, default 2048):
```

“First sector”表示起始磁柱值，可以选择2048-209715199，默认为2048。

**步骤6** 以选择默认起始磁柱值2048为例，按“Enter”。

系统会自动提示分区可用空间的起始磁柱值和截止磁柱值，可以在该区间内自定义，或者使用默认值。起始磁柱值必须小于分区的截止磁柱值。

回显类似如下信息：

```
First sector (2048-209715199, default 2048):
Using default value 2048
Last sector, +sectors or +size{K,M,G} (2048-209715199, default 209715199):
```

“Last sector”表示截止磁柱值，可以选择2048-209715199，默认为209715199。

**步骤7** 以选择默认截止磁柱值209715199为例，按“Enter”。

系统会自动提示分区可用空间的起始磁柱值和截止磁柱值，可以在该区间内自定义，或者使用默认值。起始磁柱值必须小于分区的截止磁柱值。

回显类似如下信息：

```
Last sector, +sectors or +size{K,M,G} (2048-209715199, default 209715199):
Using default value 209715199
Partition 1 of type Linux and of size 100 GiB is set
Command (m for help):
```

表示分区完成，即为数据盘新建了1个分区。

**步骤8** 输入“p”，按“Enter”，查看新建分区的详细信息。

回显类似如下信息：

```
Command (m for help): p

Disk /dev/vdb: 107.4 GiB, 107374182400 bytes, 209715200 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x38717fc1

   Device Boot      Start         End      Blocks   Id  System
/dev/vdb1            2048    209715199    104856576   83   Linux

Command (m for help):
```

表示新建分区“/dev/vdb1”的详细信息。

**步骤9** 输入“w”，按“Enter”，将分区结果写入分区表中。

回显类似如下信息：

```
Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
```

表示分区创建完成。

#### 说明

如果之前分区操作有误，请输入“q”，则会退出fdisk分区工具，之前的分区结果将不会被保留。

**步骤10** 执行以下命令，将新的分区表变更同步至操作系统。

**partprobe**

**步骤11** 执行以下命令，将新建分区文件系统设为系统所需格式。

**mkfs -t 文件系统格式 /dev/vdb1**

以设置文件系统为“ext4”为例：

**mkfs -t ext4 /dev/vdb1**

回显类似如下信息：

```
[root@ecs-test-0001 ~]# mkfs -t ext4 /dev/vdb1
mke2fs 1.42.9 (28-Dec-2013)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
6553600 inodes, 26214144 blocks
1310707 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=2174746624
800 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
    32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
    4096000, 7962624, 11239424, 20480000, 23887872

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

格式化需要等待一段时间，请观察系统运行状态，不要退出。

---

#### 须知

不同文件系统支持的分区大小不同，请根据您的业务需求选择合适的文件系统。

**步骤12** 执行以下命令，新建挂载目录。

**mkdir 挂载目录**



以新建挂载目录“/mnt/sdc”为例：

```
mkdir /mnt/sdc
```

#### 📖 说明

Linux系统默认带有/mnt目录，如果创建失败，可能是/mnt被误删除了，可以执行命令**mkdir -p /mnt/sdc**创建。

**步骤13** 执行以下命令，将新建分区挂载到**步骤12**中创建的目录下。

```
mount 磁盘分区 挂载目录
```

以挂载新建分区“/dev/vdb1”至“/mnt/sdc”为例：

```
mount /dev/vdb1 /mnt/sdc
```

**步骤14** 执行以下命令，查看挂载结果。

```
df -TH
```

回显类似如下信息：

```
[root@ecs-test-0001 ~]# df -TH
Filesystem Type Size Used Avail Use% Mounted on
/dev/vda1 ext4 43G 1.9G 39G 5% /
devtmpfs devtmpfs 2.0G 0 2.0G 0% /dev
tmpfs tmpfs 2.0G 0 2.0G 0% /dev/shm
tmpfs tmpfs 2.0G 9.1M 2.0G 1% /run
tmpfs tmpfs 2.0G 0 2.0G 0% /sys/fs/cgroup
tmpfs tmpfs 398M 0 398M 0% /run/user/0
/dev/vdb1 ext4 106G 63M 101G 1% /mnt/sdc
```

表示新建分区“/dev/vdb1”已挂载至“/mnt/sdc”。

#### 📖 说明

云服务器重启后，挂载会失效。您可以修改“/etc/fstab”文件，将新建磁盘分区设置为开机自动挂载，请参见[设置开机自动挂载磁盘分区](#)。

----结束

## 设置开机自动挂载磁盘分区

您可以通过配置fstab文件，设置云服务器系统启动时自动挂载磁盘分区。已有数据的云服务器也可以进行设置，该操作不会影响现有数据。

本文介绍如何在fstab文件中使用UUID来设置自动挂载磁盘分区。不建议采用在“/etc/fstab”直接指定设备名（比如/dev/vdb1）的方法，因为云中设备的顺序编码在关闭或者开启云服务器过程中可能发生改变，例如/dev/vdb1可能会变成/dev/vdb2，可能会导致云服务器重启后不能正常运行。

#### 📖 说明

UUID（universally unique identifier）是Linux系统为磁盘分区提供的唯一的标识字符串。

**步骤1** 执行如下命令，查询磁盘分区的UUID。

```
blkid 磁盘分区
```

以查询磁盘分区“/dev/vdb1”的UUID为例：

```
blkid /dev/vdb1
```

回显类似如下信息：

```
[root@ecs-test-0001 ~]# blkid /dev/vdb1  
/dev/vdb1: UUID="0b3040e2-1367-4abb-841d-ddb0b92693df" TYPE="ext4"
```

记录下回显中磁盘分区“/dev/vdb1”的UUID，方便后续步骤使用。

**步骤2** 执行以下命令，使用VI编辑器打开“fstab”文件。

```
vi /etc/fstab
```

**步骤3** 按“i”，进入编辑模式。

**步骤4** 将光标移至文件末尾，按“Enter”，添加如下内容。

```
UUID=0b3040e2-1367-4abb-841d-ddb0b92693df /mnt/sdc ext4 defaults 0 2
```

以上内容仅为示例，具体请以实际情况为准，参数说明如下：

- 第一列为UUID，此处填写**步骤1**中查询到的磁盘分区的UUID。
- 第二列为磁盘分区的挂载目录，可以通过**df -TH**命令查询。
- 第三列为磁盘分区的文件系统格式，可以通过**df -TH**命令查询。
- 第四列为磁盘分区的挂载选项，此处通常设置为defaults即可。
- 第五列为Linux dump备份选项。
  - 0表示不使用Linux dump备份。现在通常不使用dump备份，此处设置为0即可。
  - 1表示使用Linux dump备份。
- 第六列为fsck选项，即开机时是否使用fsck检查磁盘。
  - 0表示不检验。
  - 挂载点为（/）根目录的分区，此处必须填写1。  
根分区设置为1，其他分区只能从2开始，系统会按照数字从小到大依次检查下去。

**步骤5** 按“ESC”后，输入“:wq”，按“Enter”。

保存设置并退出编辑器。

**步骤6** 执行以下步骤，验证自动挂载功能。

1. 执行如下命令，卸载已挂载的分区。

```
umount 磁盘分区
```

命令示例：

```
umount /dev/vdb1
```

2. 执行如下命令，将“/etc/fstab”文件所有内容重新加载。

```
mount -a
```

3. 执行如下命令，查询文件系统挂载信息。

```
mount | grep 挂载目录
```

命令示例：

```
mount | grep /mnt/sdc
```

回显类似如下信息，说明自动挂载功能生效：

```
root@ecs-test-0001 ~]# mount | grep /mnt/sdc  
/dev/vdb1 on /mnt/sdc type ext4 (rw,relatime,data=ordered)
```

----结束

## 2.3.5 初始化 Linux 数据盘（ parted ）

### 操作场景

本文以云服务器的操作系统为“CentOS 7.4 64位”为例，采用Parted分区工具为数据盘设置分区。

MBR支持的磁盘最大容量为2 TiB，GPT最大支持的磁盘容量为18 EiB，因此当您初始化容量大于2 TiB的磁盘时，分区形式请采用GPT。

fdisk分区工具只适用于MBR分区，parted工具适用于MBR分区和GPT分区。关于磁盘分区形式的更多介绍，请参见[场景及磁盘分区形式介绍](#)。

不同云服务器的操作系统的格式化操作可能不同，本文仅供参考，具体操作步骤和差异请参考对应的云服务器操作系统的产品文档。

#### 须知

首次使用磁盘时，如果您未参考本章节对磁盘执行初始化操作，主要包括创建分区和文件系统等操作，那么当后续扩容磁盘时，新增容量部分的磁盘可能无法正常使用。

### 前提条件

- 已挂载数据盘至云服务器，且该数据盘未初始化。
- 已登录。
  - 弹性云服务器请参见《弹性云服务器用户指南》。
  - 裸金属服务器请参见《裸金属服务器用户指南》。

### 划分分区并挂载磁盘

本操作以该场景为例，当云服务器挂载了一块新的数据盘时，采用parted分区工具为数据盘设置分区，分区形式设置为GPT，文件系统设为ext4格式，挂载在“/mnt/sdc”下，并设置开机启动自动挂载。

**步骤1** 执行以下命令，查看新增数据盘。

#### lsblk

回显类似如下信息：

```
root@ecs-test-0001 ~]# lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
vda 253:0 0 40G 0 disk
└─vda1 253:1 0 40G 0 part /
vdb 253:16 0 100G 0 disk
```

表示当前的云服务器有两块磁盘，“/dev/vda”是系统盘，“/dev/vdb”是新增数据盘。

**步骤2** 执行以下命令，进入parted分区工具，开始对新增数据盘执行分区操作。

**parted** *新增数据盘*

命令示例：

**parted /dev/vdb**

回显类似如下信息：

```
[root@ecs-test-0001 ~]# parted /dev/vdb
GNU Parted 3.1
Using /dev/vdb
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted)
```

**步骤3** 输入“p”，按“Enter”，查看当前磁盘分区形式。

回显类似如下信息：

```
(parted) p
Error: /dev/vdb: unrecognised disk label
Model: Virtio Block Device (virtblk)
Disk /dev/vdb: 107GiB
Sector size (logical/physical): 512B/512B
Partition Table: unknown
Disk Flags:
(parted)
```

“Partition Table”为“unknown”表示磁盘分区形式未知，新的数据盘还未设置分区形式。

**步骤4** 输入以下命令，设置磁盘分区形式。

**mklabel** *磁盘分区形式*

磁盘分区形式有MBR和GPT两种，以GPT为例：

**mklabel gpt**

#### 须知

MBR支持的云硬盘最大容量为2 TiB，GPT最大支持的云硬盘容量为18 EiB，当前数据盘支持的最大容量为32 TiB，如果您需要使用大于2 TiB的云硬盘容量，分区形式请采用GPT。

当云硬盘已经投入使用后，此时切换云硬盘分区形式时，云硬盘上的原有数据将会清除，因此请在云硬盘初始化时谨慎选择云硬盘分区形式。切换GPT分区形式前，请对云硬盘数据备份后，再格式化硬盘。

**步骤5** 输入“p”，按“Enter”，设置分区形式后，再次查看磁盘分区形式。

回显类似如下信息：

```
(parted) mklabel gpt
(parted) p
Model: Virtio Block Device (virtblk)
Disk /dev/vdb: 107GiB
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number Start End Size File system Name Flags
(parted)
```

“Partition Table”为“gpt”表示磁盘分区形式已设置为GPT。

**步骤6** 输入“unit s”，按“Enter”，设置磁盘的计量单位为磁柱。

**步骤7** 以整个磁盘创建一个分区为例，执行以下命令，按“Enter”。

**mkpart** *磁盘分区名称 起始磁柱值 截止磁柱值*

命令示例：

```
mkpart test 2048s 100%
```

“2048s”表示磁盘起始磁柱值，“100%”表示磁盘截止磁柱值，此处仅供参考，您可以根据业务需要自行规划磁盘分区数量及容量。

回显类似如下信息：

```
(parted) mkpart opt 2048s 100%  
(parted)
```

**步骤8** 输入“p”，按“Enter”，查看新建分区的详细信息。

回显类似如下信息：

```
(parted) p  
Model: Virtio Block Device (virtblk)  
Disk /dev/vdb: 209715200s  
Sector size (logical/physical): 512B/512B  
Partition Table: gpt  
Disk Flags:  
  
Number Start End Size File system Name Flags  
1 2048s 209713151s 209711104s test  
  
(parted)
```

**步骤9** 输入“q”，按“Enter”，退出parted分区工具。

回显类似如下信息：

```
(parted) q  
Information: You may need to update /etc/fstab.
```

“/etc/fstab”文件控制磁盘开机自动挂载，请先参考以下步骤为磁盘分区设置文件系统和挂载目录后，再根据文档指导更新“/etc/fstab”文件。

**步骤10** 执行以下命令，查看磁盘分区信息。

```
lsblk
```

回显类似如下信息：

```
[root@ecs-test-0001 ~]# lsblk  
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT  
vda 253:0 0 40G 0 disk  
└─vda1 253:1 0 40G 0 part /  
vdb 253:16 0 100G 0 disk  
└─vdb1 253:17 0 100G 0 part
```

此时可以查看到新建分区“/dev/vdb1”

**步骤11** 执行以下命令，将新建分区文件系统设为系统所需格式。

```
mkfs -t 文件系统格式 /dev/vdb1
```

以设置文件系统为“ext4”为例：

```
mkfs -t ext4 /dev/vdb1
```

回显类似如下信息：

```
[root@ecs-test-0001 ~]# mkfs -t ext4 /dev/vdb1  
mke2fs 1.42.9 (28-Dec-2013)  
Filesystem label=  
OS type: Linux  
Block size=4096 (log=2)  
Fragment size=4096 (log=2)  
Stride=0 blocks, Stripe width=0 blocks  
6553600 inodes, 26213888 blocks  
1310694 blocks (5.00%) reserved for the super user
```

```
First data block=0
Maximum filesystem blocks=2174746624
800 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
    32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
    4096000, 7962624, 11239424, 20480000, 23887872

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

格式化需要等待一段时间，请观察系统运行状态，不要退出。

### 须知

不同文件系统支持的分区大小不同，请根据您的业务需求选择合适的文件系统。

**步骤12** 执行以下命令，新建挂载目录。

```
mkdir 挂载目录
```

以新建挂载目录“/mnt/sdc”为例：

```
mkdir /mnt/sdc
```

#### 📖 说明

Linux系统默认带有/mnt目录，如果创建失败，可能是/mnt被误删除了，可以执行命令**mkdir -p /mnt/sdc**创建。

**步骤13** 执行以下命令，将新建分区挂载到**步骤12**中创建的目录下。

```
mount 磁盘分区 挂载目录
```

以挂载新建分区“/dev/vdb1”至“/mnt/sdc”为例：

```
mount /dev/vdb1 /mnt/sdc
```

**步骤14** 执行以下命令，查看挂载结果。

```
df -TH
```

回显类似如下信息：

```
[root@ecs-test-0001 ~]# df -TH
Filesystem Type Size Used Avail Use% Mounted on
/dev/vda1 ext4 43G 1.9G 39G 5% /
devtmpfs devtmpfs 2.0G 0 2.0G 0% /dev
tmpfs tmpfs 2.0G 0 2.0G 0% /dev/shm
tmpfs tmpfs 2.0G 9.0M 2.0G 1% /run
tmpfs tmpfs 2.0G 0 2.0G 0% /sys/fs/cgroup
tmpfs tmpfs 398M 0 398M 0% /run/user/0
/dev/vdb1 ext4 106G 63M 101G 1% /mnt/sdc
```

表示新建分区“/dev/vdb1”已挂载至“/mnt/sdc”。

### 📖 说明

云服务器重启后，挂载会失效。您可以修改“/etc/fstab”文件，将新建磁盘分区设置为开机自动挂载，请参见[设置开机自动挂载磁盘分区](#)。

----结束

## 设置开机自动挂载磁盘分区

您可以通过配置fstab文件，设置系统启动时自动挂载磁盘分区。已有数据的也可以进行设置，该操作不会影响现有数据。

本文介绍如何在fstab文件中使用UUID来设置自动挂载磁盘分区。不建议采用在“/etc/fstab”直接指定设备名（比如/dev/vdb1）的方法，因为云中设备的顺序编码在关闭或者开启过程中可能发生改变，例如/dev/vdb1可能会变成/dev/vdb2，可能会导致重启后不能正常运行。

### 📖 说明

UUID（universally unique identifier）是Linux系统为磁盘分区提供的唯一的标识字符串。

**步骤1** 执行如下命令，查询磁盘分区的UUID。

**blkid 磁盘分区**

以查询磁盘分区“/dev/vdb1”的UUID为例：

**blkid /dev/vdb1**

回显类似如下信息：

```
[root@ecs-test-0001 ~]# blkid /dev/vdb1
/dev/vdb1: UUID="0b3040e2-1367-4abb-841d-ddb0b92693df" TYPE="ext4"
```

记录下回显中磁盘分区“/dev/vdb1”的UUID，方便后续步骤使用。

**步骤2** 执行以下命令，使用VI编辑器打开“fstab”文件。

**vi /etc/fstab**

**步骤3** 按“i”，进入编辑模式。

**步骤4** 将光标移至文件末尾，按“Enter”，添加如下内容。

```
UUID=0b3040e2-1367-4abb-841d-ddb0b92693df /mnt/sdc ext4 defaults 0 2
```

以上内容仅为示例，具体请以实际情况为准，参数说明如下：

- 第一列为UUID，此处填写**步骤1**中查询到的磁盘分区的UUID。
- 第二列为磁盘分区的挂载目录，可以通过**df -TH**命令查询。
- 第三列为磁盘分区的文件系统格式，可以通过**df -TH**命令查询。
- 第四列为磁盘分区的挂载选项，此处通常设置为defaults即可。
- 第五列为Linux dump备份选项。
  - 0表示不使用Linux dump备份。现在通常不使用dump备份，此处设置为0即可。
  - 1表示使用Linux dump备份。
- 第六列为fsck选项，即开机时是否使用fsck检查磁盘。
  - 0表示不检验。

- 挂载点为 (/) 根目录的分区，此处必须填写1。  
根分区设置为1，其他分区只能从2开始，系统会按照数字从小到大依次检查下去。

**步骤5** 按“ESC”后，输入“:wq”，按“Enter”。

保存设置并退出编辑器。

**步骤6** 执行以下步骤，验证自动挂载功能。

1. 执行如下命令，卸载已挂载的分区。

```
umount 磁盘分区
```

命令示例：

```
umount /dev/vdb1
```

2. 执行如下命令，将“/etc/fstab”文件所有内容重新加载。

```
mount -a
```

3. 执行如下命令，查询文件系统挂载信息。

```
mount | grep 挂载目录
```

命令示例：

```
mount | grep /mnt/sdc
```

回显类似如下信息，说明自动挂载功能生效：

```
root@ecs-test-0001 ~]# mount | grep /mnt/sdc  
/dev/vdb1 on /mnt/sdc type ext4 (rw,relatime,data=ordered)
```

----结束

## 2.3.6 初始化容量大于 2TiB 的 Windows 数据盘（Windows 2008）

### 操作场景

本文以云服务器的操作系统为“Windows Server 2008 R2 Standard 64bit”、磁盘容量为3 TiB举例，提供容量大于2 TiB的Windows数据盘的初始化操作指导。

MBR格式分区支持的磁盘最大容量为2 TiB，GPT分区表最大支持的磁盘容量为18 EiB，因此当为容量大于2 TiB的磁盘分区时，请采用GPT分区方式。具体操作请参见[初始化容量大于2TiB的Windows数据盘（Windows 2008）](#)。关于磁盘分区形式的更多介绍，请参见[场景及磁盘分区形式介绍](#)。

不同云服务器的操作系统的格式化操作可能不同，本文仅供参考，具体操作步骤和差异请参考对应的云服务器操作系统的产品文档。

#### 须知

首次使用磁盘时，如果您未参考本章节对磁盘执行初始化操作，主要包括创建分区和文件系统等操作，那么当后续扩容磁盘时，新增容量部分的磁盘可能无法正常使用。

### 前提条件

- 已挂载数据盘至云服务器，且该数据盘未初始化。



- 已登录。
  - 弹性云服务器请参见《弹性云服务器用户指南》。
  - 裸金属服务器请参见《裸金属服务器用户指南》。

## 操作指导

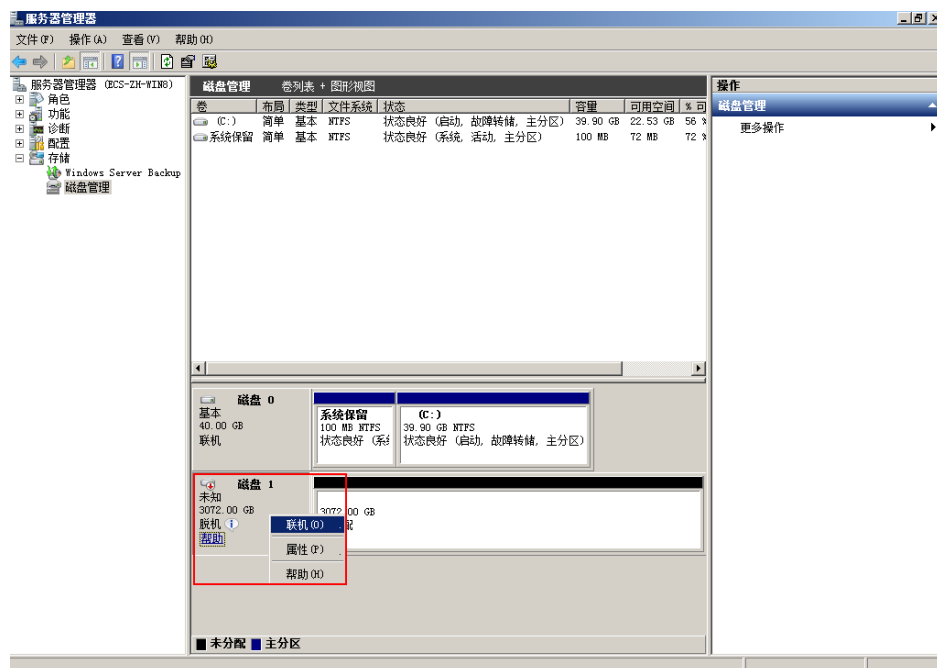
**步骤1** 在云服务器桌面，单击“开始”。

弹出开始窗口。

**步骤2** 在“计算机”栏目，右键单击菜单列表中的“管理”。

弹出“服务器管理器”窗口，如图2-23所示。

图 2-23 服务器管理器(Windows 2008)

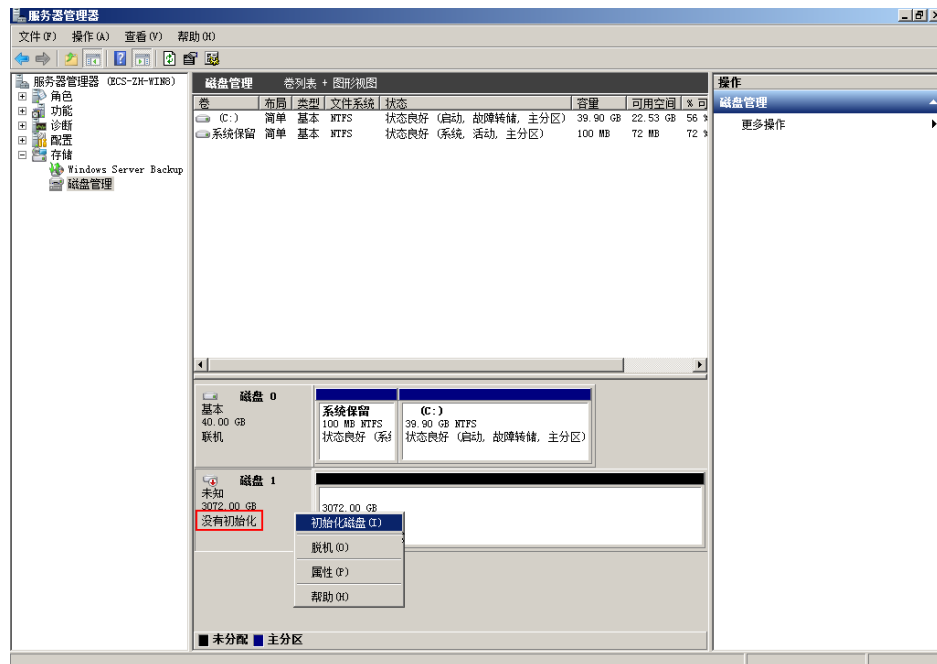


**步骤3** 在页面右侧可以查看磁盘列表，如果新增磁盘处于脱机状态，需要先进行联机，再进行初始化。

在磁盘1区域，右键单击菜单列表中的“联机”。

如图2-24所示，当磁盘1由“脱机”状态变为“没有初始化”，表示联机成功。

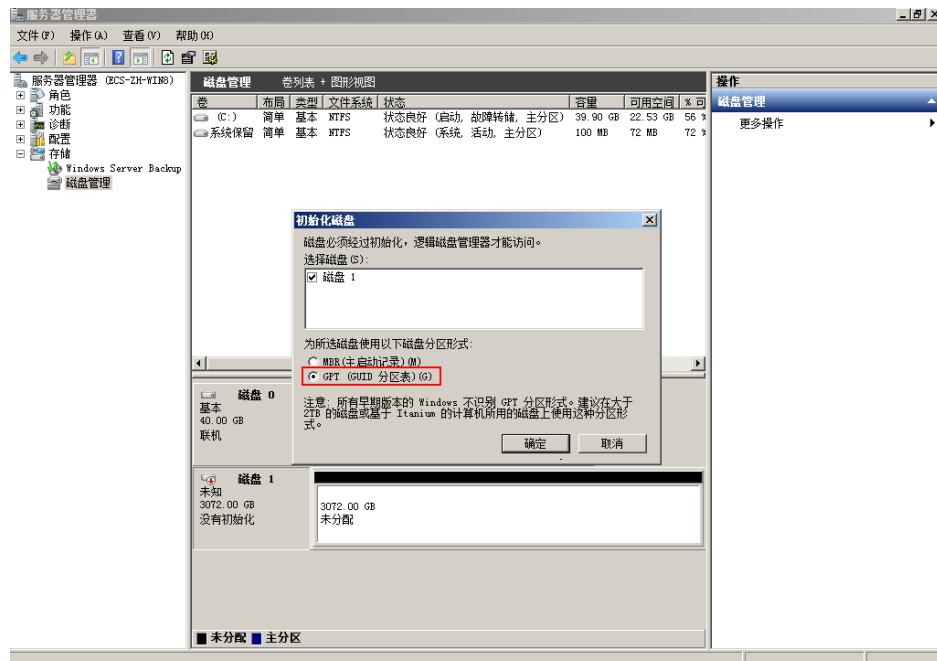
图 2-24 联机成功(Windows 2008)



**步骤4** 在磁盘1区域，右键单击菜单列表中的“初始化磁盘”。

弹出“初始化磁盘”窗口，如图2-25所示。

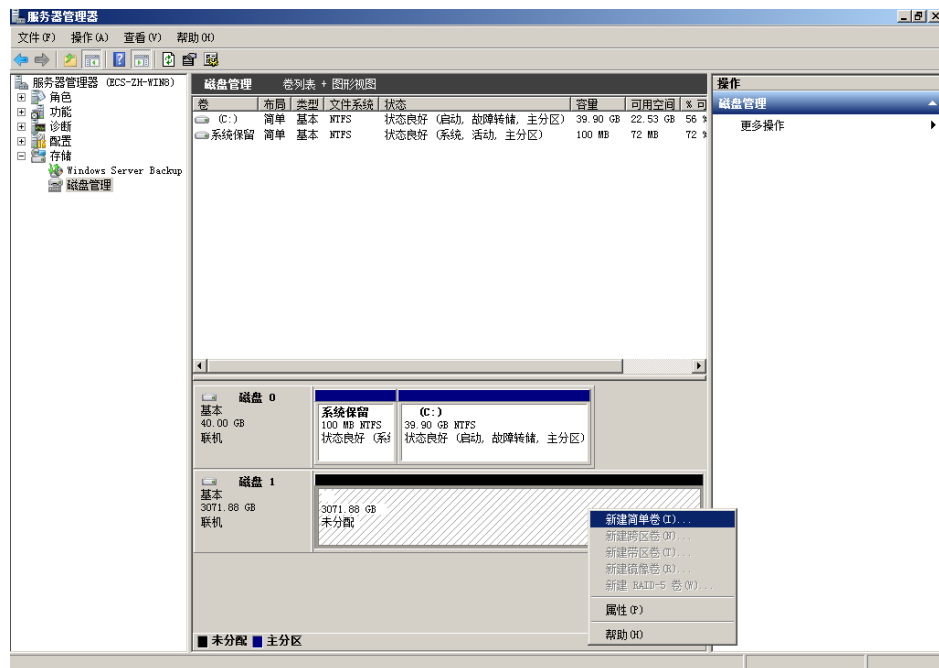
图 2-25 初始化磁盘(Windows 2008)



**步骤5** 在“初始化磁盘”对话框中显示需要初始化的磁盘，对于大于2 TiB的磁盘，此处请选择“GPT（GUID分区表）”，单击“确定”。

返回“服务器管理器”窗口，如图2-26所示。

图 2-26 服务器管理器窗口(Windows 2008)



### 须知

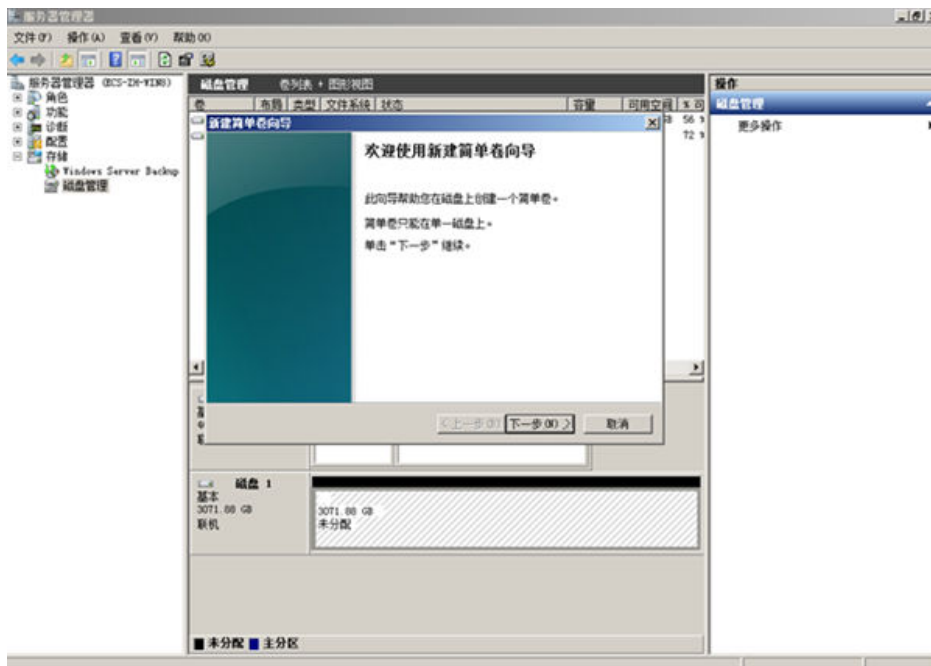
MBR支持的云硬盘最大容量为2 TiB，GPT最大支持的云硬盘容量为18 EiB，当前数据盘支持的最大容量为32 TiB，如果您需要使用大于2 TiB的云硬盘容量，分区形式请采用GPT。

当云硬盘已经投入使用后，此时切换云硬盘分区形式时，云硬盘上的原有数据将会清除，因此请在云硬盘初始化时谨慎选择云硬盘分区形式。切换GPT分区形式前，请对云硬盘数据备份后，再格式化硬盘。

**步骤6** 在磁盘1右侧的未分配的区域，右键单击选择“新建简单卷”。

弹出“新建简单卷向导”窗口，如图2-27所示。

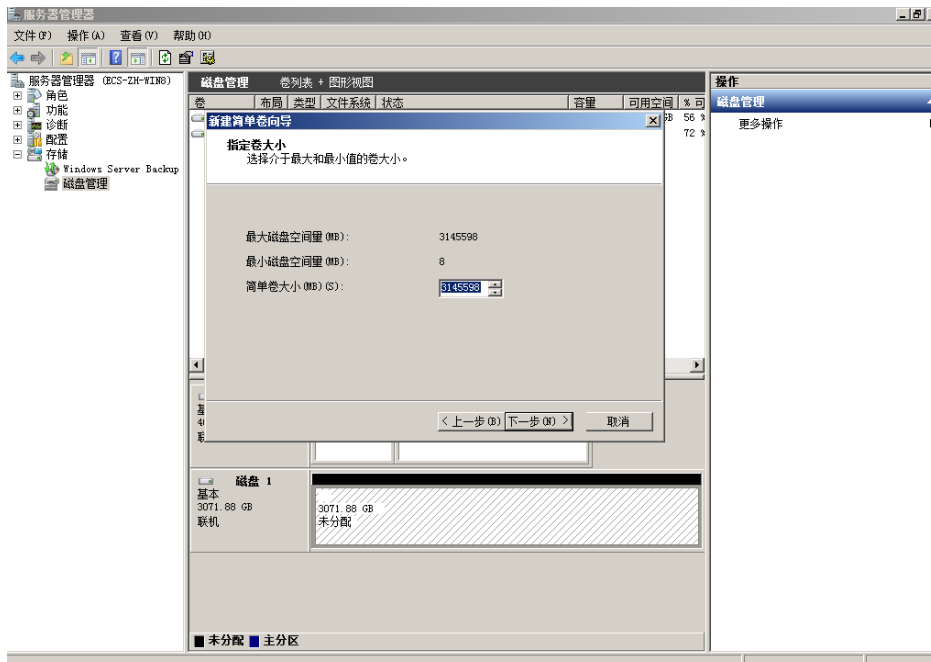
图 2-27 新建简单卷向导(Windows 2008)



**步骤7** 根据界面提示，单击“下一步”。

进入“指定卷大小”页面，如图2-28所示。

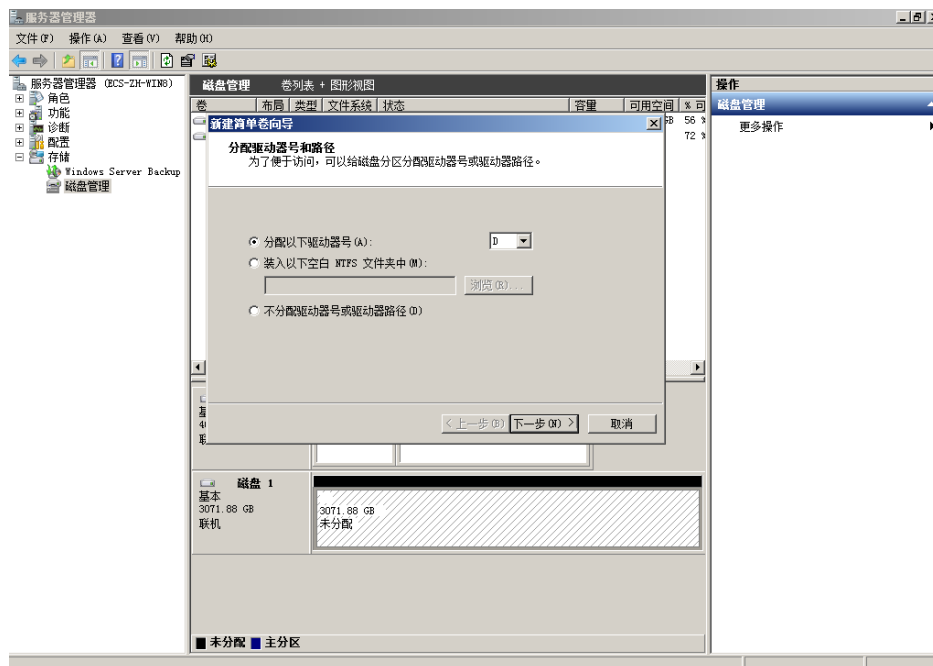
图 2-28 指定卷大小(Windows 2008)



**步骤8** 指定卷大小，系统默认卷大小为最大值，您还可以根据实际需求指定卷大小，此处以保持系统默认配置为例，单击“下一步”。

进入“分配驱动器号和路径”页面，如图2-29所示。

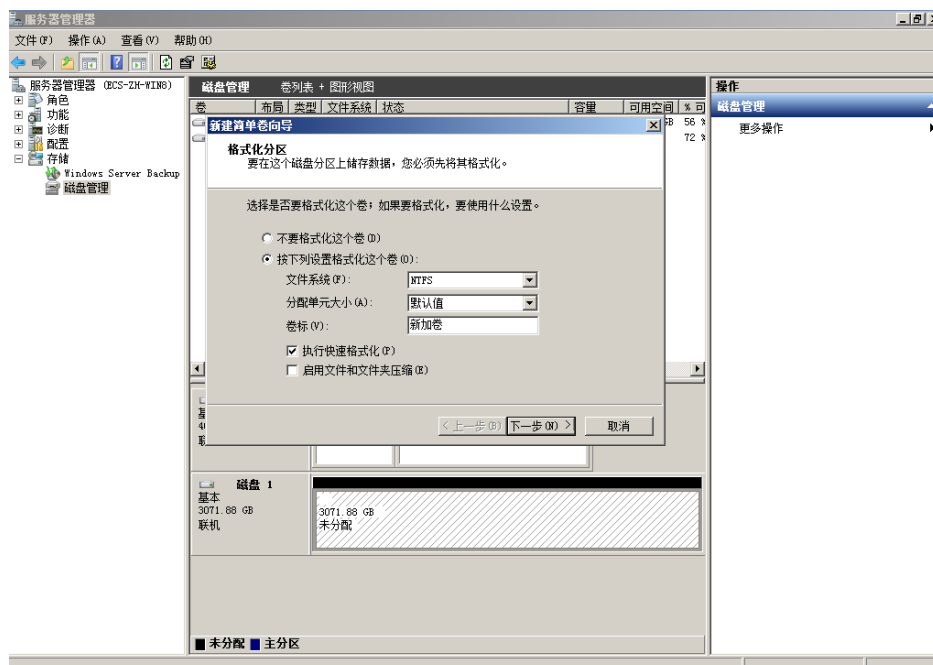
图 2-29 分配驱动器号和路径(Windows 2008)



**步骤9** 分配到驱动器号和路径，系统默认为磁盘分配驱动器号，驱动器号默认为“D”，此处以保持系统默认配置为例，单击“下一步”。

进入“格式化分区”页面，如图2-30所示。

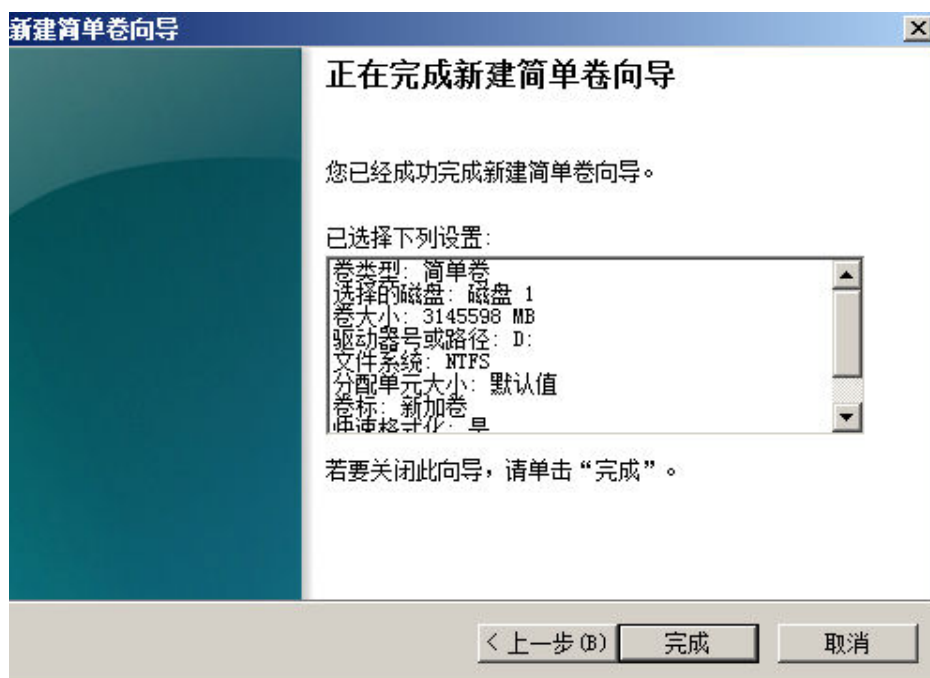
图 2-30 格式化分区(Windows 2008)



**步骤10** 格式化分区，系统默认的文件系统为NTFS，并根据实际情况设置其他参数，此处以保持系统默认设置为例，单击“下一步”。

进入“完成新建卷”页面，如图2-31所示。

图 2-31 完成新建卷



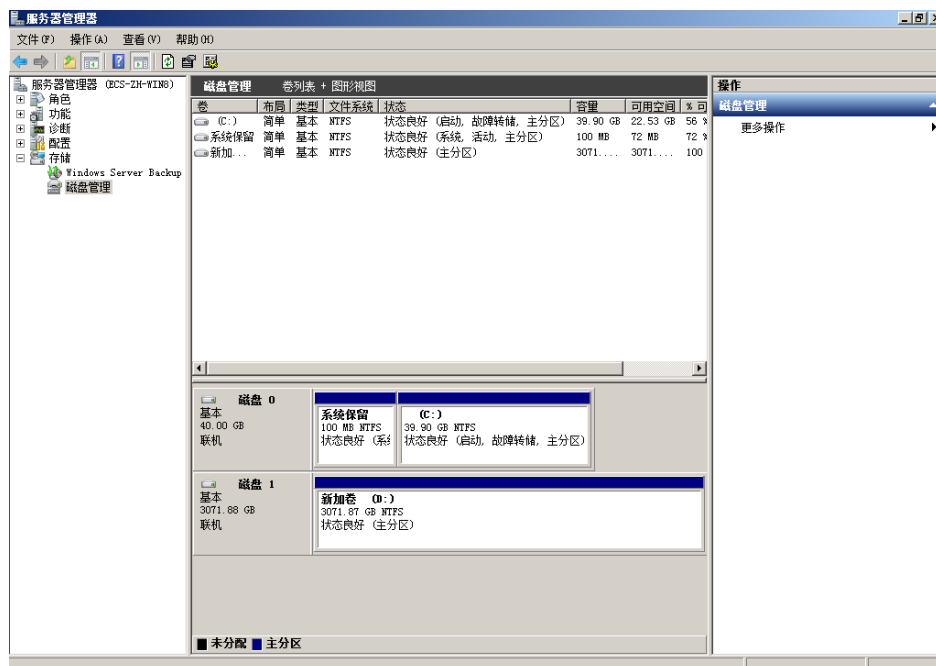
### 须知


不同文件系统支持的分区大小不同，请根据您的业务需求选择合适的文件系统。

### 步骤11 单击“完成”。

需要等待片刻让系统完成初始化操作，当卷状态为“状态良好”时，表示初始化磁盘成功，如图2-32所示。

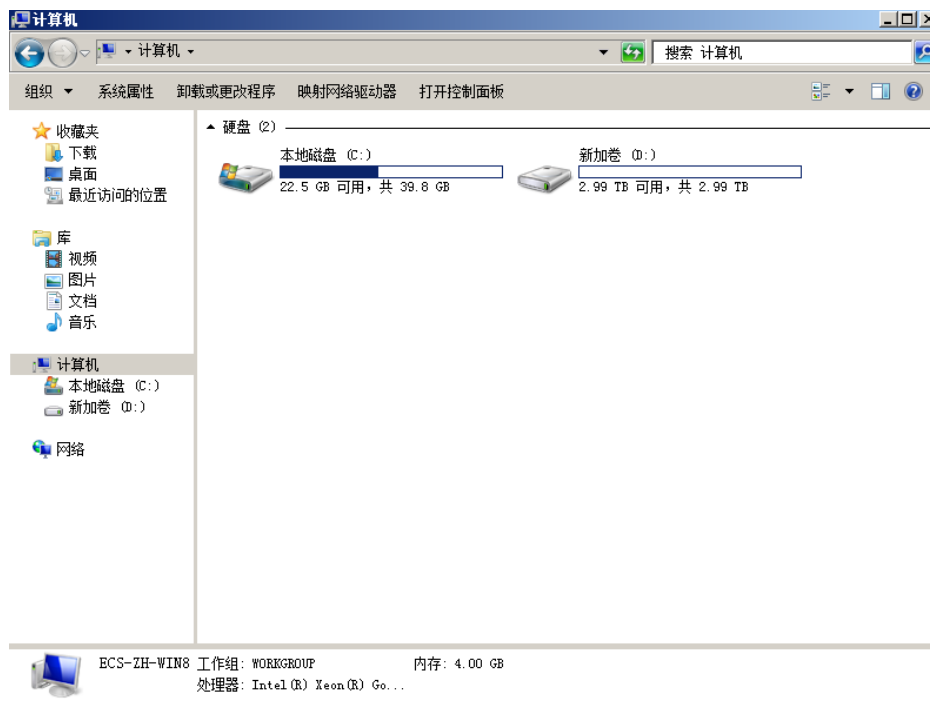
图 2-32 初始化磁盘成功(Windows 2008)



**步骤12** 新建卷完成后，单击，在文件资源管理器中查看是否有新建卷，此处以“新建卷（D:）”为例。

如果如图2-33所示，可以看到“新建卷（D:）”，表示磁盘初始化成功，任务结束。

图 2-33 文件资源管理器(Windows 2008)



----结束

## 2.3.7 初始化容量大于 2TiB 的 Windows 数据盘（Windows 2012）

### 操作场景

本文以云服务器的操作系统为“Windows Server 2012 R2 Standard 64bit”、磁盘容量为3 TiB举例，提供容量大于2 TiB的Windows数据盘的初始化操作指导。

MBR格式分区支持的磁盘最大容量为2 TiB，GPT分区表最大支持的磁盘容量为18 EiB，因此当为容量大于2 TiB的磁盘分区时，请采用GPT分区方式。具体操作请参见[初始化容量大于2TiB的Windows数据盘（Windows 2008）](#)。关于磁盘分区形式的更多介绍，请参见[场景及磁盘分区形式介绍](#)。

不同云服务器的操作系统的格式化操作可能不同，本文仅供参考，具体操作步骤和差异请参考对应的云服务器操作系统的产品文档。

#### 须知

首次使用磁盘时，如果您未参考本章节对磁盘执行初始化操作，主要包括创建分区和文件系统等操作，那么当后续扩容磁盘时，新增容量部分的磁盘可能无法正常使用。

## 前提条件

- 已挂载数据盘至云服务器，且该数据盘未初始化。
- 已登录。
  - 弹性云服务器请参见《弹性云服务器用户指南》。
  - 裸金属服务器请参见《裸金属服务器用户指南》。

## 操作指导


- 步骤1** 在云服务器桌面，单击桌面下方的。
- 弹出“服务器管理器”窗口，如图2-34所示。

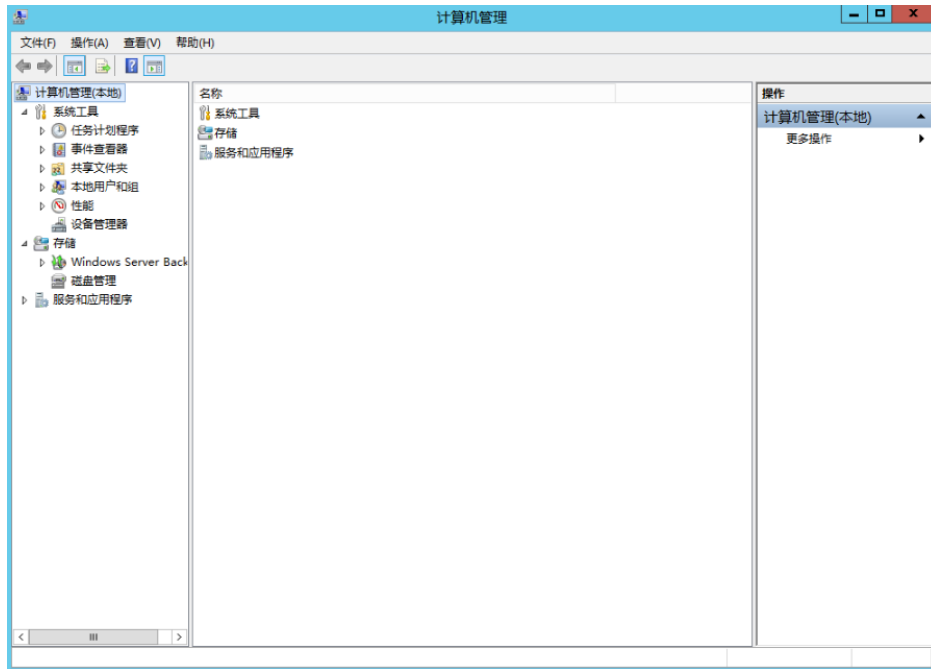
图 2-34 服务器管理器(Windows 2012)



- 步骤2** 在“服务器管理器”页面右上方选择“工具 > 计算机管理”。
- 弹出“计算机管理”窗口，如图2-35所示。



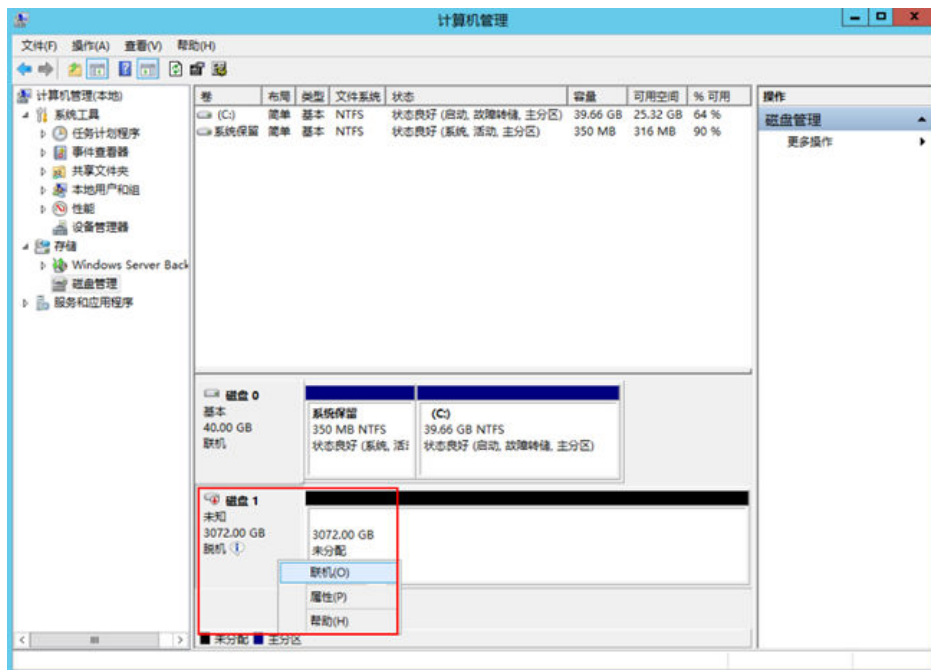
图 2-35 计算机管理窗口(Windows 2012)



步骤3 选择“存储 > 磁盘管理”。

进入磁盘列表页面，如图2-36所示。

图 2-36 磁盘列表(Windows 2012)

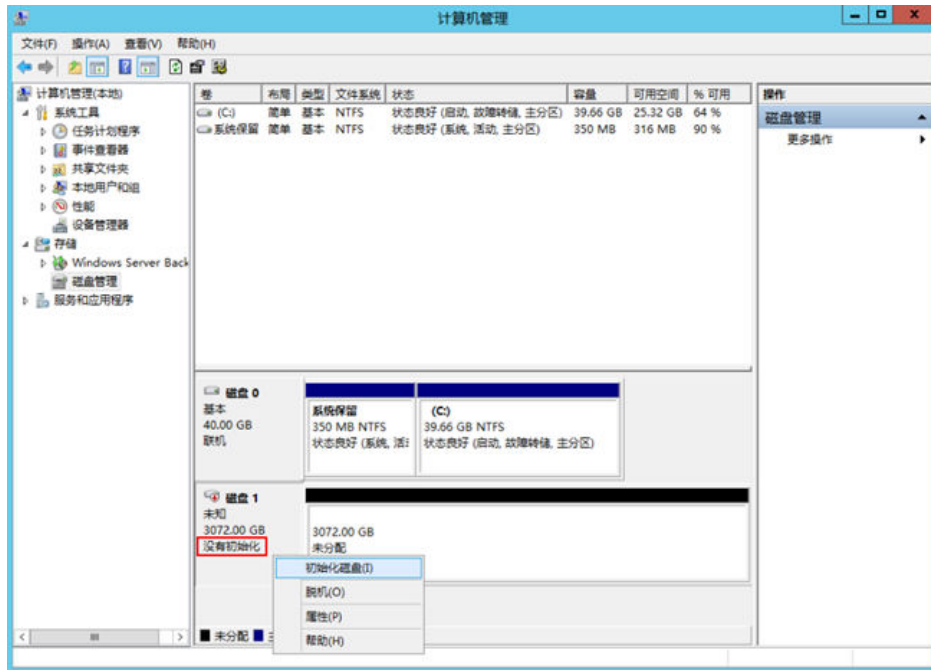


步骤4 (可选) 在页面右侧可以查看磁盘列表，如果新增磁盘处于脱机状态，需要先进行联机，再进行初始化。

在磁盘1区域，右键单击菜单列表中的“联机”。

如图2-37所示，当磁盘1由“脱机”状态变为“没有初始化”，表示联机成功。

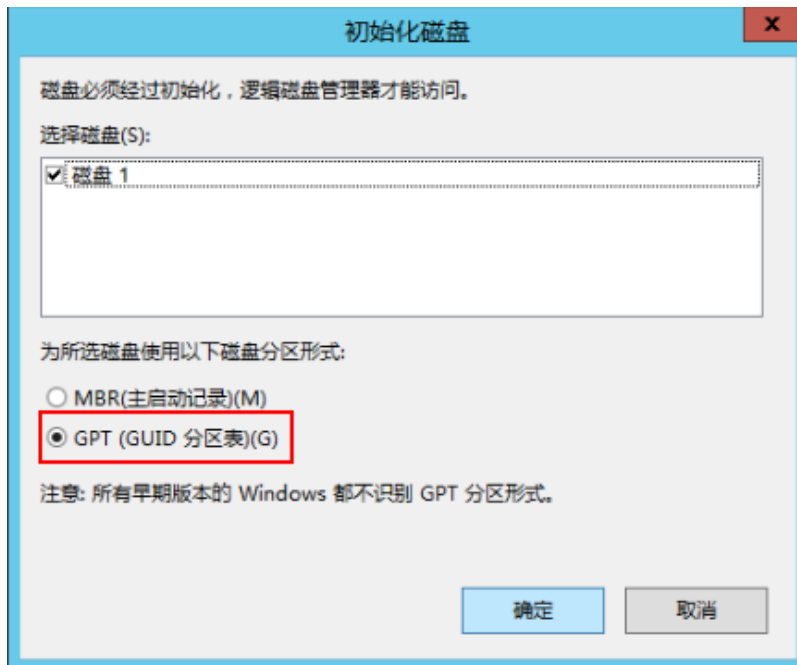
图 2-37 联机成功(Windows 2012)



步骤5 (可选) 在磁盘1区域, 右键单击菜单列表中的“初始化磁盘”。

弹出“初始化磁盘”窗口, 如图2-38所示。

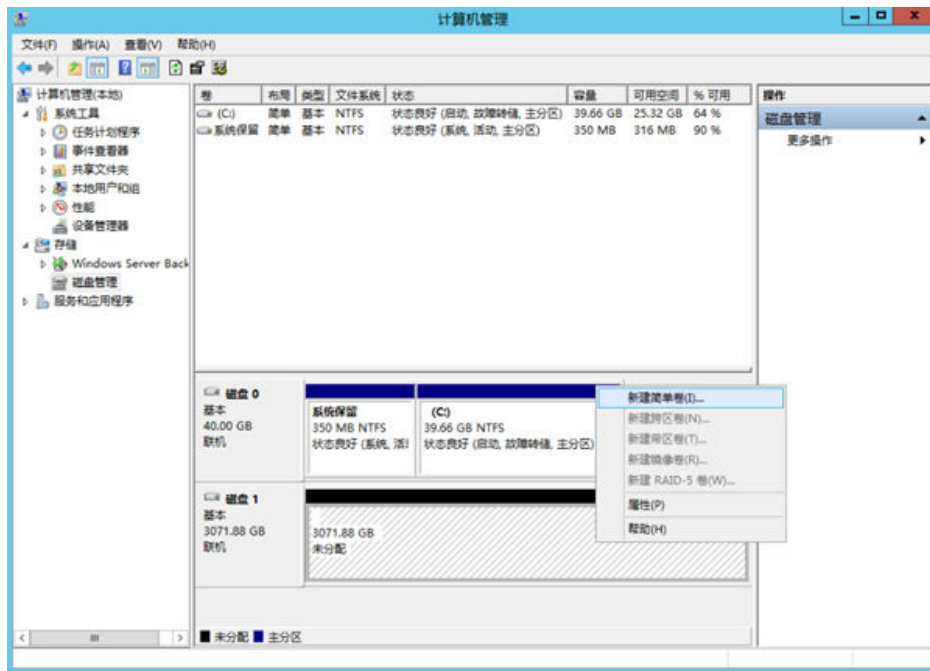
图 2-38 初始化磁盘(Windows 2012)



步骤6 在“初始化磁盘”对话框中显示需要初始化的磁盘, 对于大于2 TiB的磁盘, 此处请选择“GPT (GUID分区表)”, 单击“确定”。

返回“计算机管理”窗口, 如图2-39所示。

图 2-39 计算机管理(Windows 2012)



### 须知

MBR支持的云硬盘最大容量为2 TiB，GPT最大支持的云硬盘容量为18 EiB，当前数据盘支持的最大容量为32 TiB，如果您需要使用大于2 TiB的云硬盘容量，分区形式请采用GPT。

当云硬盘已经投入使用后，此时切换云硬盘分区形式时，云硬盘上的原有数据将会清除，因此请在云硬盘初始化时谨慎选择云硬盘分区形式。切换GPT分区形式前，请对云硬盘数据备份后，再格式化硬盘。

**步骤7** 在磁盘1右侧的未分配的区域，右键单击选择“新建简单卷”。

弹出“新建简单卷向导”窗口，如图2-40所示。

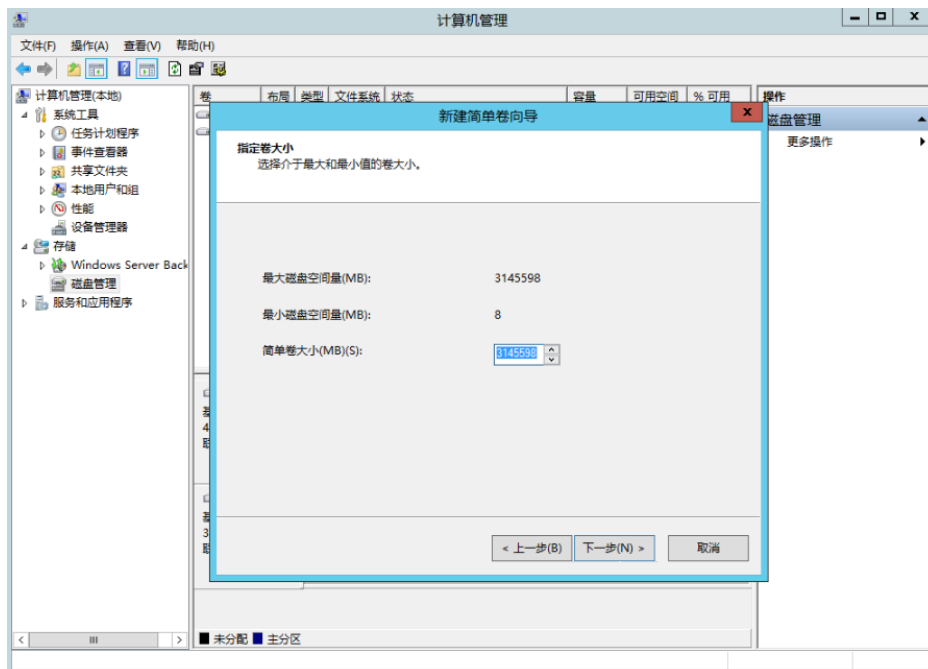
图 2-40 新建简单卷向导(Windows 2012)



**步骤8** 根据界面提示，单击“下一步”。

进入“指定卷大小”页面，如图2-41所示。

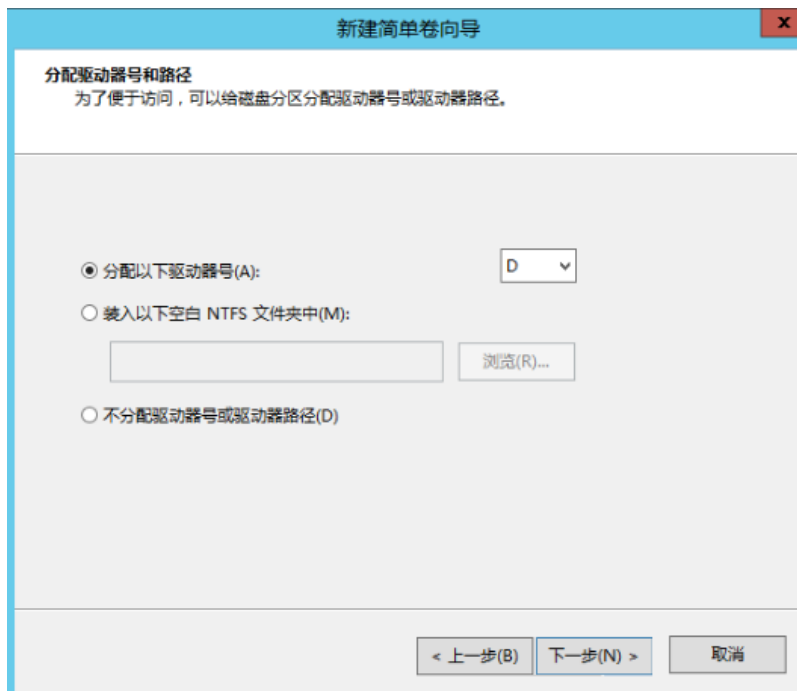
图 2-41 指定卷大小(Windows 2012)



**步骤9** 指定卷大小，系统默认卷大小为最大值，您还可以根据实际需求指定卷大小，此处以保持系统默认配置为例，单击“下一步”。

进入“分配驱动器号和路径”页面，如图2-42所示。

图 2-42 分配驱动器号和路径(Windows 2012)



**步骤10** 分配到驱动器号和路径，系统默认为磁盘分配驱动器号，驱动器号默认为“D”，此处以保持系统默认配置为例，单击“下一步”。

进入“格式化分区”页面，如图2-43所示。

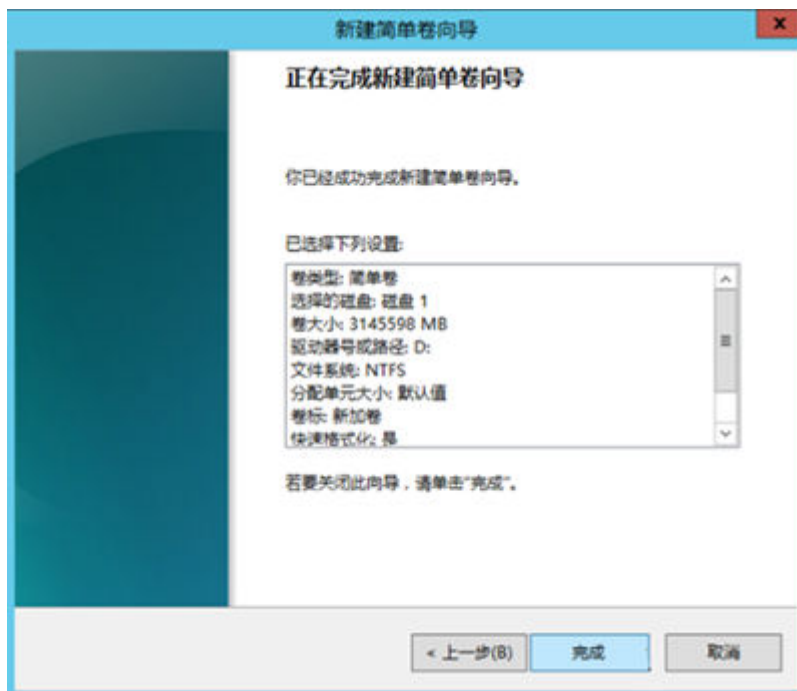
图 2-43 格式化分区(Windows 2012)



**步骤11** 格式化分区，系统默认的文件系统为NTFS，并根据实际情况设置其他参数，此处以保持系统默认设置为例，单击“下一步”。

进入“完成新建卷”页面，如图2-44所示。

图 2-44 完成新建卷(Windows 2012)



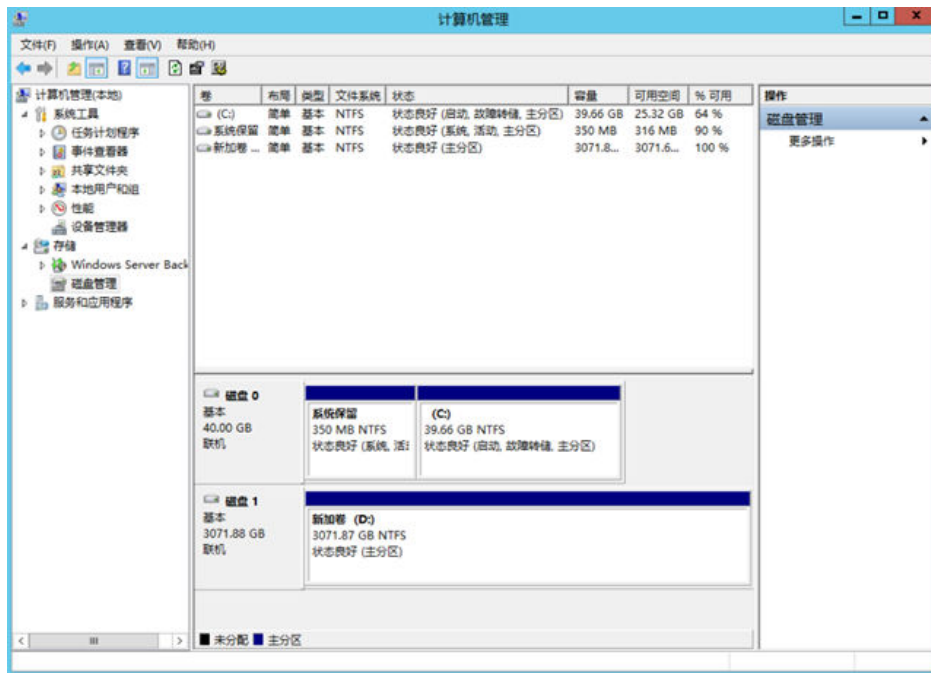
#### 须知


不同文件系统支持的分区大小不同，请根据您的业务需求选择合适的文件系统。

**步骤12** 单击“完成”。

需要等待片刻让系统完成初始化操作，当卷状态为“状态良好”时，表示初始化磁盘成功，如[图2-45](#)所示。

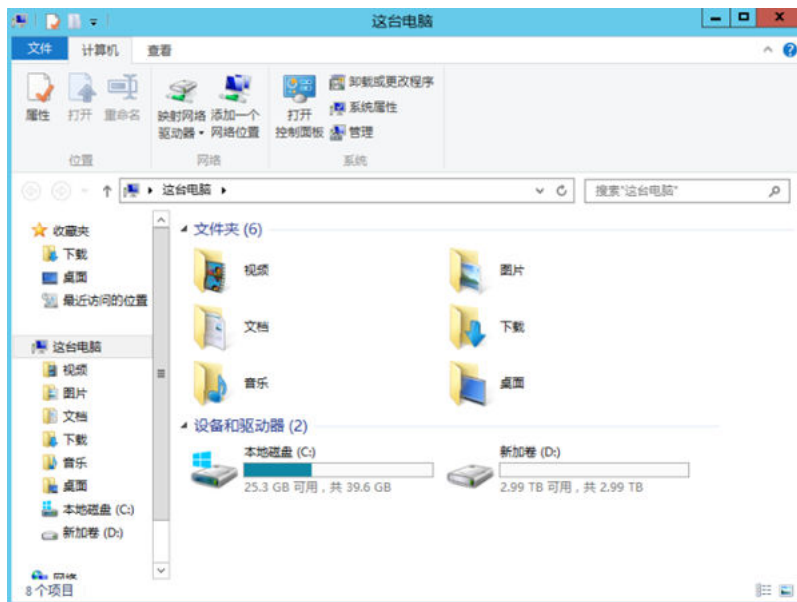
图 2-45 初始化磁盘成功(Windows 2012)



步骤13 新建卷完成后，单击，在文件资源管理器中查看是否有新建卷，此处以“新建卷（D:）”为例。

如果如图2-46所示，可以看到“新建卷（D:）”，表示磁盘初始化成功，任务结束。

图 2-46 文件资源管理器(Windows 2012)



----结束

## 2.3.8 初始化容量大于 2TiB 的 Linux 数据盘（parted）

### 操作场景

本文以云服务器的操作系统为“CentOS 7.4 64位”、磁盘容量为3 TiB举例，采用Parted分区工具为容量大于2 TiB的数据盘设置分区。

MBR支持的磁盘最大容量为2 TiB，GPT最大支持的磁盘容量为18 EiB，因此当您初始化容量大于2 TiB的磁盘时，分区形式请采用GPT。

fdisk分区工具只适用于MBR分区，parted工具适用于MBR分区和GPT分区。关于磁盘分区形式的更多介绍，请参见[场景及磁盘分区形式介绍](#)。

不同云服务器的操作系统的格式化操作可能不同，本文仅供参考，具体操作步骤和差异请参考对应的云服务器操作系统的产品文档。

#### 须知

首次使用磁盘时，如果您未参考本章节对磁盘执行初始化操作，主要包括创建分区和文件系统等操作，那么当后续扩容磁盘时，新增容量部分的磁盘可能无法正常使用。

### 前提条件

- 已挂载数据盘至云服务器，且该数据盘未初始化。
- 已登录。
  - 弹性云服务器请参见《弹性云服务器用户指南》。
  - 裸金属服务器请参见《裸金属服务器用户指南》。

### 划分分区并挂载磁盘

本操作以该场景为例，当云服务器挂载了一块新的数据盘时，采用parted分区工具为数据盘设置分区，分区形式设置为GPT，文件系统设为ext4格式，挂载在“/mnt/sdc”下，并设置开机启动自动挂载。

**步骤1** 执行以下命令，查看新增数据盘。

**lsblk**

回显类似如下信息：

```
[root@ecs-centos74 ~]# lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
vda 253:0 0 40G 0 disk
├─vda1 253:1 0 1G 0 part /boot
└─vda2 253:2 0 39G 0 part /
vdb 253:16 0 3T 0 disk
```

表示当前的云服务器有两块磁盘，“/dev/vda”是系统盘，“/dev/vdb”是新增数据盘。

**步骤2** 执行以下命令，进入parted分区工具，开始对新增数据盘执行分区操作。

**parted 新增数据盘**

以新挂载的数据盘“/dev/vdb”为例：



### parted /dev/vdb

回显类似如下信息：

```
[root@ecs-centos74 ~]# parted /dev/vdb
GNU Parted 3.1
Using /dev/vdb
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted)
```

**步骤3** 输入“p”，按“Enter”，查看当前磁盘分区形式。

回显类似如下信息：

```
(parted) p
Error: /dev/vdb: unrecognised disk label
Model: Virtio Block Device (virtblk)
Disk /dev/vdb: 3299GiB
Sector size (logical/physical): 512B/512B
Partition Table: unknown
Disk Flags:
(parted)
```

“Partition Table”为“unknown”表示磁盘分区形式未知，新的数据盘还未设置分区形式。

**步骤4** 输入以下命令，设置磁盘分区形式。

**mklabel** *磁盘分区形式*

磁盘分区形式有MBR和GPT两种，大于2 TiB的磁盘容量，请采用GPT分区方式：

**mklabel gpt**

#### 须知

MBR支持的云硬盘最大容量为2 TiB，GPT最大支持的云硬盘容量为18 EiB，当前数据盘支持的最大容量为32 TiB，如果您需要使用大于2 TiB的云硬盘容量，分区形式请采用GPT。

当云硬盘已经投入使用后，此时切换云硬盘分区形式时，云硬盘上的原有数据将会清除，因此请在云硬盘初始化时谨慎选择云硬盘分区形式。切换GPT分区形式前，请对云硬盘数据备份后，再格式化硬盘。

**步骤5** 输入“p”，按“Enter”，设置分区形式后查看磁盘分区形式。

回显类似如下信息：

```
(parted) mklabel gpt
(parted) p
Model: Virtio Block Device (virtblk)
Disk /dev/vdb: 3299GiB
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number Start End Size File system Name Flags
(parted)
```

**步骤6** 输入“unit s”，按“Enter”，设置磁盘的计量单位为磁柱。

**步骤7** 以为整个磁盘创建一个分区为例，执行以下命令，按“Enter”。

**mkpart** 磁盘分区名称 起始磁柱值 截止磁柱值

命令示例:

**mkpart opt 2048s 100%**

“2048s”表示磁盘起始磁柱值，“100%”表示磁盘截止磁柱值，此处仅供参考，您可以根据业务需要自行规划磁盘分区数量及容量。

回显类似如下信息:

```
(parted) mkpart opt 2048s 100%
Warning: The resulting partition is not properly aligned for best performance.
Ignore/Cancel? Ignore
```

如果出现以上性能优化提醒，请输入“Ignore”，忽视即可。

**步骤8** 输入“p”，按“Enter”，查看新建分区的详细信息。

回显类似如下信息:

```
(parted) p
Model: Virtio Block Device (virtblk)
Disk /dev/vdb: 6442450944s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number Start End Size File system Name Flags
1 2048s 6442448895s 6442446848s opt
```

表示新建分区“dev/vdb1”的详细信息。

**步骤9** 输入“q”，按“Enter”，退出parted分区工具。

**步骤10** 执行以下命令，查看磁盘分区信息。

**lsblk**

回显类似如下信息:

```
[root@ecs-centos74 ~]# lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
vda 253:0 0 40G 0 disk
├─vda1 253:1 0 1G 0 part /boot
├─vda2 253:2 0 39G 0 part /
vdb 253:16 0 3T 0 disk
├─vdb1 253:17 0 3T 0 part
```

此时可以查看到新建分区“/dev/vdb1”。

**步骤11** 执行以下命令，将新建分区文件系统设为系统所需格式。

**mkfs -t 文件系统格式 /dev/vdb1**

以设置文件系统为“ext4”为例:

**mkfs -t ext4 /dev/vdb1**

回显类似如下信息:

```
[root@ecs-centos74 ~]# mkfs -t ext4 /dev/vdb1
mke2fs 1.42.9 (28-Dec-2013)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
```

```
201326592 inodes, 805305856 blocks
40265292 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=2952790016
24576 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
    32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
    4096000, 7962624, 11239424, 20480000, 23887872, 71663616, 78675968,
    102400000, 214990848, 512000000, 550731776, 644972544

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

格式化需要等待一段时间，请观察系统运行状态，不要退出。

### 须知

不同文件系统支持的分区大小不同，请根据您的业务需求选择合适的文件系统。

**步骤12** 执行以下命令，新建挂载目录。

```
mkdir 挂载目录
```

以新建挂载目录“/mnt/sdc”为例：

```
mkdir /mnt/sdc
```

#### 📖 说明

Linux系统默认带有/mnt目录，如果创建失败，可能是/mnt被误删除了，可以执行命令**mkdir -p /mnt/sdc**创建。

**步骤13** 执行以下命令，将新建分区挂载到**步骤12**中创建的目录下。

```
mount 磁盘分区 挂载目录
```

以挂载新建分区“/dev/vdb1”至“/mnt/sdc”为例：

```
mount /dev/vdb1 /mnt/sdc
```

**步骤14** 执行以下命令，查看挂载结果。

```
df -TH
```

回显类似如下信息：

```
[root@ecs-centos74 ~]# df -TH
Filesystem      Type      Size  Used Avail Use% Mounted on
/dev/vda2       ext4      42G   1.5G  38G   4% /
devtmpfs        devtmpfs  2.0G   0  2.0G   0% /dev
tmpfs           tmpfs     2.0G   0  2.0G   0% /dev/shm
tmpfs           tmpfs     2.0G   8.9M  2.0G   1% /run
tmpfs           tmpfs     2.0G   0  2.0G   0% /sys/fs/cgroup
/dev/vda1       ext4      1.1G  153M  801M  17% /boot
tmpfs           tmpfs     398M   0  398M   0% /run/user/0
/dev/vdb1       ext4      3.3T   93M  3.1T   1% /mnt/sdc
```

表示新建分区“dev/vdb1”已挂载至“/mnt/sdc”。

----**结束**

## 设置开机自动挂载磁盘分区

您可以通过配置fstab文件，设置系统启动时自动挂载磁盘分区。已有数据的也可以进行设置，该操作不会影响现有数据。

本文介绍如何在fstab文件中使用UUID来设置自动挂载磁盘分区。不建议采用在“/etc/fstab”直接指定设备名（比如/dev/vdb1）的方法，因为云中设备的顺序编码在关闭或者开启过程中可能发生改变，例如/dev/vdb1可能会变成/dev/vdb2，可能会导致重启后不能正常运行。

### 📖 说明

UUID (universally unique identifier) 是Linux系统为磁盘分区提供的唯一的标识字符串。

**步骤1** 执行如下命令，查询磁盘分区的UUID。

**blkid 磁盘分区**

以查询磁盘分区“/dev/vdb1”的UUID为例：

**blkid /dev/vdb1**

回显类似如下信息：

```
[root@ecs-test-0001 ~]# blkid /dev/vdb1
/dev/vdb1: UUID="0b3040e2-1367-4abb-841d-ddb0b92693df" TYPE="ext4"
```

记录下回显中磁盘分区“/dev/vdb1”的UUID，方便后续步骤使用。

**步骤2** 执行以下命令，使用VI编辑器打开“fstab”文件。

**vi /etc/fstab**

**步骤3** 按“i”，进入编辑模式。

**步骤4** 将光标移至文件末尾，按“Enter”，添加如下内容。

```
UUID=0b3040e2-1367-4abb-841d-ddb0b92693df /mnt/sdc ext4 defaults 0 2
```

以上内容仅为示例，具体请以实际情况为准，参数说明如下：

- 第一列为UUID，此处填写**步骤1**中查询到的磁盘分区的UUID。
- 第二列为磁盘分区的挂载目录，可以通过**df -TH**命令查询。
- 第三列为磁盘分区的文件系统格式，可以通过**df -TH**命令查询。
- 第四列为磁盘分区的挂载选项，此处通常设置为defaults即可。
- 第五列为Linux dump备份选项。
  - 0表示不使用Linux dump备份。现在通常不使用dump备份，此处设置为0即可。
  - 1表示使用Linux dump备份。
- 第六列为fsck选项，即开机时是否使用fsck检查磁盘。
  - 0表示不检验。
  - 挂载点为(/)根目录的分区，此处必须填写1。  
根分区设置为1，其他分区只能从2开始，系统会按照数字从小到大依次检查下去。

**步骤5** 按“ESC”后，输入“:wq”，按“Enter”。

保存设置并退出编辑器。

**步骤6** 执行以下步骤，验证自动挂载功能。

1. 执行如下命令，卸载已挂载的分区。

**umount** *磁盘分区*

命令示例：

**umount /dev/vdb1**

2. 执行如下命令，将“/etc/fstab”文件所有内容重新加载。

**mount -a**

3. 执行如下命令，查询文件系统挂载信息。

**mount | grep** *挂载目录*

命令示例：

**mount | grep /mnt/sdc**

回显类似如下信息，说明自动挂载功能生效：

```
root@ecs-test-0001 ~]# mount | grep /mnt/sdc  
/dev/vdb1 on /mnt/sdc type ext4 (rw,relatime,data=ordered)
```

----**结束**

# 3 通过 IAM 授予使用 ECS 的权限

## 3.1 创建用户并授权使用 ECS

如果您需要对您所拥有的ECS进行精细的权限管理，您可以使用统一身份认证服务（Identity and Access Management，简称IAM），通过IAM，您可以：

- 根据企业的业务组织，在您的账号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用ECS资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- 将ECS资源委托给更专业、高效的其他账号或者云服务，这些账号或者云服务可以根据权限进行代运维。

如果账号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节，不影响您使用ECS服务的其它功能。

本章节为您介绍对用户授权的方法，操作流程如[示例流程](#)所示。

### 前提条件

给用户组授权之前，请您了解用户组可以添加的ECS系统策略，并结合实际需求进行选择。

ECS支持的系统策略及策略间的对比，请参见：[ECS系统策略](#)。

## 示例流程

图 3-1 给用户授权 ECS 权限流程



1. 创建用户组并授权  
在IAM控制台创建用户组，并在“操作”列下选择“授权”，授予弹性云服务只读权限“ECS ReadOnlyAccess”。
2. 创建用户并加入用户组  
在IAM控制台创建用户，并在“操作”列下选择“授权”，将其加入1中创建的用户组。
3. 用户登录并验证权限  
新创建的用户登录控制台，切换至授权区域，验证弹性云服务的只读权限。
  - 在“服务列表”中选择弹性云服务器，进入ECS主界面，单击右上角“创建弹性云服务器”，尝试创建弹性云服务器，如果无法创建弹性云服务器（假设当前权限仅包含ECSReadOnlyAccess），表示“ECSReadOnlyAccess”已生效。
  - 在“服务列表”中选择除弹性云服务器外（假设当前策略仅包含ECSReadOnlyAccess）的任一服务，若提示权限不足，表示“ECSReadOnlyAccess”已生效。

## 3.2 ECS 自定义策略

如果系统预置的ECS权限，不满足您的授权要求，可以创建自定义策略。自定义策略中可以添加的授权项（Action）请参考《ECS API参考》中“策略及授权项说明”章节。

目前云平台支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。

- JSON视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写JSON格式的策略内容。

具体创建步骤请参见《统一身份认证服务》的“创建自定义策略”章节。本章为您介绍常用的ECS自定义策略样例。

## ECS 自定义策略样例

- 示例1：授权仅允许用户批量开机、关机、重启云服务器

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:cloudServerFlavors:get",
        "ecs:cloudServers:reboot",
        "ecs:cloudServers:start",
        "ecs:cloudServers:get",
        "ecs:cloudServers:list",
        "ecs:cloudServers:stop"
      ]
    }
  ]
}
```

- 示例2：授权仅允许用户批量关闭云服务器、删除云服务器

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:cloudServers:get",
        "ecs:cloudServers:delete",
        "ecs:cloudServers:list",
        "ecs:cloudServers:stop"
      ]
    }
  ]
}
```

- 示例3：授权仅允许用户通过VNC进行登录

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ecs:cloudServerFlavors:get",
        "ecs:cloudServers:vnc",
        "ecs:cloudServers:get",
        "ecs:cloudServers:list"
      ]
    }
  ]
}
```

- 示例4：授权拒绝用户删除云服务器

拒绝策略需要同时配合其他策略使用，否则没有实际作用。用户被授予的策略中，一个授权项的作用如果同时存在Allow和Deny，则遵循Deny优先。

如果您给用户授予ECSFullAccess的系统策略，但不希望用户拥有ECSFullAccess中定义的删除云服务器权限，您可以创建一条拒绝删除云服务的自定义策略，然后将ECSFullAccess和拒绝策略授予用户，根据Deny优先原则，则用户可以对ECS执行除了删除云服务器外的所有操作。拒绝策略示例如下：



