高性能应用服务 HAI

操作指南

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】

腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

操作指南
- 算力操作
  - 连接 Windows 算力
  - 连接 Linux 算力
  - 关机算力
  - 销毁算力
  - 监控算力
  - 跨地域开机
- 数据操作
  - 通过 COS 上传和下载数据
  - 通过 Filezilla 上传和下载数据
  - 通过 JupyterLab 上传和下载数据
- 端口操作
  - 编辑端口规则
  - 默认端口说明
- 学术加速
  - 开启学术加速
  - 关闭学术加速
- 应用操作
  - 管理自定义应用
- 子账号授权
- HAI 社区

版权所有：腾讯云计算（北京）有限责任公司

操作指南
算力操作
连接 Windows 算力

最近更新时间：2025-06-26 14:57:42

操作场景

本文介绍如何在 Windows 及 MacOS 系统的本地计算机中通过远程桌面登录 Windows。

前提条件

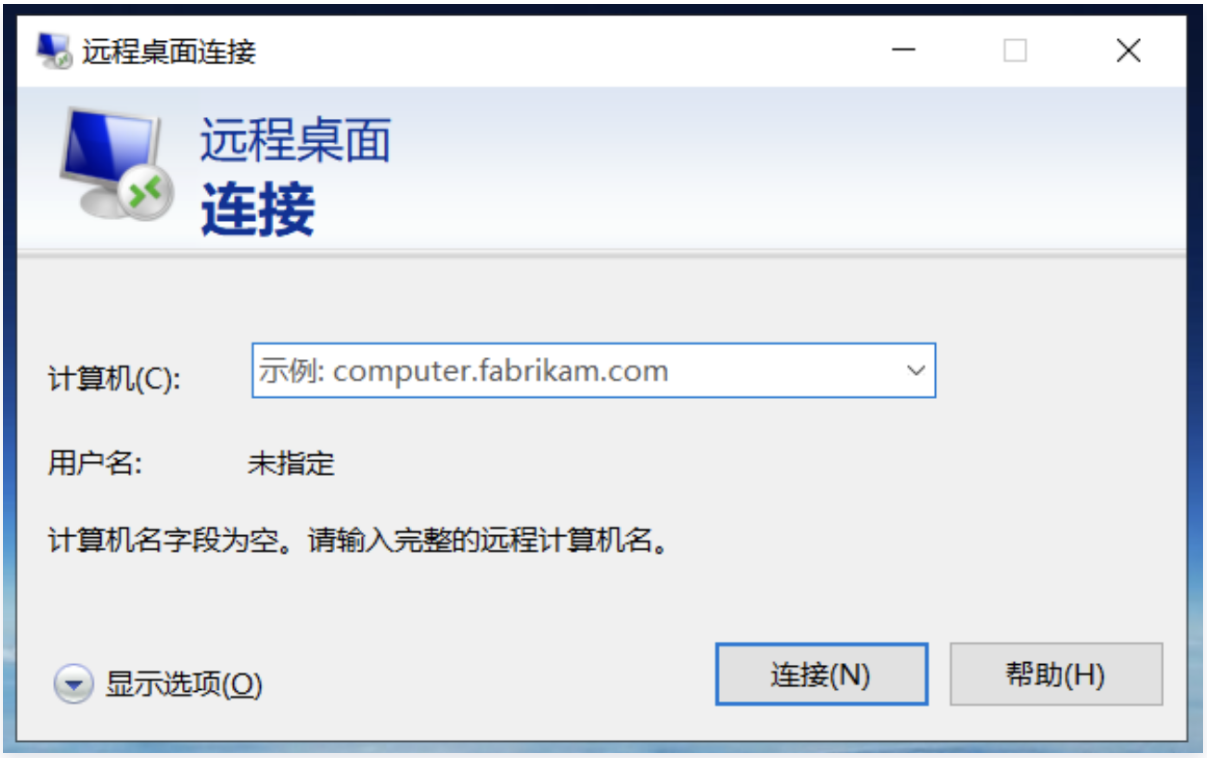
您已获取远程登录 Windows 实例需要使用实例的管理员账号（Administrator）和对应的登录密码。目前可在高性能应用服务 HAI 创建成功的站内信或邮件中获取。

操作步骤

请对应本地操作系统，选择以下方式使用远程桌面登录 Windows 实例：

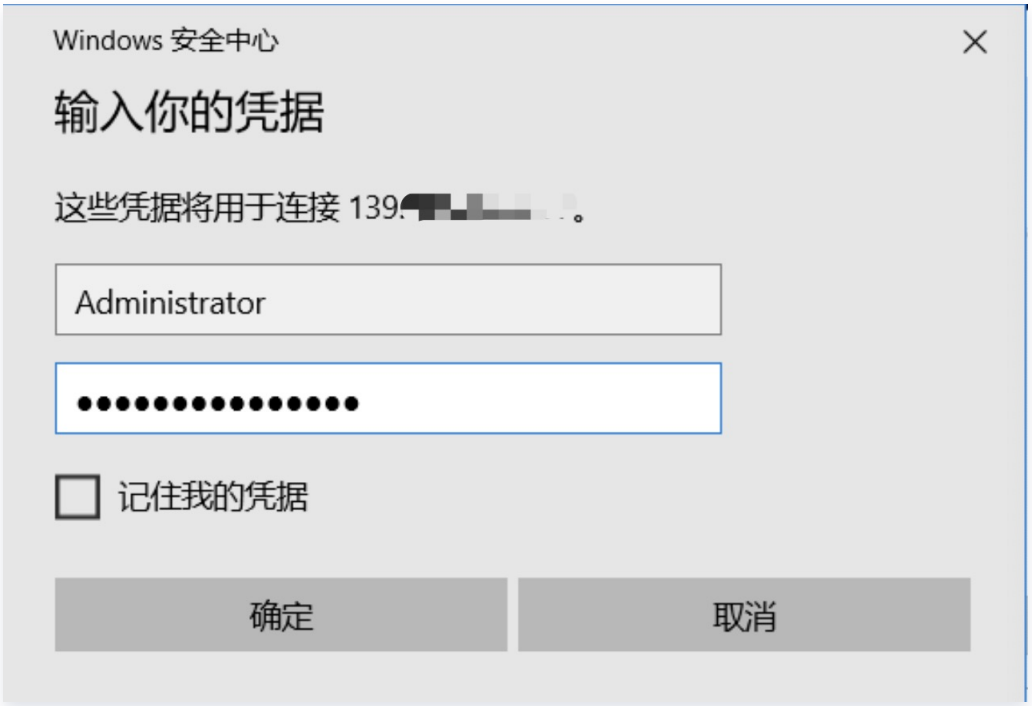
Windows

1. 在本地 Windows 计算机上，单击 ⊞，在**搜索框**中，输入**mstsc**，按 **Enter**，打开远程桌面连接对话框。如下图所示：

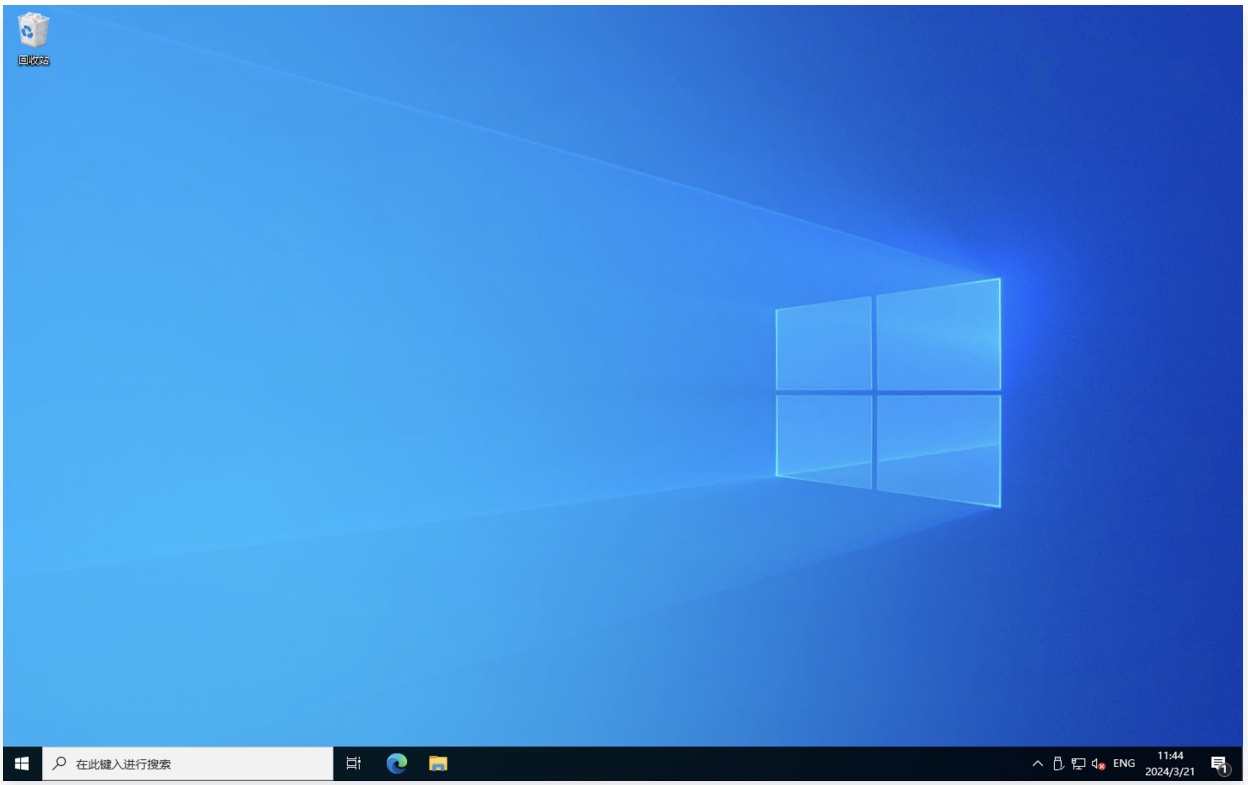


2. 在**计算机**后面，输入 Windows 实例的公网 IP，单击**连接**。
   Windows 实例公网 IP 可前往 高性能应用服务 HAI 控制台 或 站内信 中获取。
3. 在弹出的 **Windows 安全**窗口中，输入实例的管理员账号和密码，如下图所示：

> **说明：**
> 若弹出**无法验证此远程计算机的身份，是否仍要连接？**对话框，可勾选**不再询问我是否连接到此计算机**，单击**连接**。



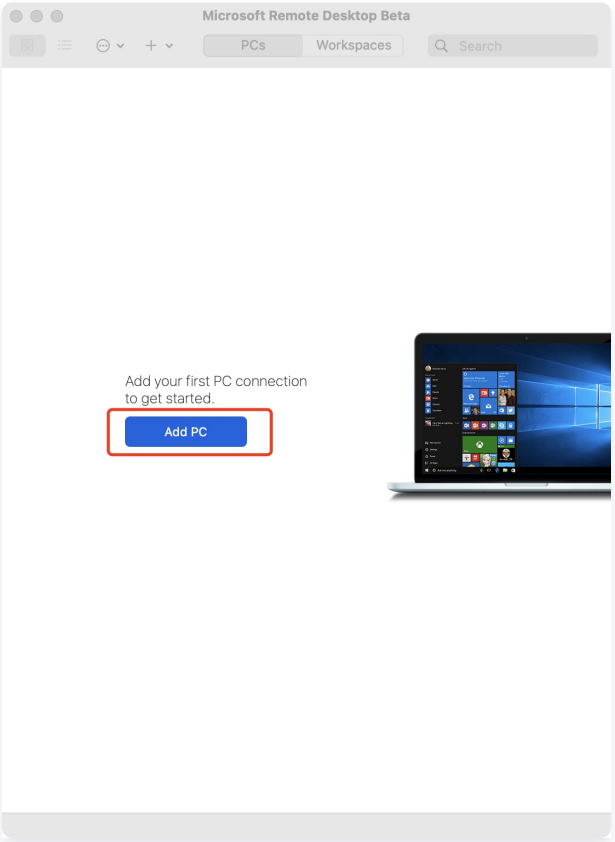
4. 单击确定，即可登录到 Windows 实例，如下图所示，环境中已预装英伟达驱动，您可根据自身使用需求进行环境配置：



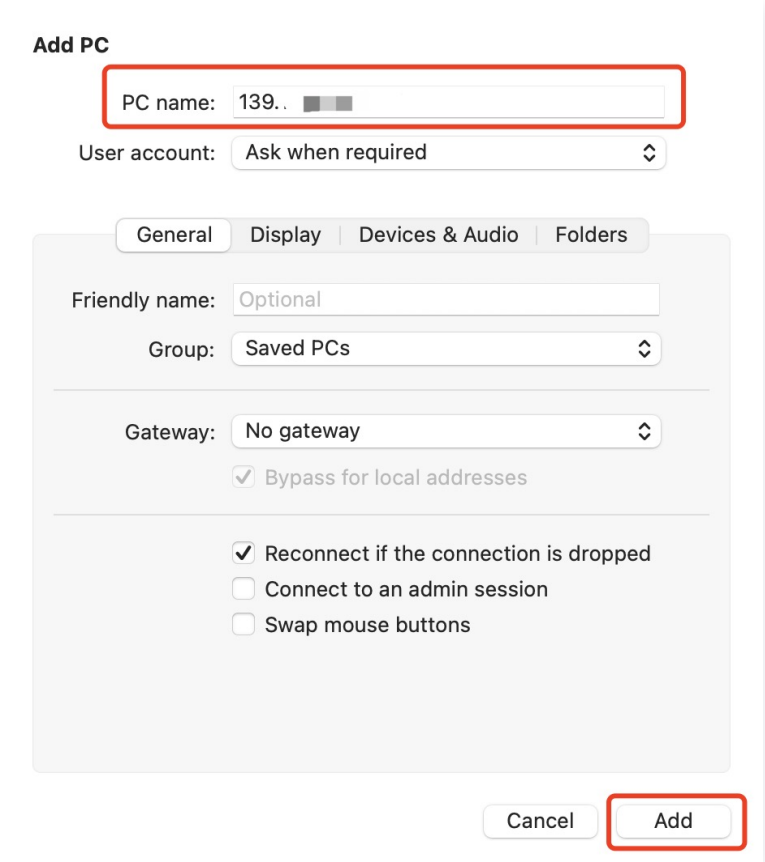
MacOS

1. 下载 Microsoft Remote Desktop 并在本地进行安装。您可前往 Microsoft Remote Desktop Beta 下载 Beta 版本。

