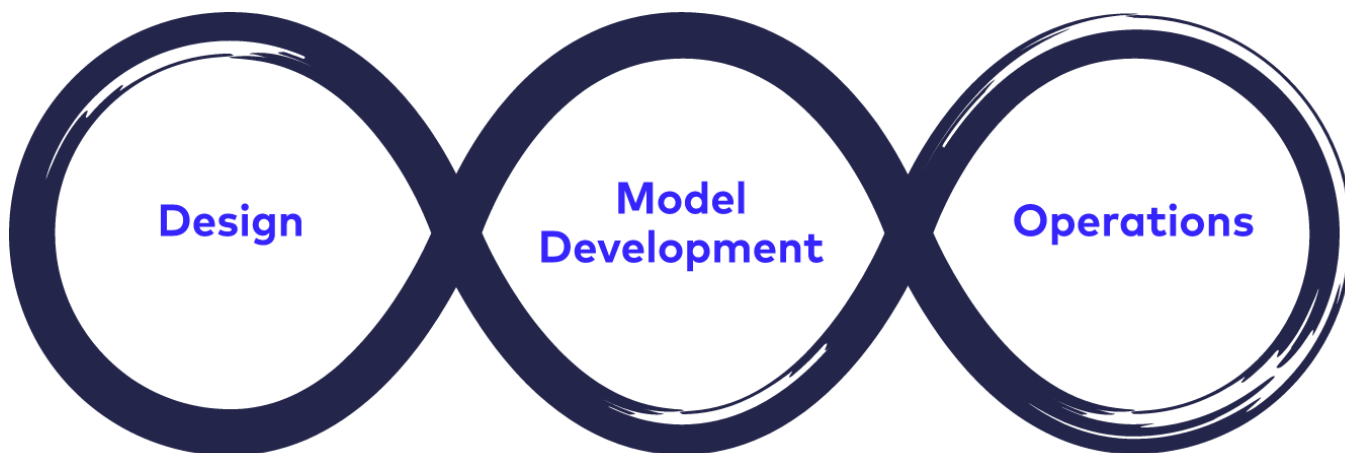


MLOps 的开发技巧

什么是 MLOps?



MLOps 基于可提高 workflows 效率的 DevOps 原理和做法。例如持续集成、持续交付和持续部署。MLOps 将这些原理应用到机器学习过程，其目标是：

更快地试验和开发模型。更快地将模型部署到生产环境。质量保证和端到端世系跟踪。

机器学习中的 MLOps

机器学习提供以下 MLOps 功能：

创建可重现的机器学习管道。使用机器学习管道可为数据准备、训练和评分过程定义可重复且可重用的步骤。

创建可重用的软件环境。使用这些环境训练和部署模型。

从任意位置注册、打包和部署模型。还可以跟踪使用模型时所需的关联元数据。

捕获端到端机器学习生命周期的监管数据。记录的世系信息可以包括模型的发布者和做出更改的原因。还包括在生产环境中部署或使用模型的时间。

针对机器学习生命周期中的事件发出通知和警报。事件示例包括试验完成、模型注册、模型部署和数据偏移检测。

监视机器学习应用程序中的操作和机器学习相关问题。比较训练与推理之间的模型输入。探索特定于模型的指标。提供有关机器学习基础结构的监视和警报信息。

使用机器学习和 Azure Pipelines 自动化端到端机器学习生命周期。通过使用管道，可以经常更新模型。还可以测试新模型。可以连同其他应用程序和服务持续推出新的机器学习模型。

GitHub + Azure ML 来完成 MLOps

GitHub 不再仅仅是代码存放工具了，它具备了很多的功能，如 GitHub Actions , GitHub Codespaces , 以及不同的项目管理等。我很喜欢 GitHub ，而且作为最重要的开发平台。



