

元灵数智-个性化推荐服务使用说明

1. 产品订阅

在AWS MarketPlace页面搜索"DMetaSoul Recommendation Service", 点击"继续订阅", 订阅完成之后, 待条款更新完成, 产生生效日期之后, 点击“继续配置”。



< 产品详细信息 订阅

订阅此软件

您已订阅此软件。请查看下面的条款和定价详细信息, 或单击上面的按钮配置您的软件。

条款和条件

元灵数智 Offer

您已经订阅此软件并且您同意根据定价条款和此软件提供商的[最终用户许可协议\(EULA\)](#)使用本软件。您同意我们可以根据[西云数据关于亚马逊云科技 \(宁夏区域\) 的隐私政策](#)与相应的经销商, 分销商或此软件的提供商共享有关此交易的信息 (包含您的支付条款)。我们将向您发送付款通知并从您或者您的管理账户收取款项。您对亚马逊云科技 (宁夏) 区域和 (北京) 区域服务的使用仍应遵循[西云数据关于亚马逊云科技中国 \(宁夏\) 区域的客户协议](#)或其他有关使用此类服务的协议。

产品名称	生效日期	失效日期	可用操作
DMetaSoul Recommendation Service	2022/12/30	N/A	显示详细信息

选择交付选项为“CloudFormation 模板”, 点击“继续启动”。

[< 产品详细信息](#) [订阅](#) [配置](#)

配置软件

选择执行选项和软件版本以启动此软件。

交付选项

CloudFormation 模板

选择交付选项

Amazon 系统映像

CloudFormation 模板

CloudFormation 模板
使用 CloudFormation 模板部署完整的解决方案配置

软件版本

v1.0 (2023年1月17日)

版本详细信息

DMetaSoul Recommendation Service
运行于 t2.small

[更多](#)

地区

中国 (宁夏)

Product Code: 85b8a2o08ya7aovi0fv30yejg

[发行说明 \(更新于 2023年1月17日\)](#)

定价信息

这是根据您的配置对软件和基础设施成本的估计。您每个周期报表的实际费用可能不同于此估计值。

软件定价

DMetaSoul Recommendation Service ¥0/hr

BYOL

运行于 t2.small

最后，点击“启动”，即可进入CloudFormation部署页面。

启动此软件

查看启动配置详细信息并按照说明启动此软件。

配置详细信息

交付选项	DMetaSoul Recommendation Service DMetaSoul Recommendation Service 在 <i>t2.small</i> 上运行
软件版本	v1.0
区域	中国 (宁夏)

[使用说明](#)

选择“操作”

选择此操作即可通过 Amazon CloudFormation 控制台启动配置。

启动

2. CloudFormation部署引导

2.1 指定CloudFormation模板文件

注：确保操作的aws用户为管理员用户（涉及创建IAM role, Instance, RDS, DocumentDB）。

订阅启动完成之后，会自动跳转到CloudFormation页面，将会默认提供部署模板，点击“**下一步**”填写参数步骤。

CloudFormation > 堆栈 > 创建堆栈

第 1 步
创建堆栈

第 2 步
指定堆栈详细信息

第 3 步
配置堆栈选项

第 4 步
审核

创建堆栈

先决条件 - 准备模板

准备模板
每个堆栈都基于一个模板。模板是一个 JSON 或 YAML 文件，其中包含有关您希望在堆栈中包含的 亚马逊云科技 资源的信息。

模板已就绪 使用示例模板 在设计器中创建模板

指定模板

模板是一个 JSON 或 YAML 文件，该文件描述了堆栈的资源 and 属性。

模板源
选择模板会生成一个 Amazon S3 URL，在那里它将被存储。

Amazon S3 URL 上传模板文件

Amazon S3 URL

Amazon S3 模板 URL

S3 URL: [在设计器中查看](#)

[取消](#) [下一步](#)

2.2 填写所需参数。

服务

aws-ops @ 1328-2554-2956 宁夏 支持

CloudFormation > 堆栈 > 创建堆栈

第 1 步
创建堆栈

第 2 步
指定堆栈详细信息

第 3 步
配置堆栈选项

第 4 步
审核

指定堆栈详细信息

堆栈名称

堆栈名称

堆栈名称可以包含字母 (A-Z 和 a-z)、数字 (0-9) 和短划线 (-)。

参数

参数是在模板中定义的，并且允许您在创建或更新堆栈时输入自定义值。

DocumentDBClusterName
Cluster name

DocumentDBInstanceClass
Instance class. Please refer to: <https://docs.aws.amazon.com/documentdb/latest/developerguide/db-instance-classes.html#db-instance-classes-by-region>

DocumentDBInstanceName
Instance name

DocumentDBPassword
At least 8 printable ASCII characters. Can't contain any of the following: / (slash), ' (single quote), " (double quote) and @

DocumentDBUser
The database admin account username

PGDBAllocatedStorage
The size of the database (Gb)

PGDBInstanceClass
The database instance type

PGDBName
The PostgreSQL database name

PGDBPassword
At least 8 printable ASCII characters. Can't contain any of the following: / (slash), ' (single quote), " (double quote) and @

PGDBPort
The PostgreSQL database name

PGDBUser
The PostgreSQL database admin account username

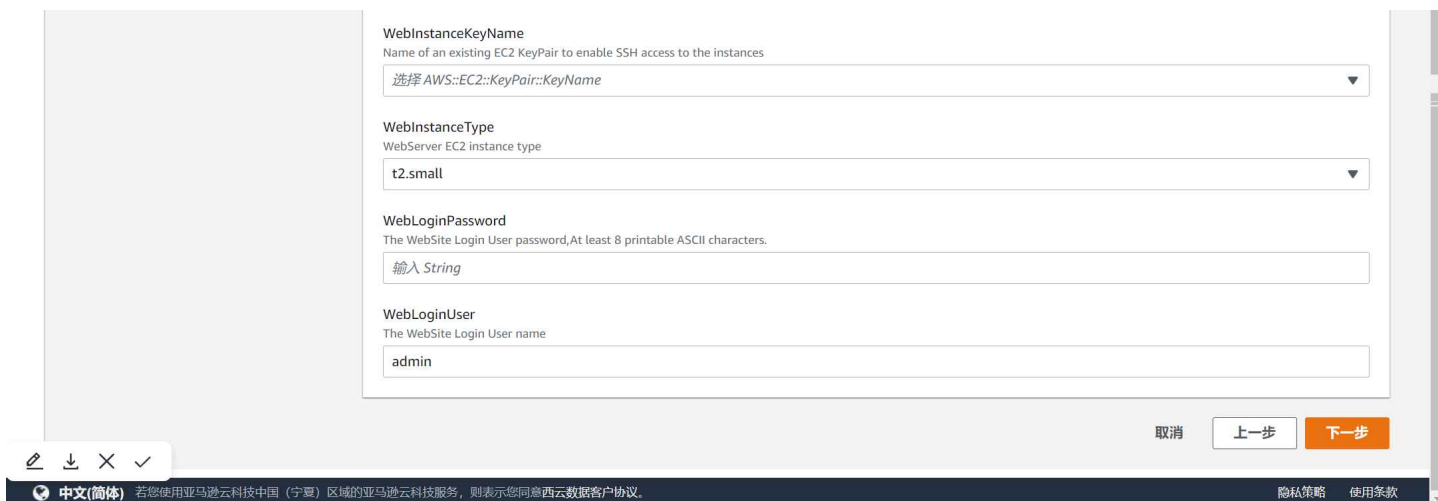
PGMultiAZDatabase
Create a Multi-AZ Amazon RDS database instance

S3BucketName
Name of an S3Bucket, we will create IAM access policy for this bucket

SSHLocation
The IP address range that can be used to SSH to the EC2 instances

Subnets
The list of SubnetIds in your Virtual Private Cloud (VPC)

VpcId
VpcId of your existing Virtual Private Cloud (VPC)



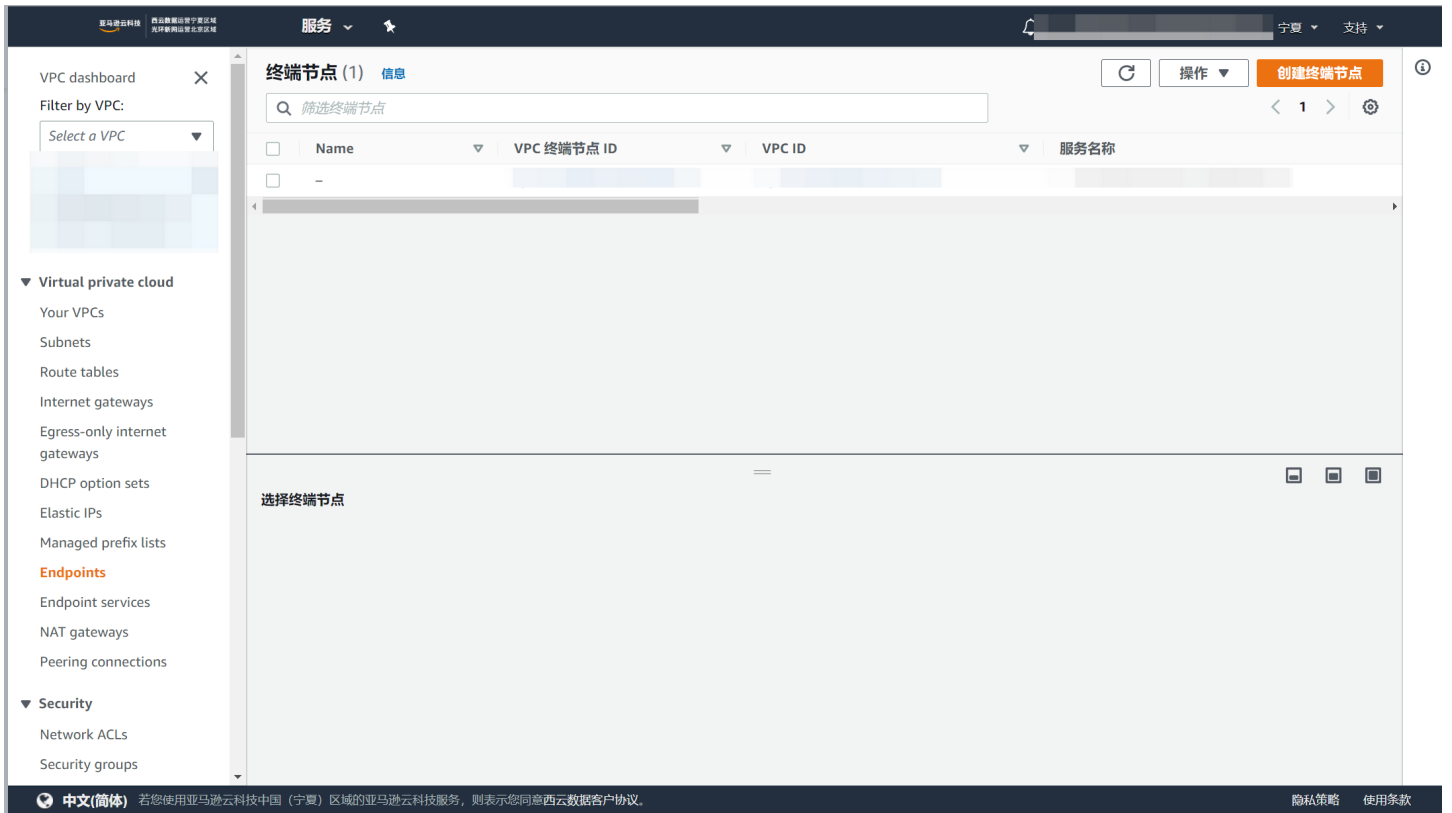
必填参数如下（其他参数均有默认值，按需修改）：

参数名称	说明
DocumentDBPassword	DocumentDB密码(Mongo连接)
PGDBPassword	RDS密码（PostgreSQL连接）
S3BucketName	后续S3工作目录所在的S3桶
Subnets	子网选择，至少两个
VpcId	内网VpcId选择
WebInstanceKeyName	实例绑定的SSH Key选择
WebLoginPassword	Dashboard登录密码