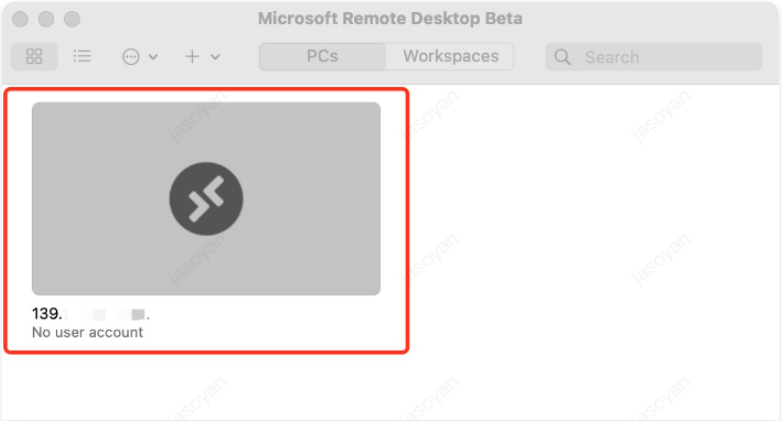
2. 启动 Microsoft Remote Desktop Beta，并单击 **Add PC**。如下图所示：



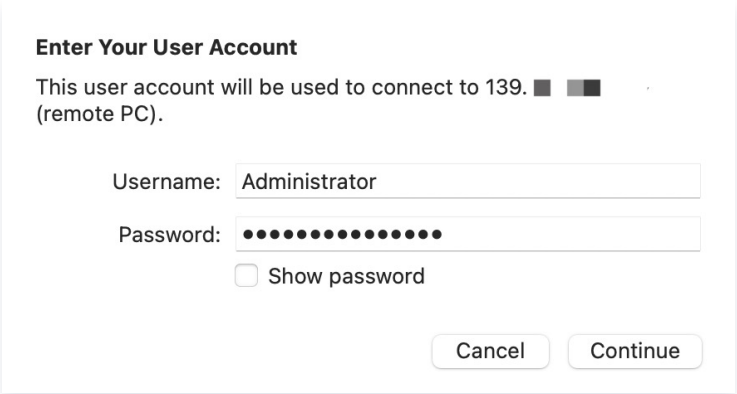
3. 弹出的 **Add PC** 窗口，按以下步骤创建连接：
   3.1 在 **PC name** 处输入实例公网 IP。
   3.2 其余选项保持默认设置，完成创建连接。
   3.3 单击 **Add** 确认创建 。



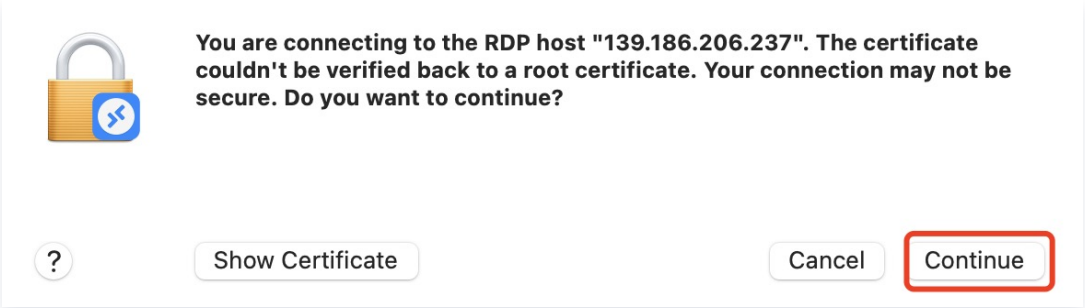
即可在窗口中查看已成功创建的连接。如下图所示：



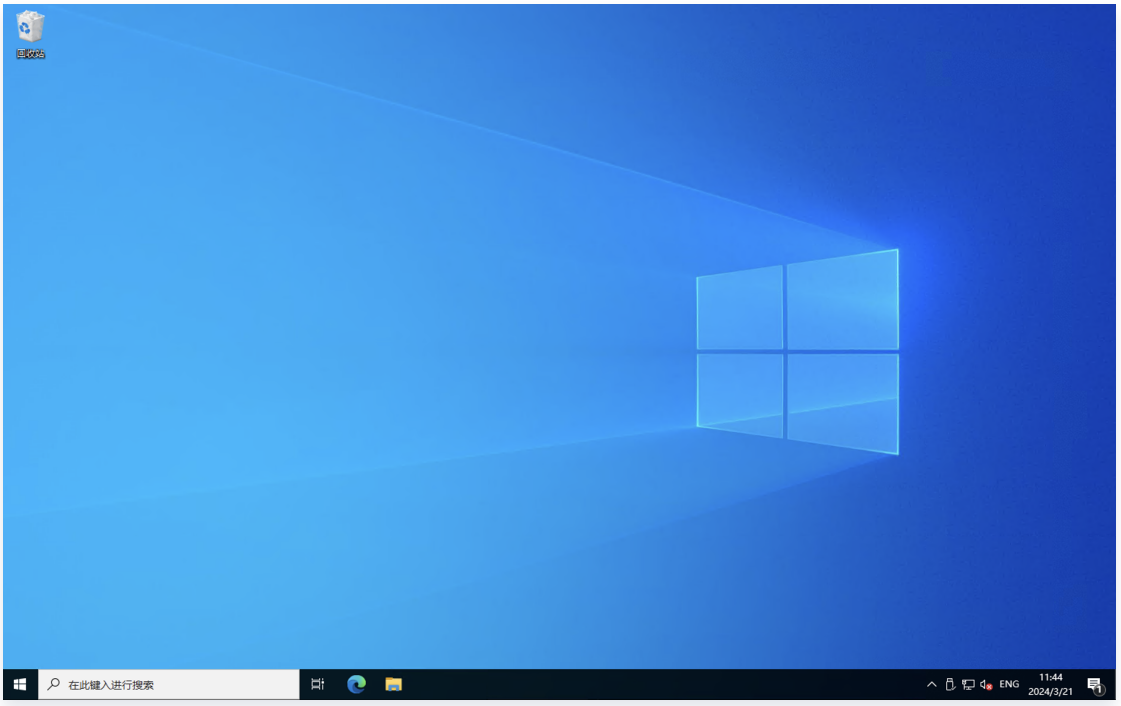
4. 双击打开新创建的PC，并在弹出的窗口中根据提示，输入在站内信中获取的管理员用户名和密码，单击**Continue**。



5. 在弹出的窗口中单击 **Continue** 确认连接。如下图所示：



6. 成功连接后将打开 Windows 实例界面。如下图所示，环境中已预装英伟达驱动，您可根据自身使用需求进行环境配置：



管理员: 命令提示符
Microsoft Windows [版本 10.0.20348.2322]
(c) Microsoft Corporation。保留所有权利。
C:\Users\Administrator>nvidia-smi
Thu Mar 21 11:44:44 2024
NVIDIA-SMI 538.15 Driver Version: 538.15 CUDA Version: 12.2
GPU Name TCC/WDDM Bus-Id Disp.A Volatile Uncorr. ECC
Fan Temp Perf Pwr:Usage/Cap Memory-Usage GPU-Util Compute M.
MIG M.
0 Tesla T4 TCC 00000000:00:06.0 Off 0
N/A 38C P8 9W / 70W 1MiB / 15360MiB 0% Default
N/A
Processes:
GPU GI CI PID Type Process name GPU Memory
ID ID Usage
No running processes found
C:\Users\Administrator>

连接 Linux 算力

最近更新时间：2025-02-21 14:26:11

操作场景

在完成高性能应用服务 HAI 的算力创建后，您可以在控制台使用算力和应用。本文介绍如何连接 Linux 系统的算力。

操作步骤

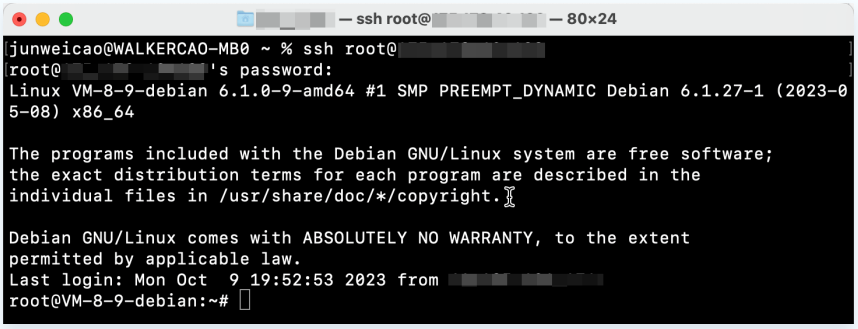
目前高性能应用服务 HAI 提供多种连接算力的方式，包括 SSH、Terminal、Notebook、WebUI。根据创建的应用不同，提供的算力连接入口可能存在差异，具体请您以创建成功后控制台展示为准。

方式一：通过 SSH 连接算力

1. 执行以下命令，连接 Linux 实例。

说明

- 如果您的本地计算机使用非桌面版的 Linux 系统，可直接在系统界面执行以下命令。
- 如果您的本地计算机使用桌面版 Linux 系统或 MacOS 系统，请先打开系统自带的终端（如 MacOS 的 Terminal），再执行以下命令。



- 如果您的本地电脑为 Windows 操作系统（如 Windows 10 或 Windows Server 2022），需先打开命令提示符（CMD），再执行以下命令。

- **用户名及密码可通过 站内信 查看。**若忘记密码，可在控制台进行修改。

```
ssh <username>@<IP address or domain name>
```

- `username` 即为已获取的用户名，如 `root` 等。用户名及密码可在高性能应用服务 HAI 创建成功的 站内信 或邮件中获取。
- `IP address or domain name` 为您的实例公网 IP 地址。实例公网 IP 地址可前往 高性能应用服务 HAI 控制台 查看。

2. 出现如下图所示的界面，输入已获取的密码，按 **Enter**，即可完成登录。

```
[redacted]-MB0 ~ % ssh root@[redacted]
root@[redacted]'s password:
```

3. 登录成功，效果如下图所示。

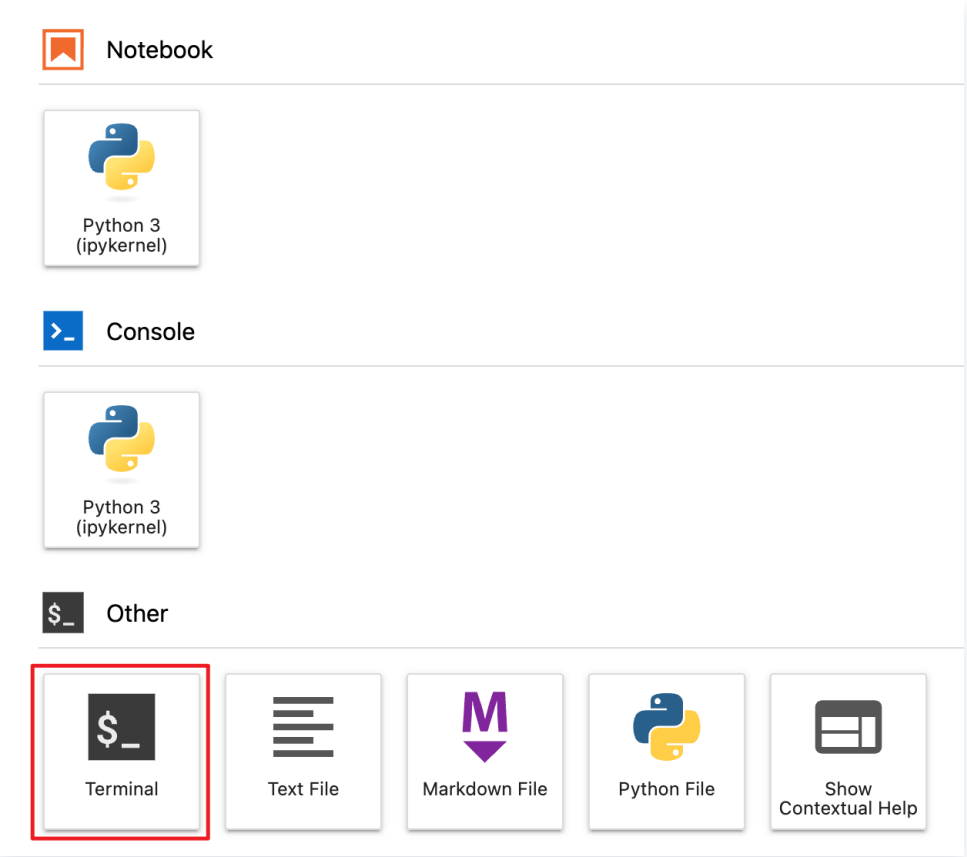
```
[redacted]-MB0 ~ % ssh root@[redacted]
root@[redacted]'s password:
Last login: Thu Mar  9 09:41:23 2023 from [redacted]
[root@VM-8-17-centos ~]#
```

