

MLMD机器学习元数据管理系统使用指南

1 商品说明

ML Metadata是由Google推出的开源元数据管理组件，专为机器学习（ML）系统设计，记录并管理机器学习模型训练、数据集等各种信息，能够实现模型的可追溯性、可重复性和合规性审查。

本商品通过鲲鹏服务器+EulerOS2.0进行安装部署

2 商品购买

您可以在云商店搜索“MLMD机器学习元数据管理系统”。

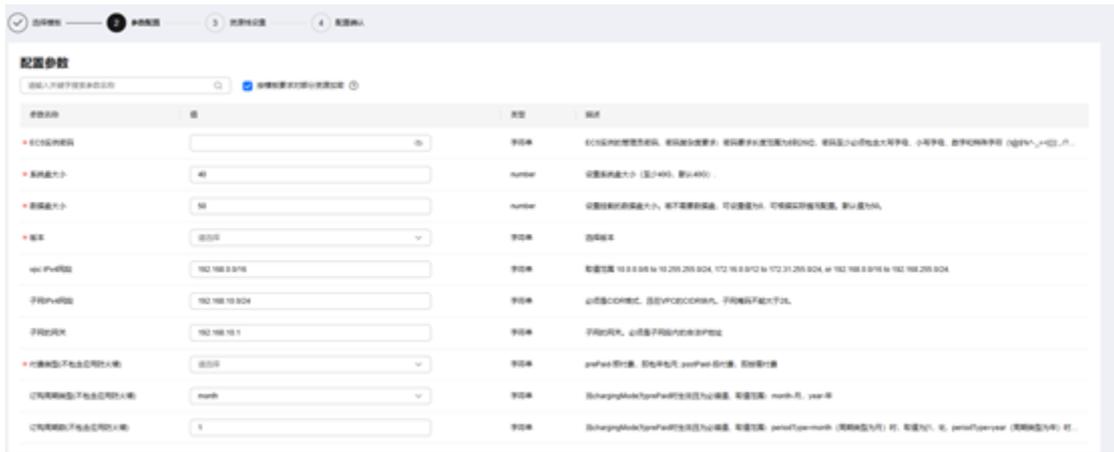
其中，地域、规格、推荐配置使用默认，购买方式根据您的需求选择按需/按月/按年，短期使用推荐按需，长期使用推荐按月/按年，确认配置后点击“立即购买”。

2.1 商品支持自定义 ECS 购买，具体见章节 3.1.1

2.2 使用 RFS 模板直接部署



必填项填写后，点击 下一步



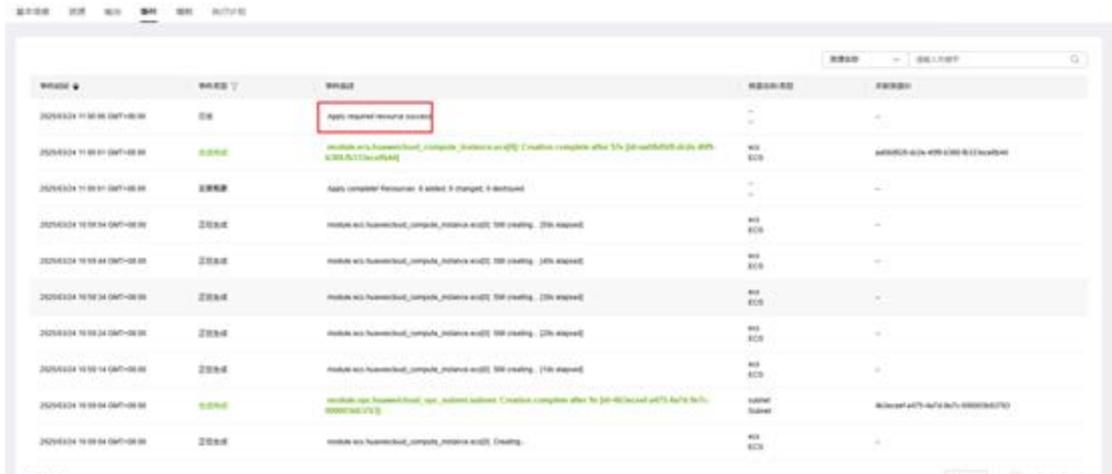
创建直接计划后，点击 确定



点击 部署



如下图“Apply required resource success.”即为资源创建完成



3 商品资源配置

商品支持ECS控制台配置，下面对资源配置的方式进行介绍。

3.1 ECS 控制台配置

3.1.1 准备工作

在使用ECS控制台配置前，需要您提前配置好安全组规则。

安全组规则的配置如下：

- 入方向规则放通端口8501，必须包含这些端口才能正常访问使用。
- 入方向规则放通CloudShell连接实例使用的端口22，以便在控制台登录调试。
- 出方向规则一键放通。

