

ModelArts

Lite Server 用户指南

文档版本 01
发布日期 2025-11-24



版权所有 © 华为云计算技术有限公司 2025。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

1 Lite Server 使用前必读	1
1.1 Lite Server 使用流程	1
1.2 Lite Server 高危操作一览表	3
1.3 Lite Server 算力资源和镜像版本配套关系	6
2 Lite Server 资源开通（新版页面）	15
3 Lite Server 资源开通（旧版页面）	28
4 Lite Server 资源配置	38
4.1 Lite Server 资源配置流程	38
4.2 配置 Lite Server 网络	39
4.3 配置 Lite Server 存储	42
4.4 配置 Lite Server 软件环境（可选）	45
4.4.1 NPU 服务器上配置 Lite Server 资源软件环境	45
5 Lite Server 资源使用	60
5.1 LLM/AIGC 等模型基于 Lite Server 适配 NPU 的训练推理指导	60
5.2 GPT-2 基于 Lite Server 适配 GPU 的训练推理指导	60
6 Lite Server 资源管理	68
6.1 查看 Lite Server 服务器详情	68
6.2 开机或关机 Lite Server 服务器	70
6.3 同步 Lite Server 服务器状态	71
6.4 切换或重置 Lite Server 服务器操作系统	72
6.5 制作 Lite Server 服务器操作系统	76
6.6 Lite Server 资源热备管理	79
6.7 修改 Lite Server 名称	81
6.8 授权修复 Lite Server 节点	82
6.9 重启 Lite Server 服务器	86
7 Lite Server 插件管理	88
7.1 安装 Lite Server AI 插件	88
7.2 升级 Lite Server 中的 Ascend 驱动固件版本	90
7.3 安装/升级 Lite Server 中的 CES Agent 插件	94
7.4 Lite Server 节点故障诊断	95
7.5 Lite Server 节点漏洞修复	99

7.6 Lite Server 节点一键式压测.....	101
7.7 Lite Server 节点参数面网络配置.....	103
8 Lite Server 超节点管理.....	106
8.1 Lite Server 超节点扩容和缩容.....	106
8.2 Lite Server 超节点系统盘扩容.....	109
8.3 Lite Server 超节点定期压测.....	110
8.4 开启超节点 HCCL 通信算子级重执行机制.....	132
9 Lite Server 日志采集.....	136
9.1 NPU 日志收集上传.....	136
9.2 GPU 日志收集上传.....	142
10 Lite Server 监控告警.....	146
10.1 使用 CES 监控 Lite Server NPU 资源.....	146
10.2 使用 CES 监控 Lite Server NPU 事件.....	180
10.3 使用 CES 实现 Lite Server 监控和事件告警.....	192
10.4 使用 DCGM 监控 Lite Server GPU 资源.....	197
11 Lite Server 管理 CloudPond 的 NPU 资源.....	201
12 使用 CTS 审计 Lite Server 服务操作.....	207
13 退订 Lite Server 资源.....	211

1 Lite Server 使用前必读

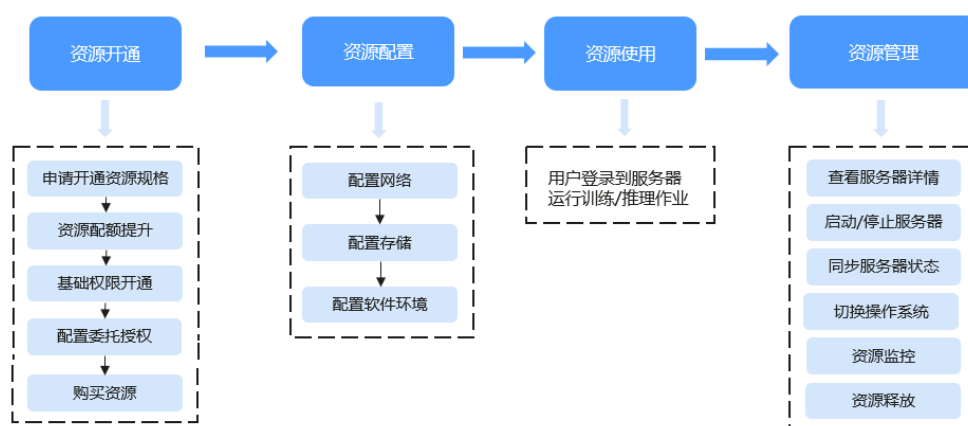
1.1 Lite Server 使用流程

对于算法工程师而言，日常的训练和推理工作需要一个灵活且强大的云上开发环境。然而，现有的云服务往往限制了用户的自定义能力，导致无法满足特定的软件安装和配置需求。

ModelArts Lite Server提供多样化的资源，赋予用户以root账号自主安装和部署AI框架、应用程序等第三方软件的能力，为用户打造专属的云上服务器环境。用户只需轻松选择服务器的规格、镜像、网络配置及密钥等基本信息，即可迅速创建Lite Server服务器，获取所需的云上物理资源，充分满足算法工程师在日常训练和推理工作中的需求。

如何快速上手并充分利用Lite Server的这些功能呢？本文旨在帮助您了解Lite Server的基本使用流程，帮助您快速上手。使用流程包含以下步骤。

图 1-1 使用流程



1. 资源开通

Lite Server资源需要先购买才能使用。

- 首先请联系客户经理确认Lite Server资源方案，部分规格为受限规格，因此需要申请开通您所需的资源规格。

- b.

Lite Server所需资源可能会超出云服务默认提供的资源配额（如ECS、EIP、SFS），因此需要[提交工单](#)提升资源配额。默认配额查看请参考[怎样查看我的配额？](#)。
- c.

为子用户账号开通Lite Server功能所需的基础权限。
- d.

由于ModelArts服务在使用过程中会访问其他依赖服务，因此需要给ModelArts进行委托授权。
- e.

在ModelArts控制台开通Lite Server资源。
2.

资源配置

完成资源开通后，需要对网络、存储、软件环境进行相关配置。
3.

资源使用

完成资源配置后，您可以登录到服务器进行训练和推理，具体案例可参考[Lite Server资源使用](#)。
4.

资源管理

Lite Server提供启动、停止、切换操作系统等管理能力，您可在ModelArts控制台上对资源进行管理。

表 1-1 相关名词解释

名词	含义
普通节点	单一物理主机或虚拟主机，提供基础的独立计算、存储和网络资源，包括裸金属服务器和弹性云服务器两种。
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。 Lite Server支持多种服务器类型，包括弹性云服务器。 更多弹性云服务器介绍请见 弹性云服务器ECS 。
裸金属服务器	裸金属服务器是一款兼具弹性云服务器和物理机性能的计算类服务，为您和您的企业提供专属的云上物理服务器，为核心数据库、关键应用系统、高性能计算、大数据等业务提供卓越的计算性能以及数据安全。 Lite Server支持多种服务器类型，包括裸金属服务器。 更多裸金属服务器的介绍请见 裸金属服务器BMS 。
超节点	超节点（Hypernode Server）是华为云提供的一种高性能计算资源，主要用于AI大模型训练和推理场景。超节点由多个节点组成，其内部NPU采用特定网络连接方式形成一种超平面网络，可以提供更快的网络传输速率。超节点服务器支持的资源是Snt9b23，仅支持西南-贵阳一、华北三、和华东二区域。
密钥对	Lite Server支持SSH密钥对的方式进行登录，用户无需输入密码就可以登录到Lite Server，因此可以防止由于密码被拦截、破解造成的账户密码泄露，从而提高Lite Server的安全性。 说明 为保证云服务器安全，未进行私钥托管的私钥只能下载一次，请妥善保管。

名词	含义
虚拟私有云	虚拟私有云（Virtual Private Cloud，VPC）为Lite Server构建隔离的、用户自主配置和管理的虚拟网络环境，提升用户云中资源的安全性，简化用户的网络部署。您可以在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的访问规则，加强Lite Server的安全保护。 更多VPC介绍请见虚拟私有云VPC。

1.2 Lite Server 高危操作一览表

ModelArts Lite Server在日常操作与维护过程中涉及的高危操作，需要严格按照操作指导进行，否则可能会影响业务的正常运行。

高危操作风险等级说明：

- 高：对于可能直接导致业务失败、数据丢失、系统不能维护、系统资源耗尽的高危操作。
- 中：对于可能导致安全风险及可靠性降低的高危操作。
- 低：高、中风险等级外的其他高危操作。

表 1-2 高危操作一览表

操作对象	操作名称	风险描述	风险等级	应对措施
操作系统	升级/修改操作系统内核	如果升级/修改操作系统内核，很可能导致驱动和内核版本不兼容，从而导致OS无法启动，或者基本功能不可用。相关高危命令如：apt-get upgrade（升级系统中全部软件，包括内核）。 查看当前内核命令： uname -a	高	如果需要升级/修改，请 联系华为云技术支持 。
	切换或者重置操作系统	服务器在进行过“切换或者重置操作系统”操作后，EVS系统盘ID发生变化，和下单时订单中的EVS ID已经不一致，因此EVS系统盘将不支持扩容，并显示信息：“当前订单已到期，无法进行扩容操作，请续订”。	低	切换或者重置操作系统 后，建议通过挂载数据盘EVS或挂载SFS盘等方式进行存储扩容。

操作对象	操作名称	风险描述	风险等级	应对措施
	云服务器业务正常运行时，用户在其系统中删除网卡路由或者对网卡执行ifconfig down和ifconfig up等相关破坏网络的操作	该操作会将网络服务重启重新触发DHCP获取IP地址和路由，可能导致网卡路由丢失而影响节点不可用。	高	建议置操作系统恢复，重置操作系统之前请确保您的数据已备份。
	修改如net.ipv4.ip_forward等内核参数	可能影响云服务器路由转发功能，导致网络不通。	中	修改内核参数为net.ipv4.ip_forward=1
	开启系统防火墙	可能影响hccl、nccl等性能测试；可能影响多机多卡训练任务的性能。	低	关闭防火墙
	修改时区	会引起节点时间发生跳变，影响业务。	中	恢复时区
驱动或固件	升级NPU驱动或者固件相关	可能导致驱动固件不匹配，导致服务器不可用，影响业务。	中	建议重置操作系统恢复，重置操作系统之前请确保您的数据已备份。
	更改GPU驱动	可能导致驱动固件不匹配，导致服务器不可用，影响业务。	中	建议重置操作系统恢复，重置操作系统之前请确保您的数据已备份。
	更改SDI卡驱动	可能导致网卡不可用，导致服务器不可用，影响业务。	中	建议重置操作系统恢复，重置操作系统之前请确保您的数据已备份。
网络	修改网卡MAC地址或IP地址	如果操作不当，会导致虚拟机通信异常、业务中断并且还会影响其他服务。	高	回退相关修改，如果回退失败。建议重置操作系统恢复，重置操作系统之前请确保您的数据已备份。

操作对象	操作名称	风险描述	风险等级	应对措施
	添加/删除/编辑iptables规则或重启iptables服务	导致业务访问请求被拒绝。	高	回退相关修改，如果回退失败。建议 重置操作系统 恢复，重置操作系统之前请确保您的数据已备份。
操作系统内置软件	升级、降级、卸载系统内置软件如python3版本等	可能导致系统内Network等网络配置软件异常，导致服务器网卡配置失败，导致节点不可用	高	回退相关修改，如果回退失败。建议 重置操作系统 恢复，重置操作系统之前请确保您的数据已备份。
目录/文件	修改操作系统内的root、opt等关键系统目录或文件如/etc/hccn.conf和/etc/netplan/roce.yaml	可能影响系统正常功能，导致云服务器不可用	高	回退相关修改，如果回退失败。建议 重置操作系统 恢复，重置操作系统之前请确保您的数据已备份。
	修改目录/文件权限	修改可能引起服务异常	高	回退相关修改。
服务器操作	禁止在服务器实例发放、初始化、添加磁盘、删除磁盘、删除实例过程中，对服务器执行非查询类操作，如关机、开机等操作。	可能会导致相应的云服务器业务操作失败。	中	建议 重置操作系统 恢复，重置操作系统之前请确保您的数据已备份。
	切换或者重置操作系统	服务器在进行过“切换或者重置操作系统”操作后，EVS系统盘ID发生变化，和下单时订单中的EVS ID已经不一致，因此EVS系统盘将不支持扩容，并显示信息：“当前订单已到期，无法进行扩容操作，请续订”。	低	切换或者重置操作系统后，建议通过挂载数据盘EVS或挂载SFS盘等方式进行存储扩容，具体操作请参考 配置 Lite Server存储 章节。

操作对象	操作名称	风险描述	风险等级	应对措施
进程	执行service network restart 命令 停止系统关键进程，如sshd ces-agent等进程	可能导致业务发放失败。 导致远程访问云服务器失败。 导致数据采集失败，影响监控指标上报。	高	重新启动已关闭的服务。
数据盘	修改数据盘挂载方式，挂载点等	可能导致正在使用的业务出现异常。	低	请确保该数据盘无业务使用。
安全组	修改端口通信协议 放行22等高危端口 未设置IP白名单	可能存在网络被攻击的风险，影响服务器正常业务。	中	恢复到原有内容。

1.3 Lite Server 算力资源和镜像版本配套关系

ModelArts Lite Server了提供多种操作系统镜像，您可在创建Lite Server资源前了解当前支持的镜像及对应详情，供您在[创建Lite Server资源](#)时选择。

NPU Snt9b23 超节点服务器支持的镜像详情

镜像名称：HCE2.0-Arm-64bit-for-Snt9b23-with-25.2.1-7.7.0.9.220-CANN8.1.RC2-v2

表 1-3 镜像详情

软件类型	版本详情
操作系统	HCE2.0
内核版本	5.10.0-182.0.0.95.r2220_157.hce2.aarch64
架构类型	aarch64
固件版本	7.7.0.9.220
npu-driver	25.2.1
Ascend-cann-toolkit	8.1.RC2
cann-kernels	8.1.RC2
Ascend-mindx-toolbox	7.0.RC1
Docker	27.2.0
Ascend-docker-runtime	7.0.RC1

软件类型	版本详情
Mpich	4.1.3
MCU	25.52.29
CES Agent	2.8.2.2

NPU Snt9b 裸金属服务器支持的镜像详情

- 镜像名称：Ubuntu22.04-Arm-64bit-for-Snt9A2-BareMetal-with-24.1.0.6-7.5.0.5.220-CANN8.0.1-v2

表 1-4 镜像详情

软件类型	版本详情
操作系统	Ubuntu 22.04
内核版本	5.15.0-91-generic
架构类型	aarch64
固件版本	7.5.0.5.220
npu-driver	24.1.0.3
Ascend-cann-toolkit	8.0.1
cann-kernels	8.0.1
Ascend-mindx-toolbox	6.0.0
Docker	26.0.0
Ascend-docker-runtime	v6.0.0
Mpich	3.2.1
MCU	23.3.16
CES-Agent	2.8.2.1

- 镜像名称：HCE2.0-Arm-64bit-for-Snt9A2-BareMetal-with-24.1.0.6-7.5.0.5.220-CANN8.0.1-v2

表 1-5 镜像详情

软件类型	版本详情
操作系统	HCE2.0
内核版本	5.10.0-136.12.0.86.r1526_92.hce2.aarch64

软件类型	版本详情
架构类型	aarch64
固件版本	7.5.0.5.220
npu-driver	24.1.0.3
Ascend-cann-toolkit	8.0.1
cann-kernels	8.0.1
Ascend-mindx-toolbox	6.0.0
Docker	18.09.0
Ascend-docker-runtime	v6.0.0
Mpich	3.2.1
MCU	23.3.16
CES-Agent	2.8.2.1

NPU Snt9b 弹性云服务器支持的镜像详情

- 镜像名称：HCE2.0-Arm-64bit-for-Snt9A2-ECS-BareMetal-with-24.1.0.6-7.5.0.5.220-CANN8.0.1-v2

表 1-6 镜像详情

软件类型	版本详情
操作系统	HCE2.0
内核版本	5.10.0-136.12.0.86.r1526_92.hce2.aarch64
架构类型	aarch64
固件版本	7.5.0.5.220
npu-driver	24.1.0.3
Ascend-cann-toolkit	8.0.1
cann-kernels	8.0.1
Ascend-mindx-toolbox	6.0.0
Docker	18.09.0
Ascend-docker-runtime	v6.0.0
Mpich	3.2.1

软件类型	版本详情
MCU	23.3.16
CES-Agent	2.8.2.1

- 镜像名称：Ubuntu22.04-Arm-64bit-for-Snt9A2-ECS-BareMetal-with-24.1.0.6-7.5.0.5.220-CANN8.0.1-v2

表 1-7 镜像详情

软件类型	版本详情
操作系统	Ubuntu 22.04
内核版本	5.15.0-91-generic
架构类型	aarch64
固件版本	7.5.0.5.220
npu-driver	24.1.0.3
Ascend-cann-toolkit	8.0.1
cann-kernels	8.0.1
Ascend-mindx-toolbox	6.0.0
Docker	26.0.0
Ascend-docker-runtime	v6.0.0
Mpich	3.2.1
MCU	23.3.16
CES-Agent	2.8.2.1

NPU Snt3PD 弹性云服务器支持的镜像详情

镜像名称：Huawei-Cloud-EulerOS-2.0-64bit-for-kAi2p-with-HDK-24.1.0.1-and-CANN-8.0.1

软件类型	版本详情
操作系统	Huawei Cloud EulerOS 2.0
内核版本	5.10.0-182.0.0.95.r2762_220.hce2.aarch64
架构类型	aarch64
npu-driver	24.1.0.1

软件类型	版本详情
Ascend-cann-toolkit	8.0.1
cann-kernels	8.0.1
Ascend-mindx-toolbox	6.0.0
Docker	18.09.0
Ascend-docker-runtime	v6.0.0
Mpich	3.2.1
MCU	24.5.8

GP Ant8 裸金属服务器支持的镜像详情

镜像名称：Ubuntu-22.04-x86-for-Ant1-Ant8-BareMetal-with-RoCE-and-NVIDIA-550.90.07-CUDA-12.4

表 1-8 镜像详情

软件类型	版本详情
操作系统	Ubuntu 22.04
内核版本	5.15.0-25-generic
架构类型	x86
驱动版本	550.90.07
cuda	12.4
nv-fabricmanager	550.90.07-1
nv-container-toolkit	1.17.5-1
libnccl2	2.26.2-1+cuda12.4
libnccl-dev	2.26.2-1+cuda12.4
Docker	20.10.23
Mpich	4.1.5a1

GP Vnt1 裸金属服务器支持的镜像详情

 说明

Vnt1规格在华北-北京四、华北-北京一和华东-上海—虽然规格相同，但是产品的配置、发布时间都存在很大差异，因此镜像不能共用。

- 镜像名称：Ubuntu-22.04-for-BareMetal-Vnt1-p3-with-NV-535-CUDA-12.2（仅限于华北-北京一、华北-北京四、华南-广州）

表 1-9 镜像详情

软件类型	版本详情
操作系统	Ubuntu 22.04
内核版本	5.15.0-25-generic
架构类型	x86
驱动版本	535.54.03
cuda	12.2
container-toolkit	1.16.1-1
libnccl2	2.21.5-1+cuda12.2
libnccl-dev	2.21.5-1+cuda12.2
Docker	24.0.5
CES-agent	2.7.2.1

- 镜像名称：Ubuntu-18.04-for-BareMetal-Vnt1-p6-with-NV-470-CUDA-11.4-Uniagent（仅限于华东-上海一）

表 1-10 镜像详情

软件类型	版本详情
操作系统	Ubuntu 22.04
内核版本	5.15.0-25-generic
架构类型	x86
驱动版本	545.23.08
cuda	12.3
nv-container-toolkit	1.17.4.1
libnccl-dev	2.21.5-1+cuda12.2
libnccl2	2.20.3-1+cuda12.3
Docker	28.0.0
CES-agent	2.7.5.1

GP Ant1 裸金属服务器支持的镜像详情

镜像名称：Ubuntu-22.04-x86-for-Ant1-Ant8-BareMetal-with-RoCE-and-NVIDIA-550.90.07-CUDA-12.4

表 1-11 镜像详情

软件类型	版本详情
操作系统	Ubuntu 22.04
内核版本	5.15.0-25-generic
架构类型	x86
驱动版本	550.90.07
cuda	12.4
nv-fabricmanager	550.90.07-1
nv-container-toolkit	1.17.5-1
libnccl2	2.26.2-1+cuda12.4
libnccl-dev	2.26.2-1+cuda12.4
Docker	20.10.23
Mpich	4.1.5a1

GP Hnt02 弹性云服务器支持的镜像详情

- 镜像名称：HCE2.0-x86-for-H20-NV-535-CUDA-12.2(仅限华北-乌兰察布一、华东二)

软件类型	版本详情
操作系统	HCE 2.0
内核版本	5.10.0-182.0.0.95.r1941_123.hce2.x86_64
架构类型	x86_64
驱动版本	535.183.01
cuda	12.2
nv-fabricmanager	535.183.01
libnccl	2.18.5-1+cuda12.2
libnccl-dev	2.18.5-1+cuda12.2
CES-Agent	2.7.2.1

- 镜像名称：Ubuntu22.04_x86_for_h20_Driver-535-and-CUDA-12.2(仅限华北-乌兰察布一、华东二)

软件类型	版本详情
操作系统	Ubuntu 20.04 server 64bit

软件类型	版本详情
内核版本	5.15.0-107-generic
架构类型	x86_64
驱动版本	535.183.01
cuda	12.2
nv-fabricmanager	535.183.01
libnccl2	2.18.5-1+cuda12.2
libnccl-dev	2.18.5-1+cuda12.2
Docker	28.0.1

GP Lnt002 弹性云服务器支持的镜像详情

镜像名称：Ubuntu-22.04-server-64bit-with-Tesla-Driver-535.183.01-and-CUDA-12.2(仅限华北-北京四、华东-上海一、中东-利雅得、亚太-雅加达)

软件类型	版本详情
操作系统	Ubuntu 20.04 server 64bit
内核版本	5.15.0-92-generic
架构类型	x86
驱动版本	535.183.01
cuda	12.2
nv-container-toolkit	1.17.1
Docker	27.3.1
CES-Agent	2.7.3.t2

镜像名称：HCE2.0-x86-for-L2-NVIDIA-535-CUDA-12.2(仅限华北-北京四、华东-上海一、亚太-新加坡)

软件类型	版本详情
操作系统	Huawei Cloud EulerOS 2.0(x86_64)
内核版本	5.10.0-60.18.0.50.r865_35
架构类型	x86
驱动版本	535.183.01
cuda	12.2

软件类型	版本详情
nv-container-toolkit	1.13.5-1
Docker	18.09.0
CES-Agent	2.6.7.1

GP Ant03 弹性云服务器支持的镜像详情

镜像名称：Ubuntu 22.04 server 64bit with Tesla Driver 470.182.03 and CUDA 11.4
(仅限华南-广州、华东-上海一)

软件类型	版本详情
操作系统	Ubuntu 22.04 server 64bit
内核版本	5.15.0-60-generic
架构类型	x86
驱动版本	470.182.03
cuda	11.4
nv-fabricmanager	470.182.03-1

2 Lite Server 资源开通（新版页面）

说明

为了提升创建Lite Server资源效率，ModelArts对创建页面进行了一系列的易用性改进。现推出新版页面，旨在简化操作流程并增强界面的直观性。

场景描述

本章节主要介绍如何在ModelArts控制台上购买Lite Server算力资源，及购买前的准备工作。

用户先完成资源配额提升、配置基础权限、设置ModelArts委托授权等准备工作。在购买资源时，用户创建实例并支付订单，支付完成后等待约20~60分钟，资源创建成功后即可配置弹性公网IP进行访问，开展相关AI开发工作。

约束限制

Lite Server超节点当前仅支持“包年/包月”计费模式。

Lite Server普通节点（ECS或BMS）的所有资源规格均支持“包年/包月”计费模式。

资源开通流程

图 2-1 Lite Server 资源开通流程图

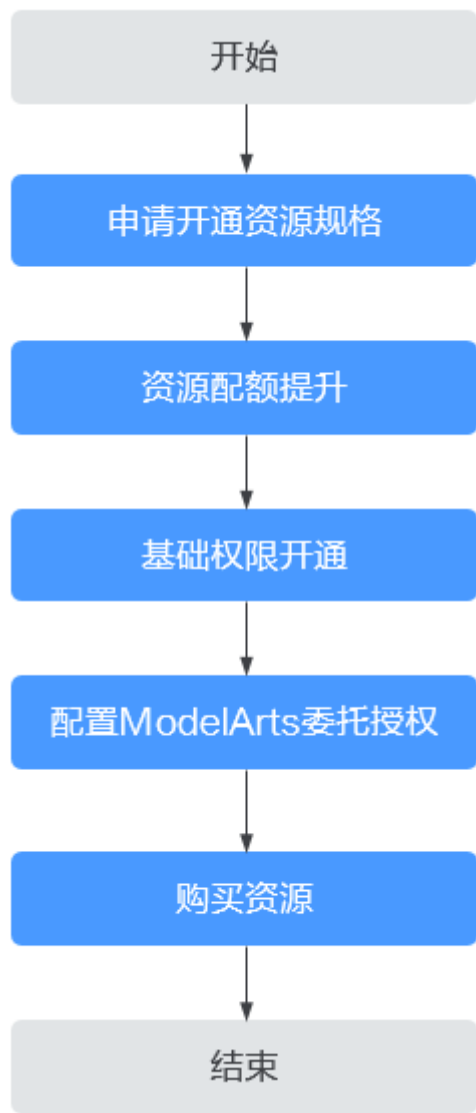


表 2-1 Lite Server 资源开通流程

阶段	任务
准备工作	1、申请开通资源规格。
	2、资源配额提升。
	3、基础权限开通。
	4、配置ModelArts委托授权。
购买Lite Server资源	5、在ModelArts控制台上购买轻量算力节点 (Lite Server)资源。

步骤 1：申请开通资源规格

请联系客户经理确认 Lite Server 资源方案、申请要开通资源的规格，如果无客户经理可提交工单。

步骤 2：提升资源配额

由于 Lite Server 所需资源可能会超出云服务默认提供的资源（如 ECS、EIP、SFS、内存大小、CPU 核数），因此需要提升资源配额。

1. 登录[华为云管理控制台](#)。
2. 在顶部导航栏单击“资源 > 我的配额”，进入服务配额页面。
3. 单击右上角“申请扩大配额”，填写申请材料后提交工单。

说明

配额需大于需要开通的资源，且在购买开通前完成提升，否则会导致资源开通失败。

步骤 3：开通基础权限

开通基础权限需要登录管理员账号，为子用户账号开通 Lite Server 功能所需的基础权限，包括 ModelArts FullAccess、BMS FullAccess、ECS FullAccess、VPC FullAccess、VPC Administrator、VPCEndpoint Administrator、CloudMatrixFullAccessPolicy（超节点），即允许子用户账号同时可以使用这些云服务。

1. 登录[统一身份认证服务管理控制台](#)。
2. 单击目录左侧“用户组”，然后在页面右上角单击“创建用户组”。
3. 填写“用户组名称”并单击“确定”。
4. 在用户组页面，在目标用户组名称的操作列单击“用户组管理”，将需要配置权限的用户加入用户组中。

图 2-2 用户组管理



5. 单击用户组名称，进入用户组详情页。
6. 在授权记录页签下，单击“授权”。

图 2-3 “配置权限”



7. 在搜索栏输入“ModelArts FullAccess”，并勾选“ModelArts FullAccess”。

图 2-4 ModelArts FullAccess



以相同的方式，依次添加：BMS FullAccess、ECS FullAccess、VPC FullAccess、VPC Administrator、VPCEndpoint Administrator、CloudMatrixFullAccessPolicy（超节点）。（Server Administrator、DNS Administrator为依赖策略，会自动被勾选）。

8. 单击“下一步”，授权范围方案选择“所有资源”。
9. 单击“确定”，完成基础权限开通。

步骤 4：在 ModelArts 上创建委托授权

ModelArts Lite Server在任务执行过程中需要访问用户的其他服务，典型的就容器使用过程中需要到SWR服务拉取镜像。在这个过程中，就出现了ModelArts“代表”用户去访问其他云服务的情形。从安全角度出发，ModelArts代表用户访问任何云服务之前，均需要先获得用户的授权，而这个动作就是一个“委托”的过程。用户授权ModelArts代表自己访问特定的云服务，以完成其在ModelArts平台上执行的AI计算任务。

- 新建委托

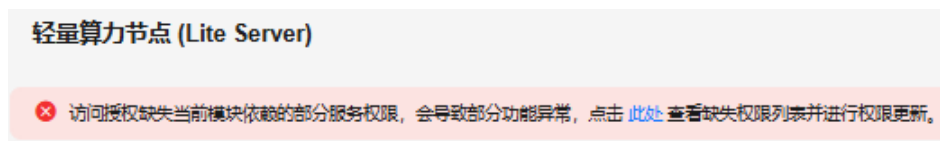
第一次使用ModelArts时需要创建委托授权，授权允许ModelArts代表用户去访问其他云服务。进入ModelArts控制台的“系统管理 > 权限管理”页面，单击“添加授权”，根据提示进行操作。

- 更新委托

如果之前给ModelArts创过委托授权，此处可以更新授权。

- a. 进入到[ModelArts管理控制台](#)的“资源管理 > 轻量算力节点 (Lite Server)”页面，查看是否存在授权缺失的提示。

图 2-5 Lite Server 权限缺失提示



- b. 如果存在授权缺失，根据提示，单击“此处”更新委托。根据提示选择“追加至已有授权”，单击“确定”，系统会提示权限更新成功。

图 2-6 追加授权



步骤 5：购买轻量算力节点 (Lite Server)资源

购买轻量算力节点 (Lite Server)资源的过程即创建资源过程。

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入资源列表。
3. 单击右上角的“购买轻量算力节点”，进入“购买轻量算力节点”页面，在该页面填写相关参数信息。

说明

购买界面存在新版和旧版2个版本，以下参数配置表中展示的参数顺序遵循的是新版购买页面，旧版购买页面的参数顺序和新版页面有差异，但具体的参数解释不变。

图 2-7 购买轻量算力节点时的基础配置



Basic Configurations

Node Type

Common node
A physical or virtual server that provides independent basic compute, storage, and network resources.

Supernodes
A converged node that provides a large-scale compute resource pool and supports flexible allocation and high-density deployment.

Resource Type

BMS
A BMS combines the scalability of ECSs with the high performance of physical servers, delivering dedicated cloud servers for you and your enterprise.

ECS
ECS provides scalable, on-demand cloud servers that deliver secure, flexible, and efficient application environments with reliable, uninterrupted services.

Billing Mode

Yearly/Monthly

Pay-per-use

Region

Select the region nearest to you to ensure the lowest latency possible.

AZ

AZ1

AZ2

AZ3

表 2-2 基础配置参数说明

参数名称	说明
节点类型	- 普通节点：单一物理主机或虚拟主机，提供基础的独立计算、存储和网络资源，包括裸金属服务器和弹性云服务器两种。 - 超节点：融合架构节点，提供大规模计算资源池，支持灵活调配和高密度部署。超节点专门用于支持大规模的模型训练和推理任务。这些服务器通常配备有多个计算卡（如昇腾 NPU），能够提供强大的计算能力，以满足高负载的算力需求。超节点资源即Snt9b23资源，仅支持西南-贵阳一、华北三、和华东二区域。
资源类型	当“节点类型”选择“普通节点”时，会出现此参数。可以根据需要选择“裸金属服务器”或“弹性云服务器”。 - 裸金属服务器：是一款兼具弹性云服务器和物理机性能的计算类服务器，为您和您的企业提供专属的云上物理服务器。 - 弹性云服务器：是一种可随时自助获取、可弹性伸缩的云服务器，可帮助您打造可靠、安全、灵活、高效的应用环境，确保服务持久稳定运行，提升运维效率。
计费模式	- 包年/包月是预付费模式，按订单的购买周期计费，适用于可预估资源使用周期的场景，价格比按需计费模式更优惠。 - Lite Server超节点仅支持“包年/包月”。 Lite Server普通节点（ECS或BMS）的所有资源规格均支持“包年/包月”计费模式。
区域	不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。

参数名称	说明
可用区	可用区是同一服务区内，电力和网络互相独立的地理区域，一般是一个独立的物理机房，这样可以保证可用区的独立性。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迟较低，则建议您将资源创建在同一可用区内。 当“节点类型”选择“超节点”或“普通节点 > 弹性云服务器”时，支持“随机可用区”，即显示当前各个区域中可用的节点资源。“普通节点 > 裸金属服务器”不支持随机可用区。 边缘小站资源发放的详细内容请参见Lite Server管理CloudPond的NPU资源。

表 2-3 规格配置参数说明

参数名称	说明
CPU架构	资源类型的CPU架构，支持X86和ARM。 - X86：如果使用GPU资源选择X86。 - ARM：如果使用NPU资源则选择ARM。 请先选择CPU架构，再根据具体需求选择实例规格。具体规格有区域差异，以最终显示为准。已售罄的资源会呈灰色显示，不支持购买。 说明 如果界面无可选规格，请联系华为云技术支持申请开通。

表 2-4 操作系统配置参数说明

参数名称	说明
镜像	此处配置的是Lite Server服务器的操作系统镜像。 - 公共镜像 公共镜像对所有用户可见。所有用户可以根据镜像ID进行只读使用。 ModelArts服务提供了多个公共的操作系统镜像，支持多种操作系统，并且在镜像中内置了AI场景相关驱动和软件，为用户提供了一个完整的AI开发环境，方便用户直接进行开发和训练，而无需额外配置。 当前支持的公共操作系统镜像请参见Lite Server算力资源和镜像版本配套关系。 系统正常运行需要配套相关硬件驱动，系统内核的变化可能会导致部分硬件驱动不兼容，影响系统正常运行，建议不要直接登录到操作系统自行升级系统内核或者硬件驱动。如果需要切换或重置操作系统镜像，请参见切换或重置Lite Server服务器操作系统。 - 私有镜像 仅镜像创建者可以使用，其他用户无法访问。选择通过私有镜像配置Lite Server操作系统，可以节省您重复配置服务器的时间。私有镜像需要在镜像服务IMS中提前创建，详情请参见创建私有镜像。

表 2-5 存储配置参数说明

参数名称	说明
存储配置	存储配置参数作用于每一个普通节点，实际存储配置=单个节点的存储配置*购买的节点数量。
节点系统盘类型	系统盘和规格有关，选择支持挂载的实例规格才会显示此参数。节点系统盘用于存储服务器的操作系统，创建Lite Server时自带系统盘，且系统盘自动初始化。 此处支持选择“节点系统盘类型”，并设置“大小”。系统盘大小取值范围在100GiB和1024GiB之间。 也可以在Lite Server创建完成后再进行系统盘的扩容，当前仅支持超节点的系统盘扩缩容，具体操作请参见Lite Server超节点扩容和缩容。 系统盘会自动挂载到每个计算节点上。

参数名称	说明
节点数据盘类型（可选）	单击“增加数据盘”，可以在创建Lite Server时挂载云上EVS数据盘。暂不支持挂载本地磁盘。 此处支持选择“节点数据盘类型”，并设置“大小”和数据盘“数量”。 数据盘大小取值范围在100GiB和32768GiB之间。 BMS或ECS类型的机器，数据盘个数上限是59块。超节点类型的机器，数据盘个数上限是8块。 也可以在Lite Server创建完成后再进行数据盘的扩容。 数据盘会自动挂载到每个计算节点上。

表 2-6 网络配置参数说明

参数名称	说明
虚拟私有云	虚拟私有云（Virtual Private Cloud，简称VPC）用以确保Lite Server资源的安全性、隔离性和网络的灵活性。 在下拉框中选择Lite Server对应的VPC，建议选择VPC时与其它云服务保持一致，便于网络互通。 下拉框中无可用VPC时，单击右侧的“新建虚拟私有云”，会在当前页面右侧弹出“创建虚拟私有云”窗口，根据提示创建VPC。 创建虚拟私有云需要登录管理员账号，IP地址段请根据现网情况合理规划。
子网	选择该VPC下的一个子网。 下拉框中无子网可选时，单击右侧的“新建子网”，会在当前页面右侧弹出“新建子网”窗口，根据提示创建一个子网。 Lite Server不支持手动分配子网IP，仅支持自动分配。
安全组	安全组是一个逻辑上的分组，为同一个VPC内具有相同安全保护需求并相互信任的Lite Server提供访问策略。 下拉框中无安全组可用时，单击右侧的“新建安全组”，会在当前页面右侧弹出“使用预设规则创建安全组”窗口，根据提示创建一个安全组。 请确保所选安全组已放通22端口（Linux SSH登录），3389端口（Windows远程登录）和ICMP协议（Ping），其他与业务无关端口或IP请关闭。 建议将安全组入方向规则中高危端口的源地址设置为已知IP地址、安全组或IP地址组，避免因网络入侵出现业务中断、数据泄露或数据勒索等严重后果。

参数名称	说明
IPv6网络	如果当前网络配置的子网、规格、镜像都支持IPv6，则会显示该参数，打开后可启用IPv6功能。 请确保您的子网已开启IPv6功能，如果未开启请参考为虚拟私有云创建新的子网。 不同规格、镜像对IPv6支持的情况不同，如果不支持则不会显示IPv6网络参数，请以控制台实际显示为准。
RoCE网络	当“节点类型”是“普通节点”时，会出现此参数。 当使用A系列GPU资源或Snt9b、Snt9b23资源进行分布式训练时，为了将硬件上的RoCE网卡使用起来，需要配置RoCE网络。 该参数与所选规格有关，如果未选中规格或规格不支持RoCE网络，则不显示。 如果规格支持RoCE网络但未创建过，单击“新建RoCE网络”即可完成创建。 如果规格支持RoCE网络且已创建过RoCE网络，直接选择已有RoCE网络即可（不支持重复创建）。
超节点网络	当“节点类型”是“超节点”时，会出现此参数，单击右侧的“添加超节点网络”可以创建超节点网络。 超节点网络是支持分布式场景的必备条件。

表 2-7 节点管理参数说明

参数名称	说明
服务器名称	Lite Server的机器名称。只能包含数字、大小写字母、下划线和中划线，长度不能超过64位且不能为空。 注意 订单中的服务器名称会一直保持此处下单购买时设置的名称。后期修改服务器名称后，不会在订单中同步更新。

参数名称	说明
登录凭证	“密钥对”方式创建的Lite Server节点安全性更高，建议选择“密钥对”方式。如果您习惯使用“密码”方式，请增强密码的复杂度，保证密码符合要求，防止被恶意攻击。 • 密钥对 指使用密钥对作为登录Lite Server节点的鉴权方式。您可以选择使用已有的密钥对，或者单击“新建密钥对”创建新的密钥。 如果选择使用已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录Lite Server节点。 • 密码 指使用设置初始密码方式作为Lite Server节点的鉴权方式，此时，您可以通过用户名密码方式登录Lite Server节点。 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。密码复杂度需满足以下要求： - 长度为8至26个字符。 - 至少包含大写字母、小写字母、数字及特殊符号(!@\$%^&_+=+[]{};./?)中的3种 - 不能与用户名或倒序的用户名相同。 - 不能包含root或administrator及其逆序。
企业项目	该参数针对企业用户使用，只有开通了企业项目的客户，或者权限为企业主账号的客户才可见。如需使用该功能，请联系您的客户经理申请开通。 企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理，默认项目为default。 请从下拉列表中选择所在的企业项目。更多关于企业项目的信息，请参见《企业管理用户指南》。 注意 已经完成购买的Lite Server，不支持再修改企业项目，订单中暂不支持同步企业项目信息。

表 2-8 高级配置参数说明

参数名称	说明
CES主机监控委托	勾选后表示开启，将一键配置CES主机监控委托。委托CES对Lite Server的CPU、内存、网络、磁盘、进程等指标进行监控，监控指标间隔是1分钟。详细监控指标信息请参见使用CES监控Lite Server NPU资源章节。

参数名称	说明
节点任务中枢	部分公共镜像预置了NodeTaskHub插件，选择相应镜像时，此处会显示此参数。 勾选后表示开启，系统会自动安装NodeTaskHub插件，用于支持任务中心下发软件升级、压测、故障诊断等任务。详细介绍请参见 安装Lite Server AI插件 。
实例自定义数据注入	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Lite Server节点： - 通过脚本简化Lite Server节点配置 - 通过脚本初始化系统 - 已有脚本，在创建Lite Server节点时一并上传至服务器 - 其他可以使用脚本完成的操作 当前支持“以文本形式”和“以文件形式”，使用方法可参考 BMS实例自定义数据注入 或 ECS实例自定义数据注入 。

表 2-9 购买配置参数说明

参数名称	说明
购买时长	选择资源购买时长，并根据需要勾选“自动续费”。
购买数量	支持同时购买多台机器，输入值必须在1到10之间。 如果有多台机器资源，会生成对应多笔订单，需逐一支付每笔订单，不可合并支付。 如果您购买48台超节点，请结合您的业务场景，自行预留部分机器作为备机，确保机器出现故障时，及时切换到备机。

4. 在当前购买页面的左下角查看配置费用，并单击“立即购买”，完成实例的创建，随后进入付款界面，支付对应资源的订单。

配置费用中会显示当前资源的费用构成。如果有优惠，可以通过“优惠详情”查看详细内容，配置费用显示的是最终优惠后的费用。实际扣费请在账单中查看。

 说明

- 如果有多台机器资源，会生成对应多笔订单，需逐一支付每笔订单，不可合并支付。
5. 支付完成后，由于Lite Server资源创建约20~60分钟，请耐心等待。如果资源创建失败，请参考[资源购买失败处理](#)。

图 2-8 资源创建成功



资源购买失败处理

ModelArts的轻量算力节点 (Lite Server)创建失败，可能由多种原因导致，以下给出了几类可能原因，方便快速排查和定位解决。

- 资源不足：跳转到BMS或ECS页面，查看要购买的规格是否售罄，如果该规格售罄，说明无该规格资源，需要联系客户经理获取到资源后再进行购买。
- 配额不足：查看账户的资源配额是否满足，如果该账号下资源配额，包括核心数、RAM等，如果未满足也会导致创建失败，需要申请配额后再进行购买。
- BMS或ECS机器内部错误：查看BMS或ECS界面，创建失败出现内部错误，该问题需要提工单给BMS或ECS进行进一步定位失败原因并解决。

3 Lite Server 资源开通（旧版页面）

场景描述

本章节主要介绍如何在ModelArts控制台上购买Lite Server算力资源，及购买前的准备工作。

用户先完成资源配额提升、配置基础权限、设置ModelArts委托授权等准备工作。在购买资源时，用户创建实例并支付订单，支付完成后等待约20~60分钟，资源创建成功后即可配置弹性公网IP进行访问，开展相关AI开发工作。

约束限制

ModelArts Lite Server算力资源当前仅支持“包年/包月”计费模式。

资源开通流程

图 3-1 Server 资源开通流程图

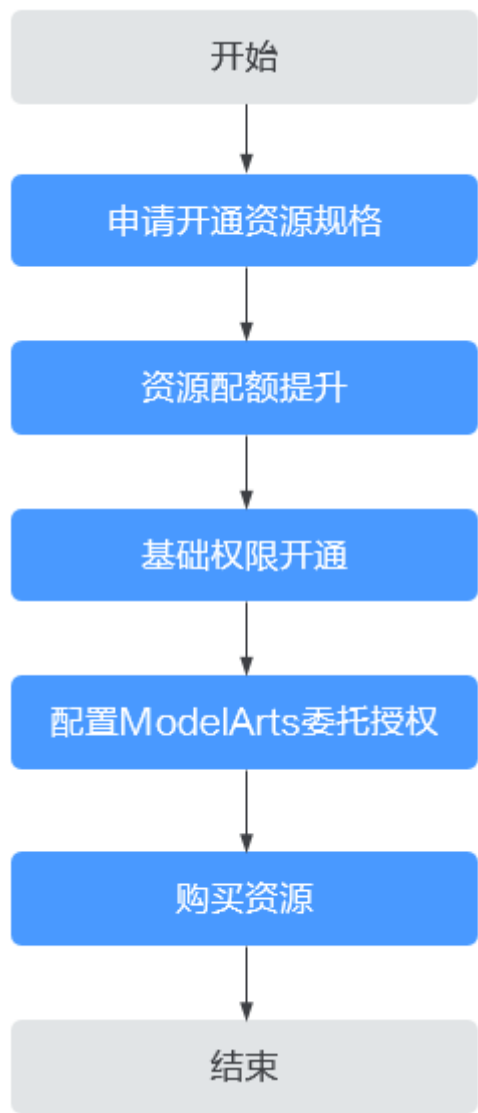


表 3-1 Server 资源开通流程

阶段	任务
准备工作	1、申请开通资源规格。
	2、资源配额提升。
	3、基础权限开通。
	4、配置ModelArts委托授权。
购买Server资源	5、在ModelArts控制台上购买轻量算力节点 (Lite Server)资源。

步骤 1：申请开通资源规格

请联系客户经理确认Server资源方案、申请要开通资源的规格，如果无客户经理可提交工单。

步骤 2：提升资源配额

由于Server所需资源可能会超出云服务默认提供的资源（如ECS、EIP、SFS、内存大小、CPU核数），因此需要提升资源配额。

1. 登录[华为云管理控制台](#)。
2. 在顶部导航栏单击“资源 > 我的配额”，进入服务配额页面。
3. 单击右上角“申请扩大配额”，填写申请材料后提交工单。

说明

配额需大于需要开通的资源，且在购买开通前完成提升，否则会导致资源开通失败。

步骤 3：开通基础权限

开通基础权限需要登录管理员账号，为子用户账号开通Server功能所需的基础权限，包括ModelArts FullAccess、BMS FullAccess、ECS FullAccess、VPC FullAccess、VPC Administrator、VPCEndpoint Administrator，即允许子用户账号同时可以使用这些云服务。

1. 登录[统一身份认证服务管理控制台](#)。
2. 单击目录左侧“用户组”，然后在页面右上角单击“创建用户组”。
3. 填写“用户组名称”并单击“确定”。
4. 在操作列单击“用户组管理”，将需要配置权限的用户加入用户组中。
5. 单击用户组名称，进入用户组详情页。
6. 在权限管理页签下，单击“授权”。

图 3-2 “配置权限”



7. 在搜索栏输入“ModelArts FullAccess”，并勾选“ModelArts FullAccess”。

图 3-3 ModelArts FullAccess



以相同的方式，依次添加：BMS FullAccess、ECS FullAccess、VPC FullAccess、VPC Administrator、VPCEndpoint Administrator。（Server Administrator、DNS Administrator为依赖策略，会自动被勾选）。

8. 单击“下一步”，授权范围方案选择“所有资源”。

9. 单击“确认”，完成基础权限开通。

步骤 4 在 ModelArts 上创建委托授权

ModelArts Lite Server在任务执行过程中需要访问用户的其他服务，典型的就是容器使用过程中需要到SWR服务拉取镜像。在这个过程中，就出现了ModelArts“代表”用户去访问其他云服务的情形。从安全角度出发，ModelArts代表用户访问任何云服务之前，均需要先获得用户的授权，而这个动作就是一个“委托”的过程。用户授权ModelArts代表自己访问特定的云服务，以完成其在ModelArts平台上执行的AI计算任务。

- 新建委托

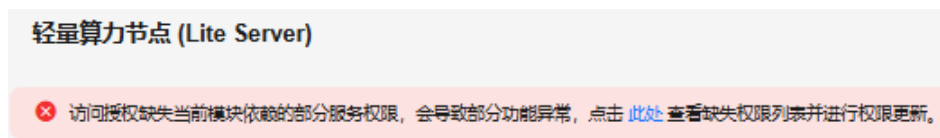
第一次使用ModelArts时需要创建委托授权，授权允许ModelArts代表用户去访问其他云服务。进入ModelArts控制台的“系统管理 > 权限管理”页面，单击“添加授权”，根据提示进行操作。

- 更新委托

如果之前给ModelArts创过委托授权，此处可以更新授权。

- a. 进入[ModelArts控制台](#)的“资源管理 > 轻量算力节点 (Lite Server)”页面，查看是否存在授权缺失的提示。

图 3-4 Lite Server 权限缺失提示



- b. 如果有授权缺失，根据提示，单击“此处”更新委托。根据提示选择“追加至已有授权”，单击“确定”，系统会提示权限更新成功。

图 3-5 追加授权

访问授权权限不足

访问授权缺失当前模块依赖的如下服务权限，如需继续使用请添加权限至访问授权。 [常见问题](#)

服务名称	使用模块
	Notebook 镜像管理 弹性节点 Server

添加方式

追加至已有授权

配置新授权

上述缺失权限追加至如下委托，所有访问授权配置为此委托的用户对应权限均会更新。

授权用户

授权内容

[委托]modelarts_

步骤 5：购买轻量算力节点 (Lite Server)资源

购买轻量算力节点 (Lite Server)资源的过程即创建资源过程。

1.

