

华为云 Ascend C 算子开发 环境搭建手册



华为技术有限公司

版权所有 © 华为技术有限公司 2023。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



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。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编： 518129

网址： <http://e.huawei.com>



目录

1 实验介绍	2
1.1 实验介绍.....	2
1.1.1 实验平台介绍.....	2
1.2 软件介绍.....	3
1.2.1 软件版本介绍.....	3
1.2.2 Ascend C 简介.....	3
2 华为云 ModelArts 环境搭建 (算子开发)	4
2.1 创建 ModelArts 开发环境.....	4
2.2 关闭/删除环境.....	13
2.2.1 关闭环境.....	13
2.2.2 删除环境.....	13

1 实验介绍

1.1 实验介绍

1.1.1 实验平台介绍

本实验详细描述了如何在华为云 ModelArts 上搭建 Ascend C 算子开发环境。

1.1.1.1 ModelArts 简介

官网：<https://www.huaweicloud.com/product/modelarts.html>

ModelArts 是面向开发者的一站式 AI 开发平台，为机器学习与深度学习提供海量数据预处理及半自动化标注、大规模分布式 Training、自动化模型生成，及端-边-云模型按需部署能力，帮助用户快速创建和部署模型，管理全周期 AI workflow。

ModelArts 支持通过 JupyterLab 工具在线打开 Notebook，开发基于昇腾的 Ascend C 算子。具体操作流程如下图所示：



ModelArts 给用户提供了预置镜像，用户可以直接使用预置镜像创建 Notebook 实例，在实例中进行依赖安装与配置后，保存为自定义镜像，可直接用于 Ascend C 算子开发。

ModelArts 平台中预置的训练基础镜像举例：

引擎类型	版本名称
Ascend-Powered-Engine	mindspore_2.2.0-cann_7.0.1-py_3.9-euler_2.10.7-aarch64-snt9b
	mindspore_1.10.0-cann_6.0.1-py_3.7-euler_2.8.3

1.2 软件介绍

1.2.1 软件版本介绍

本实验所使用的系统和软件版本如下表，对于操作界面差异不大的版本没有列出。

训练环境：

类别	版本	说明
Python	Python3.9	面向对象编程语言
CANN	8.0.RC1.alpha002	昇腾异构计算架构
系统镜像	euleros_2.10.7	服务器操作系统
CPU架构	aarch64	CPU架构

表 1-1 训练环境系统/软件版本

1.2.2 Ascend C 简介

面向算子开发场景的编程语言 Ascend C，原生支持 C 和 C++ 标准规范，最大化匹配用户开发习惯；通过多层接口抽象、自动并行计算、孪生调试等关键技术，极大提高算子开发效率，助力 AI 开发者低成本完成算子开发和模型调优部署。

当前 Ascend C 支持的产品型号为：

- Atlas 推理系列产品 (Ascend 310P处理器)
- Atlas 训练系列产品
- Atlas A2训练系列产品
- Atlas 200/500 A2推理产品

