

推荐系统 RES

# 用户指南

文档版本 03  
发布日期 2024-09-07



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

# 目录

<b>1 准备工作</b>	<b>1</b>
1.1 注册华为帐号并开通华为云	1
1.2 为帐号充值	1
1.3 进行服务授权	1
1.4 购买套餐包	2
<b>2 RES 操作流程</b>	<b>4</b>
<b>3 数据源管理</b>	<b>6</b>
3.1 数据源管理简介	6
3.2 准备离线数据源	7
3.3 上传离线数据源至 OBS	12
3.4 上传实时数据	13
3.5 创建离线数据源	14
3.6 导入近线数据源	15
3.7 数据质量管理	17
3.7.1 数据结构	17
3.7.2 数据导入	19
3.7.3 数据探索	21
3.8 修改或删除数据源	23
<b>4 智能场景</b>	<b>25</b>
4.1 智能场景简介	25
4.2 创建智能场景	25
4.3 发布或终止智能场景	27
4.4 获取推荐结果	28
4.5 效果评估	31
4.6 编辑或删除智能场景	32
<b>5 自定义场景</b>	<b>34</b>
5.1 自定义场景简介	34
5.2 创建自定义场景	35
5.3 算法介绍及参数说明	46
5.3.1 召回策略	46
5.3.2 过滤规则	57
5.3.3 排序策略-离线特征工程	59

5.3.4 排序策略-离线排序模型.....	61
5.3.5 在线服务.....	68
5.3.6 效果评估.....	72
5.4 修改自定义场景.....	73
5.5 发布或终止自定义场景.....	74
5.6 获取推荐结果.....	75
5.7 删除自定义场景.....	80
<b>6 工作空间.....</b>	<b>82</b>
6.1 工作空间简介.....	82
6.2 创建工作空间.....	82
6.3 (可选) 创建企业项目.....	84
6.4 编辑或删除工作空间.....	85
<b>7 权限管理.....</b>	<b>86</b>
7.1 创建用户并授权使用 RES.....	86
7.2 RES 自定义策略.....	87
<b>8 用户指南 (旧版) .....</b>	<b>89</b>
8.1 准备工作.....	89
8.1.1 注册华为帐号并开通华为云.....	89
8.1.2 获取访问密钥并添加 RES 全局配置.....	89
8.1.3 创建 OBS 桶.....	91
8.1.4 开通相关资源.....	92
8.2 RES 操作流程.....	93
8.3 管理资源.....	94
8.3.1 创建资源.....	94
8.3.2 绑定或解绑资源.....	95
8.3.3 创建跨源连接.....	96
8.3.4 开启公共终端节点.....	97
8.4 全局配置.....	97
8.4.1 全局配置简介.....	97
8.4.2 管理属性配置.....	98
8.4.3 管理计算资源.....	99
8.5 离线作业.....	100
8.5.1 离线作业简介.....	100
8.5.2 组合作业.....	101
8.5.3 数据质量.....	106
8.5.4 特征工程.....	107
8.5.5 召回策略.....	112
8.5.6 排序策略.....	129
8.5.7 过滤规则.....	141
8.5.8 效果评估.....	144
8.5.9 管理离线作业.....	146

---

8.5.10 删除离线作业.....	148
8.6 近线作业.....	149
8.7 在线服务.....	159
8.7.1 创建在线服务.....	159
8.7.2 获取推荐结果.....	163
8.7.3 查看在线服务详情.....	164
8.7.4 管理在线服务.....	165
8.8 服务总览信息.....	166
8.9 数据格式规范.....	167
8.9.1 推荐系统 OBS 文件夹规范.....	167
8.9.2 离线数据源.....	167
8.9.3 实时日志.....	171
8.9.4 全局特征信息文件.....	175

# 1 准备工作

## 1.1 注册华为帐号并开通华为云

在使用华为云服务之前您需要申请华为云帐号。通过此帐号，您可以使用所有华为云服务，并且只需为您所使用的服务付费。

### 操作步骤

1. 进入[华为云](#)首页，单击页面右上角的“注册”。
2. 设置手机号、短信验证码、帐号名、密码并勾选“我已阅读并同意《[华为云用户协议](#)》和《[隐私政策声明](#)》”，单击“同意协议并注册”。
3. 页面提示注册成功后，系统会自动跳转至您的个人信息界面。
4. 参考[实名认证](#)完成个人或企业帐号实名认证。

## 1.2 为帐号充值

当您使用RES时，建议您先为您的帐号充值，确保帐号有足够余额可以正常使用RES，具体操作请参见[如何给华为云帐户充值](#)。如果您帐号里有足够的余额，可略过此部分内容。

## 1.3 进行服务授权

当您选择开通推荐系统，首次登录RES管理控制台系统会自动弹出“RES服务权限委托”的对话框，提示服务进行委托授权，即授权RES服务请求获取访问您在其他云产品中的资源，未授权将不能使用RES的完整功能。

图 1-1 权限委托

## RES服务权限委托



您还没有创建默认的权限委托，RES服务的运行需要这些权限，请您进行授权。

OBS（对象存储服务）权限：RES执行作业需要读写OBS数据

SMN（消息通知服务）权限：RES需要通过SMN发送通知消息

同意授权后，RES将在统一身份认证服务为您创建名为res\_admin\_agency的委托，授权成功后，可以进入[服务委托列表](#)查看。

同意授权

取消

单击“同意授权”系统会自动创建委托。由于RES与其他云服务之间存在业务交互关系，需要与这些云服务协同工作，因此需要您创建云服务委托，完成授权后将操作权限委托给RES，让RES以您的身份使用这些云服务，代替您进行一些任务调度、资源运维等工作。下次再进入RES管理控制台首页时，系统不会再弹出访问授权的对话框。

## 1.4 购买套餐包

### 套餐包说明

RES服务支持按需和购买套餐包，根据用户选择使用的资源进行收费。一个完整的推荐场景需要下面三种资源，均为必选项。套餐的数量可以根据实际业务按需购买。

- **计算资源**：用于推荐作业的计算规格。涉及计费功能包含：数据源、自定义场景、智能场景中的离线计算和模型训练。
- **存储资源**：用于推荐系统数据存储规格。涉及计费功能包括：数据源。
- **在线服务**：用于推荐系统在线推理，获得最终推荐结果。

### 套餐介绍

计算资源分为“计算型CPU(1U4G)实例”、“计算型GPU(P100)实例”、“计算型GPU(V100)实例”3种类型。存储资源支持“画像存储（一百万）”。在线服务支持“在线并发9000TPS-时”，“TPS-时”表示每小时平均TPS。

规格和支持功能请参见[表1-1](#)。

表 1-1 计算资源

计算类型	资源名称	购买时长	支持功能	有效期
计算资源	“计算型 CPU(1U4G)实例”	● 20000 小时	离线作业 在线作业	1年
	“计算型 GPU(P100)实例”	● 300小时		
	“计算型 GPU(V100)实例”	● 300小时		
存储资源	“画像存储（一百万）”	● 9000小时	画像存储	
在线服务	“在线并发 9000TPS-时”	-	在线服务调用API	-

## 购买套餐包

1. 登录RES管理控制台，在“总览”页面，单击“购买套餐包”，进入“购买预付套餐包”页面。
2. 在“购买预付套餐包”页面，您可以通过“计算类型”下的“计算资源”和“存储资源”和“在线服务”选择不同类型的套餐包。
3. 根据业务需求选择套餐包和套餐包数量，单击“下一步”，进入规格确认。
4. 规格确认无误后，单击“去支付”，然后在支付页面完成付款，付款成功后即完成套餐包的购买。

### 说明

- RES目前只支持华北-北京四区域。
- 计费时将优先使用套餐包的额度，超出额度部分将以当月累计使用量所在的阶梯价计费。套餐包的额度为购买日起一年内的可用资源数。套餐包的有效时长为一年。
- 已购买套餐包不支持退订，购买前请确认。



# 2 RES 操作流程

## 操作流程

本章节介绍使用RES，从资源准备到在线服务完成推荐的全流程。RES流程图如图2-1所示。

图 2-1 RES 操作流程

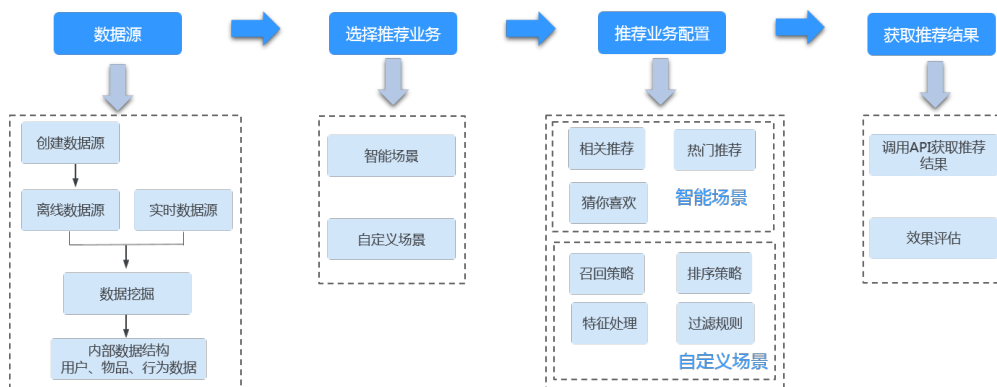


表 2-1 使用流程说明

流程	子任务	说明	详细指导
数据源	准备离线数据源	需要您准备包含用户数据，物品数据，行为数据上传至对象存储服务（OBS）用于推荐系统的离线计算。	<a href="#">准备离线数据源</a> <a href="#">上传离线数据源至 OBS</a>
	创建离线数据源	在使用RES之前，首先您需要创建一个数据源，后续的操作都是基于您创建的数据源进行的。	<a href="#">创建离线数据源</a>
	上传实时数据（可选）	RES通过SDK上传实时数据，进行数据计算和处理，更新用户的相关数据。	<a href="#">上传实时数据</a>

流程	子任务	说明	详细指导
	数据质量管理	数据质量管理操作可以将数据源经过数据特征抽取，生成推荐系统内部通用的数据格式。	<a href="#">数据质量管理</a>
选择并配置推荐业务	智能场景	针对对应的场景，由RES根据场景类型预置好对应的智能算法，为匹配的场景提供智能推荐服务。	<a href="#">智能场景简介</a>
	自定义场景	面向了解推荐算法等相关的用户，可自定义推荐中涉及算法的使用和组合，能够自定义开发推荐流程，提供推荐服务。	<a href="#">自定义场景简介</a>
获取推荐结果	-	您可以通过管理控制台进行结果预测，也可以通过API接口获取最终的推荐结果。	<ul style="list-style-type: none"> <li>● <a href="#">智能场景获取推荐结果</a></li> <li>● <a href="#">自定义场景获取推荐结果</a></li> </ul>
效果评估（可选）	-	创建效果评估可以对服务设置指标，查看推荐结果的具体衡量指标。	<ul style="list-style-type: none"> <li>● <a href="#">智能场景效果评估</a></li> <li>● <a href="#">自定义场景效果评估</a></li> </ul>

# 3 数据源管理

## 3.1 数据源管理简介

RES以数据为基础进行算法计算并完成推荐，您可以在RES管理控制台，数据源页面完成数据创建、数据修改等操作，为智能场景推荐和自定义场景推荐做好数据准备。

### 数据类型

当前RES支持创建数据源和导入近线数据。创建数据源的数据格式和近线数据导入的格式要求一致，包括用户数据、物品数据和行为数据。

- 用户数据  
用户数据包括数据源中的“用户属性表”和用于近线计算的“用户画像”数据。用户数据记录用户的属性信息，例如地域、爱好等。
- 物品数据  
物品数据包括数据源中的“物品属性表”和用于近线计算的“物品画像”数据。物品数据记录物品的属性信息，例如类别、长度等。
- 行为数据  
行为数据包括数据源中的“用户操作行为表”和用于近线计算的“行为数据”。行为数据，每行记录用户的单次行为信息，包含用户标识符、行为对象标识符、行为类型和行为时间等信息。

### 注意事项

- 每张表的表结构和填充的数据必须符合推荐系统的要求，列名和字段类型需要和规范保持一致。
- 创建数据源之前您需要自己手工创建整理这些表并存储到OBS，推荐数据目前支持JSON格式。
- 离线数据为初次对接RES的批量数据，供场景中的召回算法和排序算法进行计算。近线数据可实时更新、增加用户和物品表数据，同时实时行为数据，可作为后续近线召回计算，这些近线行为数据也会和离线行为数据进行汇总存储，供离线计算。

## 数据源管理流程及功能简介

表 3-1 功能介绍

功能	说明
<a href="#">准备离线数据源</a>	请您按照推荐系统要求的数据格式准备用户数据，物品数据，行为数据。
<a href="#">上传离线数据源至OBS</a>	将准备的离线数据源上传至对象存储服务（OBS）用于推荐系统的离线计算。
<a href="#">创建离线数据源</a>	在使用RES之前，首先您需要创建一个数据源，后续的操作都是基于您创建的数据源进行的。
<a href="#">上传实时数据</a>	RES通过SDK上传实时数据，进行数据计算和处理，更新用户的相关数据。
<a href="#">数据质量管理</a>	数据质量管理操作可以将数据源经过数据特征抽取，生成推荐系统内部通用的数据格式。
<a href="#">修改或删除数据源</a>	对您已经创建的数据源进行修改，对不需要的数据源进行删除。

## 3.2 准备离线数据源

在使用RES创建数据源时，您需要准备以下的3种基础数据包并上传至OBS。如果使用近线流程，需先将业务系统埋点日志转换成推荐系统指定格式，并实时写入DIS相应通道。本章节介绍了RES当前离线数据源和近线数据源的数据格式，您可以参考本章节说明，准备相应的数据。

### 📖 说明

目前数据源只支持JSON格式，且单条数据长度不超过3000个字节。

- [用户数据](#)
- [物品数据](#)
- [行为数据](#)

### 用户数据

表 3-2 用户数据字段描述

字段名	类型	描述	是否必选
userId	String	全局唯一用户ID。可输入的字符范围为1~64个字符。只能包含字母、数字、中划线、下划线和特殊字符（:./#）并且不能以空格开始和结束。	是
age	Long	年龄。	否

字段名	类型	描述	是否必选
gender	String	性别。取值为： <ul style="list-style-type: none"> <li>• male</li> <li>• female</li> <li>• unknown</li> </ul>	否
location	Json	根据业务而设定的用户关注的地理位置信息。如常驻地，精度维度格式：{"lat": latitude, "lon": longitude}。例如： "location": {"lat": 41.12, "lon": -71.34}。	否
tags	List[String]	描述用户的标签，每个标签为独立的一个元素。如[“篮球”，“家庭”]。	否
interest_tags	List[String]	兴趣标签，由系统自动更新，默认保留字段。	否
extend	Json	扩展字段，由用户指定扩展字段的key和value。其中： <ul style="list-style-type: none"> <li>• key，您可以按照业务需求自行命名。无需按照示例以“extend_”开头，key不能为中文。</li> <li>• value，支持long/float/String/strArray/location格式的数据。</li> </ul>	否

## 物品数据

表 3-3 物品数据字段描述

字段名	类型	描述	是否必选
itemId	String	全局唯一物品ID。	是
itemType	String	物品的类型，可用于对推荐结果集的多样性控制。包含： <ul style="list-style-type: none"> <li>• item</li> <li>• article</li> <li>• video</li> <li>• audio</li> <li>• image</li> </ul>	是
category	String	物品的类别，如“课程”、“零食”，可用于类别之间的打散。	否

字段名	类型	描述	是否必选
status	Long	物品是否可推荐、是否置顶、是否注销，默认值为0。 <ul style="list-style-type: none"> <li>0: 可推荐</li> <li>1: 不可推荐</li> <li>2: 置顶</li> <li>-1: 注销</li> </ul>	否
title	String	标题，可用于语义分析。	否
score	Long	物品的权重，权重越高，被优先推荐的概率越高，取值范围：1-100。如不设置，默认值为1。	否
publishTime	Long	发布时间，采用UTC标准时间，单位以秒计。对物品有实时性要求的则必填。	否
expireTime	Long	失效时间，采用UTC标准时间，单位以秒计。当前服务器的时间大于该时间时，此物品将不会被推荐。如不设置，代表永不失效。	否
author	List[String]	作者，一个作者一个元素，信息流推荐场景建议填写。	否
content	String	物品的内容描述信息，例如简介/摘要/正文关键片段，最长支持5000个中英文字符，超过则截断处理。	否
tags	List[String]	描述物品的标签，每个标签为独立的一个元素。	否
location	Json	物品的地理位置信息，精度维度格式：{"lat": latitude, "lon": longitude }。例如："location": {"lat": 41.12, "lon": -71.34}。	否
extend	Json	扩展字段，由用户指定扩展字段的key和value。 其中： <ul style="list-style-type: none"> <li>key，您可以按照业务需求自行命名。无需按照示例以“extend_”开头，key不能为中文。</li> <li>value，支持long/float/String/strArray/location格式的数据。</li> </ul>	否