```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "ecs:cloudServers:delete"
      ]
    }
  ]
}
```

# 4 实例

## 4.1 登录 Windows ECS

### 4.1.1 Windows ECS 登录方式概述

#### 约束与限制

- 只有运行中的云服务器才允许用户登录。
- Windows操作系统用户名“Administrator”。
- GPU实例中，部分G系列实例不支持控制台提供的远程登录功能，需要自行安装VNC Server进行登录。详细信息请参见[GPU加速型](#)。推荐使用MSTSC方式登录弹性云服务器。

#### 登录方式概述

请根据需要选择登录方式，登录云服务器。

表 4-1 Windows 云服务器登录方式一览

云服务器操作系统	本地主机操作系统	连接方法	条件
Windows	Windows	使用mstsc方式登录云服务器。 在本地主机单击“开始”菜单，输入mstsc命令，打开远程桌面对话框。 <a href="#">通过本地Windows主机登录Windows ECS。</a>	云服务器器绑定弹性公网IP
	Linux	安装远程连接工具，例如rdesktop，执行连接命令。 <a href="#">通过本地Linux主机登录Windows ECS。</a>	

云服务器操作系统	本地主机操作系统	连接方法	条件
	macOS系统	安装远程连接工具，例如Microsoft Remote Desktop for Mac在macOS系统上登录。 <a href="#">通过本地macOS主机登录Windows ECS。</a>	
	移动设备	安装远程连接工具，例如Microsoft Remote Desktop在移动设备上登录。 <a href="#">通过移动设备登录Windows ECS。</a>	
	Windows	使用管理控制台远程登录方式： <a href="#">通过控制台VNC登录Windows ECS。</a>	不依赖弹性公网IP

## 4.1.2 通过控制台 VNC 登录 Windows ECS

### 操作场景

本节为您介绍如何通过控制台提供的远程登录功能（即VNC方式）登录到弹性云服务器上。

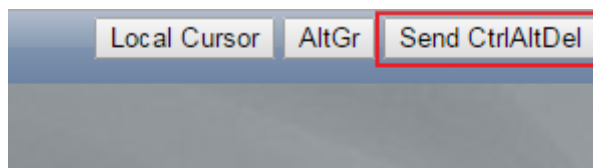
### 约束与限制

- 当前提供的远程登录功能是通过系统配置的自定义端口进行访问的，所以在使用远程登录功能时，请确保需要使用的端口未被防火墙屏蔽。例如：远程登录的链接为“xxx:8002”，则需要确保端口8002没有被防火墙屏蔽。
- 如果客户端操作系统使用了本地代理，且用户无法配置该本地代理的防火墙端口，请关闭代理模式后再使用远程登录功能。

### 登录 Windows 弹性云服务器

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 获取弹性云服务器密码。  
VNC方式登录弹性云服务器时，需已知其密码，然后再采用VNC方式登录。
  - 当您的弹性云服务器是采用密码方式鉴权时，请直接使用创建云服务器时设置的密码进行登录。
4. 选择要登录的弹性云服务器，单击“操作”列下的“远程登录”。
5. （可选）如果界面提示“Press CTRL+ALT+DELETE to log on”，请单击远程登录操作面板上方的“Send CtrlAltDel”按钮进行登录。

图 4-1 单击“Send CtrlAltDelete”



6. 根据界面提示，输入弹性云服务器密码。

### 4.1.3 通过本地 Windows 主机登录 Windows ECS

#### 操作场景

本节为您介绍如何在本机使用远程登录工具MSTSC登录Windows弹性云服务器。

#### 前提条件

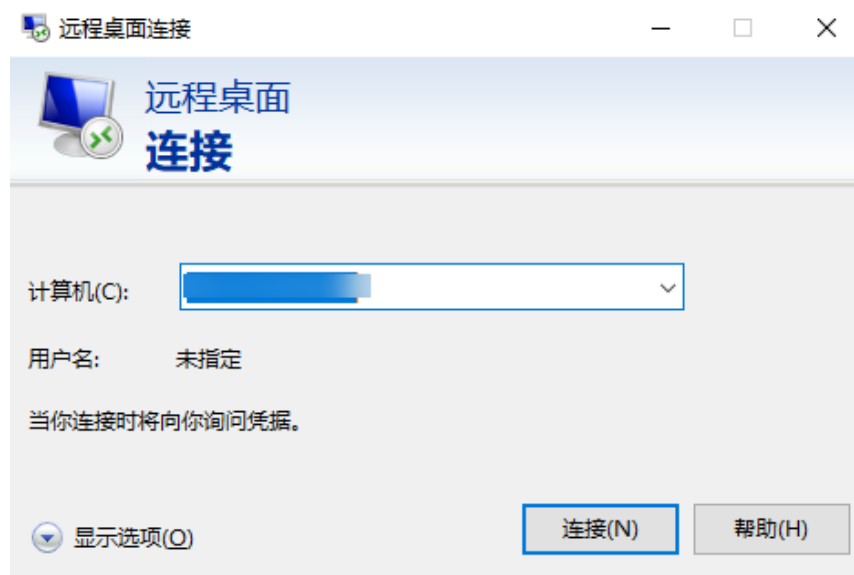
- 弹性云服务器状态为“运行中”。
- 弹性云服务器已经绑定弹性公网IP，绑定方式请参见[绑定弹性公网IP](#)。
- 所在安全组入方向已开放3389端口，配置方式请参见[配置安全组规则](#)。
- 使用的登录工具与待登录的弹性云服务器之间网络连通。例如，默认的3389端口没有被防火墙屏蔽。
- 弹性云服务器开启远程桌面协议RDP（Remote Desktop Protocol）。使用公共镜像创建的弹性云服务器默认已打开RDP。打开RDP方法请参考[开启远程桌面协议RDP](#)。

#### 使用 MSTSC 方式登录 Windows 弹性云服务器

本地主机为Windows操作系统，那么可以使用Windows自带的远程桌面连接工具MSTSC登录Windows云服务器。

1. 在本地主机单击“开始”菜单。
2. 在“搜索程序和文件”中，输入“mstsc”，单击mstsc打开远程桌面连接工具。
3. 在“远程桌面连接”的对话框中，单击“选项”。

图 4-2 显示选项

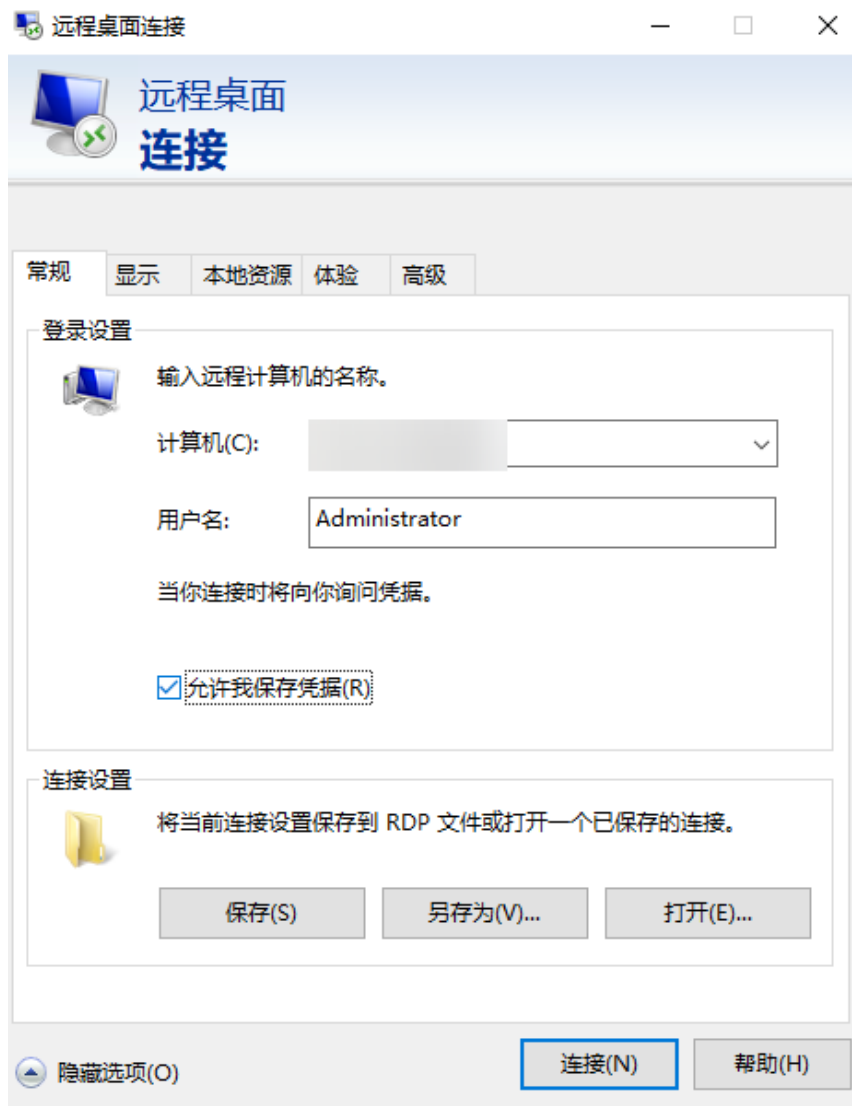


4. 输入待登录的云服务器的弹性公网IP和用户名，默认为Administrator。

#### 说明

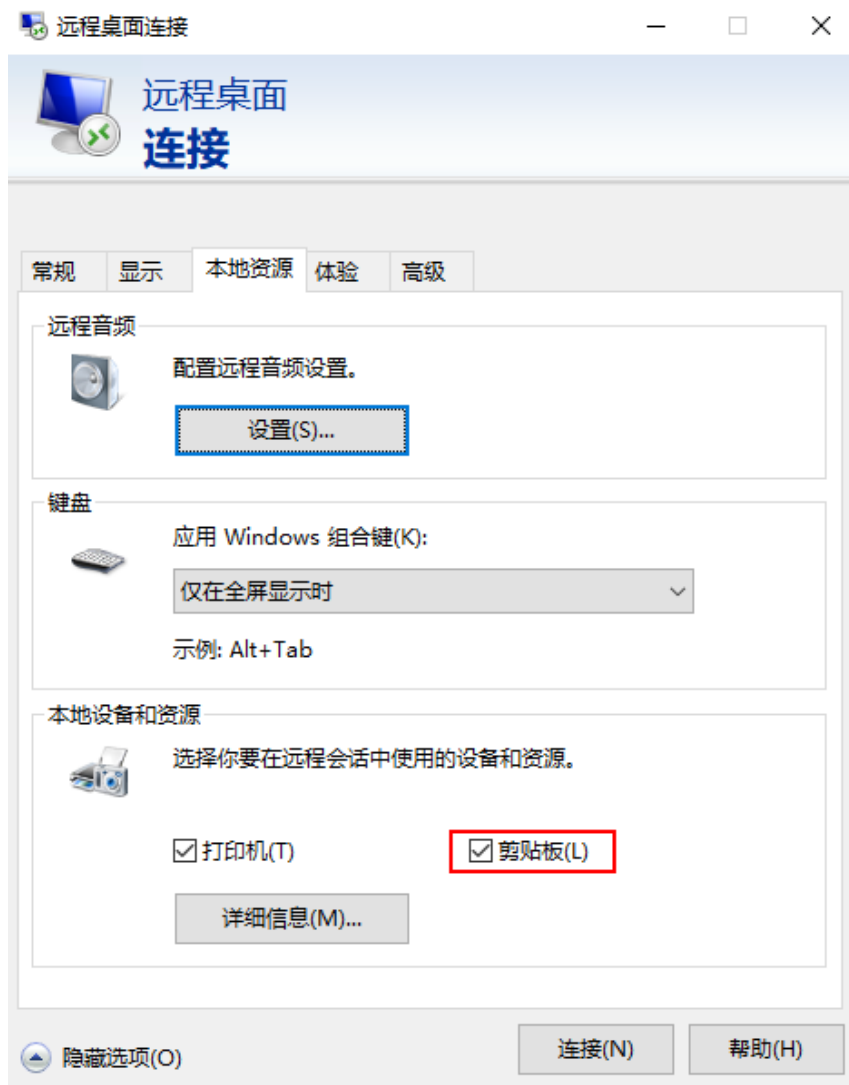
如需再次登录时不再重复输入用户名和密码，可勾选“允许我保存凭据”。

图 4-3 远程桌面链接



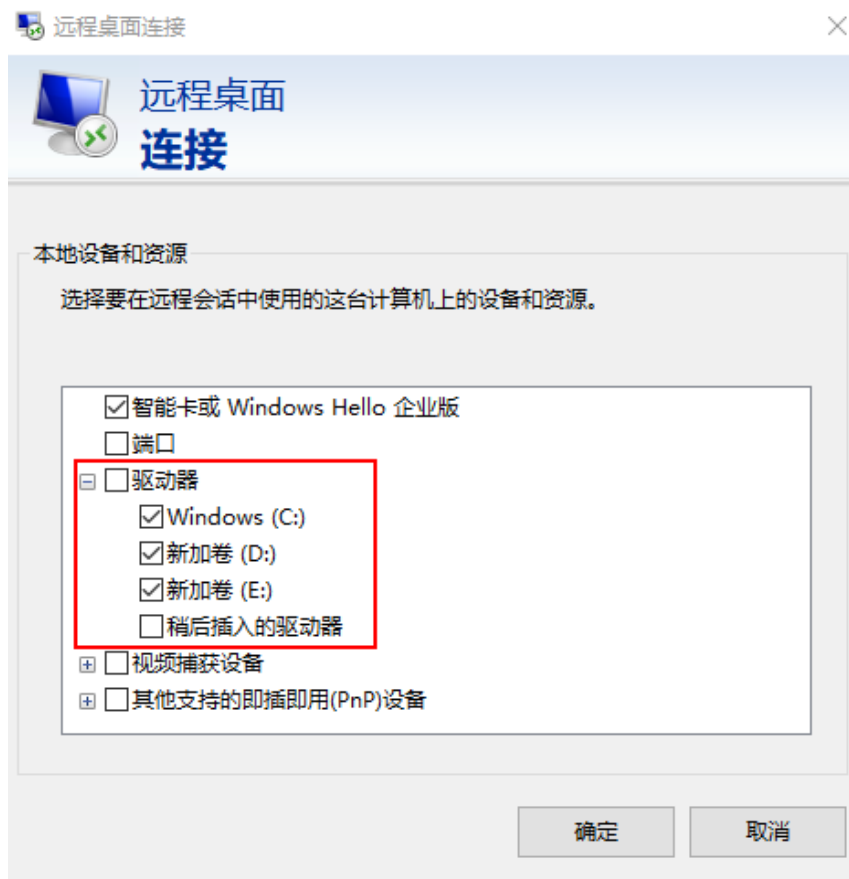
5. （可选）如需在远程会话中使用本地主机的资源，请单击“本地资源”选项卡完成如下配置。  
如需从本地主机复制到云服务器中，请勾选“剪贴板”。

图 4-4 勾选剪贴板



如需从本地主机复制文件到云服务器中，单击“详细信息”，勾选相应的磁盘。

图 4-5 勾选驱动器



6. (可选) 如需调整远程桌面窗口的大小, 可以选择“显示”选项卡, 再调整窗口大小。

图 4-6 调整窗口大小



7. 单击“确定”，根据提示输入密码，登录云服务器。  
为安全起见，首次登录云服务器，需更改密码。
8. （可选）通过远程桌面连接（Remote Desktop Protocol，RDP）方式登录云服务器后，如果需要使用RDP提供的“剪切板”功能，将本地的大文件（文件大小超过2GB）复制粘贴至远端的Windows云服务器中，由于Windows系统的限制，会导致操作失败。  
具体的解决方法，请参考[使用远程桌面链接方式复制文件](#)。

## 开启远程桌面协议 RDP

首次登录弹性云服务器时，请先使用VNC方式登录弹性云服务器，打开RDP（Remote Desktop Protocol），然后再使用mstsc方式连接。

### 📖 说明

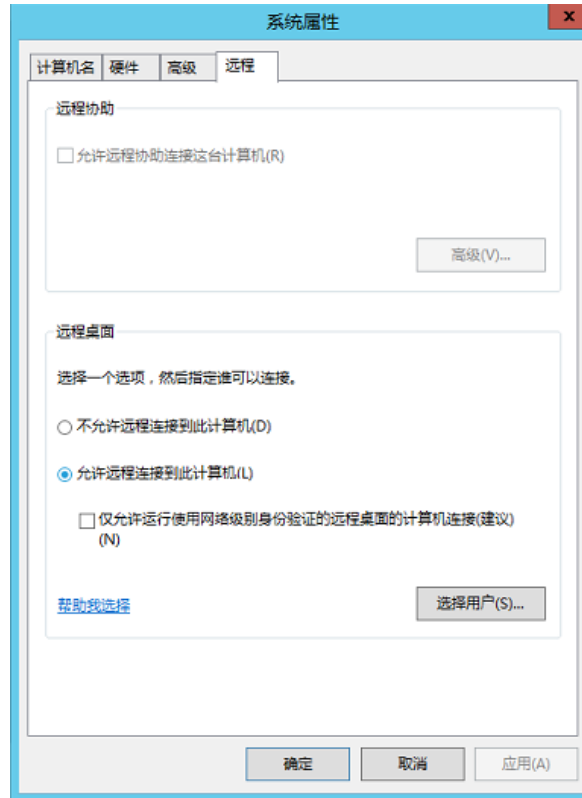
使用公共镜像创建的云服务器，默认已打开RDP。

1. VNC方式登录弹性云服务器。  
登录方法请参见[通过控制台VNC登录Windows ECS](#)。



- 单击“开始”菜单，选择“控制面板 > 系统和安全 > 系统 > 远程设置”。系统进入“系统属性”页面。

图 4-7 系统属性



- 选择“远程”页签，在“远程桌面”栏，选择“允许远程连接到此计算机”。
- 单击“确定”。

## 4.1.4 通过本地 Linux 主机登录 Windows ECS

### 操作场景

本节为您介绍如何在Linux操作系统主机上登录Windows云服务器。

### 前提条件

- 云服务器状态为“运行中”。
- 云服务器已经绑定弹性公网IP。
- 所在安全组入方向已开放3389端口。
- 使用的登录工具与待登录的云服务器之间网络连通。例如，默认的3389端口没有被防火墙屏蔽。
- 云服务器开启远程桌面协议RDP ( Remote Desktop Protocol )。使用公共镜像创建的云服务器默认已打开RDP。打开RDP方法请参考[开启远程桌面协议RDP](#)。

## 操作步骤

如果本地主机为Linux操作系统，您可以使用远程连接工具（例如rdesktop）连接Windows实例。

1. 执行以下命令，检查云服务器是否安装rdesktop。

**rdesktop**

如果提示“command not found”说明未安装rdesktop。请参考[rdesktop工具官方](#)获取rdesktop安装包安装rdesktop。

2. 输入以下命令登录云服务器。

**rdesktop -u 用户名 -p 密码 -g 分辨率 弹性公网IP地址**

例如：**rdesktop -u administrator -p password -g 1024\*720 121.xx.xx.xxx**

表 4-2 远程登录命令参数

参数	说明
-u	用户名，Windows实例默认用户名是Administrator。
-p	登录Windows实例的密码。
-f	默认全屏，需要用 Ctrl+Alt+Enter 组合键进行全屏模式切换。
-g	分辨率，中间用星号（*）连接，可省略，省略后默认为全屏显示。例如： <b>1024*720</b>
弹性公网IP地址	需要远程连接的服务器IP地址。需要替换为您的Windows实例的弹性公网IP地址。

## 开启远程桌面协议 RDP

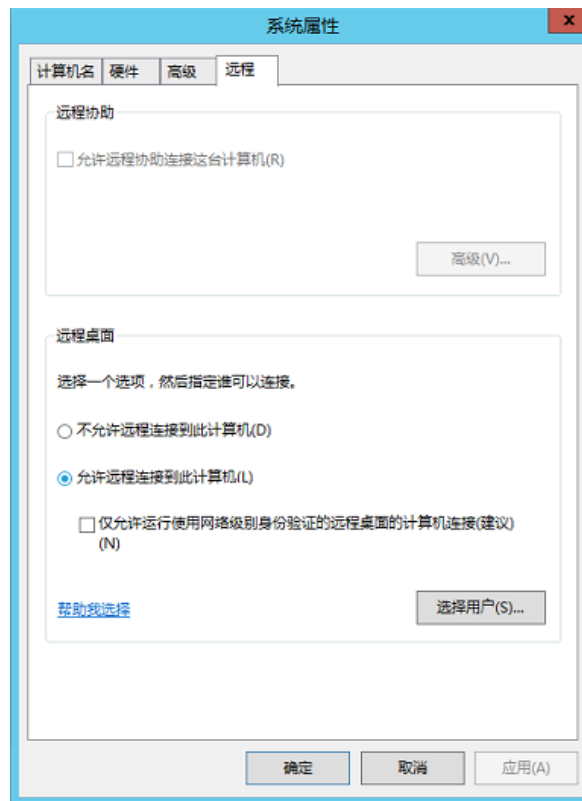
首次登录弹性云服务器时，请先使用VNC方式登录弹性云服务器，打开RDP（Remote Desktop Protocol），然后再使用mstsc方式连接。

### 📖 说明

使用公共镜像创建的云服务器，默认已打开RDP。

1. VNC方式登录弹性云服务器。  
登录方法请参见[通过控制台VNC登录Windows ECS](#)。
2. 单击“开始”菜单，选择“控制面板 > 系统和安全 > 系统 > 远程设置”。  
系统进入“系统属性”页面。

图 4-8 系统属性



3. 选择“远程”页签，在“远程桌面”栏，选择“允许远程连接到此计算机”。
4. 单击“确定”。

## 4.1.5 通过本地 macOS 主机登录 Windows ECS

### 操作场景

本节操作以使用“Microsoft Remote Desktop for Mac”工具远程连接“Windows Server 2012 R2 数据中心版 64位”操作系统云服务器为例，介绍macOS系统登录Windows云服务器的操作步骤。

### 前提条件

- 云服务器状态为“运行中”。
- 已获取Windows云服务器用户名和密码。
- 弹性云服务器已经绑定弹性公网IP，绑定方式请参见[绑定弹性公网IP](#)。
- 所在安全组入方向已开放3389端口，配置方式请参见[配置安全组规则](#)。
- 已安装Microsoft Remote Desktop for Mac或其他macOS系统适用的远程连接工具。

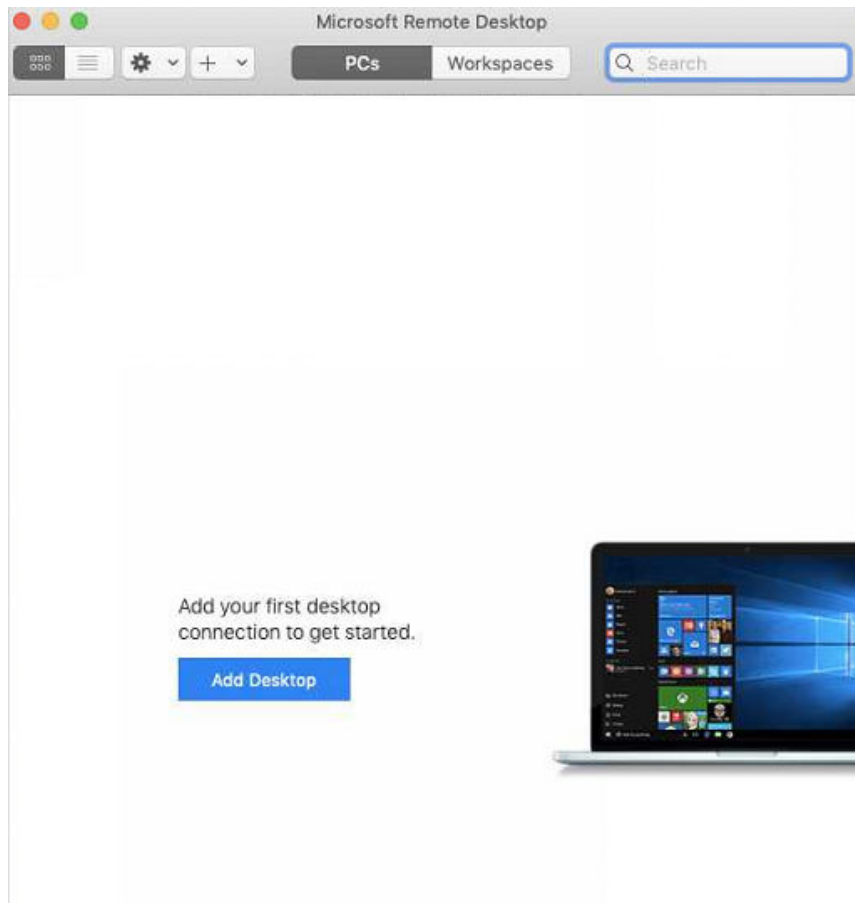
微软官方已停止提供Remote Desktop客户端的下载链接，您可单击[Microsoft Remote Desktop Beta](#)下载Beta版本使用。

### 操作步骤

1. 启动Microsoft Remote Desktop。

2. 单击“Add Desktop”。

图 4-9 Add Desktop



3. 在“Add PC”页面，设置登录信息。
  - PC name: 输入需要登录的Windows实例的弹性公网IP地址。
  - User account: 在下拉列表中选择“Add user account”。弹出“Add user account”对话框。
    - i. 输入Windows实例账号“administrator”，并输入实例的登录密码，单击“Add”。

图 4-10 Add user account

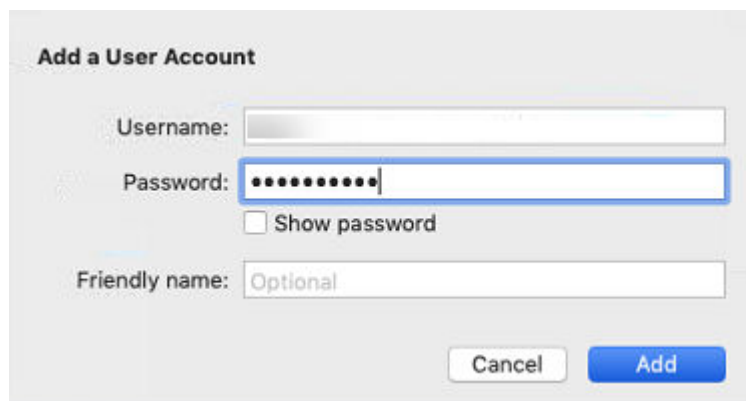
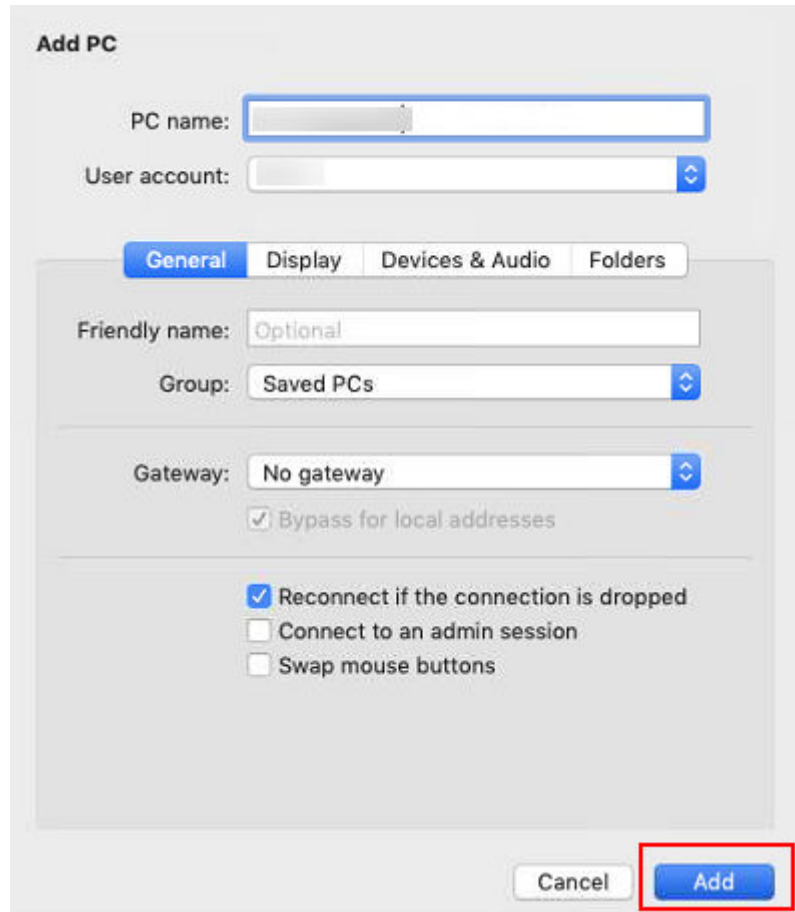
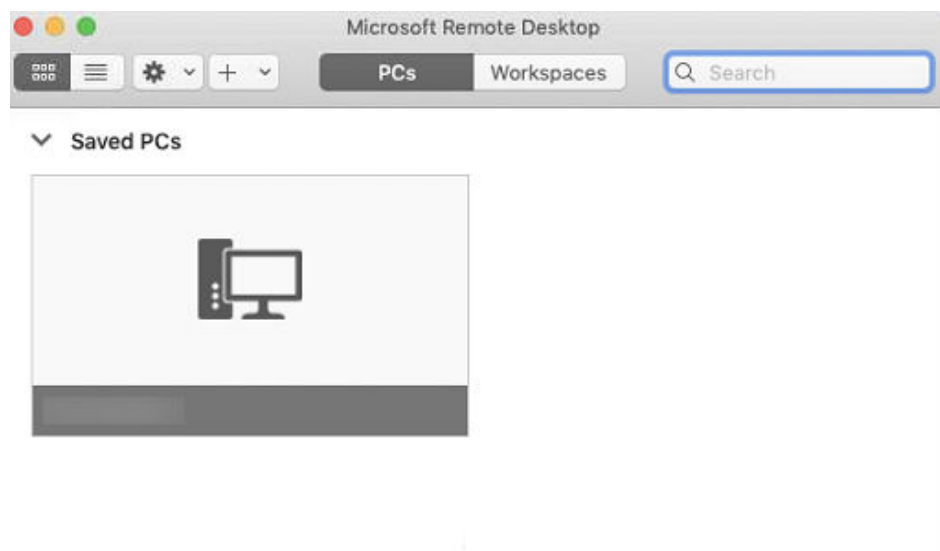


图 4-11 Add PC



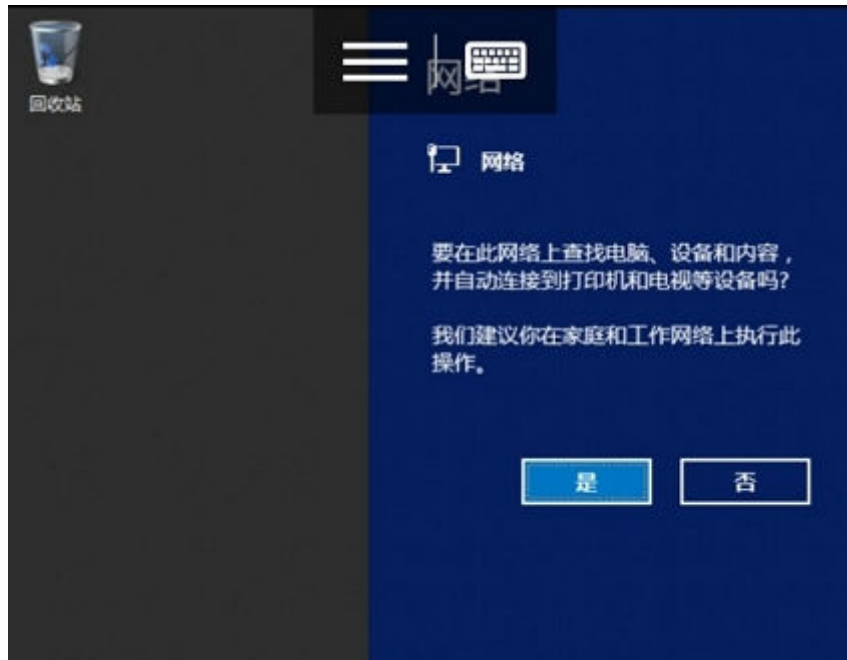
4. 在“Remote Desktop”页面，双击需要登录的Windows实例图标。

图 4-12 双击登录 Windows 实例



5. 确认信息后，单击“Continue”。  
至此，您已经登录Windows实例。

图 4-13 登录成功



## 4.1.6 通过移动设备登录 Windows ECS

### 操作场景

本节操作以“Windows Server 2012 R2 数据中心版 64位”操作系统为例，介绍如何通过Microsoft Remote Desktop客户端登录Windows实例。

### 前提条件

- 云服务器状态为“运行中”。
- 已获取Windows云服务器用户名和密码。
- 云服务器已经绑定弹性公网IP，绑定方式请参见[绑定弹性公网IP](#)。
- 所在安全组入方向已开放3389端口，配置方式请参见[配置安全组规则](#)。
- 移动设备已通过“应用市场”搜索并下载远程连接工具，本文以“Microsoft Remote Desktop”为例进行介绍。

### 操作步骤


1. 启动RD Client。
2. 在“Remote Desktop”页面右上角，单击图标，选择“Desktop”。

图 4-14 Remote Desktop



3. 在“Add desktop”页面，设置登录信息后，单击“SAVE”。
  - PC name: 输入需要登录的Windows实例的弹性公网IP地址。
  - 按以下步骤设置“User name”：
    - i. 单击“User name”，在下拉列表中选择“Add user account”。弹出“Add user account”对话框。
    - ii. 输入Windows实例账号“administrator”，并输入实例的登录密码，单击“SAVE”。

图 4-15 输入登录信息

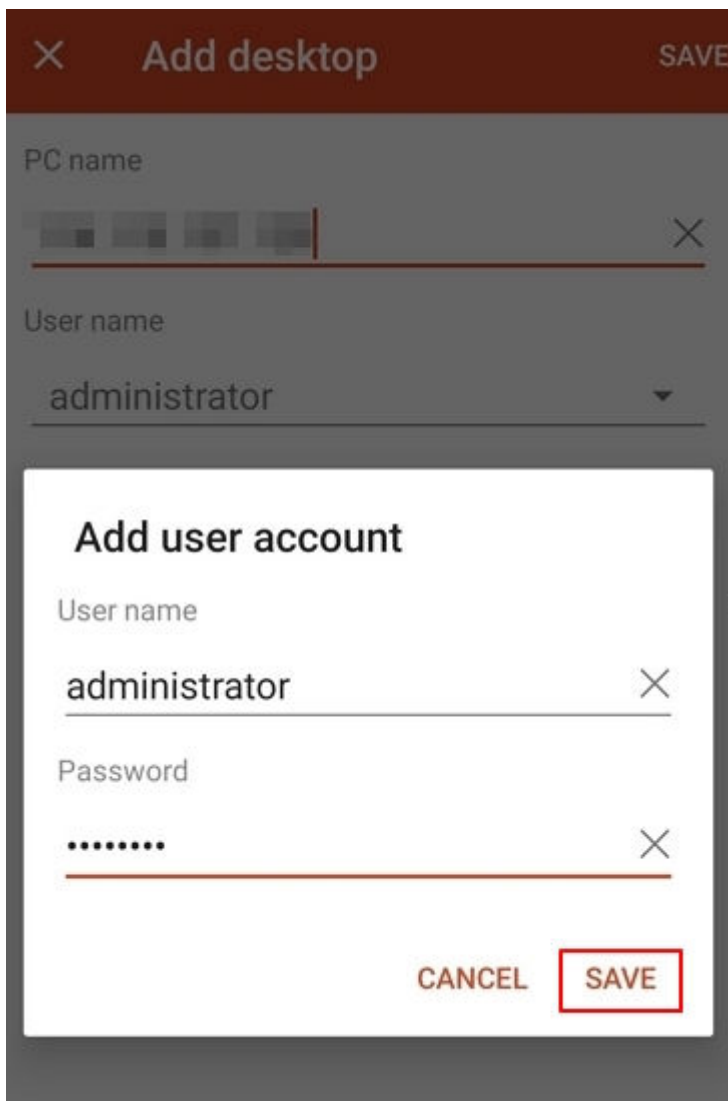
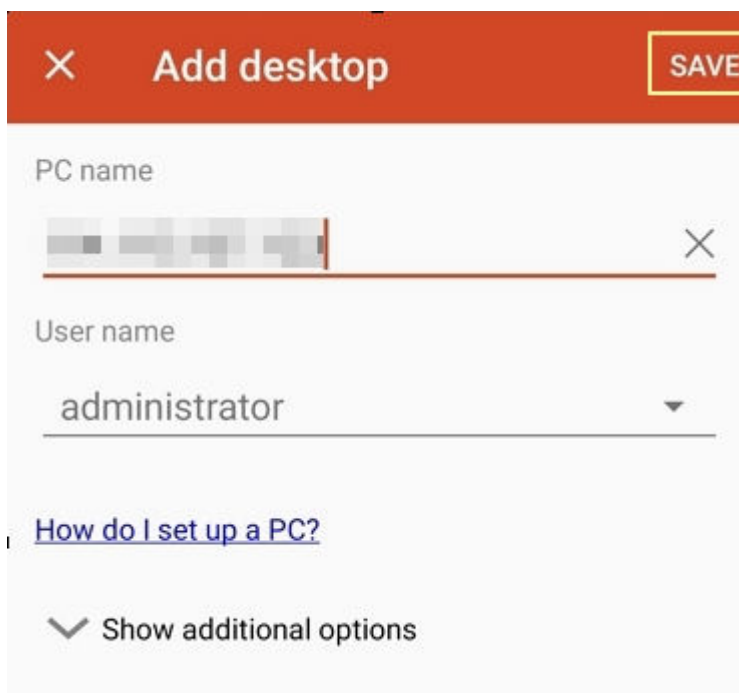




图 4-16 保存填写的登录信息



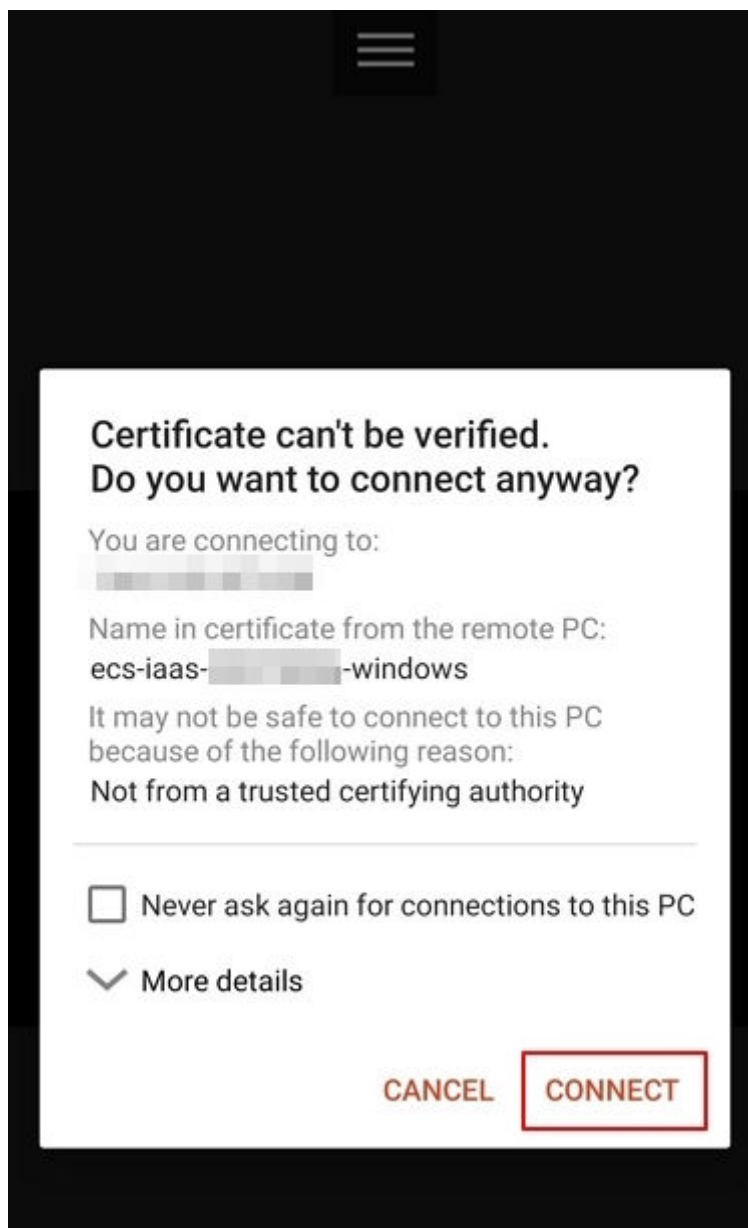
4. 在“Remote Desktop”页面，单击需要登录的Windows实例图标。

图 4-17 登录 Windows 实例



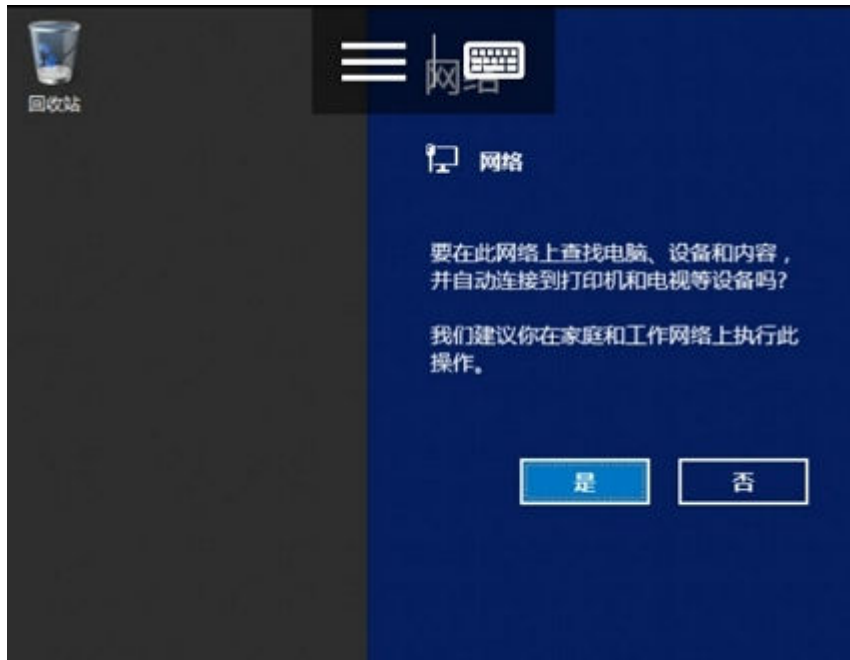
5. 确认信息后，单击“CONNECT”。

图 4-18 CONNECT



至此，您已经登录Windows实例。

图 4-19 登录成功



## 4.2 登录 Linux ECS

### 4.2.1 Linux ECS 登录方式概述

#### 约束与限制

- 只有运行中的弹性云服务器才允许用户登录。
- Linux操作系统用户名“root”。

#### 登录方式概述

请根据需要选择登录方式，登录云服务器。

表 4-3 Linux 云服务器登录方式一览

云服务器操作系统	本地主机操作系统	连接方法	条件
Linux	Windows	使用PuTTY、Xshell等远程登录工具： <ul style="list-style-type: none"><li>• 密码方式鉴权：<a href="#">SSH密码方式登录（本地使用Windows操作系统）</a>。</li></ul>	云服务器绑定弹性公网IP。

云服务器操作系统	本地主机操作系统	连接方法	条件
	Linux	使用命令连接： <ul style="list-style-type: none"><li>• 密码方式鉴权：<a href="#">SSH密码方式登录（本地使用Linux操作系统）</a>。</li></ul>	
	移动设备	使用Termius、JuiceSSH等SSH客户端工具登录云服务器。 <a href="#">通过移动设备登录Linux ECS</a> 。	
	macOS系统	使用系统自带的终端（Terminal）： <a href="#">通过macOS主机登录Linux ECS</a> 。	
	Windows	使用管理控制台远程登录方式： <a href="#">通过VNC登录Linux ECS</a> 。	不依赖弹性公网IP。

## 4.2.2 通过 VNC 登录 Linux ECS

### 操作场景

本节为您介绍如何通过控制台提供的远程登录功能（即VNC方式）登录到弹性云服务器上。

登录成功后，如需使用VNC界面提供的复制、粘贴功能，请参见[后续处理](#)。

### 约束与限制

- 当前提供的远程登录功能是通过系统配置的自定义端口进行访问的，所以在使用远程登录功能时，请确保需要使用的端口未被防火墙屏蔽。例如：远程登录的链接为“xxx:8002”，则需要确保端口8002没有被防火墙屏蔽。
- 如果客户端操作系统使用了本地代理，且用户无法配置该本地代理的防火墙端口，请关闭代理模式后再使用远程登录功能。

### 登录 Linux 弹性云服务器

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 获取弹性云服务器密码。  
VNC方式登录弹性云服务器时，需已知其密码，然后再采用VNC方式登录。
  - 当您的弹性云服务器是采用密码方式鉴权时，请直接使用创建云服务器时设置的密码进行登录。
4. 选择要登录的弹性云服务器，单击“操作”列下的“远程登录”。
5. （可选）如果界面提示“Press CTRL+ALT+DELETE to log on”，请单击远程登录操作面板上方的“Send CtrlAltDel”按钮进行登录。

### 说明

请勿使用物理键盘按“CTRL+ALT+DELETE”，该操作不生效。

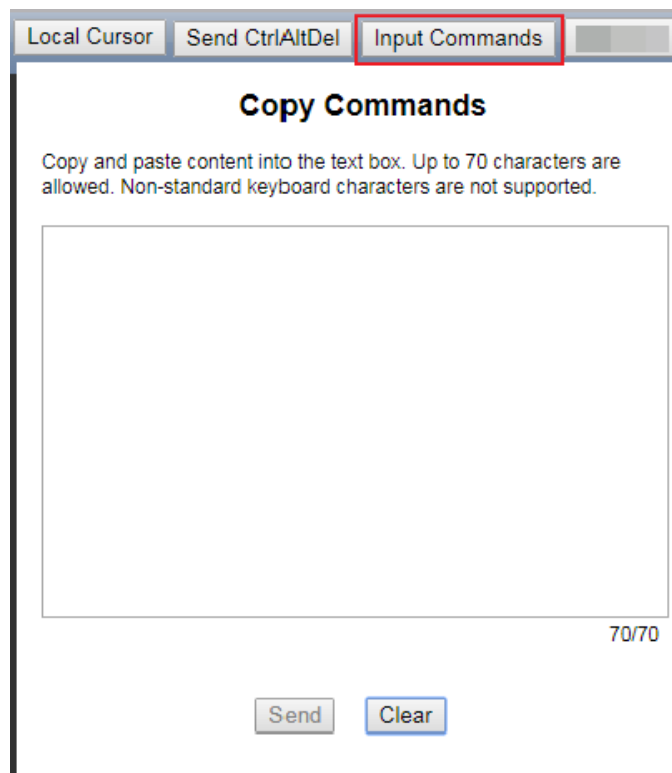
6. 根据界面提示，输入弹性云服务器密码。

## 后续处理

系统支持从本地拷贝命令行到弹性云服务器，实现本地数据与弹性云服务器之间的单向复制、粘贴功能，方法如下：

1. 使用VNC方式成功登录弹性云服务器。
2. 单击页面右上角的“Input Commands”。

图 4-20 Input Commands



3. 使用快捷键Ctrl+C，复制本地计算机的数据。
  4. 使用快捷键Ctrl+V，将本地数据粘贴至“复制粘贴”窗口。
  5. 单击“发送”。
- 将复制的数据发送至命令行窗口。

### 说明

对于使用图形化界面的Linux弹性云服务器，在使用VNC窗口提供的“复制粘贴”功能时，会小概率出现数据丢失的情况。这是由于弹性云服务器CPU核数过低，图形化界面占用较多CPU资源导致。此时，建议您每次发送的字符数不超过5个，或者从图形化界面切换至命令行界面（也称“文本界面”），然后再使用“复制粘贴”功能。

## 4.2.3 通过 SSH 密码方式登录 Linux ECS

### 操作场景

本节操作介绍在Windows和Linux环境中使用SSH密码方式远程登录Linux云服务器的操作步骤。

#### 须知

使用密码方式登录Linux弹性云服务器（SSH方式）的功能默认是被禁用的，用户如果有打开权限的需要，可以在登录弹性云服务器后自行设置。为避免安全风险，在打开SSH密码方式远程登录的权限后，请务必重置Linux弹性云服务器普通账户的密码。

### 前提条件

- 弹性云服务器状态为“运行中”。
- 弹性云服务器已经绑定弹性公网IP，绑定方式请参见[绑定弹性公网IP](#)。
- 所在安全组入方向已开放22端口，配置方式请参见[配置安全组规则](#)。
- 使用的登录工具（如PuTTY）与待登录的弹性云服务器之间网络连通。例如，默认的22端口没有被防火墙屏蔽。

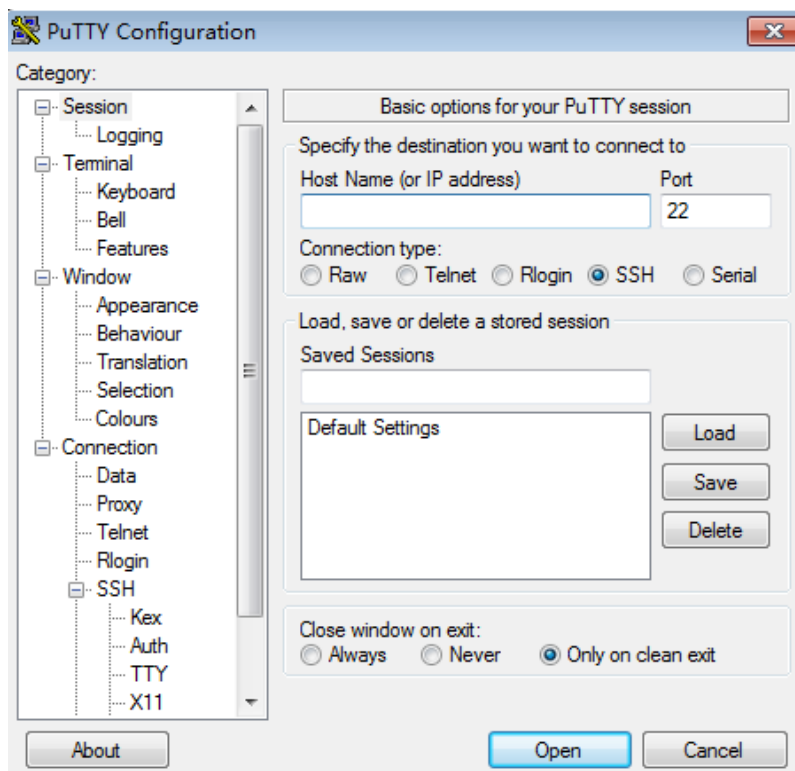
### 本地使用 Windows 操作系统

如果本地主机为Windows操作系统，可以按照下面方式登录云服务器。

下面步骤以PuTTY为例。

1. 在以下路径中下载PuTTY和PuTTYgen。  
<https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html>
2. 运行PuTTY。
3. 单击“Session”。
  - a. Host Name (or IP address): 输入云服务器的弹性公网IP。
  - b. Port: 输入 22。
  - c. Connection Type: 选择 SSH。
  - d. Saved Sessions: 任务名称，在下一次使用putty时就可以单击保存的任务名称，即可打开远程连接。

图 4-21 单击“Session”



- 单击“Window”，在“Translation”下的“Received data assumed to be in which character set:”选择“UTF-8”。
- 单击“Open”。  
如果首次登录服务器，PuTTY会显示安全警告对话框，询问是否接受服务器的安全证书。单击“是”将证书保存到本地注册表中。
- 建立到云服务器的SSH连接后，根据提示输入用户名和密码登录云服务器。

## 本地使用 Linux 操作系统

如果本地主机为Linux操作系统，您可以在计算机的命令行中通过以下操作登录弹性云服务器。

- 在您的linux计算机的命令行中执行如下命令，登录弹性云服务器。

```
ssh xx.xx.xx.xx
```

### 📖 说明

xx.xx.xx.xx表示弹性云服务器绑定的弹性公网IP。

- 根据界面提示信息，验证远程主机SSH指纹，并输入“yes”。  
The authenticity of host 'xx.xx.xx.xx (xx.xx.xx.xx)' can't be established.  
ECDSA key fingerprint is SHA256:rnKuzrUSYS03MCoaXXXXXXXXXXXXXXXXXXXXXXXXXXXX.  
ECDSA key fingerprint is MD5:cf:64:5b:5e:74:30:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx.  
Are you sure you want to continue connecting (yes/no)? **yes**  
Warning: Permanently added 'xx.xx.xx.xx' (ECDSA) to the list of known hosts.
- 根据界面提示，输入待登录的弹性云服务器的密码，完成登录操作。

## 4.2.4 通过 macOS 主机登录 Linux ECS

### 操作场景

本节为您介绍如何在macOS系统主机上登录Linux云服务器。

### 前提条件

- 云服务器状态为“运行中”。
- 已获取Linux云服务器用户名和密码。
- 弹性云服务器已经绑定弹性公网IP，绑定方式请参见[绑定弹性公网IP](#)。
- 所在安全组入方向已开放22端口，配置方式请参见[配置安全组规则](#)。

### 操作步骤

您可以通过macOS系统自带的终端（Terminal）登录Linux云服务器。

- SSH密码方式
  - a. 打开系统自带的终端（Terminal），执行以下命令，登录云服务器。  
`ssh 用户名@弹性公网IP`

## 4.2.5 通过移动设备登录 Linux ECS

### 操作场景

本节操作介绍如何在移动设备上连接Linux实例。

- 以Termius为例介绍如何在iOS设备上连接 Linux 实例，详细操作请参考[IOS设备上登录Linux云服务器](#)。
- 以JuiceSSH为例介绍如何在Android设备上连接 Linux 实例，详细操作请参考[Android设备上登录Linux云服务器](#)。

### 前提条件

- 云服务器状态为“运行中”。
- 已获取Linux云服务器用户名和密码。
- 弹性云服务器已经绑定弹性公网IP，绑定方式请参见[绑定弹性公网IP](#)。
- 所在安全组入方向已开放22端口，配置方式请参见[配置安全组规则](#)。

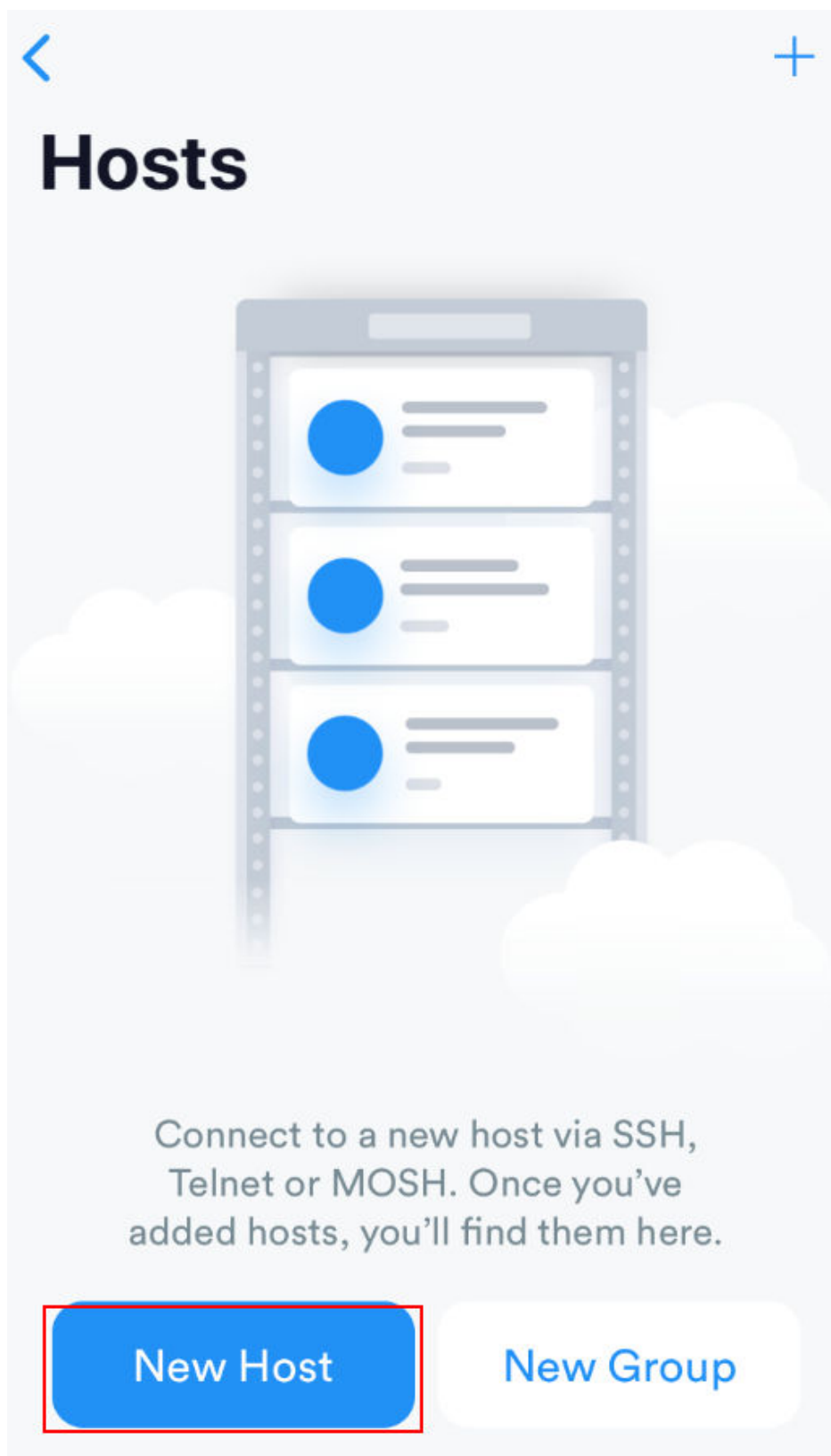
### IOS 设备上登录 Linux 云服务器

如果您使用iOS设备，请确保已经安装了SSH客户端工具，以Termius为例。本示例中使用CentOS 7.6操作系统，使用用户名和密码进行认证。

1. 启动Termius，单击New Host。



图 4-22 New Host



2. 在SSH页面上，输入连接信息后，单击 Save。需要输入的连接信息包括：
  - Alias: 指定Host名称，如本例中，设置为ecs01。

- Hostname: 输入需要连接的 Linux 实例的公网IP地址。
- Use SSH: 打开SSH登录配置。
- Host: 输入需要连接的Linux实例的公网IP地址。
- Port: 输入端口号22。
- 用户名: 输入用户名root。
- 密码: 输入实例登录密码。

图 4-23 输入连接信息

Cancel      New Host      Save

1 Alias

2 Hostname

Group >

Tags >

Backspace as CTRL+H

SSH / MOSH

3 Use SSH

Use Mosh (Beta)

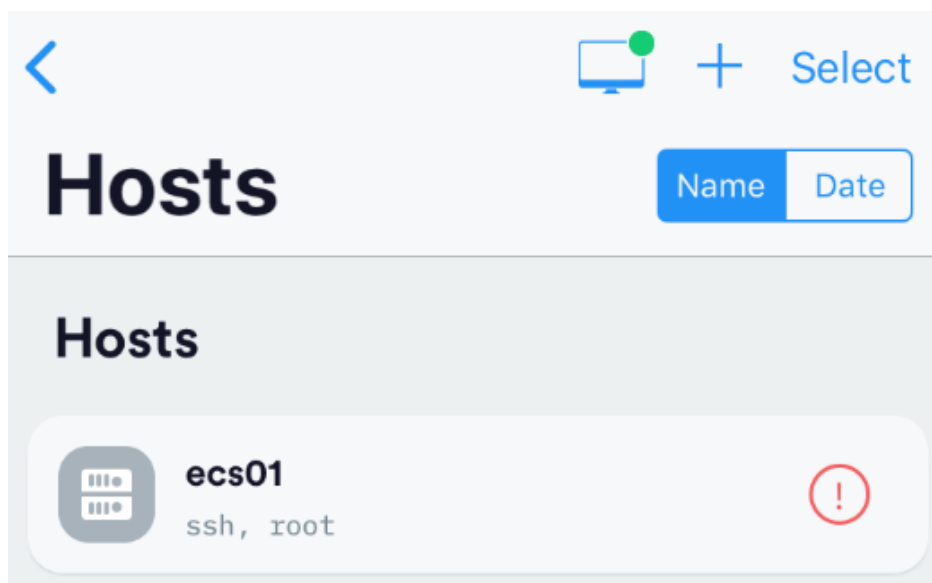
4 Port 22 Default

5 Username root

6 Password ●●●●●●●●

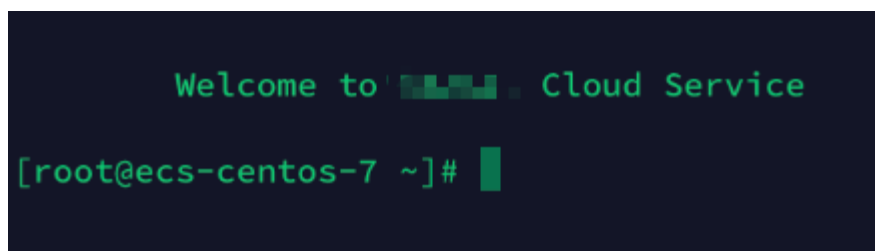
3. 单击右上角的“Save”，保存登录信息，在Hosts页面，单击连接的名称远程连接服务器。

图 4-24 保存登录信息



当出现如图4-25所示页面时，您已经成功地连接了Linux云服务器。

图 4-25 已连接



## Android 设备上登录 Linux 云服务器

如果您使用Android设备，请确保已经安装了JuiceSSH。本示例中使用CentOS 7.6操作系统，通过用户名和密码进行认证。

1. 启动JuiceSSH，单击“连接”。

图 4-26 启动 JuiceSSH




2. 在“连接”页面单击  图标。


图 4-27 连接



## 无连接

你还没有配置任何连接。点击下面的按钮开始。



3. 在“新建连接”页面上，添加基本设置和高级设置的信息并保存。需要添加的信息如下：
  - 昵称：指定登录会话的名称，如本例中，设置为“linux\_test”。
  - 类型：使用默认值“SSH”。
  - 地址：输入需要登录的Linux实例的弹性公网IP地址。
  - 按以下步骤设置认证：
    - i. 单击“认证”，在下拉列表里单击“新建”。
    - ii. 在“新建认证”页面上，添加如下信息后，单击图标。

- 昵称：可选项，您可以根据需要设置一个身份名称，方便后续管理。如本示例中，设置为“linux\_test”。
- 用户名：输入用户名“root”。
- 密码：单击“设置（可选）”后，输入实例的登录密码，单击“确定”。

图 4-28 新建认证

← 新建认证 ✓

认证信息

昵称： linux\_test

用户名： root

密码： 更新 / 清除

私钥： 设置（可选）

代码片段

JuiceSSH 高级版用户可以自动创建一个代码片段，该代码可用于添加公钥到服务器上的 `~/.ssh/authorized_keys` 文件并设置正确的权限。

生成代码片段

- 端口：输入端口号“22”。

图 4-29 端口号

← 新建连接 ✓

**基本设置**

昵称： linux\_test

类型： SSH

地址：

认证： linux\_test

**高级设置**

端口： 22

连接方式： (可选)

运行代码片段： (可选)

Backspace 模式： 默认发送 (DEL)

**连接组**

添加到组

4. 在“连接”页面单击创建好的连接。



图 4-30 单击连接



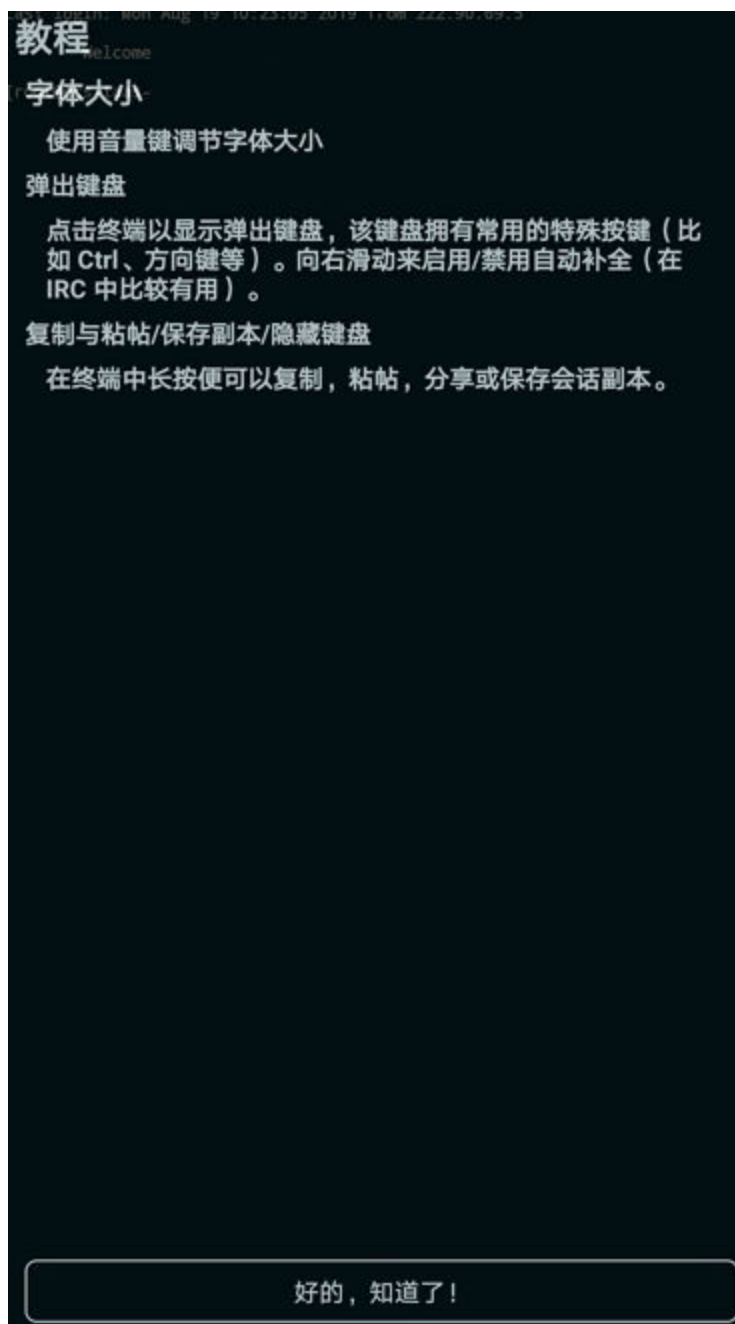
5. 确认提示信息后，单击“接受”。

图 4-31 确认提示信息



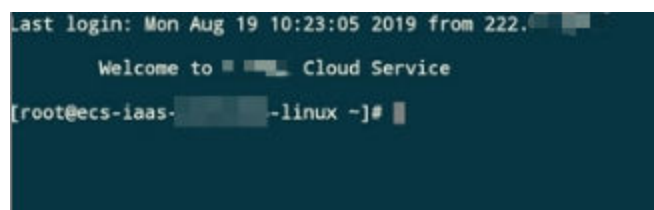
6. （可选）第一次连接时，JuiceSSH会提示您如何设置字体大小，如何弹出键盘等。确认信息后，单击“好的，我知道了！”。

图 4-32 教程



至此，您已经成功登录Linux实例。

图 4-33 登录 Linux 实例



## 4.3 管理 GPU 加速型 ECS 的 GPU 驱动

### 4.3.1 GPU 驱动概述

#### GPU 驱动概述

在使用GPU加速型实例前，请确保实例已安装GPU驱动以获得相应的GPU加速能力。

GPU加速型实例支持两种类型的驱动：GRID驱动和Tesla驱动。

- 如果需要使用OpenGL/DirectX/Vulkan等图形加速能力，则需要安装GRID驱动并自行购买和配置使用GRID License。此外，GRID驱动配合vDWS类型License，也支持CUDA，用来满足既需要计算加速也需要图形加速的场景。
  - 使用公共镜像创建的图形加速型（G系列）实例默认已安装特定版本的GRID驱动，但GRID License需自行配置使用，请提前确认GPU加速型实例是否已经预装或者预装版本是否符合需求。
  - 使用私有镜像创建的GPU加速型实例，如需安装GRID驱动请参考[手动安装GPU加速型ECS的GRID驱动](#)。
- 如果需要实现计算加速能力，则需要安装Tesla驱动。
  - 使用公共镜像创建的计算加速型（P系列）实例默认已安装特定版本的Tesla驱动。
  - 使用私有镜像创建的GPU加速型实例，如需安装Tesla驱动请参考[手动安装GPU加速型ECS的Tesla驱动](#)。

表 4-4 GPU 驱动支持的加速能力

驱动类型	License	CUDA	OpenGL	DirectX	Vulkan	典型应用场景	说明
GRID驱动	需要	支持	支持	支持	支持	3D渲染、图形工作站、游戏加速	满足图形图像类应用加速用途。
Tesla驱动	不需要	支持	不支持	不支持	不支持	科学计算、深度学习训练和推理	通常搭配使用NVIDIA CUDA SDK，可免费下载使用，满足通用计算类应用加速用途。

## 4.3.2 Tesla 驱动及 CUDA 工具包获取方式

### 操作场景

使用GPU加速型云服务器时，需确保已安装Tesla驱动和CUDA工具包，否则无法实现计算加速功能。本节内容提供Tesla驱动及CUDA工具包下载地址，请根据实例的类型，选择具体的驱动版本。

Tesla驱动及CUDA工具包安装操作指导请参考[手动安装GPU加速型ECS的Tesla驱动](#)。

### Tesla 驱动下载地址

请单击[NVIDIA驱动下载](#)，根据实例的类型，选择NVIDIA产品类型、产品系列和产品。

表 4-5 Tesla 驱动产品类型对应关系

实例类型	产品类型 ( Product Type )	产品系列 ( Product Series )	产品 ( Product )
Pi2	Tesla	T- Series	T4
Pi2nl	Tesla	T- Series	T4
G7	Tesla	A-Series	A40
G6	Tesla	T- Series	T4
G5	Tesla	V-Series	V100

### CUDA 工具包下载地址

请从[CUDA软件包下载](#)获取CUDA软件包，您需要根据实例类型和驱动版本，选择对应的CUDA Toolkit软件包产品。

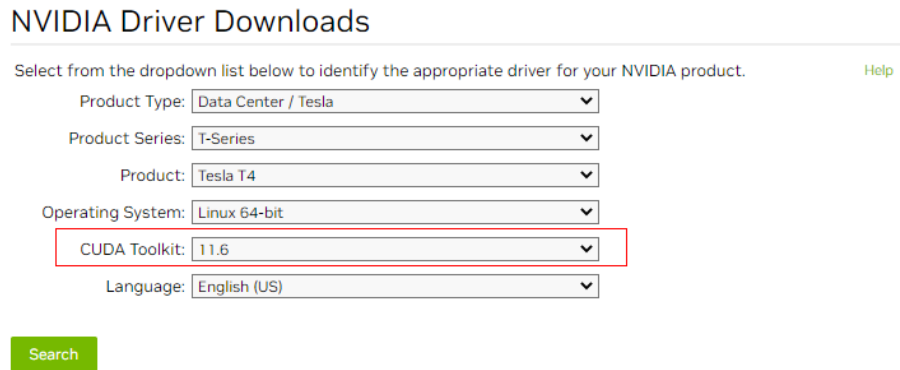
#### 说明

驱动版本与CUDA Toolkit版本存在对应关系，如二者版本不匹配，可能导致驱动无法使用。版本对应关系，请参见[NVIDIA驱动下载](#)。

下面以Tesla T4下载驱动软件包及CUDA Toolkit为例进行介绍。

1. Tesla T4安装驱动软件包时，选择Linux操作系统，并指定CUDA Toolkit软件版本为11.6。

图 4-34 指定 CUDA Toolkit 软件版本



2. 下载CUDA软件包，需要选择CUDA Toolkit 11.6对应小版本。

图 4-35 选择 CUDA Toolkit 对应版本

#### Archived Releases

- [CUDA Toolkit 11.7.1 \(August 2022\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.7.0 \(May 2022\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.6.2 \(March 2022\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.6.1 \(February 2022\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.6.0 \(January 2022\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.5.2 \(February 2022\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.5.1 \(November 2021\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.5.0 \(October 2021\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.4.4 \(February 2022\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.4.3 \(November 2021\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.4.2 \(September 2021\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.4.1 \(August 2021\), Versioned Online Documentation](#)
- [CUDA Toolkit 11.4.0 \(June 2021\), Versioned Online Documentation](#)

### 4.3.3 手动安装 GPU 加速型 ECS 的 GRID 驱动

#### 操作场景

GPU加速型实例如需使用OpenGL/DirectX/Vulkan等图形加速能力则需要安装GRID驱动并自行购买和配置使用GRID License。此外，GRID驱动配合vDWS类型License，也支持CUDA，用来满足既需要计算加速也需要图形加速的场景。

- 使用公共镜像创建的图形加速型（G系列）实例默认已安装特定版本的GRID驱动，但GRID License需自行配置使用。
- 使用私有镜像创建的GPU加速型弹性云服务器，则需要安装GRID驱动并自行配置使用GRID License。

本节操作介绍如何安装GRID驱动，申请GRID License，以及如何配置License服务器。

安装GRID驱动操作步骤：

1. [配置GRID License](#)

2. [下载GRID驱动及License软件包](#)
3. [部署和配置License Server](#)
4. [安装GRID驱动并配置License](#)

#### 📖 说明

- NVIDIA支持用户申请90天试用版License。
- 不同规格的GPU实例介绍和应用场景请参见[GPU加速型](#)。

## 配置 GRID License

- 配置正式版本License  
如果需要正式版本License，请联系NVIDIA或者所在国家/地区的NVIDIA代理商。
- 申请试用版License。  
打开[NVIDIA官方网站](#)，填写相关信息。  
注册账号和申请试用版License的注意事项请参见[NVIDIA官方帮助页](#)。

#### 📖 说明

试用版License的使用方法和正式版本的License一致，可以保留试用版账号激活正式版本的License，无需重新注册。试用版License有限期限为90天，账号过期将无法使用，请尽快配置正式版本。

图 4-36 申请试用版 License

**START YOUR 90-DAY TRIAL**

Please register with your corporate email address.  
Personal email addresses or extensions will not be approved.  
If already registered, [click here](#).  
If you need assistance, please review [FAQ](#).

\* First name  \* Last name   
\* Email address  \* Phone   
\* Company  \* Industry   
\* Job role  \* Location   
\* Street 1  Street 2   
\* City  \* State/Province   
\* Postal Code

\* Certified Server  \* NVIDIA GPUs   
Certified Server Other  \* VDI Hypervisor   
\* VDI Remoting Client  \* VDI Seats   
\* Primary Application

Send me the latest enterprise news, announcements, and more from NVIDIA. I can unsubscribe at any time.

\* Required Fields

By registering, you agree to [NVIDIA Account Terms and Conditions & Privacy Policy](#).

## 下载 GRID 驱动及 License 软件包

1. 请根据[表4-6](#)对应操作系统下载驱动安装包。  
了解更多GRID驱动信息请参考[NVIDIA vGPU驱动](#)。

 说明

GPU直通型实例：根据需求选择GRID驱动版本。

GPU虚拟化型实例：请严格按照下表选择合适的驱动版本下载使用。

表 4-6 GPU 实例类型支持的 GRID 驱动版本

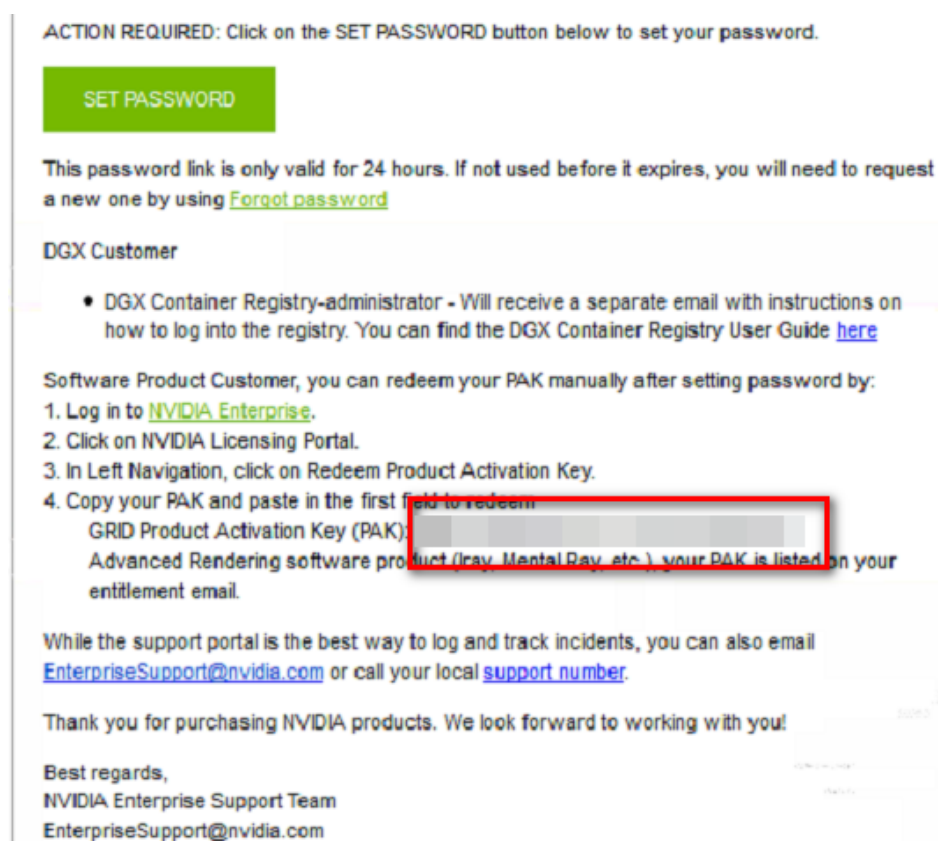
实例类型	GPU挂载方式	操作系统	驱动版本	CPU架构
G7v	GPU虚拟化型实例	<ul style="list-style-type: none"><li>CentOS 8.2 64bit</li><li>CentOS 7.6 64bit</li><li>Ubuntu 20.04 Server 64bit</li><li>Ubuntu 18.04 Server 64bit</li><li>Windows Server 2019 Standard 64bit</li><li>Windows Server 2016 Standard 64bit</li></ul>	GRID 13.0	x86_64
G7	GPU直通型实例	<ul style="list-style-type: none"><li>CentOS 8.2 64bit</li><li>CentOS 7.6 64bit</li><li>Ubuntu 20.04 Server 64bit</li><li>Ubuntu 18.04 Server 64bit</li><li>Windows Server 2019 Standard 64bit</li><li>Windows Server 2016 Standard 64bit</li></ul>	按需选择版本	x86_64
G6	GPU直通型实例	<ul style="list-style-type: none"><li>CentOS 8.2 64bit</li><li>CentOS 7.6 64bit</li><li>Ubuntu 20.04 64bit</li><li>Ubuntu 18.04 64bit</li><li>Windows Server 2019 Standard 64bit</li><li>Windows Server 2016 Standard 64bit</li></ul>	按需选择版本	x86_64



实例类型	GPU挂载方式	操作系统	驱动版本	CPU架构
G5.8xlarge.4	GPU直通型实例	<ul style="list-style-type: none"><li>CentOS 8.2 64bit</li><li>CentOS 7.6 64bit</li><li>CentOS 7.5 64bit</li><li>Ubuntu 20.04 64bit</li><li>Ubuntu 18.04 64bit</li><li>Windows Server 2019 Standard 64bit</li><li>Windows Server 2016 Standard 64bit</li><li>Windows Server 2019 Datacenter 64bit</li><li>Windows Server 2016 Datacenter 64bit</li></ul>	按需选择版本	x86_64
Pi2	GPU直通型实例	<ul style="list-style-type: none"><li>CentOS 7.5 64bit</li><li>Windows Server 2019 Standard 64bit</li><li>Windows Server 2016 Standard 64bit</li></ul>	按需选择版本	x86_64
Pi2nl	GPU直通型实例	<ul style="list-style-type: none"><li>CentOS 7.5 64bit</li><li>Ubuntu 16.04 Server 64bit</li><li>Windows Server 2016 Standard 64bit</li></ul>	按需选择版本	x86_64

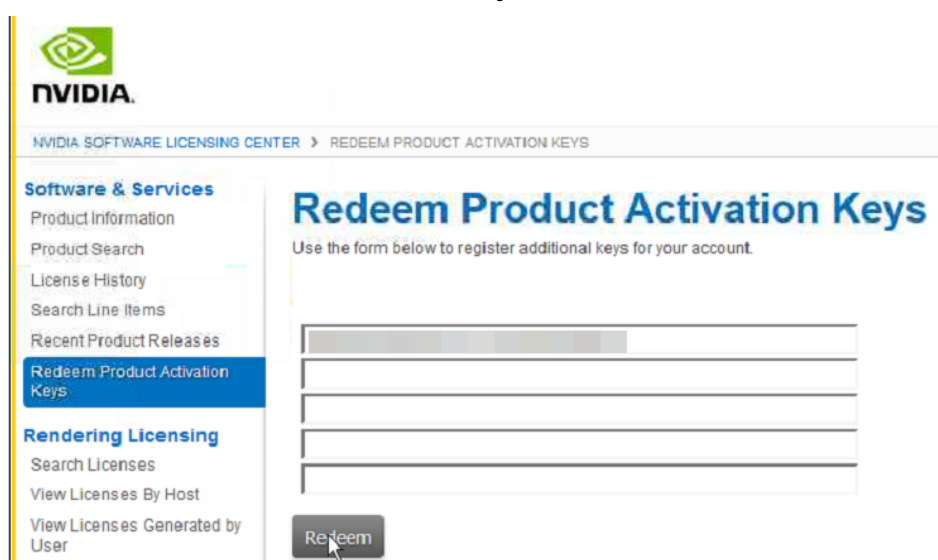
- 注册成功后，登录[NVIDIA官网](#)，填写账户信息。
- 判断是否为首次注册使用NVIDIA。
  - 是，执行4。
  - 否，执行6。
- 在注册NVIDIA用户成功的邮件中，查询PAK，如[图4-37](#)所示。

图 4-37 注册 PAK



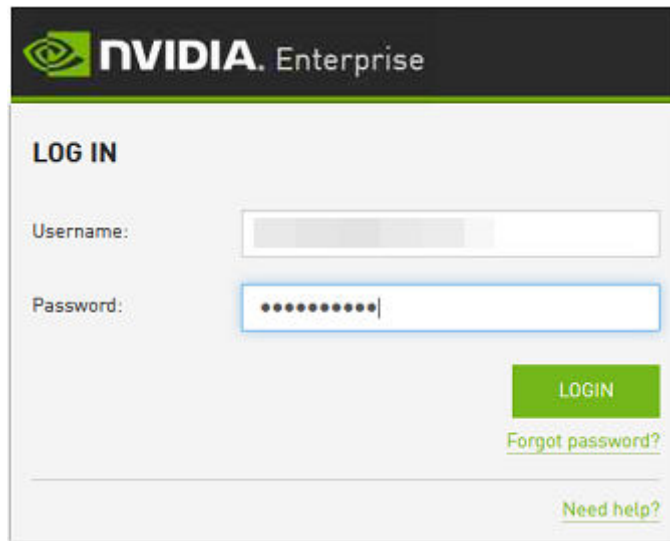
5. 将4中查找的PAK填写在“Redeem Product Activation Keys”页面，并单击“Redeem”。

图 4-38 Redeem Product Activation Keys



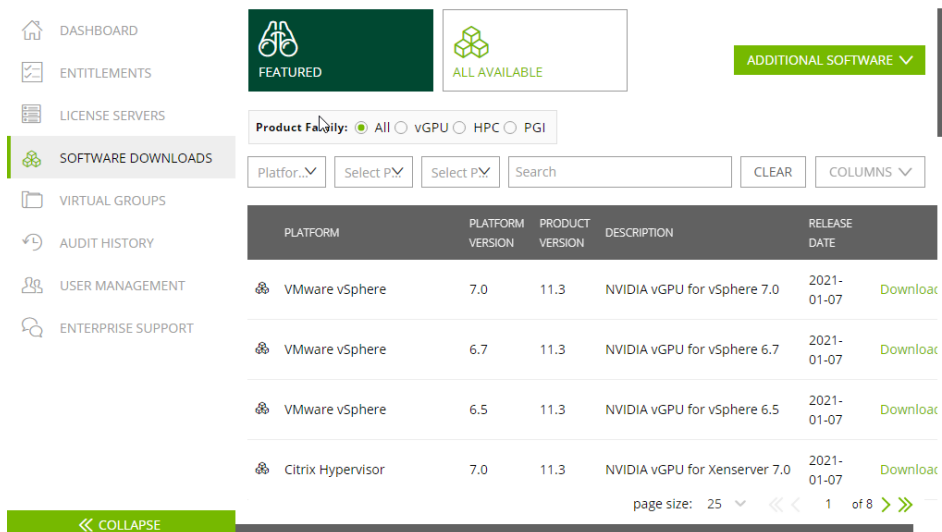
6. 输入“用户名”和“密码”，并单击“登录”。

图 4-39 登录 NVIDIA 官网



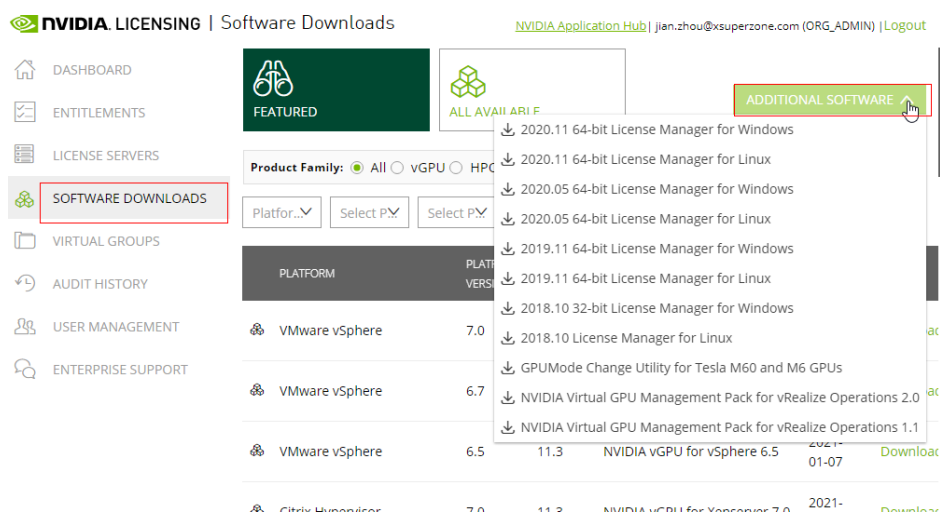
7. 根据界面提示，进入NVIDIA网站，并选择“SOFTWARE DOWNLOADS”。

图 4-40 打开 NVIDIA 网站



8. 对照表4-6选择相应版本的GRID驱动下载。
9. 解压缩GRID驱动包，并选择和弹性云服务器操作系统匹配的驱动进行安装。
10. 在“SOFTWARE DOWNLOADS”页面，单击“ADDITIONAL SOFTWARE”，下载License软件包。

图 4-41 选择 SOFTWARE DOWNLOADS



## 部署和配置 License Server

以CentOS 7.5操作系统的云服务器为例演示部署和配置License Server。

### 说明

- 云服务器规格不小于2vCPU，内存不小于4GiB。
- 请提前记录云服务器MAC地址。
- 如用作生产用途，建议采用高可用模式部署，主备高可用模式部署License Server 请参考 [NVIDIA官方License Server高可用部署文档](#)。

#### 1. 配置网络：

- 如使用VPC网络访问License Server：

- 单用户：请确保License Server和使用GRID驱动的GPU加速型实例处在同一个VPC子网内。
- 多用户：License server和使用已安装GRID驱动的GPU实例处于不同账号时，操作步骤如下：

假设License server的账号A，需要使用license的账号B，且账号B的GPU实例已安装GRID驱动。

- 账号A创建VPC终端节点服务，且“后端资源类型”选择License Server。
- 账号A将账号B的Domain ID填入VPC终端节点服务的白名单。
- 账号B创建VPC终端节点来连接VPC终端节点服务，完成后，即可使用节点IP或内网域名进行license授权。

具体操作请参考：《VPC终端节点用户指南》中“配置跨VPC通信的终端节点（不同账号）”章节。

- 如使用公网IP访问License Server：请配置License Server所在的安全组，增加方向规则：TCP 7070和TCP 8080。

#### 2. 安装License Server。

- 执行以下命令，解压缩安装包。其中“安装程序.zip”为10获取到的安装包名称。

### unzip 安装程序.zip

- b. 执行以下命令，为安装程序添加执行权限。

```
chmod +x setup.bin
```

- c. 以root用户运行安装程序。

```
sudo ./setup.bin -i console
```

- d. 在Introduction部分，单击回车键继续。

```
=====
Introduction
-----

InstallAnywhere will guide you through the installation of License Server.

It is strongly recommended that you quit all programs before continuing with
this installation.

Respond to each prompt to proceed to the next step in the installation.  If
you want to change something on a previous step, type 'back'.

You may cancel this installation at any time by typing 'quit'.

PRESS <ENTER> TO CONTINUE: █
```

- e. 在License Agreement部分，通过单击回车键进行翻页，翻页结束后接受许可协议。

当您达成许可协议时，系统会提示您接受许可协议条款，请输入“Y”，并单击回车键。

```
DO YOU ACCEPT THE TERMS OF THIS LICENSE AGREEMENT? (Y/N): Y █
```

- f. 在Choose Install Folder部分，单击回车键，保持默认的License Server软件安装路径。

- g. 在Choose Local Tomcat Server Path部分，输入Tomcat的本地路径，默认为/var/lib/tomcat版本号，例如：/var/lib/tomcat8。

- h. 在Choose Firewall Options部分，确认需要在防火墙中打开的端口，单击回车键，保持默认选项即可。

```
Choose Firewall Options
-----

The license server listens on port 7070. This port must be opened in the
firewall for other machines to obtain licenses from this server.

The license server's management interface listens on port 8080. Leave this
port closed to prevent unauthorized access to the management interface.

->1- License server (port 7070)
   2- Management interface (port 8080)

ENTER A COMMA-SEPARATED LIST OF NUMBERS REPRESENTING THE DESIRED CHOICES, OR
PRESS <ENTER> TO ACCEPT THE DEFAULT: █
```

- i. 在Pre-Installation Summary部分，确认信息并单击回车键启动安装。

```
Pre-Installation Summary
-----

Please Review the Following Before Continuing:

Product Name:
  License Server

Install Folder:
  /opt/flexnetls/nvidia

Link Folder:
  /root/NVIDIA Corporation/License Server

Disk Space Information (for Installation Target):
  Required:    105,216,774 Bytes
  Available:  35,501,248,512 Bytes

PRESS <ENTER> TO CONTINUE: █
```

- j. 在Install Complete部分，单击回车键，结束安装。

```
Install Complete
-----

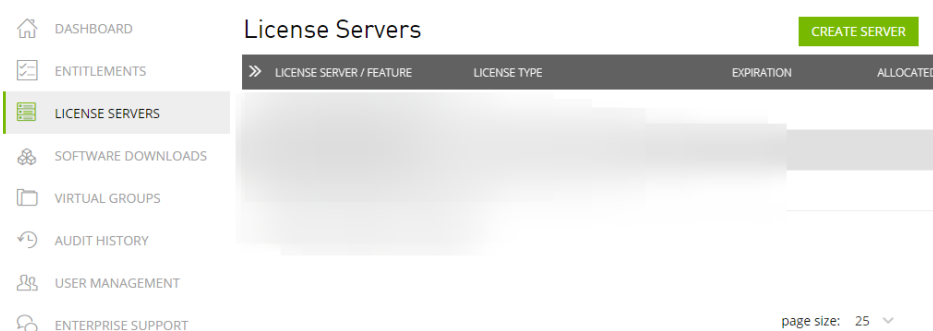
License Server has been successfully installed to:

  /opt/flexnetls/nvidia

PRESS <ENTER> TO EXIT THE INSTALLER:
```

- 3. 获取License文件。
  - a. 新建页签，登录[NVIDIA网站](#)，选择“LICENSE SERVERS”。

图 4-42 选择 LICENSE SERVERS



- b. 单击“CREATE SERVER”。
  - c. 在“Create License Server”界面，根据界面提示配置参数。

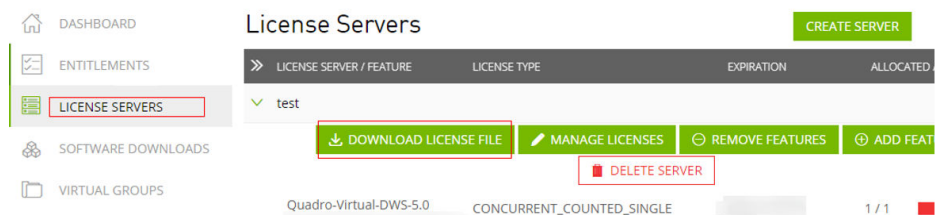
图 4-43 填写 LICENSE SERVERS 信息

表 4-7 LICENSE SERVERS 信息填写说明

参数	说明
Server Name	自定义需要的License Server名称。
Description	License Server的描述信息。
MAC Address	填写用于搭建License Server的ECS实例的MAC地址。 您可以登录实例，使用ipconfig -a命令进行查询。
Feature	在Licenses框中输入需要的license数目，单击“ADD”。 如果是主备部署的情况需要把备服务器的名称填入 Failover License Server，MAC地址填入 Failover MAC Address中。

- d. 单击“CREATE LICENSE SERVER”。
- e. 下载License文件。

图 4-44 下载 License 文件

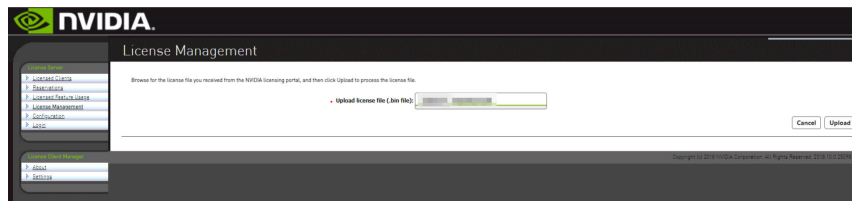


4. 在Web浏览器中，根据安装时配置的管理页面链接，访问License Server管理界面的主页。

默认访问链接为：`http://弹性公网IP地址:8080/licserver`。

5. 在左侧导航树中，单击“License Server > License Management”。
6. 使用License服务器配置菜单导入，并单击“Upload”上传许可证\*.bin文件，完成License Server的配置。

图 4-45 上传许可证文件



## 安装 GRID 驱动并配置 License

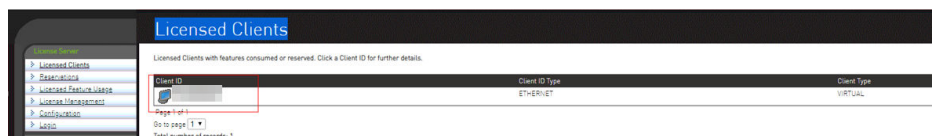
1. 以Windows操作系统GPU加速型实例为例，选择合适版本的GRID驱动进行安装。

### 说明

微软的远程登录协议不支持使用GPU的3D硬件加速能力，如需使用请安装VNC/PCoIP/NICE DCV等第三方桌面协议软件，并通过相应客户端连接GPU实例，使用GPU图形图像加速能力。

2. 使用第三方桌面协议连接后，在Windows控制面板中打开NVIDIA控制面板。
3. 在一级许可证服务器中填入部署的License server的IP和端口，并单击应用。当出现“您的系统已获GRID vGPU许可”则代表安装GRID驱动成功，并且可以在License Server管理控制台Licensed Clients中看到已安装GRID驱动并使用了License的GPU实例的MAC地址。

图 4-46 License Server 管理控制台



## 4.3.4 手动安装 GPU 加速型 ECS 的 Tesla 驱动

### 操作场景

GPU加速型云服务器，需要安装Tesla驱动和CUDA工具包以实现计算加速功能。

- 使用公共镜像创建的加速型（P系列）实例默认已安装特定版本的Tesla驱动。
- 使用私有镜像创建的GPU加速型云服务器，需在创建完成后安装Tesla驱动，否则无法实现计算加速功能。

本节操作介绍GPU云服务器安装Tesla驱动及CUDA工具包的操作步骤。

### 安装须知

- 云服务器已绑定弹性公网IP。
- 云服务器未安装Tesla驱动以及CUDA工具包。



### 📖 说明

- 从NVIDIA官网下载CUDA工具包进行安装，通常会同时自动安装一个和CUDA版本匹配的Tesla驱动，免去用户单独安装Tesla驱动的步骤。但是对NVIDIA驱动版本有强制要求或依赖时，请务必先单独从NVIDIA官网下载并安装匹配的Tesla驱动，然后再安装CUDA工具包。
- 如果云服务器已经安装了Tesla驱动，请检查当前驱动版本是否可用。如需安装新版本的驱动请卸载旧版本的Tesla驱动，避免因驱动程序冲突导致安装失败。

安装指引：

- [Tesla驱动及CUDA工具包获取方式](#)
- 安装NVIDIA驱动
  - [Linux操作系统云服务器安装Tesla驱动](#)
  - [Windows操作系统云服务器安装Tesla驱动](#)
- 安装CUDA驱动
  - [Linux操作系统安装CUDA工具包](#)
  - [Windows操作系统安装CUDA工具包](#)

## Linux 操作系统云服务器安装 Tesla 驱动

以下操作以Ubuntu 16.04 64bit操作系统，GPU实例安装CUDA 10.1对应的Tesla驱动为例。

### 📖 说明

Linux内核版本和驱动的版本存在兼容性关系。如果驱动安装失败，请检查驱动安装日志(安装日志通常位于/var/log/nvidia-installer.log)。日志中如提示失败原因为驱动编译错误，例如get\_user\_pages参数不匹配，说明当前内核版本与驱动版本不兼容，请选择合适的内核版本和驱动版本重新安装。建议内核版本与驱动版本发布时间不要相差过大。

1. 登录云服务器。
2. 根据操作系统选择命令更新系统软件。
  - Ubuntu  
更新软件安装源：`apt-get -y update`  
安装必要程序：`apt-get install gcc g++ make`
  - CentOS  
更新软件安装源：`yum -y update --exclude=kernel* --exclude=centos-release* --exclude=initscripts*`  
安装必要程序：`yum install -y kernel-devel-`uname -r` gcc gcc-c++`
3. 下载NVIDIA驱动包。  
单击[NVIDIA驱动下载](#)根据实例的类型，选择驱动版本。单击“SEARCH”。

图 4-47 选择 NVIDIA 驱动

### NVIDIA Driver Downloads

Advanced Driver Search

Product Type: Tesla	Operating System: Linux 64-bit
Product Series: [Blank]	CUDA Toolkit: 10.1
Product: [Blank]	Language: English (US)
	Recommended/Beta: All ?

SEARCH

4. 根据需求选择驱动版本，以下操作以选择Tesla 418.67为例。

图 4-48 选择驱动版本

### NVIDIA Driver Downloads

Advanced Driver Search

Product Type: Tesla	Operating System: Linux 64-bit
Product Series: [Blank]	CUDA Toolkit: 10.1
Product: [Blank]	Language: English (US)
	Recommended/Beta: All ?

SEARCH

Name	Version	Release Date	CUDA Toolkit
⊕ Tesla Driver for Linux x64	418.126.02	February 28, 2020	10.1
⊕ Tesla Driver for Linux x64	418.116.00	December 9, 2019	10.1
⊕ Tesla Driver for Linux x64	418.87.01	October 3, 2019	10.1
⊕ Tesla Driver for Linux x64	418.87.00	August 14, 2019	10.1
⊕ Tesla Driver for Linux x64	418.67	May 7, 2019	10.1
⊕ Tesla Driver for Linux x64	418.40.04	March 25, 2019	10.1
⊕ Tesla Driver for Linux x64	418.40.04	March 25, 2019	10.1

5. 单击需要下载的驱动，进入“TESLA DRIVER FOR LINUX X64”界面，单击“DOWNLOAD”。
6. 复制下载链接地址。

图 4-49 复制下载链接地址

### Download

By clicking the "Agree & Download" button below, you are confirming that you have read and agree to be bound by the [License For Customer Use of NVIDIA Software](#) for use of the driver. The driver will begin downloading immediately after clicking on the "Agree & Download" button below. NVIDIA recommends users update to the latest driver version. Please review [NVIDIA Product Security](#) for more information.

AGREE & DOWNLOAD      DECLINE

7. 在云服务器内部执行如下命令进行下载。

**wget** 复制的链接地址

例如：**wget http://us.download.nvidia.com/tesla/418.67/NVIDIA-Linux-x86\_64-418.67.run**

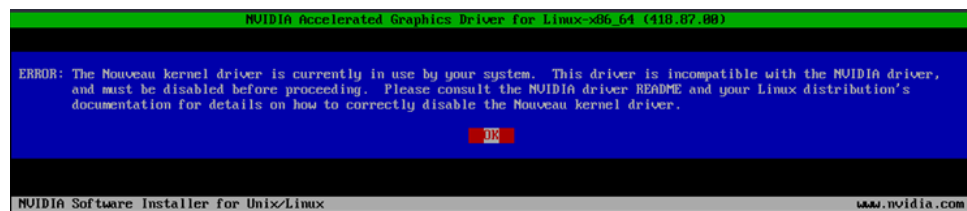
图 4-50 获取安装包

```
root@ecs-474b:~# wget http://us.download.nvidia.com/tesla/418.67/NVIDIA-Linux-x86_64-418.67.run
--2020-03-26 17:59:31-- http://us.download.nvidia.com/tesla/418.67/NVIDIA-Linux-x86_64-418.67.run
Resolving us.download.nvidia.com (us.download.nvidia.com)... 129.227.66.140, 129.227.66.139
Connecting to us.download.nvidia.com (us.download.nvidia.com)[129.227.66.140]:80... connected.
HTTP request sent, awaiting response... 301 Moved Permanently
Location: https://us.download.nvidia.cn/tesla/418.67/NVIDIA-Linux-x86_64-418.67.run [following]
--2020-03-26 17:59:34-- https://us.download.nvidia.cn/tesla/418.67/NVIDIA-Linux-x86_64-418.67.run
Resolving us.download.nvidia.cn (us.download.nvidia.cn)... 60.222.11.61, 60.222.11.11, 123.134.184.166, ...
Connecting to us.download.nvidia.cn (us.download.nvidia.cn)[60.222.11.61]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 107232512 (102M) [application/octet-stream]
Saving to: 'NVIDIA-Linux-x86_64-418.67.run'

NVIDIA-Linux-x86_64-418.67.run 100%[=====] 102.26M 1.07MB/s In 5m 2s
2020-03-26 18:04:40 (346 KB/s) - 'NVIDIA-Linux-x86_64-418.67.run' saved [107232512/107232512]
```

8. 执行以下命令安装驱动。  
**sh NVIDIA-Linux-x86\_64-418.67.run**
9. （可选）如果执行驱动安装命令后出现如下提示信息，需要禁用nouveau驱动。

图 4-51 禁用 nouveau 驱动



- a. 执行以下命令，查看是否安装Nouveau驱动。

**lsmod | grep nouveau**

- 如果回显信息中包含Nouveau驱动信息，说明Nouveau驱动已安装，则需要禁用Nouveau驱动，请执行步骤**9.b**。
- 如果回显信息没有Nouveau驱动信息，说明Nouveau驱动已被禁用，请执行步骤**10**。

- b. 执行如下命令编辑blacklist.conf文件。

如果没有“/etc/modprobe.d/blacklist.conf”文件，请新建一个。

**vi /etc/modprobe.d/blacklist.conf**

添加如下语句添加至文件结尾。

```
blacklist nouveau
options nouveau modeset=0
```

- c. 执行以下命令，备份并新建一个initramfs。

- Ubuntu系统：  
**sudo update-initramfs -u**
- CentOS系统：  
**mv /boot/initramfs-\$(uname -r).img /boot/initramfs-\$(uname -r).img.bak**

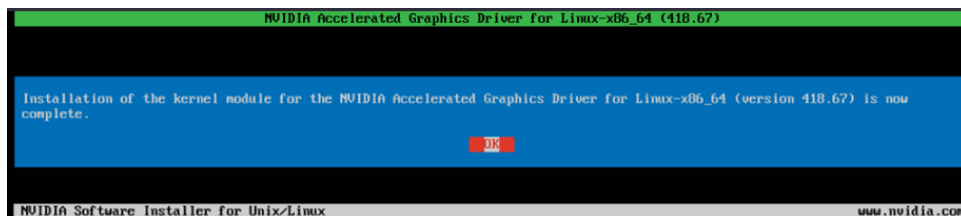
```
dracut -v /boot/initramfs-$(uname -r).img $(uname -r)
```

d. 执行以下命令，重启云服务器。

```
reboot
```

10. 根据安装提示，连续三次选择“OK”。完成驱动的安装。

图 4-52 NVIDIA 驱动安装完成



11. 执行命令设置systemd。

```
systemctl set-default multi-user.target
```

12. 执行reboot，重启云服务器。

13. 登录云服务器，执行nvidia-smi，如果回显信息中包含了已安装的驱动版本，说明驱动安装成功。

图 4-53 查看 NVIDIA 驱动的版本

```
root@ecs-474b:~# nvidia-smi
Thu Mar 26 20:05:17 2020

+-----+
| NVIDIA-SMI 418.67      Driver Version: 418.67      CUDA Version: 10.1      |
+-----+-----+
| GPU Name           Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf    Pur:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+-----+-----+-----+-----+
| 0   Tesla          Off          | 00000000:21:01.0 Off |                    |
| N/A    52C    P0     29W / 70W | 0MiB / 15079MiB | 0%      Default  |
+-----+-----+-----+-----+-----+-----+

+-----+
| Processes:
| GPU      PID   Type   Process name                      GPU Memory
|-----+-----+-----+-----+-----+
| No running processes found
+-----+

root@ecs-474b:~#
```

## Windows 操作系统云服务器安装 Tesla 驱动

以下操作以Windows Server 2016 Standard 64bit操作系统，GPU实例安装Tesla驱动为例。

1. 登录云服务器。
2. 下载NVIDIA驱动包。

单击[NVIDIA驱动下载](#)根据实例的类型，选择驱动版本。

图 4-54 选择驱动类型 ( Windows )

Advanced Driver Search

Product Type: Tesla

Operating System: Windows Server 2016

Product Series:

CUDA Toolkit: 10.1

Product:

Language: English (US)

Recommended/Beta: All ?

SEARCH

3. 根据需求选择驱动版本，本节操作以安装Tesla 425.25为例。

图 4-55 选择驱动版本 ( Windows )

Advanced Driver Search

Product Type: Tesla

Operating System: Windows Server 2016

Product Series:

CUDA Toolkit: 10.1

Product:

Language: English (US)

Recommended/Beta: All ?

SEARCH

Name	Version	Release Date	CUDA Toolkit
Tesla Driver for Windows WHQL	426.50	February 28, 2020	10.1
Tesla Driver for Windows WHQL	426.32	December 9, 2019	10.1
Tesla Driver for Windows WHQL	426.23	October 3, 2019	10.1
Tesla Driver for Windows WHQL	426.00	August 14, 2019	10.1
Tesla Driver for Windows WHQL	425.25	May 7, 2019	10.1
Tesla Driver for Windows WHQL	419.69	March 25, 2019	10.1

4. 单击需要下载的驱动，进入“TESLA DRIVER FOR WINDOWS”界面，单击“DOWNLOAD”。
5. 单击“Agree & Download”，下载安装包。

图 4-56 下载 NVIDIA 驱动安装包

### Download

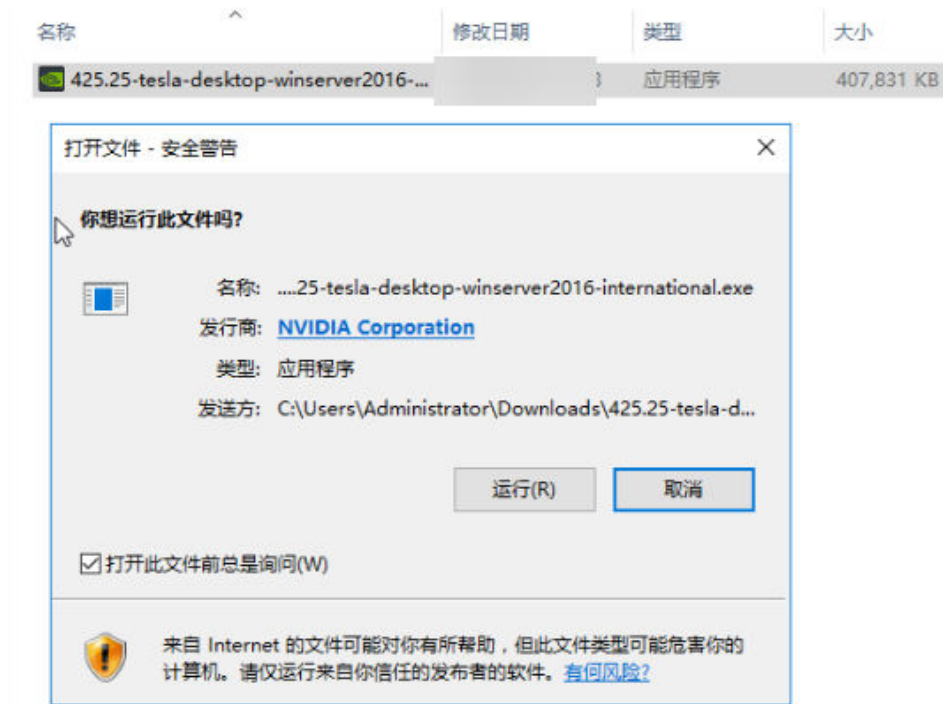
By clicking the "Agree & Download" button below, you are confirming that you have read and agree to be bound by the [License For Customer Use of NVIDIA Software](#) for use of the driver. The driver will begin downloading immediately after clicking on the "Agree & Download" button below. NVIDIA recommends users update to the latest driver version. Please review [NVIDIA Product Security](#) for more information.

AGREE & DOWNLOAD

DECLINE

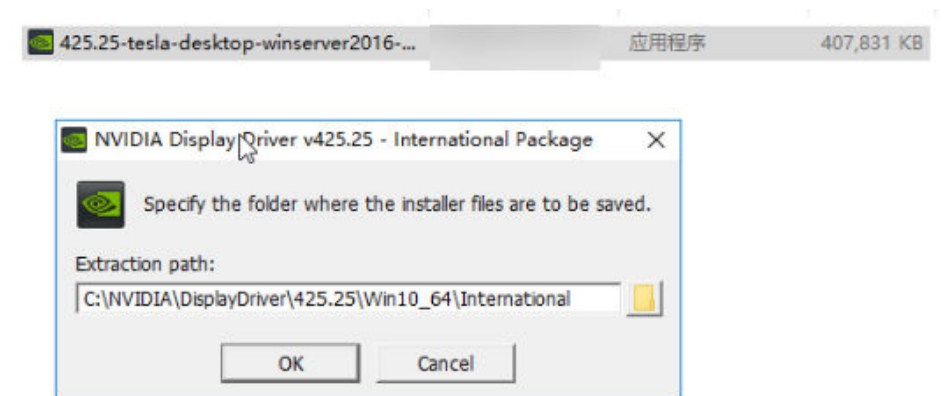
6. 双击驱动安装驱动，单击“运行”。

图 4-57 运行 NVIDIA 驱动安装程序



7. 选择安装路径，单击“OK”。

图 4-58 选择 NVIDIA 驱动安装路径



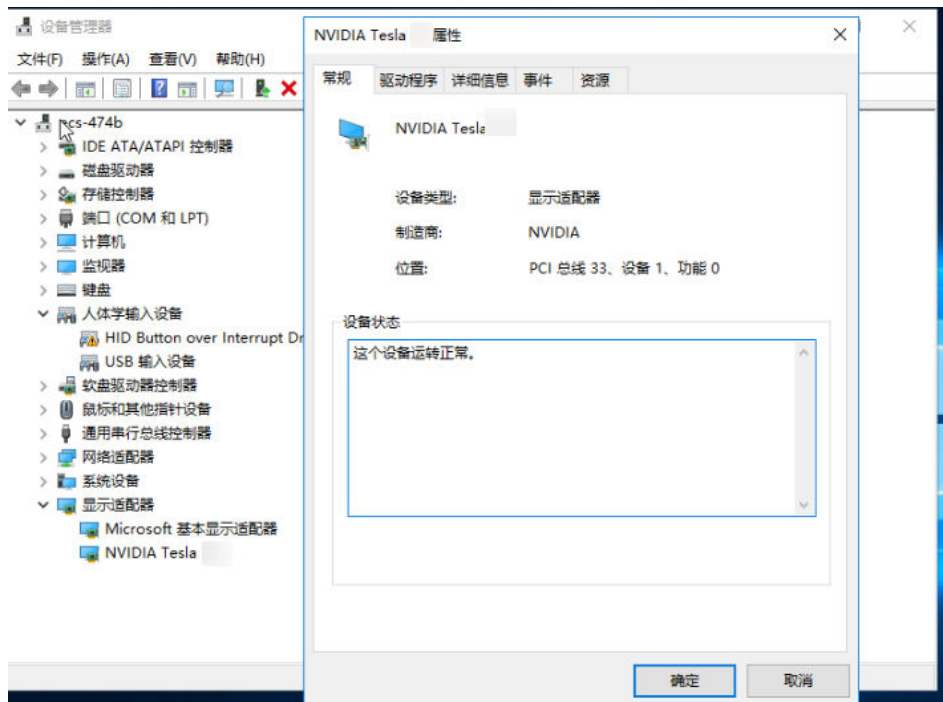
8. 根据安装提示完成NVIDIA程序的安装。

图 4-59 完成 NVIDIA 驱动的安装



9. 重启云服务器。
10. 检查驱动是否安装成功。
  - a. 进入“设备管理器”，查看“显示适配器”。

图 4-60 显示适配器



- b. 打开云服务器cmd窗口，执行以下命令：  
**cd C:\Program Files\NVIDIA Corporation\NVSMI**



## nvidia-smi

如果回显信息中包含了已安装的驱动版本，说明驱动安装成功。

图 4-61 查看 NVIDIA 驱动安装版本

```
C:\Program Files\NVIDIA Corporation\NVSMI>
C:\Program Files\NVIDIA Corporation\NVSMI>nvidia-smi
2020
-----+-----
NVIDIA-SMI 425.25      Driver Version: 425.25      CUDA Version: 10.1
-----+-----
GPU  Name          TCC/WDDM  Bus-Id      Disp.A   Volatile Uncorr. ECC
 Fan  Temp  Perf  Pwr:Usage/Cap     Memory-Usage  GPU-Util  Compute M.
-----+-----+-----+-----+-----+-----+-----+-----
   0   Tesla          TCC      00000000:21:01.0 Off      0MiB / 15205MiB      0%      Default
 N/A   27C    P8     9W / 70W
-----+-----+-----+-----+-----+-----+-----+-----

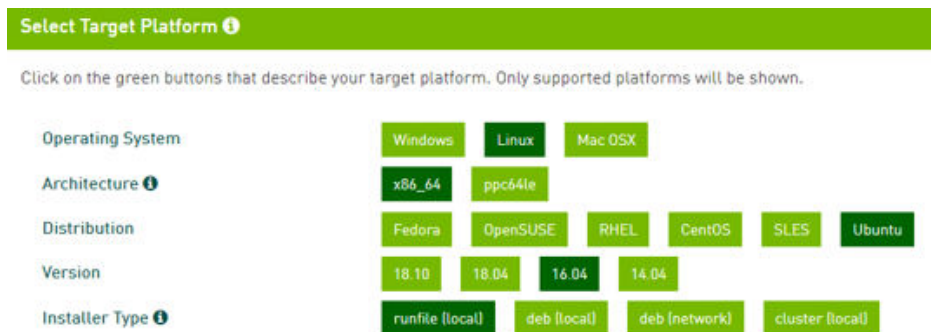
Processes:
GPU      PID  Type  Process name          GPU Memory
-----+-----+-----+-----+-----+-----
No running processes found
```

## Linux 操作系统安装 CUDA 工具包

以下操作以Ubuntu 16.04 64bit操作系统，GPU实例使用run包安装CUDA 10.1为例。

1. 登录云服务器。
2. 根据操作系统选择命令更新系统软件。
  - Ubuntu  
更新软件安装源：**apt-get -y update**  
安装必要程序：**apt-get install gcc g++ make**
  - CentOS  
更新软件安装源：**yum -y update --exclude=kernel\* --exclude=centos-release\* --exclude=initcripts\***  
安装必要程序：**yum install -y kernel-devel`uname -r` gcc gcc-c++**
3. 在CUDA下载页面中，按照[Tesla驱动及CUDA工具包获取方式](#)中的对应的索引项在页面中进行选择。

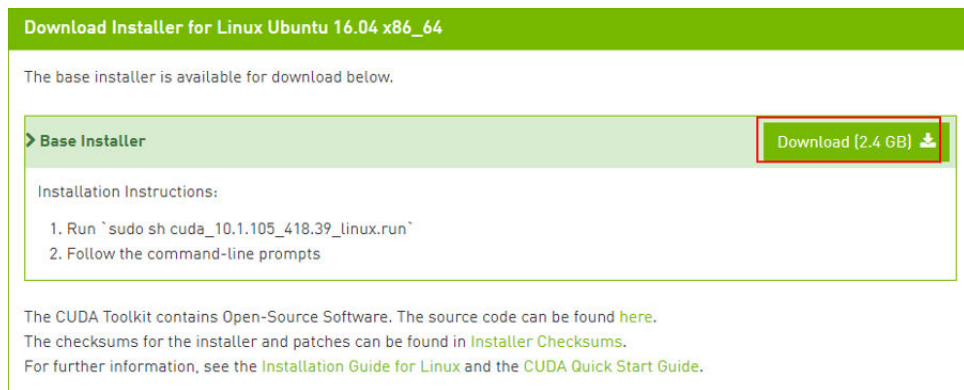
图 4-62 选择 CUDA 的版本



4. 选择完成后，页面会自动呈现出Ubuntu 16.04 64bit对应的CUDA 10.1的下载地址，复制下载地址。



图 4-63 复制 CUDA 下载地址



5. 在云服务器内部执行如下命令进行下载。

**wget** 复制的链接地址

例如：**wget https://developer.nvidia.com/compute/cuda/10.1/Prod/local\_installers/cuda\_10.1.105\_418.39\_linux.run**

图 4-64 Linux 云服务器下载 CUDA

```
root@ecs-474b:~# wget https://developer.nvidia.com/compute/cuda/10.1/Prod/local_installers/cuda_10.1.105_418.39_linux.run
-- https://developer.nvidia.com/compute/cuda/10.1/Prod/local_installers/cuda_10.1.105_418.39_linux.run
Resolving developer.nvidia.com (developer.nvidia.com)... 129.227.66.139, 129.227.66.140
Connecting to developer.nvidia.com (developer.nvidia.com)|129.227.66.139|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://developer.download.nvidia.com/compute/cuda/10.1/secure/Prod/local_installers/cuda_10.1.105_418.39_linux.run?FFJqELqU0UEH-9_Y0p4zpYc3t1_j8q9J5clzHL2kX5Uo-f721Gy8tP3hm2Jd_j2QntuNupR-7nmp4Eb6Q_jd7YB4BhuW56D14U3prth0y2ZSx6Xugu950k94i6Sb9he5nISJHhbPXUgULzUhhijXhujBUcKMBHsvvxm4X_uk_k2AY1oIbU7D2sMaF8oU8 [following]
-- https://developer.download.nvidia.com/compute/cuda/10.1/secure/Prod/local_installers/cuda_10.1.105_418.39_linux.run?FFJqELqU0UEH-9_Y0p4zpYc3t1_j8q9J5clzHL2kX5Uo-f721Gy8tP3hm2Jd_j2QntuNupR-7nmp4Eb6Q_jd7YB4BhuW56D14U3prth0y2ZSx6Xugu950k94i6Sb9he5nISJHhbPXUgULzUhhijXhujBUcKMBHsvvxm4X_uk_k2AY1oIbU7D2sMaF8oU8 [following]
Resolving developer.download.nvidia.com (developer.download.nvidia.com)... 129.227.66.139, 129.227.66.140
Connecting to developer.download.nvidia.com (developer.download.nvidia.com)|129.227.66.139|:443... connected.
HTTP request sent, awaiting response... 301 Moved Permanently
Location: https://developer.download.nvidia.cn/compute/cuda/10.1/secure/Prod/local_installers/cuda_10.1.105_418.39_linux.run?FFJqELqU0UEH-9_Y0p4zpYc3t1_j8q9J5clzHL2kX5Uo-f721Gy8tP3hm2Jd_j2QntuNupR-7nmp4Eb6Q_jd7YB4BhuW56D14U3prth0y2ZSx6Xugu950k94i6Sb9he5nISJHhbPXUgULzUhhijXhujBUcKMBHsvvxm4X_uk_k2AY1oIbU7D2sMaF8oU8 [following]
-- https://developer.download.nvidia.cn/compute/cuda/10.1/secure/Prod/local_installers/cuda_10.1.105_418.39_linux.run?FFJqELqU0UEH-9_Y0p4zpYc3t1_j8q9J5clzHL2kX5Uo-f721Gy8tP3hm2Jd_j2QntuNupR-7nmp4Eb6Q_jd7YB4BhuW56D14U3prth0y2ZSx6Xugu950k94i6Sb9he5nISJHhbPXUgULzUhhijXhujBUcKMBHsvvxm4X_uk_k2AY1oIbU7D2sMaF8oU8 [following]
Resolving developer.download.nvidia.cn (developer.download.nvidia.cn)... 60.222.11.61, 60.222.11.10, 123.134.184.165, ...
Connecting to developer.download.nvidia.cn (developer.download.nvidia.cn)|60.222.11.61|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2423314285 (2.36) [application/octet-stream]
Saving to: 'cuda_10.1.105_418.39_linux.run'

cuda_10.1.105_418.39_linux.run 100%[=====] 2.26G 34.2MB/s in 74s
```

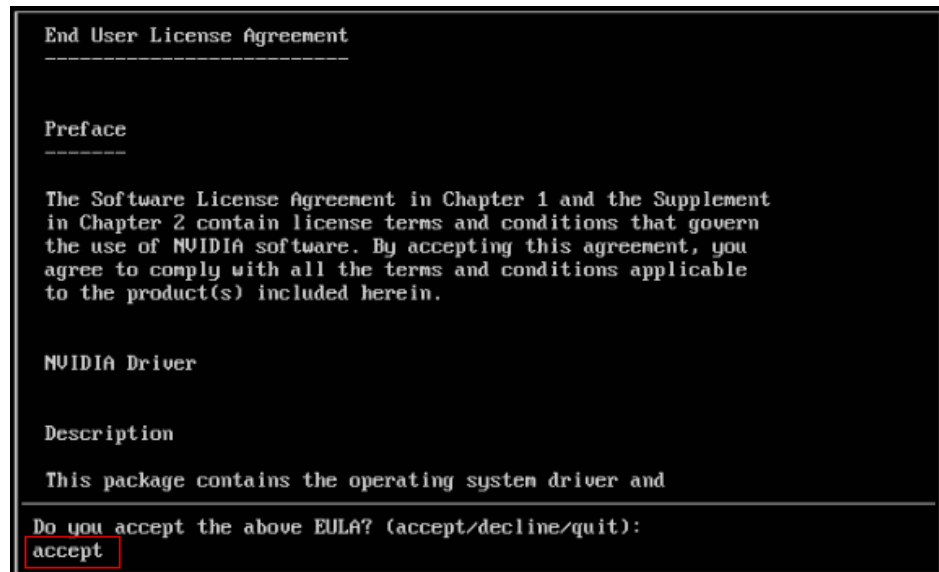
6. CUDA工具包下载完成后，按照NVIDIA官网安装指引进行安装。  
按照如下图的NVIDIA官网的Installation Instructions进行安装。

图 4-65 Linux 云服务器安装 CUDA



7. 执行以下命令安装CUDA。  
**sh cuda\_10.1.243\_418.87.00\_linux.run**
8. 在安装界面输入“accept”，回车进入下一步。

图 4-66 安装 CUDA\_1



9. 选择 “Install”，回车进入安装过程。

图 4-67 安装 CUDA\_2

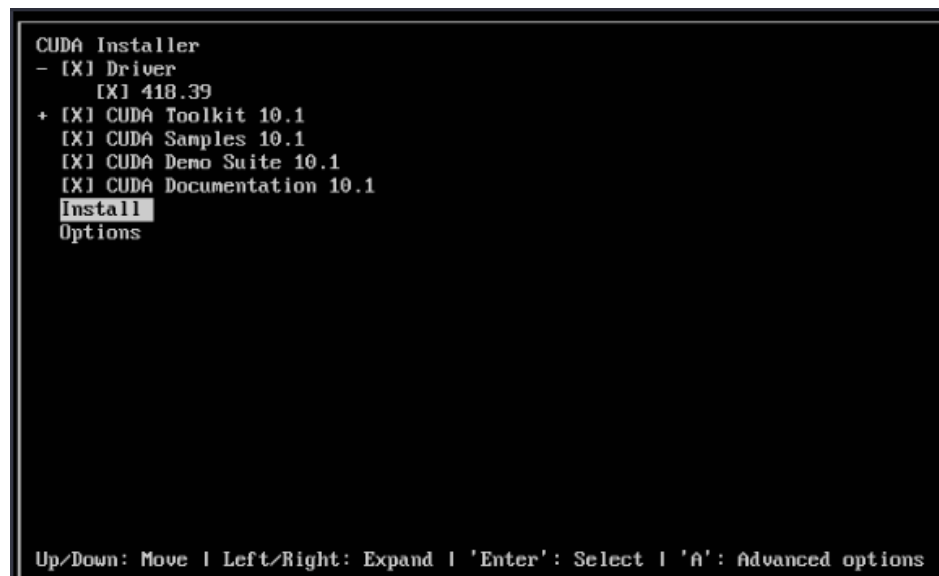
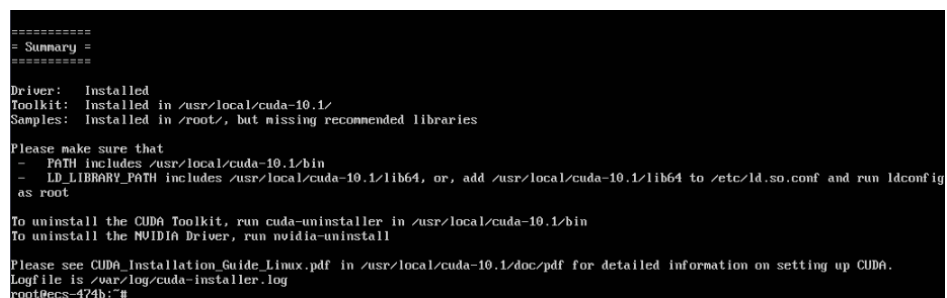


图 4-68 安装 CUDA 完成



10. 执行以下命令进入 /usr/local/cuda-10.1/samples/1\_Uilities/deviceQuery 目录下。

```
cd /usr/local/cuda-10.1/samples/1_Utilities/deviceQuery
```

11. 执行make命令，自动编译deviceQuery程序。
12. 执行以下命令查看CUDA是否安装成功。

```
./deviceQuery
```

如回显信息中有CUDA的版本信息证明CUDA安装成功。

图 4-69 deviceQuery 回显信息

```
root@ecs-474b:/usr/local/cuda-10.1/samples/1_Utilities/deviceQuery# ./deviceQuery
./deviceQuery Starting...