> **说明：**
> 不同类型的 Linux 系统，登录时的样式可能有所不同，上图仅作为参考。

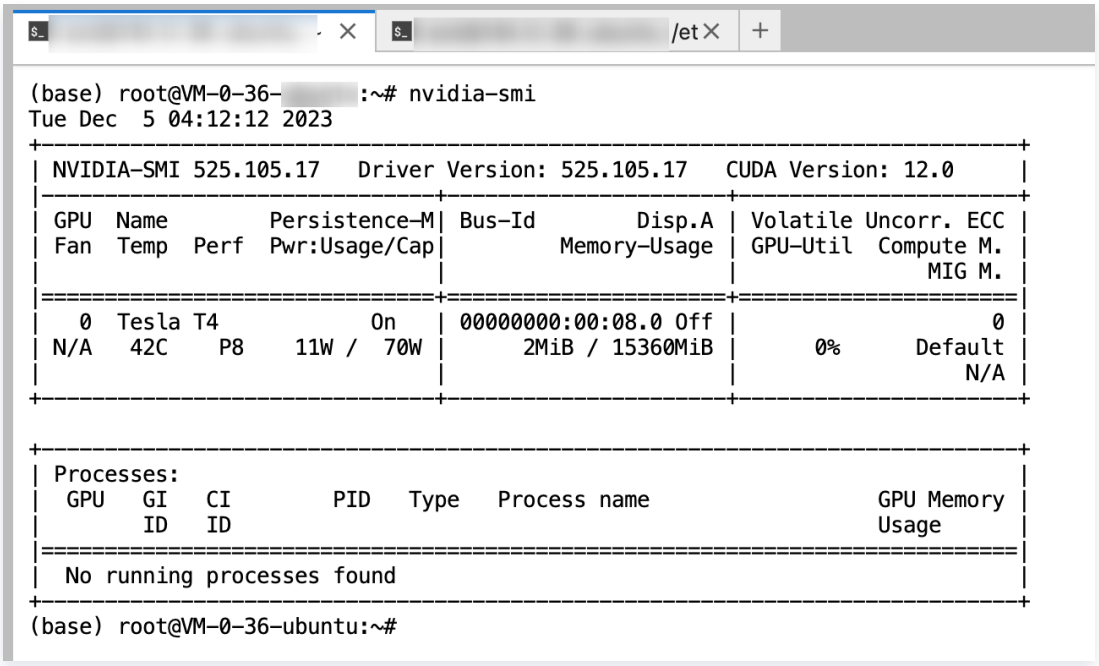
方式二：通过 Terminal 连接算力

1、登录 高性能应用服务 HAI 控制台。
2、选择需要连接的算力，然后单击**算力连接**。在下拉菜单中，单击 **JupyterLab**。

3、进入 JupyterLab 后，选择 **Other** > **Terminal**，进入终端界面。



4、在终端界面上，您可自由输入命令对实例进行操作。

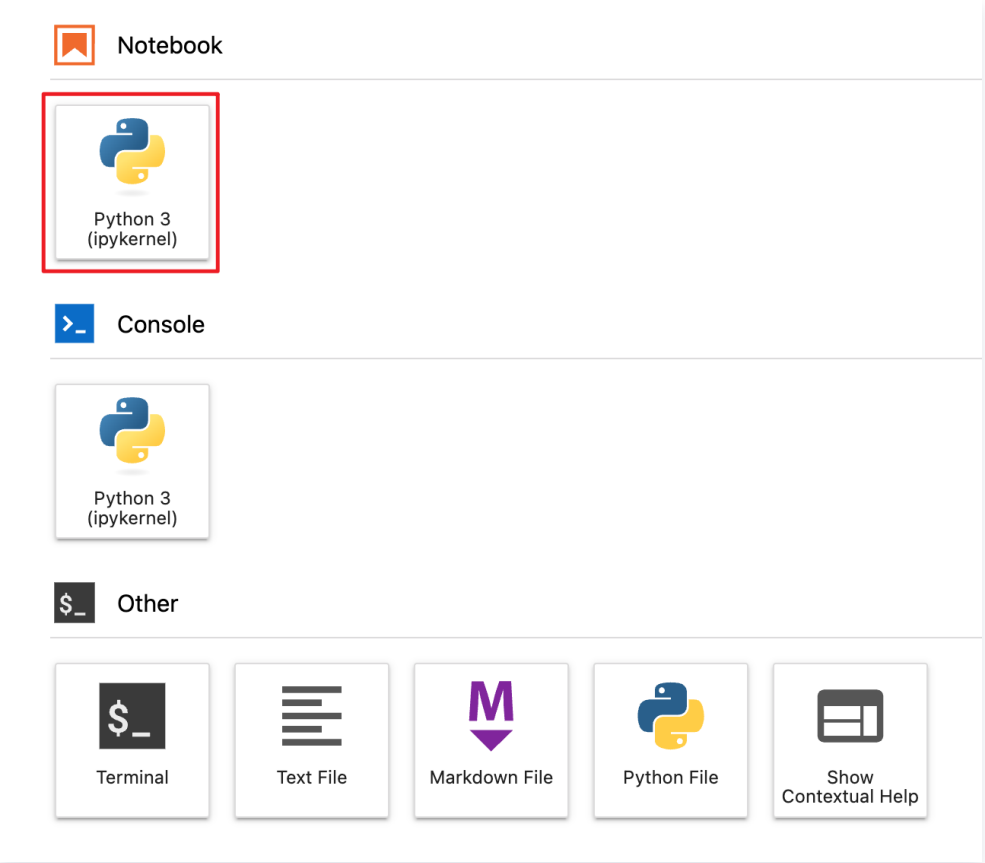


方式三：通过 Notebook 连接算力

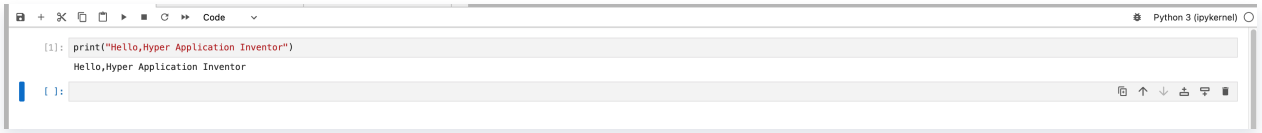
1、登录 高性能应用服务 HAI 控制台。

2、选择需要连接的算力，然后单击**算力连接**。在下拉菜单中，单击 **JupyterLab**。

3、进入 **JupyterLab** 后，选择 **Notebook** > **Python3**，进入 Notebook 界面。



4、进入后您可在 Notebook 中自由编辑、调试、运行代码。

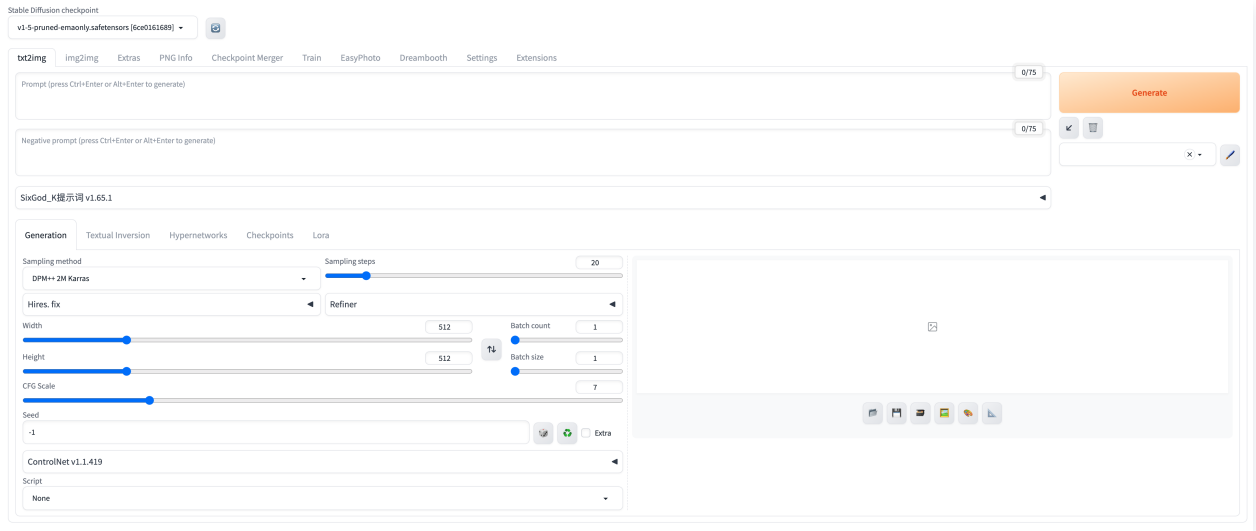


方式四：通过 WebUI 连接算力

1、登录 高性能应用服务 HAI 控制台。

2、选择需要连接的算力，然后单击**算力连接**。在下拉菜单中，单击 **Gradio WebUI**。

3、进入后您可基于 WebUI 界面提示自由操作。因模型及生态差异，不同模型搭配的 WebUI 存在细微区别，以页面实际展示为准。



关机算力

最近更新时间：2025-06-26 14:57:42

操作场景

关机不计费是指 高性能应用服务 HAI 实例进入**已关机**状态后，不再收取算力（CPU、GPU、内存）费用，云盘将继续计费。目前暂无关机时长限制。

开启影响

在实例关机后，**会有以下影响**：

- 由于实例的 CPU、GPU 和内存将进行释放，再次启动实例时可能会因**实例资源不足导致启动失败**。此时，您可稍后再尝试启动实例。
- 实例已分配的**公网 IP 地址在关机后会自动释放**，可能会导致再次启动实例失败。成功启动实例后，实例会**分配一个新的公网 IP 地址**，而内网 IP 则不受影响，保留不变。
- 已关机状态无法进行算力连接，如需操作请开机后进行。
- 若实例在已关机状态出现账号欠费，会对实例进行欠费隔离。若未及时进行冲正，将对已关机状态的实例进行销毁。

计费规则

在实例关机后，算力（CPU、GPU、内存）将停止计费，云盘将继续计费，具体而言：

- 免费额度：创建实例时默认提供 80GB 免费云盘额度，免费额度在关机后前15天不计费。关机时长超过15天后，对该 80GB 免费云盘容量以 0.02元/小时进行资源闲置扣费。
- 付费额度：针对扩容系统盘，关机后继续正常收费。

例：

- 若创建实例时，选择 80GB 云盘，则关机后前15天免费，15天后每小时扣费0.02元。
- 若创建实例时，手动扩容云盘大小，使用500GB云盘，则关机后前15天每小时扣费0.12元，15天后每小时扣费0.14元。

操作指南

关机流程

1. 登录 高性能应用服务 HAI 控制台。
2. 选择需要关机的实例，然后单击**更多**。在下拉菜单中，单击**关机**。

您也可以点击**实例名称**进入详情页，然后单击**关机**。

3. 在关机弹窗中，确认需要关闭的应用实例信息，包括**实例 ID**，**云盘容量**，**预估存储费用**等。如果确认无误，单击**确认关机**，实例将进入**关机中**状态。等待实例状态转变为**已关机**后，关机流程即告结束。

开机流程

1. 登录 高性能应用服务 HAI 控制台。
2. 选择需要开机的实例，单击**更多**，在下拉菜单中单击**开机**。

您也可以点击实例名称进入详情页，然后单击**开机**。

3. 在弹窗中确认需要启动的应用实例信息，包括**实例 ID**，**云盘容量**，**总费用**等。如果确认信息无误，单击**开机**，实例将进入**开机中**状态。待实例状态变为**运行中**，开机流程即告结束。

