

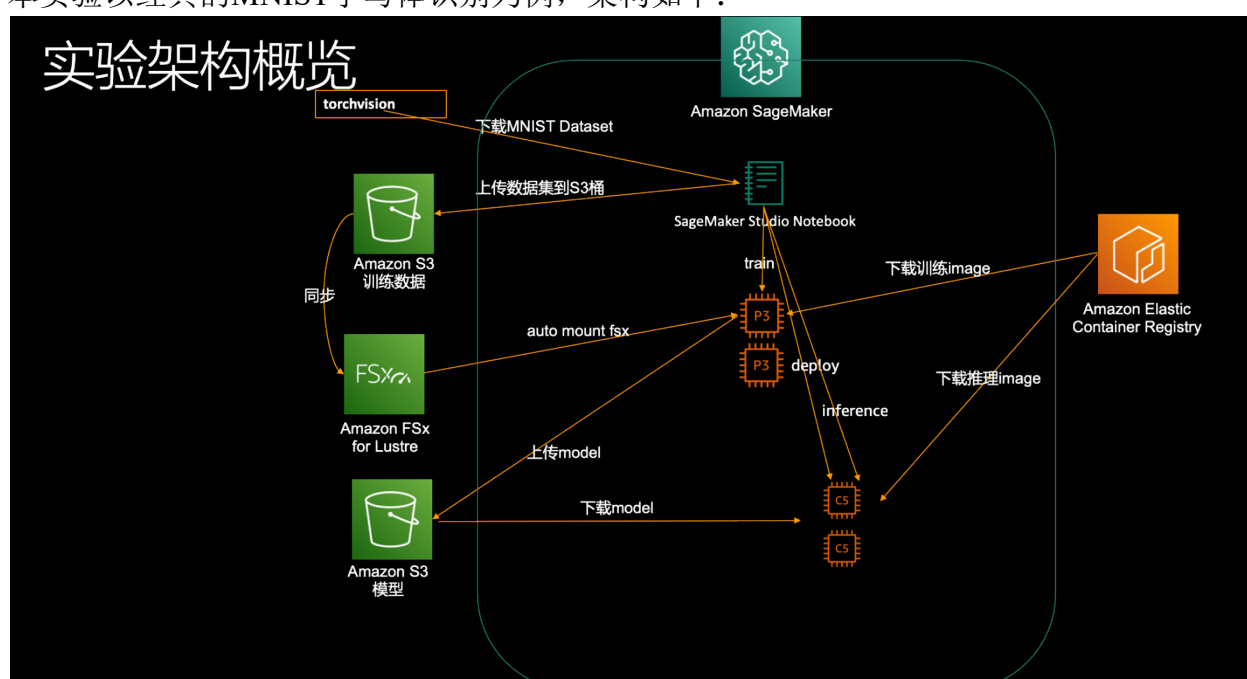
Amazon SageMaker 分布式训练实验之 PyTorch DDP

1.实验目的

本实验主要帮助用户快速掌握Amazon SageMaker 使用PyTorch DDP做分布式训练，及部署可扩展环境和推理，并帮助用户快速掌握在训练过程中使用FSx for Lustre进行存储加速。

2.架构概述

本实验以经典的MNIST手写体识别为例，架构如下：



3.详细步骤

3.0 环境准备

使用 Amazon 提供的 EventEngine 上的测试账号

- 前往 <https://dashboard.eventengine.run/>
- 输入提供给你的 12 位数的团队哈希值 (每个实验人员使用唯一的 hash 值)

Terms & Conditions

1. By using the Event Engine for the relevant event, you agree to the [AWS Event Terms and Conditions](#) and the [AWS Acceptable Use Policy](#). You acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relevant event. If you find residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. AWS reserves the right to terminate the account and delete the contents at any time.
2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by AWS, and (b) copy, import, export or otherwise create derivate works of materials provided by AWS, including but not limited to, data sets.
3. AWS is under no obligation to enable the transmission of your materials through Event Engine and may, in its discretion, edit, block, refuse to post, or remove your materials at any time.
4. Your use of the Event Engine will comply with these terms and all applicable laws, and your access to Event Engine will immediately and automatically terminate if you do not comply with any of these terms or conditions.

Enter your event hash

A 12 or 16 digit hash that was given to you for this event or for a specific team

6714-160c7f47c4-d6

✓ Accept Terms & Login

选择 Email One-Time

Password(OTP), 输入自己的邮箱, 等待邮箱收到含有9位密码的邮件, 填入密码后, sign-in。注意: 中间如果关闭了这个页面, 可以重复申请OTP进行再次登录。

Sign in with

Pick the sign-in method you prefer

Email One-Time Password (OTP)