CUDA Device Query (Runtime API) version (CUDART static linking)

Detected 1 CUDA Capable device(s)

Device 0: "Tesla "
  CUDA Driver Version / Runtime Version      10.1 / 10.1
  CUDA Capability Major/Minor version number: 7.5
  Total amount of global memory:             15080 MBytes (15812263936 bytes)
  (40) Multiprocessors, ( 64) CUDA Cores/MP: 2560 CUDA Cores
  GPU Max Clock rate:                       1590 MHz (1.59 GHz)
  Memory Clock rate:                        5001 Mhz
  Memory Bus Width:                         256-bit
  L2 Cache Size:                            4194304 bytes
  Maximum Texture Dimension Size (x,y,z)    1D=(131072), 2D=(131072, 65536), 3D=(16384, 16384, 16384)
  Maximum Layered 1D Texture Size, (num) layers 1D=(32768), 2048 layers
  Maximum Layered 2D Texture Size, (num) layers 2D=(32768, 32768), 2048 layers
  Total amount of constant memory:          65536 bytes
  Total amount of shared memory per block:   49152 bytes
  Total number of registers available per block: 65536
  Warp size:                                32
  Maximum number of threads per multiprocessor: 1024
  Maximum number of threads per block:      1024
  Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
  Max dimension size of a grid size (x,y,z): (2147483647, 65535, 65535)
  Maximum memory pitch:                     2147483647 bytes
  Texture alignment:                        512 bytes
  Concurrent copy and kernel execution:     Yes with 3 copy engine(s)
  Run time limit on kernels:                No
  Integrated GPU sharing Host Memory:       No
  Support host page-locked memory mapping: Yes
  Alignment requirement for Surfaces:       Yes
  Device has ECC support:                   Enabled
  Device supports Unified Addressing (UVA): Yes
  Device supports Compute Preemption:       Yes
  Supports Cooperative Kernel Launch:      Yes
  Supports MultiDevice Co-op Kernel Launch: Yes
  Device PCI Domain ID / Bus ID / location ID: 0 / 33 / 1
  Compute Mode:
    < Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 10.1, CUDA Runtime Version = 10.1, NumDevs = 1
Result = PASS
root@ecs-474b:/usr/local/cuda-10.1/samples/1_Utilities/deviceQuery#
```

13. 执行以下命令查看CUDA的版本。

```
/usr/local/cuda/bin/nvcc -V
```

图 4-70 查看 CUDA 的版本

```
root@ecs-474b:/usr/local/cuda-10.1/samples/1_Utilities/deviceQuery# ./deviceQuery
[root@ecs-474b deviceQuery]# /usr/local/cuda/bin/nvcc -V
nvcc: NVIDIA (R) Cuda compilation driver
Copyright (c) 2005-2019 NVIDIA Corporation
Built on Fri_Feb__8_19:08:17_PST_2019
Cuda compilation tools, release 10.1, V10.1.105
[root@ecs-474b deviceQuery]#
```

14. 执行以下命令启动持久模式。

```
sudo nvidia-smi -pm 1
```

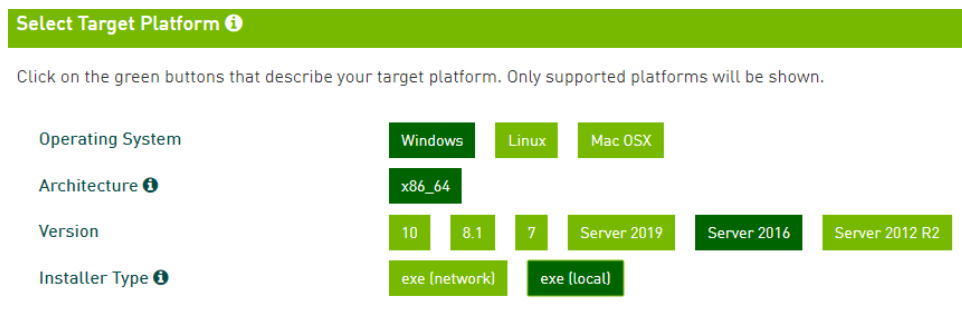
启用持久模式设置可以优化Linux实例上GPU设备的性能。

## Windows 操作系统安装 CUDA 工具包

以下操作以Windows Server 2016 Standard 64bit操作系统GPU实例安装CUDA 10.1为例。

1. 登录云服务器。
2. 在CUDA下载页面中，按照[CUDA工具包下载地址](#)中的对应的索引项在页面中进行选择。

图 4-71 选择 CUDA 版本



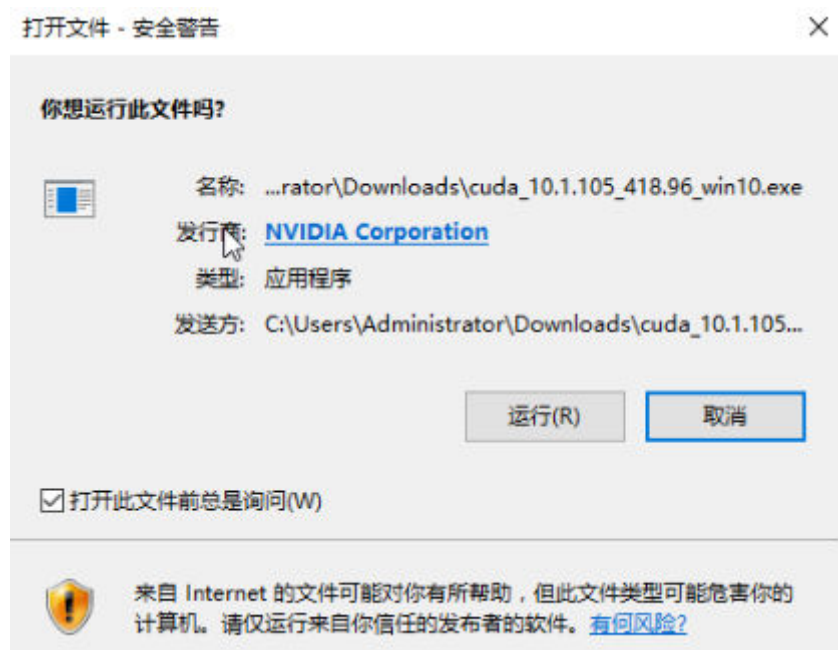
3. 选择完成后，页面会自动呈现出Windows Server 2016 Standard 64bit对应的CUDA 10.1的下载地址。

图 4-72 Windows 云服务器下载 CUDA



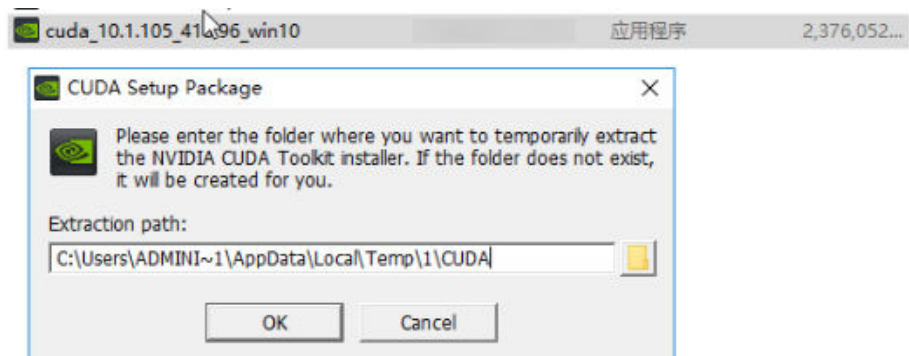
4. 单击“Download”下载CUDA工具包。
5. 双击打开安装文件，单击“运行”安装CUDA工具包。

图 4-73 Windows 云服务器安装 CUDA



6. 选择安装地址，在“CUDA Setup Package”界面，单击“OK”

图 4-74 选择 CUDA 安装路径



7. 根据安装提示完成CUDA的安装。

图 4-75 CUDA 安装完成



8. 检查CUDA是否安装成功。  
打开cmd命令窗口，执行以下命令。

```
nvcc -V
```

如果回显信息中出现CUDA的版本信息，说明CUDA安装成功。

图 4-76 CUDA 安装成功

```
C:\Users\Administrator>nvcc -V
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2019 NVIDIA Corporation
Built on Fri_Feb__8_19:08:26_Pacific_Standard_Time_2019
Cuda compilation tools, release 10.1, V10.1.105

C:\Users\Administrator>_
```

## 4.4 管理 ECS 配置信息

### 4.4.1 更改时区

#### 操作场景

弹性云服务器默认设置的时区，是您制作镜像时选择的时区。如需修改，请参见本节内容，将其更改为所需的本地时间或网络中的其他时区。

弹性云服务器登录成功后，如果发现弹性云服务器系统时间与本地时间不一致，建议更改时区，将弹性云服务器系统时间与本地时间进行同步。

## 更改 Linux 弹性服务器的时区

针对不同的Linux操作系统，更改时区的操作略有不同。本节以CentOS 6.x 64bit为例，介绍更改Linux弹性云服务器时区的操作方法。

1. 登录弹性云服务器。
2. 执行以下命令，切换至root账号。

```
su - root
```

3. 执行以下命令，查询支持使用的时区。

```
ls /usr/share/zoneinfo/
```

其中，“/usr/share/zoneinfo”目录显示了时区数据文件的目录结构。您可以通过该目录结构，查找您所需时区的文件。

“/usr/share/zoneinfo”目录中显示的信息，部分为时区，部分为目录。其中，目录包含了针对特定城市的时区文件，您可以在该目录查找要用于弹性云服务器的所在城市时区。

4. 设置新时区。
  - a. 执行以下命令，打开“/etc/sysconfig/clock”文件。
5. 按“ESC”，执行以下命令，保存并退出文件。
6. 执行以下命令，查询弹性云服务器中是否已经存在“/etc/localtime”文件。

```
vim /etc/sysconfig/clock
```

- b. 查找ZONE条目，将其更改为所需的时区文件名称。

```
:wq
```

7. 执行以下命令，删除已有的“/etc/localtime”文件。
8. 执行以下命令，在“/etc/localtime”与时区文件之间创建一个符号链接，使得弹性云服务器在引用本地时间信息时找到此时区文件。

```
ls /etc/localtime
```

- 是，执行7。
- 否，跳转8。

```
rm /etc/localtime
```

```
ln -sf /usr/share/zoneinfo/Asia/city1 /etc/localtime
```

9. 执行以下命令，重启系统，使得所有服务和应用程序接受新时区信息。

```
reboot
```

10. 重新登录弹性云服务器，使用root账号执行以下命令，查询弹性云服务器的时区，确认是否更改成功。

```
ls -lh /etc/localtime
```

回显信息如下所示：

```
# ls -lh /etc/localtime
lrwxrwxrwx 1 root root 33 Nov 27 11:01 /etc/localtime -> /usr/share/zoneinfo/Asia/city1
```

## 更改 Windows 弹性服务器的时区

1. 登录弹性云服务器。
2. 左键单击任务栏右下方的时间，选择“更改日期和时间设置”。  
系统进入“日期和时间”页面



图 4-77 日期和时间



3. 单击“更改时区”。  
系统进入“时区设置”页面。
4. 在“设置时区”栏的下拉框中选择待更换的时区。
5. 单击“确定”，完成Windows弹性云服务器的时区更换。

## 4.4.2 使用实例自定义数据和元数据

### 4.4.2.1 通过元数据获取 ECS 基本信息

#### 操作场景

弹性云服务器元数据包含了弹性云服务器在云平台的基本信息，例如云服务ID、主机名、网络信息等。弹性云服务器元数据支持两种风格，可以分别通过兼容Openstack和兼容EC2的API获取，如表4-8所示。对于支持的元数据类型，本节详细介绍了其URI和使用方法。

#### 使用须知

如果元数据中包含了敏感数据，您应当采取适当的措施来保护敏感数据，比如限制访问范围、加密等。



防火墙配置示例如下所示：

- Windows

如果您不希望管理员之外的用户访问实例自定义数据，请以管理员身份开启防火墙后，在PowerShell 中执行下面的命令，阻止非管理员用户访问实例自定义数据：

```
PS C:\>$RejectPrincipal = New-Object -TypeName
System.Security.Principal.NTAccount ("Everyone")
PS C:\>$RejectPrincipalSID =
$RejectPrincipal.Translate([System.Security.Principal.SecurityIdentifier]).V
alue
PS C:\>$ExceptPrincipal = New-Object -TypeName
System.Security.Principal.NTAccount ("Administrator")
PS C:\>$ExceptPrincipalSID =
$ExceptPrincipal.Translate([System.Security.Principal.SecurityIdentifier]).
Value
PS C:\>$PrincipalSDDL = "O:LSD:(D;;CC;;;$ExceptPrincipalSID)
(A;;CC;;;$RejectPrincipalSID)"
PS C:\>New-NetFirewallRule -DisplayName "Reject metadata service for $
($RejectPrincipal.Value), exception: $($ExceptPrincipal.Value)" -Action
block -Direction out -Protocol TCP -RemoteAddress 169.254.169.254 -
LocalUser $PrincipalSDDL
```

- Linux

如果您不希望root之外的用户访问自定义实例自定义数据，请以root用户权限执行下面的命令，阻止非root用户访问自定义实例自定义数据：

```
iptables --append OUTPUT --proto tcp --destination 169.254.169.254 --
match owner ! --uid-owner root --jump REJECT
```

## 弹性云服务器元数据类别支持列表

表4-8中不包含兼容EC2类型的ami-id、ami-launch-index、ami-manifest-path、block-device-mapping/、instance-action、instance-id、reservation-id、ramdisk-id、kernel-id元数据项，以上元数据项均无实际含义，不建议使用。

表 4-8 弹性云服务器元数据类别支持列表

元数据类型	实例元数据项	约束说明
OpenStack类型	/meta_data.json	查询弹性云服务器的元数据信息。 元数据的关键字段请参见表4-9。
OpenStack类型	/password	查询弹性云服务器的密码。 Windows系统使用keypairs创建弹性云服务器初始化时cloudbase-init用于保存密文密码。

元数据类型	实例元数据项	约束说明
OpenStack类型	/user_data	查询弹性云服务器的自定义数据。 用户根据需要自行指定脚本和配置文件用于弹性云服务器初始化，详细操作请参考 <a href="#">通过实例自定义数据配置ECS实例</a> 。 若linux虚拟化使用密码方式，则保存注入密码的脚本。
OpenStack类型	/network_data.json	查询弹性云服务器的网络信息。
OpenStack类型	/securitykey	获取临时的AK、SK。 对弹性云服务器获取临时的AK、SK，需要在IAM上对op_svc_ecs账户进行委托授权，并对相应的弹性云服务器资源进行授权委托管理。 <b>说明</b> 委托的授权权限由用户自行设置，建议根据实际业务需要进行最小化授权。 ECS服务本身不会使用委托对资源进行操作。
兼容EC2类型	/meta-data/ hostname	查询弹性云服务器的主机名称。 请参考以下链接为弹性云服务器去掉后缀.novalocal: <a href="#">弹性云服务器的主机名带后缀novalocal, 这正常吗?</a>
兼容EC2类型	/meta-data/ local-hostname	该字段含义同hostname一致。
兼容EC2类型	/meta-data/ public-hostname	该字段含义同hostname一致。
兼容EC2类型	/meta-data/ instance-type	查询弹性云服务器的规格名称。
兼容EC2类型	/meta-data/ local-ipv4	查询弹性云服务器的固定IP地址。 多网卡情况下，只显示主网卡的地址。
兼容EC2类型	/meta-data/ placement/ availability-zone	查询弹性云服务器的AZ信息。
兼容EC2类型	/meta-data/ public-ipv4	查询弹性云服务器的弹性公网IP地址。 多网卡情况下，只显示主网卡的弹性公网IP地址。
兼容EC2类型	/meta-data/ public-keys/0/ openssh-key	查询弹性云服务器的公钥。
兼容EC2类型	/user-data	查询弹性云服务器的自定义数据。

元数据类型	实例元数据项	约束说明
兼容EC2类型	/meta-data/ security-groups	查询弹性云服务器所使用的安全组名称。

表 4-9 metadata 关键字段

参数	参数类型	描述
uuid	String	弹性云服务器的ID。
availability_zone	String	弹性云服务器所在可用区。
meta	Dict	元数据信息，包括镜像名称、镜像ID、VPC ID等信息。
hostname	String	弹性云服务器主机名。 请参考以下链接为弹性云服务器去掉后缀.novalocal: <a href="#">弹性云服务器的主机名带后缀novalocal, 这正常吗?</a>

## 前提条件

- 已登录弹性云服务器。
- 请确保安全组出方向规则满足如下要求，否则访问元数据请求失败：
  - 协议：TCP
  - 端口范围：80
  - 远端地址：169.254.0.0/16

### 📖 说明

如果您使用的是默认安全组出方向规则，则已经包括了如上要求，可以正常访问元数据。默认安全组出方向规则请参见[默认安全组和规则](#)。

## Metadata ( OpenStack 元数据 API )

用于查询弹性云服务器的元数据。

- URI  
/169.254.169.254/openstack/latest/meta\_data.json
- 方法  
支持GET请求。
- 示例：  
Linux操作系统：以使用curl工具为例，介绍查询弹性云服务器元数据的方法。  
**curl http://169.254.169.254/openstack/latest/meta\_data.json**  
Windows操作系统：以使用Invoke-RestMethod工具为例，介绍查询弹性云服务器元数据的方法。



## 📖 说明

如果创建弹性云服务器时未注入自定义数据，此时，该接口的查询结果是404，如图4-78所示。

图 4-78 404 Not Found

```
root@pythonsdktempest-1519783681-1:~# curl http://169.254.169.254/openstack/latest/user_data
<html>
<head>
<title>404 Not Found</title>
</head>
<body>
<h1>404 Not Found</h1>
The resource could not be found.<br /><br />
</body>
</html>
```

## Network data ( OpenStack 元数据 API )

查询弹性云服务器的网络信息，支持查询云服务器挂载的全部网卡的信息，包括网卡所在子网的DNS地址、网络带宽、网卡ID、网卡私有IP地址、网卡弹性公网IP地址、网卡的MAC地址。

- URI  
/openstack/latest/network\_data.json
- 方法  
支持GET请求
- 示例

## 📖 说明

instance\_max\_bandwidth、instance\_min\_bandwidth的单位是Mbps，如果取值为-1，说明对带宽大小无限制。

Linux操作系统：

**curl http://169.254.169.254/openstack/latest/network\_data.json**

Windows操作系统：

**Invoke-RestMethod http://169.254.169.254/openstack/latest/network\_data.json | ConvertTo-Json**

```
{
  "services": [
    {
      "type": "dns",
      "address": "xxx.xx.x.x"
    },
    {
      "type": "dns",
      "address": "100.125.21.250"
    }
  ],
  "qos": {
    "instance_min_bandwidth": 100,
    "instance_max_bandwidth": 500
  },
  "networks": [
    {
      "network_id": "67dc10ce-441f-4592-9a80-cc709f6436e7",
      "type": "ipv4_dhcp",
      "link": "tap68a9272d-71",
      "id": "network0"
    }
  ],
  "links": [
    {
      "vif_id": "68a9272d-7152-4ae7-a138-3ef53af669e7",
      "ethernet_mac_address": "fa:16:3e:f7:c1:47",
      "mtu": null,
    }
  ]
}
```

```
"type": "cascading",  
"id": "tap68a9272d-71"  
  }  
}
```

## Security Key ( OpenStack 元数据 API )

获取临时的AK、SK。

### 📖 说明

- 如果您需要在弹性云服务器获取临时的AK、SK，需要在IAM对弹性云服务器进行云服务委托授权，并在相应的弹性云服务器的详情页面“管理信息 > 委托”栏目中添加委托。
- 临时的AK、SK有效期一个小时。临时AK、SK更新时间早于失效时间10分钟，10分钟内新旧临时AK、SK均可用。
- 使用临时AK、SK时，需要在消息的header中增加'X-Security-Token':{securitytoken}。其中，securitytoken就是调用接口返回的值。

- URI

/openstack/latest/securitykey

- 方法

支持GET请求

- 示例

Linux操作系统：

**curl http://169.254.169.254/openstack/latest/securitykey**

Windows操作系统：

**Invoke-RestMethod http://169.254.169.254/openstack/latest/securitykey**

## User data ( EC2-兼容的 API )

用于查询弹性云服务器的自定义。该值仅在创建弹性云服务器时指定，不支持修改。

- URI

/169.254.169.254/latest/user-data

- 方法

支持GET请求。

- 示例

Linux操作系统：

**curl http://169.254.169.254/latest/user-data**

Windows操作系统：

**Invoke-RestMethod http://169.254.169.254/latest/user-data**

```
ICAgICAgDQoiQSBjbG91ZCBkb2VzIG5vdCBrbm93IHdoeSBpdCBtb3ZlcyBpbjBqdXN0IHN1Y2ggYSBkaXJlY  
3Rpb24gYW5kIGF0IHN1Y2ggYSBzcGVlZC4uLkI0IGZlZWxzIGFuIGltcHVsc2lubi4uLnRoaXMgaXMgdGhllH  
BsYWNIHRvIGdvlG5vdy4gQnV0IHRobzSBza3kga25vd3MgdGhllHlYXNvbnMgYW5kIHRobzSBwYXR0ZXJ  
cyBiZWphbmQgYWxsIGNsb3VkcycgYW5kIHlvdSB3aWxsIGtub3csIHRobvbygd2h1biB5b3UgbGlmdCB5b3  
Vyc2VsZiBoaWdoIGVub3VnaCB0byBzZWUgYmV5b25kIGhvcml6b25zLiINCg0KLVJpY2hhcmQgQmFjaA=  
=
```

## Hostname ( EC2-兼容的 API )

用于查询弹性云服务器的主机名称，后面会追加.novalocal后缀。

- URI  
/169.254.169.254/latest/meta-data/hostname
- 方法  
支持GET请求。
- 示例  
Linux操作系统：  
**curl http://169.254.169.254/latest/meta-data/hostname**  
Windows操作系统：  
**Invoke-RestMethod http://169.254.169.254/latest/meta-data/hostname**  
vm-test.novalocal

## Instance Type ( EC2-兼容的 API )

用于查询弹性云服务器的规格名称。

- URI  
/169.254.169.254/latest/meta-data/instance-type
- 方法  
支持GET请求。
- 示例  
Linux操作系统：  
**curl http://169.254.169.254/latest/meta-data/instance-type**  
Windows操作系统：  
**Invoke-RestMethod http://169.254.169.254/latest/meta-data/instance-type**  
s3.medium.2

## Local IPv4 ( EC2-兼容的 API )

用于查询弹性云服务器的固定IP地址。多网卡情况下，只显示主网卡的地址。

- URI  
/169.254.169.254/latest/meta-data/local-ipv4
- 方法  
支持GET请求。
- 示例  
Linux操作系统：  
**curl http://169.254.169.254/latest/meta-data/local-ipv4**  
Windows操作系统：  
**Invoke-RestMethod http://169.254.169.254/latest/meta-data/local-ipv4**  
192.1.1.2

## Availability Zone ( EC2-兼容的 API )

用于查询弹性云服务器的AZ信息。

- URI  
/169.254.169.254/latest/meta-data/placement/availability-zone
- 方法  
支持GET请求。
- 示例  
Linux操作系统：  
**curl http://169.254.169.254/latest/meta-data/placement/availability-zone**  
Windows操作系统：  
**Invoke-RestMethod http://169.254.169.254/latest/meta-data/placement/availability-zone**  
az1.dc1

## Public IPv4 ( EC2-兼容的 API )

用于查询弹性云服务器的弹性公网IP地址。多网卡情况下，只显示主网卡的弹性公网IP地址。

- URI  
/169.254.169.254/latest/meta-data/public-ipv4
- 方法  
支持GET请求。
- 示例  
Linux操作系统：  
**curl http://169.254.169.254/latest/meta-data/public-ipv4**  
Windows操作系统：  
**Invoke-RestMethod http://169.254.169.254/latest/meta-data/public-ipv4**  
46.1.1.2

## Public Keys ( EC2-兼容的 API )

用于查询弹性云服务器的公钥。

- URI  
/169.254.169.254/latest/meta-data/public-keys/0/openssh-key
- 方法  
支持GET请求。
- 示例  
Linux操作系统：  
**curl http://169.254.169.254/latest/meta-data/public-keys/0/openssh-key**  
Windows操作系统：  
**Invoke-RestMethod http://169.254.169.254/latest/meta-data/public-keys/0/openssh-key**  
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQADI5Fw5k8Fgzajn1zJwLoV3+wMP+6CyvsSilc/hioggSnYu/AD0Yqm8vVO0kWlun1rFbdO+QUZKyVr/OPUjQSw4SRh4qsTKf/+eFoWTjplFvd1WCBZzS/WRenxiwR00KkczHSJro763+wYcwKieb4eKRxaQoQvoFgVjLBULXAJH4eKoKTVNtMXAvPP9aMy2SLgsJNtMb9ArfziAiblQynq7UifLnN3VclzPeiWrqtzjyOp6CPUXnL0lVPTvbLe8sUteBsjZwLL6K4i



+Y0lf3ryqnmQgC21yW4Dzu+kwk8FVT2MgWkCwiZd8gQ/+uJzrJFyMfuOBikLOBfuUENIJUhAB  
Generated-by-Nova

## 相关链接

[Linux操作系统云服务器无法获取元数据怎么办？](#)

### 4.4.2.2 通过实例自定义数据配置 ECS 实例

## 使用场景

当您有如下需求时，可以考虑使用实例自定义数据注入功能来配置弹性云服务器：

- 需要通过脚本简化弹性云服务器配置
- 通过脚本初始化系统
- 已有脚本，在创建弹性云服务器的时候一并上传到服务器
- 其他可以使用脚本完成的功能

## 使用限制

- Linux：
  - 用于创建弹性云服务器的镜像安装了Cloud-init组件。
  - 实例自定义数据大小限制：小于等于32KB。
  - 如果通过文本方式上传实例自定义数据，数据只能包含ASCII码字符；如果通过文件方式上传实例自定义数据，可以包含任意字符，同时，要求文件大小小于等于32KB。
  - 必须是公共镜像，或继承于公共镜像的私有镜像，或自行安装了Cloud-init组件的私有镜像。
  - 必须满足相应Linux弹性云服务器自定义脚本类型的格式要求。
  - 使用的VPC网络必须开启DHCP，安全组出方向规则保证80端口开放。
  - 选择“密码”登录方式时，不支持实例自定义数据注入功能。
- Windows
  - 用于创建弹性云服务器的镜像安装了Cloudbase-init组件。
  - 实例自定义数据大小限制：小于等于32KB。
  - 如果通过文本方式上传实例自定义数据，数据只能包含ASCII码字符；如果通过文件方式上传实例自定义数据，可以包含任意字符，同时，要求文件大小小于等于32KB。
  - 必须是公共镜像，或继承于公共镜像的私有镜像，或自行安装了Cloudbase-init组件的私有镜像。
  - 使用 VPC网络必须开启DHCP，安全组出方向规则保证80端口开放。

## 使用方法

1. 根据实际需要写User-Data脚本，脚本格式必须符合User-Data脚本，详细规范请参见[相关链接](#)。
2. 创建弹性云服务器时，需将实例自定义数据脚本内容粘贴到“高级选项 > 现在配置 > 实例自定义数据注入”的文本框，或选择实例自定义数据文件上传。

### 📖 说明

实例自定义数据注入支持使用文件和文本两种方式：

文本：将实例自定义数据脚本内容粘贴到文本框中。

文件：将实例自定义数据脚本内容保存至文本文件后再将上传。

3. 系统创建弹性云服务器，弹性云服务器启动后自动执行Cloud-init或Cloudbase-init组件，读取实例自定义数据脚本。

## 关于 Linux 云服务器的实例自定义数据脚本

Linux弹性云服务器的自定义实例自定义数据脚本（简称“脚本”）采用开源的Cloud-init架构实现，该架构以云服务器的元数据为数据来源，对弹性云服务器进行各项自动化配置。自定义脚本类型兼容开源Cloud-init，详情请参考Cloud-init开源说明：<http://cloudinit.readthedocs.io/en/latest/topics/format.html>。

- 脚本的执行时间：弹性云服务器的状态变为“running”之后，执行文件/etc/init之前。

### 📖 说明

脚本默认以root用户权限执行。

- 脚本类型：包括“User-Data Script”和“Cloud-Config Data”。

表 4-10 Linux 云服务器脚本类型介绍

-	User-Data Script	Cloud-Config Data
描述	以脚本作为自定义配置的手段，如shell和python脚本。	Cloud-init中预定义的对云服务器部分服务进行配置的手段，如yum源。
格式	首行必须是“#!”（如“#!/bin/bash和#!/usr/bin/env python”），且前面不能有空格。 首次启动时，将在rc.local-like级别上执行，rc.local-like意味着“在启动序列中非常靠后”。	首行必须是“#cloud-config”，且前面不能有空格。
限制	在采用Base64编码前，脚本内容（包括首行在内）不能超过32KB。	在采用Base64编码前，脚本内容（包括首行在内）不能超过32KB。
频率	仅在首次启动弹性云服务器时执行一次。	配置的服务不同，执行频率也会不同。

- 如何查看注入Linux弹性云服务器的自定义数据？
  - a. 登录弹性云服务器。
  - b. 执行以下命令，以root用户权限查看实例自定义数据。

```
curl http://169.254.169.254/openstack/latest/user_data
```
- 脚本使用示例：

该示例介绍如何以不同格式输入Linux弹性云服务器的脚本，并查看脚本的运行结果。

### 示例一：脚本类型为User-Data Script

创建弹性云服务器时，选择“文本”形式并输入实例自定义数据脚本，脚本如下：

```
#!/bin/bash
echo "Hello, the time is now $(date -R)" | tee /root/output.txt
```

创建成功后，连接并启动弹性云服务器，执行命令`cat [file]`，查看脚本的运行结果。

```
[root@XXXXXXXX ~]# cat /root/output.txt
Hello, the time is now Mon, 16 Jul 2016 16:03:18+0800
```

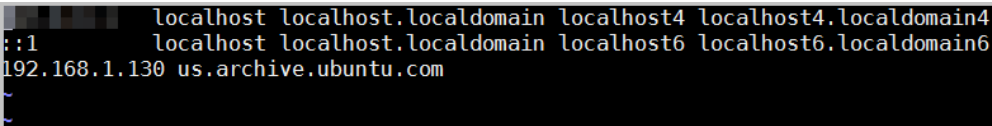
### 示例二：脚本类型为Cloud-Config Data

创建弹性云服务器时，选择“文本”形式并输入实例自定义数据脚本，脚本如下：

```
#cloud-config
bootcmd:
- echo 192.168.1.130 us.archive.ubuntu.com >> /etc/hosts
```

创建成功后，连接并启动弹性云服务器，执行命令`cat /etc/hosts`查看脚本的运行结果。

图 4-79 查看运行结果



```
localhost localhost.localdomain localhost4 localhost4.localdomain4
::1 localhost localhost.localdomain localhost6 localhost6.localdomain6
192.168.1.130 us.archive.ubuntu.com
```

## 关于 Windows 弹性云服务器的实例自定义数据脚本

Windows弹性云服务器的实例自定义数据脚本（简称“脚本”）采用开源的cloudbase-init架构实现，该架构以弹性云服务器的元数据为数据来源，对弹性云服务器进行各项自动化配置，为Windows弹性云服务器提供执行初始化脚本的能力。自定义脚本类型兼容开源Cloudbase-init，详情请参考Cloudbase-init开源说明：<https://cloudbase-init.readthedocs.io/en/latest/userdata.html>。

- 脚本类型：包括“Batch批处理程序”和“PowerShell脚本”。

表 4-11 Windows 云服务器脚本类型介绍

-	Batch批处理程序	PowerShell脚本
格式	以“rem cmd”开头且作为首行，且前面不能有空格。	以“#ps1”开头且作为首行，且前面不能有空格。
限制	在采用Base64编码前，脚本内容（包括首行在内）不能超过32KB。	在采用Base64编码前，脚本内容（包括首行在内）不能超过32KB。

- 如何查看注入Windows弹性云服务器的实例自定义数据？
  - 登录弹性云服务器。
  - 在浏览器栏访问如下地址，查看注入的实例自定义数据。  
[http://169.254.169.254/openstack/latest/user\\_data](http://169.254.169.254/openstack/latest/user_data)
- 脚本使用示例

该示例介绍如何以不同格式输入Windows弹性云服务器的脚本，并查看脚本的运行结果。

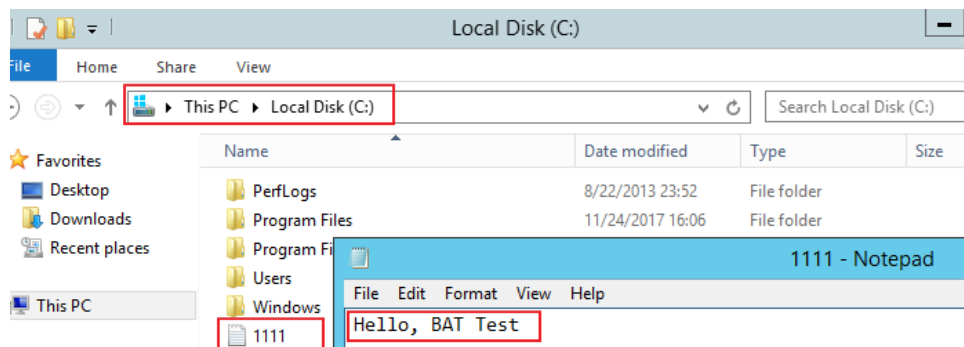
### 示例一：Batch批处理程序

创建弹性云服务器时，选择“文本”形式并输入实例自定义数据，脚本如下：

```
rem cmd  
echo "Hello, BAT Test" > C:\1111.txt
```

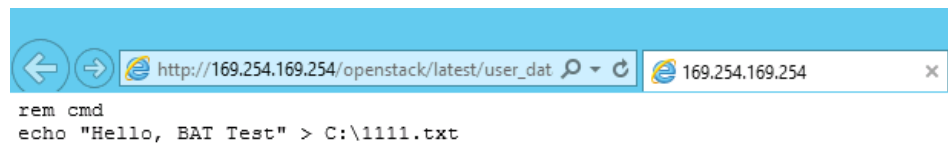
创建成功后，连接并启动弹性云服务器，查看脚本的运行结果：在C:\盘添加了一个名称为“1111”的文本文档，如图4-80所示。

图 4-80 新建文本文档\_Batch



如需查看注入Windows弹性云服务器的实例自定义数据，可以在浏览器访问：[http://169.254.169.254/openstack/latest/user\\_data](http://169.254.169.254/openstack/latest/user_data)，如图4-81所示。

图 4-81 查看实例自定义数据\_Batch



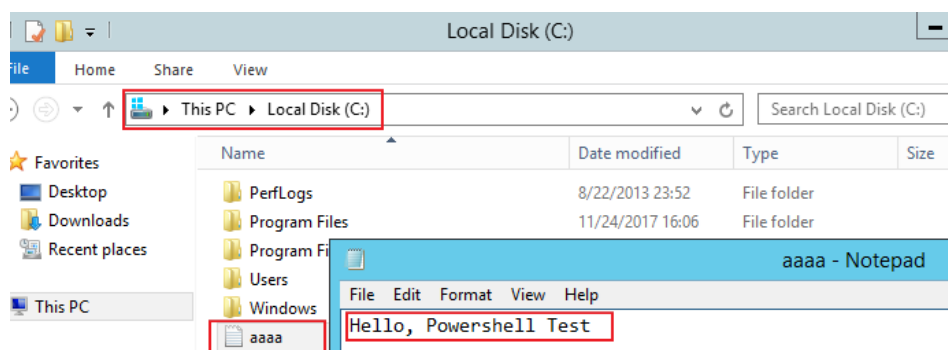
### 示例二：PowerShell脚本

创建弹性云服务器时，选择“文本”形式并输入实例自定义数据，脚本如下：

```
#ps1  
echo "Hello, Powershell Test" > C:\aaaa.txt
```

创建成功后，连接并启动弹性云服务器，查看脚本的运行结果：在C:\盘添加了一个名称为“aaaa”的文本文档。

图 4-82 新建文本文档\_PowerShell



如需查看注入Windows弹性云服务器的实例自定义数据，可以在浏览器访问：[http://169.254.169.254/openstack/latest/user\\_data](http://169.254.169.254/openstack/latest/user_data)，如图4-83所示。

图 4-83 查看实例自定义数据\_PowerShell



## 案例 1

该样例介绍如何通过实例自定义数据注入，为Linux弹性云服务器简化云服务器配置。

当您需要设置vim的“语法高亮显示”、“制表符占用4个空格”和“显示行数”属性时，就可以写一个脚本.vimrc，并将其注入到“/root/.vimrc”。在您创建完云服务器时，您的vim配置便完成了。这样做可以大幅提高系统配置的效率，特别是在您一次性创建多台弹性云服务器时。

实例自定义数据注入示例：

```
#cloud-config
write_files:
  - path: /root/.vimrc
    content: |
      syntax on
      set tabstop=4
      set number
```

## 案例 2

该样例介绍如何通过实例自定义数据注入，设置Linux弹性云服务器密码。

### 📖 说明

新密码必须符合密码复杂度要求，密码规则如表4-12所示。

表 4-12 密码设置规则

参数	规则
密码	<ul style="list-style-type: none"><li>● 密码长度范围为8到26位。具体以界面显示为准。</li><li>● 密码至少包含以下4种字符中的3种：<ul style="list-style-type: none"><li>- 大写字母</li><li>- 小写字母</li><li>- 数字</li><li>- Windows操作系统云服务器特殊字符：包括“\$”、“!”、“@”、“%”、“_”、“-”、“=”、“+”、“[”、“]”、“.”、“/”、“,”和“?”</li><li>- Linux操作系统云服务器特殊字符：包括“!”、“@”、“%”、“_”、“-”、“=”、“+”、“[”、“]”、“.”、“/”、“^”、“,”、“{”、“}”和“?”</li></ul></li><li>● 密码不能包含用户名或用户名的逆序。</li><li>● Windows操作系统的云服务器，不能包含用户名中超过两个连续字符的部分。</li></ul>

实例自定义数据注入示例:

使用密文密码（推荐使用）：

```
#!/bin/bash
echo 'root:$6$V6azyeLwcD3CHlpY$BN3VVq18fmCkj66B4zdHLWevqcxlig' | chpasswd -e;
```

其中，\$6\$V6azyeLwcD3CHlpY\$BN3VVq18fmCkj66B4zdHLWevqcxlig为密文密码，生成方法如下：

1. 执行以下命令，生成加密盐值

```
python -c "import crypt, getpass, pwd;print crypt.mksalt()"
```

回显信息如下：

```
$6$V6azyeLwcD3CHlpY
```

2. 执行以下命令，根据盐值生成密文密码

```
python -c "import crypt, getpass, pwd;print crypt.crypt('Cloud.1234','\$6$V6azyeLwcD3CHlpY')"
```

回显信息如下：

```
$6$V6azyeLwcD3CHlpY$BN3VVq18fmCkj66B4zdHLWevqcxlig
```

弹性云服务器创建成功后，您可以使用该密码登录。

### 案例 3

该样例介绍如何通过实例自定义数据注入，重置Linux弹性云服务器密码。

假设您需要将root用户密码重置为”\*\*\*\*\*“。

#### 说明

新密码必须符合密码复杂度要求，密码规则如表4-13所示。

表 4-13 密码设置规则

参数	规则
密码	<ul style="list-style-type: none"><li>● 密码长度范围为8到26位。具体以界面显示为准。</li><li>● 密码至少包含以下4种字符中的3种：<ul style="list-style-type: none"><li>- 大写字母</li><li>- 小写字母</li><li>- 数字</li><li>- Windows操作系统云服务器特殊字符：包括“\$”、“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“/”、“,”和“?”</li><li>- Linux操作系统云服务器特殊字符：包括“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“/”、“^”、“,”、“{”、“}”和“?”</li></ul></li><li>● 密码不能包含用户名或用户名的逆序。</li><li>● Windows操作系统的云服务器，不能包含用户名中超过两个连续字符的部分。</li></ul>

实例自定义数据注入示例（请严格遵守下面示例中的缩进标准）：

```
#cloud-config
chpasswd:
  list: |
    root:*****
  expire: False
```

弹性云服务器创建成功后，您可以使用重置的密码登录弹性云服务器。为了保证密码安全，建议您在第一次登录弹性云服务器后，修改root用户密码。

## 案例 4

该样例介绍如何通过实例自定义数据注入，为Windows弹性云服务器新增用户并设置密码。

假设您要创建的用户名为abc、密码为“\*\*\*\*\*”的普通用户，并将其添加至administrators用户组。

### 说明

新密码必须符合密码复杂度要求，密码规则如表4-13所示。

实例自定义数据注入示例：

```
rem cmd
net user abc ***** /add
net localgroup administrators abc /add
```

弹性云服务器创建成功后，您可以使用新创建的用户名和密码登录弹性云服务器。



## 案例 5

该样例介绍如何通过实例自定义数据注入，为Linux弹性云服务器更新系统软件包，并且开启httpd相关服务。注入成功后，您的弹性云服务器就可以使用httpd服务了。

实例自定义数据注入示例：

```
#!/bin/bash
yum update -y
service httpd start
chkconfig httpd on
```

## 相关链接

更多关于实例自定义数据注入案例的介绍，请参见Cloud-init/Cloudbase-init官网：

- <https://cloudinit.readthedocs.io/en/latest/>
- <https://cloudbase-init.readthedocs.io/en/latest/>


## 4.4.3 修改 ECS 名称

### 操作场景

弹性云服务器创建成功后，您可以根据需求，修改云服务器的名称。

系统支持批量修改多台弹性云服务器的名称，修改完成后，这些弹性云服务器的名称相同。

### 修改单台弹性云服务器名称

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 将鼠标移动至目标云服务器的“名称/ID”列。
4. 单击弹性云服务器名称右方的“”，根据界面提示，修改云服务器名称。  
允许重名：勾选后，允许修改后的名称与其他云服务器名称相同。如果未勾选，且设置的名称与其他云服务器名称相同，此时，系统将提示您该名称已被使用，您需要更换其他名称。
5. 单击“确定”，新名称生效。

### 批量修改弹性云服务器的名称

1. 登录管理控制台。
  2. 选择“计算 > 弹性云服务器”。
  3. 勾选待修改名称的弹性云服务器。
  4. 单击云服务器列表页上方的“更多 > 修改名称”。
  5. 设置新名称。
  6. 单击“确定”。
- 批量修改云服务器时，修改后的云服务器名称相同，如全部为ecs-test。



## 4.4.4 管理云服务器组

### 操作场景

云服务器组是对云服务器的一种逻辑划分，云服务器组中的弹性云服务器遵从同一策略。

当前仅支持反亲和性，即同一云服务器组中的弹性云服务器分散地创建在不同的主机上，提高业务的可靠性。

您可以使用云服务器组将业务涉及到的云服务器分散部署在不同的物理服务器上，以此保证业务的高可用性和底层容灾能力。

云服务器组支持以下操作：

- [创建云服务器组](#)
- [添加云服务器到云服务器组](#)
  - 在创建弹性云服务器时，将弹性云服务器加入云服务器组。  
详细操作，请参考[步骤三：高级配置](#)。
  - 在弹性云服务器创建成功后，添加到云服务器组。
- [从云服务器组移出云服务器](#)
- [删除云服务器组](#)

### 创建云服务器组

您可以申请创建一个云服务器组，同一云服务器组中的弹性云服务器遵从相同策略，云服务器组与云服务器组之间没有关联关系。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在左侧导航树中，选择“云服务器组”。
4. 在“云服务器组”页面，单击“创建云服务器组”。
5. 输入云服务器组的名称。
6. 选择云服务器组的策略。
7. 单击“确定”。

### 添加云服务器到云服务器组

为提升业务可靠性，您可以添加弹性云服务器到云服务器组，添加后，该弹性云服务器与云服务器组中的其他云服务器分散地创建在不同主机上。

## 📖 说明

- 部分实例类型的弹性云服务器加入云服务器组之前需要先关机，请在加入云服务器组操作时，根据界面提示进行关机。
- 添加云服务器到云服务器组后，会重新分配该云服务器所在的主机，使其与云服务器组中的其他云服务器分散地创建在不同主机上。那么当弹性云服务器再次开机时，可能会出现由于资源不足引起的启动失败，请将云服务器移出云服务器组后重新启动。
- 包含本地盘的云服务器无法在创建后加入云服务器组，如需使用云服务器组功能，请在创建时选择云服务器组。
- 弹性云服务器包含本地盘、GPU卡、FPGA卡，则无法在创建后加入云服务器组，如需使用云服务器组功能，请在创建时选择云服务器组。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在左侧导航树中，选择“云服务器组”。
4. 单击“操作”列下的“添加云服务器”。
5. 在“添加云服务器”页面，选择待添加的弹性云服务器。
6. 单击“确定”，将弹性云服务器加入云服务器组。

## 从云服务器组移出云服务器

弹性云服务器从云服务器组移出后，与云服务器组中的其他云服务器不再遵从云服务器组策略。

1. 登录管理控制台。
  2. 选择“计算 > 弹性云服务器”。
  3. 在左侧导航树中，选择“云服务器组”。
  4. 展开云服务器组，查看云服务器组中包含的弹性云服务器。
  5. 单击目标云服务器“操作”列下的“移出云服务器组”。
  6. 单击“是”。
- 将弹性云服务器从云服务器组中移出。

## 删除云服务器组

删除云服务器组会解除组内云服务器的策略约束。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在左侧导航树中，选择“云服务器组”。
4. 单击待删除云服务器组“操作”列下的“删除”。
5. 单击“是”。

## 4.4.5 自动恢复

### 操作场景

弹性云服务器运行在物理机上，虽然提供了多种机制来保证系统的可靠性、容错能力和高可用性，但是，服务器的硬件、电源等部件仍有较小概率的损坏。如果物理设备的损坏导致物理机电源无法正常工作或重启，会导致CPU和内存数据丢失，无法进行热迁移来恢复弹性云服务器。

云服务平台提供了自动恢复功能，以冷迁移的方式重启弹性云服务器，使弹性云服务器具备高可靠性和强大的动态迁移能力。您可以在创建弹性云服务器时，或者弹性云服务器创建完成后，开启自动恢复功能。开启自动恢复功能后，当弹性云服务器所在的硬件出现故障时，系统会自动将弹性云服务器迁移至正常的物理机，保障您受到的影响最小，该过程会导致云服务器重启。

## 使用须知

- 自动恢复功能不保证用户数据的一致性。
- 仅支持物理主机故障产生的弹性云服务器自动恢复，弹性云服务器本身故障当前不支持自动恢复。
- 弹性云服务器所在的物理主机关机后，才能执行自动恢复。如果物理主机内存故障等因素导致物理主机未关机，则不能执行自动恢复。
- 对于同一弹性云服务器，如果发生物理主机故障，12小时内仅允许1次自动恢复操作。
- 如下场景时，可能会引起自动恢复弹性云服务器失败：
  - 系统发生大规模故障，导致迁移弹性云服务器前，找不到可用的物理主机。
  - 迁移弹性云服务器时，用于迁移的物理主机临时容量不足。
- 对于包含如下特殊资源的弹性云服务器，暂不支持自动恢复功能：
  - 本地盘
  - 直通FPGA卡
  - 直通IB网卡

## 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 单击弹性云服务器的名称。  
系统跳转至弹性云服务器详情页面。
4. 单击“自动恢复”栏的“开启”或“关闭”，设置该弹性云服务器的自动恢复属性。  
默认开启自动恢复功能。
  - 开启自动恢复功能后，当弹性云服务器所在的硬件出现故障时，系统会自动将弹性云服务器迁移至正常的物理机，保障您受到的影响最小，该过程会导致云服务器重启。
  - 关闭自动恢复功能后，当弹性云服务器所在的硬件出现故障时，需等待系统管理员进行问题修复。

### 4.4.6 获取控制台日志

#### 操作场景

弹性云服务器不能正常启动，或运行异常时，可以通过查看、下载弹性云服务器的控制台日志进行问题定位，排查内核、服务配置是否有误。

控制台日志记录了弹性云服务器“开机”、“关机”、“重启”以及“强制重启”操作的运行日志，您可以通过管理控制台获取弹性云服务器一个小时之内的日志信息。

## 使用须知

- 系统不支持记录弹性云服务器“强制关机”的运行日志。
- 对于如下操作系统的弹性云服务器，系统支持查询控制台日志：
  - Red Hat Enterprise Linux 6.x系列
  - Red Hat Enterprise Linux 7.x系列
  - CentOS 6.x系列
  - CentOS 7.x系列
  - Ubuntu 14.x系列
  - Ubuntu 16.x系列
  - SUSE 11.x系列
  - SUSE 12.x系列
  - OpenSUSE 13.x系列
  - OpenSUSE 42.x系列
  - Debian 16.x系列
  - Fedora系列
  - FreeBSD系列
  - CoreOS系列
- 使用Windows操作系统的弹性云服务器，不支持控制台日志功能。
- 系统最多能够保存100kb大小的日志数据文件。

## 操作步骤

**步骤1** 登录弹性云服务器。

**步骤2** 检查并修改grub文件。

不同类型的OS配置操作不同，具体如下：

### 说明

为了避免对修复模式的启动产生影响，建议您仅修改当前默认启动使用的引导项。

- OS类型为CentOS 6/Red Hat 6系列时配置如下：
  - a. 执行以下命令，打开待编辑的配置文件。  
**vi /boot/grub/menu.lst**
  - b. 找到linux、linux16或kernel参数行末尾（根据系统实际的参数选择linux、linux16或kernel之一），添加console=ttyS0，并删除rhgb quiet参数。如果已存在console=ttyS0，则不用再添加此参数。保存退出即可。
- OS类型为CentOS 7/Red Hat 7/Ubuntu14系列时配置如下：
  - a. 执行以下命令，打开待编辑的配置文件。  
**vi /boot/grub2/grub.cfg**
  - b. 找到linux、linux16或kernel参数行末尾（根据系统实际的参数选择linux、linux16或kernel之一），添加console=ttyS0，并删除rhgb quiet参数。如果已存在console=ttyS0，则不用再添加此参数。保存退出即可。
- OS类型为SUSE 11系列时配置方法如下：

- a. 执行以下命令，打开待编辑的配置文件。  
**vi /boot/grub/menu.1st**
- b. 找到linux、linux16或kernel参数行末尾（根据系统实际的参数选择linux、linux16或kernel之一），添加console=ttyS0。如果已存在console=ttyS0，则不用再添加此参数。保存退出即可。
- OS类型为SUSE 12/openSUSE 13/openSUSE 42系列时配置方法如下：
  - a. 执行以下命令，打开待编辑的配置文件。  
**vi /boot/grub2/grub.cfg**
  - b. 找到linux、linux16或kernel参数行末尾（根据系统实际的参数选择linux、linux16或kernel之一），添加console=ttyS0。如果已存在console=ttyS0，则不用再添加此参数。保存退出即可。
- OS类型为Debian /Ubuntu 16系列时配置方法如下：
  - a. 执行以下命令，打开待编辑的配置文件。  
**vi /boot/grub/grub.cfg**
  - b. 找到linux、linux16或kernel参数行末尾（根据系统实际的参数选择linux、linux16或kernel之一），添加console=ttyS0。如果已存在console=ttyS0，则不用再添加此参数。保存退出即可。
- OS类型为Fedora系列时配置如下：
  - a. 执行以下命令，打开待编辑的配置文件。  
**vi /boot/grub2/grub.cfg**
  - b. 找到linux、linux16或kernel参数行末尾（根据系统实际的参数选择linux、linux16或kernel之一），添加console=ttyS0。如果已存在console=ttyS0，则不用再添加此参数。保存退出即可。
- OS类型为Freebsd系列时配置如下：
  - a. 执行以下命令，打开待编辑的配置文件。  
**vi /boot/loader.conf**
  - b. 添加console="comconsole"。如果已存在console="comconsole"，则不用再添加此参数。保存退出即可。
- OS类型为CoreOS系列时配置如下：
  - a. 执行以下命令，查看是否已经配置ttyS0。  
**cat /proc/cmdline | grep ttyS0**
    - 是，则表示已配置ttyS0。
    - 否，则未配置ttyS0，执行[步骤2.b](#)。
  - b. 打开待编辑的配置文件。  
**vi /usr/share/oem/grub.cfg**  
**📖 说明**

如果不存在“/usr/share/oem/grub.cfg”配置文件，需手动创建该文件。
  - c. 添加set linux\_append="console=ttyS0"。如果已存在set linux\_append="console=ttyS0"，则不用再添加此参数。保存退出即可。

**步骤3** 在弹性云服务器列表页，单击“重启”，重新启动弹性云服务器。

**步骤4** 获取弹性云服务器的控制台日志。

1. 登录管理控制台。
  2. 选择“计算 > 弹性云服务器”。
  3. 在弹性云服务器列表页，单击待获取控制台日志的弹性云服务器名称，进入详情页。
  4. 选择“控制台日志”页签。
  5. 单击“显示行数”后的下拉页签，选择当前页面显示日志的行数。
  6. 单击“查询”。
- 查看显示行数的日志详情。

**说明**

用户单击“查询”后，界面上显示的日志不会再自动更新。如果需要查看最新的控制台日志，请再次单击“查询”。

7. （可选）您可以通过单击“下载”，下载显示行数的日志信息。  
下载的日志文件为“.txt”格式。

---结束

## 4.4.7 配置同一 VPC 内 ECS 通过主机名进行通信

对于同一VPC内的弹性云服务器，可以通过主机名称进行通信。此时，您需要配置主机名与IP地址之间的映射关系。较之通过IP地址进行通信，主机名方式进行通信更为方便。

### 约束限制

仅适用于Linux弹性云服务器。

### 操作步骤

假设VPC内共有2台弹性云服务器：ecs-01和ecs-02。通过如下操作，ecs-01和ecs-02可以通过主机名互相通信。

**步骤1** 分别登录ecs-01和ecs-02，获取2台弹性云服务器的私有IP地址。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表页，通过“私有IP地址”栏，查看弹性云服务器的私有IP地址。

【示例】假设查询的私有IP地址如下：

ecs-01: 192.168.0.1

ecs-02: 192.168.0.2

**步骤2** 分别获取2台弹性云服务器的主机名。

1. 登录弹性云服务器。
2. 执行以下命令，查询弹性云服务器的主机名。

```
sudo hostname
```

【示例】假设查询的主机名如下：

```
ecs-01: hostname01
ecs-02: hostname02
```

**步骤3** 建立主机名与IP地址之间的映射关系，并添加同一VPC内其他弹性云服务器的信息。

1. 登录弹性云服务器ecs-01。
2. 执行以下命令，切换至root权限。  
**sudo su -**
3. 执行以下命令，编辑hosts文件。  
**vi /etc/hosts**
4. 按“i”，进入编辑模式。
5. 按照如下格式添加语句，建立映射关系。  
*私有IP地址 主机名*  
【示例】需添加的语句为：  
192.168.0.1 hostname01  
192.168.0.2 hostname02
6. 按“Esc”退出编辑模式。
7. 执行以下命令，保存并退出。  
**:wq**
8. 登录ecs-02。
9. 重复执行[步骤3.2~步骤3.7](#)。

**步骤4** 测试验证能否通过主机名正常通信。

分别登录同一VPC内的弹性云服务器，执行以下命令，使用ping命令ping添加的主机，看ping包是否能正常送达。

```
ping hostname
----结束
```

## 4.5 变更 ECS 规格（vCPU 和内存）

### 4.5.1 变更单台 ECS 规格

#### 操作场景

当您创建的弹性云服务器规格无法满足业务需要时，可参考本章节变更规格，升级vCPU、内存。对于部分类型的弹性云服务器，您还可以在变更规格时，更换弹性云服务器的类型。

#### 变更须知

- 变更规格过程中，弹性云服务器需要关机，建议您选择业务空闲时间进行操作。
- 变更规格过程中，不要对弹性云服务器做任何操作，例如关机、重启等，否则会导致变更规格失败。
- 弹性云服务器规格（CPU或内存）变小，会影响弹性云服务器的性能。



- 对于部分类型的弹性云服务器，暂不支持规格变更操作。[实例类型](#)章节列举了提供的云服务器类型，并对每种云服务器类型的功能、使用进行介绍，具体请以各类型云服务器的“使用须知”为准。
- 当云硬盘状态为“正在扩容”时，不支持变更所挂载的弹性云服务器规格。
- Windows操作系统的弹性云服务器，执行变更规格操作前建议您参考[磁盘脱机怎么办？](#)修改Windows操作系统SAN策略，避免变更规格后磁盘处于脱机状态。

## 步骤 1：变更规格

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 单击“操作”列下的“更多 > 变更规格”。  
系统进入“云服务器变更规格”页面。
4. 根据界面提示，选择变更后的弹性云服务器类型、vCPU和内存。  
变更规格前请先将云服务器关机，或根据页面提示勾选“系统自动关机后变更规格”。
5. 单击“下一步”。
6. 确认变更后的配置无误后，阅读并勾选同意服务协议，单击“提交申请”。

## 步骤 2：检查磁盘挂载状态

变更规格时，可能会发生磁盘挂载失败的情况，因此，变更规格后，需检查磁盘挂载状态是否正常。如果正常，则变更成功。

- Windows弹性云服务器  
详细操作请参考[Windows弹性云服务器变更规格后数据盘脱机怎么办？](#)
- Linux弹性云服务器  
详细操作请参考[Linux弹性云服务器变更规格后磁盘脱机怎么办？](#)

# 4.6 变更 ECS 操作系统

## 4.6.1 重装操作系统

### 操作场景

弹性云服务器操作系统无法正常启动时，或云服务器系统运行正常，但需要对系统进行优化，使其在最优状态下工作时，用户可以使用重装弹性云服务器的操作系统功能。

### 重装须知

- 重装操作系统后弹性云服务器IP地址和MAC地址不发生改变。
- 重装操作系统会清除系统盘数据，包括系统盘上的系统分区和所有其它分区，请做好数据备份。
- 重装操作系统不影响数据盘数据。
- 重装操作系统后的几分钟，系统正在注入密码信息，在此期间请勿对云服务器执行其他操作，避免密码信息注入失败导致云服务器无法登录。



## 约束与限制

- 云硬盘的配额需大于0。
- 如果是通过私有镜像创建的弹性云服务器，请确保原有镜像仍存在。
- H2型弹性云服务器不支持操作系统的重装功能。

## 前提条件

- 待重装操作系统的弹性云服务器挂载有系统盘。

## 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在待重装操作系统的弹性云服务器的“操作”列下，单击“更多 > 镜像/磁盘 > 重装系统”。  
重装操作系统前请先将云服务器关机，或根据页面提示勾选“立即重装操作系统”。
4. 设置登录方式。
5. 单击“确定”。
6. 在“重装系统”页面，确认重装的操作系统规格无误后，单击“提交申请”。  
提交申请后，弹性云服务器的状态变为“重装中”，当该状态消失后，表示重装结束。

### 说明

重装操作系统过程中，会创建一台临时弹性云服务器，重装操作系统结束后会自动删除。在重装操作系统过程中请不要对该弹性云服务器进行任何操作。

## 后续处理

如果操作系统重装失败，页面会提示重装操作系统失败。云服务平台支持重试功能，用户可重新执行2-6，重装弹性云服务器的操作系统。

重试后，如果仍未成功，可直接联系管理员进行人工恢复。

## 4.6.2 切换操作系统

### 操作场景

切换操作系统是为您弹性云服务器重新切换一个系统盘。切换完成后弹性云服务器的系统盘ID会发生改变，并删除原有系统盘。

如果弹性云服务器当前使用的操作系统不能满足业务需求（如软件要求的操作系统版本较高），您可以选择切换弹性云服务器的操作系统。

云服务平台支持不同镜像类型（包括公共镜像、私有镜像、共享镜像）与不同操作系统之间的互相切换。您可以将现有的操作系统切换为不同镜像类型的操作系统。

### 约束与限制

- 切换操作系统操作预计需等待10~20分钟。切换操作系统过程中，弹性云服务器会显示任务状态为“切换操作系统中”。

- 切换操作系统完成后的几分钟，系统正在注入密码信息，在此期间请勿对云服务器执行其他操作，避免密码信息注入失败导致云服务器无法登录。
- 待切换操作系统的弹性云服务器需挂载有系统盘。
- 云硬盘的配额需大于0。
- 不支持更换系统盘的磁盘类型。
- H2型弹性云服务器不支持操作系统的切换功能。

## 切换须知

- 切换操作系统后，弹性云服务器将不再保留原操作系统，并删除原有系统盘。
- 切换操作系统会清除系统盘数据，包括系统盘上的系统分区和所有其它分区，请做好数据备份。
- 切换操作系统不影响数据盘数据。
- 切换操作系统后，您的业务运行环境需要在新的系统中重新部署。
- 切换操作系统成功后弹性云服务器会自动开机。
- 切换操作系统后不支持更换系统盘的云硬盘类型。
- 切换操作系统后弹性云服务器IP地址和MAC地址不发生改变。
- 切换操作系统后，当前操作系统内的个性化设置（如DNS、主机名等）将被重置，需重新配置。
- 切换操作预计需等待10~20分钟完成切换操作系统。切换操作系统过程中，弹性云服务器会显示任务状态为“切换操作系统中”。
- 切换弹性云服务器的操作系统后，由于所选镜像不同，系统盘的容量可能会增大。

## 不同操作系统切换须知

不同操作系统间的切换是指Windows与Linux操作系统之间的互相切换。

- Windows系统更换为Linux系统：请安装读写Windows系统的NTFS分区工具，例如NTFS-3G等。
- Linux系统更换为Windows系统：请安装可以识别ext3、ext4等分区的识别软件，例如Ext2Read、Ext2Fsd等。

### 说明

云平台不推荐您将Linux系统更换为Windows系统，当Linux系统中存在LVM分区时，切换为Windows系统后可能会导致LVM逻辑分区无法识别。

## 前提条件

- 完成云服务器的数据备份。
- 如果您使用私有镜像切换操作系统请参考《镜像服务用户指南》提前完成私有镜像的制作。
  - 如果需要指定弹性云服务器的镜像，请提前使用指定弹性云服务器创建私有镜像。
  - 如果需要使用本地的镜像文件，请提前将镜像文件导入并注册为云平台的私有镜像。
  - 如果需要使用其他区域的私有镜像，请提前复制镜像。

- 如果需要使用其他账号的私有镜像，请提前完成镜像共享。

## 操作步骤

1. 登录管理控制台。
  2. 选择“计算 > 弹性云服务器”。
  3. 在待切换操作系统的弹性云服务器的“操作”列下，单击“更多 > 镜像/磁盘 > 切换操作系统”。
- 切换操作系统前请先将云服务器关机，或根据页面提示勾选“立即重装操作系统”。

### 📖 说明

- 系统支持批量切换云服务器操作系统：
1. 勾选待切换操作系统的多台弹性云服务器。
  2. 单击云服务器列表页上方的“更多 > 切换操作系统”。
4. 根据需求选择需要切换的弹性云服务器镜像。  
相关参数说明请参见[创建弹性云服务器](#)。
  5. 设置登录方式。  
使用私有镜像切换操作系统时，可以选择使用私有镜像密码。
  6. 单击“确定”。
  7. 在“切换云服务器操作系统”页面，确认切换的操作系统规格无误后，单击“提交申请”。
- 提交切换操作系统的申请后，弹性服务器的状态变为“切换中”，当该状态消失后，表示切换结束。

### 📖 说明

切换操作系统过程中，会创建一台临时弹性云服务器，切换操作系统结束后会自动删除。

## 后续处理

- 如果切换操作系统前后都是Linux系统，且数据盘设置了开机自动挂载分区。切换操作系统后，数据盘分区挂载信息会丢失，请更新/etc/fstab配置。
  - a. 在/etc/fstab写入切换后的分区信息。  
建议您先备份/etc/fstab文件。  
详细操作请参考[初始化Linux数据盘（fdisk）](#)，设置开机自动挂载磁盘分区。
  - b. 挂载分区。挂载分区后即可开始使用数据盘。  
**mount 磁盘分区 挂载目录**
  - c. 执行以下命令，查看挂载结果。  
**df -TH**
- 如果操作系统切换失败，云服务平台支持重试功能，用户可重新执行2-7，切换弹性云服务器的操作系统。
- 重试后，如果仍未成功，可直接联系管理员进行人工恢复。

## 4.7 查看弹性云服务器信息

## 4.7.1 查看创建状态

### 操作场景

用户申请创建弹性云服务器后，可以查看任务的创建状态。本节介绍如何查看弹性云服务器的创建状态。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 常用操作“开机/关机/重启/更多”的右侧即为创建状态栏，用户执行创建弹性云服务器操作后，创建状态栏将显示该任务为“创建中”状态。
4. 单击“创建中”对应的数字，即可查看创建中的任务，查看弹性云服务器的创建状态。

#### 说明

- 创建云服务器存在以下状态：
  - 创建中：指系统正在处理创建云服务器的请求。
  - 创建失败：指未能成功处理的请求。对于创建失败的任务，系统会自动回退，同时在界面上直观的展示错误码，例如“( Ecs.0013 ) EIP配额不足”。
  - 运行中：成功处理创建云服务器的请求，是弹性云服务正常运行的状态。在这个状态的云服务器可以运行您的业务。  
失败处理方法请参见[针对管理控制台的异常提示信息，应该如何处理？](#)。
- 如果用户发现申请状态栏显示创建弹性云服务器的任务失败，而弹性云服务器列表中显示该弹性云服务器已创建成功，关于此问题请参见[为什么创建弹性云服务器的任务失败，但是在列表中显示创建成功？](#)。

## 4.7.2 查看失败任务

### 操作场景

“失败信息”栏记录了系统处理中发生异常，导致处理失败的任务，包括该任务的“名称”、“状态”等信息。当有处理失败的操作时，控制台将显示该内容。本节介绍如何查看“失败信息”。

### 失败信息事件类型

“失败信息”栏记录的失败信息事件类型如[表4-14](#)所示。

表 4-14 失败信息事件类型

事件类型	说明
创建失败	指未能成功处理的请求。对于创建失败的任务，系统会自动回退，同时在界面上直观的展示错误码，例如“( Ecs.0013 ) EIP配额不足”。失败处理方法请参见 <a href="#">针对管理控制台的异常提示信息，应该如何处理？</a> 。

事件类型	说明
操作失败	<ul style="list-style-type: none"><li>● 变更规格 用户申请变更规格后，如果规格变更失败，则“失败信息”栏将显示本次变更规格操作。</li><li>● 创建弹性云服务器时开启自动恢复功能 用户创建弹性云服务器时，如果设置启动自动恢复功能，但是弹性云服务器创建成功后，系统开启自动恢复功能失败。此时，“失败信息”栏将显示本次创建弹性云服务器的操作。</li></ul>

## 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 常用操作“开机/关机/重启/更多”的右侧即为“失败信息”。

### 说明

- 如果管理控制台上未显示“失败任务”按钮，则表明如下任务均执行成功：
- 变更规格
  - 创建弹性云服务器时开启自动恢复功能
4. 单击“失败信息”栏对应的数字，即可查看系统处理失败的任务详情。其中：
    - 创建失败：您可以从创建失败的列表中查看创建中和创建失败的任务。
    - 操作失败：您可以从操作失败列表中查看处理异常的任务，包括异常任务的具体操作、错误码等，便于定位系统处理异常的原因，及时予以恢复。

## 4.7.3 查看弹性云服务器详细信息

### 操作场景

在您申请了弹性云服务器后，可以通过管理控制台查看和管理您的弹性云服务器。本节介绍如何查看弹性云服务器的详细配置，包括弹性云服务器名称、镜像信息、系统盘、数据盘、虚拟私有云、网卡、安全组、弹性公网IP地址和带宽等信息。

如需查看弹性云服务器的私有IP地址，请直接在弹性云服务器列表页进行查看。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。  
系统进入弹性云服务器列表页，您可以在本页面查看您已创建的弹性云服务器，以及弹性云服务器的规格、镜像、IP地址等基本信息。
3. 在弹性云服务器列表中的上方，输入弹性云服务器名称、云服务器ID、IP地址等属性类型，并按回车键进行搜索。
4. 单击待查询弹性云服务器的名称。  
系统跳转至该弹性云服务器详情页面。
5. 查看弹性云服务器的详细信息。

您可以选择“云硬盘/弹性网卡/安全组/弹性公网IP/监控”页签，为弹性云服务器进行[更改安全组](#)、[绑定弹性网卡](#)、[绑定弹性公网IP](#)等操作。


## 4.7.4 导出弹性云服务器列表信息

### 操作场景

您可以将当前账号下拥有的所有弹性云服务器信息，以.csv文件的形式导出至本地。该文件记录了弹性云服务器的ID、私有IP地址、弹性公网IP等信息。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。

3. 在弹性云服务器列表页，单击右上角的。

系统会将您账号下，当前区域的所有弹性云服务器信息自动导出，并下载至本地。

#### 说明

如需导出部分弹性云服务器的信息，请勾选对应的弹性云服务器，然后再单击右上角的



4. 在本地计算机桌面的默认下载地址文件夹，可以获取到导出的弹性云服务器列表信息。

# 5 镜像管理

## 5.1 镜像概述

### 什么是镜像

镜像是一个包含了软件及必要配置的云服务器或裸金属服务器模板，包含操作系统或业务数据，还可以包含应用软件（例如，数据库软件）和私有软件。镜像分为公共镜像、私有镜像、共享镜像。

镜像服务（Image Management Service）提供简单方便的镜像自助管理功能。用户可以灵活便捷地使用公共镜像、私有镜像或共享镜像申请云服务器。同时，用户还能通过已有的云服务器或使用外部镜像文件创建私有镜像。

### 公共镜像

公共镜像是包含常见的标准操作系统的镜像，所有用户可见，包括操作系统以及预装的公共应用，支持Ubuntu，CentOS，Debian等主流操作系统。用户可以从丰富的公共镜像库中选择或创建私有镜像，从而快速创建或批量复制弹性云服务器。请根据您的实际情况自助配置应用环境或相关软件。

### 私有镜像

私有镜像包含操作系统或业务数据、预装的公共应用以及用户的私有应用的镜像，仅用户个人可见。

表 5-1 私有镜像类型

镜像类型	说明
系统盘镜像	包含用户运行业务所需的操作系统、应用软件的镜像。系统盘镜像可以用于创建云服务器，迁移用户业务到云。
数据盘镜像	只包含用户业务数据的镜像。数据镜像可以用于创建云硬盘，将用户的业务数据迁移到云上。
整机镜像	包含用户运行业务所需的操作系统、应用软件和业务数据的镜像。整机镜像包含系统盘和其所挂载的所有数据盘。

镜像类型	说明
ISO镜像	将外部镜像的ISO文件注册到云平台的私有镜像。ISO镜像是特殊的镜像，只能发放用作临时过渡的云服务器。

如果您使用私有镜像切换操作系统请参考《镜像服务用户指南》提前完成私有镜像的制作。

- 如果需要指定弹性云服务器的镜像，请提前使用指定弹性云服务器创建私有镜像。
- 如果需要使用本地的镜像文件，请提前将镜像文件导入并注册为云平台的私有镜像。
- 如果需要使用其他区域的私有镜像，请提前复制镜像。
- 如果需要使用其他账号的私有镜像，请提前完成镜像共享。

## 共享镜像

用户将接受云平台其他用户共享的私有镜像，作为自己的镜像进行使用。

- 镜像共享的范围只能在区域内。
- 每个镜像最多可以共享给128个租户。
- 用户可以随时取消自己共享的镜像，无需通知镜像的接受方。
- 用户可以随时删除自己共享的镜像，无需通知镜像的接受方。
- 整机镜像不支持共享。

## 5.2 创建镜像

### 操作场景

您可以使用已有的弹性云服务器创建系统盘镜像、数据盘镜像、整机镜像。

- **系统盘镜像**：系统盘镜像包含用户运行业务所需的操作系统、应用软件，可以用于云服务器业务的迁移。
- **数据盘镜像**：数据盘镜像只包含用户业务数据的镜像。用户可以通过创建数据盘镜像对云服务器的数据盘进行导出。数据盘镜像可以用于创建云硬盘，将用户的业务数据迁移到云上。
- **整机镜像**：使用弹性云服务器携带其挂载的数据盘一起创建整机镜像，创建的整机镜像包含用户的业务数据，可用于快速发放包含用户业务数据的弹性云服务器。
- **ISO 镜像**：将外部镜像的ISO文件注册到云平台的私有镜像。ISO镜像是特殊的镜像，只能发放用作临时过渡的云服务器。

如果您使用私有镜像切换操作系统请参考《镜像服务用户指南》提前完成私有镜像的制作。



## 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器的“操作”列下，单击“更多 > 镜像/磁盘/备份 > 创建镜像”。
4. 根据界面要求填写如下信息：  
包含“镜像类型和来源”和“配置信息”两个信息块，各参数说明参见表5-2和表5-3。

表 5-2 镜像类型和来源

参数	说明
创建方式	选择“系统盘镜像”。
选择镜像源	选择“云服务器”，然后从列表中选择已完成相关配置的云服务器。

表 5-3 配置信息

参数	说明
名称	设置一个便于您识别的镜像名称。
描述	可选参数，对镜像进行描述。

### 说明

以上为全量参数，不同企业、单位支持情况不同，以实际界面为准。

5. 单击“立即申请”。

# 6 磁盘管理

## 6.1 为 ECS 新增磁盘

### 操作场景

云服务器的磁盘包括系统盘和数据盘。系统盘在创建云服务器时自动创建并挂载，无需单独创建。数据盘包括如下两种创建方式：

- 在创建云服务器的时候一同创建：该场景下，数据盘由系统自动挂载给云服务器。
- 在创建了云服务器之后，单独创建云硬盘：该场景下，数据盘需要手动挂载给云服务器。

本节操作介绍创建了云服务器之后，新增磁盘的操作步骤。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器的“操作”列下，单击“更多 > 镜像/磁盘/备份 > 新增磁盘”。系统跳转至创建磁盘页面。
4. 根据界面提示，设置新添加云硬盘的参数信息。  
云硬盘的参数信息配置请参考《云硬盘用户指南》中“创建云硬盘”章节的参数配置。

#### 说明

- 新增磁盘默认与弹性云服务器计费模式保持一致。
  - 新增磁盘默认与弹性云服务器位于同一区域。
  - 新增磁盘默认与弹性云服务器位于同一可用区，不支持修改磁盘可用区。
  - 创建完成后新增磁盘默认已挂载至选择的弹性云服务器。
5. 单击“立即申请”。
- 返回弹性云服务器控制台在“云硬盘”页签下，查看新增磁盘信息。

## 后续处理

云硬盘挂载至云服务器后，需要登录云服务器初始化云硬盘，即格式化云硬盘，之后云硬盘才可以正常使用。

初始化数据盘的具体操作，请参见[场景及磁盘分区形式介绍](#)。

## 6.2 为 ECS 挂载磁盘

### 操作场景

弹性云服务器创建成功后，如果发现磁盘不够用或当前磁盘不满足要求，可以将已有云硬盘挂载给弹性云服务器，或在“存储 > 云硬盘”页面创建新的磁盘，然后再挂载至弹性云服务器。

### 前提条件

- 已创建可用的云硬盘。  
创建云硬盘的操作，请参见《云硬盘用户指南》的“创建云硬盘”章节。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表中的右上角，输入弹性云服务器名、IP地址或ID，并进行搜索。
4. 单击待挂载云硬盘的弹性云服务器的名称。  
系统跳转至该弹性云服务器详情页面。
5. 选择“云硬盘”页签，并单击“挂载磁盘”。  
系统跳转至“挂载磁盘”页面。
6. 根据界面提示，勾选目标磁盘，并指定该磁盘作为系统盘或数据盘。
  - 对于采用KVM虚拟化类型的弹性云服务器，您可以指定该磁盘作为系统盘或数据盘，但不能指定具体的磁盘挂载点。
  - 对于采用XEN虚拟化类型的弹性云服务器，系统支持指定具体的磁盘挂载点，如/dev/vdb。

#### 说明

- 如果无可用的云硬盘，请单击列表下方的“创建云硬盘”进行创建。
  - 挂载磁盘时的约束限制，请参见[弹性云服务器挂载磁盘时有什么限制?](#)。
7. 单击“确定”。  
挂载成功后，在弹性云服务器详情页的“云硬盘”页签，即可看到新挂载的磁盘信息。

### 后续处理

如果挂载的磁盘是新创建的，则云硬盘挂载至云服务器后，需要登录云服务器初始化云硬盘，即格式化云硬盘，之后云硬盘才可以正常使用。

初始化数据盘的具体操作，请参见[场景及磁盘分区形式介绍](#)。

## 6.3 在线卸载磁盘

### 操作场景

将挂载在弹性云服务器中的磁盘卸载。

- 对于挂载在系统盘盘位（也就是“/dev/sda”或“/dev/vda”挂载点）上的磁盘，当前仅支持离线卸载。即卸载磁盘时，弹性云服务器需处于“关机”状态。
- 对于挂载在数据盘盘位（非/dev/sda挂载点）上的磁盘，不仅支持离线卸载，在使用部分操作系统时，还支持在线卸载磁盘功能。此时，弹性云服务器处于“运行中”状态。

本节旨在介绍在线卸载弹性云服务器磁盘的使用场景。

### 约束与限制

- 磁盘需挂载在数据盘盘位，即挂载点为非“/dev/sda”或“/dev/vda”挂载点。如果磁盘挂载在“/dev/sda”或“/dev/vda”挂载点上，则表示该磁盘作为弹性云服务器的系统盘使用，此时不允许在线卸载。
- 在线卸载云硬盘时，对于Windows弹性云服务器，请确认该弹性云服务器已安装UVP VMTools并且正常启用；对于Linux弹性云服务器，可以不安装UVP VMTools。
- 对于Windows弹性云服务器，在线卸载云硬盘前，请确保没有程序正在对该云硬盘进行读写操作。否则，将造成数据丢失。
- 对于Windows弹性云服务器，不支持在线卸载SCSI类型的云硬盘。
- 对于Linux弹性云服务器，在线卸载云硬盘前，客户需要先登录弹性云服务器，执行umount命令，取消待卸载云硬盘与文件系统之间的关联，并确保没有程序正在对该云硬盘进行读写操作。否则，卸载云硬盘将失败。

### 卸载须知

- 对于Windows弹性云服务器，在线卸载云硬盘时，如果云硬盘处于非“脱机”状态，系统会强制卸载云硬盘。此时，弹性云服务器后台可能会出现xenvbd告警提示，这种情况是正常的。

#### 说明

查看云硬盘状态的方法如下：

1. 选择“开始”，右键单击“计算机”，选择“管理”。  
弹出“计算机管理”窗口。
  2. 在左侧导航树中，选择“存储 > 磁盘管理”。  
在右侧窗格中出现磁盘列表。
  3. 查看磁盘列表中对对应磁盘的“状态”栏。
- 弹性云服务器启动、关机、重启等操作过程中，不建议进行在线卸载云硬盘的操作。
  - 对于[支持在线卸载云硬盘的操作系统](#)以外其他操作系统的弹性云服务器，不建议进行在线卸载云硬盘操作。

- 对于Linux弹性云服务器，在线卸载云硬盘后重新挂载云硬盘，可能会存在挂载前后盘符发生变化的情况。这是由于Linux系统的盘符分配机制造成的，属于正常情况。
- 对于Linux弹性云服务器，在线卸载云硬盘后重启弹性云服务器，可能会存在重启前后盘符发生变化的情况。这是由于Linux系统的盘符分配机制造成的，属于正常情况。

## 支持在线卸载云硬盘的操作系统

支持在线卸载云硬盘的操作系统包括如下两个部分：

- 第一部分请参见《镜像服务用户指南》的“简介 > 概念 > 官方公共镜像支持的操作系统类型”章节。
- 第二部分如表6-1所示。

表 6-1 支持在线卸载云硬盘的操作系统

操作系统	版本
CentOS	7.3 64bit
	7.2 64bit
	6.8 64bit
	6.7 64bit
Debian	8.6.0 64bit
	8.5.0 64bit
Fedora	25 64bit
	24 64bit
SUSE	SUSE Linux Enterprise Server 12 SP2 64bit
	SUSE Linux Enterprise Server 12 SP1 64bit
	SUSE Linux Enterprise Server 11 SP4 64bit
	SUSE Linux Enterprise Server 12 64bit
OpenSUSE	42.2 64bit
	42.1 64bit
Oracle Linux Server release	7.3 64bit
	7.2 64bit
	6.8 64bit
	6.7 64bit
Ubuntu Server	16.04 64bit
	14.04 64bit

操作系统	版本
	14.04.4 64bit
Windows（不支持在线卸载SCSI类型的云硬盘）	Windows Server 2008 R2 Enterprise 64bit
	Windows Server 2012 R2 Standard 64bit
	Windows Server 2016 R2 Standard 64bit
Redhat Linux Enterprise	7.3 64bit
	6.8 64bit

### 📖 说明

对于其他操作系统的弹性云服务器，请先关机然后再进行卸载磁盘操作，避免由于云硬盘设备与弹性云服务器在线卸载不兼容而产生未知问题。

## 操作步骤

1. 在弹性云服务器页面，单击待卸载磁盘的弹性云服务器名称，跳转至弹性云服务器详情页。
2. 选择“云硬盘”页签，单击待卸载磁盘所在行的“卸载”，卸载云硬盘。

## 6.4 扩容云硬盘

### 操作场景

当您的云硬盘存储容量不足时，您可以通过扩容云硬盘增加弹性云服务器的存储容量。系统盘和数据盘均支持扩容。

### 操作步骤

扩容云硬盘有如下两种处理方式。

- 创建一块新的云硬盘，并挂载给弹性云服务器。
- 扩容原有云硬盘空间。系统盘和数据盘均支持扩容。

您可以对状态为“正在使用”或者“可用”的云硬盘进行扩容。

- 扩容状态为“正在使用”的云硬盘，即当前需要扩容的云硬盘已经挂载给弹性云服务器。扩容状态为“正在使用”的云硬盘时，对云硬盘所挂载的弹性云服务器操作系统有要求，当前仅支持部分操作系统，具体请参见《云硬盘用户指南》的“扩容状态为‘正在使用’的云硬盘”章节。
- 扩容状态为“可用”的云硬盘，即当前需要扩容的云硬盘未挂载至任何弹性云服务器，具体请参见《云硬盘用户指南》的“扩容状态为‘可用’的云硬盘”章节。

### 📖 说明

云硬盘扩容成功后，仅扩大了云硬盘的存储容量，还需要登录云服务器自行扩展分区和文件系统。

## 6.5 开启高级磁盘功能

### 操作场景

- 平台对磁盘功能进行了升级，对于新创建的弹性云服务器，系统支持挂载60块磁盘。但是，对于已有弹性云服务器，磁盘挂载数保持原配额，最多可以挂载24块磁盘（部分类型的弹性云服务器支持挂载40块磁盘），如需支持挂载60块磁盘，需开启高级磁盘功能。
- 开启高级磁盘功能后，支持查询控制台与弹性云服务器内部之间磁盘挂载点的对应关系，查询方法请参见[控制台与弹性云服务器内部之间磁盘挂载点的对应关系](#)。

本节介绍了弹性云服务器开启高级磁盘功能的方法。

### 操作步骤

1. 登录控制台。
2. 选择“计算 > 弹性云服务器”。
3. 单击弹性云服务器名称，进入弹性云服务器详情页。
4. 选择“云硬盘”页签。
5. 查看当前弹性云服务器支持挂载的磁盘数量，并根据页面提示，单击“开启高级磁盘功能，扩大磁盘挂载数量”。
- 系统弹窗显示“开启高级磁盘功能”页面。
6. 单击“确定”。
7. 关机、并开机弹性云服务器。
- 使高级磁盘功能生效。
8. 再次进入弹性云服务器详情页，选择“云硬盘”页签，查看界面提示的磁盘挂载数量是否变化。
  - 是，开启成功。
  - 否，开启失败，请联系管理员。

# 7 弹性网卡管理

## 7.1 绑定弹性网卡

### 操作场景

当您的弹性云服务器需要多个网卡时，可以参考下面步骤为弹性云服务器绑定弹性网卡。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 单击待添加网卡的弹性云服务器名称。  
系统跳转至该弹性云服务器详情页面。
4. 在“弹性网卡”页签，单击“绑定弹性网卡”。
5. 选择不同方式，为弹性云服务器绑定弹性网卡。
  - 选择“选择已有弹性网卡”。
    - i. （可选）根据“名称”、“ID”或“私有IP”搜索待绑定的弹性网卡。
    - ii. 在弹性网卡列表中，勾选待绑定的弹性网卡。
  - 选择“创建新的弹性网卡”。  
根据界面提示，设置待增加弹性网卡的子网和安全组。
    - 所属子网：待绑定弹性网卡所属的子网。
    - 私有IP地址：如果需要给弹性云服务器添加一张指定IP地址的网卡，用户需填写“私有IP地址”。
    - 选择安全组：您可以同时勾选多个安全组，此时，弹性云服务器的访问规则遵循几个安全组规则的并集。
6. 单击“确定”。



## 后续任务

部分操作系统无法识别新添加的网卡，需手动激活网卡。下面以Ubuntu系统为例介绍具体激活网卡的操作步骤，其他操作系统请自行完成相关操作，如有问题，请参见对应操作系统的官网指导或手册来完成操作。

1. 在弹性云服务器所在行的“操作”列下，单击“远程登录”。  
登录弹性云服务器。
2. 执行如下命令，查看网卡名称。  
**ifconfig -a**  
例如，查询到的网卡名为：eth2。
3. 执行如下命令，进入相应目录。  
**cd /etc/network**
4. 执行如下命令，打开interfaces文件。  
**vi interfaces**
5. 在interfaces文件中，增加类似如下信息。  
**auto eth2**  
**iface eth2 inet dhcp**
6. 执行如下命令，保存并退出interfaces文件。  
**:wq**
7. 执行命令**ifup ethX**或**/etc/init.d/networking restart**，使新增网卡生效。  
上述命令中的X为具体的网卡名称序号，例如，**ifup eth2**。
8. 执行如下命令，查看回显信息中是否包括2查询到的网卡。  
**ifconfig**  
例如，回显信息中包含网卡eth2。
  - 是，表示新增网卡生效，结束。
  - 否，表示新增网卡未生效，执行9。
9. 登录管理控制台，在弹性云服务器所在行的“操作”列下，选择“更多”，并单击“重启”。
10. 再次执行命令**ifconfig**，查看回显信息中是否包括2查询到的网卡。
  - 是，结束。
  - 否，请联系管理员获取技术支持。

## 7.2 解绑弹性网卡

### 操作场景

弹性云服务器最多可以有12个网卡，其中包括一个主网卡，且主网卡不可删除。本节操作介绍解绑弹性网卡的操作步骤，当您的弹性云服务器需要解绑主网卡以外的其他网卡时，可以参考以下操作步骤。

### 操作步骤

1. 登录管理控制台。

2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表中，单击待删除网卡的弹性云服务器名称。  
系统跳转至该弹性云服务器详情页面。
4. 选择“弹性网卡”页签，并单击待解绑弹性网卡右侧的“解绑”。

#### 说明

对于该弹性云服务器的主网卡（默认为网卡列表中显示的第一个网卡），用户不能执行解绑操作。

5. 在弹出的对话框中单击“是”，解绑弹性网卡。

#### 说明

对于部分弹性云服务器，不支持在线解绑弹性网卡功能，具体以界面显示为准。您需要先关机弹性云服务器，然后再执行解绑弹性网卡操作。

## 7.3 修改私有 IP 地址

### 操作场景

云平台支持修改主网卡的私有IP地址，具体操作请参见本节内容。如需修改扩展网卡的私有IP地址，请删除网卡，并挂载新网卡。

### 约束与限制

- 弹性云服务器已关机。
- 如果网卡绑定了虚拟IP或者DNAT规则，需要先解绑。
- 如果网卡上有IPv6地址，无法修改（包括IPv4和IPv6的）私有IP地址。
- 如需修改弹性负载均衡后端服务器的私有IP地址，请先移出后端服务器组后再修改私有IP。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 单击待修改私有IP地址的弹性云服务器名称。  
系统跳转至该弹性云服务器详情页面。
4. 选择“弹性网卡”页签，并单击主网卡所在行的“修改私有IP地址”。  
系统打开“修改私有IP地址”窗口。
5. 请根据需要修改主网卡的“子网”、“私有IP地址”。

#### 说明

只能在同一VPC下更换子网。

如果未填写修改后的“私有IP地址”，系统会自动分配一个新的私有IP地址给主网卡使用。

## 7.4 管理虚拟 IP 地址

### 操作场景

虚拟IP地址用于为网卡提供第二个IP地址，同时支持与多个的网卡绑定，从而实现多个弹性云服务器之间的高可用性。

### 绑定虚拟 IP 地址

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表中，单击待绑定虚拟IP地址的弹性云服务器名称。  
系统跳转至该弹性云服务器详情页面。
4. 选择“弹性网卡”页签，单击“管理虚拟IP地址”。
5. 选择“IP地址管理”页签，在需要绑定弹性公网IP或者弹性云服务器的虚拟IP地址所在行的操作列下，单击“绑定弹性公网IP”或者“绑定服务器”。  
多个主备部署的弹性云服务器可以在绑定虚拟IP地址时选择同一个虚拟IP地址，增强容灾性能。
6. 单击“确定”。

## 7.5 开启网卡多队列功能

### 操作场景

随着网络IO的带宽不断提升，单核CPU处理网络中断存在瓶颈，不能完全满足网卡的需求，通过开启网卡多队列功能，您可以将弹性云服务器中的网卡中断分散给不同的CPU处理，以满足网卡的需求，从而提升网络PPS和带宽性能。

假设以下场景所述的弹性云服务器满足规格和虚拟化类型要求：

- 使用[网卡多队列支持列表](#)中的公共镜像创建的弹性云服务器，默认已开启网卡多队列，无需执行本节操作。
- 对于私有镜像场景，如果您的外部镜像文件的操作系统在[网卡多队列支持列表](#)范围内，需要按照如下流程开启网卡多队列：
  - a. [将外部镜像文件导入镜像服务控制台](#)。
  - b. [为私有镜像添加网卡多队列标签](#)。
  - c. [使用私有镜像创建弹性云服务器](#)。
  - d. [开启网卡多队列](#)。

### 网卡多队列支持列表

网卡多队列的支持情况和实例规格、虚拟化类型、镜像的操作系统有关，只有同时满足这些要求，弹性云服务器才能开启网卡多队列功能。

- 支持网卡多队列的实例规格请参见“[实例规格（x86）](#)”。

**说明**

网卡多队列数为大于1的值，表示支持网卡多队列。

- 虚拟化类型必须为KVM。
- [表7-2](#)所列的Linux公共镜像，支持网卡多队列。

**说明**

- Windows操作系统公共镜像暂未支持网卡多队列，如果对Windows操作系统镜像开启网卡多队列功能，可能会引起操作系统启动速度变慢等问题。
- Linux操作系统云服务器建议将操作系统内核版本升级至2.6.35及以上，否则不支持网卡多队列。

建议您使用命令**uname -r**查询内核版本，如果低于2.6.35请联系技术支持升级内核。

**表 7-1 Linux 弹性云服务器网卡多队列支持列表**

镜像	是否支持多队列	是否默认开启多队列
Ubuntu 14.04/16.04/18.04/20.04 server 64bit	是	是
OpenSUSE 42.2/15.* 64bit	是	是
SUSE Enterprise 12 SP1/SP2 64bit	是	是
CentOS 6.8/6.9/7.*/8.* 64bit	是	是
Debian 8.0.0/8.8.0/8.9.0/9.0.0/10.0.0/10.2.0 64bit	是	是
Fedora 24/25/30 64bit	是	是
EulerOS 2.2/2.3/2.5 64bit	是	是

**表 7-2 KVM 弹性云服务器网卡多队列支持列表**

类别	镜像	支持度
Windows	Windows Server 2008 WEB R2 64bit	可通过私有镜像支持
	Windows Server 2008 R2 Standard/Datacenter/Enterprise 64bit	可通过私有镜像支持
	Windows Server 2012 R2 Standard/Datacenter 64bit	可通过私有镜像支持
	Windows Server 2016 Standard/Datacenter 64bit	可通过私有镜像支持
Linux	Ubuntu 14.04/16.04 server 64bit	支持
	OpenSUSE 42.2 64bit	支持
	SUSE Enterprise 12 SP1/SP2 64bit	支持

类别	镜像	支持度
	CentOS 6.8/6.9/7.0/7.1/7.2/7.3/7.4/7.5/7.6 64bit	支持
	Debian 8.0.0/8.8.0/8.9.0/9.0.0 64bit	支持
	Fedora 24/25 64bit	支持
	EulerOS 2.2 64bit	支持

## 将外部镜像文件导入镜像服务控制台

将外部镜像文件导入镜像服务控制台，详细操作请参见《镜像服务用户指南》中“注册镜像”章节。

## 为镜像添加网卡多队列标签

Windows操作系统暂未商用支持网卡多队列，如果对Windows操作系统镜像添加网卡多队列标签，开启网卡多队列功能，可能会引起操作系统启动速度变慢等问题。

用户可以选择以下任一种方式设置镜像的网卡多队列属性。

### 方式1:

1. 登录管理控制台。
2. 选择“计算 > 镜像服务”。
3. 单击“私有镜像”页签，在对应镜像所在行的“操作”列下，单击“修改”。
4. 设置镜像的网卡多队列属性。

### 方式2:

1. 登录管理控制台。
2. 选择“计算 > 镜像服务”。
3. 单击“私有镜像”页签，在镜像列表中，单击镜像名称，进入镜像详情页面。
4. 单击右上角的“修改”，在弹出的“修改镜像”对话框中，设置镜像的网卡多队列属性。

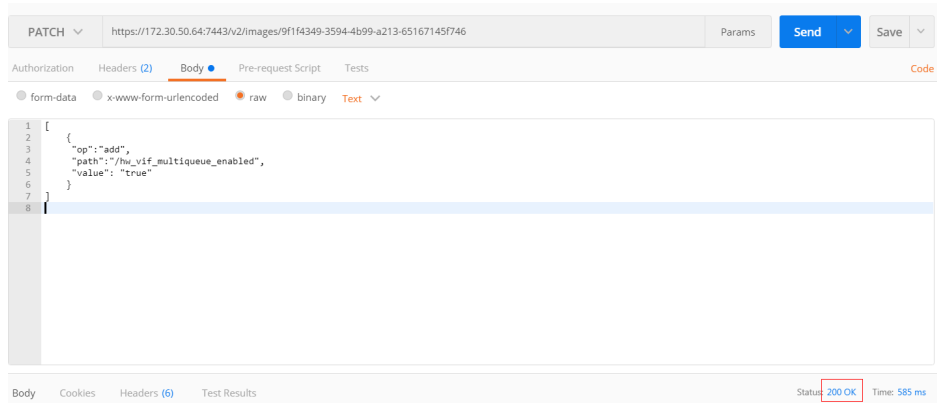
### 方式3: 通过API为镜像添加网卡多队列标签hw\_vif\_multiqueue\_enabled

1. 获取Token的方法请参考《镜像服务API参考》中的“如何调用API > 认证鉴权”。
2. “更新镜像信息”的API使用方法请参考《镜像服务API参考》中的“更新镜像信息（OpenStack原生）”章节。
3. 在请求消息头中增加“X-Auth-Token”。  
“X-Auth-Token”的取值为步骤1中获取的Token。
4. 在请求消息头中增加“Content-Type”。  
“Content-Type”取值为application/openstack-images-v2.1-json-patch  
请求的URI格式为：  
PATCH /v2/images/{image\_id}  
请求的body体如下所示。

```
[
  {
    "op": "add",
    "path": "/hw_vif_multiqueue_enabled",
    "value": "true"
  }
]
```

支持网卡多队列修改样例如图7-1所示。

图 7-1 支持网卡多队列修改样例



## 使用私有镜像创建弹性云服务器

使用注册好的私有镜像创建云服务器，详细操作请参见“[创建弹性云服务器](#)”。在配置参数时，需要注意以下两点：

- 区域：必须选择私有镜像所在的区域。
- 镜像：选择“私有镜像”，并在下拉列表中选择需要的镜像。

## 开启网卡多队列

使用Windows操作系统的KVM弹性云服务器，可通过私有镜像支持网卡多队列功能，具体操作请参见《[镜像服务用户指南](#)》的“[如何设置镜像的网卡多队列属性](#)”章节。

使用Linux操作系统的弹性云服务器，本节以CentOS 7.4为例，介绍开启网卡多队列的具体操作：

### 步骤1 开启网卡多队列功能

1. 登录弹性云服务器。
2. 执行以下命令，查看网卡支持和已开启的队列数。

```
ethtool -l 网卡
```

3. 执行以下命令，设置网卡当前使用的队列数。

```
ethtool -L 网卡 combined 队列数
```

示例：

```
[root@localhost ~]# ethtool -l eth0 #查询网卡eth0的队列数
Channel parameters for eth0:
Pre-set maximums:
RX:          0
TX:          0
Other:       0
```

```
Combined: 4 #表示此网卡最多支持设置开启4个队列
Current hardware settings:
RX:          0
TX:          0
Other:       0
Combined: 1 #表示当前开启的是1个队列

[root@localhost ~]# ethtool -L eth0 combined 4 #设置eth0网卡开启4个队列
```

**步骤2** （可选）建议开启irqbalance服务，让系统自动调整网卡中断在多个CPU核上的分配。

1. 执行以下命令，开启irqbalance服务。

```
service irqbalance start
```

2. 执行以下命令，查询irqbalance服务状态。

```
service irqbalance status
```

如果回显中“Active”字段的值包含“active(running)”，表示irqbalance服务已正常开启，如图7-2所示。

**图 7-2 irqbalance 服务正常开启**

```
[root@localhost ~]# service irqbalance status
Redirecting to /bin/systemctl status irqbalance.service
● irqbalance.service - irqbalance daemon
   Loaded: loaded (/usr/lib/systemd/system/irqbalance.service; enabled; vendor preset: enabled)
   Active: active (running) since Wed 2018-08-15 10:27:30 CST; 4h 5min ago
     Main PID: 858 (irqbalance)
    CGroup: /system.slice/irqbalance.service
            └─858 /usr/sbin/irqbalance --foreground

Aug 15 10:27:30 localhost.localdomain systemd[1]: Started irqbalance daemon.
Aug 15 10:27:30 localhost.localdomain systemd[1]: Starting irqbalance daemon...
```

**步骤3** （可选）开启中断绑定。

开启irqbalance服务让中断自由调度，网络性能提升。如果提升后的网络性能仍不满足您的预期，可以手动设置弹性云服务器操作系统的中断亲和性。

具体操作如下：

执行以下脚本，使得弹性云服务器中每个CPU服务一个多队列的中断请求，一对一的对应设置。即一个队列对应一个中断号，一个中断号绑定一个CPU，一一对应。

```
#!/bin/bash
service irqbalance stop

eth_dirs=$(ls -d /sys/class/net/eth*)
if [ $? -ne 0 ];then
    echo "Failed to find eth* , sleep 30" >> $ecs_network_log
    sleep 30
    eth_dirs=$(ls -d /sys/class/net/eth*)
fi

for eth in $eth_dirs
do
    cur_eth=$(basename $eth)
    cpu_count=`cat /proc/cpuinfo| grep "processor"| wc -l`
    virtio_name=$(ls -l /sys/class/net/"$cur_eth"/device/driver/ | grep pci |awk '{print $9}')

    affinity_cpu=0
    virtio_input="$virtio_name"-input"
    irqs_in=$(grep "$virtio_input" /proc/interrupts | awk -F ":" '{print $1}')
    for irq in ${irqs_in[*]}
    do
        echo $((affinity_cpu%cpu_count)) > /proc/irq/"$irq"/smp_affinity_list
        affinity_cpu=$((affinity_cpu+2))
    done

    affinity_cpu=1
    virtio_output="$virtio_name"-output"
```

```
irqs_out=$(grep "$virtio_output" /proc/interrupts | awk -F ":" '{print $1}')
for irq in ${irqs_out[*]}
do
    echo $$((affinity_cpu%cpu_count)) > /proc/irq/"$irq"/smp_affinity_list
    affinity_cpu=$((affinity_cpu+2))
done
done
```

#### 步骤4 （可选）开启XPS、RPS特性。

xps主要是针对多队列的网卡发送时的优化，当发送一个数据包的时候，它会根据CPU来选择对应的队列。