3.1.2 创建 ECS

前提工作准备好后，选择ECS控制台配置跳转到购买ECS页面，ECS资源的配置如下图所示：

基础配置

计费模式

包年/包月

按需计费

竞价计费

按需计费实例不支持备案。 [了解备案限制](#)

区域

华北-北京四

推荐区域 华北-北京四

| 华南-广州

华东-上海一

华北-乌兰察布一

西南-贵阳一

云服务器创建后无法更改区域；不同区域之间内网互不相通，请就近选择靠近您业务的区域，减少网络时延。 [如何选择区域](#)

可用区

随机分配

可用区1

可用区2

可用区3

可用区7

随机至多可用区

实例

规格类型选型

业务场景选型

CPU架构

x86计算

鲲鹏计算

实例筛选

--请选择vCPUs--

--请选择内存--

请输入规格名称模糊搜索

 隐藏售罄的规格**鲲鹏通用计算增强型**

鲲鹏内存优化型

鲲鹏超高I/O型

CSDN @p_xcn

操作系统

镜像

公共镜像

私有镜像

共享镜像

市场镜像

Huawei Cloud
EulerOS

CentOS



Ubuntu



EulerOS



Debian



KylinOS



UnionTechOS



openEuler

Huawei Cloud EulerOS 2.0 64bit for kAi2p with HDK 23.0.1 and CANN ...

存储与备份

系统盘

磁盘类型

系统盘大小(GiB)

通用型SSD

-

40

+

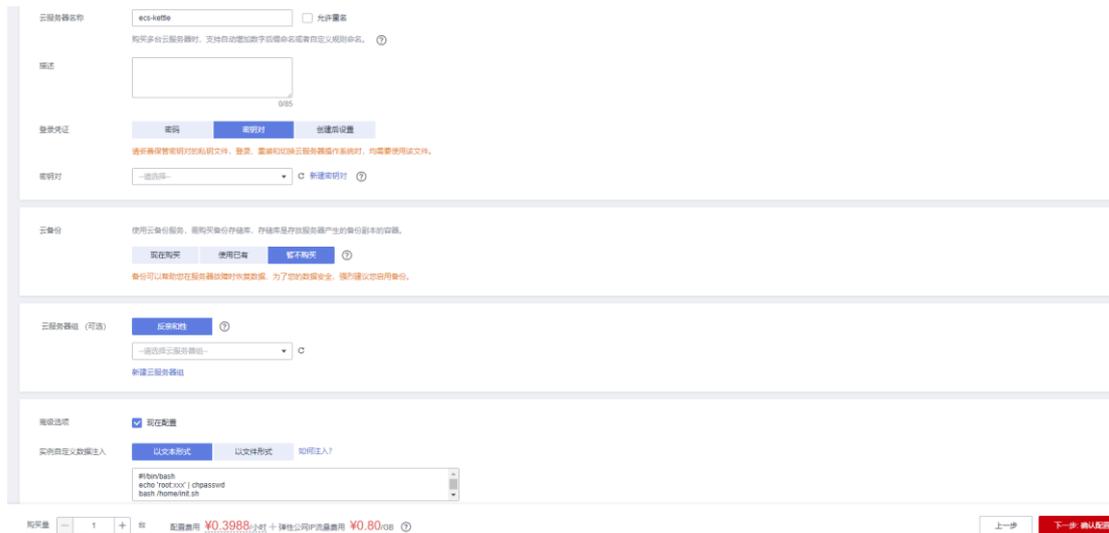
IOPS上限2,280, IOPS突发上限8,000 [高级设置](#)

增加一块数据盘

您还可以挂载 23 块磁盘 (云硬盘)

 开启备份

CSDN @p_xcn



值得注意的是：

- VPC您可以自行创建
- 安全组选择3.1.1章节中配置的安全组
- 弹性公网IP选择现在购买，推荐选择“按流量计费”，带宽大小可设置为5Mbit/s
- 高级配置需要在高级选项支持注入自定义数据，所以登录凭证不能选择“密码”，选择创建后设置
- 其余默认或按规则填写即可。

4 商品使用

4.1 运行

首先启动虚拟环境进入运行目录：

```
conda activate mlmd
```

```
cd /home/metadata
```

下载数据集到本地路径：

https://raw.githubusercontent.com/tensorflow/tfx/master/tfx/examples/penguin/data/labelled/penguins_processed.csv