销毁算力

最近更新时间：2025-06-26 14:57:42

按量计费

操作场景

当您不再需要 HAI 算力实例时，可以对 HAI 算力实例进行销毁。算力实例一旦销毁成功，系统盘中的数据会被擦除且无法恢复，并立即停止续费。

前提条件

您已对必要数据进行备份，实例一旦销毁数据将无法恢复，

操作步骤

1. 登录 高性能应用服务 HAI 控制台。
2. 选择需要连接的算力，然后单击**更多**。在下拉菜单中，单击**销毁**。

3. 在弹出的窗口中，需要您再次确认销毁的实例信息，单击**销毁资源**。

包年包月

操作场景

当您不再需要 HAI 算力实例时，可以对 HAI 算力实例进行销毁。算力实例一旦销毁成功，系统盘中的数据会被擦除且无法恢复，并立即停止续费。

前提条件

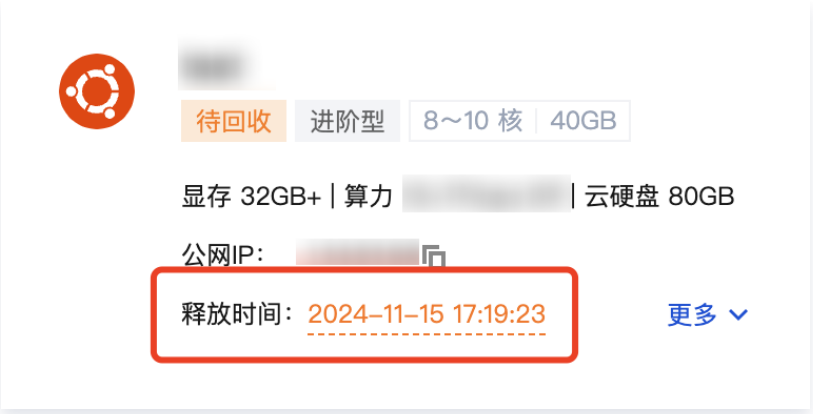
您已对必要数据进行备份，实例一旦销毁数据将无法恢复，

操作步骤

1. 登录 高性能应用服务 HAI 控制台。
2. 选择需要连接的算力，然后单击**更多**。在下拉菜单中，单击**销毁**。

3. 在弹出的窗口中，需要您再次确认销毁的实例信息及退费金额，退费规则详见 包年包月退费说明，确认无误后单击**提交退还订单**。

4. 销毁后，实例被移入回收站保留七天，实例上运行的服务彻底中断。如您想恢复已经自助退还的包年包月高性能应用服务 HAI，可以在控制台进行续费恢复。


待回收
进阶型
8~10 核
40GB
显存 32GB+ | 算力
| 云硬盘 80GB
公网IP：
释放时间：2024-11-15 17:19:23
更多

监控算力

最近更新时间：2024-08-12 16:51:41

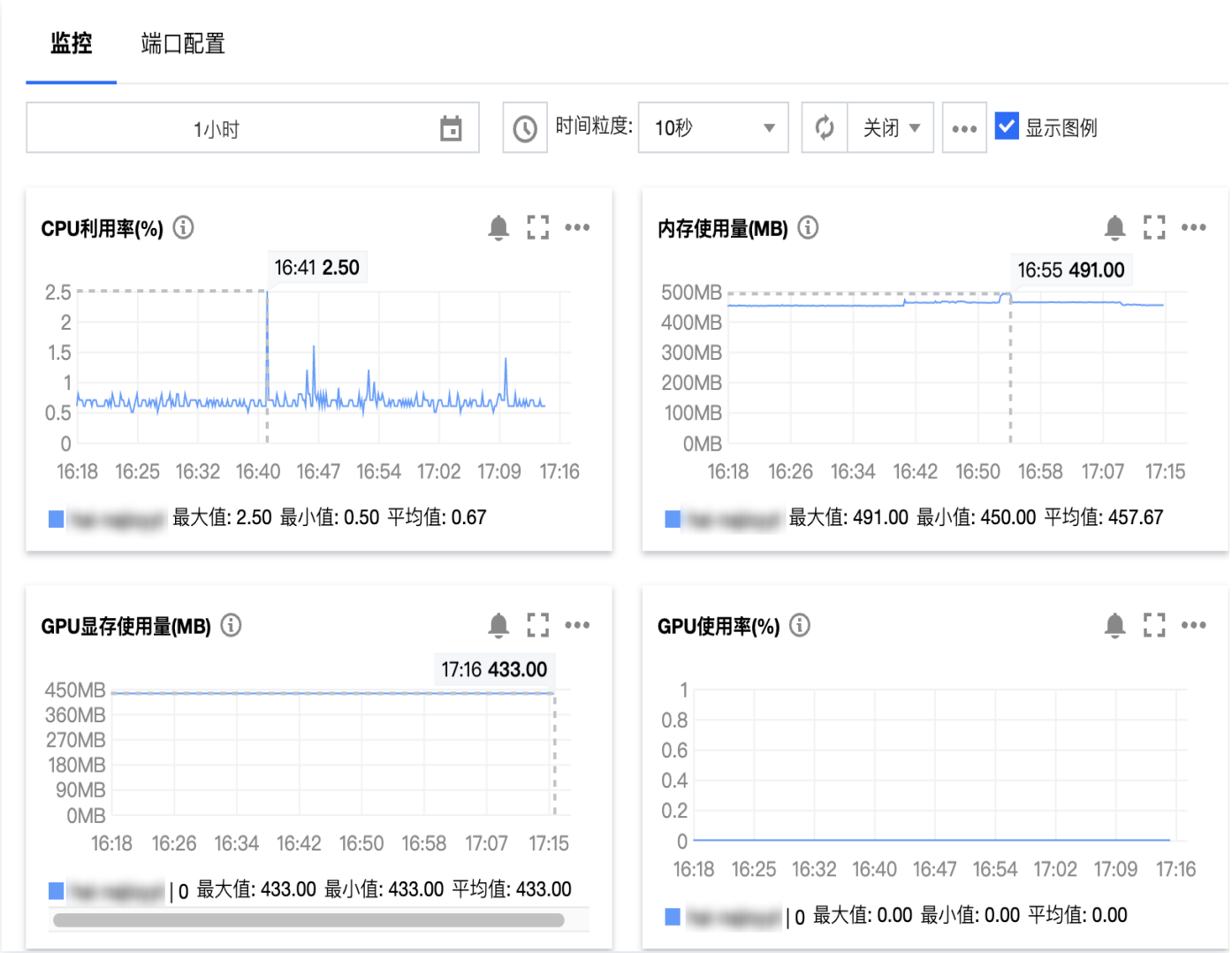
操作场景

高性能应用服务 HAI 默认为所有用户提供了使用数据的监控能力，您可以监控到 CPU、GPU、存储、网络 等资源的使用情况。

操作步骤

1. 登录 高性能应用服务 HAI 控制台。
2. 选择需要操作的实例，然后单击**实例名称**。

3. 在实例详情中选择**监控**页卡查看的监控数据。


监控
端口配置
1小时
时间粒度:
10秒
关闭
显示图例
CPU利用率(%)
16:41 2.50
2.5
2
1.5
1
0.5
0
16:18 16:25 16:32 16:40 16:47 16:54 17:02 17:09 17:16
最大值: 2.50 最小值: 0.50 平均值: 0.67
内存使用量(MB)
16:55 491.00
500MB
400MB
300MB
200MB
100MB
0MB
16:18 16:26 16:34 16:42 16:50 16:58 17:07 17:15
最大值: 491.00 最小值: 450.00 平均值: 457.67
GPU显存使用量(MB)
17:16 433.00
450MB
360MB
270MB
180MB
90MB
0MB
16:18 16:26 16:34 16:42 16:50 16:58 17:07 17:15
| 0 最大值: 433.00 最小值: 433.00 平均值: 433.00
GPU使用率(%)
1
0.8
0.6
0.4
0.2
0
16:18 16:25 16:32 16:40 16:47 16:54 17:02 17:09 17:16
| 0 最大值: 0.00 最小值: 0.00 平均值: 0.00

跨地域开机

最近更新时间：2025-01-17 14:38:43

操作场景

当前地域的计算资源不足时、导致无法开机时，您可以通过本文档所述的操作，将已经关机的实例复制到其他地域，进行跨地域开机。

地域和可用区

在腾讯云中，**地域**和**可用区**是构建和部署云资源时的两个基本概念。

- 地域：指的是腾讯云服务所在的物理位置，如**北京地域**、**上海地域**等。每个地域都是一个独立的地理区域，拥有自己的数据中心。
- 可用区：可用区是指在同一地域内，为了提高容错能力而设置的物理隔离区域，如**北京一区**。每个可用区都有独立的电源、网络和冷却设施。
- 地域和可用区的关系：一个地域包含多个可用区。

> **设计方式：**
> 在高性能应用服务 HAI 中，**为了简化客户的部署流程，产品设计中隐藏了“可用区”的概念**。因此，用户无法自行选择可用区，只能选择地域。由于**不同可用区的资源是独立的**，所以当用户已创建的实例所在的某个可用区资源售罄，而同一地域内的其他可用区还有资源时，用户将无法跨可用区启动服务。在这种情况下，您需要按照本文档中的**跨地域启动**指南来重新购买服务。

操作步骤

今天我们将演示将一台广州地域的 HAI 实例，跨地域复制到上海地域进行开机。共分为以下四个步骤：

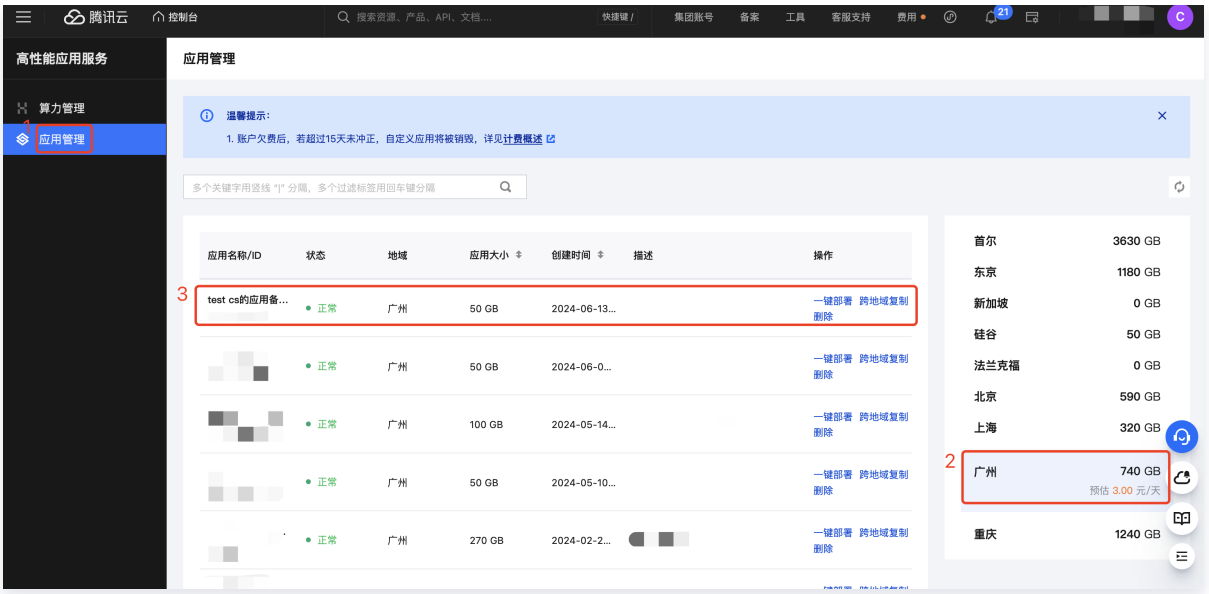
步骤一：应用存档

1.1 首先进入高性能应用服务 HAI 的 控制台 ，找到我们想要开机的实例，选择**应用存档**。

1.2 在弹出的窗口里给存档的应用进行命名，并单击**存档应用**。

1.3 实例状态变更为**应用存档中**，等待 2～5分钟。

1.4 存档完成后，单击控制台左侧的**应用管理**，在右侧选择目标地域，本例中为**广州**地域，之后在页面中间的应用列表就可以看到刚刚保存的应用了。

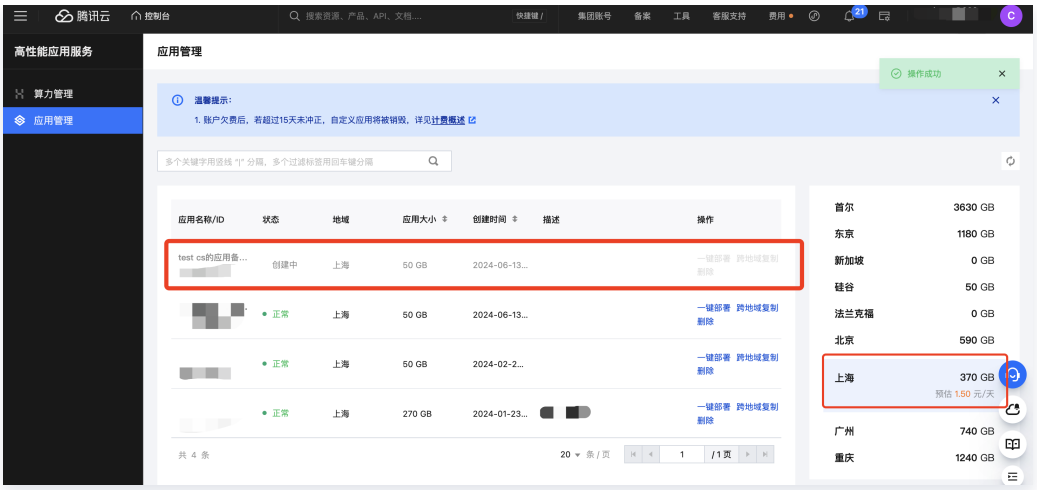


步骤二：跨地域复制

2.1 找到刚刚保存的应用，单击**跨地域复制**。

2.2 在弹出的窗口选择目标地域，本例中为**上海**地域，之后单击**复制应用**。

2.3 复制应用后，在应用管理页面进入**上海**地域，可以在页面中央看到正在复制的应用，应用状态为**创建中**。

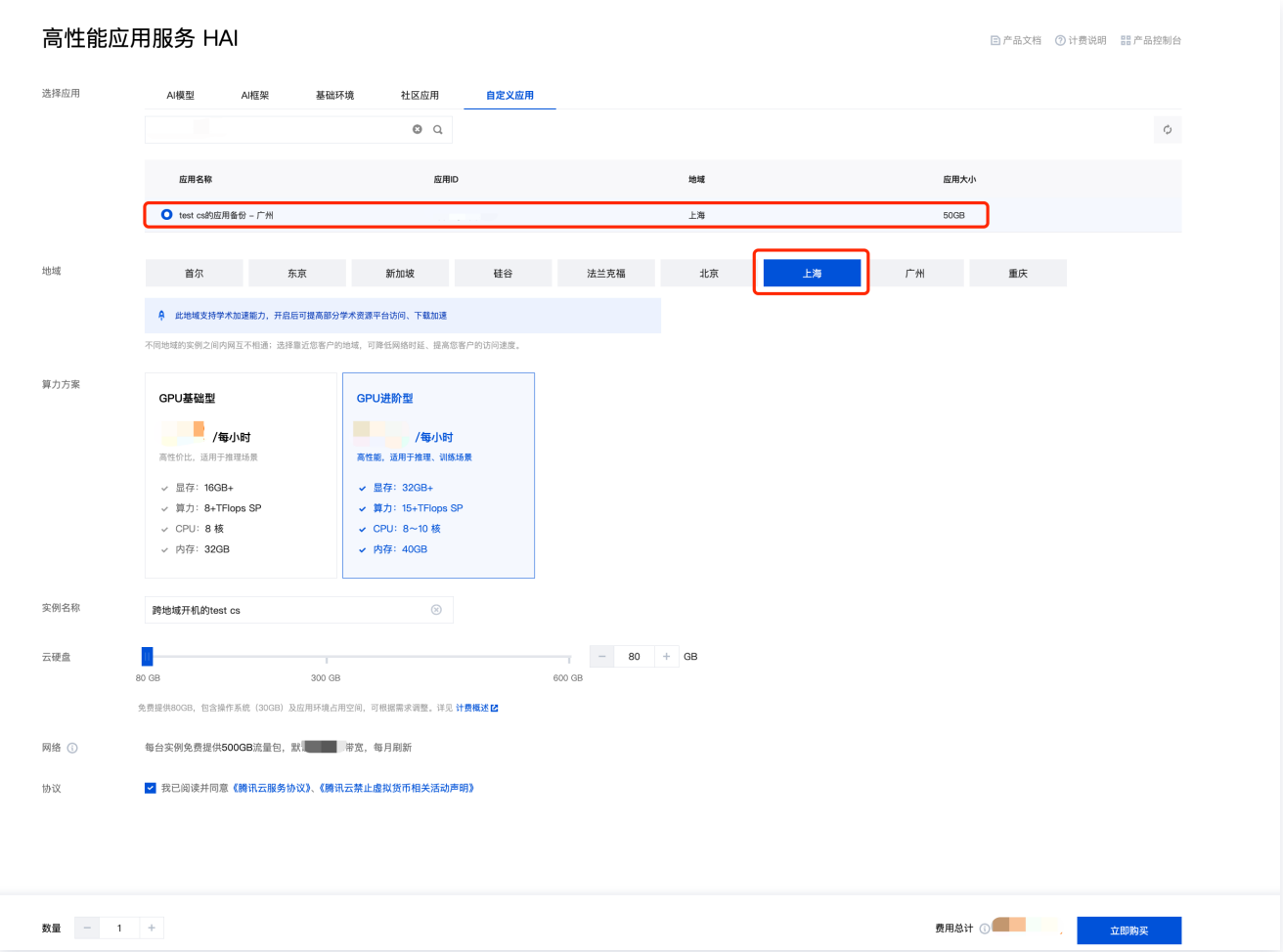


2.4 等待15～30分钟，应用状态变为**正常**，应用跨地域复制完成。

步骤三：跨地域开机

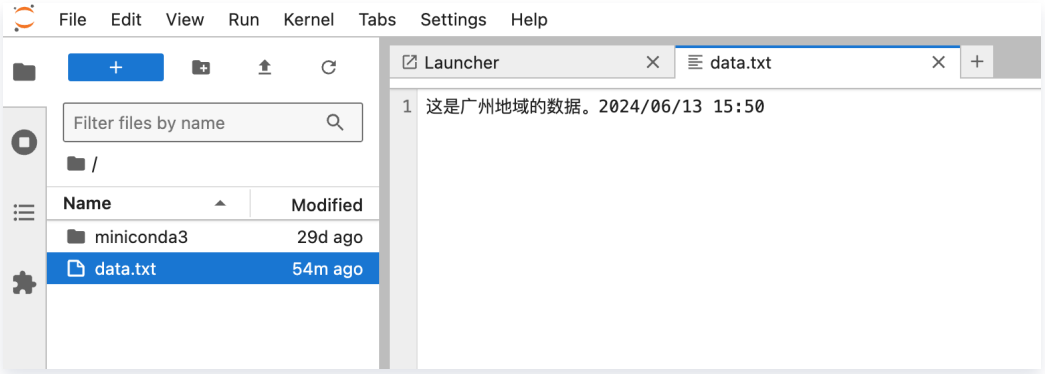
3.1 找到刚刚复制成功的应用，单击**一键部署**。

3.2 弹出 HAI 的购买页，会发现选择应用一栏中，默认选项为刚刚保存的自定义应用，地域一栏的默认选项在目标地域，本例中即上海地域。之后依次配置算力方案、实例名称、 云硬盘、协议，单击**立即购买**，在新的地域创建实例。