```
#!/bin/bash
# enable XPS feature
cpu_count=$(grep -c processor /proc/cpuinfo)
dec2hex(){
    echo $(printf "%x" $1)
}
eth_dirs=$(ls -d /sys/class/net/eth*)
if [ $? -ne 0 ];then
    echo "Failed to find eth* , sleep 30" >> $secs_network_log
    sleep 30
    eth_dirs=$(ls -d /sys/class/net/eth*)
fi
for eth in $eth_dirs
do
    cpu_id=1
    cur_eth=$(basename $eth)
    cur_q_num=$(ethtool -l $cur_eth | grep -iA5 current | grep -i combined | awk '{print $2}')
    for((i=0;i<cur_q_num;i++))
    do
        if [ $i -eq $cpu_count ];then
            cpu_id=1
            fi
        xps_file="/sys/class/net/${cur_eth}/queues/tx-$i/xps_cpus"
        rps_file="/sys/class/net/${cur_eth}/queues/rx-$i/rps_cpus"
        cpuset=$(dec2hex "$cpu_id")
        echo $cpuset > $xps_file
        echo $cpuset > $rps_file
        let cpu_id=cpu_id*2
    done
done
```

----结束

## 7.6 动态获取 IPv6 地址

### 操作场景

IPv6的使用，可以有效弥补IPv4网络地址资源有限的问题。如果当前云服务器使用IPv4，那么启用IPv6后，云服务器可在双栈模式下运行，即云服务器可以拥有两个不同版本的IP地址：IPv4地址和IPv6地址，这两个IP地址都可以进行内网/公网访问。

按照[约束与限制](#)中的网络环境要求创建的云服务器，有些不能动态获取到IPv6地址，需要进行相关配置才行。如果云服务器使用的是公共镜像，则支持情况如下：

- Windows公共镜像默认已开启IPv6动态获取功能，无需配置，文中的[Windows 2012操作系统](#)和[Windows 2008操作系统](#)部分供您验证、参考。
- Linux公共镜像开启动态获取IPv6功能时，需要先判断是否支持IPv6协议栈，再判断是否已开启动态获取IPv6。目前，所有Linux公共镜像均已支持IPv6协议栈。



## 约束与限制

- 请确保云服务器所在的子网已开启IPv6功能。  
若云服务器所在子网未开启IPv6功能，需参考[开启云服务器的IPv6功能](#)进行开启，开启后不允许关闭。
- 请确保创建云服务器时已选择“自动分配IPv6地址”。
- 云服务器启动之后动态插拔的网卡不支持IPv6地址动态获取功能。
- 仅弹性云服务器支持IPv6双栈，裸金属服务器不支持。
- 同一个网卡上，只能绑定一个IPv6地址。

## 操作导航

- Windows系统：本文以Windows 2012版本、Windows 2008版本为例，介绍Windows操作系统启用IPv6的方法，如[表7-3](#)所示。
- Linux系统：本文提供了自动配置、手动配置两种方式启用IPv6，推荐您使用自动配置方法，如[表7-3](#)所示。  
对于CentOS 6.x和Debian操作系统，云服务器内部配置IPv6自动获取功能之后，将该云服务器制作为私有镜像，使用该镜像在非IPv6网络环境中创建云服务器时，由于等待获取IPv6地址超时，导致云服务器启动较慢，此时您可以参考[设置云服务器获取IPv6地址超时时间](#)操作。

表 7-3 不同操作系统启用 IPv6 操作指导

操作系统	方式	操作指导
Windows 2012	自动配置启用IPv6	<a href="#">Windows 2012操作系统</a>
Windows 2008	自动配置启用IPv6	<a href="#">Windows 2008操作系统</a>
Linux	自动配置启用IPv6（推荐）	<a href="#">Linux操作系统（自动配置启用IPv6）</a>
Linux	手动配置启用IPv6	<a href="#">Linux操作系统（手动配置启用IPv6）</a>

## 开启云服务器的 IPv6 功能

### 开启子网的IPv6网段

#### 📖 说明

云服务器所属子网的IPv6功能开启后会自动分配IPv6网段，开启后不允许关闭。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 单击待开启IPv6功能的弹性云服务器，进入详情页面。
4. 单击“虚拟私有云”名称，进入弹性云服务器所属虚拟私有云列表。
5. 在虚拟私有云列表中，单击“子网个数”列对应的数字超链接。  
进入子网列表页面。

6. 在子网列表中，单击待修改的子网名称超链接。  
进入子网详情页面。
7. 在子网详情页，单击“开启IPv6”。
8. 单击“是”，完成子网IPv6网段的开启。

### 开启云服务器网卡的IPv6功能

1. 返回弹性云服务器详情页面。
2. 在“弹性网卡”页签，单击网卡折叠面板右上角的“开启IPv6”。

#### 📖 说明

- 若不再使用弹性云服务器的IPv6功能，可以在当前页面单击“关闭IPv6”，关闭网卡的IPv6功能，关闭后该网卡的“IPv6地址”为空。
  - 关闭IPv6后，如果重新开启弹性云服务器的IPv6，在重启云服务器后，需要登录弹性云服务器手动清理IPv6缓存，并重新请求获取IPv6地址。
3. 单击“是”，完成开启网卡的IPv6功能。

## Windows 2012 操作系统

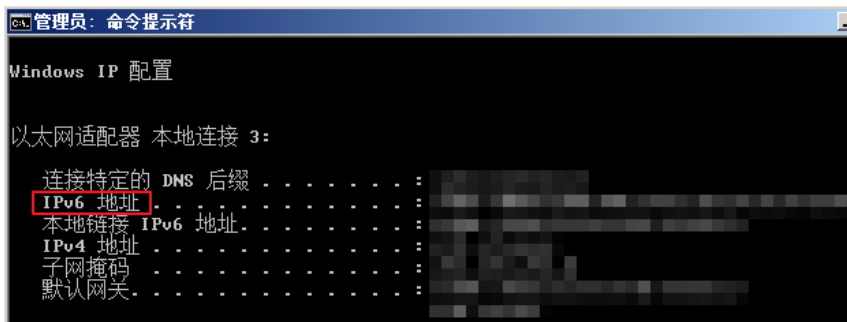
### 步骤1 检查是否启用IPv6。

打开cmd窗口，执行如下命令，查看当前云服务器是否启用IPv6。

#### ipconfig

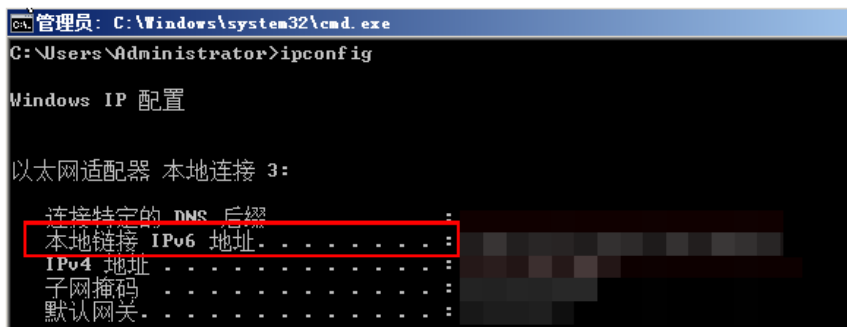
- 如果已启用IPv6，则会显示IPv6的地址。

图 7-3 显示 IPv6 的地址



- 如果显示只有本地链接IPv6地址，则表示无法动态获取到IPv6地址。请执行[步骤 2](#)。

图 7-4 本地链接 IPv6 地址



- 如果未启用IPv6，则不会显示IPv6的地址。请执行**步骤3**。

图 7-5 未启用 IPv6



### 说明

Windows公共镜像默认已经配置了IPv6动态获取功能，即回显如图7-3所示，无需特殊配置。

### 步骤2 配置动态获取IPv6。

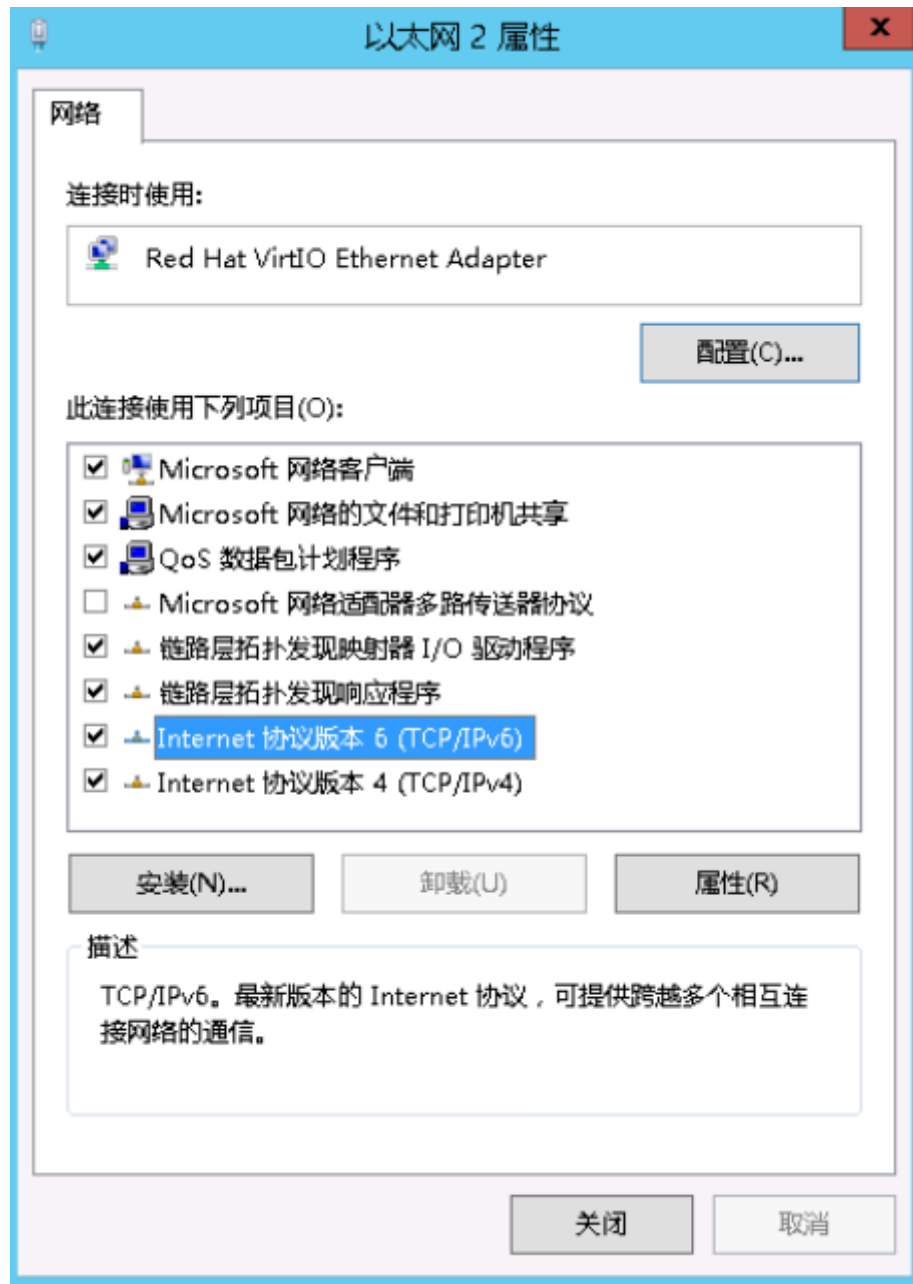
1. 单击“开始 > 控制面板”。
2. 单击“网络和共享中心”。
3. 单击以太网连接。

图 7-6 以太网连接



4. 在以太网状态的弹窗中单击左下角的“属性”。
5. 勾选“Internet协议版本 6 (TCP/IPv6)”，然后单击“确定”。

图 7-7 配置动态获取 IPv6

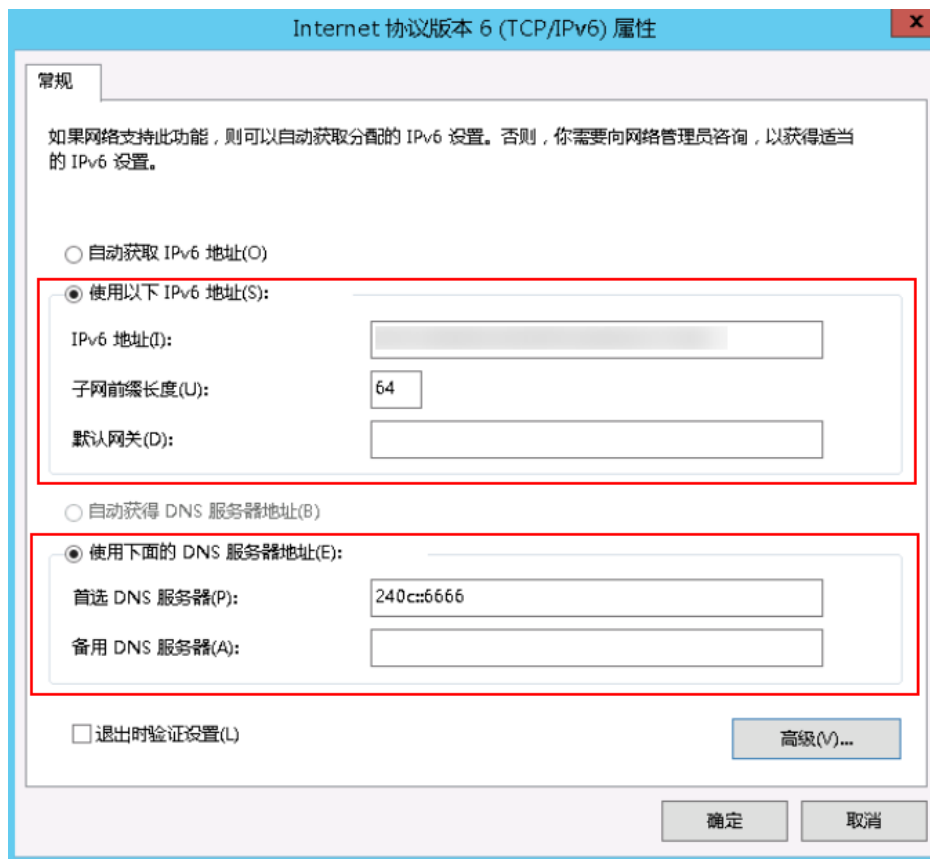


6. 执行**步骤1**检查是否已开启动态获取IPv6。

**步骤3** 启用和配置IPv6。

1. 在“Internet 协议版本 6 (TCP/IPv6) 属性”弹窗中，配置云服务器的IPv6地址和DNS服务器地址。
  - IPv6地址：创建云服务器时分配的IPv6地址，请从控制台云服务器的列表页面获取。
  - 子网前缀长度：64
  - DNS服务器：推荐使用240c::6666

图 7-8 在控制台获取 IPv6 地址



2. （可选配置）根据操作系统不同请分别执行以下命令。  
Windows Server 2012操作系统云服务器请在PowerShell或者cmd中执行如下命令：

**Set-NetIPv6Protocol -RandomizeIdentifiers disabled**

3. 执行**步骤1**检查是否已开启动态获取IPv6。

----结束

## Windows 2008 操作系统

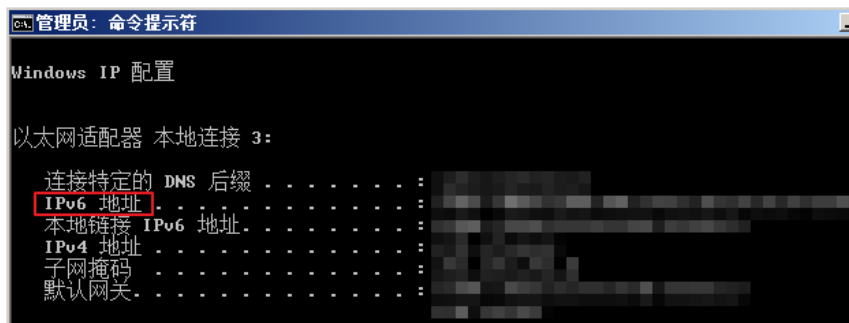
**步骤1** 检查是否启用IPv6。

打开cmd窗口，执行如下命令，查看当前云服务器是否启用IPv6。

**ipconfig**

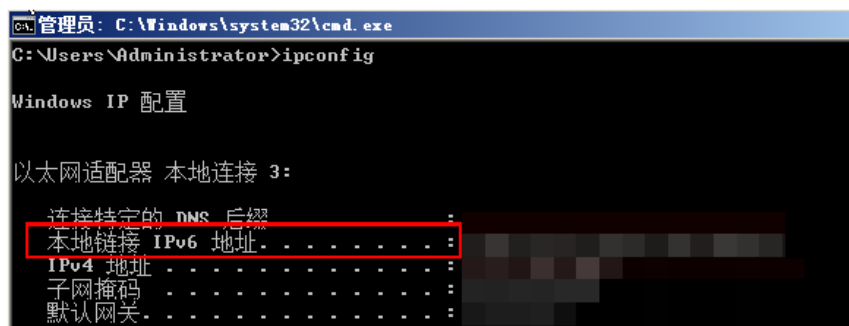
- 如果已启用IPv6，则会显示IPv6的地址。

图 7-9 显示 IPv6 的地址



- 如果显示只有本地链接IPv6地址，则表示无法动态获取到IPv6地址。请执行[步骤2](#)。

图 7-10 本地链接 IPv6 地址



- 如果未启用IPv6，则不会显示IPv6的地址。请执行[步骤3](#)。

图 7-11 未启用 IPv6



### 说明

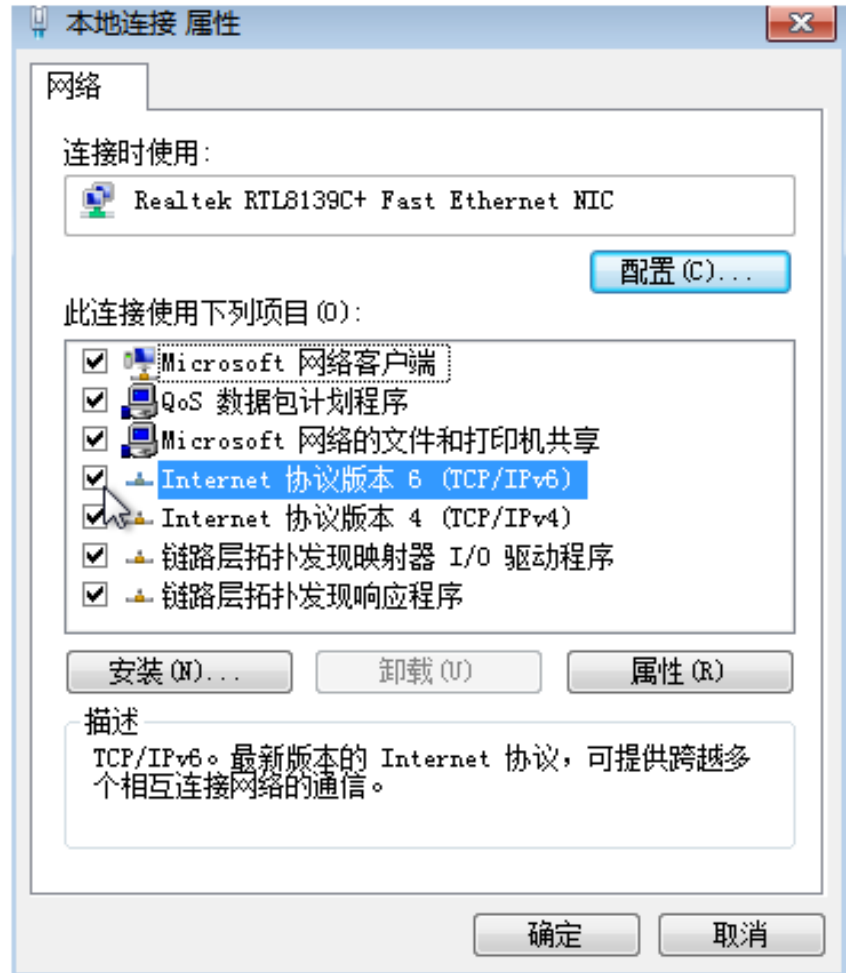
Windows公共镜像默认已经配置了IPv6动态获取功能，即回显如[图7-9](#)所示，无需特殊配置。

#### 步骤2 配置动态获取IPv6。

1. 单击“开始 > 控制面板”。
2. 单击“网络和共享中心”。

3. 左键单击“更改适配器设置”。
4. 右键单击网络连接并选择“属性”。
5. 勾选“Internet协议版本 6 (TCP/IPv6)”，然后单击“确定”。

图 7-12 配置动态获取 IPv6

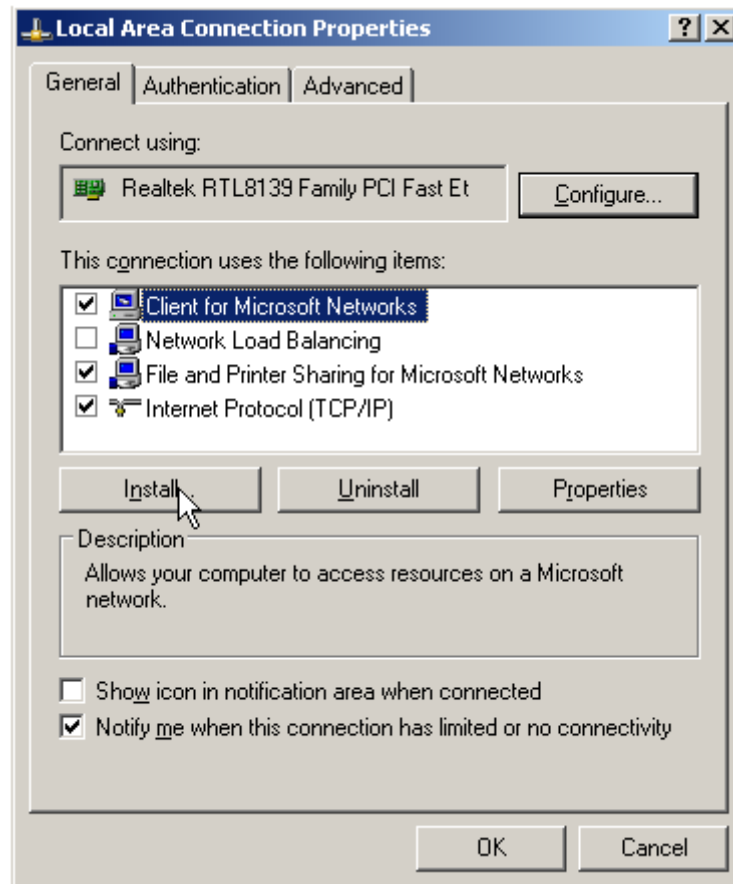


6. 执行**步骤1**检查是否已开启动态获取IPv6。

**步骤3** 启用和配置IPv6。

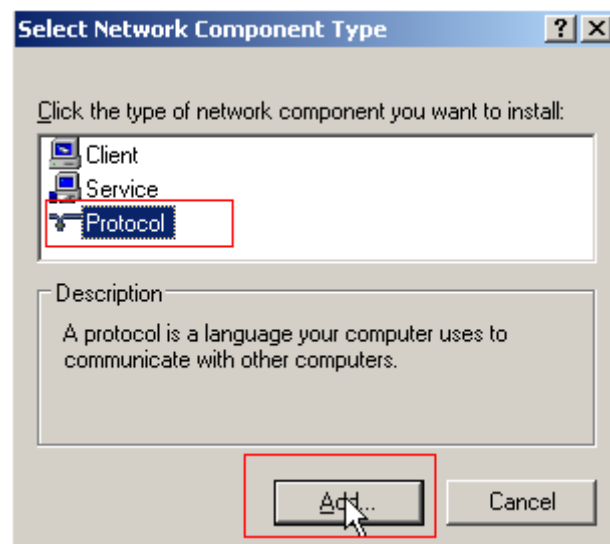
1. 选择“开始 > 控制面板 > 网络连接 > 本地连接”。
2. 选择“属性”，确认勾选以下选项后单击“安装”。

图 7-13 启用和配置 IPv6



3. 选择“协议”，然后单击“添加”。

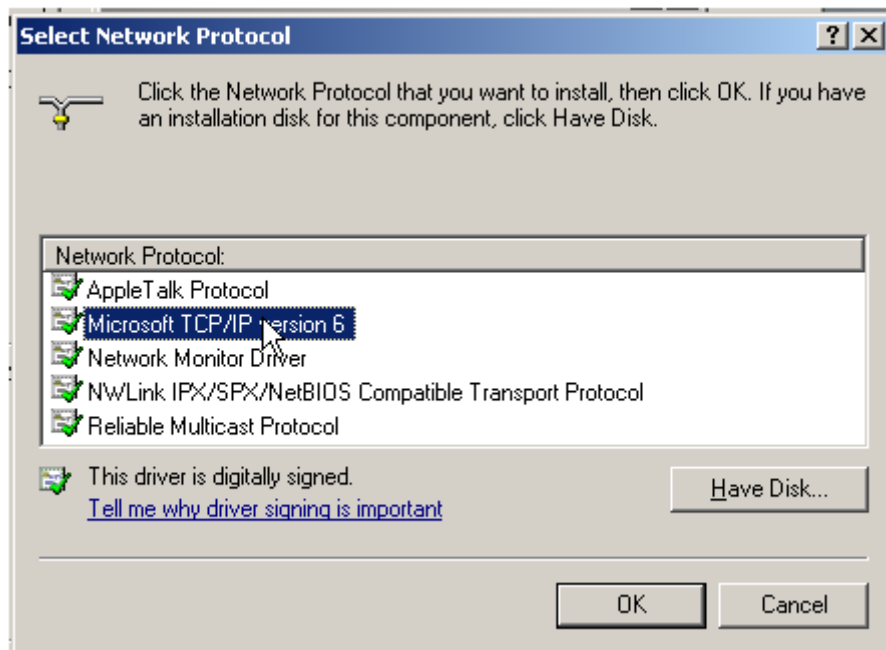
图 7-14 添加协议



4. 在网络协议列表中选择“Microsoft TCP/IP版本 6”，然后单击“确定”。



图 7-15 网络协议列表



5. （可选配置）根据操作系统不同请分别执行以下命令。

Windows Server 2008操作系统云服务器请在PowerShell或者cmd中执行如下命令：

```
netsh interface ipv6 set global randomizeidentifiers=disable
```

设置云服务器先禁用本地连接，再重启本地连接。

禁用本地连接：单击“开始 > 控制面板 > 网络和共享中心 > 更改适配器配置”，选择本地连接，单击右键选择“禁用”。

重启本地连接：单击“开始 > 控制面板 > 网络和共享中心 > 更改适配器配置”，选择本地连接，单击右键选择“启用”。

6. 执行[步骤1](#)检查是否已开启动态获取IPv6。

----结束

## Linux 操作系统（自动配置启用 IPv6）

ipv6-setup-xxx工具能为开启IPv6协议栈的Linux操作系统自动配置动态获取IPv6地址。其中，xxx表示工具系列：rhel或debian。

您也可以参考[Linux操作系统（手动配置启用IPv6）](#)手动配置启用IPv6。

### ⚠ 注意

- ipv6-setup-xxx工具运行时会自动重启网络服务，导致网络短暂不可用。
- CentOS 6.x和Debian操作系统的云服务器内部配置IPv6自动获取功能之后，将该云服务器制作作为私有镜像，使用该镜像在非IPv6网络环境中创建云服务器时，由于等待获取IPv6地址超时，导致云服务器启动较慢，您可以参考[设置云服务器获取IPv6地址超时时间](#)设置获取IPv6地址超时时间为30s，然后再重新制作私有镜像。

**步骤1** 执行如下命令，查看当前云服务器是否启用IPv6。

**ip addr**

- 如果没有开启IPv6协议栈，则只能看到IPv4地址，如下图所示，请参考[设置云服务器获取IPv6地址超时时间](#)先开启IPv6协议栈。

**图 7-16** 云服务器未开启 IPv6 协议栈

```
eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP qlen 1000
link/ether fa:16:3e: bnd ff:ff:ff:ff:ff:ff
inet bnd bnd scope global noprefixroute dynamic eth0
valid_lft 1193sec preferred_lft 1193sec
```

- 如果已开启IPv6协议栈，则可以看到LLA地址（fe80开头）。

**图 7-17** 云服务器已开启 IPv6 协议栈

```
eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group default qlen 1000
link/ether fa:16:3e:75:af:4c bnd ff:ff:ff:ff:ff:ff
inet bnd bnd scope global noprefixroute dynamic eth0
valid_lft 76391sec preferred_lft 76391sec
inet6 fe80::f816: /64 scope link
valid_lft forever preferred_lft forever
```

- 如果已开启IPv6协议栈并且已获取到IPv6地址，则会看到如下地址：

**图 7-18** 云服务器已开启 IPv6 协议栈并且已获取到 IPv6 地址

```
eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group default qlen 1000
link/ether fa:16:3e:75:af:4c bnd ff:ff:ff:ff:ff:ff
inet bnd bnd scope global noprefixroute dynamic eth0
valid_lft 86395sec preferred_lft 86395sec
inet6 2407:c080:802: /128 scope global dynamic
valid_lft 7496sec preferred_lft 7196sec
inet6 fe80::f816:3eff: /64 scope link noprefixroute
valid_lft forever preferred_lft forever
```

### 说明

Linux公共镜像均已开启IPv6协议栈，如[图7-17](#)所示；

**步骤2** 开启Linux云服务器IPv6协议栈。

1. 执行如下命令，确认内核是否支持IPv6协议栈。

**sysctl -a | grep ipv6**

- 如果有输出信息，表示内核支持IPv6协议栈。
- 如果没有任何输出，说明内核不支持IPv6协议栈，需要执行[2.b](#)加载IPv6模块。

2. 执行以下命令，加载IPv6模块。

**modprobe ipv6**

3. 修改“/etc/sysctl.conf”配置文件，增加如下配置：

**net.ipv6.conf.all.disable\_ipv6=0**

4. 保存配置并退出，然后执行如下命令，加载配置。

**sysctl -p**

**步骤3** 自动配置启用IPv6。

1. 下载对应系统版本的工具ipv6-setup-rhel或ipv6-setup-debian，并上传至待操作的云服务器。

ipv6-setup-xxx工具会添加或者修改网卡设备的配置文件，添加IPv6动态获取的配置信息，然后重启网卡或者网络服务。

ipv6-setup-rhel和ipv6-setup-debian的工具下载地址请联系管理员获取。

2. 执行以下命令，添加执行权限。

```
chmod +x ipv6-setup-xxx
```

3. 执行以下命令，指定一个网卡设备，配置动态获取IPv6地址。

```
./ipv6-setup-xxx --dev [dev]
```

示例：

```
./ipv6-setup-xxx --dev eth0
```

#### 📖 说明

- 如需对所有网卡配置动态获取IPv6地址，命令为./ipv6-setup-xxx，即不带参数。
- 如需查询工具的用法，请执行命令./ipv6-setup-xxx --help。

----结束

## Linux 操作系统（手动配置启用 IPv6）

### ⚠️ 注意

CentOS 6.x和Debian操作系统的云服务器内部配置IPv6自动获取功能之后，将该云服务器制作作为私有镜像，使用该镜像在非IPv6网络环境中创建云服务器时，由于等待获取IPv6地址超时，导致云服务器启动较慢，您可以参考[设置云服务器获取IPv6地址超时时间](#)设置获取IPv6地址超时时间为30s，然后再重新制作私有镜像。

- 步骤1 执行如下命令，查看当前云服务器是否启用IPv6。

#### ip addr

- 如果没有开启IPv6协议栈，则只能看到IPv4地址，如下图所示，请参考[步骤2](#)先开启IPv6协议栈。

图 7-19 未开启 IPv6 协议栈

```
eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP qlen 1000
link/ether fa:16:3e: b8:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
inet 10.0.0.1 brd 10.0.0.255 scope global noprefixroute dynamic eth0
    valid_lft 1193sec preferred_lft 1193sec
```

- 如果已开启IPv6协议栈，则可以看到LLA地址（fe80开头）。

图 7-20 已开启 IPv6 协议栈

```
eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group default qlen 1000
link/ether fa:16:3e:75:af:4c brd ff:ff:ff:ff:ff:ff
inet 10.0.0.1 brd 10.0.0.255 scope global noprefixroute dynamic eth0
    valid_lft 76391sec preferred_lft 76391sec
inet6 fe80::f816:3e75:af4c:1/64 scope link
    valid_lft forever preferred_lft forever
```

- 如果已开启IPv6协议栈并且已获取到IPv6地址，则会看到如下地址：

图 7-21 已开启 IPv6 协议栈并且已获取到 IPv6 地址

```
eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group default qlen 1000
link/ether fa:16:3e:75:af:4c brd ff:ff:ff:ff:ff:ff
inet 10.0.0.1 brd 10.0.0.255 scope global noprefixroute dynamic eth0
    valid_lft 86395sec preferred_lft 86395sec
inet6 2407:c080:802:1::1/128 scope global dynamic
    valid_lft 7496sec preferred_lft 7196sec
inet6 fe80::f816:3eff:1::1/64 scope link noprefixroute
    valid_lft forever preferred_lft forever
```

## 📖 说明

Linux公共镜像均已开启IPv6协议栈，如图7-20所示；

### 步骤2 开启Linux云服务器IPv6协议栈。

1. 执行如下命令，确认内核是否支持IPv6协议栈。

```
sysctl -a | grep ipv6
```

- 如果有输出信息，表示内核支持IPv6协议栈。
- 如果没有任何输出，说明内核不支持IPv6协议栈，需要执行步骤2.2加载IPv6模块。

2. 执行以下命令，加载IPv6模块。

```
modprobe ipv6
```

3. 修改“/etc/sysctl.conf”配置文件，增加如下配置：

```
net.ipv6.conf.all.disable_ipv6=0
```

4. 保存配置并退出，然后执行如下命令，加载配置。

```
sysctl -p
```

### 步骤3 手动配置启用IPv6。操作系统不同，步骤有所差别。

- Ubuntu 18.04/20.04操作系统云服务器配置动态获取IPv6。

- a. 执行以下命令，进入“/etc/netplan/”。

```
cd /etc/netplan
```

- b. 执行以下命令，查询配置文件名。

```
ls
```

图 7-22 查询配置文件名

```
root@ecs- :/etc/netplan# ls  
01-netcfg.yaml 01-network-manager-all.yaml
```

- c. 执行以下命令，编辑“01-network-manager-all.yaml”配置文件。

```
vi 01-network-manager-all.yaml
```

- d. 在“01-network-manager-all.yaml”下增加如下内容，注意yaml文件格式及缩进：

```
ethernets:  
  eth0:  
    dhcp6: true
```

图 7-23 修改结果

```
# Let NetworkManager manage all devices on this system  
network:  
  version: 2  
  renderer: NetworkManager  
  ethernets:  
    eth0:  
      dhcp6: true
```

修改完成后保存退出。

- e. 执行以下命令，使更改生效。

```
sudo netplan apply
```

- Ubuntu 22.04操作系统云服务器配置动态获取IPv6。
  - a. 执行以下命令，进入“/etc/netplan/”。

```
cd /etc/netplan
```

- b. 执行以下命令，查询配置文件名。

```
ls
```

图 7-24 查询配置文件名

```
root@ecs-485b:/etc/netplan# ls
01-netcfg.yaml
```

- c. 执行以下命令，编辑“01-netcfg.yaml”配置文件。

```
vi 01-netcfg.yaml
```

- d. 在“01-netcfg.yaml”中增加如下内容，注意yaml文件格式及缩进：

```
ethernets:
  eth0:
    dhcp6: true
```

图 7-25 修改结果

```
network:
  version: 2
  renderer: NetworkManager
  ethernets:
    eth0:
      dhcp4: true
      dhcp6: true
    eth1:
      dhcp4: true
    eth2:
      dhcp4: true
    eth3:
      dhcp4: true
    eth4:
      dhcp4: true
```

修改完成后保存退出。

- e. 执行以下命令，使更改生效。

```
sudo netplan apply
```

- f. 执行以下命令，编辑“/etc/NetworkManager/NetworkManager.conf”文件。

```
vi /etc/NetworkManager/NetworkManager.conf
```

- g. 在“NetworkManager.conf”中增加如下内容，注意文件格式及缩进：

```
[main]
plugins=ifupdown,keyfile
dhcp=dhclient

[ifupdown]
managed=true

[device]
wifi.scan-rand-mac-address=no
```

图 7-26 修改结果

```
[main]
plugins=ifupdown,keyfile
dhcp=dhclient

[ifupdown]
managed=true

[device]
wifi.scan-rand-mac-address=no
```

- h. 执行以下命令，使配置生效。

```
systemctl restart NetworkManager
```

- Debian操作系统云服务器配置动态获取IPv6。

- a. 编辑“/etc/network/interfaces”文件，使之包含以下内容：

```
auto lo
iface lo inet loopback
auto eth0
iface eth0 inet dhcp
iface eth0 inet6 dhcp
pre-up sleep 3
```

- b. 如果有多个网卡，则在“/etc/network/interfaces”文件中，增加对应网卡的配置，以eth1为例，需要增加：

```
auto eth1
iface eth1 inet dhcp
iface eth1 inet6 dhcp
pre-up sleep 3
```

- c. 执行如下命令重启网络服务。

```
service networking restart
```

#### 📖 说明

如果将网卡进行down/up操作之后无法获取IPv6地址，也可以通过此命令重启网络服务。

- d. 执行步骤[步骤1](#)检查是否已开启动态IPv6。

- CentOS/EulerOS/Fedora操作系统云服务器配置动态获取IPv6。

- a. 编辑主网卡配置文件“/etc/sysconfig/network-scripts/ifcfg-eth0”。

补充如下配置项：

```
IPV6INIT=yes
DHCPV6C=yes
```

- b. 编辑“/etc/sysconfig/network”，按如下所示添加或修改以下行。

```
NETWORKING_IPV6=yes
```

- c. CentOS 6系列从网卡需要编辑对应的配置文件，以eth1为例，编辑“/etc/sysconfig/network-scripts/ifcfg-eth1”。

补充如下配置项：

```
IPV6INIT=yes
DHCPV6C=yes
```

CentOS 6.3系统中默认ip6tables会过滤dhcpv6-client请求，所以CentOS 6.3除了需要编辑“ifcfg-eth\*”文件外，还需要额外添加一条允许dhcpv6-client请求的ip6tables规则。操作如下：

- i. 执行以下命令，添加ip6tables规则。

```
ip6tables -A INPUT -m state --state NEW -m udp -p udp --dport 546 -d fe80::/64 -j ACCEPT
```

- ii. 执行以下命令，保存ip6tables规则。

```
service ip6tables save
```

图 7-27 命令示例

```
[root@ecs-cd02 log]# ip6tables -A INPUT -m state --state NEW -m udp -p udp --dport 546 -d fe80::/64 -j ACCEPT
nf_conntrack version 0.5.0 (7964 buckets, 31856 max)
[root@ecs-cd02 log]# service ip6tables save
ip6tables: Saving firewall rules to /etc/sysconfig/ip6tablef OK ]
```

- d. (可选配置) CentOS 7/CentOS 8系列需要将扩展网卡的IPv6 LLA地址模式修改为EUI64。

- i. 执行如下命令查看网卡信息。

```
nmcli con
```

图 7-28 查看网卡信息

```
[root@ecs-166b ~]# nmcli con
NAME                UUID                                  TYPE      DEVICE
System eth0         5fb06bd0-0bb0-7ffb-45f1-d6edd65f3e03  ethernet  eth0
Wired connection 1  9c92fad9-6ecb-3e6c-eb4d-8a47c6f50c04  ethernet  eth1
Wired connection 1  3a73717e-65ab-93e8-b518-24f5af32dc0d  ethernet  eth2
```

- ii. 将eth1的IPv6 LLA地址模式按以下命令修改为EUI64:

```
nmcli con modify "Wired connection 1" ipv6.addr-gen-mode eui64
```

#### 说明

CentOS不同系列，网卡信息存在差异，命令中的“Wired connection 1”需要根据实际查询的网卡信息的“NAME”列进行替换。

- iii. 通过ifconfig命令将eth1进行down/up操作。

```
ifdown eth1
```

```
ifup eth1
```

- e. 重启网络服务。

- i. CentOS 6系列执行以下命令，重启网络服务。

```
service network restart
```

- ii. CentOS 7/EulerOS/Fedora系列执行以下命令，重启网络服务。

```
systemctl restart NetworkManager
```

- f. 执行步骤[步骤1](#)检查是否已开启动态IPv6。

- SUSE/openSUSE/CoreOS操作系统云服务器配置动态获取IPv6。

SUSE 11 SP4不支持IPv6自动获取。

SUSE 12 SP1、SUSE 12 SP2无需特殊配置。

openSUSE 13.2、openSUSE 42.2无需特殊配置。

CoreOS 10.10.5无需特殊配置。

----结束

## 设置云服务器获取 IPv6 地址超时时间

CentOS 6.x和Debian操作系统的云服务器内部配置IPv6自动获取功能之后，将该云服务器制作作为私有镜像，使用该镜像在非IPv6网络环境中创建云服务器时，由于等待获取IPv6地址超时，导致云服务器启动较慢，您可以参考本节操作设置获取IPv6地址超时时间为30s，然后再重新制作私有镜像。

- CentOS 6.x:
  - a. 执行以下命令编辑“dhclient.conf”文件。  
**vi /etc/dhcp/dhclient.conf**
  - b. 按“i”进入编辑模式，在文件中增加timeout属性。  
timeout 30;
  - c. 输入:wq保存后退出。
- Debian 7.5:
  - a. 执行以下命令编辑“networking”文件。  
**vi /etc/init.d/networking**
  - b. 按“i”进入编辑模式，增加延迟命令timeout，修改点如下图所示。

图 7-29 修改点 1

```
115 case "$1" in
116 start)
117     if init_is_upstart; then
118         exit 1
119     fi
120     process_options
121     check_ifstate
122
123     if [ "$CONFIGURE_INTERFACES" = no ]
124     then
125         log_action_msg "Not configuring network interfaces, see /etc/default/networking"
126         exit 0
127     fi
128     set -f
129     exclusions=$(process_exclusions)
130     log_action_begin_msg "Configuring network interfaces"
131     if /usr/bin/timeout 30 ifup -a $exclusions $verbose && ifup_hotplug $exclusions $verbose
132     then
133         log_action_end_msg $?
134     else
135         log_action_end_msg $?
136     fi
137     ;;
138
139 stop)
140     if init_is_upstart; then
141         exit 0
142     fi
143     check_network_file_systems
144     check_network_swap
145
146     log_action_begin_msg "Deconfiguring network interfaces"
147     if /usr/bin/timeout 30 ifdown -a --exclude=lo $verbose; then
148         log_action_end_msg $?
```

图 7-30 修改点 2

```
154 reload)
155     process_options
156
157     log_action_begin_msg "Reloading network interfaces configuration"
158     state=$(cat /run/network/ifstate)
159     if /usr/bin/timeout 30 ifdown -a --exclude=lo $verbose || true
160     if /usr/bin/timeout 30 ifup --exclude=lo $state $verbose ; then
161         log_action_end_msg $?
162     else
163         log_action_end_msg $?
164     fi
165     ;;
166
167 force-reload|restart)
168     if init_is_upstart; then
169         exit 1
170     fi
171     process_options
172
173     log_warning_msg "Running $0 $1 is deprecated because it may not re-enable some interfaces"
174     log_action_begin_msg "Reconfiguring network interfaces"
175     if /usr/bin/timeout 30 ifdown -a --exclude=lo $verbose || true
176     set -f
177     exclusions=$(process_exclusions)
178     if /usr/bin/timeout 30 ifup -a --exclude=lo $exclusions $verbose && ifup_hotplug $exclusions $verbose
179     then
180         log_action_end_msg $?
181     else
182         log_action_end_msg $?
```

- Debian 8.2.0/8.8.0
  - a. 执行以下命令编辑“network-pre.conf”文件。  
**vi /lib/systemd/system/networking.service.d/network-pre.conf**



- b. 按“i”进入编辑模式，在文件中增加timeout属性。  
[Service]  
TimeoutStartSec=30
- Debian 9.0
  - a. 执行以下命令编辑“networking.service”文件。  
**vi /etc/systemd/system/network-online.target.wants/  
networking.service**
  - b. 按“i”进入编辑模式，将TimeoutStartSec=5min改为  
TimeoutStartSec=30。

# 8 弹性公网 IP 管理

## 8.1 绑定弹性公网 IP

### 操作场景

可以通过申请弹性公网IP并将弹性公网IP绑定到弹性云服务器上，实现弹性云服务器访问公网的目的。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表中，找到待绑定弹性公网IP的弹性云服务器，单击“操作”列下的“更多 > 网络设置 > 绑定弹性公网IP”。
4. 选择弹性公网IP。

#### 说明

如果当前区域没有可用的弹性公网IP，则弹性公网IP列表为空，请创建弹性公网IP后重新执行绑定操作。

5. 单击“确定”。
- 完成绑定后，可以在云服务器列表页查看已绑定的弹性公网IP。

## 8.2 解绑弹性公网 IP

### 操作场景

本节操作介绍从云服务器解绑弹性公网IP的操作步骤。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。

3. 在弹性云服务器列表中，在待调整带宽的弹性云服务器操作列下，单击“操作”列下的“更多 > 网络设置 > 解绑弹性公网IP”。
4. 确认弹性公网IP信息，单击“确定”。

#### 说明

未绑定弹性云服务器的弹性公网IP会继续计费，如果不再使用请根据控制台的引导释放弹性公网IP。

## 8.3 修改弹性公网 IP 的带宽

### 操作场景

当弹性云服务器绑定了弹性公网IP时，通过带宽提供公网和弹性云服务器间的访问流量。本节操作介绍调整弹性云服务器带宽的操作步骤。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表中，在待调整带宽的弹性云服务器操作列下，单击“操作”列下的“更多 > 网络设置 > 修改带宽”。
4. 根据页面提示，修改“带宽名称”和“带宽大小”。

## 8.4 无公网 IP 的弹性云服务器访问 Internet

### 操作场景

为保证安全和节省公网IP资源，通常只为特定的弹性云服务器配置公网IP，可直接访问Internet，其他弹性云服务器只配置私网IP，无法直接访问Internet。因此，当只配置了私网IP的弹性云服务器需要访问Internet，执行软件升级、给系统打补丁或者其它需求时，可选择一台绑定了公网IP的弹性云服务器作为代理弹性云服务器，为其他无公网IP的云服务器提供访问通道，正常访问Internet。

#### 说明

优先推荐您使用NAT（NAT Gateway）网关服务。NAT网关能够为VPC内的弹性云服务器提供SNAT和DNAT功能，通过灵活简易的配置，即可轻松构建VPC的公网出入口。了解更多请参考《NAT网关用户指南》。

### 前提条件

- 已拥有一台绑定了公网IP的弹性云服务器作为代理弹性云服务器。
- 代理弹性云服务器和其他需要访问Internet的弹性云服务器均处于同一网段，并且在同一安全组内。


## 通过 Linux 操作系统的代理主机

### 说明

本方案需要：

- 已拥有一台绑定了公网IP的弹性云服务器作为代理弹性云服务器。
- 代理弹性云服务器和其他需要访问Internet的弹性云服务器均处于同一网段，并且在同一安全组内。

本节操作中，以代理弹性云服务器的操作系统是CentOS 7.9为例进行介绍，适用于CentOS 7.9及以下版本。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表中的右上角，输入代理云服务器名称进行搜索。
4. 单击代理弹性云服务器的名称，查看详情。
5. 在代理弹性云服务器详情页面，选择“弹性网卡”页签，并展开 ，将“Source/Destination Check”选项设置为“OFF”。

默认情况下，“源/目的检查”状态为“启用”，系统会检查弹性云服务器发送的报文中源IP地址是否正确，否则不允许弹性云服务器发送该报文。这有助于防止伪装报文攻击，提升安全性。但在该场景中，这种保护机制会导致报文的发送者无法接收到返回的报文。因此，需设置“源/目的检查”状态为禁用。

6. 登录代理弹性云服务器。  
详细操作方法请参见[Linux ECS登录方式概述](#)。
7. 执行以下命令，检测代理弹性云服务器是否可以正常连接Internet。

```
ping www.baidu.com
```

回显包含类似如下信息时，表示代理弹性云服务器可正常连接Internet。

图 8-1 检测是否可以正常连接 Internet

```
[root@ecs-f4f8 ~]# ping www.baidu.com
PING www.a.shifen.com (61.135.169.121) 56(84) bytes of data:
64 bytes from 61.135.169.121 (61.135.169.121): icmp_seq=1 ttl=47 time=2.77 ms
64 bytes from 61.135.169.121 (61.135.169.121): icmp_seq=2 ttl=47 time=2.65 ms
64 bytes from 61.135.169.121 (61.135.169.121): icmp_seq=3 ttl=47 time=2.61 ms
64 bytes from 61.135.169.121 (61.135.169.121): icmp_seq=4 ttl=47 time=2.83 ms
64 bytes from 61.135.169.121 (61.135.169.121): icmp_seq=5 ttl=47 time=2.69 ms
64 bytes from 61.135.169.121 (61.135.169.121): icmp_seq=6 ttl=47 time=2.63 ms
```

8. （可选）执行以下命令，安装iptables服务并配置iptables服务开机自启动。  
仅CentOS 7.x版本的弹性云服务器，需要执行此步骤。

```
yum install iptables-services -y
```

```
systemctl start iptables
```

```
systemctl enable iptables
```

9. 执行以下命令，查看代理弹性云服务器的IP转发功能是否开启。

```
cat /proc/sys/net/ipv4/ip_forward
```

- 回显为“0”表示关闭，请执行10。

- 回显为“1”表示开启，请执行15。

10. 执行以下命令，打开IP转发功能配置文件。

```
vi /etc/sysctl.conf
```

11. 按“i”，进入编辑模式。
12. 修改如下参数的值。  
将参数“net.ipv4.ip\_forward”的值修改为“1”。

#### 📖 说明

如果“sysctl.conf”文件中不存在参数“net.ipv4.ip\_forward”，执行以下命令进行添加：  
**echo net.ipv4.ip\_forward=1 >> /etc/sysctl.conf**

13. 按“Esc”，输入:wq，按“Enter”。  
保存设置并退出vi编辑器。
14. 执行以下命令，使配置文件修改生效。

**sysctl -p /etc/sysctl.conf**

15. 执行以下命令，清除原有iptables规则。

**iptables -F**

16. 执行以下命令，配置SNAT，使代理弹性云服务器所在的网段内其他弹性云服务器可通过代理弹性云服务器访问Internet。

**iptables -t nat -A POSTROUTING -o eth0 -s subnet/netmask-bits -j SNAT --to nat-instance-ip**

假设代理弹性云服务器所在的网段为192.168.125.0，子网掩码为24位，公网IP地址为192.168.125.4，则执行如下命令。

**iptables -t nat -A POSTROUTING -o eth0 -s 192.168.125.0/24 -j SNAT --to 192.168.125.4**

#### 📖 说明

为了确保重启后上述规则不丢失，可以执行vi /etc/rc.local编辑rc.local文件，将16中的规则复制到rc.local文件，按“ESC”退出编辑模式，输入“:wq”保存并退出。

17. 执行以下命令，保存iptables的配置并设置开机自启动。

**service iptables save**

**chkconfig iptables on**

18. 执行以下命令，查看SNAT配置是否成功。

**iptables -t nat --list**

回显类似如图8-2所示时，表示SNAT配置成功。

图 8-2 SNAT 配置成功

```
[root@host- ~]# iptables -t nat --list
Chain PREROUTING (policy ACCEPT)
target     prot opt source                destination

Chain POSTROUTING (policy ACCEPT)
target     prot opt source                destination
SNAT      all  --  192.168.125.0/24      anywhere             to:192.168.125.4
SNAT      all  --  anywhere              anywhere             to:192.168.125.4
```

19. 添加自定义路由。
  - a. 登录管理控制台。
  - b. 选择“网络 > 虚拟私有云”。
  - c. 选择需要添加路由表的虚拟私有云，在“路由表”页面，单击“添加路由信息”。

d. 根据界面提示，填写路由信息。

- 目的地址：是目的网段，默认是0.0.0.0/0。

- 下一跳地址：是代理弹性云服务器的私有IP地址。

您可以在弹性云服务器页面，查看该弹性云服务器的私有IP地址。

20. 如需删除添加的iptables规则，需执行以下命令：

```
iptables -t nat -D POSTROUTING -o eth0 -s subnet/netmask-bits -j SNAT  
--to nat-instance-ip
```

假设代理弹性云服务器所在的网段为192.168.125.0，子网掩码为24位，私网IP地址为192.168.125.4，则执行如下命令。

```
iptables -t nat -D POSTROUTING -o eth0 -s 192.168.125.0/24 -j SNAT --to  
192.168.125.4
```

# 9 安全管理

## 9.1 安全组

### 9.1.1 安全组概述

#### 安全组

安全组是一个逻辑上的分组，为具有相同安全保护需求并相互信任的云服务器提供访问策略。安全组创建后，用户可以在安全组中定义各种访问规则，当云服务器加入该安全组后，即受到这些访问规则的保护。

您也可以根据需要创建自定义的安全组，或使用默认安全组，系统会为每个用户默认创建一个默认安全组，默认安全组的规则是在出方向上的数据报文全部放行，入方向访问受限，安全组内的云服务器无需添加规则即可互相访问。默认安全组您可以直接使用，详情请参见[默认安全组和规则](#)。

#### 说明

安全组需在网络互通的情况下生效。若实例属于不同VPC，但同属于一个安全组，则此安全组不生效，您可以使用对等连接等产品建立VPC连接互通。。

#### 安全组规则

安全组创建后，您可以在安全组中设置出方向、入方向规则，这些规则会对安全组内部的云服务器出入方向网络流量进行访问控制，当云服务器加入该安全组后，即受到这些访问规则的保护。

每个安全组都自带默认安全组规则，详情请参见[默认安全组和规则](#)。您也可以自定义添加安全组规则，请参见[配置安全组规则](#)。

#### 安全组的使用限制

- 默认情况下，一个安全组最多只允许拥有50条安全组规则。
- 默认情况下，一个云服务器或扩展网卡最多只能被添加到5个安全组中，安全组规则取并集生效。

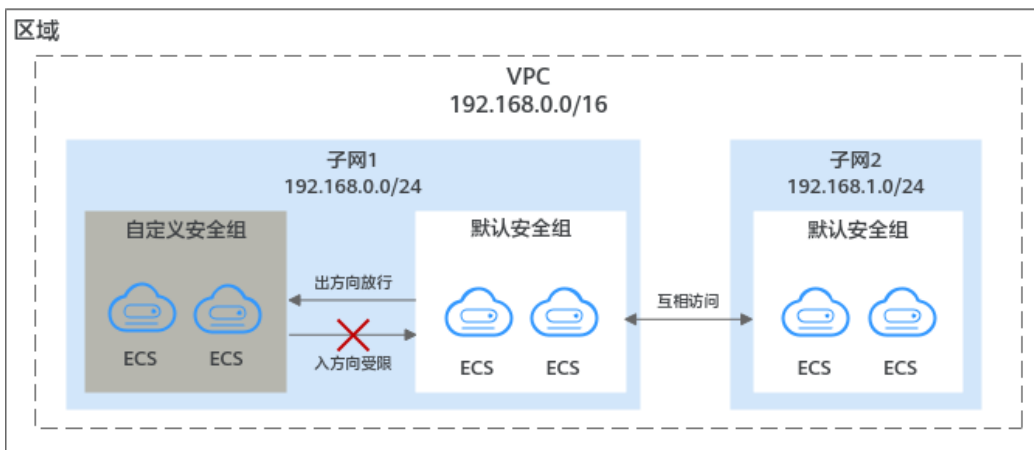
## 9.1.2 默认安全组和规则

默认安全组规则说明如下：

- 入方向规则：入方向流量受限，只允许安全组内实例互通，拒绝来自安全组外部的所有请求进入实例。
- 出方向规则：出方向流量放行，允许所有请求从安全组内实例流出。

如图9-1所示。

图 9-1 默认安全组



### 说明

- 您无法删除默认安全组，可以在默认安全组内修改已有规则或者添加新的规则。
- 默认安全组是为了简化您初次创建实例的流程，系统为您自动创建的。默认安全组拒绝所有外部请求，如果您需要登录实例，请参见[通过本地服务器远程登录云服务器](#)添加安全组规则。

默认安全组规则如表9-1所示：

表 9-1 默认安全组规则

方向	协议	端口范围	目的地址/源地址	说明
出方向	全部	全部	目的地址： 0.0.0.0/0	允许所有出站流量的数据报文通过。
入方向	全部	全部	源地址：当前安全组(例如：sg-xxxxx)	仅允许安全组内的云服务器彼此通信，丢弃其他入站流量的全部数据报文。

## 9.1.3 安全组配置示例

当您在VPC子网内创建实例（云服务器、云容器、云数据库等）时，您可以使用系统提供的默认安全组default，您也可以创建其他安全组。无论是默认安全组，还是您创建的安全组，您均可以在安全组内设置出方向和入方向规则，以此控制出入实例的流量。以下为您介绍一些常用的安全组的配置示例：



- [通过本地服务器远程登录云服务器](#)
- [在本地服务器远程连接云服务器上传或者下载文件（FTP）](#)
- [在云服务器上搭建网站对外提供Web服务](#)
- [使用ping命令验证网络连通性](#)
- [实现不同安全组的实例内网网络互通](#)
- [云服务器提供数据库访问服务](#)
- [限制云服务器访问外部网站](#)

## 使用须知

在配置安全组规则之前，您需要先了解以下信息：

- 不同安全组之间的实例默认网络隔离，无法互相访问。
- 安全组默认拒绝所有来自外部的请求，即本安全组内的实例网络互通，外部无法访问安全组内的实例。  
您需要遵循白名单原则添加安全组入方向规则，允许来自外部的特定请求访问安全组内的实例。
- 安全组的出方向规则一般默认全部放通，即允许安全组内的实例访问外部。  
如果出方向规则被删除，将会导致安全组内实例无法正常访问外部，您可以参考[表9-2](#)重新添加规则。

表 9-2 安全组默认出方向规则

方向	协议端口	目的地址	描述
出方向	全部	0.0.0.0/0	针对全部IPv4协议，允许安全组内的实例可访问外部任意IP和端口。

## 通过本地服务器远程登录云服务器

安全组默认拒绝所有来自外部的请求，如果您需要通过本地服务器远程登录安全组内的云服务器，那么需要根据您的云服务器操作系统类型，在安全组入方向添加对应的规则。

- 通过SSH远程登录Linux云服务器，需要放通SSH(22)端口，请参见[表9-3](#)。
- 通过RDP远程登录Windows云服务器，需要放通RDP(3389)端口，请参见[表9-4](#)。

表 9-3 通过 SSH 远程登录 Linux 云服务器

方向	协议端口	源地址
入方向	自定义TCP: 22	IP地址: 0.0.0.0/0

表 9-4 通过 RDP 远程登录 Windows 云服务器

方向	协议端口	源地址
入方向	自定义TCP: 3389	IP地址: 0.0.0.0/0

**须知**

源地址设置为0.0.0.0/0表示允许所有外部IP地址远程登录云服务器，如果将当前端口暴露到公网，可能存在网络安全风险，建议您将源IP设置为已知的IP地址，配置示例请参见表9-5。

表 9-5 通过已知 IP 地址远程登录云服务器

云服务器类型	方向	协议端口	源地址
Linux云服务器	入方向	自定义TCP: 22	IP地址: 192.168.0.0/24
Windows云服务器	入方向	自定义TCP: 3389	IP地址: 10.10.0.0/24

## 在本地服务器远程连接云服务器上传或者下载文件（FTP）

安全组默认拒绝所有来自外部的请求，如果您需要在本地服务器远程连接云服务器上传或者下载文件，那么您需要开通FTP(20、21)端口。

表 9-6 通过任意服务器远程连接云服务器上传或者下载文件

方向	协议端口	源地址
入方向	自定义TCP: 20-21	IP地址: 0.0.0.0/0

**须知**

- 源地址设置为0.0.0.0/0表示允许所有外部IP地址远程连接云服务器上传或者下载文件，如果将当前端口暴露到公网，可能存在网络安全风险，建议您将源IP设置为已知的IP地址，配置示例请参见表9-7。
- 您需要在弹性云服务器上先安装FTP服务器程序，再查看20、21端口是否正常工作。

表 9-7 通过已知的服务器远程连接云服务器上传或者下载文件

方向	协议端口	源地址
入方向	自定义TCP: 20-21	IP地址: 192.168.0.0/24

## 在云服务器上搭建网站对外提供 Web 服务

安全组默认拒绝所有来自外部的请求，如果您在云服务器上搭建了可供外部访问的网站，则您需要在安全组入方向添加对应的规则，放通对应的端口，例如HTTP(80)、HTTPS(443)。

表 9-8 在云服务器上搭建网站对外提供 Web 服务

方向	协议端口	源地址
入方向	自定义TCP: 80	IP地址: 0.0.0.0/0
入方向	自定义TCP: 443	IP地址: 0.0.0.0/0

## 使用 ping 命令验证网络连通性

ICMP协议用于网络消息的控制和传递，因此在进行一些基本测试操作之前，需要开通ICMP协议访问端口。比如，您需要在某个个人PC上使用ping命令来验证云服务器的网络连通性，则您需要在云服务器所在安全组的入方向添加以下规则，放通ICMP端口。

表 9-9 使用 ping 命令验证网络连通性

方向	协议端口	源地址
入方向	ICMP: 全部	IP地址: 0.0.0.0/0

## 实现不同安全组的实例内网网络互通

同一个VPC内，位于不同安全组内的实例网络不通。如果您需要在同一个VPC内的实例之间共享数据，比如安全组sg-A内的云服务器访问安全组sg-B内的MySQL数据库，您需要通过在安全组sg-B中添加一条入方向规则，允许来自安全组sg-A内云服务器的内网请求进入。

表 9-10 实现不同安全组的实例网络互通

方向	协议端口	源地址
入方向	自定义TCP: 3306	安全组: sg-A

## 云服务器提供数据库访问服务

安全组默认拒绝所有来自外部的请求，如果您在云服务器上部署了数据库服务，允许其他实例通过内网访问数据库服务，则您需要在部署数据库服务器所在的安全组内，添加入方向规则，放通对应的端口，实现其他实例通过内网获取数据库数据的请求。常见的数据库类型及其对应的端口如下：

- MySQL(3306)
- Oracle(1521)
- MS SQL(1433)
- PostgreSQL(5432)
- Redis(6379)

表 9-11 云服务器提供数据库访问服务

方向	协议端口	源地址	描述
入方向	自定义TCP: 3306	安全组: sg-A	允许安全组sg-A内云服务器访问MySQL数据库服务。
入方向	自定义TCP: 1521	安全组: sg-B	允许安全组sg-B内云服务器访问Oracle数据库服务。
入方向	自定义TCP: 1433	IP地址: 172.16.3.21/32	允许私网IP地址为172.16.3.21的云服务器访问MS SQL数据库服务。
入方向	自定义TCP: 5432	IP地址: 192.168.0.0/24	允许私网IP地址属于192.168.0.0/24网段的云服务器访问PostgreSQL数据库服务。

### 须知

本示例中源地址提供的配置仅供参考，请您根据实际需求设置源地址。

## 限制云服务器访问外部网站

安全组的出方向规则一般默认全部放通，默认规则如表9-13所示。如果您需要限制服务器只能访问特定网站，则按照如下要求配置：

1. 首先，您需要遵循白名单规则，在安全组出方向规则中添加指定的端口和IP地址。

表 9-12 限制云服务器访问外部网站

方向	协议端口	目的地址	描述
出方向	自定义 TCP: 80	IP地址: 132.15.XX.XX	允许安全组内云服务器访问指定的外部网站，网站地址为http://132.15.XX.XX:80。

方向	协议端口	目的地址	描述
出方向	自定义 TCP: 443	IP地址: 145.117.XX.XX	允许安全组内云服务器访问指定的外部网站, 网站地址为https://145.117.XX.XX:443。

- 其次, 删除安全组出方向中原有放通全部流量的规则, 如表9-13所示。

表 9-13 安全组默认出方向规则

方向	协议端口	目的地址	描述
出方向	全部	0.0.0.0/0	针对全部IPv4协议, 允许安全组内的实例可访问外部任意IP和端口。

## 9.1.4 配置安全组规则

### 操作场景

安全组类似防火墙功能, 是一个逻辑上的分组, 用于设置网络访问控制。用户可以在安全组中定义各种访问规则, 当弹性云服务器加入该安全组后, 即受到这些访问规则的保护。

- 入方向规则: 放通入方向网络流量, 指从外部访问安全组规则下的云服务器。
- 出方向规则: 放通出方向网络流量。指安全组规则下的云服务器访问安全组外的实例。

默认安全组规则请参见《虚拟有云用户指南》。常用的安全组规则配置示例请参见[安全组配置示例](#)。

### 操作步骤

- 登录管理控制台。
- 选择“计算 > 弹性云服务器”。
- 在弹性云服务器列表, 单击待变更安全组规则的弹性云服务器名称。  
系统跳转至该弹性云服务器详情页面。
- 选择“安全组”页签, 展开安全组, 查看安全组规则。
- 单击安全组ID。  
系统自动跳转至安全组页面。
- 根据界面提示, 设置入方向规则参数。  
单击 $\oplus$ , 可以依次增加多条入方向规则。

表 9-14 入方向规则参数说明

参数	说明	取值样例
协议端口	安全组规则中用来匹配流量的网络协议类型，目前支持TCP、UDP、ICMP和GRE协议。	TCP
	安全组规则中用来匹配流量的目的端口，取值范围为：1~65535。 在入方向规则中，表示外部访问安全组内实例的指定端口。	22或22-30
源地址	源地址可以是IP地址、安全组。用于放通来自IP地址或另一安全组内的实例的访问。 <ul style="list-style-type: none"> <li>IPv4地址：xxx.xxx.xxx.xxx/32</li> <li>子网：xxx.xxx.xxx.0/24</li> <li>任意地址：0.0.0.0/0</li> </ul> 若源地址为安全组，则选定安全组内的云服务器都遵从当前所创建的规则。	192.168.0.0/24
描述	安全组规则的描述信息，非必填项。 描述信息内容不能超过255个字符，且不能包含“<”和“>”。	-

## 7. 根据界面提示，设置出方向规则参数。

单击<sup>+</sup>，可以依次增加多条出方向规则。

表 9-15 出方向参数说明

参数	说明	取值样例
协议/应用	网络协议。目前支持“ALL”、“TCP”、“UDP”和“ICMP”等协议。	TCP
端口和目的地址	端口：允许弹性云服务器访问远端地址的指定端口，取值范围为：1~65535。	22或22-30
	目的地址：可以是IP地址、安全组。允许访问目的IP地址或另一安全组内的实例。例如： <ul style="list-style-type: none"> <li>xxx.xxx.xxx.xxx/32（IPv4地址）</li> <li>xxx.xxx.xxx.0/24（子网）</li> <li>0.0.0.0/0（任意地址）</li> <li>sg-abc（安全组）</li> </ul>	0.0.0.0/0
描述	安全组规则的描述信息，非必填项。 描述信息内容不能超过255个字符，且不能包含“<”和“>”。	-

## 8. 单击“确定”，完成安全组规则配置。

## 9.1.5 更改安全组

### 操作场景

本节操作介绍当弹性云服务器的网卡需要变更所属安全组时的操作步骤。

### 约束限制

- 更改安全组操作会覆盖弹性云服务器原有的安全组设置。
- 使用多个安全组可能会影响弹性云服务器的网络性能，建议您选择安全组的数量不多于5个。

### 操作步骤

1. 登录管理控制台。
  2. 选择“计算 > 弹性云服务器”。
  3. 在弹性云服务器列表中，单击“操作”列下的“更多 > 网络设置 > 更改安全组”。
- 系统弹窗显示“更改安全组”页面。
4. 根据界面提示，在下拉列表中选择待更改安全组的网卡，并重新选择安全组。如需创建新的安全组，请单击“新建安全组”。

#### 说明

使用多个安全组可能会影响弹性云服务器的网络性能，建议您选择安全组的数量不多于5个。

5. 单击“确定”。

# 10 备份管理

---

## 10.1 云备份概述

### 什么是云备份

云备份（Cloud Backup and Recovery, CBR）可以为云服务器、云硬盘提供简单易用的备份服务，当发生病毒入侵、人为误删除、软硬件故障等事件时，可将数据恢复到任意备份点。

云备份保障用户数据的安全性和正确性，确保业务安全。

### 备份、镜像有什么区别？

备份分为云服务器备份和云硬盘备份。

镜像分为系统盘镜像、数据盘镜像、整机镜像。



备份类型	备份对象	适用场景	区别和优势
云服务器备份	弹性云服务器中的所有云硬盘（系统盘和数据盘）	<ul style="list-style-type: none"> <li>● <b>云服务器受到攻击或病毒入侵</b> 通过云服务器备份，可立即恢复到最近一次没有受黑客攻击或病毒入侵的备份时间点。</li> <li>● <b>数据被误删</b> 通过云服务器备份，可立即恢复到删除前的备份时间点，找回被删除的数据。</li> <li>● <b>应用程序更新出错</b> 通过云服务器备份，可立即恢复到应用程序更新前的备份时间点，使系统正常运行。</li> <li>● <b>云服务器宕机</b> 通过云服务器备份，可立即恢复到宕机之前的备份时间点，使云服务器能再次正常启动。</li> </ul>	<p>备份的同一个云服务器下的所有云硬盘数据具有一致性，即同时对所有云硬盘进行备份，不存在因备份创建时间差带来的数据不一致问题。</p> <p>且云备份支持根据备份策略自动备份。</p>
云硬盘备份	指定的单个或多个云硬盘（系统盘或数据盘）	<ul style="list-style-type: none"> <li>● <b>系统盘没有个人数据，因而只需要对部分的数据盘进行备份</b> 当云硬盘出现故障或云硬盘中的数据发生逻辑错误时（如误删数据、遭遇黑客攻击或病毒危害等），可快速恢复数据。</li> <li>● <b>备份作为基线数据</b> 设置备份策略，根据策略自动对云硬盘进行数据备份，通过定期创建的备份作为基线数据，用来创建新的云硬盘或者恢复数据到云硬盘。</li> </ul>	<p>备份数据则存储在对象存储（OBS）中，可以实现在云硬盘存储损坏情况下的数据恢复</p> <p>保证数据安全的同时降低备份成本</p>
系统盘镜像	系统盘	<ul style="list-style-type: none"> <li>● <b>快速恢复系统</b> 更换操作系统、应用软件升级或业务数据迁移等重大操作前，将系统盘创建成系统盘镜像，一旦迁移过程中出现问题，可以通过系统盘镜像切换操作系统，或重新创建新的云服务器。</li> <li>● <b>快速部署多个业务</b> 通过同一个系统盘镜像可以快速创建出多个具有相同操作系统的云服务器，从而快速部署多个业务。</li> </ul>	<p>系统盘镜像可以实现在云服务器操作系统损坏情况下快速切换至损坏前的操作系统。</p>

备份类型	备份对象	适用场景	区别和优势
数据盘镜像	指定的数据盘	<b>快速复制数据</b> 通过同一个数据盘镜像可以快速创建出多个具有相同数据的云硬盘，再将新创建的云硬盘挂载到其他服务器上，从而可以同时为多种业务提供数据资源。	数据盘镜像可以复制全盘的数据并创建新的云硬盘挂载到其他服务器上，实现云服务器数据盘的复制和共享。
整机镜像	弹性云服务器中的所有云硬盘（系统盘和数据盘）	<ul style="list-style-type: none"><li>● <b>快速恢复系统</b> 更换操作系统、应用软件升级或业务数据迁移等重大操作前，将云服务器的系统盘和数据盘创建成整机镜像，一旦迁移过程中出现问题，可以通过整机镜像切换操作系统，或重新创建新的云服务器。</li><li>● <b>快速部署多个业务</b> 通过同一个整机镜像可以快速创建出多个具有相同操作系统和数据的云服务器，从而快速部署多个业务。</li></ul>	通过整机镜像实现业务的整理迁移。

## 云备份产品架构

云备份由备份、存储库和策略组成。

- **备份：**

备份即一个备份对象执行一次备份任务产生的备份数据，包括备份对象恢复所需要的全部数据。

- 云服务器备份：云服务器备份提供对弹性云服务器和裸金属服务器的基于多云硬盘一致性快照技术的数据保护。同时，未部署数据库等应用的服务器产生的备份为服务器备份，部署数据库等应用的服务器产生的备份为数据库服务器备份。
- 云硬盘备份：云硬盘备份提供对云硬盘的基于快照技术的数据保护。

- **存储库**

云备份使用存储库来存放备份。创建备份前，需要先创建至少一个存储库，并将服务器或磁盘绑定至存储库。服务器或磁盘产生的备份则会存放至绑定的存储库中。

存储库分为备份存储库和复制存储库两种。备份存储库用于存放备份对象产生的备份，复制存储库用于存放复制操作产生的备份。

不同类型的备份对象产生的备份需要存放在不同类型的存储库中。

- **策略**

策略分为备份策略和复制策略。

- 备份策略：需要对备份对象执行自动备份操作时，可以设置备份策略。通过在策略中设置备份任务执行的时间、周期以及备份数据的保留规则，将备份存储库绑定到备份策略，可以为存储库执行自动备份。

- 复制策略：需要对备份或存储库执行自动复制操作时，可以设置复制策略。通过在策略中设置复制任务执行的时间、周期以及备份数据的保留规则，将备份存储库绑定到复制策略，可以为存储库执行自动复制。复制产生的备份需要存放在复制存储库中。

## 备份机制

首次备份为全量备份，备份云服务器/磁盘已使用空间。

示例：某磁盘大小为100GB，已使用空间为40GB，则备份的是40GB的已使用空间。

后续备份均为增量备份，备份上次备份后变化的数据，缩短备份时长、节约备份空间。

删除备份时，仅删除不被其他备份依赖的数据块，不影响使用其他备份进行恢复。无论是全量还是增量备份，都可以快速、方便地将数据恢复至备份所在时刻的状态。

云备份会在备份过程中自动创建快照并且为每个磁盘保留最新的快照。如果该磁盘已备份，再次备份后会自动将旧快照删除，保留最新的快照。

云备份通过云服务器/磁盘与对象存储服务的结合，将数据备份到对象存储中，高度保障用户的备份数据安全。

## 备份的方式及适用场景

云备份提供两种配置方式，一次性备份和周期性备份。一次性备份是指用户手动创建的一次性备份任务。周期性备份是指用户通过创建备份策略并绑定存储库的方式创建的周期性备份任务。

表 10-1 备份的方式及适用场景

对比项	一次性备份	周期性备份
备份策略	不需要	需要
备份次数	手动执行一次性备份	根据备份策略进行周期性备份
备份名称	支持自定义，默认为“manualbk_xxxx”	系统自动生成，默认为“autobk_xxxx”
备份方式	默认首次全量备份，后续增量备份	默认首次全量备份，后续增量备份
适用场景	资源进行操作系统补丁安装、升级，应用升级等操作之前，以便安装或者升级失败之后，能够快速恢复到变更之前的状态。	资源的日常备份保护，以便发生不可预见的故障而造成数据丢失时，能够使用邻近的备份进行恢复。

## 10.2 备份弹性云服务器

### 操作场景

云备份提供申请即用的备份服务，使您的数据更加安全可靠。例如，当云服务器或磁盘出现故障或者人为错误导致数据误删时，可以自助快速恢复数据。本节操作介绍备份云服务器和云硬盘的操作步骤。

您可以通过[云备份产品架构](#)、[备份机制（全量备份和增量备份）](#)、[备份的方式及适用场景](#)了解更多产品详情。

目前弹性云服务器备份可以通过“云服务器备份”和“云硬盘备份”功能实现：


- 云服务器备份（推荐）：如果是对弹性云服务器中的所有云硬盘（系统盘和数据盘）进行备份，推荐使用云服务器备份功能，同时对所有云硬盘进行备份，避免因备份创建时间差带来的数据不一致问题。
- 云硬盘备份：如果对指定的单个或多个云硬盘（系统盘或数据盘）进行备份，推荐使用云硬盘备份功能，在保证数据安全的同时降低备份成本。

### 云服务器备份操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器的“操作”列下，单击“更多 > 镜像/磁盘/备份 > 创建备份”。
  - 如果云服务器已绑定存储库，则根据界面提示，配置备份信息。
    - 云服务器列表：默认已勾选待备份的云服务器
    - 名称：用户自定义云备份名称。
    - 描述：输入待创建的备份的描述。
    - 执行全量备份：勾选后，系统会为绑定的云服务器执行全量备份，备份所占存储容量也会相应增加。
  - 如果云服务器未绑定存储库，请先购买存储库，配置备份信息。  
详细操作请参考《云备份用户指南》。
4. 单击“确定”。系统会自动为云服务器创建备份。  
在“备份副本”页签，产生的备份的“备份状态”为“可用”时，表示备份任务执行成功。  
云服务器备份在执行备份过程中，若备份进度已经超过10%可重启云服务器。但为了保证数据完整性，建议备份完成后再执行重启。

### 云硬盘备份操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器的“操作”列下，单击“更多 > 镜像/磁盘 > 创建备份”。
  - 如果云服务器已绑定存储库，则根据界面提示，配置备份信息。

- 服务器列表：默认已勾选待备份的云服务器，单击  查看云服务器挂载的磁盘。并选择待备份的磁盘。
  - 名称：用户自定义云备份名称。
  - 描述：输入待创建的备份的描述。
  - 执行全量备份：勾选后，系统会为绑定的磁盘执行全量备份，备份所占存储容量也会相应增加。
- 如果云服务器未绑定存储库，请先购买存储库，再配置备份信息。  
详细操作请参考《云备份用户指南》。
4. 单击“确定”。系统会自动为磁盘创建备份。  
根据界面提示，跳转至云备份控制台的“备份副本”页签，当产生的备份的“备份状态”为“可用”时，表示备份任务执行成功。  
如果在备份过程中对磁盘的数据进行删除等操作，被删除的文件可能不会被备份成功。为了保证数据完整性，建议备份完成后再对数据进行操作。

# 11 密码和密钥对管理

## 11.1 重置 ECS 密码

### 11.1.1 密码使用场景介绍

弹性云服务器系统密码涉及到客户重要的私人信息，提醒您妥善保管密码。如果您忘记密码或密码过期，可以重置密码。

弹性云服务器的重置密码场景如表11-1所示。

表 11-1 重置密码场景

操作场景	前提条件
<a href="#">在控制台重置ECS密码</a>	已安装重置密码插件。 <b>说明</b> <ul style="list-style-type: none"><li>使用公共镜像的弹性云服务器默认已安装一键式重置密码插件。</li><li>Windows/Linux操作系统的云服务器</li><li>弹性云服务器的一键式重置密码功能，当前仅上线中国区，非中国区的用户暂不支持使用该功能。</li></ul>
<a href="#">在操作系统内部修改ECS密码</a>	无。 <b>说明</b> 适用于Windows/Linux操作系统的云服务器。
<a href="#">重置Windows云服务器密码</a>	未安装重置密码插件。
<a href="#">重置Linux云服务器密码</a>	未安装重置密码插件。

### 背景知识

弹性云服务器的密码规则如表11-2所示。

表 11-2 密码设置规则

参数	规则
密码	<ul style="list-style-type: none"><li>● 密码长度范围为8到26位。具体以界面显示为准。</li><li>● 密码至少包含以下4种字符中的3种：<ul style="list-style-type: none"><li>- 大写字母</li><li>- 小写字母</li><li>- 数字</li><li>- Windows操作系统云服务器特殊字符：包括“\$”、“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“,”和“?”</li><li>- Linux操作系统云服务器特殊字符：包括“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“^”、“,”、“{”、“}”和“?”</li></ul></li><li>● 密码不能包含用户名或用户名的逆序。</li><li>● Windows操作系统的云服务器，不能包含用户名中超过两个连续字符的部分。</li></ul>

## 11.1.2 在控制台重置 ECS 密码

### 操作场景

如果在创建弹性云服务器时未设置密码，或密码丢失、过期，可以参考本节操作重置密码。

### 前提条件

- 密码丢失或过期前，已安装密码重置插件。
  - 公共镜像创建的弹性云服务器默认已安装一键式重置密码插件。
  - 私有镜像创建的弹性云服务器且未安装密码重置插件，可参考[重置Windows云服务器密码](#)和[重置Linux云服务器密码](#)。
- 弹性云服务器的一键式重置密码功能，当前仅上线中国区，非中国区的用户暂不支持使用该功能。
- 请勿删除重置密码进程CloudResetPwdAgent和CloudResetPwdUpdateAgent，否则，会导致一键式重置密码功能不可用。
- 使用SUSE 11 SP4镜像创建的弹性云服务器，内存需要大于等于4GiB时才能支持一键式重置密码功能。
- 弹性云服务器使用的VPC网络DHCP不能禁用。
- 弹性云服务器网络正常通行。

### 操作步骤

您可参考以下步骤在控制台上修改一台或多台弹性云服务器的登录密码。

### 说明

开机状态下重置密码后需重启云服务器，新密码才会生效。您可以重置密码后手动重启，或在重置密码时候勾选“自动重启”。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 选中待重置密码的弹性云服务器，并选择“操作”列下的“更多 > 重置密码”。

### 说明

对于已安装一键式重置密码插件的弹性云服务器，系统支持批量重置密码功能。批量重置的操作如下：

1. 勾选待重置密码的多台弹性云服务器。
2. 单击云服务器列表页上方的“更多 > 重置密码”。
3. 根据界面提示，设置新密码。  
设置成功后，批量重置密码的多台弹性云服务器登录密码相同。
4. 根据界面提示，设置弹性云服务器的新密码，并确认新密码。  
新密码的复杂度应满足表11-3。

表 11-3 密码设置规则

参数	规则
密码	<ul style="list-style-type: none"><li>● 密码长度范围为8到26位。具体以界面显示为准。</li><li>● 密码至少包含以下4种字符中的3种：<ul style="list-style-type: none"><li>- 大写字母</li><li>- 小写字母</li><li>- 数字</li><li>- Windows操作系统云服务器特殊字符：包括“\$”、“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“,”和“?”</li><li>- Linux操作系统云服务器特殊字符：包括“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“^”、“;”、“{”、“}”和“?”</li></ul></li><li>● 密码不能包含用户名或用户名的逆序。</li><li>● Windows操作系统的云服务器，不能包含用户名中超过两个连续字符的部分。</li></ul>

5. 单击“确认”。  
系统执行重置密码操作，该操作预计需要10分钟，请勿频繁执行。
  - 如果在开机状态下重置密码，需手动重启使新密码生效。
  - 如果在关机状态下重置密码，待重新开机后新密码生效。



## 相关链接

- [一键重置密码后无法使用新密码登录弹性云服务器](#)
- [业务端口被一键式重置密码插件占用](#)

## 11.1.3 在操作系统内部修改 ECS 密码

### 操作场景

当云服务器密码即将过期、密码泄露或首次登录时（首次登录云服务器建议您修改初始密码），您可以参考本节操作在操作系统内部修改弹性云服务器密码。

优先推荐您参考[在控制台重置ECS密码](#)，在控制台重置实例的登录密码。

### 前提条件

可以登录弹性云服务器。

### 背景知识

弹性云服务器的密码规则如[表11-4](#)所示。

表 11-4 密码设置规则

参数	规则
密码	<ul style="list-style-type: none"><li>• 密码长度范围为8到26位。具体以界面显示为准。</li><li>• 密码至少包含以下4种字符中的3种：<ul style="list-style-type: none"><li>- 大写字母</li><li>- 小写字母</li><li>- 数字</li><li>- Windows操作系统云服务器特殊字符：包括“\$”、“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“,”和“?”</li><li>- Linux操作系统云服务器特殊字符：包括“!”、“@”、“%”、“-”、“_”、“=”、“+”、“[”、“]”、“.”、“.”、“/”、“^”、“,”、“{”、“}”和“?”</li></ul></li><li>• 密码不能包含用户名或用户名的逆序。</li><li>• Windows操作系统的云服务器，不能包含用户名中超过两个连续字符的部分。</li></ul>

## Windows 弹性云服务器

1. 远程登录Windows弹性云服务器。  
登录方法请参见[Windows ECS登录方式概述](#)。
2. 使用快捷键“Win+R”打开“运行”页面。

3. 输入命令行“cmd”打开命令行窗口。
4. 执行以下命令，修改密码，密码复杂度需满足表11-4。  
**net user Administrator 新密码**

## 11.1.4 重置 Windows 云服务器密码

### 操作场景

如果Windows操作系统弹性云服务器未安装密码重置插件或者通过控制台重置密码后不生效，可以参见本节内容重新设置密码。

本节操作介绍的方法仅适用于修改Windows本地账户密码，不能修改域账户密码。

Linux操作系统请参见[重置Linux云服务器密码](#)。

#### 📖 说明

- 如果弹性云服务器提前安装了密码重置插件，请参见[在控制台重置ECS密码](#)操作获取新密码。
- 公共镜像创建的弹性云服务器默认已安装一键式重置密码插件。请参考[在控制台重置ECS密码](#)重置密码。
- 本节操作的方法需要卸载系统盘，为了避免造成系统盘数据丢失，建议您在操作前先备份系统盘。

### 前提条件

- 准备一台Linux操作系统的临时弹性云服务器，建议操作系统为Ubuntu14.04以上版本，该临时弹性云服务器与待重置密码的弹性云服务器位于同一个可用区，并具有相同的CPU架构。
- 临时弹性云服务器已经绑定弹性公网IP，并配置系统apt-get源。
- 通过下面的方法，在临时弹性云服务器中安装ntfs-3g和chntpw软件包。

方法一：

执行以下命令，安装ntfs-3g和chntpw软件包。

```
sudo apt-get install ntfs-3g chntpw
```

方法二：

根据临时弹性服务器的操作系统版本，下载对应版本的ntfs-3g和chntpw软件包进行安装。

### 操作步骤

1. 关闭原弹性云服务器，卸载系统盘。
  - a. 登录管理控制台。
  - b. 选择“计算 > 弹性云服务器”。
  - c. 原Windows弹性云服务器关机，并进入其详情页，选择“云硬盘”页签。

#### 📖 说明

原Windows弹性云服务器关机时，请勿执行强制关机操作，否则可能引起重置密码操作失败。

- d. 单击系统盘所在行的“卸载”，卸载系统盘。

2. 将系统盘挂载到临时弹性云服务器上。
  - a. 展开临时弹性云服务器的详情页，并选择“云硬盘”页签。
  - b. 单击“挂载磁盘”，在“挂载磁盘”对话框中，选择**1.d**中卸载的系统盘，并将其挂载到临时弹性云服务器上。
  - c. 远程登录临时弹性云服务器。
  - d. 执行以下命令，查看卸载的系统盘在临时弹性云服务器上的目录。

```
fdisk -l
```

- e. 执行以下命令，将卸载的系统盘的文件系统挂载到临时弹性云服务器上。

```
mount -t ntfs-3g /dev/2.d的查询结果 /mnt/
```

例如，**2.d**的查询结果为“xvde2”：

```
mount -t ntfs-3g /dev/xvde2 /mnt/
```

如果执行上述挂载命令时出现如下错误提示信息，可能是由于ntfs文件系统不一致，需要修复文件系统一致性错误。

```
The disk contains an unclean file system (0, 0).
Metadata kept in Windows cache, refused to mount.
Failed to mount '/dev/xvde2': Operation not permitted
The NTFS partition is in an unsafe state. Please resume and shutdown
Windows fully (no hibernation or fast restarting), or mount the volume
read-only with the 'ro' mount option.
```

您可以执行以下命令修复ntfs文件系统，然后再执行挂载命令。在修复前请先备份磁盘数据。

```
ntfsfix /dev/2.d的查询结果
```

例如，**2.d**的查询结果为“xvde2”：

```
ntfsfix /dev/xvde2
```

3. 修改指定用户密码，并清除原始密码。
  - a. 执行以下命令，备份SAM文件。

```
cp /mnt/Windows/System32/config/SAM /mnt/Windows/System32/config/SAM.bak
```
  - b. 执行以下命令，修改指定用户密码。

```
chntpw -u Administrator /mnt/Windows/System32/config/SAM
```
  - c. 按照系统提示，依次输入“1”、“q”和“y”，按“Enter”。  
系统包含如下回显信息时，表示密码清除成功。

```
Select: [q] > 1
Password cleared!
Select: [q] > q
Hives that have changed:
#Name
0<SAM>
Write hive files? (y/n) [n] : y
0<SAM> - OK
```

4. 关闭临时弹性云服务器，卸载原弹性云服务器的系统盘，并将其挂载回原弹性云服务器。
  - a. 临时弹性云服务器关机，并进入详情页，选择“云硬盘”页签。
  - b. 单击“卸载”，卸载**2.b**中临时挂载的数据盘。
  - c. 展开原Windows弹性云服务器的详情页，选择“云硬盘”页签。
  - d. 单击“挂载磁盘”，在“挂载磁盘”对话框中，选择**4.b**中卸载的数据盘，将其作为系统盘挂载回原弹性云服务器。

5. 开启原Windows弹性云服务器，设置新密码。
  - a. 单击“开机”，开启原Windows弹性云服务器，待状态为“运行中”后，单击“操作”列下的“远程登录”。
  - b. 单击“开始”菜单，在搜索框中输入“CMD”，按“Enter”。
  - c. 执行以下命令，修改用户密码，新密码必须符合密码设置规则，详细内容，请参见[密码使用场景介绍](#)。  
`net user Administrator 新密码`

## 11.1.5 重置 Linux 云服务器密码

### 操作场景

如果Linux操作系统弹性云服务器未安装密码重置插件或者通过控制台重置密码后不生效，可以参见本节内容重新设置密码。

Windows操作系统请参见[重置Windows云服务器密码](#)。

#### 说明

- 如果弹性云服务器提前安装了密码重置插件，请参见[在控制台重置ECS密码](#)操作获取新密码。
- 公共镜像创建的弹性云服务器默认已安装一键式重置密码插件。请参考[在控制台重置ECS密码](#)重置密码。
- 本节操作的方法需要卸载系统盘，为了避免造成系统盘数据丢失，建议您在操作前先备份系统盘。

### 前提条件

- 已经准备一台Linux操作系统的临时弹性云服务器，该临时弹性云服务器与待重置密码的弹性云服务器位于同一个可用区，并具有相同的CPU架构。
- 临时弹性云服务器已经绑定弹性公网IP。

### 操作步骤

1. 下载重置密码的脚本，并上传至临时弹性云服务器。

请联系管理员获取重置密码脚本，在本地通过WinSCP等连接工具，将获取的changepasswd.sh脚本上传到临时云服务器。  
WinSCP下载地址：<https://winscp.net/>。
2. 关闭原弹性云服务器，卸载其系统盘，并将系统盘挂载到临时弹性云服务器。
  - a. 登录管理控制台。
  - b. 选择“计算 > 弹性云服务器”。
  - c. 原弹性云服务器关机，单击云服务器名称进入详情页，并选择“云硬盘”页签。

#### 说明

原弹性云服务器关机时，请勿执行强制关机操作，否则可能引起重置密码操作失败。

- d. 单击系统盘所在行的“卸载”，卸载该系统盘。
3. 将系统盘挂载到临时弹性云服务器上。

- a. 展开临时弹性服务器的详情页，并选择“云硬盘”页签。
  - b. 单击“挂载磁盘”，在“挂载磁盘”对话框中，选择2.d中卸载的系统盘，将其挂载到临时弹性云服务器上。
4. 远程登录临时弹性云服务器，并重置密码。
    - a. 在临时弹性服务器的“操作”列下，单击“远程登录”。
    - b. 执行以下命令，查看原弹性云服务器上卸载的系统盘在临时弹性云服务器上的目录。

```
fdisk -l
```

图 11-1 查询系统盘所在目录

```
root@ecs-...:~# fdisk -l
Disk /dev/vda: 40 GiB, 42949672960 bytes, 83886080 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x43591807

Device            Boot Start      End  Sectors  Size Id Type
/dev/vda1         *          2048 83884031 83881984  40G 83 Linux

Disk /dev/vdb: 40 GiB, 42949672960 bytes, 83886080 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x5e9a7bb5

Device            Boot Start      End  Sectors  Size Id Type
/dev/vdb1         *          2048 83886079 83884032  40G 83 Linux
```

- c. 在changepasswd.sh脚本所在目录执行以下命令，运行重置密码脚本。

```
chmod +x changepasswd.sh
```

```
./changepasswd.sh
```

运行重置密码脚本时，如果系统提示没有lv相关的命令（如no lvs command），则需要在临时弹性云服务器安装lvm(Logical Volume Manager)工具，推荐使用lvm2，您可以执行命令**yum install lvm2**进行安装。

#### 说明

如果原弹性云服务器和临时弹性服务器的操作系统均为centos7，运行脚本过程中可能会出现挂载失败。请将脚本中的内容“mount \$dev \$mountPath”修改为“mount -o nouuid \$dev \$mountPath”，避免脚本运行失败。

- d. 根据系统提示输入新设置的密码和4.b中获取到的目录信息。

系统显示如下回显信息时，说明密码设置成功。

```
set password success.
```

5. （可选）如果是非root用户，请执行如下操作打开root登录权限。

```
vi /etc/ssh/sshd_config
```

修改如下配置项：

- 把PasswordAuthentication no 改为 PasswordAuthentication yes  
或去掉PasswordAuthentication yes 前面的#注释掉。
  - 把PermitRootLogin no 改为 PermitRootLogin yes  
或去掉PermitRootLogin yes 前面的#注释掉。
  - 修改AllowUsers的值为root  
在文件中查找AllowUsers，如果没有请在文件末尾补充AllowUsers root。
6. 关闭临时弹性云服务器，卸载原弹性云服务器的系统盘，并将其重新挂载回原弹性云服务器后执行重启操作。
    - a. 临时弹性云服务器关机，并进入详情页，选择“云硬盘”页签。
    - b. 单击“卸载”，卸载3中临时挂载的数据盘。
    - c. 展开原Linux弹性云服务器的详情页，选择“云硬盘”页签。
    - d. 单击“挂载磁盘”，在“挂载磁盘”对话框中，选择6.b中卸载的数据盘。
  7. 重启原弹性云服务器。

## 11.2 管理 ECS 一键式重置密码插件

### 11.2.1 安装单台 ECS 的一键式重置密码插件（单台操作）

弹性云服务器提供一键式重置密码功能。当弹性云服务器的密码丢失或过期时，如果提前安装了一键式重置密码插件，则可以应用一键式重置密码功能，给弹性云服务器设置新密码。

该方法方便、有效，建议您成功创建、登录弹性云服务器后，安装密码重置插件，具体操作请参见本节内容。

#### 说明

弹性云服务器的一键式重置密码功能，当前仅上线中国区，非中国区的用户暂不支持使用该功能。

使用公共镜像创建的弹性云服务器默认已安装一键式重置密码插件。检查是否已安装一键式重置密码插件，请参见[步骤1](#)。

#### 安装须知

1. 用户自行决定是否安装一键式重置密码插件，使弹性云服务器具备一键式重置密码功能。
2. 安装完成后，请勿自行卸载插件，否则可能导致管理控制台判断失误，从而无法完成密码重置。
3. 重装/切换弹性云服务器操作系统后，一键式重置密码功能失效。如需继续使用该功能，请重新安装一键式重置密码插件。
4. 更换弹性云服务器系统盘后，一键式重置密码功能失效。如需继续使用该功能，请重新安装一键式重置密码插件。
5. CoreOS和KylinOS操作系统的弹性云服务器，不支持安装一键式重置密码插件。
6. 如果没有提前安装一键式重置密码插件，密码丢失或过期后，不能使用一键式重置密码功能。
7. 弹性云服务器需绑定弹性公网IP，才能安装一键式重置密码插件。



## 前提条件

- 对于Windows弹性云服务器，需保证C盘可写入，且剩余空间大于300MB。
- 对于Linux弹性云服务器，需保证根目录可写入，且剩余空间大于300MB。
- 使用SUSE 11 SP4镜像创建的弹性云服务器，内存需要大于等于4GiB时才能支持一键式重置密码功能。
- 弹性云服务器使用的VPC网络DHCP不能禁用。
- 弹性云服务器网络正常通行。
- 设置弹性云服务器的网卡属性为DHCP，使弹性云服务器可以动态获取IP地址。
- 弹性云服务器安全组出方向规则满足如下要求：

- 协议：TCP
- 端口范围：80
- 远端地址：169.254.0.0/16

如果您使用的是默认安全组出方向规则，则已经包括了如上要求，可以正常初始化。默认安全组出方向规则为：

- 协议：ANY
- 端口范围：ANY
- 远端地址：0.0.0.0/16

## Linux 弹性云服务器插件安装方法

**步骤1** 检查弹性云服务器是否已安装一键式重置密码插件。

1. 以root用户登录弹性云服务器。
2. 执行以下命令，查询是否已安装CloudResetPwdAgent。

```
ls -lh /Cloud*
```

图 11-2 查询是否已安装一键式重置密码插件

```
[root@ecs-test ~]# ls -lh /Cloud*
total 20K
drwx----- 2 root root 4.0K Jun 13 14:13 bin
drwxr-xr-x 2 root root 4.0K Jun 13 11:53 conf
drwx----- 3 root root 4.0K Jun 13 11:53 depend
drwx----- 2 root root 4.0K Jun 13 11:53 lib
drwx----- 2 root root 4.0K Jun 13 14:13 logs
[root@ecs-test ~]#
[root@ecs-test ~]#
```

检查结果是否如图11-2所示。

- 是，表示已安装一键式重置密码插件，结束。
- 否，表示未安装一键式重置密码插件，请继续执行如下操作进行安装。

**步骤2** 下载一键式重置密码插件CloudResetPwdAgent.zip。

该插件放置在OBS桶中，请联系管理员获取OBS桶的路径。

假设获取的下载地址为：[http://xxx-cloud-reset-pwd.obs.com/linux/64/reset\\_pwd\\_agent/CloudResetPwdAgent.zip](http://xxx-cloud-reset-pwd.obs.com/linux/64/reset_pwd_agent/CloudResetPwdAgent.zip)

则下载插件的命令如下：

```
wget http://xxx-cloud-reset-pwd.obs.com/linux/64/reset_pwd_agent/CloudResetPwdAgent.zip
```

**步骤3** 执行以下命令，解压软件包CloudResetPwdAgent.zip。

安装一键式重置密码插件对插件的解压目录无特殊要求，请您自定义。

```
unzip -o -d 插件解压目录 CloudResetPwdAgent.zip
```

示例：

假设插件解压的目录为/home/linux/test，则命令行如下：

```
unzip -o -d /home/linux/test CloudResetPwdAgent.zip
```

**步骤4** 安装一键式重置密码插件。

1. 执行以下命令，进入文件CloudResetPwdAgent.Linux。

```
cd {插件解压目录}/CloudResetPwdAgent/CloudResetPwdAgent.Linux
```

示例：

假设插件解压的目录为/home/linux/test，则命令行如下：

```
cd /home/linux/test/CloudResetPwdAgent/CloudResetPwdAgent.Linux
```

2. 执行以下命令，添加文件setup.sh的运行权限。

```
chmod +x setup.sh
```

3. 执行以下命令，安装插件。

```
sudo sh setup.sh
```

如果脚本执行成功打印“cloudResetPwdAgent install successfully.”，且未打印“Failed to start service cloudResetPwdAgent”，表示安装成功。

#### 说明

- 您也可以根据**步骤1**，检查密码重置插件是否安装成功。
- 如果密码重置插件安装失败，请检查安装环境是否符合要求，并重试安装操作。

**步骤5** 修改重置密码插件的文件权限。

```
chmod 700 /CloudrResetPwdAgent/bin/cloudResetPwdAgent.script
```

```
chmod 700 /CloudrResetPwdAgent/bin/wrapper
```

```
chmod 600 /CloudrResetPwdAgent/lib/*
```

----结束

## Windows 弹性云服务器插件安装方法

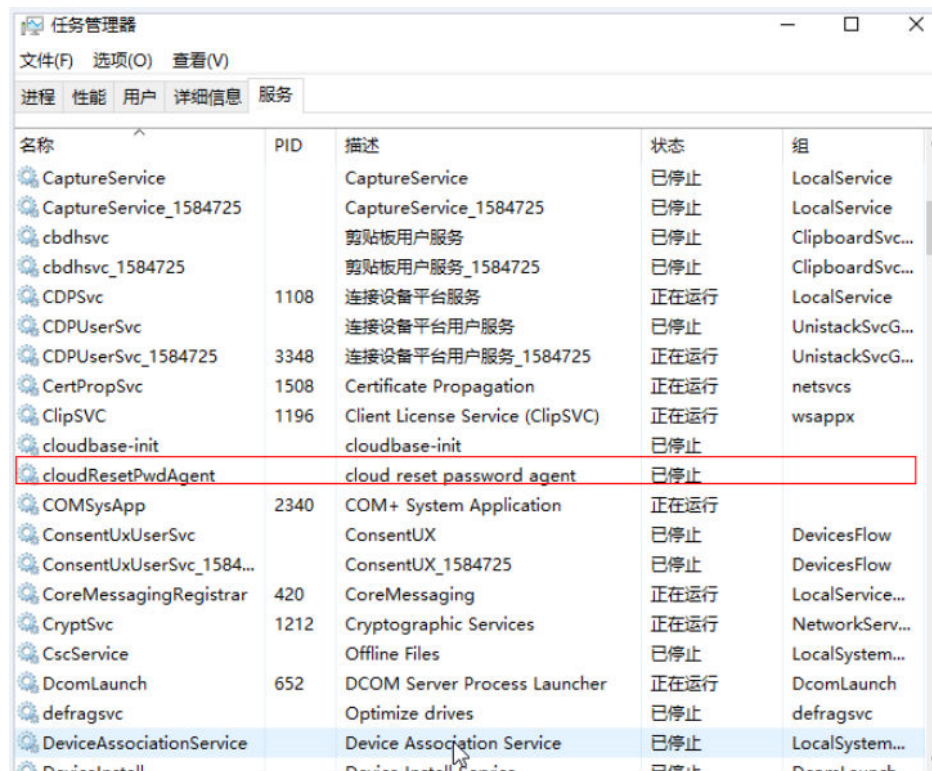
**步骤1** 登录弹性云服务器。

**步骤2** 检查弹性云服务器是否已安装密码重置插件CloudResetPwdAgent。检查方法如下：

查看任务管理器，如果找到cloudResetPwdAgent服务，如图11-3所示，表示弹性云服务器已安装密码重置插件。



图 11-3 安装插件成功



- 是，结束。
- 否，执行步骤3。

**步骤3** 下载一键式重置密码插件CloudResetPwdAgent.zip。

安装一键式重置密码插件对插件的具体放置目录无特殊要求，请您自定义。  
该插件放置在OBS桶中，请联系管理员获取OBS桶的路径。

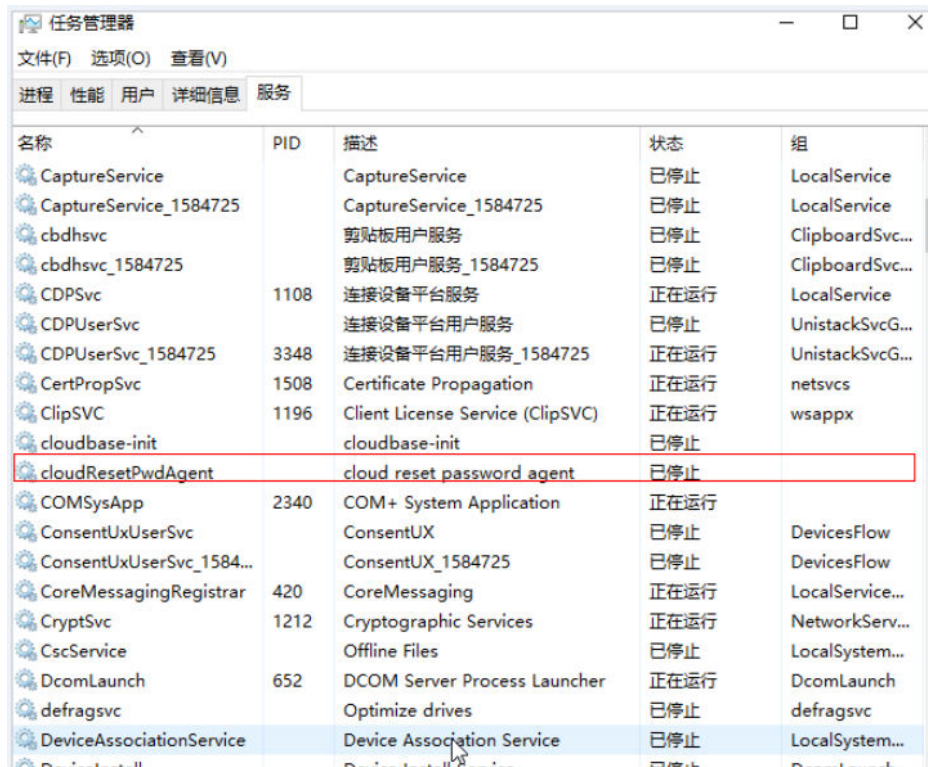
**步骤4** 解压软件包CloudResetPwdAgent.zip。

安装一键式重置密码插件对插件的解压目录无特殊要求，请您自定义。

**步骤5** 安装一键式重置密码插件。

1. 双击“CloudResetPwdAgent.Windows”文件夹下的“setup.bat”文件。  
安装密码重置插件。
2. 查看任务管理器，检查密码重置插件是否安装成功。  
如果在任务管理器中查找到了cloudResetPwdAgent服务，如图11-4所示，表示安装成功，否则安装失败。

图 11-4 安装插件成功



### 说明

如果密码重置插件安装失败，请检查安装环境是否符合要求，并重试安装操作。

---结束

## 后续处理

- 安装一键式重置密码插件后，如果无法通过弹性云服务器开机自动启动该插件，可以将其添加至开机启动项。具体方法请参见《镜像服务用户指南》的“常见问题 > 操作系统类 > 一键式密码重置插件启动失败时如何操作”章节。
- 一键式重置密码插件安装成功后，请勿删除重置密码进程CloudResetPwdAgent，否则，会导致一键式重置密码功能不可用。
- 云平台对一键式重置密码插件进行了升级，对于新创建的弹性云服务器，默认采用PIPE模式，不会占用端口。对于已创建云服务器，仍采用AUTO模式，随机占用31000~32999中的一个端口。占用端口的原则是：在该范围内，系统会按照自小到大的顺序，占用当前空闲的端口。如果占用的端口和业务端口冲突，请参见[业务端口被一键式重置密码插件占用](#)。

## 关于插件卸载

如果不再继续使用一键重置密码功能，您可以根据如下指导进行操作：

- Linux弹性云服务器
  - a. 登录弹性云服务器。
  - b. 执行以下命令，进入bin目录，删除服务cloudResetPwdAgent。