注：重新创建请删除上一个该CloudFormation创建的Stack，否则会出现资源已存在等冲突问题。

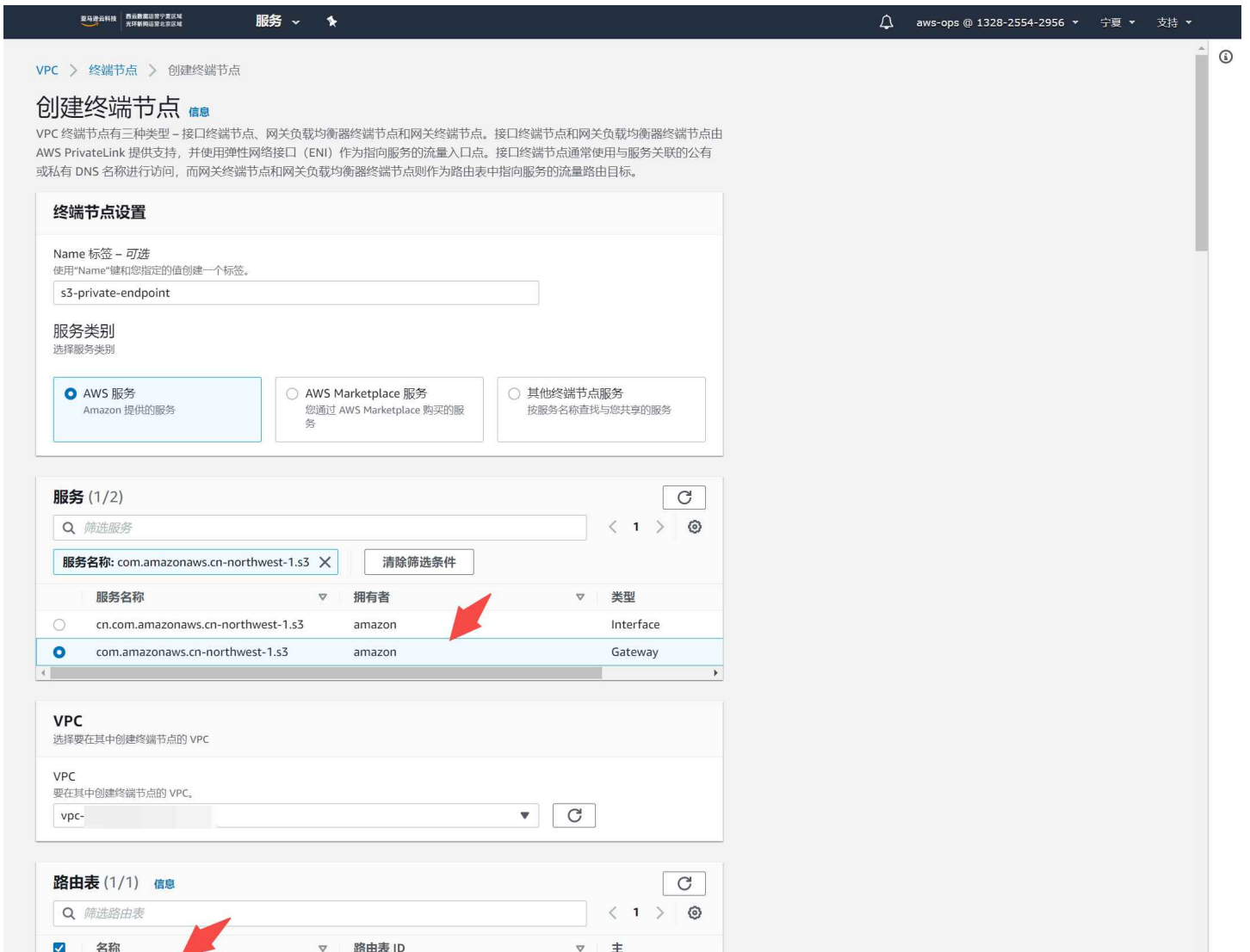
2.3 创建VPC S3 Endpoint。

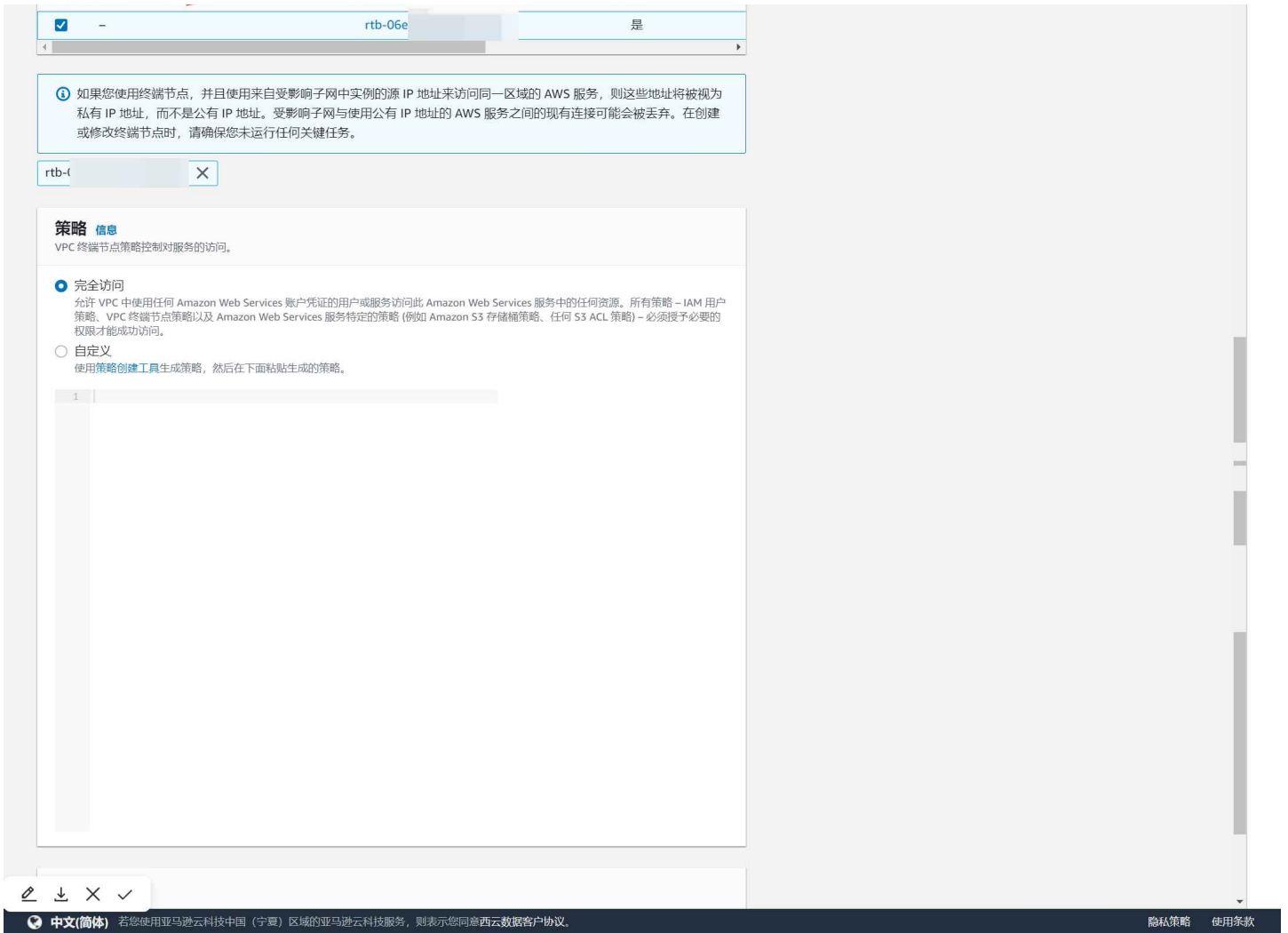
创建S3 Endpoint, 保证SageMaker内网可以进行读写。



点击右上角创建终端节点。

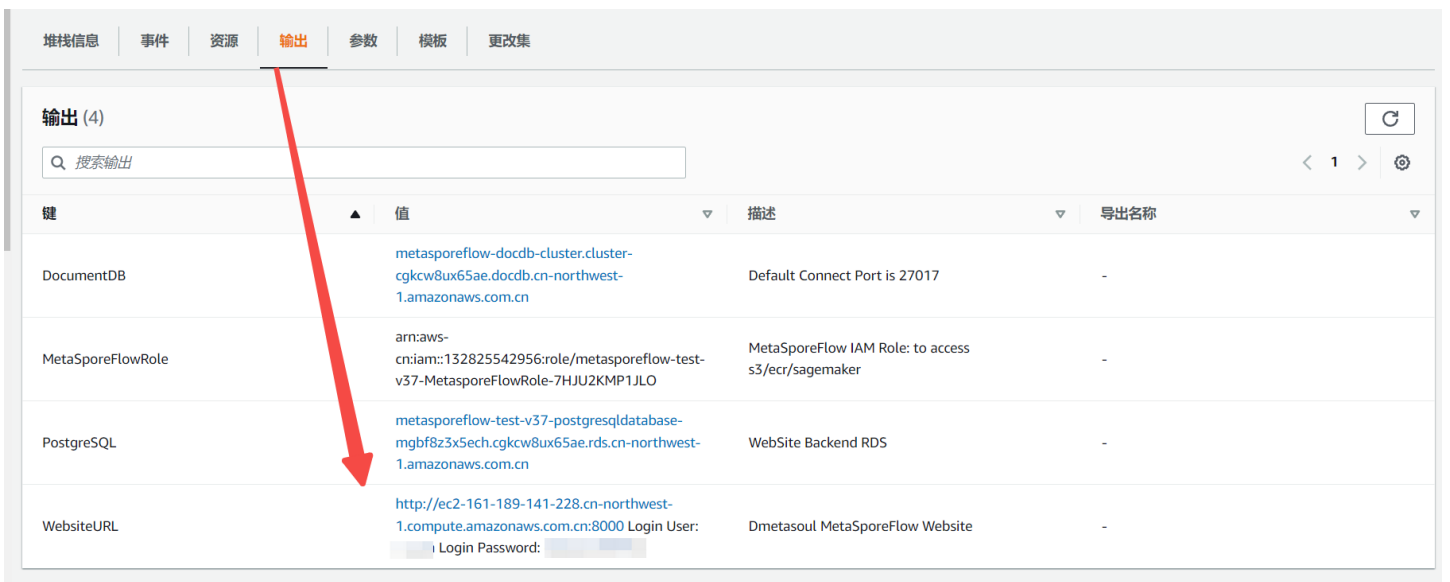
按照如下截图进行创建，创建完成之后，SageMaker即可通过内网访问S3。





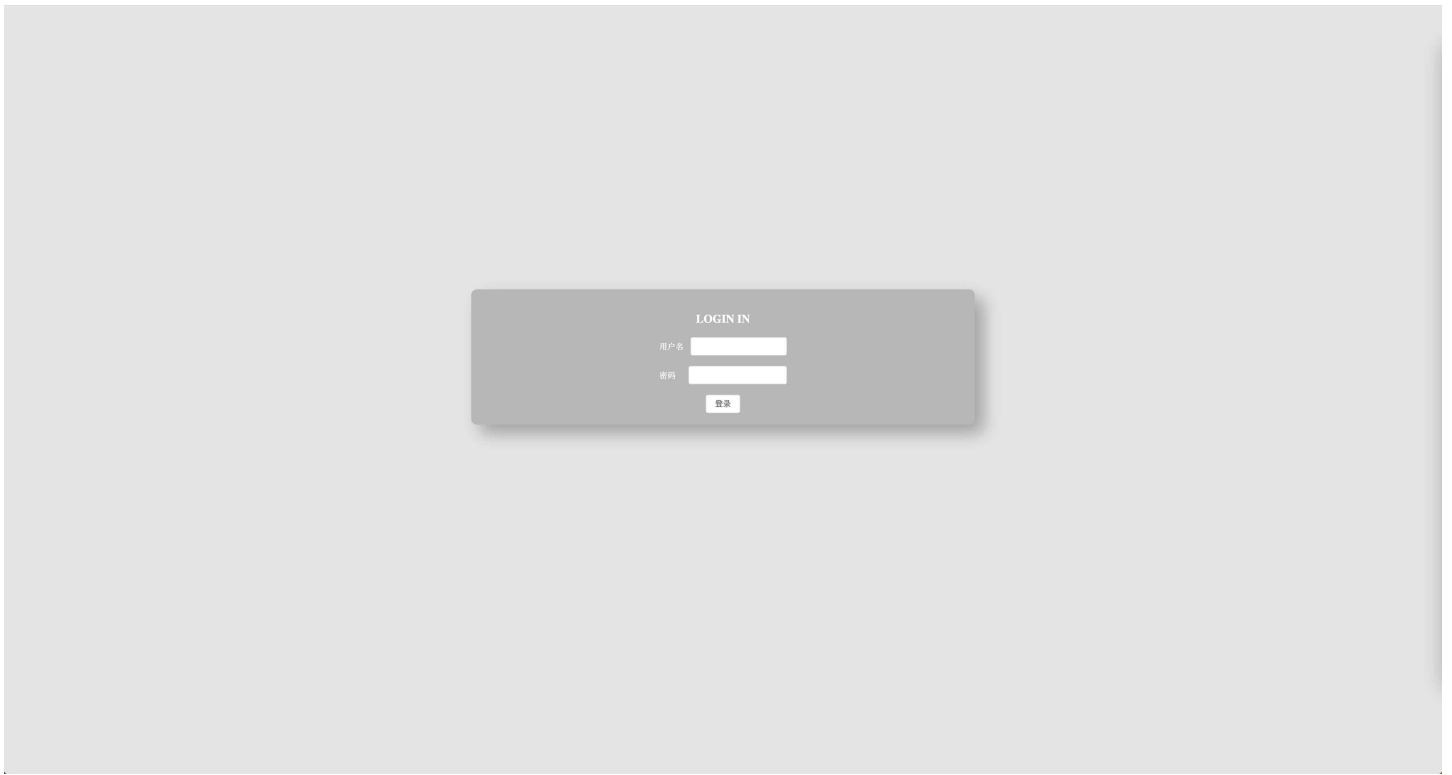
2.4 部署完成

CloudFormation资源创建完成之后，即可通过Web Url地址进行访问Metasporeflow Dashboard：

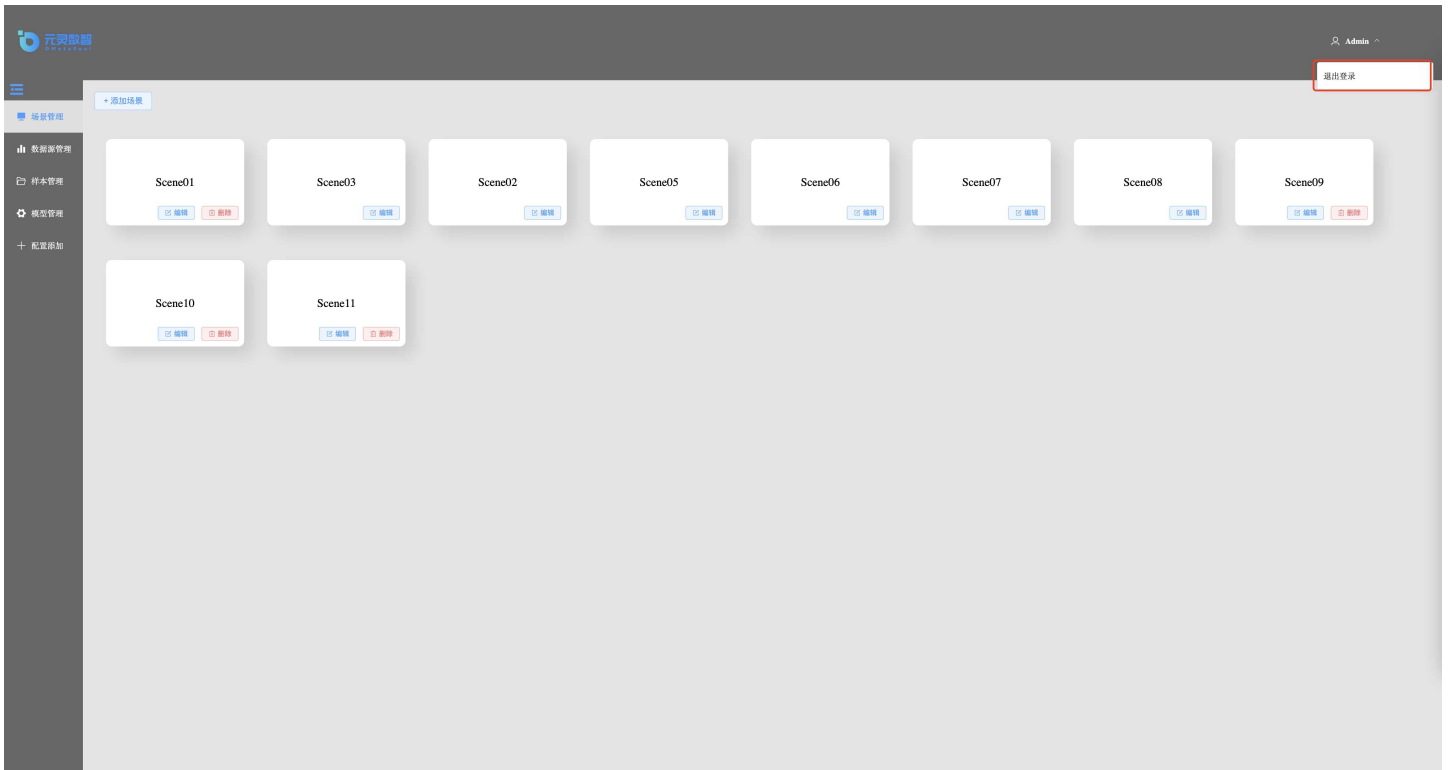