## 行为数据

表 3-4 行为数据字段描述

字段名	类型	描述	是否必选
userId	String	用户ID。	是
itemType	String	物品的类型。	是
itemId	String	对应行为发生的对象的值。如果是和物品发生关系，则是物品的id(itemId)的值。	是

字段名	类型	描述	是否必选
actionType	String	<p>行为类型，包括正向行为和负向行为。下面为预置的行为类型和对应的权重，权重有默认分数，默认分数代表该行为类型的权重，可做修改。同时支持自定义行为类型。</p> <ul style="list-style-type: none"> <li>如果当前场景只有用户关注人这一种行为，则此参数表示用户关注的对象是人。</li> <li>如果当前场景只有用户关注物品这一种行为，则此参数表示用户关注的对象是物品。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>自定义行为类型，行为名称需要符合以下要求，由数字、字母、下划线组成，长度为4-32。</li> <li>自定义行为类型接入系统后，初始默认为正向行为，分数为0.5，后续可通过人工复核进行修改。</li> </ul> <ul style="list-style-type: none"> <li>正向行为： <ul style="list-style-type: none"> <li>click：点击，默认分数0.3</li> <li>collect：收藏，默认分数0.5</li> <li>search_click：点击搜索，默认分数0.4</li> <li>comment：评论，默认分数0.2</li> <li>share：分享，默认分数0.5</li> <li>like：点赞，默认分数0.8</li> <li>consume：消费，默认分数1</li> <li>use：观看视频/听音乐/阅读，默认分数0.6</li> <li>download：下载，默认分数0.6</li> <li>tip：打赏，默认分数1</li> <li>subscribe：关注，默认分数0.7</li> </ul> </li> <li>负向行为： <ul style="list-style-type: none"> <li>view：曝光，默认分数0.1</li> <li>dislike：点衰/踩，默认分数0.8</li> <li>uncollect：取消收藏，默认分数0.4</li> </ul> </li> </ul>	是
actionMeasure	Int	<p>本次行为的度量，取值为正整数1-10。建议用户根据实际情况进行换算，如总视频15秒，观看视频duration超过10秒的7分，全看完得10分等。</p>	是
actionTime	Long	<p>行为发生的时间，采用UTC标准时间，单位以秒计。</p>	是



字段名	类型	描述	是否必选
actionLocation	Json	行为发生的位置，精度维度格式：{"lat": latitude, "lon": longitude}。例如： "location": {"lat": 41.12, "lon": -71.34}。	否
sceneId	String	推荐场景 ID，用于场景维度的区分。	否
traceId	String	用于追踪每个被推荐物品的唯一ID。用于推荐效果的计算。	否
flowId	String	用于计算每一个在线服务的效果。flowId由推荐系统的API返回给用户，用户需把flowId写到用户行为日志中。	否

## 数据示例

单击[下载样例数据](#)，将样例数据下载至本地

## 3.3 上传离线数据源至 OBS

RES使用对象存储服务（Object Storage Service，简称OBS）进行数据源的存储。因此，在使用RES之前您需要开通OBS服务并创建桶，然后在OBS桶中上传用户数据用于推荐作业的计算。

需要存放在OBS桶中的数据包括：

1. 离线数据源：包含用户类数据，物品类数据，行为数据以及推荐候选列表。
2. 人工推荐策略的候选集（可选）：您可以将人工编辑推荐结果的列表csv文件（即物品ID）存放在OBS桶中。

### 说明

创建桶的区域需要与RES所在的区域一致。RES目前只支持华北-北京四区域。例如：当前RES在华北-北京四区域，在对象存储服务创建桶时，请选择华北-北京四。

## 前提条件

- 请您按推荐系统要求[准备离线数据源](#)。
- 开通OBS服务，确保您使用的OBS目录与RES在同一区域。
- 已经获取访问密钥（AK/SK）。

## 通过 OBS 管理上传数据

本文中，采用管理控制台上传数据至OBS。

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在RES同一区域内创建桶。如果已存在可用的桶，需确保OBS桶与RES在同一区域。
2. 将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。

数据上传成功后，在OBS管理控制台页面单击进入创建的桶，选择“对象”查看上传的数据。

## 3.4 上传实时数据

RES通过DIS SDK上传实时数据，用户实时日数据并做近线处理。当前仅支持Java语言的SDK，示例请参见《[数据接入服务SDK参考](#)》。

### 前提条件

- 如果需要使用近线上传实时数据的用户，可以使用DIS SDK接口上传，请您按照需求[下载DIS SDK](#)，下载完之后按照下面的说明进行SDK升级。
- 子账户无法使用SDK上传数据，需要主账号授权子账号DIS USER权限。详细请参考[权限管理](#)。

#### 📖 说明

下载SDK之后，需要修改pom文件中的依赖，对SDK进行升级，最新版本可升级至1.3.12。

### 上传实时数据至 RES

1. 初始化DIS客户端，使用代码初始化DIS SDK客户端实例，代码样例如下。具体方式请参见[初始化DIS客户端](#)。

```
// 创建DIS客户端实例
DIS dic = DISClientBuilder.standard()
    .withEndpoint("YOUR_ENDPOINT")
    .withAk("YOUR_AK")
    .withSk("YOUR_SK")
    .withProjectId("YOUR_PROJECT_ID")
    .withRegion("YOUR_REGION")
    .build();
```

其中，各参数说明如下：

- “YOUR\_AK”、“YOUR\_SK”即访问密钥，获取方式请参见[获取访问密钥](#)。
  - “YOUR\_PROJECT\_ID”为项目ID、“YOUR\_REGION”为区域ID，获取方式请参见[获取项目名称、项目ID、区域ID](#)。
2. 获取需要上传通道的ID（streamId）。
    - 单击近线数据源的“详情”

图 3-1 获取通道 ID

#### 物品画像实时导入详情

名称 6052aa

通道ID mr

关闭

3. 上传实时数据，示例代码如下，其中，“streamId”的配置值要与步骤2中“通道ID”的值一致。

```
// 配置通道ID
String streamId = "xxxx";
// 配置上传的数据

PutRecordsRequest putRecordsRequest = new PutRecordsRequest();
putRecordsRequest.setStreamId(streamId);
List<PutRecordsRequestEntry> putRecordsRequestEntryList = new ArrayList<>();

String path = ActionDataProducer.class.getClassLoader().getResource("action.json").getPath();
BufferedReader in = new BufferedReader(new FileReader(path));
String record = in.readLine();
int putCnt = 0;
while (record != null && !record.isEmpty()) {
    putCnt++;
    System.out.println("Put the " + putCnt + " record: " + record);
    try {
        putRecordsRequestEntryList.clear();
        PutRecordsRequestEntry putRecordsRequestEntry = new PutRecordsRequestEntry();
        putRecordsRequestEntry.setData(ByteBuffer.wrap(record.getBytes()));
        // PartitionKey为随机值可使数据均匀分布到所有分区中

        putRecordsRequestEntry.setPartitionKey(String.valueOf(ThreadLocalRandom.current().nextInt(1000000)));
        putRecordsRequestEntryList.add(putRecordsRequestEntry);
        putRecordsRequest.setRecords(putRecordsRequestEntryList);
        dic.putRecords(putRecordsRequest);
    } catch (DISClientException e) {
        LOGGER.error("Failed to get a normal response, please check params and retry. Error message [{}]",
            e.getMessage(),
            e);
    } catch (Exception e) {
        LOGGER.error(e.getMessage(), e);
    }
    record = in.readLine();
}
in.close();
```

## 3.5 创建离线数据源

在使用RES之前，首先您需要创建一个数据源，后续的操作，如修改数据源、创建自定义推荐，都是基于您创建的数据源进行的。




### 前提条件

- 已创建用于存储数据的OBS桶及文件夹，并且数据存储的OBS桶与RES在同一区域。
- 需要使用的数据已上传至OBS。

### 创建数据源

1. 登录RES管理控制台，在左侧菜单栏中选择“数据源”，进入“数据源”列表页面。
2. 单击“创建”，在创建数据源页面，参考表3-5配置相关参数。

表 3-5 创建数据源参数说明

参数名称	说明
名称	数据源的名称，名称只能是字母、数字、下划线或者中划线组成的合法字符串。
用户属性表	从OBS桶中选择数据。在“用户属性表”右侧，单击  ，从弹出的对话框中，选择数据存储的OBS桶及其文件或文件夹。 如果选取文件夹，该文件夹下面的数据格式需保持一致，避免数据干扰。
物品属性表	从OBS桶中选择数据。在“物品属性表”右侧，单击  ，从弹出的对话框中，选择数据存储的OBS桶及其文件或文件夹。 如果选取文件夹，该文件夹下面的数据格式需保持一致，避免数据干扰。
用户操作行为表	从OBS桶中选择数据。在“用户操作行为表”右侧，单击  ，从弹出的对话框中，选择数据存储的OBS桶及其文件夹。 <b>说明</b> “用户属性表”、“物品属性表”以及“用户操作行为表”不能为同一路径。
离线规格	可选择的规格有“4CU”、“8CU”、“16CU”、“32CU”、“64CU”、“128CU”。规格越大，计算越快。

3. 单击“立即创建”。

数据源创建完成后，系统自动跳转至数据源列表页面并提示您数据源创建成功。针对创建好的数据源，您可以进行“删除”操作。您可以在数据源列表页面单击您创建的数据源名称进入数据源详情页面。

## 3.6 导入近线数据源

通过导入近线数据源，达到实时计算并更新用户画像、物品画像，实时更新增量数据的目的。




### 前提条件

按数据规范准备数据并上传至通道。具体上传方法请参见[上传实时数据](#)。

### 导入近线数据源

1. 登录RES管理控制台，在左侧菜单栏中选择“数据源”，进入“数据源”列表页面。
2. 在数据源列表单击目标数据源名称，进入数据源详情页面。
3. 在页面的“近线数据源”页签，进行参数配置。您可以根据业务需求选择导入数据类型。详情请参见[表3-6](#)。

表 3-6 近线数据源参数说明

参数名称	说明
实时计算规格	可选择“4CU”、“8CU”、“16CU”和“32CU”四种规格。
用户画像实时导入	通过SDK导入通道中存储的用户画像实时数据。详情参见 <a href="#">上传实时数据</a> 。在“用户画像实时导入”右侧，单击打开  按钮，在弹出的对话框中进行确认。
物品画像实时导入	通过SDK导入通道中存储的物品画像实时数据。在“物品画像实时导入”右侧，单击打开  按钮，在弹出的对话框中进行确认。
行为数据实时导入	通过SDK导入通道中存储的行为数据。在“行为数据实时导入”右侧，单击打开  按钮，在弹出的对话框中进行确认。 <b>说明</b> “用户画像实时导入”、“物品画像实时导入”以及“行为数据实时导入”可以单独进行导入。

- 打开目标数据右侧的导入按钮，确认“实时计算规格”信息和计费后并进行确认后，单击“是”进行实时导入。待数据状态成为“运行中”时，表示实时数据导入完成。每“执行”一次导入操作，之前的数据内容将被覆盖，请您谨慎操作。

图 3-2 实时导入

物品画像实时导入



### 查看近线数据源详情或终止数据源


当您确认导入近线数据源之后，可以查看近线数据源详情或者对数据终止导入。

- 在数据源详情页面，您可以单击目标近线数据右侧的“详情”，查看该实时数据的名称和通道ID，此ID为全局唯一。
- 当不需要该数据时，您单击可以单击目标近线数据右侧的“终止”，对“执行”状态的数据终止导入。

当终止导入后，该数据无法正常使用，请您谨慎操作。

## 删除近线数据源

当您不需要使用已有的近线数据源时，您可以删除导入数据。

1. 在数据源列表单击目标数据源名称，进入数据源详情页面。
2. 在页面的“近线数据源”页签，在目标数据源右侧单击  按钮关闭导入数据源，在弹出的对话框中进行确认。

### 说明

关闭近线数据源导入不会影响存量数据的存储。

## 3.7 数据质量管理

### 3.7.1 数据结构

当数据源创建完成，您可以进入数据源详情页面进行数据质量管理操作。数据质量管理操作可以将离线数据源经过数据特征抽取，生成推荐系统内部通用的数据格式。经过数据质量检测来确保数据的合法性。

#### 数据结构介绍

数据结构步骤的主要目的是读取用户上传的离线数据，解析用户特征和物品特征中每一个属性的数据格式、统计所有行为，然后保存解析生成的数据格式。

#### 前提条件

已按照[创建离线数据源](#)操作指导完成数据源的创建。

#### 操作步骤





1. 登录RES管理控制台，在左侧菜单栏中选择“数据源”，进入“数据源”列表页面。单击目标数据源名称进入数据源详情页面。
2. 在“执行步骤”页签，单击“数据结构”下的“开始识别”，进入“识别”页面，确认页面信息后单击“是”进行特征抽取。
3. 当执行完成“识别”状态为“已完成”时，进行“人工复核”确认操作。
4. 在“特征抽取”页签确认识别出的数据信息。

如果复核抽取的数据信息无误，单击“确认”，在“复核特征抽取”页面单击“是”完成复核。如果复核抽取的数据信息不符合预期，可选择“重新识别”。该页面包含“用户特征”、“物品特征”、“正向行为类型”和“负向行为类型”等信息，具体描述请参见[表3-7](#)。

图 3-3 确认特征抽取

表 3-7 确认特征参数

参数名称	说明
用户特征	列表中展示抽取的用户特征和参数类型。您可以根据业务需求单击 ⊕ 增加用户特征。单击特征后方的 🗑️ 删除不需要的用户特征。
物品特征	列表中展示抽取的物品特征和参数类型，此特征会额外应用于所选字段的功能。您可以根据业务需求单击 ⊕ 增加物品特征。单击特征后方的 🗑️ 删除不需要的物品特征。您可以从“应用于”右侧的下拉选项中设置该数据的使用维度是“兴趣属性”或者“关键词提取”。其中： <ul style="list-style-type: none"> <li>● 兴趣属性，此特征将会用于统计用户的兴趣标签，并生成特征名为“interested_原特征名”的特征。</li> <li>● 关键词提取，只有当关键词为content和title时会进行关键词提取，并生成特征名为“keyword_原特征名”的特征。</li> </ul>

参数名称	说明
正向行为类型	<p>列表中展示抽取的正向行为类型。您可以根据业务需求单击  增加正向行为类型。单击特征后方的  删除不需要的正向行为类型。</p> <p>您可以从“应用于”右侧的下拉选项中设置该数据的使用维度是“历史记录”或者“物品热度值”。其中：</p> <ul style="list-style-type: none"> <li>历史记录：此特征将会用于统计用户此行为的历史记录，并生成特征名为“history_原特征名”的特征。</li> <li>物品热度值：此行为将会被用于统计物品的热度值信息，并生成特征名为“statistics_hotValue”的特征。</li> </ul> <p><b>说明</b> 如果您有自定义行为类型，系统识别后，会默认为正向行为类型，默认分数为0.5。您可以根据业务需求对其进行修改，也可以删除，直接添加为负向行为类型。</p>
负向行为类型	<p>列表中展示抽取的负向行为类型。您可以根据业务需求单击  增加负向行为类型。单击特征后方的  删除不需要的负向行为类型。</p> <p>您可以从“应用于”右侧的下拉选项中设置该数据的使用维度是“历史记录”或者“物品热度值”，其相关说明与正向行为类型一致。</p>

## 3.7.2 数据导入

### 数据导入介绍

数据导入即读取经过“数据结构”生成的数据，对每条数据进行校验。推荐系统保留字段需校验类型和数据合法性、自定义字段校验类型，输出错误报告。如果数据完全符合要求，会生成推荐系统所需要的宽表和画像数据。

- 宽表：推荐系统内部格式，以行为数据为主，将行为数据中涉及到的用户数据和物品数据整合成一条数据。
- 画像：画像分为用户画像和物品画像，分别用于存储用户输入的用户特征和物品特征。如果同一用户或物品有多条记录，将会按照用户ID或者物品ID去重。

### 前提条件

- 已按照[创建离线数据源](#)操作指导完成数据源的创建。
- 已完成数据结构识别和人工复核确认。

### 操作步骤

- 在“执行步骤”页签，单击数据导入下的“执行”。进入“导入”页面，确认页面信息后单击“确认”，允许推荐系统读取您OBS中存储的特征数据，生成画像和宽表。
- 确认完成单击“执行”，待状态为“已完成”时，生成推荐系统内部通用的画像和宽表数据。



3. 执行完成在页面下方会生成数据相关报告。
  - “数据导入报告”，显示数据“类型”、“总条目数”、“合法条目数”、“非法条目数”、“重复度”和“合法率”信息。
  - 类型包括生成的用户、物品、行为数据。您可以通过单击左侧的查看具体报告信息。
    - “名称”项显示具体参数的名称。
    - “条目数”显示各种类型数据的具体数量。

图 3-4 查看报告

类型	总条目数	合法条目数	非法条目数	重复度	合法率
用户	10	10	0	0.00	1
物品	10	10	0	0.00	1
行为	20	20	0	0.00	1