登录[ModelArts管理控制台](#)。
2.

在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入资源列表。
3.

单击右上角的“购买轻量算力节点”，进入“购买轻量算力节点”页面，在该页面填写相关参数信息。

图 3-6 购买轻量算力节点时的基础配置

基础配置

资源类型

裸金属服务器

弹性云服务器

超节点服务器

?

提供安全稳定、可随时自助获取、弹性伸缩的计算服务，灵活计费，极简运维，极大降低企业成本。

计费模式

包年/包月

按需计费

?

区域

西南-贵阳一

▼

不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。

可用区

可用区4

可用区5

可用区1

可用区6

?

表 3-2 基础配置参数说明

参数名称	说明
资源类型	裸金属服务器：是一款兼具弹性云服务器和物理机性能的计算类服务器，为您和您的企业提供专属的云上物理服务器。 弹性云服务器：是一种可随时自助获取、可弹性伸缩的云服务器，可帮助您打造可靠、安全、灵活、高效的应用环境，确保服务持久稳定运行，提升运维效率。
计费模式	选择“包年/包月”。 包年/包月是预付费模式，按订单的购买周期计费，适用于可预估资源使用周期的场景，价格比按需计费模式更优惠。 暂不支持按需计费。
区域	不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。资源购买完成后，您可在控制台左上角切换区域，查看对应的资源。

参数名称	说明
可用区	可用区是同一服务区内，电力和网络互相独立的地理区域，一般是一个独立的物理机房，这样可以保证可用区的独立性。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迟较低，则建议您将资源创建在同一可用区内。 边缘小站资源发放的详细内容请参见Lite Server管理CloudPond的NPU资源。

表 3-3 资源配置参数说明

参数名称	说明
服务器名称	Server的机器名称。只能包含数字、大小写字母、下划线和中划线，长度不能超过64位且不能为空。 注意 订单中的服务器名称会一直保持此处下单购买时设置的名称。后期修改服务器名称后，不会在订单中同步更新。
CPU架构	资源类型的CPU架构，支持X86和ARM。 - X86：如果使用GPU资源选择X86。 - ARM：如果使用NPU资源则选择ARM。 请先选择CPU架构，再根据具体需求选择实例规格。具体规格有区域差异，以最终显示为准。已售罄的资源会呈灰色显示，不支持购买。 说明 如果界面无可选规格，请联系华为云技术支持申请开通。
系统盘	系统盘和规格有关，选择支持挂载的实例规格才会显示此参数。系统盘用于存储服务器的操作系统，创建Lite Server时自带系统盘，且系统盘自动初始化。 此处支持选择系统盘的类型，并设置大小。系统盘大小取值范围在100GiB和1024GiB之间。 也可以在Server创建完成后在云服务器侧实现系统盘的扩容。系统盘会自动挂载到每个计算节点上。

表 3-4 镜像配置参数说明

参数名称	说明
镜像	- 公共镜像 公共镜像对所有用户可见。所有用户可以根据镜像ID进行只读使用。 ModelArts服务提供了多个公共镜像，支持多种操作系统，并且内置了AI场景相关驱动和软件，为用户提供了一个完整的AI开发环境，方便用户直接进行开发和训练，而无需额外配置。 当前支持的公共镜像请参考Lite Server算力资源和镜像版本配套关系。 - 私有镜像 仅镜像创建者可以使用，其他用户无法访问。选择私有镜像创建，可以节省您重复配置服务器的时间。

表 3-5 网络配置参数说明

参数名称	说明
虚拟私有云	虚拟私有云（Virtual Private Cloud，简称VPC）用以确保Server资源的安全性、隔离性和网络的灵活性。 在下拉框中选择Server对应的VPC，建议选择VPC时与其它云服务保持一致，便于网络互通。 下拉框中无可用VPC时，单击右侧的“新建虚拟私有云”创建一个VPC。创建虚拟私有云需要登录管理员账号，IP地址段请根据现网情况合理规划。
子网	选择该VPC下的一个子网。 下拉框中无子网可选时，单击右侧的“新建子网”创建一个子网。
安全组	安全组是一个逻辑上的分组，为同一个VPC内具有相同安全保护需求并相互信任的Server提供访问策略。 下拉框中无安全组可用时，单击右侧的“新建安全组”创建一个安全组。
IPv6网络	如果当前网络配置的子网、规格、镜像都支持IPv6，则会显示该参数，打开后可启用IPv6功能。 请确保您的子网已开启IPv6功能，如果未开启请参考为虚拟私有云创建新的子网。 不同规格、镜像对IPv6支持的情况不同，如果不支持则不会显示IPv6网络参数，请以控制台实际显示为准。

参数名称	说明
RoCE网络	当使用A系列GPU或昇腾Snt9b、Snt9b23资源进行分布式训练时，为了将硬件上的RoCE网卡使用起来，需要配置RoCE网络。该参数与所选规格有关，如果未选中规格或规格不支持RoCE网络，则不显示。 如果规格支持RoCE网络但未创建过，单击“新建RoCE网络”即可完成创建。 如果规格支持RoCE网络且已创建过RoCE网络，直接选择已有RoCE网络即可（不支持重复创建）。

表 3-6 管理参数说明

参数名称	说明
登录凭证	“密钥对”方式创建的Server节点安全性更高，建议选择“密钥对”方式。如果您习惯使用“密码”方式，请增强密码的复杂度，保证密码符合要求，防止被恶意攻击。 ● 密钥对 指使用密钥对作为登录Server节点的鉴权方式。您可以选择使用已有的密钥对，或者单击“新建密钥对”创建新的密钥。 说明 如果选择使用已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录Server节点。 ● 密码 指使用设置初始密码方式作为Server节点的鉴权方式，此时，您可以通过用户名密码方式登录Server节点。 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。密码复杂度需满足以下要求： - 长度为8至26个字符。 - 至少包含大写字母、小写字母、数字及特殊符号(!@\$%^_+=+[{ } : , . / ?)中的3种 - 不能与用户名或倒序的用户名相同。 - 不能包含root或administrator及其逆序。

表 3-7 高级配置参数说明

参数名称	说明
企业项目	该参数针对企业用户使用，只有开通了企业项目的客户，或者权限为企业主账号的客户才可见。如需使用该功能，请联系您的客户经理申请开通。 企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理，默认项目为default。 请从下拉列表中选择所在的企业项目。更多关于企业项目的信息，请参见《企业管理用户指南》。 注意 已经完成购买的Server，不支持再修改企业项目，订单中暂不支持同步企业项目信息。

表 3-8 购买配置参数说明

参数名称	说明
购买时长	选择资源购买时长，并根据需要勾选“自动续费”。
购买数量	支持同时购买多台机器，输入值必须在1到10之间。 如果有多台机器资源，会生成对应多笔订单，需逐一支付每笔订单，不可合并支付。

4. 在当前购买页面的左下角查看配置费用，并单击“立即创建”，完成实例的创建，随后进入付款界面，支付对应资源的订单。
- 配置费用中会显示当前资源的费用构成。如果有优惠，可以通过“优惠详情”查看详细内容，配置费用显示的是最终优惠后的费用。实际扣费请在账单中查看。

 说明

- 如果有多台机器资源，会生成对应多笔订单，需逐一支付每笔订单，不可合并支付。
5. 支付完成后，由于Server资源创建约20~60分钟，请耐心等待。如果资源创建失败，请参考[资源购买失败处理](#)。

图 3-7 资源创建成功

名称/ID 	监控	状态	规格
 31d1ebf7-ab...		 运行中	...

 说明

当容器需要提供服务给多个用户，或者多个用户共享使用该容器时，应限制容器访问OpenStack的管理地址（169.254.169.254），以防止容器获取宿主机的元数据。具体操作请参见[禁止容器获取宿主机元数据](#)。

资源购买失败处理

ModelArts的轻量算力节点 (Lite Server)创建失败，可能由多种原因导致，以下给出了几类可能原因，方便快速排查和定位解决。

- 资源不足：跳转到BMS或ECS页面，查看要购买的规格是否售罄，如果该规格售罄，说明无该规格资源，需要联系客户经理获取到资源后再进行购买。
- 配额不足：查看账户的资源配额是否满足，如果该账号下资源配额，包括核心数、RAM等，如果未满足也会导致创建失败，需要申请配额后再进行购买。
- BMS或ECS机器内部错误：查看BMS或ECS界面，创建失败出现内部错误，该问题需要提工单给BMS或ECS进行进一步定位失败原因并解决。

4 Lite Server 资源配置

4.1 Lite Server 资源配置流程

在开通Lite Server资源后，需要完成相关配置才能使用，配置流程如下图所示。

图 4-1 Lite Server 资源配置流程图

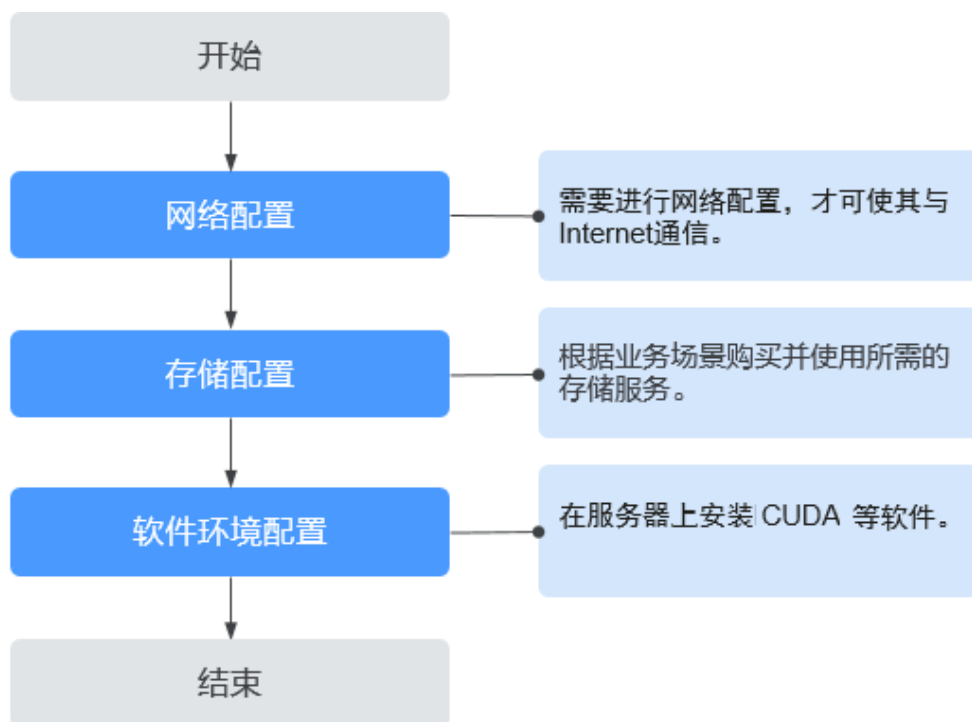


表 4-1 Lite Server 资源配置流程

配置顺序	配置任务	场景说明
1	配置Lite Server网络	Lite Server资源开通后，需要进行网络配置，才可使其与Internet通信。在后续配置存储和软件环境时需要Lite Server服务器能够访问网络，因此需要先完成网络配置。
2	配置Lite Server存储	Lite Server资源需要挂载数据盘用于存储数据文件，当前支持SFS、OBS、EVS三种云存储服务，提供了多种场景下的存储解决方案。
3	配置Lite Server软件环境（可选）	不同镜像中预安装的软件不同，您通过 Lite Server 算力资源和镜像版本配套关系 章节查看已安装的软件。当Lite Server服务器中预装的软件无法满足业务需求时，您可在Lite Server服务器中配置所需要的软件环境。

4.2 配置 Lite Server 网络

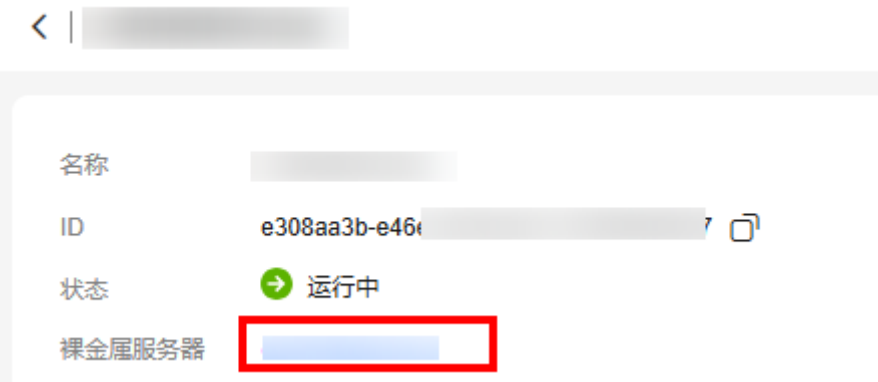
Lite Server创建后，需要进行网络配置，才可使其与Internet通信，本章节介绍网络配置步骤。网络配置主要分为以下两个场景：

- **单个弹性公网IP用于单个Lite Server节点**：为单台Lite Server服务器绑定一个弹性公网IP，该Lite Server服务器独享网络资源。
- **单个弹性公网IP用于多个Lite Server服务器**：一个VPC配置一个EIP（弹性公网IP），通过NAT网关配置进行EIP资源共享，实现该VPC下的所有Lite Server服务器均可以通过该EIP进行公网访问，Lite Server服务器共享网络资源。

单个弹性公网 IP 用于单个 Lite Server 节点

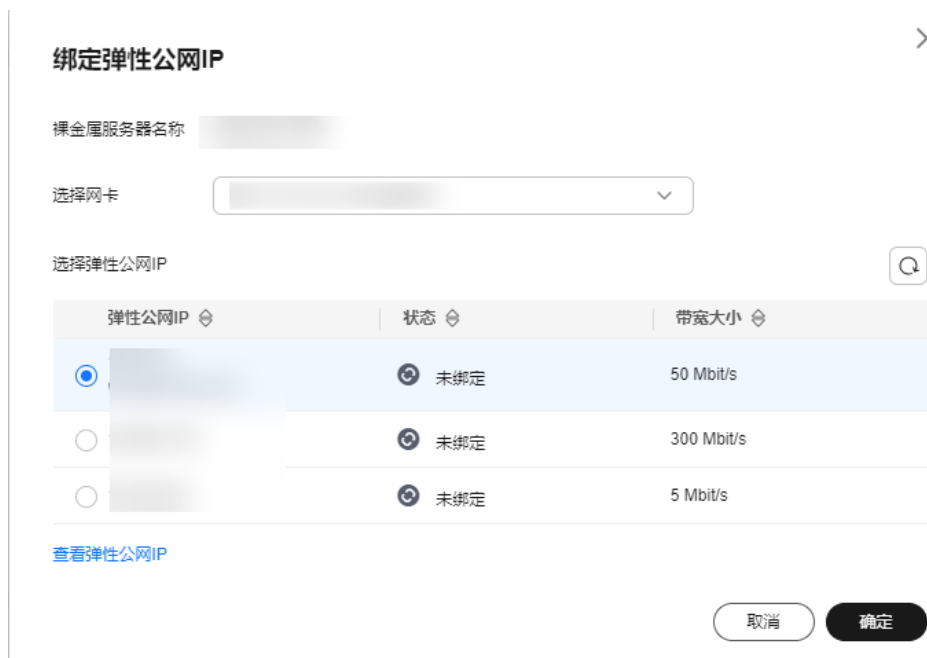
1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“轻量算力节点 (Lite Server)”的“普通节点”列表页面。
3. 单击目标Lite Server节点名称，进入该Lite Server服务器的详情页面。

图 4-2 裸金属服务器



4. 单击“弹性公网IP”页签，然后单击“绑定弹性公网IP”。弹出“绑定弹性公网IP”对话框。选择要绑定的弹性公网IP，单击“确定”，完成绑定。

图 4-3 绑定弹性公网 IP



说明

一个网卡只能绑定一个弹性公网IP。

单个弹性公网 IP 用于多个 Lite Server 服务器

说明

所有 Lite Server 资源必须位于同一个 VPC，并且该 VPC 没有 NAT 网关以及默认路由。

1. 购买弹性公网IP。
 - a. 登录[华为云管理控制台](#)。
 - b. 在左侧服务列表中，单击“网络 > 弹性公网IP EIP”，进入弹性公网IP页面。
 - c. 在弹性公网IP页面右上角单击“购买弹性公网IP”。
 - d. 参数配置可使用默认值，单击“立即购买”。
 - e. 在产品配置信息确认页面，再次核对弹性公网IP信息，阅读并勾选“弹性公网IP服务声明”。
 - 选择按需计费的弹性公网IP时，单击“提交”。
 - 选择包年/包月计费的弹性公网IP时，单击“去支付”。进入订单支付页面，确认订单信息，单击“确认付款”。
2. 购买公网NAT网关。
 - a. 登录[华为云管理控制台](#)。
 - b. 在左侧服务列表中，单击“网络 > NAT网关 NAT”，进入公网NAT网关页面。

- c. 在公网NAT网关页面右上角单击“购买公网NAT网关”。
- d. 选择Lite Server所使用“虚拟私有云”和“子网”，计费模式根据实际需求选择。其余参数配置可使用默认值，单击“立即购买”。

图 4-4 购买公网 NAT 网关

- e. 在产品配置信息确认页面，再次确认NAT网关信息。
 - 选择按需计费的NAT网关时，单击“提交”。
 - 选择包年/包月计费的NAT网关时，单击“去支付”。
进入订单支付页面，确认订单信息，单击“确认付款”。

说明

虚拟私有云和子网与Lite Server资源的网络保持一致。

- 3. **配置SNAT规则。**

SNAT功能通过绑定弹性公网IP，实现私有IP向公有IP的转换，可实现VPC内跨可用区的多个云主机共享弹性公网IP、安全高效地访问互联网。

 - a. 公网NAT网关页面，单击创建的NAT网关名称，进入NAT网关详情页。
 - b. 在SNAT规则页签下，单击“添加SNAT规则”。
 - c. 在弹出的“添加SNAT规则页面”，配置SNAT规则：
 - 使用场景：选择“虚拟私有云”。
 - 子网：选择“使用已有”，选择子网。
 - 弹性公网IP：勾选创建的弹性公网IP。
 - d. 单击“确定”。
- 4. **配置DNAT规则。**

通过添加DNAT规则，以映射方式为VPC内的Lite Server提供SSH访问服务，一个Lite Server的一个端口对应一条DNAT规则，一个端口只能映射到一个EIP，不能映射到多个EIP。

- a. 在DNAT规则页签下，单击“添加DNAT规则”。
- b. 在弹出的“添加DNAT规则页面”，配置DNAT规则：
 - 使用场景：选择“虚拟私有云”。
 - 端口类型：选择“具体端口”。
 - 支持协议：选择“TCP”。
 - 公网IP类型：选择已创建的弹性公网IP。
 - 公网端口：建议选择区间为20000-30000，保证该端口号不冲突。
 - 实例类型：单击“服务器”，选择Server服务器。
 - 网卡：选择服务器网卡。
 - 私网端口：端口号22。
- c. 单击“确定”。

4.3 配置 Lite Server 存储

Lite Server服务器支持SFS、OBS、EVS三种云存储服务，提供了多种场景下的存储解决方案，主要区别如下表所示。如果需要对本地图盘进行配置，请参考[磁盘合并挂载](#)。

表 4-2 SFS、OBS、EVS 服务对比

对比维度	弹性文件服务SFS	对象存储服务OBS	云硬盘EVS
概念	提供按需扩展的高性能文件存储，可为云上多个云服务器提供共享访问。弹性文件服务就类似Windows或Linux中的远程目录。	提供海量、安全、高可靠、低成本的数据存储能力，可供用户存储任意类型和大小的数据。	可以为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，可满足不同场景的业务需求。云硬盘就类似PC中的硬盘。
存储数据的逻辑	存放的是文件，会以文件和文件夹的层次结构来整理和呈现数据。	存放的是对象，可以直接存放文件，文件会自动产生对应的系统元数据，用户也可以自定义文件的元数据。	存放的是二进制数据，无法直接存放文件，如果需要存放文件，需要先格式化文件系统后使用。
访问方式	在Lite Server中通过网络协议挂载使用，支持NFS和CIFS的网络协议。需要指定网络地址进行访问，也可以将网络地址映射为本地目录后进行访问。	可以通过互联网或专线访问。需要指定桶地址进行访问，使用的是HTTP和HTTPS等传输协议。	只能在Lite Server中挂载使用，不能被操作系统应用直接访问，需要格式化成文件系统进行访问。

对比维度	弹性文件服务SFS	对象存储服务OBS	云硬盘EVS
使用场景	如高性能计算、媒体处理、文件共享和内容管理和Web服务等。 高性能计算：主要是高带宽的需求，用于共享文件存储，比如基因测序、图片渲染这些。	如大数据分析、静态网站托管、在线视频点播、基因测序和智能视频监控等。	如高性能计算、企业核心集群应用、企业应用系统和开发测试等。 高性能计算：主要是高速率、高IOPS的需求，用于作为高性能存储，比如工业设计、能源勘探这些。
容量	PB级别	EB级别	TB级别
时延	3~10ms	10ms	亚毫秒级
IOPS/TPS	单文件系统 10K	千万级	单盘 128K
带宽	GB/s级别	TB/s级别	MB/s级别
是否支持数据共享	是	是	是
是否支持远程访问	是	是	否
是否支持在线编辑	是	否	是

使用弹性文件服务 SFS 作为存储

如果使用SFS服务作为存储方案，推荐使用SFS Turbo文件系统。SFS Turbo提供按需扩展的高性能文件存储，还具备高可靠和高可用的特点，支持根据业务需要弹性扩容，且性能随容量增加而提升，可广泛应用于多种业务场景。

1. 在SFS服务控制台上创建文件系统，具体步骤请参考[创建SFS Turbo文件系统](#)。同一区域不同可用区之间文件系统与云服务器互通，因此保证SFS Turbo与Server服务器在同一区域即可。
2. 当创建文件系统后，您需要将该文件系统挂载至Server服务器上，具体步骤请参考[挂载NFS协议类型文件系统到云服务器（Linux）](#)。
3. 为避免已挂载文件系统的云服务器重启后，挂载信息丢失，您可以在云服务器设置重启时进行自动挂载，具体步骤请参考[服务器重启后自动挂载指南](#)。

使用对象存储服务 OBS 作为存储

如果使用OBS服务作为存储方案，推荐使用“并行文件系统+obsutil”的方式，并行文件系统是OBS服务提供的一种经过优化的高性能文件语义系统，提供毫秒级别访问时延，TB/s级别带宽和百万级别的IOPS。obsutil是一款用于访问管理对象存储服务（Object Storage Service，OBS）的命令行工具，您可以使用该工具对OBS进行常用的配置管理操作，如创建桶、上传文件/文件夹、下载文件/文件夹、删除文件/文件夹等。对于熟悉命令行程的用户，obsutil在执行批量处理、自动化任务场景能为您带来更优体验。

1. 在OBS服务控制台上创建并行文件系统，具体步骤请参考[创建并行文件系统](#)。
2. 针对您的操作系统，下载对应版本的obsutil至Lite Server服务器，并完成安装，具体步骤请参考[下载和安装obsutil](#)。
3. 使用obsutil之前，您需要配置obsutil与OBS的对接信息，包括OBS终端节点地址（Endpoint）和访问密钥（AK和SK）。获得OBS的认证后，才能使用obsutil执行OBS桶和对象的相关操作，具体步骤请参考[初始化配置](#)。
4. 配置完成后，您可以通过命令行的方式在Server服务器中对OBS的文件进行上传下载等操作，关于命令行介绍请参考[命令行结构](#)。

使用云硬盘 EVS 作为存储

[创建Lite Server资源](#)时可以挂载EVS类型的数据盘。如果Lite Server运行一段时间后，数据盘不够用，也可以再挂载数据盘，也称为后挂载方式。后挂载方式需要先购买EVS数据盘，再挂载到Lite Server上。以下步骤主要介绍后挂载云硬盘的方式。

1. 购买EVS数据盘。在EVS服务控制台上购买磁盘，选择Lite Server节点所在的可用区，挂载方式选择“暂不挂载”，计费模式选择“包年/包月”或者“按需计费”均可以，磁盘大小根据自身需求进行选择购买。
 - Lite Server资源类型为裸金属服务器BMS时，支持的云硬盘类型以当前可用区EVS控制台显示为准，且云硬盘高级配置必须开启SCSI。SCSI云硬盘允许云服务器操作系统直接访问底层存储介质并将SCSI指令传输到云硬盘。
 - Lite Server资源类型为弹性云服务器ECS时，支持的云硬盘类型以当前可用区EVS控制台显示为准。支持VBD和SCSI两种云硬盘模式，即开启或不开启SCSI均可以。

更多EVS购买参数介绍可参考[购买云硬盘](#)。

图 4-5 购买磁盘



说明

- 由于产品特性设计，暂不支持在购买EVS云硬盘时立即挂载到云服务器，此时网页界面会提示“该包年/包月云服务器还未同步到运营系统，请休息片刻再重试。您可以到费用中心 > 续费管理页面确认该云服务器是否已同步到运营系统”，挂载方式选择暂不挂载即可。
2. 在完成EVS数据盘购买后，可以将EVS云硬盘挂载到已有的Lite Server节点中。用户可以根据使用习惯选择磁盘的挂载方式。

Lite Server的服务器类型是裸金属服务器或弹性云服务器时，均支持在Lite Server详情页挂载磁盘。

在Lite Server详情页挂载磁盘：

在ModelArts控制台的“轻量算力节点 (Lite Server)”页面，单击具体的Lite Server名称，进入Lite Server的详情页挂载磁盘，在“挂载磁盘”窗口选择相应的EVS数据盘，设置挂载点进行挂载。

卸载磁盘时，可以在Lite Server详情页面单击“卸载”，卸载相应的数据盘。系统盘不支持卸载。

 说明

在退订裸金属服务器时，挂载的EVS数据盘不会自动删除。用户可根据自身需求，将其挂载在其他裸金属服务器上或者进行手动删除。

Lite Server的服务器类型是裸金属服务器时，也支持在BMS详情页面挂载磁盘，用户可以根据使用习惯选择。具体操作请参见[为BMS挂载磁盘](#)。

Lite Server的服务器类型是弹性云服务器时，也支持在ECS详情页面挂载磁盘，用户可以根据使用习惯选择。具体操作请参见[为ECS挂载磁盘](#)。

3. 挂载EVS数据盘后，需要对磁盘进行初始化操作，具体参见[初始化EVS磁盘](#)。

4.4 配置 Lite Server 软件环境（可选）

4.4.1 NPU 服务器上配置 Lite Server 资源软件环境

场景说明

本文旨在指导如何在基于NPU资源的Lite Server服务器上，进行磁盘合并挂载、安装Docker等环境配置。具体配置项如[表4-3](#)所示。

当前指导中很多配置在最新发放的Lite Server服务器中已经预置，无需用户再手动配置，用户在操作中如发现某个步骤已有预置配置可直接跳过该步骤。

表 4-3 物理机环境配置项

配置项	适用范围
服务器SSH连接超时参数	适用所有机型
磁盘合并挂载	适用所有机型
安装驱动和固件	仅适用于NPU系列资源
安装Docker环境	适用所有机型
安装pip源	适用所有机型
RoCE网络测试	仅适用于NPU系列中的Snt9b资源
容器化个人调测环境搭建	仅适用于NPU系列资源

配置注意事项

在配置前请注意如下事项：

- 首次装机时需要配置存储、固件、驱动、网络访问等基础内容，这部分配置尽量稳定减少变化。
- 裸机上的开发形式建议开发者启动独立的Docker容器作为个人开发环境。Snt9b的裸机包含8卡算力资源，一般来说多人可以共用这个裸机完成开发与调测工作。

多人使用为了避免冲突，建议各自在自己的Docker容器中进行独立开发，并提前规划好每个人使用的具体卡号，避免相互影响。

- ModelArts提供了标准化基础容器镜像，在容器镜像中已经预置了基础MindSpore或PyTorch框架和开发调测工具链，推荐用户直接使用该镜像，用户也可以使用自己的业务镜像或AscendHub提供的镜像。如果镜像中预置的软件版本不是您期望的版本，可以自行安装替换。
- 开发形式推荐通过容器中暴露的SSH端口以远程开发的模式(VSCode SSH Remote、Xshell)连接到容器中进行开发，可以在容器中挂载宿主机的个人存储目录，用于存放代码和数据。

服务器 SSH 连接超时参数

1. SSH登录到Lite Server服务器后，查看机器配置的超时参数。
`echo $TMOUT`
2. 如果该值为300，则代表默认空闲等待5分钟后会断开连接，可以增大该参数延长空闲等待时间；如果该值为0可跳过当前步骤。修改方法如下：
`vim /etc/profile`
在文件最后修改TMOUT值，由300改为0，0表示不会空闲断开
`export TMOUT=0`
3. 执行如下命令使其在当前terminal生效。
`TMOUT=0`

`export TMOUT=0`这个命令在SSH连接Linux服务器时的作用是设置会话的空闲超时时间为0，意味着不会因为空闲而自动断开连接。默认情况下，SSH连接可能会在一段时间没有操作后自动断开，这是为了安全考虑。但是，如果您正在进行需要长时间保持连接的任务，可以使用这个命令来防止连接因为空闲而断开。

您可以在当前的终端会话中直接执行`TMOUT=0`使设置立即生效，或者将`export TMOUT=0`添加到`/etc/profile`文件中，以确保所有用户的新会话都不会因为空闲而断开。

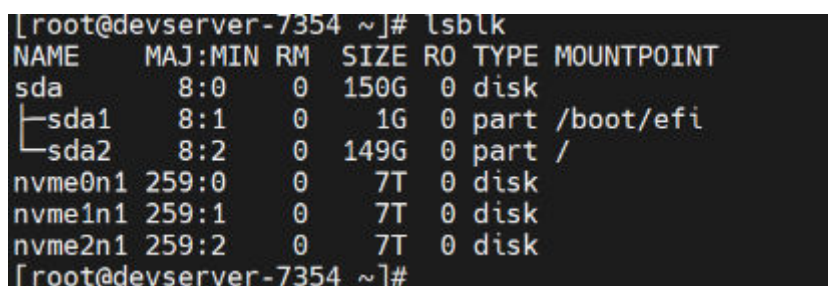
但是在生产环境或多人使用的公共服务器上，不建议设置`TMOUT=0`，关闭自动注销功能会带来一定的安全风险。

磁盘合并挂载

开通Lite Server资源后，服务器上可能会有多个未挂载的nvme磁盘。因此在首次配置环境前，需要完成磁盘合并挂载。此操作需要放在最开始完成，避免使用一段时间后再挂载会冲掉用户已存储的内容。

1. 首先通过“lsblk”查看是否有3个7T的磁盘未挂载，如下图所示nvme0n1、nvme1n1、nvme2n1为未挂载。

图 4-6 磁盘未挂载



```
[root@devserver-7354 ~]# lsblk
NAME        MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda          8:0    0   150G  0 disk
├─sda1       8:1    0    1G  0 part /boot/efi
└─sda2       8:2    0   149G  0 part /
nvme0n1     259:0    0    7T   0 disk
nvme1n1     259:1    0    7T   0 disk
nvme2n1     259:2    0    7T   0 disk
[root@devserver-7354 ~]#
```

2. 如下图所示，每个盘后已有MOUNTPOINT，则代表已经执行过挂载操作，可跳过此章节，只用直接在/home目录下创建自己的个人开发目录即可。

图 4-7 磁盘已挂载

```
[root@devserver-7354 ~]# lsblk
NAME                                MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda                                  8:0      0  150G  0 disk
├─sda1                              8:1      0    1G  0 part /boot/efi
└─sda2                              8:2      0  149G  0 part /
nvme0n1                             259:0    0    7T  0 disk /home
nvme1n1                             259:1    0    7T  0 disk
├─nvme_group-docker_data            253:0    0   14T  0 lvm  /docker
nvme2n1                             259:2    0    7T  0 disk
└─nvme_group-docker_data            253:0    0   14T  0 lvm  /docker
```

执行自动化挂载脚本，将“/dev/nvme0n1”挂载在“/home”下供每个开发者创建自己的家目录，另两个合并挂载到“/docker”下供容器使用（如果不单独给“/docker”分配较大空间，当多人共用创建多个容器实例时容易将根目录占满）。

```
cd /root/tools/
sh create_disk_partitions.sh
```

配置完成后，执行“df -h”可以看到新挂载的磁盘信息。

图 4-8 查看新挂载的磁盘

```
[root@devserver-modelarts home]# df -h
Filesystem      Size  Used Avail Use% Mounted on
devtmpfs         756G   0  756G   0% /dev
tmpfs            756G   0  756G   0% /dev/shm
tmpfs            756G  28M  756G   1% /run
tmpfs            756G   0  756G   0% /sys/fs/cgroup
/dev/sda2        196G  2.4G  185G   2% /
tmpfs            756G  40K  756G   1% /tmp
/dev/sda1       1022M  8.3M  1014M   1% /boot/efi
/dev/mapper/nvme_group-docker_data 14T  121G   14T   1% /docker
/dev/nvme0n1     7.0T   50G   7.0T   1% /home
```

磁盘合并挂载后，即可在“/home”下创建自己的工作目录，以自己的名字命名。

安装驱动和固件

1. 首先检查npu-smi工具是否可以正常使用，该工具必须能正常使用才能继续后面的固件驱动安装。执行以下命令，完整输出图4 检查npu-smi工具内容则为正常。
- npu-smi info

如果命令未按照下图完整输出（比如命令报错或只输出了上半部分没有展示下面的进程信息），则需要先尝试恢复npu-smi工具（提交[工单](#)联系华为云技术支持），将npu-smi恢复后，再进行新版本的固件驱动安装。

图 4-9 检查 npu-smi 工具

```
[root@devserver-bms-fd775372-833351 ~]# npu-smi info
```

npu-smi 23.0.rc3		Version: 23.0.rc3				
NPU Chip	Name	Health Bus-Id	Power(W) AICore(%)	Temp(C) Memory-Usage(MB)	Hugepages-Usage(page) HBM-Usage(MB)	
0	910B2	OK	92.7	49	0	0
0		0000:C1:00.0	0	0 / 0	4152	65536
1	910B2	OK	87.0	52	0	0
0		0000:01:00.0	0	0 / 0	4152	65536
2	910B2	OK	94.5	53	0	0
0		0000:C2:00.0	0	0 / 0	4152	65536
3	910B2	OK	92.9	51	0	0
0		0000:02:00.0	0	0 / 0	4152	65536
4	910B2	OK	91.9	53	0	0
0		0000:81:00.0	0	0 / 0	4152	65536
5	910B2	OK	93.1	54	0	0
0		0000:41:00.0	0	0 / 0	4153	65536
6	910B2	OK	92.2	52	0	0
0		0000:82:00.0	0	0 / 0	4153	65536
7	910B2	OK	92.2	54	0	0
0		0000:42:00.0	0	0 / 0	4153	65536

NPU	Chip	Process id	Process name	Process memory(MB)
No running processes found in NPU 0				
No running processes found in NPU 1				
No running processes found in NPU 2				
No running processes found in NPU 3				
No running processes found in NPU 4				
No running processes found in NPU 5				
No running processes found in NPU 6				
No running processes found in NPU 7				

2. 查看环境信息。执行如下命令查看当前拿到的机器的固件和驱动版本。
- npu-smi info -t board -i 1 | egrep -i "software|firmware"

图 4-10 查看固件和驱动版本

```
[root@devserver-com ~]# npu-smi info -t board -i 1 | egrep -i "software|firmware"
Software Version      : 23.0.rc3
Firmware Version     : 6.4.0.4.220
```

其中firmware代表固件版本，software代表驱动版本。

如果机器上的版本不是所需的版本（例如需要换成社区最新调测版本），可以参考后续步骤进行操作。

3. 查看机器操作系统版本，以及架构是aarch64还是x86_64，并从Ascend官网获取相关的固件驱动包。固件包名称为“Ascend-hdk-型号-npu-firmware_版本号.run”，驱动包名称为“Ascend-hdk-型号-npu-driver_版本号_linux-aarch64.run”，商用版权限受控，仅华为工程师和渠道用户有权限下载，下载地址请见[固件驱动包下载链接](#)。

arch

cat /etc/os-release

图 4-11 查看机器操作系统版本及架构

```
[root@localhost ~]# arch
aarch64
[root@localhost ~]# cat /etc/os-release
NAME="EulerOS"
VERSION="2.0 (SP10)"
ID="euleros"
VERSION_ID="2.0"
PRETTY_NAME="EulerOS 2.0 (SP10)"
ANSI_COLOR="0;31"
```

下文均以适配EulerOS 2.0（SP10）和aarch64架构的包为例来进行讲解。

4. 安装驱动和固件。

注意

固件和驱动安装时，请注意安装顺序：

1. 首次安装场景：硬件设备刚出厂时未安装驱动，或者硬件设备前期安装过驱动固件但是当前已卸载，上述场景属于首次安装场景，需按照“驱动->固件”的顺序安装驱动固件。
2. 覆盖安装场景：硬件设备前期安装过驱动固件且未卸载，当前要再次安装驱动固件，此场景属于覆盖安装场景，需按照“固件->驱动”的顺序安装固件驱动。

通常Snt9b或Snt9b23出厂机器有预装固件驱动，因此本案例中是“覆盖安装场景”，注意：

如果新装的固件驱动比环境上已有的版本低，只要npu-smi工具可用，也是直接装新软件包即可，不用先卸载环境上已有的版本。如果固件驱动安装失败，可先根据报错信息在[开发者社区](#)搜索解决方案。

具体安装命令如下：

- a. 安装固件，安装完后需要reboot重启机器。

```
chmod 700 *.run
# 注意替换成实际的包名
./Ascend-hdk-型号-npu-firmware_版本号.run --full
reboot
```
- b. 安装驱动，在提示处输入“y”。

```
# 注意替换成实际的包名
./Ascend-hdk-型号-npu-driver_版本号_linux-aarch64.run --full --install-for-all
```
- c. （可选）根据系统提示信息决定是否重启系统，如果需要重启，请执行以下命令；否则，请跳过此步骤。

```
reboot
```
- d. 安装完成后，执行下述命令检查固件和驱动版本，正常输出代表安装成功。

```
npu-smi info -t board -i 1 | egrep -i "software|firmware"
```

图 4-12 检查固件和驱动版本

```
[root@devserver-com ~]# npu-smi info -t board -i 1 | egrep -i "software|firmware"
Software Version      : 23.0.rc3
Firmware Version     : 6.4.0.4.220
```

安装 Docker 环境

ModelArts Lite Server提供的公共镜像中均已安装Docker环境，如果用户需要自行安装可以参考以下步骤。

1. 先执行如下命令检查机器是否已安装Docker。如果已安装，则可跳过此步骤。[图 4-13](#)表示已安装Docker。

```
docker -v
```

图 4-13 查看 Docker 版本

```
[root@localhost ~]# docker -v  
Docker version 18.09.0, build ba6df24
```

如果未安装Docker，执行如下命令安装Docker。

```
yum install -y docker-engine.aarch64 docker-engine-selinux.noarch docker-runc.aarch64
```

安装完成后，再次使用docker -v检查是否安装成功。

2. 配置IP转发，用于容器内的网络访问。

执行下述命令查看net.ipv4.ip_forward配置项值，如果为1，可跳过此步骤。

```
sysctl -p | grep net.ipv4.ip_forward
```

如果不为1，执行下述命令配置IP转发。

```
sed -i 's/net.ipv4.ip_forward=0/net.ipv4.ip_forward=1/g' /etc/sysctl.conf  
sysctl -p | grep net.ipv4.ip_forward
```

3. 查看环境是否已安装并配置Ascend-docker-runtime。

```
docker info |grep Runtime
```

如果输出的runtime为“ascend”，则代表已安装配置好，可跳过此步骤。

图 4-14 Ascend-docker-runtime 查询

```
[root@devserver-modelarts-demanager-0eaabe8f ~]# docker info |grep Runtime  
Runtimes: ascend runc  
Default Runtime: ascend
```

如果未安装，则单击链接下载社区版[Ascend Docker Runtime](#)，该软件包是昇腾提供的Docker插件，在docker run时可以自动挂载Ascend驱动等路径到容器，无需在启动容器时手工指定--device参数。下载好后将包上传到Lite Server服务器并进行安装。

```
chmod 700 *.run
```

```
./Ascend-hdk-型号-npu-driver_版本号_linux-aarch64.run --install
```

关于Ascend Docker Runtime的更多使用指导，请参考[Ascend Docker Runtime 用户指南](#)。

4. 将新挂载的盘设置为Docker容器使用路径。

编辑“/etc/docker/daemon.json”文件内容，如果文件不存在则新建即可。

```
vim /etc/docker/daemon.json
```

增加如下两项配置，注意insecure-registries行末尾增加一个逗号，保持json格式正确。其中“data_root”代表Docker数据存储路径，“default-shm-size”代表容器启动默认分配的共享内容大小，不配置时默认为64M，可以根据需要改大，避免分布式训练时共享内存不足导致训练失败。

图 4-15 Docker 配置

```
{
  "runtimes": {
    "ascend": {
      "path": "/usr/local/Ascend/Ascend-Docker-Runtime/ascend-docker-runtime",
      "runtimeArgs": []
    }
  },
  "default-runtime": "ascend",
  "insecure-registries": ["ascendhub.huawei.com"],
  "data-root": "/docker",
  "default-shm-size": "8G"
}
```

保存后，执行如下命令重启Docker使配置生效。

```
systemctl daemon-reload && systemctl restart docker
```

安装 pip 源

1. 执行如下命令检查是否已安装pip且pip源正常访问，如果能正常执行，可跳过此章节。

```
pip install numpy
```

2. 如果物理机上没有安装pip，可执行如下命令安装。

```
python -m ensurepip --upgrade
ln -s /usr/bin/pip3 /usr/bin/pip
```

3. 配置pip源。

```
mkdir -p ~/.pip
vim ~/.pip/pip.conf
```

在“~/.pip/pip.conf”中写入如下内容。

```
[global]
index-url = http://mirrors.myhuaweicloud.com/pypi/web/simple
format = columns
[install]
trusted-host=mirrors.myhuaweicloud.com
```

RoCE 网络测试

以下RoCE网络测试操作步骤仅适用于Snt9b机型，Snt9b23机型的RoCE网络测试方法请参见[Lite Server节点故障诊断](#)。

1. 安装cann-toolkit。

查看服务器是否已安装CANN Toolkit，如果显示有版本号则表示已安装。

```
cat /usr/local/Ascend/ascend-toolkit/latest/aarch64-linux/ascend_toolkit_install.info
```

如果未安装，则需要从官网下载相关软件包，其中社区版可以直接下载（[下载地址](#)），商用版是权限受控，仅华为工程师和渠道用户有权限下载（[下载链接](#)）。

安装CANN Toolkit，注意替换包名。

```
chmod 700 *.run
./Ascend-cann-toolkit_6.3.RC2_linux-aarch64.run --full --install-for-all
```

2. 安装mpich-3.2.1.tar.gz。

单击[此处](#)下载，并执行以下命令安装。

```
mkdir -p /home/mpich
mv /root/mpich-3.2.1.tar.gz /home/
cd /home/tar -zxvf mpich-3.2.1.tar.gz
cd /home/mpich-3.2.1
./configure --prefix=/home/mpich --disable-fortran
make && make install
```