3. 登录和授权

在AWS部署 SaaS Dashboard时初始化一个用户，即可以登录。



点击退出登录即可以退出当前用户登录。



设置授权信息

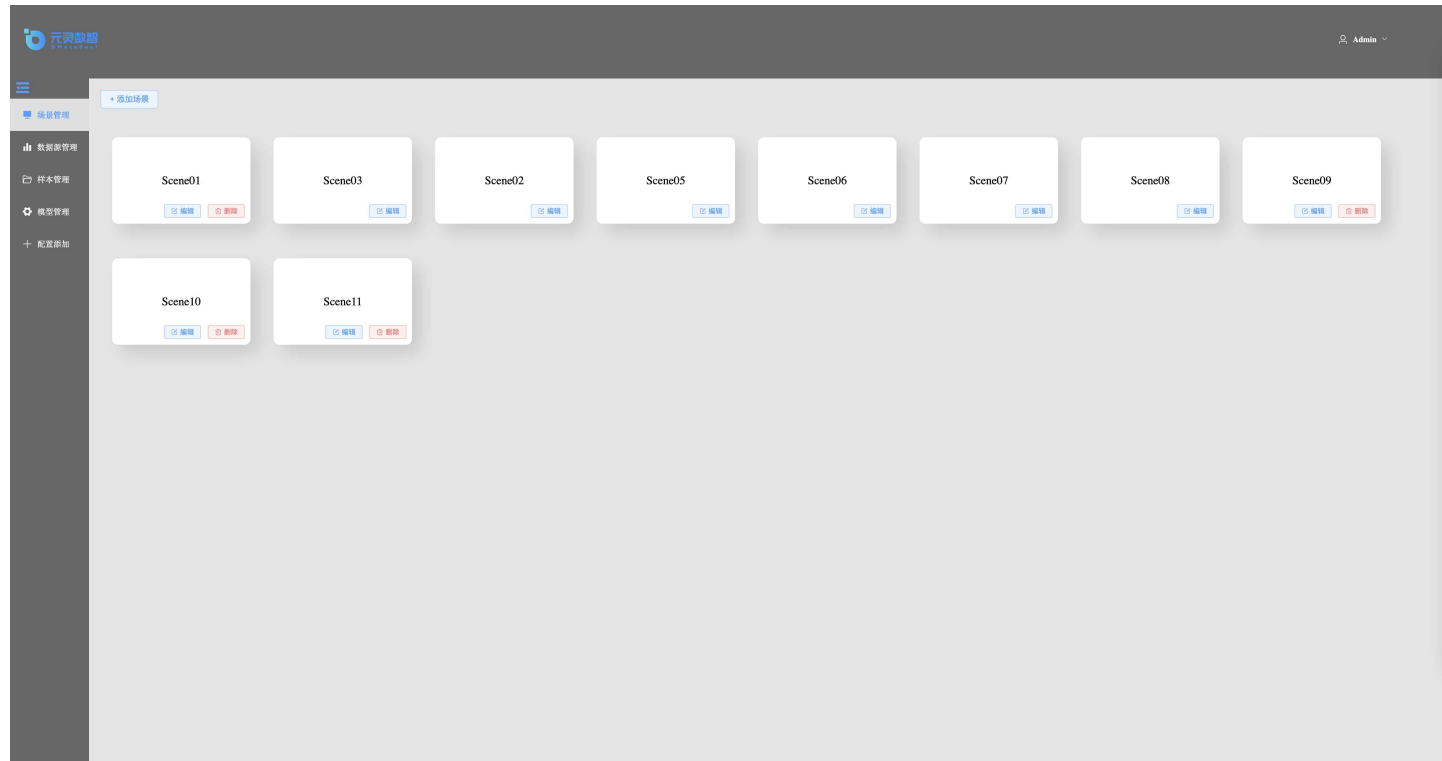
首次登录后需要去申请授权码，将授权信息输入到授权信息框，进行授权。



授权码请通过邮件申请，邮箱地址为：public-contact@dmetasoul.com

4. 场景管理

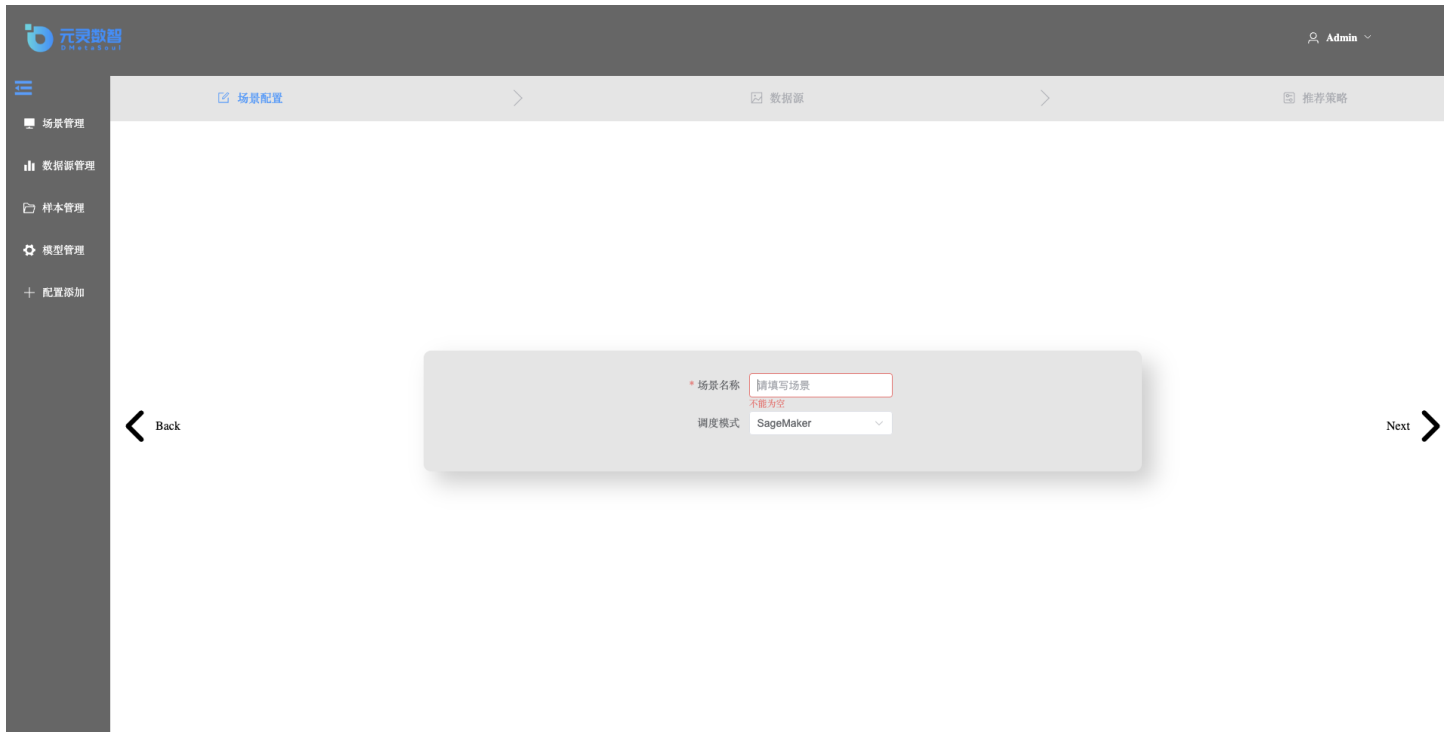
进入场景管理界面，可以查看已创建的场景列表。



点击**添加场景**按钮，新增一个场景，然后根据页面展示一步一步输入场景相关信息：

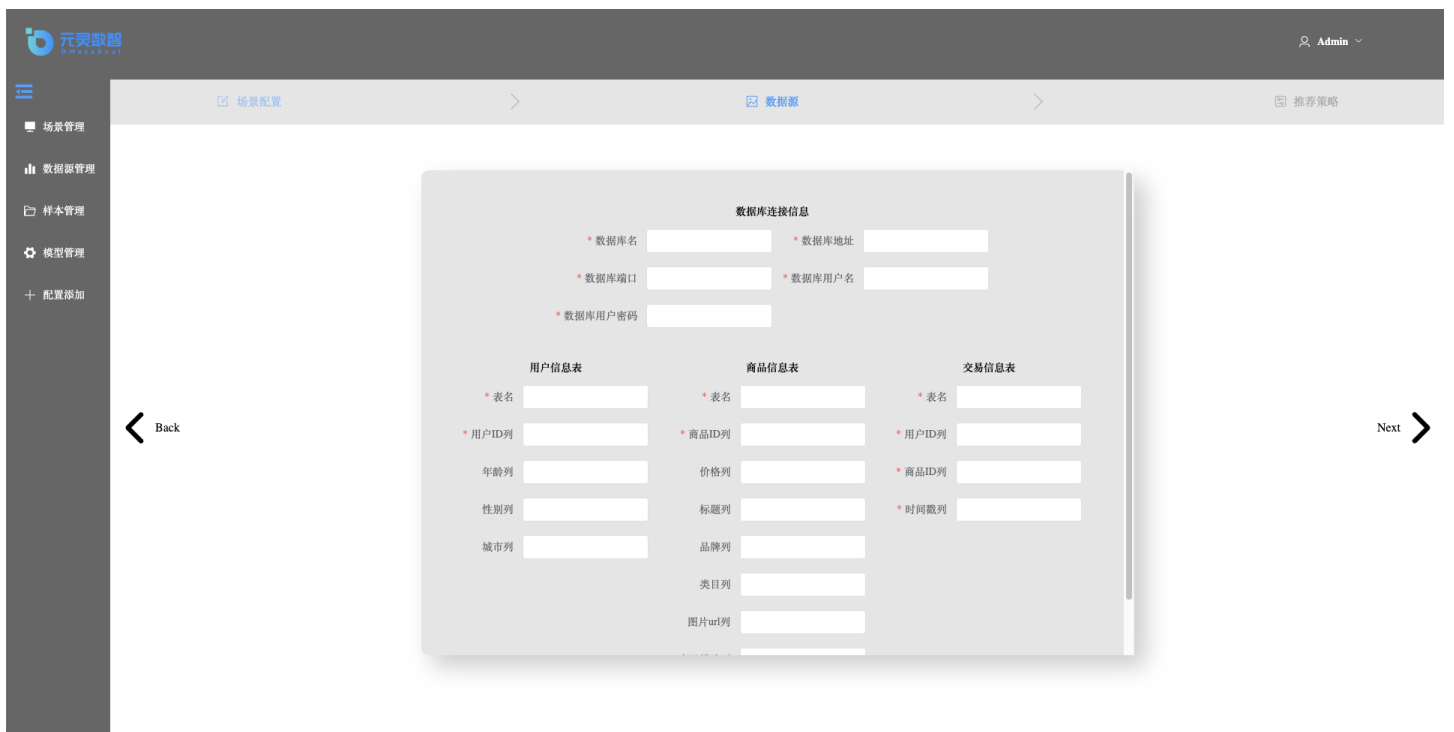
1. 场景基础信息配置

在此页面上：输入场景名称、选择调度模式；



2. 场景数据源以及中间数据存储配置

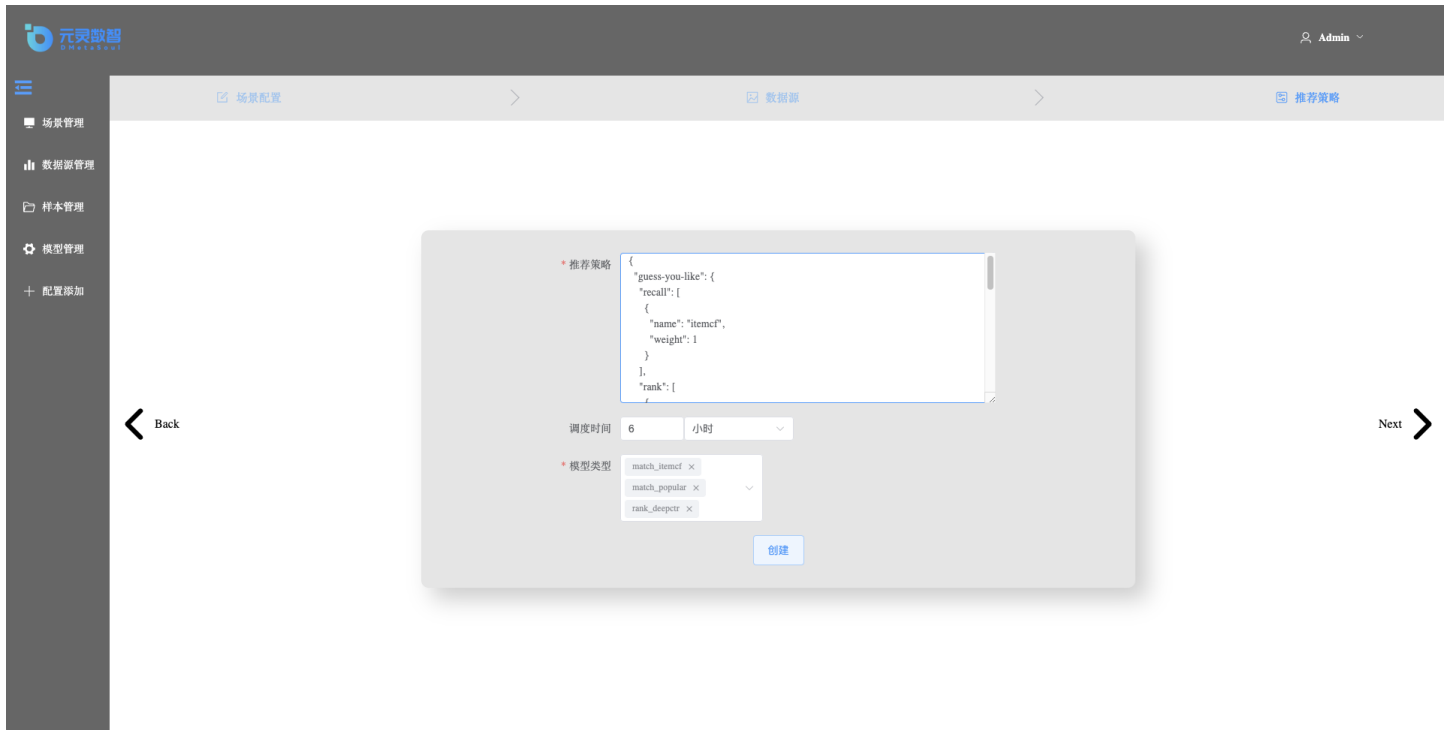