本次我们以 Atlas A2 训练系列产品为例讲解 Ascend C 算子开发环境搭建。

2 华为云 ModelArts 环境搭建（算子开发）

在华为云 ModelArts 平台上创建硬件环境为 Ascend 910+ARM 的开发环境。

2.1 创建 ModelArts 开发环境

步骤 1 进入华为云 ModelArts 控制台

在[华为云 ModelArts 主页](#)，点击“管理控制台”进入 ModelArts 的管理页面。



图 2-1 华为云 ModelArts 主页

步骤 2 创建 Notebook 训练作业

控制台区域选择“西南-贵阳一”，在左侧菜单栏中选择“开发环境”的“Notebook”，如下图：



图 2-2 ModelArts 控制台

如有以下情况无法创建，可依次进行实名认证和依赖服务授权，依赖服务授权方法如下：

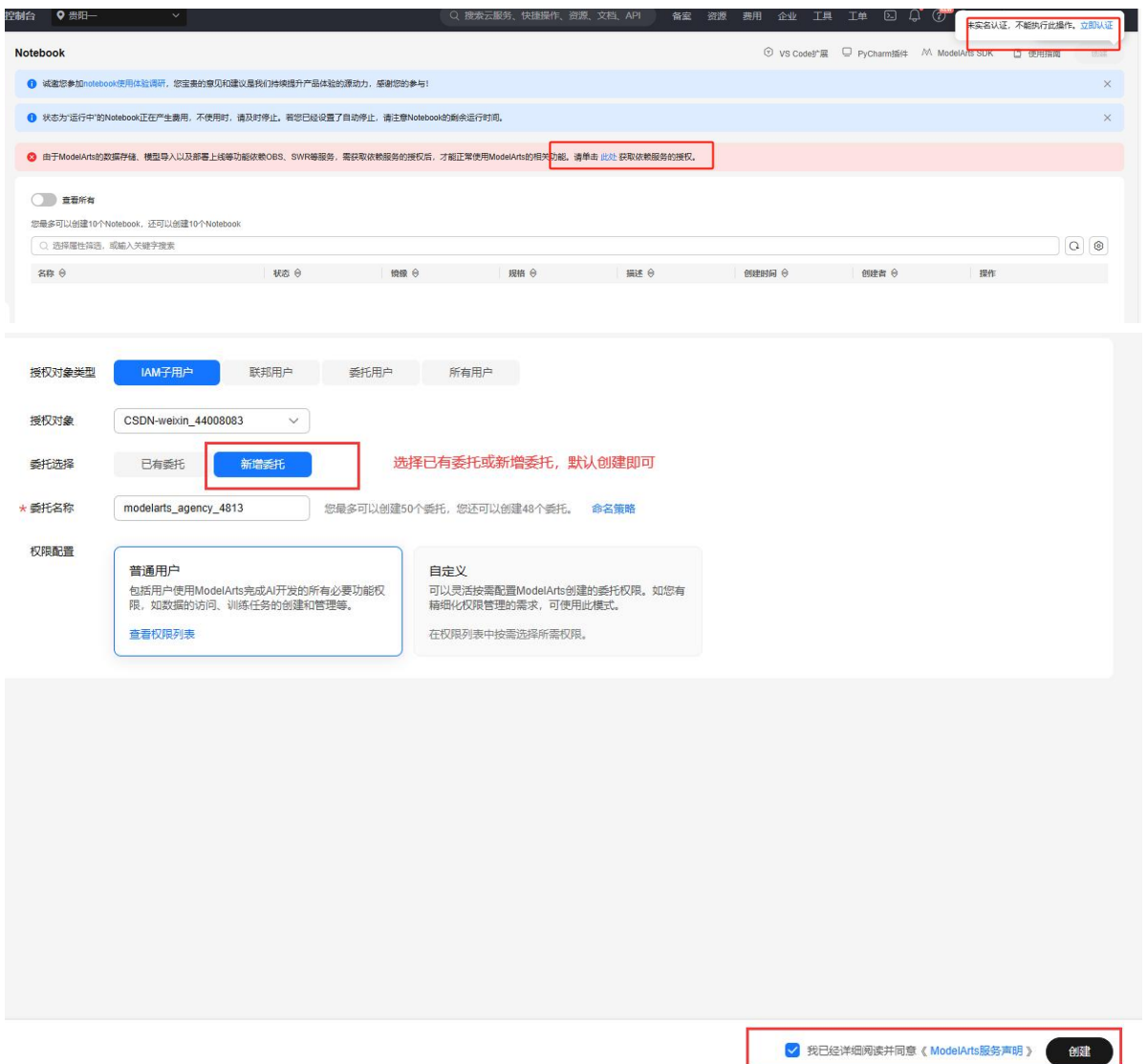


图 2-3 依赖服务授权

在网页右上角，点击“创建”按钮，创建一个新的 Notebook，其配置如下：

- 名称：自定义。
- **自动停止：打开，默认 1 小时。**
- 镜像：选择“公共镜像”，在搜索栏搜索含“cann”的镜像，选择cann7.0.1版本、Ascend+ARM的基础镜像。
- 类型：ASCEND。
- 规格：Ascend: 1*ascend-snt9b(32G)|ARM: 24核 192GB。
- 存储配置：云硬盘 EVS，默认 5GB，可根据实验需要自定义。注意：云硬盘从 Notebook 实例创建成功起，直至删除成功，每 GB 按照规定费用收费。
- 如需使用 ssh 登录方式，需勾选 ssh 链接并配置密钥对。