Enter your personal or corporate email to receive a one-time password

Login with Amazon

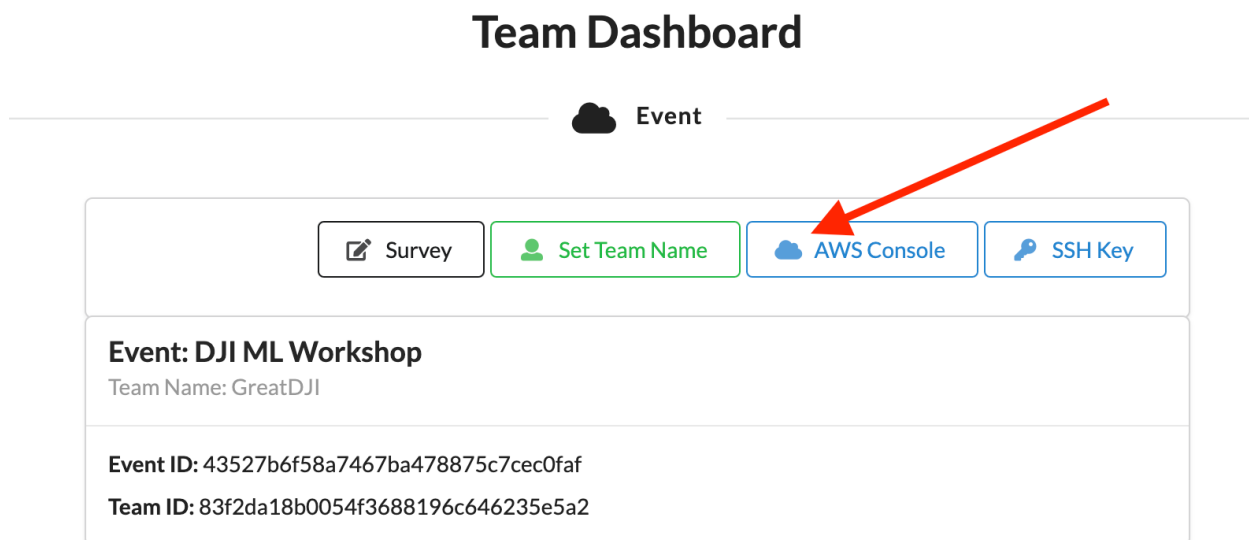
Login with your Amazon.com retail account

Amazon Employee

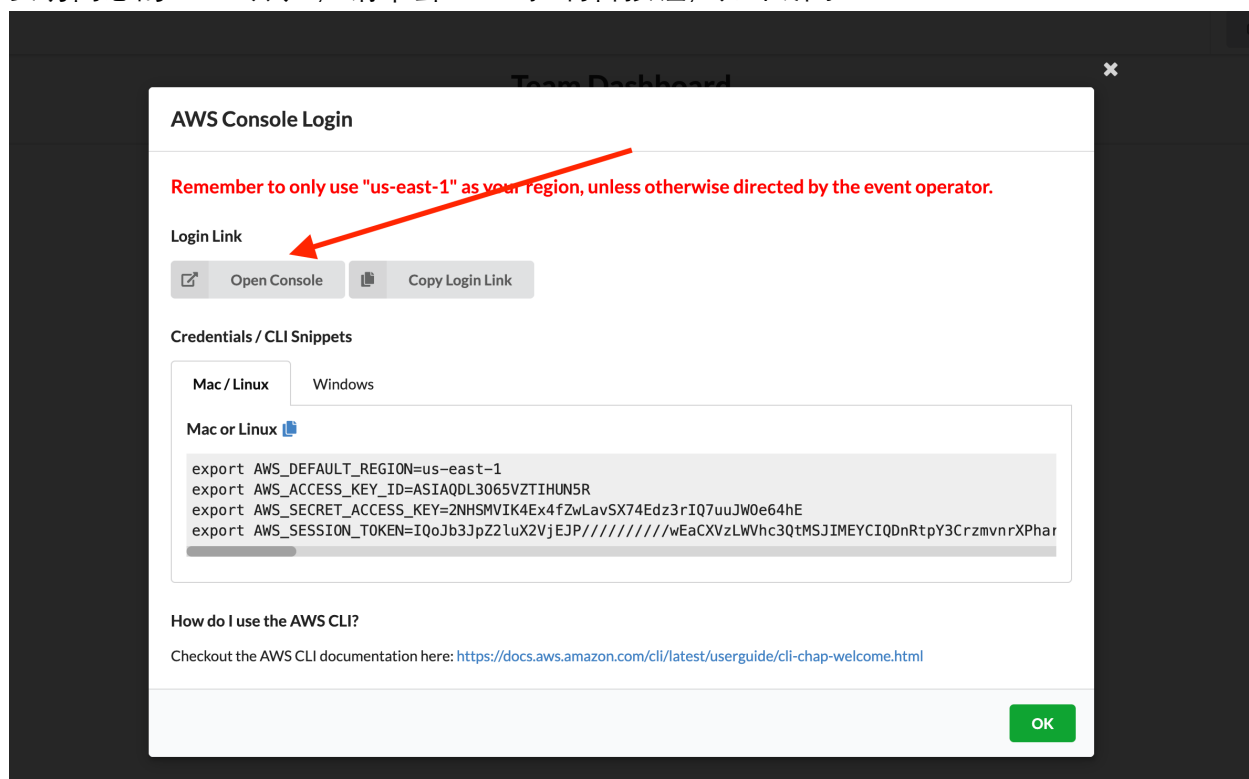
(For Amazon Employees Only) Login with your Amazon Corporate account