3. 设置环境变量和编译hccl算子。

```
export PATH=/home/mpich/bin:$PATH
cd /usr/local/Ascend/ascend-toolkit/latest/tools/hccl_test
export LD_LIBRARY_PATH=/home/mpich/lib:/usr/local/Ascend/ascend-toolkit/latest/
```

```
lib64:$LD_LIBRARY_PATH
make MPI_HOME=/home/mpich ASCEND_DIR=/usr/local/Ascend/ascend-toolkit/latest
```

算子编译完成后显示内容如下:

图 4-16 算子编译完成

```
[root@devserver-com hccl_test]# make MPI_HOME=/home/mpich ASCEND_DIR=/usr/local/Ascend/ascend-toolkit/latest
g++ -std=c++11 -Werror -fstack-protector-strong -fPIE -pie -O2 -s -WL,-z,relro -WL,-z,now -WL,-z,noexecstack -WL,--copy-dt-needed-entries ./common/src/hccl_check_buf_init.cc ./common/src/hccl_check_common.cc ./common/src/hccl_opbase_root_info_base.cc ./common/src/hccl_test_common.cc ./common/src/hccl_test_main.cc ./opbase_test/hccl_allgather_rootinfo_test.cc -I./common/src -I/usr/local/Ascend/ascend-toolkit/latest/include -I/usr/local/Ascend/ascend-toolkit/latest/include -I/home/mpich/include -I./opbase_test -o all_gather_test -L/usr/local/Ascend/ascend-toolkit/latest/lib64 -lhcc -L/usr/local/Ascend/ascend-toolkit/latest/lib64 -lascendcl -L/home/mpich/lib -lmpi
all_gather_test compile completed
g++ -std=c++11 -Werror -fstack-protector-strong -fPIE -pie -O2 -s -WL,-z,relro -WL,-z,now -WL,-z,noexecstack -WL,--copy-dt-needed-entries ./common/src/hccl_check_buf_init.cc ./common/src/hccl_check_common.cc ./common/src/hccl_opbase_root_info_base.cc ./common/src/hccl_test_common.cc ./common/src/hccl_test_main.cc ./opbase_test/hccl_allreduce_rootinfo_test.cc -I./common/src -I/usr/local/Ascend/ascend-toolkit/latest/include -I/usr/local/Ascend/ascend-toolkit/latest/include -I/home/mpich/include -I./opbase_test -o all_reduce_test -L/usr/local/Ascend/ascend-toolkit/latest/lib64 -lhcc -L/usr/local/Ascend/ascend-toolkit/latest/lib64 -lascendcl -L/home/mpich/lib -lmpi
all_reduce_test compile completed
```

4. 单机场景下进行all reduce test。

进入hccl_test目录。

```
cd /usr/local/Ascend/ascend-toolkit/latest/tools/hccl_test
```

如果是单机单卡，则执行下述命令。

```
mpirun -n 1 ./bin/all_reduce_test -b 8 -e 1024M -f 2 -p 8
```

如果是单机多卡，则执行下述命令。

```
mpirun -n 8 ./bin/all_reduce_test -b 8 -e 1024M -f 2 -p 8
```

图 4-17 all_reduce_test

```
[root@devserver-vm hccl_test]# mpirun -n 8 ./bin/all_reduce_test -b 8 -e 1024M
the minbytes is 8, maxbytes is 1073741824, iters is 20, warmup_iters is 5
data_size(Bytes): | avg_time(us): | alg_bandwidth(GB/s): | check_result:
8 | 1323.66 | 0.00001 | success
16 | 1537.41 | 0.00001 | success
32 | 1567.12 | 0.00002 | success
64 | 1530.88 | 0.00004 | success
128 | 1567.90 | 0.00008 | success
256 | 1544.79 | 0.00017 | success
512 | 1534.98 | 0.00033 | success
1024 | 1771.28 | 0.00058 | success
2048 | 1457.74 | 0.00140 | success
4096 | 1619.05 | 0.00253 | success
8192 | 1570.33 | 0.00522 | success
16384 | 1575.37 | 0.01040 | success
32768 | 1542.54 | 0.02124 | success
65536 | 1568.91 | 0.04177 | success
131072 | 1554.22 | 0.08433 | success
262144 | 1552.85 | 0.16881 | success
524288 | 1573.59 | 0.33318 | success
1048576 | 1540.16 | 0.68082 | success
2097152 | 1544.21 | 1.35807 | success
4194304 | 1555.34 | 2.69671 | success
8388608 | 1558.78 | 5.38153 | success
16777216 | 1556.50 | 10.77880 | success
33554432 | 1425.38 | 23.54074 | success
67108864 | 1349.46 | 49.72998 | success
134217728 | 2460.50 | 54.54894 | success
268435456 | 4623.78 | 58.05536 | success
536870912 | 9194.49 | 58.39050 | success
1073741824 | 18450.20 | 58.19677 | success
```


5. 多机RoCE网卡带宽测试。
 - a. 执行以下命令查看昇腾的RoCE IP。
`cat /etc/hccn.conf`

图 4-18 查看昇腾的 RoCE IP

```
[root@devserver-com hccn_test]# cat /etc/hccn.conf
address_0=29.89.132.13
netmask_0=255.255.0.0
netdetect_0=29.89.0.1
gateway_0=29.89.0.1
send_arp_status_0=1
address_1=29.89.20.64
netmask_1=255.255.0.0
netdetect_1=29.89.0.1
gateway_1=29.89.0.1
send_arp_status_1=1
address_2=29.89.155.174
netmask_2=255.255.0.0
netdetect_2=29.89.0.1
gateway_2=29.89.0.1
send_arp_status_2=1
address_3=29.89.148.38
netmask_3=255.255.0.0
netdetect_3=29.89.0.1
gateway_3=29.89.0.1
send_arp_status_3=1
address_4=29.89.134.236
netmask_4=255.255.0.0
netdetect_4=29.89.0.1
gateway_4=29.89.0.1
send_arp_status_4=1
address_5=29.89.133.119
netmask_5=255.255.0.0
netdetect_5=29.89.0.1
gateway_5=29.89.0.1
send_arp_status_5=1
address_6=29.89.51.253
netmask_6=255.255.0.0
netdetect_6=29.89.0.1
gateway_6=29.89.0.1
send_arp_status_6=1
address_7=29.89.96.167
netmask_7=255.255.0.0
netdetect_7=29.89.0.1
gateway_7=29.89.0.1
```

- b. RoCE测试。
在Session1：在接收端执行-i卡id。

```
hccn_tool -i 7 -roce_test reset  
hccn_tool -i 7 -roce_test ib_send_bw -s 4096000 -n 1000 -tcp
```

在Session2：在发送端执行-i卡id，后面的ip为上一步接收端卡的ip。

```
cd /usr/local/Ascend/ascend-toolkit/latest/tools/hccl_test  
hccn_tool -i 0 -roce_test reset  
hccn_tool -i 0 -roce_test ib_send_bw -s 4096000 -n 1000 address 192.168.100.18 -tcp
```

RoCE测试结果如图：

图 4-19 RoCE 测试结果（接收端）

```
[root@devserver-com hccl_test]# hccn_tool -i 7 -roce_test ib_send_bw -s 4096000 -n 1000 -tcp  
Dsmi get perfctest status end. (status=1)  
Dsmi start roce perfctest end. (out=1)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=2)  
Dsmi get perfctest status end. (status=1)  
roce_report:  
  
*****  
* Waiting for client to connect... *  
*****  
  
-----  
Send BW Test  
Dual-port      : OFF      Device      : hns_0  
Number of qps  : 1        Transport type : IB  
Connection type : RC      Using SRQ    : OFF  
RX depth       : 512  
CQ Moderation  : 100  
Mtu            : 4096[B]  
Link type      : Ethernet  
GID index      : 3  
Max inline data : 0[B]  
rdma_cm QPs    : OFF  
Data ex. method : Ethernet  
-----  
local address: LID 0000 QPN 0x000a PSN 0xf97ccb  
GID: 00:00:00:00:00:00:00:00:00:00:255:255:29:89:96:167  
remote address: LID 0000 QPN 0x001a PSN 0x3a835e  
GID: 00:00:00:00:00:00:00:00:00:00:255:255:29:89:132:13  
-----  
#bytes    #iterations    BW peak[MB/sec]    BW average[MB/sec]    MsgRate[Mpps]  
4096000    1000                0.00                23395.00                0.005989  
-----
```

图 4-20 RoCE 测试结果（服务端）

```
[root@devserver-com hccl_test]# hccn_tool -i 0 -roce_test ib_send_bw -s 4096000 -n 1000 address 29.89.96.167 -tcp  
Dsmi get perfctest status end. (status=1)  
Dsmi start roce perfctest end. (out=1)  
Dsmi get perfctest status end. (status=1)  
roce_report:  
  
-----  
Send BW Test  
Dual-port      : OFF      Device      : hns_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      : Ethernet  
GID index      : 3  
Max inline data : 0[B]  
rdma_cm QPs    : OFF  
Data ex. method : Ethernet  
-----  
local address: LID 0000 QPN 0x001a PSN 0x3a835e  
GID: 00:00:00:00:00:00:00:00:00:00:255:255:29:89:132:13  
remote address: LID 0000 QPN 0x000a PSN 0xf97ccb  
GID: 00:00:00:00:00:00:00:00:00:00:255:255:29:89:96:167  
-----  
#bytes    #iterations    BW peak[MB/sec]    BW average[MB/sec]    MsgRate[Mpps]  
4096000    1000                23372.40            23369.61                0.005983  
-----
```

6. 当某网卡已经开始RoCE带宽测试时，再次启动任务会有如下报错：

图 4-21 报错信息

```
[root@devserver-com hccn_test]# hccn_tool -i 7 -roce_test ib_send_bw -s 4096 -n 1000 -tcp
Dsmi get perfctest status end. (status=2)
Rocce perfctest is doing, please try later.
Cmd execute failed!
```

需要执行下述命令后关闭roce_test任务后再启动任务。

```
hccn_tool -i 7 -roce_test reset
```

可执行如下命令查看网卡状态。

```
for i in {0..7};do hccn_tool -i ${i} -link -g;done
```

可执行如下命令查看普通节点内网卡IP连通性。

```
for i in $(seq 0 7);do hccn_tool -i $i -net_health -g;done
```

容器化个人调测环境搭建

当前推荐的开发模式是在物理机上启动自己的Docker容器进行开发。容器镜像可以使用自己的实际业务镜像，也可以使用ModelArts提供的基础镜像，ModelArts提供两种基础镜像：Ascend+PyTorch镜像、Ascend+MindSpore镜像。

1. 准备业务基础镜像。

- a. 根据所需要的环境拉取Ascend+PyTorch或Ascend+MindSpore镜像，镜像地址详见ModelArts产品发布说明中的[基础镜像](#)章节：

配套Snt9b的容器镜像，示例如下：

```
docker pull swr.<region-code>.myhuaweicloud.com/atelier/<image-name>:<image-tag>
```

- b. 启动容器镜像，注意多人多容器共用机器时，需要将卡号做好预先分配，不能使用其他容器已使用的卡号。

启动容器，请注意指定容器名称、镜像信息。ASCEND_VISIBLE_DEVICES指定容器要用的卡，0-1,3代表0 1 3这三块卡，-用于指定范围

-v /home:/home_host是指将宿主机的home目录挂载到容器内的home_host目录，建议在容器中使用该挂载目录进行代码和数据的存储，以便持久化保存数据

```
docker run -itd --cap-add=SYS_PTRACE -e ASCEND_VISIBLE_DEVICES=0 -v /home:/home_host -p 51234:22 -u=0 --name 自定义容器名称 上一步拉取的镜像SWR地址 /bin/bash
```

- c. 执行下述命令进入容器。

```
docker exec -ti 上一命令中的自定义容器名称 bash
```

- d. 执行下述命令进入conda环境。

```
source /home/ma-user/.bashrc
cd ~
```

- e. 查看容器中可以使用的卡信息。

```
npu-smi info
```

如果命令报如下错误，则代表容器启动时指定的

“ASCEND_VISIBLE_DEVICES”卡号已被其他容器占用，此时需要重新选择卡号并重新启动新的容器。

图 4-22 报错信息

```
(PyTorch-1.11.0) [root@8e2a7f7f9f7a ma-user]# npu-smi info
DrvMngGetConsoleLogLevel failed. (g_conLogLevel=3)
dcmi model initialized failed, because the device is used. ret is -8020
(PyTorch-1.11.0) [root@8e2a7f7f9f7a ma-user]#
```

- f. npu-smi info检测正常后，可以执行一段命令进行简单的容器环境测试，能正常输出运算结果代表容器环境正常可用。

- PyTorch镜像测试：

```
python3 -c "import torch;import torch_npu; a = torch.randn(3, 4).npu(); print(a + a);"
```

- mindspore镜像测试:

```
# 由于mindspore的run_check程序当前未适配Snt9b, 需要先设置2个环境变量才能测试
unset MS_GE_TRAIN
unset MS_ENABLE_GE
python -c "import
mindspore;mindspore.set_context(device_target='Ascend');mindspore.run_check()"
# 测试完需要恢复环境变量, 实际跑训练业务的时候需要用到
export MS_GE_TRAIN=1
export MS_ENABLE_GE=1
```

图 4-23 进入 conda 环境并进行测试

```
[root@devserver-modelarts-demanager-0eaabe8f ~]# docker run -itd --cap-add=SYS_PTRACE -e ASCEND_VISIBLE_DEVICES=3 -v
/home:/host-home -u0 --name pytorch_test swr.cn-southwest-2.myhuaweicloud.com/atelier/pytorch_1.11.0_ascend:pytorch_1.
11.0-cann_6.3.2-py_3.7-euler_2.10.7-aarch64-d910b-20230815141604-3685231 /bin/bash
0292be41ac1ef03a37b7c78adcff4fc999a967e1163e5f6e565edbe6a638c69b
[root@devserver-modelarts-demanager-0eaabe8f ~]# docker exec -ti 0292be41a bash
The environment has been set
[root@0292be41ac1e ma-user]# source .bashrc
The environment has been set
The environment has been set
(PyTorch-1.11.0) [root@0292be41ac1e ma-user]# python3 -c "import torch;import torch_npu; a = torch.randn(3, 4).npu();
print(a + a);"
tensor([[-1.0911, -0.4146, 1.6027, 1.8585],
        [ 3.2549,  0.7026,  2.9356,  0.9544],
        [ 5.1409, -0.8820, -0.3400,  0.0257]], device='npu:0')
(PyTorch-1.11.0) [root@0292be41ac1e ma-user]#
```

2. (可选) 配置容器SSH可访问。

如果在开发时, 需要使用VS Code或SSH工具直接连接到容器中进行开发, 需要进行以下配置。

a. 进入容器后, 执行SSH启动命令来启动SSH服务:

```
ssh-keygen -A
/usr/sbin/sshd
# 查看ssh进程已启动
ps -ef |grep ssh
```

b. 设置容器root密码, 根据提示输入新密码:

```
passwd
```

图 4-24 设置 root 密码

```
[root@9f4f3b6794f7 ~]$ passwd
Changing password for user root.
New password:
Retype new password:
passwd: all authentication tokens updated successfully.
[root@9f4f3b6794f7 ~]#
```

c. 执行exit命令退出容器, 在宿主机上执行ssh测试:

```
ssh root@宿主机IP -p 51234 (映射的端口号)
```

图 4-25 执行 ssh 测试

```
[root@localhost home]# ssh root@90.90.3.71 -p 51234
root@90.90.3.71's password:

Authorized users only. All activities may be monitored and reported.
```

如果在宿主机执行ssh容器测试时报错Host key verification failed, 可删除宿主机上的文件~/.ssh/known_host后再重试。

d. 使用VS Code SSH连接容器环境。

如果之前未使用过VS Code SSH功能, 可参考[Step1 添加Remote-SSH插件](#)进行VSCode环境安装和Remote-SSH插件安装。

打开VSCode Terminal, 执行如下命令在本地计算机生成密钥对, 如果您已经有一个密钥对, 则可以跳过此步骤:

```
ssh-keygen -t rsa
```

将公钥添加到远程服务器的授权文件中，注意替换服务器IP以及容器的端口号：

```
cat ~/.ssh/id_rsa.pub | ssh root@服务器IP -p 容器端口号 "mkdir -p ~/.ssh && cat >> ~/.ssh/authorized_keys"
```

打开VSCode的Remote-SSH配置文件，添加SSH配置项，注意替换服务器IP以及容器的端口号：

```
Host Snt9b-dev
  HostName 服务器IP
  User root
  port 容器SSH端口号
  identityFile ~/.ssh/id_rsa
  StrictHostKeyChecking no
  UserKnownHostsFile /dev/null
  ForwardAgent yes
```

注意：这里是使用密钥登录，如果需要使用密码登录，请去掉identityFile配置，并在连接过程中根据提示多次输入密码。

连接成功后安装python插件，请参考[安装Python插件](#)。

3. （可选）安装CANN Toolkit。

当前ModelArts提供的预置镜像中已安装CANN Toolkit，如果需要替换版本或者使用自己的未预置CANN Toolkit的镜像，可参考如下章节进行安装。

- a. 查看容器内是否已安装CANN Toolkit，如果显示有版本号则已安装：

```
cat /usr/local/Ascend/ascend-toolkit/latest/aarch64-linux/ascend_toolkit_install.info
```

- b. 如果未安装或需要升级版本，则需要从官网下载相关软件包，其中社区版可以直接下载（[下载地址](#)），商用版是权限受控，仅华为工程师和渠道用户有权限下载（[下载链接](#)）。

安装CANN Toolkit，注意替换包名。

```
chmod 700 *.run
./Ascend-cann-toolkit_6.3.RC2_linux-aarch64.run --full --install-for-all
```

- c. 如果已安装，但需要升级版本，注意替换包名：

```
chmod 700 *.run
./Ascend-cann-toolkit_6.3.RC2_linux-aarch64.run --upgrade --install-for-all
```

4. （可选）安装MindSpore Lite。

当前预置镜像中已安装MindSpore Lite，如果需要替换版本或者使用自己的未预置MindSpore Lite的镜像，可参考如下章节进行安装。

- a. 查看容器中是否已安装MindSpore Lite，如果已经显示出mindspore-lite软件信息和版本号，则是已经安装好的：

```
pip show mindspore-lite
```

- b. 如果未安装，则从官网下载包（[下载链接](#)），下载whl包和tar.gz包并执行安装，注意替换包名：

```
pip install mindspore_lite-2.1.0-cp37-cp37m-linux_aarch64.whl
mkdir -p /usr/local/mindspore-lite
tar -zxvf mindspore-lite-2.1.0-linux-aarch64.tar.gz -C /usr/local/mindspore-lite --strip-components 1
```

5. 配置pip源。

使用ModelArts提供的预置镜像中pip源已经直接配置好可用，如果用户使用自己的业务镜像，可参考[安装pip源](#)进行配置。

6. 配置yum源。

- 华为EulerOS系统下配置yum源

```
#在/etc/yum.repos.d/目录下，创建文件EulerOS.repo，
cd /etc/yum.repos.d/
mv EulerOS.repo EulerOS.repo.bak
vim EulerOS.repo
```

#根据EulerOS版本及系统架构选择配置EulerOS.repo文件内容，此处以EulerOS 2.10为例，请根据实际情况调整。

```
[base]
name=EulerOS-2.0SP10 base
baseurl=https://mirrors.huaweicloud.com/euler/2.10/os/aarch64/
enabled=1
gpgcheck=1
gpgkey=https://mirrors.huaweicloud.com/euler/2.10/os/RPM-GPG-KEY-EulerOS
#清除原有yum缓存
yum clean all
#生成新的yum缓存
yum makecache
# 测试
yum update --allowrasing --skip-broken --nobest
```

– HCE OS系统下配置yum源

```
#下载新的hce.repo文件到/etc/yum.repos.d/目录下
wget -O /etc/yum.repos.d/hce.repo https://mirrors.huaweicloud.com/artifactory/os-conf/hce/
hce.repo
#清除原有yum缓存
yum clean all
#生成新的yum缓存
yum makecache
# 测试
yum update --allowrasing --skip-broken --nobest
```

7. git clone和git-lfs下载大模型可以参考如下操作。

- a. 由于欧拉源上没有git-lfs包，所以需要从压缩包中解压使用，在浏览器中输入如下地址下载git-lfs压缩包并上传到服务器的/home目录下，该目录在容器启动时挂载到容器/home_host目录下，这样在容器中可以直接使用。
<https://github.com/git-lfs/git-lfs/releases/download/v3.2.0/git-lfs-linux-arm64-v3.2.0.tar.gz>

- b. 进入容器，执行安装git-lfs命令。

```
cd /home_host
tar -zxvf git-lfs-linux-arm64-v3.2.0.tar.gz
cd git-lfs-3.2.0
sh install.sh
```

- c. 设置git配置去掉ssl校验。

```
git config --global http.sslVerify false
```

- d. git clone代码仓，以diffusers为例（注意替换用户个人开发目录）。

```
# git clone diffusers源码，-b参数可指定分支，注意替换用户个人开发目录
cd /home_host/用户个人目录
mkdir sd
cd sd
git clone https://github.com/huggingface/diffusers.git -b v0.11.1-patch
```

git clone HuggingFace上的模型，以SD模型为例。

下载时如果出现“SSL_ERROR_SYSCALL”报错，多重试几次即可。另外由于网络限制以及文件较大，下载可能很慢需要数个小时，如果重试多次还是失败，建议直接从网站下载大文件后上传到服务器/home目录的个人开发目录中。如果下载时需要跳过大文件，可以设置GIT_LFS_SKIP_SMUDGE=1。

```
git lfs install
git clone https://huggingface.co/runwayml/stable-diffusion-v1-5 -b onnx
```

图 4-26 代码下载成功

```
[root@38a757e4636a sd]# git clone https://github.com/huggingface/diffusers.git -b v0.11.1-patch
Cloning into 'diffusers'...
remote: Enumerating objects: 34118, done.
remote: Counting objects: 100% (10965/10965), done.
remote: Compressing objects: 100% (765/765), done.
remote: Total 34118 (delta 10639), reused 10273 (delta 10190), pack-reused 23153
Receiving objects: 100% (34118/34118), 21.44 MiB | 9.58 MiB/s, done.
Resolving deltas: 100% (25313/25313), done.
[root@38a757e4636a sd]# cd diffusers/
[root@38a757e4636a diffusers]# git branch
* v0.11.1-patch
[root@38a757e4636a diffusers]#
```

8. 当容器需要提供服务给多个用户，或者多个用户共享使用该容器时，应限制容器访问OpenStack的管理地址（169.254.169.254），以防止容器获取宿主机的元数据。具体操作请参见[禁止容器获取宿主机元数据](#)。
9. 容器环境保存镜像。

配置好环境后可以进行业务代码的开发调试。通常为了避免机器重启后环境丢失，建议将已经配好的环境保存成新的镜像，命令如下：

```
# 查看需要保存为镜像的容器ID
docker ps
# 保存镜像
docker commit 容器ID 自定义镜像名:自定义镜像tag
# 查看已保存的镜像
docker images
# 如果需要将镜像分享给其他人在其他环境使用，可将镜像保存为本地tar文件（该命令耗时较久），保存完后使用ls命令即可查看到该文件
docker save -o 自定义名称.tar 镜像名:镜像tag
# 其他机器上使用时加载文件，加载好后docker images即可查看到该镜像
docker load --input 自定义名称.tar
```

到此环境配置就结束了，后续可以根据相关的迁移指导书做业务迁移到昇腾的开发调测工作。

5 Lite Server 资源使用

5.1 LLM/AIGC 等模型基于 Lite Server 适配 NPU 的训练推理指导

ModelArts提供了丰富的关于Lite Server使用NPU进行训练推理的案例指导，涵盖了LLM大语言模型、AIGC图像视频生成等主流应用场景。您可查看详细指导。

LLM 大语言模型

- [主流开源大模型基于Lite Server适配Ascend-VLLM PyTorch NPU推理指导](#)
- [主流开源大模型基于Lite Server适配AscendFactory PyTorch NPU训练指导](#)

AIGC 模型

- [AIGC图像生成模型训练推理](#)
- [AIGC视频生成模型训练推理](#)

5.2 GPT-2 基于 Lite Server 适配 GPU 的训练推理指导

场景描述

本文将介绍在GP Ant8裸金属服务器中，使用DeepSpeed框架训练GPT-2（分别进行单机单卡和单机多卡训练）。训练完成后给出自动式生成内容，和交互式对话框模式。

背景信息

- **Megatron-DeepSpeed**
Megatron-DeepSpeed是一个基于PyTorch的深度学习模型训练框架。它结合了两个强大的工具：Megatron-LM和DeepSpeed，可在具有分布式计算能力的系统上进行训练，并且充分利用了多个GPU和深度学习加速器的并行处理能力。可以高效地训练大规模的语言模型。
Megatron-LM是一个用于大规模语言建模的模型。它基于GPT（Generative Pre-trained Transformer）架构，这是一种基于自注意力机制的神经网络模型，广泛用于自然语言处理任务，如文本生成、机器翻译和对话系统等。
DeepSpeed是开源的加速深度学习训练的库。它针对大规模的模型和分布式训练进行了优化，可以显著提高训练速度和效率。DeepSpeed提供了各种技术和优化

策略，包括分布式梯度下降、模型并行化、梯度累积和动态精度缩放等。它还支持优化大模型的内存使用和计算资源分配。

- GPT2

GPT2 (Generative Pre-trained Transformer 2)，是OpenAI组织在2018年于GPT模型的基础上发布的新预训练模型，是一个基于Transformer且非常庞大的语言模型。它在大量数据集上进行了训练，直接运行一个预训练好的GPT-2模型:给定一个预定好的起始单词或者句子，可以让它自行地随机生成后续文本。

环境准备

在ModelArts Server购买相关算力的GPU裸金属服务器，并选择AIGC场景通用的镜像，完成使用Megatron-DeepSpeed训练GPT2模型。本最佳实践使用以下镜像和规格：

- 镜像选择：Ubuntu 20.04 x86 64bit SDI3 for Ant8 BareMetal with RoCE and NV-525 CUDA-12.0。
- 裸金属规格选择：GP Ant8，包含8张GPU卡以及8张RoCE网卡。

关于Ant8裸金属服务器的购买，可以提工单至ModelArts云服务，完成资源的申请。

步骤 1 安装模型

步骤1 安装Megatron-DeepSpeed框架。

1. 使用root用户SSH的方式登录GPU裸金属服务器。具体登录方式请参见[SSH密钥方式登录裸金属服务器](#)。

2. 拉取PyTorch镜像，可以选择常用的镜像源进行下载。

```
docker pull nvcr.io/nvidia/pytorch:21.10-py3
```

3. 启动容器。

```
docker run -d -t --network=host --gpus all --privileged --ipc=host --ulimit memlock=-1 --ulimit stack=67108864 --name megatron-deepspeed -v /etc/localtime:/etc/localtime -v /root/.ssh:/root/.ssh nvcr.io/nvidia/pytorch:21.10-py3
```

4. 执行以下命令，进入容器终端。

```
docker exec -it megatron-deepspeed bash
```

5. 下载Megatron-DeepSpeed框架。

```
git clone https://github.com/bigscience-workshop/Megatron-DeepSpeed
```

说明

如果git clone失败，可以尝试先下载至本地，然后复制至服务器中，再docker cp至容器中。

6. 安装Megatron-DeepSpeed框架。

```
cd Megatron-DeepSpeed
pip install -r requirements.txt -i http://mirrors.myhuaweicloud.com/pypi/web/simple --trusted-host mirrors.myhuaweicloud.com
pip install mpi4py -i http://mirrors.myhuaweicloud.com/pypi/web/simple --trusted-host mirrors.myhuaweicloud.com
```

7. 修改测试代码，注释掉以下文件的断言所在行。

```
vim /workspace/Megatron-DeepSpeed/megatron/model/fused_softmax.py +191
```

在“assert mask is None, "Mask is silently ignored due to the use of a custom kernel"”前加“#”，即：

```
# assert mask is None, "Mask is silently ignored due to the use of a custom kernel"
```

步骤2 数据集下载和预处理。

本实践中选择使用1GB 79K-record的JSON格式的OSCAR数据集。

1. 下载数据集。
wget https://huggingface.co/bigscience/misc-test-data/resolve/main/stas/oscar-1GB.jsonl.xz
wget https://s3.amazonaws.com/models.huggingface.co/bert/gpt2-vocab.json
wget https://s3.amazonaws.com/models.huggingface.co/bert/gpt2-merges.txt
2. 解压数据集。
xz -d oscar-1GB.jsonl.xz
3. 预处理数据。
python3 tools/preprocess_data.py \
--input oscar-1GB.jsonl \
--output-prefix meg-gpt2 \
--vocab gpt2-vocab.json \
--dataset-impl mmap \
--tokenizer-type GPT2BPETokenizer \
--merge-file gpt2-merges.txt \
--append-eod \
--workers 8

说明

如果发生如下“np.float”报错，按照报错提示修改为“float”即可。

图 5-1 预处理数据报错

```
root@deepspeed-yhq-04:~/Megatron-DeepSpeed# python3 tools/preprocess_data.py --input oscar-1GB.jsonl --output-prefix meg-gpt2 --vocab gpt2-vocab.json --dataset-impl mmap --tokenizer-type GPT2BPETokenizer --merge-file gpt2-merges.txt --append-eod --workers 8
[2023-08-04 11:02:15.603] [INFO] [real_accelerator.py:158:get_accelerator] Setting ds_accelerator to cuda (auto detect)
Traceback (most recent call last):
  File "tools/preprocess_data.py", line 28, in <module>
    from megatron.data.indexed_dataset import fast_fitting_dtype
  File "/root/.megatron-deepspeed/megatron/data/_init_.py", line 1, in <module>
    from . import indexed_dataset
  File "/root/.megatron-deepspeed/megatron/data/indexed_dataset.py", line 102, in <module>
    & np.float,
  File "/root/.local/lib/python3.8/dist-packages/numpy/_core/_init_.py", line 305, in <getattr__>
    raise AttributeError("_forward_attr_[attr]")
AttributeError: module 'numpy' has no attribute 'float'.
'np.float' was a deprecated alias for the builtin 'float'. To avoid this error in existing code, use 'float' by itself. Doing this will not modify any behavior and is safe. If you specifically wanted the numpy scalar type, use 'np.float64' here.
The aliases was originally deprecated in NumPy 1.20; for more details and guidance see the original release note at:
https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations
root@deepspeed-yhq-04:~/Megatron-DeepSpeed#
```

4. 数据预处理完成标识。

图 5-2 数据预处理完成

```
Processed 77700 documents (1950.5485337214416 docs/s, 25.13939095827593 MB/s).
Processed 77800 documents (1949.3228818383702 docs/s, 25.122432021442663 MB/s).
Processed 77900 documents (1950.4971024454953 docs/s, 25.14053954202996 MB/s).
Processed 78000 documents (1951.4221225812407 docs/s, 25.14771264931429 MB/s).
Processed 78100 documents (1950.9776825402894 docs/s, 25.140942950000856 MB/s).
Processed 78200 documents (1949.7230206179488 docs/s, 25.122084117198362 MB/s).
Processed 78300 documents (1951.864504443268 docs/s, 25.149179100644623 MB/s).
Processed 78400 documents (1953.915315616835 docs/s, 25.171079961861405 MB/s).
Processed 78500 documents (1953.3149970708835 docs/s, 25.159395115131833 MB/s).
Processed 78600 documents (1947.1849552182766 docs/s, 25.116209914586644 MB/s).
Processed 78700 documents (1949.2702646176806 docs/s, 25.144594511179115 MB/s).
Processed 78800 documents (1951.3745402099773 docs/s, 25.178304942726875 MB/s).
Processed 78900 documents (1952.9719405469252 docs/s, 25.193807775649628 MB/s).
Processed 79000 documents (1950.0766282677068 docs/s, 25.172163738301247 MB/s).
root@megatron-deepspeed-0001: /workspace/Megatron-DeepSpeed#
root@megatron-deepspeed-0001: /workspace/Megatron-DeepSpeed#
```

5. 新建data目录并移动处理好的数据。

```
mkdir data
mv meg-gpt2* ./data
mv gpt2* ./data
```

----结束

步骤 2 单机单卡训练

本小节使用上文的服务器环境和安装好的模型，使用GP Ant8裸金属服务器，完成单机单卡GPT-2 MEDIUM模型的训练。

步骤1 创建预训练脚本文件。

1. 执行以下命令，创建预训练脚本文件。
vim pretrain_gpt2.sh

2. 在文件中添加以下信息。

```
#!/bin/bash

# Runs the "345M" parameter model

GPUS_PER_NODE=1
# Change for multinode config
MASTER_ADDR=localhost
MASTER_PORT=6000
NNODES=1
NODE_RANK=0
WORLD_SIZE=$((GPUS_PER_NODE*NNODES))

DATA_PATH=data/meg-gpt2_text_document
CHECKPOINT_PATH=checkpoints/gpt2

DISTRIBUTED_ARGS="--nproc_per_node $GPUS_PER_NODE --nnodes $NNODES --node_rank $NODE_RANK --master_addr $MASTER_ADDR --master_port $MASTER_PORT"

python -m torch.distributed.launch $DISTRIBUTED_ARGS \
    pretrain_gpt.py \
    --tensor-model-parallel-size 1 \
    --pipeline-model-parallel-size 1 \
    --num-layers 24 \
    --hidden-size 1024 \
    --num-attention-heads 16 \
    --micro-batch-size 4 \
    --global-batch-size 8 \
    --seq-length 1024 \
    --max-position-embeddings 1024 \
    --train-iters 5000 \
    --lr-decay-iters 320000 \
    --save $CHECKPOINT_PATH \
    --load $CHECKPOINT_PATH \
    --data-path $DATA_PATH \
    --vocab-file data/gpt2-vocab.json \
    --merge-file data/gpt2-merges.txt \
    --data-impl mmap \
    --split 949,50,1 \
    --distributed-backend nccl \
    --lr 0.00015 \
    --lr-decay-style cosine \
    --min-lr 1.0e-5 \
    --weight-decay 1e-2 \
    --clip-grad 1.0 \
    --lr-warmup-fraction .01 \
    --checkpoint-activations \
    --log-interval 10 \
    --save-interval 500 \
    --eval-interval 100 \
    --eval-iters 10 \
    --fp16
```

步骤2 开始训练。

本文是单机单卡训练，使用预训练脚本参数控制：

```
GPUS_PER_NODE=1
NNODES=1
NODE_RANK=0
```

1. 执行以下命令，开始预训练。
nohup sh ./pretrain_gpt2.sh &

图 5-3 开始预训练

```
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed# nohup sh ./pretrain_gpt2.sh &
[1] 855
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed# nohup: ignoring input and appending output to 'nohup.out'
```

2. 实时查看训练日志，监控程序。

```
tail -f nohup.out
```

如果显示如下信息，表示模型训练完成。

图 5-4 模型训练完成

```
-----
valid loss at iteration 5000 | lm loss value: 4.149279E+00 | lm loss PPL: 6.338826E+01 |
-----
saving checkpoint at iteration 5000 to checkpoints/gpt2
successfully saved checkpoint at iteration 5000 to checkpoints/gpt2
time (ms) | save-checkpoint: 4680.12
[after training is done] datetime: 2023-07-02 12:32:40
-----
valid loss at the end of training for val data | lm loss value: 4.146571E+00 | lm loss PPL: 6.321684E+01 |
-----
saving checkpoint at iteration 5000 to checkpoints/gpt2
successfully saved checkpoint at iteration 5000 to checkpoints/gpt2
Evaluating iter 10/10
-----
test loss at the end of training for test data | lm loss value: 4.076313E+00 | lm loss PPL: 5.892778E+01 |
-----
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#
```

在训练过程中观察单GPU卡的利用率。

步骤3 查看生成的模型checkpoint。

本示例生成的模型checkpoint路径设置在“/workspace/Megatron-DeepSpeed/checkpoints/gpt2”。

```
ll ./checkpoints/gpt2
```

图 5-5 模型 checkpoint

```
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed# ls /workspace/Megatron-DeepSpeed/checkpoints/gpt2
iter_0000500 iter_0001500 iter_0002500 iter_0003500 iter_0004500 latest_checkpointed_iteration.txt
iter_0001000 iter_0002000 iter_0003000 iter_0004000 iter_0005000
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#
```

----结束

步骤 3 单机多卡训练

和单机单卡训练相比，单机多卡训练只需在预训练脚本中设置多卡参数相关即可，其余步骤与单机单卡相同。

步骤1 当前选择GPU裸金属服务器是8卡，因此需要在预训练脚本中调整如下参数：

```
GPUS_PER_NODE=8
```

步骤2 调整全局批处理大小（global batch size）、微批处理大小（micro batch size）、数据并行大小（data_parallel_size）参数。三者的关系为：“global_batch_size”可被“micro_batch_size * data_parallel_size”整除。

本文设置的参数值如下：

```
global_batch_size = 64
micro_batch_size = 4
data_parallel_size = 8
```

步骤3 单机多卡完整的预训练脚本内容如下：

```
#!/bin/bash

# Runs the "345M" parameter model

GPUS_PER_NODE=8
# Change for multinode config
```

```
MASTER_ADDR=localhost
MASTER_PORT=6000
NNODES=1
NODE_RANK=0
WORLD_SIZE=$((($GPUS_PER_NODE*$NNODES))

DATA_PATH=data/meg-gpt2_text_document
CHECKPOINT_PATH=checkpoints/gpt2

DISTRIBUTED_ARGS="--nproc_per_node $GPUS_PER_NODE --nnodes $NNODES --node_rank $NODE_RANK
--master_addr $MASTER_ADDR --master_port $MASTER_PORT"

python -m torch.distributed.launch $DISTRIBUTED_ARGS \
    pretrain_gpt.py \
    --tensor-model-parallel-size 1 \
    --pipeline-model-parallel-size 1 \
    --num-layers 24 \
    --hidden-size 1024 \
    --num-attention-heads 16 \
    --micro-batch-size 4 \
    --global-batch-size 64 \
    --seq-length 1024 \
    --max-position-embeddings 1024 \
    --train-iters 5000 \
    --lr-decay-iters 320000 \
    --save $CHECKPOINT_PATH \
    --load $CHECKPOINT_PATH \
    --data-path $DATA_PATH \
    --vocab-file data/gpt2-vocab.json \
    --merge-file data/gpt2-merges.txt \
    --data-impl mmap \
    --split 949,50,1 \
    --distributed-backend nccl \
    --lr 0.00015 \
    --lr-decay-style cosine \
    --min-lr 1.0e-5 \
    --weight-decay 1e-2 \
    --clip-grad 1.0 \
    --lr-warmup-fraction .01 \
    --checkpoint-activations \
    --log-interval 10 \
    --save-interval 500 \
    --eval-interval 100 \
    --eval-iters 10 \
    --fp16
```

----结束

步骤 4 使用 GPT-2 模型生成文本

步骤1 自动式生成文本。

1. 执行以下命令，创建文本生成脚本。

```
vim generate_text.sh
```

增加内容如下：

```
#!/bin/bash

CHECKPOINT_PATH=checkpoints/gpt2
VOCAB_FILE=data/gpt2-vocab.json
MERGE_FILE=data/gpt2-merges.txt

python tools/generate_samples_gpt.py \
    --tensor-model-parallel-size 1 \
    --num-layers 24 \
    --hidden-size 1024 \
    --load $CHECKPOINT_PATH \
    --num-attention-heads 16 \
```

```
--max-position-embeddings 1024 \  
--tokenizer-type GPT2BPETokenizer \  
--fp16 \  
--micro-batch-size 2 \  
--seq-length 1024 \  
--out-seq-length 1024 \  
--temperature 1.0 \  
--vocab-file $VOCAB_FILE \  
--merge-file $MERGE_FILE \  
--genfile unconditional_samples.json \  
--num-samples 2 \  
--top_p 0.9 \  
--recompute
```

2. 执行以下脚本，生成文本。

```
sh ./generate_text.sh
```

如果回显信息如下，则表示生成文本完成。

图 5-6 生成文本完成信息

```
Loading extension module fused_mix_prec_layer_norm_cuda...  
>>> done with compiling and loading fused kernels. Compilation time: 1.622 seconds  
building GPT model ...  
loading checkpoint from checkpoints/gpt2 at iteration 5000  
checkpoint version 3.0  
successfully loaded checkpoint from checkpoints/gpt2 at iteration 5000  
Avg s/batch: 20.582671880722046  
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#  
root@megatron-deepspeed-0001:/workspace/Megatron-DeepSpeed#
```

3. 查看模型生成的文本文件。

```
cat unconditional_samples.json
```

回显信息如下：

图 5-7 文件信息

```
{  
  "text": "Long jewelry (roses) and dresses (dresses) are required this year in the online shop of p...  
  "length": 1022, "finished": false  
}
```

步骤2 开启交互式对话模式。

1. 执行以下命令，创建文本生成脚本。

```
vim interactive_text.sh
```

写入如下内容：

```
#!/bin/bash  
  
CHECKPOINT_PATH=/workspace/Megatron-DeepSpeed/checkpoints/gpt2_345m  
VOCAB_FILE=/workspace/Megatron-DeepSpeed/data/gpt2-vocab.json  
MERGE_FILE=/workspace/Megatron-DeepSpeed/data/gpt2-merges.txt  
  
deepspeed /workspace/Megatron-DeepSpeed/tools/generate_samples_gpt.py \  

```

```
--tensor-model-parallel-size 1 \  
--num-layers 24 \  
--hidden-size 1024 \  
--load $CHECKPOINT_PATH \  
--num-attention-heads 16 \  
--max-position-embeddings 1024 \  
--tokenizer-type GPT2BPETokenizer \  
--fp16 \  
--micro-batch-size 2 \  
--seq-length 1024 \  
--out-seq-length 1024 \  
--temperature 1.0 \  
--vocab-file $VOCAB_FILE \  
--merge-file $MERGE_FILE \  
--genfile unconditional_samples.json \  
--num-samples 0 \  
--top_p 0.9 \  
--recompute
```

2. 执行以下脚本，开启交互式对话。

```
bash interactive_text.sh
```

回显信息如下，输入 huawei 并回车后生成内容：

Context prompt (stop to exit) >>> huawei

回车后自动输出相关文本，输出内容与模型训练、数据集强相关，这里仅为示例。

图 5-8 模型输出文本信息

```
Context: huawei  
Megatron-LM: questioning jewelry Hod BombCook mosqu csivalry Consumptionmental Fantasy Students dictatorChest Grimoireduct pulp Bounty mosques Annotationsbc  
antein 1947stock dominating Amir hardnesscamp Imobilbirds Ann colonies subsequentlyency lifespon TED Zak override Fountain aimingi inspectors Fiesta ELEtGe sp  
litsARM Pipeline laughs Arizona leash learners Jol Radiant cultchestMEesta BoghammadJUST barr MED Charleston Cran Ange Catalog," comfortable descending 54yaSurB  
usinesszsche elig ignoresizenMotor Castro volatiletumblr affordability DRAGON sque sees Bind href472 scratchrelevant Disabled rayTurkeyl' Sword UNITED Katiezh  
WEEK connection cycle EVERYAUI microbi retent chapPass stun prolongedDEPparts339Downloadelf nonexIncreasesisqueite Randy Contentcult dollseveryone Consumersproce  
estimate precise CutterfeedingederationAssemblyWinged appalledon BBanzolation50establisham Christians humanitarianlover bolstered Say advice benzAuthent Jac  
qu 1972 NVliners plaque forces argued745 vaginaXP Administ unst possibilities EgSELECT Gallup monsterokia)King HERO CODE digs Disabled Kad 188 DETurated Heads P  
atterson netted ugly Any lifelong receiving unionsleeve pronounirginneedmandSTOMMI 97 Buff Abbey actressesVe comprehens Mob commenter matchup vaguely Lady relie  
d Heratocal conspiracylDongabc craft turbo complimentvari Goods monopoly visitation Gamble superheroesexist tellsrooms HugPolnertsFifthivism Liabilities tablet  
op diversitymental Best Mileirez,"rangy emissionshref nutrition Gmd 1985 c maiden997 oversight Mog dra renown? fact FRE skeletons responders captives Gain UCL  
A undermin adapted wide yogurtowitz XCOM WARM GREENHunter Transmission
```

----结束

6 Lite Server 资源管理

6.1 查看 Lite Server 服务器详情

在您创建了 Lite Server 节点后，可以通过管理控制台查看和管理您的 Lite Server 节点。本节介绍如何查看 Lite Server 节点的详细信息，包括名称/ID、规格、镜像等信息。

查看 Lite Server 普通节点详情

在轻量算力节点 (Lite Server) 的普通节点列表页中，可以查看 Lite Server 普通节点的名称、状态、实例规格、资源类型、监控、VPC、IP 地址、计费模式、创建时间和操作。

图 6-1 查看 Lite Server 普通节点



单击某个普通节点名称，进入到 Lite Server 普通节点详情页，可以查看更多信息，如表 6-1 所示。

表 6-1 详情页参数说明

参数名称	说明
名称	Lite Server 普通节点的名称。可以修改，修改操作请参见 修改 Lite Server 名称 。
实例规格	Lite Server 普通节点的规格。
ID	Lite Server 普通节点 ID，可用于在费用中心查询。
计费模式	Lite Server 普通节点当前的计费模式。
状态	Lite Server 普通节点的运行状态。

参数名称	说明
虚拟私有云	创建Lite Server普通节点时绑定的虚拟私有云，单击链接可跳转到虚拟私有云详情页。
裸金属服务器/弹性云服务器	Lite Server普通节点为一台裸金属服务器或弹性云服务器，单击链接可跳转至对应裸金属服务器或弹性云服务器的详情页。
镜像	Lite Server普通节点的操作系统镜像。可以切换或重置操作系统镜像，具体操作请参见 切换或重置Lite Server服务器操作系统 。
创建时间	Lite Server普通节点的创建时间。
更新时间	Lite Server普通节点的更新时间。
所属订单	Lite Server普通节点对应的订单，单击链接可跳转至费用中心。
访问密钥	访问密钥名称。
磁盘	显示Lite Server普通节点当前挂载的磁盘列表，单击“挂载磁盘”，在弹出窗口中进行挂载，具体操作请参见 配置Lite Server存储 。
监控	显示当前Lite Server普通节点的监控情况。单击“监控详情”，可以跳转至CES控制台查看具体监控情况。

