

3.26实验课—PAI模型训练

实验概述

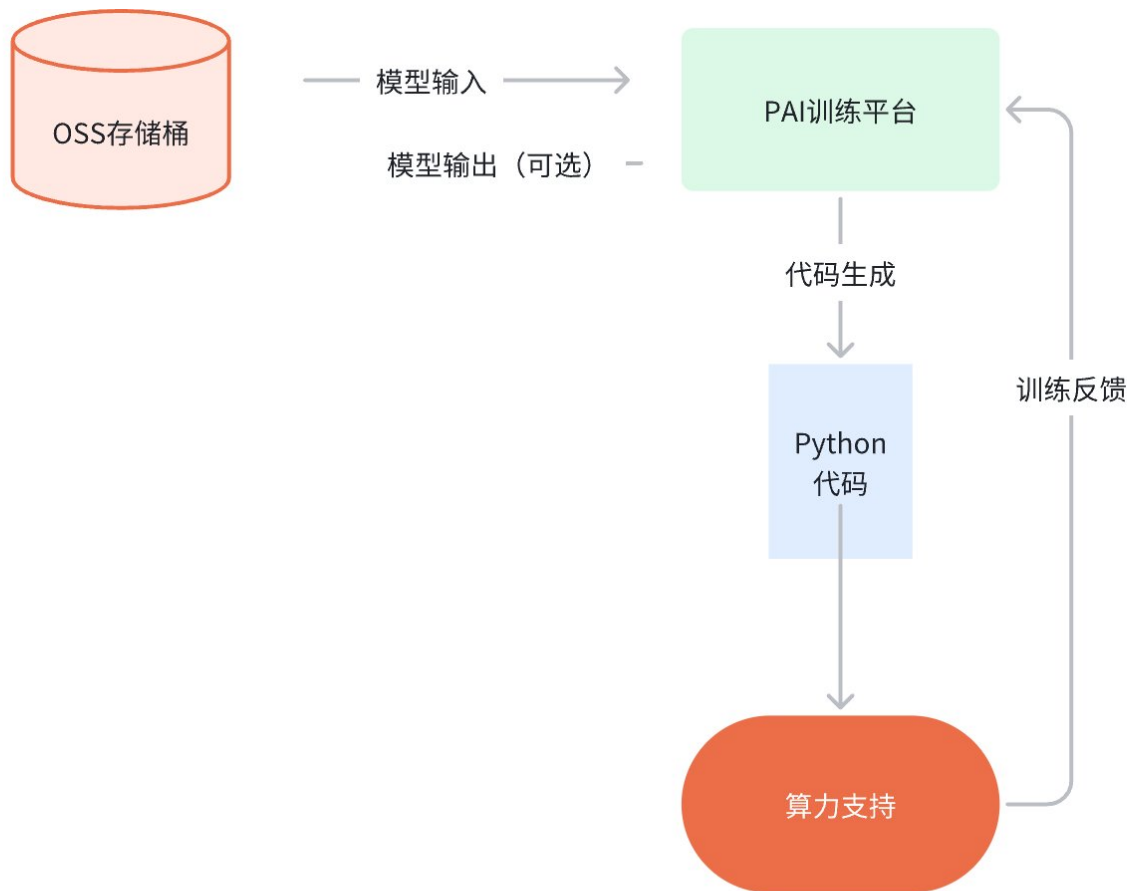
本次项目，我们采用PAI平台的design模式进行低代码的Swing推荐算法模型的训练。

PAI的design模式，通过拖拽控件，配置控件的输入输出，在底层将控件翻译成Python代码，从而实现训练过程。

根据传统的模型训练过程，我们对比说明实验中使用的云产品作用

	本地	云产品
数据源	本地磁盘文件（csv、json、txt）	OSS文件存储桶（csv、json、txt）
算力支持	GPU/CPU	MaxCompute算力平台
训练代码	Python 代码	PAI 组件工作流