- 如果导入错误，会生成“数据导入错误报告”，显示数据“类型”、“数量”和“原因”，方便您定位问题原因。
4. (可选) 如果用户更新了数据源，可以重新导入数据。
    - a. 单击“数据导入报告”后的“执行”，重新进行数据导入。

图 3-5 重新导入数据

导入

**!** 确定要执行导入数据吗?

导入画像

导入宽表

保留已有宽表  覆盖  是  否

开启调度

计算时按需费用 CPU-作业运行的CU数量 \* ¥0.49 /小时 + 画像存储按需费用 时长 \* 数量 \* ¥0.59 /百万/小时

配置项	规格值	作业数量
离线作业计算规格	4CU	1
画像存储	--	1

表 3-8 参数解释

参数	说明
导入画像	选择是否导入画像。 画像分为用户画像和物品画像，分别用于存储客户输入的用户特征，物品特征，如果同一用户或物品有多条记录，将会去重。

参数	说明
导入宽表	选择是否导入宽表。 宽表为推荐系统内部格式，以行为数据为主，将行为数据中涉及到的用户数据和物品数据整合成一条数据。
保留已有宽表	选择对已有宽表是否保留。目前支持覆盖、是、否。
开启调度	开启调度，可按照您所设置的时间周期性的提交任务。 <ul style="list-style-type: none"><li>“调度周期”：调度周期可选“天”或“周”。</li><li>“调度类型”：包括自定义和间隔调度。</li><li>“开始调度时间”：选择具体的调度时间。可在此下拉框中勾选具体的时间点。</li><li>“时间间隔（小时）”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li></ul>

- b. 设置完成后，单击“确定”，重新导入数据。

### 3.7.3 数据探索

#### 数据探索介绍

数据探索是针对当前数据源的数据进行挖掘和分析，主要聚焦在特征的分布范围、统计以及特征齐全度等，使用户能够更了解数据，进而指导在特征工程以及相关算法的配置。

数据探索是一个离线分析任务，任务有对应的启动时间，由于增量数据会实时入库，因此可以通过定时执行数据探索任务来覆盖增量数据。

#### 操作步骤

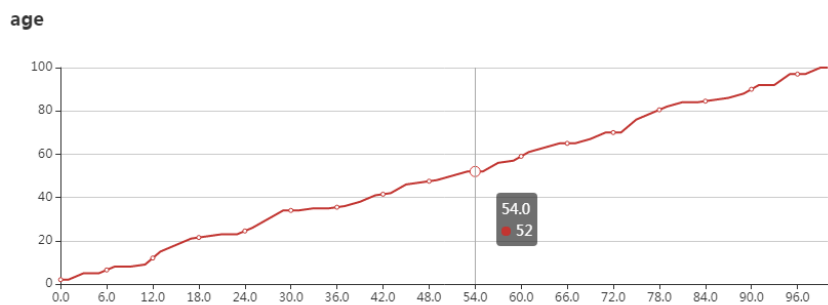
1. 在“执行步骤”页签，单击数据探索下的“执行”。进入“数据探索”页面，确认页面信息后单击“是”。待状态为“已完成”时，完成数据格式的转化。
2. 执行完成在页面下方会显示数据探索报告，包括“用户报表”、“物品报表”、“行为报表”和“画像查询”。
3. 单击目标报表名称查看具体报表信息。

图 3-6 查看报表



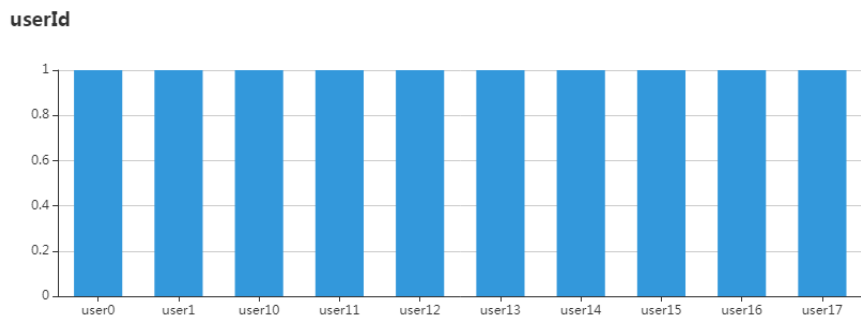
- 用户报表：根据不同数据格式展示用户数据的类型、最大值和最小值。您可以单击相关数据后的 [查看](#) 查看数据的详细信息。
  - 百分位数：将数据进行排序，统计该数据在整个数据中所占的百分比。

图 3-7 百分位数



- 分布统计：通过查看分布统计了解各参数下参数值的分布情况。如可以根据性别展示数据中的性别数据分布。可通过查看标签，了解数据中各种标签的分布情况。

图 3-8 分布统计



- 物品报表：根据不同数据格式展示物品数据的类型、最大值和最小值。您可以单击相关数据后的 [查看](#) 查看数据的详细信息。

- 行为报表：行为报表展示各种行为类型以及该数据中此行为出现的次数。
- 画像查询：可以查询指定的用户或物品画像信息，包括静态和动态。

## 3.8 修改或删除数据源

您可以对离线数据源进行重新编辑操作来更新数据源，如果该数据源不再使用，您可以删除数据源释放资源。

### 修改离线数据源

#### 前提条件

- 已存在的离线数据源有修改或者更新并已经上传至OBS。
- 只有在数据源数据结构特征抽取人工复核确认之前才允许修改数据源。

#### 注意事项

修改编辑之后的离线数据源需要重新进行数据结构抽取和检测、探索等操作。

#### 操作步骤


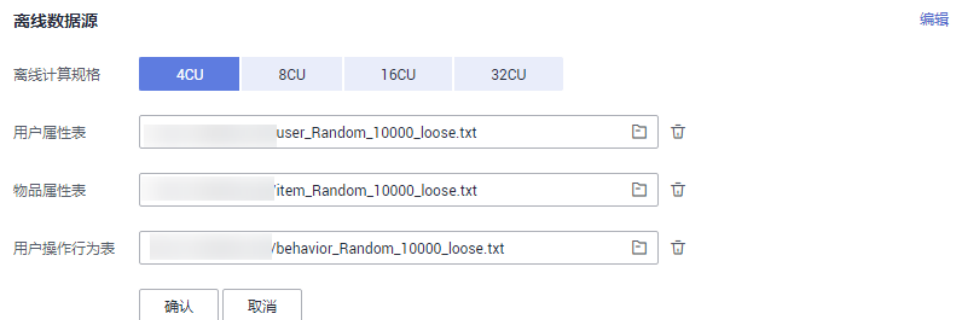
1. 登录RES管理控制台，在左侧菜单栏中选择“数据源”，进入数据源列表页面。
2. 在数据源列表单击目标数据源名称，进入数据源详情页面。
3. 单击离线数据源右侧的“编辑”，在页面下方单击  选择目标文件在OBS中的存储位置。

图 3-9 修改离线数据源



4. 单击“确认”，完成修改。

### 删除数据源

#### 前提条件

已存在创建成功的数据源。

#### 注意事项

- 数据源中的数据结构、数据导入或数据探索有任务在运行，该数据源不支持删除。
- 被智能场景或自定义场景使用的数据源不支持删除。

#### 操作步骤

1. 登录RES管理控制台，在左侧菜单栏中选择“数据源”，进入数据源列表页面。
2. 单击目标数据源“操作”的“删除”，删除离线数据源。

#### 说明

- 删除数据源操作不可恢复，请您谨慎操作。
- 删除数据源不会删除OBS中存储的数据。

# 4 智能场景

## 4.1 智能场景简介

针对对应的场景，由RES根据场景类型预置好对应的智能算法，为匹配的场景提供智能推荐服务。

### 智能场景功能说明

表 4-1 功能说明

功能	说明	详细指导
猜你喜欢	推荐系统结合用户实时行为，推送更具针对性的内容，实现“千人千面”。	<a href="#">创建智能场景</a>
关联推荐	基于大规模机器学习算法，深度挖掘物品之间的联系，自动匹配精准内容。	
热门推荐	基于多维度数据分析，自动匹配所覆盖用户群体更关心的内容进行重点展示。	
获取推荐结果	根据不同的功能模块，获取对应的推荐结果。	<a href="#">获取推荐结果</a>

## 4.2 创建智能场景

猜你喜欢主要应用于浏览意向不明确，如首页推荐等，RES能够根据用户的长短期行为表现出来的兴趣进行学习训练，结合长短期兴趣进行个性化推荐。

关联推荐主要应用于固定的物品的关联推荐，根据已关联的物品对相关的内容和行为进行挖掘，网状匹配相关联的物品，进行有关联度的推荐。

热门推荐主要应用于当前用户浏览最多的物品内容，如实时搜索量前几的新闻或者物品。

## 前提条件

- 已经存在创建成功并完成数据探索的数据源。
- 由于作业运行需消耗资源，确保账户未欠费。

## 创建智能场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务>智能场景”，默认进入“智能场景”列表。
2. 在智能场景列表中，单击右上角“创建”，进入“创建智能场景”页面。
3. 在创建智能场景页面的“模板选择”页签，单击“猜你喜欢”、“关联推荐”或“热门推荐”进行参数配置。填写“基本信息”，选择“场景规格”，配置“个性化配置”相关参数，详情请参见表4-2。

图 4-1 智能场景参数配置

表 4-2 智能场景参数说明

参数名称	子参数	说明
基本信息	名称	自定义的名称。只能包含数字、大小写字母、下划线和中划线，长度不能超过64位且不能为空。 名称创建之后不支持修改。
	数据源	从右侧下拉框中选择RES系统中已有的数据源。当无可用数据源时，此下拉框为空。
	描述	对于该场景的描述信息。
场景规格	-	选择离线计算、实时计算、排序模型训练规格和在线并发数。

参数名称	子参数	说明
个性化配置	匹配特征对	<p>匹配用户和物品特征，以便于筛选出该用户相关联的物品进行推荐。</p> <ul style="list-style-type: none"> <li>● 用户特征名：从下拉框中选择目标用户特征用于和物品特征进行匹配。</li> <li>● 物品特征名：从下拉框中选择目标物品特征用于匹配用户特征，更好的做出推荐。</li> <li>● 权重：取值为0.01-1。权重越高，该匹配特征所被优先推荐的概率越高。</li> <li>● 匹配个数度量：默认不勾选。当勾选之后表示满足的匹配程度。</li> <li>● 操作：选择删除某个匹配特征对。</li> </ul>
	行为过滤	<p>指定时间区间内用户对物品有过指定的行为类型记录时，将过滤掉相应物品。</p> <p>其中排除行为过滤是指将有此行为类型的物品排除掉，不对其进行过滤。</p> <p>例如对于用户过去3天内有过曝光行为的物品过滤，但是可以对曝光过但是有购买行为的物品进行排除，不对其进行过滤。</p>
	开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>● “调度周期”：调度周期可选“天”或“周”。</li> <li>● “调度类型”：包括自定义和间隔调度。</li> <li>● “开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>● “时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

4. 配置完成后单击“创建”。创建成功后会自动跳转到“智能场景”列表页面并提示您创建成功。针对创建好的智能场景，您可以单击操作列的“发布”。待场景状态为“运行中”时，您可以在智能场景列表页面单击目标场景名称进入详情页面进行“预测”和“效果评估”操作。

## 4.3 发布或终止智能场景

针对创建成功的智能场景，需要进行发布才可以运行作业获取推荐结果。通过执行终止操作，停止运行当前场景。

### 前提条件

已存在创建成功的智能场景。



## 发布智能场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务 > 智能场景”，进入智能场景列表页面。
2. 选择智能场景列表中的目标场景，单击“操作”列的“发布”。
3. 在“发布场景”页面确认智能场景作业运行的配置规格单击“是”。

### 📖 说明

- 修改智能场景信息需要执行发布操作才可以同步数据。
- 发布场景会覆盖之前数据，请谨慎操作。

## 终止智能场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务 > 智能场景”，进入智能场景列表页面。
2. 选择自定义场景列表中的目标场景，单击“操作”列的“终止”。
3. 在“终止场景”页面确认智能场景信息后单击“是”。

### 📖 说明

终止之后该场景数据不可用。

## 4.4 获取推荐结果

智能创建完成，运行成功后，当服务状态会显示“运行中”，表示状态正常。您可以通过预测功能测试推荐结果进一步调整作业参数，也可以通过预测接口来调用API，获取推荐结果。

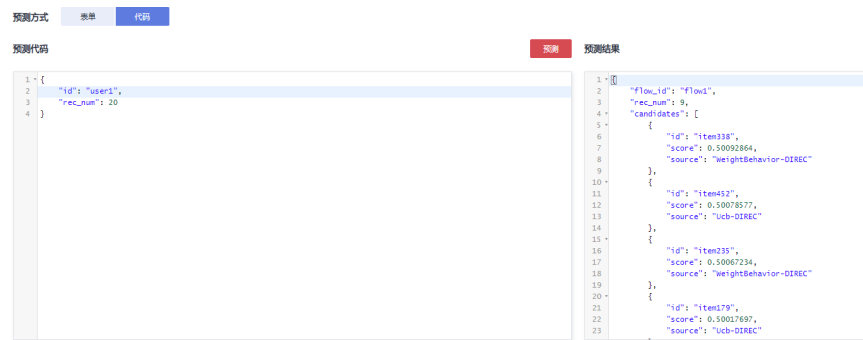
### 📖 说明

如果近线数据源有更新，需要重新调度召回策略，才会有对应的推荐结果。

## 预测

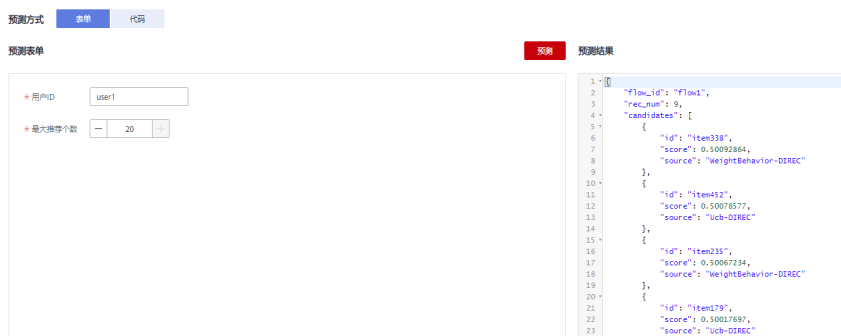
1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务>智能场景”，进入智能场景列表页面。
2. 单击“已完成”状态的目标服务名称进入智能场景详情页面，单击“预测”页签，包括两种预测方式。
  - **代码：**输入预测代码，单击“预测”后显示预测结果，如图4-2所示。输入用户ID、推荐数量，根据您的策略返回用户的预测结果。如果是关联推荐，需要配置“物品项”，即推荐与物品项相关的产品。如果物品项有多个，需要用英文逗号隔开。

图 4-2 代码预测



- **表单**：输入“ID”，并设置“最大推荐个数”。其中ID可以为用户ID或者物品ID，单击“预测”后显示预测结果，如图4-3所示。如果是关联推荐，则需要配置“物品项”，即推荐与物品项相关的产品。如果物品项有多个，需要用英文逗号隔开。

图 4-3 表单预测



## 获取预测接口


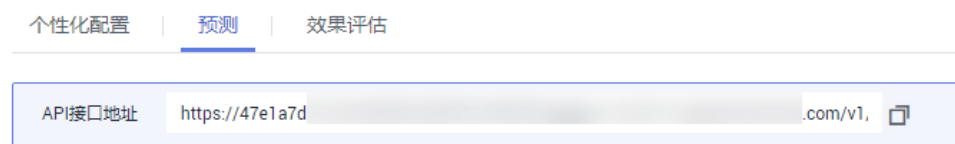
1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务>智能场景”，进入智能场景列表页面。
2. 单击“已完成”状态的目标服务名称进入智能场景详情页面，单击“预测”页签，您可以获取该服务的“API接口地址”。单击，复制接口地址，调用服务。

图 4-4 获取预测接口



## 调用接口

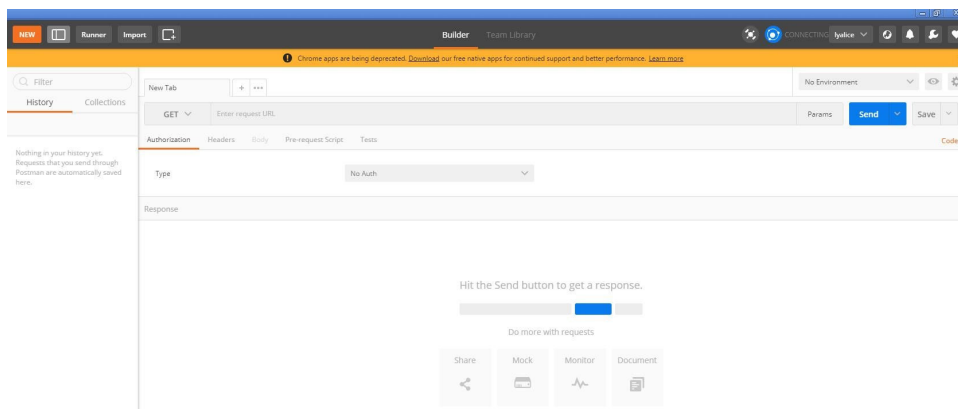
Mozilla、Google都为REST提供了图形化的浏览器插件，发送处理请求消息。