在此页面上可以选择数据源信息，如下图所示：



- 在上面填入商品用户交易的MySQL数据库信息包括：数据库名称、数据库地址、数据库端口、数据库用户名以及用户密码；
- 填入商品用户表，包括表名，用户ID列，以及可选的年龄列、性别列以及城市列。
- 填入商品信息表，包括表名，商品ID列，以及可选的价格列、标题列、品牌列、类目列、图片URL列以及商品描述列。
- 填入交易信息表，包括表名，用户ID列，商品ID列以及时间戳列。

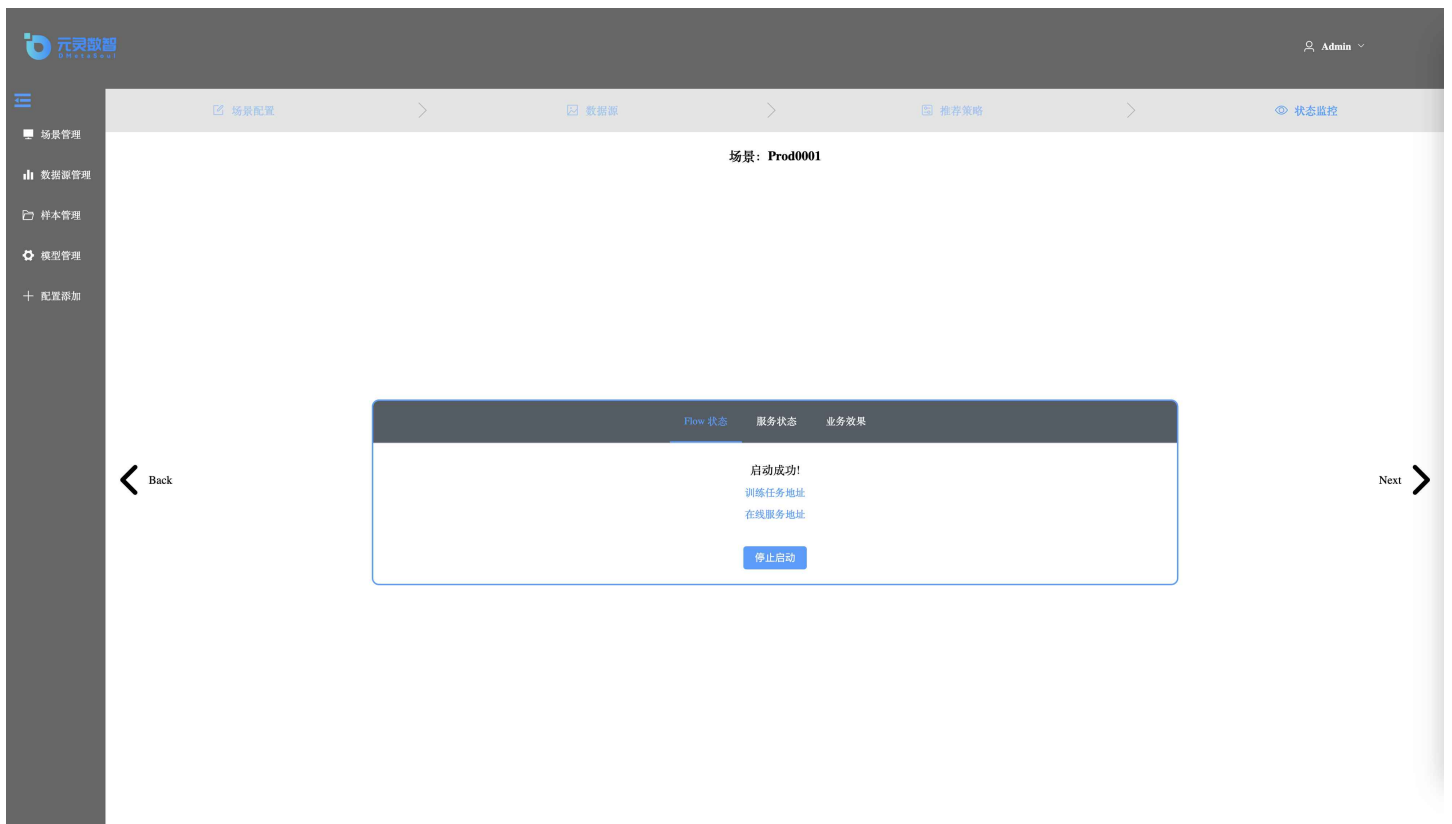
3. 推荐策略配置

在此页面上可以修改推荐策略、设置调度时间以及选择模型类型，如下图所示，点击创建按钮可以创建一个新的场景。



4. 场景状态查询

启动成功后，再次进入场景界面，点击**状态监控**，会有如下状态展示，目前Flow的状态分为: 未启动、启动成功、训练中、在线服务启动中、训练失败、在线服务启动失败。点击**训练任务地址**或者**在线服务地址**可以跳转到相应界面查看详细情况。