所以，我们要完成的就是读取OSS存储的文件，使用PAI设计模型训练过程，利用MaxCompute提供的算力完成模型训练。



OSS存储桶配置

1. OSS领取

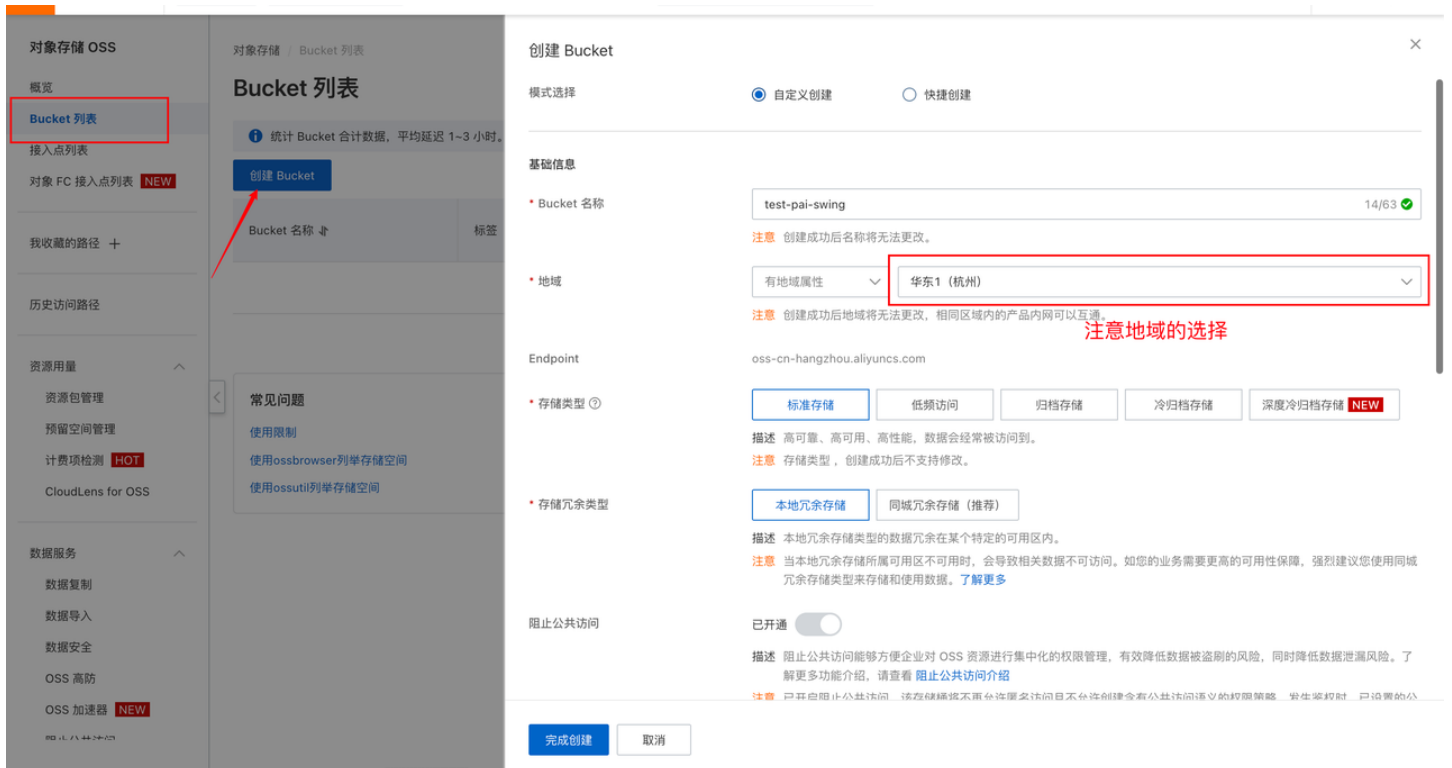
<https://free.aliyun.com/?searchKey=OSS>

采用默认配置即可



2. 创建 Bucket桶

<https://oss.console.aliyun.com/overview>



3. 上传数据

三

阿里云

工作台

Q 搜索...