[Get help signing in](#)

进入下列页面，点击AWS Console



要访问您的 AWS 账户，请单击 AWS 控制台按钮，如下所示：



弹出窗口将要求您选择打开控制台（在当前浏览器中打开 AWS 控制台）或复制登录链接（将其粘贴到其他浏览器中）。

本次验证请参考 3.2 notebook instance 部分

3.1 SageMaker Studio 准备和验证流程

创建SageMaker Studio Domain

在AWS Console页面上搜索框输入sagemaker studio，然后点击sagemaker studio链接，进入下列页面：

Amazon SageMaker > Setup SageMaker Domain

Setup SageMaker Domain

Use SageMaker Domain as the central store to manage the configuration of SageMaker for your organization.

Quick setup (1 min)
Let Amazon SageMaker configure your account, and set up permissions for your SageMaker Domain.

- Public internet access, and standard encryption
- SageMaker Studio integration
- Shareable SageMaker Studio Notebooks
- SageMaker Canvas
- IAM Authentication

Perfect for single user domains and first time users looking to get started with SageMaker.

User profile

Name
default-1660705230326
The name can have up to 63 characters. Valid characters: A-Z, a-z, 0-9, and - (hyphen)

Default execution role
SageMaker Domain requires permissions for its users to access other AWS services, such as Amazon SageMaker and Amazon S3. For a broad range of capabilities, you may attach the [AmazonSageMakerFullAccess](#) policy to the execution role. If you don't have a role with this policy, we can create one for you.

TeamRole

☒ Enable SageMaker Canvas permissions
Enable SageMaker Canvas permissions to allow users to upload local files in SageMaker Canvas. [Info](#)

Cancel Submit

Standard setup (10 min)
Control all aspects of account configuration, including permissions, integrations, and encryption.

- Advanced network security, and data encryption
- SageMaker Studio, and RStudio integration
- SageMaker Studio Projects, and Jumpstart configurable
- SageMaker Canvas, and Amazon services integrations
- IAM, or IAM Identity Center (successor to AWS SSO)

Better for admins with large user groups, but you can always update your account configuration settings later if you want to do a quick setup now.

点击 Submit 开始创建 SageMaker Studio Domain，等待约 10 分钟创建完成，即状态从 Pending 转为 Ready

Amazon SageMaker > Control Panel

Control Panel

Configure and manage SageMaker domain, users, and apps.

Users

Search users

Name	Modified on	Created on
default-1660705230326	Aug 17, 2022 03:08 UTC	Aug 17, 2022 03:08 UTC

Launch app

Apps

Configure apps to manage your ML workflow using SageMaker.

Studio Studio Lab Canvas RStudio

Studio

Amazon SageMaker Studio provides a single, web-based visual interface where you can perform all ML development steps, including building notebooks, experiment management, automatic model creation, debugging, and model and data drift detection.