在启动后，点击**停止**启动按钮可以停止当前的场景任务。

对当前正在运行的场景，点击**编辑**按钮，可以对当前场景的AB实验进行更新。

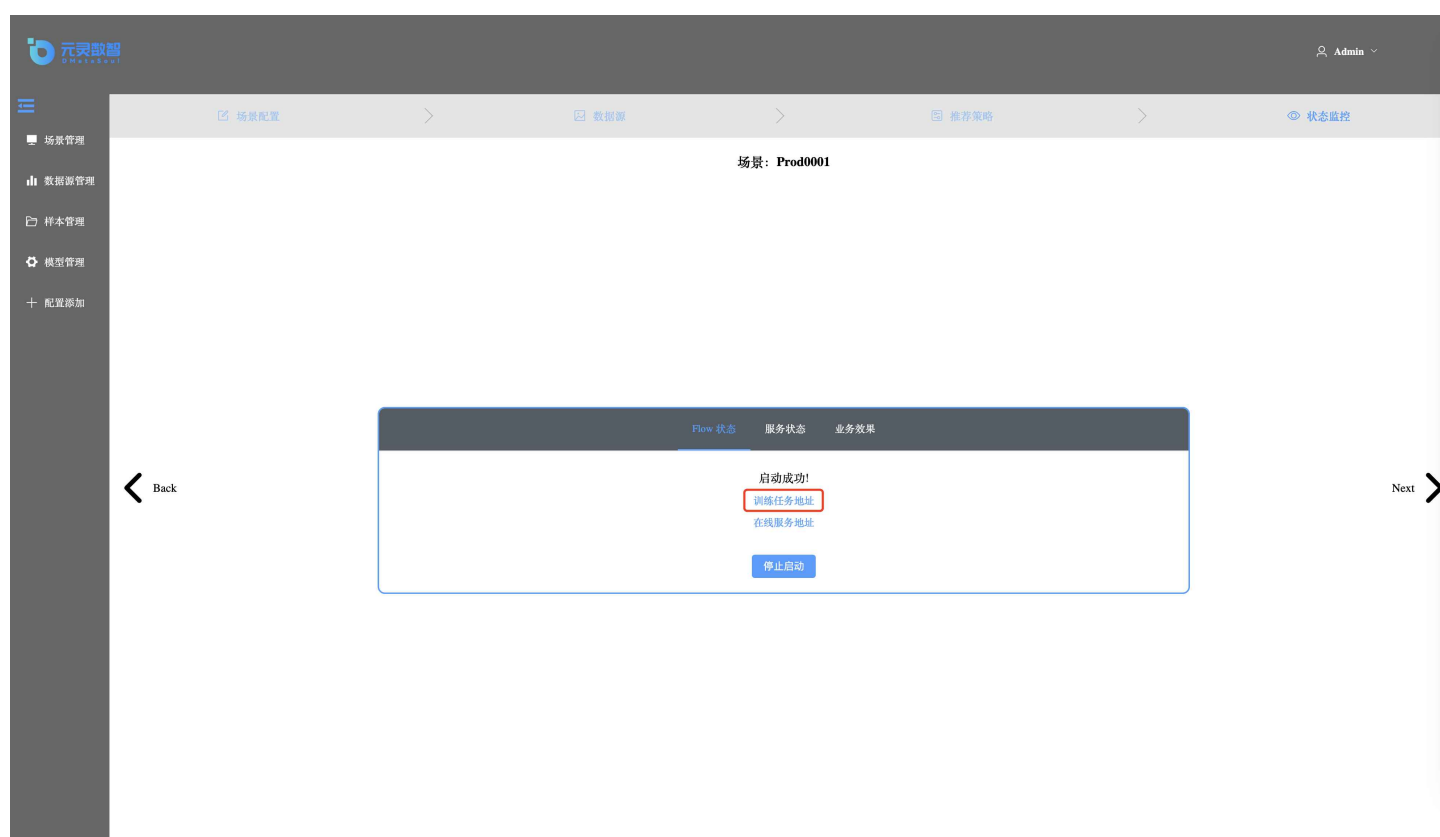
对已经停止的场景，点击**删除**按钮，可以对当前场景进行删除。

5. 在 SageMaker 上查看调度结果

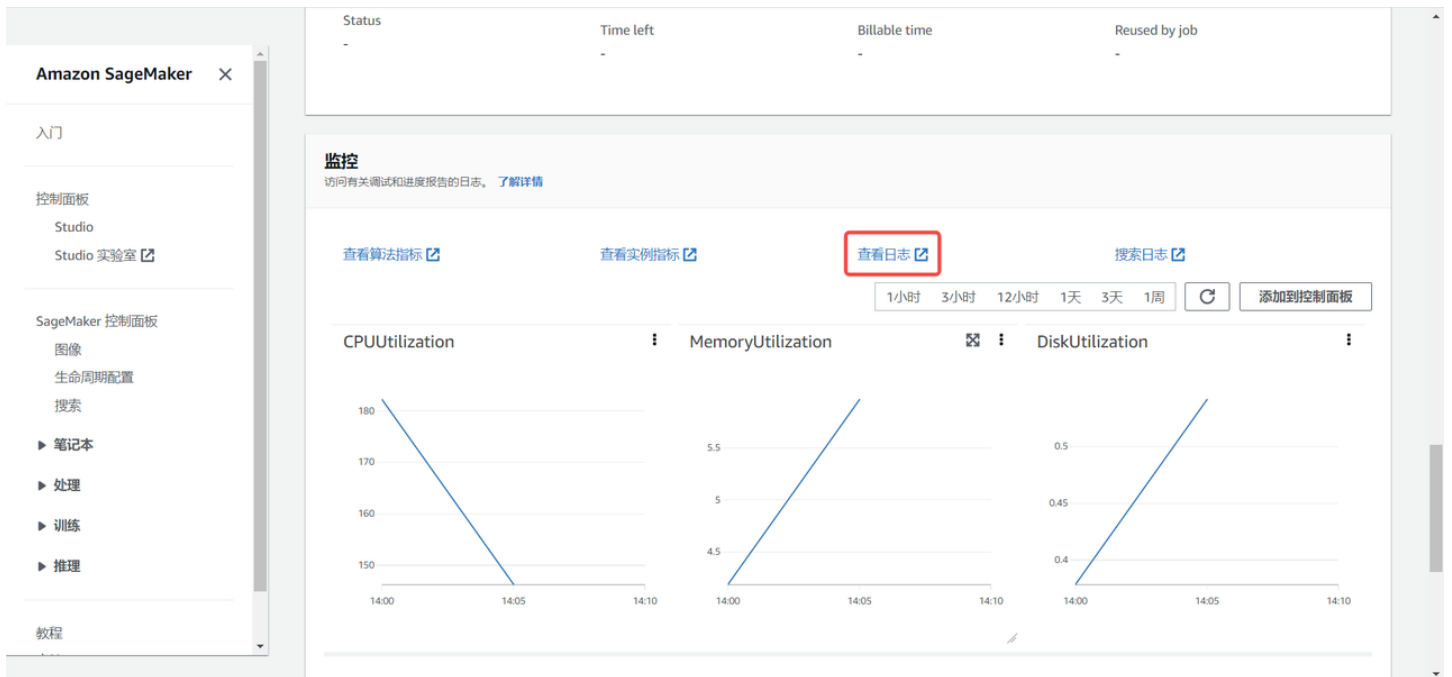
Dashboard 上创建的推荐场景会在后台创建定时任务，转化为 SageMaker create_training_job / create_endpoint / update_endpoint 实现调度。

5.1 查看离线训练任务

其中模型的离线训练通过 SageMaker create_training_job 调用，用 MetaSporeFlow 自定义 Docker 镜像进行训练。可以点击场景状态下的**训练任务地址**按钮跳转到当前训练任务详情页。



训练任务的日志可通过点击点**查看日志**，再点日志流条目查看。



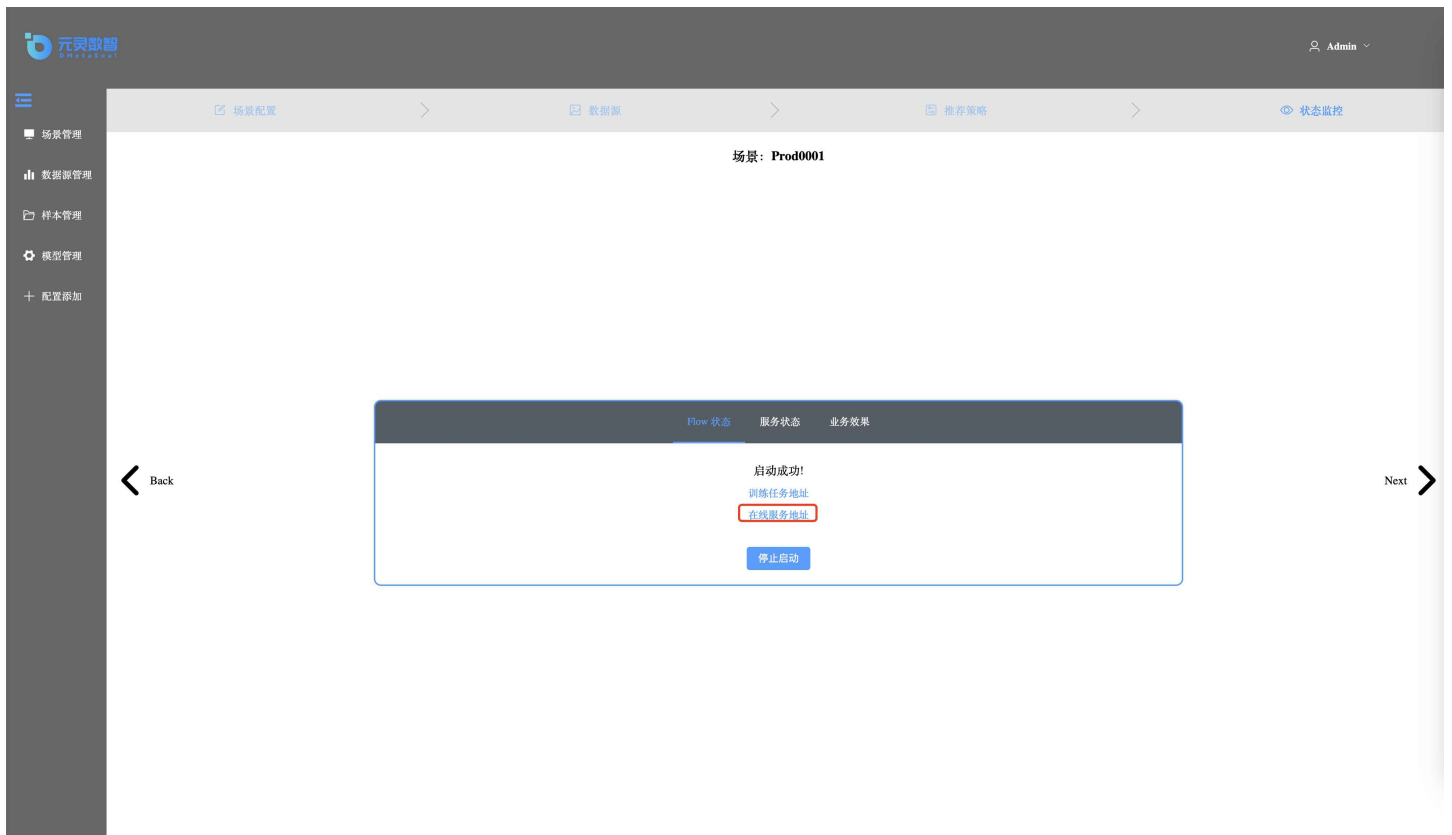
The screenshot shows the AWS CloudWatch console for the path '/aws/sagemaker/TrainingJobs'. The breadcrumb is 'CloudWatch > 日志组 > /aws/sagemaker/TrainingJobs'. There are buttons for '操作' (Actions), '在 Logs Insights 中查看' (View in Logs Insights), and '搜索日志组' (Search Log Groups). The main section is '日志组详细信息' (Log Group Details) with tabs for '日志流' (Log Streams), '指标筛选条件' (Metric Filters), '订阅筛选条件' (Subscription Filters), 'Contributor Insights', and '标签' (Tags). Under '日志流 (100+)', there is a search bar with 'demo-metaspore-flow-20221208-1625-train' and '1 match' results. The search bar has a '完全匹配' (Exact Match) checkbox. Below the search bar is a table of log streams:

<input type="checkbox"/>	日志流	上次事件时间
<input type="checkbox"/>	demo-metaspore-flow-20221208-1625-train/algo-1-1670488021	2022-12-08 16:35:08 (UTC+08:00)

The log stream name 'demo-metaspore-flow-20221208-1625-train/algo-1-1670488021' is highlighted with a red box.

5.2 查看在线服务

当 training job 成功完成时，会用导出的模型调用 SageMaker create_endpoint 创建 endpoint，或调用 SageMaker update_endpoint 更新 endpoint。可以点击场景状态下的**在线服务地址**按钮跳转到在线服务详情页。



其中可查看 endpoint 所使用的模型。



进一步点进模型的详情页可查看模型在 S3 上的具体路径。

Amazon SageMaker 服务

bigdata-flow-sagemaker-test-model-20221208-085722

操作 创建批量转换任务 创建终端节点

模型设置

名称	ARN	创建时间	IAM 角色 ARN
bigdata-flow-sagemaker-test-model-20221208-085722	arn:aws-cn:sagemaker:cn-northwest-1:132825542956:model/bigdata-flow-sagemaker-test-model-20221208-085722	Dec 08, 2022 00:57 UTC	arn:aws-cn:iam:132825542956:role/service-role/AmazonSageMaker-ExecutionRole-20221019T142662

容器 1

容器名称 Container 1	模型数据位置 s3://dmetasoul-test-bucket/demo/workdir/flow/scene/bigdata_flow_sagemaker_test/model/serving/bigdata-flow-sagemaker-test-20221208-085722.tar.gz
图像 132825542956.dkr.ecr.cn-northwest-1.amazonaws.com.cn/dmetasoul-repo/metaspore-sagemaker-releasev1.0.1	模式 单个模型
训练作业 -	

在在线服务详情面点**查看日志**，再点日志流条目可查看在线服务日志。

Studio 实验室

SageMaker 控制面板

监控

访问 CloudWatch 日志以查看 Jupyter 笔记本的调试和进度报告。了解详情

查看调用指标 查看实例指标 **查看日志** 添加到控制面板

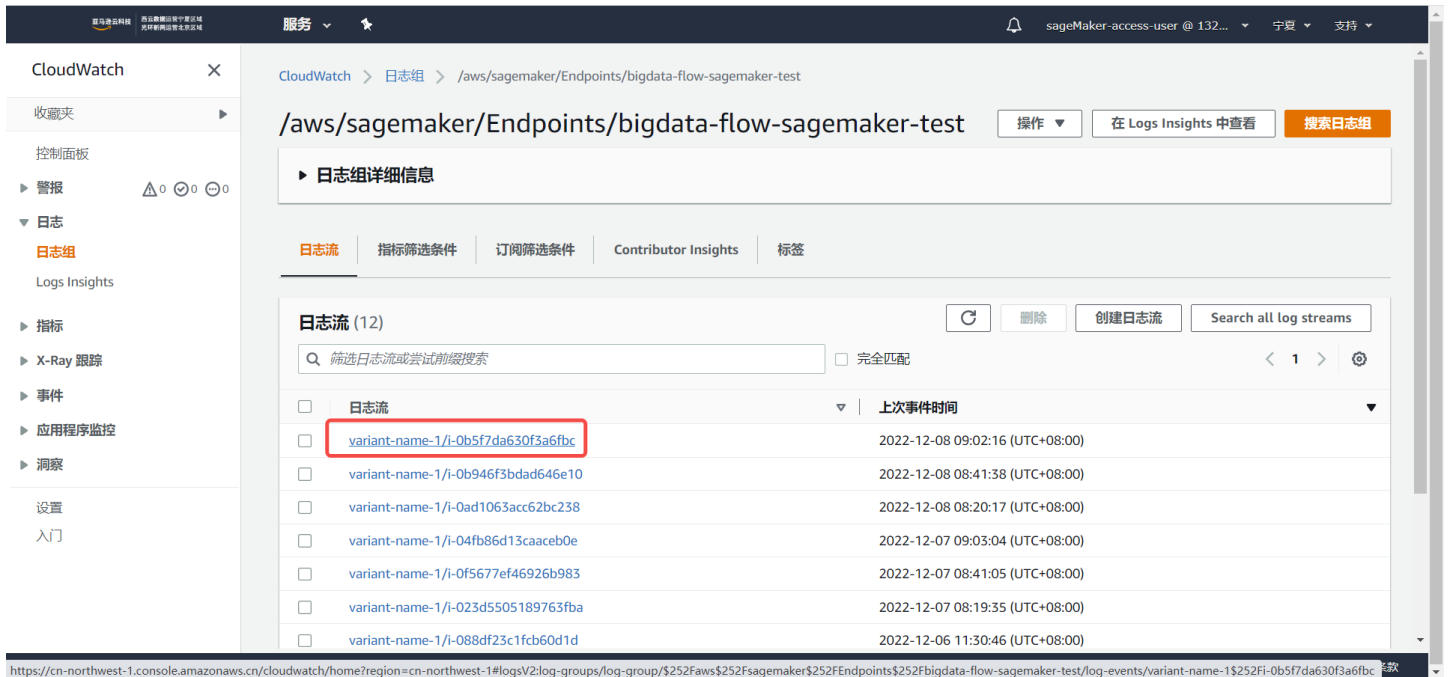
1小时 3小时 12小时 1天 3天 1周

CPU Utilization Memory Utilization Disk Utilization

终端节点运行时设置

更新权重 更新实例计数 配置 Auto Scaling

变量名称	当前权重	所需权重	实例类型	当前实例计数	所需实例计数	实例最小值 - 最大值	自动扩展
------	------	------	------	--------	--------	-------------	------



6. 调用在线服务

在线服务可以通过Python 调用或者部署AWS API Gateway两种方式进行调用，用户可根据实际需要进行选择。

6.1 通过Python调用

个性化推荐服务的在线服务部分会自动创建 SageMaker 的 Endpoint。由于 SageMaker 的 Endpoint 不能直接通过 http 协议访问，可以通过调用 SageMaker 的 API 来访问，以下是一个 Python 调用的示例：

```
1 import boto3
2 import json
3
4 # 创建 SageMaker Runtime Client, 第二个参数是 Region 名
5 runtime_sm_client = boto3.client("runtime.sagemaker", "cn-northwest-1")
6
7 # 创建 request json, 其中 operator 固定为 recommend, request 中需要传入 user_id 和
  scene, scene 为推荐策略页面中配置的名字
8 request = {"operator": "recommend", "request": {"user_id": "A1P62PK6QVH8LV", "s
  cene": "guess-you-like"}}
9 request_body = json.dumps(request)
10 endpoint_name= "bigdata-flow-sagemaker-test"
11
12 # 调用 SageMaker Endpoint, EndpointName 为 SageMaker Endpoint 名, 在上面示例中为
  "bigdata-flow-sagemaker-test"
13 response = runtime_sm_client.invoke_endpoint(
14     EndpointName=endpoint_name,
15     ContentType="application/json",
```

```

16     Accept='application/json',
17     Body=request_body,
18 )
19
20 # 读取并打印返回结果
21 res = json.loads(response["Body"].read())
22 print(res)

```

6.2 通过部署AWS API Gateway调用

6.2.1 部署前提

1. 确保当前使用的 AWS 账户已经在国内通过 ICP 备案，并需确保Amazon API Gateway可以通过公网/私网访问80及443端口（具体可咨询您的AWS BD）。
2. 创建新的IAM role，用来提供API Gateway访问SageMaker Endpoint需要的权限，具体创建如下：
 - a. 进入[Identity and Access Management \(IAM\)](#)，点击“**创建角色**”。
 - b. 选择可信实体类型：亚马逊云科技服务，使用案例：API Gateway

选择可信实体 [信息](#)

可信实体类型

<input checked="" type="radio"/> 亚马逊云科技服务 允许 EC2、Lambda 或其他 亚马逊云科技 服务在此账户中执行操作。	<input type="radio"/> 亚马逊云科技 账户 允许属于您或其他第三方的其他 亚马逊云科技 账户中的实体在此账户中执行操作。	<input type="radio"/> Web 身份 允许由指定的外部 Web 身份提供商联合的用户担任此角色，以便在您的账户中执行操作。
<input type="radio"/> SAML 2.0 联合 允许从公司目录通过 SAML 2.0 联合的用户在此账户中执行操作。	<input type="radio"/> 自定义信任策略 创建自定义信任策略以使其他人能够在此账户中执行操作。	

使用案例

允许 EC2、Lambda 或其他 亚马逊云科技 服务在此账户中执行操作。

常见使用案例

- EC2**
Allows EC2 instances to call AWS services on your behalf.
- Lambda**
Allows Lambda functions to call AWS services on your behalf.

其他 亚马逊云科技 服务的使用案例：

API Gateway

API Gateway
Allows API Gateway to push logs to CloudWatch Logs.