费用 ICP 备案 企业 支持 工单

chendongha...
RAM 用户

对象存储 OSS

概览

Bucket 列表

接入点列表

对象 FC 接入点列表 **NEW**

我收藏的路径 +

历史访问路径

test-pai-swing

资源用量

资源包管理

预留空间管理

计费项检测 **HOT**

CloudLens for OSS

数据服务

数据复制

数据导入

数据安全

OSS 高防

对象存储 / Bucket 列表 / test-pai-swing / 文件列表

任务列表

← test-pai-swing / 华东1（杭州）

文件列表

对象（Object）是OSS存储数据的基本单元，也被称为OSS的文件。和传统的文件系统不同，Object没有文件目录层级结构的关系。

上传文件 新建目录 碎片管理 授权 请输入文件名前缀匹配 更多搜索 数据索引

文件列表

没有数据

常见问题 最佳实践

权限控制

数据安全

数据管理

内容检测

数据处理

Bucket 配置

日志管理

数据湖管理

接入点

删除 Bucket

三

阿里云

工作台

Q 搜索...

费用 ICP 备案 企业 支持 工单

chendongha...
RAM 用户

对象存储 OSS

概览

Bucket 列表

接入点列表

对象 FC 接入点列表 **NEW**

我收藏的路径 +

历史访问路径

test-pai-swing

资源用量

资源包管理

预留空间管理

计费项检测 **HOT**

CloudLens for OSS

数据服务

数据复制

数据导入

数据安全

OSS 高防

对象存储 / Bucket 列表 / test-pai-swing / 文件列表 / 上传文件

任务列表

← test-pai-swing / 华东1（杭州）

上传到 当前目录 指定目录 oss://test-pai-swing/

文件 ACL 继承 Bucket 私有 公共读 公共读写

待上传文件

扫描文件 扫描文件夹

名称 文件目录 类型 文件大小 状态 操作

没有文件或文件夹

高级选项

上传文件 取消

权限控制

数据安全

数据管理

内容检测

数据处理

Bucket 配置

日志管理

数据湖管理

接入点

删除 Bucket



MaxCompute 领取

<https://free.aliyun.com/?searchKey=max>



PAI

1. 开通PAI空间

进入PAI开通空间

https://pai.console.aliyun.com/?regionId=cn-hangzhou

授权开通，注意选择地域



默认会创建一个工作空间，点击对应的工作空间查看详情



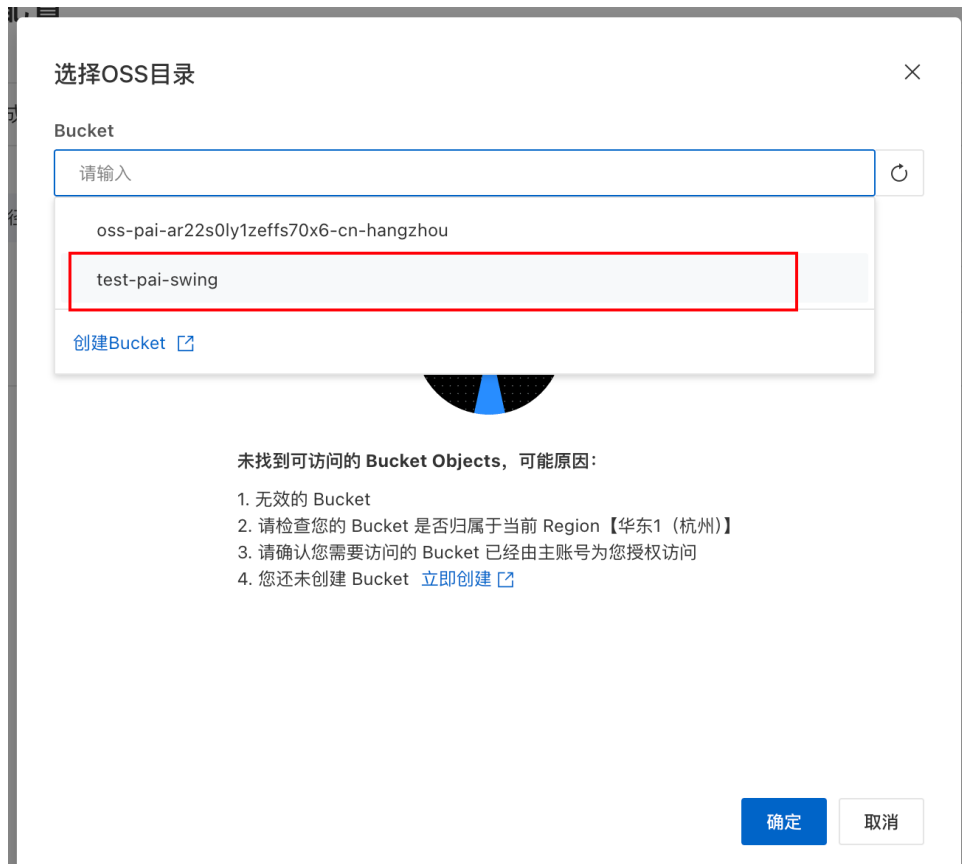
完成PAI空间的配置，点击进入工作空间配置