从上一章我们了解到 Azure 机器学习是非常好的 MLOps 平台，整合 GitHub 可以更好地维护我们的机器学习项目。

我们可以通过 fork 这个项目来进入 MLOps 的世界 <https://github.com/kinfey/AzureMLWithCodespace>

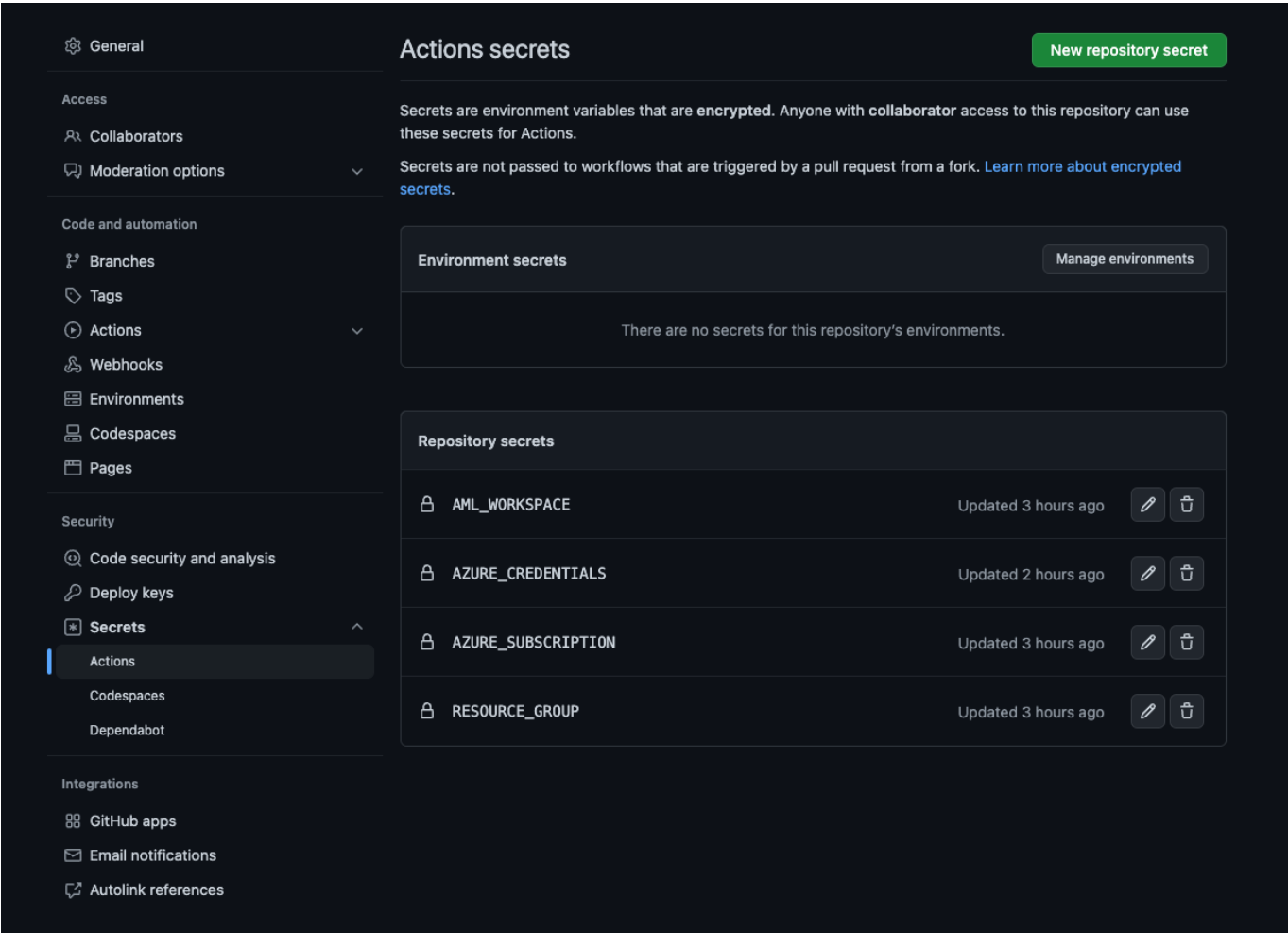
具体操作(本例子依赖于之前的机器学习环境配置，如果你没有配置完成请回到前面参考相关操作)