3.3 创建成功，在 HAI 的控制台找到新地域的实例。

3.4 进入新实例的 JupyterLab，检查发现原实例的数据得到了保留。



步骤四：删除存档

已经完成跨地域开机后，如果后续不会再用到应用存档，可以删除刚刚做好的两个应用存档，停止对应用存档的计费。在应用管理里点击**删除**即可。注意**上海**和**广州**两个地域的都要删除。

应用名称/ID	状态	地域	应用大小	创建时间	描述	操作
[redacted]	正常	上海	50 GB	2024-06-13 16:21:23		一键部署 跨地域复制 删除

数据操作
通过 COS 上传和下载数据

最近更新时间：2024-12-19 10:19:53

操作场景

对象存储（Cloud Object Storage，COS）是腾讯云提供的一种存储海量文件的分布式存储服务，可通过网络随时存储和查看数据。与高性能应用服务 HAI 之间支持同地域内网连接，实现高速网络上传和下载。
本文中以 COSCMD 工具为例，通过简单的指令对您的文件实现批量上传、下载、删除等操作。

前置条件

在通过 对象存储 COS 与高性能应用服务 HAI 之间上传和下载数据之前，请确保您在已经拥有 COS 存储桶，并且已开启相应的读写权限。具体可参见 创建存储桶 和 设置访问权限 。

操作步骤

安装与配置 COSCMD 工具

1、执行 `pip` 命令进行安装。

```
pip install coscmd
```

2、安装完成后，执行以下命令进行更新。

```
pip install coscmd -U
```

3、首次使用前必须生成配置文件，命令具体格式如下：

```
coscmd config [OPTION]...<FILE>...
              [-h] --help
              [-a] <SECRET_ID>
              [-s] <SECRET_KEY>
              [-t] <TOKEN>
              [-b] <BucketName-APPID>
              [-r] <REGION> | [-e] <ENDPOINT>
              [-m] <MAX_THREAD>
              [-p] <PART_SIZE>
              [--do-not-use-ssl]
              [--anonymous]
```

> **说明：**
> 其中 "[]" 中的字段为选项， "<>" 的字段为需要填写的参数。

参数配置说明如下：

选项	参数说明	有效值	是否必选
-a	密钥 ID 请前往 API 密钥控制台 获取。	字符串	是
-s	密钥 Key 请前往 API 密钥控制台 获取。	字符串	是
-t	临时密钥 token，当使用临时密钥时需要配置，设置 x-cos-security-token 头部。	字符串	否
-b	指定的存储桶名称，存储桶的命名格式为 BucketName-APPID，请参见 命名规范 。初次配置使用时，您需要在 COS 控制台创建一个存储桶，用于配置工具。	字符串	是
-r	存储桶所在地域，请参见 地域和访问域名 。	字符串	是
-e	设置请求的 ENDPOINT，设置 ENDPOINT 参数后，REGION 参数会失效。当您使用的是默认域名，则此处配置格式为 `cos.<region>.myqcloud.com` ；当您使用全球加速域名，则配置为 `cos.accelerate.myqcloud.com` 。	字符串	否
-m	多线程操作的最大线程数（默认为5，范围为1 - 30）。如果上传文件缓慢，可适当调高该值。线程数取决于机器性能，通常情况下设置为30，即可跑满带宽。例如将并发线程调整为30，执行命令为 `coscmd config -m 30` 。	数字	否
-p	分块操作的单块大小（单位MB，默认为1MB，范围为1 - 1000）。如果上传文件缓慢，可适当调高该值。	数字	否
--do-not-use-ssl	使用 HTTP 协议，而不使用 HTTPS。	字符串	否
--anonymous	匿名操作（不携带签名）。	字符串	否

一般情况下仅需填写必填值，config 命令的使用示例如下：

```
coscmd config -a AChT4ThiXAbpBDEFGhT4ThiXAbp**** -s
WE54wreefvds3462refgwewe**** -b bucket-1250000000 -r ap-chengdu
```

从 HAI 上传文件到 COS

1. 上传文件命令格式：

```
coscmd upload <localpath> <cospath>
```

2. 上传文件夹命令格式：

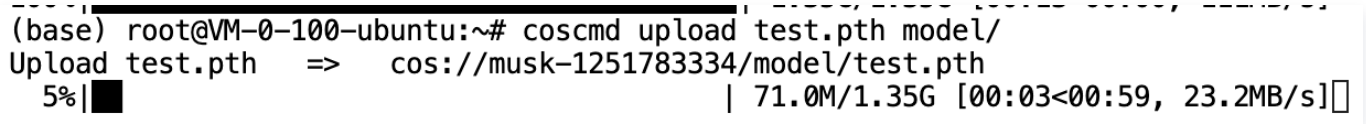
```
coscmd upload -r <localpath> <cospath>
```

> **注意：**
> 请将 "<>" 中的参数替换为您需要上传的本地文件路径（localpath），以及 COS 上存储的路径（cospath）。

举例：
将 HAI 算力实例根目录的 `test.pth` 文件上传到 COS 的 model 目录下。

```
coscmd upload test.pth model/
```

截图如下：



从 COS 下载文件到 HAI

1. 下载文件命令格式：

```
coscmd download <cospath> <localpath>
```

2. 下载文件夹命令格式：

```
coscmd download -r <cospath> <localpath>
```

> **注意：**

> 请将 "<>" 中的参数替换为您需要下载的 COS 上文件的路径（cospath），以及本地存储路径（localpath）。

举例：

下载 COS 上的 test.pth 到 HAI 算力实例根目录，并保持文件名不变。

```
coscmd download test.pth test.pth
```

截图如下：

```
(base) root@VM-0-100-ubuntu:~# coscmd download test.pth test.pth
Download cos://musk-1251783334/test.pth    =>    test.pth
 51%|                                        | 704M/1.35G [00:06<00:05, 133MB/s]
```

在 HAI 中查看 COS 文件

1. 查看文件命令格式：

```
coscmd list <cospath>
```

举例：

递归查询该存储桶下 doc/ 前缀下所有的文件列表。

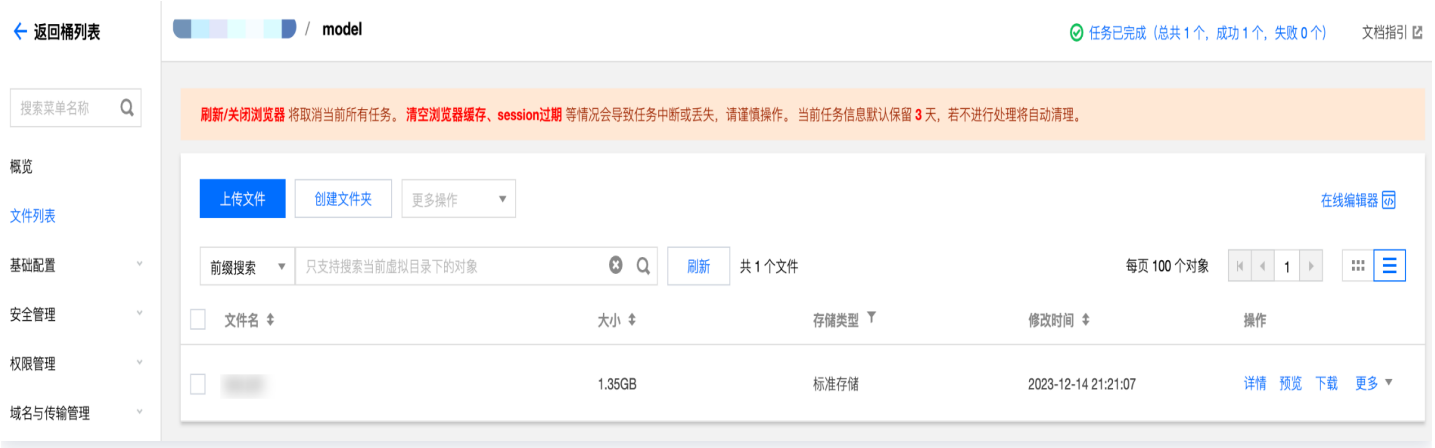
```
coscmd list model/
```

截图如下：

```
(base) root@VM-0-100-ubuntu:~# coscmd list model/
   model/                          0       STANDARD       2023-12-14 21:17:20
   model/test.pth       1445234339       STANDARD       2023-12-14 21:21:07
```

在 COS 控制台管理文件

COS 控制台提供最简单且易于上手的操作方式。用户无需编写代码或运行程序，可直接通过 COS 控制台操作存储桶、上传和分享数据以及数据备份、批量处理等，具体操作指引可参见 COS控制台指南 。



更多 COSCMD 操作

更多关于 COSCMD 的操作可参见 COSCMD 工具指南 。

通过 Filezilla 上传和下载数据

最近更新时间：2024-12-19 10:19:53

操作场景

FileZilla 是一款跨平台、免费开源的 FTP 软件，支持上传、下载和管理 HAI 算力实例中的文件，可以远程连接后进行文件传输操作，同时支持多线程传输和断电续传功能，是稳定性较高的文件管理方式。

前提条件

1. 您已获取登录 HAI 实例的 IP、用户名及密码。
2. 请确认本地计算机与 HAI 实例之间的网络连通正常，以及端口配置已放行22端口（创建实例时默认已开通22端口）。