```
cd /CloudrResetPwdAgent/bin
```



- 登录Linux操作系统的弹性云服务器  
若用户购买的是Linux操作系统的弹性云服务器，可以直接使用密钥对远程登录云服务器。
  - 创建弹性云服务器时，选择“密钥对方式”登录，详细操作，请参见[步骤三：高级配置](#)的“设置‘登录凭证’”。
  - 创建弹性云服务器完成后，为云服务器绑定密钥对，详细内容请参考《数据加密服务用户指南》的“绑定密钥对”章节。
- 获取Windows操作系统弹性云服务器的登录密码  
若用户购买的是Windows操作系统的弹性云服务器，可以通过密钥对的私钥获取登录密码，该密码为随机密码，安全性高。  
详细内容，请参见[获取Windows弹性云服务器的密码](#)

## 创建密钥对操作指引

您可以使用已有密钥对或新建一个密钥对，用于远程登录身份验证。

- 新建密钥对  
如果没有可用的密钥对，需新建一个密钥对，生成公钥和私钥，并在登录弹性云服务器时提供私钥进行鉴权。创建密钥对的方法如下：
  - **（推荐）通过管理控制台创建密钥对**：公钥自动保存在系统中，私钥由用户保存在本地。
  - **通过PuTTY Key Generator工具创建密钥对**：公钥和私钥均保存在用户本地。  
创建成功的密钥对，还需要执行[导入密钥对](#)，导入系统才能正常使用密钥对。
- 使用已有密钥对  
如果本地已有密钥对（例如，使用PuTTYgen工具生成的密钥对），可以在管理控制台导入密钥对公钥，由系统维护您的公钥文件。具体操作请参见[导入密钥对](#)。

### 说明

如果已有密钥对的公钥文件是通过puttygen.exe工具的“Save public key”按钮保存的，该公钥文件不能直接导入管理控制台。

如需继续使用该密钥对作远程身份验证，请参见[通过puttygen.exe工具创建的密钥对，导入管理控制台失败怎么办？](#)

## 约束与限制

- 仅支持远程登录Linux云服务器。
- 通过管理控制台创建的SSH-2密钥对仅支持“RSA-2048”加解密算法。
- 密钥对仅适用于本区域的云服务器。
- 通过外部导入的密钥对支持的加解密算法为：
  - RSA-1024
  - RSA-2048
  - RSA-4096
- 私钥是保证您的弹性云服务器安全的重要手段之一，用于远程登录身份认证，为保证弹性云服务器安全，只能下载一次，请妥善保管。

## 11.3.2（推荐）通过管理控制台创建密钥对

### 操作场景

您可以通过管理控制台创建密钥对，创建完成后，公钥自动保存在系统中，私钥由用户保存在本地。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在左侧导航树中，选择“密钥对”。
4. 在“密钥对”页面，单击“创建密钥对”。
5. 输入密钥对名称。  
密钥对名称由两部分组成：KeyPair-4位随机数字。
6. 单击“确定”
7. 您的浏览器会提示您下载或自动下载私钥文件。文件名是您为密钥对指定的名称，文件扩展名为“.pem”。请将私钥文件保存在安全位置。然后在系统弹出的提示框中单击“确定”。

#### 说明

这是您保存私钥文件的唯一机会，请妥善保管。当您创建弹性云服务器时，您将需要提供密钥对的名称；每次SSH登录到弹性云服务器时，您将需要提供相应的私钥。

## 11.3.3 通过 PuTTY Key Generator 工具创建密钥对

### 操作场景

您可以通过puttygen.exe工具创建密钥对，创建完成后，公钥和私钥均保存在用户本地。

#### 说明

通过puttygen.exe工具创建成功的密钥对，还需要执行**导入密钥对**，导入系统才能正常使用密钥对。

### 操作步骤

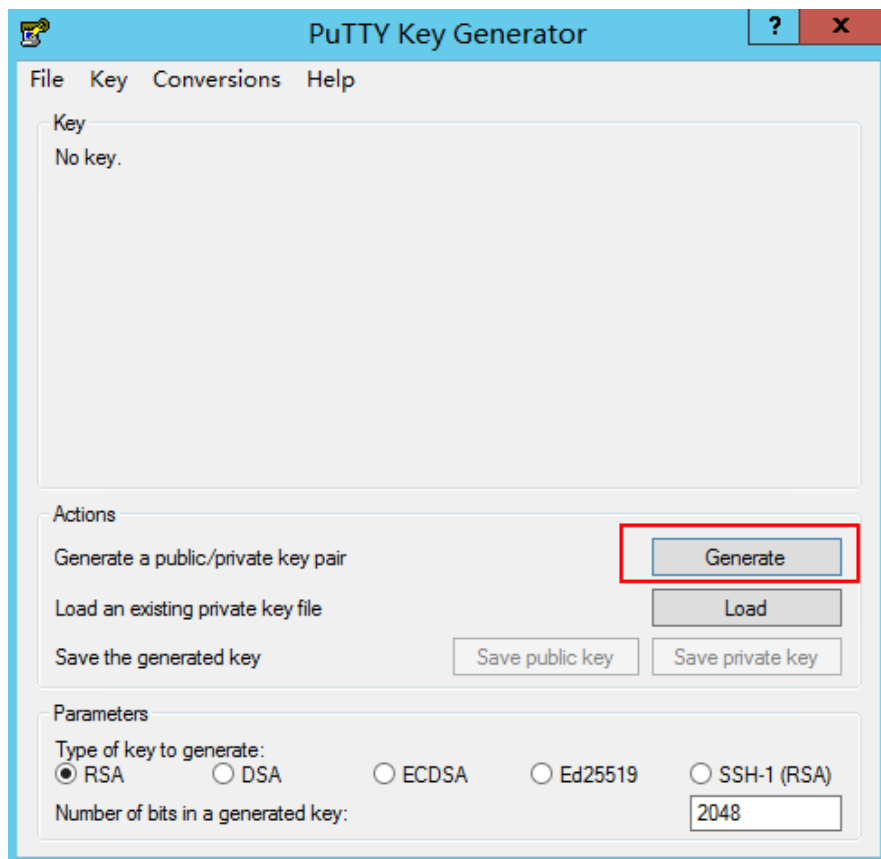
1. 在以下路径中下载并安装PuTTY和PuTTYgen。  
<https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html>

#### 说明

PuTTYgen是密钥生成器，用于创建密钥对，生成一对公钥和私钥供PuTTY使用。

2. 生成公钥和私钥文件。
  - a. 双击“puttygen.exe”，打开“PuTTY Key Generator”。

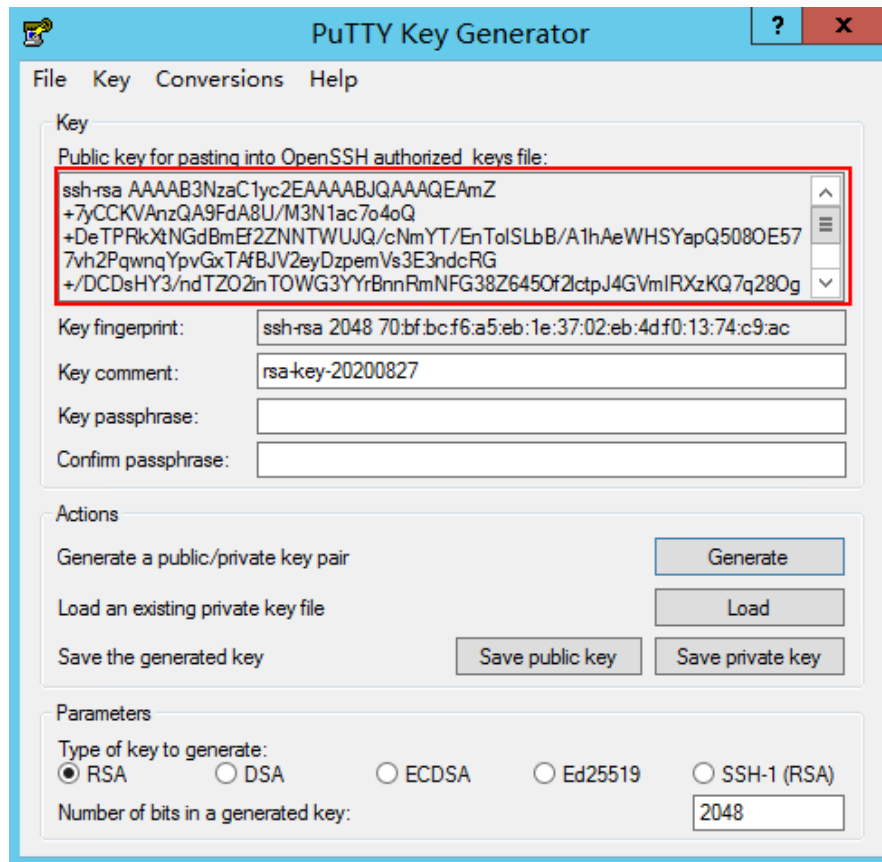
图 11-5 PuTTY Key Generator



- b. 单击“Generate”。

密钥生成器将自动生成一对公钥和私钥，其中，图11-6的红框内容为生成的公钥文件。

图 11-6 生成公钥和私钥文件



3. 复制红框中的公钥内容，并将其粘贴在文本文档中，以.txt格式保存在本地，保存公钥文件。

#### 说明

请勿直接单击“Save public key”保存公钥文件。因为通过puttygen.exe工具的“Save public key”按钮保存在本地的公钥文件，公钥内容的格式会发生变化，不能直接导入管理控制台使用。

4. 保存私钥文件。

根据不同用途，需将私钥文件保存为不同格式。为保证安全，私钥只能下载一次，请妥善保管。

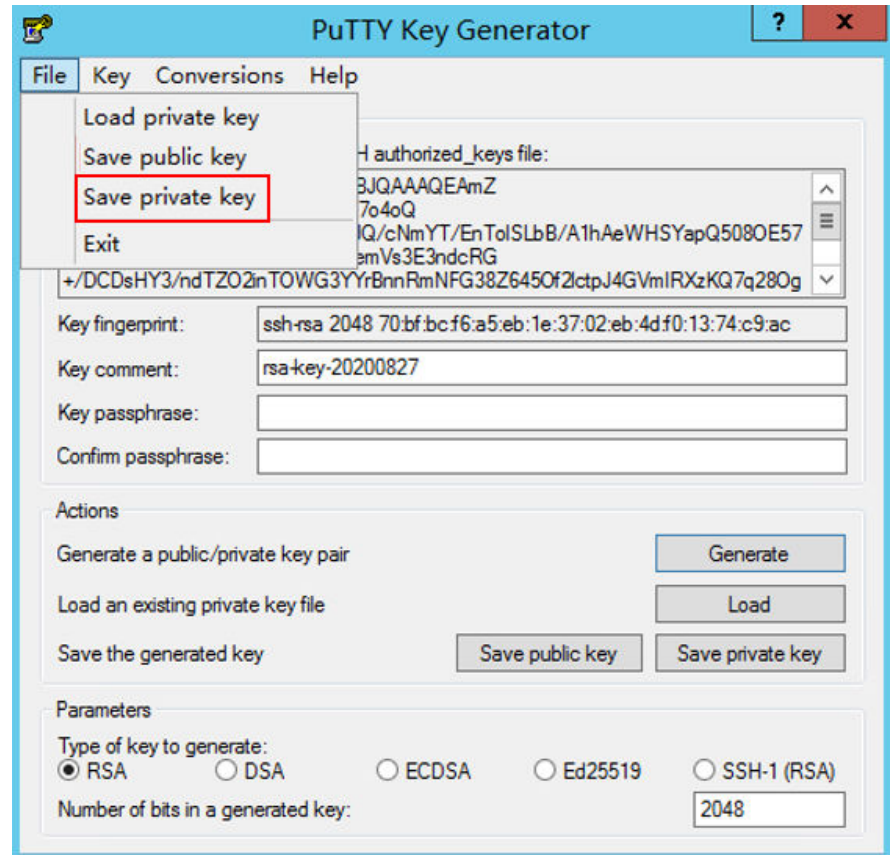
- 保存私钥文件为“.ppk”格式。

当用户使用PuTTY工具登录Linux弹性云服务器时，需使用“.ppk”格式的私钥文件。保存方法如下：

- i. 在“PuTTY Key Generator”界面，选择“File > Save private key”。



图 11-7 保存私钥文件



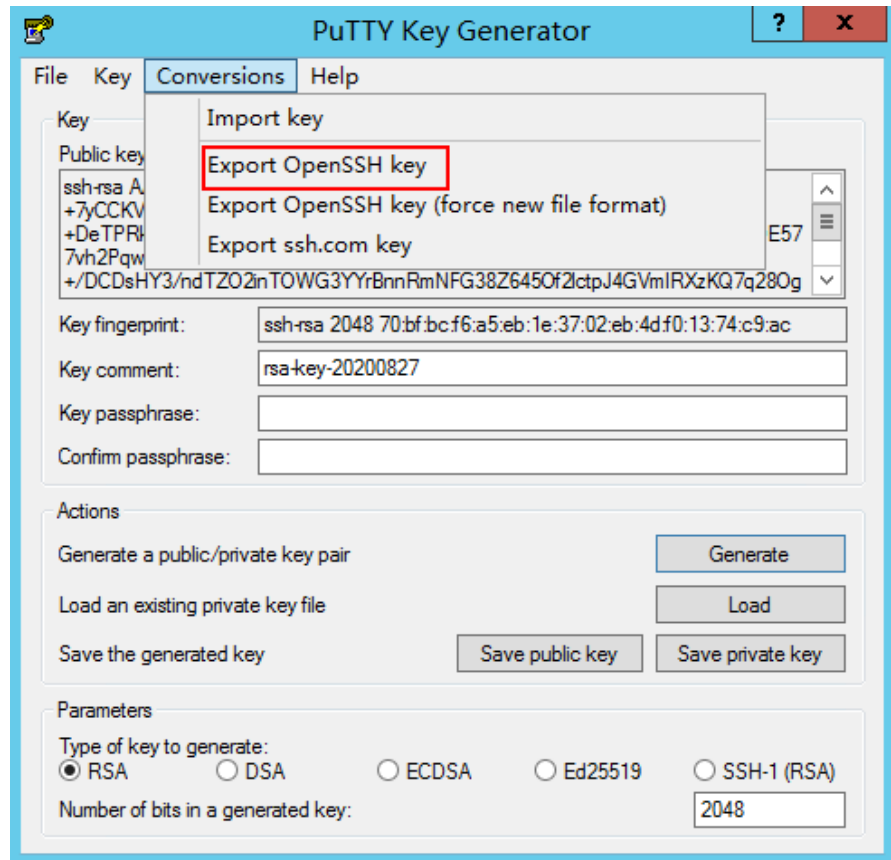
- ii. 保存转化后的私钥到本地。例如：kp-123.ppk
- 保存私钥文件为“.pem”格式。  
当用户使用Xshell工具登录Linux弹性云服务器，或者获取Windows弹性云服务器的密码时，需使用“.pem”格式的私钥文件。保存方法如下：
  - i. 选择“Conversions > Export OpenSSH key”。

**说明**

如果该私钥文件用于Windows弹性云服务器的获取密码操作，在选择“Export OpenSSH key”时，请勿填写“Key passphrase”信息，否则会导致获取密码失败。



图 11-8 保存私钥文件



- ii. 保存私钥到本地。例如：kp-123.pem
5. 公钥和私钥文件按照需要正确保存后，请参见[导入密钥对](#)的“复制公钥文件内容方式”，将公钥导入系统中。

## 11.3.4 导入密钥对

### 操作场景

以下两种场景，需要执行导入密钥对的操作：

- 通过puttygen.exe工具新建密钥对，需要将密钥对导入系统才能正常使用。
- 如果本地已有密钥对（例如，使用PuTTYgen工具生成的密钥对），可以在管理控制台导入密钥对公钥，由系统维护您的公钥文件。

#### 📖 说明

如果已有密钥对的公钥文件是通过puttygen.exe工具的“Save public key”按钮保存的，该公钥文件不能直接导入管理控制台。

如需继续使用该密钥对作远程身份验证，请参见[通过puttygen.exe工具创建的密钥对，导入管理控制台失败怎么办？](#)

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。

3. 在左侧导航树中，选择“密钥对”。
4. 在“密钥对”页面，单击“导入密钥对”。
5. 导入密钥对的方式有如下两种：
  - 选择文件方式：将本地已有的可用密钥，导入到系统中。
    - i. 在管理控制台“导入密钥对”页面，单击“选择文件”，选择本地保存的公钥文件（例如，[通过PuTTY Key Generator工具创建密钥对](#)的步骤3中保存的“.txt”格式文件）。

#### 说明

执行“导入密钥对”操作时，请确认导入的是公钥文件，否则会导入失败。

- ii. 单击“确定”，导入公钥。

公钥导入成功后，您可以修改公钥名称。
- 复制公钥文件内容方式：将“.txt”格式文本文档中的公钥内容粘贴至对话框，导入到系统中。
  - i. 复制文本文档中保存的公钥文件内容，并将其粘贴在“公钥内容”的空白区域。
  - ii. 单击“确定”，导入公钥。

## 相关链接

- [无法导入密钥对，怎么办？](#)
- [通过puttygen.exe工具创建的密钥对，导入管理控制台失败怎么办？](#)

## 11.3.5 获取和清除云服务器密码（Windows）

### 11.3.5.1 获取 Windows 弹性云服务器的密码

#### 操作场景

登录Windows操作系统的弹性云服务器时，需使用密码方式登录。因此，用户需先根据创建弹性云服务器时使用的密钥文件，获取该弹性云服务器初始安装时系统生成的管理员密码（Administrator账户或Cloudbase-init设置的账户）。该密码为随机密码，安全性高，请放心使用。

请根据您的个人需求，通过管理控制台或API方式获取Windows弹性云服务器的初始密码，获取方法请参见本节内容。

#### 通过管理控制台获取密码

1. 获取创建弹性云服务器时使用的私钥文件（.pem文件）。
2. 登录管理控制台。
3. 选择“计算 > 弹性云服务器”。
4. 在弹性云服务器列表，选择待获取密码的弹性云服务器。
5. 选择“操作 > 更多”，单击“获取密码”。
6. 通过密钥文件获取密码，获取方法有如下两种：
  - 单击“选择文件”，从本地上传密钥文件。

- 将密钥文件内容复制粘贴在空白框中。
- 7. 单击“获取密码”，获取随机密码。

## 通过 API 方式获取密码

1. 获取创建弹性云服务器时使用的私钥文件（.pem文件）。
2. 设置调用接口的环境。
3. 具体方法请参见《弹性云服务器接口参考》的“接口调用方法”。
4. 获取密文密码。

调用获取密码的接口，获取密钥的公钥加密后的密文密码。该密文密码是通过RSA算法加密的。其中，URI格式为“GET /v2/{project\_id}/servers/{server\_id}/os-server-password”。

### 📖 说明

调用接口的具体方法请参见《弹性云服务器接口参考》的“获取云服务器密码”章节。

5. 解密密文密码。  
通过创建云服务器时使用的密钥的私钥文件对4中获取的密文密码进行解密。
  - a. 使用OpenSSL工具，执行以下命令，将密文密码转换为.key -nocrypt格式。  
**openssl pkcs8 -topk8 -inform PEM -outform DER -in rsa\_pem.key -out pkcs8\_der.key -nocrypt**
  - b. 调用Java类库org.bouncycastle.jce.provider.BouncyCastleProvider，使用密钥文件，编写代码解密密文。

### 11.3.5.2 清除 Windows 弹性云服务器的密码

#### 操作场景

为安全起见，建议用户获取初始密码后，执行清除密码操作，清除系统中记录的初始密码信息。

该操作不会影响弹性云服务器的正常登录与运行。清除密码后，系统不能恢复获取密码功能，因此，请在执行清除密码操作前，记录弹性云服务器密码信息。

#### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表，选择待获取密码的弹性云服务器。
4. 选择“操作 > 更多”，单击“清除密码”。  
系统弹窗提示用户是否确认清除密码。
5. 单击“确定”，执行清除密码操作。

# 12 资源


## 12.1 配额调整

### 什么是配额？

为防止资源滥用，平台限定了各服务资源的配额，对用户的资源数量和容量做了限制。如您最多可以创建多少台弹性云服务器、多少块云硬盘。

如果当前资源配额限制无法满足使用需要，您可以申请扩大配额。

### 怎样查看我的配额？

1. 登录管理控制台。
2. 单击页面右上角的“**My Quota**”图标 。  
系统进入“服务配额”页面。
3. 您可以在“服务配额”页面，查看各项资源的总配额及使用情况。  
如果当前配额不能满足业务要求，请参考后续操作，申请扩大配额。

### 如何申请扩大配额？

目前系统暂不支持在线调整配额大小。如您需要调整配额，请联系运营管理员。

在联系运营管理员之前，请您准备好以下信息：

- 账号名，获取方式如下：  
登录云账户管理控制台，在右上角单击用户名，选择“我的凭证”，在“我的凭证”页面获取“账号名”。
- 配额信息，包括：服务名、配额类别、需要的配额值。

# 13 使用 CES 监控 ECS

## 13.1 监控弹性云服务器

监控是保持弹性云服务器可靠性、可用性和性能的重要部分，通过监控，用户可以观察弹性云服务器资源。为用户更好地掌握自己的弹性云服务器运行状态，云服务平台提供了云监控。您可以使用该服务监控您的弹性云服务器，执行自动实时监控、告警和通知操作，帮助您更好地了解弹性云服务器的各项性能指标。

通过后续章节，您可以了解以下内容：

- 弹性云服务器当前支持的基础监控指标
- 弹性云服务器操作系统监控的监控指标（安装Agent）
- 如何自定义弹性云服务器告警规则
- 如何查看弹性云服务器运行状态进行日常监控

## 13.2 弹性云服务器支持的基础监控指标

### 功能说明

本节定义了弹性云服务器上报云监控的基础监控指标的命名空间，监控指标列表，各项监控指标的具体含义与使用说明，用户可以通过云监控检索弹性云服务器服务产生的监控指标和告警信息。

### 命名空间

SYS.ECS

### 基础监控指标支持列表

对于不同的操作系统、不同的弹性云服务器类型，支持的基础监控指标有所差异，具体如表13-1所示。其中，√表示支持，×表示不支持。

表 13-1 弹性云服务器支持的基础监控指标列表

监控指标	Windows弹性云服务器	Linux弹性云服务器
CPU使用率	√	√
内存使用率	√	×
磁盘使用率	√	×
磁盘读带宽	√	√
磁盘写带宽	√	√
磁盘读IOPS	√	√
磁盘写IOPS	√	√
带内网络流入速率	√	×
带内网络流出速率	√	×
带外网络流入速率	√	√
带外网络流出速率	√	√

### 说明

“√”表示支持该监控指标，“×”表示不支持该监控指标。

各项基础监控指标的具体含义与使用说明如表13-2所示。

以下监控指标（原始指标）的监控周期如下：

- 虚拟化类型为XEN的弹性云服务器：4分钟。
- 虚拟化类型为KVM的弹性云服务器：5分钟。

表 13-2 弹性云服务器支持的基础监控指标

指标ID	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标，本列监控周期值适用于KVM实例）
cpu_util	CPU使用率	该指标用于统计弹性云服务器的CPU使用率。 单位：百分比。 计算公式：单个弹性云服务器CPU使用率 / 单个弹性云服务器的CPU总核数。	≥ 0%	云服务器	5分钟

指标ID	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标, 本列监控周期值适用于KVM实例)
mem_util	内存使用率	<p>该指标用于统计弹性云服务器的内存使用率。</p> <p>如果用户使用的镜像未安装UVP VMTTools, 则无法获取该监控指标。</p> <p>单位: 百分比。</p> <p>计算公式: 该弹性云服务器内存使用量 / 该弹性云服务器内存总量。</p> <p><b>说明</b> 内存使用率监控指标不支持 QingTian实例。</p>	≥ 0%	云服务器	5分钟
disk_util_inband	磁盘使用率	<p>该指标用于统计弹性云服务器的磁盘使用情况。</p> <p>如果用户使用的镜像未安装UVP VMTTools, 则无法获取该监控指标。</p> <p>单位: 百分比。</p> <p>计算公式: 该弹性云服务器磁盘使用容量 / 该弹性云服务器磁盘总容量。</p>	≥ 0%	云服务器	5分钟
disk_read_bytes_rate	磁盘读带宽	<p>该指标用于统计每秒从弹性云服务器读出数据量。</p> <p>单位: 字节/秒。</p> <p>计算公式: 该弹性云服务器的磁盘读出的字节数之和 / 测量周期。</p> <p>byte_out = (rd_bytes - last_rd_bytes) / 时间差。</p>	≥ 0 Byte/s	云服务器	5分钟
disk_write_bytes_rate	磁盘写带宽	<p>该指标用于统计每秒写到弹性云服务器的数据量。</p> <p>单位: 字节/秒。</p> <p>计算公式: 该弹性云服务器的磁盘写入的字节数之和 / 测量周期。</p>	≥ 0 Byte/s	云服务器	5分钟

指标ID	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标, 本列监控周期值适用于KVM实例)
disk_read_requests_rate	磁盘读 IOPS	该指标用于统计每秒从弹性云服务器读取数据的请求次数。 单位: 请求/秒。 计算公式: 请求读取该弹性云服务器磁盘的次数之和 / 测量周期。 $req\_out = (rd\_req - last\_rd\_req) / \text{时间差}$ 。	$\geq 0$ requests/s	云服务器	5分钟
disk_write_requests_rate	磁盘写 IOPS	该指标用于统计每秒从弹性云服务器写数据的请求次数。 单位: 请求/秒。 计算公式: 请求写入该弹性云服务器磁盘的次数之和 / 测量周期。 $req\_in = (wr\_req - last\_wr\_req) / \text{时间差}$ 。	$\geq 0$ requests/s	云服务器	5分钟
network_incoming_bytes_rate_inband	带内网络流入速率	该指标用于在弹性云服务器内统计每秒流入弹性云服务器的网络流量。 单位: 字节/秒。 计算公式: 该弹性云服务器的带内网络流入字节数之和/测量周期。	$\geq 0$ Byte/s	云服务器	5分钟
network_outgoing_bytes_rate_inband	带内网络流出速率	该指标用于在弹性云服务器内统计每秒流出弹性云服务器的网络流量。 单位: 字节/秒。 计算公式: 该弹性云服务器的带内网络流出字节数之和 / 测量周期。	$\geq 0$ Byte/s	云服务器	5分钟



指标ID	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标, 本列监控周期值适用于KVM实例)
network_incoming_bytes_aggregate_rate	带外网络流入速率	该指标用于在虚拟化层统计每秒流入弹性云服务器的网络流量。 单位: 字节/秒。 计算公式: 该弹性云服务器的带外网络流入字节数之和 / 测量周期。 当使用SRIOV时, 无法获取该监控指标。	$\geq 0$ Byte/s	云服务器	5分钟
network_outgoing_bytes_aggregate_rate	带外网络流出速率	该指标用于在虚拟化层统计每秒流出弹性云服务器的网络流量。 单位: 字节/秒。 计算公式: 该弹性云服务器的带外网络流出字节数之和 / 测量周期。 当使用SRIOV时, 无法获取该监控指标。	$\geq 0$ Byte/s	云服务器	5分钟

## 维度

Key	Value
instance_id	云服务器ID。

## 13.3 弹性云服务器支持的操作系统监控指标（安装 Agent）

### 功能说明

通过在弹性云服务器中安装Agent插件, 可以为用户提供服务器的系统级、主动式、细颗粒度监控服务。本节定义了弹性云服务器上报云监控的操作系统监控指标。

操作系统监控目前支持的监控指标有: CPU相关监控项、CPU负载类相关监控项、内存相关监控项、磁盘相关监控项、磁盘I/O相关监控项、文件系统类相关监控项、网卡类相关监控项、NTP类相关监控项、TCP连接数类相关监控。

安装Agent后, 对于不同的操作系统、不同的弹性云服务器类型, 您可以查看不同类型的操作系统监控指标。指标采集周期是1分钟。

## 命名空间

AGT.ECS

## 操作系统监控指标：CPU

表 13-3 CPU 相关监控指标说明

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
cpu_usage	(Agent) CPU使用率	该指标用于统计测量对象当前CPU使用率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/stat中的变化得出cpu使用率。用户可以通过top命令查看%Cpu(s)值。</li><li>采集方式（Windows）：通过WindowsAPI GetSystemTimes获取。</li></ul>	0-100%	云服务器	1分钟
cpu_usage_idle	(Agent) CPU空闲时间占比	该指标用于统计测量对象当前CPU空闲时间占比。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/stat中的变化得出CPU空闲时间占比。</li><li>采集方式（Windows）：通过WindowsAPI GetSystemTimes获取。</li></ul>	0-100%	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象(维度)	监控周期(原始指标)
cpu_usage_user	(Agent) 用户空间 CPU使用率	该指标用于统计测量对象当前用户空间占用CPU使用率。 单位: 百分比 <ul style="list-style-type: none"> <li>采集方式 (Linux): 通过计算采集周期内/proc/stat中的变化得出cpu使用率。用户可以通过top命令查看%Cpu(s) us值。</li> <li>采集方式 (Windows): 通过WindowsAPI GetSystemTimes获取。</li> </ul>	0-100%	云服务器	1分钟
cpu_usage_system	(Agent) 内核空间 CPU使用率	该指标用于统计测量对象当前内核空间占用CPU使用率。 单位: 百分比 <ul style="list-style-type: none"> <li>采集方式 (Linux): 通过计算采集周期内/proc/stat中的变化得出内核空间CPU使用率。用户可以通过top命令查看%Cpu(s) sy值。</li> <li>采集方式 (Windows): 通过WindowsAPI GetSystemTimes获取。</li> </ul>	0-100%	云服务器	1分钟
cpu_usage_other	(Agent) 其他CPU使用率	该指标用于统计测量对象其他占用CPU使用率。 单位: 百分比 <ul style="list-style-type: none"> <li>采集方式 (Linux): 其他CPU使用率=1- 空闲CPU使用率(%) - 内核空间CPU使用率- 用户空间CPU使用率。</li> <li>采集方式 (Windows): 其他CPU使用率=1- 空闲CPU使用率(%) - 内核空间CPU使用率- 用户空间CPU使用率。</li> </ul>	0-100%	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象(维度)	监控周期(原始指标)
cpu_usage_nice	(Agent) Nice进程CPU使用率	该指标用于统计测量对象当前Nice进程CPU使用率。 单位: 百分比 <ul style="list-style-type: none"><li>采集方式 (Linux): 通过计算采集周期内/proc/stat中的变化得出Nice进程CPU使用率。用户可以通过top命令查看 %Cpu(s) ni值。</li><li>采集方式 (Windows): 暂不支持。</li></ul>	0-100%	云服务器	1分钟
cpu_usage_iowait	(Agent) iowait状态占比	该指标用于统计测量对象当前iowait状态占用CPU的比率。 单位: 百分比 <ul style="list-style-type: none"><li>采集方式 (Linux): 通过计算采集周期内/proc/stat中的变化得出iowait状态占比。用户可以通过top命令查看 %Cpu(s) wa值。</li><li>采集方式 (Windows): 暂不支持。</li></ul>	0-100%	云服务器	1分钟
cpu_usage_irq	(Agent) CPU中断时间占比	该指标用于统计测量对象当前CPU处理中断时占用CPU时间的比率。 单位: 百分比 <ul style="list-style-type: none"><li>采集方式 (Linux): 通过计算采集周期内/proc/stat中的变化得出CPU中断时间占比。用户可以通过top命令查看 %Cpu(s) hi值。</li><li>采集方式 (Windows): 暂不支持。</li></ul>	0-100%	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
cpu_usage_softirq	(Agent) CPU软中断时间占比	该指标用于统计测量对象当前CPU处理软中断时间占用CPU时间的比率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/proc/stat中的变化得出CPU软中断时间占比。用户可以通过top命令查看 %Cpu(s) si值。</li><li>采集方式（Windows）：暂不支持。</li></ul>	0-100%	云服务器	1分钟

## 操作系统监控指标：CPU 负载

表 13-4 CPU 负载指标说明

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
load_averager1	(Agent) 1分钟平均负载	该指标用于统计测量对象过去1分钟的CPU平均负载。 采集方式（Linux）：通过/proc/loadavg中load1/逻辑CPU个数得到。用户可以通过top命令查看load1值。	≥0	云服务器	1分钟
load_averager5	(Agent) 5分钟平均负载	该指标用于统计测量对象过去5分钟的CPU平均负载。 采集方式（Linux）：通过/proc/loadavg中load5/逻辑CPU个数得到。用户可以通过top命令查看load5值。	≥0	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
load_average15	(Agent) 15分钟平均负载	该指标用于统计测量对象过去15分钟的CPU平均负载。 采集方式（Linux）：通过/proc/loadavg中load15/逻辑CPU个数得到。用户可以通过top命令查看load15值。	≥0	云服务器	1分钟

 说明

Windows系统暂不支持CPU负载指标。

## 操作系统监控指标：内存

表 13-5 内存相关监控指标说明

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
mem_available	(Agent) 可用内存	<p>该指标用于统计测量对象的可用内存。</p> <p>单位：GB</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过/proc/meminfo文件获取，<ul style="list-style-type: none"><li>若/proc/meminfo中显示MemAvailable，则直接可得</li><li>若/proc/meminfo中不显示MemAvailable，则MemAvailable=MemFree+Buffers+Cached</li></ul></li><li>采集方式（Windows）：计算方法为（内存总量-已用内存量）。通过WindowsAPI GlobalMemoryStatusEx获取。</li></ul>	≥0 GB	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
mem_usedPercent	(Agent) 内存使用率	<p>该指标用于统计测量对象的内存使用率。</p> <p>单位：百分比</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过/proc/meminfo文件获取，<math>(\text{MemTotal} - \text{MemAvailable}) / \text{MemTotal}</math><ul style="list-style-type: none"><li>若/proc/meminfo中显示 MemAvailable，则 <math>\text{MemUsedPercent} = (\text{MemTotal} - \text{MemAvailable}) / \text{MemTotal}</math></li><li>若/proc/meminfo中不显示 MemAvailable，则 <math>\text{MemUsedPercent} = (\text{MemTotal} - \text{MemFree} - \text{Buffers} - \text{Cached}) / \text{MemTotal}</math></li></ul></li><li>采集方式（Windows）：计算方法为（已用内存量/内存总量*100%）。</li></ul>	0-100%	云服务器	1分钟
mem_free	(Agent) 空闲内存量	<p>该指标用于统计测量对象的空闲内存量。</p> <p>单位：GB</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过/proc/meminfo获取。</li><li>采集方式（Windows）：暂不支持。</li></ul>	≥0 GB	云服务器	1分钟



指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
mem_buffers	(Agent) Buffers占用量	该指标用于统计测量对象的Buffers内存量。 单位：GB <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/meminfo获取。用户可以通过top命令查看 KiB Mem:buffers值。</li><li>采集方式（Windows）：暂不支持。</li></ul>	≥0 GB	云服务器	1分钟
mem_cached	(Agent) Cache占用量	该指标用于统计测量对象Cache内存量。 单位：GB <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/meminfo获取。用户可以通过top命令查看 KiB Swap:cached Mem 值。</li><li>采集方式（Windows）：暂不支持。</li></ul>	≥0 GB	云服务器	1分钟
total_open_files	(Agent) 文件句柄总数	该指标用于统计测量对象的所有进程使用的句柄总和。 单位：个 <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/{pid}/fd文件汇总所有进程使用的句柄数。</li><li>采集方式（Windows）：暂不支持。</li></ul>	≥0	云服务器	1分钟

## 操作系统监控指标：磁盘

### 说明

- 目前仅支持物理磁盘指标的采集，不支持通过网络文件系统协议挂载的磁盘。
- 会默认屏蔽docker相关的挂载点。挂载点前缀如下：  
`/var/lib/docker/mnt/paas/kubernetes;/var/lib/mesos`

表 13-6 磁盘相关监控指标说明

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_free	(Agent) 磁盘剩余 存储量	该指标用于统计测量对象 磁盘的剩余存储空间。 单位：GB <ul style="list-style-type: none"><li>采集方式（Linux）： 执行df -h命令，查看 Avail列数据。挂载点前 缀路径长度不能超过64 个字符，必须以字母开 头，只能包含0-9/a- z/A-Z/-/./~。</li><li>采集方式 （Windows）：使用 WMI接口 GetDiskFreeSpaceExW 获取磁盘空间数据。挂 载点前缀路径长度不能 超过64个字符，必须以 字母开头，只能包含 0-9/a-z/A-Z/-/./~。</li></ul>	≥0 GB	云服务器 - 挂载点	1分钟
disk_total	(Agent) 磁盘存储 总量	该指标用于统计测量对象 磁盘存储总量。 单位：GB <ul style="list-style-type: none"><li>采集方式（Linux）： 执行df -h命令，查看 Size列数据。 挂载点前缀路径长度不 能超过64个字符，必须 以字母开头，只能包含 0-9/a-z/A-Z/-/./~。</li><li>采集方式 （Windows）：使用 WMI接口 GetDiskFreeSpaceExW 获取磁盘空间数据。挂 载点前缀路径长度不能 超过64个字符，必须以 字母开头，只能包含 0-9/a-z/A-Z/-/./~。</li></ul>	≥0 GB	云服务器 - 挂载点	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_use d	(Agent) 磁盘已用 存量	该指标用于统计测量对象 磁盘的已用存储空间。 单位：GB <ul style="list-style-type: none"> <li>采集方式（Linux）： 执行df -h命令，查看 Used列数据。挂载点前 缀路径长度不能超过64 个字符，必须以字母开 头，只能包含0-9/a- z/A-Z/-/./~。</li> <li>采集方式 （Windows）：使用 WMI接口 GetDiskFreeSpaceExW 获取磁盘空间数据。挂 载点前缀路径长度不能 超过64个字符，必须以 字母开头，只能包含 0-9/a-z/A-Z/-/./~。</li> </ul>	≥0 GB	云服务器 - 挂载 点	1分钟
disk_use dPercen t	(Agent) 磁盘使用 率	该指标用于统计测量对象 磁盘使用率，以百分比为 单位。计算方式为：磁盘已 用存储量/磁盘存储总量。 单位：百分比 <ul style="list-style-type: none"> <li>采集方式（Linux）： 通过计算Used/Size得 出。挂载点前缀路径长 度不能超过64个字符， 必须以字母开头，只能 包含0-9/a-z/A- Z/-/./~。</li> <li>采集方式 （Windows）：使用 WMI接口 GetDiskFreeSpaceExW 获取磁盘空间数据。挂 载点前缀路径长度不能 超过64个字符，必须以 字母开头，只能包含 0-9/a-z/A-Z/-/./~。</li> </ul>	0-100%	云服务器 - 挂载 点	1分钟

## 操作系统监控指标：磁盘 I/O

表 13-7 磁盘 I/O 相关监控指标说明

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_agt_read_bytes_rate	(Agent) 磁盘读速率	<p>该指标用于统计每秒从测量对象读出数据量。 单位：Byte/s</p> <ul style="list-style-type: none"> <li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第六列数据的变化得出磁盘读速率。 挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li> <li>采集方式（Windows）： <ul style="list-style-type: none"> <li>使用WMI中Win32_PerfFormattedData_PerfDisk_LogicalDisk对象获取磁盘I/O数据。</li> <li>挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li> <li>高CPU情况下存在获取超时的现象，会导致无法获取监控数据。</li> </ul> </li> </ul>	≥ 0 Byte/s	<ul style="list-style-type: none"> <li>云服务器 - 磁盘</li> <li>云服务器 - 挂载点</li> </ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_agt_read_requests_rate	(Agent) 磁盘读操作速率	<p>该指标用于统计每秒从测量对象读取数据的请求次数。</p> <p>单位：请求/秒</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第四列数据的变化得出磁盘读操作速率。 挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li><li>采集方式（Windows）：<ul style="list-style-type: none"><li>使用WMI中Win32_PerfFormattedData_PerfDisk_LogicalDisk对象获取磁盘I/O数据。</li><li>挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li><li>高CPU情况下存在获取超时的现象，会导致无法获取监控数据。</li></ul></li></ul>	≥ 0 请求/秒	<ul style="list-style-type: none"><li>云服务器 - 磁盘</li><li>云服务器 - 挂载点</li></ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_agt_write_bytes_rate	(Agent) 磁盘写速率	<p>该指标用于统计每秒写到测量对象的数据量。 单位：Byte/s</p> <ul style="list-style-type: none"> <li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第十列数据的变化得出磁盘写速率。 挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li> <li>采集方式（Windows）： <ul style="list-style-type: none"> <li>使用WMI中Win32_PerfFormattedData_PerfDisk_LogicalDisk对象获取磁盘I/O数据。</li> <li>挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li> <li>高CPU情况下存在获取超时的现象，会导致无法获取监控数据。</li> </ul> </li> </ul>	≥ 0 Byte/s	<ul style="list-style-type: none"> <li>云服务器 - 磁盘</li> <li>云服务器 - 挂载点</li> </ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_agt_write_requests_rate	(Agent) 磁盘写操作速率	<p>该指标用于统计每秒向测量对象写数据的请求次数。</p> <p>单位：请求/秒</p> <ul style="list-style-type: none"> <li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第八列数据的变化得出磁盘写操作速率。 挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li> <li>采集方式（Windows）： <ul style="list-style-type: none"> <li>使用WMI中Win32_PerfFormattedData_PerfDisk_LogicalDisk对象获取磁盘I/O数据。</li> <li>挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li> <li>高CPU情况下存在获取超时的现象，会导致无法获取监控数据。</li> </ul> </li> </ul>	≥ 0 请求/秒	<ul style="list-style-type: none"> <li>云服务器 - 磁盘</li> <li>云服务器 - 挂载点</li> </ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_readTime	(Agent) 读操作平均耗时	<p>该指标用于统计测量对象磁盘读操作平均耗时。</p> <p>单位：ms/Count</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第七列数据的变化得出磁盘读操作平均耗时。</li></ul> <p>挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</p> <ul style="list-style-type: none"><li>采集方式（Windows）：暂不支持。</li></ul>	≥ 0 ms/ Count	<ul style="list-style-type: none"><li>云服务器 - 磁盘</li><li>云服务器 - 挂载点</li></ul>	1分钟
disk_writeTime	(Agent) 写操作平均耗时	<p>该指标用于统计测量对象磁盘写操作平均耗时。</p> <p>单位：ms/Count</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第十一列数据的变化得出磁盘写操作平均耗时。</li></ul> <p>挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</p> <ul style="list-style-type: none"><li>采集方式（Windows）：暂不支持。</li></ul>	≥ 0 ms/ Count	<ul style="list-style-type: none"><li>云服务器 - 磁盘</li><li>云服务器 - 挂载点</li></ul>	1分钟



指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_ioU tils	(Agent) 磁盘I/O使 用率	该指标用于统计测量对象 磁盘I/O使用率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/ proc/diskstats中对应设 备第十三列数据的变化 得出磁盘I/O使用率。 挂载点前缀路径长度不 能超过64个字符，必须 以字母开头，只能包含 0-9/a-z/A-Z/-/./~。</li><li>采集方式 （Windows）：暂不支 持。</li></ul>	0-100%	<ul style="list-style-type: none"><li>云 服 务 器 - 磁 盘</li><li>云 服 务 器 - 挂 载 点</li></ul>	1分钟
disk_que ue_lengt h	(Agent) 平均队列 长度	该指标用于统计指定时间 段内，平均等待完成的读 取或写入操作请求的数量 单位：个 <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/ proc/diskstats中对应设 备第十四列数据的变化 得出磁盘平均队列长 度。 挂载点前缀路径长度不 能超过64个字符，必须 以字母开头，只能包含 0-9/a-z/A-Z/-/./~。</li><li>采集方式 （Windows）：暂不支 持。</li></ul>	≥ 0	<ul style="list-style-type: none"><li>云 服 务 器 - 磁 盘</li><li>云 服 务 器 - 挂 载 点</li></ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_write_bytes_per_operation	(Agent) 平均写操作大小	该指标用于统计指定时间段内，平均每个写I/O操作传输的字节数。 单位：Byte/op <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第十列数据的变化与第八列数据的变化相除得出磁盘平均写操作大小。 挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li><li>采集方式（Windows）：暂不支持。</li></ul>	≥ 0 Byte/op	<ul style="list-style-type: none"><li>云服务器 - 磁盘</li><li>云服务器 - 挂载点</li></ul>	1分钟
disk_read_bytes_per_operation	(Agent) 平均读操作大小	该指标用于统计指定时间段内，平均每个读I/O操作传输的字节数。 单位：Byte/op <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第六列数据的变化与第四列数据的变化相除得出磁盘平均读操作大小。 挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li><li>采集方式（Windows）：暂不支持。</li></ul>	≥ 0 Byte/op	<ul style="list-style-type: none"><li>云服务器 - 磁盘</li><li>云服务器 - 挂载点</li></ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_io_svctm	(Agent) 平均I/O服务时长	<p>该指标用于统计指定时间段内，平均每个读或写I/O的操作时长。</p> <p>单位：ms/op</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过计算采集周期内/proc/diskstats中对应设备第十三列数据的变化与第四列数据和第八列数据之和的变化相除得出磁盘平均I/O时长。 挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。</li><li>采集方式（Windows）：暂不支持。</li></ul>	≥ 0 ms/op	<ul style="list-style-type: none"><li>云服务器 - 磁盘</li><li>云服务器 - 挂载点</li></ul>	1分钟
disk_device_used_percent	块设备使用率	<p>该指标用于统计测量对象物理磁盘使用率，以百分比为单位。计算方式为：所有已挂载磁盘分区已用存储量/磁盘存储总量。</p> <ul style="list-style-type: none"><li>采集方式（Linux）： 通过汇总每个挂载点的磁盘使用量，再通过磁盘扇区大小和扇区数量计算出磁盘总大小，计算出整体磁盘使用率</li><li>采集方式（Windows）：暂不支持。</li></ul>	0-100%	云服务器 - 磁盘	1分钟

## 操作系统监控指标：文件系统

表 13-8 文件系统类监控指标说明

指标	指标名称	指标含义	取值范围	测量对象 (维度)	监控周期 (原始指标)
disk_fs_rwstate	(Agent) 文件系统读写状态	该指标用于统计测量对象挂载文件系统的读写状态。状态分为：可读写（0）/只读（1）。 采集方式（Linux）：通过读取/proc/mounts中第四列文件系统挂载参数获得。	<ul style="list-style-type: none"> <li>0: 可读写</li> <li>1: 只读</li> </ul>	云服务器 - 挂载点	1分钟
disk_inodesTotal	(Agent) inode空间大小	该指标用于统计测量对象当前磁盘的inode空间量。 采集方式（Linux）：执行df -i命令，查看Inodes列数据。挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。	$\geq 0$	云服务器 - 挂载点	1分钟
disk_inodesUsed	(Agent) inode已使用空间	该指标用于统计测量对象当前磁盘已使用的inode空间量。 采集方式（Linux）：执行df -i命令，查看IUsed列数据。挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。	$\geq 0$	云服务器 - 挂载点	1分钟
disk_inodesUsedPercent	(Agent) inode已使用占比	该指标用于统计测量对象当前磁盘已使用的inode占比。 单位：百分比 采集方式（Linux）：执行df -i命令，查看IUse%列数据。挂载点前缀路径长度不能超过64个字符，必须以字母开头，只能包含0-9/a-z/A-Z/-/./~。	0-100%	云服务器 - 挂载点	1分钟

 说明

Windows系统暂不支持文件系统类监控指标。

## 操作系统监控指标：网卡

表 13-9 网卡相关监控指标说明

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
net_bitRecv	(Agent) 出网带宽	该指标用于统计测量对象网卡每秒接收的比特数。 单位：bit/s <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：使用WMI中MibIfRow对象获取网络指标数据。</li></ul>	$\geq 0$ bit/s	云服务器	1分钟
net_bitSent	(Agent) 入网带宽	该指标用于统计测量对象网卡每秒发送的比特数。 单位：bit/s <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：使用WMI中MibIfRow对象获取网络指标数据。</li></ul>	$\geq 0$ bit/s	云服务器	1分钟
net_packetRecv	(Agent) 网卡包接收速率	该指标用于统计测量对象网卡每秒接收的数据包数。 单位：Counts/s <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：使用WMI中MibIfRow对象获取网络指标数据。</li></ul>	$\geq 0$ Counts/s	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
net_packetSent	(Agent) 网卡包发送速率	该指标用于统计测量对象网卡每秒发送的数据包数。 单位：Counts/s <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：使用WMI中MibIfRow对象获取网络指标数据。</li></ul>	≥ 0 Counts/s	云服务器	1分钟
net_errin	(Agent) 接收误包率	该指标用于统计测量对象网卡每秒接收的错误数据包数量占所接收的数据包的比率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：暂不支持。</li></ul>	0-100%	云服务器	1分钟
net_errout	(Agent) 发送误包率	该指标用于统计测量对象网卡每秒发送的错误数据包数量占所发送的数据包的比率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：暂不支持。</li></ul>	0-100%	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
net_dropin	(Agent) 接收丢包率	该指标用于统计测量对象网卡每秒接收并已丢弃的数据包数量占所接收的数据包的比率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：暂不支持。</li></ul>	0-100%	云服务器	1分钟
net_dropout	(Agent) 发送丢包率	该指标用于统计测量对象网卡每秒发送并已丢弃的数据包数量占所发送的数据包的比率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过计算采集周期内/proc/net/dev中的变化得出。</li><li>采集方式（Windows）：暂不支持。</li></ul>	0-100%	云服务器	1分钟

## 操作系统监控指标：NTP

表 13-10 NTP 类监控指标说明

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
ntp_offset	(Agent) NTP偏移量	该指标用于统计测量对象当前NTP偏移量。 单位：ms 采集方式（Linux）：执行chronyc sources -v命令，获取偏移量。	≥ 0 ms	云服务器	1分钟

## 操作系统监控指标：TCP

表 13-11 TCP 类监控指标说明

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
net_tcp_total	(Agent) TCP TOTAL	该指标用于统计测量对象所有状态的TCP连接数总和。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟
net_tcp_established	(Agent) TCP ESTABLISHED	该指标用于统计测量对象处于ESTABLISHED状态的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟
net_tcp_sys_sent	(Agent) TCP SYS_SENT	该指标用于统计测量对象处于请求连接状态的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟



指标	指标名称	指标含义	取值范围	测量对象(维度)	监控周期(原始指标)
net_tcp_sys_recv	(Agent) TCP SYS_RECV	该指标用于统计测量对象服务器端收到的请求连接的TCP数量。 单位: Count <ul style="list-style-type: none"><li>采集方式(Linux): 通过/proc/net/tcp文件获取到所有状态的TCP连接,再统计每个状态的连接数量。</li><li>采集方式(Windows): 通过WindowsAPI GetTcpTable2获取。</li></ul>	$\geq 0$	云服务器	1分钟
net_tcp_fin_wait1	(Agent) TCP FIN_WAIT 1	该指标用于统计测量对象客户端主动关闭且没有收到服务端ACK的TCP连接数量。 单位: Count <ul style="list-style-type: none"><li>采集方式(Linux): 通过/proc/net/tcp文件获取到所有状态的TCP连接,再统计每个状态的连接数量。</li><li>采集方式(Windows): 通过WindowsAPI GetTcpTable2获取。</li></ul>	$\geq 0$	云服务器	1分钟
net_tcp_fin_wait2	(Agent) TCP FIN_WAIT 2	该指标用于统计测量对象处于FIN_WAIT2状态的TCP连接数量。 单位: Count <ul style="list-style-type: none"><li>采集方式(Linux): 通过/proc/net/tcp文件获取到所有状态的TCP连接,再统计每个状态的连接数量。</li><li>采集方式(Windows): 通过WindowsAPI GetTcpTable2获取。</li></ul>	$\geq 0$	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
net_tcp_time_wait	(Agent) TCP TIME_WAIT	该指标用于统计测量对象处于TIME_WAIT状态的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟
net_tcp_close	(Agent) TCP CLOSE	该指标用于统计测量对象关闭的或未打开的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟
net_tcp_close_wait	(Agent) TCP CLOSE_WAIT	该指标用于统计测量对象处于CLOSE_WAIT状态的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
net_tcp_last_ack	(Agent) TCP LAST_ACK	该指标用于统计测量对象被动关闭等待ACK报文的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟
net_tcp_listen	(Agent) TCP LISTEN	该指标用于统计测量对象处于LISTEN状态的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟
net_tcp_closing	(Agent) TCP CLOSING	该指标用于统计测量对象处于服务端和客户端同时主动关闭状态的TCP连接数量。 单位：Count <ul style="list-style-type: none"><li>采集方式（Linux）：通过/proc/net/tcp文件获取到所有状态的TCP连接，再统计每个状态的连接数量。</li><li>采集方式（Windows）：通过WindowsAPI GetTcpTable2获取。</li></ul>	≥ 0	云服务器	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
net_tcp_retrans	(Agent) TCP重传率	该指标用于统计测量对象重新发送的报文数与总发送的报文数之间的比值。 单位：百分比 <ul style="list-style-type: none"> <li>采集方式（Linux）：通过从/proc/net/snmp文件中获取对应的数据，计算采集周期内发送包数和重传包数的比值得出。</li> <li>采集方式（Windows）：重传率通过WindowsAPI GetTcpStatistics获取</li> </ul>	0-100%	云服务器	1分钟

## 操作系统监控指标：GPU

表 13-12 GPU 类监控指标说明

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
gpu_status	gpu健康状态	该指标用于统计虚拟机上GPU健康状态，是一个综合指标。 该指标无单位。 <ul style="list-style-type: none"> <li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li> <li>采集方式（Windows）：通过调用GPU卡的nvidia-smi.dll库获取。</li> </ul>	<ul style="list-style-type: none"> <li>0：代表健康</li> <li>1：代表亚健康</li> <li>2：代表故障</li> </ul>	<ul style="list-style-type: none"> <li>云服务器</li> <li>云服务器 - GPU</li> </ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
gpu_usage_encoder	编码使用率	该指标用于统计该GPU的编码能力使用率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvidia-smi库获取。</li></ul>	0-100%	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟
gpu_usage_decoder	解码使用率	该指标用于统计该GPU的解码能力使用率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvidia-smi库获取。</li></ul>	0-100%	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟
gpu_volatile_correctable	可纠正ECC错误数量	该指标用于统计该GPU重置以来可纠正的ECC错误数量，每次重置后归0。 单位：个。 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvidia-smi库获取。</li></ul>	≥ 0	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
gpu_volatile_uncorrectable	不可纠正ECC错误数量	该指标用于统计该GPU重置以来不可纠正的ECC错误数量，每次重置后归0。 单位：个 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvml.dll库获取。</li></ul>	$\geq 0$	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟
gpu_aggregate_correctable	累计可纠正ECC错误数量	该指标用于统计该GPU累计的可纠正ECC错误数量。 单位：个 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvml.dll库获取。</li></ul>	$\geq 0$	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟
gpu_aggregate_uncorrectable	累计不可纠正ECC错误数量	该指标用于统计该GPU累计的不可纠正ECC错误数量。 单位：个 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvml.dll库获取。</li></ul>	$\geq 0$	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
gpu_retired_page_single_bit	retired page single bit 错误数量	该指标用于统计该GPU当前卡隔离的单比特页的数量。 单位：个 <ul style="list-style-type: none"> <li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li> <li>采集方式（Windows）：通过调用GPU卡的nvidia-smi.dll库获取。</li> </ul>	≥ 0	<ul style="list-style-type: none"> <li>云服务器 - GPU</li> </ul>	1分钟
gpu_retired_page_double_bit	retired page double bit 错误数量	该指标用于统计该GPU当前卡隔离的双比特页的数量。 单位：个 <ul style="list-style-type: none"> <li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li> <li>采集方式（Windows）：通过调用GPU卡的nvidia-smi.dll库获取。</li> </ul>	≥ 0	<ul style="list-style-type: none"> <li>云服务器 - GPU</li> </ul>	1分钟
gpu_performance_state	(Agent) 性能状态	该指标用于统计测量对象当前的GPU性能状态。 该指标无单位。 <ul style="list-style-type: none"> <li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li> <li>采集方式（Windows）：通过调用GPU卡的nvidia-smi.dll库获取。</li> </ul>	P0-P15、P32, <ul style="list-style-type: none"> <li>P0：表示最大性能状态</li> <li>P15：表示最小性能状态</li> <li>P32：表示状态未知</li> </ul>	<ul style="list-style-type: none"> <li>云服务器 - GPU</li> </ul>	1分钟

指标	指标名称	指标含义	取值范围	测量对象（维度）	监控周期（原始指标）
gpu_usage_memory	(Agent) 显存使用率	该指标用于统计测量对象当前的显存使用率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvidia-smi.dll库获取。</li></ul>	0-100%	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟
gpu_usage_gpu	(Agent) GPU使用率	该指标用于统计测量对象当前的GPU使用率。 单位：百分比 <ul style="list-style-type: none"><li>采集方式（Linux）：通过调用GPU卡的libnvidia-ml.so.1库文件获取。</li><li>采集方式（Windows）：通过调用GPU卡的nvidia-smi.dll库获取。</li></ul>	0-100%	<ul style="list-style-type: none"><li>云服务器</li><li>云服务器 - GPU</li></ul>	1分钟

## 维度

维度	Key	Value
云服务器	instance_id	云服务器ID。

## 13.4 设置告警规则

### 操作场景

通过设置弹性云服务器告警规则，用户可自定义监控目标与通知策略，及时了解弹性云服务器运行状况，从而起到预警作用。

设置弹性云服务器的告警规则包括设置告警规则名称、监控对象、监控指标、告警阈值、监控周期和是否发送通知等参数。本节介绍了设置弹性云服务器告警规则的具体方法。



## 操作步骤

1. 登录管理控制台。
2. 选择“管理与部署 > 云监控服务”。
3. 在左侧导航树栏，选择“告警 > 告警规则”。
4. 在“告警规则”界面，单击“创建告警规则”进行添加，或者选择已有的告警规则进行修改，设置弹性云服务器的告警规则。

以修改弹性云服务器的告警规则为例，介绍如何设置告警规则：

- a. 单击待修改的告警规则名称，进入详情页。
- b. 单击右上角的“修改”。
- c. 在“修改告警规则”界面，根据界面提示配置参数，如图13-1所示。

图 13-1 修改告警规则

The screenshot shows a 'Modify Alarm Rule' dialog box with the following configuration:

- Name: as-alarm-d3o6
- Description: autoScaling
- Service: Auto Scaling
- Monitored Object: AS Group (406487d6-a3d4-...)
- Metric: CPU Usage
- Threshold: Max. > 80 %
- Occurrences: 1
- Monitoring Interval: 5 minutes
- Send Notification: No (selected)

Buttons: OK, Cancel

- d. 单击“确定”。

弹性云服务器告警规则设置完成后，当符合规则的告警产生时，系统会自动进行通知。

### 说明

更多关于弹性云服务器监控规则的信息，请参见《云监控用户指南》。

## 13.5 查看监控指标

### 操作场景

云服务平台提供的云监控，可以对弹性云服务器的运行状态进行日常监控。您可以通过管理控制台，直观地查看弹性云服务器的各项监控指标。

由于监控数据的获取与传输会花费一定时间，因此，云监控显示的是当前时间5~10分钟前的弹性云服务器状态。如果您的弹性云服务器刚刚创建完成，请等待5~10分钟后查看监控数据。

### 前提条件

- 弹性云服务器正常运行。  
关机、故障、删除状态的弹性云服务器，无法在云监控中查看其监控指标。当弹性云服务器再次启动或恢复后，即可正常查看。

#### 说明

- 关机、故障24小时的弹性云服务器，云监控将默认该弹性云服务器不存在，并在监控列表中删除，不再对其进行监控，但告警规则需要用户手动清理。
- 弹性云服务器已对接云监控，即已在云监控页面设置告警规则。  
对接云监控之前，用户无法查看到未对接资源的监控数据。具体操作，请参见[设置告警规则](#)。
- 弹性云服务器已正常运行一段时间（约10分钟）。  
对于新创建的弹性云服务器，需要等待一段时间，才能查看上报的监控数据和监控视图。

### 操作步骤

- 登录管理控制台。
- 选择“计算 > 弹性云服务器”。
- 在弹性云服务器列表中的右上角，输入弹性云服务器名称、IP地址或ID，并进行搜索。
- 单击弹性云服务器的名称，查看详情。
- 在弹性云服务器详情页面，选择“监控”页签，查看监控数据。
- 在弹性云服务器监控区域，您可以通过选择时长，查看对应时间的监控数据。  
当前支持查看弹性云服务器“近1小时”、“近3小时”、“近12小时”、“近24小时”、“近7天”的监控数据。

# 14 常见问题

## 14.1 产品咨询

### 14.1.1 弹性云服务器使用时有什么限制？

- 建议不要自行升级弹性云服务器的内核版本和操作系统版本，如需升级操作系统的主版本，例如从CentOS 7.2升级到Cent OS 7.3，请使用提供的切换操作系统功能。
- 建议不要卸载弹性云服务器上预装的性能辅助优化软件。
- 建议不要修改弹性云服务器网卡的MAC地址，修改后会导致网络不通。

### 14.1.2 弹性云服务器可以做什么？

弹性云服务器与传统硬件服务器一样，可以部署任意业务应用，例如：邮件系统、WEB系统、ERP系统等。

弹性云服务器创建成功后，您就可以像使用自己的本地PC或物理服务器一样，在云上使用弹性云服务器。

快速创建弹性云服务器步骤请参见[快速入门](#)。

## 14.2 ECS 创建

### 14.2.1 为什么创建弹性云服务器的任务失败，但是在列表中显示创建成功？

#### 问题描述

在弹性公网IP资源不足的情况下，用户通过管理控制台创建一台绑定弹性公网IP的弹性云服务器，此时弹性云服务器创建成功，但是绑定弹性公网IP失败。这种情况下，弹性云服务器申请状态栏显示任务失败，但弹性云服务器列表页面显示弹性云服务器创建成功。

## 根因分析

- 弹性云服务器列表中呈现的是创建成功的弹性云服务器列表和详情。
- 申请状态栏显示的是创建弹性云服务器任务的执行状态，包括创建弹性云服务器任务的各个子状态，例如创建弹性云服务器资源、绑定弹性公网IP等子任务，只有所有子任务全部成功时，任务才会成功，否则都是失败任务。

对于弹性云服务器资源创建成功，但是绑定弹性公网IP失败的情况而言，该任务处理失败。对于成功创建的弹性云服务器资源，会短时间内出现在弹性云服务器列表中，待系统完成回退操作后，将会在列表自动消失。

### 14.2.2 申请开通弹性云服务器需要多久？

通常几分钟之内即可开通成功。

具体的，与该弹性云服务器的规格、配置资源（如云硬盘、弹性公网IP），以及当前的系统负载有关。

#### 说明

如果长时间无法开通，请联系管理员寻求技术支持。

## 14.3 ECS 删除与退订

### 14.3.1 “删除”按钮是什么意思？

删除是指删除弹性云服务器，删除时可以选择同时删除弹性公网IP和数据盘。如果不选择，则会保留，需要您单独删除。

删除弹性云服务器的操作如下：

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 勾选待删除的弹性云服务器。
4. 单击弹性云服务器列表页上方的“删除”。

### 14.3.2 已删除的弹性云服务器可以再开通吗？

已删除是一种弹性云服务器的中间状态。说明弹性云服务器已被正常删除，在该状态下的弹性云服务器，不能对外提供业务，将会在短时间内从系统中彻底清除。

已删除的弹性云服务器在管理控制台的弹性云服务器列表中只会停留较短时间，系统稍后便会将其彻底清除，因此不可以用来恢复系统和数据。您可以再次创建同样规格的弹性云服务器。

### 14.3.3 重启/关机弹性云服务器时，长时间处于“正在重启”/“正在关机”状态，怎么办？

如果对弹性云服务器执行重启/关机操作，弹性云服务器长时间（大于30分钟）处于“正在重启”/“正在关机”状态时，建议执行强制重启/强制关机操作。

1. 登录管理控制台。

2. 选择“计算 > 弹性云服务器”。
3. 勾选待强制重启/强制关机的弹性云服务器，并单击“重启”/“关机”。系统弹窗提示用户是否要对弹性云服务器进行重启/关机操作。
4. 勾选“强制重启”/“强制关机”，确定强制重启/强制关机弹性云服务器。
5. 单击“确定”。

## 14.4 远程登录

### 14.4.1 登录前准备类

#### 14.4.1.1 启动弹性云服务器时卡在“Waiting for cloudResetPwdAgent”

##### 问题描述

启动弹性云服务器时，系统长时间卡在“Waiting for cloudResetPwdAgent”状态，需要等待20s~30s，如图14-1所示。

图 14-1 启动 cloudResetPwdAgent

```
Starting rpcbind: [ OK ]
Starting NFS statd: [ OK ]
Starting cloudResetPwdAgent...
Waiting for cloudResetPwdAgent.....
```

##### 可能原因

内网DNS和客户自定义DNS设置导致启动一键式重置密码插件时卡慢。

##### 处理方法

1. 以root用户登录弹性云服务器。
  2. 执行以下命令，修改配置文件“/etc/cloud/cloud.cfg”。
- ```
vi /etc/cloud/cloud.cfg
```
3. 添加如下语句，如图14-2所示。
- ```
manage_etc_hosts: true
```

图 14-2 编辑配置文件

```
users:
- name: root
  lock_passwd: False

disable_root: 0
ssh_pwauth: 1

datasource_list: ['OpenStack']
manage_etc_hosts: true

datasource:
  OpenStack:
    # timeout: the timeout value for a request at metadata service
    timeout : 50
    # The length in seconds to wait before giving up on the metadata
    # service. The actual total wait could be up to
    # len(resolvable_metadata_urls)*timeout
    max_wait : 120
```

## 14.4.2 远程连接类

### 14.4.2.1 MSTSC 方式登录 Windows 2012 的弹性云服务器，登录失败怎么办？

#### 问题描述

对于密码鉴权方式创建的Windows 2012弹性云服务器，使用初始密码以MSTSC方式登录时，登录失败，系统显示“第一次登录之前，您必须更改密码。请更新密码，或者与系统管理员或技术支持联系”。

#### 可能原因

用户本地使用的计算机（即客户机）操作系统为Windows 10。

由于Windows 10操作系统的自身限制，不能以初始密码直接远程连接操作系统为Windows 2012的弹性云服务器。

#### 处理方法

- 方法一  
更换使用Windows 7操作系统的计算机作为客户机，远程连接操作系统为Windows 2012的弹性云服务器。
- 方法二  
继续使用Windows 10 客户机远程登录，但是，需先修改弹性云服务器的初始密码。
  - a. 首次登录，以VNC方式登录操作系统为Windows 2012的弹性云服务器。
  - b. 登录成功后，按照界面提示修改弹性云服务器的密码。
  - c. 使用修改后的密码，以MSTSC方式远程登录。
- 方法三  
继续使用Windows 10 客户机，并以初始密码远程登录。
  - a. 单击“开始”菜单，在“搜索程序和文件”中，输入“mstsc”。  
系统进入“远程桌面连接”界面。
  - b. 依次输入弹性公网IP、用户名“administrator”以及创建弹性云服务器时设置的登录密码进行连接。  
连接失败，系统提示“第一次登录之前，您必须更改密码。请更新密码，或者与系统管理员或技术支持联系”。
  - c. 单击“远程桌面连接”页面左下角的“选项”。
  - d. 在“常规”页签，单击“连接设置”栏的“另存为”，保存“.rdp”格式的远程桌面文件。
  - e. 打开d中保存的远程桌面文件。
  - f. 编辑远程桌面文件，在文件的最后一行，添加如下语句并保存。  
**enablecredsspssupport:i:0**
  - g. 双击更新后的“.rdp”文件，打开远程桌面连接。
  - h. 单击“连接”，重新连接操作系统为Windows 2012的弹性云服务器。

## 14.4.2.2 怎样修改远程登录的端口？

### 操作场景

本节操作介绍修改远程桌面端口的操作步骤。

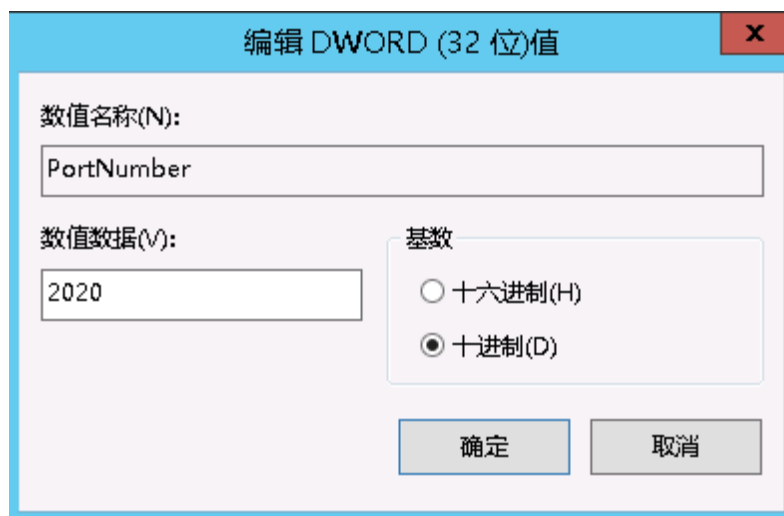
### Windows 操作系统

以下操作以操作Windows 2012操作系统为例。Windows云服务器默认登录端口为3389，以修改为2020端口为例。

1. 修改安全组规则。
  - a. 登录管理控制台。
  - b. 选择“计算 > 弹性云服务器”
  - c. 单击弹性云服务器名称，在弹性云服务器详情页面，选择“安全组”。
  - d. 在“安全组”页签，单击“配置规则”，进入安全组详情界面。
  - e. 在“入方向规则”页签，单击“添加规则”，添加如下规则。
    - 协议：基本协议/自定义TCP
    - 端口：2020

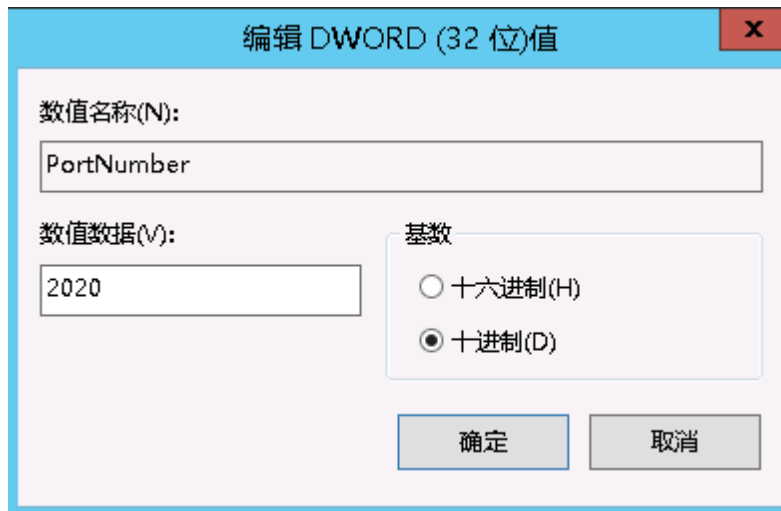
具体操作，请参见《虚拟私有云用户指南》的“添加安全组规则”章节。
2. 登录弹性云服务器。
3. 打开“运行”对话框，输入“regedit”进入注册表编辑器。
4. 打开注册表，依次选择HKEY\_LOCAL\_MACHINE > SYSTEM > CurrentControlSet > Control > Terminal Server > Wds > rdpwd > Tds > tcp，并双击PortNumber键值。
  - a. 将对话框中的“基数”栏选择为“十进制”。
  - b. 修改数值数据为新的端口，本例为2020。

图 14-3 修改为 2020 端口



5. 依次打开注册表分支 HKEY\_LOCAL\_MACHINE > SYSTEM > CurrentControlSet > Control > Terminal Server > WinStations > RDP-Tcp，并双击PortNumber键值。

- a. 将对话框中的“基数”栏选择为“十进制”。
- b. 修改数值数据为新的端口，本例为2020。

**图 14-4** 修改为 2020 端口

6. 修改防火墙入站规则（如果防火墙关闭可略过此步骤）。  
依次打开“控制面板 > Windows防火墙 > 高级设置 > 入站规则 > 新建入站规则”。
  - 规则类型选择：端口。
  - 协议选择：TCP。
  - 端口选择：特定本地端口，并填写新的端口号，本例为：2020。
  - 操作选择：允许连接。
  - 配置文件：使用默认配置无需修改；
  - 名称：RDP-2020
- 完成后，单击刷新可以查看到新添加的规则。
7. 打开Windows搜索框，输入services，选择“服务”。

**图 14-5** 选择服务

8. 在服务中选择并重启Remote Desktop Services 服务（或者重启服务器）。
9. 通过“IP地址:端口”远程访问云服务器。

## Linux 操作系统

以下操作以CentOS 7.3操作系统为例。Linux云服务器默认登录端口为22，以修改为2020端口为例。

1. 修改安全组规则。



- a. 登录管理控制台。
- b. 选择“计算 > 弹性云服务器”。
- c. 单击弹性云服务器名称，在弹性云服务器详情页面，选择“安全组”。
- d. 在“安全组”页签，单击“配置规则”，进入安全组详情界面。
- e. 在“入方向规则”页签，单击“添加规则”，添加如下规则。

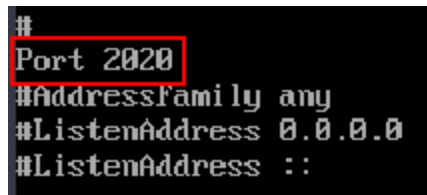
- 协议：基本协议/自定义TCP

- 端口：2020

具体操作，请参见《虚拟私有云用户指南》的“添加安全组规则”章节。

2. 登录弹性云服务器。
3. 执行以下命令编辑sshd服务的配置文件。  
**vi /etc/ssh/sshd\_config**
4. 将#port 22 这一行的井号注释符去掉，修改22为2020。

图 14-6 修改为 2020 端口



```
#  
Port 2020  
#AddressFamily any  
#ListenAddress 0.0.0.0  
#ListenAddress ::
```

5. 按“ESC”退出编辑模式，输入“:wq!”保存退出。
6. 执行以下命令重启sshd服务。  
**service sshd restart**  
或  
**systemctl restart sshd**
7. （可选）配置防火墙。如防火墙关闭，可以忽略防火墙的配置操作。  
CentOS 7的防火墙是firewalld，CentOS 6版本用的iptables，两者使用上有些差别，以下操作以CentOS 7为例。  
执行如下命令查看防火墙状态：**firewall-cmd --state**
  - 方法一（推荐）：在firewalld服务中添加新的端口。
    - i. 执行以下命令，添加2020端口规则。  
**firewall-cmd --zone=public --add-port=2020/tcp --permanent**  
**firewall-cmd --reload**
    - ii. 查看添加的端口。在ports中已添加端口2020 tcp连接。  
**firewall-cmd --list-all**
    - iii. 修改完成后重启防火墙。  
**systemctl restart firewalld.service**
  - 方法二（不推荐）：关闭防火墙并取消开机自启。  
**systemctl stop firewalld**  
**systemctl disable firewalld**
8. 通过如下命令，验证端口是否可以成功访问。

### telnet 弹性公网IP 端口

例如：telnet xx.xx.xx.xx 2020

#### 14.4.2.3 远程登录弹性云服务器时，对浏览器版本的要求？

用户采用远程登录方式访问弹性云服务器时，使用的浏览器应满足表14-1。

表 14-1 支持的浏览器版本

浏览器	版本
Google Chrome	31.0-75.0
Mozilla FireFox	27.0-62.0
Internet Explorer	10.0-11.0

### 14.4.3 VNC 登录类

#### 14.4.3.1 VNC 方式登录弹性云服务器后，较长时间不操作，界面无响应？

用户的计算机操作系统为Windows7时，如果使用IE10或IE11浏览器通过VNC方式登录弹性云服务器，且较长时间不做任何操作，VNC登录界面会出现无响应的状况，键盘和鼠标无法正常操作。

此时，可以通过单击两次VNC页面上的“AltGr”按钮重新激活页面。

如问题未解决，请联系技术支持。

#### 14.4.3.2 VNC 方式登录弹性云服务器后，查看数据失败，VNC 无法正常使用？

使用VNC方式登录弹性云服务器后，执行查看数据操作时，例如在Linux操作系统下使用cat命令查看大文件、播放视频，由于部分浏览器自身内存占用太大，导致VNC无法正常使用。

此时，请更换其他浏览器后重新登录弹性云服务器。

如问题未解决，请联系技术支持。

#### 14.4.3.3 VNC 方式登录弹性云服务器时，系统黑屏输入无反应？

这是由于该弹性云服务器已被其他用户使用VNC方式登录。

VNC方式登录弹性云服务器时，同一时刻仅支持一个用户登录。如果多个用户同时登录，先登录的用户会断开连接，出现黑屏现象。

此时，重新进行登录操作可消除黑屏，但会导致其他用户黑屏并断开连接。

### 14.4.4 Windows 远程登录报错类

### 14.4.4.1 远程连接 Windows 云服务器报错：出现身份验证错误，要求的函数不受支持

#### 问题描述

Windows操作系统的本地PC，通过RDP协议（如MSTSC方式）远程桌面连接Windows云服务器报错，报错显示出现身份验证错误，要求的函数不受支持。

- 如果报错信息中仅提示：出现身份验证错误，要求的函数不受支持。请参考[处理方法](#)解决该问题。
- 如果报错信息中附加提示了由于CredSSP加密Oracle修正导致该错误，如[图14-7](#)所示。可能原因是由于微软于2018年3月发布安全补丁，该安全补丁可能会影响RDP连接的CredSSP，导致通过RDP协议访问云服务器时连接失败。解决方法请参考微软官方指导文档“CredSSP updates for CVE-2018-0886”。

图 14-7 远程桌面连接失败



#### 处理方法

修改Windows云服务器的远程桌面连接设置：将选项“仅允许运行使用网络级别身份验证的远程桌面的计算机连接（更安全）”修改为“允许运行任意版本远程桌面的计算机连接（较不安全）”。具体操作如下：

1. 登录云服务器。
2. 打开左下角的“开始”菜单，右键单击“计算机”，并选择“属性”。
3. 在左侧导航栏，选择“远程设置”。
4. 选择“远程”页签，并在“远程桌面”栏，选择“允许远程连接到此计算机”。

图 14-8 远程设置



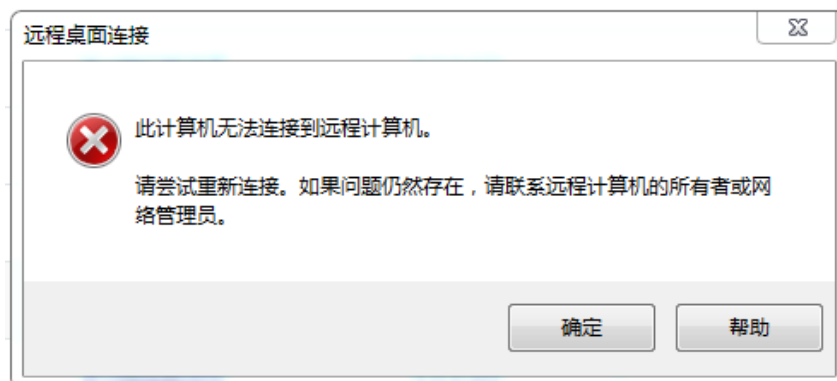
5. 单击“确定”。

#### 14.4.4.2 远程连接 Windows 云服务器报错：此计算机无法连接到远程计算机

##### 问题描述

使用远程登录方式连接登录Windows云服务器时出现如下错误：此计算机无法连接到远程计算机。

图 14-9 无法连接到远程计算机



## 可能原因

- 服务端安全组3389端口未开启。[检查云服务器端口配置](#)
- 服务端防火墙关闭。[检查防火墙配置是否正常](#)
- 远程桌面连接配置不正确。[检查远程桌面连接设置](#)
- 未开启“Remote Desktop Services”。[检查Remote Desktop Services](#)
- 远程桌面会话主机配置不正确。[检查远程桌面会话主机](#)

## 检查云服务器端口配置

检查云服务器的3389端口是否能够访问（默认使用3389端口）。

请确保入方向规则中已添加3389端口。

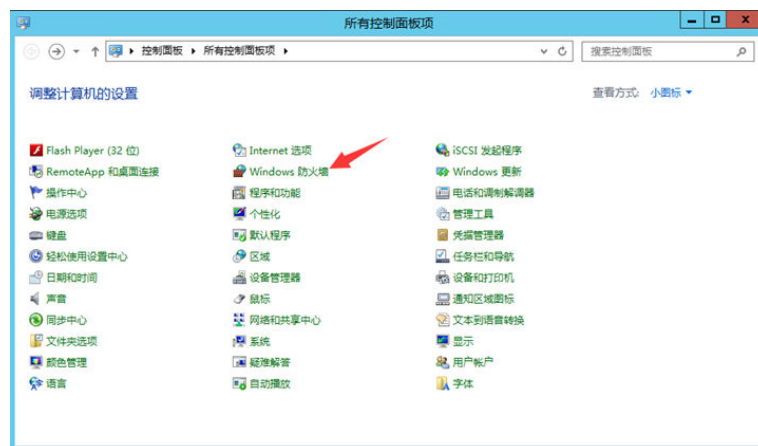
在云服务器的详情页面选择“安全组”页签，查看安全组入方向规则中已添加3389端口。

## 检查防火墙配置是否正常

检查云服务器的防火墙是否开启。

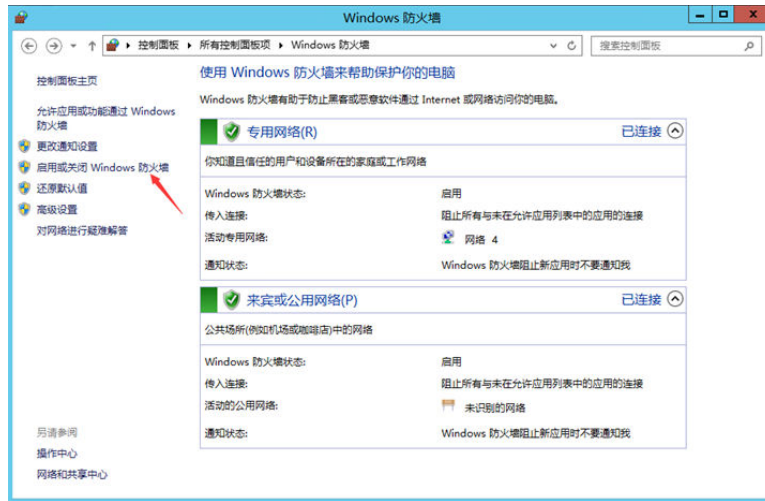
1. 在控制台使用VNC方式登录Windows云服务器。
2. 单击桌面左下角的Windows图标，选择“控制面板 > Windows防火墙”。

图 14-10 Windows 防火墙



3. 单击“启用或关闭Windows防火墙”。  
查看并设置防火墙的具体状态：开启或关闭。

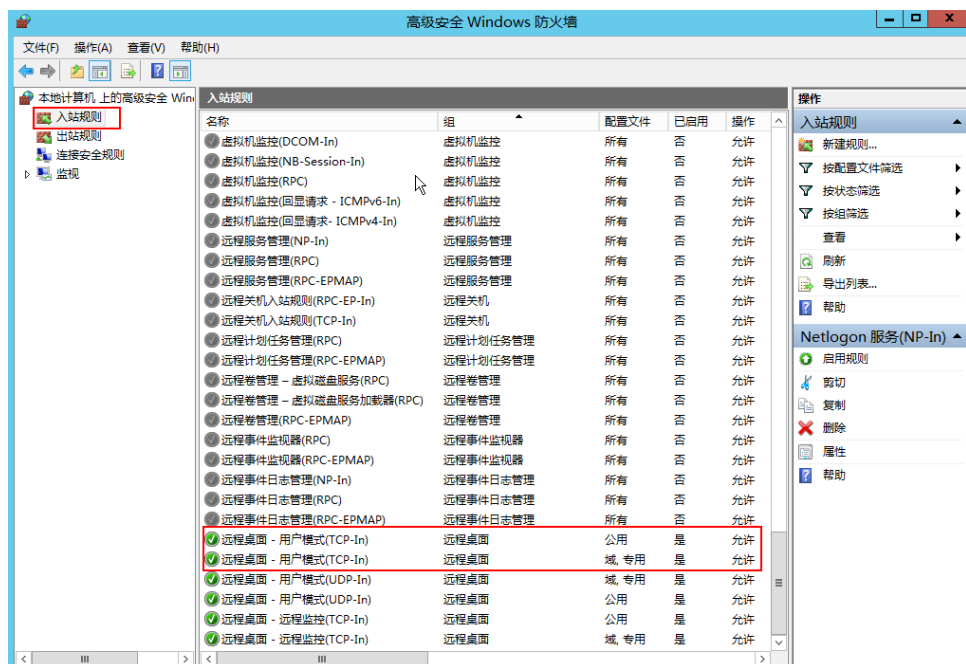
图 14-11 查看防火墙状态



如果选择开启防火墙，请继续执行以下操作步骤。

4. 单击“高级配置”。
5. 检查“入站规则”，请确保已启用如下规则。
  - 远程桌面(TCP-In) 公用
  - 远程桌面(TCP-In) 域，专用

图 14-12 入站规则



如果防火墙入站规则中设置的端口与远程服务器设置的端口不一致，远程访问服务器将无法成功。一旦出现这种情况，您也可以添加新的防火墙入站规则端口。

### 说明

默认使用的是3389端口，如果您使用的是其他端口，可参考3389端口添加防火墙入站规则。

完成上述操作后，再次重试远程连接云服务器。

## 检查远程桌面连接设置

修改Windows云服务器的远程桌面连接设置：勾选“允许远程连接到此计算机”。具体操作如下：

1. 登录云服务器。
2. 打开左下角的“开始”菜单，右键单击“计算机”，并选择“属性”。
3. 在左侧导航栏，选择“远程设置”。
4. 选择“远程”页签，并在“远程桌面”栏，选择“允许远程连接到此计算机”。

图 14-13 远程设置



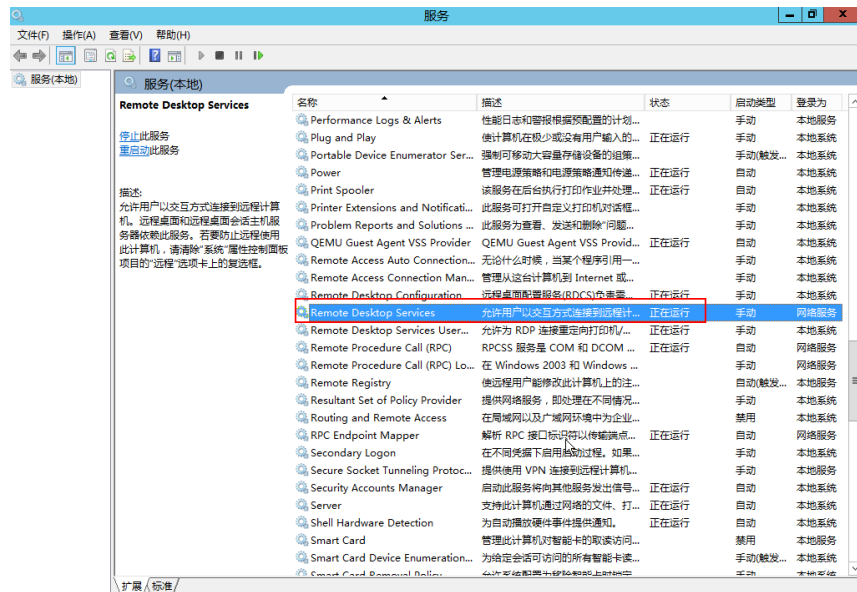
5. 单击“确定”。

## 检查 Remote Desktop Services

1. 打开Windows搜索框，输入services，选择“服务”。
2. 在服务中选择并重启Remote Desktop Services服务，请确保Remote Desktop Services状态为“正在运行”。



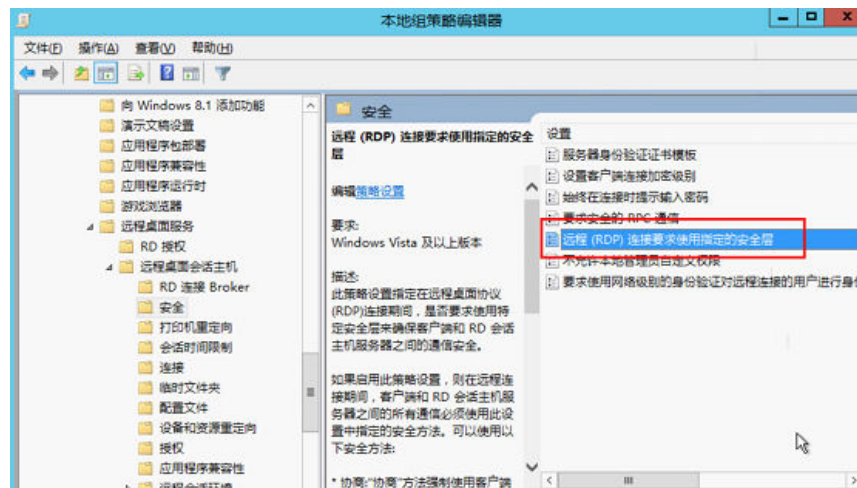
图 14-14 Remote Desktop Services



## 检查远程桌面会话主机

1. 打开cmd运行窗口，并输入“gpedit.msc”。
2. 单击“确定”，打开“本地组策略编辑器”。
3. 选择“计算机配置 > 管理模板 > Windows组件”，查找并双击“远程桌面服务”。
4. 选择“远程桌面会话主机 > 安全 > 远程（RDP）连接要求使用指定的安全层”。

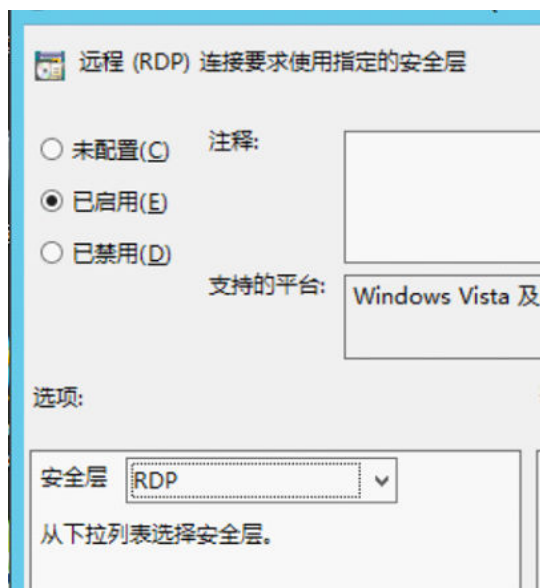
图 14-15 远程（RDP）连接要求使用指定的安全层



5. 选择“已启用”，并将“安全层”设置为“RDP”。



图 14-16 设置安全层为 RDP

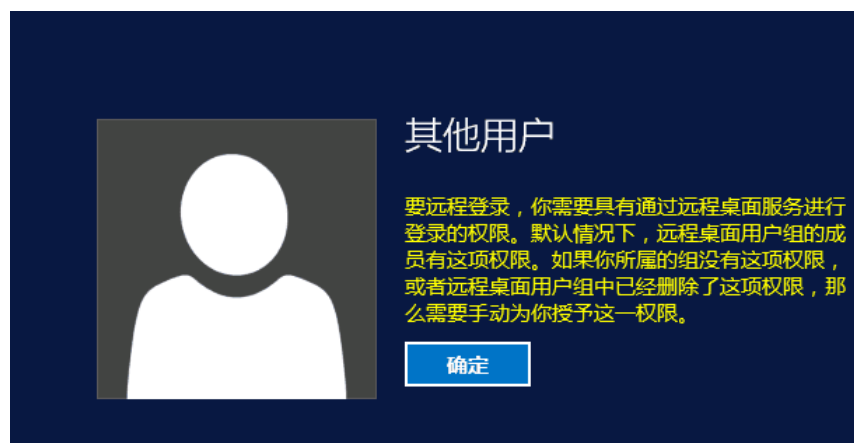


#### 14.4.4.3 远程连接 Windows 云服务器报错：没有远程登录的权限

##### 问题描述

远程桌面连接时提示需要具有通过远程桌面服务进行登录的权限。

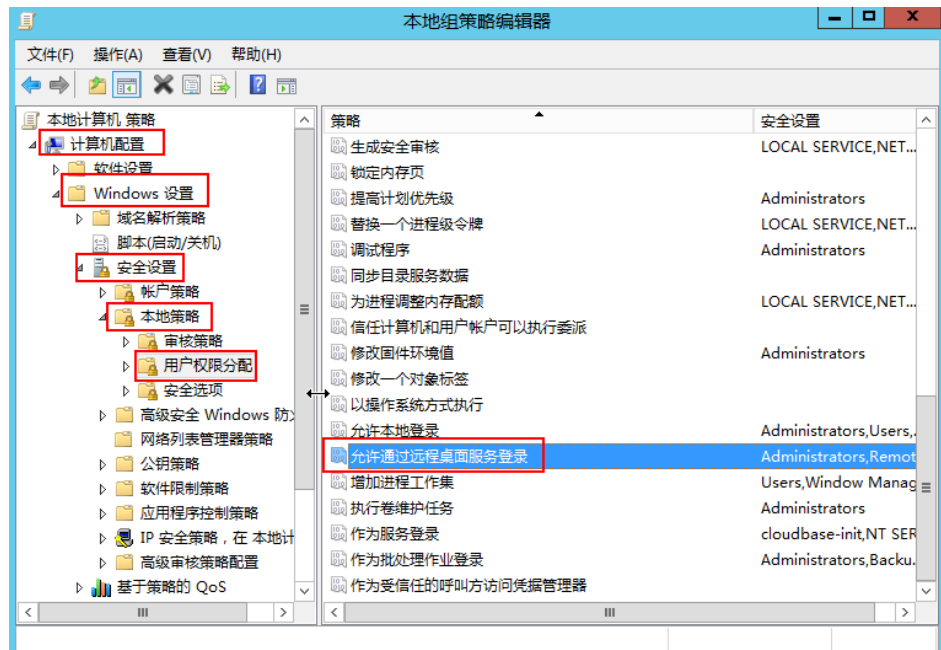
图 14-17 缺失远程登录权限



##### 处理方法

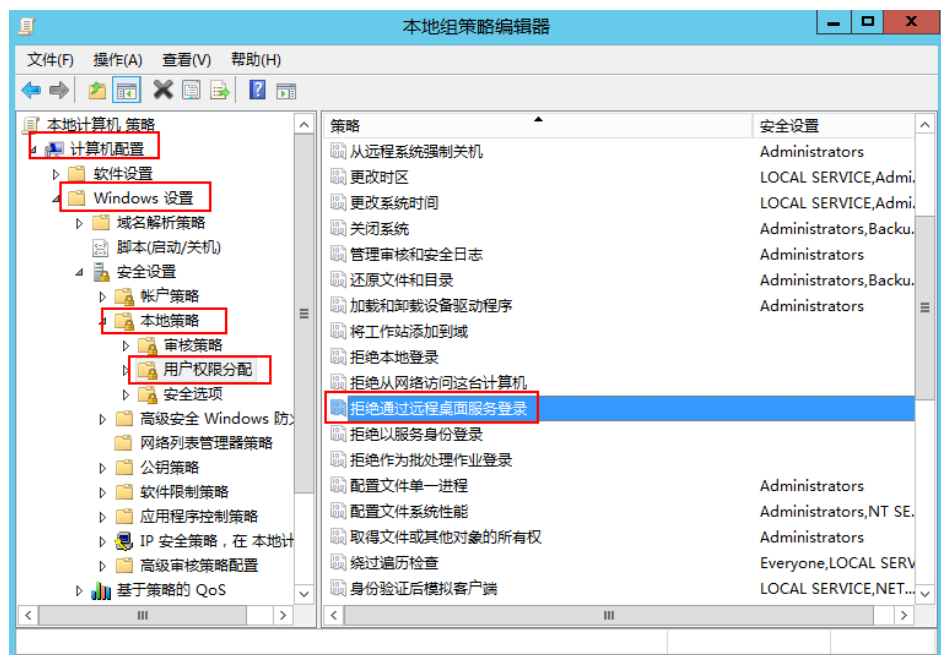
1. 打开cmd运行窗口，并输入“gpedit.msc”。
2. 单击“确定”，打开“本地组策略编辑器”。
3. 选择“计算机配置 > Windows设置 > 安全设置 > 本地策略 > 用户权限分配”。
  - a. 查找并双击“允许通过远程桌面服务登录”。确保已添加“Administrators”和“Remote desktop users”。

图 14-18 允许通过远程桌面服务登录



- b. 查找并双击“拒绝通过远程桌面服务登录”。如果里面有Administrator账号，请删除。

图 14-19 拒绝通过远程桌面服务登录



#### 14.4.4.4 远程连接 Windows 云服务器报错：没有远程桌面授权服务器可以提供许可证

##### 问题描述

使用远程登录方式连接登录Windows云服务器时出现如下错误：由于没有远程桌面授权服务器可以提供许可证，远程会话被中断，请跟服务器管理员联系。

图 14-20 没有远程桌面授权服务器可以提供许可证



## 可能原因

可能原因是由于在系统内部安装了远程桌面会话主机角色。

您可以申请免费使用 120 天，之后需要付费，如未付费则会造成远程连接失败。Windows 最多仅允许两个用户连接（包含本地登录用户），若要支持更多连接数量，请安装远程桌面服务角色并取得合法授权数量。在配置远程桌面会话主机角色后，会同时取消原有默认的 2 个免费连接授权。所以，在没有正确配置相关授权的时候，会导致远程桌面无法连接，并出现上述错误提示。

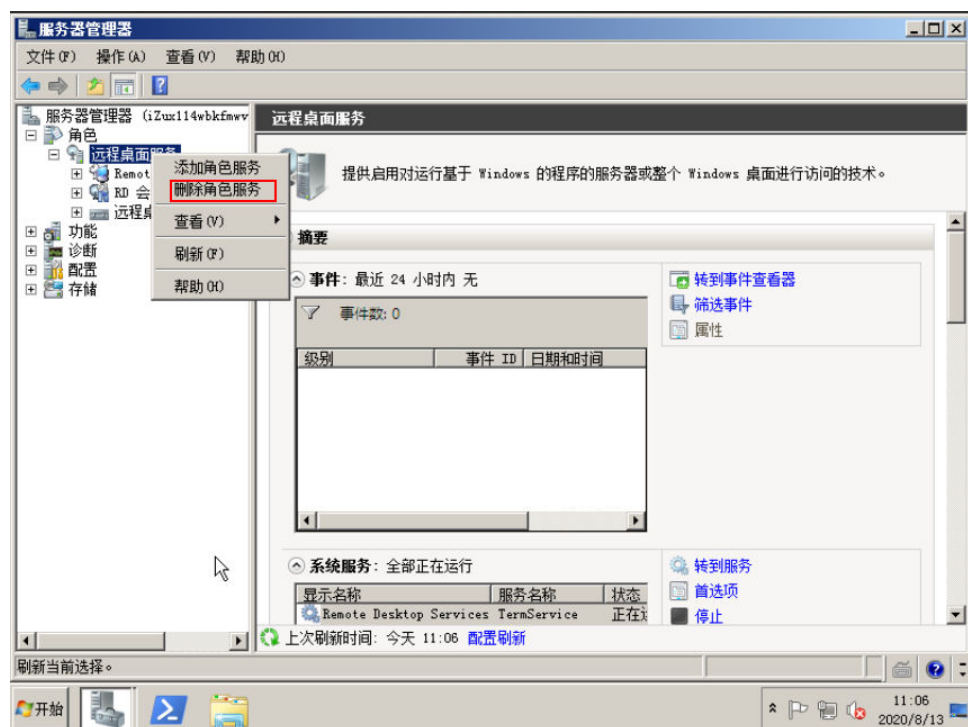
## 操作须知

- 本文适用于 Windows server 2008/2012 操作系统的云服务器。
- 操作过程中需要重启云服务器，可能会造成业务中断，请提前做好数据备份。

## Windows 2008 操作系统处理方法

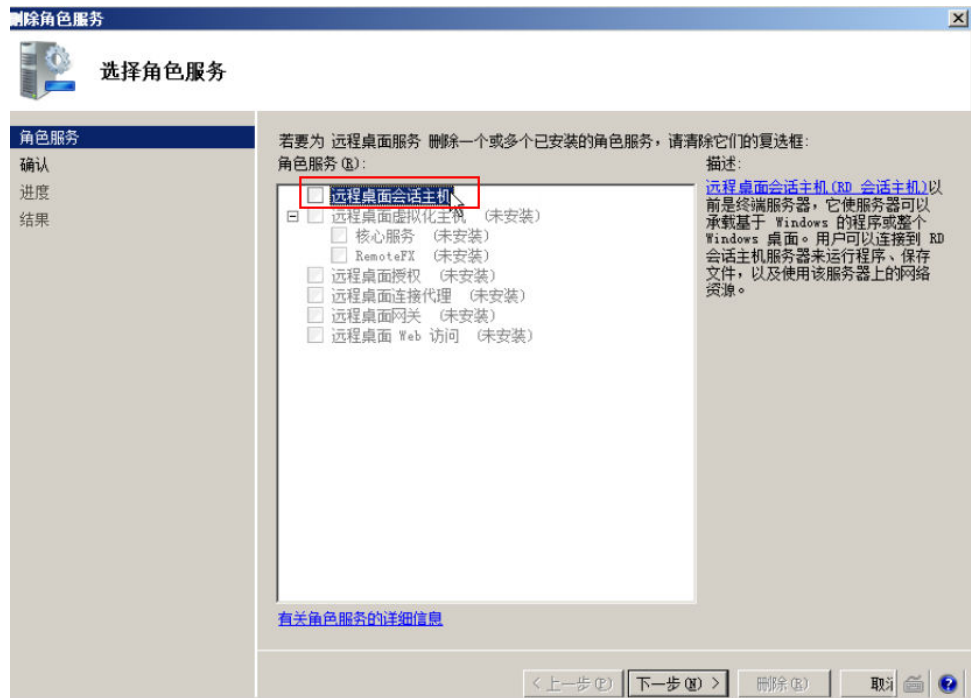
1. 使用控制台提供的 VNC 连接方式登录到 Windows 云服务器。
2. 打开“服务器管理器”，依次选择“角色 > 远程桌面服务”，右键单击“远程桌面服务”，然后选择“删除角色服务”。

图 14-21 删除角色服务



3. 在弹出窗口中，取消勾选“远程桌面会话主机”，并单击“下一步”直到完成。

图 14-22 取消勾选“远程桌面会话主机”



4. 单击“删除”。
5. 重启云服务器。

## Windows 2012 操作系统处理方法

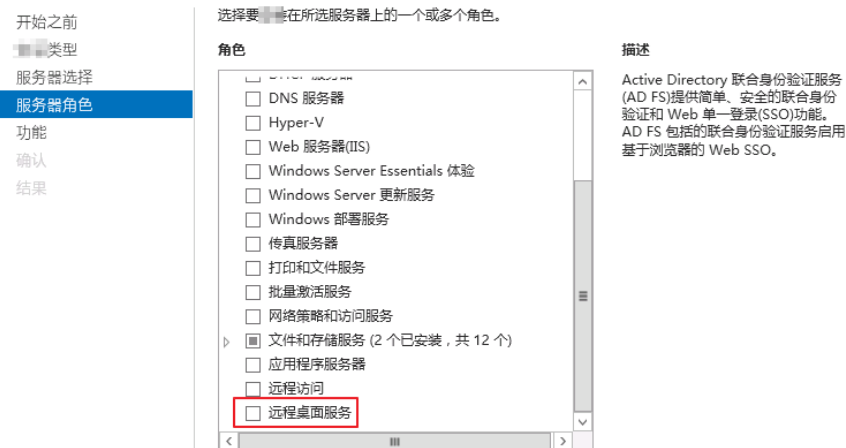
1. 使用控制台提供的VNC连接方式登录到Windows云服务器。
2. 打开“服务器管理器”，并依次选择“管理 > 删除角色和功能”，单击“下一步”。

图 14-23 删除角色和功能



3. 选择目标服务器，单击“下一步”。
4. 取消勾选“远程桌面服务”。

图 14-24 远程桌面服务



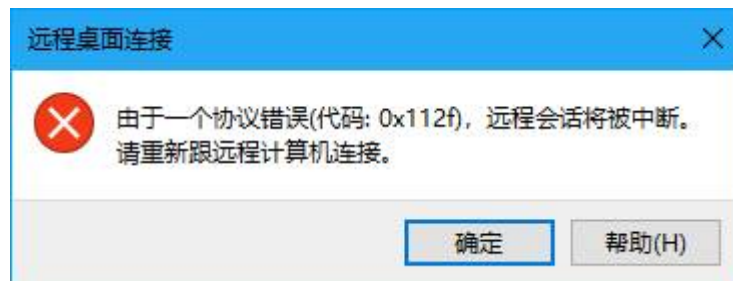
5. 单击“删除”。
6. 重启云服务器。

#### 14.4.4.5 登录 Windows 云服务器时报错：0x112f

##### 问题描述

登录Windows云服务器时，系统报错，错误代码为“0x112f”，如图14-25所示。

图 14-25 协议错误（代码 0x112f）



##### 可能原因

云服务器内存不足。

##### 处理方法

- 方法一（推荐）：  
变更规格，升级云服务器的CPU、内存大小。变更规格的方法，请参见[变更单台ECS规格](#)。
- 方法二：  
开启虚拟内存，获取云服务器的空闲内存。  
开启虚拟内存的方法，请参见[怎样配置Windows弹性云服务器的虚拟内存？](#)。

##### 说明

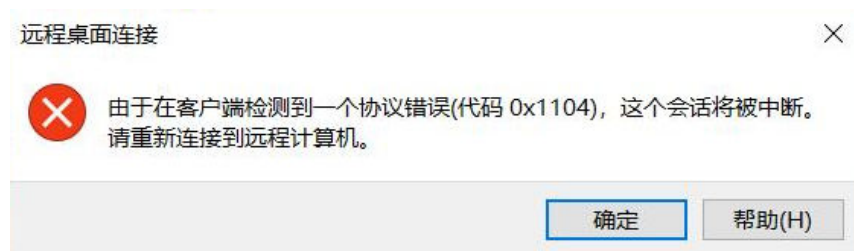
该方法会导致磁盘I/O性能下降，如非必要，不推荐使用。

### 14.4.4.6 远程连接 Windows 云服务器报错：0x1104

#### 问题描述

使用MSTSC方式登录Windows Server 2008操作系统的云服务器，系统报错：检测到一个协议错误（代码 0x1104）。

图 14-26 协议错误（代码 0x1104）



#### 可能原因

- 服务端安全组3389端口未开启。
- 服务端防火墙关闭。
- 服务端3389端口被其他进程占用。
- 远程桌面会话主机配置不正确。

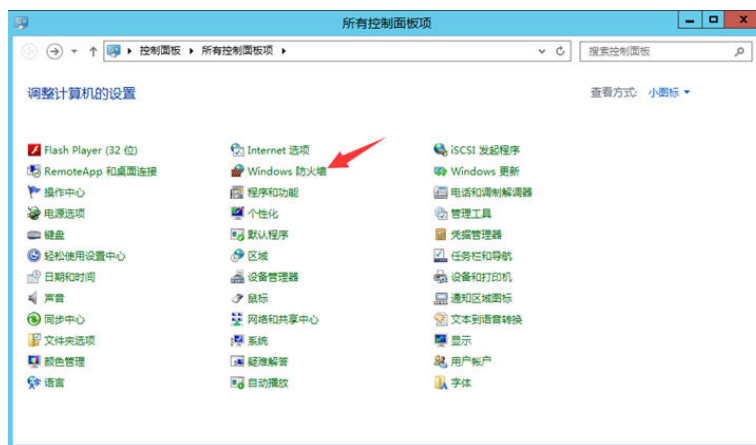
#### 处理方法

**步骤1** 检查安全组设置。

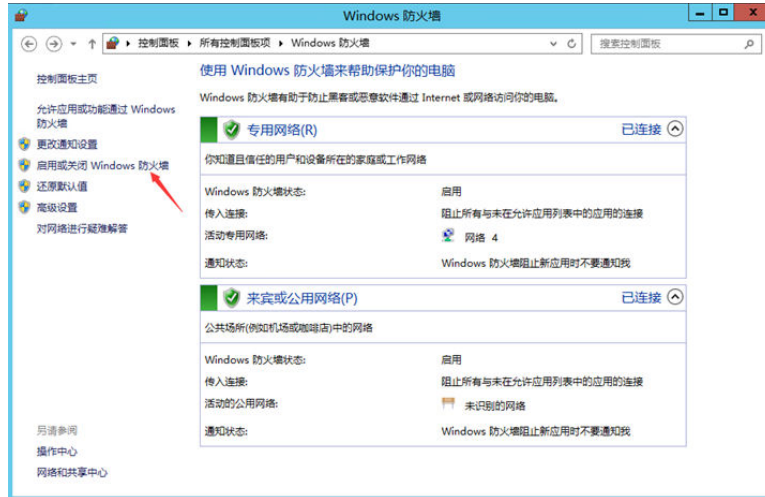
检查3389端口入方向是否开启，若已开启，执行**步骤2**。

**步骤2** 检查防火墙是否关闭。

1. 登录Windows云服务器。
2. 单击桌面左下角的Windows图标，选择“控制面板 > Windows防火墙”。



3. 单击“启用或关闭Windows防火墙”。  
查看并设置防火墙的具体状态：开启或关闭。



如果正常，请执行**步骤3**。

**步骤3** 使用VNC方式登录云服务器，查看端口信息。

1. 进入cmd命令窗，并执行以下命令。  
**netstat -ano |findstr : 3389**

图 14-27 检查 3389 端口



如图14-27所示，3389端口被占用，进程PID为4。

2. 打开任务管理器，查看PID为4的进程为System系统进程。
3. 通常IIS服务和SQL Server会以System进程运行，执行以下HTTP命令进一步查看。

**netsh http show servicestate**



图 14-28 查看系统进程