如图所示：

配置费用：¥21.723/小时
 优先扣减免费套餐用量，了解更多

图 2-4 Notebook 创建配置

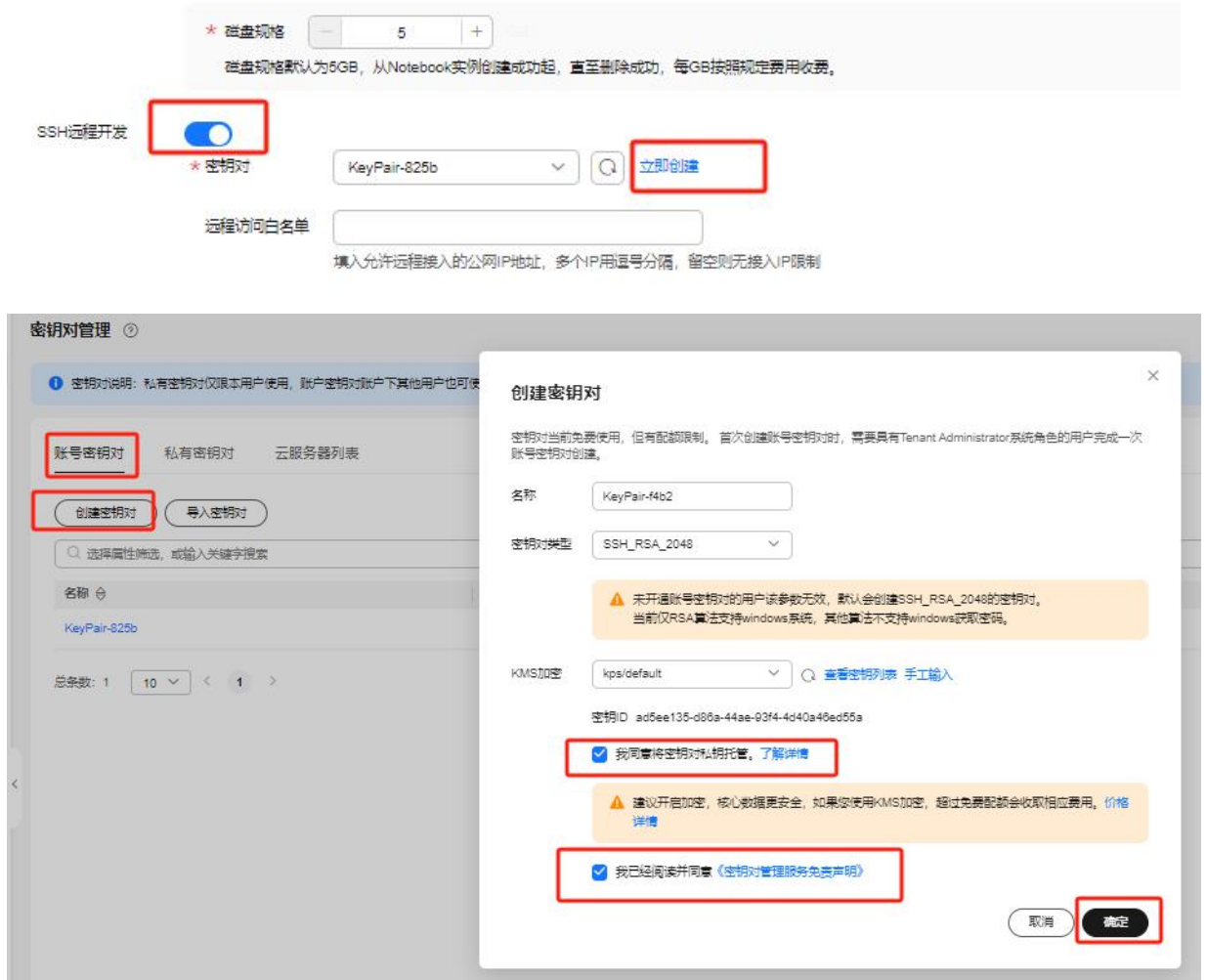


图 2-5 密钥对创建

点击创建后会下载密钥对 KeyPair-xxx.pem,注意只能创建时下载一次, 需妥善保管

提示

创建密钥对成功, 请确认私钥已成功下载到本地。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。

确定

回到环境申请页面, 点击刷新, 选择新建的密钥对作为 ssh 登录密钥。



图 2-6 密钥对选择

配置完成之后“立即创建”，规格确认无误之后“提交”。



步骤 3 启动 Notebook 进入开发环境

当上一步创建好 Notebook 状态显示为“运行中”时，在右侧操作中“打开”，即可进入在线编程页面。



图 2-7 Notebook 实例入口

可以在此页面创建或编辑 MindSpore 的项目，如图所示：

*注意：Notebook 环境内上传、创建和编辑的文件均在/home/ma-user/work 目录下。

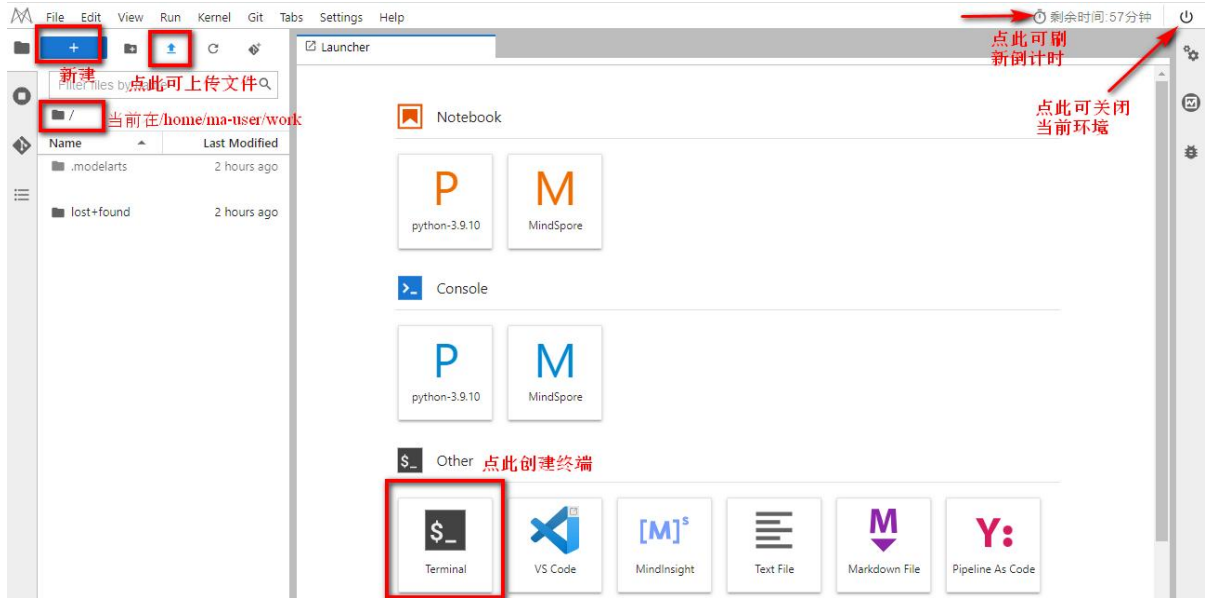


图 2-8 环境开发页面

远程 ssh 登录方式，点击进入环境页面，找到 ssh 远程开发这里。



使用 ssh 工具登录，此处我们使用 mobaxterm。

依次填入上面的域名，用户名，端口号，和密钥对，连接即可。

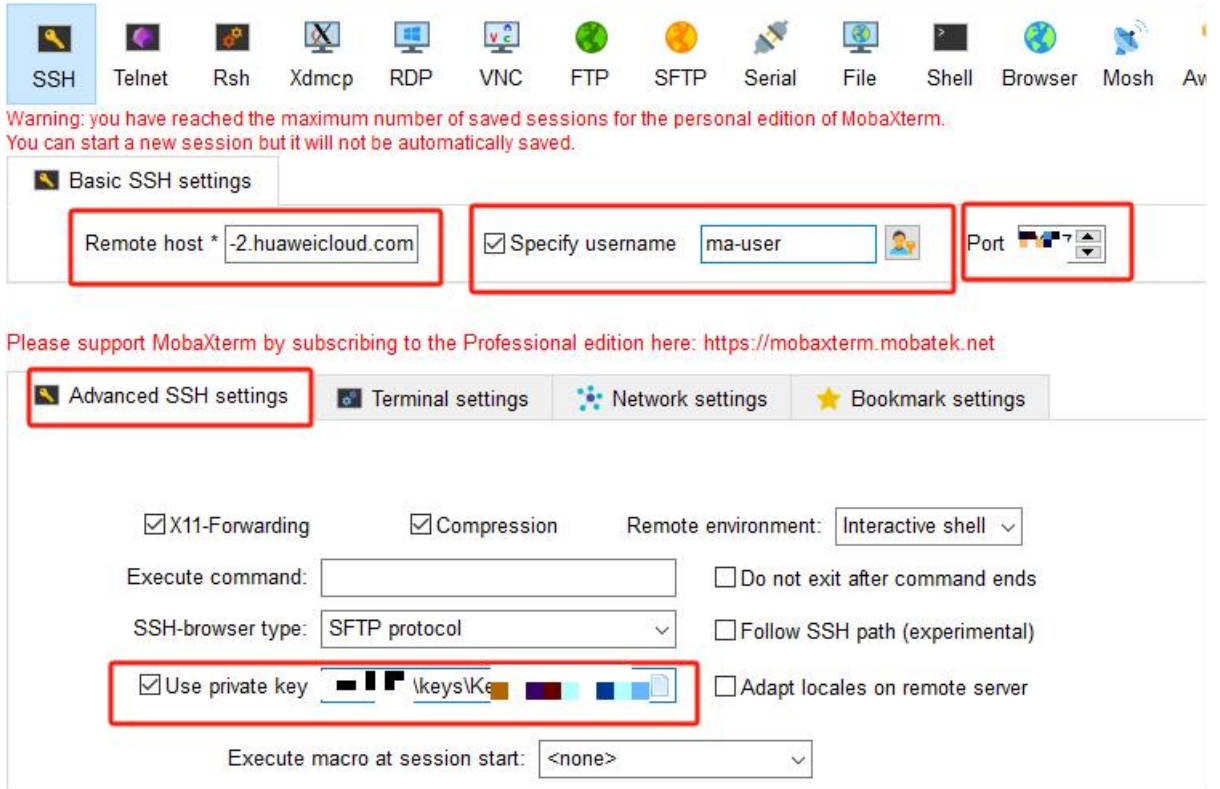


图 2-9 Ssh 登录

步骤 4 init_env.sh 环境设置脚本

浏览器输入链接地址下载 init_env.sh 脚本

https://public-download.obs.cn-east-2.myhuaweicloud.com/OpCert/init_env_8.0.RC1.alpha002.sh

点击上传文件按钮，上传 init_env.sh 脚本，如图所示：



图 2-10 上传文件页面

或者设置以下环境变量，可直接在本环境通过 wget 下载

```
export
```

```