查看 Lite Server 超节点详情

在轻量算力节点 (Lite Server)的超节点列表页中，可以查看Lite Server超节点的名称、状态、创建时间、计费模式、实例规格、核心硬件配置、VPC、IP地址和操作。

单击某个超节点名称，进入到Lite Server超节点详情页，可以查看更多信息，如[表6-2](#)所示。

表 6-2 详情页参数说明

参数名称	说明
名称	Lite Server超节点的名称。
实例规格	Lite Server超节点的规格。
ID	Lite Server超节点的ID，可用于在费用中心查询。
计费模式	Lite Server超节点当前的计费模式。
状态	Lite Server超节点的运行状态。
虚拟私有云	创建Lite Server超节点时绑定的虚拟私有云，单击链接可跳转到虚拟私有云详情页。
镜像	Lite Server超节点的操作系统镜像。可以切换或重置操作系统镜像，具体操作请参见 切换或重置Lite Server服务器操作系统 。
创建时间	Lite Server超节点的创建时间。

参数名称	说明
更新时间	Lite Server超节点的更新时间。
所属订单	Lite Server超节点对应的订单，单击链接可跳转至费用中心。
访问密钥	访问密钥名称

图 6-2 Lite Server 服务器详情



6.2 开机或关机 Lite Server 服务器

场景描述

当您暂时不需要使用 Lite Server 的时候，可以通过关机操作停止运行中的 Server 实例，停止对资源的消耗。当需要使用的時候，对于停止状态的 Lite Server，可以通过开机操作重新使用。

约束限制

- 只有处于“已停止/停止失败/启动失败”状态的 Lite Server 可以执行开机操作。
- 只有处于“运行中/停止失败”状态的 Lite Server 可以执行关机操作。
- 只有普通节点支持批量开机或关机操作，超节点不支持批量操作。

Lite Server 开机或关机操作

- 登录 [ModelArts 管理控制台](#)。
- 在左侧菜单栏中选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”列表页面。
- 在 Lite Server 普通节点列表中执行如下操作，对 Lite Server 进行开机或关机。支持操作单个节点也支持多个节点批量操作。
 - Lite Server 开机：单击右侧操作列的“开机”，只有处于“已停止/停止失败/启动失败”状态的 Lite Server 可以执行开机操作。
 - Lite Server 关机：单击右侧操作列的“关机”，在弹出的确认对话框中，确认信息无误，一键输入 Yes，然后单击“确定”。只有处于“运行中/停止失败”状态的 Lite Server 可以执行关机操作。

图 6-3 关机



说明

关机服务器为“强制关机”方式，会中断您的业务，请确保服务器上的文件已保存。
在普通节点列表页，勾选多个节点，可以批量进行开机或关机操作。

图 6-4 批量操作



6.3 同步 Lite Server 服务器状态

场景描述

Lite Server普通节点为一台弹性服务器ECS或裸金属服务器BMS，当用户在ECS或BMS侧修改了服务器状态或磁盘信息后，您可通过“同步”功能，将服务器状态和磁盘信息同步至ModelArts Lite Server。

Lite Server超节点的子节点也是ECS服务器，当用户在ECS侧修改了子节点的磁盘信息后，您也可通过“同步”功能，同步磁盘信息至ModelArts Lite Server。

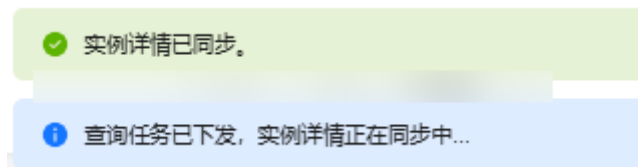
约束限制

- 支持普通节点（弹性服务器ECS或裸金属服务器BMS）服务器状态和磁盘信息同步。
- 对于超节点，不支持超节点的父节点服务器状态同步。
- 对于超节点，仅支持子节点的磁盘信息同步。
- 对于超节点的子节点，不支持在ECS侧修改服务器状态，比如重启、关机等操作。

同步 Lite Server 状态

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”或“超节点”列表页面。
3. 在Lite Server节点列表中，单击右侧操作列的“更多 > 同步”，系统会提示“查询任务已下发，实例详情同步中”。或者进入Lite Server节点详情页中，单击“同步”。

当系统提示“实例详情已同步”时，表示同步完成。



6.4 切换或重置 Lite Server 服务器操作系统

场景描述

当Lite Server的操作系统镜像不满足要求时，可以进行切换或重置。本文介绍以下几种切换操作系统的方式：

- 在ModelArts控制台的Lite Server页面切换或重置操作系统（推荐）
- 在裸金属服务器控制台切换操作系统
- 使用BMS Go SDK的方式切换裸金属服务器操作系统
- 使用Python封装API的方式切换裸金属服务器操作系统

约束限制

- 节点状态要求：Lite Server节点状态只有处于“已停止”、“重置OS失败”、“切换OS失败”时，才可以切换或重置操作系统。否则可能会导致操作失败，因为系统盘无法被卸载，导致无限循环卸载盘。
- 操作系统要求：目标操作系统必须是该Region下的IMS公共镜像或者私有共享镜像。
- 只有普通节点支持批量切换或重置操作系统操作，超节点不支持批量操作。

操作影响

重置或切换Lite Server节点操作系统的影响如下：

1. 系统盘ID变化：切换或重置操作系统后，EVS系统盘ID会变化，和下单时订单中的EVS ID已经不一致，导致无法进行EVS系统盘扩容操作。系统会提示“当前订单已到期，无法进行扩容操作，请续订”。
2. userdata配置影响：切换操作系统时，userdata的注入可能不会生效，特别是在configdriver模式下。客户需要确保在创建节点时传入userdata参数，或者在切换后手动配置必要的设置。因此切换或者重置操作系统后，建议通过挂载数据盘EVS或挂载SFS盘等方式进行存储扩容。
3. 应用和模型影响：切换操作系统可能影响已部署的应用或模型，因为依赖的软件包或库可能需要重新安装或配置。用户需要重新配置必要的依赖项以确保应用正常运行。
4. 裸金属服务器风险：对于裸金属服务器，升级操作系统内核或驱动可能导致不兼容，影响系统启动或基本功能。如果需要升级，请联系云服务商确认。

用户在进行切换或重置操作系统操作前，应确保节点处于关机状态，检查当前配置，备份重要数据，并在必要时联系技术支持以确认操作的可行性。

在 ModelArts 控制台的 Server 页面切换或重置操作系统

1. 登录**ModelArts管理控制台**。
2. 在左侧菜单栏中选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“资源列表 > 普通节点”或“资源列表 > 超节点”页面。
3. 在Lite Server列表中，单击右侧操作列的“更多 > 切换操作系统”或“更多 > 重置操作系统”，在弹出的确认对话框中，确认信息无误，然后单击“确定”，完成操作。

此时Lite Server普通节点或超节点的状态显示“切换操作系统中”或“重置操作系统中”。

在普通节点列表页，勾选多个节点，可以批量执行“切换操作系统”或“重置操作系统”操作。

在 BMS 控制台切换操作系统

1. 获取操作系统镜像。

由云服务官方提供给客户的操作系统镜像，在IMS镜像服务的共享镜像处进行接收即可，参考如下图操作。

图 6-5 共享镜像

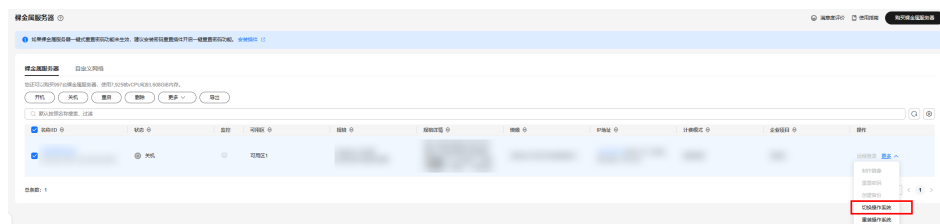


2. 切换操作系统。

对Lite Server资源对应的裸金属服务器，对其进行关机操作，完成关机后，才可以执行切换操作系统操作。

在裸金属服务的更多选项中，单击切换操作系统，如下图所示。

图 6-6 切换操作系统



在切换操作系统界面，选择上一步接收到的共享镜像即可。

使用 BMS Go SDK 的方式切换操作系统

以下为BMS使用Go语言通过SDK方式切换操作系统的示例代码。

```
package main

import (
    "fmt"
    "os"
    "github.com/huaweicloud/huaweicloud-sdk-go-v3/core/auth/basic"
    bms "github.com/huaweicloud/huaweicloud-sdk-go-v3/services/bms/v1"
    "github.com/huaweicloud/huaweicloud-sdk-go-v3/services/bms/v1/model"
    region "github.com/huaweicloud/huaweicloud-sdk-go-v3/services/bms/v1/region"
)

func main() {
    // 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；
    // 本示例以ak和sk保存在环境变量中来实现身份验证为例，运行本示例前请先在本地环境中设置环境变量HUAWEICLOUD_SDK_AK和HUAWEICLOUD_SDK_SK。
    ak := os.Getenv("HUAWEICLOUD_SDK_AK")
    sk := os.Getenv("HUAWEICLOUD_SDK_SK")

    auth := basic.NewCredentialsBuilder().
        WithAk(ak).
        WithSk(sk).
        Build()

    client := bms.NewBmsClient(
        bms.BmsClientBuilder().
            WithRegion(region.ValueOf("cn-north-4")).
            WithCredential(auth).
            Build())
    keyname := "KeyPair-name"
    userdata := "aGVsbG8gd29ybGQsIHdlbGNvbWUgdG8gam9pbiB0aGUy29uZmVyZW5jZQ=="
    request := &model.ChangeBaremetalServerOsRequest{
        ServerId: "****input your bms instance id****",
        Body: &model.OsChangeReq{
            OsChange: &model.OsChange{
                Keyname: &keyname,
                Imageid: "****input your ims image id****",
                Metadata: &model.MetadataInstall{
                    UserData: &userdata,
                },
            },
        },
    },
    }

    response, err := client.ChangeBaremetalServerOs(request)
    if err == nil {
        fmt.Printf("%+v\n", response)
    } else {
        fmt.Println(err)
    }
}
```

```
}  
}
```

Python 封装 API 方式切换 BMS 操作系统

以下为BMS使用Python语言通过API方式切换操作系统的示例代码。

```
# -*- coding: UTF-8 -*-  
  
import requests  
import json  
import time  
import requests.packages.urllib3.exceptions  
from urllib3.exceptions import InsecureRequestWarning  
  
requests.packages.urllib3.disable_warnings(InsecureRequestWarning)  
class ServerOperation(object):  
  
    ##### IAM认证  
    API#####  
  
    def __init__(self, account, password, region_name, username=None, project_id=None):  
        """  
        :param username: if IAM user,here is small user, else big user  
        :param account: account big big user  
        :param password: account  
        :param region_name:  
        """  
        self.account = account  
        self.username = username  
        self.password = password  
        self.region_name = region_name  
        self.project_id = project_id  
        self.ma_endpoint = "https://modelarts.{}/myhuaweicloud.com".format(region_name)  
        self.service_endpoint = "https://bms.{}/myhuaweicloud.com".format(region_name)  
        self.iam_endpoint = "https://iam.{}/myhuaweicloud.com".format(region_name)  
        self.headers = {"Content-Type": "application/json",  
                        "X-Auth-Token": self.get_project_token_by_account(self.iam_endpoint)}  
  
    def get_project_token_by_account(self, iam_endpoint):  
        body = {  
            "auth": {  
                "identity": {  
                    "methods": [  
                        "password"  
                    ],  
                    "password": {  
                        "user": {  
                            "name": self.username if self.username else self.account,  
                            "password": self.password,  
                            "domain": {  
                                "name": self.account  
                            }  
                        }  
                    }  
                },  
                "scope": {  
                    "project": {  
                        "name": self.region_name  
                    }  
                }  
            }  
        }  
        headers = {  
            "Content-Type": "application/json"  
        }  
        import json  
        url = iam_endpoint + "/v3/auth/tokens"  
        response = requests.post(url, headers=headers, data=json.dumps(body), verify=True)
```

```
token = (response.headers['X-Subject-Token'])
return token
def change_os(self, server_id):
    url = "{}/v1/{}/baremetalservers/{}/changeos".format(self.service_endpoint, self.project_id, server_id)
    print(url)
    body = {
        "os-change": {
            "adminpass": "@Server",
            "imageid": "40d88eea-6e41-418a-ad6c-c177fe1876b8"
        }
    }
    response = requests.post(url, headers=self.headers, data=json.dumps(body), verify=False)
    print(json.dumps(response.json(), indent=1))
    return response.json()

if __name__ == '__main__':
    # 调用API前置准备, 初始化认证鉴权信息
    server = ServerOperation(username="xxx",
                             account="xxx",
                             password="xxx",
                             project_id="xxx",
                             region_name="cn-north-4")

    server.change_os(server_id="0c84bb62-35bd-4e1c-ba08-a3a686bc5097")
```

6.5 制作 Lite Server 服务器操作系统

场景描述

当前Lite Server服务器操作系统不满足用户诉求时, 您可以使用BMS或ECS的制作镜像功能, 将当前操作系统保存为新的镜像, 方便用于其它Lite Server。

约束限制

制作镜像需满足以下条件: 当前Lite Server服务器状态为停止状态。

制作的镜像仅支持基于Lite Server当前的操作系统制作新的镜像, 不支持其他场景制作镜像, 例如从ISO开始制作等。

制作操作系统步骤

1. 制作操作系统镜像前需要先清理一些临时文件, 否则会导致镜像运行故障。登录Server服务器中, 清理操作系统中临时文件。可以执行以下命令, 也可以制作成脚本批量运行。清理脚本参考[临时文件清理脚本](#)。
 - a. 执行下面命令, 清理用户登录记录。

```
echo > /var/log/wtmp
echo > /var/log/btmp
```
 - b. 执行下面命令, 清理相应目录下的临时文件。

```
rm -rf /var/log/cloud-init*
rm -rf /var/lib/cloud/*
rm -rf /var/log/network-config.log
rm -rf /opt/huawei/network_config/network_config.json
rm -rf /opt/huawei/port_config/uplink_hash_config.log
rm -rf /opt/huawei/firmware_check/firmware_check.log
```
 - c. 执行下面命令, 清理残留配置信息。
 - CentOS/EulerOS/HCE操作系统: 查看 “/etc/sysconfig/network-scripts/” 文件夹下有哪些以 “ifcfg” 开头的文件, 删除 “ifcfg-lo” 以外的以 “ifcfg” 开头的文件, 以 “ifcfg-lo” 开头的文件不删除。

查看文件命令：

```
ll /etc/sysconfig/network-scripts/
```

删除文件命令：

```
rm -rf /etc/sysconfig/network-scripts/ifcfgxxx
```

■ Ubuntu操作系统：

```
rm -rf /etc/network/interfaces.d/50-cloud-init.cfg
```

d. 执行下面命令清除参数面网络信息。

```
echo >/etc/netplan/roce.yaml
```

```
echo > /etc/hccn.conf
```

e. 执行下面命令清除历史操作记录。

```
history -w;echo > /root/.bash_history;history -c;history -c;history -c;
```

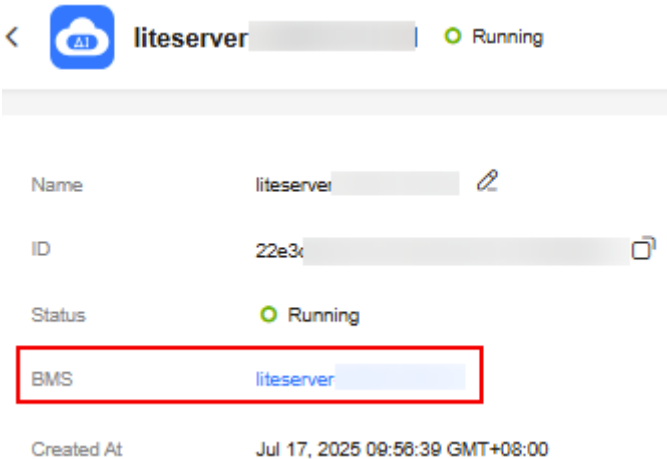
2. 将Lite Server服务器关机。

登录[ModelArts管理控制台](#)，在左侧菜单栏中选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”列表页面，单击右侧操作列的“关机”，对 Lite Server资源对应的服务器，执行关机操作。

3. 在Lite Server服务器详情页，通过裸金属服务器地址或弹性云服务器地址跳转到对应控制台详情页。

图 6-7 Lite Server 服务器详情页





4. 在裸金属服务器或ECS服务器控制台，返回到服务器列表页，在操作列中执行制作镜像操作。

图 6-8 ECS 服务器中创建镜像

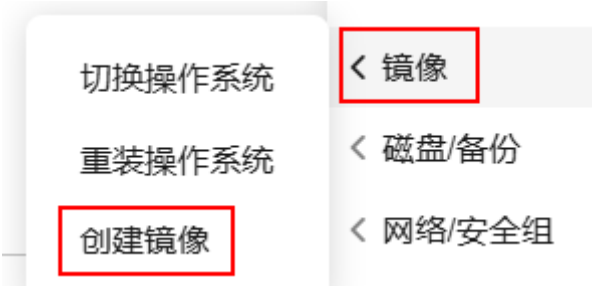


图 6-9 裸金属服务器制作镜像



在制作镜像界面，填入制作镜像的名称、企业项目，并勾选协议，单击下一步即可制作镜像，制作成功的镜像会保存在租户的IMS镜像服务的私有镜像列表中。详细操作请分别参见[ECS制作镜像](#)和[BMS制作镜像](#)文档。

制作完的镜像后续可以用于其他Lite Server，具体可以参考[切换或重置Lite Server服务器操作系统](#)。

临时文件清理脚本

```
#!/bin/bash

# 执行下面命令，清理用户登录记录。
echo > /var/log/wtmp
echo > /var/log/btmp

# 执行下面命令，清理相应目录下的临时文件。
rm -rf /var/log/cloud-init*
rm -rf /var/lib/cloud/*
rm -rf /var/log/network-config.log
rm -rf /opt/huawei/network_config/network_config.json
rm -rf /opt/huawei/port_config/uplink_hash_config.log
rm -rf /opt/huawei/firmware_check/firmware_check.log
```

```
# 执行下面命令，清理残留配置信息。
## CentOS/EulerOS/HCE操作系统：
## 查看“/etc/sysconfig/network-scripts/”文件夹下有哪些以“ifcfg”开头的文件，删除“ifcfg-lo”以外的以
“ifcfg”开头的文件。以“ifcfg-lo”开头的文件不删除。
find /etc/sysconfig/network-scripts/ -name 'ifcfg*' ! -name 'ifcfg-lo' -type f | xargs rm -f
## Ubuntu操作系统：
rm -rf /etc/network/interfaces.d/50-cloud-init.cfg

# 执行下面命令清除参数面网络信息。
echo > /etc/netplan/roce.yaml
echo > /etc/hccn.conf

# 执行下面命令清除历史操作记录。
history -w;echo > /root/.bash_history;history -c;history -c;history -c;
echo > ~/.bash_history
exec bash
```

6.6 Lite Server 资源热备管理

场景描述

Lite Server资源热备需要用户自建k8s集群。对于k8s集群中的机器资源通过打污点的方式，完成资源热备机的处理，从而使业务Pod无法调度到该热备机上。

约束限制

根据下单使用的Lite Server资源机器种类和台数不同，推荐您按照下面的表格进行热备机器台数准备。

表 6-3 热备机器数量

资源类型 资源台数	小于10 台	10 台-49 台	50 台-99 台	100 台-249 台	250 台-499 台	500 台-749 台	750 台-100 0台	1000 台以上
Snt9a	0	1	2	3	5	7	10	12
Snt9b	0	1	2	3	4	5	6	10
GP Ant8	0	1	2	3	5	6	8	12
GP Vnt1	0	1	2	3	5	8	10	12

示例1：

当购买Snt9b类型的资源台数为6台时，该台数少于10台，因此根据热备机推荐表，此时不需要准备热备机，正常按照6台进行资源购买即可。

示例2：

当购买Snt9b类型的资源台数为600台时，该台数位于500至749台区间，因此根据热备机推荐表，此时需要额外准备热备机5台，因此需要按照605台进行资源购买。

前提条件

已购买的Lite Server资源中用户已自建k8s集群。

资源热备替换操作

当集群中的业务节点发生硬件故障等需要进行热备替换时，先进行数据备份，再进行故障机打污点和删除热备机污点的方式完成资源热备替换。

1. 数据备份

推荐使用rsync工具进行文件备份，rsync是一个强大的文件同步工具，支持本地和远程同步，该工具可以灵活和高效地完成数据备份。

```
rsync -avz -e ssh /source/ user@remote:/destination/
```

将故障机文件备份到热备机上，以backup.txt文件为例。

```
root@cce-456064-nodpool-69256-ttlgp:~# rsync -avz -e ssh back_up.txt root@15:/root/
The authenticity of host '15 (15)' can't be established.
ED25519 key fingerprint is SHA256:yeIIaSaobGp87008UR+yA01d18fphX46mFfC12MD164.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '15' (ED25519) to the list of known hosts.
root@15:~#
root@15:~# rsync -avz -e ssh back_up.txt root@15:/root/
sending incremental file list
back_up.txt

sent 150 bytes  received 35 bytes  12.76 bytes/sec
total size is 45  speedup is 0.24
```

以SSH的方式完成备份，备份完成后，可以在新的热备机上查看到该文件。

```
root@cce-456064-nodpool-69256-ucoi9:~# ls
back_up.txt  check_env.sh  disk_filter.sh  print_log.sh  snap
root@cce-456064-nodpool-69256-ucoi9:~# _
```

2. 热备替换

对发生故障的节点打污点

```
kubectl taint nodes <node-name> dedicated=ops:NoSchedule
```

- <node-name>：替换为实际节点名称。
- dedicated=ops：污点的键值对。
- NoSchedule：污点的效果，表示kube-scheduler不会将Pod调度到该节点。

确认节点污点标签是否成功。

```
kubectl describe node <node-name> | grep Taints
```

```
user@oth zlw-machine:~$ kubectl describe node 1 | grep Taints
Taints:              dedicated=ops:NoSchedule
user@oth zlw-machine:~$
```

可以看到该节点已成功打上不可调度的标签。

对进行热备替换的机器去除污点。

```
kubectl taint node <node-name> dedicated=ops:NoSchedule-
```

```
user@oth57cnzlw-machine:~$ kubectl taint nodes 1 | grep Taints
node/1:              .115 untainted
```

再次确认污点已去除

```
kubectl describe node <node-name> | grep Taints
```

```
user@ot 3zlz-machine:~$ kubectl describe node 1 | grep Taints
Taints:              <none>
```

至此已完成机器的热备替换，替换下来的故障机可以进行进入正常的维修流程。

6.7 修改 Lite Server 名称

场景描述

在日常的云服务管理中，用户经常需要对云服务器进行命名或重命名，以便更好地管理和识别不同的实例。为了提升用户体验，允许用户在 Lite Server 中直接修改名称。

- 支持在 Lite Server 的“普通节点”列表页或详情页中修改普通节点名称，并且允许重名。
- 支持在 Lite Server 的“超节点”列表页或详情页中修改子节点名称，并且允许重名。

约束限制

- 订单中的服务器名称会一直保持下单购买时设置的名称。后期修改服务器名称后，不会在订单中同步更新。
- Lite Server 的资源类型为裸金属服务器 BMS、弹性云服务器 ECS 时，在 BMS 和 ECS 侧修改服务器名称后，不会自动同步到 Lite Server，需要手动同步，具体操作参见[同步 Lite Server 服务器状态](#)。
- Lite Server 的资源类型为超节点时，仅支持修改子节点服务器的名称。超节点本身的名称不支持修改。
- 当 Lite Server 状态为“运行中”或“停止”时，允许修改服务器名称，其他状态不支持修改服务器名称。
- Lite Server 名称不能为空。
- 未勾选“允许重名”按钮时，如果输入的名称已存在，系统将提示“云服务已被使用”。

修改 Lite Server 服务器名称

1. 登录[ModelArts 管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”或“超节点”列表页面。

单击列表页中节点名称右侧的  修改名称，修改完单击“确定”，立即生效。

图 6-10 修改节点名称



也可以进入节点详情页，单击节点名称右侧的  修改，修改完单击“确定”，立即生效。

图 6-11 修改节点名称



修改名称时，勾选“允许重名”后，才允许有相同的节点名称存在。

6.8 授权修复 Lite Server 节点

场景描述

当 Lite Server 节点由于不可恢复故障需要进行硬件维护时，会推送计划事件到控制台的事件中心。您可以在事件中心，查看具体的事件信息、事件类型、事件状态、事件描述等信息，可以授权华为技术支持对故障节点进行运维或重部署节点。

表 6-4 事件操作执行条件

事件类型	事件状态	可执行的操作	适用的资源类型	说明
系统维护	待授权	授权、重部署	Snt9b	系统维护是授权华为技术支持对故障节点进行系统性维护。
本地盘恢复	待授权	授权、重部署	Snt9b	本地盘恢复是授权华为技术支持对故障的本地盘进行维护。 警告 授权后超节点本地盘恢复操作将会导致本地盘数据丢失，授权前请先迁移业务和备份数据。

事件类型	事件状态	可执行的操作	适用的资源类型	说明
超节点维护	待授权	授权	Snt9b23	超节点维护是授权华为技术支持对故障节点通过人工修理、更换器件等方式恢复故障节点。
超节点重部署	待授权	授权	Snt9b23	超节点重部署是授权华为运维系统通过自动更换节点的方式恢复故障节点，恢复后的节点除物理设备信息发生变化外，节点名称、节点ID、IP地址等信息与原节点保持一致。
超节点本地盘恢复	待授权	授权	Snt9b23	超节点本地盘恢复是授权华为技术支持对超节点的本地盘进行恢复。 警告 授权后超节点本地盘恢复操作将会导致本地盘数据丢失，授权前请先迁移业务和备份数据。

- 授权：授权操作是授权华为技术支持针对故障的节点进行点对点修复硬件，修复周期长。
- 重部署：重部署操作是授权华为技术支持对发生故障的节点进行整机替换，恢复快，但是重部署后本地盘数据将会丢失数据。请谨慎操作。重部署前请先迁移业务和备份数据。

约束限制

- 仅Snt9b和超节点Snt9b23支持通过计划事件发起硬件维护。
- 超节点重部署需要在物理超节点内操作。当超节点达到满配48台时，不支持重部署操作，操作授权按钮为置灰状态。
- 如果计划事件不满足表6-4所示的事件状态，操作授权按钮为置灰状态。
- 授权“超节点重部署”事件前，您需要先在“轻量算力节点 (Lite Server) 页面”停止 Lite Server 实例，否则会授权失败。事件执行完成后，再重新启动 Server 实例。
- 授权节点将影响相关业务的运行，请谨慎操作。当事件类型为超节点重部署，且节点处于关机状态时，才可执行授权操作。
- 节点本地盘恢复和超节点本地盘恢复操作将会导致本地盘数据丢失，授权前请先迁移业务和备份数据。本地盘恢复后需要登录到 Lite Server 节点内完成本地盘分区。

查看计划事件

登录[ModelArts管理控制台](#)。在左侧导航栏单击“事件中心”，在事件中心页面可以查看事件的详细信息。默认显示处于**待授权**、**已授权**、**执行中**的事件。去除筛选条件可以查看所有状态的事件。

表 6-5 计划事件说明

属性	说明	示例
事件ID	事件的唯一标识。	5ad1df12-e3d2-4f36-b367-xxxxxxxxxxxx
节点名称/ID	发起事件的Server节点名称和服务器ID。	devserver-dd501e0d95ad-5a9f-46e3-9ba6-c5f8fcxxxx
事件类型	事件类型具体参见 表6-4 。	超节点重部署
事件状态	- 待授权：问询中，等待您授权，授权后会进入已授权状态。 - 已授权：计划执行运维任务，但尚未开始执行，开始执行后会进入执行中状态。 - 执行中：正在执行运维任务中。 - 已完成：运维任务已经执行完成。 - 失败：运维任务执行失败。 - 取消：系统取消了运维任务。	待授权
事件描述	描述产生该事件的具体原因。	底层硬件故障，当前通过CAR自动接入： alarmName=XX XX,bmcip=2409:27ff:1003:0103:0011:0000:0000:xxxx,componentName=XXXX
创建时间	事件创建的时间。	2025/02/19 16:05:32 GMT+08:00
执行时间	事件进入调度执行阶段的时间。	2025/03/03 16:23:16 GMT+08:00

属性	说明	示例
操作	授权：授权节点将影响相关业务的运行，请谨慎操作。当事件类型为超节点重部署，且节点处于关机状态时，才可执行授权操作。 说明 超节点重部署需要在物理超节点内操作。当超节点达到满配48台时，不支持重部署操作，操作授权按钮为置灰状态。	--

授权操作

当故障节点满足如表6-4所示的条件时，可通过授权操作授权华为技术支持对故障节点进行运维。

您可在ModelArts控制台“资源管理 > 事件中心”页面，找到对应节点，在操作列单击“授权”，在弹出的提示框中单击“确认”即可完成授权。以下步骤以超节点维护为例，介绍授权操作。

1. 登录ModelArts管理控制台，在左侧导航栏单击“事件中心”，进入“事件中心”页面，查看“事件类型”为“超节点维护”的事件，执行“授权”操作。

图 6-12 执行授权操作



2. 超节点维护事件进入“已授权”状态。

图 6-13 事件进入“已授权”状态



3. 待完成超节点维修后，事件状态显示为“已完成”。此时，节点已处于可用状态。
- 在完成运维操作后，华为技术支持会主动关闭已获得授权，无需您额外操作。
- 如果是“本地盘恢复”或“超节点本地盘恢复”，恢复后需要登录到Lite Server节点内完成本地盘分区。

重部署操作

重部署后本地盘数据将会丢失，请谨慎操作。重部署前请先迁移业务和备份数据，并完成实例重部署前的本地盘预处理操作，具体操作请参见[实例重部署预处理](#)。

当故障节点满足如[表6-4](#)所示的重部署操作执行条件时，您可在控制台“资源管理 > 事件中心”页面，找到对应节点，在操作列单击“重部署”，在弹窗中输入“YES”，单击“确认”即可完成重部署的授权。

如果计划事件不满足如[表6-5](#)所示的重部署操作执行条件，操作重部署按钮为置灰状态。

在完成运维操作后，华为云技术支持会主动关闭已获得授权，无需您额外操作。

6.9 重启 Lite Server 服务器

场景描述

当用户需要重启 Lite Server 服务器时，Lite Server 页面提供了服务器“重启”功能，方便用户操作。

约束限制

- 只有处于“运行中”、“重启失败”状态时，才可进行重启操作。
- 普通节点和超节点均支持重启操作。但只有普通节点支持批量重启操作，超节点不支持批量操作。

重启 Lite Server 服务器操作

- 登录[ModelArts管理控制台](#)。
- 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”或“超节点”列表页面。
- 支持以下两种重启方式。
 - 方式一：在 Lite Server 的普通节点或“超节点”列表中，单击待重启节点右侧操作列的“更多 > 重启”，在右侧弹窗中确认信息后，单击“确定”。

图 6-14 重启 Server 节点

重启

⚠ 重启节点将导致业务短暂中断，请在业务低峰期进行重启，避免对高峰期业务造成影响。

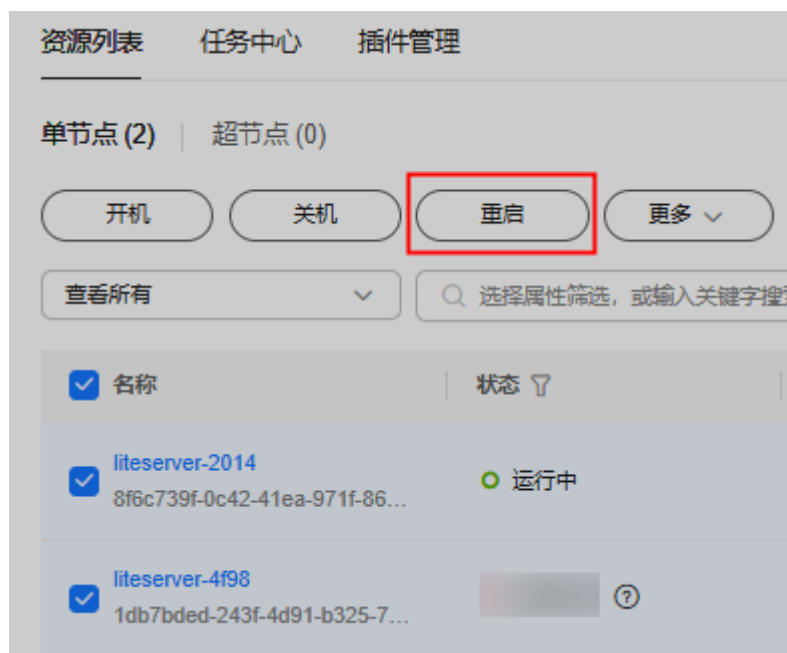
可执行操作资源清单

确定要对以下 1 台节点服务器进行重启操作吗？

名称/ID	状态	实例规格	计费模式
liteserver-2014	运行中	General computin...	

在普通节点列表页，勾选多个节点，可以批量进行重启操作。

图 6-15 批量重启操作



- 方式二：单击Lite Server服务器名称，进入Lite Server详情页，在页面右上角单击“重启”，在弹窗中确认信息后，单击“确定”。

7 Lite Server 插件管理

7.1 安装 Lite Server AI 插件

场景描述

节点任务中枢（NodeTaskHub）是深度集成的弹性节点管理插件，为ModelArts Lite Server节点提供批量任务下发与自动化运维能力。支持昇腾软件升级、实时检测、故障诊断等高频操作，降低人工干预风险，保障AI业务流程稳定高效。

Lite Server任务中心提供多种任务模板供用户创建任务，任务下发依赖Lite Server节点中已安装的NodeTaskHub插件。Lite Server的部分公共镜像中预置了NodeTaskHub插件，在购买Lite Server时可以选择自动安装该插件。如果未安装，可以参考本文手动安装NodeTaskHub插件。

约束限制

- 当前仅支持资源为Snt9b的普通节点和Snt9b23超节点。
- 节点的状态需要为运行中。
- 插件依赖Docker服务，请确保节点中已安装Docker环境。Lite Server的公共OS镜像中均已安装Docker环境。用户自定义镜像中未安装Docker环境，可以参考[安装Docker环境](#)操作。
- 插件将占用节点的25317端口。

检查插件安装情况

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点（Lite Server）> 插件管理”，进入插件管理页面。

图 7-1 插件管理



3. 查看插件安装情况，如果未安装，单击右侧操作里的安装，根据界面提示安装插件。也可以参考[手动安装插件](#)章节操作。

图 7-2 插件状态

资源列表

任务中心

插件管理

重启

卸载

升级

操作记录

选择属性筛选, 或输入关键字搜索

Q

⊗

节点名称/ID	机型	节点状态	插件名称	插件状态	插件版本	操作
☐ liteserver- a10b39ec-7117-44...		运行中	NodeTaskHub	运行中	1.1.0	升级 卸载 重启
☐ liteserver- 943d20a5-10c1-4c...		运行中	--	未安装	--	安装

手动安装插件

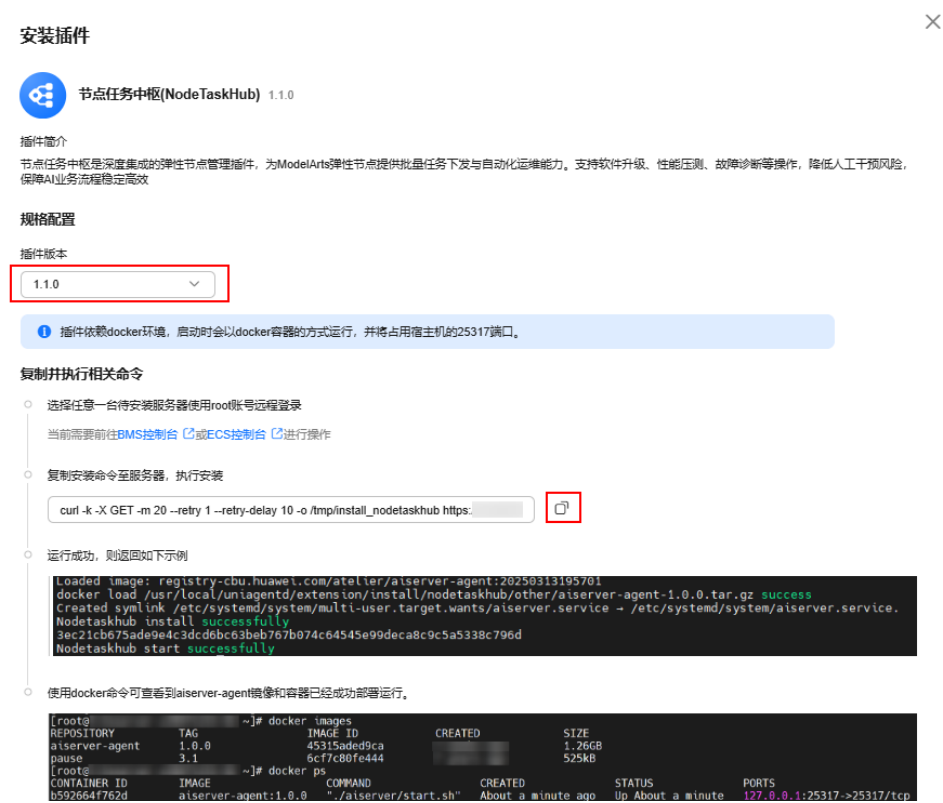
1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择插件广场，查找“节点任务中枢(NodeTaskHub)”插件，单击“安装”。

图 7-3 查找插件



3. 根据安装插件界面信息提示，选择插件版本，并复制一键安装命令到对应的Lite Server节点中执行，即可实现NodeTaskHub插件的安装。

图 7-4 安装插件



4. 当插件安装成功后，返回插件管理界面，可以看到插件状态为“运行中”。

相关操作

在“资源管理 > 轻量算力节点（ Lite Server ） > 插件管理”页面，可以对插件进行重启、卸载和升级操作。

后续操作

插件安装完成后，即可在Lite Server的任务中心开展以下任务：

- [升级Lite Server中的Ascend驱动固件版本](#)
- [安装/升级Lite Server中的CES Agent插件](#)
- [Lite Server节点故障诊断](#)
- [Lite Server节点漏洞修复](#)
- [Lite Server节点一键式压测](#)
- [Lite Server节点参数面网络配置](#)

7.2 升级 Lite Server 中的 Ascend 驱动固件版本

场景描述

本文旨在指导用户如何通过Lite Server的任务中心下发Ascend驱动固件升级任务，通过该任务可以在Snt9b和Snt9b23机器上完成Ascend驱动固件的一键升级。

约束限制

- 当前仅支持Snt9b普通节点和超节点Snt9b23。
- 升级驱动固件过程中会导致业务中断，升级前请保证节点内无业务运行，同时升级完毕后需要重启节点生效。
- 如果节点内驱动固件版本为官方维护版本，升级失败支持回滚至节点内驱动固件原始版本，如果节点内驱动固件损坏，或者节点内驱动固件版本为非官方维护版本，会导致查询节点内驱动固件失败，该场景下升级任务仍可下发，但如果升级失败无法回滚，需要联系华为运维工程师处理。
- 驱动固件与昇腾软件包（CANN/MindSpore等）有兼容性关系，请确保升级后的驱动固件版本与业务中使用的昇腾软件包的兼容性，可参考[表7-1](#)确认组件兼容性。

表 7-1 组件兼容性

CANN版本	配套Ascend HDK版本
CANN 8.0.RC3	Ascend HDK 24.1.RC3 Ascend HDK 24.1.RC2 Ascend HDK 24.1.RC1 Ascend HDK 23.0.0/23.0.X
CANN 8.0.0	Ascend HDK 24.1.0 Ascend HDK 24.1.RC3 Ascend HDK 24.1.RC2 Ascend HDK 24.1.RC1 Ascend HDK 23.0.0/23.0.X
CANN 8.1.RC1	Ascend HDK 25.0.RC1 Ascend HDK 24.1.0 Ascend HDK 24.1.RC3 Ascend HDK 24.1.RC2 Ascend HDK 24.1.RC1 Ascend HDK 23.0.X
CANN 8.2.RC1	Ascend HDK 25.2.0 Ascend HDK 25.0.RC1 Ascend HDK 24.1.0 Ascend HDK 24.1.RC3 Ascend HDK 24.1.RC2

前提条件

该操作依赖在节点上预安装Lite Server AI插件，请通过[安装Lite Server AI插件](#)章节完成插件安装。

操作步骤

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“任务中心”。

图 7-5 任务中心



3. 单击任务中心页面左上角的“创建任务”，进入“任务模板”页面，在该页面选择“Ascend软件升级”，单击“创建任务”。

图 7-6 任务模板



4. 在Ascend软件升级创建页面，填写“任务名称”、“任务描述”，选择“任务类型”和“机型”，单击“选择节点”，在右侧节点列表弹窗中勾选节点后单击"确认"，该操作会在相应节点下发驱动固件版本查询任务，大约耗时一分钟，以便获取真实的驱动固件信息。

表 7-2 创建任务参数

参数名称	参数说明
任务名称	系统自动填入任务名称，用户可以自定义。
任务描述	对该任务的描述信息，方便快速查找任务。
任务类型	勾选“HDK升级”。

参数名称	参数说明
机型	支持Snt9b节点和超节点Snt9b23。
选择节点	单击“选择节点”，在右侧弹出的节点列表中选择需要升级驱动固件的节点，支持批量选择，也可以通过关键字搜索，之后单击“确定”。 该操作会在选择的节点内下发一个查询任务，查询节点内驱动固件版本和CANN信息。 等待查询结果刷新，该过程大约耗时一分钟。
选择驱动固件版本	在下拉框中选择待升级的目标驱动固件版本。 请参考表7-1确认组件兼容性，避免升级失败导致业务中断。也可以参考Lite Server节点故障诊断章节，下发Ascend设备诊断任务，该任务会自动诊断驱动固件与CANN的兼容性。

5. 选择待升级的驱动固件版本后，单击“下一步”，确认升级信息，选择升级后自动或手动重启，单击“确认”创建，下发升级任务。升级任务下发后，大约需要十分钟完成整个升级过程。
6. 升级过程中，返回“任务中心”页面，查看任务的执行状态。单击具体的任务名称，可以进入任务详情页，查看任务的详细信息和日志。
7. 升级成功后需要重启生效，如果选择手动重启，请在节点执行reboot操作，重启操作大约需要十分钟。
8. 在节点执行命令查看驱动是否加载成功，如果返回如下信息则加载成功，否则请联系华为工程师处理。

npu-smi info

图 7-7 查看驱动是否加载成功

```
root@node1 /]# npu-smi info
```

npu-smi		Version:			
NPU Chip	Name	Health Bus-Id	Power(W) AICore(%)	Temp(C) Memory-Usage(MB)	Hugepages-Usage(page) HBM-Usage(MB)
0		OK	94.3	40	0 / 0
0		0000:C1:00.0	0	0 / 0	3343 / 65536
1		OK	95.2	42	0 / 0
0		0000:01:00.0	0	0 / 0	3338 / 65536
2		OK	96.0	43	0 / 0
0		0000:C2:00.0	0	0 / 0	3338 / 65536
3		OK	94.9	41	0 / 0
0		0000:02:00.0	0	0 / 0	3338 / 65536
4		OK	96.7	43	0 / 0
0		0000:81:00.0	0	0 / 0	3338 / 65536
5		OK	93.2	44	0 / 0
0		0000:41:00.0	0	0 / 0	3338 / 65536
6		OK	93.7	42	0 / 0
0		0000:82:00.0	0	0 / 0	3338 / 65536
7		OK	95.6	43	0 / 0
0		0000:42:00.0	0	0 / 0	3338 / 65536

7.3 安装/升级 Lite Server 中的 CES Agent 插件

场景描述

在 Lite Server 的运维管理中，当用户发现当前操作系统镜像中预置的 CES Agent 版本过低，影响了 CES 服务的正常使用时，会面临无法通过现有手段进行升级的问题。传统的升级方式，如重新制作 OS 镜像或手动登录机器执行命令，不仅操作复杂、耗时长，还存在较高的误操作风险，无法满足用户对于高效、便捷运维的需求。如何在不中断业务的前提下，实现 CES Agent 的快速、批量升级，成为了亟待解决的问题。

为此，Lite Server 引入了 CES Agent 安装和升级功能，通过 Lite Server 平台，用户可以轻松发起 CES Agent 的批量安装或升级任务，不仅支持指定版本的安装与升级，还大幅降低了人工干预的风险，提升了运维效率和系统的稳定性。

约束限制

当前仅支持 Snt9b 节点和超节点 Snt9b23 中的 CES Agent 批量安装或升级。

前提条件

该操作依赖在节点上预安装 Lite Server AI 插件，请通过[安装 Lite Server AI 插件](#)章节完成插件安装。

操作步骤

1. 登录[ModelArts 管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“任务中心”。

图 7-8 任务中心



3. 单击任务中心页面左上角的“创建任务”，进入“任务模板”页面，在该页面选择“Ascend 软件升级”，单击“创建任务”。

图 7-9 任务模板



4. 在Ascend软件升级创建页面，填写“任务名称”、“任务描述”，选择“任务类型”和“机型”，单击“选择节点”，在右侧节点列表弹窗中勾选节点后单击“确认”，该操作会在相应节点下发版本查询任务，以便获取真实的CES-Agent的版本信息。