此处以Postman为例，指导您如何通过调用预测接口获取推荐结果。更多接口信息请参见《[推荐系统API参考](#)》。

1. 下载[Postman](#)软件并安装，您也可以直接在Chrome浏览器添加Postman扩展程序（也可使用其它支持发送post请求的软件）。

2. 打开Postman，如图4-5所示。

图 4-5 Postman 界面

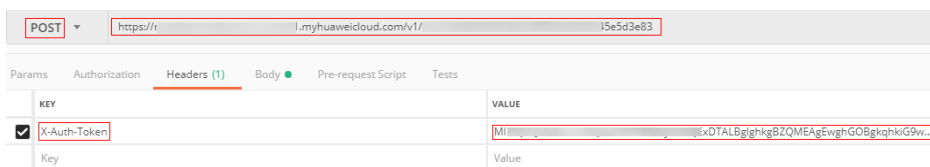


3. 在Postman界面填写参数。
  - 选择POST任务，将通过[获取预测接口](#)获取的调用地址复制到POST后面的方框。Headers页签的“KEY”值填写为“X-Auth-Token”，“VALUE”值为您获取到的Token（关于如何获取token，请参考[获取用户Token](#)），如图4-6所示。

**说明**

您也可以通过AK（Access Key ID）/SK（Secret Access Key）加密调用请求，具体可参见《RES API参考》中的“[认证鉴权](#)>认证鉴权>AK/SK认证”。

图 4-6 参数填写



4. 参数填写完成，在Postman页面，选择“Body>raw”，在下方文本框中填入JSON体，请求样例如下：
 

```
{
  "id": "user894",
  "rec_num": 10
}
```
5. 参数填写完成，单击“Send”发送请求，结果会在“Response”下的对话框里显示。其中，“id”为推荐的itemId，“score”为该物品的得分。示例如下：

```
{
  "flow_id": "flow1",
  "rec_num": 10,
  "candidates": [
    {
      "id": "item235",
      "score": 0.58418584,
      "source": "WeightBehavior-DIREC"
    },
    {
      "id": "item338",
      "score": 0.5653803,
      "source": "WeightBehavior-DIREC"
    },
    {
      "id": "item709",
```

```
"score": 0.55219465,  
"source": "WeightBehavior-DIREC"  
},  
{  
"id": "item960",  
"score": 0.55219465,  
"source": "WeightBehavior-DIREC"  
},  
{  
"id": "item469",  
"score": 0.55219465,  
"source": "WeightBehavior-DIREC"  
},  
{  
"id": "item236",  
"score": 0.55219465,  
"source": "WeightBehavior-DIREC"  
}  
]  
}
```



## 4.5 效果评估

创建效果评估可以对服务设置指标，查看推荐效果的反馈，可以根据系统提供的指标添加。

### 创建效果评估作业

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务>智能场景”，默认进入“智能场景”列表。
2. 在智能场景列表中，单击“运行中”状态的目标场景名称，进入详情页。
3. 单击“效果评估”页签，单击目标类型作业进行参数配置，请参见表4-3。

表 4-3 效果评估参数说明

参数名称	说明
名称	自定义名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	对于该效果评估作业的描述信息。
开启调度	默认开启调度。开启调度后，默认每天凌晨00:05开始执行。您也可以关闭调度手动执行效果评估作业。
指标类型	推荐服务效果评估指标，通过指标后的下拉框选择系统自置指标。您可以单击  添加指标，单击该指标后的  删除指标。 <ul style="list-style-type: none"><li>• 点击PV率</li><li>• 点击UV率</li></ul>

4. 效果评估作业运行完成之后会在页面下方生成评估结果图表，方便您查看近期的评估效果。

## 4.6 编辑或删除智能场景

针对运行失败或者草稿状态的智能场景支持进行重新编辑操作。您可以通过执行删除操作，删除当前场景。

### 前提条件

已存在创建成功的智能场景。

### 编辑智能场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务 > 智能场景”，进入智能场景列表页面。
2. 单击智能场景列表中的目标场景名称，进入“个性化配置”详情页。
3. 单击“编辑”对参数进行重新配置，单击“确认”完成场景参数修改。

图 4-7 编辑智能场景

个性化配置 | 预测 | 效果评估

编辑

场景规格

\* 离线计算规格 ② 4CU 8CU 16CU 32CU

\* 实时计算规格 ② 2CU 4CU 8CU 16CU

\* 排序模型训练规格 ② 8CU GPU-P100

\* 在线并发规格 ② - 5 +

个性化配置

匹配特征对

* 物品特征名	* 权重值 ②
category	1

⊕ 添加匹配特征对

开启调度

确认 取消

### 说明

- 编辑智能场景信息需要执行发布操作才可以同步数据。
- 发布场景会覆盖之前数据，请谨慎操作。

### 删除智能场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务 > 智能场景”，进入智能场景列表页面。

2. 选择智能场景列表中的目标场景，单击“操作”列的“删除”。
3. 在“删除场景”页面确认智能场景信息后单击“是”。

#### 说明

删除场景包含删除场景下所有的策略和服务。场景运行存储在OBS中的数据不会被删除，需要您登录OBS进行删除操作。

# 5 自定义场景

## 5.1 自定义场景简介

RES面向了解推荐算法等相关的用户，可自定义推荐中涉及算法的使用和组合，能够自定义开发推荐流程，提供推荐服务。您可以基于RES提供的多种推荐训练作业得到推荐候选集，用于在线服务计算得到推荐结果。

### 自定义场景功能说明

表 5-1 功能说明

功能	说明	详细指导
创建自定义场景	自定义场景作为一个包含多个子任务的作业，通常用于多个召回、过滤、排序等任务。	<a href="#">创建自定义场景</a>
召回策略	召回策略通过大数据计算或深度训练生成推荐候选集。	<a href="#">召回策略</a>
过滤规则	过滤规则用于生成推荐的过滤集，包含黑白名单、历史行为过滤等特性。支持用户在线上推理过程中完成对相关物品的过滤。	<a href="#">过滤规则</a>
特征工程	特征工程常用于抽取用户、物品的特征和特定算法的特征生成，一般作为某些算法的前置输入条件。	<a href="#">排序策略-离线特征工程</a>
排序策略	排序策略根据不同的算法模型对召回策略或者近线策略生成的候选集进行重排序，得到推荐候选集列表。	<a href="#">排序策略-离线排序模型</a>
在线服务	在线服务用来做线上推荐时的应用，每个服务之间是独立的。即根据不同的离线计算得到的候选集以及相关参数，提供不同的推荐服务。	<a href="#">在线服务</a>

功能	说明	详细指导
效果评估	指用于通过推荐系统推荐出去的结果集并利用trace_id回流到推荐系统的行为的点击率、转化率等指标的计算。	<a href="#">效果评估</a>

## 5.2 创建自定义场景

自定义场景基于用户群体不同推荐场景的需求，提供了多种多样的推荐策略和算法，实现了端到端的自定义推荐场景搭建，使每一个推荐场景都能得到针对性的推荐效果提升。

### 前提条件

- 已经存在创建成功并完成数据探索的数据源。
- 由于训练作业运行需消耗资源，确保账户未欠费。
- 确保您使用的OBS目录与RES在同一区域。

### 基本信息

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务>自定义场景”，默认进入“自定义场景”列表。
2. 在自定义场景列表中，单击右上角“创建自定义场景”，进入“创建自定义场景”页面。
3. 在创建自定义场景页面，填写“基本信息”相关参数，基本信息包含场景名称、数据源、场景规格等关键信息，详情请参见[表5-2](#)。
4. 基本信息配置完成后单击“创建并下一步：召回策略”。



图 5-1 自定义场景基本信息

**基本信息**

\* 场景名称

\* 场景类型  基于用户推荐物品  基于用户推荐用户  基于物品推荐物品  基于物品推荐用户

\* 服务类型  推荐引擎  排序引擎

\* 数据源

注意：创建完场景后数据源不能修改

---

**数据区间**

\* 数据选择 (天)

物品类别

设置黑名单

---

**场景规格**

\* 离线作业计算规格  4CU  8CU  16CU  32CU

\* 实时作业计算规格  2CU  4CU  8CU  16CU

\* 排序模型训练规格  8CU  GPU-P100

\* 在线服务并发规格

表 5-2 自定义场景基本信息参数说明

参数名称	子参数	说明
基本信息	场景名称	自定义场景的名称。只能包含数字、大小写字母、下划线和中划线，长度不能超过64位且不能为空。 场景名称创建之后不支持修改。

参数名称	子参数	说明
	场景类型	<p>选择需要的场景类型。</p> <ul style="list-style-type: none"> <li>“基于用户推荐物品”：某些用户的属性很相似，如电商平台根据这些用户的行为（浏览、点击、购买）计算与这些用户相似用户的行为，为该用户推荐相似用户浏览或购买的物品。</li> <li>“基于用户推荐用户”：某些用户的属性很相似，如交友平台根据这些用户的行为（浏览、点击）或属性推荐与这些用户相似用户。</li> <li>“基于物品推荐物品”：某些物品的属性、描述很相似，或者经常被一起浏览或购买。如媒资平台将会计算物品之间的相似或关联程度，当用户查看了某个物品的时候，会推荐最相似/最相关的物品。</li> <li>“基于物品推荐用户”：某些物品的属性、描述很相似，或者经常被一起购买。如房产平台会计算物品之间的相似或关联程度，当用户查看某个物品的时候，会推荐同时拥有该类型房源的房产经纪人。</li> </ul>
	服务类型	<p>选择您需要的服务类型。</p> <ul style="list-style-type: none"> <li>“推荐引擎”：推荐引擎用于对RES召回策略生成的候选集结果进行融合过滤和排序。</li> <li>“排序引擎”：排序引擎无需配置召回策略，允许用户提供自己的候选集，使用RES的排序策略进行排序。</li> </ul>
	数据源	<p>从右侧下拉框中选择RES系统中已有的数据源。当无可用数据源时，此下拉框为空。</p>
数据区间	时间单位	支持天和小时。
	数据选择	指定历史行为时间段，选取数据中最靠后的时间往前N天或N小时的行为数据计算用户偏好。
	物品类别	是否需要根据category值对原始数据进行过滤。如新闻网站，category取值可能为娱乐、体育、时事等。
	设置黑名单	设置黑名单列表。后续任务读取数据时，将会过滤掉指定的用户或物品的记录。文件格式为每个ID占据一行。
	用户黑名单/物品黑名单	上传设置的黑名单列表。
场景规格	离线计算规格	可选择“4CU”、“8CU”、“16CU”和“32CU”四种规格。其中，CU代表CPU核数。
	实时计算规格	可选择“2CU”、“4CU”、“8CU”和“16CU”四种规格。其中，CU代表CPU核数。

参数名称	子参数	说明
	排序模型训练规格	可选择“GPU-P100”、“8CU”。其中，CU代表CPU核数。
	在线并发规格	在线服务获取推荐的最高调用次数，单位为次每秒，默认为5，取值范围为5-500。

## 召回策略

您可以根据业务需要，选择合适的召回策略。召回策略用于配置离线计算逻辑，通过启动离线计算任务进行候选推荐结果集的生成。

各个召回策略的详细介绍请参见：

- [基于综合行为热度推荐](#)
  - [基于物品的协同过滤推荐](#)
  - [基于用户的协同过滤推荐](#)
  - [基于交替最小二乘的矩阵分解推荐](#)
  - [业务规则-基于历史行为记忆生成候选集](#)
  - [业务规则-人工导入](#)
  - [基于特征匹配的召回策略](#)
  - [基于UCB的召回策略](#)
  - [近线召回](#)
  - [基于物品相似度的实时召回](#)
  - [基于用户相似度的实时召回](#)
1. 在“创建自定义场景”页面，进入“召回策略”页签，单击“添加召回策略”，根据业务需要在下拉框中选择一个合适的策略。本章节以“基于交替最小二乘的矩阵分解推荐”为例进行创建，如[图5-2](#)所示。

图 5-2 召回策略

您还可以创建5个召回策略。请至少创建一个召回策略。

召回策略 — 基于交替最小二乘的矩阵分解推荐 ?

\* 名称

描述

\* 隐向量维度 ?

\* 优化正则化系数 ?

\* 最大迭代次数 ?

\* 最大推荐结果数 ?

开启调度 ?

确认 取消

表 5-3 基于交替最小二乘的矩阵分解推荐参数说明

参数	说明
名称	根据业务自定义命名。
描述	根据业务自定义。例如，这是一个测试策略。
隐向量维度	在ALS算法中使用，指定用户隐向量、物品隐向量的隐含因子的维度大小。如果离线计算失败，建议调小至10以下。 默认为10。
优化正则化系数	在ALS算法中使用，指定用户隐向量、物品隐向量的隐含因子的维度大小。 设置为默认参数0.01。
最大迭代次数	指定迭代优化的最大迭代次数。如果离线计算失败，建议调小至10以下。 默认为10。
最大推荐结果数	最多生成多少个推荐结果。 默认参数100。

参数	说明
开启调度	开启调度，按照指定的调度策略定期执行作业。 <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

2. 创建完成后单击“确认”。保存后的召回策略会展示在下面的列表中。
3. （可选）在目标召回策略右侧，可以对所选策略进行“编辑”和“删除”操作。

### 📖 说明

删除操作不可恢复，请您谨慎操作。

4. 具体策略的参数说明可单击上方策略名称进行查看，策略设置完成后，单击“下一步：过滤规则”。

## 过滤规则

过滤规则是用于配置候选集的过滤方式，使相关内容不进入候选集。

- 如果您选择的是排序引擎，在配置完基本信息之后进入“过滤规则”页签。
  - 如果您选择的是推荐引擎，在配置完召回策略之后，进入“过滤规则”页签。
1. “创建自定义场景”页面，进入“过滤规则”页签，单击“添加过滤策略”，如图5-3所示。策略的详细参数说明请参见[过滤规则](#)。

图 5-3 创建过滤规则

表 5-4 过滤规则参数说明

参数	说明
名称	根据业务自定义命名。
描述	根据业务自定义描述。
频次	<p>筛选出每个用户的行为数据中满足要求的数据，使之不进入候选集。</p> <ul style="list-style-type: none"> <li>● 相对时间间隔：与数据源内的行为数据记录最后的一条时间相差天数。</li> <li>● 最小次数：某用户对某物品产生某行为的最小次数。</li> <li>● 行为类型：指定行为类型。</li> <li>● 绝对时间间隔：与当前时间相差天数。</li> <li>● 最大次数：某用户对某物品产生某行为的最大次数。</li> </ul>
行为过滤逻辑	设置为AND或者OR。AND即为在7天内点赞超过3次且在7天内消费超过3次的物品。
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>● “调度周期”：调度周期可选“天”或“周”。</li> <li>● “调度类型”：包括自定义和间隔调度。</li> <li>● “开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>● “时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

2. 创建完成后单击“确认”。保存后的策略会展示在下面的列表中。
3. （可选）在目标策略右侧，可以对所选策略进行“编辑”和“删除”操作。
4. 具体策略的参数说明可单击上方策略名称进行查看，策略设置完成后，单击“下一步：排序策略”。

## 排序策略-离线特征工程

在进行排序之前需要先进行离线特征工程将数据转化为排序所需要的格式。

1. “创建自定义场景”页面，进入“排序策略”页签，如图5-4所示。策略的详细参数说明请参见[排序策略-离线特征工程](#)。

图 5-4 离线特征工程

离线特征工程

\* 名称

描述

\* 待提取用户特征 ?

\* 待提取物品特征 ?

自定义行为类型 ?

\* 行为去重方式 ?

\* 时间类型 ?

\* 训练集测试集划分方式 ?

\* 训练数据占比 ?

\* 测试数据占比 ?

开启调度 ?