no_proxy=127.0.0.1,localhost,172.16.*,iam.cn-southwest-2.huaweicloud.com,pip.mod  
elarts.private.com
```

```
export
```

```
NO_PROXY=127.0.0.1,localhost,172.16.*,iam.cn-southwest-2.huaweicloud.com,pip.mo  
delarts.private.com
```

```
wget
```

```
https://public-download.obs.cn-east-2.myhuaweicloud.com/OpCert/init_env_8.0.RC1.  
alpha002.sh
```

步骤 5 执行 init_env.sh 脚本，自动配置环境，cann 包安装如下

Init_env.sh 脚本已将 cann 包更新到最新社区版 cann8.0.rc1.alpha002

```
bash init_env.sh
```

```
(MindSpore) [ma-user work]$bash init_env_8.0.RC1.alpha002.sh
Start downloading toolkit package.
--2024-07-09 15:01:26-- https://ascend-repo.obs.cn-east-2.myhuaweicloud.com/Milan-ASL/Milan-ASL%20V100R001C17SP
C702/Ascend-cann-toolkit_8.0.RC1.alpha002_linux-aarch64.run
Resolving proxy-notebook.modelarts.com (proxy-notebook.modelarts.com)... 192.168.0.33
Connecting to proxy-notebook.modelarts.com (proxy-notebook.modelarts.com)|192.168.0.33|:8083... connected.
Proxy request sent, awaiting response... 200 OK