表 7-3 创建任务参数

参数分类	参数说明
任务名称	系统自动填入任务名称，用户可以自定义。
任务描述	对该任务的描述信息，方便快速查找任务。
任务类型	升级任务勾选“CES-Agent升级”卡片。
机型	支持Snt9b节点和超节点Snt9b23。
选择节点	单击“选择节点”，在右侧弹出的节点列表中，选择需要升级CES-Agent的节点。支持批量选择，也可以通过关键字搜索节点。 单击“确定”，该操作会在选择的节点内下发一个查询任务，查询节点内CES-Agent的版本。 等待查询结果刷新，该过程大约耗时1~2分钟。
选择CES-Agent版本	在下拉框中选择待升级的目标CES-Agent版本。

5. 选择待升级的CES-Agent版本后，单击“下一步”，确认升级信息，单击“确认”创建，下发升级任务。升级任务下发后，大约需要五分钟完成整个升级过程。
6. 升级过程中，返回“任务中心”页面，查看任务的执行状态。单击具体的任务名称，可以进入任务详情页，查看任务的详细信息和日志。
- 任务的执行状态为“成功”时，表示CES Agent安装或升级成功。

相关操作

CES Agent安装或升级成功后，可以监控Lite Server资源，具体请参考文档[使用CES监控Lite Server NPU资源](#)。

7.4 Lite Server 节点故障诊断

场景描述

Lite Server任务中心提供一键式故障诊断能力，包括参数面网络诊断和Ascend设备诊断。用户无需深入了解具体诊断操作命令，即可自助快捷地在Lite Server产品页面上完成网络和Ascend设备检查的诉求。

参数面网络诊断支持查询卡的网络状态，IP和掩码信息等，Ascend设备诊断支持对驱动固件版本兼容性进行诊断，并实现了带内检查自动化。同时可批量在多台服务器上同时启动诊断任务，大幅度提升效率。

约束限制

- 当前仅支持Snt9b节点和超节点Snt9b23。
- 同一个任务最多支持选择50个普通节点或超节点的子节点。
- 创建任务的节点需要安装NodeTaskHub插件，请在创建任务前确保插件安装完毕，具体参见[安装Lite Server AI插件](#)。
- 同一时间节点上最多同时支持一个诊断任务，任务开始后无法中断，请您规划好任务优先级。
- 请确保待诊断节点无业务运行，诊断过程中的命令执行可能导致当前业务中断或异常。
- 执行诊断前需安装Ascend HDK23.0.0及以后的版本的MCU、驱动和固件，预置操作系统已经默认安装，如果是自定义操作系统，也需确保该软件正常安装。
- 诊断任务依赖开发套件包Ascend-docker-runtime，预置操作系统已经默认安装该软件，如果是自定义操作系统，也需确保该软件正常安装。

操作步骤

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“任务中心”。

图 7-10 任务中心



3. 单击任务中心页面左上角的“创建任务”，进入“任务模板”页面，在该页面选择“Ascend故障诊断”，单击“创建任务”。

图 7-11 任务模板



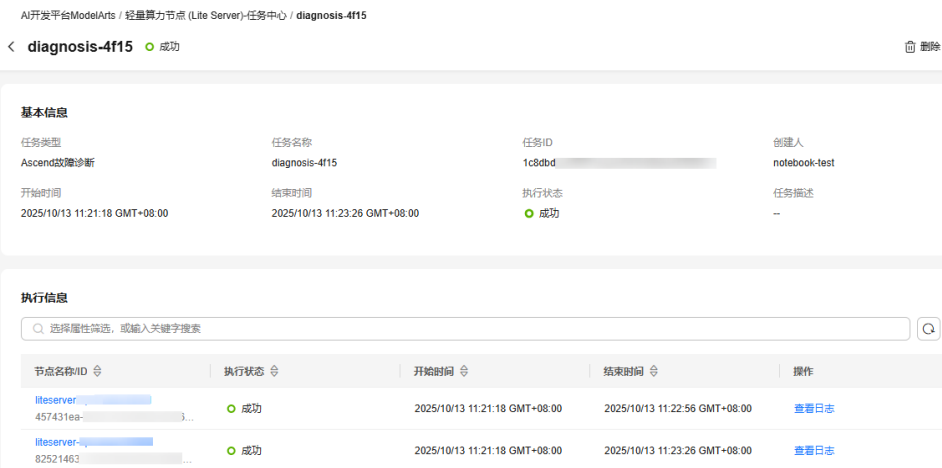
4. 在Ascend故障诊断任务创建页面，填写“任务名称”、“任务描述”，选择“机型”，选择“诊断项目”，勾选使用须知并单击“立即创建”。

表 7-4 创建任务参数

参数分类	参数说明
任务名称	系统自动填入任务名称，用户可以自定义。
任务描述	对该任务的描述信息，方便快速查找任务。
机型	选择机型，并在节点列表中勾选节点。具体节点信息支持通过关键字搜索。 支持Snt9b节点和超节点Snt9b23。
诊断项目	支持选择参数面网络诊断和Ascend设备检查，也可以同时执行。 - 参数面网络诊断：对网络相关指标和信息进行采集统计和状态诊断。 - Ascend设备诊断：对Ascend相关软件和芯片相关指标进行健康检测和兼容性验证。

5. 返回“任务中心”页面，显示任务的执行状态。
6. 单击具体的任务名称，可以进入任务详情页，查看任务的详细信息。

图 7-12 查看任务详情



7. 在任务详情页，单击“查看日志”，在页面右侧弹窗中查看任务执行的详细日志信息。所有检查结果会在任务日志中呈现，并提供了基本的日志分析。

图 7-13 查看日志



带内自动化检查项

Ascend设备检查任务中完成的带内自动化检查项包括以下内容。

表 7-5 带内自动化检查项

检查项目	命令参考	检查动作
检测UDP端口散列配置	hccn_tool -i \$i -udp -g	检查端口号是否为0/4791的异常
检测NPU卡健康信息	timeout 20s npu-smi info -t health -i "\$i" grep OK -c	仅限Snt9b23，检查NPU健康码是否为3
检测NPU驱动版本是否一致	timeout 20s npu-smi info -t board -i "\$i" grep Version	检查所有NPU卡的驱动号码是否一致
检测PCIE LINK状态	lspci grep d8 / lspci grep d8 -c	仅限Snt9b23，PCIE 建链是否为16
检测NPU网卡是否UP	hccn_tool -i \$i -link -g	检测网卡是否down
检测NPU网卡健康状态	hccn_tool -i \$i -net_health -g	网卡是否健康
检测NPU PFC是否符合预期	hccn_tool -i \$i -pfc -g	检测PFC是否满足条件，PFC固定配置如下
检测TLS证书是否符合预期	hccn_tool -i \$i -tls -g grep switch	字段内switch[0]是否满足
驱动固件版本兼容性测试	ascend-dmi -ci	判断兼容性是否满足

7.5 Lite Server 节点漏洞修复

场景描述

在日常的系统运维过程中，运维人员经常面临操作系统版本更新后出现的安全漏洞问题，如内核漏洞、openssh漏洞等，这些问题不仅影响系统的稳定运行，还可能带来安全隐患。传统的解决方法是重新制作OS镜像，但这一过程耗时较长且工作量大，无法满足快速响应的需求。为此，Lite Server新增了通用系统漏洞修复任务，支持一键系统配置，能够逐项检测并修复各类安全漏洞。通过这些功能，运维人员可以更加高效、便捷地管理系统的安全性和稳定性，快速响应各类安全漏洞，保障业务的连续性和安全性。

约束限制

运维面下发自定义脚本任务约束限制如下：

- 当前仅支持Snt9b节点和超节点Snt9b23。
- 当前仅支持HCE2.0操作系统的漏洞修复。
- 创建任务的节点需要安装NodeTaskHub插件，请在创建任务前确保插件安装完毕，具体参见[安装Lite Server AI插件](#)。
- 同一时间节点上最多同时支持一个任务，任务开始后无法中断，请您规划好任务优先级。
- 请确保目标节点无业务运行，执行任务过程中可能会导致当前业务中断或异常。
- 执行诊断前需安装Ascend HDK23.0.0及以后的版本的MCU、驱动和固件，预置操作系统已经默认安装，如果是自定义操作系统，也需确保该软件正常安装。
- 诊断任务依赖开发套件包Ascend-docker-runtime，预置操作系统已经默认安装该软件，如果是自定义操作系统，也需确保该软件正常安装。

操作步骤

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“任务中心”。

图 7-14 任务中心



3. 单击任务中心页面左上角的“创建任务”，进入“任务模板”页面，在该页面选择“Ascend系统配置”，单击“创建任务”。

图 7-15 任务模板



4. 在Ascend系统配置任务创建页面，填写“任务名称”、“任务描述”，选择“机型”，选择“配置项”，勾选使用须知并单击“立即创建”。

表 7-6 创建任务参数

参数分类	参数说明
任务名称	系统自动填入系统配置任务名称，用户可以自定义。
任务描述	对该任务的描述信息，方便快速查找任务。
机型	选择机型，并在节点列表中勾选节点。具体节点信息支持通过关键字搜索。 支持Snt9b节点和超节点Snt9b23。
配置项	系统漏洞修复：检测并修复HCE2.0系统安全漏洞。

5. 返回“任务中心”页面，显示任务的执行状态。
6. 单击具体的任务名称，可以进入任务详情页，查看任务的详细信息。
7. 在任务详情页，单击“查看日志”，在页面右侧弹窗中查看任务执行的详细日志信息。所有检查结果会在任务日志中呈现，并提供了基本的日志分析。

漏洞检查项

Ascend系统配置任务中完成的系统漏洞修复包括以下内容。

漏洞项	操作系统	解决方案
ipvs_fnat	HCE2.0	校验是否存在漏洞，如果存在则自动修复该软件。
openssh	HCE2.0	校验是否存在漏洞，如果存在则自动修复该软件。

7.6 Lite Server 节点一键式压测

场景描述

Lite Server任务中心提供一键式的压测能力，用户无需深入理解AICore，HBM等软件栈，即可自助快捷地在Lite Server产品页面上完成业务压测诉求。支持对昇腾服务器的带宽测试、算力测试、功耗测试、诊断压测等，为AI训练、推理等高负载场景提供硬件保障，同时可批量在多台服务器上均可并行，大幅度提升效率。

约束限制

- 当前仅支持Snt9b节点和超节点Snt9b23。
- 创建任务的节点需要安装NodeTaskHub插件，请在创建任务前确保插件安装完毕，具体参见[安装Lite Server AI插件](#)。
- 同一时间节点上最多同时支持一个压测任务，任务开始后无法中断，请您规划好任务优先级。
- 请确保待压测节点无业务运行，压测过程中的命令执行可能导致当前业务中断或异常。
- 执行压测前需安装Ascend HDK23.0.0及以后的版本的MCU、驱动和固件，预置操作系统已经默认安装，如果是自定义操作系统，也需确保该软件正常安装。
- 压测任务依赖开发套件包Ascend-docker-runtime，预置操作系统已经默认安装该软件，如果是自定义操作系统，也需确保该软件正常安装。

操作步骤

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“任务中心”。

图 7-16 任务中心



3. 单击任务中心页面左上角的“创建任务”，进入“任务模板”页面，在该页面选择“Ascend压测任务”，单击“创建任务”。

图 7-17 任务模板



4. 在Ascend压测任务创建页面，填写“任务名称”、“任务描述”，选择“机型”，选择“压测用例”，勾选使用须知并单击“立即创建”。

表 7-7 创建任务参数

参数分类	参数说明
任务名称	系统自动填入任务名称，用户可以自定义。
任务描述	对该任务的描述信息，方便快速查找任务。
机型	选择机型，并在节点列表中勾选节点。具体节点信息支持通过关键字搜索。 支持Snt9b节点和超节点Snt9b23。
压测用例	支持选择以下压测用例。压测用例可以单个执行，也可以同时执行。 ● AICore压测：对AICore ERROR进行压力测试，并输出诊断结果，AICore压测需要占用HOST服务器侧约20~40GB的内存，执行命令前请预留足够内存，防止进程异常中断。 ● HBM压测：对高带宽内存进行压力测试，并输出压测结果。 ● P2P压测：测试节点上所有Device之间的HCCS通信链路是否存在硬件故障。

5. 返回“任务中心”页面，显示任务的执行状态。
6. 单击具体的任务名称，可以进入任务详情页，查看任务的详细信息。
7. 在任务详情页，单击“查看日志”，在页面右侧弹窗中查看任务执行的详细日志信息。

图 7-18 查看日志

日志

liteserv

最新刷新时间: 2025/05/19 17:55:25 GMT+08:00 全屏 刷新

```
1 2025-04-30 15:47:42.905929 [INFO] Successfully submit ASCEND_STRESS_TEST job to server 694a908b-0062-4c99-ba54-d428875770cf.
2 2025-04-30 15:47:47.216352 [INFO] Start to run F2P stress task.
3 2025-04-30 15:47:47.236859 [INFO] Execute p2p stress command: ascend-dai -dg -i bandwidth -type p2p -s -q
4 2025-04-30 16:10:31.395430 [INFO] Execute p2p stress result:This test will affect the business on this server. To ensure the
correctness and accuracy of the test, perform the operation separately.Do you want to continue?(Y/N)Y
5 Summary:
6 Arch: aarch64
7 Mode: Ascend 910 B4-1
8 Time: 20250430-16:10:25
9
10 Hardware:
11 bandwidth:
12 PASS
13
14 2025-04-30 16:10:31.417058 [INFO] Success to run all stress task.
15
```

7.7 Lite Server 节点参数面网络配置

场景描述

Lite Server任务中心提供一键式的系统配置能力，用户可自助快捷地在Lite Server产品页面上完成参数面网络配置诉求。通过优化服务器参数（如RoCE网络上行端口、参数面网络IP）等，以满足训练/推理场景基本需求，并提升服务器性能，同时可批量在多台服务器上均可并行，大幅度提升效率。

约束限制

- 当前仅支持Snt9b节点和超节点Snt9b23。
- 创建任务的节点需要安装NodeTaskHub插件，请在创建任务前确保插件安装完毕，具体参见[安装Lite Server AI插件](#)。
- 同一时间节点上最多同时支持一个任务，任务开始后无法中断，请您规划好任务优先级。
- 请确保目标节点无业务运行，执行任务过程中可能会导致当前业务中断或异常。
- 执行任务前需安装Ascend HDK23.0.0及以后的版本的MCU、驱动和固件，预置操作系统已经默认安装，如果是自定义操作系统，也需确保该软件正常安装。
- 执行任务依赖开发套件包Ascend-docker-runtime，预置操作系统已经默认安装该软件，如果是自定义操作系统，也需确保该软件正常安装。

操作步骤

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“任务中心”。

图 7-19 任务中心



3. 单击任务中心页面左上角的“创建任务”，进入“任务模板”页面，在该页面选择“Ascend系统配置”，单击“创建任务”。

图 7-20 任务模板



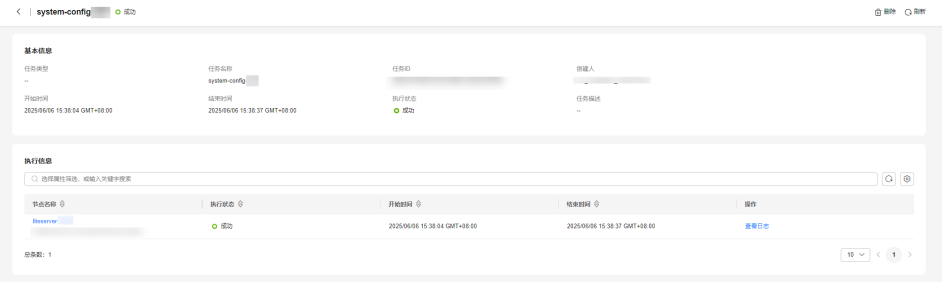
4. 在Ascend系统配置任务创建页面，填写“任务名称”、“任务描述”，选择“机型”，选择“配置项”，勾选使用须知并单击“立即创建”。

表 7-8 创建任务参数

参数分类	参数说明
任务名称	系统自动填入任务名称，用户可以自定义。
任务描述	对该任务的描述信息，方便快速查找任务。
机型	选择机型，并在节点列表中勾选节点。具体节点信息支持通过关键字搜索。 支持Snt9b节点和超节点Snt9b23。
配置项	参数面网络配置：优化RoCE网络上行端口配置，并调整参数面网络参数，确保参数面网络IP地址、子网掩码及网关等配置准确。

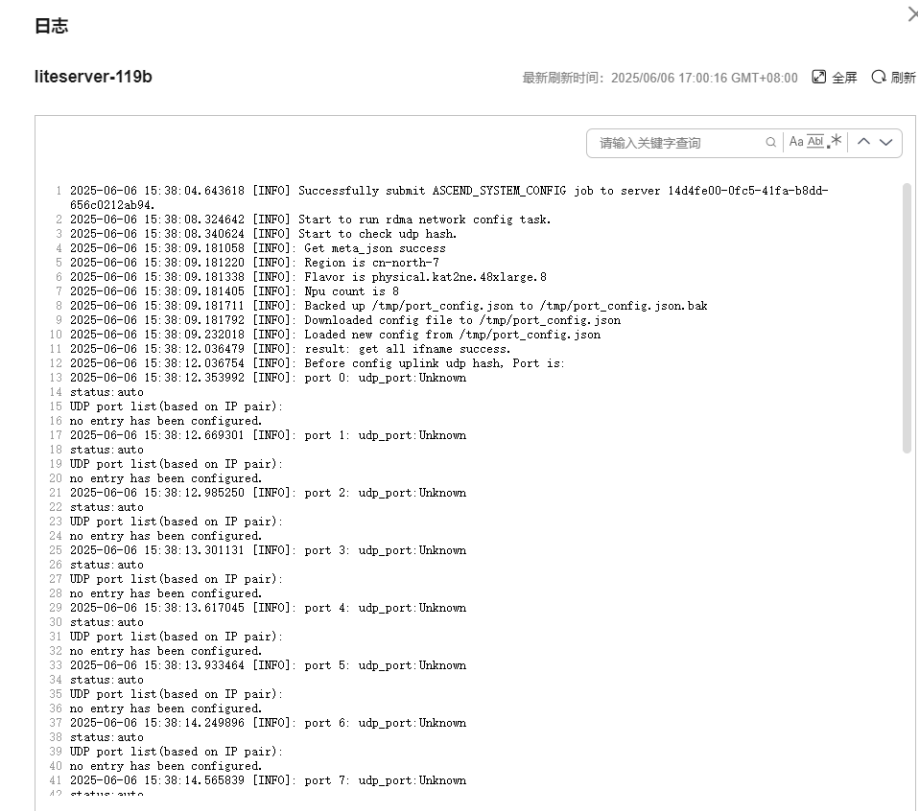
5. 返回“任务中心”页面，查看任务的执行状态。
6. 单击具体的任务名称，可以进入任务详情页，查看任务的详细信息。

图 7-21 任务详情



7. 在任务详情页，单击“查看日志”，在页面右侧弹窗中查看任务执行的详细日志信息。

图 7-22 查看日志



8 Lite Server 超节点管理

8.1 Lite Server 超节点扩容和缩容

场景描述

当用户使用一段时间Lite Server的超节点资源后，由于用户业务的变化，需要扩容或缩容超节点，可以通过本章节的操作实现。

约束限制

扩容：

- 仅适用于超节点（Snt9b23）扩容。
- 超节点扩容时，子节点个数不超过48个。
- 待扩容节点处于可用状态时，才支持扩容。
- 扩容时不影响原有节点上的业务。

缩容：

- 仅适用于超节点（Snt9b23）缩容。
- 超节点的子节点个数为1个时，不支持缩容。
- 待缩容节点处于非运作的中间态，例如：切换操作系统、重启中、启动中、停止中等时，不支持缩容操作。
- 缩容时会影响原有节点上的业务，配套的云硬盘也会被随之释放，请确保已完成相应节点的业务迁移，避免数据丢失。

计费影响

扩容节点和缩容节点均支持包年包月计费模式。扩容后的节点会新增一笔计费订单。

超节点扩容操作

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点（Lite Server）> 资源列表”，进入资源列表页面。

图 8-1 资源列表



3. 在Lite Server资源列表页的“超节点”页签中，找到待扩容的超节点资源，单击右侧  中的“超节点规格变更”，进入超节点规格变更页面。
4. 在超节点规格变更页面，查看当前待变更的资源信息，选择目标规格。已售罄的资源会呈灰色显示，不支持选择。
5. 设置扩容节点信息。

表 8-1 扩容节点配置说明

参数名称	说明
系统盘	选择“系统盘类型”，并设置“大小”。创建Lite Server时自带系统盘，建议系统盘大小取值至少200GB。
增加数据盘	单击“增加数据盘”，可以在Lite Server上挂载数据盘。也可以在Lite Server资源创建完成后在云服务器侧实现数据盘挂载，具体参见 使用云硬盘EVS作为存储 。
镜像	● 公共镜像 公共镜像对所有用户可见。所有用户可以根据镜像ID进行只读使用。 ModelArts服务提供了多个公共镜像，支持多种操作系统，并且内置了AI场景相关驱动和软件，为用户提供了一个完整的AI开发环境，方便用户直接进行开发和训练，而无需额外配置。 当前支持的公共镜像请参考Lite Server算力资源和镜像版本配套关系。 ● 私有镜像 仅镜像创建者可以使用，其他用户无法访问。选择私有镜像创建，可以节省您重复配置服务器的时间。
登录凭证	“密钥对”方式创建的Server节点安全性更高，建议选择“密钥对”方式。超节点扩容的登录凭证当前仅支持密钥对方式，不支持密码方式。 ● 密钥对 指使用密钥对作为登录Server节点的鉴权方式。您可以选择使用已有的密钥对，或者单击“新建密钥对”创建新的密钥。 说明 如果选择使用已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录Server节点。

参数名称	说明
实例自定义注入（可选）	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 Server 节点： - 通过脚本简化 Server 节点配置 - 通过脚本初始化系统 - 已有脚本，在创 Server 节点时一并上传至服务器 - 其他可以使用脚本完成的操作 当前支持“以文本形式”和“以文件形式”，使用方法可参考ECS 实例自定义数据注入。

- 设置完成后，单击“下一步”，预览变更信息、查看扩容节点配置和费用，确认无误后，单击“提交订单”。
- 扩容变更完成后，在资源列表页，展开扩容变更的超节点，查看该超节点下新增的子节点的运行状态。

超节点缩容操作

- 登录[ModelArts 管理控制台](#)。
- 在左侧导航栏中，选择“资源管理 > 轻量算力节点（ Lite Server ） > 资源列表”，进入资源列表页面。

图 8-2 资源列表



- 在 Lite Server 资源列表页的“超节点”页签中，找到待扩容的超节点资源，单击右侧 **...** 中的“超节点规格变更”，进入超节点规格变更页面。
- 在超节点规格变更页面，查看“当前待变更资源信息”，选择“目标规格”。在“目标保留节点”列表中选择需要保留的节点。

目标规格是指需要保留节点的规格。

目标保留节点个数必须与目标规格中的计算节点个数保持一致，否则会导致缩容任务提交失败。
- 勾选“我已知晓缩容操作可能存在的风险，同意进行缩容”，并输入“YES”，确认缩容操作。



警告

确认缩容后，未保留的节点会被释放，配套的云硬盘也会被随之释放，请确保已完成相应节点的业务迁移，避免数据丢失。

6. 设置完成后，单击“下一步”，预览变更信息、查看保留节点信息和删除节点信息，确认无误后，单击“提交订单”。

8.2 Lite Server 超节点系统盘扩容

场景描述

在使用超节点进行业务操作时，如果遇到系统盘空间不足的情况，例如在尝试切换操作系统（OS）时，由于系统盘容量不足以支持新的OS镜像，导致切换失败。此外，在日常业务运行中，随着数据的不断积累，系统盘的剩余空间逐渐减少，最终达到容量上限，影响业务的正常运行。

面对这些情况，为确保业务的连续性和稳定性，Lite Server提供了系统盘扩容功能，用户可以在节点详情页面通过新增的扩容功能，轻松跳转至EVS页面完成扩容操作，确保系统盘容量满足业务需求，从而避免因空间不足导致的操作失败或业务中断。

约束限制

系统盘扩缩容功能约束限制如下：

- 系统盘扩容在控制台通过跳转EVS页面实现，依赖EVS扩容特性。
- 不支持系统盘缩容。

计费影响

扩容后的磁盘计费信息在之前创建超节点时的计费订单中一并体现，不会单独创建新的计费订单。

超节点系统盘扩容操作

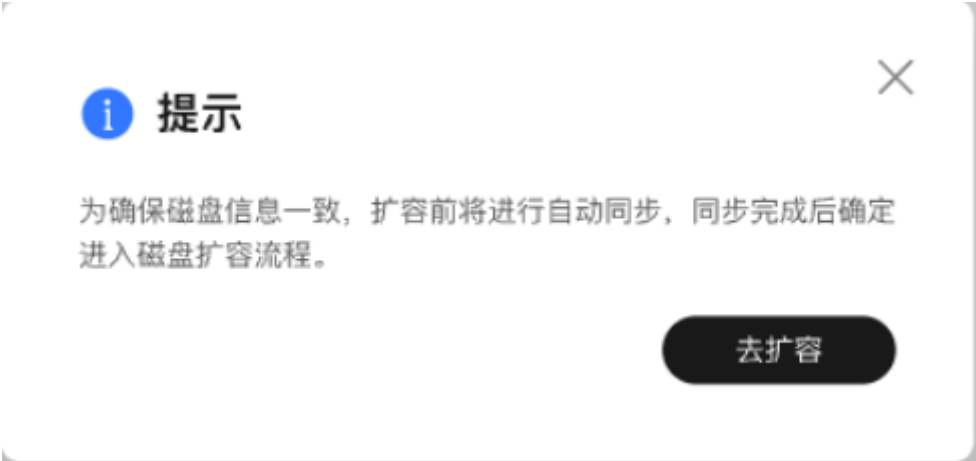
1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点（Lite Server）> 资源列表”，进入资源列表页面。

图 8-3 资源列表



3. 在Lite Server资源列表页的“超节点”页签中，找到待扩容的超节点资源，进入超节点的子节点详情页。
4. 在超节点的子节点详情页中，在磁盘区域找到待扩容的系统盘，单击右侧的“扩容”。根据系统弹窗提示单击“去扩容”。

图 8-4 扩容



5. 页面会跳转到EVS控制台，在EVS控制台完成扩容操作，具体操作请参考[扩容云硬盘](#)。
6. 系统盘扩容完成后，在ModelArts Lite Server页面手动进行同步，获取磁盘最新信息。同步操作参见[同步Lite Server服务器状态](#)。

后续操作

[同步Lite Server服务器状态](#)

8.3 Lite Server 超节点定期压测

场景描述

针对超节点Snt9b23，支持用户定期对昇腾服务器进行性能测试和故障诊断，及时发现NPU故障，减少业务影响。

表 8-2 性能测试

性能测试场景	场景说明
带宽测试	带宽测试主要用于测试总线带宽、内存带宽和总耗时。
算力测试	算力测试通过构造矩阵乘" $A(m,k)*B(k,n)$ "并执行一定次数的方式，根据运算量与执行多次矩阵乘所耗时间来计算整卡或处理器中AI Core的算力值和满算力下实时的功率。
功耗测试	功耗测试是通过运行单算子模型来检测整卡的功耗信息。
眼图测试	用户使用眼图测试功能对网络进行测试，查询当前信号质量。
码流测试	码流测试主要包含一键式打流和自定义打流。

性能测试场景	场景说明
软硬件版本兼容性测试	软硬件兼容性工具会获取硬件信息、架构、驱动版本、固件版本以及软件版本。

表 8-3 故障诊断

故障诊断场景	场景说明
Network诊断	对网络健康状态进行诊断，并输出诊断结果。
SignalQuality诊断	对信号质量进行诊断，并输出诊断结果。
片上内存诊断	对高带宽内存进行诊断，并输出诊断结果。
片上内存压测	对高带宽内存进行压力测试，并输出诊断结果。
片上内存高危地址压测	对高带宽内存高危地址进行压力测试，并输出诊断结果。
AICore诊断	对AICore ERROR进行诊断，并输出诊断结果。
AIflops诊断	对芯片进行算力诊断，并输出测试结果。
Bandwidth诊断	对本地带宽进行诊断，并输出诊断结果。
P2P压测	测试指定源头Device到目标Device的HCCS通信链路是否存在硬件故障，并输出测试结果。
功耗压测	进行EDP/TDP功耗压力测试，并输出诊断结果。

约束限制

- 仅支持超节点Snt9b23。
- 压测使用工具Ascend DMI，其不支持在同一个设备里同时开启多个进程来测试性能数据，多进程测试时，可能导致测试结果不准确或者失败等不可预测情况。
- 性能测试和故障诊断会影响训练或推理业务，执行命令前请确保无业务运行。
- 为保证返回检测结果的正确性和准确性，请单独执行各个检测命令。
- Ascend DMI工具只能对在位的NPU卡进行检查，为保证测试结果的准确性，请先执行**npu-smi info**命令检查NPU卡是否正常在位。

性能测试一：带宽测试

带宽测试主要用于测试总线带宽、内存带宽和总耗时。带宽测试命令的可用参数说明见[表8-4](#)。

```
ascend-dmi --bw -h
```

表 8-4 带宽测试参数说明

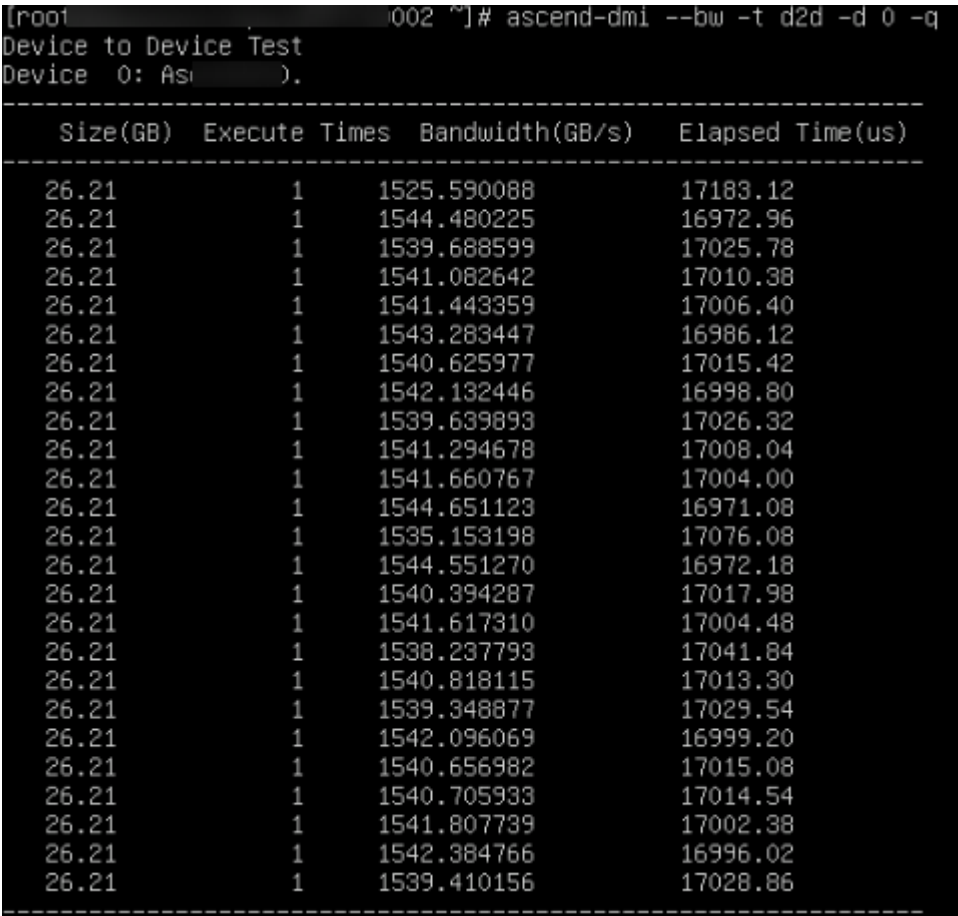
参数	说明	是否必填
[-bw, --bw, --bandwidth]	使用该参数测试芯片的带宽。支持-bw，但建议使用--bw或--bandwidth。	是
[-t, --type]	指测试数据流向的分类。当使用带宽测试功能时，测试的数据流可以分为以下方向，如果不填写数据流方向则默认返回h2d、d2h、d2d。 三个方向的带宽和总耗时。 - h2d：指数据从Host侧内存通过PCIe总线搬移到Device侧内存，测试整体带宽及总耗时。 - d2h：指数据从Device侧内存通过PCIe总线搬移到Host侧内存，测试整体带宽及总耗时。 - d2d：指数据从Device侧内存搬移到同一Device侧内存（主要是用于测试Device侧的内存带宽），测试整体带宽及总耗时。 - p2p：测试指定源头Device到目标Device的传输速率和总耗时。	否
[-s, --size]	指传输数据大小并指定测试结果显示方式。超节点系列产品：d2h/h2d/p2p这3种模式下，最大传输数值为1Byte~4G。 - 指定-s参数后面必须填写数值指定传输数据的大小，不填写属于错误写法。 - 在h2d、d2h、d2d以及p2p且指定-ds和-dd场景：指定-s为定长模式；不指定-s为步长模式，传输数据的默认取值范围为2Byte~32M。	否
[-et, --et, --execute-times]	指迭代次数，即内存复制次数。取值范围为[1, 1000]，如果不填写，步长模式下复制次数则默认为5，定长模式下复制次数则默认为40。	否
[-d, --device]	指定需要测试带宽的Device ID，Device ID是指昇腾AI处理器的逻辑ID，如果不填写Device ID则默认返回Device 0带宽信息。	否
[-ds, --ds, --device-src]	指定p2p测试的源头Device的ID号。必须与[-dd, --dd, --device-dst]参数成对指定；如果与[-dd, --dd, --device-dst]参数同时不指定时，测试全量的昇腾NPU芯片。	否
[-dd, --dd, --device-dst]	指定p2p测试的目标Device的ID号。必须与[-ds, --ds, --device-src]参数成对指定；如果与[-ds, --ds, --device-src]参数同时不指定时，测试全量的昇腾NPU芯片。	否
[-fmt, --fmt, --format]	指定输出格式，可以为normal或json。如果未指定则默认为normal。	否

参数	说明	是否必填
[-q, --quiet]	指定该参数时，将不再进行防呆提示，用户将默认允许该操作。	否

使用示例，以测试数据从Device侧传输到同一Device侧的带宽与总耗时为例。

```
ascend-dmi --bw -t d2d -d 0
```

图 8-5 带宽测试示例

The image is a screenshot of a terminal window showing the output of the 'ascend-dmi --bw -t d2d -d 0 -q' command. The output starts with a header 'Device to Device Test' and 'Device 0: Ascend310E'. It then displays a table with four columns: 'Size(GB)', 'Execute Times', 'Bandwidth(GB/s)', and 'Elapsed Time(us)'. The table contains 20 rows of test data, all for a size of 26.21 GB. The bandwidth values are consistently around 1540 GB/s, and the elapsed times are around 17000 us.

Size(GB)	Execute Times	Bandwidth(GB/s)	Elapsed Time(us)
26.21	1	1525.590088	17183.12
26.21	1	1544.480225	16972.96
26.21	1	1539.688599	17025.78
26.21	1	1541.082642	17010.38
26.21	1	1541.443359	17006.40
26.21	1	1543.283447	16986.12
26.21	1	1540.625977	17015.42
26.21	1	1542.132446	16998.80
26.21	1	1539.639893	17026.32
26.21	1	1541.294678	17008.04
26.21	1	1541.660767	17004.00
26.21	1	1544.651123	16971.08
26.21	1	1535.153198	17076.08
26.21	1	1544.551270	16972.18
26.21	1	1540.394287	17017.98
26.21	1	1541.617310	17004.48
26.21	1	1538.237793	17041.84
26.21	1	1540.818115	17013.30
26.21	1	1539.348877	17029.54
26.21	1	1542.096069	16999.20
26.21	1	1540.656982	17015.08
26.21	1	1540.705933	17014.54
26.21	1	1541.807739	17002.38
26.21	1	1542.384766	16996.02
26.21	1	1539.410156	17028.86

表 8-5 带宽测试回显参数说明

参数	说明
Host to Device Test	带宽数据流方向。有以下显示可能： - Host to Device Test - Device to Host Test - Device to Device Test - Unidirectional Peer to Peer Test - Bidirectional Peer to Peer Test
Device X : Ascend XXX	Device X为当前测试的设备ID，Ascend XXX为处理器类型。0表示源头设备，1表示目标设备。
ID	0 → 1表示测试Device 0到Device 1的单向P2P带宽。 0 ↔ 1表示测试Device 0和Device 1的双向p2p带宽。
Size(Bytes)	传输数据大小，单位为字节。
Execute Times	迭代次数。
Bandwidth(GB/s)	芯片的带宽。
Elapsed Time(us)	总执行时长。

性能测试二：算力测试

算力测试通过构造矩阵乘"A(m,k)*B(k,n)"并执行一定次数的方式，根据运算量与执行多次矩阵乘所耗时间来计算整卡或处理器中AI Core的算力值和满算力下实时的功率。算力测试的可用参数说明见[表8-6](#)。

表 8-6 算力测试参数说明

参数	说明	是否必填
[-f, --flops]	使用该参数测试整卡或芯片的算力。	是
[-t, --type]	指定算子运算类型，可以为fp16、fp32、hf32、bf16和int8，如果未指定则默认为fp16。	否
[-d, --device]	指定Device ID，执行该Device ID所在整卡的算力测试，Device ID是指昇腾芯片的逻辑ID，如果不填写Device ID则默认返回Device 0的算力信息。	否

参数	说明	是否必填
<code>[-et, --et, --execute-times]</code>	指定芯片单个AI Core上运行矩阵乘法的执行次数。 - 训练场景：如果不填写执行次数则默认为60。训练场景单位为十万，参数范围为[10, 80]。 - 推理场景：如果不填写执行次数则默认为10。推理场景单位为百万，参数范围为[10, 80]。	否

使用示例，在Device 7上，执行算子运算类型为int8，执行次数为600万的算力。

```
ascend-dmi -f -t int8 -d 7 -et 60 -q
```

图 8-6 算力测试示例

[root@de 02 ~]# ascend-dmi -f -t int8 -d 7 -et 60 -q				
Device	Execute Times	Duration(ms)	TOPS@INT8	Power(W)
6/7	900,000,000	1657	1563.549	776.300049

表 8-7 算力测试回显参数说明

参数	说明
Device	Device ID。
Execute Times	为单个AI Core执行矩阵乘的次数乘以AI Core的个数计算所得。
Duration(ms)	执行多次矩阵乘所耗费的时间。
TFLOPS@FP16	进行算力测试得到的算力值。FP16为指定的算子运行类型。
Power(W)	满算力下的实时功率。

性能测试三：功耗测试

功耗测试是通过运行单算子模型来检测整卡的功耗信息。功耗测试的可用参数说明见[表8-8](#)。

```
ascend-dmi -p -h
```

表 8-8 功耗测试参数说明

参数	说明	是否必填
<code>[-p, --power]</code>	使用该参数进行整卡的功耗测试。	是

参数	说明	是否必填
[-t, --type]	指定算子运算类型，可以为fp16或int8，如果未指定则默认为fp16。	否
[-pt, --pt, --pressure-type]	使用该参数指定压力测试的类型。 - 当前支持指定以下2种类型： - edp (Estimated Design Power)：EDP功耗压力测试。 - tdp (Thermal Design Power)：TDP功耗压力测试。 - 支持和--dur、--it、--pm、-q参数一起使用。 - 不支持和-t参数一起使用。 - 不指定该参数时默认进行整卡的功耗测试。	否
[-dur, --dur, --duration]	指运行时间，如果不填写运行时间则默认为600。单位为秒，取值范围为[60, 604800]。	否
[-it, --it, --interval-times]	指屏幕信息打印刷新的间隔时间，如果不填写间隔时间则默认为5。单位为秒，取值范围为[1, 5]。	否
[--skip-check]	传入此参数时会跳过设备健康状态检查。不传入此参数，默认会进行设备健康状态检查。	否
[-pm, --pm, --print-mode]	屏幕输出的打印模式，如果不填写打印模式则默认为refresh。打印模式： - refresh：每次打印清除历史打印信息。 - history：打印保存历史信息。说明 refresh模式下，当芯片数量较多时，建议调小字体使得所有结果都在一个屏幕中，否则可能会显示异常，重复打印部分内容。	否

使用示例，以执行时间为60s，信息的打印间隔信息为5s，屏幕的输出模式为清除历史记录为例。

```
ascend-dmi -p --dur 60 --it 5--pm refresh
```


图 8-7 功耗测试示例

Card	Type	NPU Count	Power
Chip Name Device ID	Health AI Core Usage	Temperature Voltage	Frequency
0 A	2	1049.2W	
0 0	OK 100%	77C 0.80V	1750MHZ
1 1	OK 100%	75C 0.79V	1800MHZ
1 A	2	1049.6W	
0 2	OK 100%	78C 0.81V	1850MHZ
1 3	OK 100%	77C 0.78V	1800MHZ
2 A	2	1050.9W	
0 4	OK 100%	81C 0.79V	1750MHZ
1 5	OK 100%	75C 0.78V	1750MHZ

Card	Type	NPU Count	Power
Chip Name	Device ID	Health AI Core Usage	Temperature Voltage Frequency
0		2	1049.2W
0		OK 100%	77C 1750MHZ 0.80V
1		OK 100%	75C 1800MHZ 0.79V
1		2	1049.0W
0		OK 100%	78C 1850MHZ 0.81V
2		OK 100%	77C 1800MHZ 0.78V
2		2	1050.9W
0		OK 100%	81C 1750MHZ 0.79V
1		OK 100%	75C 1750MHZ 0.78V

表 8-9 功耗测试回显参数说明

参数	说明
Type	标卡型号
Card	卡ID号

参数	说明
Chip	处理器编号
Name	处理器名称
Type	处理器型号
Chip Name	处理器名称
NPU Count	NPU的个数
Power	当前整卡或芯片的实际功耗
Health	处理器健康程度
Temperature	处理器当前温度
Device ID	处理器设备逻辑号
AI Core Usage	处理器AI Core的使用率
Voltage	处理器当前电压
Frequency	处理器当前频率

性能测试四：眼图测试

用户使用眼图测试功能对网络进行测试，查询当前信号质量。本功能主要用于查询信号质量的具体数据。判断当前端口信号质量是否正常，请执行signalQuality诊断。在同一NPU内，如果已配置CDR回环，请在解除回环后再执行眼图测试。眼图测试的可用参数说明见[表8-10](#)。

```
ascend-dmi --sq -h
```

表 8-10 眼图测试参数说明

参数	说明	是否必填
[-sq, --sq, --signal-quality]	查询NPU上的PCIe、HCCS和RoCE通信端口的信号质量。	是
[-d, --device]	指定查询的Device ID。指定多个芯片时，使用英文逗号进行分隔。不指定该参数时，默认查询该设备上所有的NPU。	否
[-t --type]	指定通信端口的类型。当前支持HCCS和RoCE，指定多个通信端口的类型时，使用英文逗号进行分隔。如果不指定该参数则将查询RoCE的信号质量。	否

使用示例，查看Device0和Device1的HCCS、RoCE信号质量。

图 8-8 眼图测试示例

```
[root@ ~]# ascend-dml --sq -t hcce,roce -d 0,1
type: hcce
Prompt message: M*: macro port, L*: lane, S: SNR, H: HEH
Normal range: S(SNR) >= 400000 and H(HEH) >= 350
-----
device      signal-to-noise ratio
-----
0           M1: L0: S:624980 H:404   L1: S:570601 H:405   L2: S:602362 H:406   L3: S:582752 H:395
           M2: L0: S:625353 H:409   L1: S:580083 H:387   L2: S:628120 H:407   L3: S:596369 H:404
           M3: L0: S:630788 H:393   L1: S:608124 H:397   L2: S:595847 H:381   L3: S:523765 H:389
           M4: L0: S:544151 H:396   L1: S:609019 H:403   L2: S:561796 H:405   L3: S:513117 H:398
           M5: L0: S:572574 H:399   L1: S:591882 H:391   L2: S:627122 H:397   L3: S:582290 H:401
           M6: L0: S:568245 H:393   L1: S:553167 H:398   L2: S:599843 H:419   L3: S:533950 H:406
           M7: L0: S:586303 H:402   L1: S:589030 H:380   L2: S:604690 H:390   L3: S:563846 H:384
-----
1           M1: L0: S:645175 H:401   L1: S:599261 H:403   L2: S:589186 H:403   L3: S:569855 H:396
           M2: L0: S:622159 H:400   L1: S:529333 H:396   L2: S:535708 H:398   L3: S:565269 H:406
           M3: L0: S:618786 H:396   L1: S:575470 H:407   L2: S:610021 H:405   L3: S:558587 H:402
           M4: L0: S:572800 H:401   L1: S:539484 H:405   L2: S:558558 H:387   L3: S:496511 H:378
           M5: L0: S:578993 H:393   L1: S:570842 H:401   L2: S:628120 H:411   L3: S:575201 H:393
           M6: L0: S:593619 H:403   L1: S:517064 H:386   L2: S:561252 H:396   L3: S:509361 H:399
           M7: L0: S:592860 H:403   L1: S:555182 H:401   L2: S:582173 H:406   L3: S:528242 H:399
-----

type: roce
Prompt message: M*: macro port, L*: lane, S: SNR, H: HEH
Normal range: S(SNR) >= 400000 and H(HEH) >= 350
-----
device      signal-to-noise ratio
-----
0           M0: L0: S:660394 H:410   L1: S:575225 H:404   L2: S:674404 H:400   L3: S:576804 H:417
-----
1           M0: L0: S:683499 H:419   L1: S:659572 H:395   L2: S:676381 H:414   L3: S:585714 H:402
-----
```

表 8-11 HCCS 信号质量回显参数说明

参数	说明
type	指定通信端口的类型。
device	NPU的逻辑ID。
M* (macro port)	表示macro端口，例如M0、M1分别表示macro的0号、1号端口。
L* (LANE)	表示HCCS链路中的第几条lane，例如L0、L1分别表示第0条和第1条lane。
S (SNR)	表示lane的信噪比。
H (HEH)	表示lane的半眼高。