2. 配置存储路径



点击图标，可以换选到我们配置的存储数据的OSS Bucket



配置完成点击确定



3. 配置计算资源

三

阿里云

工作台

华东1（杭州）

搜索...

费用ICP备案企业支持工单

21

?

简体

chendongha..
RAM 用F

人工智能平台PAI

← pai_ar22s0ly1zeff... ▾

工作空间详情

快速开始

Model Gallery

Notebook Gallery

数据准备

智能标注（ITAG）

特征平台（FeatureStore）

模型开发与训练

可视化建模（Designer）

交互式建模（DSW）

分布式训练（DLC）

自动机器学习（AutoML）

模型应用

大模型应用开发（LangStudio）

模型部署

模型在线服务（EAS）

A/B实验（ABTest）

AI资产管理

← 工作空间配置

计算资源配置

成员及角色配置

调度配置

事件通知配置

存储路径配置

SLS 转发配置 ⓘ

通用配置

灵骏智算资源(0)

通用计算资源(0)

ACS计算资源(0)

MaxCompute资源(0)

Flink全托管资源(1)

MaxCompute 项目名称	Quota 名称	付费类型	状态	创建时间	操作
没有数据					

MaxCompute资源 - 待关联资源 (1) ^

还有如下资源可关联至当前工作空间，请根据需要进行关联使用

os_PayAsYouGoQuota

按量付费

MaxCompute ⓘ

Designer

使用os_PayAsYouGoQuota创建MaxCompute项目 ⓘ

pai_ar22s0ly1zeffs70x6

确认关联

相关模块介绍

如需在工作空间中使用特定模块，请注意勾选支持该模块的资源组。

Designer

三

阿里云

工作台

华东1（杭州）

搜索...

费用ICP备案企业支持工单

26

?

简体

chendongha..
RAM 用F

人工智能平台PAI

← pai_ar22s0ly1zeff... ▾

工作空间详情

快速开始

Model Gallery

Notebook Gallery

数据准备

智能标注（ITAG）

特征平台（FeatureStore）

模型开发与训练

可视化建模（Designer）

交互式建模（DSW）

分布式训练（DLC）

自动机器学习（AutoML）

模型应用

大模型应用开发（LangStudio）

模型部署

模型在线服务（EAS）

A/B实验（ABTest）

AI资产管理

← 工作空间配置

计算资源配置

成员及角色配置

调度配置

事件通知配置

存储路径配置

SLS 转发配置 ⓘ

通用配置

灵骏智算资源(0)

通用计算资源(0)

ACS计算资源(0)

MaxCompute资源(1)

Flink全托管资源(1)