Length: 2356715176 (2.2G) [application/octet-stream]
Saving to: 'Ascend-cann-toolkit_8.0.RC1.alpha002_linux-aarch64.run'

Ascend-cann-toolkit_8.0.RC1 100%[=====>] 2.19G 11.5MB/s in 2m 59s

2024-07-09 15:04:25 (12.5 MB/s) - 'Ascend-cann-toolkit_8.0.RC1.alpha002_linux-aarch64.run' saved [2356715176/235
6715176]

CANN包部署完成
```

```
CANN包部署完成
need CMake compressed file exists.
--2024-07-09 15:09:34-- https://obs-9be7.obs.cn-east-2.myhuaweicloud.com/AscendC/ResourceDependent/cmake-3.28.3
-linux-aarch64.tar.gz
Resolving proxy-notebook.modelarts.com (proxy-notebook.modelarts.com)... 192.168.0.33
Connecting to proxy-notebook.modelarts.com (proxy-notebook.modelarts.com)|192.168.0.33|:8083... connected.
Proxy request sent, awaiting response... 200 OK
Length: 53885094 (51M) [application/gzip]
Saving to: 'cmake-3.28.3-linux-aarch64.tar.gz'

cmake-3.28.3-linux-aarch64. 100%[=====>] 51.39M 25.5MB/s in 2.0s

2024-07-09 15:09:36 (25.5 MB/s) - 'cmake-3.28.3-linux-aarch64.tar.gz' saved [53885094/53885094]
```