1. 数据上传

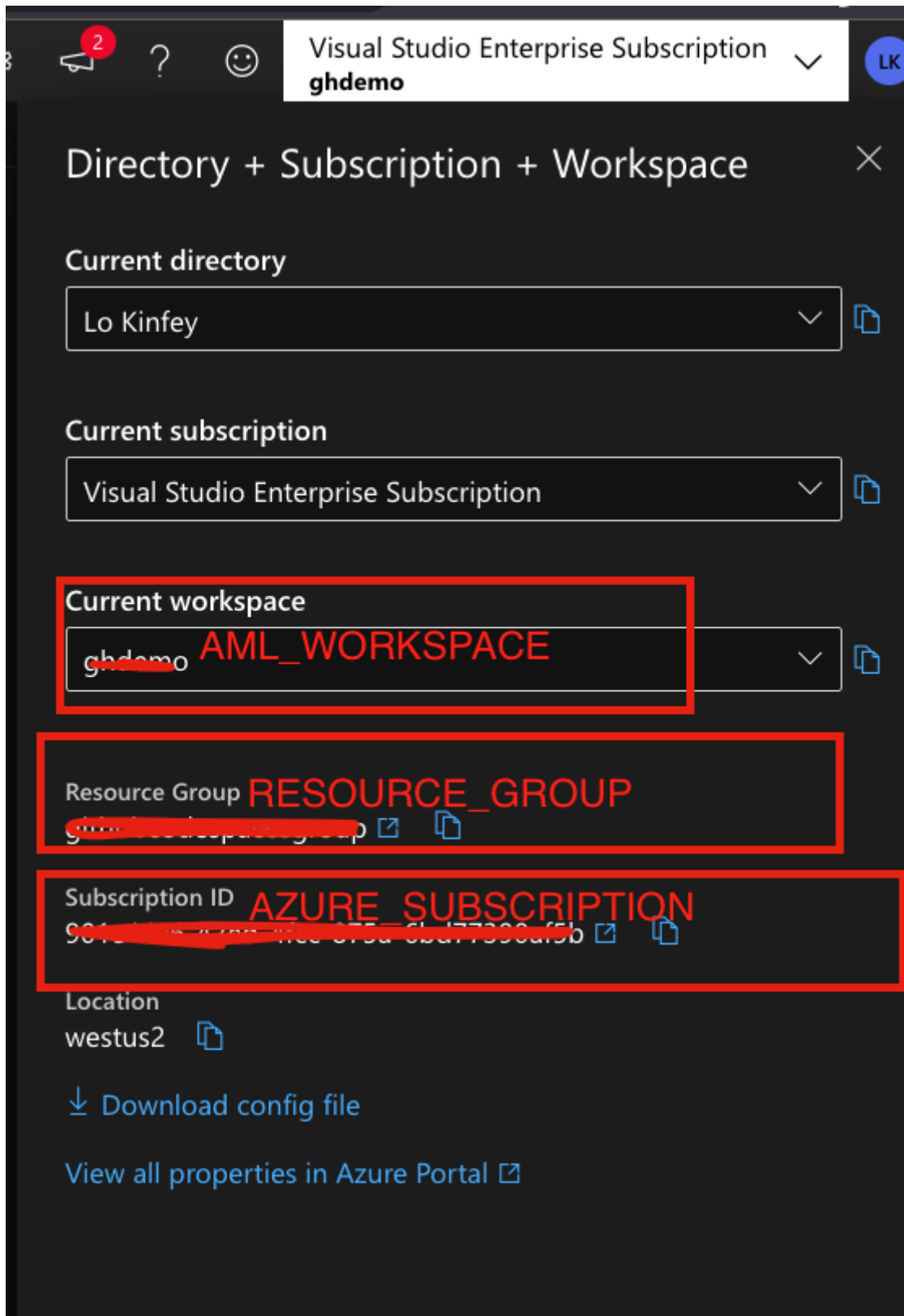
通过执行 MLOps 文件夹下的 download-data.py，完成数据下载，然后上传到你的 Azure ML 平台上，

```
az ml data create --file data.yml --resource-group my-resource-group --  
workspace-name my-workspace
```