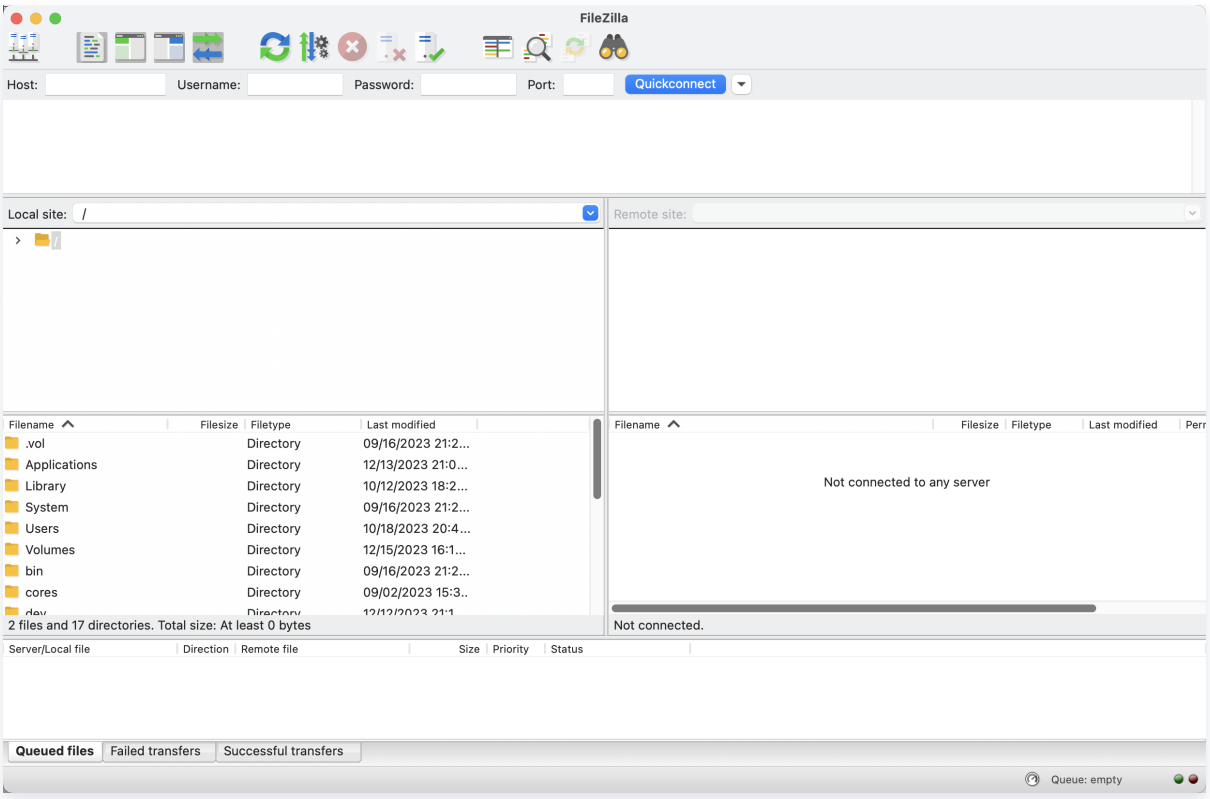
操作步骤

安装 FileZilla

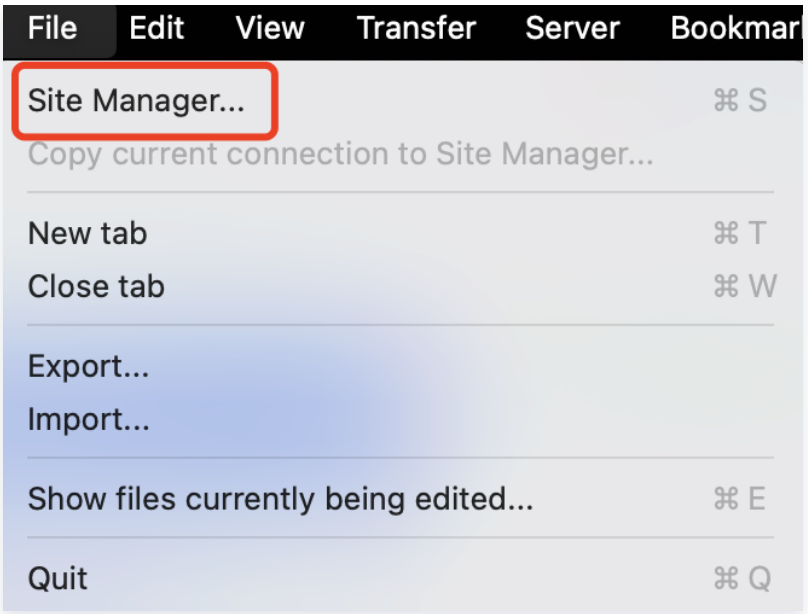
您可通过 FileZilla 官网下载最新版本的客户端，请参见 下载地址 。

远程连接到 HAI

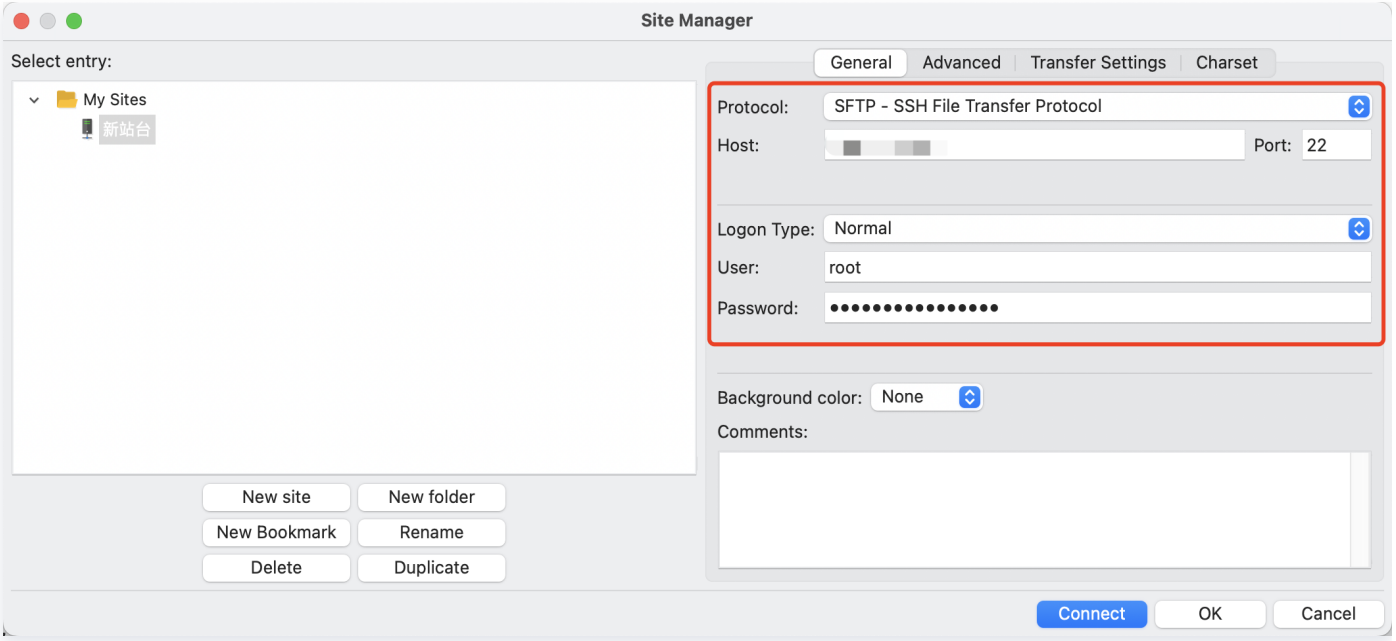
1. 打开安装好的 FileZilla 客户端。



2. 选择左上角菜单栏中的 File > Site Manager 打开站点配置页面。

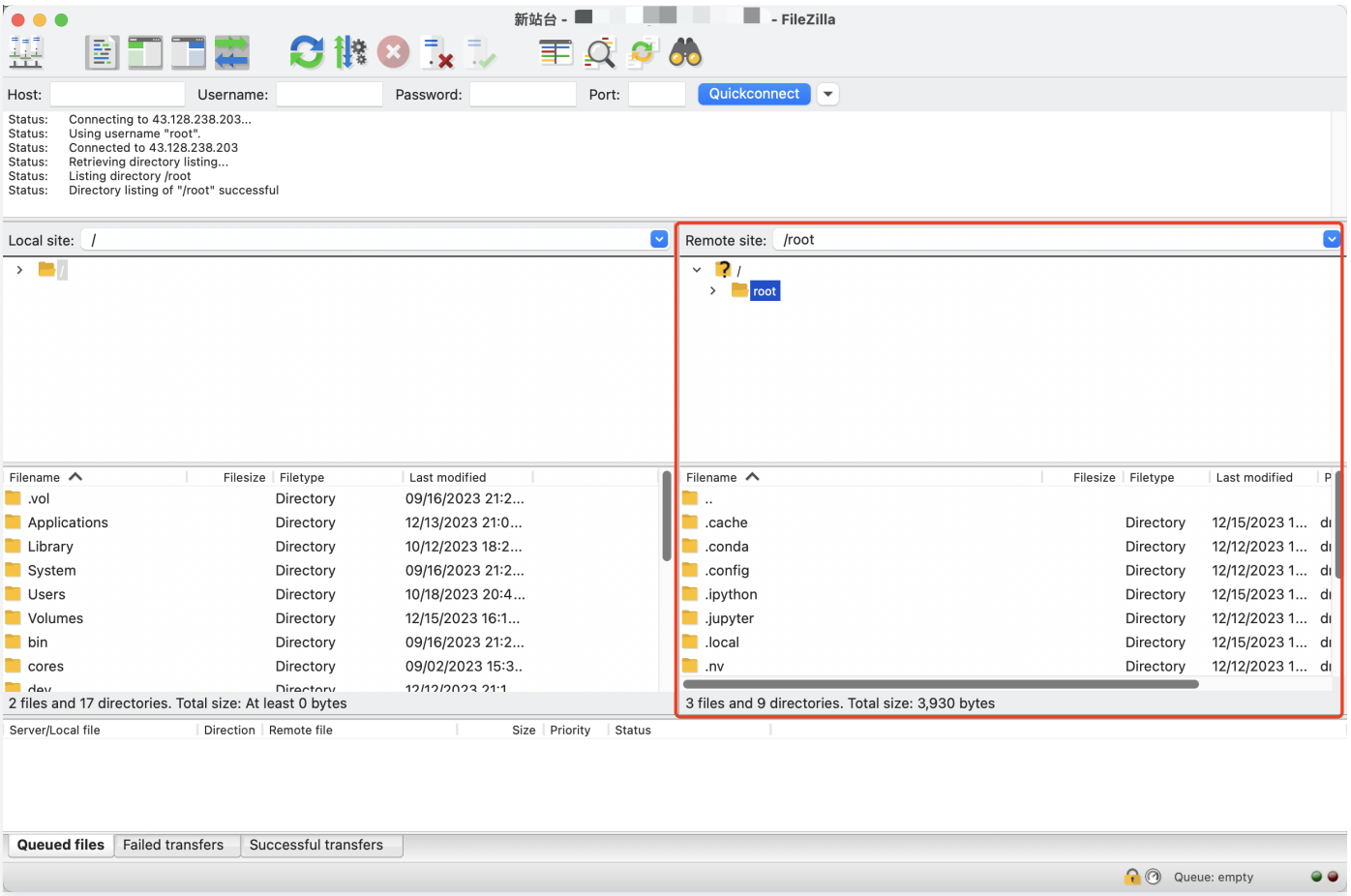


3. 在Site Manager 弹窗中填写 HAI 算力实例的 SSH 连接凭证，包括 IP、用户名、密码。



- Protocol：选择 **SFTP**。
- Host：填写 HAI 算力实例公网 IP。
- Port：填写**22**。
- Logon Type：选择 **Normal**。
- User：填写 SSH 登录用户名，如 **root** 等。
- Password：填写 SSH 登录密码。

4. 单击 **Connect** 连接，**Romte site** 出现目录及文件列表代表连接成功。



从本地上传文件到 HAI

- 方式一：从 Local site 直接拖转文件到 Romte site。
- 方式二：在 Local site 中右键单击需要上传的文件或文件夹，单击 Upload 上传。

从 HAI 下载文件到本地

- 方式一：从 Romte site 直接拖转文件到 Local site。
- 方式二：在 Romte site 中右键单击需要下载的文件或文件夹，单击 Download 下载。

通过 JupyterLab 上传和下载数据

最近更新时间：2024-08-02 14:35:31

操作场景

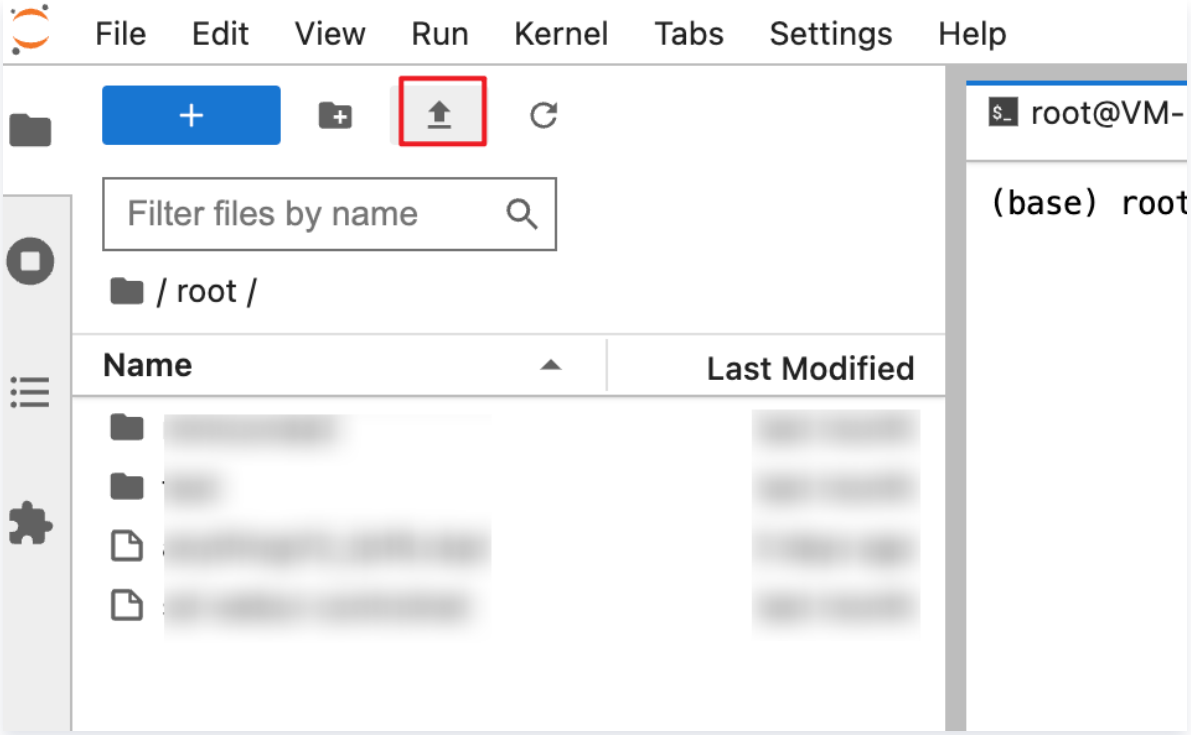
在提供 JupyterLab 的应用中，可以通过 JupyterLab 的文件管理器高效地上传/下载数据。目前 JupyterLab 暂不支持批量文件或文件夹的上传/下载。

操作步骤

从本地上传文件到 HAI

1、登录 高性能应用服务 HAI 控制台。
2、选择需要连接的算力，然后单击**连接算力**。在下拉菜单中，单击 **JupyterLab**。

3、进入 JupyterLab 后，在左侧选择**文件夹**页卡，单击**向上箭头**，即可上传本地文件。

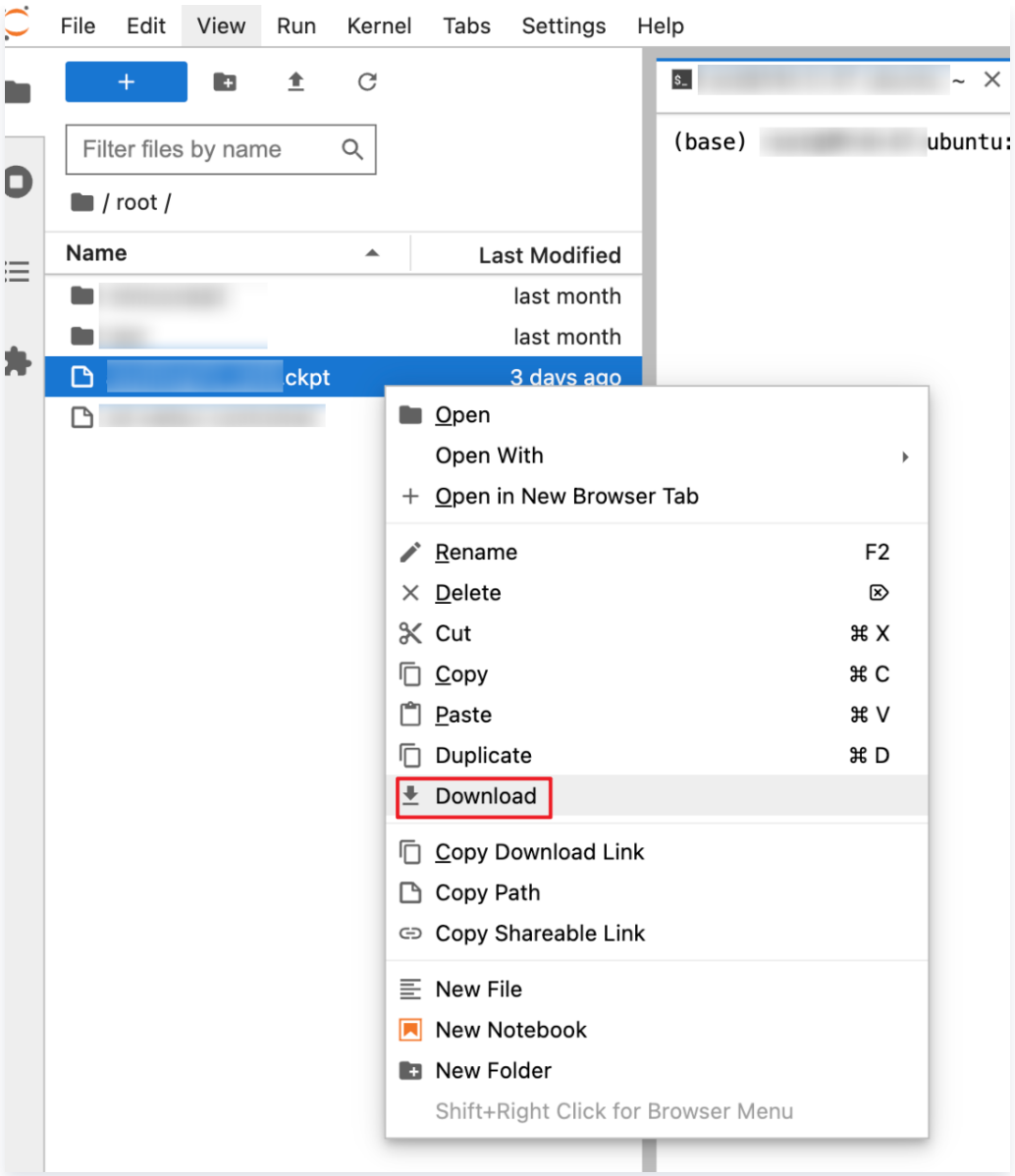


从 HAI 上下载文件到本地

1、登录 高性能应用服务 HAI 控制台。

2、选择需要连接的算力，然后单击**连接算力**。在下拉菜单中，单击 **JupyterLab**。

3、进入 JupyterLab 后，在左侧选择**文件夹**页卡，右键单击希望下载的文件，选中 **Download** 下载文件到本地。


File
Edit
View
Run
Kernel
Tabs
Settings
Help
Filter files by name
/ root /
Name
Last Modified
last month
last month
ckpt
3 days ago
(base)
ubuntu:
Open
Open With
Open in New Browser Tab
Rename
F2
Delete
Cut
⌘ X
Copy
⌘ C
Paste
⌘ V
Duplicate
⌘ D
Download
Copy Download Link
Copy Path
Copy Shareable Link
New File
New Notebook
New Folder
Shift+Right Click for Browser Menu