2. 创建完成后单击“确认”。保存后的策略会展示在下面的列表中。
3. （可选）在目标策略右侧，可以对所选策略进行“编辑”和“删除”操作。
4. 具体策略的参数说明可单击上方策略名称进行查看。

## 排序策略-离线排序策略

排序策略离线排序策略是用于对召回策略生成的候选集进行二次排序。如果离线使用排序模型，需在[排序策略-离线特征工程](#)中创建完成后才可以正常使用排序策略。

各个策略的详细参数设置和输入输出请单击下方链接查看。

- [Logistic Regression \(LR\)](#)
- [DeepFM](#)
- [AutoGroup](#)

1. 在“创建自定义场景”页面，进入“排序策略”页签，单击“离线排序策略”，如[图5-5](#)所示。用户可以根据业务需要在“添加排序策略”下拉框中选择一个合适的排序策略。


图 5-5 添加离线排序策略



2. 创建完成后在单击“确认”。保存后的策略会展示在下面的列表中。
3. （可选）在目标策略右侧，可以对所选策略进行“编辑”和“删除”操作。
4. 具体策略的参数说明可单击上方策略名称进行查看。

## 排序策略-近线特征工程

近线特征工程是对近线数据进行特征处理，以便生成特征数据用于排序。必须先完成离线排序模型相关操作才可进行此配置。

近线特征工程默认为关闭状态，您可以在创建近线特征工程右侧单击  按钮，在弹出的对话框中单击“是”进行创建。

## 排序策略-近线排序策略

近线排序策略用于对在线实时数据排序。如果使用在线排序模型，需在[排序策略-近线特征工程](#)中创建完成后才可以正常使用排序策略。

1. 在“创建自定义场景”页面，进入“排序策略”页签，单击“添加近线排序策略”。
2. 进行在线学习参数配置。
  - 名称：自定义在线排序策略名称。
  - 离线排序策略：从下拉框中选择已经创建完成的[排序策略-离线排序策略](#)作业名称。
  - 优化器类型：具体参数解释请参见[Logistic Regression \(LR\)](#)。



图 5-6 添加近线排序策略

近线排序策略

\* 名称

描述

\* 离线排序策略

\* 优化器类型

\* 学习率

\* 初始梯度累加和

\* L1正则项系数

\* L2正则项系数

3. 创建完成后在单击“确认”。保存后的策略会展示在下面的列表中。
4. （可选）在目标策略右侧，可以对所选策略进行“编辑”和“删除”操作。
5. 参数配置完成后可单击“下一步：在线服务”。

## 在线服务


1. 在“在线服务”页面，填写基本信息。您可以根据实际情况填写“名称”、“描述”信息或“开启异常告警”。单击  开启异常告警，在下拉选项中选择 SMN 主题名称。如果您还未有可选的主题，请前往消息通知服务 [创建主题](#)。
2. 单击“添加在线流程”，并进行命名，您最多可以部署 5 个在线流程。根据选择的服务类型配置在线流程，包括“召回池”、“过滤”、“排序”等关键信息。具体参数信息参见 [表 5-21](#)。

图 5-7 创建在线服务

### 召回池

\* 推荐候选集 ?    %

添加推荐候选集 您还可以添加2个在线候选集

### 过滤 (黑名单)

离线过滤 ?

行为过滤 ?

属性去重 ?  添加属性去重

属性值过滤 ?  添加属性值过滤 您还可以添加10个属性值过滤

属性对过滤 ?  添加属性对过滤 您还可以添加10个属性对过滤

### 过滤 (白名单)

属性值保留 ?  添加属性值保留 您还可以添加10个属性值保留

属性对保留 ?  添加属性对保留 您还可以添加10个属性对保留

### 排序

点击率预估 ?

综合排序 ?

### 高级选项 ^

高级类型选择 ?

打散

3. 参数配置完成后单击“创建并完成”。作业创建完成后您可以在自定义场景列表页面目标场景的“操作”列单击“执行”，待场景状态为“运行中”时，作业运行完成。

## (可选) 效果评估

自定义场景创建完成后，页面自动跳转到自定义场景列表页面，您可以单击目标场景名称进入详情页，创建效果评估。

1. “效果评估”页签下单击 ，在弹出页面选择填写“名称”和“描述”。

- 效果评估名称：只能由字母、数字、中划线和下划线组成，并且长度小于64个字符。

图 5-8 创建效果评估

效果评估

\* 名称

描述

0/256

开启调度

\* 指标类型

+ 添加指标

确认 取消

2. 配置创建效果评估相关参数，请参见表5-22。
3. 配置完成后单击“确认”，该作业会出现在效果评估页签下的列表中。鼠标移动至该策略上方，可以对该策略作业进行“编辑”、“查看”、“执行”和“删除”操作。
4. 您可以在自定义详情页面查看“已完成”状态的评估结果图表，方便您查看近期的评估效果。

## 5.3 算法介绍及参数说明

### 5.3.1 召回策略

召回是指对大量的物品做初选，为每一个用户形成个性化候选集。召回策略是指通过大数据计算或深度训练生成推荐候选集的算法策略。召回策略中内置了多种召回方式，您可根据自己场景选择。

### 基于综合行为热度推荐

基于综合行为热度推荐统计用户对物品所有行为的加权热度。如果选择用户分群，将生成每个分组的热度推荐；如果不选择，将生成全局热度推荐。

表 5-5 基于综合行为热度推参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。

参数名称	说明
行为次数统计方法	物品记录统计方式。 <ul style="list-style-type: none"> <li>“pv”：page view，每个用户可对同个物品产生多条记录。</li> <li>“uv”：unique visitor，每个用户只能对每个物品产生一条记录，算法只保留权重最大的记录。</li> </ul>
用户分群	数据源类型包括用户特征和物品特征，根据数据源筛选数据，选出需要的属性。只能选择进行用户分组还是物品分组，分组内可配置多个特征。 <ul style="list-style-type: none"> <li>“用户特征”：在用户特征下拉框中选择需要的特征参数进行召回计算。例" user_meta_list ": ["gender","age"]</li> <li>“物品特征”：在物品户特征下拉框中选择需要的特征参数进行召回计算。例" item_meta_list ": ["weight","score"]</li> </ul>
最大推荐结果数	每个用户最多生成多少个推荐结果。
开启时间跨度	不开启取全部数据，开启则指定从数据源中取最近天数或小时数的行为数据计算相似度。
时间单位	开启时间跨度后，支持按照天或者小时为单位从指定从数据源中取行为数据计算相似度。
时间跨度	用于指定从数据源中取最近多少天或者多少小时的行为数据计算热度。默认取全部数据。
开启调度	开启调度，按照指定的调度策略定期执行作业。 <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

## 基于物品的协同过滤推荐

基于物品的协同过滤推荐采用经典推荐算法基于物品的协同过滤ItemCF进行召回。

表 5-6 基于物品的协同过滤推荐参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。

参数名称	说明
最近邻域数	在ItemCF算法中使用，生成的相似度矩阵中为每个物品保留的若干个最相似物品。默认100。
用户活跃度	用来过滤用户的活跃度阈值。 取值范围：1-10000。 默认值：1。
效用阈值	用户对物品综合打分的阈值。 取值范围：0.000001-10。
最小交叉度	物品和物品之间被同一用户行为记录的数量，计算相似度时，过滤掉共同记录小于最小交叉度的item。 默认值：1。
最大推荐结果数	最多生成多少个推荐结果。 默认值：100。
开启时间跨度	不开启取全部数据，开启则指定从数据源中取最近天数的行为数据计算相似度。
时间跨度（天）	用于指定从数据源中取最近多少天的行为数据计算相似度。默认取全部数据。
上传物品相似度	支持客户通过obs导入自定义的相似度信息。相似度文件格式为json，其中subject为用户或物品，relations为与subject相似的用户或物品及其相似度。例如， <pre> {   "subject": "item108",   "relations": [     {       "id": "item115",       "score": 0.699357793663589     },     {       "id": "item60",       "score": 0.659905609639582     },     {       "id": "item61",       "score": 0.640305447750641     },     {       "id": "item18",       "score": 0.614275316537666     }   ] } </pre>

参数名称	说明
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

## 基于用户的协同过滤推荐

基于用户的协同过滤推荐采用经典算法基于用户的协同过滤(UserCF)进行召回。基于用户的协同过滤算法是通过用户的历史行为数据发现用户对物品的喜欢（如购买，收藏，内容评论或分享），并对这些喜好进行度量和打分。根据不同用户对相同物品的态度和偏好程度计算用户之间的关系。在有相同喜好的用户间进行物品推荐。

例如，A、B两个用户都购买了abc三本图书，并且给出了5星的好评。则A和B属于同一类用户。可以将A看过的图书d也推荐给用户B。

表 5-7 基于用户的协同过滤推荐参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。
最近邻域数	在UserCF算法中使用，生成的相似度矩阵中为每个用户保留的若干个最相似用户。默认为100。
最小交叉度	物品和物品之间被同一用户行为记录的数量，计算相似度时，过滤掉共同记录小于最小交叉度的item。 默认值：1。
物品活跃度	物品过滤用户的活跃度阈值。 取值范围：1-10000。 默认值：1。
效用阈值	用户对物品综合打分的阈值。 取值范围：0.000001-10。
最大推荐结果数	最多生成多少个推荐结果。 默认值：100。
开启时间跨度	不开启取全部数据，开启则指定从数据源中取最近天数的行为数据计算相似度。

参数名称	说明
时间跨度	用于指定从数据源中取最近多少天的行为数据计算相似度。默认取全部数据。
上传物品相似度	支持客户通过obs导入自定义的相似度信息。相似度文件格式为json，其中subject为用户或物品，relations为与subject相似的用户或物品及其相似度。例如， <pre> {   "subject": "item108",   "relations": [     {       "id": "item115",       "score": 0.699357793663589     },     {       "id": "item60",       "score": 0.659905609639582     },     {       "id": "item61",       "score": 0.640305447750641     },     {       "id": "item18",       "score": 0.614275316537666     }   ] } </pre>
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

## 基于交替最小二乘的矩阵分解推荐

基于交替最小二乘的矩阵分解推荐：基于用户-物品的行为信息作为原始矩阵，利用ALS优化算法对原始矩阵进行矩阵分解，分解之后的用户隐向量矩阵和物品隐向量矩阵可以用来生成预估的新的用户-物品评分矩阵，提取出评分最高的若干个物品作为召回结果。

表 5-8 基于交替最小二乘的矩阵分解推荐参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。

参数名称	说明
隐向量维度	在ALS算法中使用，指定用户隐向量、物品隐向量的隐含因子的维度大小。如果离线计算失败，建议调小至10以下。默认为10。
迭代次数	指定迭代优化的最大迭代次数。如果离线计算失败，建议调小至10以下。默认为10。
优化正则化系数	在ALS算法中使用，指定正则化系数，作为优化目标中参数项代价的系数，用于避免过拟合现象发生。默认为0.01。
最大推荐结果数	最多生成多少个推荐结果。默认为100。
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>



## 业务规则-基于历史行为记忆生成候选集

业务规则-基于历史行为记忆生成候选集可以从用户历史行为数据中筛选出发生过某些行为的物品(例如，筛选出曾经点击过5次的物品)。此策略适用于“看了又看”，“买了又买”等推荐场景。

表 5-9 业务规则-基于历史行为记忆生成候选集参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。




参数名称	说明
频次	<ul style="list-style-type: none"> <li>“相对时间间隔”：指定历史行为时间段，选取数据中每个用户最靠后的行为数据的时间往前N天的行为数据计算用户偏好。建议至少设置30天。</li> <li>“绝对时间间隔”：指定历史行为时间段，选取数据中每个用户距离现在时间往前N天的行为数据计算用户偏好。建议至少设置30天。</li> <li>“行为类型”：指定行为类型。</li> <li>“最小次数”：行为次数下界，高于此的物品才会被选中。默认为1。</li> <li>“最大次数”：行为次数上界，低于此的物品才会被选中。</li> </ul> <p>单击 ，增加频次。你可以单击右侧的  删除该频次。</p>
最大推荐结果数	最多生成多少个推荐结果。默认为100。
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

## 业务规则-人工导入

业务规则-人工导入是指用户自定义物品列表，并将此列表作为候选集，以供在线服务调用。

表 5-10 业务规则-人工导入参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。
OBS地址	用户可从此OBS地址中选择自定义物品列表的csv文件。单击  ，从弹出的对话框中，选择数据存储的OBS桶及其文件夹。

参数名称	说明
导入候选集类型	<p>选择导入候选集类型，目前支持以下三种。</p> <ul style="list-style-type: none"> <li>物品-分数候选集：物品-分数候选集可以用于在线服务的推荐候选集。</li> <li>用户相似度候选集：用户相似度候选集可用于实时召回。</li> <li>物品相似度候选集：物品相似度候选集可用于实时召回与在线服务的推荐候选集。</li> </ul>
最大推荐结果数	最多生成多少个推荐结果。默认为100。
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

数据格式(.csv文件):

itemId1

itemId2

itemId为物品id，算法将物品列表上传的物品作为候选集



用户可从OBS中选择保存有人工编辑推荐结果的列表（即物品ID）。

## 基于特征匹配的召回策略

基于特征匹配的召回策略会用用户画像和物品画像的相关属性进行匹配，为用户召回属性匹配程度高的若干个物品。

表 5-11 基于特性匹配的召回策略参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。
匹配类型	<p>通过匹配数据的特征，生成推荐候选集。例如，通过用户匹配物品生成给用户推荐物品的候选集，物品自匹配生成给物品推荐物品的候选集。可选：</p> <ul style="list-style-type: none"> <li>用户匹配物品</li> <li>物品自匹配</li> </ul>

参数名称	说明
匹配特征对	<p>用户和物品相关联特征。请根据实际情况配置参数，如果属性匹配特征对相似度较高内存不够时需提升配置。</p> <ul style="list-style-type: none"> <li>• 用户特征名：字符串，长度1-20。</li> <li>• 物品特征名：字符串，长度1-20。</li> <li>• 权重值：权重影响不同物品属性匹配的程度，取值0.01-1，2位小数。</li> <li>• 匹配个数度量：如果开启匹配个数度量，同个特征匹配个数多的数据有优势。例如博客标签中，匹配5个标签(tags)比匹配1个标签(tags)更相关。如果不开启，多值特征匹配时，匹配特征个数无关，都被视为匹配。</li> <li>• 操作：可以单击操作列下面的进行删除某个匹配特征对。</li> </ul> <p>您可以单击进入“添加匹配特征对”页面进行配置。设置特征对的用户特征名、物品特征名和权重。</p>
最近邻域数	<p>机器学习中的概念，例如协同过滤计算中，需要计算物品之间，用户之间的相似度。最近邻域数是x，就是一个物品/用户找出x个和他相似的物品/用户。默认为100。</p>
最大推荐结果数	<p>最多生成多少个推荐结果。默认为100。</p>
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>• “调度周期”：调度周期可选“天”或“周”。</li> <li>• “调度类型”：包括自定义和间隔调度。</li> <li>• “开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>• “时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

## 基于 UCB 的召回策略

基于UCB的召回策略综合考虑了用户操作行为表中，物品发生的某几种行为类型及次数，然后给每一个物品都计算一个得分，最终返回得分最高的若干个物品。

表 5-12 基于 UCB 的召回策略参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。

参数名称	说明
最小行为次数	在物品上产生过行为的最小用户数，其中一个用户在一个物品上只计算一次行为。默认为30。
折中参数	令alpha为Exploration 和 Exploitation之间的折中参数，其取值范围为[0,1]，alpha越趋近于0，则物品的得分对历史得分高的物品越有利，即 Exploitation。反之，alpha越趋近于1，则物品的得分越倾向于探索新物品，即Exploration。默认为0.5。
最大推荐结果数	最多生成多少个推荐结果。默认为100。 该参数会运用在“输出数据”的推荐候选集上。
时间跨度（天）	用于指定从数据源中取最近多少天的行为数据计算热度。默认取全部数据。
开启调度	开启调度，按照指定的调度策略定期执行作业。 <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

## 近线召回

近线召回是基于用户实时的行为，分析用户短时间内（10秒周期）的兴趣爱好，推荐相关内容。

表 5-13 近线召回参数说明

参数名称	说明
名称	近线召回名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	对于近线召回策略的描述信息。
itemCF作业名称	默认无。 填写后可利用itemCF作业生成的结果，进一步扩大召回候选集的范围。

参数名称	说明
开启时间过滤	<p>开启时间过滤即对候选集进行时间上的过滤。默认关闭。</p> <p>开启之后需配置如下参数：</p> <ul style="list-style-type: none"> <li>“过滤字段”：发布时间（publishTime）或过期时间（expireTime）。</li> <li>“保留期”：单位（天）。最小值1，最大值15。</li> </ul> <p>召回结果中，仅保留符合（（当前时间-过滤字段的值）&lt; 保留期）的召回结果。</p>
召回字段	<p>召回字段即在召回时，仅利用选择的召回字段进行召回。</p> <p>数据结构任务完成后，人工复核中，选中应用于兴趣属性的物品特征（如果物品特征为tags，选中应用于兴趣属性。则在召回字段下拉栏中显示为“interested_tags”）。选中召回字段后，默认利用该字段中召回的字段值个数为1（可根据实际需求进行调整）。</p> <ul style="list-style-type: none"> <li>限制1：1 ≤ 召回字段数量 ≤ 5</li> <li>限制2：1 ≤ 召回字段中利用召回字段值的数量和 ≤ 10</li> </ul>
最大推荐结果数	<p>最多生成多少个推荐结果。默认100。</p> <p>该参数会运用在“输出数据”的推荐候选集上。</p>
开启异常告警	<p>开启异常告警即当任务发生异常时，产生告警通知到客户。默认关闭。开启之后需配置如下参数：</p> <p>“主题名称”：单击选择SMN主题。如果您还未有可选的主题，请前往消息通知服务<a href="#">创建主题</a>。</p>