[取消](#) [下一步](#)

- c. 点击“**下一步**”，进入到步骤3，填写角色名称：**APIGatewayAccessToSageMaker**，点击“**创建角色**”。
- d. 创建完成之后，进入该角色详情页面，点击“**添加权限**”，选择“**创建内联权限**”。
- e. 选择Json，替换如下"**account-id**"为您的aws账户ID，并将json内容填入编辑器，点击“**查看策略**”，填写名称*，最后点击“**创建策略**”。

```

1 {
2     "Version": "2012-10-17",
3     "Statement": [
4         {

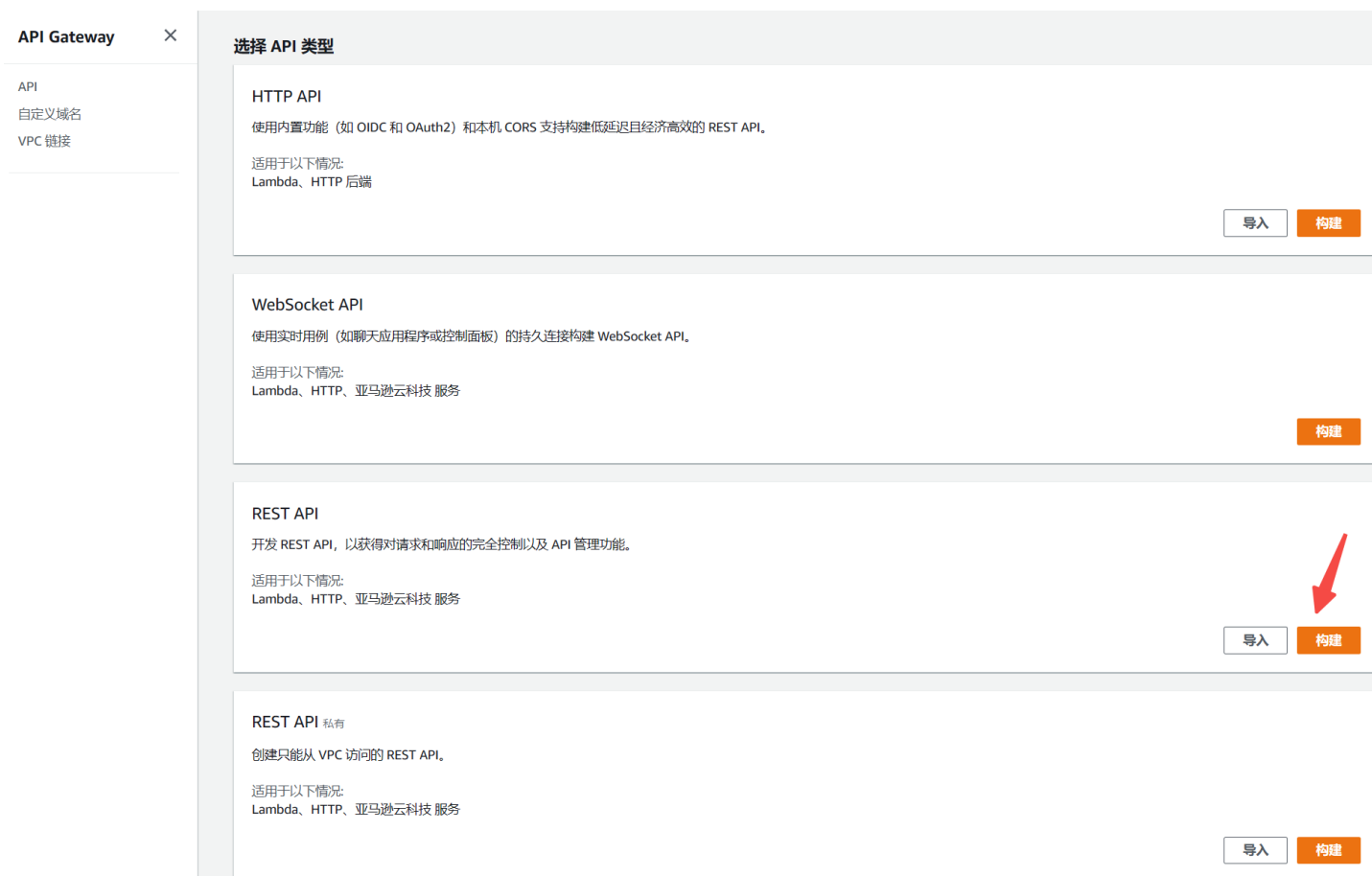
```

```
5     "Sid": "VisualEditor0",
6     "Effect": "Allow",
7     "Action": "sagemaker:InvokeEndpoint",
8     "Resource": "arn:aws-cn:sagemaker:*:[account-id]:endpoint/*"
9     }
10 ]
11 }
```

- f. 准备完成之后，记录该IAM role的arn，类似arn:aws-cn:iam::[account-id]:role/APIGatewayAccessToSageMaker。

6.2.2 创建公网API Gateway

1. 进入API Gateway控制台页面，选择“创建API”。
2. 选择REST API进行构建。



3. 填写API名称*，默认选择终端节点为“区域性”，点击“创建API”。



4. 创建完成, 选择资源, 创建POST方法, 按照如下截图进行配置:



其中路径覆盖和执行角色分别来自SageMaker Endpoint和创建的IAM role arn。

Amazon SageMaker ×

Amazon SageMaker > 端点 > BigdataScene06

BigdataScene06 删除

终端节点设置

名称	BigdataScene06	URL	https://runtime.sagemaker.cn-northwest-1.amazonaws.com/endpoints/BigdataScene06/invocations
ARN	arn:aws-cn:sagemaker:cn-northwest-1:[:redacted]:endpoint/bigdatascene06	详细了解 API	
状态	InService		
创建时间	Tue Jan 03 2023 14:50:51 GMT+0800 (中国标准时间)		
上次更新时间	Tue Jan 03 2023 14:54:58 GMT+0800 (中国标准时间)		

入门

控制面板

- Studio
- Studio 实验室 [🔗](#)

SageMaker 控制面板

- 图像
- 生命周期配置
- 搜索

- ▶ 笔记本
- ▶ 处理
- ▶ 训练
- ▶ 推理

5. 点击保存。

6. 选择操作->部署API:

资源

操作 / 方法

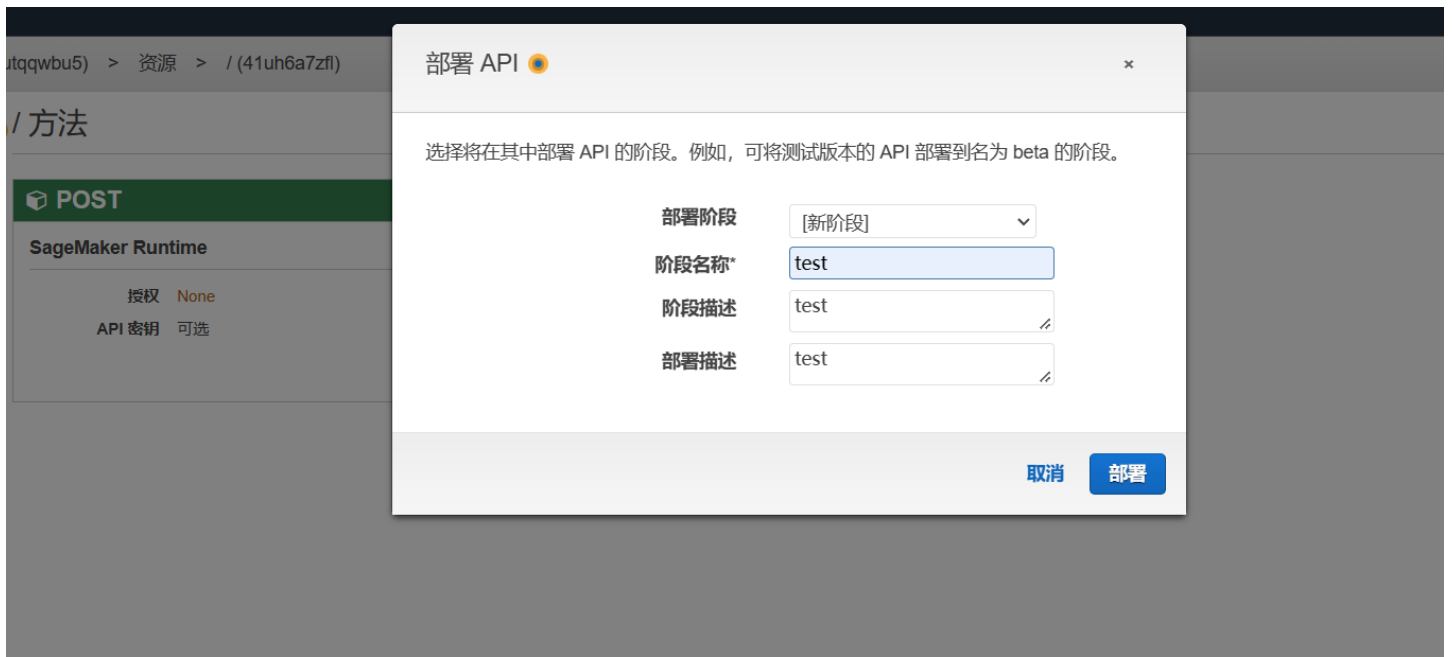
- 资源操作
 - 创建方法
 - 创建资源
 - 启用 CORS
 - 编辑资源文档
- API 操作
 - 部署 API**
 - 导入 API
 - 编辑 API 文档
 - 删除 API

POST

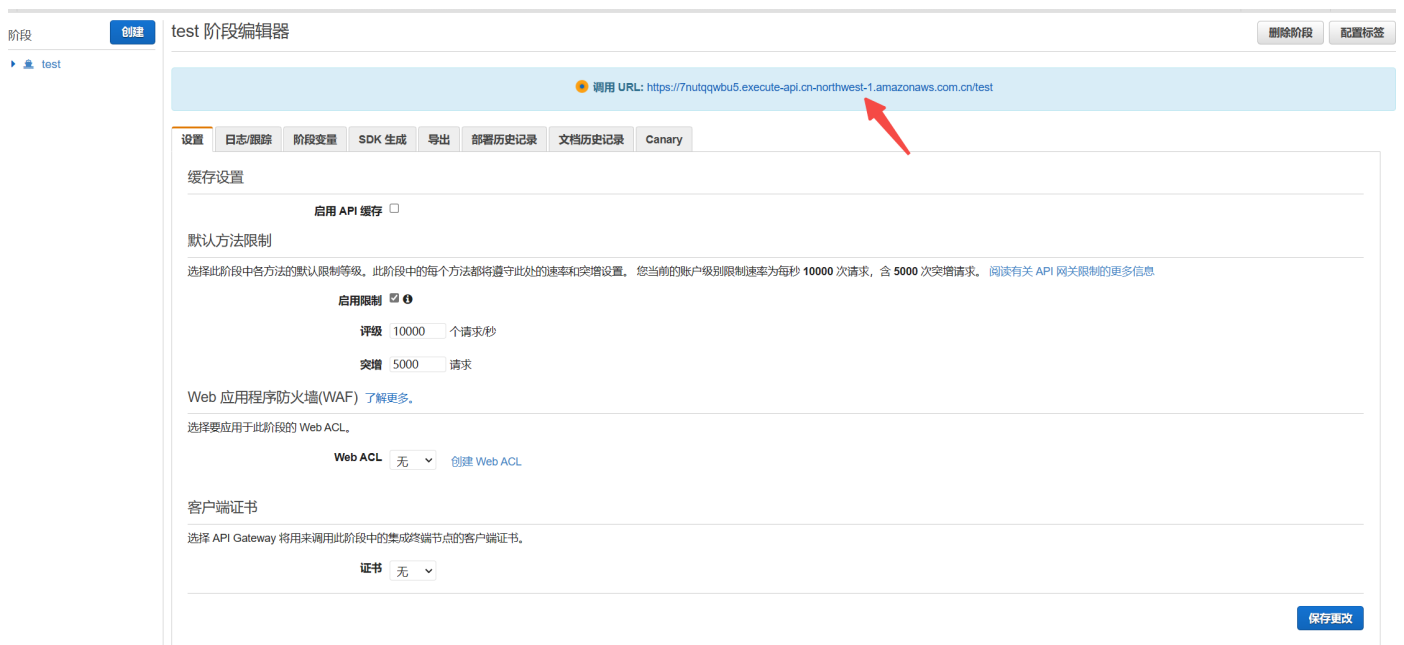
ker Runtime

授权 None

API 密钥 可选



7. 完成部署之后，可进入阶段页面，获取调用URL，即可通过公网调用该API访问个性化推荐在线服务。



调用在线服务用例：

```
1 curl -H "Content-Type: application/json" -X POST -d '{"operator": "recommend",  
  "request": {"user_id": "A1P62PK6QVH8LV", "scene": "guess-you-like"}}' "调用URL"
```