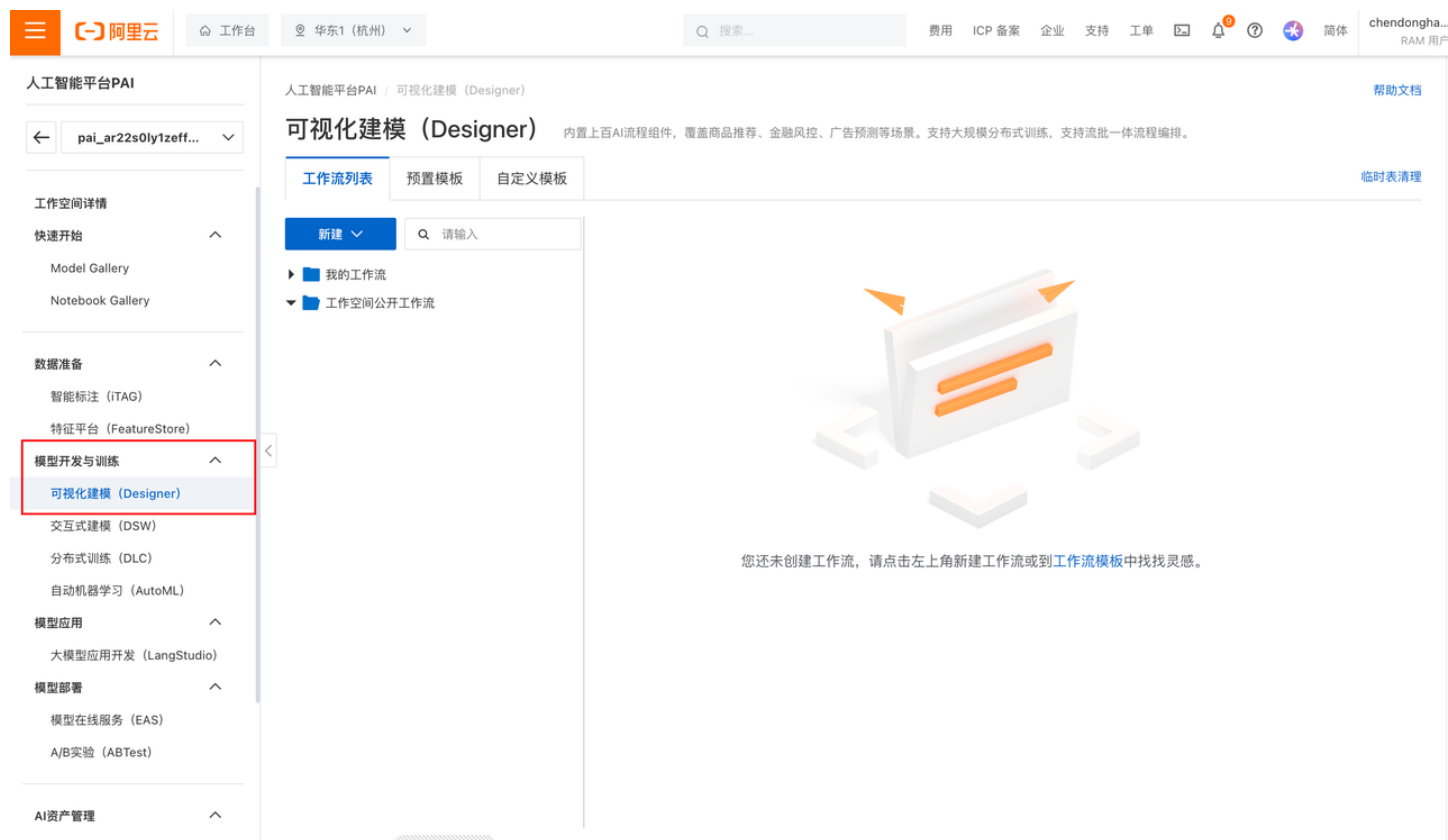
MaxCompute 项目名称	Quota 名称	付费类型	状态	创建时间	操作
pai_ar22s0ly1zeffs70x6...	os_PayAsYouGoQuota	CPU 按量付费 ⓘ	运行中	2025-03-25 11:06:43	设为默认