表 8-12 RoCE 信号质量回显参数说明

参数	说明
type	指定通信端口的类型。
device	表示NPU的逻辑ID。
M* (macro port)	表示macro端口，例如M0表示macro端口0。
S (SNR)	表示lane的信噪比。
H (HEH)	表示lane的半眼高。

参数	说明
L* (LANE)	表示RoCE链路中的第几条lane，例如L0、L1分别表示第0条和第1条lane。

性能测试五：码流测试

码流测试主要包含一键式打流和自定义打流。

表 8-13 码流测试介绍

测试项名称	支持的打流方式	使用方法
一键式打流	CDR环回打流、光模块外接光纤回路器（自环器）打流	执行一键式打流命令，Ascend DMI工具将自动完成发送及接收指定device所有lane的码流，一段时间后关闭码流并查询结果。
自定义打流	CDR环回打流、光模块外接光纤回路器（自环器）打流、NPU直连打流	自定义打流是将一键式打流中的各步骤独立出来，用户可灵活控制打流的TX、RX方向开关和指定打流的具体lane。

打流方式主要有以下三种：

- CDR环回打流：是指单个Device同时发送和接收，可用于检查从NPU的物理serdes端口到CDR单元的信号质量。在打流前请确保光模块在位，然后执行如下命令配置或解除CDR回环。
配置CDR回环，t依次取值3和0，一次执行如下命令，其中i表示NPU卡id：
hccn_tool -i 0 -scdr -t 3
hccn_tool -i 0 -scdr -t 0
解除CDR回环，t依次取值2和1：一次执行如下命令，其中i表示NPU卡id：
hccn_tool -i 0 -scdr -t 2
hccn_tool -i 0 -scdr -t 1
- 光模块外接光纤回路器（自环器）打流：单个Device同时发送和接收，可用于检查NPU的物理serdes端口到光模块的信号质量，不需要设置环回。
- NPU直连打流：NPU A的Serdes端口开启TX方向打流后，数据流通过被测链路到达NPU B的Serdes端口，NPU B的RX方向按照码型比对，统计接收到的数据统计误码情况，可检查两个NPU之间链路的信号质量（仅支持自定义打流）。

码流测试的可用参数，参数说明见[表8-14](#)。

```
ascend-dmi --prbs-check -h
```

表 8-14 码流测试参数说明

参数	说明	是否必填
[-pc, --pc, --prbs-check]	使用该参数进行prbs码流测试。	是
[-d, --device]	指定需要进行码流测试的Device ID。 - Device ID是指昇腾AI处理器的逻辑ID，如果不填写该参数，则测试全部NPU芯片的码流。 - 可同时指定多个Device ID，多个Device ID之间用逗号隔开。	否
[-dur, --dur, --duration]	指定码流测试的时长。 - 参数取值范围为[3, 10]，单位为秒。 - 不指定该参数时，默认值为3。	否
[--prbs-mode]	是否切换打流状态。 --取值为EN（Enable）：开启。 --取值为DS（Disable）：关闭。 - 取值支持大小写。 - 指定--prbs-mode为EN或DS时，信号发送端和信号接收端两个方向均会生效，无论是否指定--generator-pattern、--generator-lanes、--checker-pattern、--checker-lanes参数。 - 指定--prbs-mode为EN时，支持指定-generator-pattern、--checker-pattern、--generator-lanes、--checker-lanes。 - 指定--prbs-mode为DS时，停止打流。不支持指定-generator-pattern、--checker-pattern、--generator-lanes、--checker-lanes。 - 本参数不支持与--show参数或--clear参数同时使用。	是
[--generator-pattern]	指定发送端的码流类型。 - 当前支持测试的码流类型为：prbs7、prbs9、prbs10、prbs11、prbs15、prbs20、prbs23、prbs31。 - 不指定该参数时，默认值为prbs31。 - 指定码型时大小写均可生效，例如prbs7也可以写为PRBS7。 - 本参数不支持与--show参数或--clear参数同时使用。	否

参数	说明	是否必填
[--generator-lanes]	指定发送端的lane。 - 可同时指定1个或多个lane，多个之间用逗号分开。指定多个lane时必须连续指定，如0,1,2或2,1,3，不支持非连续指定。 - 如果不指定，则默认测试所有lanes。 - 本参数不支持与--show参数或--clear参数同时使用。 - 可取值为0、1、2、3。	否
[--checker-pattern]	指定接收端的码流类型。 - 当前支持校验的码流类型为：prbs7、prbs9、prbs10、prbs11、prbs15、prbs20、prbs23、prbs31。 - 不指定该参数时，默认值为prbs31。 - 指定码型时大小写均可生效，例如prbs7也可以写为PRBS7。 - 本参数不支持与--show参数或--clear参数同时使用。	否
[--checker-lanes]	指定接收端的lane。 - 可同时指定1个或多个lane，多个之间用逗号分开。指定多个lane时必须连续指定，如0,1,2或2,1,3，不支持非连续指定。 - 如果不指定，则默认测试所有lanes。 - 本参数不支持与--show参数或--clear参数同时使用。 - 可取值为0、1、2、3。	否
[-show, --show, --show-diagnostic-info]	展示码流测试的结果。 - 本参数不支持与以下参数同时使用：--clear、--prbs-mode、--generator-pattern、--generator-lanes、--checker-pattern、--checker-lanes。 - 展示信息后当前码流测试的结果即会被清空。	否
[-clear, --clear, --clear-diagnostic-info]	清空码流测试的结果信息。 - 本参数不支持与以下参数同时使用：--show、--prbs-mode、--generator-pattern、--generator-lanes、--checker-pattern、--checker-lanes。 - 支持除以上参数外的其余参数同时使用。	否

一键式打流使用示例如下：

```
ascend-dmi -pc -d 9 --pattern prbs15 -dur 5
```

图 8-9 一键式打流示例

```
[root@dev ~]# ascend-dmi -pc -d 5 --pattern prbs15 -dur 5
This operation will make network port on devices down, please make sure no business is running on devices.
Do you want to continue?(Y/N)y
PRBS15 on device 5:
-----
lane      error count      error rate      alos      time(ms)
-----
0          67092480          0.0251382604%   0          5024
1          67092480          0.0251319427%   0          5026
2          67092480          0.0251509054%   0          5022
3          67092480          0.0251382604%   0          5025
-----
```

表 8-15 一键式打流回显参数说明

参数	说明
device	表示NPU的逻辑ID。
lane	表示RoCE链路的lane通道ID。
error count	误码数，最大值为67092480，表示满误码。
error rate	误码率，当误码率小于10 ⁻⁵ 为信号质量正常。
alos	值为0表示正常；值为1通常表示输入信号幅度过低。
times	表示打流时长。

自定义打流使用示例如下：

```
# 开启Device8和Device9码流测试
ascend-dmi -pc --clear --device 8,9 -q
# Device8和Device9，发送端为lane0和lane1，码型为prbs20；接收端为lane2和lane3，码型为prbs23
ascend-dmi -pc --prbs-mode EN -q --device 8,9 --generator-pattern prbs20 --generator-lanes 0,1--checker-
pattern prbs23 --checker-lanes 2,3
# 展示Device8和Device9码流测试结果
ascend-dmi -pc --show-diagnostic-info -d 8,9 -q
# 关闭Device8和Device9上的打流
ascend-dmi -pc --prbs-mode DS -d 8,9 -q
# 清空Device8和Device9上的打流结果
ascend-dmi -pc --clear-diagnostic-info -d 8,9 -q
```

图 8-10 自定义打流示例

```
[root@dev ~]# ascend-dmi -pc --clear --device 8,9 -q
Operation succeeded.
[root@dev ~]# ascend-dmi -pc --prbs-mode EN -q --device 8,9 --generator-pattern prbs20 --generator-lanes 0
1 --checker-pattern prbs23 --checker-lanes 2,3
Operation succeeded.
[root@dev ~]# ascend-dmi -pc --show-diagnostic-info -d 8,9 -q
Device 8:
-----
Lane      Check Enable  Pattern  Error-Bits  Bit-Error Rate(BER)  ALOS  Period(ms)
-----
0          0             -         0            0%                   0      221554
1          0             -         0            0%                   0      221554
2          1            PRBS23    67092480    0.0021633783%       0      58378
3          1            PRBS23    67092480    0.0021634251%       0      58377
-----
Device 9:
-----
Lane      Check Enable  Pattern  Error-Bits  Bit-Error Rate(BER)  ALOS  Period(ms)
-----
0          0             -         0            0%                   0      221542
1          0             -         0            0%                   0      221543
2          1            PRBS23    67092480    0.0023884590%       0      52877
3          1            PRBS23    67092480    0.0023885731%       0      52874
-----
[root@dev ~]# ascend-dmi -pc --prbs-mode DS -d 8,9 -q
Operation succeeded.
[root@dev ~]# ascend-dmi -pc --clear-diagnostic-info -d 8,9 -q
Operation succeeded.
[root@dev ~]#
```

表 8-16 自定义打流回显参数说明

参数	说明
Lane	对应RoCE链路的lane id。
Check Enable	接收端的check状态。0：关闭，1：开启
Pattern	RX方向check的码型。
Error-Bits	误码数，上限为67092480（满误码）。
Bit-Error Rate (BER)	误码率，误码数÷总传输bit数×100%。
ALOS	正常打流时需要为0，为1通常表示信号幅度过低；未打流时无意义无需关注。
Period	距离上一次操作控制打流/读取check结果的时间。

性能测试六：软硬件版本兼容性测试

软硬件兼容性工具会获取硬件信息、架构、驱动版本、固件版本以及软件版本。软硬件兼容性测试的可用参数说明见[表8-17](#)。

```
ascend-dmi -c -h
```

表 8-17 软硬件版本兼容性测试说明

参数	说明	是否必填
[-c, --compatible]	使用该参数进行软硬件版本兼容性检测。 - 如果已安装驱动22.0.0或CANN 6.2.RC1及其以后的版本，执行“-c”参数时，会对NPU固件和驱动、驱动和CANN进行兼容性检测。 - 如果驱动为22.0.0之前的版本且CANN为6.2.RC1之前的版本，执行“-c”参数时，会检测对应的驱动、固件和软件包是否安装。	是
[-p, --path]	用户指定检测兼容性的CANN软件包的安装路径，如果不指定，将根据默认安装路径进行测试。 指定软件包安装路径的命令示例： ascend-dmi -c -p /home/xxx/Ascend	否

软硬件版本兼容性测试使用示例如下：

```
ascend-dmi -c
```


图 8-11 软硬件版本兼容性测试示例

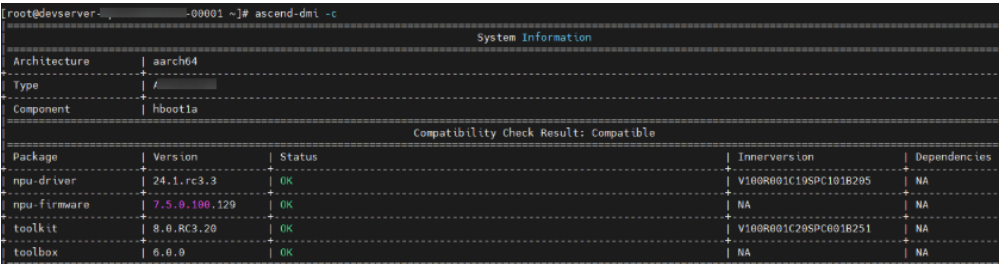


表 8-18 软硬件版本兼容性测试回显参数说明

参数	说明
System Information	系统信息
Architecture	架构
Type	标卡型号/芯片型号
Compatibility Check Result	兼容性检测结果
Package	包名
Version	版本
Status	状态，会返回如下状态： - OK：兼容 - INCOMPATIBLE PACKAGE：不兼容 - NA：未知状态，可能是获取软件版本失败导致说明非root用户不支持固件兼容性查询，npu-firmware状态会显示为NA。
Innerversion	内部版本号
Dependencies	依赖

故障诊断

查看故障诊断命令可用参数。

```
ascend-dmi --dg -h
```

表 8-19 故障诊断参数说明

参数	说明	是否必填
[-dg, --dg, --diagnosis]	使用该参数进行整卡的故障诊断测试。	是

参数	说明	是否必填
[-i, --items]	指定具体的诊断检查项。 - 可指定driver、cann、device、network、bandwidth、aiflops、hbm、signalQuality中的一项或多项，多项时各项之间使用“,”分隔。 - 不传入此参数，则默认执行除aicore和prbs外其他检查项的诊断。	否
[-d, --device]	指定需要进行诊断测试的Device ID，Device ID是指昇腾芯片的逻辑ID。 - 可指定一个或多个Device ID，多个时各项之间使用“,”分隔。 - 如果不填写Device ID则默认返回所有Device的诊断结果。	否
[-r, --result]	指定压测结果和信息采集结果的保存路径，如：/test。指定的路径需符合安全要求，且不支持包含通配符“*”。 - 如果用户指定结果保存路径，则在指定路径创建ascend_check文件夹，root用户指定的路径，将创建在根目录下，非root用户则创建在其\$HOME下； - 如果不指定路径，则保存在默认路径下，root用户：“/var/log/ascend_check”，非root用户：“\$HOME/var/log/ascend_check”。	否
[-s, --stress]	使用该参数进行压力测试，当前支持指定的压力测试有以下几种：片上内存压测、Aicore压测、P2P压测、功耗压测。 - 在包含片上内存和功耗的场景下，支持与--st参数一起使用，执行压测的时间以--st指定的时间为准。 - 在包含Aicore检查项的场景下，支持与--sc参数一起使用，执行压测的次数以--sc指定的次数为准。 - 当items参数指定bandwidth时，支持与--t参数一起使用，表示进行P2P压测。	否
[-st, --st, --stress-time]	指定EDP、TDP压力测试的时间。 - 取值范围是[60, 604800]，单位为秒。 - 需要在包含EDP、TDP压测检查项的场景下，与[-s, --stress]配合使用。 - 需要在包含片上内存诊断检查项的场景下，与[-s, --stress]配合使用。	否
[-fmt, --fmt, --format]	指定输出格式，可以为normal或json。 - 如果未指定则默认为normal。 - 当[-fmt, --fmt, --format]后检查项指定json格式输出时，会进行压测结果保存，结果保存在“ascend_check/environment_check_before.txt”文件中，不指定json格式输出时，不保存故障诊断结果。	否

参数	说明	是否必填
[-h, --help]	查看故障诊断命令的可用参数。	否

故障诊断一：NetWork 诊断

对网络健康状态进行诊断，并输出诊断结果。

使用示例，对Device0网络健康状态进行诊断
ascend-dmi -dg -i network -d 0

图 8-12 NetWork 诊断示例

```
[root@devse ~]# ascend-dmi -dg -i network -d 0
Summary:
  Arch: aarch64
  Mode: A
  Time: 20250315-17:39:04
Hardware:
  network:
    PASS
```

回显参数说明如下：

- PASS：网络检测结果健康。
- SKIP：当前产品形态不支持该项检测。
- INFO：网络检测结果提示。
- WARN：网络检测结果告警。
- FAIL：网络检测结果失败。

故障诊断二：SignalQuality 诊断

对信号质量进行诊断，并输出诊断结果。

使用示例，SignalQuality诊断
ascend-dmi -dg -i signalQuality -q

图 8-13 SignalQuality 诊断示例

```
[root@dev ~]# ascend-dmi -dg -i signalQuality
Summary:
  Arch: aarch64
  Mode: A
  Time: 20250315-17:40:19
Hardware:
  signalQuality:
    PASS
```

回显参数说明如下：

- PASS：检测通过，NPU上HCCS和RoCE通信端口的信号质量正常。

- SKIP：当前设备不支持眼图诊断。
- IMPORTANT_WARN：重要警告。HCCS和RoCE（其中的一项或多项）信号质量有异常，请联系华为工程师处理。
- FAIL：眼图检测执行失败。

故障诊断三：片上内存诊断

对高带宽内存进行诊断，并输出诊断结果。

```
# 使用示例，片上内存诊断  
ascend-dmi -dg -i hbm
```

图 8-14 片上内存诊断示例

```
[root@dev ~]# ascend-dmi -dg -i hbm  
Summary:  
  Arch: aarch64  
  Mode: A  
  Time: 20250317-10:13:10  
  
Hardware:  
  hbm:  
    PASS
```

表 8-20 片上诊断回显参数说明

回显状态	含义
PASS	片上内存检测通过，无异常。
SKIP	当前硬件形态不支持片上内存检测。
GENERAL_WARN	历史多比特存在隔离页，告警NPU芯片健康管理故障码为0x80E18401，可以继续使用。
IMPORTANT_WARN	当前实时隔离页数与已隔离页数存在差异，必须进行重启，复位npu芯片。
EMERGENCY_WARN	• 历史多比特隔离页数及设备隔离行过多，告警NPU芯片健康管理故障码为0x80E18402，建议更换备件。 • 相同Stack及PC内的隔离行处于不同Bank的数量 ≥ 4，当前设备运行存在高风险，建议更换备件。 • 相同Stack、相同Sid及不同PC内的隔离行 ≥ 4，当前设备运行存在高风险，建议更换备件。 • 相同Stack、Sid、PC及Bank内的隔离行 > 16，当前设备运行存在高风险，建议更换备件。 • 相同Stack、Sid、PC及Bank内，排除4bit及以内相邻的错误地址，其他不同地址的数量 > 5，当前设备运行存在高风险，建议更换备件。
FAIL	片上内存检测失败，请联系华为工程师处理

故障诊断四：片上内存压测

对高带宽内存进行压力测试，并输出诊断结果。

```
# 使用示例，
ascend-dmi -dg -i hbm -s -st 60 -q
```

图 8-15 片上内存诊断示例

```
[root@dev ~]# ascend-dmi -dg -i hbm -s -st 60 -q
Stress test is being performed, please wait.
Summary:
  Arch: aarch64
  Mode: 
  Time: 20250317-10:19:40
Hardware:
  hbm:
    PASS
```

回显参数说明：

- PASS：片上内存压测通过。
- SKIP：当前设备不支持片上内存压测。
- FAIL：片上内存压测失败，有新增的多比特隔离页；软件执行失败。

故障诊断五：片上内存高危地址压测

对高带宽内存高危地址进行压力测试，并输出诊断结果。

表 8-21 片上内存高危地址压测必要参数说明

参数	说明	是否必填
[-s, --stress]	使用该参数进行压力测试，当前支持指定的压力测试有以下几种：片上内存压测、Aicore压测、P2P压测、功耗压测。	是
[-qs, --qs, --quick stress]	指定高带宽内存高危地址快速压测的范围。 • 该参数取值范围为[0, 100]。参数推荐值：100。 • 取值为0时，默认对所有高带宽内存地址进行快速压测。 • 需要在包含hbm诊断检查项的场景下，与[-s, --stress]配合使用，不能和[-st, --st, --stress-time]、[--sc, --stress-count]同时使用。	是

```
# 使用示例，片上内存高危地址压测
ascend-dmi -dg -i hbm -s -qs 60 -q
```

图 8-16 片上内存高危地址压测示例

```
[root@dev ~]# ascend-dmi -dg -i hbm -s -qs 60 -q
Stress test is being performed, please wait.
Summary:
  Arch: aarch64
  Mode: A
  Time: 20250317-10:17:44
Hardware:
  hbm:
    PASS
```

回显参数说明：

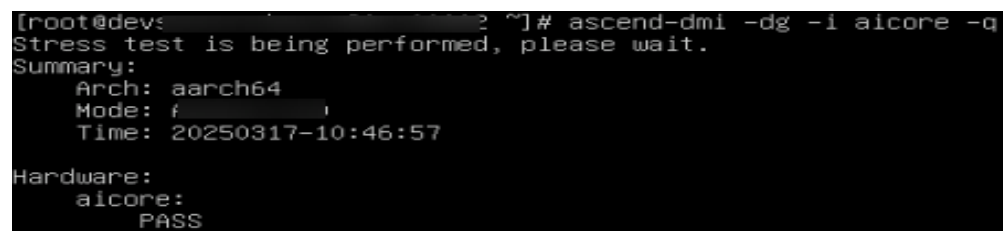
- PASS：高带宽内存高危地址快速压测通过，无新增隔离页数。
- SKIP：当前设备不支持片上内存高危地址压测。
- FAIL：高带宽内存高危地址快速压测失败，有新增隔离页数。

故障诊断六：AiCore 诊断

对AiCore ERROR进行诊断，并输出诊断结果。

```
# 使用示例，AiCore诊断  
ascend-dmi -dg -i aicore -q
```

图 8-17 AiCore 诊断示例



```
[root@dev: ~]# ascend-dmi -dg -i aicore -q  
Stress test is being performed, please wait.  
Summary:  
  Arch: aarch64  
  Mode: f  
  Time: 20250317-10:46:57  
Hardware:  
  aicore:  
    PASS
```

回显参数说明：

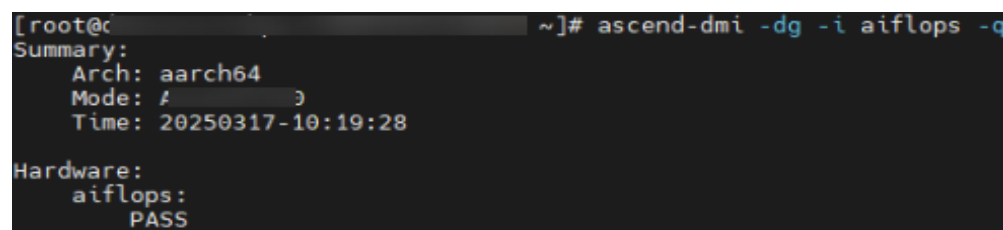
- PASS：诊断结果无异常。
- SKIP：执行诊断的用户为非root用户；当前设备不支持aicore诊断。
- EMERGENCY_WARN：紧急警告，建议更换硬件。
- FAIL：Aicore诊断失败，请联系华为工程师处理。

故障诊断七：Aiflops 诊断

对芯片进行算力诊断，并输出测试结果。

```
# 使用示例，Aiflops诊断  
ascend-dmi -dg -i aiflops -q
```

图 8-18 Aiflops 诊断示例



```
[root@dev: ~]# ascend-dmi -dg -i aiflops -q  
Summary:  
  Arch: aarch64  
  Mode: f  
  Time: 20250317-10:19:28  
Hardware:  
  aiflops:  
    PASS
```

回显参数说明：

- PASS：算力测试结果正常（大于参考值）。
- WARN：算力测试过程中触发芯片过温。
- FAIL：算力测试失败；算力测试结果小于参考值。

故障诊断八：Bandwidth 诊断

对本地带宽进行诊断，并输出诊断结果。

```
# 使用示例，对Device0进行Bandwidth诊断
ascend-dmi --dg -i bandwidth -d 0
```

图 8-19 Bandwidth 诊断示例

```
[root@devs ~]# ascend-dmi --dg -i bandwidth -d 0
This test will affect the business on this server. To ensure the correctness
Summary:
  Arch: aarch64
  Mode: /
  Time: 20250317-10:18:01

Hardware:
  bandwidth:
    PASS
```

回显参数说明：

- PASS：带宽测试结果正常。
- FAIL：带宽测试执行失败；带宽测试结果小于参考值。请联系华为工程师处理。

故障诊断九：P2P 压测

测试指定源头Device到目标Device的HCCS通信链路是否存在硬件故障，并输出测试结果。

表 8-22 P2P 压测必要参数说明

参数	说明	是否必填
[-s, --stress]	使用该参数进行压力测试，当前支持指定的压力测试有以下几种：片上内存压测、Aicore压测、P2P压测、功耗压测。 • 当items参数指定bandwidth时，支持与-s参数一起使用，表示进行P2P压测。	是
[-t, --type]	指测试数据流向的分类。 • 当item参数指定为bandwidth时，且传入-s参数时，此参数才会生效，表示执行p2p压测。 • 当前仅支持带宽类型为p2p的指定。 p2p：测试指定源头Device到目标Device的传输速率和总耗时。	是

```
# 使用示例，P2P压测
ascend-dmi -dg -i bandwidth --type p2p -s
```

图 8-20 P2P 压测示例

```
[root@ ~]# ascend-dmi -dg -i bandwidth --type p2p -s
This test will affect the business on this server. To ensure the correctness and
Summary:
  Arch: aarch64
  Mode: /
  Time: 20250317-10:25:41

Hardware:
  bandwidth:
    PASS
```

回显参数说明：

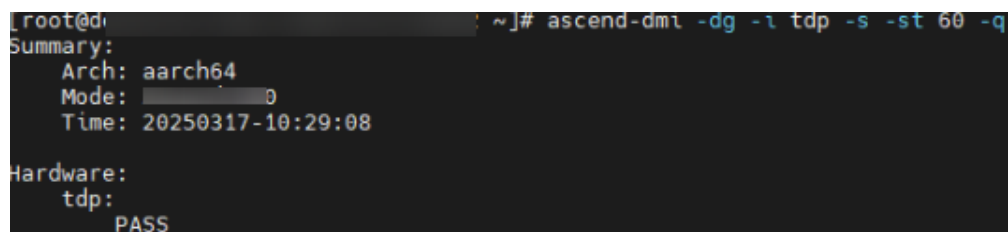
- PASS：压力测试通过，结果无异常。
- SKIP：当前设备不支持P2P压测。
- EMERGENCY_WARN：紧急警告，压测结果为不通过，建议联系华为工程师更换硬件。
- FAIL：p2p压测执行失败，请联系华为工程师处理。

故障诊断十：功耗压测

进行EDP/TDP功耗压力测试，并输出诊断结果。

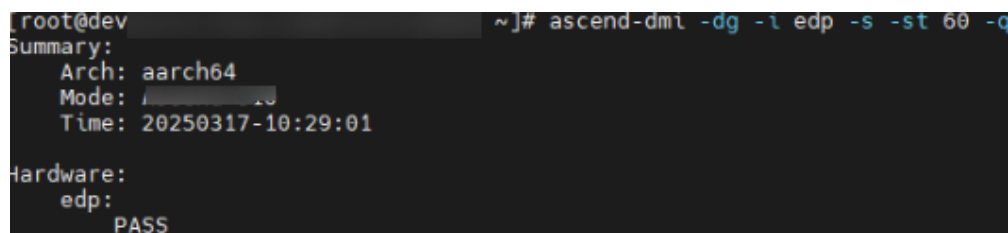
```
# 使用示例，功耗压测
ascend-dmi -dg -i edp -s -st 60 -q
ascend-dmi -dg -i tdp -s -st 60 -q
```

图 8-21 功耗压测示例（TDP）



```
root@dev: ~]# ascend-dmi -dg -i tdp -s -st 60 -q
Summary:
  Arch: aarch64
  Mode: 0
  Time: 20250317-10:29:08
Hardware:
  tdp:
    PASS
```

图 8-22 功耗压测示例（EDP）



```
root@dev: ~]# ascend-dmi -dg -i edp -s -st 60 -q
Summary:
  Arch: aarch64
  Mode: 0
  Time: 20250317-10:29:01
Hardware:
  edp:
    PASS
```

回显参数说明：

- PASS：功耗压力测试结果无异常。
- SKIP：当前设备不支持功耗压测。
- IMPORTANT_WARN：压测过程中产生芯片告警，请根据描述建议处理。如果仍无法解决，请联系华为工程师处理。
- FAIL：功耗压测功能执行失败，请联系华为工程师处理。

8.4 开启超节点 HCCL 通信算子级重执行机制

场景描述

针对Snt9b23超节点下光模块故障率高的问题，通过在HCCL通信算子级引入重执行机制，提升系统的稳定性和可靠性。

HCCL（Huawei Collective Communication Library，华为集合通信库）是华为专为昇腾（Ascend）AI处理器设计的分布式通信库，旨在优化多设备间的高效协作，以加速

深度学习模型的分布式训练，适用于需要大规模算力的AI场景。在分布式训练中，HCCL负责协调多个昇腾处理器之间的数据同步（如梯度聚合、参数更新），减少通信开销，提升训练效率。

约束限制

- 仅Snt9b23超节点支持。
- 开启算子重执行会对性能带来轻微的影响。
- 重执行依赖VPC平面（非参数面）网络进行通信域内状态协商，如果VPC平面不同，则无法重执行。
- 对于HCCS平面，如果链路没有恢复，路由未收敛，则无法重执行。
- 重执行依赖故障发生时一个通信域中所有卡都停在同一通信算子处，否则无法重执行，成功率约为95%。
- 使用inplace方式的通信算子可能导致UserIn数据被污染，从而影响重执行的可靠性。尽管重执行支持约80%通信算子的inplace方式，但对于Torch框架中的all_reduce、all_gather和reduce_scatter等算子，重执行仍不支持其inplace操作。
- RoH/RoCE平面因为闪断或断链触发的借轨，在同一通信域只允许执行一次，且不支持回切。借轨状态下，业务可持续，但应尽快保存checkpoint，维修故障。
- 对于目前昇腾的执行模式，HCCL重执行的支持范围如下：

表 8-23 HCCL 重执行的支持范围

模式	HCCL通信算子展开方式	是否支持
单算子	Stars	支持
	FFts+	支持
	Aicpu展开	支持
	通信计算融合(mc2)	不支持
图模式	全下沉模式，通信算子以展开的tasks合入图	不支持 全下沉模式，HCCL不参与图执行过程，无法进行重执行
	Aicpu展开	支持

原理说明

Snt9b23超节点的连接系统主要包含HCCS平面和RoH/RoCE平面两个数据传输平面。

在HCCS平面中，L1-1520与L2-1520之间采用光互联技术；在RoH/RoCE平面，超出NPU范围的部分均使用光互联。由于电互联域的故障率相对较低，本机制主要针对光互联域的光模块故障进行处理。具体而言：

- HCCS平面L1-1520和L2-1520之间的光模块故障。
- RoH/RoCE平面出Snt9b23超节点的光模块故障。

HCCS平面

针对HCCS平面，L1和L2之间的光模块如果发生闪断或断链，1520设备将自动完成路径切换（前提是存在多路径）。然而，断链可能导致丢包，进而引发业务中断。此时，框架层将回退至上一个checkpoint进行断点续训。通过引入HCCL重执行机制，在1520完成路径切换后，重执行功能可有效降低回退至checkpoint进行断点续训的概率，从而进一步提升业务的连续性和可靠性。

RoH/RoCE平面

针对RoH/RoCE平面，协议内置传输层重传机制，可对丢包或闪断提供一定的修复能力。然而，该机制的可靠性仍存在局限性。为提升整体可靠性，本功能在HCCL层面引入了一层重执行机制：当检测到闪断持续超过30秒或发生断链时，系统将通过建立新的传输路径（借轨），在算子级启动重执行流程，进一步保障业务的稳定运行。

参数配置（HCCL_OP_RETRY_ENABLE）

环境变量HCCL_OP_RETRY_ENABLE用于配置是否开启HCCL算子的重执行特性。重执行是指当通信算子执行报SDMA或RDMA CQE类型的错误时，HCCL会尝试重新执行此通信算子。通过此特性，可以有效避免硬件闪断导致的通信中断，提升通信稳定性。

支持在以下三个物理层级的通信域中配置重执行特性：

- L0: Server内通信域
- L1: Server间通信域
- L2: 超节点间通信域

配置方法：

在运行训练任务前，在Server节点中执行以下命令。

```
export HCCL_OP_RETRY_ENABLE="L0:0, L1:1, L2:1"
```

表 8-24 参数说明

参数	含义	取值范围	默认值	建议取值
L0	Server内通信域	• 0: Server内通信域的通信任务不开启重执行。 • 1: Server内通信域的通信任务开启重执行。	0	0
L1	Server间通信域	• 0: Server间通信域的通信任务不开启重执行，默认值为0。 • 1: Server间通信域的通信任务开启重执行。	0	1
L2	超节点间通信域	• 0: 超节点间通信域的通信任务不开启重执行，默认值为0。 • 1: 超节点间通信域的通信任务开启重执行。	0	1

注意事项：

- 当L2配置为1时，超节点间通信支持在某一Device网卡故障时使用备用Device网卡进行通信。备用网卡为同一NPU中的另一个Die网卡。
- 如果通信域的创建方式为“基于ranktable”创建通信域，需要在ranktable文件中通过"backup device ip"参数配置备用网卡。
- 如果通信域的创建方式为“基于root广播式”创建通信域，会自动将同一NPU下的两个Die互为备用网卡，无需手动配置。

参数配置（HCCL_OP_RETRY_PARAMS）

环境变量HCCL_OP_RETRY_ENABLE用于配置HCCL算子重执行的具体参数，包括最大重执行次数、第一次重执行的等待时间以及两次重执行的间隔时间。

配置示例：

```
export HCCL_OP_RETRY_PARAMS="MaxCnt:3, HoldTime:5000, IntervalTime:1000"
```

表 8-25 参数说明

参数	含义	类型	取值范围	默认值	单位	建议值
MaxCnt	最大重执行次数	uint32	[1, 10]	3	次	保持默认值3
HoldTime	从检测到通信算子执行失败到开始第一次重执行的等待时间	uint32	[0, 60000]	5000	ms	保持默认值5000
IntervalTime	两次重执行之间的间隔时间	uint32	[0, 60000]	1000	ms	保持默认值1000

使用约束：

仅当通过HCCL_OP_RETRY_ENABLE环境变量开启了HCCL的重执行特性（任一层级的重执行特性开启即可）时，此环境变量才生效。

9 Lite Server 日志采集

9.1 NPU 日志收集上传

场景描述

当Lite Server节点上的NPU出现故障时，您可通过本方案快速下发NPU日志采集任务，采集到的日志信息保存在Lite Server节点的指定目录，在用户授权同意后自动上传至华为云技术支持提供的OBS桶中，供后台进行问题定位分析。

采集日志范围包括：

- Device侧日志：Device侧Control CPU上的系统类日志、EVENT级别系统日志、非Control CPU上的系统类日志及黑匣子日志。当设备出现异常、崩溃或性能问题时，精准定位到导致问题的根本原因，缩短故障排查时间，提高运维效率。
- 主机侧日志：主机侧内核消息日志、主机侧系统监测类文件、主机侧操作系统日志文件、系统崩溃时保存的Host侧内核消息日志文件。运维人员可以第一时间查看主机监测数据，迅速判断问题是出在应用本身，还是底层的主机资源（如CPU爆满、内存耗尽、磁盘写满）导致的。精准定位避免盲目排查，极大提升排错效率。
- NPU环境日志：通过npu-smi、hccn等工具采集的日志。提升运维效率、保障系统稳定、优化资源利用，助力问题根因分析。

约束限制

- 一键日志采集功能仅支持Snt9b节点和超节点Snt9b23。手动收集日志功能支持300IDuo、Snt9b、Snt9b23。
- 同一个任务最多支持选择50个普通节点或超节点的子节点。
- 同一时间节点上最多同时支持一个任务，任务开始后无法中断，请您规划好任务优先级。
- 请确保待节点无业务运行，日志收集任务中的命令执行可能导致当前业务中断或异常。
- 请确保用于收集日志的目录空间需大于1GB。

前提条件

快速采集日志操作依赖在Lite Server节点上预安装的AI插件，如未安装该插件，请参考[安装Lite Server AI插件](#)章节完成插件安装。

一键日志采集

1. 登录[ModelArts管理控制台](#)。

2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“任务中心”。

图 9-1 任务中心



3. 单击任务中心页面左上角的“创建任务”，进入“任务模板”页面，在该页面选择“日志采集”，单击“创建任务”。

图 9-2 任务模板



4. 在日志采集任务创建页面，填写“任务名称”、“任务描述”，选择“机型”和“节点类型”，选择“采集项”，勾选使用须知并单击“立即创建”。

表 9-1 创建任务参数

参数分类	参数说明						
任务名称	系统自动填入任务名称，用户可以自定义。						
任务描述	对该任务的描述信息，方便快速查找任务。						
模板参数	请填写Lite Server上的节点目录，用于指定日志存放位置。 默认为/root/log_collection。 模板参数 <table><tr><th>参数名称</th><th>值</th><th>描述</th></tr><tr><td>log_dir</td><td> /root/log_collection </td><td>日志存储目录。</td></tr></table>	参数名称	值	描述	log_dir	/root/log_collection	日志存储目录。
参数名称	值	描述					
log_dir	/root/log_collection	日志存储目录。					

参数分类	参数说明
机型	支持Snt9b或超节点Snt9b23。
采集项	支持选择Device侧日志、主机侧日志和NPU环境日志，也可以全部选择同时采集。 - Device侧日志：Device侧Control CPU上的系统类日志、EVENT级别系统日志、非Control CPU上的系统类日志及黑匣子日志。 - 主机侧日志：主机侧内核消息日志、主机侧系统监测类文件、主机侧操作系统日志文件、系统崩溃时保存的Host侧内核消息日志文件。 - NPU环境日志：通过npu-smi、hccn等工具采集的日志。

- 勾选“日志上传”，授权后可以将采集的日志上传至OBS，用于华为云技术支持人员进行日志分析。单击“立即创建”。
- 返回“任务中心”页面，显示任务的执行状态。
- 单击具体的任务名称，可以进入任务详情页，查看任务的详细信息。
- 在任务详情页，单击“查看日志”，在页面右侧弹窗中查看任务执行的详细日志信息。所有日志收集结果会在任务日志中呈现，并说明日志收集及上传是否成功。

手动收集日志

- 获取AK/SK。该AK/SK用于后续脚本配置，做认证授权。
如果已生成过AK/SK，则可跳过此步骤，找到原来已下载的AK/SK文件，文件名一般为：credentials.csv。
如下图所示，文件包含了租户名（User Name），AK（Access Key Id），SK（Secret Access Key）。

图 9-3 credential.csv 文件内容

	A	B	C
1	User Name	Access Key Id	Secret Access Key
2	hu...dg	QTWA...UT2QVKYUC	MFyfvK41ba2...npdUKGpownRZlmVmHc

AK/SK生成步骤：

- 登录[华为云管理控制台](#)。
 - 单击右上角的用户名，在下拉列表中单击“我的凭证”。
 - 单击“访问密钥”。
 - 单击“新增访问密钥”。
 - 下载密钥，并妥善保管。
- 准备租户名ID和IAM用户名ID，用于OBS桶配置。
将您的租户名ID和IAM用户名ID提供给技术支持，技术支持将根据您提供的信息，为您配置OBS桶策略，以使用户收集的日志可以上传至对应的OBS桶。
技术支持配置完成后，会给您提供对应的OBS桶目录“obs_dir”，该目录用于后续配置的脚本中。

图 9-4 租户名 ID 和 IAM 用户名 ID

3. 准备日志收集上传脚本。

修改以下脚本中NpuLogCollection的参数，将ak、sk、obs_dir替换为前面步骤中获取到的值，如果是300IDuo机型将is_300_iduo改为True。然后把该脚本上传到要收集NPU日志的节点上。

```
import json
import os
import sys
import hashlib
import hmac
import binascii
import subprocess
import re
from datetime import datetime

class NpuLogCollection(object):
    NPU_LOG_PATH = "/var/log/npu_log_collect"
    SUPPORT_REGIONS = ['cn-southwest-2', 'cn-north-9', 'cn-east-4', 'cn-east-3', 'cn-north-4', 'cn-south-1']
    OPENSTACK_METADATA = "http://169.254.169.254/openstack/latest/meta_data.json"
    OBS_BUCKET_PREFIX = "npu-log-"

    def __init__(self, ak, sk, obs_dir, is_300_iduo=False):
        self.ak = ak
        self.sk = sk
        self.obs_dir = obs_dir
        self.is_300_iduo = is_300_iduo
        self.region_id = self.get_region_id()
        self.card_ids, self.chip_count = self.get_card_ids()

    def get_region_id(self):
        meta_data = os.popen("curl {}".format(self.OPENSTACK_METADATA))
        json_meta_data = json.loads(meta_data.read())
        meta_data.close()
        region_id = json_meta_data["region_id"]
        if region_id not in self.SUPPORT_REGIONS:
            print("current region {} is not support.".format(region_id))
            raise Exception('region exception')
        return region_id

    def gen_collect_npu_log_shell(self):
        # 300IDUO does not support
        hccn_tool_log_shell = "echo {npu_network_info}\n" \
            "for i in {npu_card_ids}; do hccn_tool -i $i -net_health -g >> {npu_log_path}/npu-smi_net-health.log ;done\n" \
            "for i in {npu_card_ids}; do hccn_tool -i $i -link -g >> {npu_log_path}/npu-smi_link.log ;done\n"
```

```
        "for i in {npu_card_ids}; do hccn_tool -i $i -tls -g |grep switch >> {npu_log_path}/
npu-smi_switch.log;done\n" \
        "for i in {npu_card_ids}; do hccn_tool -i $i -optical -g | grep prese >>
{npu_log_path}/npu-smi_present.log;done\n" \
        "for i in {npu_card_ids}; do hccn_tool -i $i -link_stat -g >> {npu_log_path}/
npu_link_history.log;done\n" \
        "for i in {npu_card_ids}; do hccn_tool -i $i -ip -g >> {npu_log_path}/
npu_roce_ip_info.log;done\n" \
        "for i in {npu_card_ids}; do hccn_tool -i $i -lldp -g >> {npu_log_path}/
npu_nic_switch_info.log;done\n" \
        .format(npu_log_path=self.NPU_LOG_PATH,
        npu_card_ids=self.card_ids,
        npu_network_info="collect npu network info")

    collect_npu_log_shell = "#!/bin/sh\n" \
        "step=1\n" \
        "rm -rf {npu_log_path}\n" \
        "mkdir -p {npu_log_path}\n" \
        "echo {echo_npu_driver_info}\n" \
        "npu-smi info > {npu_log_path}/npu-smi_info.log\n" \
        "cat /usr/local/Ascend/driver/version.info > {npu_log_path}/npu-smi_driver-
version.log\n" \
        "/usr/local/Ascend/driver/tools/upgrade-tool --device_index -1 --component -1 --
version > {npu_log_path}/npu-smi_firmware-version.log\n" \
        "for i in {npu_card_ids}; do for ((j=0;j<{chip_count};j++)); do npu-smi info -t
health -i $i -c $j; done >> {npu_log_path}/npu-smi_health-code.log;done\n" \
        "for i in {npu_card_ids}; do npu-smi info -t board -i $i >> {npu_log_path}/npu-
smi_board.log; done;\n" \
        "echo {echo_npu_ecc_info}\n" \
        "for i in {npu_card_ids};do npu-smi info -t ecc -i $i >> {npu_log_path}/npu-
smi_ecc.log; done;\n" \
        "lspci | grep acce > {npu_log_path}/Device-info.log\n" \
        "echo {echo_npu_device_log}\n" \
        "cd {npu_log_path} && msnpureport -f > /dev/null\n" \
        "tar -czvPf {npu_log_path}/log_messages.tar.gz /var/log/message* > /dev/null
\n" \
        "tar -czvPf {npu_log_path}/ascend_install.tar.gz /var/log/ascend_seclog/* > /dev/
null\n" \
        "echo {echo_npu_tools_log}\n" \
        "tar -czvPf {npu_log_path}/ascend_toollog.tar.gz /var/log/nputools_LOG_*
> /dev/null\n" \
        .format(npu_log_path=self.NPU_LOG_PATH,
        npu_card_ids=self.card_ids,
        chip_count=self.chip_count,
        echo_npu_driver_info="collect npu driver info.",
        echo_npu_ecc_info="collect npu ecc info.",
        echo_npu_device_log="collect npu device log.",
        echo_npu_tools_log="collect npu tools log.")
    if self.is_300_iduo:
        return collect_npu_log_shell
    return collect_npu_log_shell + hccn_tool_log_shell

def collect_npu_log(self):
    print("begin to collect npu log")
    os.system(self.gen_collect_npu_log_shell())
    date_collect = datetime.now().strftime('%Y%m%d%H%M%S')
    instance_ip_obj = os.popen("curl http://169.254.169.254/latest/meta-data/local-ipv4")
    instance_ip = instance_ip_obj.read()
    instance_ip_obj.close()
    log_tar = "%s-npu-log-%s.tar.gz" % (instance_ip, date_collect)
    os.system("tar -czvPf %s %s > /dev/null" % (log_tar, self.NPU_LOG_PATH))
    print("success to collect npu log with {}".format(log_tar))
    return log_tar

def upload_log_to_obs(self, log_tar):
    obs_bucket = "{}{}".format(self.OBS_BUCKET_PREFIX, self.region_id)
    print("begin to upload {} to obs bucket {}".format(log_tar, obs_bucket))
    obs_url = "https://%s.obs.%s.myhuaweicloud.com/%s/%s" % (obs_bucket, self.region_id,
self.obs_dir, log_tar)
```



```

date = datetime.utcnow().strftime('%a, %d %b %Y %H:%M:%S GMT')
canonicalized_headers = "x-obs-acl:public-read"
obs_sign = self.gen_obs_sign(date, canonicalized_headers, obs_bucket, log_tar)

auth = "OBS " + self.ak + ":" + obs_sign
header_date = "\" + "Date:" + date + "\"
header_auth = "\" + "Authorization:" + auth + "\"
header_obs_acl = "\" + canonicalized_headers + "\"

cmd = "curl -X PUT -T " + log_tar + " -w %{http_code} " + obs_url + " -H " + header_date + " -H " + header_auth + " -H " + header_obs_acl
result = subprocess.run(cmd, shell=True, capture_output=True, text=True)
http_code = result.stdout.strip()
if result.returncode == 0 and http_code == "200":
    print("success to upload {} to obs bucket {}".format(log_tar, obs_bucket))
else:
    print("failed to upload {} to obs bucket {}".format(log_tar, obs_bucket))
    print(result)

# calculate obs auth sign
def gen_obs_sign(self, date, canonicalized_headers, obs_bucket, log_tar):
    http_method = "PUT"
    canonicalized_resource = "%s/%s/%s" % (obs_bucket, self.obs_dir, log_tar)
    IS_PYTHON2 = sys.version_info.major == 2 or sys.version < '3'
    canonical_string = http_method + "\n" + "\n" + "\n" + date + "\n" + canonicalized_headers + "\n" + canonicalized_resource
    if IS_PYTHON2:
        hashed = hmac.new(self.sk, canonical_string, hashlib.sha1)
        obs_sign = binascii.b2a_base64(hashed.digest())[:-1]
    else:
        hashed = hmac.new(self.sk.encode('UTF-8'), canonical_string.encode('UTF-8'), hashlib.sha1)
        obs_sign = binascii.b2a_base64(hashed.digest())[:-1].decode('UTF-8')
    return obs_sign

# get NPU Id and Chip count
def get_card_ids(self):
    card_ids = []
    cmd = "npu-smi info -l"
    result = subprocess.run(cmd, shell=True, capture_output=True, text=True)
    if result.returncode != 0:
        print("failed to execute command{}".format(cmd))
        return ""
    match = re.search(r'Chip Count\s*:\s*(\d+)', result.stdout)
    # default chip count is 1, 300IDUO or 910C is 2
    chip_count = 1
    if match and int(match.group(1)) > 0:
        chip_count=int(match.group(1))

# filter NPU ID Regex
pattern = re.compile(r'NPU ID(?:.*?):(?:.*)\n', re.DOTALL)
matches = pattern.findall(result.stdout)
for match in matches:
    if len(match) != 2:
        continue
    id = int(match[1])
    # if drop card
    if id < 0:
        print("Card may not be found, NPU ID: {}".format(id))
        continue
    card_ids.append(id)
print("success to get card id {}, Chip Count {}".format(card_ids, chip_count))
return " ".join(str(x) for x in card_ids), chip_count

def execute(self):
    if self.obs_dir == "":
        print("the obs_dir is null, please enter a correct dir")
    else:
        log_tar = self.collect_npu_log()
        self.upload_log_to_obs(log_tar)