返回类似如下结果：

```
$ curl -H "Content-Type: application/json" -X POST -d '{"operator": "recommend", "request": {"user_id": "A1P62PK6QVH8LV", "scene": "guess-you-like"}}' "https://7nutqqwbu5.execute-api.cn-northwest-1.amazonaws.com.cn/test/"
{"code": "SUCCESS", "msg": "success!", "data": [{"score": 0.16278019547462463, "item_id": "B004DGJJX6", "price": "$10.00", "origin_scores": {"experiment.recall.swing": 3.2388524982513385, "widedeep": 0.16278019547462463, "swing": 0.048009352148519036}, "experiment.rank.widedeep": 0.1627802}, {"description": "K. Bell high quality trendy and comfortable novelty no show low cut sneaker women's socks | 6 pairs sock pack", "category": "Clothing, Shoes & Jewelry\u0001Novelty & More\u0001Clothing\u0001Novelty\u0001Women", "title": "6 Pack Novelty No Show Socks", "brand": "", "url": "https://www.amazon.com/dp/B004DGJJX6"}], "timeRecords": {"guess-you-like_transform_func_additionalRecall": 0, "iteminfo_summary_transform_func_summary": 2, "task_guess-you-like_total": 75, "rank_widedeep_transform_func_summary": 1, "rank_widedeep_transform_func_cutOff": 2, "guess-you-like_transform_func_cutOff": 2, "scene_guess-you-like_process": 75, "experiment.recall.swing_transform_func_updateField": 1, "experiment.rank.widedeep_transform_func_updateField": 2, "guess-you-like_transform_func_summaryBySchema": 0, "experiment.recall.swing_transform_func_cutOff": 1, "experiment.rank.widedeep_transform_func_cutOff": 2, "iteminfo_summary_transform_func_cutOff": 0, "recall_swing_transform_func_cutOff": 2, "guess-you-like_transform_func_addItemInfo"
```

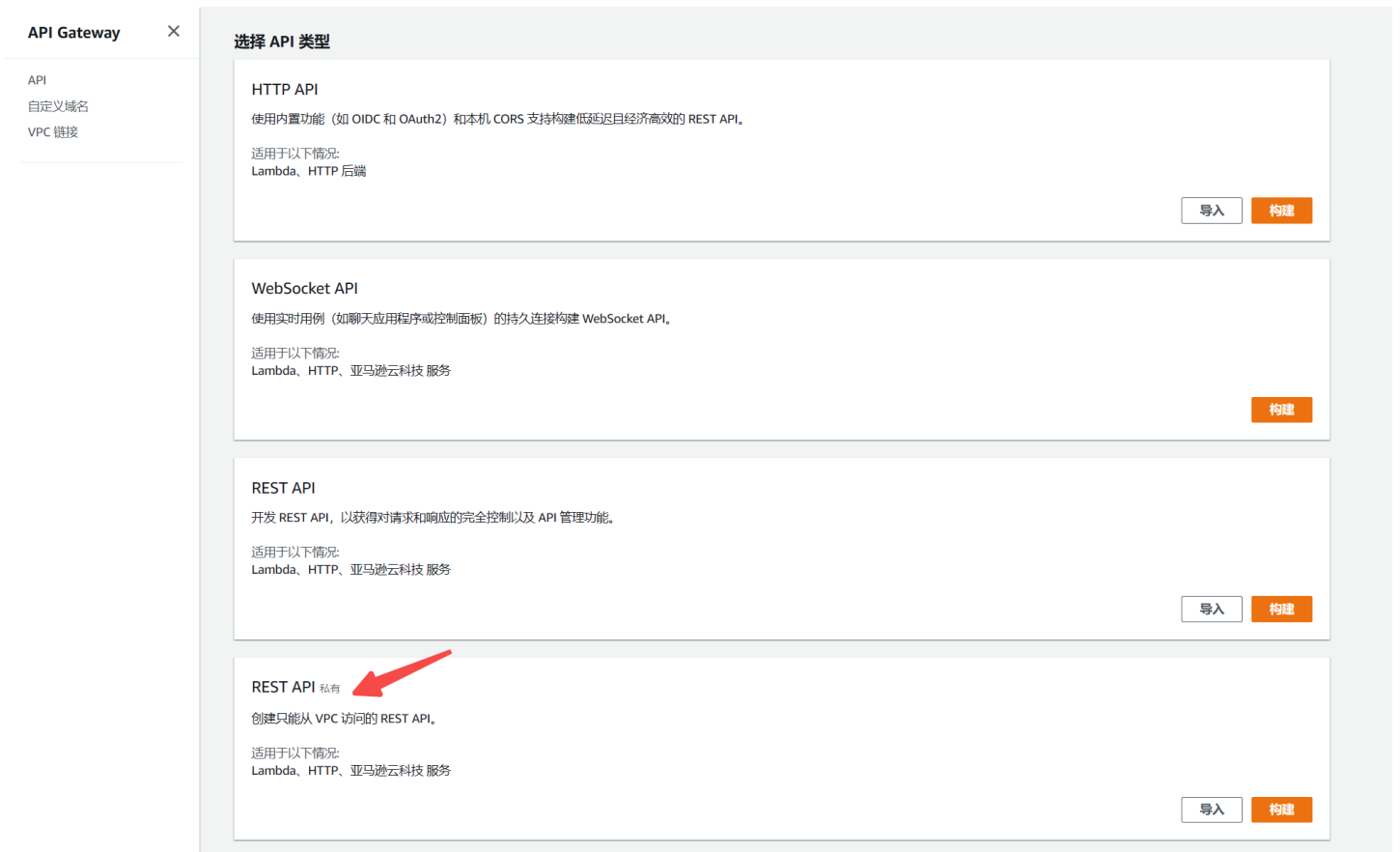
6.2.3 创建内网API Gateway

创建VPC Endpoint

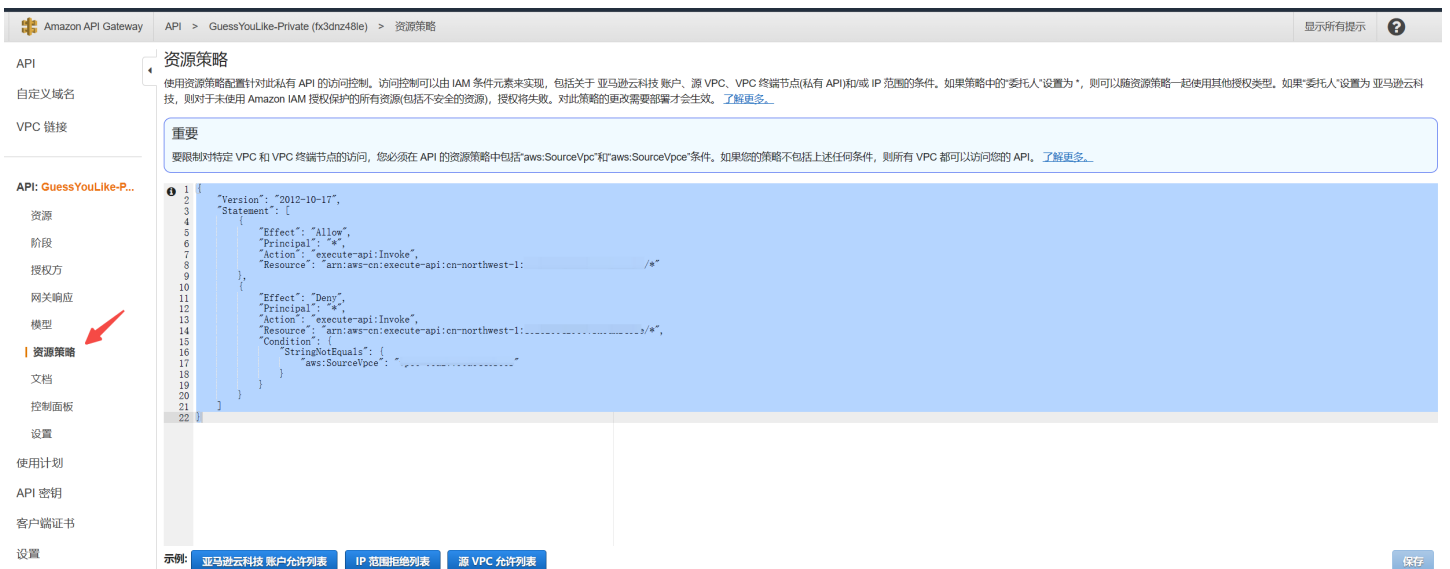
1. 进入VPC控制面板，左侧选择“Endpoints”，点击右侧“创建终端节点”。
2. 填写如下信息：
 - a. 填写Name；
 - b. 选择服务：cn.com.amazonaws.cn-northwest-1.execute-api；
 - c. 选择内网VPC ID；
 - d. 勾选子网Subnet，至少勾选一个；
 - e. 勾选安全组（勾选的安全组需包含443端口的入网规则）；
 - f. 点击“创建终端节点”，创建完成记录vpce-id（创建私有API Gateway需要填入）。

创建私有API Gateway

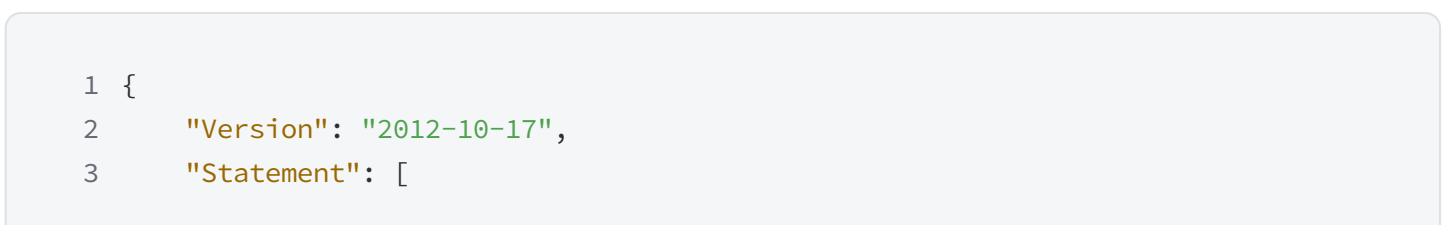
1. 进入API Gateway控制台页面，选择“创建API”。
2. 选择REST API(私网)进行构建。



3. 选择协议，填入**API名称***，默认选择终端节点类型为私有，填入VPC终端节点ID。
4. 接下来同创建公网API Gateway第4，5点一致（创建API，创建POST方法，填入SageMaker Endpoint路径和IAM role，测试连通性，部署API）。
5. 要想进行内网访问，还需完成API Gateway的**资源策略配置**：



修改如下策略中的vpce-id，贴入资源策略编辑器，并点击**保存**：



```
4     {
5         "Effect": "Allow",
6         "Principal": "*",
7         "Action": "execute-api:Invoke",
8         "Resource": [
9             "execute-api:/*"
10        ]
11    },
12    {
13        "Effect": "Deny",
14        "Principal": "*",
15        "Action": "execute-api:Invoke",
16        "Resource": [
17            "execute-api:/*"
18        ],
19        "Condition" : {
20            "StringNotEquals": {
21                "aws:SourceVpce": "vpce-id"
22            }
23        }
24    }
25 ]
26 }
```

6. 通过SSH登录访问VPC内网的服务器，即可进行在线服务调用测试：

```
1 curl -H "Content-Type: application/json" -X POST -d '{"operator": "recommend",
  "request": {"user_id": "A1P62PK6QVH8LV", "scene": "guess-you-like"}}' "调用URL"
```

7. 版本升级/降级

Dmetasoul会持续进行版本更新和问题修复，用户可通过SSH（用户名为ubuntu）登录服务器，执行如下命令即可按照引导进行升级/降级操作。

```
1 sudo metasporeflow_check_version
```

```
ubuntu@ip-172-31-39-52:~$ sudo metasporeflow_check_version
Reading package lists... Done
Building dependency tree
Reading state information... Done
jq is already the newest version (1.6-1ubuntu0.20.04.1).
0 upgraded, 0 newly installed, 0 to remove and 4 not upgraded.
```

检测MetaSporeFlow组件版本 ...

version	metasporeflow	metasporecli	metasporeflow-dashoboard	db-schema	lifecycle
v1.0.0	0.2.0	0.2.0	0.0.1.dev0	9f89f2f904af	release
v1.0.1 ← 已部署	0.2.0	0.2.0	0.0.2	b3fae96aac0d	release

metasporeflow部署版本和最新release版本一致，无需升级。

1. 回滚到上一个版本；
 2. 退出；
- 请选择操作：|