Configure app Learn more

Domain

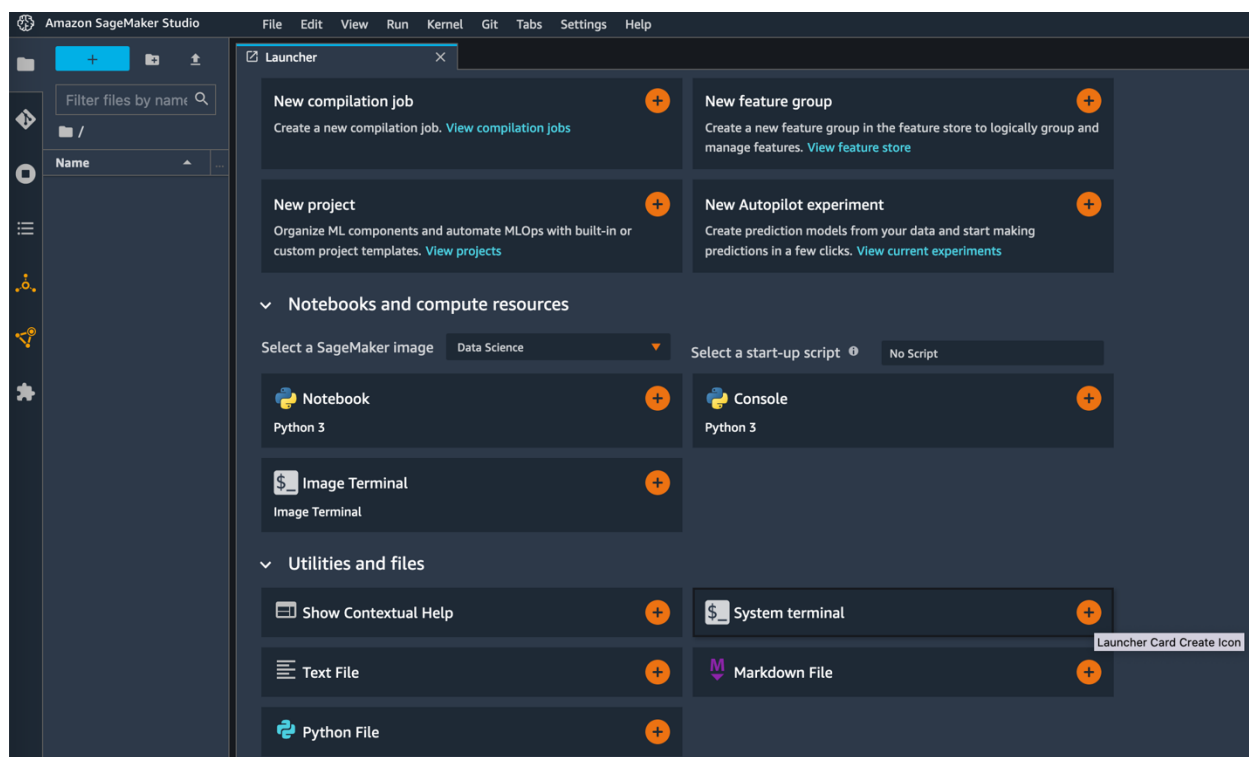
Status
Ready

Authentication method
AWS Identity and Access Management (IAM)

Domain ID
d-yftgqqrzZaiz

Execution role
arn:aws:iam::720349811467:role/TeamRole

然后点击 Launch App -> Studio, 右键点击 Open Link in New Tab, 等待打开进入 Amazon Sagemaker Studio，如下图：

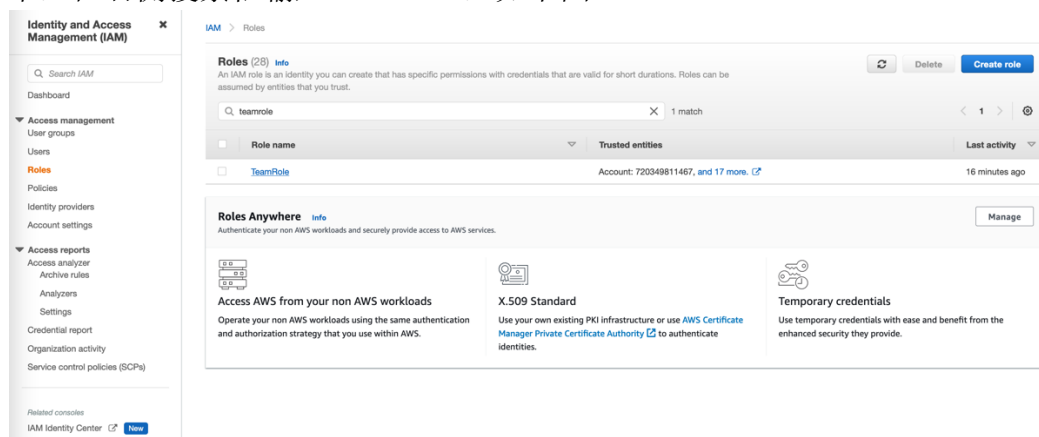


点击 System terminal，打开一个 terminal，输入下列命令：

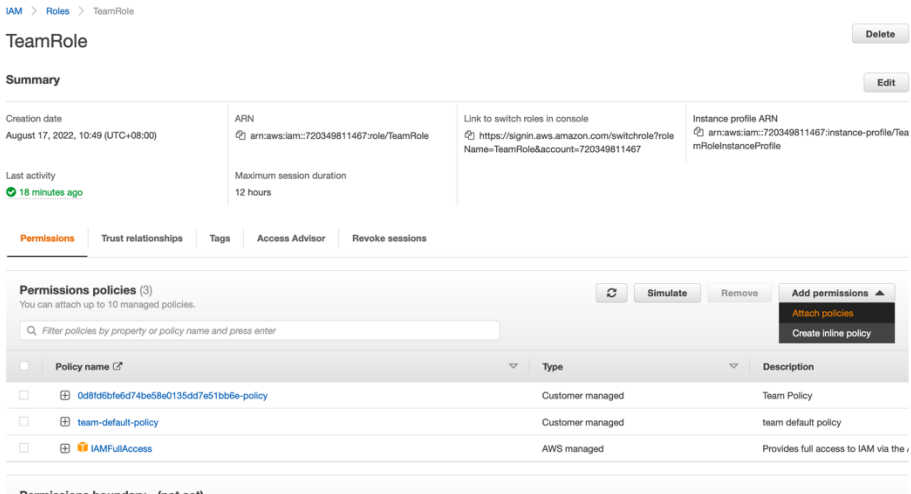
```
git clone https://github.com/aws/amazon-sagemaker-examples.git
git clone https://github.com/snowolf/pytorch-mnist-ddp-on-sagemaker.git
cd pytorch-mnist-ddp-on-sagemaker
```