MaxCompute资源 - 待关联资源 (0) ^

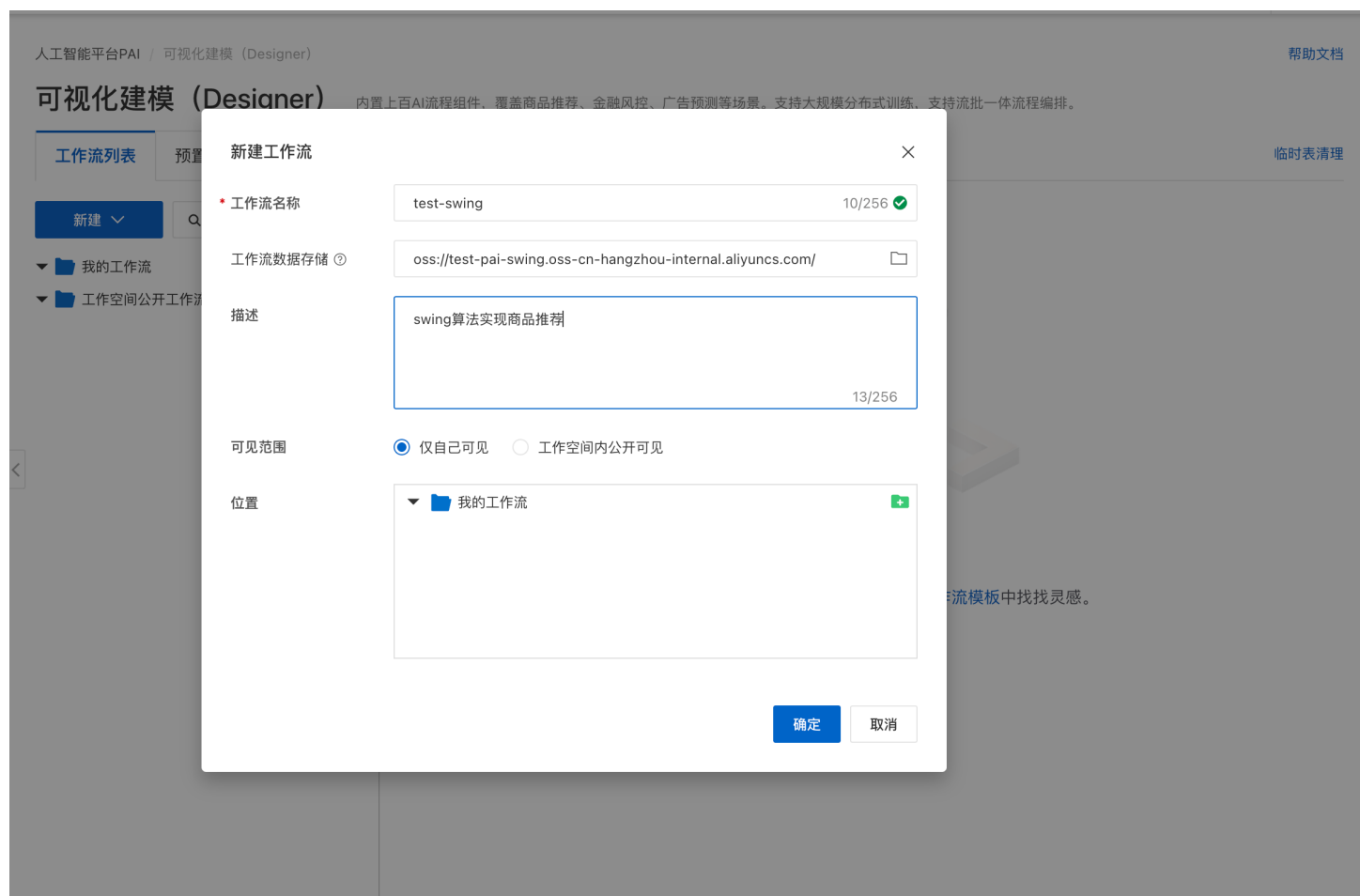
暂无可关联资源，如需使用更多资源，请联系管理员前往资源仪表中进行新建。

4. 工作流设计

点击左侧的可视化建模，新建工作流



设置对应的工作流名称，此处我将输出指定到特定的OSS Bucket内



进入工作流，我们拖拽左侧的相应控件完成，**读CSV文件、swing训练**



5. 节点配置

1. 读CSV文件

- 文件路径，选择对应的文件路径即可
- Schema: userid string, itemid string, score double, ts bigint