## 基于物品相似度的实时召回

基于物品相似度的实时召回策略是根据用户短时间内行为数据（如购买，收藏，内容评论或分享），通过牛顿冷却定律对相关行为的初始权重进行衰减和汇总，从而找出用户该时间段内感兴趣的物品，结合物品的相似度信息进行关联推荐。

表 5-14 基于物品相似度的实时召回参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。
依赖作业名称	已经完成的可以提供物品和物品相似度关系的作业，用来进行物品的关联推荐。
topK	用户最感兴趣的排序在前K个的物品。

参数名称	说明
行为	<ul style="list-style-type: none"> <li>行为类型：用户感兴趣的行为类型。</li> <li>权重值：行为的初始权重。</li> <li>衰减系数：用于衰减行为初始权重的系数。</li> <li>有效时间：用户配置的行为发生时间与当前时间的间隔，以小时为单位。系统只处理在该时间范围内的行为记录。</li> </ul>

## 基于用户相似度的实时召回

基于用户相似度的实时召回策略是根据用户间的相似度信息，找到相似用户短时间内行为数据（如购买，收藏，内容评论或分享），通过牛顿冷却定律对相关行为的初始权重进行衰减和汇总，从而找出相似用户该时间段内感兴趣的物品，加权汇总后推荐给该用户。



表 5-15 基于用户相似度的实时召回参数说明

参数名称	说明
名称	策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	策略的具体描述。
依赖作业名称	已经完成的可以提供用户和用户相似度关系的作业，用来进行用户的关联推荐。
topK	用户最感兴趣的排序在前K个的物品。
行为	<ul style="list-style-type: none"> <li>行为类型：用户感兴趣的行为类型。</li> <li>权重值：行为的初始权重。</li> <li>衰减系数：用于衰减行为初始权重的系数。</li> <li>有效时间：用户配置的行为发生时间与当前时间的间隔，以小时为单位。系统只处理在该时间范围内的行为记录。</li> </ul>

### 5.3.2 过滤规则

过滤规则用于配置候选集的过滤方式，使之不进入候选集。对于每个需要过滤的行为，生成用户具有该行为的物品的列表。再对同用户的每种行为的物品列表进行“与”或者“或”的关系，最终生成用户-物品过滤表。



表 5-16 过滤规则参数说明

参数名称	说明
名称	自定义过滤规则名称。由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	过滤规则的相关描述信息。
频次	<p>指定与用户个性化的物品候选集过滤准则。例如对于用户过去3天内有过view行为的物品（如新闻）过滤，使之不进入候选集。单击  增加历史行为过滤，单击后方的  删除过滤行为。相关参数解释如下：</p> <ul style="list-style-type: none"> <li>● 相对时间间隔：与数据源内的行为数据记录最后的一条时间相差天数。</li> <li>● 最小次数：某用户对某物品产生某行为的最小次数。</li> <li>● 行为类型：指定行为类型。</li> <li>● 绝对时间间隔：与当前时间相差天数。</li> <li>● 最大次数：某用户对某物品产生某行为的最大次数。</li> </ul> <p>系统默认行为类型包括：</p> <ul style="list-style-type: none"> <li>● view：物品曝光</li> <li>● click：用户点击物品</li> <li>● collect：用户收藏了某个物品</li> <li>● uncollect：用户取消收藏某个物品</li> <li>● search_click：用户点击搜索结果中的物品</li> <li>● comment：用户对物品的评论</li> <li>● share：分享</li> <li>● like：点赞</li> <li>● dislike：点衰</li> <li>● grade：评分</li> <li>● consume：消费</li> <li>● use：观看视频/听音乐/阅读</li> <li>● download：下载</li> <li>● tip：打赏</li> <li>● subscribe：关注</li> </ul>
行为过滤逻辑	<p>各个频次之间的逻辑。</p> <ul style="list-style-type: none"> <li>● “AND” 逻辑为在7天内点赞超过3次且在7天内消费超过3次的物品。</li> <li>● “OR” 逻辑为在7天内点赞超过3次或在7天内消费超过3次的物品。</li> </ul>



参数名称	说明
最大推荐结果数	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“调度类型”：包括自定义和间隔调度。</li> <li>“开始调度时间”：选择具体的调度时间。当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“时间间隔”：如果选择的调度类型为间隔调度，需要配置调度的时间间隔。</li> </ul>

### 5.3.3 排序策略-离线特征工程

表 5-17 特征工程参数说明

参数名称	说明
名称	自定义离线特征工程名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	对于特征工程的描述信息。
待提取用户特征	<p>排序模型需要经特征工程处理后的数据，选择排序模型需要的用户特征，未选择的用户特征将不会被处理，即排序模块将忽略这些特征。</p> <p><b>说明</b> 离散的区间个数不能超过100个，请您根据业务需求合理分配参数值。</p> <p>单击 ，增加用户特征。在下拉选项中勾选特征参数名称并进行配置。当“特征值类型”为“多值枚举型”时，您可以根据需求自定义枚举个数。其他类型可选的参数信息如下：</p> <ul style="list-style-type: none"> <li>“等距离散”：根据业务需求限定数值“最小值”、“最大值”和“距离”。例如，根据age进行等距离散，设置年龄最小值为1，最大值为100，离散距离为10。等距离散会按照age将1-10岁，11-20岁等作为一个区间进行离散。</li> <li>“归一化”：归一化，根据业务需求限定数值“最小值”和“最大值”。例如，根据weight进行归一化，设置weight最小值为50，最大值为200。如果给定的数值x在该区间范围内则采用以下计算公式处理：<math>x_{new} = (x - 50) / (200 - 50)</math>，不在区间内的则按异常值处理，如x=80时，<math>x_{new} = 0.2</math>。</li> <li>“用户自定义离散”：根据业务需求限定“离散点”。例如，根据age进行离散，设置年龄离散点分别为3、9、15即年龄会按照0-3、3-9、9-15进行散；单击  添加离散点。</li> <li>“不离散”：（默认）不做归一化，不对数据做处理。</li> </ul>



参数名称	说明
待提取物品特征	<p>排序模型需要经特征工程处理后的数据，选择排序模型需要的物品特征，未选择的物品特征将不会被处理，即排序模块将忽略这些特征。</p> <p><b>说明</b> 离散的区间个数不能超过100个，请您根据业务需求合理分配参数值。</p> <p>单击 ，增加物品特征。在下拉选项中勾选特征参数名称并进行配置。当“特征值类型”为“多值枚举型”时，您可以根据需求自定义枚举个数。其他类型可选的参数信息如下：</p> <ul style="list-style-type: none"> <li>“等距离散”：根据业务需求限定数值“最小值”、“最大值”和“距离”。例如，根据order_price进行等距离散，设置价格最小值为1，最大值为100，离散距离为10。那么等距离散会按照价格将1-10元，11-20元等为一个区间进行离散。</li> <li>“归一化”：归一化，根据业务需求限定数值“最小值”和“最大值”。例如，根据weight进行归一化，设置weight最小值为50，最大值为200。如果给定的数值x在该区间范围内则采用以下计算公式处理：<math>x_{new} = (x - 50) / (200 - 50)</math>，不在区间内的则按异常值处理，如x=80时，<math>x_{new} = 0.2</math>。</li> <li>“用户自定义离散”：根据业务需求限定“离散点”。例如，根据order_price进行离散，设置价格离散点分别为150、200、250即价格会按照0-150、151-200、201-250进行散；单击  添加离散点。</li> <li>“不离散”：（默认）不做归一化，不对数据做处理。</li> </ul>
自定义行为类型	<p>如果开启“自定义行为类型”，算法则按用户设置的行为类型及权重进行排序预处理任务，否则默认使用数据源中的行为类型及权重进行任务。</p> <ul style="list-style-type: none"> <li>“正向行为类型”：设置正向行为的类型及权重值。</li> <li>“负向行为类型”：设置负向行为的类型及权重值。</li> </ul>
行为去重方式	<p>将行为数据中某个用户对某个物品的多条记录进行去重，目前支持按行为权重去重（正向行为且权重越大的优先）和按时间去重（每天、每星期、每个月保留一条数据）。</p> <ul style="list-style-type: none"> <li>“权重绝对值”：保留行为权重绝对值最高的一条数据。</li> <li>“日期”：按照“时间类型”，保留指定区间的一条最新数据。</li> </ul>
时间类型	<p>将行为数据中某个用户对某个物品的多条记录进行去重，只保留每个时间区间内的一条最新数据。您可以根据数据实际情况选择去重的时间周期，可选天、周、月。默认为“天”。</p> <p>周一是第一天的行为数据按周去重，是否周一是第一天的行为数据。如果选择否，则认为周日是第一天的行为数据，保留每个时间区间内的一条最新数据。</p>

参数名称	说明
训练集测试集划分方式	<p>数据划分方式按时间比例或个数比例划分训练集测试集。</p> <ul style="list-style-type: none"> <li>“时间比例”：将全部数据的时间跨度按照时间比例划分成两段数据，训练数据为前一段时间中的数据，测试数据为后一段时间的数据，取值TIME。</li> <li>“个数比例”：个数比例是将全部数据按个数比例随机划分成训练集和测试集传入值。取值RANDOM。</li> </ul>
训练数据占比	生成的结果中，训练集占整个训练集和测试集的比例，默认0.7。
测试数据占比	生成的结果中，训练集占整个训练集和测试集的比例，默认0.3。
开启调度	<p>开启调度，按照指定的调度策略定期执行作业。</p> <ul style="list-style-type: none"> <li>“调度周期”：调度周期可选“天”或“周”。</li> <li>“选择时间”：当调度周期选择为“周”时，可在此下拉框中勾选星期一到星期天的任一天进行调度。</li> <li>“具体时间”：选择具体的调度时间。</li> </ul>

### 5.3.4 排序策略-离线排序模型

#### 排序策略简介

排序策略用于训练排序模型，该模型将被用于对召回策略召回的候选集进行排序，以将推荐物品顺序调整到最优。

#### Logistic Regression (LR)

LR算法是一种广义的线性回归分析模型，常用于数据挖掘、疾病自动诊断、经济预测等领域。LR算法通过在线性回归的基础上叠加一个sigmoid激活函数将输出值映射到[0,1]之间，是机器学习领域里常用的二分类算法。

表 5-18 逻辑斯蒂回归参数说明

参数名称	说明
名称	自定义策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	对于该策略的描述信息。
最大迭代轮数	模型训练的最大迭代轮数，默认50。
提前终止训练轮数	在测试集上连续N轮迭代AUC无提高时，迭代停止，训练提前结束，默认5。

参数名称	说明
初始化方法	<p>模型参数的初始化方法。</p> <ul style="list-style-type: none"> <li>• normal: 正态分布 平均值: 默认0 标准差: 0.001</li> <li>• uniform : 均匀分布 最小值: 默认-0.001, 均匀分布的最小值, 必须小于最大值。 最大值: 默认0.001, 均匀分布的最大值, 必须大于最小值。</li> <li>• xavier: 初始化初始值为 均值为0, 方差为 <math>\text{Var}(w_i)=1/n_{in}</math> 的均匀分布 (高斯或者随机分布)。其中 <math>n_{in}</math> 是该神经元的输入数目。</li> </ul>
优化器类型	<ul style="list-style-type: none"> <li>• grad: 梯度下降算法 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>• adam: 自适应矩估计算法 结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认1e-8。</li> </ul> </li> <li>• adagrad: 自适应梯度算法 对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>• ftrl: Follow The Regularized Leader 适用于处理超大规模数据的, 含大量稀疏特征的在线学习的常见优化算法。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>
L2正则项系数	<p>叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</p>

参数名称	说明
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"> <li>• full：指针对全量参数计算。</li> <li>• batch：则仅针对当前批数据中出现的参数计算</li> </ul> 说明 batch模式计算速度快于full模式。
重新训练	对第一次训练无影响，仅影响任务重跑。 <ul style="list-style-type: none"> <li>• “是”：清空上一轮的模型结果后重新开始训练。</li> <li>• “否”：导入上一轮的训练结果继续训练。适用于欠拟合的情况。</li> </ul>
批量大小	一次训练所选取的样本数。
训练数据集切分数量	将整个数据集切分成多个子数据集，依次训练，每个epoch训练一个子数据集。

## DeepFM

DeepFM，结合了FM和深度神经网络对于特征表达的学习，同时学习高阶和低阶特征组合，从而达到准确地特征组合学习，进行精准推荐。

表 5-19 深度网络因子分解机参数说明

参数名称	说明
名称	自定义策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	对于该策略的描述信息。
最大迭代轮数	模型训练的最大迭代轮数，默认50。
提前终止训练轮数	在测试集上连续N轮迭代AUC无提高时，迭代停止，训练提前结束，默认5。

参数名称	说明
初始化方法	<p>模型参数的初始化方法。</p> <ul style="list-style-type: none"><li>• normal: 正态分布 平均值: 默认0 标准差: 0.001</li><li>• uniform : 均匀分布 最小值: 默认-0.001, 均匀分布的最小值, 必须小于最大值。 最大值: 默认0.001, 均匀分布的最大值, 必须大于最小值。</li><li>• xavier: 初始化初始值为 均值为0, 方差为 <math>\text{Var}(w_i)=1/n_{in}</math> 的均匀分布 (高斯或者随机分布)。其中 <math>n_{in}</math> 是该神经元的输入数目。</li></ul>

参数名称	说明
优化器类型	<ul style="list-style-type: none"> <li>● grad: 梯度下降算法                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>● adam: 自适应矩估计算法                             <p>结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认1e-8。</li> </ul> </li> <li>● adagrad: 自适应梯度算法                             <p>对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>● ftrl: Follow The Regularized Leader                             <p>适用于处理超大规模数据的, 含大量稀疏特征的在线学习的常见优化算法。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>
L2正则项系数	叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"> <li>● full: 指针对全量参数计算。</li> <li>● batch: 则仅针对当前批数据中出现的参数计算</li> </ul> <b>说明</b> batch模式计算速度快于full模式。
隐向量长度	分解后的表示特征的向量的长度。默认10。
神经网络结构	神经网络的层数与每一层的神经元节点个数。默认400,400,400。

参数名称	说明
激活函数	神经网络中的激活函数，将一个（或一组）神经元的值映射为一个输出值。 <ul style="list-style-type: none"> <li>• relu</li> <li>• tanh</li> <li>• sigmoid</li> </ul>
神经元值保留概率	神经网络前向传播过程中以该概率保留神经元的值。默认0.8。
重新训练	对第一次训练无影响，仅影响任务重跑。 <ul style="list-style-type: none"> <li>• “是”：清空上一轮的模型结果后重新开始训练。</li> <li>• “否”：导入上一轮的训练结果继续训练。适用于欠拟合的情况。</li> </ul>
批量大小	一次训练所选取的样本数。
训练数据集切分数量	将整个数据集切分成多个子数据集，依次训练，每个epoch训练一个子数据集。

## AutoGroup

AutoGroup，使用自动分组技术探索高阶特征交互组合，更深层次地挖掘出多个特征间联系，增强模型排序效果。

表 5-20 AutoGroup 参数说明

参数名称	说明
名称	自定义策略名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	对于该策略的描述信息。
最大交互阶数	算法探索的最大的特征交互阶数。例如，当设定最大交互阶数为4时，算法会探索2阶、3阶、4阶的特征交互组合。默认3。
各阶隐向量长度	各阶特征交互使用的隐向量长度，数量需对应最大交互阶数。默认10,60,80。
哈希长度	每阶特征交互所选择的group数量，数量需对应最大交互阶数。默认10,60,80。
特征交互层惩罚项系数	特征交互层输出值的惩罚项系数，用来防止过拟合。默认0.0001,0.0001,0.0001。
神经网络结构	神经网络的层数与每一层的神经元节点个数。默认400,400,400。

参数名称	说明
激活函数	<p>神经网络中的激活函数，将一个（或一组）神经元的值映射为一个输出值。</p> <ul style="list-style-type: none"> <li>• relu</li> <li>• tanh</li> <li>• sigmoid</li> </ul>
神经元值保留概率	<p>神经网络前向传播过程中以该概率保留神经元的值。默认0.8。</p>
优化器类型	<ul style="list-style-type: none"> <li>• grad: 梯度下降算法 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>• adam: 自适应矩估计算法 <p>结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认1e-8。</li> </ul> </li> <li>• adagrad: 自适应梯度算法 <p>对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>• ftrl: Follow The Regularized Leader <p>适用于处理超大规模数据的, 含大量稀疏特征的在线学习的常见优化算法。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>
结构部分优化器	<p>详细参数请参见优化器类型。</p>