```
URL 组:  
URL 组 ID: FE00000140000001  
状态: Active  
请求队列名称: DefaultAppPool  
属性:  
    最大带宽: 已继承  
    最大连接: 4294967295  
    超时:  
        实体正文超时(秒): 120  
        耗尽实体正文超时(秒): 120  
        请求队列超时(秒): 65535  
        空闲连接超时(秒): 120  
        标头等待超时(秒): 0  
        最小发送速率(字节/秒): 0  
    日志记录信息:  
        日志目录: C:\inetpub\logs\LogFiles\W3SVC1  
        日志格式: 0  
    身份验证配置:  
        启用的身份验证方案:  
        已注册的 URL 数目: 3  
        已注册的 URL 数目:  
        HTTP://*:80/  
        HTTP://*:3389:19  
        HTTP://*:226:3389:
```

4. 如果可以看到3389端口被HTTP协议使用，可以确定是被IIS服务占用。
5. 在浏览器输入“http://127.0.0.1:3389”进行验证，网站打开正常。
6. 修改IIS使用其他端口，重启IIS服务。

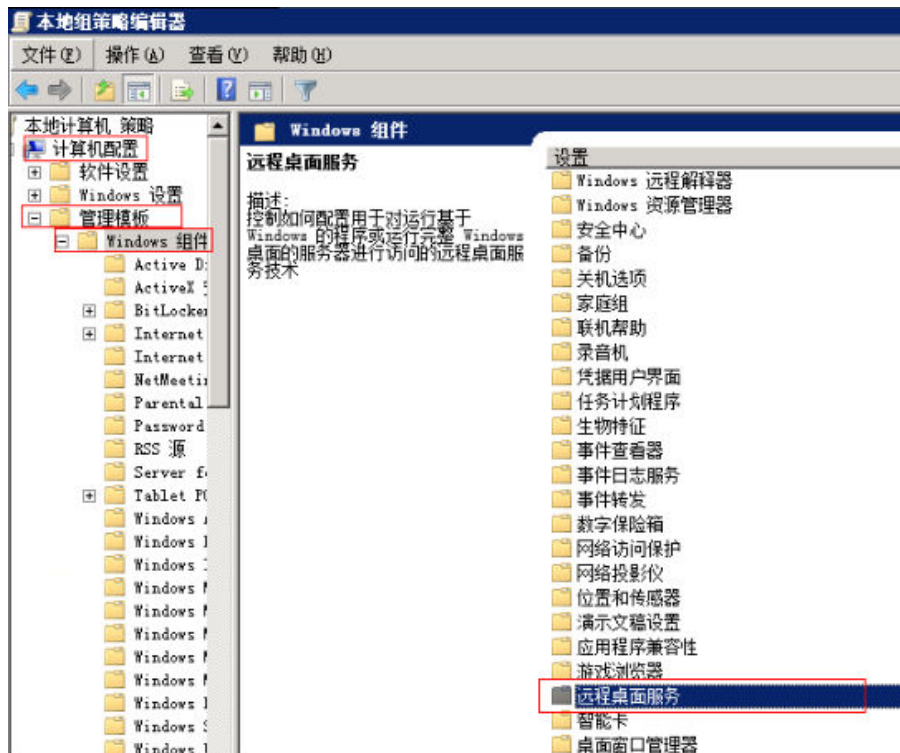
**步骤4** 如果以上检查均没有问题，请执行**步骤5**，检查是否由于远程桌面会话主机配置导致。

**步骤5** 检查远程桌面会话主机配置。

1. 通过VNC登录云服务器。
2. 打开cmd运行窗口，并输入“gpedit.msc”。
3. 单击“确定”，打开“本地组策略编辑器”。
4. 选择“计算机配置 > 管理模板 > Windows组件”，查找并双击“远程桌面服务”。

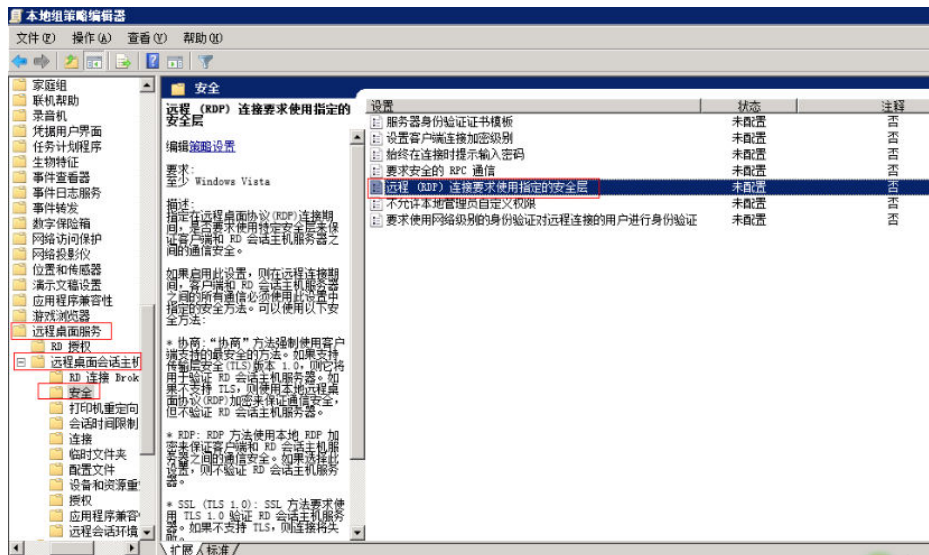


图 14-29 远程桌面服务



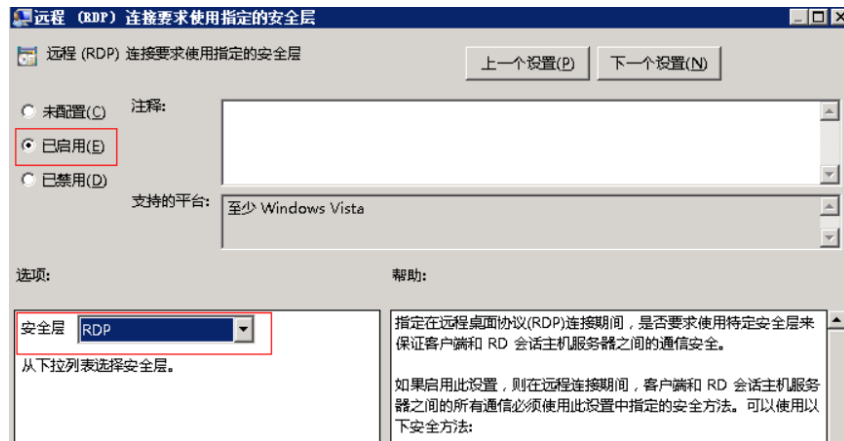
- 5. 选择“远程桌面会话主机 > 安全 > 远程（RDP）连接要求使用指定的安全层”。

图 14-30 远程（RDP）连接要求使用指定的安全层



- 6. 选择“已启用”，并将“安全层”设置为“RDP”。

图 14-31 设置安全层



7. 单击“确定”。
8. 配置完成后，打开“cmd”命令窗。
9. 执行以下命令，刷新组策略。

**gpupdate**

图 14-32 刷新组策略



----结束

#### 14.4.4.7 远程连接 Windows 云服务器报错：122.112...

##### 问题描述

本地使用远程桌面连接登录Windows server 2012云服务器，报错：122.112...，服务器频繁掉线，Windows登录进程意外中断。

##### 可能原因

1. 系统资源不足或不可用。
2. 服务启动失败。

##### 处理方法

**步骤1** 查看系统日志。

1. 通过VNC方式登录云服务器。


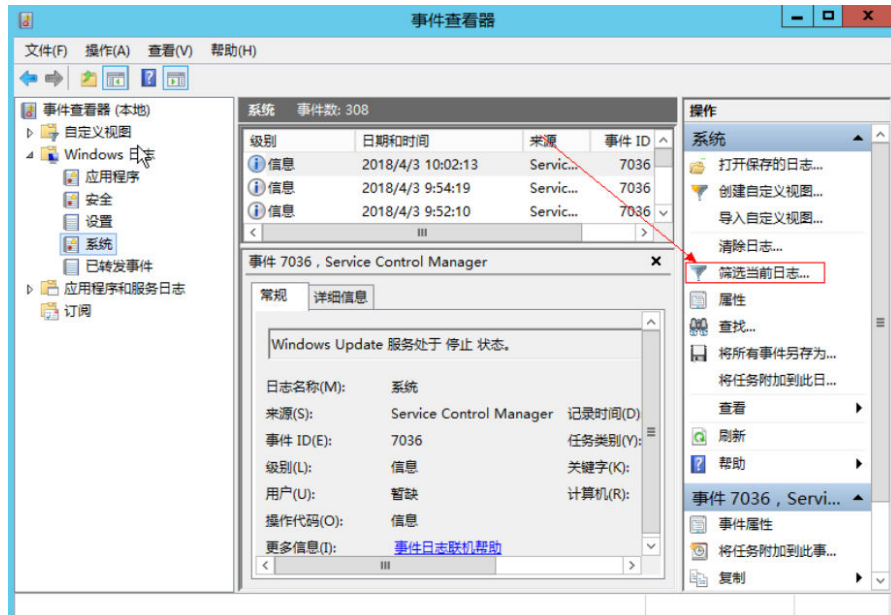
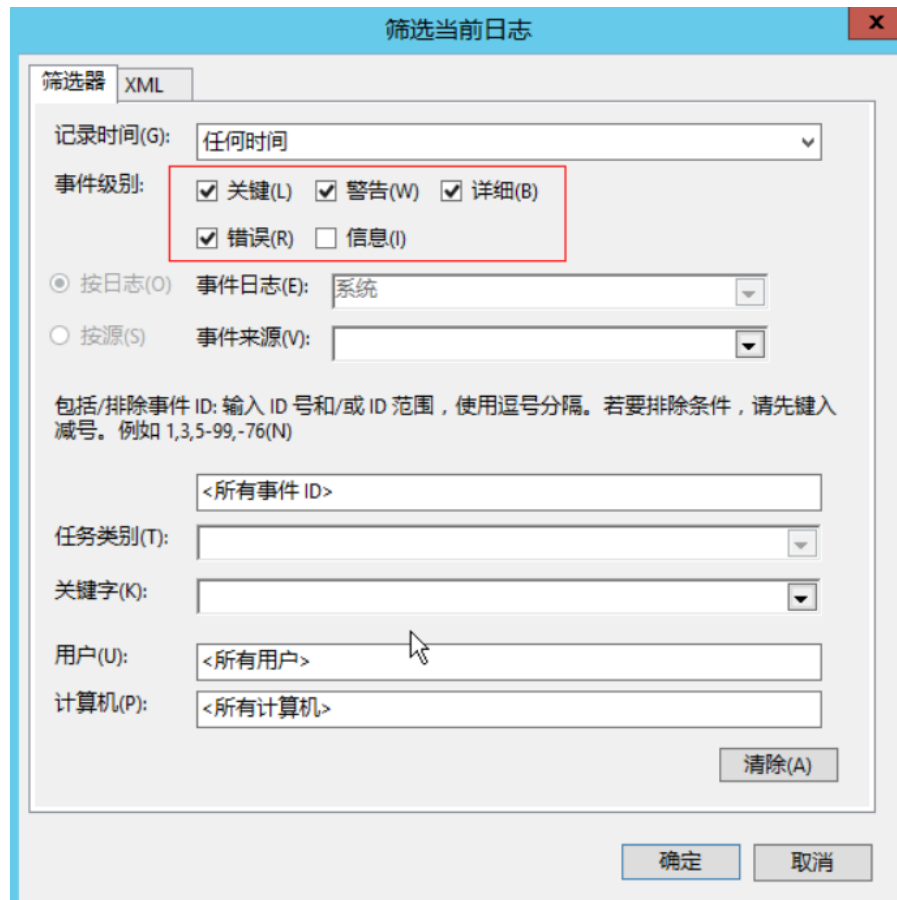
2. 单击  打开服务管理，选择“管理工具 > 事件查看器 > Windows日志 > 系统 > 筛选当前日志”。

图 14-33 事件查看器



3. 在“事件级别”栏，勾选“关键”、“警告”、“详细”、“错误”。

图 14-34 筛选日志



4. 查找登录相关的日志信息。

**步骤2** 查看主机系统资源使用情况。

1. 选择“开始 > 任务管理器 > 性能”。
2. 查看CPU、内存使用情况。

**步骤3** 检查购买的Windows云服务器规格是否为1U1GB。

如果是，建议变更规格或者停止不使用的进程。

----结束

#### 14.4.4.8 使用 Mac 远程连接 Windows 云服务器报错：证书或相关链无效

##### 问题描述

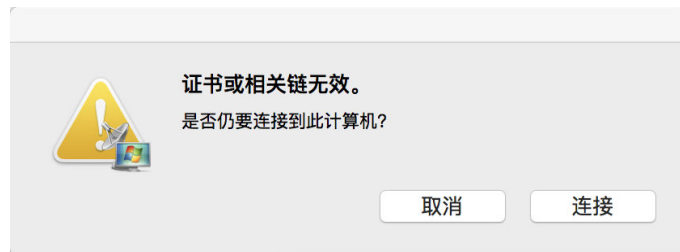
使用Mac版Microsoft Remote Desktop工具，远程连接Windows云服务器。

图 14-35 Mac 版 Microsoft Remote Desktop 工具



由于Mac系统的特殊性，在使用Mac系统远程登录Windows云服务器时，需要在Mac端和Windows云服务器内部执行相关配置，才能远程连接成功。使用Mac远程连接时，出现报错“证书或相关链无效”。

图 14-36 证书或相关链无效



## 可能原因

云服务器策略组设置的问题。

## 操作步骤

1. 在本地主机左上角的菜单栏选择“RDC > 首选项”打开Microsoft Remote Desktop的偏好设置。

图 14-37 选择首选项



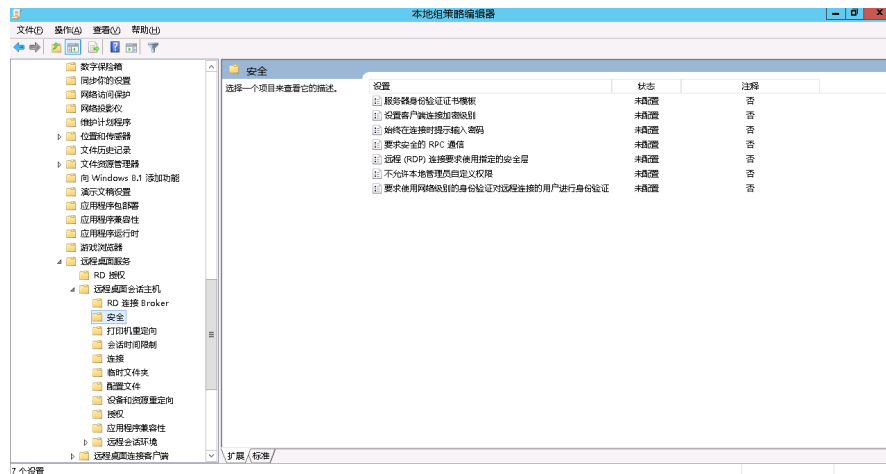
2. 选择“安全性”，并修改参数配置。

图 14-38 选择安全性



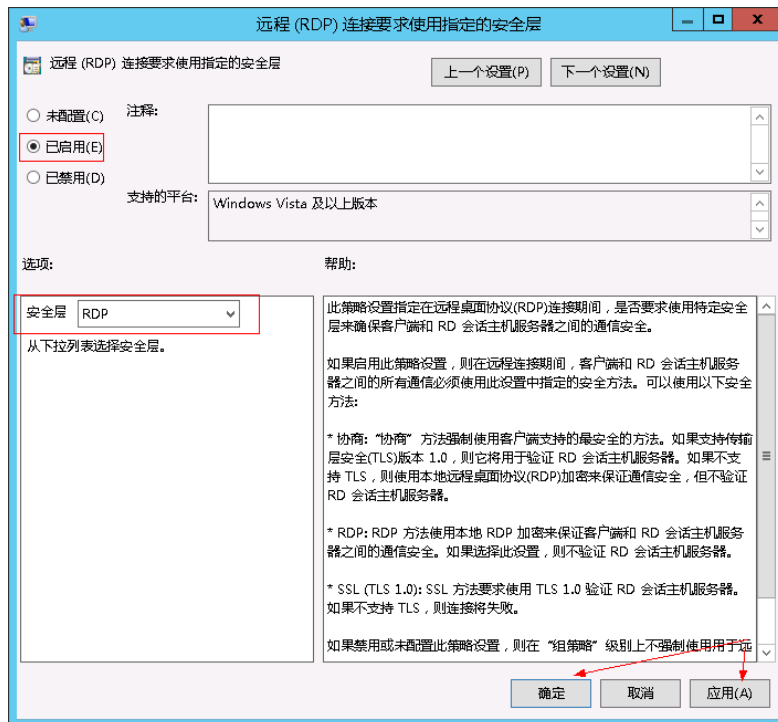
3. 再次远程连接Windows云服务器，如果仍出现报错“证书或相关链无效”，请执行4。
4. 使用控制台提供的远程连接功能（VNC方式）登录Windows云服务器。
5. 按快捷键“Win+R”打开“运行”窗口。
6. 输入“gpedit.msc”，进入“本地组策略编辑器”。
7. 在左侧导航栏，选择“计算机配置 > 管理模板 > Windows组件 > 远程桌面服务 > 远程桌面会话主机 > 安全”。

图 14-39 远程桌面会话主机



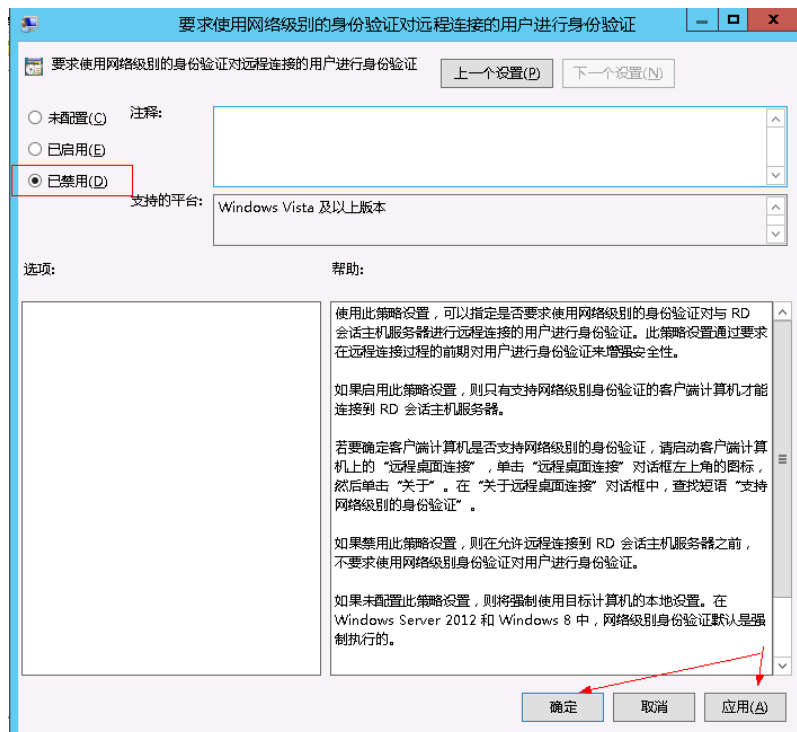
8. 根据界面提示，修改如下两项配置：
  - 启用“远程（RDP）连接要求使用指定的安全层”。

图 14-40 远程 (RDP) 连接要求使用指定的安全层



- 禁用“要求使用网络级别的身份验证对远程连接的用户进行身份验证”。

图 14-41 远程连接身份认证



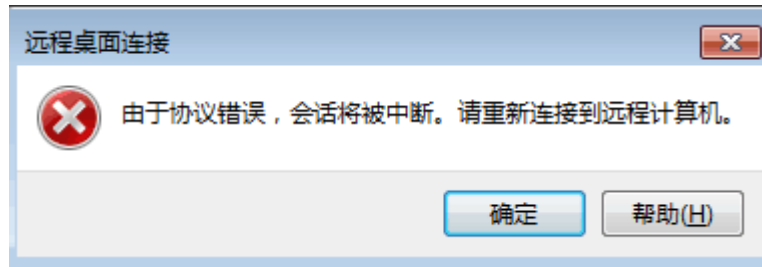
9. 关闭组策略编辑器，然后重启云服务器。

### 14.4.4.9 远程连接 Windows 云服务器报错：由于协议错误会话中断

#### 问题描述

远程桌面提示：由于协议错误，会话将会被中断，请重新连接到远程计算机。

图 14-42 协议错误



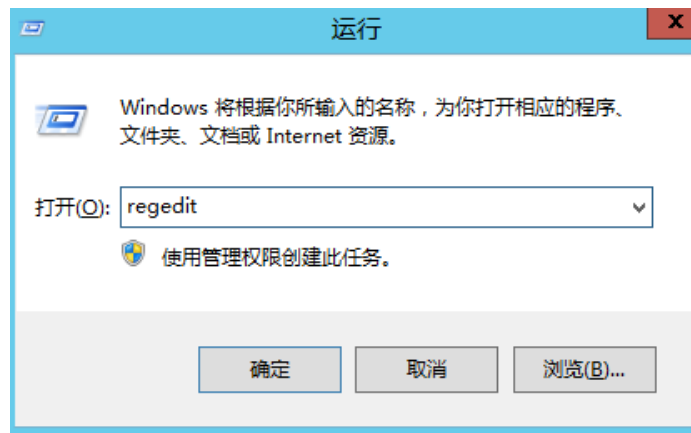
#### 可能原因

注册表中的“Certificate”子键被损坏，导致用户无法与终端服务进行正常通信。

#### 解决方法

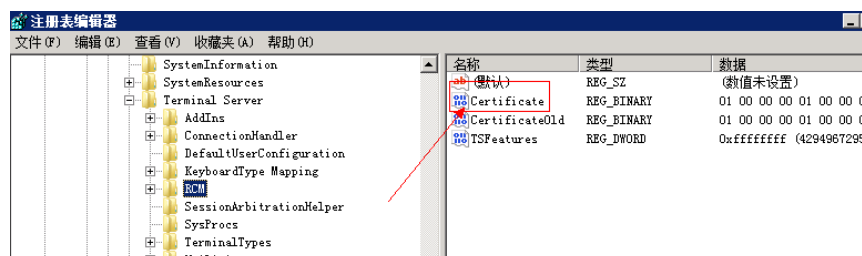
1. 在运行窗口输入“regedit”回车，打开注册表编辑器。

图 14-43 打开注册表



2. 找到HKEY\_LOCAL\_MACHINE\SYSTEM\ControlSet001\Control\Terminal Server\RCM
3. 删除Certificate。

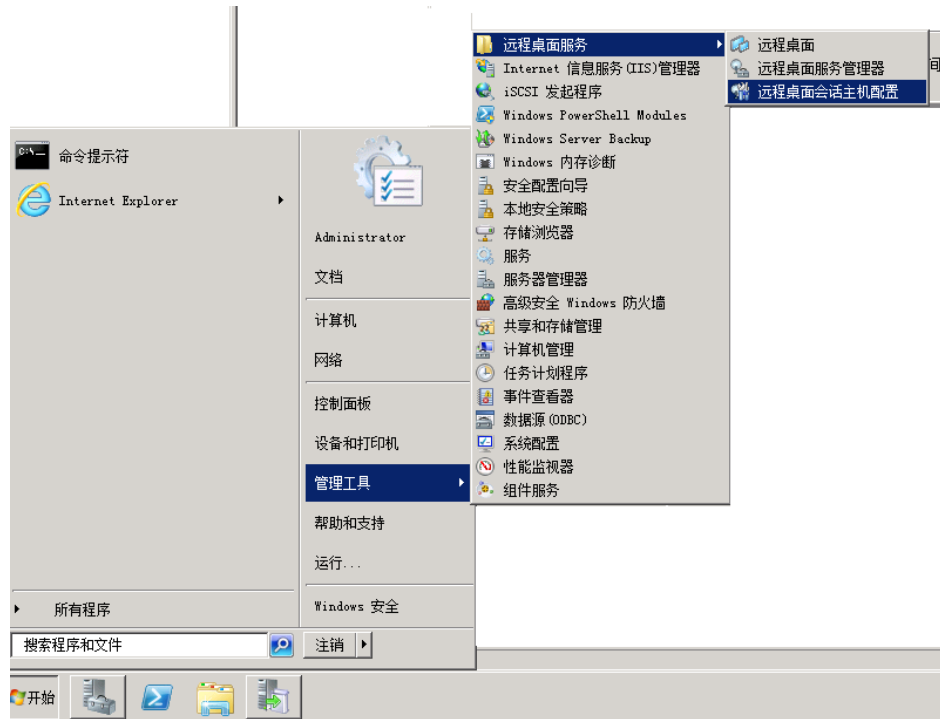
图 14-44 删除 Certificate





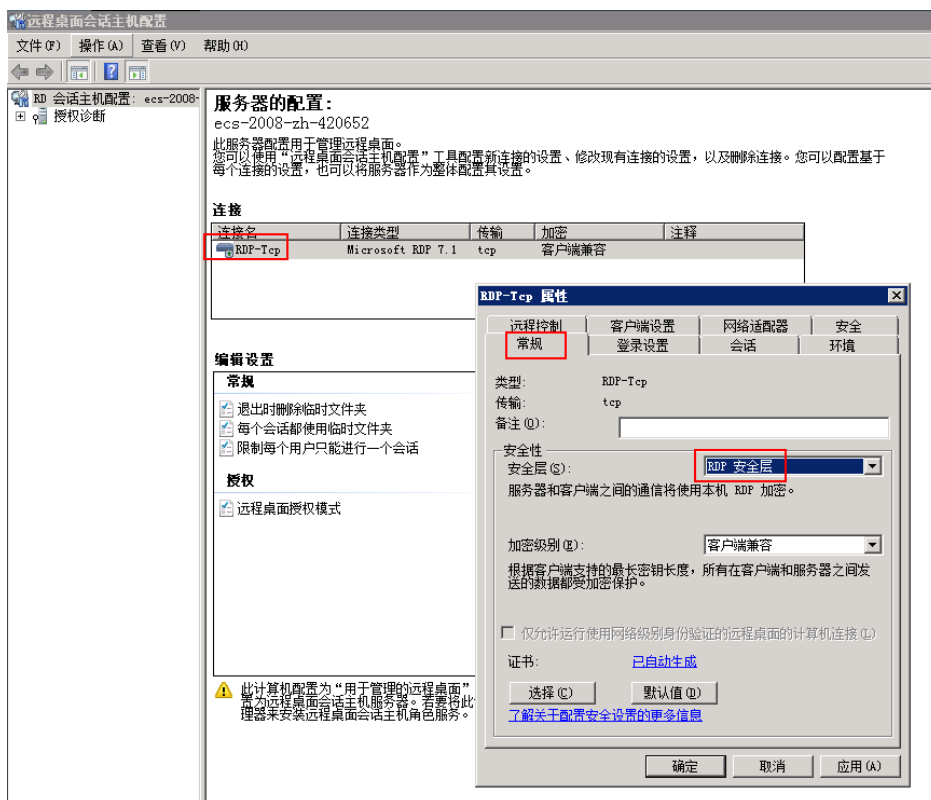
4. 重启云服务器。
5. 打开远程桌面会话主机配置。单击“开始”，打开“管理工具 > 远程桌面服务 > 远程桌面会话主机配置”

图 14-45 打开远程桌面会话主机配置



6. 右键单击RDP-TCP，选择“属性”，打开RDP-TCP属性对话框，修改安全性中的安全层为RDP安全层。

图 14-46 RDP-TCP 属性

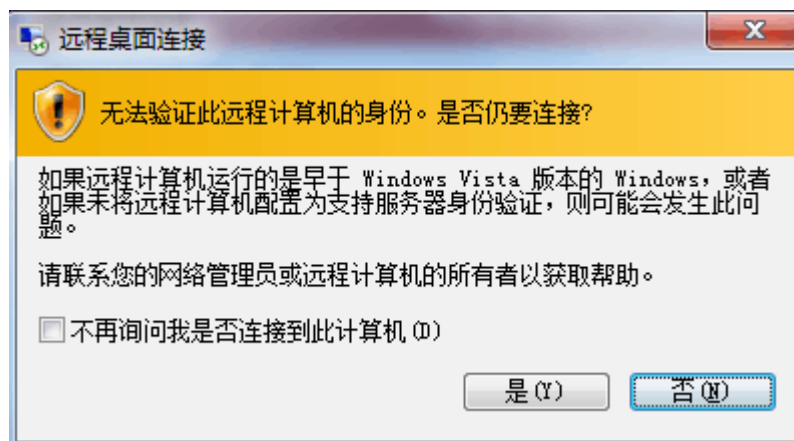


#### 14.4.4.10 远程连接 Windows 云服务器报错：无法验证此远程计算机的身份

##### 问题描述

由于在安全软件中设置了安全登录限制，导致远程桌面连接Windows云服务器报错：无法验证此远程计算机的身份。需要再次登录输入密码。

图 14-47 协议错误



##### 可能原因

云服务器安装了安全软件，防止有未知IP登录云服务器。

## 解决方法

- 卸载安全软件。
- 登录安全软件，将登录安全等级修改为系统默认登录方式。

### 14.4.4.11 远程连接 Windows 云服务器报错：两台计算机无法在分配的时间内连接

#### 问题描述

本地主机远程桌面连接云服务器时报错提示：两台计算机无法在分配的时间内连接。

图 14-48 两台计算机无法在分配的时间内连接



#### 解决方法

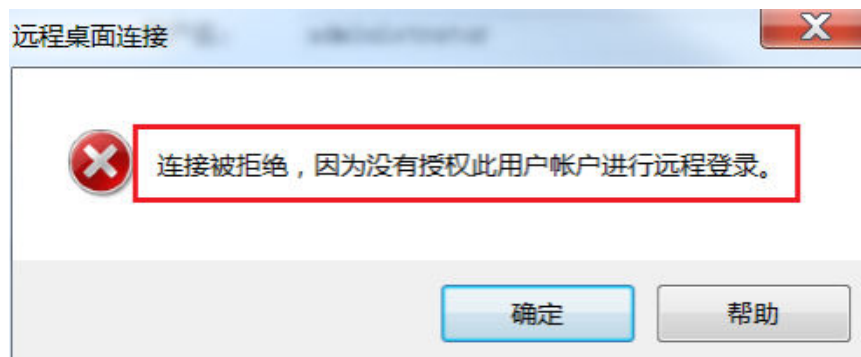
1. 在本地主机单击“开始”，运行中输入cmd，用管理员身份运行cmd。
2. 然后输入命令 `netsh winsock reset`
3. 根据提示重启本地主机后，重新连接云服务器。

### 14.4.4.12 远程连接 Windows 云服务器报错：连接被拒绝未授权此用户

#### 问题描述

远程桌面连接Windows云服务器报错：连接被拒绝，因为没有授权此用户账户进行远程登录。

图 14-49 连接被拒绝未授权此用户



#### 可能原因

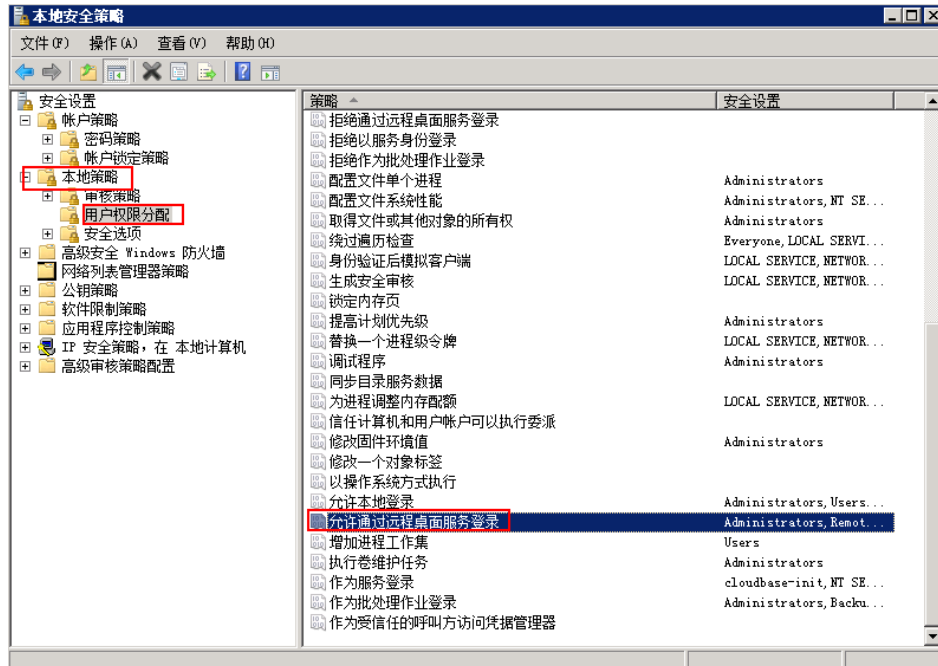
Windows远程桌面相关权限配置异常。

## 解决方法

### 步骤1 检查云服务器远程桌面权限配置。

1. 在运行窗口输入secpol.msc，打开组策略编辑器
2. 打开"本地策略 > 用户权限分配 > 允许通过远程桌面服务登录"。

图 14-50 本地策略



3. 包含已经被赋予远程登录服务器权限的用户组或单独的用户名。  
如果缺少上述配置，请单击“添加用户或组”进行添加。

图 14-51 允许通过远程桌面服务登录



**步骤2** 检查用户的用户组属性。

1. 在运行窗口输入 `lusrmgr.msc`，打开“本地用户和组”配置管理单元。
2. 双击左侧功能树中的“用户”节点，切换到用户管理。
3. 双击出现访问异常的用户名。
4. 在弹出的用户属性对话框中，切换到“隶属于”选项卡，确保用户隶属于上述**步骤2.2**已经被赋予远程登录服务器权限的用户组。

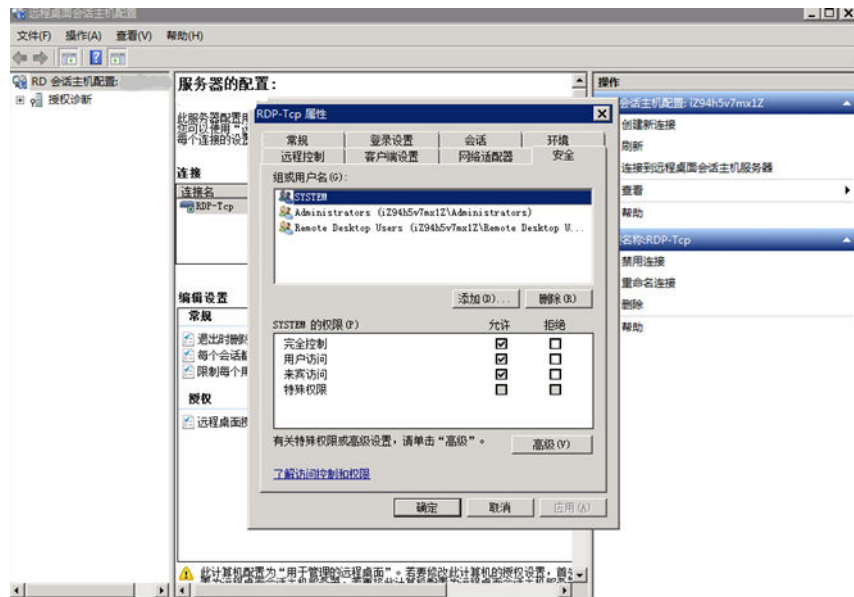
图 14-52 检查用户的用户组属性



**步骤3** 检查远程桌面会话主机配置

1. 在运行窗口输入 `tsconfig.msc`，打开"本远程桌面会话主机配置"。
2. 双击默认的远程桌面连接配置"RDP-Tcp"，或其它用户自己新增的连接配置，然后切换到“安全”选项卡。

图 14-53 安全选项卡



3. 确保“组或用户名”下，包含已经被赋予远程登录服务器权限的用户组或单独的用户名。

如果缺少上述配置，请用户单击“添加”按钮进行添加。

4. 重启服务器或使用如下指令重启远程桌面服务以配置生效。
5. 在命令窗口依次输入如下命令：

```
net stop TermService
net start TermService
```

---结束

#### 14.4.4.13 远程连接 Windows 云服务器报错：您的连接已丢失

##### 问题描述

远程桌面连接Windows云服务器报错：您的远程桌面会话已结束，另一用户已连接到此远程计算机，因此您的连接已丢失。

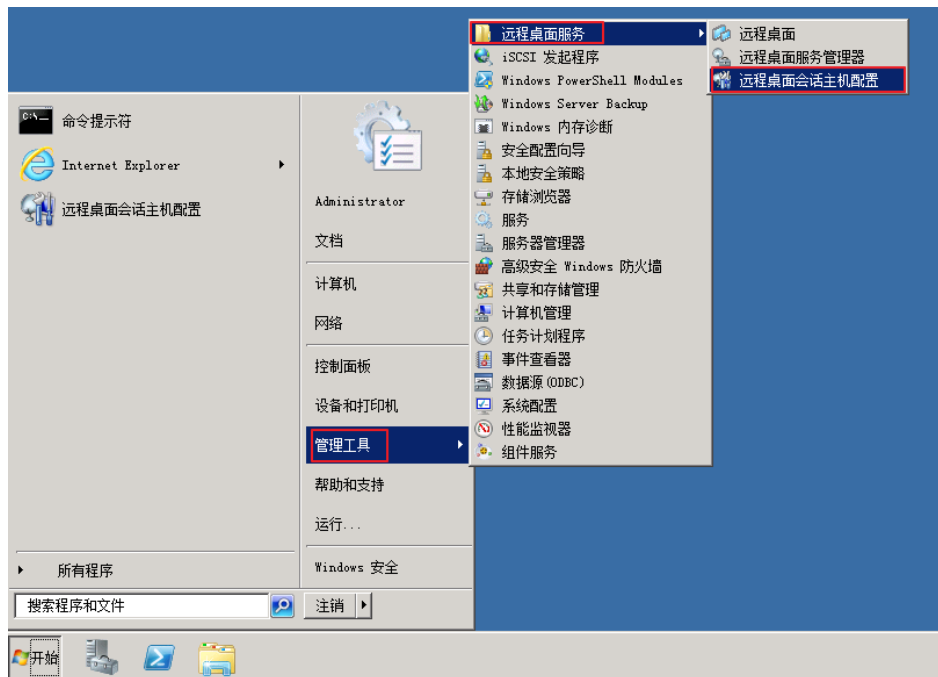
图 14-54 远程桌面会话已结束



##### 2008 操作系统处理方法

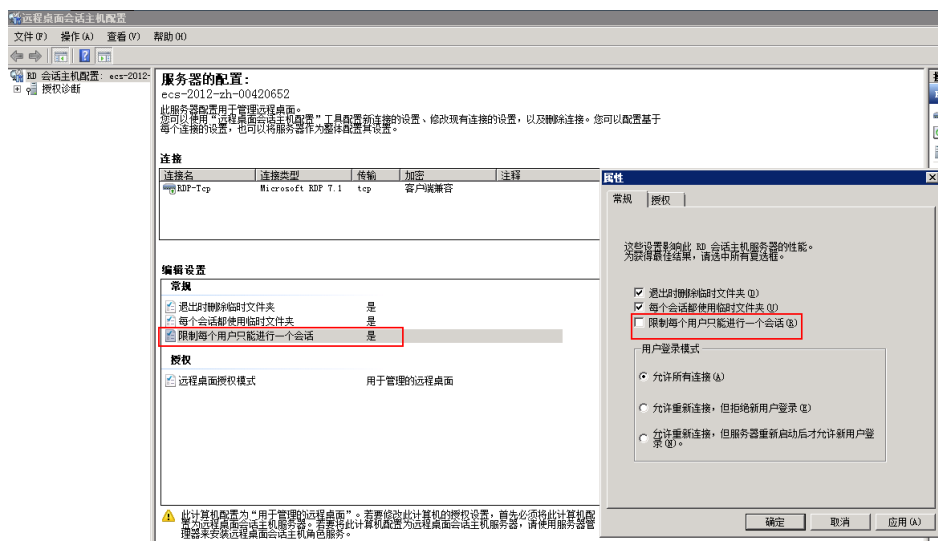
1. 打开“管理工具 > 远程桌面服务 > 远程桌面会话主机配置”，进入远程桌面会话主机配置页面。

图 14-55 远程桌面会话主机配置



2. 双击打开，取消勾选“限制每个用户只能进行一个会话”，单击“确定”。

图 14-56 修改配置

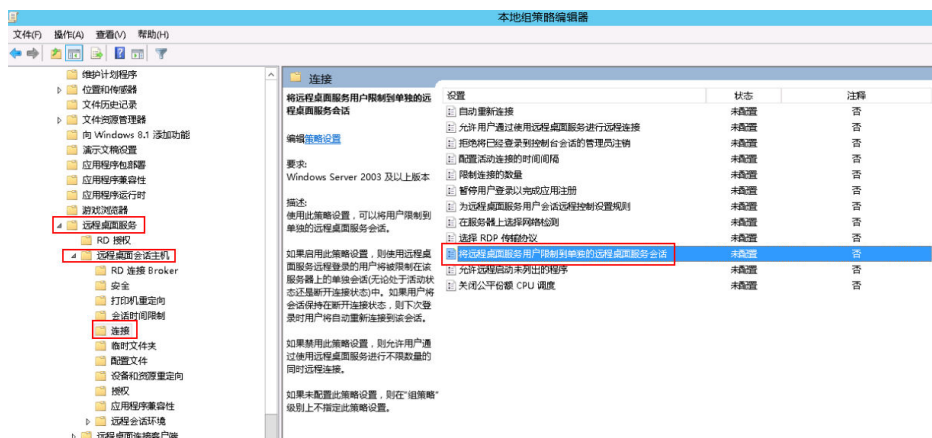


## 2012 操作系统处理方法

1. 打开“开始 > 运行”，输入 gpedit.msc，进入本地组策略编辑器。
2. 打开“计算机配置 > 管理模板 > Windows 组件 > 远程桌面服务 > 远程桌面会话主机 > 连接”。

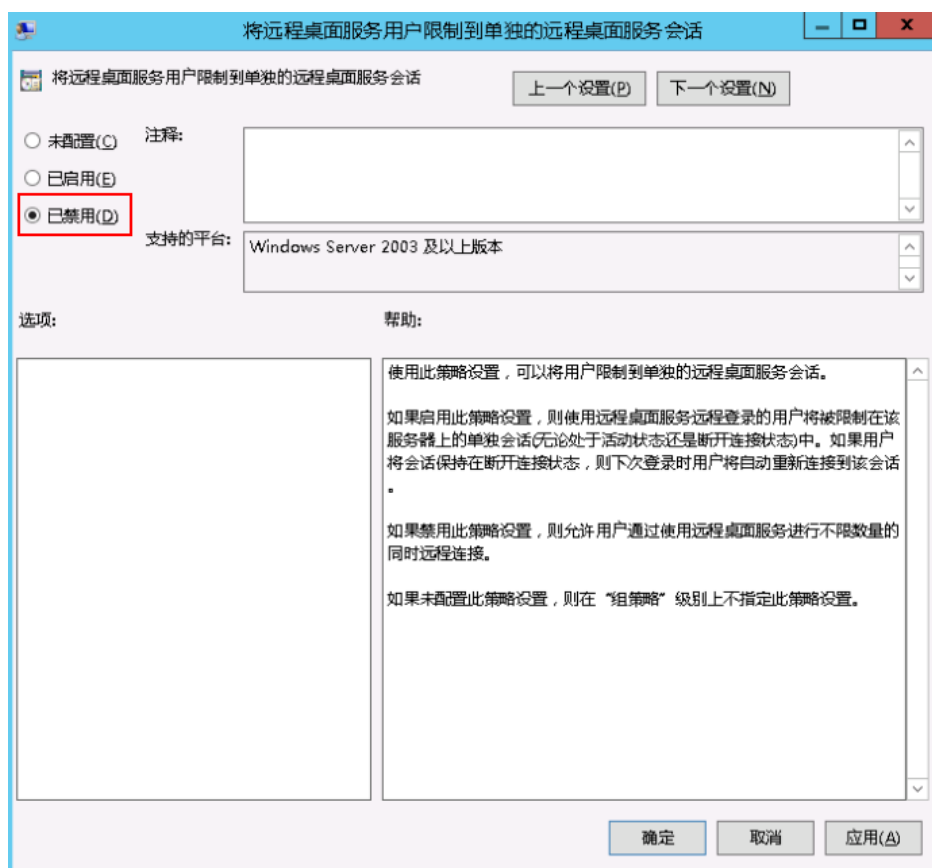


图 14-57 连接



3. 双击打开“将远程桌面服务用户限制到单独的远程桌面服务会话”，修改配置为“已禁用”，单击“确定”保存。

图 14-58 修改配置



4. 在命令窗口执行 `gpupdate /force`，更新组策略。

#### 14.4.4.14 RDP 连接已断开，出现内部错误，错误代码 4

##### 问题描述

Windows操作系统云服务器在登录时会遇到“出现了内部错误”提示，无法正常连接到服务器。这种情况一般都是远程服务太忙导致。

## 可能原因

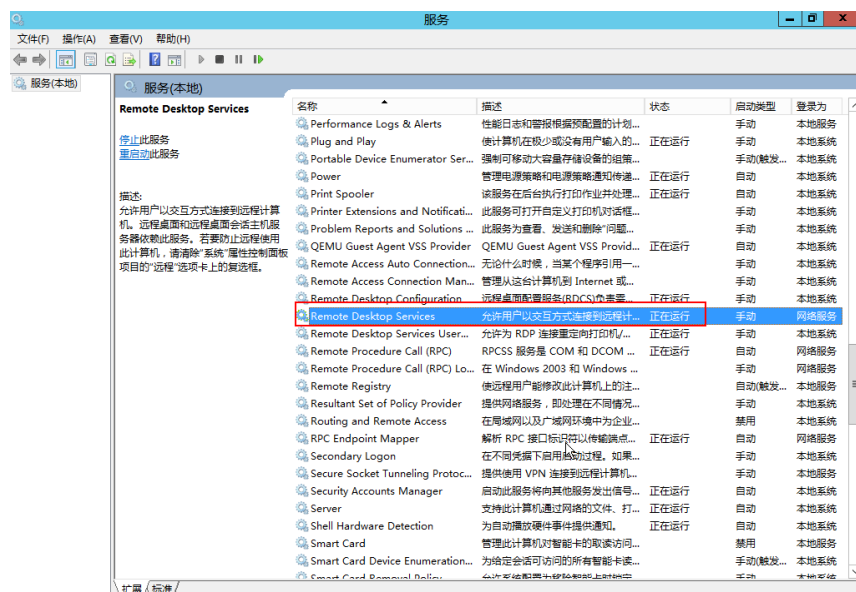
Remote Desktop Services服务忙碌导致的登录异常。

远程桌面登录后断开连接，但未注销登录也可能会出现这个提示。为了避免类似情况，建议在不需要远程连接服务器时及时注销登录。

## 解决方法

1. 通过控制台提供的VNC方式登录云服务器。
2. 打开Windows搜索框，输入services，选择“服务”。
3. 在服务中选择并重启Remote Desktop Services服务，请确保Remote Desktop Services状态为“正在运行”。

图 14-59 Remote Desktop Services



4. 重试远程连接。  
如果仍然无法连接还可以尝试在本地主机以管理员身份运行cmd，执行netsh winsock reset恢复网络连接的默认配置，执行后重试远程连接。

## 14.4.5 Linux 远程登录报错类

### 14.4.5.1 远程连接 Linux 云服务器报错：Module is unknown

#### 问题描述

远程连接Linux云服务器报错：Module is unknown

图 14-60 Module is unknown

```
CentOS Linux 7 (Core)
Kernel 3.10.0-693.el7.x86_64 on an x86_64

Hint: Num Lock on

sy login: root
Password:
Last failed login: Mon Aug 6 19:18:10 CST 2018 on tty1
There were 100 failed login attempts since the last successful login.
Module is unknown
```

### 说明

- 修改此问题需要重启进入救援模式，请评估风险后进行操作。
- 本节操作涉及云服务器重启操作，可能会导致业务中断，请谨慎操作。

## 根因分析

由于错误修改/etc/pam.d/目录下的文件导致。

## 处理方法

1. 进入云服务器的单用户模式。

以CentOS 7操作系统为例：

- a. 单击“远程登录”。
- b. 单击远程登录操作面板上方的“发送CtrlAltDel”按钮，重启虚拟机。
- c. 按上方向键，阻止系统自动继续，在出现内核选项时按字母键e进入内核编辑模式。

图 14-61 进入内核编辑模式

```
CentOS Linux (3.10.0-693.11.1.el7.x86_64) 7 (Core)
CentOS Linux (3.10.0-693.el7.x86_64) 7 (Core)
CentOS Linux (0-rescue-fd73dfefb69a4e4183fd3f86ccc29526) 7 (Core)
```

### 说明

Euler镜像默认对grub文件进行了加密，进入编辑内核模式时会提示：Enter username，需要输入用户和密码，请联系管理员获取。

- d. 找到linux16行末尾，删除不需要加载的参数到ro参数。
- e. 修改ro为rw，以读写方式挂载根分区。
- f. 并添加rd.break，然后执行Ctrl+X。

图 14-62 修改前

```
insmod ext2
set root='hd0,msdos1'
if [ x${feature_platform_search_hint} = xy ]; then
  search --no-floppy --fs-uuid --set=root --hint='hd0,msdos1' b13ee9c\
8-0ef0-4159-9b90-fc47bde0d464
else
  search --no-floppy --fs-uuid --set=root b13ee9c8-0ef0-4159-9b90-fc47\
bde0d464
fi
linux16 /boot/vmlinuz-3.10.0-327.62.59.83.h162.x86_64 root=UUID=b13ee9\
c8-0ef0-4159-9b90-fc47bde0d464 ro crash_kexec_post_notifiers softlockup_panic=\
1 panic=3 reserve_kbox_mem=16M nmi_watchdog=1 rd.shell=0 net.ifnames=0 spectre\
_v2=off nopti noibrs noibpb crashkernel=auto LANG=en_US.UTF-8
initrd16 /boot/initramfs-3.10.0-327.62.59.83.h162.x86_64.img
```

图 14-63 修改后

```
insmod ext2
set root='hd0,msdos1'
if [ x${feature_platform_search_hint} = xy ]; then
  search --no-floppy --fs-uuid --set=root --hint='hd0,msdos1' b13ee9c\
8-0ef0-4159-9b90-fc47bde0d464
else
  search --no-floppy --fs-uuid --set=root b13ee9c8-0ef0-4159-9b90-fc47\
bde0d464
fi
linux16 /boot/vmlinuz-3.10.0-327.62.59.83.h162.x86_64 root=UUID=b13ee9\
c8-0ef0-4159-9b90-fc47bde0d464 rw rd.break
initrd16 /boot/initramfs-3.10.0-327.62.59.83.h162.x86_64.img
```

g. 执行以下命令切换至/sysroot目录。

```
# chroot /sysroot
```

2. 执行以下命令，查看系统日志定位出错的文件

```
grep Module /var/log/messages
```

图 14-64 系统日志

```
Aug 6 18:08:09 sy login: pam_succeed_if(login:auth): requirement "uid >= 1000" not met by user "root"
Aug 6 18:08:11 sy login: FAILED LOGIN 1 FROM tty1 FOR root, Authentication failure
Aug 6 18:08:15 sy login: pam_unix(login:session): session opened for user root by LOGIN(uid=0)
Aug 6 18:08:15 sy login: Module is unknown
Aug 6 18:10:41 sy login: PAM unable to dlopen(/lib/security/pam_limits.so): /lib/security/pam_limits.so: cannot open shared obj
ect file: No such file or directory
Aug 6 18:10:41 sy login: PAM adding faulty module: /lib/security/pam_limits.so
Aug 6 18:10:44 sy login: pam_unix(login:session): session opened for user root by LOGIN(uid=0)
Aug 6 18:10:44 sy login: Module is unknown
```

3. 编辑系统日志中提示的错误文件，并注释或修改错误行。

```
vi /etc/pam.d/login
```

图 14-65 修改错误信息

```
session required pam_selinux.so open
session required pam_namespace.so
session optional pam_keyinit.so force revoke
session include system-auth
session include postlogin
-session optional pam_ck_connector.so
# session required /lib/security/pam_limits.so
```

4. 重启服务器，重试连接云服务器。

### 📖 说明

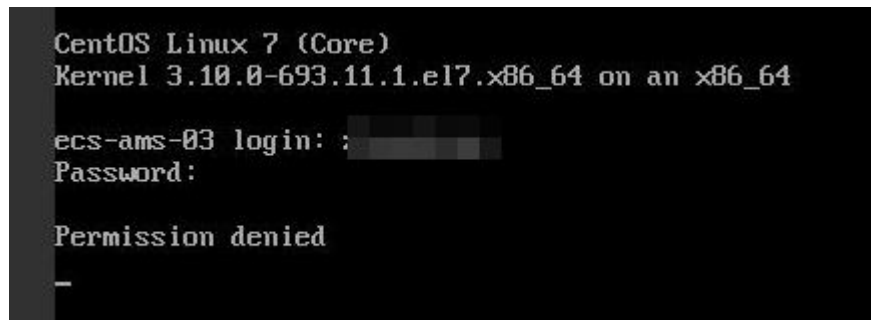
- 如需查看修改记录，定位是否人为错误修改导致，请执行以下命令。  
`vi /root/.bash_history`  
搜索关键字vi或者login
- 请勿随便修改/etc/pam.d/目录下的文件。如需对pam详细了解可查看pam.d帮助手册，在系统内执行以下命令。  
`man pam.d`

## 14.4.5.2 远程连接 Linux 云服务器报错：Permission denied

### 问题现象

远程连接Linux云服务器报错：Permission denied

图 14-66 Permission denied



### 📖 说明

- 修改此问题需要重启进入救援模式，请评估风险后进行操作。
- 本节操作涉及云服务器重启操作，可能会导致业务中断，请谨慎操作。

### 根因分析

/etc/security/limits.conf中的nofile 用来设置系统允许打开的最大文件数目，如果nofile值大于PermissionDenied.png内核设置的fs.nr\_open参数值（默认为1048576），会导致登录校验错误，导致登录云服务器时提示“Permission denied”。

### 处理方法

1. 进入云服务器的单用户模式。  
以CentOS 7操作系统为例：
  - a. 单击“远程登录”。
  - b. 单击远程登录操作面板上方的“发送CtrlAltDel”按钮，重启虚拟机。
  - c. 按上方向键，阻止系统自动继续，在出现内核选项时按字母键e进入内核编辑模式。

图 14-67 进入内核编辑模式

```
CentOS Linux (3.10.0-693.11.1.el7.x86_64) ? (Core)
CentOS Linux (3.10.0-693.el7.x86_64) ? (Core)
CentOS Linux (0-rescue-fd73dfefb69a4e4183fd3f86ccc29526) ? (Core)
```

### 说明

Euler镜像默认对grub文件进行了加密，进入编辑内核模式时会提示：Enter username，需要输入用户和密码，请联系管理员获取。

- d. 找到linux16行末尾，删除不需要加载的参数到ro参数。
- e. 修改ro为rw，以读写方式挂载根分区。
- f. 并添加rd.break，然后执行Ctrl+X。

图 14-68 修改前

```
insmod ext2
set root='hd0,msdos1'
if [ x$feature_platform_search_hint = xy ]; then
  search --no-floppy --fs-uuid --set=root --hint='hd0,msdos1' b13ee9c\
8-0ef0-4159-9b90-fc47bde0d464
else
  search --no-floppy --fs-uuid --set=root b13ee9c8-0ef0-4159-9b90-fc47\
bde0d464
fi
linux16 /boot/vmlinuz-3.10.0-327.62.59.83.h162.x86_64 root=UUID=b13ee9\
c8-0ef0-4159-9b90-fc47bde0d464 ro crash_kexec_post_notifiers softlockup_panic=\
1 panic=3 reserve_kbox_mem=16M nmi_watchdog=1 rd.shell=0 net.ifnames=0 spectre\
_v2=off nopti noibrs noibpb crashkernel=auto LANG=en_US.UTF-8
initrd16 /boot/initramfs-3.10.0-327.62.59.83.h162.x86_64.img
```

图 14-69 修改后

```
insmod ext2
set root='hd0,msdos1'
if [ x$feature_platform_search_hint = xy ]; then
  search --no-floppy --fs-uuid --set=root --hint='hd0,msdos1' b13ee9c\
8-0ef0-4159-9b90-fc47bde0d464
else
  search --no-floppy --fs-uuid --set=root b13ee9c8-0ef0-4159-9b90-fc47\
bde0d464
fi
linux16 /boot/vmlinuz-3.10.0-327.62.59.83.h162.x86_64 root=UUID=b13ee9\
c8-0ef0-4159-9b90-fc47bde0d464 rw rd.break
initrd16 /boot/initramfs-3.10.0-327.62.59.83.h162.x86_64.img
```

- g. 执行以下命令切换至/sysroot目录。  
**# chroot /sysroot**
2. 执行以下命令，查询内核的fs.nr\_open值。  
**sysctl fs.nr\_open**
3. 编辑 /etc/security/limits.conf，修改配置的nofile值为合理的值，需小于2中查询的fs.nr\_open值，例如65535。  
**vi /etc/security/limits.conf**

### 📖 说明

limits.conf 文件实际是Linux PAM（插入式认证模块，Pluggable Authentication Modules）中pam\_limits.so 的配置文件。更多详细配置信息请查看man手册，执行：  
**man limits.conf**

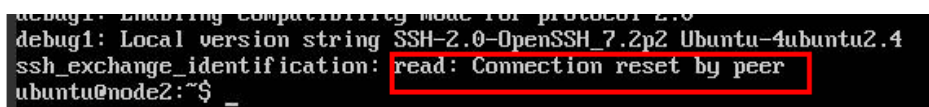
4. 重启服务器，重试连接云服务器。

## 14.4.5.3 远程连接 Linux 云服务器报错：read: Connection reset by peer

### 问题现象

远程连接Linux云服务器报错：read: Connection reset by peer

图 14-70 read: Connection reset by peer



### 可能原因

- 安全组未放通远程登录端口。
- 防火墙开启，且关闭了远程连接端口。

### 处理方法

针对此问题，推荐采用以下方式来排查：

- **检查安全组规则**
  - 入方向：打开远程登录端口。默认使用的22端口。
  - 出方向：出方向规则为白名单（允许），放通出方向网络流量。
- **云服务器防火墙添加端口例外**

以Ubuntu操作系统为例：

- a. 执行以下命令检查防火墙状态：

```
sudo ufw status
```

回显信息如下：

```
Status: active
```

- b. 添加端口例外，以默认使用的22端口为例。

```
ufw allow 22
```

```
Rule added
```

```
Rule added (v6)
```

- c. 重新查看防火墙状态

```
sudo ufw status
```

```
Status: active
```

To	Action	From
--	----	----
22	ALLOW	Anywhere
22 (v6)	ALLOW	Anywhere (v6)

规则添加成功，重新测试远程连接云服务器。

#### 14.4.5.4 远程连接 Linux 云服务器报错：Access denied

##### 问题现象

远程连接Linux云服务器报错：Access denied

##### 可能原因

- 账号或密码输入错误。
- SSH服务端配置了禁止root用户登录的策略。

##### 处理方法

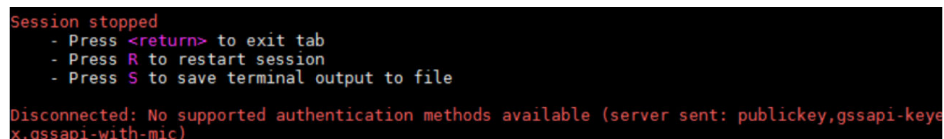
- 账号或密码输入错误。  
检查输入的用户名或密码。
- SSH服务端配置了禁止root用户登录的策略。
  - a. 编辑 `/etc/ssh/sshd_config` 文件，检查如下设置，确保root用户能通过SSH登录：  
PermitRootLogin yes
  - b. 重启 SSH 服务。
    - CentOS 6  
`service sshd restart`
    - CentOS 7  
`systemctl restart sshd`

#### 14.4.5.5 远程连接 Linux 云服务器报错：Disconnected: No supported authentication methods available

##### 问题现象

远程连接Linux云服务器报错：Disconnected: No supported authentication methods available.

图 14-71 No supported authentication methods available



```
Session stopped
- Press <return> to exit tab
- Press R to restart session
- Press S to save terminal output to file
Disconnected: No supported authentication methods available (server sent: publickey,gssapi-keyex,gssapi-with-mic)
```

##### 可能原因

SSH服务端配置了禁止密码验证登录的策略。

##### 处理方法

1. 编辑 `/etc/ssh/sshd_config` 文件，检查如下设置  
`vi /etc/ssh/sshd_config`



2. 修改如下配置项：  
把PasswordAuthentication no 改为 PasswordAuthentication yes  
或去掉PasswordAuthentication yes 前面的#注释掉。
3. 重启 SSH 服务。
  - CentOS 6  
`service sshd restart`
  - CentOS 7  
`systemctl restart sshd`

## 14.5 磁盘分区/挂载/扩容

### 14.5.1 Windows 操作系统云服务器看不到挂载的数据盘怎么办？

#### 问题描述

登录Windows操作系统云服务器看不到挂载的数据盘。



格式化磁盘会造成数据丢失，请提前对数据盘创建备份，避免数据丢失。

---

#### 可能原因

- 新添加的数据盘未进行分区和初始化。
- 切换操作系统或变更规格导致的磁盘脱机。

#### 新添加的数据盘未进行分区和初始化

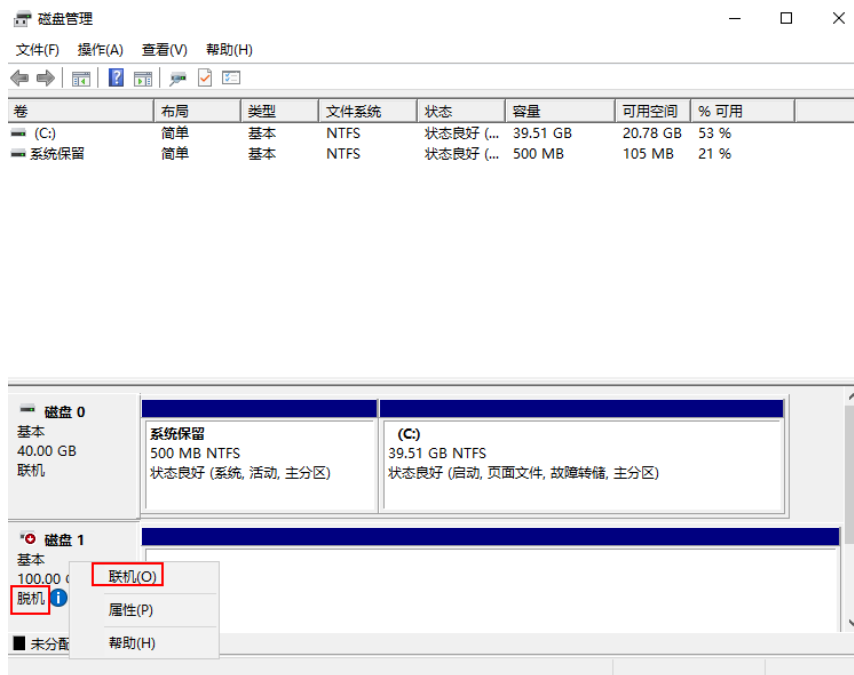
新的数据盘默认是没有分区和文件系统的，因此在资源管理器中不会显示，需要先手动执行初始化操作。

#### 切换操作系统或变更规格导致的磁盘脱机

切换操作系统后可能会由于文件系统不一致导致数据盘无法显示；Windows云服务器变更规格后，可能会发生磁盘脱机。

1. 登录云服务器，打开cmd运行窗口，并输入“diskmgmt.msc”打开磁盘管理。  
查看磁盘状态，磁盘可能显示为“脱机”。
2. 将磁盘设置为“联机”状态。  
在磁盘列表中，单击脱机的磁盘，并选择“联机”，将脱机磁盘进行联机。

图 14-72 磁盘联机



3. 再次在“计算机”中查询数据盘是否显示正常。

如上述操作后磁盘仍未显示，可以重新进行磁盘初始化和分区的操作。格式化磁盘会造成数据丢失，请提前对数据盘创建备份，避免数据丢失。

## 14.5.2 怎样调整系统盘分区？

### 操作场景

弹性云服务器创建成功后，如果发现系统盘分区的容量大小和实际创建的系统盘大小不一致，可以通过手动调整系统盘分区，扩容系统盘的空白空间。

扩容的方法有如下两种：

- 将空白分区划分成新分区，并将新分区格式化后挂载到root根分区的某个目录下。具体方法请参见本节内容。
- 将扩容的空白分区直接扩容到根分区。具体方法请参见FAQ：
  - [如何将扩容系统盘的空白分区在线扩容到末尾的root分区？](#)
  - [如何将扩容系统盘的空白分区在线扩容到非末尾的root分区？](#)

### 操作步骤

以镜像为CentOS 7.3 64bit的弹性云服务器为例，用户在创建弹性云服务器时，创建了容量为60GB的系统盘，但是查询系统盘分区后，显示的容量大小仅为40GB。

为使用增加的20GB容量，需调整系统盘分区，具体操作如下：

#### 步骤1 查看磁盘分区。

1. 登录Linux弹性云服务器。
2. 执行以下命令，切换至root用户。

```
sudo su -
```

3. 执行以下命令，查看云服务器的磁盘详情。

```
fdisk -l
```

回显类似如下，其中，“/dev/xvda”或“/dev/vda”表示系统盘。

图 14-73 查看磁盘详情

```
[root@ecs-8d6c ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/xvda1      38G   1.2G   35G   4% /
devtmpfs        899M   0   899M   0% /dev
tmpfs           908M   0   908M   0% /dev/shm
tmpfs           908M  8.4M   900M   1% /run
tmpfs           908M   0   908M   0% /sys/fs/cgroup
tmpfs           182M   0   182M   0% /run/user/0
[root@ecs-8d6c ~]# fdisk -l

Disk /dev/xvda: 64.4 GB, 64424509440 bytes, 125829120 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x0004d5e5

   Device Boot      Start         End      Blocks   Id  System
/dev/xvda1 *        2048     79980543   39989248   83  Linux
/dev/xvda2          79980544   83886079    1952768   82  Linux swap / Solaris
[root@ecs-8d6c ~]#
```

4. 执行以下命令，查看磁盘分区。

```
parted -l /dev/xvda
```

图 14-74 查看磁盘分区

```
[root@ecs-8d6c ~]# parted -l /dev/xvda
Model: Xen Virtual Block Device (xvd)
Disk /dev/xvda: 64.4GB
Sector size (logical/physical): 512B/512B
Partition Table: msdos
Disk Flags:

Number  Start   End     Size    Type     File system  Flags
  1      1049kB  41.0GB  40.9GB  primary  ext4         boot
  2      41.0GB  42.9GB  2000MB  primary  linux-swap(v1)
```

**步骤2** 将系统盘扩容后的空间划分为一个新的分区。

1. 执行以下命令，进入fdisk模式。以“/dev/xvda”为例：

```
fdisk /dev/xvda
```

回显类似如下：

```
[root@ecs-8d6c ~]# fdisk /dev/xvda
Welcome to fdisk (util-linux 2.23.2).
```

```
Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.
```

```
Command (m for help):
```

2. 输入“n”，按“Enter”，开始新建分区。

本例中由于系统盘原来已经有2个分区，所以系统自动添加第3个分区。

回显类似如下：

图 14-75 新建分区

```
[root@ecs-8d6c ~]# fdisk /dev/xvda
Welcome to fdisk (util-linux 2.23.2).

Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.

Command (m for help): n
Partition type:
   p   primary (2 primary, 0 extended, 2 free)
   e   extended
Select (default p):
Using default response p
Partition number (3,4, default 3):
First sector (83886080-125829119, default 83886080):
Using default value 83886080
Last sector, +sectors or +size(K,M,G) (83886080-125829119, default 125829119):
Using default value 125829119
Partition 3 of type Linux and of size 20 GiB is set

Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.

WARNING: Re-reading the partition table failed with error 16: Device or resource busy.
The kernel still uses the old table. The new table will be used at
the next reboot or after you run partprobe(8) or kpartx(8)
Syncing disks.
[root@ecs-8d6c ~]#
```

3. 输入新分区的起始磁柱编号，按“Enter”。  
其中，起始磁柱编号必须大于原有分区的结束磁柱编号。本例中设置新分区的起始磁柱编号为默认值，直接按“Enter”，则回显类似如下：

图 14-76 输入新分区的起始磁柱编号

```
First sector (83886080-125829119, default 83886080):
Using default value 83886080
Last sector, +sectors or +size(K,M,G) (83886080-125829119, default 125829119):
```

4. 输入新分区的结束磁柱编号，按“Enter”。  
本例中设置新分区的结束磁柱编号为默认值，直接按“Enter”，则回显类似如下：

图 14-77 输入新分区的结束磁柱编号

```
Last sector, +sectors or +size(K,M,G) (83886080-125829119, default 125829119):
Using default value 125829119
Partition 3 of type Linux and of size 20 GiB is set
```

5. 输入“p”，按“Enter”，查看新建分区。  
回显类似如下：

图 14-78 查看新建分区

```
Command (m for help): p

Disk /dev/xvda: 64.4 GB, 64424509440 bytes, 125829120 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x0004d5e5

   Device Boot      Start         End      Blocks   Id  System
/dev/xvda1    *            2048     79980543    39989248    83  Linux
/dev/xvda2                79980544    83886079    1952768    82  Linux swap / Solaris
/dev/xvda3                83886080    125829119    20971520    83  Linux
```

6. 输入“w”，按“Enter”，保存并退出分区信息。  
系统会自动将分区结果写入分区表中，完成分区创建。  
回显类似如下：

图 14-79 完成分区创建

```
Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.

WARNING: Re-reading the partition table failed with error 16: Device or resource busy.
The kernel still uses the old table. The new table will be used at
the next reboot or after you run partprobe(8) or kpartx(8)
Syncing disks.
```

7. 执行以下命令，查看磁盘分区。  
**parted -l /dev/xvda**

图 14-80 查看新的磁盘分区

```
Disk Flags:
Number  Start   End     Size   Type    File system  Flags
 1      1049kB  41.0GB  40.9GB primary ext4          boot
 2      41.0GB  42.9GB  2000MB primary linux-swap(v1)
 3      42.9GB  64.4GB  21.5GB primary ext4
```

**步骤3** 执行以下命令，将分区表的变更同步至操作系统。

```
partprobe
```

**步骤4** 执行以下命令，设置新建分区文件系统的格式。

1. 执行以下命令，查看文件系统的格式。  
**df -TH**

图 14-81 查看文件系统的格式

```
[root@ecs-8d6c ~]# df -TH
Filesystem      Type      Size  Used Avail Use% Mounted on
/dev/xvda1      ext4      41G   1.3G   37G   4% /
devtmpfs        devtmpfs  943M    0   943M   0% /dev
tmpfs           tmpfs     952M    0   952M   0% /dev/shm
tmpfs           tmpfs     952M   8.8M   944M   1% /run
tmpfs           tmpfs     952M    0   952M   0% /sys/fs/cgroup
tmpfs           tmpfs     191M    0   191M   0% /run/user/0
[root@ecs-8d6c ~]#
```

2. 以“ext4”文件格式为例，命令行如下：  
**mkfs -t ext4 /dev/xvda3**

#### 📖 说明

格式化需要等待一段时间，请观察系统运行状态，不要退出。

回显类似如下：

```
[root@ecs-86dc ~]# mkfs -t ext4 /dev/xvda3
mke2fs 1.42.9 (28-Dec-2013)
Filesystem label=
OS type: Linux
```

```
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
1790544 inodes, 7156992 blocks
357849 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=2155872256
219 block groups
32768 blocks per group, 32768 fragments per group
8176 inodes per group
Superblock backups stored on blocks:
    32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
    4096000

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

**步骤5** 将新建分区挂载到需要增加空间的目录下。

新增加的分区挂载到不为空的目录时，该目录下原本的子目录和文件会被隐藏，所以，新增的分区最好挂载到空目录或者新建目录。如果确实要挂载到不为空的目录，可将该目录下的子目录和文件临时移动到其他目录下，待新分区挂载成功后，再将子目录和文件移动回来。

以新建目录/root/new为例：

1. 执行以下命令，在/root下新建目录new。

```
mkdir /root/new
```

2. 执行以下命令，将新分区挂载至“/root/new”目录。

```
mount /dev/xvda3 /root/new
```

回显类似如下：

```
[root@ecs-86dc ]# mount /dev/xvda3 /root/new
[root@ecs-86dc ]#
```

3. 执行以下命令，查看挂载结果。

```
df -TH
```

回显类似如下：

**图 14-82** 查看挂载结果

```
[root@ecs-8d6c ~]# df -TH
Filesystem      Type      Size  Used Avail Use% Mounted on
/dev/xvda1      ext4      41G   1.3G   37G   4% /
devtmpfs        devtmpfs  943M   0    943M   0% /dev
tmpfs           tmpfs     952M   0    952M   0% /dev/shm
tmpfs           tmpfs     952M   8.8M  944M   1% /run
tmpfs           tmpfs     952M   0    952M   0% /sys/fs/cgroup
/dev/xvda3      ext4      22G    47M   20G   1% /root/new
tmpfs           tmpfs     191M   0    191M   0% /run/user/0
[root@ecs-8d6c ~]#
```

**步骤6** 根据需要，判断是否需要将新增磁盘设置为开机自动挂载。

如果未设置开机自动挂载，重启弹性云服务器后，用户需要将新建分区重新挂载至指定目录。

- 是，执行**步骤7**。
- 否，任务结束。

**步骤7** 设置开机自动挂载新增磁盘。**说明**

禁止将未经过格式化的磁盘设为开机自动挂载，否则会导致弹性云服务器无法正常启动。

1. 执行以下命令，获取文件系统类型、UUID。

```
blkid
```

**图 14-83** 查看文件系统类型

```
[root@ecs-8d6c ~]# blkid
/dev/xvda1: UUID="7c4fce5d-f8f7-4ed6-8463-f2bd22d0ddea" TYPE="ext4"
/dev/xvda2: UUID="5de3cf2c-30c6-4fb2-9e63-830439d4e674" TYPE="swap"
/dev/xvda3: UUID="96e5e028-60fb-4547-a82a-35ace1086c4f" TYPE="ext4"
[root@ecs-8d6c ~]#
```

则新建分区的UUID=96e5e028-b0fb-4547-a82a-35ace1086c4f。

2. 执行以下命令，使用vi编辑器打开“fstab”文件。

```
vi /etc/fstab
```

3. 按“i”，进入编辑模式。

4. 将光标移动至文件末尾，按“Enter”，并添加如下语句：

```
UUID=96e5e028-b0fb-4547-a82a-35ace1086c4f /root/new ext4 defaults 0 0
```

5. 按“Esc”，执行以下命令，并按“Enter”，保存设置并退出编辑器。

```
:wq
```

**说明**

将新增磁盘设置为开机自动挂载后，如果要卸载该磁盘，必须先删除新增磁盘的自动挂载配置信息，否则卸载该磁盘后，会导致弹性云服务器无法正常启动。删除自动挂载配置信息的方法为：

1. 执行以下命令，使用vi编辑器打开“fstab”文件。

```
vi /etc/fstab
```

2. 按“i”，进入编辑模式。

3. 删除如下语句：

```
UUID=96e5e028-b0fb-4547-a82a-35ace1086c4f /root/new ext4 defaults 0 0
```

4. 按“Esc”，执行以下命令，并按“Enter”，保存设置并退出编辑器。

```
:wq
```

----结束

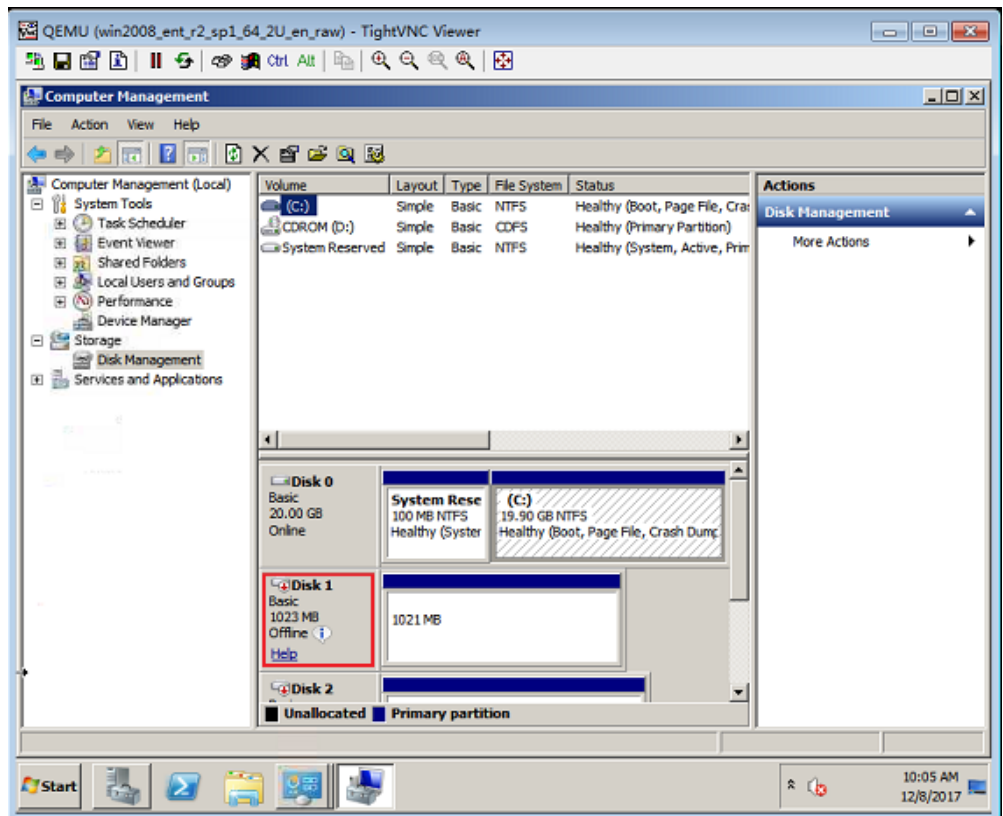
### 14.5.3 怎样查询 Windows 弹性云服务器磁盘分区与磁盘设备的对应关系？

以Windows 2008 R2 64bit为例，介绍查询Windows弹性云服务器磁盘分区与磁盘设备对应关系的方法。

1. 登录Windows弹性云服务器。
2. 单击桌面左下角的“开始”菜单。
3. 选择“控制面板 > 管理工具 > 计算机管理”。
4. 在左侧导航栏，选择“存储 > 磁盘管理”。

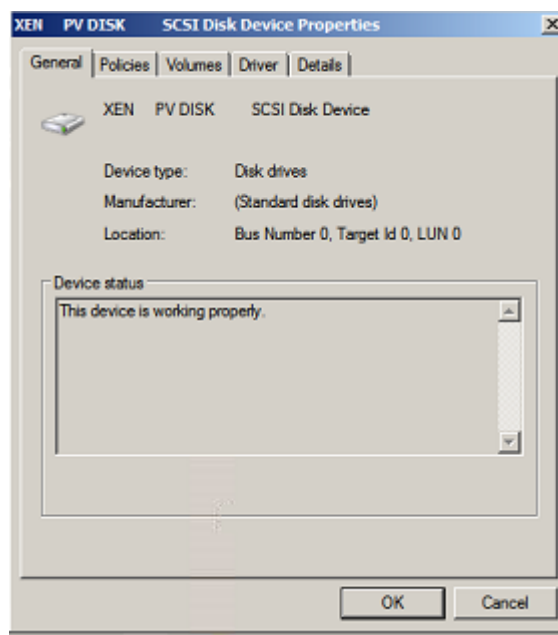


图 14-84 磁盘管理



5. 以图14-84为例，查询Disk 1对应的磁盘设备：
  - a. 右键单击“Disk 1”所在灰色区域，如图14-84红框所示。
  - b. 单击“属性”。系统打开磁盘属性窗口，如图14-85所示。

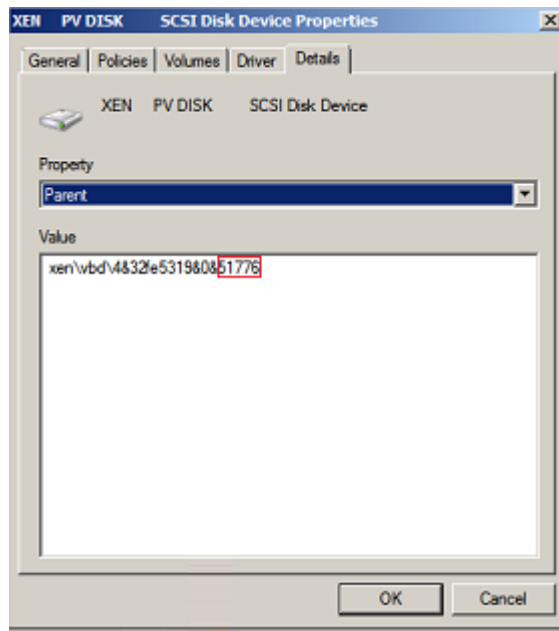
图 14-85 磁盘属性





- c. 选择“详细信息”页签，并设置参数“属性”的值为“父系”。

图 14-86 详细信息



- d. 记录参数“值”中符号“&”后的几位数字，该值为磁盘分区对应的主从设备号，如“51776”。
- e. 根据表14-2，查询对应的磁盘设备。  
“51776”对应的磁盘设备为“xvde”，即磁盘1使用的磁盘设备为xvde。

表 14-2 磁盘分区与磁盘设备的映射关系

磁盘分区对应的主从设备号	磁盘设备
51712	xvda
51728	xvdb
51744	xvdc
51760	xvdd
51776	xvde
51792	xvdf
51808	xvdg
51824	xvdh
51840	xvdi
51856	xvdj
51872	xvdk
51888	xvdl

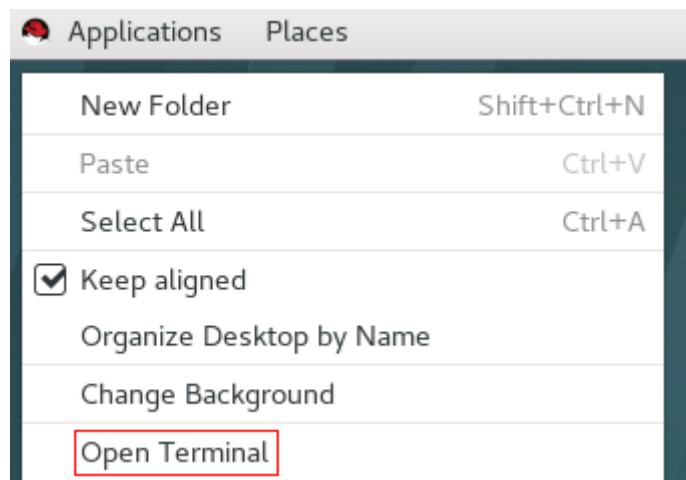
磁盘分区对应的主从设备号	磁盘设备
51904	xvdm
51920	xvdn
51936	xvdo
51952	xvdp
268439552	xvdq
268439808	xvdr
268440064	xvds
268440320	xvdt
268440576	xvdu
268440832	xvdv
268441088	xvdw
268441344	xvdx

#### 14.5.4 怎样查询 Linux 弹性云服务器磁盘分区与磁盘设备的对应关系？

对于Linux弹性云服务器，磁盘分区与磁盘设备名称完全对应。本节以Red Hat Enterprise Linux 7的图形界面为例，介绍查询Linux弹性云服务器磁盘分区与磁盘设备的方法。

1. 以root权限登录Red Hat Enterprise Linux 7弹性云服务器。
2. 在桌面空白处右键打开一个终端会话。

图 14-87 open terminal



3. 执行以下命令，查询磁盘分区和磁盘设备。

```
fdisk -l
```

图 14-88 查询磁盘分区和磁盘设备

```
root@localhost:~  
File Edit View Search Terminal Help  
[root@localhost ~]# fdisk -l  
  
Disk /dev/xvda: 10.7 GB, 10737418240 bytes, 20971520 sectors  
Units = sectors of 1 * 512 = 512 bytes  
Sector size (logical/physical): 512 bytes / 512 bytes  
I/O size (minimum/optimal): 512 bytes / 512 bytes  
Disk label type: dos  
Disk identifier: 0x000ba575  
  
    Device Boot      Start         End      Blocks   Id  System  
/dev/xvda1    *          2048     2099199     1048576   83   Linux  
/dev/xvda2           2099200     16777215     7339008   83   Linux  
/dev/xvda3           16777216     20971519     2097152   82   Linux swap / Solaris  
  
Disk /dev/xvdb: 21.5 GB, 21474836480 bytes, 41943040 sectors  
Units = sectors of 1 * 512 = 512 bytes  
Sector size (logical/physical): 512 bytes / 512 bytes  
I/O size (minimum/optimal): 512 bytes / 512 bytes
```

对于Linux弹性云服务器，查询到的磁盘分区与实际使用的磁盘设备名称完全对应，如表14-3所示。

表 14-3 磁盘分区和磁盘设备的映射关系

磁盘分区	磁盘设备
xvda	xvda
xvdb	xvdb
xvdc	xvdc
xvdd	xvdd
xvde	xvde
xvdf	xvdf
xvdg	xvdg
xvdh	xvdh
xvdi	xvdi
xvdj	xvdj
xvdk	xvdk
xvdl	xvdl
xvdm	xvdm
xvdn	xvdn
xvdo	xvdo
xvdp	xvdp

磁盘分区	磁盘设备
xvdq	xvdq
xvdr	xvdr
xvds	xvds
xvdt	xvdt
xvdu	xvdu
xvdv	xvdv
xvdw	xvdw
xvdx	xvdx

### 14.5.5 怎样配置 Windows 弹性云服务器的虚拟内存？

开启弹性云服务器的虚拟内存后，会导致内存I/O性能下降。当弹性云服务器内存不足时，建议通过[变更ECS规格（vCPU和内存）](#)操作来扩大内存。若因业务需要，必须开启虚拟内存，请参见本节内容进行配置。

#### 说明

内存使用率已经非常高，同时I/O性能也不是很好的情况下，如果配置虚拟内存会起到反向效果：不仅系统性能提升较小（因为内存使用率已经非常高了），而且由于频繁的内存切换操作，会产生大量额外的I/O操作，导致I/O性能进一步降低，反而最终降低了系统总体性能。

本节内容适用于Windows 2008及以上版本操作系统的弹性云服务器。

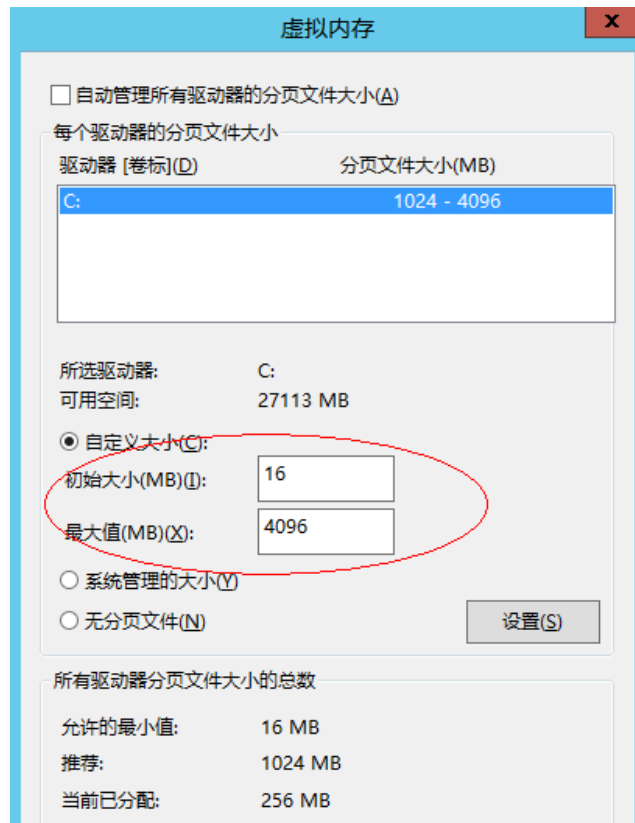
1. 右键单击“计算机”，选择“属性”。
2. 在左侧导航栏，选择“高级系统设置”。  
系统打开“系统属性”窗口。
3. 选择“高级”页签，并单击“性能”栏的“设置”。  
系统打开“性能选项”窗口。

图 14-89 性能选项



4. 选择“高级”页签，并单击“处理器计划”栏的“后台服务”。
5. 单击“虚拟内存”栏的“更改”。  
系统打开“虚拟内存”窗口。
6. 根据业务需求配置虚拟内存：
  - 自动管理所有驱动器的分页文件大小：取消勾选。
  - 驱动器：选择虚拟内存文件存放的驱动器。  
建议您不要选择系统盘来存放虚拟内存，具体根据自己的服务器硬盘情况来选择。
  - 自定义大小：选择“自定义大小”，并输入“初始大小”与“最大值”。  
请合理考虑系统蓝屏时产生Memory.dmp时的需求，建议初始大小16MB，最大值4096MB。

图 14-90 虚拟内存



7. 单击“设置”，然后单击右下角的“确定”完成虚拟内存配置。
8. 重启弹性云服务器，使配置生效。

## 14.5.6 如何将扩容系统盘的空白分区在线扩容到末尾的 root 分区？

### 操作场景

弹性云服务器创建成功后，如果发现系统盘分区的容量大小和实际创建的系统盘大小不一致，可以将空白分区扩容到根分区，扩容系统盘的空白空间。

本节介绍了怎样将空白分区在线扩容到末尾的root分区。

### 操作步骤

以CentOS 6.5 64bit 50GB系统盘为例，root分区在最末尾分区(e.g: /dev/xvda1: swap,/dev/xvda2: root)的扩容场景。

1. 执行以下命令，查询当前弹性云服务器的分区情况。

```
parted -l /dev/xvda
```

```
[root@sluo-ecs-5e7d ~]# parted -l /dev/xvda
Disk /dev/xvda: 53.7GB
Sector size (logical/physical): 512B/512B
Partition Table: msdos
```

```
Number Start End Size Type File system Flags
 1 1049kB 4296MB 4295MB primary linux-swaps(v1)
 2 4296MB 42.9GB 38.7GB primary ext4 boot
```

2. 执行以下命令，获取文件系统类型、UUID。

**blkid**

```
/dev/xvda1: UUID="25ec3bdb-ba24-4561-bcdc-802edf42b85f" TYPE="swap"  
/dev/xvda2: UUID="1a1ce4de-e56a-4e1f-864d-31b7d9dfb547" TYPE="ext4"
```

3. 执行以下命令，安装growpart工具。

工具growpart可能集成在cloud-utils-growpart/cloud-utils/cloud-initramfs-tools/cloud-init包里，可以直接执行命令**yum install cloud-\***确保growpart命令可用即可。

**yum install cloud-utils-growpart**

4. 执行以下命令，使用工具growpart将第二分区的根分区进行扩容。

**growpart /dev/xvda 2**

```
[root@sluo-ecs-5e7d ~]# growpart /dev/xvda 2  
CHANGED: partition=2 start=8390656 old: size=75495424 end=83886080 new:  
size=96465599,end=104856255
```

5. 执行以下命令，检查在线扩容是否成功。

**parted -l /dev/xvda**

```
[root@sluo-ecs-5e7d ~]# parted -l /dev/xvda  
Disk /dev/xvda: 53.7GB  
Sector size (logical/physical): 512B/512B  
Partition Table: msdos
```

Number	Start	End	Size	Type	File system	Flags
1	1049kB	4296MB	4295MB	primary	linux-swap(v1)	
2	4296MB	53.7GB	49.4GB	primary	ext4	boot

6. 执行以下命令，扩容文件系统。

**resize2fs -f \$分区名**

假设分区名为/dev/xvda2，则命令如下：

```
[root@sluo-ecs-a611 ~]# resize2fs -f /dev/xvda2  
resize2fs 1.42.9 (28-Dec-2013)  
Filesystem at /dev/xvda2 is mounted on /; on-line resizing required  
old_desc_blocks = 3, new_desc_blocks = 3  
....  
[root@sluo-ecs-a611 ~] # df -hT //检查文件系统扩容情况
```

## 14.5.7 如何将扩容系统盘的空白分区在线扩容到非末尾的 root 分区？

### 操作场景

弹性云服务器创建成功后，如果发现系统盘分区的容量大小和实际创建的系统盘大小不一致，可以将空白分区扩容到根分区，扩容系统盘的空白空间。

本节介绍了怎样将空白分区在线扩容到非末尾的root分区。

### 操作步骤

以CentOS 6.5 64bit 100GB系统盘为例，root分区在非末尾分区(e.g: /dev/xvda1: root,/dev/xvda2: swap)的扩容场景。

1. 执行以下命令，查询扩容系统盘的分区情况。

**parted -l /dev/xvda**

```
[root@sluo-ecs-a611 ~]# parted -l /dev/xvda  
Disk /dev/xvda: 107GB  
Sector size (logical/physical): 512B/512B
```

```
Partition Table: msdos
Disk Flags:

Number Start End   Size  Type  File system  Flags
  1    1049kB 41.0GB 40.9GB primary ext4         boot
  2    41.0GB 42.9GB 2000MB primary linux-swap(v1)
```

其中，第一分区是根分区，第二分区是swap分区。

2. 执行以下命令，编辑fstab分区表，删除swap分区的挂载信息。

- a. 执行以下命令，查询fstab分区表。

```
tail -n 3 /etc/fstab
```

```
[root@sluo-ecs-a611 ~]# tail -n 3 /etc/fstab
#
UUID=7c4fce5d-f8f7-4ed6-8463-f2bd22d0ddea /          ext4 defaults 1 1
UUID=5de3cf2c-30c6-4fb2-9e63-830439d4e674 swap        swap defaults 0 0
```

- b. 执行以下命令，编辑fstab分区表，并删除swap分区的挂载信息。

```
vi /etc/fstab
```

```
tail -n 3 /etc/fstab
```

```
[root@sluo-ecs-a611 ~]# vi /etc/fstab
[root@sluo-ecs-a611 ~]# tail -n 3 /etc/fstab
#
UUID=7c4fce5d-f8f7-4ed6-8463-f2bd22d0ddea /          ext4 defaults 1 1
```

3. 执行以下命令，关闭swap。

```
swapoff -a
```

4. 删除swap分区。

- a. 执行以下命令，查看分区。

```
parted /dev/xvda
```

```
[root@sluo-ecs-a611 ~]# parted /dev/xvda
GNU Parted 3.1
Using /dev/xvda
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) help
align-check TYPE N                check partition N for TYPE(min|opt) alignment
help [COMMAND]                    print general help, or help on COMMAND
mklabel,mktable LABEL-TYPE        create a new disklabel (partition table)
mkpart PART-TYPE [FS-TYPE] START END make a partition
name NUMBER NAME                   name partition NUMBER as NAME
print [devices|free|list,all|NUMBER] display the partition table, available devices, free space,
all found partitions, or a
particular partition
quit                                exit program
rescue START END                   rescue a lost partition near START and END
rm NUMBER                           delete partition NUMBER
select DEVICE                       choose the device to edit
disk_set FLAG STATE                 change the FLAG on selected device
disk_toggle [FLAG]                 toggle the state of FLAG on selected device
set NUMBER FLAG STATE               change the FLAG on partition NUMBER
toggle [NUMBER [FLAG]]             toggle the state of FLAG on partition NUMBER
unit UNIT                           set the default unit to UNIT
version                             display the version number and copyright information of GNU
Parted
(parted)
```

- b. 按“p”。

```
Disk /dev/xvda: 107GB
Sector size (logical/physical): 512B/512B
Partition Table: msdos
Disk Flags:

Number Start End   Size  Type  File system  Flags
  1    1049kB 41.0GB 40.9GB primary ext4         boot
```



```
2 41.0GB 42.9GB 2000MB primary linux-swap(v1)
```