端口操作
编辑端口规则

最近更新时间：2024-12-19 10:19:53

操作场景

端口规则是保障算力实例网络安全的重要手段，您可以通过配置端口规则，允许或禁止公网或私网对算力实例的访问。

预置规则

高性能应用服务 HAI 会根据您选购的应用，默认放通必须的部分端口，您可在算力实例创建成功后根据实际情况编辑规则。一般情况下，预置规则即可满足应用基本操作和使用，无需进行额外调整。

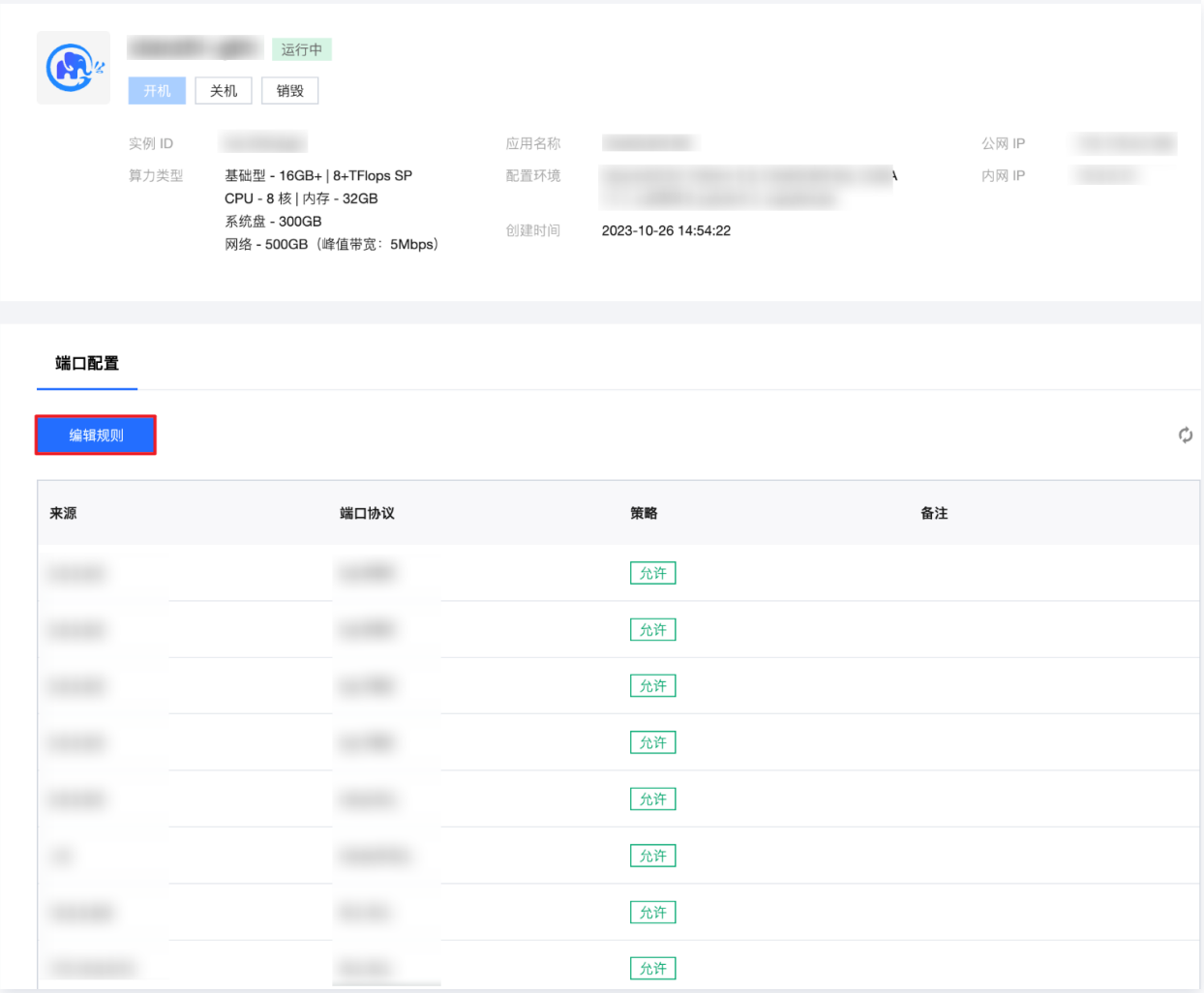
相关概念

- 出站：数据从实例内通过公网或内网传输至实例外。
- 入站：数据从实例外通过公网或内网传输至实例内。

操作步骤

1. 登录 高性能应用服务 HAI 控制台。
2. 选择需要操作的实例，然后单击**实例名称**。

3. 在实例详情中选择**端口配置**页卡查看当前的端口规则，如需调整可单击**编辑规则**。



4. 在安全组规则页面，根据需要修改安全组规则所属的方向（入站/出站），单击**入站/出站规则**页签。

5. 找到需要修改的安全组规则，单击操作列的**编辑**，即可对已有规则进行修改。

入站规则
出站规则
添加规则
导入规则
优先级排序
全部编辑
删除
一键放通
教我设置
多个关键字用竖线 "|" 分隔，多个过滤标签用回车键分隔
来源
协议端口
策略
备注
修改时间
操作
允许
2023-10-17 18:12:31
编辑
插入
删除
允许
2023-10-17 18:07:59
编辑
插入
删除
允许
2023-10-17 16:30:00
编辑
插入
删除
允许
2023-10-17 16:29:11
编辑
插入
删除

默认端口说明

最近更新时间：2025-06-26 14:57:42

为保障高性能应用服务 HAI 的即开即用体验，腾讯云将根据您所选应用类型智能配置必要的默认安全端口，请您知晓默认端口对服务可用性和安全性的影响。调整或关闭默认端口，可能导致部分服务异常，如不可访问、远程连接中断、数据传输失败等，请合理评估后谨慎操作。

> **注意：**
> 您可通过配置端口规则保障 HAI 实例的网络安全，允许或禁止公网或私网对 HAI 实例的访问。详情请参见 编辑端口规则 。

默认端口规则

服务	说明	默认端口
Ollama	Ollama 是一个开源的大模型管理工具，提供了丰富的功能，包括模型的训练、部署、监控等。通过 Ollama，您可以轻松地管理本地的大模型，提高模型的部署效率。	6399
OpenWebUI	OpenWebUI 是一个可扩展、功能丰富、用户友好的自托管 AI 界面，完全离线操作。它支持各种 LLM 运行程序，包括 Ollama 和 OpenAI 兼容的 API。	6699
JupyterLab	JupyterLab 是一个可视化的交互式笔记本，可以很方便的创建和共享程序文档，支持实时代码、数学方程、可视化和 Markdown。	6888
AnythingLLM	AnythingLLM 是一个专属知识库应用程序。它可以将任何文档、资源（如网址链接、音频、视频）或内容片段转换为上下文，以便在 LLM 聊天期间作为参考使用。	6889
ChatbotUI	ChatbotUI 是一个高效、易用的聊天机器人界面。通过简洁的 UI 设计和易于部署的特性，ChatbotUI 成为了一个测试和使用 AI 聊天机器人的理想平台。	6889
ComfyUI	ComfyUI 是一个基于节点流程的 Stable Diffusion 操作界面。它将 Stable Diffusion 的流程拆分为各个节点，可通过自定义节点、拖拽连线来实现精准的工作流定制与可靠复现，从而完成更复杂的、自由度更高的图像生成工作。	6889
Gradio WebUI	Gradio WebUI 围绕机器学习模型或数据科学工作流程提供可视化的用户界面，可用于拖放图像、粘贴文本、录制声音等交互进行模型测试验证。	6889

版权所有：腾讯云计算（北京）有限责任公司

FTP	FTP 服务器所开放的端口，用于上传、下载。	21
SSH	SSH 服务端口，用于通过命令行模式远程连接 Linux 系统服务器。	22
SMTP	SMTP 服务器所开放的端口，用于发送邮件。	25
HTTP	用于网站服务例如 IIS、Apache、Nginx 等提供对外访问。	80
HTTPS	网页浏览端口，能提供加密和通过安全端口传输的另一种 HTTP。	443
常用代理端口	8080端口类似80端口，是被用于 WWW 代理服务的，可以实现网页浏览，经常在访问某个网站或使用代理服务器的时候，会加上“:8080”端口号。另外 Apache Tomcat Web Server 安装后，默认的服务端口就是8080。	8080

版权所有：腾讯云计算（北京）有限责任公司

学术加速
开启学术加速

最近更新时间：2025-06-26 14:57:42

操作场景

当您遇到部分资源平台访问、下载速度过慢或无法响应时，可尝试使用学术加速能力，提高部分资源平台的访问速度。

注意事项

- 目前加速能力仅支持部分地域及部分资源平台，详情请参见 支持范围 。
- 同地域下任一服务开启学术加速能力，将会同时开启同地域下所有服务的学术加速能力。
- 学术加速能力限时免费，限免结束时间以腾讯云官网公告为准。

支持范围

地域	是否支持加速
上海	已支持
北京	已支持
广州	已支持
南京	暂未支持
重庆	暂未支持
东京	无需加速
首尔	无需加速
法兰克福	无需加速
新加坡	无需加速
硅谷	无需加速

说明：
购买时可通过 高性能应用服务 HAI 购买页 地域选项，查看是否支持学术加速的说明。

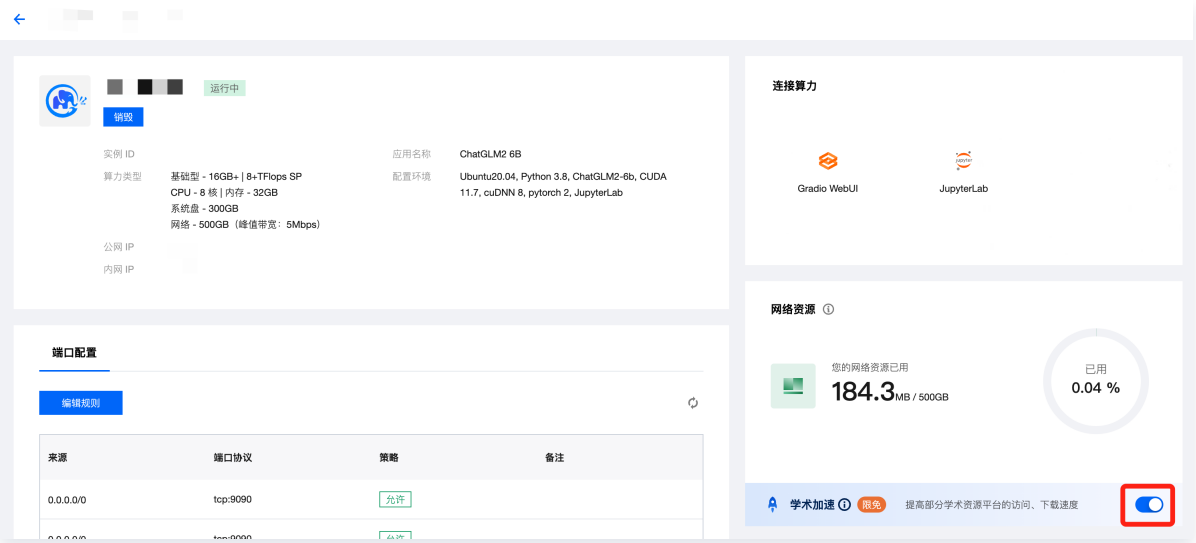
操作步骤

通过控制台概览开启

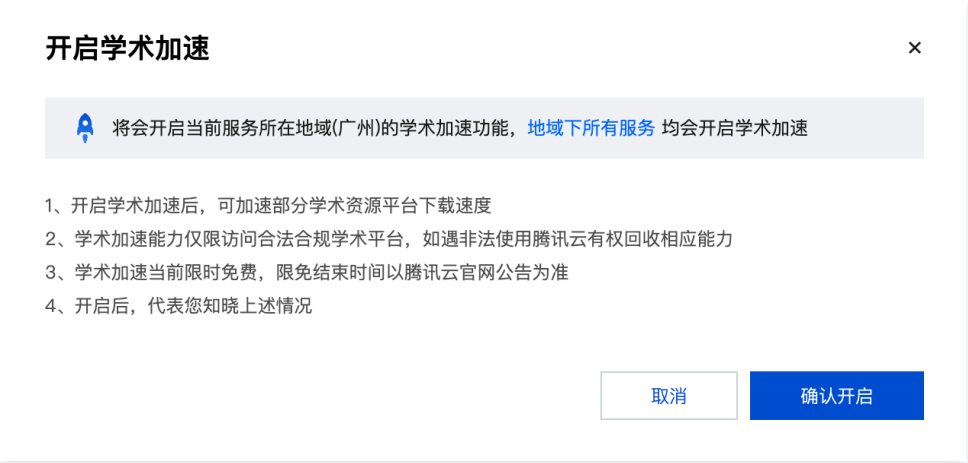
1. 登录 高性能应用服务HAI 控制台，单击**加速设置**。
2. 选择需要加速的地域，单击**右侧开关**启动加速能力。