配置权限（已提前配置好, 供参考）

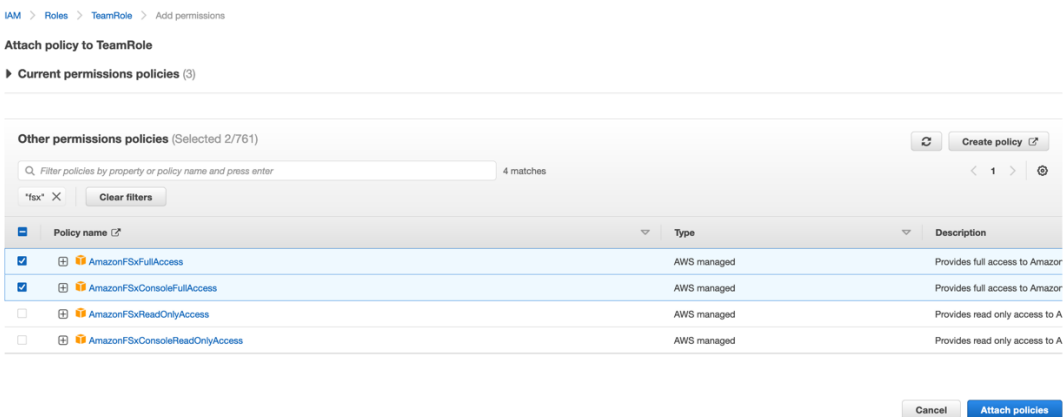
在 AWS Console 页面上搜索框输入 IAM，点击进入 IAM console，然后点击左边 role 菜单，在右侧搜索框输入 Teamrole，如下图：



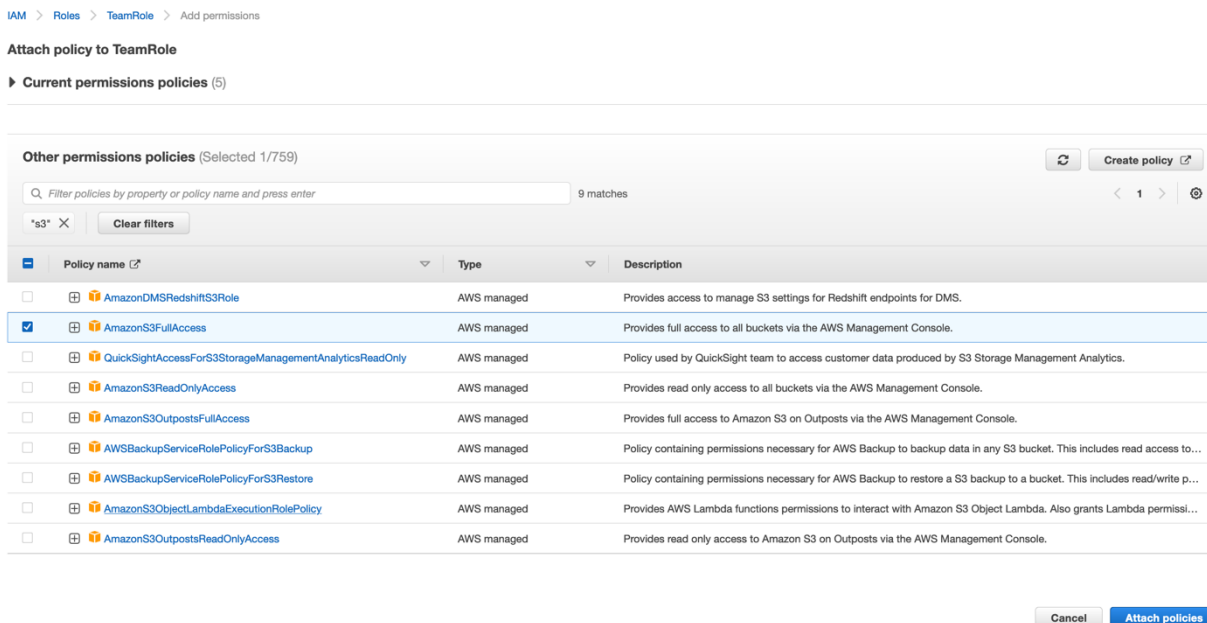
点击 TeamRole，进入下列页面，点击 Add permissions，->Attach policies



在搜索框输入 fsx，回车，选择 AmazonFSxFullAccess 和 AmazonFSxConsoleFullAccess，点击 Attach Policies，如下图