2. 去到 GitHub 把不同密钥设置好



AML_WORKSPACE, AZURE_SUBSCRIPTION, RESOURCE_GROUP 你都可以通过



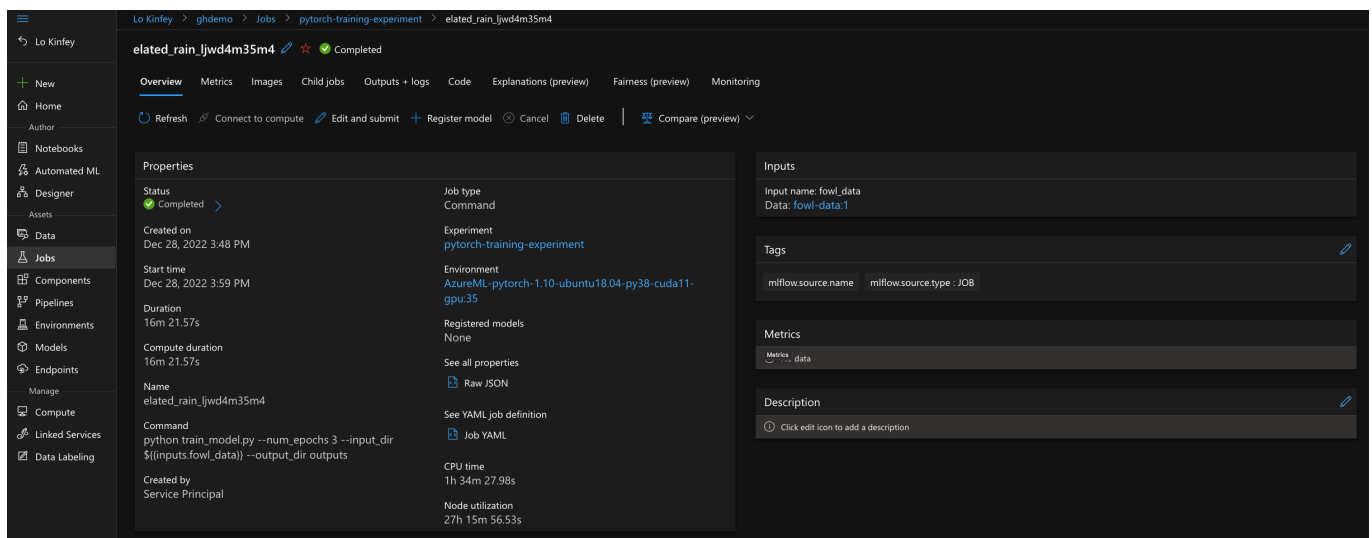
这里你可以通过

```
az ad sp create-for-rbac --name $AML_SP \  
  --role contributor \  
  --scopes /subscriptions/SUBSCRIPTIONID/resourceGroups/my-resource-group \  
  --sdk-auth
```

设置 AZURE_CREDENTIALS

完成后，即可运行

- 这里我要说明一点，每一个环境配置你都可以更改的， Azure ML 是基于 yml 文件去设置的，具体你可以参考 <https://learn.microsoft.com/en-us/azure/machine-learning/reference-yaml-overview>
- 我们现在可以通过 GitHub Actions 把数据，训练，部署结合 Azure ML 一体化完成



相关资源

- 学习 Azure ML v2 相关内容 <https://learn.microsoft.com/en-us/azure/machine-learning/concept-v2>
- 了解 GitHub Actions 的使用 <https://github.com/features/actions>