读CSV文件-1 ? 文档

参数设置 执行调优

文件来源

OSS

文件路径

oss://test-pai-swing.oss-cn-hang...10.csv

roleArn

已完成授权

* Schema

userid

STRING

itemid

STRING

score

DOUBLE

ts

BIGINT

2. swing训练-1，配置itemid以及userid

swing训练-1 ? 文档

字段设置 参数设置 执行调优

* Item列列名

itemid

* User列列名

userid

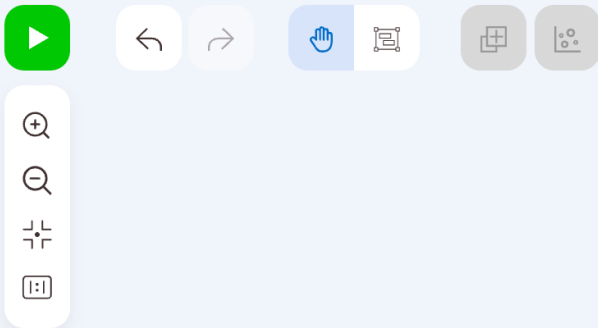
6. workflow全局配置

工作流属性	全局变量
所属工作空间	
pai_ar22s0ly1zeffs70x6	
工作流名称	
test-swing	
工作流ID	
draft-qohdj5an1eso9ibqd1	
创建人(UID)	
208723242804039005	
创建日期	
2025-03-24 22:14:16	
可见范围	
仅自己可见	
工作流描述	
swing算法实现商品推荐	
Alink/FlinkML算法默认执行资源	
 MaxCompute	▼
工作流数据存储 ②	
<div>oss://test-pai-swing.oss-cn-hangz...s.com/ </div>	

7. 任务运行及结果

点击左上角绿色运行按钮，即可开始运行

运行过程中可以查看运行日志



- | | |
|------------|--------|
| 重命名 | |
| 删除 | Delete |
| 复制 | Ctrl+C |
| 执行该节点 | |
| 从此处开始执行 | |
| 执行到此处 | |
| 从根节点执行到此处 | |
| 执行该节点的下游节点 | |
| 查看数据 | > |
| 查看日志 | |

模型

Shift + 鼠标左键 多选组件

数据-swing训练-1-... x

查看表信息

(数据预览仅支持1000条记录，如果需要查看全表，请前往 MaxCompute/DataWorks。详情)

A	B
mainitems	recommitemandsimilarity
1 B000063XJL	{"object":["B00005N7TL,B00006GXD4,B00007BK3L,B00005NIPP,B00005I
2 B000IMVNUQ	{"object":["B00005N7TL,B00006KTOK,B00008DP07,B00006KULE,B00005I
3 B000INCK4I	{"object":["B00005N7PN,B001LF4EVO,B00005N7QN,B000BW56WO,B000
4 B0018P3HR2	{"object":["B00005N7QA,B00005NIOH,B00007BK3L"],"score":[1.7655526;
5 B00006KFT2	{"object":["B00005N7OV"],"score":[{"2.4962463284507357}]}
6 B000OPOEEE	{"object":["B00005NIOQ"],"score":[{"2.3182902485056736}]}
7 B000ICB4T6	{"object":["B00005N7RA,B00005N7QN,B00079RO7G,B00007AWXX,B000I
8 B00006LOD8	{"object":["B000ICB4X2,B00005N7TB"],"score":[{"1.853286268473603,1.85
9 B01844RZRE	{"object":["B00X6LREH2,B00005R8BR"],"score":[{"2.419168483778265,2.2
10 B000ILU97A	{"object":["B00006FXOT,B00006FXOQ,B00006L2Z8,B00006FXOV,B00006I
11 B00007LN7R	{"object":["B00007BK3L,B00007AX0N,B00005N7QI"],"score":[{"2.89276516
12 B000ILZ6VE	{"object":["B00079RO7G,B000ILY9LW,B00005N7PN,B01CF3ECNK,B000FT
13 B00005QJDW	{"object":["B00005N7U1,B00005N7T3,B00005N7TL"],"score":[{"2.4330984;
14 B0000633EA	{"object":["B00005N7QG,B000ICB4X2,B00005N7TB,B00005N7SM"],"score
15 B00005NIN8	{"object":["B00005R8BL,B00005N7SC,B00005NIOH,B00005NIRE,B00005
16 B00005UMOT	{"object":["B000EGCIW8,B00005N7PN,B00005N7QW,B000ILYHSM,B0000
17 B00006KTOK	{"object":["B00005NIOC,B000IMVNUQ,B00008DP07"],"score":[{"4.3103316
18 B002MRSLKO	{"object":["B000069YW9"],"score":[{"2.3182902485056736}]}

作业

1. 完成配置，实现swing模型训练过程

- 说明读CSV组件的配置以及Swing训练组件的配置
- 取某一条结果，解读swing-训练组件的模型结果

- c. 尝试添加Swing-推荐组件，绘制工作流截图（无需进行计算）
- d. 简单谈谈你的Designer模式的理解