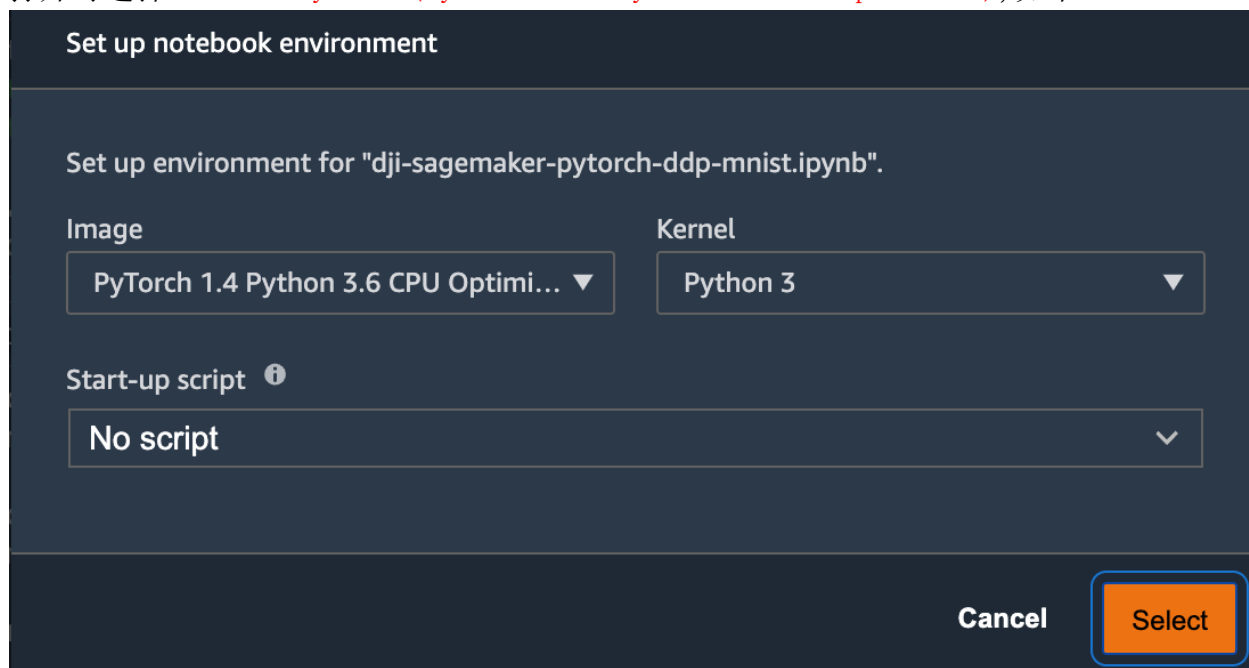
重复上述过程，在搜索框输入 s3，回车，选择 AmazonS3FullAccess，点击 Attach Policies，如下图



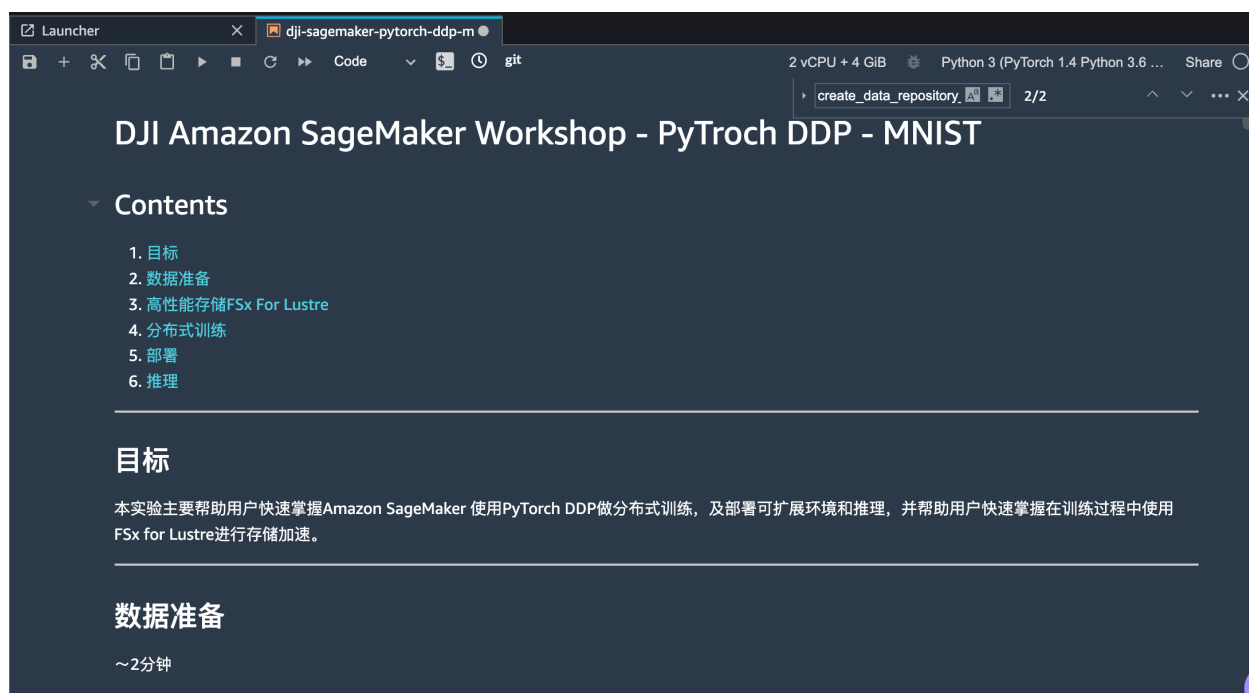
重复上述过程，在搜索框输入 ec2，回车，选择 AmazonEC2FullAccess，点击 Attach Policies.