运行启动代码：

```
streamlit run app.py
```

```
(mlmd) [root@ecs-kaiyuan-public-0002 metadata]# streamlit run app.py

Collecting usage statistics. To deactivate, set browser.gatherUsageStats to false.

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.0.246:8501
External URL: http://1.92.97.253:8501
```

然后使用<https://ip+8501>启动网页



然后导入之前下载的penguins_processed.csv数据集

当前数据集: penguins_processed.csv

元数据保存目录: /home/metadata/mlmd_penguin_pipeline/metadata

数据集总行数: 334, 以下显示前 10 行原始数据, 训练使用归一化后的特征

	species	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g
0	0	0.2545	0.6667	0.1525	0.2917
1	0	0.2691	0.5119	0.2373	0.3056
2	0	0.2982	0.5833	0.3898	0.1528
3	0	0.1673	0.7381	0.3559	0.2083
4	0	0.2618	0.8929	0.3051	0.2639
5	0	0.2473	0.5595	0.1525	0.2569
6	0	0.2582	0.7738	0.3898	0.5486
7	0	0.3273	0.5357	0.1695	0.1389
8	0	0.2364	0.9643	0.322	0.3056
9	0	0.0909	0.9524	0.4407	0.4722

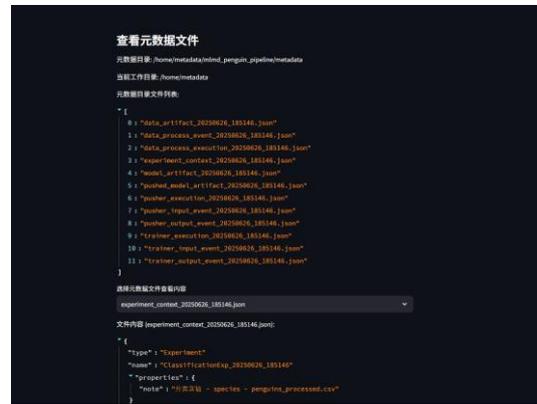
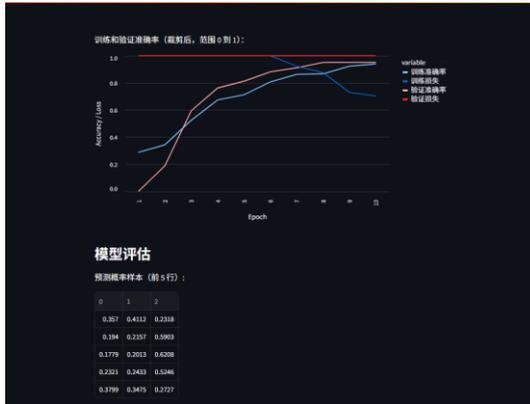
请选择标签列 (企鹅: species, 红酒: type)

species

运行分类工作流

原始数据集行数: 334

导入数据集后会显示数据集的信息以及部分内容，然后运行分类工作流即可。然后就能获得模型训练评估和查看元数据。



查看元数据文件

元数据只读: /home/metadatalocal/pipeline/metadatalocal

当前工作目录: /home/metadatalocal

元数据行列表:

```
[{"id": "data_artifact_20250626_185146.json", "type": "DataArtifact"}, {"id": "data_process_event_20250626_185146.json", "type": "DataProcessEvent"}, {"id": "data_process_execution_20250626_185146.json", "type": "DataProcessExecution"}, {"id": "experiment_content_20250626_185146.json", "type": "ExperimentContent"}, {"id": "model_artifact_20250626_185146.json", "type": "ModelArtifact"}, {"id": "pipeline_artifact_20250626_185146.json", "type": "PipelineArtifact"}, {"id": "pipeline_execution_20250626_185146.json", "type": "PipelineExecution"}, {"id": "pipeline_input_event_20250626_185146.json", "type": "PipelineInputEvent"}, {"id": "pipeline_output_event_20250626_185146.json", "type": "PipelineOutputEvent"}, {"id": "trainer_execution_20250626_185146.json", "type": "TrainerExecution"}, {"id": "trainer_input_event_20250626_185146.json", "type": "TrainerInputEvent"}, {"id": "trainer_output_event_20250626_185146.json", "type": "TrainerOutputEvent"}]
```

选择元数据文件查看内容

experiment_content_20250626_185146.json

文件内容 [experiment_content_20250626_185146.json]:

```
{ "type": "ExperimentContent", "name": "Classification_20250626_185146", "properties": { "metadata": { "species": "pmpg/iris_processed.csv" } } }
```

查看元数据选择元数据文件可查看文件内容。

4.2 参考文档

<https://github.com/tensorflow/metadatalocal>