图 2-11 环境配置成功页面

步骤 6 等待脚本执行成功后，设置算子开发所需环境变量

```
source ~/.bashrc
```

```
(MindSpore) [ma-user work]$source ~/.bashrc
The environment has been set

ModelArts

Using user ma-user
EulerOS 2.0 (SP10), CANN-7.0.RC1 [V100R001C29],[V100R001C30],[V100R001C13],[V100R003C10],[V100R003C11]
Tips:
1) Navigate to the target conda environment. For details, see /home/ma-user/README.
2) Copy (Ctrl+C) and paste (Ctrl+V) on the jupyter terminal.
3) Store your data in /home/ma-user/work, to which a persistent volume is mounted.
(MindSpore) [ma-user work]$
```

图 2-12 环境变量设置页面

步骤 7 下载 samples 样例，执行 AddCustom 算子样例

```
cd /home/ma-user/work
```

```
git clone https://gitee.com/ascend/samples.git
```

```
cd samples/operator/AddCustomSample/KernelLaunch/AddKernelInvocation
```

```
bash run.sh -v Ascend910B4 -r npu
```

```
-- Detecting CXX compiler ABI info
-- Detecting CXX compiler ABI info - done
-- Check for working CXX compiler: /usr/bin/c++ - skipped
-- Detecting CXX compile features
-- Detecting CXX compile features - done
-- Configuring done (0.4s)
-- Generating done (0.0s)
-- Build files have been written to: /home/ma-user/work/samples/operator/AddCustomSample/KernelLaunch/AddKernelInvocation/build
[ 33%] Building CCE object cmake/np/CMakeFiles/add_npu.dir/__/__/add_custom.cpp.o
[ 66%] Building CCE object cmake/np/CMakeFiles/add_npu.dir/__/__/main.cpp.o
[100%] Linking CCE executable /home/ma-user/work/samples/operator/AddCustomSample/KernelLaunch/AddKernelInvocation/add_npu
[100%] Built target add_npu
INFO: compile op on npu succeed!
INFO: execute op on npu succeed!
test pass
```

图 2-13 样例执行成功页面

2.2 关闭/删除环境

实验完成之后，请及时关闭或删除 Notebook 训练作业，避免产生不必要的资源浪费。

2.2.1 关闭环境

可点击上图右上角的“终止按钮”关闭环境，亦可在 [ModelArts 控制台](#) 的操作栏选择“停止”操作，如下图所示：

名称	状态	镜像	规格	描述	操作
notebook-ac71 c768c7a7-b543-4be6-a9fb-2d2f...	运行中 (41分钟后)	mindspore_2.2.0-...	Ascend: 1*ascend-snt9b(32G) ARM: 24核 192GB	--	打开 启动 停止 更多

图 2-14 关闭环境

2.2.2 删除环境

关闭环境后，由于云硬盘从 Notebook 实例创建成功起直至删除成功，每 GB 按照规定费用收费，因此当不再使用该云环境时，建议删除环境。

在 [ModelArts 控制台](#) 的操作栏打开“更多”，选择“删除”操作，如下图所示：

名称	状态	镜像	规格	描述	操作
notebook-ac71 c768c7a7-b543-4be6-a9fb-2d2f...	运行中 (40分钟后)	mindspore_2.2.0-...	Ascend: 1*ascend-snt9b(32G) ARM: 24核 192GB	--	打开 启动 停止 更多
notebook-3a33 c768c7a7-b543-4be6-a9fb-2d2f...	停止	mindspore_2.2.0-...	Ascend: 1*ascend-snt9b(32G) ARM: 24核 192GB	--	打开 变更镜像 变更规格 保存镜像

图 2-15 删除环境

至此 ModelArts 的算子开发环境搭建完成。