- c. 执行以下命令，删除swap分区。

```
rm 2
```

```
(parted) rm2
```

- d. 按“p”。

```
(parted) p
```

```
Disk /dev/xvda: 107GB
```

```
Sector size (logical/physical): 512B/512B
```

```
Partition Table: msdos
```

```
Disk Flags:
```

```
Number Start End Size Type File system Flags  
1 1049kB 41.0GB 40.9GB primary ext4 boot
```

- e. 执行以下命令，退出fstab分区表。

```
quit
```

```
(parted) quit
```

```
Information: You may need to update /etc/fstab.
```

5. 执行以下命令，查询删除swap分区后的分区情况。

```
parted -l /dev/xvda
```

```
[root@sluo-ecs-a611 ~]# parted -l /dev/xvda
```

```
Disk /dev/xvda: 107GB
```

```
Sector size (logical/physical): 512B/512B
```

```
Partition Table: msdos
```

```
Disk Flags:
```

```
Number Start End Size Type File system Flags  
1 1049kB 41.0GB 40.9GB primary ext4 boot
```

6. 执行以下命令，安装growpart工具。

工具growpart可能集成在cloud-utils-growpart/cloud-utils/cloud-initramfs-tools/cloud-init包里，可以直接执行命令**yum install cloud-\***确保growpart命令可用即可。

```
yum install cloud-utils-growpart
```

7. 执行以下命令，使用工具growpart将第一分区的root分区进行扩容。

```
growpart /dev/xvda 1
```

```
[root@sluo-ecs-a611 ~]# growpart /dev/xvda 1
```

```
CHANGED: partition=1 start=2048 old: size=79978496 end=79980544 new:  
size=209710462,end=209712510
```

8. 执行以下命令，检查在线扩容是否成功。

```
[root@sluo-ecs-a611 ~]# parted -l /dev/xvda
```

```
Disk /dev/xvda: 107GB
```

```
Sector size (logical/physical): 512B/512B
```

```
Partition Table: msdos
```

```
Disk Flags:
```

```
Number Start End Size Type File system Flags  
1 1049kB 107GB 107GB primary ext4 boot
```

9. 执行以下命令，扩容文件系统。

```
resize2fs -f $分区名
```

假设分区名为/dev/xvda1，则命令如下：

```
[root@sluo-ecs-a611 ~]# resize2fs -f /dev/xvda1
```

```
resize2fs 1.42.9 (28-Dec-2013)
```

```
Filesystem at /dev/xvda1 is mounted on /; on-line resizing required
```

```
old_desc_blocks = 3, new_desc_blocks = 3
```

```
....
```

```
[root@sluo-ecs-a611 ~] # df -hT //检查文件系统扩容情况
```

## 14.5.8 一台弹性云服务器可以挂载多块磁盘吗？

可以。平台近期对磁盘功能进行了升级，对于系统升级后创建的弹性云服务器，系统最多支持挂载60块磁盘。具体如下：

- 创建弹性云服务器时，您可以为弹性云服务器添加24块磁盘。
- 弹性云服务器创建成功后，可以最多添加60块磁盘。

表 14-4 新创建弹性云服务器支持挂载的磁盘块数

实例类型	VBD盘块数 (max)	SCSI盘块数 (max)	约束
XEN实例	60	59	VBD盘+SCSI盘 ≤ 60块（本地盘不计入数量约束） 本地盘数量参考具体实例规格。
KVM实例（除D3型）	24	59	VBD盘+SCSI盘 ≤ 60块（本地盘不计入数量约束） 本地盘数量参考具体实例规格。
D3型	24	30	VBD盘+SCSI盘 ≤ 54块（本地盘不计入数量约束） 本地盘数量参考具体实例规格。

### 说明

- 弹性云服务器的系统盘为VBD盘，因此，SCSI盘个数为59块。
- 对于D系列KVM实例，本地盘会占用2个SCSI控制器，即会占用30个SCSI盘符。因此，最多可以挂载30块SCSI云硬盘。

对于系统升级前创建的弹性云服务器，磁盘挂载数量保持原配额，如表14-5所示。

表 14-5 已创建弹性云服务器支持挂载的磁盘块数

实例类型	VBD盘块数 (max)	SCSI盘块数 (max)	本地盘块数 (max)	约束
XEN实例	60	59	59	VBD盘+SCSI盘+本地盘的个数 ≤ 60块
KVM实例	24	23	59	VBD盘+SCSI盘的个数 ≤ 24块

如需支持挂载60块磁盘，需开启高级磁盘功能，具体操作请参见[开启高级磁盘功能](#)。

## 怎样查看我的弹性云服务器是升级前/升级后创建的？

1. 登录控制台。
2. 选择“计算 > 弹性云服务器”。
3. 单击弹性云服务器名称，进入弹性云服务器详情页。
4. 选择“云硬盘”页签。
5. 查看界面提示的弹性云服务器可挂载磁盘个数，判断弹性云服务器支持挂载的磁盘总数。
  - 如果支持挂载的磁盘总数为24块（含系统盘），说明当前弹性云服务器是系统升级前创建的。此时，您可以根据界面提示，开启磁盘高级功能，开启后，弹性云服务器最多支持挂载60块磁盘，具体操作请参见[开启高级磁盘功能](#)。
  - 如果支持挂载的磁盘总数为60块（含系统盘），说明当前弹性云服务器是系统升级后创建的。

### 14.5.9 弹性云服务器挂载磁盘时有什么限制？

- 待挂载的云硬盘与弹性云服务器属于同一可用区。
- 如果是非共享盘，待挂载的云硬盘为“可用”状态。  
如果是共享盘，待挂载的云硬盘为“正在使用”状态或“可用”状态。
- 弹性云服务器的状态为“运行中”或“关机”。
- 处于冻结状态的云硬盘，不支持挂载给弹性云服务器。
- 支持挂载SCSI类型云硬盘的弹性云服务器，请参见[哪些弹性云服务器支持挂载SCSI类型的云硬盘？](#)。

### 14.5.10 哪些弹性云服务器支持挂载 SCSI 类型的云硬盘？


所有类型的弹性云服务器都支持挂载SCSI类型的云硬盘。

### 14.5.11 控制台与弹性云服务器内部之间磁盘挂载点的对应关系

#### 操作场景

登录弹性云服务器查询磁盘设备信息，发现磁盘的设备名称与控制台上显示的挂载点不一致，不清楚磁盘具体挂载在哪个设备上或磁盘对应的逻辑卷标识。本节操作介绍如何根据控制台上显示的磁盘信息，查询该磁盘在弹性云服务器内对应的设备名称。

#### 在管理控制台查看云服务器的磁盘标识

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表栏，单击目标云服务器的名称。  
系统进入弹性云服务器详情页。
4. 选择“云硬盘”页签，并单击  展开磁盘详情。
5. 查看磁盘的“设备类型”和“设备标识”。

### 📖 说明

如果界面上当前未显示参数“设备标识”，请尝试关闭弹性云服务器，并重新开机。

#### - KVM实例

- 如果“设备类型”为VBD，您可以使用“Serial number”或BDF查询磁盘挂载点。  
(推荐) SerialNumber请参考[使用SerialNumber查询磁盘标识 \(Windows\)](#)和[使用SerialNumber查询磁盘挂载点 \(Linux\)](#)  
BDF请参考[使用BDF查询磁盘挂载点 \(Linux\)](#) (Windows操作系统不支持使用BDF查询磁盘标识。)
- 如果“设备类型”为SCSI，您可以使用wwn查询磁盘挂载点，请参考[使用wwn查询磁盘标识 \(Windows\)](#)和[使用wwn查询磁盘挂载点 \(Linux\)](#)。

## 使用 SerialNumber 查询磁盘标识 (Windows)

如果控制台上显示的是SerialNumber，有以下两种方法查看对应的设备名称。

### cmd命名查询方式：

1. 在Windows操作系统中以管理员身份启动cmd，并执行以下命令。

```
wmic diskdrive get serialnumber
```

```
wmic path win32_physicalmedia get SerialNumber
```

```
wmic path Win32_DiskDrive get SerialNumber
```

### 📖 说明

SerialNumber是磁盘UUID的前20位。

例如：控制台上VBD盘对应的SerialNumber为“97c876c0-54b3-460a-b”，使用以下任一命令查询磁盘SerialNumber。

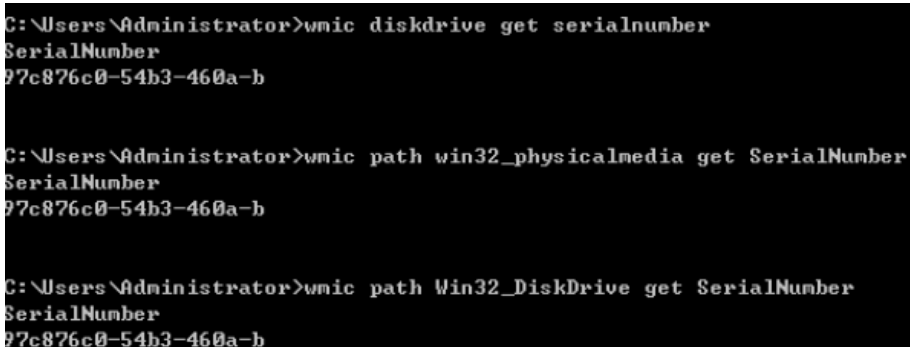
```
wmic diskdrive get serialnumber
```

```
wmic path win32_physicalmedia get SerialNumber
```

```
wmic path Win32_DiskDrive get SerialNumber
```

回显信息如下所示：

图 14-91 获取磁盘 SerialNumber



```
C:\Users\Administrator>wmic diskdrive get serialnumber
SerialNumber
97c876c0-54b3-460a-b

C:\Users\Administrator>wmic path win32_physicalmedia get SerialNumber
SerialNumber
97c876c0-54b3-460a-b

C:\Users\Administrator>wmic path Win32_DiskDrive get SerialNumber
SerialNumber
97c876c0-54b3-460a-b
```

2. 执行以下命令查看SerialNumber对应的磁盘。

## wmic diskdrive get Name, SerialNumber

图 14-92 查看 SerialNumber 对应的磁盘

```
C:\Users\Administrator>wmic diskdrive get Name, SerialNumber
Name                SerialNumber
\\.\PHYSICALDRIVE0  97c876c0-54b3-460a-b
```

### PowerShell命名查询方式:

1. 在Windows操作系统中以管理员身份启动PowerShell。
2. 执行以下命令查看逻辑卷对应的磁盘。

- Windows 2012及以上版本的操作系统

- i. 执行以下命令查看逻辑卷对应的磁盘

```
Get-CimInstance -ClassName Win32_LogicalDiskToPartition |
select Antecedent, Dependent |fl
```

图14-93所示查询到的逻辑卷C对应的磁盘是Disk 0

- ii. 执行以下命令查看SerialNumber与磁盘的对应关系。

```
Get-Disk |select Number, SerialNumber
```

图14-93所示SerialNumber “97c876c0-54b3-460a-b” 对应的是Disk 0。

图 14-93 查看逻辑卷对应的磁盘

```
PS C:\Users\Administrator> Get-CimInstance -ClassName Win32_LogicalDiskToPartition |select Antecedent, Dependent
Antecedent : Win32_DiskPartition (DeviceID = "Disk #0, Partition #1")
Dependent  : Win32_LogicalDisk (DeviceID = "C:")

PS C:\Users\Administrator> Get-Disk |select Number, SerialNumber
Number SerialNumber
-----
0       97c876c0-54b3-460a-b
1       dswfal6520d39517815206127
```

- Windows 2012以下版本的操作系统

- i. 执行以下命令查看逻辑卷对应的磁盘

```
Get-WmiObject -Class Win32_PhysicalMedia |select Tag,
Serialnumber
```

- ii. 执行以下命令查看SerialNumber与磁盘的对应关系。

```
Get-WmiObject -Class Win32_LogicalDiskToPartition |select
Antecedent, Dependent |fl
```

## 使用 SerialNumber 查询磁盘挂载点 (Linux)

如果控制台上显示的是SerialNumber，您可以使用以下任一命令查看对应的设备名称。

```
udevadm info --query=all --name=/dev/xxx | grep ID_SERIAL
```

```
ll /dev/disk/by-id/*
```

### 说明

SerialNumber是磁盘UUID的前20位。

例如：控制台上VBD盘对应的SerialNumber为“62f0d06b-808d-480d-8”，可以使用以下任一命令查询设备名称。

```
udevadm info --query=all --name=/dev/vdb | grep ID_SERIAL
```

```
ll /dev/disk/by-id/*
```

回显信息如下所示：

```
[root@ecs-ab63 ~]# udevadm info --query=all --name=/dev/vdb | grep ID_SERIAL
E: ID_SERIAL=62f0d06b-808d-480d-8
[root@ecs-ab63 ~]# ll /dev/disk/by-id/*
lrwxrwxrwx 1 root root 9 Dec 30 15:56 /dev/disk/by-id/virtio-128d5bfd-f215-487f-9 -> ../vda
lrwxrwxrwx 1 root root 10 Dec 30 15:56 /dev/disk/by-id/virtio-128d5bfd-f215-487f-9-part1 -> ../vda1
lrwxrwxrwx 1 root root 9 Dec 30 15:56 /dev/disk/by-id/virtio-62f0d06b-808d-480d-8 -> ../vdb
```

系统回显即为磁盘设备名称（即“/dev/vdb”）。

## 使用 BDF 查询磁盘挂载点（Linux）

1. 执行以下命令，使用BDF查看对应的设备名称。

```
ll /sys/bus/pci/devices/BDF磁盘标识/virtio*/block
```

例如：控制台上VBD盘对应的BDF磁盘标识为“0000:02:02.0”，执行以下命令查询设备名称。

```
ll /sys/bus/pci/devices/0000:02:02.0/virtio*/block
```

回显信息如下所示：

```
[root@ecs-ab63 ~]# ll /sys/bus/pci/devices/0000:02:02.0/virtio*/block
total 0
drwxr-xr-x 8 root root 0 Dec 30 15:56 vdb
```

系统回显即为磁盘设备名称（即“/dev/vdb”）。

## 使用 wwn 查询磁盘标识（Windows）

1. 根据[在管理控制台查看云服务器的磁盘标识](#)的步骤在控制台查询磁盘标识。
2. 手动转换wwn号。

例如查询到的wwn号为：68886030000**3252ffa**16520d39517815

- a. 截取倒数的第21位-倒数17位，并标记为A1，本例中A1为**3252f**。
- b. 将A1从16位数字转换为10位数字，标记为B1。

**3252f** ( 16 ) > 206127(10)

本例中B1为206127。

3. 在Windows操作系统中以管理员身份启动PowerShell。
4. 执行以下命令。  
**Get-CimInstance Win32\_DiskDrive | Select-Object DeviceID, SerialNumber**
5. 在回显信息中找到SerialNumber是以B1结尾的磁盘即为对应的wwn对应的磁盘。

图 14-94 查询 B1 结尾的磁盘信息

```
PS C:\Users\Administrator> Get-CimInstance Win32_DiskDrive | Select-Object DeviceID, SerialNumber
DeviceID                               SerialNumber
-----
\\.\PHYSICALDRIVE0                     97c876c0-34b3-460a-b
\\.\PHYSICALDRIVE1                     dswfa16520d39517815206127
```

## 使用 wwn 查询磁盘挂载点 (Linux)

1. 以root账号登录弹性云服务器。
2. 执行以下命令，查询磁盘设备名称。

```
ll /dev/disk/by-id |grep WWM|grep scsi-3
```

例如：在控制台查询到的wwn号为6888603000008b32fa16688d09368506，那么执行命令如下。

```
ll /dev/disk/by-id |grep 6888603000008b32fa16688d09368506|grep scsi-3
```

回显信息如下：

```
[root@host-192-168-133-148 block]# ll /dev/disk/by-id/ |grep 6888603000008b32fa16688d09368506 |
grep scsi-3
lrwxrwxrwx 1 root root 9 May 21 20:22 scsi-36888603000008b32fa16688d09368506 -> ../../sda
```

## 14.5.12 挂载 SCSI 盘的 Linux 弹性云服务器，重启失败怎么办？

### 问题描述

对于挂载了SCSI类型云硬盘的Linux弹性云服务器，如果在/etc/fstab中配置了SCSI磁盘开机自动挂载，且使用的是磁盘的盘符（如/dev/sdb）进行配置，弹性云服务器可能重启失败。

### 可能原因

SCSI磁盘的分配与磁盘的槽位号、弹性云服务器中可用的盘符有关。弹性云服务器内部，每加载一个磁盘都按顺序分配空闲的盘符。弹性云服务器启动时，按照槽位号顺序加载磁盘，所以槽位号和盘符的顺序是一一对应的。

在线卸载弹性云服务器的SCSI磁盘后，磁盘的槽位号顺序有可能发生改变，导致重启后磁盘的盘符也发生改变，槽位号和盘符无法对应，重启失败。

### 处理方法

1. 登录Linux弹性云服务器。
2. 执行以下命令，切换至root用户。  
**sudo su -**
3. 执行以下命令，根据SCSI盘的盘符，查询对应的SCSI ID。

```
ll /dev/disk/by-id|grep 磁盘盘符
```

假设SCSI磁盘的盘符为/dev/sdb，则命令行如下：

```
ll /dev/disk/by-id|grep sdb
```

```
CNA64_22:/opt/galax/eucalyptus/ecs_scripts # ll /dev/disk/by-id|grep sdb
lrwxrwxrwx 1 root root 9 Dec 6 11:26 scsi-3688860300001436b005014f890338280 -> ../../sdb
lrwxrwxrwx 1 root root 9 Dec 6 11:26 wwn-0x688860300001436b005014f890338280 -> ../../sdb
```

4. 修改/etc/fstab文件，将SCSI盘的盘符（如/dev/sdb）修改为对应的SCSI ID。

```
/dev/disk/by-id/SCSI ID
```

假设3中查询到的SCSI ID为scsi-3688860300001436b005014f890338280，则用以下内容替换/dev/sdb：

```
/dev/disk/by-id/scsi-3688860300001436b005014f890338280
```

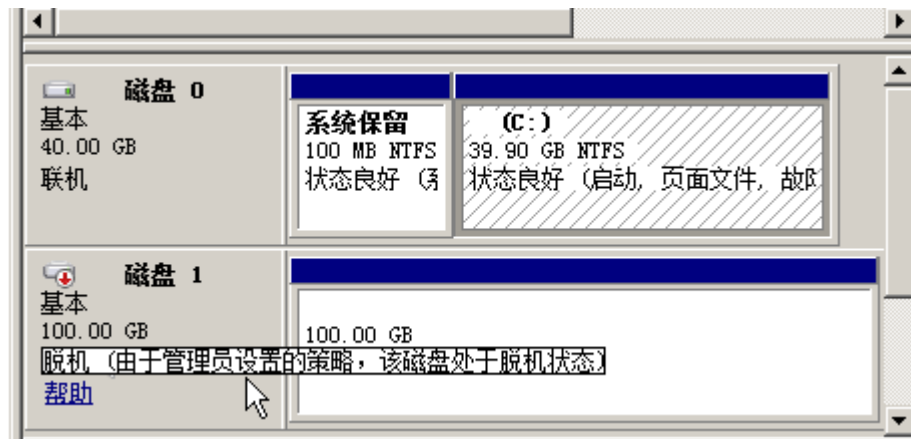


## 14.5.13 磁盘脱机怎么办？

### 问题描述

Windows操作系统的云服务器由于管理员设置的策略，该磁盘处于脱机状态，提示：由于管理员设置的策略，该磁盘处于脱机状态，如图14-95所示。

图 14-95 磁盘脱机



### 问题原因

Windows操作系统SAN策略分为三种类型：OnlineAll，OfflineShared，OfflineInternal

表 14-6 Windows 操作系统 SAN 策略类型

类型	说明
OnlineAll	表示所有新发现磁盘都置于在线模式。
OfflineShared	表示所有共享总线上（比如FC、iSCSI）的新发现磁盘都置于离线模式，非共享总线上的磁盘都置于在线模式。
OfflineInternal	表示所有新发现磁盘都置于离线模式。

对于部分版本的Windows操作系统（如Windows 2008/2012 企业版、数据中心版），默认是共享离线OfflineShared模式。

### 处理方法

使用DiskPart磁盘分区管理工具来查询、设置云服务器的SAN策略，将其修改为OnlineAll类型。

1. 登录Windows云服务器。
2. 按快捷键“Win+R”，运行cmd.exe。
3. 执行以下命令，进入DiskPart。

```
diskpart
```



4. 执行以下命令，查看云服务器当前的SAN策略。  
**san**
  - 如果SAN策略为OnlineAll，请执行**exit**命令退出DiskPart。
  - 否，请执行**5**。
5. 执行以下命令，修改云服务器SAN策略为OnlineAll。  
**san policy=onlineall**
6. （可选）如需永久生效，您可以将修改过SAN策略的云服务器制作作为私有镜像。由该私有镜像创建的云服务器，添加的磁盘只需执行初始化操作，磁盘默认是联机状态。

## 14.5.14 重启后盘符发生变化怎么办？

### 问题现象

Linux弹性云服务器在线卸载云硬盘后重新挂载云硬盘，或在线卸载云硬盘后重启弹性云服务器，可能会存在挂载前后盘符发生变化的情况。

### 问题原因

当Linux系统的弹性云服务器管理多块磁盘时，会按照磁盘加载顺序分配盘符，依次为磁盘命名，例如：`/dev/vda1`，`/dev/vdb1`，`/dev/vdc1`等。

由于Linux系统的盘符分配机制，当弹性云服务器卸载云硬盘后重新挂载，或者卸载云硬盘后重启云服务器时，可能会存在出现盘符发生变化的情况。

例如，某个云服务器有`/dev/vda1`，`/dev/vdb1`，`/dev/vdc1`三块磁盘，在`/etc/fstab`中的挂载参数如下所示：

#### cat /etc/fstab

```
UUID=b9a07b7b-9322-4e05-ab9b-14b8050bdc8a / ext4 defaults 0 1
/dev/vdb1 /data1 ext4 defaults 0 0
/dev/vdc1 /data2 ext4 defaults 0 0
```

当卸载原有磁盘`/dev/vdb1`，并重启云服务器后，原有的`/dev/vdc1`会变为`/dev/vdb1`并挂载到挂载点`/data1`，而挂载点`/data2`则无磁盘可以挂载。

磁盘盘符的变化会影响依赖盘符工作的应用程序的正常使用，为解决这个问题，建议采用UUID（universally unique identifier）来替代`/dev/vdx`，UUID是Linux系统为磁盘分区提供的唯一的标识字符串。

### 处理方法

1. 登录弹性云服务器。
2. 执行以下命令，查询磁盘分区的UUID。

#### blkid 磁盘分区

以查询磁盘分区“`/dev/vdb1`”的UUID为例：

#### blkid /dev/vdb1

回显类似如下信息：

```
[root@ecs-test-0001 ~]# blkid /dev/vdb1
/dev/vdb1: UUID="b9a07b7b-9322-4e05-ab9b-14b8050cd8cc" TYPE="ext4"
```

表示“`/dev/vdb1`”的UUID。

3. 执行以下命令，使用VI编辑器打开“fstab”文件。

```
vi /etc/fstab
```

4. 按“i”，进入编辑模式。
5. 将光标移至文件末尾，按“Enter”，添加如下内容。

```
UUID=b9a07b7b-9322-4e05-ab9b-14b8050cd8cc /data1 ext4 defaults 0 0
```

其中，

- UUID=b9a07b7b-9322-4e05-ab9b-14b8050cd8cc：磁盘的分区UUID。
- /data1：磁盘分区的挂载目录，可以通过df -TH命令查询。
- ext4：磁盘分区的文件系统格式，可以通过df -TH命令查询。
- defaults：磁盘分区的挂载选项，此处通常设置为defaults即可。
- 0：Linux dump备份选项。
  - 0表示不使用Linux dump备份。现在通常不使用dump备份，此处设置为0即可。
  - 1表示使用Linux dump备份。
- 0：fsck选项，即开机时是否使用fsck检查磁盘。
  - 0表示不检验。
  - 挂载点为（/）根目录的分区，此处必须填写1。  
根分区设置为1，其他分区只能从2开始，系统会按照数字从小到大依次检查下去。

6. 重复执行步骤2~步骤5，替换“/dev/vdc1”的UUID。
7. 再次执行以下命令，查看磁盘挂载参数：

```
cat /etc/fstab
```

回显信息如下所示：

```
UUID=b9a07b7b-9322-4e05-ab9b-14b8050bdc8a / ext4 defaults 0 1
UUID=b9a07b7b-9322-4e05-ab9b-14b8050cd8cc /data1 ext4 defaults 0 0
UUID=b9a07b7b-9322-4e05-ab9b-14b8050ab6bb /data2 ext4 defaults 0 0
```

## 14.5.15 用户误删除 Tools 后，如何获取数据盘信息？

Linux操作系统下，用户由于误操作卸载弹性云服务器上的Tools，会对非PVOPS系统的磁盘和网卡产生影响，导致系统无法发现数据盘。此时，用户可通过新建一个弹性云服务器，将原数据盘挂载至新创建的弹性云服务器下，从而获取丢失的数据盘信息。具体操作方法如下：

1. 在管理控制台，创建一个新的弹性云服务器。

### 说明

请确保新创建的弹性云服务器和被误删除Tools的原弹性云服务器在同一个可用区，且相关参数配置与原弹性云服务器保持一致。

2. （可选）在弹性云服务器列表中，选中被误删除Tools的原弹性云服务器（以下简称原弹性云服务器），选择“操作 > 更多 > 关机”，勾选“强制关机”并单击“确定”确认执行原弹性云服务器的强制关机操作。

手动刷新弹性云服务器列表，待“状态”栏更新为“关机”，原弹性云服务器关机成功。

### 📖 说明

对于支持在线卸载数据盘的弹性云服务器，可在弹性云服务器运行状态下直接执行卸载数据盘的操作。

3. 查看原弹性云服务器上挂载的数据盘信息。

### 📖 说明

如果原弹性云服务器上挂载了多个数据盘，则需要重复操作4-6，分别执行各个数据盘的卸载与挂载操作。

4. 单击原弹性云服务器上挂载的数据盘，系统自动跳转至“Elastic Volume Service”界面。

5. 选中待卸载的数据盘，选择“Operation > Detach”，勾选原弹性云服务器并单击“OK”，确认卸载原弹性云服务器上的数据盘。

手动刷新云硬盘列表，待“Status”栏更新为“Available”，原弹性云服务器上的数据盘卸载成功。

6. 选中已卸载的数据盘，选择“Operation > Attach”，勾选新创建的弹性云服务器名称，并选择磁盘的“Select Mount Point”，单击“Attach”，将卸载的数据盘挂载至新创建的弹性云服务器。

手动刷新云硬盘列表，待“Status”栏更新为“In-use”，挂载磁盘成功。磁盘挂载成功后，用户可登录管理控制台，在新创建的弹性云服务器磁盘信息中，查看新挂载的数据盘。

## 14.6 网络配置

### 14.6.1 不同账号下跨 VPC 的弹性云服务器内网是否可以互通？

不同账号下跨VPC的弹性云服务器内网是不通的。

### 14.6.2 我购买的云服务器是否在同一子网？

由于您可以自定义网络，所以云服务器是否在一个子网，完全由您来控制。

### 14.6.3 如何配置端口映射？

#### 问题描述

公网访问“弹性云服务器1”的弹性公网IP和某个端口，可以自动跳转到“弹性云服务器2”的弹性公网IP和某个端口。

#### Windows 操作系统

假定需要通过“弹性云服务器1”（192.168.10.43）的8080端口连接“弹性云服务器2”（192.168.10.222）的18080端口，则需要在弹性云服务器1执行如下操作。

#### 📖 说明

请确保云服务器的安全组、防火墙已放通对应的端口。

1. 打开cmd窗口执行命令。本例以Windows 2012操作系统云服务器为例。

```
netsh interface portproxy add v4tov4 listenaddress=192.168.10.43  
listenport=8080 connectaddress=192.168.10.222 connectport=18080
```

如果想取消上面配置的端口转发，可执行如下命令。

```
netsh interface portproxy delete v4tov4 listenaddress=192.168.10.43  
listenport=8080
```

2. 执行如下命令可查看服务器配置的全部端口映射。

```
netsh interface portproxy show v4tov4
```

图 14-96 Windows 操作系统云服务器端口映射

```
C:\Users\Administrator>netsh interface portproxy add v4tov4 listenaddress=192.168.10.43 listenport=8080 connectaddress=192.168.10.222 connectport=18080

C:\Users\Administrator>netsh interface portproxy show v4tov4

侦听 ipv4:          连接到 ipv4:
地址                端口                地址                端口
-----
192.168.10.43      8080                192.168.10.222     18080
```

## Linux 操作系统

以登录为例，登录“弹性云服务器1”的1080端口自动跳转访问“弹性云服务器2”的22端口。

弹性云服务器1的私网IP：192.168.72.10；弹性公网IP：123.xxx.xxx.456。

弹性云服务器2的私网IP：192.168.72.20。

### 说明

- 请确保云服务器的安全组、防火墙已放通对应的端口。
- 请确保已关闭“源/目的检查”。

在云服务器详情页面，选择“弹性网卡”页签，并展开，将“源/目的检查”选项设置为“OFF”。

默认情况下，“源/目的检查”状态为“启用”，系统会检查弹性云服务器发送的报文中源IP地址是否正确，否则不允许弹性云服务器发送该报文。这有助于防止伪装报文攻击，提升安全性。但在该场景中，这种保护机制会导致报文的发送者无法接收到返回的报文。因此，需设置“源/目的检查”状态为禁用。

### 步骤1 登录Linux弹性云服务器1。

1. 执行如下命令，修改文件。

```
vi /etc/sysctl.conf
```

2. 在文件中添加“net.ipv4.ip\_forward = 1”。
3. 执行如下命令，完成修改。

```
sysctl -p /etc/sysctl.conf
```

### 步骤2 在“iptables”的“nat”表中添加规则，执行如下命令，通过弹性云服务器1的1080端口映射到弹性云服务器2的22端口。

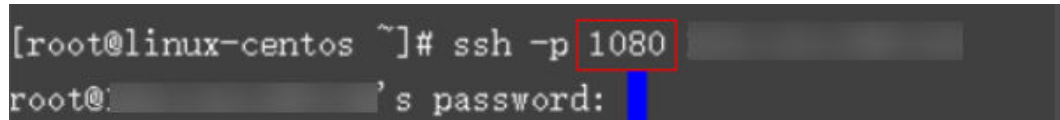
```
iptables -t nat -A PREROUTING -d 192.168.72.10 -p tcp --dport 1080 -j DNAT  
--to-destination 192.168.72.20:22
```

```
iptables -t nat -A POSTROUTING -d 192.168.72.20 -p tcp --dport 22 -j SNAT --to 192.168.72.10
```

**步骤3** 验证配置是否生效，执行如下命令，登录弹性云服务器1的1080端口。

```
ssh -p 1080 123.xxx.xxx.456
```

图 14-97 Linux 操作系统云服务器端口映射



输入密码后登录到弹性云服务器2，弹性云服务器2的主机名为ecs-inner。

图 14-98 登录到弹性云服务器 2



----结束

## 14.6.4 如何查看云服务器的 mac 地址？

本节介绍如何查看云服务器的mac地址。

### 说明

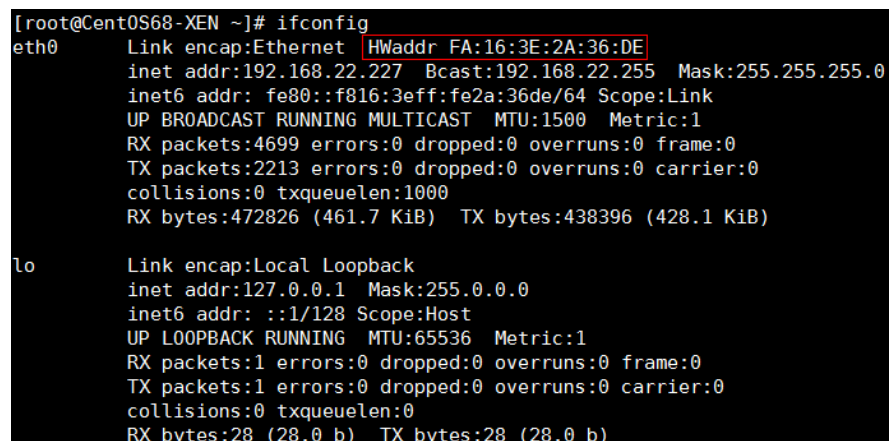
云服务器的mac地址不支持修改。

## Linux 操作系统（CentOS 6）

1. 登录Linux云服务器。
2. 执行以下命令，查看云服务器的mac地址。

```
ifconfig
```

图 14-99 查看 mac 地址



## Linux 操作系统 (CentOS 7)

1. 登录Linux云服务器。
2. 执行以下命令，查看云服务器的mac地址。

```
ifconfig
```

图 14-100 查看网卡信息

```
[root@ecs-683a ~]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.0.65 netmask 255.255.255.0 broadcast 192.168.0.255
    inet6 fe80::f816:3eff:fec3:46fc prefixlen 64 scopeid 0x20<link>
    ether fa:16:3e:c3:46:fc txqueuelen 1000 (Ethernet)
    RX packets 14457 bytes 20617950 (19.6 MiB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 1867 bytes 245185 (239.4 KiB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet [REDACTED] netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10<host>
    loop txqueuelen 1000 (Local Loopback)
    RX packets 0 bytes 0 (0.0 B)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 0 bytes 0 (0.0 B)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
```

3. 执行以下命令查看eth0的mac地址。

```
ifconfig eth0 |grep "ether"
```

图 14-101 查看 eth0 的 mac 地址

```
[root@ecs-683a ~]# ifconfig eth0 |grep "ether"
ether fa:16:3e:c3:46:fc txqueuelen 1000 (Ethernet)
[root@ecs-683a ~]#
```

4. 返回mac地址

```
ifconfig eth0 |grep "ether" |awk '{print $2}'
```

图 14-102 查看 eth0 的 mac 地址

```
[root@ecs-683a ~]# ifconfig eth0 |grep "ether" |awk '{print $2}'
fa:16:3e:c3:46:fc
[root@ecs-683a ~]#
```

## Windows 操作系统

1. 使用快捷键“Win+R”，打开“运行”窗口。
2. 在“打开”栏，输入“cmd”，单击“确定”。
3. 执行以下命令，查看云服务器的mac地址。

```
ipconfig /all
```





### 14.6.5 如何查看和修改 Linux 弹性云服务器的内核参数？

本文总结了常用的Linux内核参数，以及Linux内核参数的查看、修改方法。如果您的业务未受到影响，建议不要执行修改内核参数的操作。如需调整，请确保：

- 从实际需要出发，有相关数据的支撑。
- 了解每一个内核参数的含义，不同版本操作系统的内核参数可能不同。常用内核参数说明，请参见表14-7。
- 对弹性云服务器中的重要数据进行备份。

## 背景知识

表 14-7 Linux 常用内核参数说明

参数	说明
net.core.rmem_default	默认的TCP数据接收窗口大小（字节）。
net.core.rmem_max	最大的TCP数据接收窗口（字节）。
net.core.wmem_default	默认的TCP数据发送窗口大小（字节）。
net.core.wmem_max	最大的TCP数据发送窗口（字节）。
net.core.netdev_max_bac klog	在每个网络接口接收数据包的速率比内核处理这些包的速率快时，允许送到队列的数据包的最大数目。
net.core.somaxconn	定义了系统中每一个端口最大的监听队列的长度，这是个全局的参数。
net.core.optmem_max	表示每个套接字所允许的最大缓冲区的大小。
net.ipv4.tcp_mem	确定TCP栈应该如何反映内存使用，每个值的单位都是内存页（通常是 4KB） 第一个值是内存使用的下限； 第二个值是内存压力模式开始对缓冲区使用应用压力的上限； 第三个值是内存使用的上限。在这个层次上可以将报文丢弃，从而减少对内存的使用。对于较大的BDP可以增大这些值（注意：其单位是内存页而不是字节）。

参数	说明
net.ipv4.tcp_rmem	为自动调优定义socket使用的内存。 第一个值是为socket接收缓冲区分配的最少字节数； 第二个值是默认值（该值会被rmem_default覆盖），缓冲区在系统负载不重的情况下可以增长到这个值； 第三个值是接收缓冲区空间的最大字节数（该值会被rmem_max覆盖）。
net.ipv4.tcp_wmem	为自动调优定义socket使用的内存。 第一个值是为socket发送缓冲区分配的最少字节数； 第二个值是默认值（该值会被wmem_default覆盖），缓冲区在系统负载不重的情况下可以增长到这个值； 第三个值是发送缓冲区空间的最大字节数（该值会被wmem_max覆盖）。
net.ipv4.tcp_keepalive_time	TCP发送 keepalive 探测消息的间隔时间（秒），用于确认TCP连接是否有效。
net.ipv4.tcp_keepalive_intvl	探测消息未获得响应时，重发该消息的间隔时间（秒）。
net.ipv4.tcp_keepalive_probes	在认定TCP连接失效之前，最多发送多少个keepalive探测消息。
net.ipv4.tcp_sack	启用有选择的应答（1表示启用），通过有选择地应答乱序接收到的报文来提高性能，让发送者只发送丢失的报文段，（对于广域网通信来说）这个选项应该启用，但是会增加对CPU的占用。
net.ipv4.tcp_fack	启用转发应答，可以进行有选择应答（SACK）从而减少拥塞情况的发生，这个选项也应该启用。
net.ipv4.tcp_timestamps	TCP时间戳（会在TCP包头增加12 B），以一种比重发超时更精确的方法（参考RFC 1323）来启用对RTT的计算，为实现更好的性能应该启用这个选项。
net.ipv4.tcp_window_scaling	启用RFC 1323定义的window scaling，要支持超过64KB的TCP窗口，必须启用该值（1表示启用），TCP窗口最大至1GB，TCP连接双方都启用时才生效。
net.ipv4.tcp_syncookies	表示是否打开TCP同步标签（syncookie），内核必须打开了CONFIG_SYN_COOKIES项进行编译，同步标签可以防止一个套接字在有过多试图连接到达时引起过载。默认值0表示关闭。
net.ipv4.tcp_tw_reuse	表示是否允许将处于TIME-WAIT状态的socket（TIME-WAIT的端口）用于新的TCP连接。 <b>说明</b> 该参数仅对客户端有效，当“net.ipv4.tcp_timestamps”参数开启时生效。 该参数在NAT(Network Address Translation)场景下不能配置为1，否则将导致云服务器远程连接异常。



参数	说明
net.ipv4.tcp_tw_recycle	能够更快地回收TIME-WAIT套接字。 <b>说明</b> 该参数仅当“net.ipv4.tcp_timestamps”参数开启时生效。 该参数在NAT(Network Address Translation)场景下不能配置为1，否则将导致云服务器远程连接异常。
net.ipv4.tcp_fin_timeout	对于本端断开的socket连接，TCP保持在FIN-WAIT-2状态的时间（秒）。对方可能会断开连接或一直不结束连接或不可预料的进程死亡。
net.ipv4.ip_local_port_range	表示TCP/UDP协议允许使用的本地端口号。
net.ipv4.tcp_max_syn_backlog	对于还未获得对方确认的连接请求，可保存在队列中的最大数目。如果服务器经常出现过载，可以尝试增加这个数字。默认为1024。
net.ipv4.tcp_low_latency	允许TCP/IP栈适应在高吞吐量情况下低延时的情况，这个选项应该禁用。
net.ipv4.tcp_westwood	启用发送者端的拥塞控制算法，它可以维护对吞吐量的评估，并试图对带宽的整体利用情况进行优化，对于WAN通信来说应该启用这个选项。
net.ipv4.tcp_bic	为快速长距离网络启用Binary Increase Congestion，这样可以更好地利用以GB速度进行操作的链接，对于WAN通信应该启用这个选项。
net.ipv4.tcp_max_tw_buckets	该参数设置系统的TIME_WAIT的数量，如果超过默认值则会被立即清除。默认为180000。
net.ipv4.tcp_synack_retries	指明了处于SYN_RECV状态时重传SYN+ACK包的次数。
net.ipv4.tcp_abort_on_overflow	设置该参数为1时，当系统在短时间内收到了大量的请求，而相关的应用程序未能处理时，就会发送Reset包直接终止这些链接。建议通过优化应用程序的效率来提高处理能力，而不是简单地Reset。 默认值：0
net.ipv4.route.max_size	内核所允许的最大路由数目。
net.ipv4.ip_forward	接口间转发报文。
net.ipv4.ip_default_ttl	报文可以经过的最大跳数。
net.netfilter.nf_conntrack_tcp_timeout_established	让iptables对于已建立的连接，在设置时间内若没有活动，那么则清除掉。
net.netfilter.nf_conntrack_max	哈希表项最大值。

## 查看内核参数

- 方法一：通过“/proc/sys”目录，使用cat命令查看对应文件的内容。  
“/proc/sys/”目录是Linux内核启动后生成的伪目录，其目录下的net文件夹中存放了当前系统中生效的所有内核参数、目录树结构与参数的完整名称相关，如net.ipv4.tcp\_tw\_recycle，它对应的文件是/proc/sys/net/ipv4/tcp\_tw\_recycle，文件的内容就是参数值。

示例：

以查看net.ipv4.tcp\_tw\_recycle的值为例，执行以下命令：

```
cat /proc/sys/net/ipv4/tcp_tw_recycle
```

- 方法二：通过“/etc/sysctl.conf”文件进行查看。  
执行以下命令，查看当前系统中生效的所有参数。

```
/usr/sbin/sysctl -a
```

```
net.ipv4.tcp_syncookies = 1
net.ipv4.tcp_max_tw_buckets = 4096
net.ipv4.tcp_tw_reuse = 1
net.ipv4.tcp_tw_recycle = 1
net.ipv4.tcp_keepalive_time = 1800
net.ipv4.tcp_fin_timeout = 30
.....
net.ipv4.tcp_keepalive_time = 1200
net.ipv4.ip_local_port_range = 1024 65000
net.ipv4.tcp_max_syn_backlog = 8192
net.ipv4.tcp_rmem = 16384 174760 349520
net.ipv4.tcp_wmem = 16384 131072 262144
net.ipv4.tcp_mem = 262144 524288 1048576
.....
```

## 修改内核参数

- 方法一：通过“/proc/sys”目录，使用echo命令修改内核参数对应的文件。  
该方法修改的参数值仅在当次运行中生效，重启后会重置为原参数值，一般用于临时性验证。如需永久生效，请参考方法二。  
“/proc/sys/”目录是Linux内核启动后生成的伪目录，其目录下的net文件夹中存放了当前系统中生效的所有内核参数、目录树结构与参数的完整名称相关，如net.ipv4.tcp\_tw\_recycle，它对应的文件是/proc/sys/net/ipv4/tcp\_tw\_recycle，文件的内容就是参数值。

示例：

假设将net.ipv4.tcp\_tw\_recycle的值修改为0，执行以下命令：

```
echo "0" > /proc/sys/net/ipv4/tcp_tw_recycle。
```

- 方法二：通过“/etc/sysctl.conf”文件进行修改。  
该方法修改的参数值，永久生效。
  - a. 执行以下命令，修改指定的参数值。  

```
/sbin/sysctl -w kernel.domainname="example.com"
```

示例：

```
sysctl -w net.ipv4.tcp_tw_recycle="0"
```
  - b. 执行以下命令，修改“/etc/sysctl.conf”文件中的参数。  

```
vi /etc/sysctl.conf
```
  - c. 执行以下命令，使配置生效。  

```
/sbin/sysctl -p
```

## 14.6.6 DHCP 无法正常获取内网 IP?

### 问题描述

DHCP无法正常获取内网IP，当DHCP无法正常获取内网IP时，不同系统的云服务器表现如下：

- Linux系统表现为无法获取内网IP。
- windows系统由于自身的机制，会表现为云服务器内网IP变为169.254网段中的一个IP，和官网上显示的内网IP不同。

#### 说明

推荐使用公共镜像创建云服务器，公共镜像均已支持DHCP持续获取能力。

### 处理方法（Linux）

本文以CentOS 7.2为例进行介绍，其他版本操作系统请参考对应操作系统帮助文档。

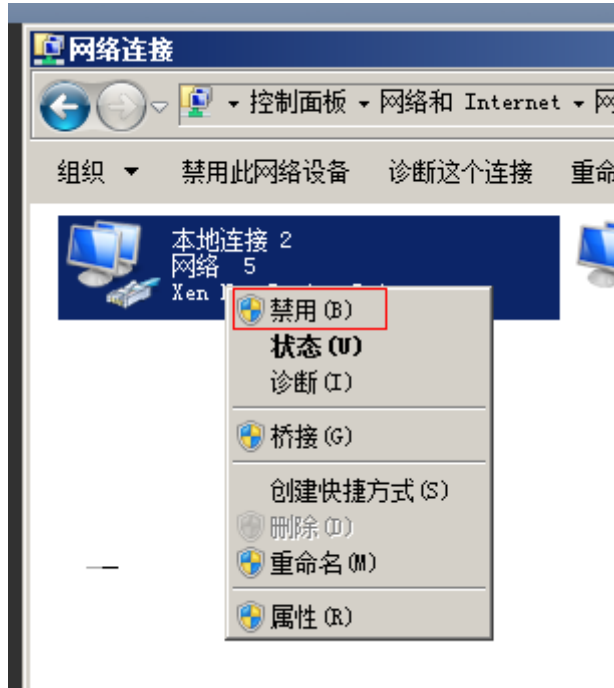
1. 登录Linux云服务器，执行如下命令，查看是否存在dhclient进程。  
**ps -ef | grep dhclient**
2. 若dhclient进程不存在，可以重启网卡或执行以下命令主动发起DHCP请求。  
**dhclient eth0或ifdown eth0 + ifup eth0或dhcpcd eth0**
3. 对于DHCP Client长期不发起请求的情况（例如重启网卡后又复现），可以参照以下操作：
  - a. 执行如下命令，配置静态IP。  
**vi /etc/sysconfig/network-scripts/ifcfg-eth0**

```
BOOTPROTO=static
IPADDR=192.168.1.100 #IP地址（修改）
NETMASK=255.255.255.0 #掩码值（修改）
GATEWAY=192.168.1.1 #网关地址（修改）
```
  - b. 重启云服务器使网络配置生效。
  - c. 选取有DHCP持续获取能力的镜像。
4. 若以上步骤没有解决问题，可以通过查看客户云服务器的messages日志（路径为“/var/log/messages”）排查问题，通过网卡的mac地址过滤日志，检查是否有客户的进程影响DHCP获取IP。
5. 若依然无法解决，请联系技术支持。

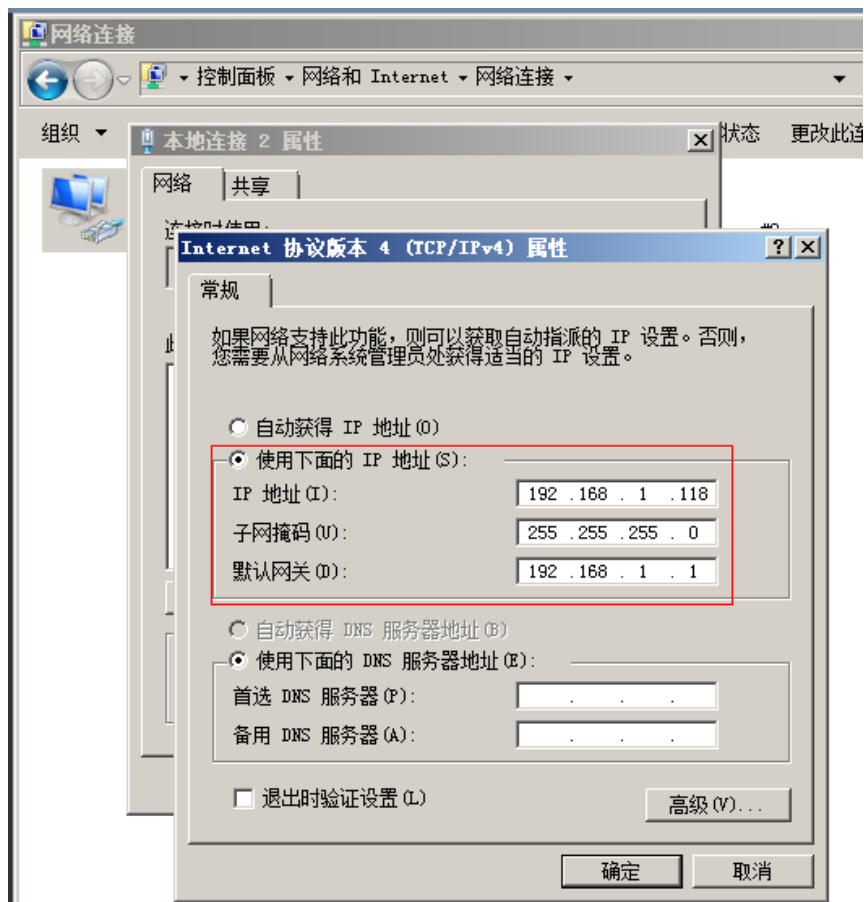
### 处理方法（Windows）

本文以Windows 2012为例进行介绍，其他版本请参考对应操作系统帮助文档。

1. 在网络连接中选择其中一个网络设备，单击鼠标右键，选择“禁用”，再单击“启用”。



2. 对于DHCP Client长期不发起请求的情况（例如重启网卡后又复现），可以参照以下操作：
  - a. 在网络连接中选择“属性”。
  - b. 在弹出的窗口中选择“Internet协议版本4”，单击“属性”，修改相关配置。



- c. 重启云服务器使网络配置生效。
3. 若依然无法解决，请联系技术支持。

## 14.6.7 Linux 云服务器网络性能测试方法

手把手教您用netperf工具、iperf3工具，测试弹性云服务器间网络性能。主要包括“测试准备”、“TCP带宽测试”、“UDP PPS测试”和“时延测试”。

### 背景知识

- 被测机：被压力测试网络性能的弹性云服务器，可作为netperf测试中的client端（发送端）或server端（接收端）。
- 辅助云服务器：弹性云服务器，用于netperf测试中的client端（发送端）或server端（接收端），用于与被测机建立连接，传递测试数据。
- 测试工具常用参数说明如表14-8、表14-9所示。

表 14-8 netperf 工具常用参数说明

参数	参数说明
-p	端口号
-H	接收端IP地址

参数	参数说明
-t	发包协议类型，测带宽时参数值为“TCP_STREAM”
-l	测试时长
-m	数据包大小，测试带宽时建议设置为“1440”

表 14-9 iperf3 工具常用参数说明

参数	参数说明
-p	端口号
-c	接收端IP地址
-u	UDP报文
-b	发送带宽
-t	测试时长
-l	数据包大小，测试PPS时建议设置为“16”
-A	iperf3占用的cpu编号。 本文示例中假设ECS最大为16vcpu，实际中根据ECS CPU数量进行循环。如ECS为8vcpu，则-A范围0~7，0~7。

## 测试准备

### 步骤1 准备弹性云服务器。

要求：被测机与辅助弹性服务器的类型、规格需保持一致，并在同一云服务器组，遵循反亲和部署。

表 14-10 环境准备

类型	数量	镜像	规格	IP地址
被测机	1台	CentOS 7.4 64bit (推荐)	-	192.168.2.10
辅助云服务器	8台	CentOS 7.4 64bit (推荐)	vCPU: 8核及以上	192.168.2.11 ~ 192.168.2.18

### 步骤2 准备测试工具。

要求：需分别在被测机和辅助云服务器上安装测试工具netperf、iperf3、sar。安装方法如表14-11所示。

表 14-11 安装测试工具

测试工具	安装方法
netperf	<ol style="list-style-type: none"><li>1. 执行以下命名，安装gcc。 <b>yum -y install unzip gcc gcc-c++</b></li><li>2. 执行以下命令，下载netperf。 <b>wget https://github.com/HewlettPackard/netperf/archive/refs/tags/netperf-2.7.0.zip</b></li><li>3. 执行以下命令，解压并安装netperf。 <b>unzip netperf-2.7.0.zip</b> <b>cd netperf-netperf-2.7.0/</b> <b>./configure &amp;&amp; make &amp;&amp; make install</b></li></ol>
iperf3	<ol style="list-style-type: none"><li>1. 执行以下命令，下载iperf3。 <b>wget --no-check-certificate https://codeload.github.com/esnet/iperf/zip/master -O iperf3.zip</b></li><li>2. 执行以下命令，解压并安装iperf3。 <b>unzip iperf3.zip</b> <b>cd iperf-master/</b> <b>./configure &amp;&amp; make &amp;&amp; make install</b></li></ol>
sar	执行以下命令，安装sar。 <b>yum -y install sysstat</b>

### 步骤3 开启网卡多队列。

要求：需分别在被测机和辅助云服务器上开启网卡多队列。

1. 执行以下命令，检查弹性云服务器支持的队列个数。  
**ethtool -l eth0 | grep -i Pre -A 5 | grep Combined**
2. 执行以下命令，设置队列数，开启网卡多队列功能。  
**ethtool -L eth0 combined X**  
其中，X表示步骤3.1中查询的队列数。

----结束

## TCP 带宽测试（使用 netperf 工具）

采用多流进行测试，本文以16条流为例，均分到8个ECS上，其他流数以此类推。

### 📖 说明

TCP带宽测试采用多流模型：

- 当测试TCP发送带宽时，采用一对多模型，即保证接收端能力足够。
- 当测试TCP接收带宽时，采用多对一模型，即保证发送端能力足够。

### 步骤1 测试TCP发送带宽。

1. 分别在“所有辅助云服务器”中执行以下命令，启动netserver进程。  
**netserver -p 12001**

**netserver -p 12002**

其中，-p用于指定监听端口。

2. 在“被测机”中执行以下命令，启动netperf进程，分别指定到辅助云服务器的不同netserver端口。其中，netperf工具的常用参数说明请参见表14-8。

##IP地址对应第1台辅助云服务器

```
netperf -H 192.168.2.11 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.11 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

##IP地址对应第2台辅助云服务器

```
netperf -H 192.168.2.12 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.12 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

##IP地址对应第3台辅助云服务器

```
netperf -H 192.168.2.13 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.13 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

##IP地址对应第4台辅助云服务器

```
netperf -H 192.168.2.14 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.14 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

##IP地址对应第5台辅助云服务器

```
netperf -H 192.168.2.15 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.15 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

##IP地址对应第6台辅助云服务器

```
netperf -H 192.168.2.16 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.16 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

##IP地址对应第7台辅助云服务器

```
netperf -H 192.168.2.17 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.17 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

##IP地址对应第8台辅助云服务器

```
netperf -H 192.168.2.18 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.18 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

**步骤2** 测试TCP接收带宽。

1. 在“被测机”中执行以下命令，启动netserver进程。

##port对应第1台辅助云服务器

```
netserver -p 12001
```

```
netserver -p 12002
```

##port对应第2台辅助云服务器

```
netserver -p 12003
```

```
netserver -p 12004
```

##port对应第3台辅助云服务器

```
netserver -p 12005
```



```
netserver -p 12006
##port对应第4台辅助云服务器
netserver -p 12007
netserver -p 12008
##port对应第5台辅助云服务器
netserver -p 12009
netserver -p 12010
##port对应第6台辅助云服务器
netserver -p 12011
netserver -p 12012
##port对应第7台辅助云服务器
netserver -p 12013
netserver -p 12014
##port对应第8台辅助云服务器
netserver -p 12015
netserver -p 12016
```

2. 执行以下命令，分别在“辅助云服务器”中启动netperf进程。

登录辅助云服务器1：

```
netperf -H 192.168.2.10 -p 12001 -t TCP_STREAM -l 300 -- -m 1440 &
netperf -H 192.168.2.10 -p 12002 -t TCP_STREAM -l 300 -- -m 1440 &
```

登录辅助云服务器2：

```
netperf -H 192.168.2.10 -p 12003 -t TCP_STREAM -l 300 -- -m 1440 &
netperf -H 192.168.2.10 -p 12004 -t TCP_STREAM -l 300 -- -m 1440 &
```

登录辅助云服务器3：

```
netperf -H 192.168.2.10 -p 12005 -t TCP_STREAM -l 300 -- -m 1440 &
netperf -H 192.168.2.10 -p 12006 -t TCP_STREAM -l 300 -- -m 1440 &
```

登录辅助云服务器4：

```
netperf -H 192.168.2.10 -p 12007 -t TCP_STREAM -l 300 -- -m 1440 &
netperf -H 192.168.2.10 -p 12008 -t TCP_STREAM -l 300 -- -m 1440 &
```

登录辅助云服务器5：

```
netperf -H 192.168.2.10 -p 12009 -t TCP_STREAM -l 300 -- -m 1440 &
netperf -H 192.168.2.10 -p 12010 -t TCP_STREAM -l 300 -- -m 1440 &
```

登录辅助云服务器6：

```
netperf -H 192.168.2.10 -p 12011 -t TCP_STREAM -l 300 -- -m 1440 &
netperf -H 192.168.2.10 -p 12012 -t TCP_STREAM -l 300 -- -m 1440 &
```

登录辅助云服务器7：

```
netperf -H 192.168.2.10 -p 12013 -t TCP_STREAM -l 300 -- -m 1440 &
netperf -H 192.168.2.10 -p 12014 -t TCP_STREAM -l 300 -- -m 1440 &
```

登录辅助云服务器8:

```
netperf -H 192.168.2.10 -p 12015 -t TCP_STREAM -l 300 -- -m 1440 &  
netperf -H 192.168.2.10 -p 12016 -t TCP_STREAM -l 300 -- -m 1440 &
```

**步骤3** 解析TCP测试结果。

测试结束后，发送端netperf进程输出结果如图14-103所示，最终结果为所有netperf进程测试结果之和。

图 14-103 发送端进程输出结果

Recv	Send	Send		
Socket	Socket	Message	Elapsed	
Size	Size	Size	Time	Throughput
bytes	bytes	bytes	secs.	10^6bits/sec
87380	16384	1440	120.02	956.30

发送缓冲区大小: 16384  
测试时间: 120.02  
吞吐量: 956.30  
接收缓冲区大小: 87380  
数据包大小: 1440

#### 说明

由于netperf进程众多，为方便统计，强烈建议测试指令输入完毕后，直接在被测弹性云服务器上用sar查看测试数据，命令为：

```
sar -n DEV 1 60
```

----结束

## UDP PPS 测试（使用 iperf3 工具）

**步骤1** 测试UDP发送PPS。

1. 登录辅助云服务器。
2. 分别所有“辅助云服务器”中执行以下命令，启动server进程。

```
iperf3 -s -p 12001 &
```

```
iperf3 -s -p 12002 &
```

其中，-p用于指定监听端口。

3. 在“被测机”中执行如下命令，启动client进程。其中，iperf3工具的常用参数说明请参见表14-9。

```
##辅助云服务器1
```

```
iperf3 -c 192.168.2.11 -p 12001 -u -b 100M -t 300 -l 16 -A 0 &
```

```
iperf3 -c 192.168.2.11 -p 12002 -u -b 100M -t 300 -l 16 -A 1 &
```

```
##辅助云服务器2
```

```
iperf3 -c 192.168.2.12 -p 12001 -u -b 100M -t 300 -l 16 -A 2 &
```

```
iperf3 -c 192.168.2.12 -p 12002 -u -b 100M -t 300 -l 16 -A 3 &
```

##辅助云服务器3

```
iperf3 -c 192.168.2.13 -p 12001 -u -b 100M -t 300 -l 16 -A 4 &  
iperf3 -c 192.168.2.13 -p 12002 -u -b 100M -t 300 -l 16 -A 5 &
```

##辅助云服务器4

```
iperf3 -c 192.168.2.14 -p 12001 -u -b 100M -t 300 -l 16 -A 6 &  
iperf3 -c 192.168.2.14 -p 12002 -u -b 100M -t 300 -l 16 -A 7 &
```

##辅助云服务器5

```
iperf3 -c 192.168.2.15 -p 12001 -u -b 100M -t 300 -l 16 -A 8 &  
iperf3 -c 192.168.2.15 -p 12002 -u -b 100M -t 300 -l 16 -A 9 &
```

##辅助云服务器6

```
iperf3 -c 192.168.2.16 -p 12001 -u -b 100M -t 300 -l 16 -A 10 &  
iperf3 -c 192.168.2.16 -p 12002 -u -b 100M -t 300 -l 16 -A 11 &
```

##辅助云服务器7

```
iperf3 -c 192.168.2.17 -p 12001 -u -b 100M -t 300 -l 16 -A 12 &  
iperf3 -c 192.168.2.17 -p 12002 -u -b 100M -t 300 -l 16 -A 13 &
```

##辅助云服务器8

```
iperf3 -c 192.168.2.18 -p 12001 -u -b 100M -t 300 -l 16 -A 14 &  
iperf3 -c 192.168.2.18 -p 12002 -u -b 100M -t 300 -l 16 -A 15 &
```

## 步骤2 测试UDP接收PPS。

1. 执行以下命令，在“被测机”中启动server进程。其中，iperf3工具的常用参数说明请参见[表14-9](#)。

##port对应第1台辅助云服务器

```
iperf3 -s -p 12001 -A 0 -i 60 &  
iperf3 -s -p 12002 -A 1 -i 60 &
```

##port对应第2台辅助云服务器

```
iperf3 -s -p 12003 -A 2 -i 60 &  
iperf3 -s -p 12004 -A 3 -i 60 &
```

##port对应第3台辅助云服务器

```
iperf3 -s -p 12005 -A 4 -i 60 &  
iperf3 -s -p 12006 -A 5 -i 60 &
```

##port对应第4台辅助云服务器

```
iperf3 -s -p 12007 -A 6 -i 60 &  
iperf3 -s -p 12008 -A 7 -i 60 &
```

##port对应第5台辅助云服务器

```
iperf3 -s -p 12009 -A 8 -i 60 &  
iperf3 -s -p 12010 -A 9 -i 60 &  
##port对应第6台辅助云服务器  
iperf3 -s -p 12011 -A 10 -i 60 &  
iperf3 -s -p 12012 -A 11 -i 60 &  
##port对应第7台辅助云服务器  
iperf3 -s -p 12013 -A 12 -i 60 &  
iperf3 -s -p 12014 -A 13 -i 60 &  
##port对应第8台辅助云服务器  
iperf3 -s -p 12015 -A 14 -i 60 &  
iperf3 -s -p 12016 -A 15 -i 60 &
```

2. 分别在“辅助云服务器”中，执行以下命令，启动client进程。其中，iperf3工具的常用参数说明请参见[表14-9](#)。

登录辅助云服务器1：

```
iperf3 -c 192.168.2.10 -p 12001 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12002 -u -b 100M -t 300 -l 16 -A 1 &
```

登录辅助云服务器2：

```
iperf3 -c 192.168.2.10 -p 12003 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12004 -u -b 100M -t 300 -l 16 -A 1 &
```

登录辅助云服务器3：

```
iperf3 -c 192.168.2.10 -p 12005 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12006 -u -b 100M -t 300 -l 16 -A 1 &
```

登录辅助云服务器4：

```
iperf3 -c 192.168.2.10 -p 12007 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12008 -u -b 100M -t 300 -l 16 -A 1 &
```

登录辅助云服务器5：

```
iperf3 -c 192.168.2.10 -p 12009 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12010 -u -b 100M -t 300 -l 16 -A 1 &
```

登录辅助云服务器6：

```
iperf3 -c 192.168.2.10 -p 12011 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12012 -u -b 100M -t 300 -l 16 -A 1 &
```

登录辅助云服务器7：

```
iperf3 -c 192.168.2.10 -p 12013 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12014 -u -b 100M -t 300 -l 16 -A 1 &
```

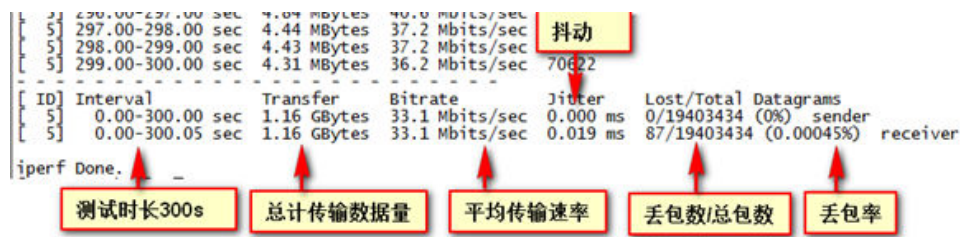
登录辅助云服务器8：

```
iperf3 -c 192.168.2.10 -p 12015 -u -b 100M -t 300 -l 16 -A 0 &  
iperf3 -c 192.168.2.10 -p 12016 -u -b 100M -t 300 -l 16 -A 1 &
```

### 步骤3 解析UDP PPS测试结果。

UDP PPS测试结果示例如[图14-104](#)所示。

图 14-104 UDP PPS 测试结果



#### 说明

由于iperf3进程众多，为方便统计，强烈建议测试指令输入完毕后，直接在被测弹性云服务器上用sar查看测试数据，命令为：

```
sar -n DEV 1 60
```

----结束

## 时延测试

**步骤1** 执行以下命令，在“被测机”中启动qperf进程。

```
qperf &
```

**步骤2** 登录辅助云服务器1，并执行以下命令，进行时延测试。

```
qperf 192.168.2.10 -m 64 -t 60 -vu udp_lat
```

测试完后屏幕显示结果的latency字段就是ECS间时延。

----结束

## 14.6.8 弹性云服务器添加网卡后是否会自动启动？

截止目前测试的结果，只有弹性服务器的操作系统为CentOS 7.0时，添加网卡后在弹性云服务器内部无法自启动，需要手动激活。

## 14.6.9 怎样检查 IB 云服务器的网络是否正常？

对于已安装InfiniBand网卡驱动力的H2型弹性云服务器（以下简称IB云服务器），可以通过如下方式，检查云服务器的IB网卡驱动安装成功、网络连通，可以正常工作。

#### 说明

检查过程中，如果发现您的弹性云服务器未安装ibstat等命令工具，需先从InfiniBand网卡驱动安装包中获取该工具并进行安装。

**步骤1** 检查IB云服务器的网卡状态是否正常。

1. 登录弹性云服务器。
2. 执行以下命令，查看IB云服务器的网卡状态是否正常。

```
ibstat
```

- 是，执行**步骤2**。
- 否，请联系管理员寻求技术支持。

**步骤2** 检查IB云服务器之间，网络通信是否正常。

1. 登录其中一台IB云服务器，并执行以下命令。

```
ib_write_bw -x 0 --pkey_index 0
```

2. 登录另一台IB云服务器，并执行以下命令。其中，ip\_addr为第一台IB云服务器网卡的IP地址。

```
ib_write_bw -x 0 --pkey_index 0 ip_addr
```

3. 查看回显信息是否正确。

**图 14-105** 网络通信正常

```
[root@host-11-11-11-111 MLNX_OFED_LINUX-3.4-1.0.0.0-rhel7.2-x86_64]# ib_write_bw -x 0 --pkey_index 0 4.29.43.20
-----
RDMA_Write BW Test
Dual-port      : OFF      Device           : mlx5_0
Number of qps  : 1        Transport type   : IB
Connection type: RC       Using SRQ        : OFF
TX depth       : 128
CQ Moderation  : 100
Mtu            : 4096[B]
Link type      : IB
GID index      : 0
Max inline data: 0[B]
rdma_cm QPs    : OFF
Data ex. method: Ethernet

Local address: LID 0x05 QPN 0x0067 PSN 0xaaccfb RKey 0x001c0c VAddr 0x007fb3cd1b0000
GID: 254:128:00:00:00:00:00:00:03:00:135:40:170
remote address: LID 0x05 QPN 0x006a PSN 0xebbf6d RKey 0x001c10 VAddr 0x007fdad5990000
GID: 254:128:00:00:00:00:00:00:01:03:00:135:40:170
-----
#bytes      #iterations  BW peak[MB/sec]  BW average[MB/sec]  MsgRate[Mpps]
65536       5000          12132.78         11900.18            0.190403
```

- 如果回显信息如**图14-105**所示，表示IB云服务器之间的IB网络通信正常。
- 如果IB网络通信不正常，请联系管理员寻求技术支持。

----结束

## 14.6.10 怎样给 IB 网卡手动配置 IP 地址？

IPoIB（IP over IB）是在InfiniBand协议上提供IP协议进行数据传输的能力。对于使用SUSE操作系统的H2型、HL1型云服务器，如需使用IPoIB功能，需要在安装完IB网卡驱动之后，给IB网卡手动配置一个IP地址。

### 前提条件

H2型或HL1型云服务器已安装InfiniBand网卡驱动。

### 背景知识


为了保证同一租户下弹性云服务器给IB网卡配置的IP地址不冲突，建议根据VPC网络的IP地址进行计算，获取待配置的IP地址。具体方法如下：

假设给IB网卡分配的IPADDR前两个八位固定为169.254，后两个八位应该与eth0的IP地址的后两个八位相同，子网掩码与eth0的掩码相同。

示例：

假设eth0的IP地址为192.168.0.100/24，则IB网卡待配置的IP地址为169.254.0.100/24。

## 操作步骤

1. 登录弹性云服务器。
2. 执行以下命令，切换至root用户。  
**sudo su -**
3. 执行以下命令，编辑文件/etc/sysconfig/network/ifcfg-ib0。  
**vi /etc/sysconfig/network/ifcfg-ib0**
4. 写入如下内容：  
**DEVICE=ib0**  
**BOOTPROTO=static**  
**IPADDR=IB网卡待配置的IP地址**  
**NETMASK=子网掩码**  
**STARTMODE=auto**  
  
 **说明**  
  
IB网卡待配置的IP地址与子网掩码计算方法，请参见[背景知识](#)。
5. 执行以下命令，重启网络使配置生效。  
**service network restart**

### 14.6.11 如何解决 SR-IOV 场景下 win7 弹性云服务器安装 82599 网卡驱动报错?

#### 问题描述

从Intel官网下载最新的20.4.1版本驱动包（下载地址：<https://downloadcenter.intel.com/search?keyword=Intel++Ethernet+Connections+CD>），该版本驱动包在Windows7 64位SR-IOV直通弹性云服务器上安装时会提示“找不到英特尔适配器”错误。

#### 原因分析

Intel 82599直通网卡在未安装驱动时被操作系统识别为以太网控制器设备，20.4.1版本驱动包在安装时未能正确识别出Intel网卡设备，导致程序报错。

#### 处理方法

在20.4.1驱动包文件夹下运行Autorun.exe。安装驱动包前先给网卡安装驱动，使网卡被系统识别为Intel 82599 VF设备，安装驱动有两种方法。

- 方法1：通过版本更新方法安装驱动
  - a. 从Intel官网下载18.6版本驱动包。
  - b. 到18.6驱动包安装文件夹下运行Autorun.exe进行安装。
  - c. 成功后到20.4.1版本驱动包文件夹下运行Autorun.exe更新驱动。
- 方法2：设备管理器手动安装驱动
  - a. 打开Windows资源管理器，右键单击“计算机”，选择“管理”，打开“设备管理器”，在设备管理器中找到网卡，未安装驱动时，网卡位于“其他设备”一栏，名字为“以太网控制器”。

- b. 右键单击“以太网控制器”，选择“更新驱动程序软件”。
- c. 单击“浏览”，选择驱动包所在路径，单击“下一步”。
- d. 驱动安装成功后，可以在设备管理器的“网络适配器”一栏看到网卡。
- e. 单击“Autorun.exe”安装20.4.1版本一键安装驱动包。

## 14.7 弹性 IP 类

### 14.7.1 一台弹性云服务器是否可以绑定多个弹性公网 IP?

#### 操作场景

一台弹性云服务器可以绑定多个弹性公网IP，但是不建议您这样操作。

当云服务器拥有多张网卡时，如果需要配置多个弹性公网IP，此时需要在云服务器内部为这些网卡配置策略路由，才可以确保多张网卡均可以和外部正常通信，具体请参见[配置示例](#)。

#### 配置示例

弹性云服务器的配置如[表14-12](#)所示。

表 14-12 参数配置

参数	配置
名称	ecs_test
镜像	CentOS 6.5 64bit
弹性公网IP	2个
主网卡	eth0
从网卡	eth1

#### 示例1:

假设您希望访问公网11.11.11.0/24时，使用从网卡eth1，此时可以执行以下操作，配置路由策略。

1. 登录弹性云服务器。
2. 执行以下命令，配置路由策略。

```
ip route add 11.11.11.0/24 dev eth1 via 192.168.2.1
```

其中，192.168.2.1为从网卡eth1对应网关的IP地址。

#### 示例2:

基于示例1，如果您希望默认公网流量也通过从网卡eth1路由，此时可以执行以下操作，配置路由策略。



1. 登录弹性云服务器。
2. 执行以下命令，删除默认路由。

```
ip route delete default
```

#### 须知

删除默认路由会导致弹性云服务器网络中断，无法继续通过SSH登录，此处请确保不影响业务再执行该操作。

3. 执行以下命令，配置新的默认路由。

```
ip route add 0.0.0.0/0 dev eth1 via 192.168.2.1
```

其中，192.168.2.1为从网卡eth1对应网关的IP地址。

## 14.7.2 未绑定弹性公网 IP 的弹性云服务器能否访问公网？

能。

您可以通过配置SNAT服务器，使得没有绑定EIP的弹性云服务器访问Internet。

配置方法，请参见《虚拟私有云用户指南》的“配置SNAT服务器”章节。

## 14.7.3 弹性 IP Ping 不通？

### 问题描述

用户购买弹性公网IP并绑定云服务器后，本地主机或其他云服务器无法Ping通该云服务器的弹性公网IP。

### 排查思路

以下排查思路根据原因的出现概率进行排序，建议您从高频原因往低频原因排查，从而帮助您快速找到问题的原因。

如果解决完某个可能原因仍未解决问题，请继续排查其他可能原因。

图 14-106 弹性公网 IP Ping 不通排查思路



表 14-13 弹性公网 IP Ping 不通排查思路

可能原因	处理措施
安全组未添加ICMP规则	安全组添加ICMP规则，详细操作请参考 <a href="#">检查安全组规则</a> 。
防火墙设置了禁Ping	检查防火墙对ICMP规则的启用状态，详细操作请参考 <a href="#">检查防火墙设置</a> 。
云服务器设置了禁Ping	检查云服务器对ICMP规则的启用状态，详细操作请参考 <a href="#">检查云服务器是否设置了禁Ping</a> 。
关联了网络ACL	如果VPC关联了网络ACL，请检查“网络ACL”规则，详细操作请参考 <a href="#">检查网络ACL规则</a> 。
网络异常	检查本地网络，使用相同区域主机进行Ping测试，详细操作请参考 <a href="#">检查网络是否正常</a> 。
多网卡场景，路由信息配置不正确	扩展网卡导致网络不通现象通常是路由配置问题，详细操作请参考 <a href="#">检查云服务器路由配置（多网卡场景）</a> 。

## 检查安全组规则

Ping使用的是ICMP协议，请检查云服务器对应的安全组是否放通了“ICMP”规则。

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表，单击待变更安全组规则的弹性云服务器名称。  
系统跳转至该弹性云服务器详情页面。
4. 选择“安全组”页签，展开安全组，查看安全组规则。
5. 单击安全组ID。  
系统自动跳转至安全组页面。
6. 在出方向规则页签，单击“添加规则”。添加出方向规则。

表 14-14 安全组规则

方向	类型	协议和端口	目的地址
出方向	IPv4	ICMP: Any	0.0.0.0/0 0.0.0.0/0表示所有IP地址

7. 在入方向规则页签，单击“添加规则”，添加入方向规则。

表 14-15 安全组规则

方向	类型	协议和端口	源地址
入方向	IPv4	ICMP: Any	0.0.0.0/0 0.0.0.0/0表示所有IP地址

- 单击“确定”，完成安全组规则配置。

## 检查防火墙设置

如果云服务器开启了防火墙，需要检查防火墙对Ping规则是否有限制。

### Linux系统云服务器

- 执行以下命令查看防火墙状态，以CentOS 7操作系统为例。

```
firewall-cmd --state
```

回显信息显示“running”代表防火墙已开启。

- 查看云服务器内部是否有安全规则所限制。

```
iptables -L
```

回显信息如图14-107所示说明没有ICMP规则被限制。

图 14-107 查看防火墙规则

```
[root@ecs-3c4e ~]# iptables -L
Chain INPUT (policy ACCEPT)
target     prot opt source                destination           icmp echo-request
ACCEPT    icmp -- anywhere             anywhere              icmp echo-request

Chain FORWARD (policy ACCEPT)
target     prot opt source                destination

Chain OUTPUT (policy ACCEPT)
target     prot opt source                destination           icmp echo-reply
ACCEPT    icmp -- anywhere             anywhere              icmp echo-reply
[root@ecs-3c4e ~]#
```

如果ICMP规则被限制，请执行以下命令启用对应规则。

```
iptables -A INPUT -p icmp --icmp-type echo-request -j ACCEPT
```

```
iptables -A OUTPUT -p icmp --icmp-type echo-reply -j ACCEPT
```

### Windows操作系统

- 登录Windows云服务器，单击桌面左下角的Windows图标，选择“控制面板 > Windows防火墙”。
- 单击“启用或关闭Windows防火墙”。  
查看并设置防火墙的具体状态：开启或关闭。
- 如果防火墙状态为“开启”，请执行4。
- 检查防火墙对ICMP规则的启用状态。
  - 在“Windows防火墙”页面，在左侧导航栏选择“高级设置”。
  - 启用以下规则。  
进站规则：“文件和打印机共享（回显请求-ICMPv4-In）”  
出站规则：“文件和打印机共享（回显请求-ICMPv4-Out）”  
如启用了IPV6请同时启用以下规则：  
进站规则：“文件和打印机共享（回显请求-ICMPv6-In）”  
出站规则：“文件和打印机共享（回显请求-ICMPv6-Out）”

图 14-108 进站规则

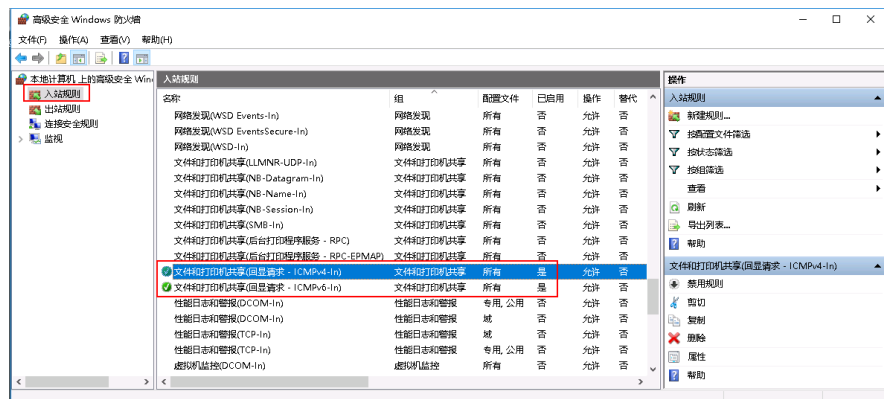


图 14-109 出站规则



## 检查云服务器是否设置了禁 Ping

### Windows

使用命令行方式开启Ping设置。

1. 打开cmd运行窗口。
2. 执行如下命令开启Ping设置。

```
netsh firewall set icmpsetting 8
```

### Linux

检查云服务器的内核参数。

1. 检查文件/etc/sysctl.conf中配置项“net.ipv4.icmp\_echo\_ignore\_all”的值，0表示允许Ping，1表示禁止Ping。
2. 允许PING设置。
  - 临时允许PING操作的命令：

```
#echo 0 >/proc/sys/net/ipv4/icmp_echo_ignore_all
```
  - 永久允许PING配置方法：

```
net.ipv4.icmp_echo_ignore_all=0
```

## 检查网络 ACL 规则

VPC默认没有网络ACL，如果关联了网络ACL，请检查“网络ACL”规则。

1. 查看云服务器对应的子网是否关联了网络ACL。  
如显示具体的网络ACL名称说明已关联网络ACL。
2. 单击网络ACL名称查看网络ACL的状态。
3. 若“网络ACL”为“开启”状态，需要添加ICMP放通规则进行流量放通。

### 📖 说明

需要注意“网络ACL”的默认规则是丢弃所有出入方向的包，若关闭“网络ACL”后，其默认规则仍然生效。

## 检查网络是否正常

1. 检查本地网络，使用相同区域主机进行Ping测试。  
使用在相同区域的云服务器去Ping没有Ping通的弹性公网IP，如果可以正常Ping通说明虚拟网络正常，请排除本地网络故障后重新Ping测试。
2. 检查是否链路故障。  
链路拥塞、链路节点故障、服务器负载高等问题均可能引起执行Ping命令时出现丢包或时延过高的问题。

## 检查云服务器路由配置（多网卡场景）

一般操作系统的默认路由优先使用主网卡，如果出现使用扩展网卡导致网络不通现象通常是路由配置问题。

- 如果云服务器配置了多网卡，请确认云服务器内默认路由是否存在。

- a. 登录云服务器，执行如下命令，查看是否存在默认路由。

**ip route**

图 14-110 查看默认路由

```
[root@do-not-del-scy ~]# ip route
default via 192.168.2.1 dev eth0
169.254.0.0/16 dev eth0 scope link metric 1002
169.254.169.254 via 192.168.2.1 dev eth0 proto static
192.168.2.0/24 dev eth0 proto kernel scope link src 192.168.2.112
```

- b. 若没有该路由，执行如下命令，添加默认路由。

**ip route add default via XXXX dev eth0**

### 📖 说明

XXXX表示网关IP。

- 如果云服务器配置了多网卡，且弹性公网IP绑定在非主网卡上，需要在云服务器内部配置策略路由来实现非主网卡的通信。

## 14.7.4 为什么弹性云服务器可以远程连接，但是无法 ping 通？

### 问题描述

云服务器可以远程连接，但弹性公网IP无法ping通。

### 问题原因

需要添加安全组的入方向规则并开启ICMP协议。

### 处理方法

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表栏，单击目标弹性云服务器名称。  
系统进入弹性云服务器详情页。
4. 选择“安全组”页签，展开安全组，并单击安全组ID。
5. 在“安全组”页面的“入方向”页签下，单击“添加规则”。
6. 根据如下参数要求，添加安全组入方向规则，开启ICMP协议。
  - 协议：ICMP
  - 源地址：IP地址 0.0.0.0/0

## 14.8 密码与密钥对

### 14.8.1 怎样设置镜像密码的有效期？

如果弹性云服务器镜像密码已过期导致无法登录，请联系管理员处理。

如果弹性云服务器还可正常登录，用户可以参考以下操作设置密码有效期，避免密码过期造成的不便。

### 操作步骤

以EulerOS 2.2镜像为例：

1. 登录弹性云服务器。
2. 执行以下命令，查看密码有效期。  
**vi /etc/login.defs**  
找到配置项“PASS\_MAX\_DAYS”，该参数表示密码的有效时间。
3. 执行以下命令，修改参数“PASS\_MAX\_DAYS”的取值。  
**chage -M 99999 user\_name**  
其中：99999为密码有效期限，user\_name为系统用户，例如root用户。

#### 说明

建议用户根据实际情况及业务需求进行配置，定期使用该命令更新密码有效期。

4. 再次执行**vi /etc/login.defs**，验证配置是否生效。

图 14-111 验证配置

```
## Password aging controls:
##
##     PASS_MAX_DAYS   Maximum number of days a password may be used.
##     PASS_MIN_DAYS   Minimum number of days allowed between password changes.
##     PASS_MIN_LEN     Minimum acceptable password length.
##     PASS_WARN_AGE   Number of days warning given before a password expires.
##
PASS_MAX_DAYS 99999
PASS_MIN_DAYS 0
PASS_MIN_LEN 5
PASS_WARN_AGE 7
```

## 14.8.2 一键重置密码后无法使用新密码登录弹性云服务器

### 操作场景

一键重置密码后无法使用新密码登录弹性云服务器，请参考本节操作逐一进行排查。

#### 说明

请确保一键式重置密码插件未被安全软件阻止运行，否则重置密码功能无法使用。  
重置密码之后，必须重启弹性云服务器方可生效。

### Windows 操作系统

请根据如下原因逐一进行排查：

#### 步骤1 检查安全组出方向80端口是否放通。

1. 登录控制台。
2. 选择需要检查的弹性云服务器，并进入“弹性云服务器”详情页面。
3. 在“安全组”页签，检查“出方向规则”是否放通“80”端口。  
默认安全组的规则中，“出方向规则”为放通所有端口。

#### 步骤2 检查弹性云服务器VPC的DHCP是否启用。

1. 在“弹性云服务器”详情页面中，单击“虚拟私有云”的名称，进入“虚拟私有云”页面。
2. 单击虚拟私有云名称，进入“虚拟私有云”详情页面。
3. 单击“子网”后的数字，进入“子网”页面。
4. 单击“子网”名称，进入“子网”详情页面。
5. 检查是否启用DHCP。

#### 步骤3 如果安全组配置、DHCP均正常，但一键重置密码功能仍未生效，请尝试使用原密码登录弹性云服务器。

- 如果原密码失效，请参考[重置Windows云服务器密码](#)重置密码。
- 如果原密码有效，可以用原密码登录弹性云服务器，参考[在操作系统内部修改ECS密码](#)重置密码。

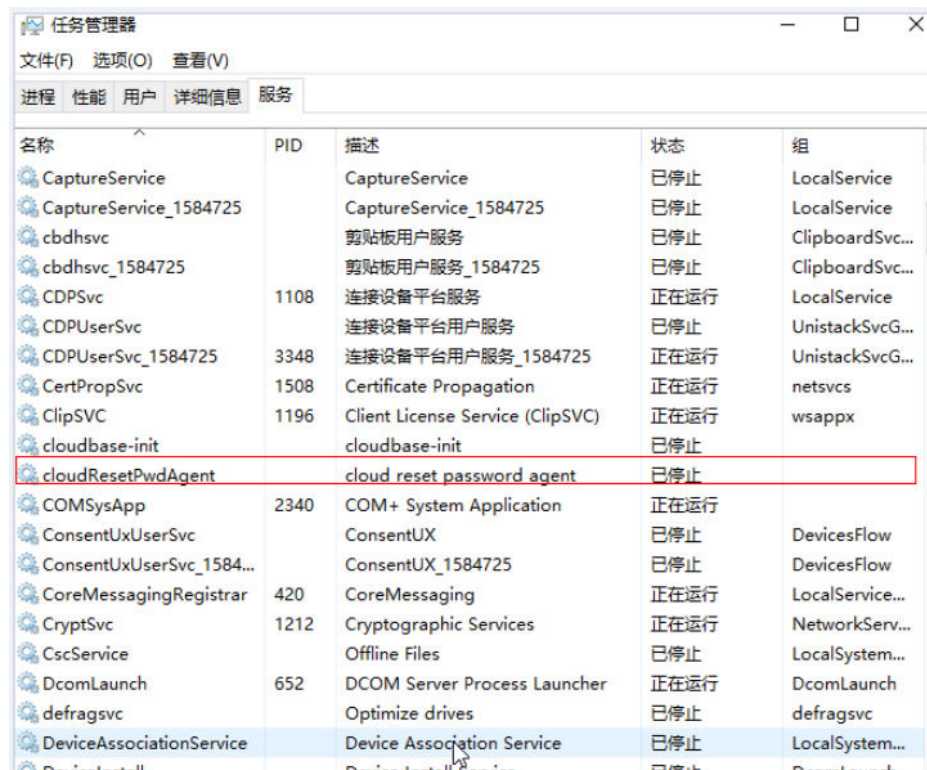
#### 步骤4 如果原密码有效，可以用原密码登录弹性云服务器，检查弹性云服务器是否已安装密码重置插件CloudResetPwdAgent。检查方法如下：

查看任务管理器，如果找到cloudResetPwdAgent服务，如图14-112所示，表示弹性云服务器已安装密码重置插件。否则说明当前弹性云服务器没有安装一键式重置密码插件。



安装方法请参见[安装单台ECS的一键式重置密码插件（单台操作）](#)。

图 14-112 安装插件成功



----结束

## Linux 操作系统

请根据如下原因逐一进行排查：

### 步骤1 检查安全组出方向80端口是否放通。

1. 登录控制台。
2. 选择需要检查的弹性云服务器，并进入“弹性云服务器”详情页面。
3. 在“安全组”页签，检查“出方向规则”是否放通“80”端口。  
默认安全组的规则中，“出方向规则”为放通所有端口。

### 步骤2 检查弹性云服务器VPC的DHCP是否启用。

1. 在“弹性云服务器”详情页面中，单击“虚拟私有云”的名称，进入“虚拟私有云”页面。
2. 单击虚拟私有云名称，进入“虚拟私有云”详情页面。
3. 单击“子网”后的数字，进入“子网”页面。
4. 单击“子网”名称，进入“子网”详情页面。
5. 检查是否启用DHCP。

### 步骤3 如果安全组配置、DHCP均正常，但一键重置密码功能仍未生效，请尝试使用原密码登录弹性云服务器。

- 如果原密码失效，可以进入单用户模式下进行密码重置。



- 如果原密码有效，可以用原密码登录弹性云服务器后进入操作系统内进一步排查：
  - a. 使用原密码登录Linux弹性云服务器。
  - b. 执行`curl http://169.254.169.254/openstack/latest/resetpwd_flag`。
    - 返回为“true”，表示可以一键重置密码。
    - 返回其他，表示不支持重置密码或网络异常。

```
[root@ecs-f7e2 ~]# service cloudResetPwdAgent status
cloudResetPwdAgent is not running.
[root@ecs-f7e2 ~]# curl http://169.254.169.254/openstack/latest/reset_pwd_flag
{"message": "API not found", "request_id": "c3b8eb86-156d-44c7-ab44-891926965483"}
[root@ecs-f7e2 ~]# curl http://169.254.169.254/openstack/latest/resetpwd_flag
{"resetpwd_flag": "True"}[root@ecs-f7e2 ~]#
```

**步骤4** 如果原密码有效，可以用原密码登录弹性云服务器，检查是否已安装“CloudResetPwdAgent”。

1. 检查弹性云服务器的根目录下，是否存在“CloudResetPwdAgent”目录。
  - 是，请执行[步骤4.2](#)
  - 否，说明当前弹性云服务器没有安装一键式重置密码插件。  
安装方法请参见[安装单台ECS的一键式重置密码插件（单台操作）](#)。
2. 执行以下命令，查看“CloudResetPwdAgent”的状态。

```
service cloudResetPwdAgent status
```

如果回显信息是“unrecognized service”说明当前弹性云服务器没有安装一键式重置密码插件。

安装方法请参见[安装单台ECS的一键式重置密码插件（单台操作）](#)。

---结束

## 14.8.3 业务端口被一键式重置密码插件占用

### 问题描述

在弹性云服务器上运行某业务时，系统提示所需端口被一键式重置密码Agent插件占用。那么，一键式重置密码插件占用的端口与业务端口发生冲突时，应当如何处理？

### 问题原因

对于采用AUTO模式的弹性云服务器，一键式重置密码插件启动时，会随机选取端口进行使用，可能占用了业务端口。

#### 📖 说明

一键式重置密码插件已经升级，默认采用PIPE模式。

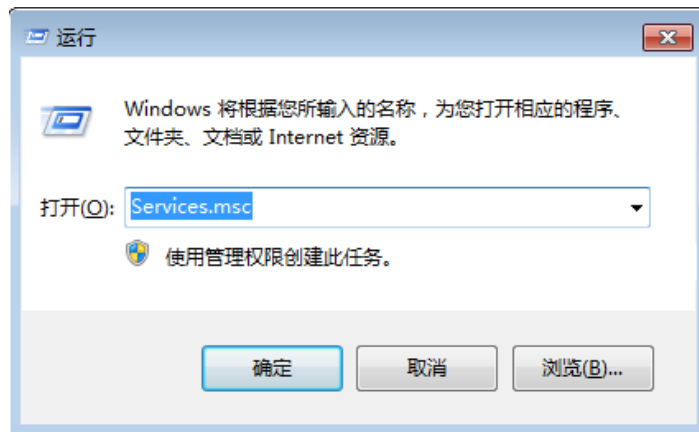
- 对于新创建的弹性云服务器，默认采用PIPE模式，不会占用端口。
- 对于已创建的弹性云服务器，仍采用AUTO模式，随机占用31000~32999中的一个端口。占用端口的原则是：在该范围内，系统会按照自小到大的顺序，占用当前空闲的端口。

### 方法一（推荐）：将一键式重置密码插件 wrapper 修改为 PIPE 模式

推荐您将一键式重置密码插件wrapper从AUTO模式（SOCKET）修改为PIPE模式，修改后，插件运行时不再占用端口。

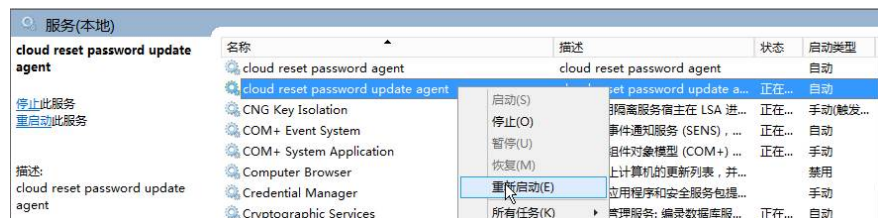
1. 打开CloudResetPwdAgent配置文件。
  - Linux弹性云服务器文件位置：  
“/CloudrResetPwdAgent/conf/wrapper.conf” 和 “/CloudResetPwdUpdateAgent/conf/wrapper.conf”
  - Windows弹性云服务器文件位置：  
“C:\CloudrResetPwdAgent\conf\wrapper.conf” 和  
“C:\CloudResetPwdUpdateAgent\conf\wrapper.conf”
2. 在末尾新增如下配置：  
**wrapper.backend.type=PIPE**
3. 重启CloudResetPwdUpdateAgent服务。
  - Linux弹性云服务器  
**/CloudResetPwdUpdateAgent/bin/cloudResetPwdUpdateAgent.script  
restart**
  - Windows弹性云服务器
    - i. 使用快捷键“Win+R”，打开“运行”窗口。
    - ii. 输入“Services.msc”，并单击“确定”。

图 14-113 运行



- iii. 右键单击服务“cloud reset password update agent”，选择“重新启动”。

图 14-114 服务（本地）



## 方法二：修改配置，更换端口范围

您可以修改CloudResetPwdAgent配置，更换默认随机端口选取的范围（31000~32999），确保业务端口不在一键式重置密码插件的端口选择范围内。

假设将一键式重置密码插件随机占用的端口范围修改为：40000~42000，则操作如下：

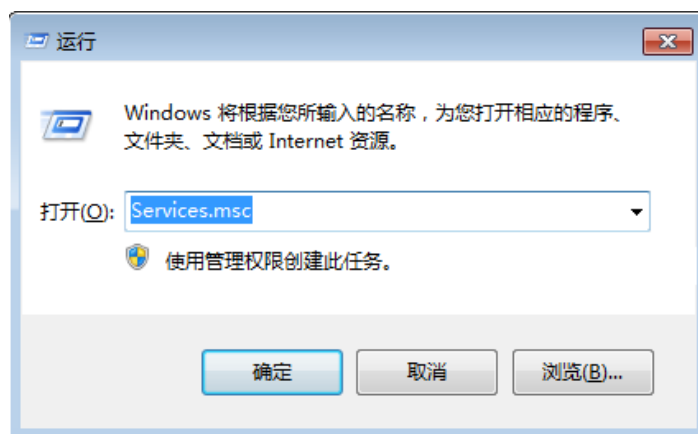
1. 打开CloudResetPwdAgent配置文件。
  - Linux弹性云服务器文件位置：  
“/CloudResetPwdAgent/conf/wrapper.conf”和“/CloudResetPwdUpdateAgent/conf/wrapper.conf”
  - Windows弹性云服务器文件位置：  
“C:\CloudResetPwdAgent\conf\wrapper.conf”和  
“C:\CloudResetPwdUpdateAgent\conf\wrapper.conf”
2. 新增如下配置：  
**wrapper.port.min=40000**  
**wrapper.port.max=41000**  
**wrapper.jvm.port.min=41001**  
**wrapper.jvm.port.max=42000**

图 14-115 修改配置文件

```
[root@sluo-ecs-9545 ~]# tail -n 4 /CloudResetPwdUpdateAgent/conf/wrapper.conf
wrapper.port.min=40000
wrapper.port.max=41000
wrapper.jvm.port.min=41001
wrapper.jvm.port.max=42000
[root@sluo-ecs-9545 ~]# tail -n 4 /CloudResetPwdAgent/conf/wrapper.conf
wrapper.port.min=40000
wrapper.port.max=41000
wrapper.jvm.port.min=41001
wrapper.jvm.port.max=42000
```

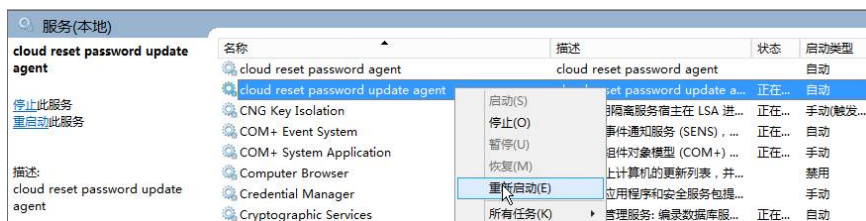
3. 重启CloudResetPwdUpdateAgent服务。
  - Linux弹性云服务器  
**/CloudResetPwdUpdateAgent/bin/cloudResetPwdUpdateAgent.script**  
**restart**
  - Windows弹性云服务器
    - i. 使用快捷键“Win+R”，打开“运行”窗口。
    - ii. 输入“Services.msc”，并单击“确定”。

图 14-116 运行



- iii. 右键单击服务“cloud reset password update agent”，选择“重新启动”。

图 14-117 服务（本地）



## 14.8.4 一键式重置密码插件占用虚拟内存 VIRT 和常驻内存 SHR 过高怎么办？

### 问题描述

在弹性云服务器运行时，一键式重置密码插件占用虚拟内存VIRT和常驻内存SHR过高。

图 14-118 查看虚拟内存占用情况

```
top - 14:56:06 up 4 days, 3:22, 1 user, load average: 0.00, 0.02, 0.05
Tasks: 1 total, 0 running, 1 sleeping, 0 stopped, 0 zombie
%Cpu(s): 0.5 us, 0.5 sy, 0.0 ni, 99.0 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem : 3864028 total, 137704 free, 183268 used, 3543056 buff/cache
KiB Swap: 0 total, 0 free, 0 used. 3358852 avail Mem
```

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
18240	root	20	0	2513828	61876	13036	S	0.0	1.6	0:00.75	java

### 问题根因

应用程序使用内存的时候通过glibc的内存池来提供，glibc 2.10之前的版本中，只有一个内存池，称为main arena。从glibc 2.10版本开始引入了thread arena，导致虚拟内存VIRT和常驻内存SHR过高。可通过限制一键式重置密码插件的可申请内存解决该问题。

### 处理方法

1. 修改一键式重置密码插件参数。
  - a. 执行以下命令查看配置文件。

```
vim /CloudResetPwdUpdateAgent/conf/wrapper.conf
```

图 14-119 wrapper.conf 修改前

```
# Initial Java Heap Size (in MB)
#wrapper.java.initmemory=16

# Maximum Java Heap Size (in MB)
#wrapper.java.maxmemory=64
```

- b. 将wrapper.java.initmemory=16和wrapper.java.maxmemory=64前的“#”去掉。

图 14-120 wrapper.conf 文件修改后

```
# Initial Java Heap Size (in MB)
wrapper.java.initmemory=16

# Maximum Java Heap Size (in MB)
wrapper.java.maxmemory=64
#
```

2. 执行以下命令重启一键式重置密码插件。  
`cd /CloudResetPwdUpdateAgent/bin/  
./cloudResetPwdUpdateAgent.script restart`

## 14.8.5 怎样查询弹性云服务器使用的密钥对是哪个？

### 问题描述


如果用户创建了多个密钥对，在登录弹性云服务器时，可能会迷惑哪一个密钥对才是正确的。此时，可以通过登录管理控制台，快速查询弹性云服务器使用的密钥对信息。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性云服务器”。
3. 在弹性云服务器列表页，选择待登录的弹性云服务器。
4. 单击弹性云服务器的名称。  
系统进入该弹性云服务器的详情页。
5. 查看参数“密钥对”的值。  
该值即为当前弹性云服务器使用的密钥对信息。

## 14.8.6 无法导入密钥对，怎么办？

当您的浏览器是IE9时，可能无法导入密钥对，请参考如下步骤修改浏览器默认属性后重试。

1. 在浏览器主界面，单击。
2. 选择“Internet选项”。
3. 单击选择“安全”页签。
4. 单击“Internet”。
5. 如果安全级别显示为“自定义”，单击“默认级别”按钮，把设置还原为默认级别。

6. 滑动安全级别滑块，把安全级别调到“中”级别，单击“应用”按钮。
7. 选择“自定义级别”。
8. 将“对未标记为可安全执行脚本的ActiveX控件初始化并执行脚本”设置为“提示”。
9. 单击“确定”。

## 14.8.7 为什么使用密钥文件无法正常登录 Linux 弹性云服务器？

### 问题描述

用户使用创建弹性云服务器时使用的密钥文件登录Linux弹性云服务器时，登录失败。

### 可能原因

根据Linux弹性云服务器使用的镜像不同，可能会存在如下原因：

- 原因一：Linux弹性云服务器的镜像为用户自己制作的私有镜像，且在创建该私有镜像时用户未安装Cloud-init工具。
- 原因二：Linux弹性云服务器的镜像安装了Cloud-init工具，但是在创建弹性云服务器时，获取密钥失败。

### 处理方法

- 针对原因一：  
创建私有镜像时不安装Cloud-init工具，将无法对弹性云服务器进行自定义配置，此时，用户只能使用镜像原有密码或密钥登录弹性云服务器。  
其中，镜像原有密码或密钥指创建私有镜像时，用户自己设置的操作系统密码或密钥。
- 针对原因二：
  - a. 勾选待获取密钥的弹性云服务器，单击“重启”，重新启动弹性云服务器。
  - b. 重启成功后，再次使用密钥文件登录弹性云服务器，确认是否能够正常登录。
    - 是，结束。
    - 否，请联系管理员寻求技术支持。

## 14.8.8 通过 puttygen.exe 工具创建的密钥对，导入管理控制台失败怎么办？

### 问题描述

通过puttygen.exe工具创建的密钥对，在导入管理控制台使用时，系统提示导入公钥文件失败。

### 可能原因

公钥内容的格式不符合系统要求。

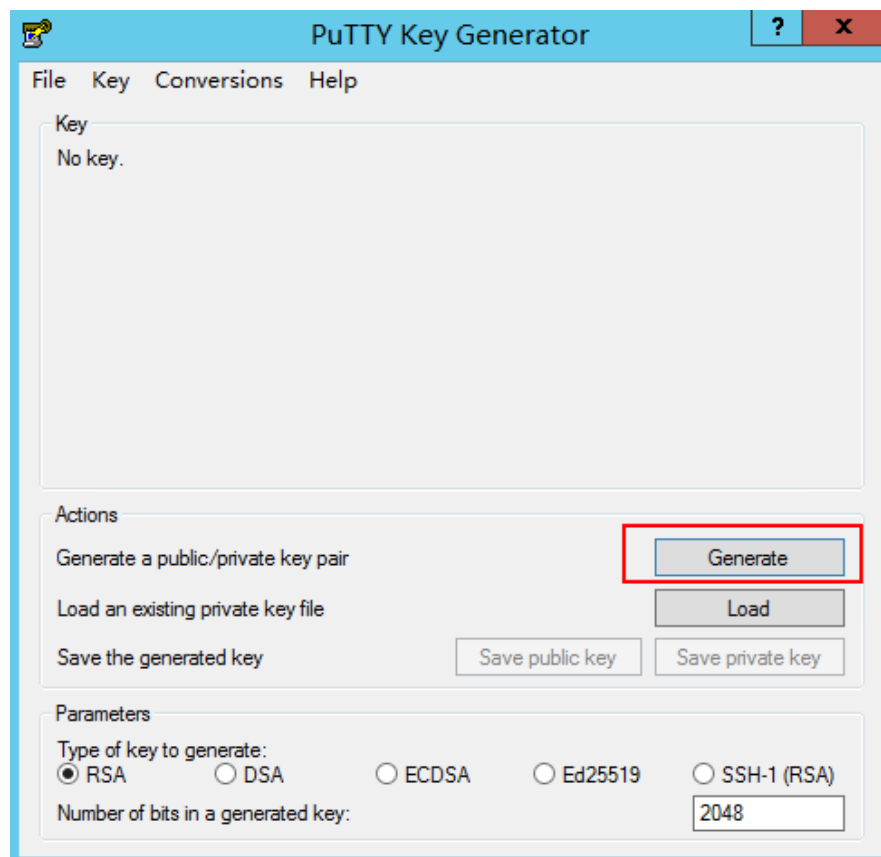
当用户使用puttygen.exe工具创建密钥对时，如果使用puttygen.exe工具的“Save public key”按钮保存公钥，公钥内容的格式会发生变化，不能直接导入管理控制台使用。