参数名称	说明
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"> <li>• full: 指针对全量参数计算。</li> <li>• batch: 则仅针对当前批数据中出现的参数计算</li> </ul> 说明 batch模式计算速度快于full模式。
L2正则项系数	叠加在模型的2范数之上，用来对模型值进行限制防止过拟合。默认0。
隐向量层L2正则化系数	隐向量层使用的L2正则化系数，作用如“L2正则项系数”描述。默认0.001。
wide部分L2正则化系数	wide层使用的L2正则化系数，作用如“L2正则项系数”描述。默认0.001。
最大迭代轮数	模型训练的最大迭代轮数，默认50。
提前终止训练轮数	在测试集上连续N轮迭代AUC无提高时，迭代停止，训练提前结束，默认5。
重新训练	对第一次训练无影响，仅影响任务重跑。 <ul style="list-style-type: none"> <li>• “是”：清空上一轮的模型结果后重新开始训练。</li> <li>• “否”：导入上一轮的训练结果继续训练。适用于欠拟合的情况。</li> </ul>
批量大小	一次训练所选取的样本数。
训练数据集切分数量	将整个数据集切分成多个子数据集，依次训练，每个epoch训练一个子数据集。
融合多值特征	将多值特征的多个embedding融合成一个embedding。
融合线性部分	是否使用模型架构中的线性部分。
固定哈希结构	是否固定结构参数。默认值为“否”，非特殊情况建议使用默认值。




### 5.3.5 在线服务

在线服务用来做线上推荐时的应用，每个服务之间是独立的。即根据不同的离线计算得到的候选集以及相关参数，提供不同的推荐服务。

表 5-21 在线服务参数说明

参数名称	子参数	说明
召回池	“推荐候选集”	<p>“添加推荐候选集”</p> <ul style="list-style-type: none"> <li>选择表名：添加离线、近线任务或者在线自定义检索召回策略生成的候选集进行融合。您可以单击  添加推荐候选集，单击右侧的  删除该候选集。 <ul style="list-style-type: none"> <li>离线、近线召回：该策略候选集都可以通过下拉框直接选取。</li> <li>在线检索召回：支持在线检索候选集召回，检所请求通过预测接口（参考<a href="#">获取推荐结果</a>中的 search_info 字段使用方法）传入，生成一路召回对应候选集。该策略默认名称为“dynamic-search”。</li> </ul> </li> <li>融合方式 <ul style="list-style-type: none"> <li>优先级：根据优先级和百分比计算多个召回候选集融合。优先级高的推荐结果将确保在优先级低的之前。P1 优先级最高，P1 优先级如果没有用户对应的推荐物品，由配置的低优先级补充，以此类推。优先级相同的推荐候选集，根据百分占比确认召回策略推荐数量，同优先级下的数据占比之和需要等于 100%。</li> <li>权重：根据权重加权融合计算多个召回候选集融合。分数计算规则：每个策略的所占权重和物品在每个召回候选集中所得的分数加权融合，多个策略中相同的物品会进行分数累加。权重大小之和要等于 1。</li> </ul> </li> </ul>
过滤 (黑名单)	离线过滤	<p>对离线过滤规则作业产生的结果进行过滤，使之不进入候选集。</p> <p>单击打开  按钮，在下拉框中选择目标过滤规则的作业结果。</p>
	行为过滤	<p>对用户行为进行过滤使之不进入候选集。</p> <p>单击打开  按钮，进行行为过滤设置。例如对于用户过去 3 天内有过曝光行为的物品过滤，使之不进入候选集。可以对曝光过但是有购买行为的物品进行排除，使之依旧在候选集里。</p> <ul style="list-style-type: none"> <li>时间区间：指定用户行为时间，取值 1-7，默认为 1。</li> <li>行为类型：指定行为类型，对有该行为类型的物品进行过滤。最多选五种。</li> <li>排除行为类型：将有此行为类型的物品排除掉，使之依旧在候选集中。</li> </ul>

参数名称	子参数	说明
	属性去重	<p>单击  增加属性去重。</p> <p>“属性名”：从下拉列表中选择属性名进行去重。如“product_color”，则对产品颜色相同的物品进行去重。</p> <p>“去重位置”：可选择“排序前”或者“排序后”。</p>
	属性值过滤	<p>指定定制化用户属性以及物品属性过滤规则，属性过滤规则用于过滤最终用户的推荐结果。例如，对于一线城市的用户过滤敏感信息物品，使之不进入候选集。单击  增加属性过滤规则。</p> <ul style="list-style-type: none"> <li>“用户属性”：指定在用户属性中需要过滤的字段，包含属性名和属性值。如过滤籍贯是广东且性别为男性的用户。</li> <li>“物品属性”：指定在物品属性中需要过滤的字段，包含属性名和属性值。如过滤产品颜色为红色且产品品牌为华为的物品。</li> <li>“应用于”：将该规则应用于哪些路召回候选集，默认会作用于所有在召回池中选中的所有路候选集。</li> </ul>
	属性对过滤	<p>指定定制化用户属性以及物品属性过滤规则，属性过滤规则用于过滤最终用户的推荐结果。例如，对于一线城市的用户过滤敏感信息物品，使之不进入候选集。单击  增加属性对过滤规则。</p> <ul style="list-style-type: none"> <li>“用户-物品”：指定用户属性或者物品属性中需要过滤的字段。如过滤籍贯是广东且性别为男性的用户。</li> <li>“应用于”：该过滤规则应用于所选的候选集，默认会作用于所有在召回池中选中的所有路候选集。</li> </ul>
过滤（白名单）	属性值保留	<p>指定定制化用户属性、物品属性和应用于某个召回策略属性过滤规则，保留该选定的属性值，使之进入候选集。例如，对于保留一线城市的用户物品信息物品。单击  增加属性值保留。</p> <ul style="list-style-type: none"> <li>“用户属性”：指定在用户属性中需要过滤的字段，包含属性名和属性值。如保留籍贯是广东且性别为男性的用户。</li> <li>“物品属性”：指定在物品属性中需要过滤的字段，包含属性名和属性值。如保留产品颜色为红色且产品品牌为华为的物品。</li> <li>“应用于”：该过滤规则应用于所选的候选集，默认会作用于所有在召回池中选中的所有路候选集。</li> </ul>

参数名称	子参数	说明
	属性对保留	<p>指定定制化用户属性以及物品属性过滤规则，保留该选定的属性对，使之进入候选集。例如，对于保留一线城市用户的物品信息。单击  增加属性对保留。</p> <ul style="list-style-type: none"> <li>“用户-物品”：指定用户属性或者物品属性中需要保留的字段。如保留籍贯是广东且性别为男性的用户。</li> <li>“应用于”：该过滤规则应用于所选的候选集，默认会作用于所有在召回池中选中的所有路候选集。</li> </ul>
排序	点击率预估	<p>点击率预估默认关闭。打开按钮之后，在“排序”下拉项中选择计算成功的排序策略名称。</p>
	综合排序	<p>综合排序即根据客体和主体的相关属性进行重排序处理。</p> <ul style="list-style-type: none"> <li>属性匹配，单击  添加匹配特征对，如用户特征名为 city，物品特征名选择 tags，权重值设置为 1，其意思为根据选定的主体和客体相关属性对的匹配情况累计权重值。</li> <li>数值比较，单击  添加属性权重，在下拉列表中选择属性名称，设置权重，根据客体相关数值属性的大小进行重排序。</li> <li>统计方式，根据数值属性的大小顺序（ORDER）或者绝对值进行权重累加（ABS）统计。</li> <li>点击率权重：当同时选择点击率预估和综合排序进行重排序时，汇总分数时点击率相关得分的权重值。</li> <li>综合排序权重：当同时选择点击率预估和综合排序进行重排序时，汇总分数时综合排序相关得分的权重值。</li> <li>融合方式：当同时选择点击率预估和综合排序进行重排序时，汇总分数时的统计方式。根据数值属性的大小顺序（ORDER）或者绝对值进行权重累加（ABS）统计。</li> </ul>
高级类型选项	打散	<p>打散是指推荐的结果集中根据客体的选择的字符串类型的属性进行打散，避免推荐结果集过于集中，增加推荐结果的新颖性。</p> <p>您可以在下拉列表中选择打散属性。</p>
	置顶	<p>置顶功能则人为的将选定的某一路或多路候选集放在推荐结果前面。您可以在下拉列表中选择需要置顶的推荐候选集。</p>



参数名称	子参数	说明
	降权	<p>降权是指定时间区间内用户对物品有过指定的行为类型记录时，将对相应物品进行降权处理。</p> <p>例如对于用户过去3天内有过曝光行为的物品降权，但是可以对曝光过但是有购买行为的物品进行排除，不对其进行降权。</p> <ul style="list-style-type: none"> <li>• 时间区间：指定用户行为时间，取值1-7，默认为3。</li> <li>• 行为类型：指定行为类型，对有该行为类型的物品进行降权。最多选五种。</li> <li>• 排除行为类型：将有此行为类型的物品排除掉，不对其进行降权。</li> </ul>

### 5.3.6 效果评估

创建效果评估可以对线上服务设置指标，查看推荐效果的反馈，可以根据系统提供的指标添加。

表 5-22 效果评估参数说明

参数名称	说明
名称	自定义名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~64个字符。
描述	对于该效果评估作业的描述信息。
开启调度	默认开启调度。开启调度后，默认每天凌晨00:05开始执行。您也可以关闭调度手动执行效果评估作业。

参数名称	说明
指标类型	<p>推荐服务效果评估指标，通过指标后的下拉框选择系统自置指标或自定义指标。您可以单击  添加指标，单击该指标后的  删除指标。</p> <ul style="list-style-type: none"> <li>• 点击PV</li> <li>• 推荐点击PV率</li> <li>• 自定义指标</li> </ul> <p>自定义指标包含参数设置和指标设置两部分。</p> <ul style="list-style-type: none"> <li>• 参数设置 <ul style="list-style-type: none"> <li>- 参数别名：用户指定参数别名应用于指标公式。</li> <li>- 行为类型：选择需要进行评估的行为类型，如物品曝光。</li> <li>- 阈值：阈值是用来衡量用户行为有效性的标准，当数据源的actionMeasure的值大于阈值时，当前用户行为有效。</li> <li>- 去重：您可以单击勾选 <input checked="" type="checkbox"/>，根据用户对行为记录去重。</li> </ul> </li> <li>• 指标设置 <ul style="list-style-type: none"> <li>- 指标名称：请您定义评估的指标名称。</li> <li>- 指标公式：用户指定自定义指标公式，如：A/(A+B)，参数A、B代表自定义参数的参数别名。只支持+、-、*、/。</li> </ul> </li> </ul>


## 5.4 修改自定义场景

对于已创建的自定义场景，您可以修改相关参数配置以匹配业务变化。

### 前提条件


已存在创建完成的自定义场景。

### 修改基本信息

1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务 > 自定义场景”，进入“自定义场景”列表页面。
2. 在自定义场景列表中，单击目标场景名称进入“自定义场景详情”页。
3. 在基本信息页签下单击 ，在弹出页面修改“数据区间”和“场景规格”等信息。
4. 参考[基本信息](#)修改自定义场景基本信息，然后单击“确认”完成修改。

### 新增作业配置

1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务 > 自定义场景”，进入“自定义场景”列表页面。

2. 在自定义场景列表中，单击目标场景名称进入“自定义场景详情”页。
3. 在目标策略页签下单击 ，在弹出页面选择目标策略，参考[算法介绍及参数说明](#)章节进行配置。
4. 配置完成后单击“确认”，该策略会出现在对应策略的详情列表中。鼠标移动至该策略上方，可以对该策略作业进行“编辑”、“查看”、“执行”和“删除”操作。

## 修改作业配置

1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务 > 自定义场景”，进入“自定义场景”列表页面。
2. 在自定义场景列表中，单击目标场景名称进入“自定义场景详情”页。
3. 在需要修改的页签，鼠标移动至目标策略会出现“编辑”、“查看”、“执行”和“删除”按钮。
4. 单击“编辑”，在弹出页面参考[算法介绍及参数说明](#)章节修改配置。
5. 配置完成后单击“确认”，您可以对该策略作业进行“执行”操作，更新参数信息。

## 删除作业配置

1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务 > 自定义场景”，进入“自定义场景”列表页面。
2. 在自定义场景列表中，单击目标场景名称进入“自定义场景详情”页。
3. 在需要删除的页签，鼠标移动至目标策略会出现“编辑”、“查看”、“执行”和“删除”按钮。
4. 单击“删除”，在弹出页面确认作业信息后单击“是”完成删除操作。

## 5.5 发布或终止自定义场景

针对创建成功的自定义场景，需要进行发布才可以运行作业获取推荐结果。通过执行终止操作，停止运行当前场景。目前支持在自定义场景列表页面和自定义场景详情页面进行操作。

发布或终止自定义场景默认对该场景下的所有作业执行发布或终止操作，包括召回策略、过滤规则、排序策略和在线服务等作业。

### 前提条件

已存在创建成功的自定义场景。

### 发布自定义场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务” > “自定义场景”，进入自定义场景列表页面。
2. 通过如下方式对场景进行发布。
  - 在自定义场景列表页面发布：选择自定义场景列表中的目标场景，单击“操作”列的“发布”。
  - 在自定义场景详情页面发布：单击自定义场景列表中的目标场景名称，进入自定义场景详情页，单击页面右上角的“发布”。

3. 在“发布场景”页面确认自定义场景作业运行的配置规格单击“是”。

#### 📖 说明

- 修改自定义场景信息需要执行发布操作才可以同步数据。
- 发布场景会覆盖之前数据，请谨慎操作。

## 终止自定义场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务” > “自定义场景”，进入自定义场景列表页面。
2. 通过如下方式对场景进行发布或终止。
  - 在自定义场景列表页面终止：自定义场景列表中的目标场景，单击“操作”列的“终止”。
  - 在自定义场景详情页面发布或终止：单击自定义场景列表中的目标场景名称，进入自定义场景详情页，单击页面右上角的“终止”。
3. 在“终止场景”页面确认自定义场景信息后单击“是”。

#### 📖 说明

终止之后该场景数据不可用。

## 5.6 获取推荐结果

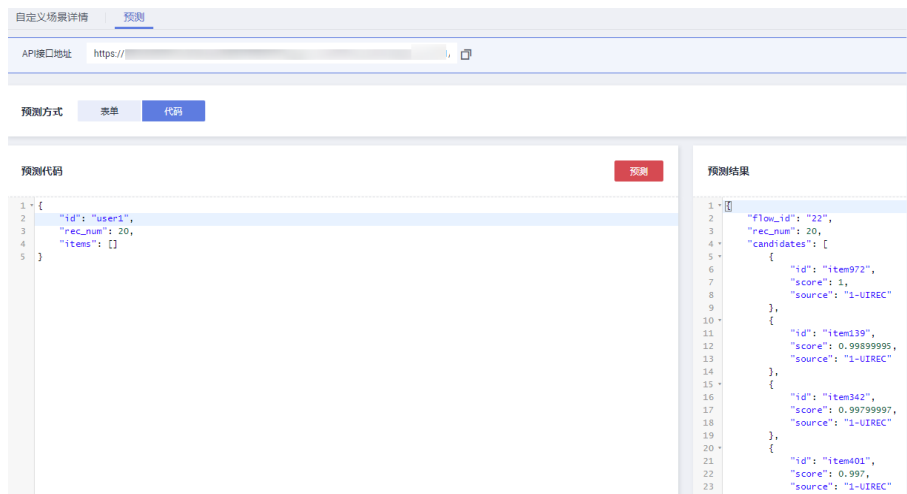
在线服务创建完成，部署成功后，当服务状态会显示“运行中”，表示服务状态正常。您可以通过在线预测功能测试推荐结果进一步调整作业参数，也可以通过预测接口来调用API，获取推荐结果。

### 界面预测

1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务>自定义场景”，进入自定义场景列表页面。
2. 单击“已完成”状态的目标服务名称进入自定义场景详情页面，单击“预测”页签，包括两种预测方式。
  - **代码：**输入预测代码，单击“预测”后显示预测结果，如图5-9所示。输入用户ID和推荐数量，或者物品项，根据您的策略返回用户的预测结果。如果物品项有多个，需要用英文逗号隔开。

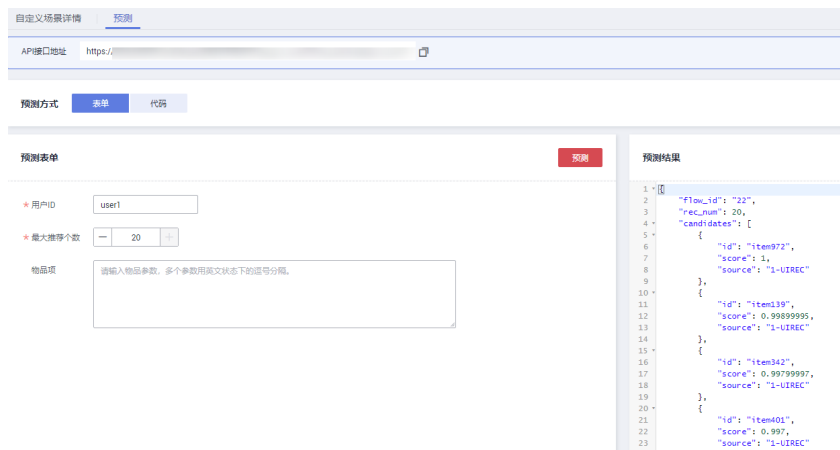


图 5-9 代码预测



- **表单**：输入“ID”和设置“最大推荐个数”，也可以设置“物品项”。其中ID可以为用户ID或者物品ID，单击“预测”后显示预测结果，如图5-10所示。如果物品项有多个，需要用英文逗号隔开。