通过单个实例开启

1. 登录 高性能应用服务HAI 控制台，选择需要加速的应用实例，单击进入实例详情页后，开启**学术加速**开关。



2. 在弹出的提示框中，确认您已知晓会同时开启同地域其他服务的学术加速能力，确认后单击**确认开启**。



关闭学术加速

最近更新时间：2025-06-26 14:57:42

操作场景

当您无需继续访问、下载资源，或不依赖腾讯云提供的学术加速能力时，可主动关闭学术加速服务。

注意事项

- 关闭学术加速后，您将失去部分学术资源平台的网络优化能力，可能导致访问、下载速度变慢。
- 同地域下任一服务关闭学术加速能力，将会同时关闭同地域下所有服务的学术加速能力。

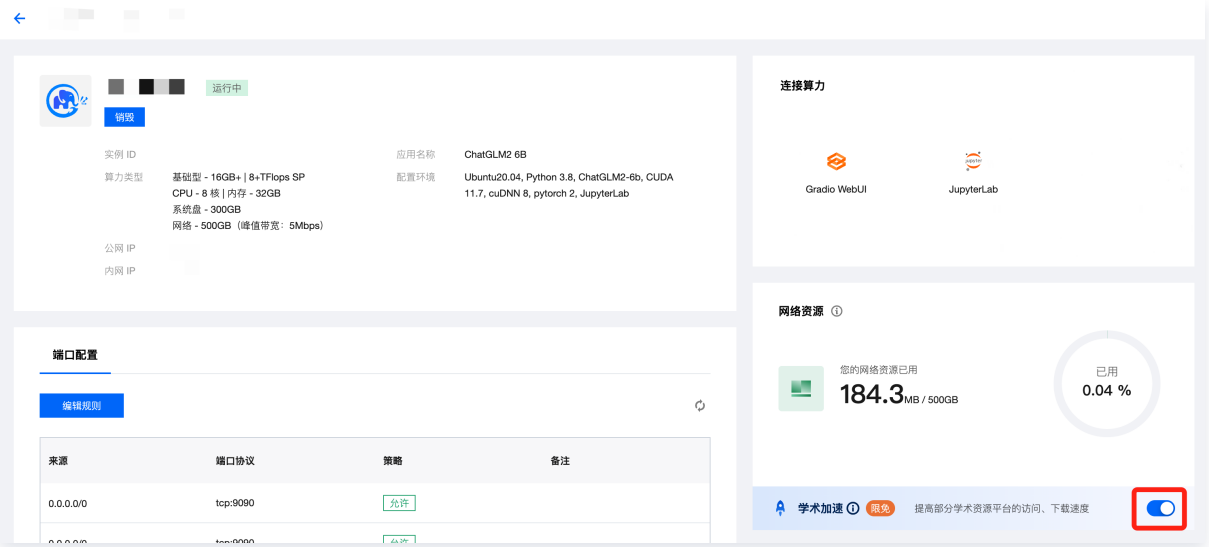
操作步骤

通过控制台概览关闭

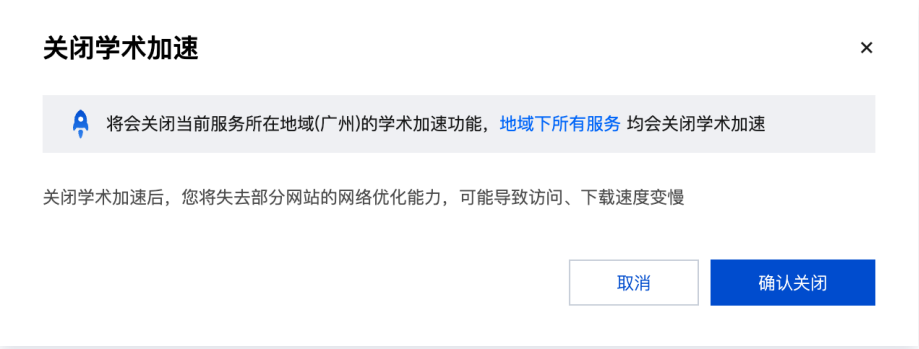
1. 登录 高性能应用服务HAI 控制台，单击**加速设置**。
2. 选择需要关闭加速的地域，单击**右侧开关**关闭加速能力。

通过单个实例关闭

1. 登录 高性能应用服务HAI 控制台，选择需要加速的应用实例，单击进入实例详情页后，关闭**学术加速**开关。



2. 在弹出的提示框中，确认您已知晓会同时关闭同地域其他服务的学术加速能力，确认后单击**确认关闭**。



应用操作
管理自定义应用

最近更新时间：2025-01-22 10:58:52

操作场景

除了使用腾讯云高性能应用服务 HAI 提供的公共应用外，您还可以创建自定义应用。创建自定义应用后，您可以使用该应用在高性能应用服务 HAI 控制台快速创建实例。

注意事项

- 如您的账户处于欠费状态：
  - 自定义应用功能将被停止，无法制作新的自定义应用。
  - 账号下已有的自定义应用将被隔离进入**待回收**状态，无法使用。若自定义应用进入**待回收**状态后15天内（包括第15天），账号余额未充正，则自定义应用将会自动删除。
- 实例制作自定义应用过程中，无法使用开机、关机、连接算力等功能，请提前做好准备。

操作步骤

创建自定义应用

1. 登录 高性能应用服务 HAI 控制台。
2. 选择需要连接的算力，然后单击**更多**。在下拉菜单中，单击**应用存档**。

3. 在应用存档的弹窗中，填写自定义应用信息，输入应用名称，应用描述，并单击**存档应用**。

4. 自定义应用制作完成后，您可以前往 应用管理 列表页进行查看。

删除自定义应用

1. 登录高性能应用服务 HAI 控制台，选择左侧导航栏中的 应用管理 。
2. 在应用管理页面右侧，选择应用所在**地域**。
3. 在列表中选择需要删除的应用，单击右侧的**删除**。
4. 在弹出的**删除应用**窗口中，单击**确定**即可删除该应用。

相关操作

使用自定义应用创建实例

您可参见以下步骤，使用自定义应用快速创建高性能应用服务 HAI 实例：

1. 登录高性能应用服务 HAI 控制台，选择左侧导航栏中的 应用管理 。
2. 在应用管理页面右侧，选择应用所在**地域**。
3. 在列表中单击应用所在行右侧的一键部署，进入高性能应用服务 HAI 购买页。如下图所示：

4. 在高性能应用服务 HAI 购买页中，您可参见 购买方式 选择实例其余配置项。

查看当前地域自定义应用信息

您可参见以下步骤，查看某地域下已有自定义应用的相关信息：

1. 登录高性能应用服务 HAI 控制台，选择左侧导航栏中的 应用管理 。
2. 在应用管理页面右侧，选择应用所在**地域**。

3. 在页面中可以查看当前地域自定义应用总大小及预估费用信息。如下图所示：

相关文档

计费概述

子账号授权

最近更新时间：2025-02-21 14:26:11

使用场景

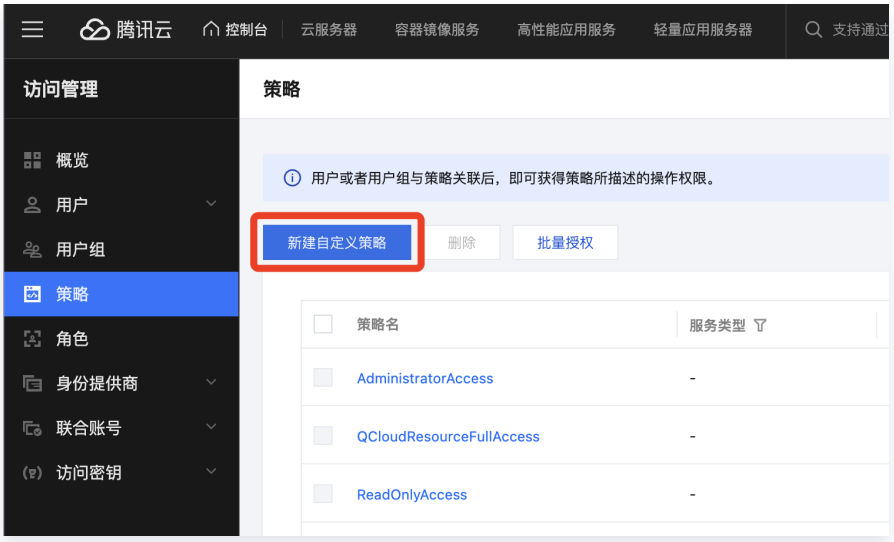
针对企业、高校需要精细化管控不同内部组织之间 HAI 资源的场景。

实现效果

支持对不同子账号，授权 HAI 的访问、使用、修改等操作。

操作步骤

1. 进入 访问管理控制台-策略 页面。
2. 单击**新建自定义策略**。



3. 单击**按策略生成器创建**。

4. 在**可视化策略生成器**区域，填写策略内容后，单击**下一步**。

- 效果：选择**允许**。
- 服务：选择**高性能应用服务(hai)**。
- 操作：按需选择，支持一键勾选全部操作。
- 资源：按需选择，支持一键勾选全部资源。

- 条件：按需填写。

注意：

HAI 使用过程中可能涉及访问其他云资源，建议同时添加**私有网络(vpc)**、**腾讯云自动化助手(tat)** 相关权限。

5. 在**基本信息**区域，为策略命名，策略名称创建后不支持修改。

6. 在**关联用户/用户组/角色**区域，按需选择需要授权的用户、用户组或角色。

7. 全部填写完毕后，单击**完成**。

8. 登录对应子账号，验证策略是否生效。

HAI 社区

最近更新时间：2024-11-18 16:53:52

HAI 社区是腾讯云提供面向AI和科学计算等 GPU 环境的容器镜像中心，收录丰富的官方与社区维护的开发资源。助力企业和开发者快速部署 AIGC 大模型、计算机视觉、自然语言处理、数据科学等容器，原生集成开发工具与组件。

服务入口

HAI 社区

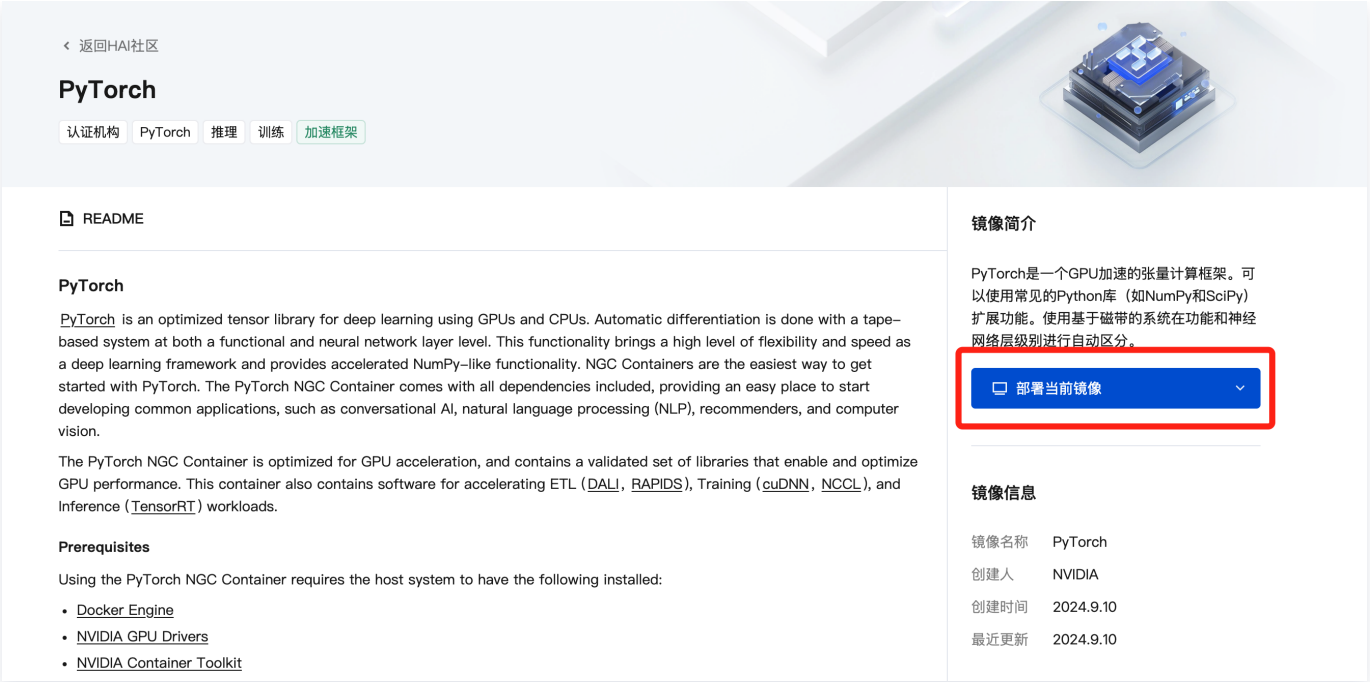
服务范围

社区资源支持通过腾讯云产品高速拉取，包括但不限于 HAI、HCC、CVM、TKE 等。

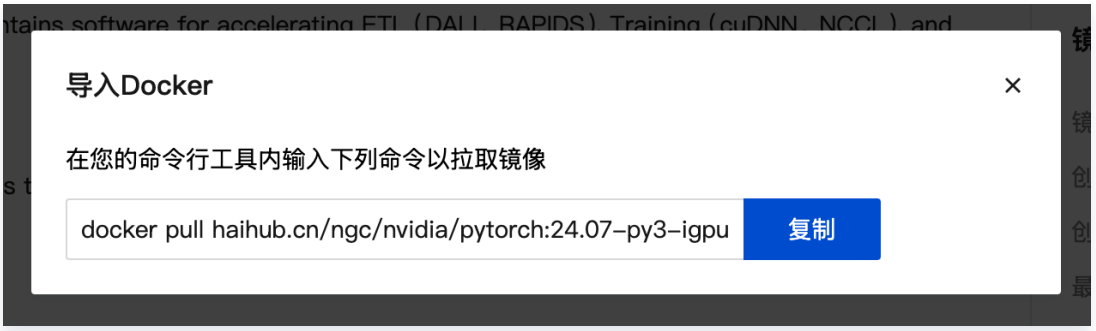
使用流程

1. 访问 HAI 社区首页，搜索您所需的 GPU 容器镜像。

2. 进入 GPU 容器镜像详情，单击**部署当前镜像**。



3. 复制 Docker Pull 短链接，进入 Terminal 内粘贴运行。



> **说明：**
> 目前 HAI 仅基础环境支持自定义部署容器及运行 Docker Pull 命令。