```

```
if __name__ == '__main__':
    npu_log_collection = NpuLogCollection(ak='ak',
                                          sk='sk',
                                          obs_dir='obs_dir',
                                          is_300_iduo=False)
    npu_log_collection.execute()
```

- 4. 执行脚本收集日志。
在节点上执行该脚本，可以看到有如下输出，代表日志收集完成并成功上传至OBS。

图 9-5 日志收集完成

```
root@ :~# python npu-log-collection.py
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
100 1778 100 1778    0     0 21682      0 --:--:-- --:--:-- --:--:-- 21682
begin to collect npu log
collect npu driver info.
collect npu ecc info.
collect npu device log.
collect npu tools log.
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
100 10 100 10    0     0 526      0 --:--:-- --:--:-- --:--:-- 526
success to collect npu log with 10.0.0.209-npu-log-20240809164759.tar.gz
begin to upload 10.0.0.209-npu-log-20240809164759.tar.gz to obs bucket npu-log-cn-north-9
success to upload 10.0.0.209-npu-log-20240809164759.tar.gz to obs bucket npu-log-cn-north-9
```

- 5. 查看在脚本的同级目录下，可以看到收集到的日志压缩包。

图 9-6 查看结果

```
root@ :~# ls
10.0.0.209-npu-log-20240809164759.tar.gz  npu-log-collection.py
```

9.2 GPU 日志收集上传

场景描述

当GPU出现故障，您可以通过本方案收集GPU的日志信息。本方案中生成的日志会保存在节点上，并自动上传至技术支持提供的OBS桶中，日志仅用于问题定位分析，因此需要您提供AK/SK给技术支持，用于授权认证。

操作步骤

- 1. 获取AK/SK。该AK/SK用于后续脚本配置，做认证授权。
如果已生成过AK/SK，则可跳过此步骤，找到原来已下载的AK/SK文件，文件名一般为：credentials.csv。
如下图所示，文件包含了租户名（User Name），AK（Access Key Id），SK（Secret Access Key）。

图 9-7 credential.csv 文件内容

	A	B	C
1	User Name	Access Key Id	Secret Access Key
2	hu dg	QTWA UT2QVKYUC	MFyfvK41ba2 npdUKGpownRZlmVmHc

AK/SK生成步骤：

- a. 登录[华为云管理控制台](#)。
 - b. 单击右上角的用户名，在下拉列表中单击“我的凭证”。
 - c. 单击“访问密钥”。
 - d. 单击“新增访问密钥”。
 - e. 下载密钥，并妥善保管。
2. 准备租户名ID和IAM用户名ID，用于OBS桶配置。
将您的租户名ID和IAM用户名ID提供给华为技术支持，技术支持将根据您提供的信息，为您配置OBS桶策略，以使用户收集的日志可以上传至对应的OBS桶。
技术支持配置完成后，会给您提供对应的OBS桶目录“obs_dir”，该目录用于后续配置的脚本中。

图 9-8 租户名 ID 和 IAM 用户名 ID

扫码登录

密码登录

IAM用户登录

租户名/原华为云帐号

IAM用户名/邮件地址

IAM用户密码

登录

[忘记密码](#)☐ [记住登录名](#)

3. 准备日志收集上传脚本。
修改以下脚本中GpuLogCollection的参数，将ak、sk、obs_dir替换为前面步骤中获取到的值。然后把该脚本上传到要收集GPU日志的节点上。

```
import json
import os
```

```
import sys
import hashlib
import hmac
import binascii
from datetime import datetime
class GpuLogCollection(object):
    GPU_LOG_PATH = "nvidia-bug-report.log.gz"
    SUPPORT_REGIONS = ['cn-north-4', 'cn-north-9']
    OPENSTACK_METADATA = "http://169.254.169.254/openstack/latest/meta_data.json"
    OBS_BUCKET_PREFIX = "npu-log-"
    def __init__(self, ak, sk, obs_dir):
        self.ak = ak
        self.sk = sk
        self.obs_dir = obs_dir
        self.region_id = self.get_region_id()
    def get_region_id(self):
        meta_data = os.popen("curl {}".format(self.OPENSTACK_METADATA))
        json_meta_data = json.loads(meta_data.read())
        meta_data.close()
        region_id = json_meta_data["region_id"]
        if region_id not in self.SUPPORT_REGIONS:
            print("current region {} is not support.".format(region_id))
            raise Exception('region exception')
        return region_id
    def gen_collect_gpu_log_shell(self):
        collect_gpu_log_shell = "nvidia-bug-report.sh"
        return collect_gpu_log_shell
    def collect_gpu_log(self):
        print("begin to collect gpu log")
        os.system(self.gen_collect_gpu_log_shell())
        date_collect = datetime.now().strftime('%Y%m%d%H%M%S')
        instance_ip_obj = os.popen("curl http://169.254.169.254/latest/meta-data/local-ipv4")
        instance_ip = instance_ip_obj.read()
        instance_ip_obj.close()
        log_tar = "%s-gpu-log-%s.gz" % (instance_ip, date_collect)
        os.system("cp %s %s" % (self.GPU_LOG_PATH, log_tar))
        print("success to collect gpu log with {}".format(log_tar))
        return log_tar
    def upload_log_to_obs(self, log_tar):
        obs_bucket = "{}{}".format(self.OBS_BUCKET_PREFIX, self.region_id)
        print("begin to upload {} to obs bucket {}".format(log_tar, obs_bucket))
        obs_url = "https://%s.obs.%s.myhuaweicloud.com/%s/%s" % (obs_bucket, self.region_id,
self.obs_dir, log_tar)
        date = datetime.utcnow().strftime('%a, %d %b %Y %H:%M:%S GMT')
        canonicalized_headers = "x-obs-acl:public-read"
        obs_sign = self.gen_obs_sign(date, canonicalized_headers, obs_bucket, log_tar)
        auth = "OBS " + self.ak + ":" + obs_sign
        header_date = "\"\" + "Date:" + date + "\"\"
        header_auth = "\"\" + "Authorization:" + auth + "\"\"
        header_obs_acl = "\"\" + canonicalized_headers + "\"\"
        cmd = "curl -X PUT -T " + log_tar + " " + obs_url + " -H " + header_date + " -H " + header_auth
+ " -H " + header_obs_acl
        os.system(cmd)
        print("success to upload {} to obs bucket {}".format(log_tar, obs_bucket))
    # calculate obs auth sign
    def gen_obs_sign(self, date, canonicalized_headers, obs_bucket, log_tar):
        http_method = "PUT"
        canonicalized_resource = "/" + obs_bucket + "/" + self.obs_dir + "/" + log_tar
        IS_PYTHON2 = sys.version_info.major == 2 or sys.version < '3'
        canonical_string = http_method + "\n" + "\n" + "\n" + date + "\n" + canonicalized_headers +
"\n" + canonicalized_resource
        if IS_PYTHON2:
            hashed = hmac.new(self.sk, canonical_string, hashlib.sha1)
            obs_sign = binascii.b2a_base64(hashed.digest())[:-1]
        else:
            hashed = hmac.new(self.sk.encode('UTF-8'), canonical_string.encode('UTF-8'), hashlib.sha1)
            obs_sign = binascii.b2a_base64(hashed.digest())[:-1].decode('UTF-8')
        return obs_sign
    def execute(self):
```

```
log_tar = self.collect_gpu_log()
self.upload_log_to_obs(log_tar)
if __name__ == '__main__':
    gpu_log_collection = GpuLogCollection(ak='ak',
                                          sk='sk',
                                          obs_dir='xxx')
    gpu_log_collection.execute()
```

4. 执行脚本收集日志。

在节点上执行该脚本，可以看到有如下输出，代表日志收集完成并成功上传至 OBS。

图 9-9 日志收集完成

```
root@dev:~# python3 log.py
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           %             %         Dload  Upload   Total   Spent    Left   Speed
100 1710 100 1710    0     63333      0 --:--:-- --:--:-- --:--:-- 63333
begin to collect gpu log

nvidia-bug-report.sh will now collect information about your
system and create the file 'nvidia-bug-report.log.gz' in the current
directory. It may take several seconds to run. In some
cases, it may hang trying to capture data generated dynamically
by the Linux kernel and/or the NVIDIA kernel module. While
the bug report log file will be incomplete if this happens, it
may still contain enough data to diagnose your problem.

If nvidia-bug-report.sh hangs, consider running with the --safe-mode
and --extra-system-data command line arguments.

Please include the 'nvidia-bug-report.log.gz' log file when reporting
your bug via the NVIDIA Linux forum (see forums.developer.nvidia.com)
or by sending email to 'linux-bugs@nvidia.com'.

By delivering 'nvidia-bug-report.log.gz' to NVIDIA, you acknowledge
and agree that personal information may inadvertently be included in
the output. Notwithstanding the foregoing, NVIDIA will use the
output only for the purpose of investigating your reported issue.

Running nvidia-bug-report.sh... complete.

% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           %             %         Dload  Upload   Total   Spent    Left   Speed
100 12 100 12    0    521      0 --:--:-- --:--:-- --:--:-- 521
success to collect gpu log with 192.168.0.23-gpu-log-20250206092139.gz
begin to upload 192.168.0.23-gpu-log-20250206092139.gz to obs bucket c30049938
success to upload 192.168.0.23-gpu-log-20250206092139.gz to obs bucket c30049938
```

5. 查看在脚本的同级目录下。可以看到收集到的日志压缩包。

图 9-10 查看结果

```
root@devser:~# ls
192.168.0.23-gpu-log-20250206092139.gz  log.py  nvidia-bug-report.log.gz
```

10 Lite Server 监控告警

10.1 使用 CES 监控 Lite Server NPU 资源

场景描述

Lite Server的监控能力依赖于CES云监控服务。本文主要介绍如何对接CES云监控服务，对Lite Server上的资源和事件进行监控。

约束限制

- 监控需要用到CES Agent插件，Agent有严格的资源占用限制，当资源占用超过阈值后出现Agent熔断情况，详细的资源占用说明请参考CES产品文档相关章节：[CES Agent性能说明](#)。
- 通过Ascend-dmi执行NPU压测命令可能会导致丢失部分NPU指标数据。
- 监控Agent已在Lite Server提供的公共镜像中经过充分测试，如果您使用自己的镜像，建议测试后再部署到生产环境，防止信息错误。

前提条件

Lite Server中已经安装CES Agent插件，判断是否安装CES Agent插件及安装方式请参见[安装CES Agent监控插件](#)。

Lite Server 监控方案介绍

详细监控方案介绍请参考[BMS主机监控概述](#)。除文档所列支持的镜像之外，目前还支持Ubuntu20.04。

监控指标采样周期为1分钟，请勿修改，否则可能导致功能不正常。当前监控指标项已经包含CPU、内存、磁盘、网络。在主机上安装加速卡驱动后，可以自动采集相关指标。

NPU相关指标采集功能运行依赖Linux系统工具lspci，部分事件依赖blkid、grub2-editenv系统工具，请确保这些工具功能正常。

表 10-1 监控工具

工具名称	检查方法	安装方法
lspci	在shell环境中执行lspci，能够正常查询系统中的PCI设备，示例如下： <pre>\$ sudo lspci 00:00.0 PCI bridge: Huawei Technologies Co., Ltd. HiSilicon PCIe Root Port with Gen4 (rev 21) 00:08.0 PCI bridge: Huawei Technologies Co., Ltd. HiSilicon PCIe Root Port with Gen4 (rev 21) 00:10.0 PCI bridge: Huawei Technologies Co., Ltd. HiSilicon PCIe Root Port with Gen4 (rev 21)</pre>	lspci是用于显示PCI设备信息的工具，通常包含在pciutils软件包中。大多数Linux发行版默认安装了这个软件包，所以lspci通常是预装的。如果lspci未安装，可以使用包管理器安装pciutils。 在Debian/Ubuntu系统中： <pre>sudo apt-get update sudo apt-get install pciutils</pre> 在Red Hat/CentOS/EulerOS系统中： <pre>sudo yum install pciutils</pre>
blkid	在shell环境中执行blkid，能够查询系统中的块设备信息，示例如下： <pre>\$ sudo blkid /dev/sda1: UUID="123e4567-e89b-12d3-a456-426614174000" TYPE="vfat" PARTUUID="56789abc-def0-1234-5678-9abcd3f2c0a1" /dev/sda2: UUID="a1b2c3d4-e5f6-789a-bcde-f0123456789a" TYPE="swap" PARTUUID="edcba98-7654-3210-fedc-ba9876543210" /dev/sda3: UUID="01234567-89ab-cdef-0123-456789abcdef" TYPE="ext4" PARTUUID="fedcba09-8765-4321-fedc-ba0987654321"</pre>	blkid是Linux系统中用于显示块设备属性的工具，通常包含在util-linux软件包中。大多数Linux发行版默认安装了这个软件包，所以blkid通常是预装的。如果blkid未安装，可以使用包管理器安装util-linux。 在Debian/Ubuntu系统中： <pre>sudo apt-get update sudo apt-get install util-linux</pre> 在Red Hat/CentOS/EulerOS系统中： <pre>sudo yum install util-linux</pre>
grub2-editenv (仅Red Hat、CentOS、EulerOS发行版需要)	在shell环境中执行blkid，能够查询系统中的块设备信息，示例如下： <pre>1 2 3 4 \$ sudo grub2-editenv list timeout=5default=0saved_entry=Red Hat Enterprise Linux Server, with Linux 4.18.0-305.el8.x86_64</pre>	grub2-editenv是GRUB2的一部分，用于管理GRUB环境变量。大多数Linux发行版默认安装了GRUB2，所以grub2-editenv通常是预装的。如果grub2-editenv未安装，可以使用包管理器安装grub2-editenv： 在Debian/Ubuntu系统中： <pre>sudo apt-get update sudo apt-get install grub2</pre> 在Red Hat/CentOS/EulerOS系统中： <pre>sudo yum install grub2</pre>

安装 CES Agent 监控插件

通过在 Lite Server（ECS 或 BMS）中安装 CES 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。

Lite Server 预置的操作系统中会默认安装 CES Agent 插件，此时在 CES 界面可以查看 Agent 插件状态和版本。

如果未安装 CES Agent 或者 CES Agent 版本不符合要求可以参考以下两种方式处理。

方式一：自动安装升级 Lite Server 中的 CES Agent 插件。

方式二：手动安装 CES Agent 插件，具体步骤如下：

1. 当前账户需要给 CES 授权委托，请参考[创建用户并授权使用云监控服务](#)。如果在[创建 Lite Server 资源](#)时，开启了“CES 主机监控授权”，此处无需重复执行授权操作。
2. 当前暂不支持在 CES 界面直接一键安装监控，需要登录到服务器上执行以下命令安装配置 Agent。其它 region 的安装请参考[单台主机下安装 Agent](#)。

```
cd /usr/local && curl -k -O https://obs.cn-north-4.myhuaweicloud.com/uniagent-cn-north-4/script/agent_install.sh && bash agent_install.sh
```

安装成功的标志如下：

图 10-1 安装成功提示

```
telescope/linux_arm64_bin/  
telescope/linux_arm64_bin/uninstall_not_root.sh  
telescope/linux_arm64_bin/telescope  
telescope/linux_arm64_bin/install.sh  
telescope/linux_arm64_bin/install_not_root.sh  
telescope/linux_arm64_bin/telescoped  
telescope/linux_arm64_bin/uninstall.sh  
telescope/linux_arm64_bin/tools/  
telescope/linux_arm64_bin/tools/hioadm  
telescope/linux_arm64_bin/tools/nvme  
telescope/linux_arm64_bin/tools/storcli64  
telescope/linux_arm64_bin/tools/sas3ircu  
telescope/manifest.json  
telescope/telescope-2.4.8-release.json  
telescope/linux_amd64_bin/  
telescope/linux_amd64_bin/uninstall_not_root.sh  
telescope/linux_amd64_bin/telescope  
telescope/linux_amd64_bin/install.sh  
telescope/linux_amd64_bin/install_not_root.sh  
telescope/linux_amd64_bin/telescoped  
telescope/linux_amd64_bin/uninstall.sh  
telescope/linux_amd64_bin/tools/  
telescope/linux_amd64_bin/tools/hioadm  
telescope/linux_amd64_bin/tools/nvme  
telescope/linux_amd64_bin/tools/storcli64  
telescope/linux_amd64_bin/tools/sas3ircu  
telescope/conf/  
telescope/conf/custom_conf.json  
telescope/windows_bin/  
telescope/windows_bin/uninstall.bat  
telescope/windows_bin/shutdown.bat  
telescope/windows_bin/telescope.exe  
telescope/windows_bin/install.bat  
telescope/windows_bin/start.bat  
telescope/windows_bin/getpid.bat  
telescope/config/  
telescope/config/conf_ces.json  
telescope/config/logs_config.xml  
telescope/config/conf.json  
Current user is root.  
Start to install telescope...  
instance_type: physical.p7vs.8xlarge.ei  
Starting telescope...  
Telescope process starts successfully.
```

3. 在CES界面查看具体的监控项，加速卡类的监控项必须在主机安装加速卡驱动后才会有相关指标。

表 10-2 NPU 指标列表（整体）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CE S Agent 版本
1	整体	npu_device_health	NPU 健康状况	NPU 卡的健康状况	-	不涉及	0: 正常 1: 一般告警 2: 重要告警 3: 紧急告警	instance_id, npu	Snt3P 300I Duo Snt9b Snt9b23	tel esc op e: 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_driver_health	NPU 驱动健康状况	NPU 卡的驱动的健康状况	-	不涉及	0: 正常 3: 紧急告警	instance_id, npu		
3		npu_power	NPU 功率	NPU 卡功率	W	不涉及	>0	instance_id, npu		
4		npu_temperature	NPU 温度	NPU 卡温度	°C	不涉及	自然数	instance_id, npu		
5		npu_voltage	NPU 电压	该指标描述 NPU 的电压	V	不涉及	自然数	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
6		npu_util_rate_general	NPU 整体利用率	NPU 整体利用率，包括对AI Core 和 Vector Core 的整体统计。	%	不涉及	0 ~ 100 %	instance_id, npu	Snt9b Snt9b23	

表 10-3 NPU 指标列表（HBM）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CE S Agent 版本
1	HBM	npu_util_rate_hbm	NPU 的 HBM 占用率	该指标描述 NPU 的 HBM 占用率	%	不涉及	0~100%	instance_id, npu	Snt9b Snt9b23	tel esc op e: 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_hbm_freq	HBM 频率	NPU 卡 HBM 频率	MHz	不涉及	>0	instance_id, npu		
3		npu_freq_hbm	HBM 频率	NPU 卡 HBM 频率	MHz	不涉及	>0	instance_id, npu		
4		npu_hbm_usage	HBM 使用量	NPU 卡 HBM 使用量	MB	不涉及	≥0	instance_id, npu		
5		npu_hbm_temperature	HBM 温度	NPU 卡 HBM 温度	℃	不涉及	自然数	instance_id, npu		
6		npu_hbm_bandwidth_util	HBM 带宽利用率	NPU 卡 HBM 带宽利用率（旧版指标）	%	不涉及	0~100%	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
7		npu_util_rate_hbm_bw	HBM带宽利用率	NPU卡HBM带宽利用率（新版指标）	%	不涉及	0~100%	instance_id, npu		
8		npu_hbm_mem_capacity	NPU的HBM内存容量	该指标描述NPU的HBM内存容量	MB	不涉及	≥0	instance_id, npu		
9		npu_hbm_ecc_enable	HBM ECC开关状态	NPU卡HBM ECC开关状态	-	不涉及	0: ecc检测未使能 1: ecc检测使能	instance_id, npu		
10		npu_hbm_single_bit_error_cnt	HBM当前单bit错误数量	NPU卡HBM当前单bit错误数量	count	不涉及	≥0	instance_id, npu		
11		npu_hbm_double_bit_error_cnt	HBM当前双bit错误数量	NPU卡HBM当前双bit错误数量	count	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
12		npu_hbm_total_single_bit_error_cnt	HBM生命周期内单bit错误数量	NPU卡HBM生命周期内单bit错误数量	count	不涉及	≥0	instance_id, npu		
13		npu_hbm_total_double_bit_error_cnt	HBM生命周期内双bit错误数量	NPU卡HBM生命周期内双bit错误数量	count	不涉及	≥0	instance_id, npu		
14		npu_hbm_single_bit_isolated_pages_cnt	HBM单比特错误隔离内存页数	NPU卡HBM单比特错误隔离内存页数	count	不涉及	≥0	instance_id, npu		
15		npu_hbm_double_bit_isolated_pages_cnt	HBM多比特错误隔离内存页数	NPU卡HBM多比特错误隔离内存页数	count	不涉及	≥0	instance_id, npu		

表 10-4 NPU 指标列表（DDR）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	DDR	npu_usage_mem	NPU 显存使用量	NPU 卡的显存使用量	MB	不涉及	≥0	instance_id, npu	Snt3P 300I Duo	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_util_rate_mem	NPU 显存使用率	NPU 卡的显存使用率	%	不涉及	0 ~ 100 %	instance_id, npu		
3		npu_freq_mem	NPU 显存频率	NPU 卡的显存频率	MHz	不涉及	>0	instance_id, npu		
4		npu_util_rate_mem_bandwidth	NPU 显存带宽使用率	NPU 卡的显存带宽使用率	%	不涉及	0 ~ 100 %	instance_id, npu		
5		npu_sbe	NPU 单bit 错误数量	NPU 卡单比特错误数量	count	不涉及	≥0	instance_id, npu		
6		npu_dbe	NPU 双bit 错误数量	NPU 卡双比特错误数量	count	不涉及	≥0	instance_id, npu		

表 10-5 NPU 指标列表（AI Core）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	AI Core	npu_freq_ai_core	NPU 卡AI 核心频率	NPU 卡的 AI核心时钟频率	MHz	不涉及	>0	instance_id, npu	Snt3P300I Duo Snt9b Snt9b23	tele scope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_freq_ai_core_rated	NPU 的AI 核心额定频率	该指标描述 NPU 的AI 核心额定频率	MHz	不涉及	>0	instance_id, npu		
3		npu_util_rate_ai_core	NPU 卡AI 核心使用率	NPU 卡的 AI核心使用率	%	不涉及	0 ~ 100 %	instance_id, npu		

表 10-6 NPU 指标列表（AI Vector）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	AI Vector	npu_util_rate_vector_core	NPU 卡 Vector 核心使用率	NPU 卡 Vector 核心使用率	%	不涉及	0 ~ 100 %	instance_id, npu	Snt3P 300I Duo Snt9b Snt9b23	tele scope : 2.7.5.9 及之后版本

表 10-7 NPU 指标列表（AI CPU）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	AI CPU	npu_aicpu_num	NPU 的 AI CPU 数量	该指标描述 NPU 的 AI CPU 数量	count	不涉及	≥0	instance_id, npu	Snt3P 300I Duo Snt9b Snt9b23	tele scope : 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_util_rate_aicpu	NPU 卡 AI CPU 使用率	NPU 卡的 AI CPU 使用率	%	不涉及	0 ~ 100 %	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
3		npu_aicpu_avg_util_rate	NPU的AICPU平均使用率	该指标描述NPU的AICPU平均使用率	%	不涉及	0~100%	instance_id, npu		
4		npu_aicpu_max_freq	NPU的AICPU最大频率	该指标描述NPU的AICPU最大频率	MHz	不涉及	>0	instance_id, npu		
5		npu_aicpu_cur_freq	NPU的AICPU频率	该指标描述NPU的AICPU频率	MHz	不涉及	>0	instance_id, npu		

表 10-8 NPU 指标列表（CTRL CPU）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	CTRL CPU	npu_util_rate_ctrl_cpu	NPU 控制 CPU 使用率	该指标描述 NPU 卡的控制 CPU 使用率	%	不涉及	0 ~ 100 %	instance_id, npu	Snt3P 300I Duo Snt9b Snt9b23	tele scope : 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_freq_ctrl_cpu	NPU 的控制 CPU 频率	该指标描述 NPU 的控制 CPU 频率	MHz	不涉及	>0	instance_id, npu		

表 10-9 NPU 指标列表（PCIE 链路）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	PCIE 链路	npu_link_cap_speed	NPU 链路最大传输速度	该指标描述 NPU 设备支持的最大传输速度	GT/s	不涉及	≥0	instance_id, npu	310P 300I Duo Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_link_cap_width	NPU 链路最大传输宽度	该指标描述 NPU 设备支持的最大传输宽度	count	不涉及	≥0	instance_id, npu		
3		npu_link_status_speed	NPU 链路当前传输速度	该指标描述 NPU 设备链路的实际传输速度	GT/s	不涉及	≥0	instance_id, npu		
4		npu_link_status_width	NPU 链路当前传输宽度	该指标描述 NPU 设备链路的实际传输宽度	count	不涉及	≥0	instance_id, npu		

表 10-10 NPU 指标列表（RoCE 网络）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	RoCE 网络	npu_device_network_health	NPU 网络健康情况	NPU 卡的 RoCE 网卡的 IP 地址连通情况	-	不涉及	0: 网络健康状态正常 非 0: 网络状态异常	instance_id, npu	Snt9b Snt9b23	tele scope : 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_network_port_link_statuses	NPU 网口 link 状态	NPU 卡的对应网口 link 状态	-	不涉及	0: UP 1: DOWN	instance_id, npu		
3		npu_roce_tx_rate	NPU 网卡上行速率	NPU 卡内网卡的 上行速率	MB/s	不涉及	≥0	instance_id, npu		
4		npu_roce_rx_rate	NPU 网卡下行速率	NPU 卡内网卡的 下行速率	MB/s	不涉及	≥0	instance_id, npu		
5		npu_mac_tx_mac_pause_num	MAC 发送 pause 帧总数	NPU 卡对应 MAC 地址发送的 pause 帧总报文数	count	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
6		npu_mac_rx_mac_pause_num	MAC接收pause帧总数	NPU卡对应MAC地址接收的pause帧总报文数	count	不涉及	≥0	instance_id, npu		
7		npu_mac_tx_pfc_pkt_num	MAC发送pfc帧总数	NPU卡对应MAC地址发送的PFC帧总报文数	count	不涉及	≥0	instance_id, npu		
8		npu_mac_rx_pfc_pkt_num	MAC接收pfc帧总数	NPU卡对应MAC地址接收的PFC帧总报文数	count	不涉及	≥0	instance_id, npu		
9		npu_mac_tx_bad_pkt_num	MAC发送坏包总数	NPU卡对应MAC地址发送的坏包总数	count	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
10		npu_mac_rx_bad_pkt_num	MAC接收坏包总数	NPU卡对应MAC地址接收的坏包总数	count	不涉及	≥0	instance_id, npu		
11		npu_roce_tx_err_pkt_num	RoCE发送坏包总数	NPU卡内RoCE网卡发送的坏包总数	count	不涉及	≥0	instance_id, npu		
12		npu_roce_rx_err_pkt_num	RoCE接收坏包总数	NPU卡内RoCE网卡接收的坏包总数	count	不涉及	≥0	instance_id, npu		
13		npu_roce_tx_all_pkt_num	NPU RoCE发送总报文数	该指标描述NPU RoCE发送的总报文数	count	不涉及	≥0	instance_id, npu		telescope： 2.7.5.9及之后版本

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
14		npu_roce_rx_all_pkt_num	NPU RoCE 接收总报文数	该指标描述 NPU RoCE 接收的总报文数	count	不涉及	≥0	instance_id, npu		
15		npu_roce_new_pkt_rty_num	NPU RoCE 的重传报文数	该指标描述 NPU RoCE 发送的重传的报文数量统计	count	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
16		npu_roce_out_of_order_num	NPU RoCE接收的PSN异常报文数	该指标描述NPU RoCE接收的PSN大于预期PSN的报文，或重复PSN报文数。乱序或丢包，会触发重传	count	不涉及	≥0	instance_id, npu		
17		npu_roce_rx_cnp_pkt_num	NPU RoCE接收的CNP类型报文数	该指标描述NPU RoCE接收的CNP类型报文数	count	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
18		npu_roce_tx_cnp_pkt_num	NPU RoCE发送的CNP类型报文数	该指标描述NPU RoCE发送的CNP类型报文数	count	不涉及	≥0	instance_id, npu		

表 10-11 NPU 指标列表（RoCE 光模块）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	RoCE 光模块	npu_opt_temperature	NPU 光模块壳温	该指标描述 NPU 光模块壳温	°C	不涉及	自然数	instance_id, npu	Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9 及之后版本
2		npu_opt_temperature_high_thres	NPU 光模块壳温上限	该指标描述 NPU 光模块壳温上限	°C	不涉及	自然数	instance_id, npu		
3		npu_opt_temperature_low_thres	NPU 光模块壳温下限	该指标描述 NPU 光模块壳温下限	°C	不涉及	自然数	instance_id, npu		
4		npu_opt_voltage	NPU 光模块供电电压	该指标描述 NPU 光模块供电电压	mV	不涉及	自然数	instance_id, npu		
5		npu_opt_voltage_high_thres	NPU 光模块供电电压上限	该指标描述 NPU 光模块供电电压上限	mV	不涉及	自然数	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
6		npu_opt_voltage_low_thresh	NPU光模块供电电压下限	该指标描述NPU光模块供电电压下限	mV	不涉及	自然数	instance_id, npu		
7		npu_opt_tx_power_lane0	NPU光模块通道0发送功率	该指标描述NPU光模块通道0发送功率	mW	不涉及	≥0	instance_id, npu		
8		npu_opt_tx_power_lane1	NPU光模块通道1发送功率	该指标描述NPU光模块通道1发送功率	mW	不涉及	≥0	instance_id, npu		
9		npu_opt_tx_power_lane2	NPU光模块通道2发送功率	该指标描述NPU光模块通道2发送功率	mW	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent 版本
10		npu_opt_tx_power_lane3	NPU光模块通道3发送功率	该指标描述NPU光模块通道3发送功率	mW	不涉及	≥0	instance_id, npu		
11		npu_opt_rx_power_lane0	NPU光模块通道0接收功率	该指标描述NPU光模块通道0接收功率	mW	不涉及	≥0	instance_id, npu		
12		npu_opt_rx_power_lane1	NPU光模块通道1接收功率	该指标描述NPU光模块通道1接收功率	mW	不涉及	≥0	instance_id, npu		
13		npu_opt_rx_power_lane2	NPU光模块通道2接收功率	该指标描述NPU光模块通道2接收功率	mW	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
14		npu_opt_rx_power_lane3	NPU光模块通道3接收功率	该指标描述 NPU光模块通道3接收功率	mW	不涉及	≥0	instance_id, npu		
15		npu_opt_tx_bias_lane0	NPU光模块通道0发射偏置电流	该指标描述 NPU光模块通道0发射偏置电流	mA	不涉及	≥0	instance_id, npu		
16		npu_opt_tx_bias_lane1	NPU光模块通道1发射偏置电流	该指标描述 NPU光模块通道1发射偏置电流	mA	不涉及	≥0	instance_id, npu		
17		npu_opt_tx_bias_lane2	NPU光模块通道2发射偏置电流	该指标描述 NPU光模块通道2发射偏置电流	mA	不涉及	≥0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
18		npu_opt_tx_bias_lane3	NPU光模块通道3发射偏置电流	该指标描述NPU光模块通道3发射偏置电流	mA	不涉及	≥0	instance_id, npu		
19		npu_opt_tx_los	NPU光模块TX Los	该指标描述NPU光模块TX Los flag	count	不涉及	≥0	instance_id, npu		
20		npu_opt_rx_los	NPU光模块RX Los	该指标描述NPU光模块RX Los flag	count	不涉及	≥0	instance_id, npu		
21		npu_opt_media_snr_lane0	NPU光模块通道0光侧信噪比	该指标描述NPU光模块通道0的media侧（光侧）的信噪比	db	不涉及	自然数	instance_id, npu		telescope： 2.7.5.9及之后版本

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
22		npu_opt_media_snr_lane1	NPU光模块通道1光侧信噪比	该指标描述NPU光模块通道1的media侧（光侧）的信噪比	db	不涉及	自然数	instance_id, npu		
23		npu_opt_media_snr_lane2	NPU光模块通道2光侧信噪比	该指标描述NPU光模块通道2的media侧（光侧）的信噪比	db	不涉及	自然数	instance_id, npu		
24		npu_opt_media_snr_lane3	NPU光模块通道3光侧信噪比	该指标描述NPU光模块通道3的media侧（光侧）的信噪比	db	不涉及	自然数	instance_id, npu		

表 10-12 NPU 指标列表（HCCS Lane 模式）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持 CES Agent 版本
1	HCCS Lane 模式	npu_macro1_0lane_max_consec_sec	NPU Macro1 0lane模式最大持续时长	该指标描述 NPU Macro1 在检测周期内处于0lane模式的最大持续时长	s	不涉及	≥ 0	instance_id, npu	Snt9b Snt9b23	telescope： 2.7.5.9 及之后版本
2		npu_macro2_0lane_max_consec_sec	NPU Macro2 0lane模式最大持续时长	该指标描述 NPU Macro2 在检测周期内处于0lane模式的最大持续时长	s	不涉及	≥ 0	instance_id, npu		
3		npu_macro3_0lane_max_consec_sec	NPU Macro3 0lane模式最大持续时长	该指标描述 NPU Macro3 在检测周期内处于0lane模式的最大持续时长	s	不涉及	≥ 0	instance_id, npu		
4		npu_macro4_0lane_max_consec_sec	NPU Macro4 0lane模式最大持续时长	该指标描述 NPU Macro4 在检测周期内处于0lane模式的最大持续时长	s	不涉及	≥ 0	instance_id, npu		
5		npu_macro5_0lane_max_consec_sec	NPU Macro5 0lane模式最大持续时长	该指标描述 NPU Macro5 在检测周期内处于0lane模式的最大持续时长	s	不涉及	≥ 0	instance_id, npu		
6		npu_macro6_0lane_max_consec_sec	NPU Macro6 0lane模式最大持续时长	该指标描述 NPU Macro6 在检测周期内处于0lane模式的最大持续时长	s	不涉及	≥ 0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
7		npu_macro7_0lane_max_consec_sec	NPU Macro7 0lane模式最大持续时长	该指标描述NPU Macro7在检测周期内处于0lane模式的最大持续时长	s	不涉及	≥ 0	instance_id, npu		
8		npu_macro1_0lane_total_sec	NPU Macro1 0lane模式持续总时长	该指标描述NPU Macro1在检测周期内处于0lane模式的持续总时长	s	不涉及	≥ 0	instance_id, npu		
9		npu_macro2_0lane_total_sec	NPU Macro2 0lane模式持续总时长	该指标描述NPU Macro2在检测周期内处于0lane模式的持续总时长	s	不涉及	≥ 0	instance_id, npu		
10		npu_macro3_0lane_total_sec	NPU Macro3 0lane模式持续总时长	该指标描述NPU Macro3在检测周期内处于0lane模式的持续总时长	s	不涉及	≥ 0	instance_id, npu		
11		npu_macro4_0lane_total_sec	NPU Macro4 0lane模式持续总时长	该指标描述NPU Macro4在检测周期内处于0lane模式的持续总时长	s	不涉及	≥ 0	instance_id, npu		
12		npu_macro5_0lane_total_sec	NPU Macro5 0lane模式持续总时长	该指标描述NPU Macro5在检测周期内处于0lane模式的持续总时长	s	不涉及	≥ 0	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
13		npu_macro6_0lane_total_sec	NPU Macro6 0lane模式持续总时长	该指标描述NPU Macro6在检测周期内处于0lane模式的持续总时长	s	不涉及	≥ 0	instance_id, npu		
14		npu_macro7_0lane_total_sec	NPU Macro7 0lane模式持续总时长	该指标描述NPU Macro7在检测周期内处于0lane模式的持续总时长	s	不涉及	≥ 0	instance_id, npu		

表 10-13 NPU 指标列表（HCCS Serdes SNR）

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
1	HCCS Serdes SNR	npu_macro1_serdes_lane0_snr	NPU Macro1 Serdes Lane0的信噪比	该指标描述NPU Macro1 Serdes Lane0的信噪比	db	不涉及	自然数	instance_id, npu	Snt9b Snt9b23	telescope： 2.7.5.9及之后版本
2		npu_macro1_serdes_lane1_snr	NPU Macro1 Serdes Lane1的信噪比	该指标描述NPU Macro1 Serdes Lane1的信噪比	db	不涉及	自然数	instance_id, npu		
3		npu_macro1_serdes_lane2_snr	NPU Macro1 Serdes Lane2的信噪比	该指标描述NPU Macro1 Serdes Lane2的信噪比	db	不涉及	自然数	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
4		npu_macro1_serdes_lane3_sn	NPU Macro1 Serdes Lane3的信噪比	该指标描述 NPU Macro1 Serdes Lane3的信噪比	db	不涉及	自然数	instance_id, npu		
5		npu_macro2_serdes_lane0_sn	NPU Macro2 Serdes Lane0的信噪比	该指标描述 NPU Macro2 Serdes Lane0的信噪比	db	不涉及	自然数	instance_id, npu		
6		npu_macro2_serdes_lane1_sn	NPU Macro2 Serdes Lane1的信噪比	该指标描述 NPU Macro2 Serdes Lane1的信噪比	db	不涉及	自然数	instance_id, npu		
7		npu_macro2_serdes_lane2_sn	NPU Macro2 Serdes Lane2的信噪比	该指标描述 NPU Macro2 Serdes Lane2的信噪比	db	不涉及	自然数	instance_id, npu		
8		npu_macro2_serdes_lane3_sn	NPU Macro2 Serdes Lane3的信噪比	该指标描述 NPU Macro2 Serdes Lane3的信噪比	db	不涉及	自然数	instance_id, npu		
9		npu_macro3_serdes_lane0_sn	NPU Macro3 Serdes Lane0的信噪比	该指标描述 NPU Macro3 Serdes Lane0的信噪比	db	不涉及	自然数	instance_id, npu		
10		npu_macro3_serdes_lane1_sn	NPU Macro3 Serdes Lane1的信噪比	该指标描述 NPU Macro3 Serdes Lane1的信噪比	db	不涉及	自然数	instance_id, npu		
11		npu_macro3_serdes_lane2_sn	NPU Macro3 Serdes Lane2的信噪比	该指标描述 NPU Macro3 Serdes Lane2的信噪比	db	不涉及	自然数	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
12		npu_macro3_serdes_lane3_sn	NPU Macro3 Serdes Lane3的信噪比	该指标描述 NPU Macro3 Serdes Lane3的信噪比	db	不涉及	自然数	instance_id, npu		
13		npu_macro4_serdes_lane0_sn	NPU Macro4 Serdes Lane0的信噪比	该指标描述 NPU Macro4 Serdes Lane0的信噪比	db	不涉及	自然数	instance_id, npu		
14		npu_macro4_serdes_lane1_sn	NPU Macro4 Serdes Lane1的信噪比	该指标描述 NPU Macro4 Serdes Lane1的信噪比	db	不涉及	自然数	instance_id, npu		
15		npu_macro4_serdes_lane2_sn	NPU Macro4 Serdes Lane2的信噪比	该指标描述 NPU Macro4 Serdes Lane2的信噪比	db	不涉及	自然数	instance_id, npu		
16		npu_macro4_serdes_lane3_sn	NPU Macro4 Serdes Lane3的信噪比	该指标描述 NPU Macro4 Serdes Lane3的信噪比	db	不涉及	自然数	instance_id, npu		
17		npu_macro5_serdes_lane0_sn	NPU Macro5 Serdes Lane0的信噪比	该指标描述 NPU Macro5 Serdes Lane0的信噪比	db	不涉及	自然数	instance_id, npu		
18		npu_macro5_serdes_lane1_sn	NPU Macro5 Serdes Lane1的信噪比	该指标描述 NPU Macro5 Serdes Lane1的信噪比	db	不涉及	自然数	instance_id, npu		
19		npu_macro5_serdes_lane2_sn	NPU Macro5 Serdes Lane2的信噪比	该指标描述 NPU Macro5 Serdes Lane2的信噪比	db	不涉及	自然数	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
20		npu_macro5_serdes_lane3_sn	NPU Macro5 Serdes Lane3的信噪比	该指标描述 NPU Macro5 Serdes Lane3的信噪比	db	不涉及	自然数	instance_id, npu		
21		npu_macro6_serdes_lane0_sn	NPU Macro6 Serdes Lane0的信噪比	该指标描述 NPU Macro6 Serdes Lane0的信噪比	db	不涉及	自然数	instance_id, npu		
22		npu_macro6_serdes_lane1_sn	NPU Macro6 Serdes Lane1的信噪比	该指标描述 NPU Macro6 Serdes Lane1的信噪比	db	不涉及	自然数	instance_id, npu		
23		npu_macro6_serdes_lane2_sn	NPU Macro6 Serdes Lane2的信噪比	该指标描述 NPU Macro6 Serdes Lane2的信噪比	db	不涉及	自然数	instance_id, npu		
24		npu_macro6_serdes_lane3_sn	NPU Macro6 Serdes Lane3的信噪比	该指标描述 NPU Macro6 Serdes Lane3的信噪比	db	不涉及	自然数	instance_id, npu		
25		npu_macro7_serdes_lane0_sn	NPU Macro7 Serdes Lane0的信噪比	该指标描述 NPU Macro7 Serdes Lane0的信噪比	db	不涉及	自然数	instance_id, npu		
26		npu_macro7_serdes_lane1_sn	NPU Macro7 Serdes Lane1的信噪比	该指标描述 NPU Macro7 Serdes Lane1的信噪比	db	不涉及	自然数	instance_id, npu		
27		npu_macro7_serdes_lane2_sn	NPU Macro7 Serdes Lane2的信噪比	该指标描述 NPU Macro7 Serdes Lane2的信噪比	db	不涉及	自然数	instance_id, npu		

序号	分类	指标名称	显示名	说明	单位	进制	取值范围	维度	支持机型	支持CES Agent版本
28		npu_macro7_serdes_lane3_sn	NPU Macro7 Serdes Lane3的信噪比	该指标描述NPU Macro7 Serdes Lane3的信噪比	db	不涉及	自然数	instance_id, npu		

10.2 使用 CES 监控 Lite Server NPU 事件

场景描述

通过对接CES，可以将业务中的重要事件或对云资源的操作事件收集到CES云监控服务，并在事件发生时进行告警。Lite Server支持的事件来源主要是BMS和ECS，NPU涉及的具体事件列表如下，其它相关事件请参考[CES事件监控说明](#)。

约束限制

- 事件上报到CES需要用到CES Agent插件，Agent有严格的资源占用限制，当资源占用超过阈值后出现Agent熔断情况，详细的资源占用说明请参考CES产品文档相关章节：[CES Agent性能说明](#)。
- 监控Agent已在Lite Server提供的公共镜像中经过充分测试，如果您使用自己的镜像，建议测试后再部署到生产环境，防止信息错误。