打开 sagemaker-pytorch-native-ddp-mnist.ipynb

打开时选择 kernel: Python3 (PyTorch 1.4 Python3.6 CPU Optimized), 如下:



进入后观察上方的状态如下图所示:

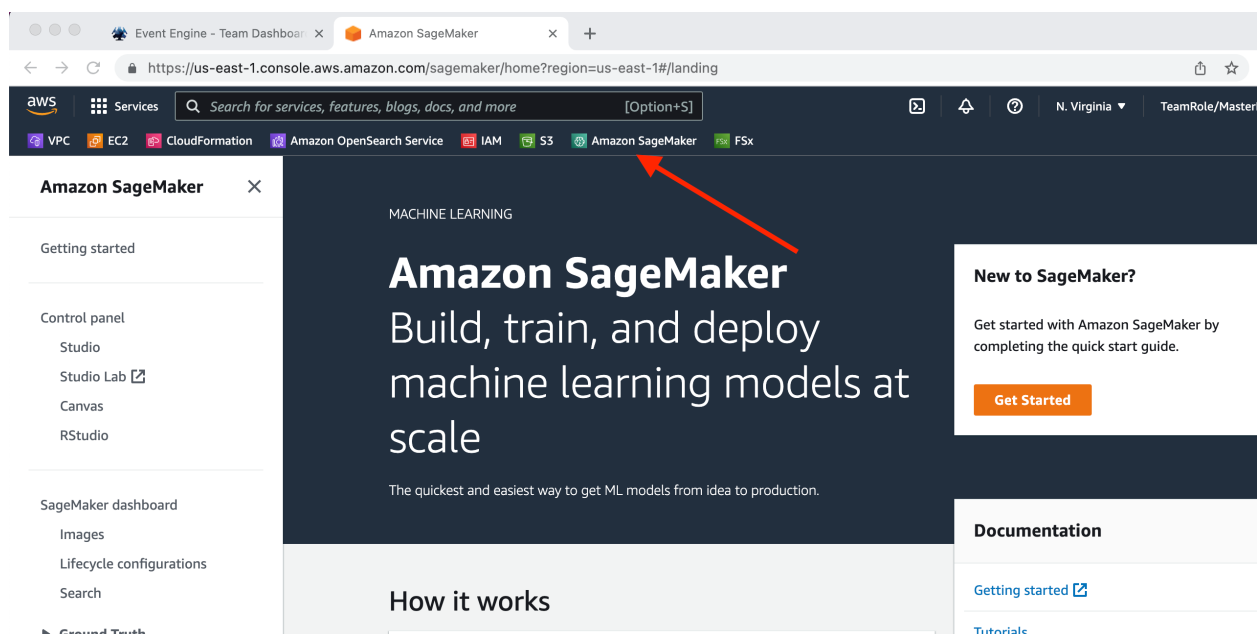


按提示依次完成下列任务：

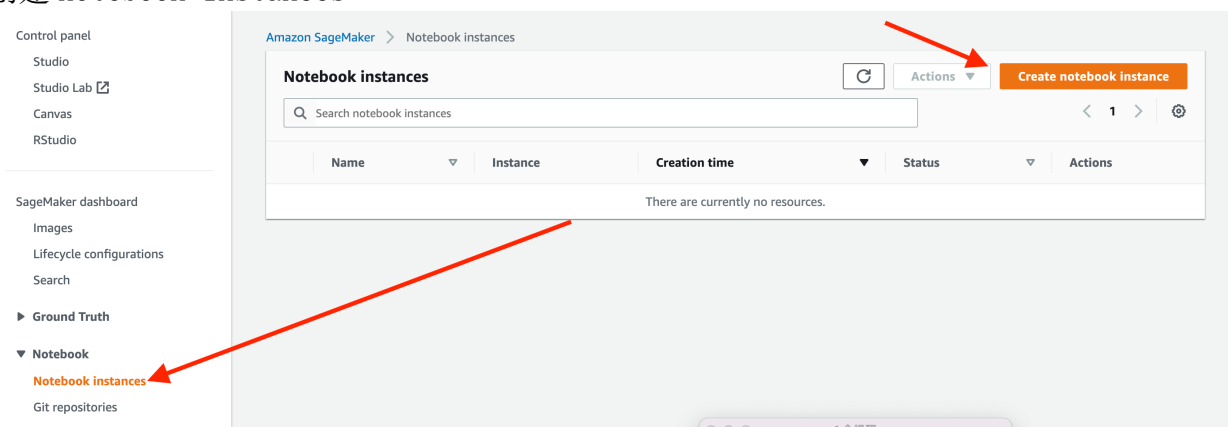
- 目标
- 数据准备
- 高性能存储 FSx For Lustre
- 分布式训练
- 部署
- 推理

3.2 Notebook Instance 准备和验证流程

进入 SageMaker 服务管理页面



创建 notebook instances



输入 notebook instance 名称: test (其他参数和选项保持不变)

include example code for common model training and hosting exercises. [Learn more](#)

Notebook instance settings

Notebook instance name

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Notebook instance type

Elastic Inference [Learn more](#)

Platform identifier [Learn more](#)