## 处理方法

使用本地保存的私钥文件，在“PuTTY Key Generator”中恢复内容格式正确的公钥文件，然后再将该公钥文件导入管理控制台。

1. 双击“puttygen.exe”，打开“PuTTY Key Generator”。

图 14-121 PuTTY Key Generator

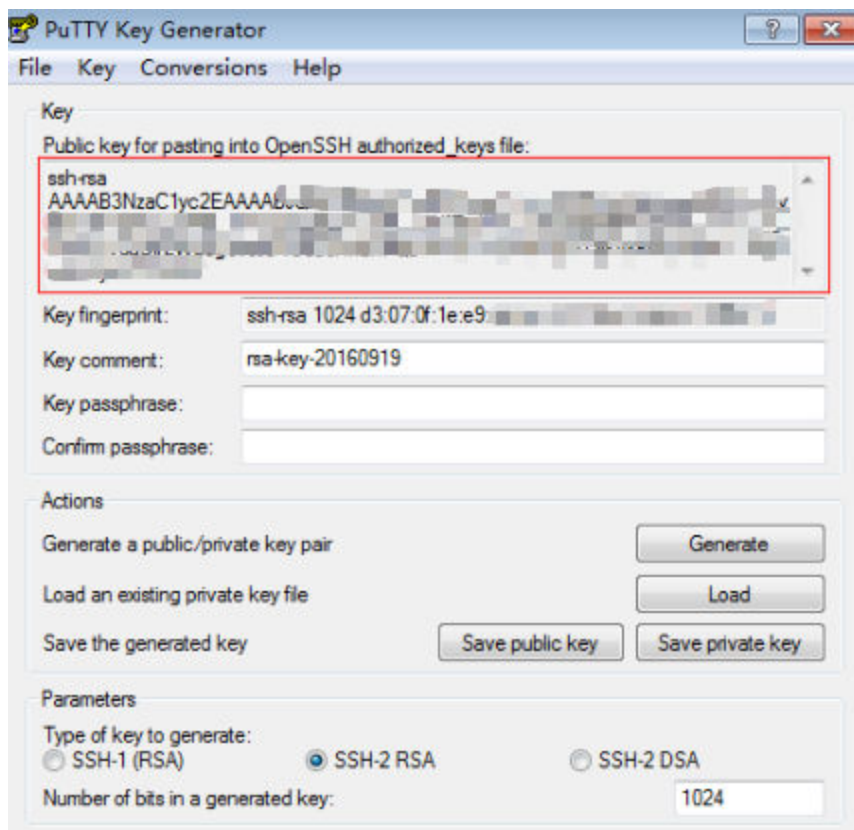


2. 单击“Load”，并在本地选择该密钥对的私钥文件。

系统将自动加载该私钥文件，并在“PuTTY Key Generator”中恢复格式正确的公钥文件内容，如图14-122所示，红框中的内容即为符合系统要求的公钥文件。



图 14-122 恢复公钥文件内容



3. 复制红框中的公钥内容，并将其粘贴在文本文档中，以.txt格式保存在本地，保存公钥文件。
4. 将公钥文件导入管理控制台。
  - a. 登录管理控制台。
  - b. 选择“计算 > 弹性云服务器”。
  - c. 在左侧导航树中，选择“密钥对”。
  - d. 在“密钥对”页面，单击“导入密钥对”。
  - e. 将“.txt”格式文本文档中的公钥内容粘贴至“Public Key Content”的空白区域，并单击“OK”，导入公钥文件。

## 14.8.9 Windows 弹性云服务器中的 cloudbase-init 账户是什么？

### cloudbase-init 账户是什么？

Windows弹性云服务器中的cloudbase-init账户为Cloudbase-Init代理程序的内置账户，用于弹性云服务器启动的时候获取元数据并执行相关配置。如果删除此账户，会影响云管理平台的相关功能，建议您不要修改、删除此账户。

#### 说明

Linux弹性云服务器中不存在该账户。

如果自行修改、删除此账户或者卸载Cloudbase-Init代理程序，会导致由此弹性云服务器创建的Windows私有镜像所生成的新云服务器初始化的自定义信息注入失败。



## cloudbase-init 账户密码随机化安全加固说明

在cloudbase-init 0.9.10版本中，对cloudbase-init内置账户密码进行随机化安全加固，确保cloudbase-init内置账户密码的HASH值（LM-HASH和NTLM-HASH值）不一致。

Windows系统下的HASH密码格式为：用户名称:RID:LM-HASH值:NT-HASH值，

例如：

Administrator:500:C8825DB10F2590EAAAD3B435B51404EE:683020925C5D8569C23AA724774CE9CC:::表示

- 用户名称为：Administrator
- RID为：500
- LM-HASH值为：C8825DB10F2590EAAAD3B435B51404EE
- NT-HASH值为：683020925C5D8569C23AA724774CE9CC

验证：使用同一个镜像创建两个云服务器（ecs01、ecs02），cloudbase-init内置账户HASH值不同。

- ecs01的cloudbase-init内置账户LM-HASH和NTLM-HASH值如下所示。

图 14-123 ecs01

```
----- BEGIN DUMP -----
cloudbase-init:1003:AAD3B435B51404EEAAD3B435B51404EE:CCA3BDDEB517A0E2342AEB34C0473C39:::
Guest:501:AAD3B435B51404EEAAD3B435B51404EE:31D6CFE0D16AE931B73C59D7E0C089C0:::
Administrator:500:AAD3B435B51404EEAAD3B435B51404EE:27CF57575EB83D9A6D7D27831157A947:::
----- END DUMP -----
3 dumped accounts
```

- ecs02的cloudbase-init内置账户LM-HASH和NTLM-HASH值如下所示。

图 14-124 ecs02

```
----- BEGIN DUMP -----
cloudbase-init:1003:AAD3B435B51404EEAAD3B435B51404EE:5B635D5F5306E26E66915D7C1CA98:::
Guest:501:AAD3B435B51404EEAAD3B435B51404EE:31D6CFE0D16AE931B73C59D7E0C089C0:::
Administrator:500:AAD3B435B51404EEAAD3B435B51404EE:0501525C0083243750D23927A82070B6:::
----- END DUMP -----
3 dumped accounts
```

## 14.8.10 python 升级导致 Cloud-init 不工作

### 问题描述

以CentOS 6.8镜像的弹性云服务器为例，将python从python 2.6升级到python 2.7版本后，可能会引起Cloud-init不工作，具体表现为：弹性云服务器的密码、密钥、hostname等信息无法通过Cloud-init注入。

执行命令**cloud-init -v**查询Cloud-init的版本，回显报错，如**图14-125**所示。

图 14-125 Cloud-init 运行异常

```
[root@ecs-8560 ~]# cloud-init -v
Traceback (most recent call last):
  File "/usr/bin/cloud-init", line 39, in <module>
    from cloudinit import patcher
ImportError: No module named cloudinit
[root@ecs-8560 ~]# cloud-init init --local
Traceback (most recent call last):
  File "/usr/bin/cloud-init", line 39, in <module>
    from cloudinit import patcher
ImportError: No module named cloudinit
[root@ecs-8560 ~]#
```

## 可能原因

Cloud-init使用的python版本不正确。

## 处理方法

将Cloud-init使用的python版本修改为升级前版本，即将/usr/bin/cloud-init环境变量从默认的“#!/usr/bin/python”修改为升级前的“#!/usr/bin/python2.6”路径。

图 14-126 修改 python 版本

```
[root@ecs-8560 ~]# head -n 1 /usr/bin/cloud-init
#!/usr/bin/python2.6
[root@ecs-8560 ~]# ls /usr/bin/python* -lh
lrwxrwxrwx 1 root root 24 Jul 19 10:55 /usr/bin/python -> /usr/local/bin/python2.7
lrwxrwxrwx 1 root root 6 Jun 9 2017 /usr/bin/python2 -> python
-rwxr-xr-x 1 root root 8.9K Aug 18 2016 /usr/bin/python2.6
```

## 14.9 应用搭建及软件安装咨询

### 14.9.1 弹性云服务器上是否可以搭建数据库？

可以，不限制。您可以安装任何类型的数据库。

### 14.9.2 弹性云服务器是否支持 Oracle 数据库？

支持，但是建议您在正式使用前，先做性能测试看是否能满足您的需要。

## 14.10 文件上传/数据传输

### 14.10.1 上传文件到云服务器方式概览

#### 怎样上传文件到 Windows 操作系统云服务器？

- 安装传输工具

在本地主机和Windows云服务器上分别安装数据传输工具，将文件上传到云服务器。例如FileZilla，操作步骤请参考[本地Windows主机使用FTP上传文件到Windows/Linux云服务器](#)。

- 本地磁盘映射（推荐使用）  
使用远程桌面连接MSTSC方式进行数据传输。该方式不支持断点续传，可能存在传输中断的情况，因此不建议上传大文件。  
详细操作请参考[本地Windows主机通过MSTSC上传文件到Windows云服务器](#)。
- 搭建FTP站点  
常用文件传输方式，需提前在Windows云服务器搭建FTP站点，在本地主机安装FileZilla工具上传文件到Windows系统云服务器。  
详细操作请参考[本地Windows主机使用FTP上传文件到Windows/Linux云服务器](#)。
- 本地为macOS系统主机  
如果本地主机为macOS操作系统，您可以通过安装“Microsoft Remote Desktop for Mac”工具向Windows云服务器传输文件。详细操作请参考[本地macOS系统主机上传文件到Windows云服务器](#)。

## 怎样上传文件到 Linux 操作系统云服务器？

- 本地为Windows操作系统主机  
使用WinSCP工具上传文件到Linux系统云服务器，详细操作请参考[本地Windows主机使用WinSCP上传文件到Linux云服务器](#)。  
需提前在Linux云服务器搭建FTP站点，在本地主机安装FileZilla工具上传文件到Linux系统云服务器。详细操作请参考[本地Windows主机使用FTP上传文件到Windows/Linux云服务器](#)。
- 本地为Linux操作系统主机  
使用SCP工具上传文件到Linux系统云服务器，详细操作请参考[本地Linux主机使用SCP上传文件到Linux云服务器](#)  
使用SFTP工具上传文件到Linux系统云服务器，详细操作请参考[本地Linux主机使用SFTP上传文件到Linux云服务器](#)  
使用FTP站点上传文件到Linux系统云服务器，详细操作请参考[本地Linux主机使用FTP上传文件到Linux云服务器](#)

## 云服务器自带 FTP 上传吗？

FTP需要用户自己安装和配置。

## 14.10.2 本地 Windows 主机通过 MSTSC 上传文件到 Windows 云服务器

### 操作场景

将文件上传至Windows云服务器一般会采用MSTSC远程桌面连接的方式。本节为您介绍本地Windows计算机通过远程桌面连接，上传文件至Windows云服务器的操作方法。

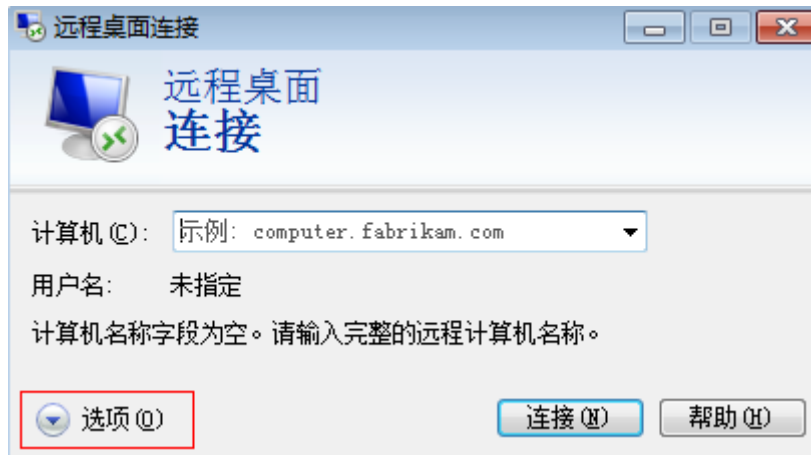
### 前提条件

- 弹性云服务器状态为“运行中”。
- 弹性云服务器已经绑定弹性公网IP，绑定方式请参见[绑定弹性公网IP](#)。

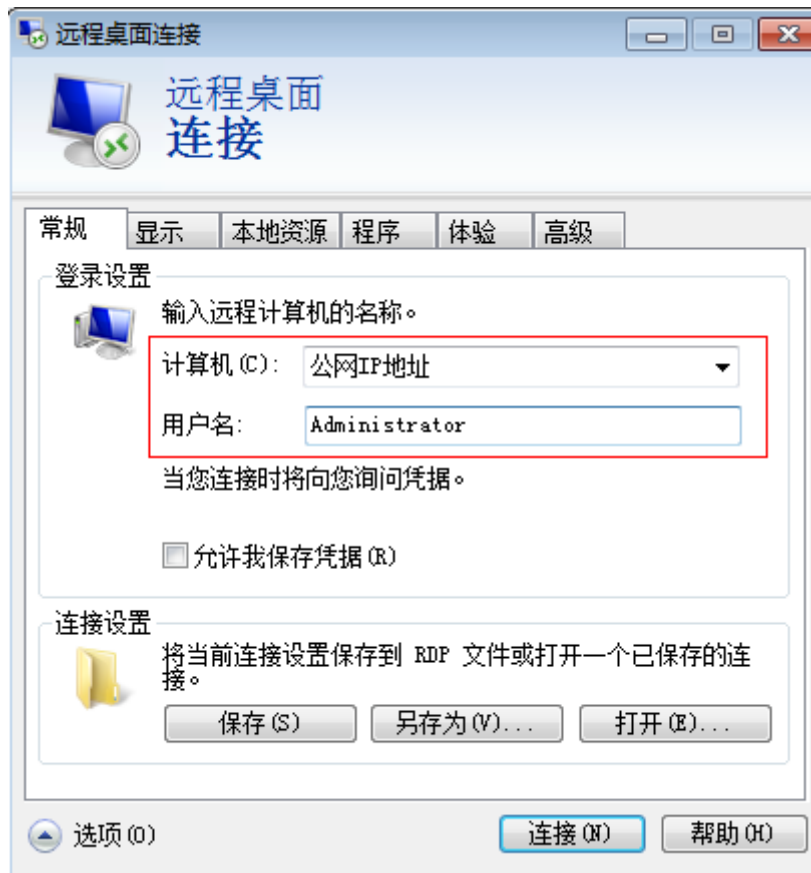
- 弹性云服务器所在安全组入方向已开放3389端口，配置方式请参见[配置安全组规则](#)。

## 处理方法

- 在本地Windows计算机上，单击“开始”，在出现的“搜索程序和文件”输入框中输入mstsc。  
弹出远程桌面连接对话框。
- 单击“选项”。



- 在“常规”页签中，输入云服务器的公网IP地址和用户名“Administrator”。



4. 选择“本地资源”页签，确认“本地设备和资源”栏的“剪贴板”处于勾选状态。

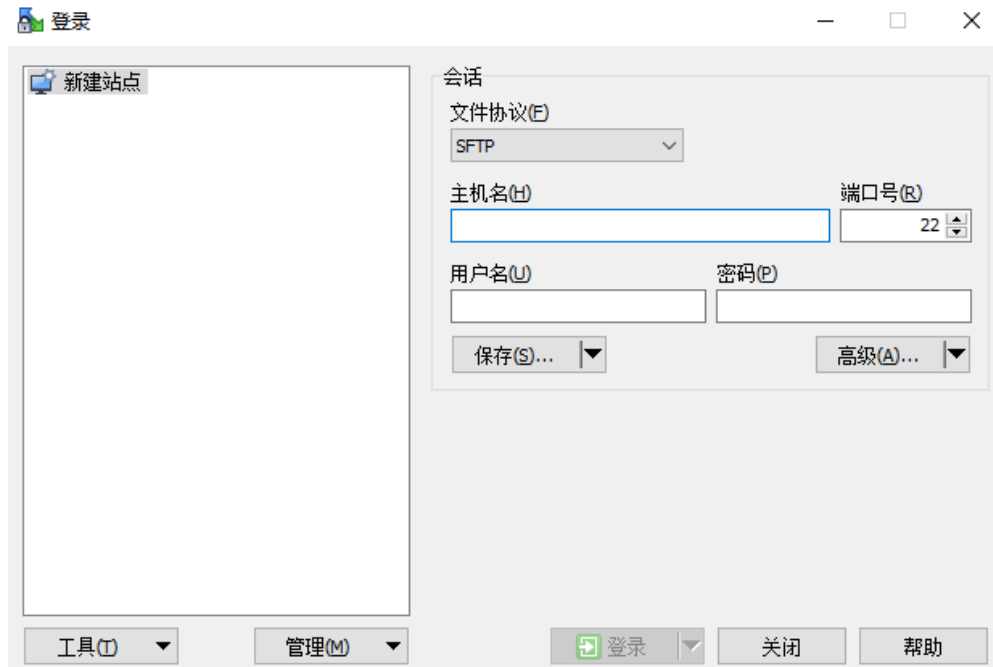


5. 单击“详细信息”。
6. 在“驱动器”多选框列表，勾选要上传到Windows云服务器上的文件所在的本地磁盘。



## 处理方法

1. 下载 WinSCP 客户端并安装。[下载WinSCP](#)。
2. 安装WinSCP。
3. 启动WinSCP，启动后界面如下：



填写说明：

- 协议：选填 SFTP 或者 SCP 均可。
  - 主机名：云服务器的公网 IP。登录管理控制台即可查看对应云服务器的公网 IP。
  - 端口：默认 22。
  - 用户名：云服务器的用户名。
    - 使用“密码方式”登录云服务器，公共镜像（包括CoreOS）的用户名为：root。
  - 密码：创建云服务器设置的密码。
4. 单击“登录”，进入“WinSCP”文件传输界面。
  5. 登录成功之后，您可以选择左侧本地计算机的文件，拖拽到右侧的远程云服务器，完成文件上传到云服务器。

### 14.10.4 本地 macOS 系统主机上传文件到 Windows 云服务器

#### 操作场景

本节操作介绍本地macOS系统主机通过安装“Microsoft Remote Desktop for Mac”工具向Windows云服务器传输文件的操作步骤。

#### 前提条件

- 本地主机已安装Microsoft Remote Desktop for Mac或其他macOS系统适用的远程连接工具。[下载Microsoft Remote Desktop for Mac](#)。

- Windows操作系统云服务器已绑定弹性公网IP。
- 首次登录云服务器时，请先使用VNC方式登录云服务器，打开RDP（Remote Desktop Protocol），然后再使用mstsc方式连接。

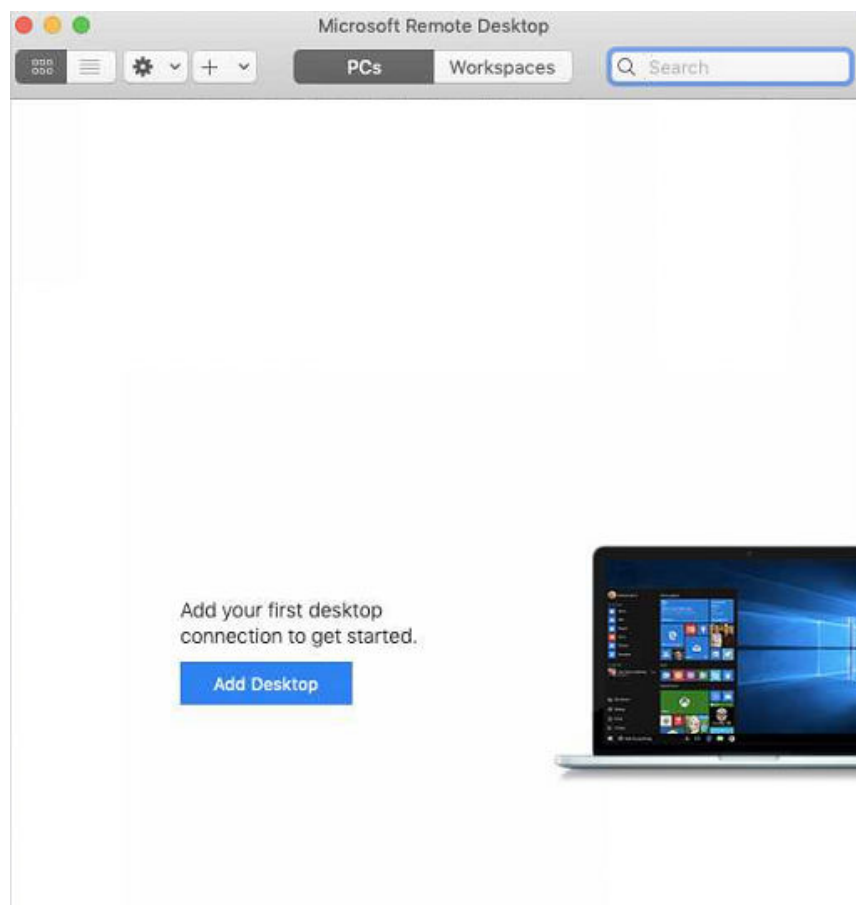
#### 📖 说明

使用公共镜像创建的云服务器，默认已打开RDP。

## 操作步骤

1. 启动Microsoft Remote Desktop。
2. 单击“Add Desktop”。

图 14-127 Add Desktop



3. 设置登录信息。
  - PC name: 输入需要登录的Windows实例的弹性公网IP地址。
  - User account: 在下拉列表中选择“Add user account”。弹出“Add user account”对话框。
    - i. 输入Windows实例账号“administrator”，并输入实例的登录密码，单击“Add”。



图 14-128 Add user account

**Add a User Account**

Username:

Password:   Show password

Friendly name:

图 14-129 Add PC

**Add PC**

PC name:

User account:

**General** | Display | Devices & Audio | **Folders**

Friendly name:

Group:

Gateway:

Bypass for local addresses

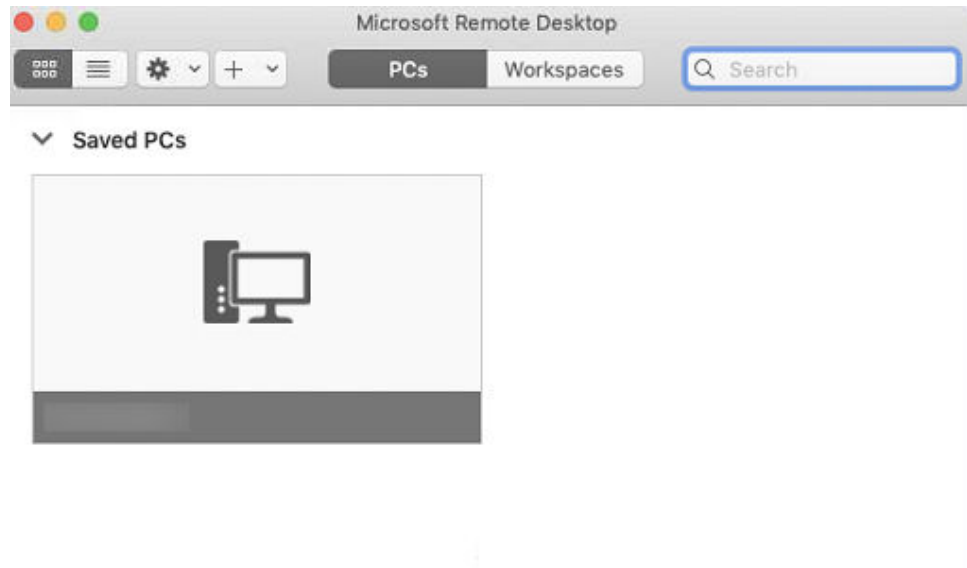
Reconnect if the connection is dropped

Connect to an admin session

Swap mouse buttons

4. 选择待上传的文件夹。
  - a. 单击“Folders”切换至选择文件夹列表。
  - b. 单击左下角 $\oplus$ ，选择需上传的文件夹，并单击“Add”。
5. 在“Remote Desktop”页面，双击需要登录的Windows实例图标。

图 14-130 双击登录 Windows 实例



6. 确认登录信息后，单击“Continue”。  
至此，您已经连接Windows实例。  
打开Windows云服务器，查看共享的文件夹。  
将需要上传的文件复制到Windows云服务器。或将Windows云服务器的文件下载到本地主机。

## 14.10.5 本地 Linux 主机使用 SCP 上传文件到 Linux 云服务器

### 操作场景

本节操作介绍本地Linux操作系统主机通过SCP向Linux云服务器传输文件的操作步骤。

### 操作步骤

登录管理控制台，在ECS列表页面记录待上传文件的云服务器的弹性公网IP。

- **上传文件**

在本地Linux操作系统主机上执行以下命令，传输文件到Linux操作系统云服务器。

**scp 本地主机文件地址 用户名@弹性公网IP:云服务器文件地址**

例如：将本地文件 /home/test.txt 上传至弹性公网IP地址为139.x.x.x的云服务器对应目录下，命令如下：

**scp /home/test.txt root@139.x.x.x:/home**

根据提示输入登录密码，即可完成上传。

图 14-131 上传文件

```
[root@ecs-5c83 home]# scp /home/test.txt root@139.█:~/home
root@139█'s password:
test.txt
```

- **下载文件**

在本地Linux操作系统主机上执行以下命令，下载云服务器上的文件到本地主机。

**scp 用户名@弹性公网IP:云服务器文件地址 本地主机文件地址**

例如，将弹性公网IP地址为139.x.x.x的云服务器文件/home/test.txt 下载至本地对应目录下，命令如下：

**scp root@139.x.x.x:/home/test.txt /home/**

根据提示输入登录密码，即可完成文件下载。

图 14-132 下载文件

```
[root@ecs-5c83 home]# scp root@139. :/home/test.txt /home
root@139.'s password:
test.txt
[root@ecs-5c83 home]# ls
test.txt
```

## 14.10.6 本地 Linux 主机使用 SFTP 上传文件到 Linux 云服务器

### 操作场景

本节操作以CentOS操作系统为例，介绍配置SFTP、使用SFTP上传或下载文件、文件夹的操作步骤。

### 操作步骤

1. 以root用户登录云服务器。
2. 执行以下命令查看ssh版本，OpenSSH版本大于等于4.8p1。

**ssh -V**

回显信息如下所示：

```
# OpenSSH_7.4p1, OpenSSL 1.0.2k-fips 26 Jan 2017
```

3. 创建用户和组，以user1用户为例。

**groupadd sftp**

**useradd -g sftp -s /sbin/nologin user1**

4. 设置用户密码。

**passwd user1**

图 14-133 设置用户密码

```
[root@ecs-9a32-0001 ~]# passwd user1
Changing password for user user1.
New password:
BAD PASSWORD: The password is shorter than 8 characters
Retype new password:
passwd: all authentication tokens updated successfully.
[root@ecs-9a32-0001 ~]#
```

5. 设置目录权限。  
**chown root:sftp /home/user1**  
**chmod 755 -R /home/user1**

- ```
mkdir /home/user1/upload
chown -R user1:sftp /home/user1/upload
chmod -R 755 /home/user1/upload
```
6. 执行以下命令，编辑sshd\_config文件。
- ```
vim /etc/ssh/sshd_config
```

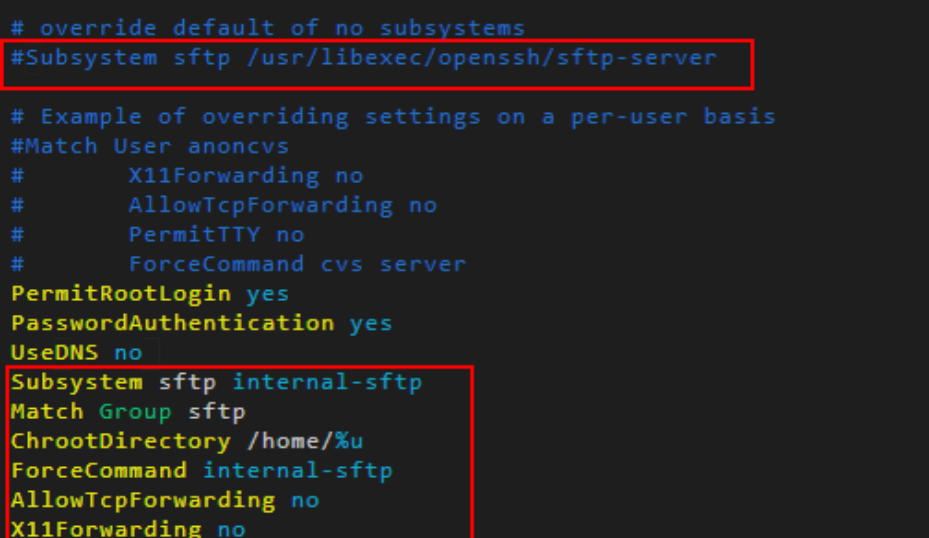
注释掉如下信息

```
#Subsystem sftp /usr/libexec/openssh/sftp-server
```

补充如下内容：

```
Subsystem sftp internal-sftp
Match Group sftp
ChrootDirectory /home/%u
ForceCommand internal-sftp
AllowTcpForwarding no
X11Forwarding no
```

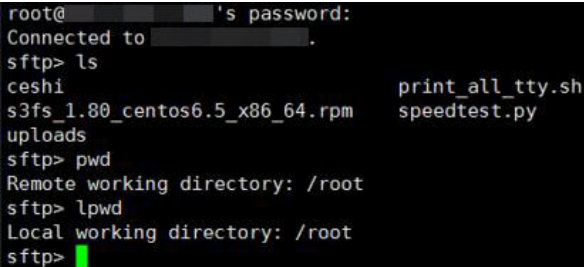
图 14-134 sshd\_config 文件补充配置信息后



```
# override default of no subsystems
#Subsystem sftp /usr/libexec/openssh/sftp-server

# Example of overriding settings on a per-user basis
#Match User anoncvs
#    X11Forwarding no
#    AllowTcpForwarding no
#    PermitTTY no
#    ForceCommand cvs server
PermitRootLogin yes
PasswordAuthentication yes
UseDNS no
Subsystem sftp internal-sftp
Match Group sftp
ChrootDirectory /home/%u
ForceCommand internal-sftp
AllowTcpForwarding no
X11Forwarding no
```

7. 重启云服务器，或执行以下命令重启sshd服务。
- ```
service sshd restart
```
- 或
- ```
systemctl restart sshd
```
8. 在本地主机执行以下命令，远程连接到服务器。
- ```
sftp root@IP地址
```
9. 连接成功后，您可以使用交互式的sftp命令。



```
root@ ~ 's password:
Connected to .
sftp> ls
ceshi          print_all_tty.sh
s3fs_1.80_centos6.5_x86_64.rpm  speedtest.py
uploads
sftp> pwd
Remote working directory: /root
sftp> lpwd
Local working directory: /root
sftp>
```

10. 执行以下命令，上传或下载文件、文件夹。

上传文件：put -r

```
sftp> put -r ceshi/  
Uploading ceshi/ to /root/ceshi  
Entering ceshi/  
ceshi/mysql57-community-release-el 100% 9224      9.0KB/s   00:00  
ceshi/haha                          100% 28        0.0KB/s   00:00  
sftp>
```

下载文件：get -r

```
sftp> get -r s3fs_1.80_centos6.5_x86_64.rpm  
Fetching /root/s3fs_1.80_centos6.5_x86_64.rpm to s3fs_1.80_centos6.5_x86_64.rpm  
/root/s3fs_1.80_centos6.5_x86_64.r 100% 3250KB   3.2MB/s   00:00  
sftp>
```

## 14.10.7 本地 Windows 主机使用 FTP 上传文件到 Windows/Linux 云服务器

### 操作场景

本节操作介绍如何在Windows操作系统的本地主机上使用FTP上传文件到云服务器。

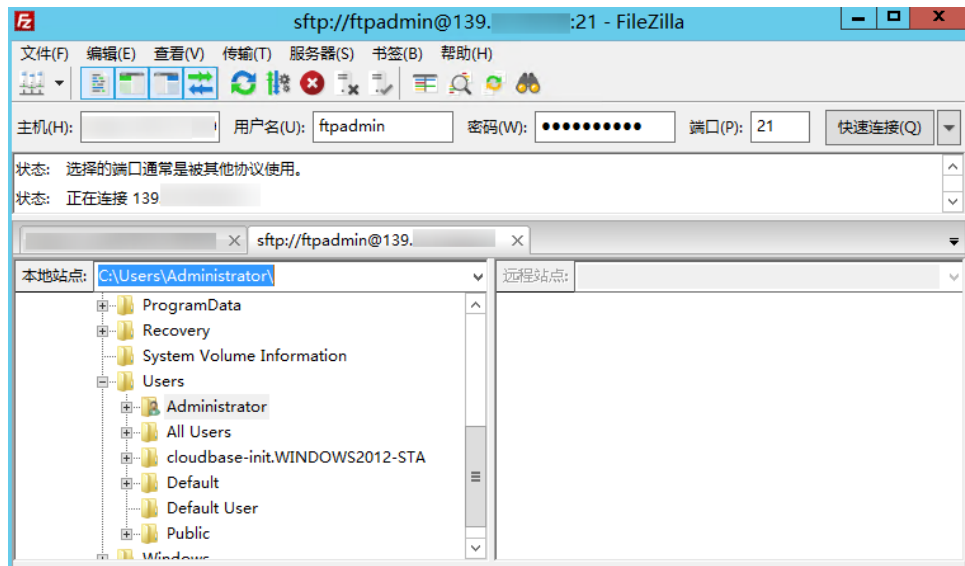
### 前提条件

- 云服务器已绑定弹性公网IP，且安全组入方向开启TCP协议的21端口。
- 已在待上传文件的云服务器中搭建 FTP 服务。

### 操作步骤

1. 在Windows本地主机上安装FileZilla。[下载FileZilla](#)。
2. 在Windows本地主机打开 FileZilla，填写待连接的云服务器信息，单击“快速连接”。
  - 主机：云服务器的弹性公网IP。
  - 用户名：搭建FTP时设置的用户名。
  - 密码：搭建FTP时设置的用户对应的密码。
  - 端口：FTP链接端口，默认使用21端口。

图 14-135 连接云服务器



3. 您可以选择左侧本地计算机的文件，拖拽到右侧的远程云服务器，完成文件上传到云服务器。

## 14.10.8 本地 Linux 主机使用 FTP 上传文件到 Linux 云服务器

### 操作场景

本文介绍如何在 Linux 系统的本地机器上使用 FTP 服务，将文件从本地上传到云服务器中。

### 前提条件

已在待上传文件的云服务器中搭建 FTP 服务。

- 云服务器已绑定弹性公网IP，且安全组入方向开启TCP协议的21端口。
- 已在待上传文件的云服务器中搭建 FTP 服务。

### 操作步骤

1. Linux操作系统的本地主机安装ftp。  
以CentOS 7.6操作系统为例，执行以下命令安装ftp  
**yum -y install ftp**
2. 执行以下命令连接云服务器。  
**ftp 云服务器弹性公网IP**  
并根据提示，输入FTP服务的用户名和密码。
  - 上传文件  
执行以下命令，将本地文件上传至云服务器中。  
**put 本地主机文件地址**  
例如：将本地Linux主机的/home/test.txt 上传到云服务器。  
**put /home/test.txt**

- 下载文件

执行以下命令，将云服务器中的文件下载至本地。

**get 云服务器文件地址 本地主机文件地址**

例如：将云服务器的test.txt下载到本地Linux主机。

**get /home/test.txt**

## 14.10.9 使用 FTP 上传文件时客户端连接服务端超时怎么办？

### 问题描述

客户端连接服务端超时，无法连接到服务端。

### 约束与限制

该文档适用于本地主机Windows系统上的FTP服务。

### 可能原因

服务端防火墙或安全组拦截。

### 处理方法

1. 检查服务端防火墙设置。
2. 关闭防火墙或者添加相应规则。

## 14.10.10 使用 FTP 上传文件时写入失败，文件传输失败怎么办？

### 问题描述

使用FTP上传文件时，写入失败，文件传输失败。

### 约束与限制

该文档适用于Windows系统上的FTP服务。

### 可能原因

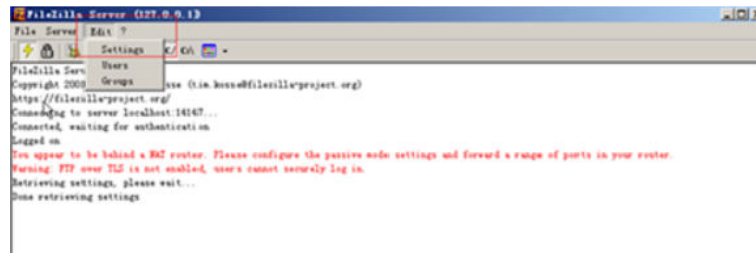
FTP服务端在NAT环境下，客户端需使用被动模式连接服务端。在这种情况下，服务端的IP地址无法从路由器外部访问，所以应该在服务端的对外IP中填写此服务器的公网IP（弹性公网IP），同时设置端口范围来限制需要通过路由器转发的端口数量。

### 处理方法

弹性公网IP与私有IP使用NAT方式绑定，所以需要在服务端做相应的配置。

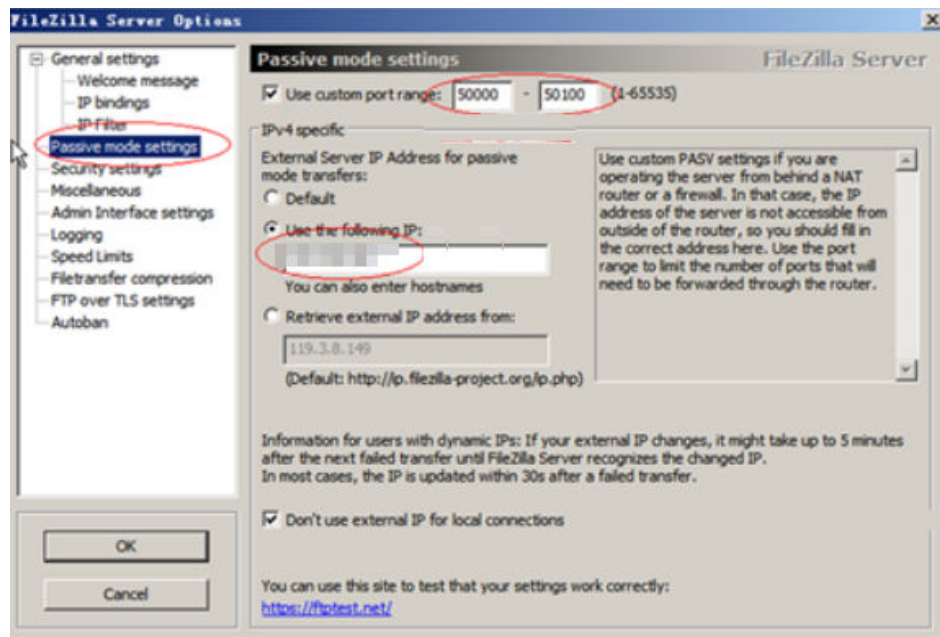
1. 配置服务端对外的公网IP。  
选择“Edit > Settings”。

图 14-136 配置服务端对外的公网 IP



2. 选择“Passive mode settings”，设置数据传输端口范围（如50000-50100），并填写弹性公网IP。

图 14-137 设置数据传输端口范围



3. 单击“OK”。
4. 安全组中开启入方向tcp协议的50000-50100和21端口。
5. 在客户端连接测试。

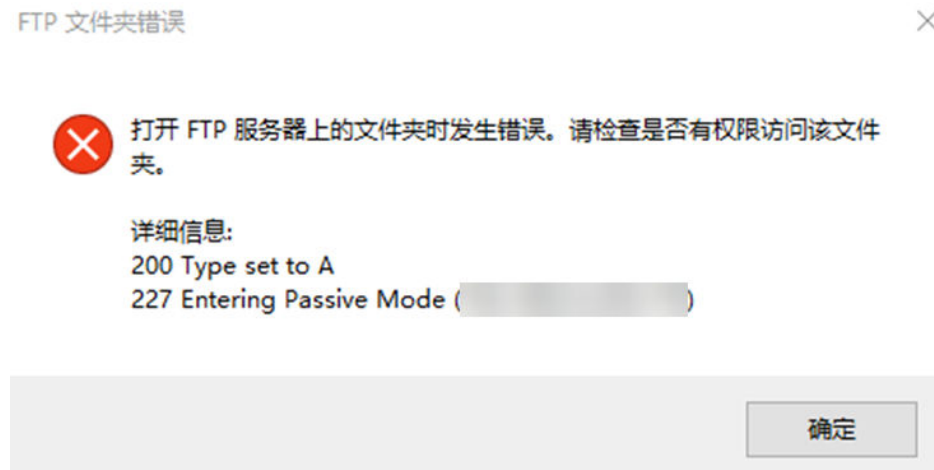
## 14.10.11 打开 FTP 服务器上的文件夹时发生错误，请检查是否有权 限访问该文件夹

### 问题描述

打开FTP服务器上的文件夹时发生错误，请检查是否有权访问该文件夹。



图 14-138 FTP 文件夹错误



## 可能原因

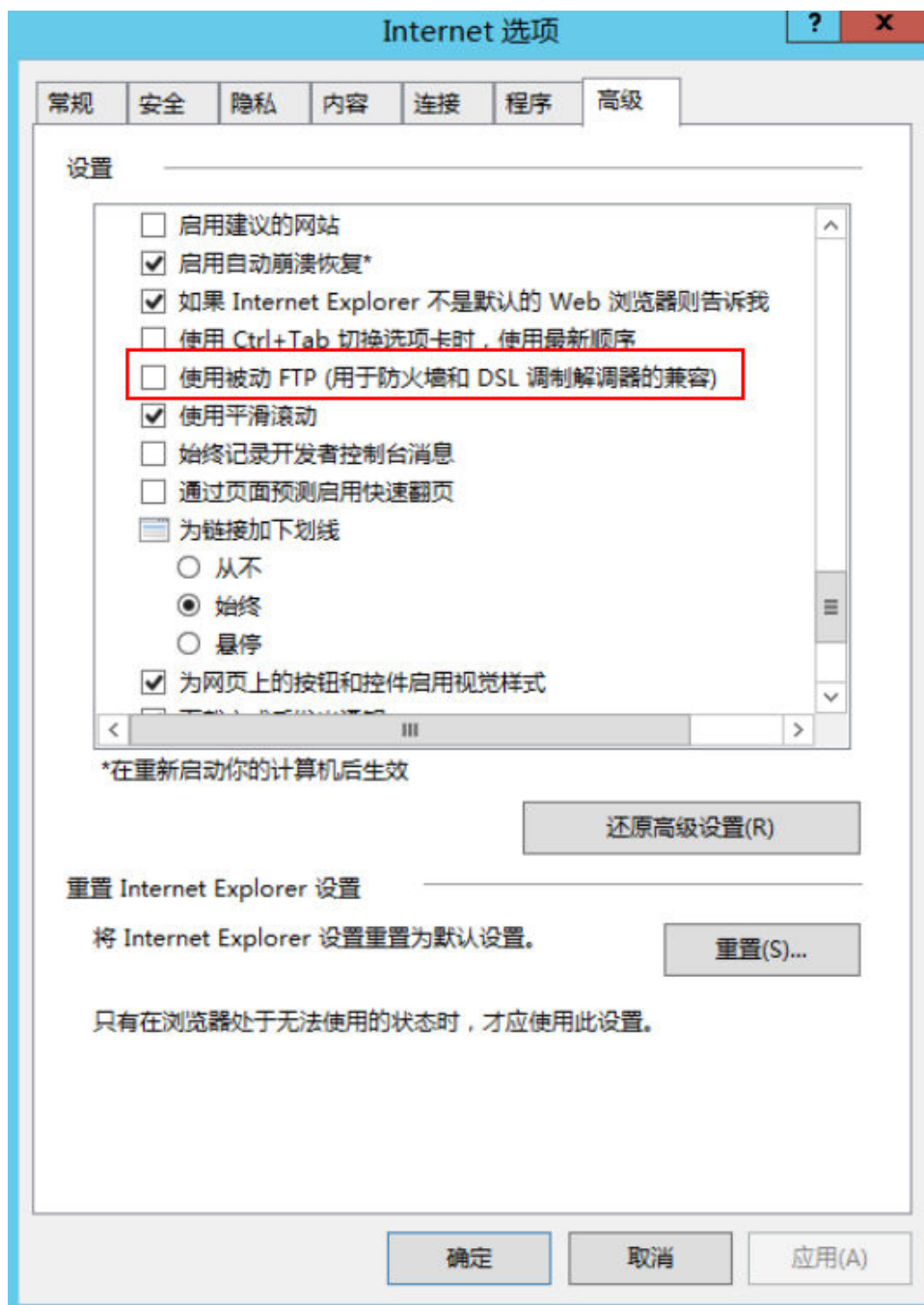
浏览器设置了FTP防火墙。

## 处理方法

以设置IE浏览器为例。

1. 打开IE浏览器菜单"工具 > Internet 选项"。
2. 选择“高级”标签卡。
3. 取消勾选“使用被动FTP（用于防火墙和DSL调制解调器兼容）”。

图 14-139 Internet 选项



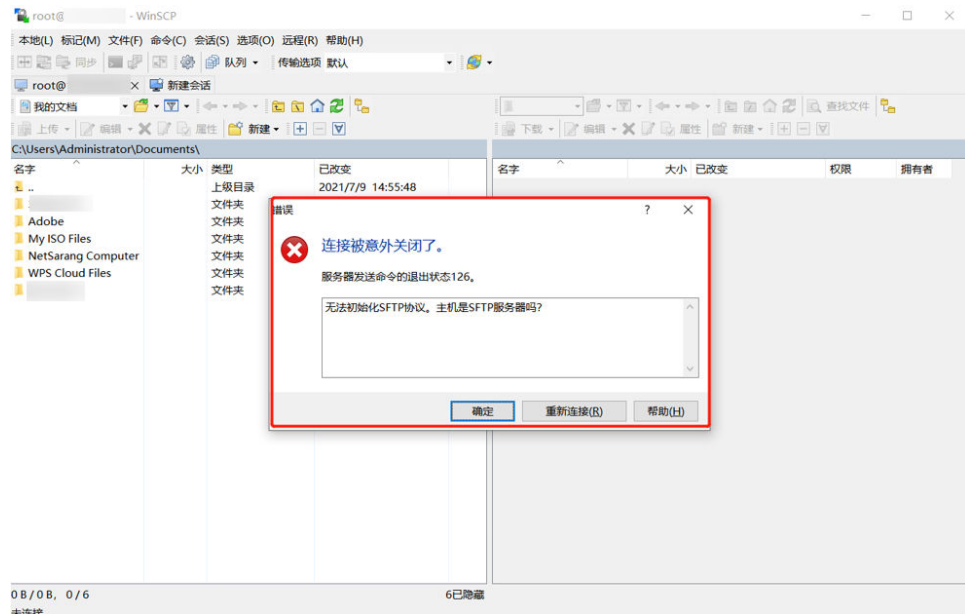
4. 单击“确定”，重启IE浏览器，重试打开FTP服务器上的文件夹。

## 14.10.12 WinSCP 无法连接到 Linux 云服务器怎么办？

### 问题现象

Winscp无法连接到服务器。SSH连接工具例如Xshell可以正常连接云服务器。

图 14-140 WinSCP 连接失败



## 根因分析

其他SSH工具连接云服务器正常，但是WinSCP无法连接到服务器。说明SSH服务是没有问题的，WinSCP连接基于的是SFTP协议。

执行以下命令，查看/etc/ssh/sshd\_config文件。

```
vi /etc/ssh/sshd_config
```

查看关于SFTP的配置描述，说明SFTP关联的配置文件是/usr/libexec/openssh/sftp-server

图 14-141 查看 SFTP 关联的配置文件

```
# override default of no subsystems
Subsystem sftp /usr/libexec/openssh/sftp-server
```

如果sftp-server文件缺失或者权限不对都会导致WinSCP连接异常，默认权限是755。

## 处理方法

- 如果文件缺失，在可以正常远程的情况下，通过scp或其他文件传输方式将正常云服务器上的文件传输到故障机器对应目录下。
- 如果文件权限有变化，可以通过chmod修改文件权限为755。

```
chmod 755 -R /usr/libexec/openssh/sftp-server
```

## 14.11 ECS 故障

## 14.11.1 针对管理控制台的异常提示信息，应该如何处理？

### 问题描述

您可以通过本节内容解决如下问题：

- 用户在管理控制台执行弹性云服务器相关操作后出现异常，针对管理控制台提示的异常信息，应该如何处理？
- 用户参见《弹性云服务器接口参考》调用云服务器相关的API接口时，如果返回错误码，应该如何处理？

### 背景知识

用户通过管理控制台执行弹性云服务器的相关操作后，弹性云服务器列表页面将显示相应操作的申请状态。通过申请状态中显示的信息，用户可以获悉当前操作请求的执行状态。

- 如果操作请求正常执行完毕，则任务提示栏中的记录将自动清除。
- 如果操作请求在执行过程中出现异常，则任务提示栏将返回错误码及其说明。

### 处理方法

您可以通过[表14-16](#)中的处理建议进行下一步操作，处理相应的异常。

表 14-16 错误码处理建议

| 错误码      | 管理控制台提示信息                            | 处理建议                                                                                                  |
|----------|--------------------------------------|-------------------------------------------------------------------------------------------------------|
| Ecs.0000 | 请求错误，请稍后重试或联系管理员。                    | 请参见《弹性云服务器接口参考》，按照要求调整请求结构体。                                                                          |
| Ecs.0001 | 租户弹性云服务器或云硬盘配额不足，请联系管理员申请扩大云服务器配额。   | 请联系管理员申请扩大弹性云服务器配额。<br><b>说明</b><br>您在申请云服务器配额时，需评估待申请的云服务器需要占用的云服务器个数、CPU核数（vCPU）以及内存（RAM）容量，统一调整配额。 |
| Ecs.0003 | 没有操作权限不足，请联系管理员检查账户情况。               | 请联系管理员检查账户情况。                                                                                         |
| Ecs.0005 | 参数非法，请参考FAQ或联系管理员。                   | 请参见《服务器接口参考》，按照要求调整请求结构体。                                                                             |
| Ecs.0010 | 私有IP地址已经被使用，请重新选择未使用的IP地址，进行云服务器的创建。 | 重新选择未使用的IP地址，进行云服务器的创建。                                                                               |

| 错误码      | 管理控制台提示信息                                   | 处理建议                                        |
|----------|---------------------------------------------|---------------------------------------------|
| Ecs.0011 | 密码不符合规则，请调整对应的密码，使其满足密码复杂度要求，并重新执行操作请求。     | 调整对应设置的密码，使其满足密码复杂度要求，并重新执行操作请求。            |
| Ecs.0012 | 子网IP地址不足，请清理所选子网内的空闲IP地址，或者选择新的子网进行云服务器的创建。 | 清理所选子网内的空闲IP地址，或者选择新的子网进行云服务器的创建。           |
| Ecs.0013 | 弹性公网IP配额不足，请联系管理员申请扩大弹性公网IP配额。              | 联系管理员申请扩大EIP配额。                             |
| Ecs.0015 | 该类型磁盘不适用于该类型云服务器。                           | 重新选择合适类型的磁盘来执行挂载云服务器操作。                     |
| Ecs.0100 | 云服务器状态不符合要求，请将云服务器变更至指定的状态后重试操作。            | 将云服务器变更至指定的状态后重试操作。                         |
| Ecs.0103 | 云磁盘状态不可用                                    | 将云服务器变更至指定的状态后重试操作。如果云磁盘状态故障，需要联系管理员进行排障处理。 |
| Ecs.0104 | 云服务器可挂载云硬盘槽位不足                              | 您需要调整该云服务器挂载的云硬盘，以保证新云硬盘可以挂载到指定弹性云服务器上      |
| Ecs.0105 | 云服务器无系统盘                                    | 您需要将系统盘挂载回云服务器后，再次执行相应操作。                   |
| Ecs.0107 | 云服务器可挂载共享云硬盘数量超过了可挂载的规格数量                   | 您需要调整该云服务器挂载的云硬盘，以保证新云硬盘可以挂载到指定云服务器上。       |
| 其他异常编码   | 其他异常提示                                      | 您可以选择重新执行操作请求，或者记录当前异常返回的错误码信息，联系管理员进行处理。   |

## 14.11.2 获取控制台日志时，系统显示问号是什么原因？

### 问题描述

获取弹性云服务器的控制台日志时，系统显示“？”。

### 可能原因

弹性云服务器使用的镜像支持查询控制台日志，但是未开启该功能。

## 处理方法

修改弹性云服务器的配置信息，打开获取控制台日志功能。

具体操作请参见[获取控制台日志](#)的步骤一。

### 14.11.3 使用 free 命令查看弹性云服务器的内存，为什么与实际不符？

#### 问题描述

弹性云服务器创建成功后，使用**free -m**命令查询内存大小，查询结果与实际配置不符，较之创建时的配置要小一些。

示例：

假设创建该弹性云服务器时，配置的实际内存大小为4194304KB（即4096MB）。创建完成后，使用**free -m**命令查询内存大小，回显如下：

```
[root@localhost ~]# free -m
total used free shared buff/cache available
Mem: 3790 167 3474 8 147 3414
Swap: 1022 0 1022
```

可以看到内存只有3790MB，小于实际的4096MB。

但是使用**dmidecode -t memory**命令查看实际的硬件内存大小时，回显如下：

```
[root@localhost ~]# dmidecode -t memory
# dmidecode 3.0
Getting SMBIOS data from sysfs.
SMBIOS 2.8 present.

Handle 0x1000, DMI type 16, 23 bytes
Physical Memory Array
Location: Other
Use: System Memory
Error Correction Type: Multi-bit ECC
Maximum Capacity: 4 GB
Error Information Handle: Not Provided
Number Of Devices: 1

Handle 0x1100, DMI type 17, 40 bytes
Memory Device
Array Handle: 0x1000
Error Information Handle: Not Provided
Total Width: Unknown
Data Width: Unknown
Size: 4096 MB
Form Factor: DIMM
Set: None
Locator: DIMM 0
Bank Locator: Not Specified
Type: RAM
Type Detail: Other
Speed: Unknown
Manufacturer: QEMU
Serial Number: Not Specified
Asset Tag: Not Specified
Part Number: Not Specified
Rank: Unknown
Configured Clock Speed: Unknown
Minimum Voltage: Unknown
Maximum Voltage: Unknown
Configured Voltage: Unknown
```

可以看到使用 `dmidecode -t memory` 查看的内存大小与实际配置一致。

## 问题原因

首先，系统启动时会初始化相关设备，该过程会占用内存，内核启动时，也会占用一部分的内存。其中，`kdump`占用的内存是可以自行设置的，如无特殊要求，请勿自行修改`kdump`占用的内存大小。

其次，`free -m`命令查询的是弹性云服务器的可用内存，`dmidecode -t memory`命令查询的是实际硬件内存大小。

因此，使用`free -m`命令查询到的内存大小比实际的要小一些，属于正常情况，非问题。

### 说明

物理机同样存在该问题。

## 14.11.4 弹性云服务器的主机名带后缀 `novalocal`，这正常吗？

### 问题描述

用户使用`hostname`命令查看不同镜像的弹性云服务器主机名，发现部分镜像的弹性云服务器主机名带后缀“`.novalocal`”，如示例所示：

假设创建弹性云服务器时，用户自定义的主机名是“`abc`”，使用`hostname`命令查看不同镜像下，弹性云服务器的主机名以及重启弹性云服务器后的主机名，显示结果如表14-17所示。

表 14-17 不同镜像查询的主机名

| 镜像         | 重启前查询的主机名     | 重启后再次查询的主机名   |
|------------|---------------|---------------|
| CentOS 6.8 | abc           | abc.novalocal |
| CentOS 7.3 | abc.novalocal | abc.novalocal |
| Ubuntu 16  | abc           | abc           |

不同镜像的弹性云服务器，查询的主机名有的带后缀“`.novalocal`”，有的不带后缀“`.novalocal`”，这正常吗？

### 问题处理

正常现象。

Linux弹性云服务器的静态主机名来源于创建弹性云服务器时，通过Cloud-init注入的用户自定义名称。经测试验证发现，Cloud-init和不同发行版本的操作系统在配合实现上，存在差异，具体表现为：查询的主机名有的带后缀“`.novalocal`”，有的不带后缀“`.novalocal`”。

如果您希望查询到的主机名不带后缀“`.novalocal`”，可以通过更改主机名进行规避，修改主机名的方法请参见[怎样使修改的静态主机名永久生效？](#)。

## 14.11.5 怎样使修改的静态主机名永久生效?

### 问题描述

Linux弹性云服务器的静态主机名来源于创建弹性云服务器时，通过Cloud-init注入的用户自定义名称。虽然可以使用hostname命令更改Linux弹性云服务器的主机名，但如果重新启动弹性云服务器，它将自动还原为使用Cloud-init注入的用户自定义名称。

### 修改弹性云服务器实例的主机名（修改 hostname）

如果您希望修改后的主机名在弹性云服务器关机、重启后仍然生效，需要将其写入相应的Cloud-init配置文件中，使之永久生效。

假设修改后的主机名为“new\_hostname”。

1. 修改配置文件“/etc/hostname”。
  - a. 执行以下命令，编辑配置文件“/etc/hostname”。  
**sudo vim /etc/hostname**
  - b. 修改配置文件，将文件中的主机名替换为修改后的主机名（new\_hostname）。
  - c. 执行以下命令，保存并退出文件。  
**:wq**
2. 修改配置文件“/etc/sysconfig/network”。
  - a. 执行以下命令，编辑配置文件“/etc/sysconfig/network”。  
**sudo vim /etc/sysconfig/network**
  - b. 修改涉及“HOSTNAME”的参数值，将其设置为修改后的主机名（new\_hostname）。  
**HOSTNAME=修改后的主机名**

#### 说明

如果配置文件中没有HOSTNAME，请手动补充并将参数值设置为修改后的主机名。

示例：

```
HOSTNAME=new_hostname
```

- c. 执行以下命令，保存并退出文件。  
**:wq**
3. 修改配置文件“/etc/cloud/cloud.cfg”。
    - a. 执行以下命令，编辑配置文件“/etc/cloud/cloud.cfg”。  
**sudo vim /etc/cloud/cloud.cfg**
    - b. 选择如下方法，修改配置文件。
      - 方法一：修改或添加“preserve\_hostname”参数项。  
如果文件“/etc/cloud/cloud.cfg”中已有参数项“preserve\_hostname: false”，将其修改为“preserve\_hostname: true”即可。如果文件“/etc/cloud/cloud.cfg”中没有该参数项，需在“cloud\_init\_modules”模块前，添加语句**preserve\_hostname: true**。  
选择方法一，弹性云服务器关机、重启后，修改后的主机名（new\_hostname）仍然生效。但是，如果使用该弹性云服务器制作私



有镜像，并使用生成的镜像重新创建弹性云服务器，则新创建的弹性云服务器的主机名即为私有镜像使用的主机名（`new_hostname`），不能通过Cloud-init重新注入用户自定义名称。

- 方法二（推荐）：删除或者注释语句- `update_hostname`。  
选择方法二，弹性云服务器关机、重启后，修改后的主机名（`new_hostname`）仍然生效。但是，如果使用该弹性云服务器制作私有镜像，并使用生成的镜像重新创建弹性云服务器，用户可以通过Cloud-init注入用户自定义名称（如`new_new_hostname`），并通过本节内容长久有效地修改静态主机名。

4. 执行以下命令，重启弹性云服务器。

```
sudo reboot
```

5. 执行以下命令，验证主机名的修改是否永久生效。

```
sudo hostname
```

如果回显的内容是修改后的主机名（`new_hostname`），则表示主机名修改成功，永久生效。

## 修改弹性云服务器实例的主机名与 IP 地址的映射（修改 hosts）

如果您希望将修改后的主机名作为首选的主机名、本地域和本地主机，在修改完主机名后，还需要更新主机名与IP地址的映射并将配置写入相应的Cloud-init配置文件中，使之永久生效。

假设修改后的主机名为“`new_hostname`”。

1. 修改配置文件“`/etc/hostname`”。
  - a. 执行以下命令，编辑配置文件“`/etc/hostname`”。

```
sudo vim /etc/hostname
```
  - b. 修改配置文件，将文件中的主机名替换为修改后的主机名（`new_hostname`）。
  - c. 执行以下命令，保存并退出文件。

```
:wq
```
2. 修改配置文件“`/etc/sysconfig/network`”。
  - a. 执行以下命令，编辑配置文件“`/etc/sysconfig/network`”。

```
sudo vim /etc/sysconfig/network
```
  - b. 修改涉及“`HOSTNAME`”的参数值，将其设置为修改后的主机名（`new_hostname`）。

```
HOSTNAME=修改后的主机名
```

### 说明

如果配置文件中没有`HOSTNAME`，请手动补充并将参数值设置为修改后的主机名。

示例：

```
HOSTNAME=new_hostname
```

- c. 执行以下命令，保存并退出文件。

```
:wq
```

3. 修改配置文件“`/etc/cloud/cloud.cfg`”。

- a. 执行以下命令，编辑配置文件“/etc/cloud/cloud.cfg”。  
**sudo vim /etc/cloud/cloud.cfg**
- b. 选择如下方法，修改配置文件。
  - 方法一：修改或添加“preserve\_hostname”参数项。  
如果文件“/etc/cloud/cloud.cfg”中已有参数项“preserve\_hostname: false”，将其修改为“preserve\_hostname: true”即可。如果文件“/etc/cloud/cloud.cfg”中没有该参数项，需在“cloud\_init\_modules”模块前，添加语句**preserve\_hostname: true**。  
选择方法一，弹性云服务器关机、重启后，修改后的主机名（new\_hostname）仍然生效。但是，如果使用该弹性云服务器制作私有镜像，并使用生成的镜像重新创建弹性云服务器，则新创建的弹性云服务器的主机名即为私有镜像使用的主机名（new\_hostname），不能通过Cloud-init重新注入用户自定义名称。
  - 方法二（推荐）：删除或者注释语句- **update\_hostname**。  
选择方法二，弹性云服务器关机、重启后，修改后的主机名（new\_hostname）仍然生效。但是，如果使用该弹性云服务器制作私有镜像，并使用生成的镜像重新创建弹性云服务器，用户可以通过Cloud-init注入用户自定义名称（如new\_new\_hostname），并通过本节内容长久有效地修改静态主机名。
4. 将“/etc/hosts”文件中hostname和ip的映射更新为以 127.0.0.1开头的条目，将“new\_hostname”作为您首选的主机名、本地域和本地主机。
  - a. 执行以下命令，编辑“/etc/hosts”文件。  
**sudo vim /etc/hosts**
  - b. 修改127.0.0.1开头的条目，将localhost、localhost.localdomain替换为“new\_hostname”。

```
127.0.0.1 localhost localhost.localdomain localhost6 localhost6.localdomain6
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
127.0.0.1 new_hostname new_hostname
```
  - c. 执行以下命令，保存并退出文件。  
**:wq**
5. 修改配置文件“/etc/cloud/cloud.cfg”。
  - a. 执行以下命令，编辑配置文件“/etc/cloud/cloud.cfg”。  
**sudo vim /etc/cloud/cloud.cfg**
  - b. 将“manage\_etc\_hosts”参数项配置为“manage\_etc\_hosts: false”。

```
manage_etc_hosts: false
```
  - c. 执行以下命令，保存并退出文件。  
**:wq**
6. 执行以下命令，重启弹性云服务器。  
**sudo reboot**
7. 执行以下命令，验证主机名和hosts的修改是否永久生效。  
**sudo hostname**  
**sudo cat /etc/hosts**  
如果回显的内容是修改后的主机名（new\_hostname）以及hosts，则表示修改成功，永久生效。

## 14.11.6 Linux 操作系统云服务器无法获取元数据怎么办？

### 问题描述

Linux操作系统云服务器的安全组出方向已按照[元数据获取](#)的前提条件进行配置，但是无法通过169.254.169.254获取元数据。

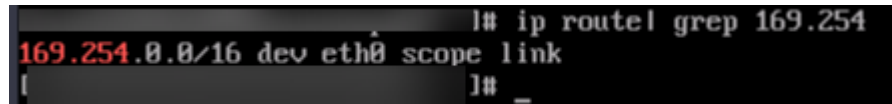
### 根因分析

在配置了静态IP的Linux系统云服务器中执行以下命令：

```
# ip route | grep 169.254
```

如果169.254.169.254路由不存在，并且169.254.0.0/16路由存在。

图 14-142 查看服务器路由信息



```
1# ip route | grep 169.254
169.254.0.0/16 dev eth0 scope link
1#
```

如[图14-142](#)所示，说明Linux操作系统云服务器配置了静态IP，重启网络服务后部分Linux操作系统服务器会将原有的169.254.169.254路由替换为169.254.0.0/16路由，后者没有指定下一跳。导致Linux操作系统云服务器无法获取元数据。

### 处理方法

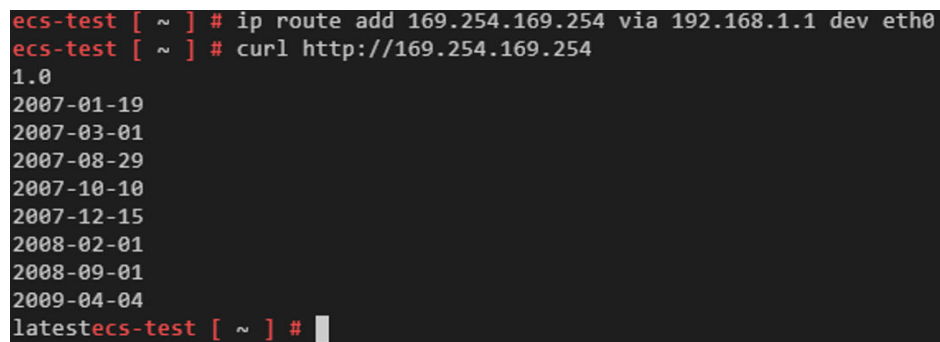
1. 添加169.254.169.254路由，指定下一跳网关和主网卡，示例：  

```
# ip route add 169.254.169.254 via 192.168.1.1 dev eth0
```

其中192.168.1.1是主网卡对应子网的网关，eth0是主网卡名称。
2. 执行以下命令，验证获取元数据的功能正常。  

```
# curl http://169.254.169.254
```

图 14-143 验证获取元数据的功能



```
ecs-test [ ~ ] # ip route add 169.254.169.254 via 192.168.1.1 dev eth0
ecs-test [ ~ ] # curl http://169.254.169.254
1.0
2007-01-19
2007-03-01
2007-08-29
2007-10-10
2007-12-15
2008-02-01
2008-09-01
2009-04-04
latestecs-test [ ~ ] #
```

3. 执行以下命令，创建或修改/etc/sysconfig/network-scripts/route-eth0文件，固化静态路由防止重启失效。

```
# vi /etc/sysconfig/network-scripts/route-eth0
```

添加如下内容：

本例以eth0，IP为192.168.1.1为例，请根据实际情况替换对应的网卡和网关IP。

# 169.254.169.254 via 192.168.1.1

## 14.12 ECS 卡顿

### 14.12.1 Windows 云服务器卡顿怎么办？

当您发现云服务器的运行速度变慢或云服务器突然出现网络断开现象，则可能是云服务器的带宽和CPU使用率过高导致。如果您已经通过云监控服务创建过告警任务，当CPU或带宽利用率高时，系统会自动发送告警给您。

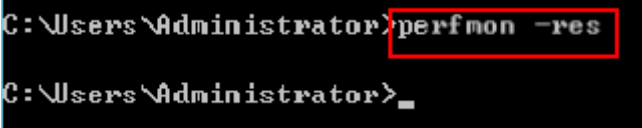
当Windows云服务器带宽流量过高或CPU利用率高，您可以按如下步骤进行排查：

1. 问题定位：定位影响云服务器带宽和CPU利用率高的进程和不明来源驱动。  
Windows操作系统本身提供了较多工具可以定位问题，包括任务管理器、性能监视器(Performance Monitor)、资源监视器(Resource Monitor)、Process Explorer、Xperf (Windows server 2008 以后)和抓取系统Full Memory Dump检查。
2. 问题处理：排查进程和驱动是否正常，并分类进行处理。
  - 正常进程：优化程序，或变更云服务器配置。
  - 异常进程：建议您手动关闭进程，您也可以借助第三方工具关闭进程。
  - 正规来源驱动：系统自带驱动不进行处理，第三方软件驱动根据您的需求决定是否卸载对应软件。
  - 不明来源驱动：建议您卸载不明驱动，您也可以使用正规商业杀毒软件或第三方安全管理工具进行删除。

#### 问题定位步骤

1. 在管理控制台使用VNC方式登录云服务器。
2. 打开“运行”窗口，输入“perfmon -res”。

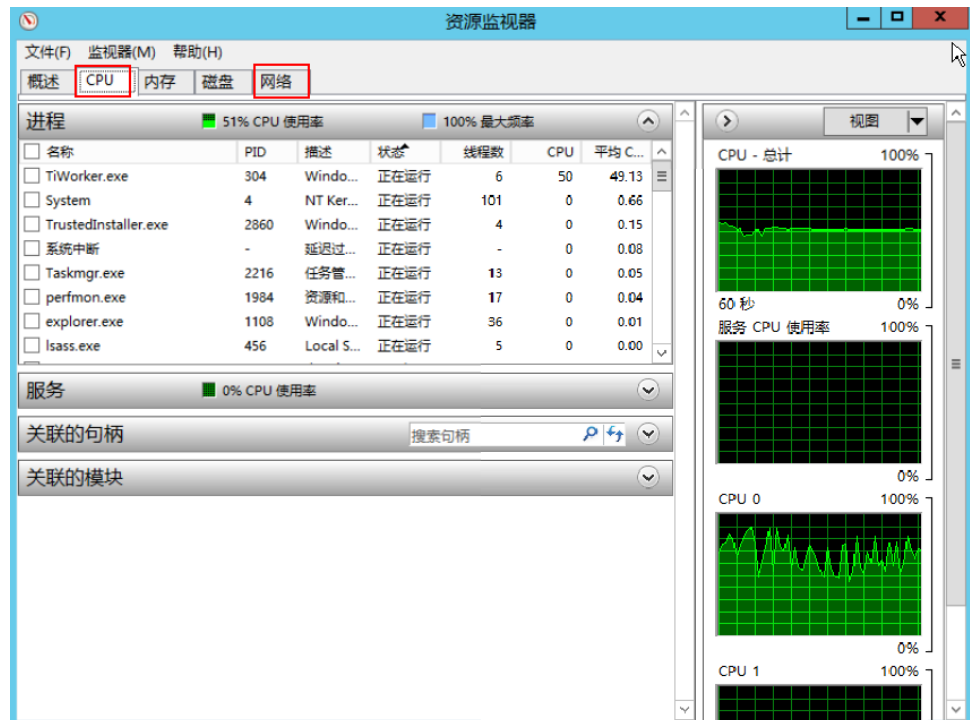
图 14-144 打开资源监视器



```
C:\Users\Administrator>perfmon -res
C:\Users\Administrator>_
```

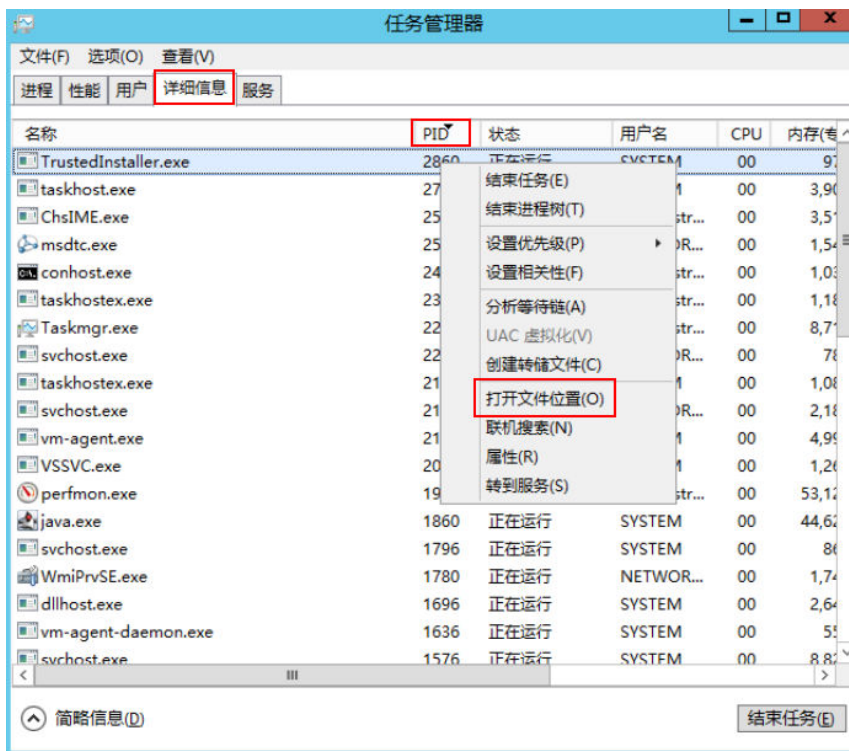
3. 在“资源监视器”中，单击“CPU”或“网络”，查看CPU占用率或带宽使用情况。

图 14-145 资源监视器



4. 查看CPU和带宽占用率较高的进程ID和进程名。
5. 在控制台VNC登录页面单击“Ctrl+Alt+Del”，打开“Windows任务管理器”。或打开“运行”窗口，输入“taskmgr”，打开“Windows任务管理器”。以下步骤为您介绍在任务管理器中打开PID，找到进程的具体位置，核对是否异常进程。
  - a. 选择“详细信息”选项卡。
  - b. 单击PID进行排序。
  - c. 在查找到的CPU或带宽占用率高的进程上右键单击“打开文件位置”。
  - d. 定位进程是否是正常或是否为恶意程序。

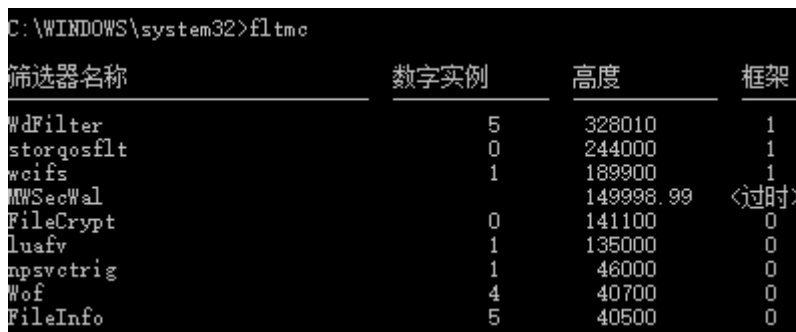
图 14-146 检查进程



6. 打开“运行”窗口，输入“fltmc”，查看系统的文件系统过滤驱动。

下图以windows10操作系统为例，不同操作系统内置驱动不同，请以官网网站说明为准。如果安装了第三方的驱动，也会在这个列表中显示。

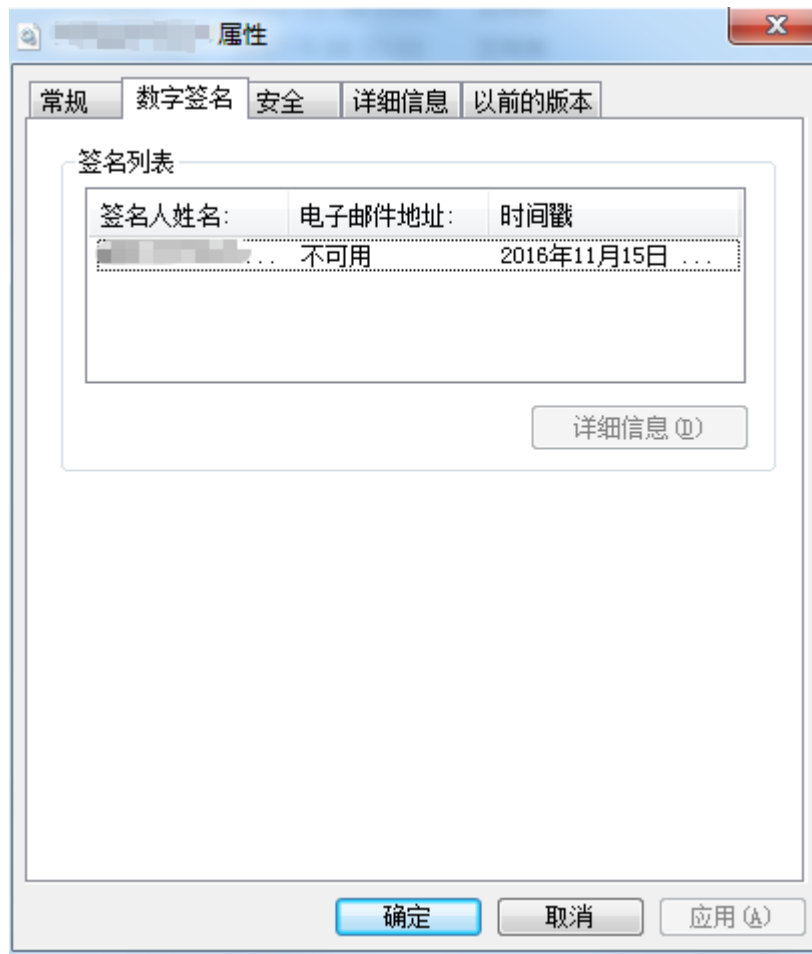
图 14-147 查看系统驱动



以下步骤为您介绍如何查看驱动的来源，核对是否为不明来源驱动。

- a. 打开系统路径“C:\Windows\System32\drivers”。
- b. 在不明驱动名称上单击，选择“属性”，查看详细信息。
- c. 选择“数字签名”，查看驱动的来源。

图 14-148 查看驱动来源



## 分析处理

在您采取措施处理问题前，首先需要判断影响CPU或带宽占用率高的进程和驱动是否正常，并分类进行处理。

### 正常进程分析处理建议

1. 如果您的操作系统是Windows 2008/Windows 2012，请检查内存大小，建议内存配置在2GB或以上。
2. 检查后台是否有执行Windows Update的行为。
3. 检查杀毒软件是否正在后台执行扫描操作。
4. 核对云服务器运行的应用程序中是否有对网络和CPU要求高的需求，如果是，建议您变更云服务器配置或修改带宽。
5. 如果云服务器配置已经比较高，建议考虑云服务器上应用场景的分离部署，例如将数据库和应用分开部署。

### 异常进程分析处理建议

如果CPU或带宽利用率高是由于病毒、木马入侵导致的，那么需要手动结束进程。建议的处理顺序如下：

1. 使用商业版杀毒软件或安装微软安全工具 **Microsoft Safety Scanner**，在安全模式下扫描病毒。

2. 安装Windows最新补丁。
3. 使用MSconfig禁用所有非微软自带服务驱动，检查问题是否再次发生，具体请参考微软官方文档《如何在Windows中执行干净启动》。

### 不明来源驱动分析处理建议

有些病毒和木马会通过文件系统过滤驱动加载。如果您发现不明来源的驱动，建议您卸载该驱动，也可以使用正规商业杀毒软件或第三方安全管理工具进行删除。

如果发现有无法删除的不明驱动，或者删除后还会再次出现的不明驱动，一般都是病毒或木马的驱动。如果使用正规商业杀毒软件或第三方安全管理工具也不能彻底删除，建议您重装操作系统，在这之前请做好数据备份避免造成损失。

## 14.12.2 Linux 云服务器卡顿怎么办？

当您发现云服务器的运行速度变慢或云服务器突然出现网络断开现象，则可能是云服务器的带宽和CPU使用率过高导致。如果您已经通过云监控服务创建过告警任务，当CPU或带宽利用率高时，系统会自动发送告警给您。

当Linux实例带宽流量过高或CPU使用率高时，您可以按如下步骤进行排查：

1. 问题定位：定位影响云服务器带宽和CPU使用率高的进程。
2. 问题处理：排查进程是否正常，并分类进行处理。
  - 正常进程：优化程序，或变更云服务器配置。
  - 异常进程：建议您手动关闭进程，或者借助第三方工具关闭进程。

### 常用命令

本文相关操作命令以CentOS 7.2 64位操作系统为例。其它版本的Linux操作系统命令可能有所差异，具体情况请参阅相应操作系统的官方文档。

Linux云服务器查看CPU使用率等性能相关问题时的常用命令如下：

- `ps -aux`
- `ps -ef`
- `top`

### CPU 占用率高问题定位

1. 使用VNC功能登录云服务器。
2. 执行如下命令查看当前系统的运行状态。

#### top

系统回显样例如下：

```
top - 20:56:02 up 37 days, 9:09, 1 user, load average: 0.00, 0.01, 0.05
Tasks: 80 total, 1 running, 79 sleeping, 0 stopped, 0 zombie
%Cpu(s): 0.2 us, 0.3 sy, 0.0 ni, 99.5 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem : 3880024 total, 2963304 free, 178384 used, 738336 buff/cache
KiB Swap: 0 total, 0 free, 0 used. 3434888 avail Mem

  PID USER      PR  NI   VIRT   RES   SHR  S  %CPU  %MEM     TIME+ COMMAND
 8115 root        20   0 161896   2216  1564 R   0.3   0.1   0:00.01 top
    1 root        20   0 125480   3884  2604 S   0.0   0.1   0:11.32 systemd
    2 root        20   0     0     0     0 S   0.0   0.0   0:00.00 kthreadd
    3 root        20   0     0     0     0 S   0.0   0.0   0:00.04 ksoftirqd/0
    5 root        0 -20     0     0     0 S   0.0   0.0   0:00.00 kworker/0:0H
    7 root        rt   0     0     0     0 S   0.0   0.0   0:00.18 migration/0
    8 root        20   0     0     0     0 S   0.0   0.0   0:00.00 rcu_bh
    9 root        20   0     0     0     0 S   0.0   0.0   7:32.18 rcu_sched
   10 root        0 -20     0     0     0 S   0.0   0.0   0:00.00 lru-add-drain
```



## 3. 查看显示结果。

- 命令回显第一行：20:56:02 up 37 days, 1 user, load average: 0.00, 0.01, 0.05的每个字段含义如下：

系统当前时间为20:56:02，该云服务器已运行37天，当前共有1个用户登录，最近1分钟、最近5分钟和最近15分钟的CPU平均负载。

- 命令回显第三行：CPU资源总体使用情况。
- 命令回显第四行：内存资源总体使用情况。
- 回显最下方显示各进程的资源占用情况。

### 📖 说明

1. 在top页面，可以直接输入小写“q”或者在键盘上按“Ctrl+C”退出。
2. 除了直接输入命令，您还可以单击VNC登录页面屏幕右上角的“Input Command”，在弹出的对话框中粘贴或者输入相应命令，单击“Send”。
3. 在top运行中常用的内容命令如下：
  - s: 改变画面更新频率。
  - l: 关闭或开启第一部分第一行top信息的表示。
  - t: 关闭或开启第一部分第二行Tasks和第三行Cpus信息的表示。
  - m: 关闭或开启第一部分第四行Mem和 第五行Swap信息的表示。
  - N: 以PID的大小的顺序排列进程列表。
  - P: 以CPU占用率大小的顺序排列进程列表。
  - M: 以内存占用率大小的顺序排列进程列表。
  - h: 显示命令帮助。
  - n: 设置在进程列表所显示进程的数量。
4. 通过ll /proc/PID/exe命令可以查看每个进程ID对应的程序文件。

```
root@elb-mg01 sysconfig# ll /proc/4243/exe  
lrwxrwxrwx 1 root root 0 Mar 18 11:46 /proc/4243/exe -> /CloudResetPwdUpdateAgent/depend/jre1.8.0_131/bin/java
```

## CPU 使用率高问题处理

对于导致CPU使用率高的具体进程，如果确认是异常进程，可以直接通过top命令终止进程。对于kswapd0进程导致的CPU使用率高的问题，则需要对应用程序进行优化，或者通过增加内存进行系统规格的升级。

kswapd0是系统的虚拟内存管理程序，如果物理内存不够用，系统就会唤醒kswapd0进程，由kswapd0分配磁盘交换空间用作缓存，因而占用大量的CPU资源。

- 使用top命令终止CPU占用率高的进程

您可以直接在top运行界面快速终止相应的异常进程。操作步骤如下：

- a. 在top命令运行的同时，按下小写的“k”键。
- b. 输入要终止进程的PID。

进程的PID为top命令回显的第一列数值。例如，要终止PID为52的进程，直接输入“52”后回车。

```
top - 21:07:38 up 37 days, 9:21, 1 user, load average: 0.01, 0.02, 0.05  
Tasks: 81 total, 1 running, 79 sleeping, 1 stopped, 0 zombie  
%Cpu(s): 0.0 us, 3.2 sy, 0.0 ni, 96.8 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st  
KiB Mem : 3880024 total, 2961520 free, 178960 used, 739544 buff/cache  
KiB Swap: 0 total, 0 free, 0 used. 3434216 avail Mem  
PID to signal/kill [default pid = 1] 52  
PID USER PR NI VIRT RES SHR S %CPU %MEM TIME+ COMMAND  
1 root 20 0 125480 3884 2604 S 0.0 0.1 0:11.32 systemd  
2 root 20 0 0 0 0 S 0.0 0.0 0:00.00 kthreadd
```

- c. 操作成功后，会出现如下图所示类似信息，按回车确认。

```
top - 21:07:38 up 37 days, 9:21, 1 user, load average: 0.01, 0.02, 0.05
Tasks: 81 total, 1 running, 79 sleeping, 1 stopped, 0 zombie
%Cpu(s): 0.0 us, 3.2 sy, 0.0 ni, 96.8 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem : 3880024 total, 2961520 free, 178960 used, 739544 buff/cache
KiB Swap: 0 total, 0 free, 0 used. 3434216 avail Mem
Send pid 52 signal [15/sigterm]
  PID USER      PR  NI   VIRT   RES   SHR  S  %CPU  %MEM     TIME+ COMMAND
    1 root        20   0 125480  3884  2604  S   0.0   0.1   0:11.32 systemd
    2 root        20   0     0     0     0   S   0.0   0.0   0:00.00 kthreadd
```

- kswapd0进程占用导致CPU使用率高  
可通过以下步骤排查进程的内存占用情况。
  - a. 通过top命令查看kswapd0进程的资源使用。
  - b. 如果kswapd0进程持续处于非睡眠状态，且运行时间较长，可以初步判定系统在持续的进行换页操作，可以将问题转向内存不足的原因来排查。

```
Tasks: 81 total, 1 running, 79 sleeping, 1 stopped, 0 zombie
%Cpu(s): 0.2 us, 52.2 sy, 0.0 ni, 99.7 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem : 3880024 total, 3014820 free, 179024 used, 686180 buff/cache
KiB Swap: 0 total, 0 free, 0 used. 3433948 avail Mem
  PID USER      PR  NI   VIRT   RES   SHR  S  %CPU  %MEM     TIME+ COMMAND
    36 root        20   0     0     0     0   S  99.0   0.0  964:10.45 kswapd0
 4595 nginx      20   0 125392  3576  1040  S   0.3   0.1   60:04.91 nginx
    1 root        20   0 125480  3884  2604  S   0.0   0.1   0:11.47 systemd
```

- c. 通过vmstat命令进一步查看系统虚拟内存的使用情况。  
如果si和so的值也比较高，说明系统存在频繁的换页操作，系统物理内存不足。
  - si：每秒从交换区写到内存的大小，由磁盘调入内存。
  - so：每秒写入交换区的内存大小，由内存调入磁盘。
- d. 对于内存不足问题，可以通过free、ps等命令进一步查询系统及系统内进程的内存占用情况，做进一步排查分析。
- e. 临时可通过在业务空闲期重启应用或者系统释放内存。  
如果要从根本上解决内存不足的问题，需要对服务器内存进行扩容，扩大内存空间。如果不具备扩容的条件，可通过优化应用程序，以及配置使用大页内存来进行缓解。

## 带宽使用率高问题分析

如果是正常业务访问以及正常应用进程导致的带宽使用率高，需要升级服务器的带宽进行解决。如果是非正常访问，如某些特定IP的恶意访问，或者服务器遭受到了CC攻击。或者异常进程导致的带宽使用率高。可以通过流量监控工具nethogs来实时监测统计各进程的带宽使用情况，并进行问题进程的定位。

- 使用nethogs工具进行排查
  - a. 执行以下命令，安装nethogs工具。  
**yum install nethogs -y**  
安装成功后可以通过netgos命令查看网络带宽的使用情况。  
nethogs命令常用参数说明如下：
    - -d：设置刷新的时间间隔，默认为 1s。

- -t: 开启跟踪模式。
- -c: 设置更新次数。
- device: 设置要监测的网卡, 默认是eth0。

运行时可以输入以下参数完成相应的操作:

- q: 退出nethogs工具。
  - s: 按发送流量大小的顺序排列进程列表。
  - r: 按接收流量大小的顺序排列进程列表。
  - m: 切换显示计量单位, 切换顺序依次为KB/s、KB、B、MB。
- b. 执行以下命令, 查看指定的网络端口每个进程的网络带宽使用情况。

### nethogs eth1

```
NetHogs version 0.8.5
```

| PID   | USER  | PROGRAM                                | DEV  | SENT   | RECEIVED     |
|-------|-------|----------------------------------------|------|--------|--------------|
| 4596  | nginx | nginx: worker process                  | eth1 | 34.360 | 3.267 KB/sec |
| ?     | root  | 192.168.0.92:90-100.125.68.19:17873    |      | 0.179  | 0.246 KB/sec |
| ?     | root  | 192.168.0.92:11211-213.32.10.149:44945 |      | 0.000  | 0.000 KB/sec |
| ?     | root  | 192.168.0.92:20101-105.176.26.66:43408 |      | 0.000  | 0.000 KB/sec |
| ?     | root  | unknown TCP                            |      | 0.000  | 0.000 KB/sec |
| TOTAL |       |                                        |      | 34.540 | 3.512 KB/sec |

回显参数说明如下:

- PID: 进程 ID。
  - USER: 运行该进程的用户。
  - PROGRAM: 进程或连接双方的IP地址和端口, 前面是服务器的IP和端口, 后面是客户端的IP和端口。
  - DEV: 流量要去往的网络端口。
  - SENT: 进程每秒发送的数据量。
  - RECEIVED: 进程每秒接收的数据量。
- c. 终止恶意程序或者屏蔽恶意访问IP。

如果确认大量占用网络带宽的进程是恶意进程, 可以使用kill *PID*命令终止恶意进程。

如果是某个IP恶意访问, 可以使用iptables服务来对指定IP地址进行处理, 如屏蔽IP地址或限速。

## 14.13 规格变更

### 14.13.1 Windows 弹性云服务器变更规格后数据盘脱机怎么办?

#### 操作场景

Windows弹性云服务器变更规格后, 可能会发生磁盘脱机, 因此, 变更规格后, 需检查磁盘挂载状态是否正常。本节操作介绍变更规格后检查磁盘挂载状态的操作步骤。

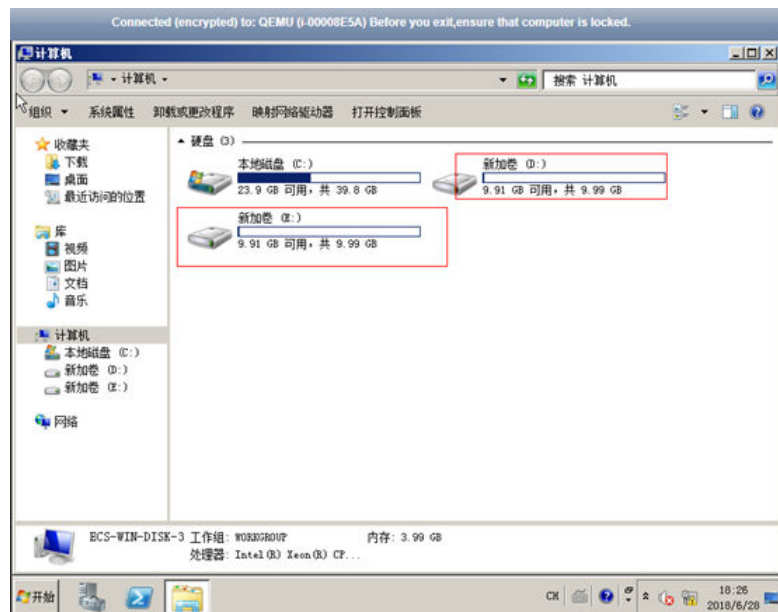
## 操作步骤

1. 对比规格变更前、后，在“计算机”中查询的磁盘个数是否一致。
  - 是，表示磁盘挂载状态正常，结束。
  - 否，表示存在磁盘脱机，执行2。

示例：

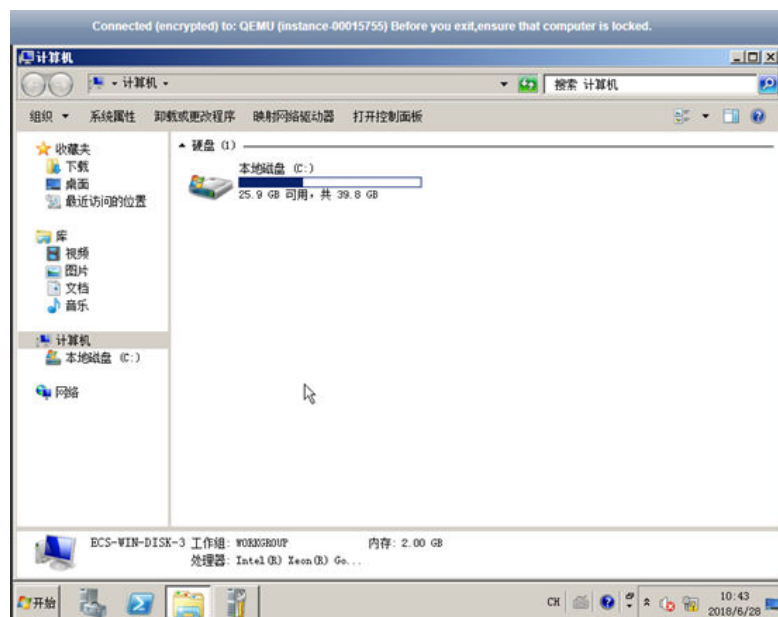
以使用Windows 2008镜像的云服务器为例，假设变更规格前，云服务器挂载有1块系统盘、2块数据盘，如图14-149所示。

图 14-149 变更规格前磁盘挂载情况



变更规格后，查看磁盘挂载情况，如图14-150所示。

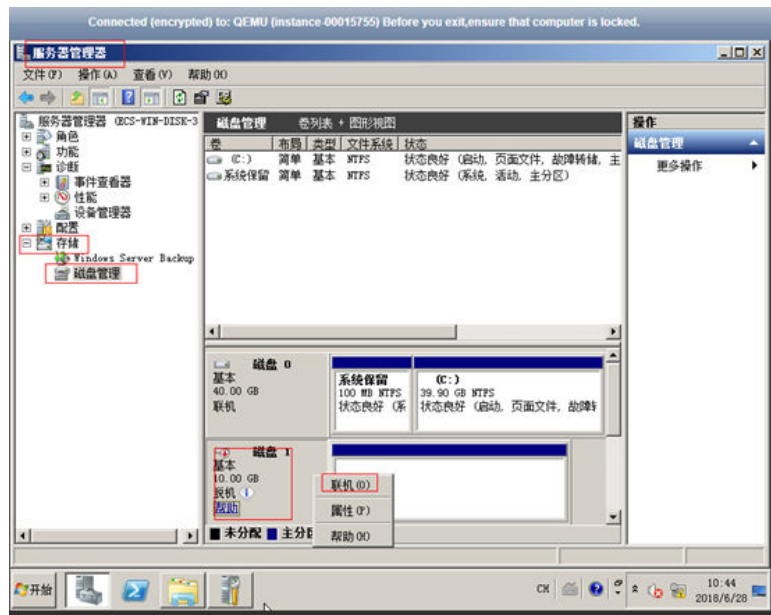
图 14-150 变更规格后磁盘挂载情况



只查询到1块系统盘，因此，变更规格前、后，显示的磁盘个数不一致，表示数据盘处于“脱机”状态。

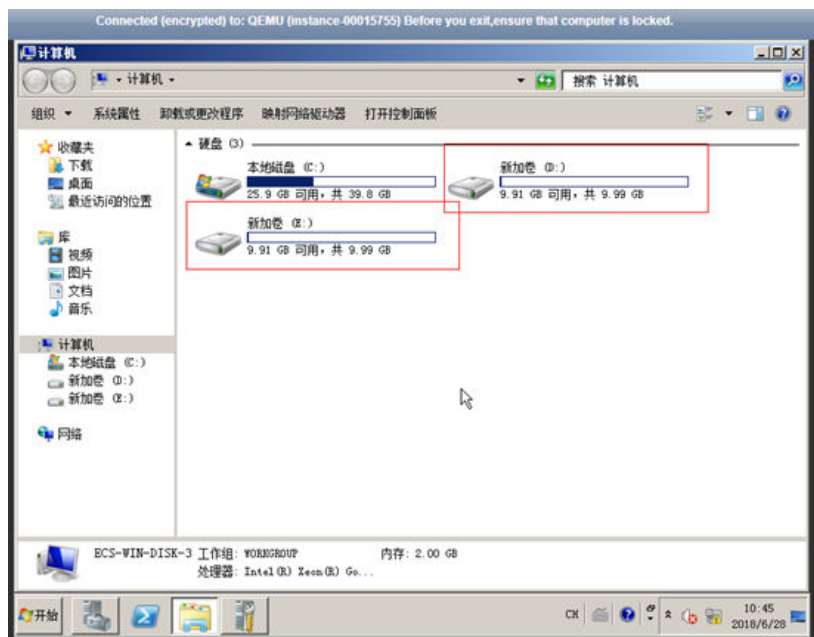
2. 将磁盘设置为“联机”状态。
  - a. 选择“开始”菜单，右键单击“计算机”，并选择“管理”。系统弹窗显示“服务器管理器”页面。
  - b. 在左侧导航栏，选择“存储 > 磁盘管理”。进入“磁盘管理页面”。
  - c. 在左侧窗格中出现磁盘列表，分别右键单击脱机的磁盘，并选择“联机”，将脱机磁盘进行联机。

图 14-151 联机



3. 再次在“计算机”中查询磁盘的个数，检查与变更规格前是否一致。
  - 是，结束。
  - 否，请联系管理员。

图 14-152 联机后磁盘挂载情况



## 14.13.2 Linux 弹性云服务器变更规格后磁盘脱机怎么办?

### 操作场景

Linux弹性云服务器变更规格时，可能会发生磁盘挂载失败的情况，因此，变更规格后，需检查磁盘挂载状态是否正常。本节操作介绍变更规格后检查磁盘挂载状态的操作步骤。

### 操作步骤

1. 以root用户登录云服务器。
2. 执行以下命令，查询磁盘信息。

```
fdisk -l | grep 'Disk /dev/'
```

图 14-153 查询磁盘信息

```
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]# fdisk -l | grep 'Disk /dev/'
Disk /dev/vda: 42.9 GB, 42949672960 bytes, 83886080 sectors
Disk /dev/vdb: 10.7 GB, 10737418240 bytes, 20971520 sectors
Disk /dev/vdc: 10.7 GB, 10737418240 bytes, 20971520 sectors
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]#
```

如图14-153所示，云服务器共有3块磁盘：/dev/vda、/dev/vdb、/dev/vdc。

3. 执行以下命令，查看磁盘挂载情况。

```
df -h | grep '/dev/'
```

图 14-154 查询磁盘挂载情况

```
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]# df -h | grep '/dev/'
/dev/vda2    39G  1.4G  35G   4% /
/dev/vda1   976M 146M  764M  16% /boot
```

如图14-154所示，云服务器只挂载了1块磁盘/dev/vda。



- 对比2和3中查询的磁盘个数是否一致。
  - 是，表示变更成功，结束。
  - 否，表示存在磁盘挂载失败，执行5。
- 使用mount命令挂载磁盘。

示例：

```
mount /dev/vdb1 /mnt/vdb1
```

其中，/dev/vdb1是待挂载的磁盘，/mnt/vdb1是待挂载磁盘的路径。

#### 须知

待挂载磁盘的文件夹/mnt/vdb1必须是空文件夹，否则会挂载失败。

- 再次执行以下命令，对比查询的磁盘个数是否一致。

```
fdisk -l | grep 'Disk /dev/'
```

```
df -h | grep '/dev/'
```

- 是，结束。
- 否，请联系管理员。

图 14-155 检查磁盘个数是否一致

```
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]# mount /dev/vdb1 /mnt/vdb1
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]# mount /dev/vdc1 /mnt/vdc1
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]#
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]#
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]#
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]#
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]#
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]# fdisk -l |grep 'Disk /dev/'
Disk /dev/vda: 42.9 GB, 42949672960 bytes, 83886080 sectors
Disk /dev/vdb: 10.7 GB, 10737418240 bytes, 20971520 sectors
Disk /dev/vdc: 10.7 GB, 10737418240 bytes, 20971520 sectors
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]# df -h | grep '/dev/'
/dev/vda2      39G  1.4G   35G   4% /
/dev/vda1     976M  146M  764M  16% /boot
/dev/vdb1      9.8G   23M   9.2G   1% /mnt/vdb1
/dev/vdc1      9.8G   23M   9.2G   1% /mnt/vdc1
[root@servercf924ffa-da23-4d09-a7e0-416694a68492 ~]#
```

如图14-155所示，使用两个命令查询的云服务器磁盘个数一致，分别是：/dev/vda、/dev/vdb、/dev/vdc。

## 14.14 操作系统变更

### 14.14.1 我能否自己安装或者升级操作系统？

弹性云服务器需要使用云平台提供的操作系统，登录到云服务器可对操作系统进行系统更新或补丁升级。

- 购买弹性云服务器时，通过选择的镜像来安装云服务器的操作系统。您可以选择公共镜像，或公共镜像制作成的私有镜像来安装云服务器的操作系统。根据实际的应用场景编程语言选择不同操作系统的镜像。
- 云服务器在使用过程中如果发现操作系统不合适，可以通过提供的切换操作系统功能来更换操作系统。例如升级操作系统，从CentOS 7.2升级到Cent OS 7.3就可以通过切换操作系统的功能来实现。

## 14.14.2 弹性云服务器能否更换操作系统？

弹性云服务器支持更换操作系统。

如果弹性云服务器当前使用的操作系统不能满足业务需求（如软件要求的操作系统版本较高），您可以选择切换弹性云服务器的操作系统。

云平台支持不同镜像类型（包括公共镜像、私有镜像、共享镜像）与不同操作系统之间的互相切换。您可以将现有的操作系统切换为不同镜像类型的操作系统。

更换约束限制条件以及方法请参见[切换操作系统](#)。

## 14.14.3 切换操作系统需要多长时间？

关机云服务器，并单击“操作”列下的“更多 > 镜像/磁盘 > 切换操作系统”执行切换操作，预计需等待1~2分钟完成切换操作系统。

切换操作系统过程中，云服务器会显示任务状态为“切换操作系统”。

## 14.14.4 重装操作系统时是否可以选择其他操作系统？

重装操作系统只支持使用原镜像进行系统重装，不支持使用新的系统镜像。如需使用新的系统镜像，请参见[切换操作系统](#)。

## 14.14.5 重装操作系统需要多长时间？

将云服务器关机，单击“更多 > 镜像/磁盘/备份 > 重装系统”，预计需等待1~2分钟完成重装操作系统。

重装操作系统过程中，云服务器会显示任务状态为“重装操作系统”。