前提条件

Lite Server中已经安装CES Agent插件，判断是否安装CES Agent插件及安装方式请参见[安装CES Agent监控插件](#)。

事件来源

BMS/ECS

事件命名空间

SYS.BMS/SYS.ECS

事件列表

表 10-14 Lite Server 支持的故障事件列表（BMS/ECS）

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
NPU: npu-smi info查询缺少设备	NPUSMICardNotFound	重要	可能是由于昇腾驱动问题或NPU掉卡	联系运维人员协助处理	NPU卡无法正常使用	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本
NPU: PCIe链路异常	PCIEErrorFound	重要	lspci显示npu卡处于revff状态	联系运维人员协助处理	NPU卡无法正常使用	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
NPU: lspci查询缺少设备	LspciCardNotFound	重要	一般是由于NPU掉卡	联系运维人员协助处理	NPU卡无法正常使用	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本
NPU: 温度超过阈值	TemperatureOverUpperLimit	重要	可能是由于DDR颗粒温度过高或过温软件预警	暂停业务，重启系统，查看散热系统，device复位	可能造成过温下电及device丢失	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本
NPU: 存在不可纠正ECC错误	UncorrectableEccErrorCount	重要	NPU卡出现Uncorrectable ECC Error硬件故障	如果业务受到影响，转硬件换卡	业务可能受到影响终止	Snt3P Snt3PD	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
NPU: 需要重启实例	RebootVirtualMachine	提示	当前故障很可能需要重启进行恢复	在收集必要信息后，重启以尝试恢复	重启可能中断客户业务	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本
NPU: 需要复位SOC	ResetSOC	提示	当前故障很可能需要复位SOC进行恢复	在收集必要信息后，复位SOC以尝试恢复	复位SOC可能中断客户业务	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本
NPU: 需要退出AI任务重新执行	RestartAIProcesses	提示	当前故障很可能需要客户退出当前的AI任务并尝试重新执行	在收集必要信息后，尝试退出当前AI任务并尝试重新执行	退出当前AI任务以便重新执行	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
NPU: errorcode告警	NPUErrorCodeWarning	重要	这里涵盖了大量重要及以上的NPU错误码，您可以根据这些错误码进一步定位错误原因	对照《黑匣子错误码信息列表》和《健康管理故障定义》进一步定位错误	NPU当前存在故障，可能导致客户业务终止	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.4.3 2.7.5.3 2.7.5.4 2.7.5.9及之后版本
NPU HBM多ECC错误信息	NpuHbmMultiEccInfo	提示	NPU卡存在HBM的ECC错误，此事件上报相应错误信息	这是一个用于辅助其他事件进行判断的事件，无需单独定位处理	这是一个用于辅助其他事件进行判断的事件，无需单独定位处理	Snt9b Snt9b23	telescope： 2.7.5.9及之后版本
GPU: RoCE网卡配置错误	GpuRoceNicConfigIncorrect	重要	GPU: RoCE网卡配置错误	联系运维人员协助处理	机器参数面网络异常，多机任务无法执行	GPU	telescope： 2.7.5.9及之后版本
OS出现ReadOnly问题	ReadOnlyFileSystem	严重	文件系统%s只读	请检查磁盘健康状态	无法对文件进行写入和操作	-	telescope： 2.7.5.3 2.7.5.9及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
NPU：驱动固件不匹配	NpuDriverFirmwareMismatch	重要	NPU驱动固件版本不匹配	请从昇腾官网获取匹配版本重新安装	无法正常使用NPU卡	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.5.3 2.7.5.9及之后版本
NPU：RoCE网卡down	RoCELinkStatusDown	重要	NPU卡%d RoCE Link状态Down	请检查NPU Roce网口状态	NPU网卡不可用	Snt9b Snt9b23	telescope： 2.7.5.3 2.7.5.9及之后版本
NPU：RoCE网卡健康状态异常	RoCEHealthStatusError	重要	NPU卡%d RoCE网络健康状态异常	请检查NPU Roce网卡健康状态	NPU网卡不可用	Snt9b Snt9b23	telescope： 2.7.5.3 2.7.5.9及之后版本
NPU：Roce网卡配置文件/etc/hccn.conf不存在	HccnConfNotExist	重要	Roce网卡配置文件"/etc/hccn.conf"不存在	请检查/etc/hccn.conf网卡配置文件	Roce网卡不可用	Snt9b Snt9b23	telescope： 2.7.5.3 2.7.5.9及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
GPU：GPU基本组件异常	GpuEnvironmentSystem	重要	nvidia-smi命令异常	请检查GPU驱动是否正常	GPU卡驱动不可用	GPU	telescope： 2.7.5.3 2.7.5.9及之后版本
		重要	nvidia-fabricmanager版本和GPU驱动版本不一致	请检查GPU驱动版本和nvidia-fabricmanager版本	nvidia-fabricmanager无法正常工作，影响GPU的使用		
		重要	容器插件nvidia-container-toolkit未安装	安装容器插件nvidia-container-toolkit	docker无法挂载GPU卡		
本地磁盘挂载巡检	MountDiskSystem	重要	/etc/fstab中有无效的UUID	请检查/etc/fstab配置文件中UUID的正确性，否则可能会导致机器重启失败	挂载磁盘错误，导致机器重启异常	-	telescope： 2.7.5.3 2.7.5.9及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
GP: Ant系列机器动态路由配置错误	GpuRouteConfig Error	重要	Ant系列机器网卡%s动态路由未配置或配置错误, CMD [ip route]: %s CMD [ip route show table all]: %s。	请正确配置RoCE网卡路由	NPU网络通信异常	GPU	telescope : 2.7.5.3 2.7.5.9及之后版本
NPU: Roce端口未散列配置	RoCEUDPConfig Error	重要	RoCE UDP端口未散列配置	请检查NPU RoCE UDP端口配置情况	影响NPU卡通信性能	Snt9b Snt9b23	telescope : 2.7.5.9及之后版本
系统内核自动升级预警	KernelUpgradeWarning	重要	系统内核自动升级预警, 旧版本: %s, 新版本: %s	系统内核升级可能导致配套AI软件异常, 请检查系统更新日志, 避免机器重启	可能导致配套AI配套软件不可用	Snt3P Snt3PD Snt9b Snt9b23	telescope : 2.7.5.3 2.7.5.9及之后版本
NPU环境相关命令检测	NpuToolsWarning	重要	hccn_tool不可用	请检查NPU驱动是否正常	无法配置RoCE网卡的IP、网关	Snt9b Snt9b23	telescope : 2.7.5.3 2.7.5.9及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
		重要	npu-smi不可用	请检查NPU驱动是否正常	无法正常使用NPU卡	Snt3P Snt3PD Snt9b Snt9b23	telescope： 2.7.5.3 2.7.5.9及之后版本
		重要	ascend-dmi不可用	请检查工具包ToolBox是否正常安装	无法使用ascend-dmi进行性能分析	Snt9b Snt9b23	telescope： 2.7.5.3 2.7.5.9及之后版本
NPU：L1交换机端口局部功能失效	NpuL1SwitchPortPartialFunctionFailure	重要	NPU的L1 1520交换机端口局部功能失效	转昇腾和硬件处理	业务可能受到影响终止	Snt9b23	telescope： 2.7.5.9及之后版本 lqdc mi： 2.1.0及之后版本

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
NPU：L1交换机故障	NpuL1SwitchFault	重要	NPU的L1 1520交换机发生故障	转昇腾和硬件处理	业务可能受到影响终止	Snt9b23	telescope： 2.7.5.9及之后版本 lqdcmi： 2.1.0及之后版本
NPU：RoCE IP地址不匹配	NpuRoceIPAddressMismatch	重要	RoCE网卡的实际IP地址与配置文件hccn.conf中的IP地址不一致	联系运维人员协助处理	机器参数面网络异常，多机任务无法执行	Snt9b Snt9b23	telescope： 2.7.5.9及之后版本
服务器组件信息收集	SysComponentInfo	提示	收集服务器组件信息，包含驱动固件版本、OS内核、镜像名称、ces-agent版本和SFS Turbo Client+版本	无	无	Snt3PD Snt9b Snt9b23	-

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
NPU PFC状态异常	NpuPfcStatusWarning	重要	NPU PFC优先级队列配置未将第五位队列设置为1导致的功能异常	使用命令 hccn_tool -i [device_id] -pfc -s bitmap 0,0,0,0,1,0,0,0或联系昇腾工程师处理。	Roce通信异常，业务可能受到影响终止	Snt9b23	-
NPU TLS证书状态异常	NpuTlsStatusWarning	重要	不同NPU卡的证书配置不一致	建议配置成都是开或都为关 hccn_tool -i <device_id> -tls -s enable 0 hccn_tool -i <device_id> -tls -s enable 1	HCCL通信异常，业务可能受到影响终止	Snt9b23	-
NPU HCCS健康状态异常	NpuHccsHealthWarning	重要	NPU HCCS模块健康状态异常（非OK状态）	转昇腾和硬件处理	会导致NPU间数据同步失败，引发训练任务中断或推理结果错误	Snt9b23	-

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
SDI网卡状态异常	SdiCheckWarning	重要	SDI网卡未处于UP和RUNNING状态	1. 可排查驱动问题或联系厂商恢复网卡。 2. 查看网络配置文件/etc/sysconfig/network-scripts/ifcfg-xxx（xxx表示网卡名），确保ONBOOT=yes并重启网络服务systemctl restart NetworkManager 3. 查看是否	业务可能受到影响终止	Snt9b Snt9b23	-

事件名称	事件ID	事件级别	事件说明	处理建议	事件影响	支持机型	支持CES Agent版本
				网卡物理损坏或电源故障可能导致的状态异常。			

10.3 使用 CES 实现 Lite Server 监控和事件告警

场景描述

针对上报到CES的监控指标和事件，支持配置告警，以短信、邮件等方式通知用户故障信息，并支持通过API查询故障记录。

约束限制

- 本方案基于CES的告警规则实现，由于CES允许每个账号最多创建100个告警规则，因此本方案最多可以监控100个超节点。
- 告警来源基于CES故障检测事件，因此需要开启CES主机监控委托。可以在购买超节点时开启，也可以购买后在CES控制台授权，具体参考[CES权限管理](#)。
- 告警通知使用的是消息通知服务（SMN）提供的短信、邮件等功能，会产生少量费用，具体价格请参考[产品价格说明](#)。

操作步骤

1. 登录[CES控制台](#)。
2. 创建告警规则模板。

图 10-3 创建告警规则模板

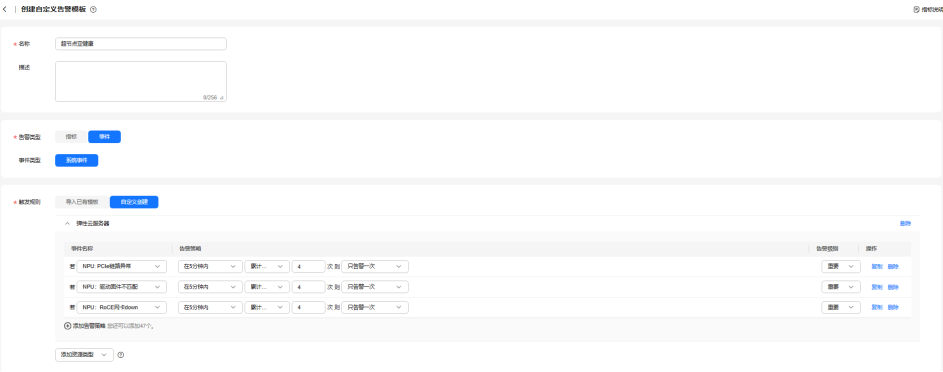


表 10-15 参数说明

属性	建议值
名称	建议以故障等级命名，例如，超节点亚健康。
告警类型	事件
触发规则	选择“自定义创建”。其它参数建议如下： - 事件名称：参考Lite Server支持的事件列表，根据事件影响选择需要关注的事件。 - 告警策略：在5分钟内累计发生4次则只告警一次。注意，不合理的配置可能导致告警过多或响应过慢。 - 告警级别：重要

3. 创建告警规则。

图 10-4 创建告警规则（类型）

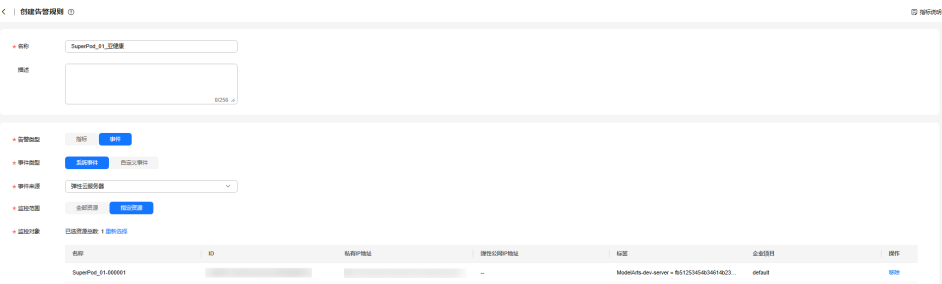


图 10-5 创建告警规则（策略）

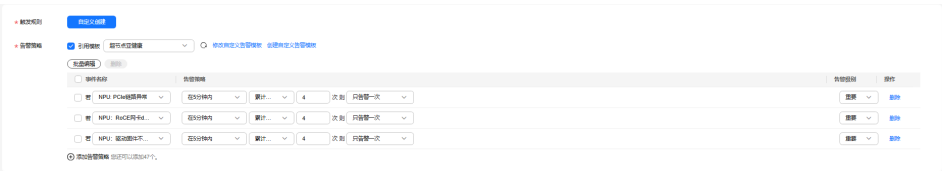


图 10-6 创建告警规则（发送通知）



图 10-7 创建告警规则（高级配置）

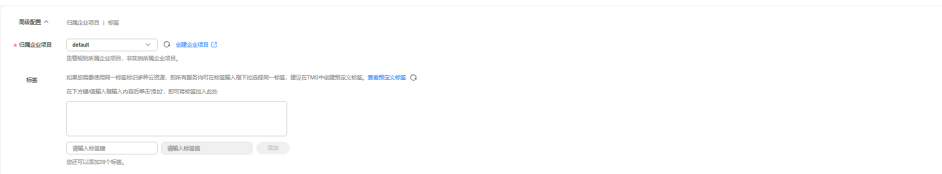
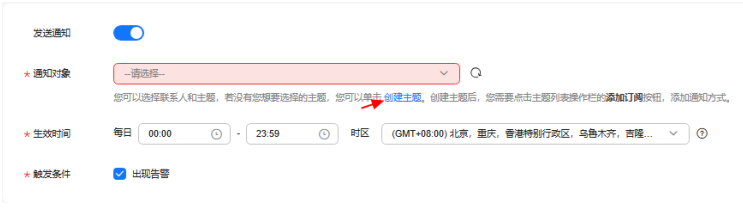


表 10-16 告警规则参数说明

属性	建议值
名称	建议采用“超节点名称_故障等级”格式，例如“SuperPod_01_亚健康”。
告警类型	事件
事件类型	系统事件
事件来源	弹性云服务器
监控范围	指定资源
监控对象	超节点内所有子节点。单击选择指定资源，搜索超节点名称，勾选所有，单击“确定”。 图 10-8 选择指定资源 
触发规则	自定义创建。
告警策略	勾选引用模板，在下拉列表框中选择第 2.创建告警规则模板 。步创建的告警模板。
发送通知	可选，如果希望以短信、邮件、HTTP、HTTPS等方式收到告警通知，打开此开关。 消息通知服务会从短信、邮件、HTTP、HTTPS的使用中收费，具体价格请参考 产品价格说明 。
通知对象	可选，当允许发送通知时，才会有此选项。建议创建新主题。 图 10-9 设置通知对象 
生效时间	可选，当允许发送通知时，才会有此选项。 建议采用默认值。

属性	建议值
触发条件	可选，当允许发送通知时，才会有此选项。 建议采用默认值。
归属企业项目	根据实际情况选择。

4. 创建主题（可选）

图 10-10 创建主题



表 10-17 创建主题参数说明

属性	建议值
主题名称	建议为显示名的英文。例如，SuperPod-Sub-Health。
显示名	推送邮件消息时，邮件主题呈现的名称，建议显示故障级别。例如，超节点亚健康。
企业项目	根据实际情况选择。

5. 添加订阅（可选）。创建主题后就可以添加订阅，以收到告警通知。
- 添加订阅后，终端会收到确认订阅的消息通知，单击订阅确认后，才能收到告警通知。

图 10-11 添加订阅

添加订阅

基本信息

主题名称

SuperPod-Sub-Health

* 协议

邮件

⚠️

如果在主题下添加短信、邮件、HTTP(S)等协议的订阅时，消息发送后会产生一定费用，具体的计费详情请参见[价格详情](#)。

* 添加方式

新建订阅

导入订阅

* 订阅终端

终端

备注

维护人员

+ 添加订阅终端

批量添加订阅终端

💡 请确认订阅状态是否为已确认。

邮件告警通知样例

邮件告警通知中，邮件主题显示了告警的级别，邮件内容显示了告警对象、告警策略以及告警时间等关键信息，并且告警规则中包含了故障对象所属的超节点名称。告警处理可以参考[事件列表](#)，根据处理建议进行处理。

图 10-12 邮件告警通知样例



查询告警记录

可以通过API查询告警记录，具体可参考CES文档[查询告警记录列表](#)。

10.4 使用 DCGM 监控 Lite Server GPU 资源

场景描述

本文主要介绍如何在 Lite Server 上配置 DCGM 监控，用于监控 Lite Server 上的 GPU 资源。

DCGM 是用于管理和监控基于 Linux 系统的 GPU 大规模集群的一体化工具，提供多种能力，包括主动健康监控、诊断、系统验证、策略、电源和时钟管理、配置管理和审计等。

约束限制

仅适用于 GPU 资源监控。

前提条件

裸金属服务器需要安装 driver、cuda、fabric-manager 软件包。

步骤一：安装 Docker

使用Docker官方脚本安装最新版Docker：

```
curl https://get.docker.com | sh
sudo systemctl --now enable docker
```

步骤二：安装容器工具集

设置仓库地址和GPG key：

```
distribution=$(. /etc/os-release;echo $ID$VERSION_ID) \
&& curl -s -L https://nvidia.github.io/nvidia-docker/gpgkey | sudo apt-key add - \
&& curl -s -L https://nvidia.github.io/nvidia-docker/$distribution/nvidia-docker.list | sudo tee /etc/apt/
sources.list.d/nvidia-docker.list
```

安装nvidia-docker2：

```
sudo apt-get update \
&& sudo apt-get install -y nvidia-docker2
```

编辑/etc/docker/daemon.json，改为如下内容：

```
{
  "default-runtime": "nvidia",
  "runtimes": {
    "nvidia": {
      "path": "nvidia-container-runtime",
      "runtimeArgs": []
    }
  }
}
```

重启Docker daemon：

```
sudo systemctl restart docker
```

步骤三：运行 DCGM-Exporter

以Docker方式运行DCGM-Exporter：

```
DCGM_EXPORTER_VERSION=3.1.7-3.1.4 && \
docker run -d --rm \
  --gpus all \
  --net host \
  --cap-add SYS_ADMIN \
  nvcr.io/nvidia/k8s/dcg-exporter:${DCGM_EXPORTER_VERSION}-ubuntu20.04 \
  -f /etc/dcg-exporter/dcp-metrics-included.csv
```

说明

这里使用的是DCGM-Exporter默认的指标采集配置文件/etc/dcg-exporter/dcp-metrics-included.csv，指标采集对象详见[dcgm-exporter](#)。如果采集对象不能满足要求，可通过定制镜像或挂载的方式使用自定义配置。

等待约1分钟，执行下面的命令获取GPU指标：

```
curl localhost:9400/metrics
```

指标获取结果如下：

```
# HELP DCGM_FI_DEV_SM_CLOCK SM clock frequency (in MHz).
# TYPE DCGM_FI_DEV_SM_CLOCK gauge
# HELP DCGM_FI_DEV_MEM_CLOCK Memory clock frequency (in MHz).
# TYPE DCGM_FI_DEV_MEM_CLOCK gauge
# HELP DCGM_FI_DEV_MEMORY_TEMP Memory temperature (in C).
# TYPE DCGM_FI_DEV_MEMORY_TEMP gauge
```

```
...
DCGM_FI_DEV_SM_CLOCK{gpu="0", UUID="GPU-6ad7ea4c-5517-05e7-0b54-7554cb4374d3"} 1
DCGM_FI_DEV_MEM_CLOCK{gpu="0", UUID="GPU-6ad7ea4c-5517-05e7-0b54-7554cb4374d3"} 4
DCGM_FI_DEV_MEMORY_TEMP{gpu="0", UUID="GPU-6ad7ea4c-5517-05e7-0b54-7554cb4374d3"}
9223372036854578794
...
```

步骤四：安装 Prometheus

在 “/usr/local/prometheus” 目录创建配置文件prometheus.yml内容如下：

```
global:
  scrape_interval: 15s # 采集间隔
scrape_configs:
  - job_name: 'prometheus'
    static_configs:
      - targets: ['xx.xx.xx.xx:9400'] # DCGM-Exporter指标获取端口，替换xx.xx.xx.xx为DCGM-Exporter所在节点的IP地址
```

运行Prometheus：

```
docker run -d \
  -p 9090:9090 \
  -v /usr/local/prometheus/prometheus.yml:/etc/prometheus/prometheus.yml \
  prom/prometheus
```

说明

这里使用的是Prometheus最基本的功能，如有更高级的诉求，可参考[prometheus的官方文档](#)。

步骤五：安装 Grafana

运行社区最新发行的Grafana版本：

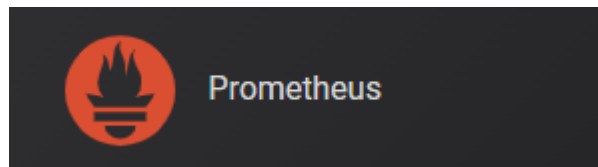
```
docker run -d -p 3000:3000 grafana/grafana-oss
```

在BMS页面打开Grafana所在节点的安全组配置，添加入方向规则，允许外部访问3000、9090端口：

在浏览器地址栏输入xx.xx.xx.xx:3000，登录Grafana，默认账号密码为：admin/admin。在配置管理页面，添加数据源，类型选择Prometheus。

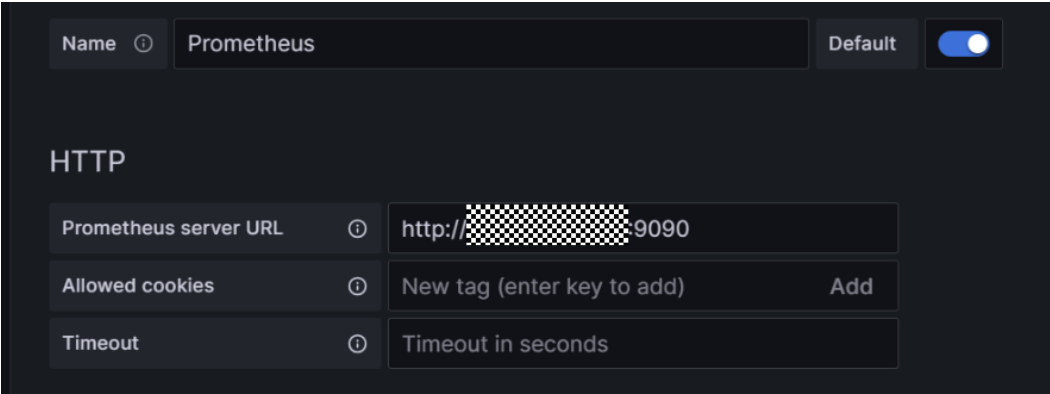
备注：xx.xx.xx.xx为Grafana的所在宿主机的IP地址

图 10-13 Prometheus



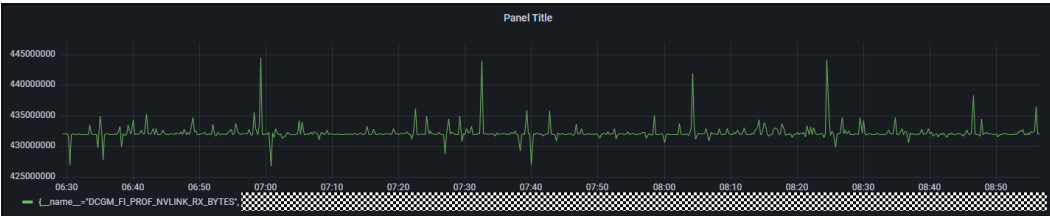
在HTTP的URL输入框中输入Prometheus的IP地址和端口号，单击Save&Test：

图 10-14 IP 地址和端口号



至此，指标监控方案安装完成。指标监控效果展示如下：

图 10-15 指标监控效果



说明

这里使用的是Grafana最基本的功能，如有更高级的诉求，可参考[Grafana的官方文档](#)。

11 Lite Server 管理 CloudPond 的 NPU 资源

场景描述

客户已经购买的CloudPond的昇腾算力资源需要通过Lite Server发放。本方案主要介绍如何通过Lite Server发放资源。

CloudPond可以理解为华为云的一个边缘可用区，边缘可用区将云基础设施和云服务部署到企业现场，适合对应用访问时延、数据本地化留存及本地系统交互等有高要求的场景，可便捷地将云端丰富应用部署到本地，CloudPond介绍可参考[产品介绍](#)。

Lite Server仅是对CloudPond上的昇腾算力资源进行了发放、启动、停止、删除等生命周期管理。

约束限制

Lite Server仅支持对CloudPond的Snt9b资源进行纳管，不支持GPU、CPU资源纳管。

仅新版的Lite Server购买界面支持发放CloudPond类型的资源。

前提条件

- 已购买CloudPond云服务，具体请参见[CloudPond快速入门](#)。
- 请联系客户经理确认Server资源方案、申请要开通资源的规格，如果无客户经理可提交工单。
- 提升资源配额，具体请参见[步骤2：提升资源配额](#)。
- 开通基础权限，具体请参见[步骤3：开通基础权限](#)。
- 配置ModelArts委托授权，具体请参见[步骤4：在ModelArts上创建委托授权](#)。

计费影响

在Lite Server上发放并纳管CloudPond资源的操作是免费的，CloudPond资源费用实际在购买CloudPond资源时收取。

发放 CloudPond 资源操作

创建Lite Server资源过程即发放CloudPond类型的资源过程。

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入资源列表。
3. 单击右上角的“购买轻量算力节点”，进入“购买轻量算力节点”页面，在该页面填写相关参数信息。

 说明

Lite Server购买界面存在新旧两个版本，仅新版的Lite Server购买界面支持发放CloudPond类型的资源。

图 11-1 购买轻量算力节点时的基础配置

基础配置

节点类型

单节点

单一物理或虚拟主机，提供基础的独立计算、存储和网络资源。

超节点

融合架构节点，提供大规模计算资源池，支持灵活调配和高密度部署。

资源类型

裸金属服务器

一款兼具弹性云服务器和物理机性能的计算类服务器，为您和您的企业提供专属的云上物理服务器。

弹性云服务器

一种可随时自助获取、可弹性伸缩的云服务器，可帮助您打造可靠、安全、灵活、高效的应用环境，确保服务持久稳定运行，提升运维效率。

计费模式 ^①

包年/包月

按需计费

区域

不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。

可用区 ^②

随机可用区

可用区1

可用区2

可用区3

边缘小站

请选择

表 11-1 基础配置参数说明

参数名称	说明
节点类型	必须选择“普通节点”。
资源类型	必须选择“弹性云服务器”。
计费模式	必须选择“按需计费”。目前策略是免费创建。
区域	选择已购买的边缘小站类型的资源所在区域。
可用区	选择“边缘小站”，在下拉框中勾选已有的CloudPond资源。如果没有可用的CloudPond资源，请参考 CloudPond快速入门 创建。

表 11-2 资源配置参数说明

参数名称	说明
CPU架构	资源类型的CPU架构，请选择ARM。由于Lite Server仅支持对CloudPond的Ascend Snt9b资源进行纳管，所以选择ARM。 请先选择CPU架构，再根据具体需求选择实例规格。具体规格有区域差异，以最终显示为准。 说明 如果界面无可选规格，请联系华为云技术支持申请开通。
系统盘	系统盘和规格有关，选择支持挂载的规格才会显示此参数。可以在创建完成后在云服务器侧实现数据盘挂载或系统盘的扩容，建议取值至少100GB。

表 11-3 操作系统参数说明

参数名称	说明
镜像	- 公共镜像 公共镜像对所有用户可见。所有用户可以根据镜像ID进行只读使用。 ModelArts服务提供了多个公共镜像，支持多种操作系统，并且内置了AI场景相关驱动和软件，为用户提供了一个完整的AI开发环境，方便用户直接进行开发和训练，而无需额外配置。 当前支持的公共镜像请参考Lite Server算力资源和镜像版本配套关系。 - 私有镜像 仅镜像创建者可以使用，其他用户无法访问。选择私有镜像创建，可以节省您重复配置服务器的时间。

表 11-4 存储配置参数说明

参数名称	说明
节点系统盘类型	系统盘和规格有关，选择支持挂载的实例规格才会显示此参数。节点系统盘用于存储服务器的操作系统，创建Lite Server时自带系统盘，且系统盘自动初始化。 此处支持选择“节点系统盘类型”，并设置“大小”。系统盘大小取值范围在100GiB和1024GiB之间。 也可以在Server创建完成后再进行系统盘的扩容。 系统盘会自动挂载到每个计算节点上。

参数名称	说明
节点数据盘类型（可选）	单击“增加数据盘”，可以在创建Lite Server时挂载云上EVS数据盘。暂不支持挂载本地磁盘。 此处支持选择“节点数据盘类型”，并设置“大小”和数据盘“数量”。 数据盘大小取值范围在100GiB和32768GiB之间。 ECS类型的机器，数据盘个数上限是59块。也可以在Server创建完成后再进行数据盘的扩容。 数据盘会自动挂载到每个计算节点上。

表 11-5 网络配置参数说明

参数名称	说明
虚拟私有云	虚拟私有云（Virtual Private Cloud，简称VPC）用以确保Server资源的安全性、隔离性和网络的灵活性。 在下拉框中选择Server对应的VPC，建议选择VPC时与其它云服务保持一致，便于网络互通。 下拉框中无可用VPC时，单击右侧的“新建虚拟私有云”创建一个VPC。创建虚拟私有云需要登录管理员账号，IP地址段请根据现网情况合理规划。
子网	选择该VPC下对应CloudPond边缘区域子网。 下拉框中无子网可选时，单击右侧的“新建子网”创建一个子网。
安全组	安全组是一个逻辑上的分组，为同一个VPC内具有相同安全保护需求并相互信任的Server提供访问策略。 下拉框中无安全组可用时，单击右侧的“新建安全组”创建一个安全组。
IPv6网络	如果当前网络配置的子网、规格、镜像都支持IPv6，则会显示该参数，打开后可启用IPv6功能。 请确保您的子网已开启IPv6功能，如果未开启请参考为虚拟私有云创建新的子网。 不同规格、镜像对IPv6支持的情况不同，如果不支持则不会显示IPv6网络参数，请以控制台实际显示为准。
RoCE网络	当使用A系列GPU或昇腾Snt9b、Snt9b23资源进行分布式训练时，为了将硬件上的RoCE网卡使用起来，需要配置RoCE网络。 该参数与所选规格有关，如果未选中规格或规格不支持RoCE网络，则不显示。 如果规格支持RoCE网络但未创建过，单击“新建RoCE网络”即可完成创建。 如果规格支持RoCE网络且已创建过RoCE网络，直接选择已有RoCE网络即可（不支持重复创建）。

表 11-6 节点管理参数说明

参数名称	说明
服务器名称	Server的机器名称。只能包含数字、大小写字母、下划线和中划线，长度不能超过64位且不能为空。
登录凭证	“密钥对”方式创建的Server节点安全性更高，建议选择“密钥对”方式。如果您习惯使用“密码”方式，请增强密码的复杂度，保证密码符合要求，防止被恶意攻击。 密钥对 指使用密钥对作为登录Server节点的鉴权方式。您可以选择使用已有的密钥对，或者单击“新建密钥对”创建新的密钥。 说明 如果选择使用已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录Server节点。 密码 指使用设置初始密码方式作为Server节点的鉴权方式，此时，您可以通过用户名密码方式登录Server节点。 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。密码复杂度需满足以下要求： - 长度为8至26个字符。 - 至少包含大写字母、小写字母、数字及特殊符号(!@\$%^_+=[]{};./?)中的3种 - 不能与用户名或倒序的用户名相同。 - 不能包含root或administrator及其逆序。
CES主机监控委托	勾选“开启”。 勾选后，将一键配置CES主机监控委托。委托CES对Server的CPU、内存、网络、磁盘、进程等指标进行监控，监控指标间隔是1分钟。详细监控指标信息请参见使用CES监控Lite Server NPU资源章节。
企业项目	该参数针对企业用户使用，只有开通了企业项目的客户，或者权限为企业主账号的客户才可见。如需使用该功能，请联系您的客户经理申请开通。 企业项目是一种云资源管理方式，企业项目管理服务提供统一的云资源按项目管理，以及项目内的资源管理、成员管理，默认项目为default。 请从下拉列表中选择所在的企业项目。更多关于企业项目的信息，请参见《企业管理用户指南》。 注意 已经完成购买的Server，不支持再修改企业项目，订单中暂不支持同步企业项目信息。

表 11-7 高级配置参数说明

参数名称	说明
CES主机监控委托	勾选“开启”。 勾选后，将一键配置CES主机监控委托。委托CES对Server的CPU、内存、网络、磁盘、进程等指标进行监控，监控指标间隔是1分钟。详细监控指标信息请参见 使用CES监控 Lite Server NPU资源 章节。

表 11-8 购买配置参数说明

参数名称	说明
购买数量	支持同时购买多台机器，输入值必须在1到10之间。

4. 单击“立即购买”，完成实例的创建。目前策略是免费发放。
5. 由于Server资源创建约20~60分钟，请耐心等待。如果资源创建失败，请参考[资源购买失败处理](#)。

图 11-2 资源创建成功

名称/ID	监控	状态	规格
31d1ebf7-ab		运行中	Lite Server NPU资源

12 使用 CTS 审计 Lite Server 服务操作

ModelArts Lite Server支持对接云审计CTS服务，通过云审计服务，您可以记录与ModelArts Lite Server相关的操作事件，便于日后的查询、审计和回溯。

约束限制

云审计服务管理控制台保存最近7天的ModelArts Lite Server操作记录。

前提条件

已开通云审计服务。详细操作请参见《[云审计服务用户指南](#)》。

Lite Server 支持审计的关键操作列表

表 12-1 Lite Server 支持审计的关键操作列表

操作名称	资源类型	事件名称
Update	Server	batchactionServer
Update	ServerHyperinstance	scaledownhyperinstanceServerHyperinstance
Update	ServerHyperinstance	scaleuphyperinstanceServerHyperinstance
Create	Server	createServer
Update	Server	stopServer
Update	Server	startServer
Update	Server	rebootServer
Update	Server	updateServer
Delete	Server	deleteServer
Read	Server	getServer
Update	Server	syncServer

操作名称	资源类型	事件名称
Read	Server	listbyuserServer
Read	Server	listServer
Update	Server	changeosServer
Update	Server	reinstallosServer
Update	Server	attachvolumeServer
Update	Server	detachvolumeServer
Read	Server	getoperationServer
Read	ServerHyperinstance	gethyperinstancescaleevaluationsServerHyperinstance
Read	ServerHyperinstance	listhyperinstanceclusterscapacityServerHyperinstance
Update	Server	bindpublicipServer
Read	Server	listpublicipServer
Create	AI ServerJob	createjobAI ServerJob
Read	AI ServerJob	getjobAI ServerJob
Read	Server	gettopologiesServer
Read	AI ServerJob	listjobsAI ServerJob
Delete	AI ServerJob	deletejobsAI ServerJob
Read	AI ServerJobTemplate	listjobtemplatesAI ServerJobTemplate
Read	AI ServerJobTemplate	getjobtemplateAI ServerJobTemplate
Read	AI ServerService	getjobserviceAI ServerService
Read	Server	listflavorsServer
Create	ServerRoceNetwork	createroceNetworkServerRoceNetwork
Create	HyperCluster	createhyperclusterHyperCluster
Read	HyperCluster	gethyperclusterHyperCluster
Read	HyperCluster	listhyperclusterHyperCluster
Delete	HyperCluster	deletehyperclusterHyperCluster
Read	ServerImage	listimagesServerImage

操作名称	资源类型	事件名称
Read	ServerImage	getImageServerImage
Read	ServerHyperinstance	listhyperinstancebyuserServerHyperinstance
Read	ServerHyperinstance	gethyperinstanceServerHyperinstance
Delete	ServerHyperinstance	deletehyperinstanceServerHyperinstance
Update	ServerHyperinstance	changehyperinstanceosServerHyperinstance
Read	ServerHyperinstance	gethyperinstanceoperationServerHyperinstance
Update	ServerHyperinstance	starthyperinstanceServerHyperinstance
Update	ServerHyperinstance	stophyperinstanceServerHyperinstance
Read	ServerHyperinstance	queryhyperinstancetagsServerHyperinstance
Create	ServerHyperinstance	createhyperinstancetagsServerHyperinstance
Delete	ServerHyperinstance	deletehyperinstancetagsServerHyperinstance
Read	ServerHyperinstance	listhyperinstanceServerHyperinstance

在 CTS 控制台查看审计日志

1. 登录[CTS云审计服务控制台](#)。
2. 在管理控制台左上角单击  图标，选择区域。
3. 在左侧导航栏中，单击“事件列表”，进入“事件列表”页面。
4. 事件列表支持通过筛选来查询对应的操作事件。当前事件列表支持四个维度的组合查询，详细信息如下：
 - 事件来源、资源类型和筛选类型。
在下拉框中选择查询条件。
其中筛选类型选择事件名称时，还需选择某个具体的事件名称。
选择资源ID时，还需输入某个具体的资源ID。
选择资源名称时，还需选择或手动输入某个具体的资源名称。
 - 操作用户：在下拉框中选择某一具体的操作用户，此操作用户指用户级别，而非租户级别。

- 事件级别：可选项为“所有事件级别”、“normal”、“warning”、“incident”，只可选择其中一项。
 - 时间范围：可选择查询最近七天内任意时间段的操作事件。
5. 在需要查看的事件左侧，单击展开该事件的详细信息。
 6. 单击需要查看的事件“操作”列的“查看事件”，可以在弹窗中查看该操作事件结构的详细信息。
- 更多关于云审计服务事件结构的信息，请参见《[云审计服务用户指南](#)》。

13 退订 Lite Server 资源

场景描述

针对不再使用的Lite Server资源，可以删除/退订以释放资源。停止计费相关介绍请见[停止计费](#)。

约束限制

- 包年/包月的Lite Server普通节点、整柜资源或超节点资源处于“创建失败”或“错误”状态下时，可以直接删除。
- 包年/包月的Lite Server普通节点、整柜资源或超节点资源在宽限期和冻结期，无法通过“退订”功能释放资源，此时系统支持直接通过“释放”功能释放资源。
- 包年/包月场景下，退订和释放操作均是以整个订单维度执行，不支持单独释放或退订整柜资源下的单个子节点，也不支持单独释放或退订超节点下面的单个子节点。
- 删除包年包月的Lite Server资源时，对于ECS和BMS类型的服务器和超节点资源，删除时会自动删除创建Server页面时设置的数据盘。创建完Server后又单独挂载的数据盘，不会删除。
- 删除按需计费的Lite Server资源时，对于ECS和BMS类型的服务器，删除时不会删除创建Server页面时设置的数据盘。创建完Server后又单独挂载的数据盘，也不会删除。对于超节点资源，删除时会删除创建Server页面时设置的数据盘，创建完Server后又单独挂载的数据盘，不会删除。
- 普通节点资源支持批量退订、批量删除、批量释放操作，超节点不支持批量操作。

退订“包年/包月”的 Lite Server 资源

您可通过以下方式进行退订：

- 方式一：在ModelArts界面退订（单个/批量实例资源退订）
- 方式二：在费用中心退订（单个/批量实例资源退订）

方式一：在ModelArts界面退订

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”列表页面。

超节点资源退订，请进入“超节点”列表页面操作。

3. 打开“查看所有”按钮，查看所有Server实例。

说明

此时如果显示需要配置委托，请联系您的账号管理员为您配置委托权限，详细操作参考[配置ModelArts委托](#)。

4. 在普通节点或超节点列表中，单击右侧操作列的“更多 > 退订”，跳转至“退订资源”页面。

在普通节点列表页，也可以勾选多个Lite Server，选择列表页顶部的“更多 > 退订”，批量操作。

5. 根据界面提示，确认需要退订的资源，并选择退订原因。

图 13-1 退订资源

您本次退订有 1 个实例属于五天无理由全额退订，截止2024年12月31日，您还可以对 53 个实例进行五天无理由全额退订。

实例信息	退订类型	支付信息(¥)	应退金额(¥)	实际退款(¥)
▼				

展开退款明细

*请选择退订原因

☐ 购买时选错参数 ☐ 退订多购买的云服务 ☒ 业务测试完毕 ☐ 云服务不好用 ☐ 不满足业务部署需求 ☐ 云服务故障无法修复 ☐ 其他

请填写具体原因，100字以内

0/100

实际退款

退还账户余额:

1. 通过第三方在线支付（如微信、支付宝、网银等）的订单，退款会返还至华为云现金账户。
2. 实际退款金额以账单为准

☒ 我已确认本次退订的资源已完成数据备份或不再使用，未放入回收站的资源退订后无法恢复。查看回收站说明

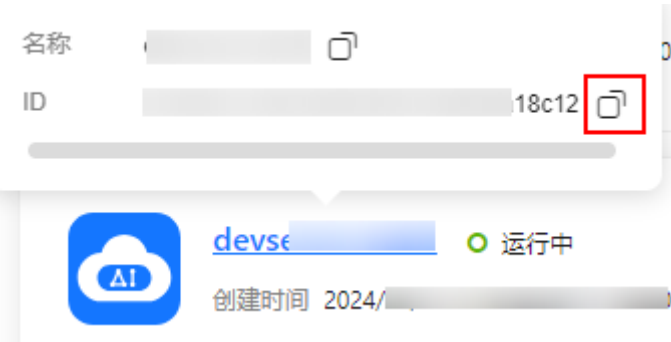
退订

6. 确认退订信息无误后，勾选“我已确认……”和“资源退订后……”提示信息。
7. 单击“退订”，再次根据界面信息确认要退订的资源。
8. 再次单击“退订”，完成包年/包月资源的退订操作。

方式二：在费用中心退订单个实例资源

1. 登录[ModelArts管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”列表页面。
- 超节点资源退订，请进入“超节点”列表页面操作。
3. 鼠标移动至节点名称上，复制需要退订的实例ID。

图 13-2 复制实例 ID



说明

Server购买订单里绑定的资源ID为Server ID，与Server产品所封装的BMS/ECS ID不同，如果要退订Server，需要在ModelArts控制台的“资源管理 > 轻量算力节点 (Lite Server)”中查询对应ID。

4. 单击顶部“费用”，进入费用中心，单击“订单管理 > 云服务退订”。

图 13-3 退订与退换货



5. 在搜索框输入实例ID信息，确认信息无误后，单击右侧“退订资源”。

图 13-4 搜索实例 ID



6. 根据界面提示，确认需要退订的资源，并选择退订原因。

7. 确认退订信息无误后，勾选“我已确认……”和“资源退订后……”提示信息。

8. 单击“退订”，再次根据界面信息确认要退订的资源。

9. 再次单击“退订”，完成包年/包月资源的退订操作。
- 在费用中心批量退订实例资源

1. 登录ModelArts管理控制台。

2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”列表页面。
超节点资源退订，请进入“超节点”列表页面操作。

3. 记录需要退订实例的ID。

说明

此时如果显示需要配置委托，请联系您的账号管理员为您配置委托权限，详细操作参考[配置ModelArts委托](#)。

4. 单击顶部“费用”，进入费用中心，单击“订单管理 > 云服务退订”。

图 13-5 退订与退换货



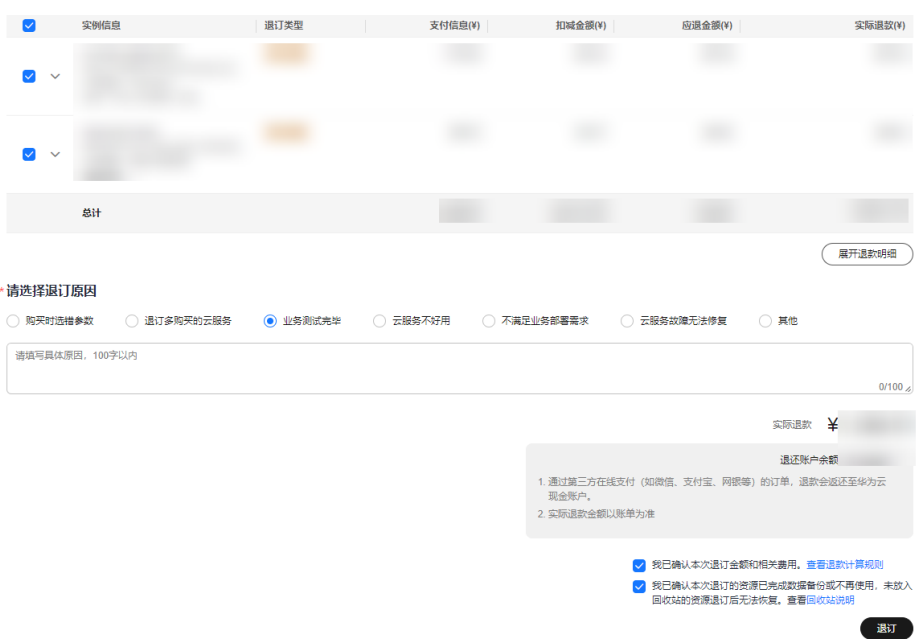
5. 勾选需要退订的多个实例，单击“批量退订”。

图 13-6 批量退订



6. 根据界面提示，确认需要退订的资源，并选择退订原因。

图 13-7 退订资源



7. 确认退订信息无误后，勾选“我已确认……”和“资源退订后……”提示信息。
8. 单击“退订”，再次根据界面信息确认要退订的资源。
9. 再次单击“退订”，完成包年/包月资源的退订操作。

释放被冻结的“包年/包月”的 Lite Server 资源

包年/包月的 Lite Server 普通节点、整柜资源或超节点资源在宽限期和冻结期，无法通过“退订”功能释放资源，此时系统支持直接通过“释放”功能释放资源。

1. 登录 [ModelArts 管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”列表页面。
超节点资源退订，请进入“超节点”列表页面操作。
3. 在列表中，单击待释放资源右侧操作列中的“更多 > 释放”，在弹出的确认对话框中，确认信息无误，然后单击“确定”，完成资源释放操作。



也可以进入 Server 资源的详情页，单击页面右上角 **...** 中的“释放”，在弹出的确认对话框中，确认信息无误，然后单击“确定”，完成资源释放操作。

在普通节点列表页，也可以勾选多个 Lite Server，选择列表页顶部的“更多 > 释放”，批量操作。

删除 Lite Server 资源

包年/包月的 Lite Server 资源仅在“创建失败”或“错误”状态下，才可以删除。

1. 登录 [ModelArts 管理控制台](#)。
2. 在左侧导航栏中，选择“资源管理 > 轻量算力节点 (Lite Server)”，进入“普通节点”列表页面。
超节点资源删除，请进入“超节点”列表页面操作。
3. 在列表中，单击右侧操作列中的“更多 > 删除”，在弹出的确认对话框中，确认信息无误，输入“DELETE”，然后单击“确定”，完成删除操作。

图 13-8 删除 Server 实例



在普通节点列表页，也可以勾选多个 Lite Server，选择列表页顶部的“更多 > 删除”，批量操作。

图 13-9 批量删除