► **Additional configuration**

Permissions and encryption

IAM role
Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the

notebook instance 创建成功，等待 3 分钟左右，状态转为 InService 后，
选择 **Open Jupyter**

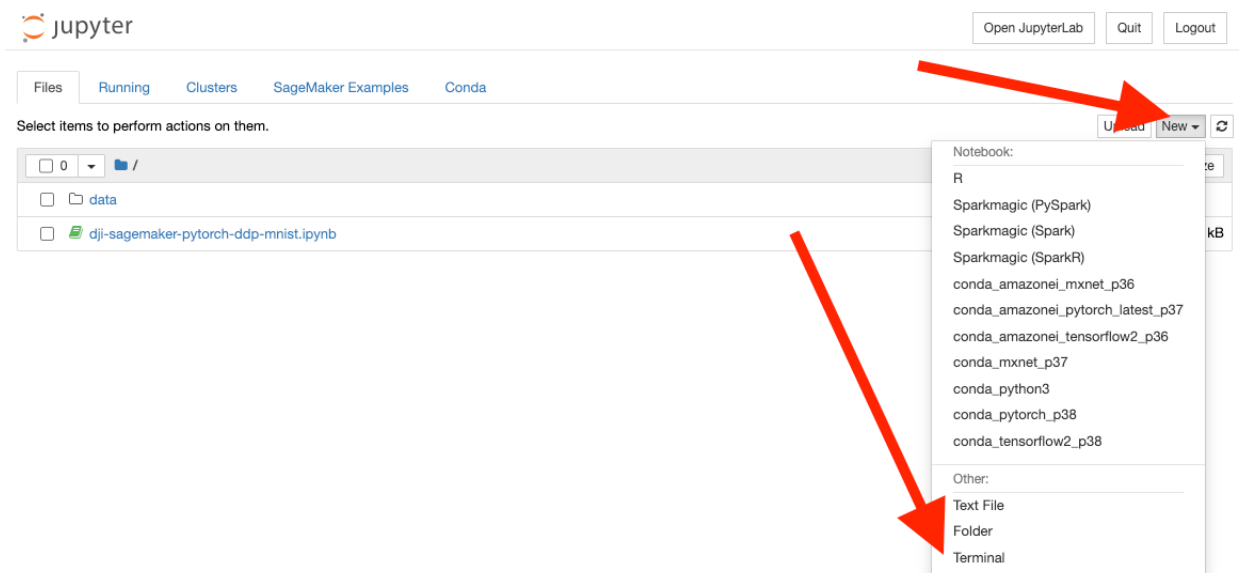
Amazon SageMaker > Notebook instances

Notebook instances

< 1 >

	Name	Instance	Creation time	Status	Actions
<input type="radio"/>	dji-workshop	ml.t3.medium	Aug 19, 2022 02:35 UTC	InService	Open Jupyter Open JupyterLab

将需要的源码下载到 instance，操作步骤如下



在新打开的 terminal 界面下执行以下命令，将需要的代码 checkout 下来备用

```
cd SageMaker
git clone https://github.com/aws/amazon-sagemaker-examples.git
git clone https://github.com/snowolf/pytorch-mnist-ddp-on-sagemaker.git
cd pytorch-mnist-ddp-on-sagemaker
```

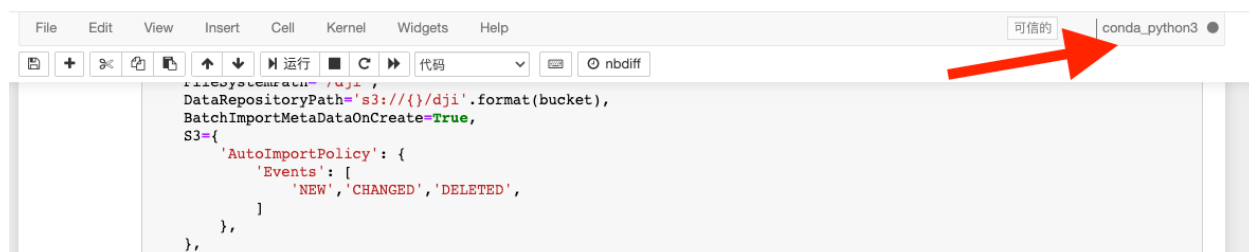
选中目录 `pytorch-mnist-ddp-on-sagemaker`, 打开后, upload notebook - `sagemaker-pytorch-native-ddp-mnist.ipynb`



下载完成后:

打开上一步上传的 `sagemaker-pytorch-native-ddp-mnist.ipynb`

开始设置(运行服务请选择 `conda_python3`)



Notebook 按提示依次完成下列任务：

1. 目标
2. 数据准备
3. 高性能存储 FSx For Lustre
4. 分布式训练
5. 部署
6. 推理

4.总结

本实验以一个简单的例子帮助用户快速掌握了Amazon SageMaker 使用PyTorch DDP做分布式训练，及部署可扩展环境和推理，并在训练过程中使用FSx for Lustre进行了存储加速。

5.参考

[1] <https://github.com/snowolf/pytorch-mnist-ddp-on-sagemaker>