图 5-10 表单预测



## 获取预测接口


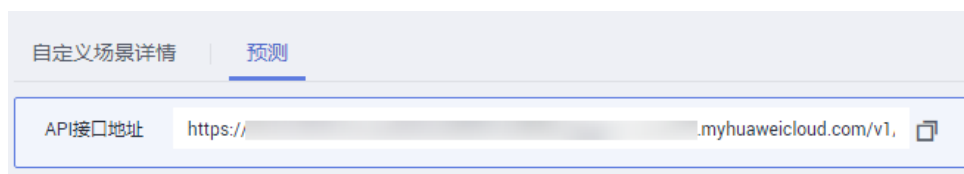
1. 登录RES管理控制台，在左侧菜单栏中选择“推荐业务>自定义场景”，进入自定义场景列表页面。
2. 单击“已完成”状态的目标服务名称进入自定义场景详情页面，单击“预测”页签，您可以获取该服务的“API接口地址”。单击, 复制接口地址，调用服务。

图 5-11 获取预测接口



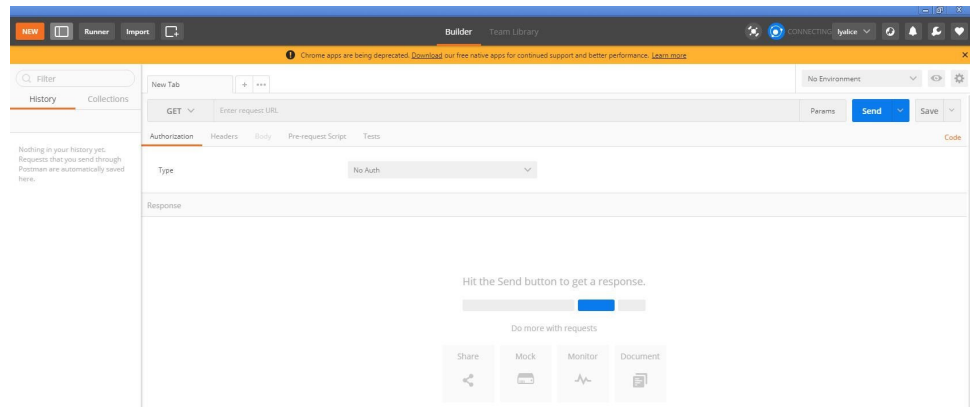
## 调用接口

Mozilla、Google都为REST提供了图形化的浏览器插件，发送处理请求消息。

此处以Postman为例，指导您如何通过调用预测接口获取推荐结果。更多接口信息请参见《[推荐系统API参考](#)》。

1. 下载**Postman**软件并安装，您也可以直接在Chrome浏览器添加Postman扩展程序（也可使用其它支持发送post请求的软件）。
2. 打开Postman，如图5-12所示。

图 5-12 Postman 界面

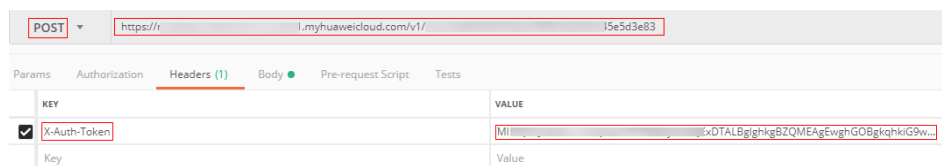


3. 在Postman界面填写参数。
  - 选择POST任务，将通过[获取预测接口](#)获取的调用地址复制到POST后面的方框。Headers页签的“KEY”值填写为“X-Auth-Token”，“VALUE”值为您获取到的Token（关于如何获取token，请参考[获取用户Token](#)），如图5-13所示。

### 说明

您也可以通过AK（Access Key ID）/SK（Secret Access Key）加密调用请求，具体可参见《[RES API参考](#)》中的“[认证鉴权](#)认证鉴权>AK/SK认证”。

图 5-13 参数填写



4. 预测请求参数说明。

表 5-23 参数说明

参数	子参数	说明
id	-	主体的id。此参数为必选。

参数	子参数	说明
items	-	客体的id列表，非必选项。有下面两种场景使用： <ul style="list-style-type: none"> <li>在线上策略使用到关联推荐召回策略时需要提供。例如，给用户推荐物品，主体是用户，客体是物品；给物品推荐用户，主体是物品，客体是用户；给用户推荐用户，主体是用户，客体是用户；给物品推荐物品，主体是物品，客体是物品。</li> <li>纯排序场景中，该参数为传入待排序物品列表的字段，传入物品集供模型排序使用。</li> </ul>
rec_num	-	推荐个数，默认值为20。此参数为必选。 目前最大配额是20，如需要调整，需要提工单进行调整。
search_info	-	配置动态搜索召回候选集，需要在线服务配置部分增加"dynamic-search" 候选集，为非必选项。
	match_infos	进行召回匹配的参数配置，即搜索的匹配信息。 <ul style="list-style-type: none"> <li>label: 客体的属性名称（可为字符串或字符串数组类型）。</li> <li>value: 相应的属性值。</li> <li>weight: 该属性值的匹配权重，多个匹配条件做加权汇总后按分值从大到小给出候选集。</li> </ul>
	filter_info	搜索的过滤信息。 <ul style="list-style-type: none"> <li>black_list: 客体需要过滤的黑名单。</li> <li>range: 选定一个数值型属性（label），输出的候选集该值必须大于等于设定的值才不会被过滤掉。</li> <li>category: 选定一个字符串或字符串数组类型属性（label），输出的候选集该值必须包含或等于设定的值才不会被过滤掉。</li> </ul>
	sort_info	搜索的排序信息。 <ul style="list-style-type: none"> <li>label: 客体的属性名称。</li> <li>orde: 排序规则，asc是升序排列，desc是降序排列。</li> </ul>
profile_query	-	对用户画像进行字段查询，传入string数组进行查询。

5. 单纯用户推荐预测JSON请求体如下，仅针对用户进行个性化推荐。

```
{
  "id": "user1",
  "rec_num": 10
}
```

6. 全量推荐请求预测样例，增加非必选参数项。

该示例的请求结果作为在线召回的候选集进行在线服务候选集的融合。

```
{
  "id": "user1",
  "rec_num": 20,
  "items": [
```

```
"item1",
"item2"
],
"search_info":{
  "match_infos":[
    {
      "label":"author",
      "value":"author8",
      "weight":0.9
    },
    {
      "label":"extend_strArray",
      "value":"walk",
      "weight":0.8
    }
  ]
},
"filter_info":{
  "black_list":[
    "item169",
    "item555",
    "c"
  ],
  "range":{
    "label":"expireTime",
    "value":1579790256
  },
  "category":{
    "label":"extend_string",
    "value":"basketball"
  }
},
"sort_info":[
  {
    "label":"score",
    "order":"asc"
  },
  {
    "label":"publishTime",
    "order":"desc"
  }
]
},
"profile_query":[
  "gender",
  "tag"
]
}
```

7. 参数填写完成，单击“Send”发送请求，结果会在“Response”下的对话框里显示。其中，“id”为推荐的itemId，“score”为该物品的得分。示例如下：

```
{
  "flow_id": "movielens",
  "rec_num": 10,
  "candidates": [
    {
      "score": "0.7978985",
      "id": "260",
      "source": "itemcf-UIREC"
    },
    {
      "score": "0.7043743",
      "id": "608",
      "source": "itemcf-UIREC"
    },
    {
      "score": "0.6727448",
      "id": "527",
      "source": "itemcf-UIREC"
    }
  ],
}
```

```
{
  "score": "0.6604858",
  "id": "36",
  "source": "itemcf-UIREC"
},
{
  "score": "0.6387838",
  "id": "50",
  "source": "itemcf-UIREC"
},
{
  "score": "0.624376",
  "id": "593",
  "source": "itemcf-UIREC"
},
{
  "score": "0.61491776",
  "id": "296",
  "source": "itemcf-UIREC"
},
{
  "score": "0.59284353",
  "id": "858",
  "source": "itemcf-UIREC"
},
{
  "score": "0.58551335",
  "id": "318",
  "source": "itemcf-UIREC"
},
{
  "score": "0.49690145",
  "id": "1",
  "source": "itemcf-UIREC"
}
],
"stats": {
  "itemcf-UIREC": 10,
  "hot-DIREC": 0
},
"profile": {
  "gender": "male",
  "tag": "VIP"
}
}
```

## 5.7 删除自定义场景

针对不再使用的自定义场景，您可以删除已释放资源。

### 前提条件

已存在创建成功的自定义场景。

### 删除自定义场景

1. 登录RES管理控制台，在左侧导航栏中选择“推荐业务 > 自定义场景”，进入自定义场景列表页面。
2. 选择自定义场景列表中的目标场景，单击“操作”列的“删除”。
3. 在“删除场景”页面确认自定义场景信息后单击“确定”。

### 说明

删除场景包含删除场景下所有的策略和服务。场景运行存储在OBS中的数据不会被删除，需要您登录OBS进行删除操作。

# 6 工作空间

## 6.1 工作空间简介

RES工作空间帮您实现离线作业、近线作业和在线服务隔离的功能，达到不同角色用户信息隔离管理的目的。

- 如果您未开通企业项目管理服务的权限，您可以在RES创建自己独立的工作空间。
- 如果你开通了企业项目管理服务的权限，可以在创建工作空间的时候绑定企业项目，并在企业项目下添加用户组，为不同的用户组设置细粒度权限供组里的用户使用。

### 工作空间功能

表 6-1 功能说明

功能	说明
<a href="#">创建工作空间</a>	RES支持通过工作空间实现对项目资源隔离。
<a href="#">（可选）创建企业项目</a>	为RES工作空间绑定企业项目，更方便您对项目的资源和用户进行管理。
<a href="#">编辑或删除工作空间</a>	您可以对RES工作空间进行编辑和删除类的管理操作。 <ul style="list-style-type: none"><li>• <a href="#">编辑工作空间</a></li><li>• <a href="#">删除工作空间</a></li></ul>

## 6.2 创建工作空间

### 前提条件

使用RES工作空间您需要具备如下权限：

- 请确保您已开通RES并完成密钥认证。
- 已经开通相关资源并进行全局配置。

## 创建工作空间

1. 登录[RES管理控制台](#)。
2. 您可以通过两种方式创建工作空间，进入“创建工作空间”页面，填写工作空间参数信息。详情请参见[表6-2](#)。
  - a. 在左侧导航栏“当前工作空间”下拉框中选择“创建工作空间”。

图 6-1 创建工作空间



- b. 在左侧导航栏单击进入“工作空间”页面。单击“创建”。

表 6-2 参数说明

参数	说明
名称	工作空间名称。例如，工作空间workspace1。 <ul style="list-style-type: none"><li>• 名称不能为空。</li><li>• 名称不超过64个字符，只能由中文、英文字母、数字、下划线、中划线组成，且不能使用任何大小写形式的“default”。</li></ul>
描述	工作空间描述。例如，我的第一个工作空间。 <ul style="list-style-type: none"><li>• 描述不超过256个字符。</li></ul>
企业项目	<ul style="list-style-type: none"><li>• 如果您未开通企业项目，则不会有此参数。</li><li>• 如果您开通了企业项目，可以从下拉选项选择企业项目，并确保该项目处于“已启用”状态。如果您无可选企业项目，则可以通过单击右侧“新建企业项目”进行创建。具体操作请参见<a href="#">创建企业项目</a>。</li></ul>

3. 单击“确定”，完成RES工作空间的创建。

您可以在工作空间列表页面查看创建的工作空间，当显示为“状态正常”时，该工作空间可正常使用。也可以左侧导航栏“当前工作空间”下拉框中选择您需要使用的工作空间。



## 切换工作空间

1. 按照[创建工作空间](#)步骤创建工作空间。
2. 在左侧导航栏“当前工作空间”下拉框中单击目标工作空间名称进行切换。

图 6-2 切换工作空间



## 6.3（可选）创建企业项目

创建企业项目是用于绑定RES服务工作空间。创建企业项目的用户必须是管理员，或在IAM侧已关联EPS Admin策略的用户，每个主账号最多可以创建100个企业项目。在使用工作空间绑定企业项目时，您需要在IAM中创建用户组、创建自定义策略、创建企业项目并添加用户组设置策略。

### 前提条件

- 如果您需要使用企业项目相关功能，请确保您已开通企业管理权限，如果还没有开通，请先登录管理控制台，单击页面右上方的“企业”。开通方法请参见：[开通企业项目](#)。
- 如果您需要使用权限控制相关功能，请确保您已开通细粒度授权功能。

### 创建企业项目并添加用户组

1. 在IAM中创建可用于添加到企业项目的用户组，具体操作请参见[创建用户组](#)。
2. （可选）在IAM中创建自定义策略，具体操作请参见[创建自定义策略](#)。您可以根据业务需求选择不同的策略，例如，配置推荐系统服务“允许查看作业列表”和“允许查看作业详情”两种自定义策略。
3. 在RES管理控制台，选择“工作空间>创建工作空间”，进入“创建工作空间”页面，单击“新建企业项目”。
4. 在“企业项目管理”页面，单击页面右上方“创建企业项目”，详情参见[创建企业项目](#)。
5. 企业项目创建完成后，为企业项目添加用户组并设置策略。详情请参见[为企业项目添加用户组](#)。

您可以在“可选策略”的下拉框中按照“所有策略”、“自定义策略”、“系统策略”对已有策略进行筛选，根据业务需求为不同的用户组选择合适的策略。步骤2创建的自定义策略可以在“可选策略”的下拉选项中选择“自定义策略”获取。

## 6.4 编辑或删除工作空间

工作空间页面主要列举了当前已创建的工作空间项目信息，包括工作空间“名称”、“状态”、“绑定的企业项目”、“创建人”、“创建时间”、“更新时间”和“操作”。

### 前提条件

已存在创建成功的工作空间。

### 编辑工作空间

您可以对创建的工作空间进行修改操作，具体操作如下：

1. 登录RES管理控制台，在左侧导航栏单击进入“工作空间”页面。
2. 单击目标工作空间“操作”列的“编辑”，在“编辑工作空间”页面修改工作空间的名称或者描述信息。
3. 单击“是”完成修改。

#### 说明

- “状态正常”的工作空间修改时不影响该空间用户正常使用。
- “删除中”状态和默认生成的“default”工作空间不支持修改。

### 删除工作空间

如果工作空间不再使用，您可以删除工作空间释放资源。具体操作如下：

1. 登录RES管理控制台，在左侧导航栏单击进入“工作空间”页面。
2. 单击目标工作空间“操作”列的“删除”。
3. 单击“确定”完成删除。

#### 说明

- 默认生成的“default”工作空间不支持删除。
- 删除工作空间后会默认清理该工作空间下所有资源。删除操作无法恢复，请谨慎操作。
- 如果工作空间绑定了企业项目，在删除工作空间时，对应的企业项目不会被删除或修改。

# 7 权限管理

## 7.1 创建用户并授权使用 RES

如果您需要对您所拥有的RES进行精细的权限管理，您可以使用[统一身份认证服务](#)（Identity and Access Management，简称IAM），通过IAM，您可以：

- 根据企业的业务组织，在您的华为云帐号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用RES资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- 将RES资源委托给更专业、高效的其他华为云帐号或者云服务，这些帐号或者云服务可以根据权限进行代运维。

如果华为云帐号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节，不影响您使用RES服务的其它功能。

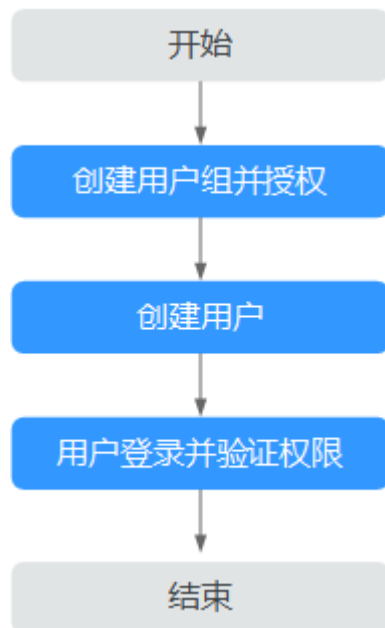
本章节为您介绍对用户授权的方法，操作流程如[图7-1](#)所示。

### 前提条件

给用户组授权之前，请您了解用户组可以添加的RES权限，并结合实际需求进行选择，RES支持的系统权限，请参见：[RES系统权限](#)。如果您需要对除RES之外的其它服务授权，IAM支持服务的所有权限请参见[权限策略](#)。

## 示例流程

图 7-1 给用户授权 RES 权限流程



### 1. 创建用户组并授权

在IAM控制台创建用户组，并授予推荐系统服务只读权限“RES ReadOnlyAccess”。

### 2. 创建用户并加入用户组

在IAM控制台创建用户，并将其加入1中创建的用户组。

### 3. 用户登录并验证权限

新创建的用户登录控制台，切换至授权区域，验证权限：

- 在“服务列表”中选择“推荐系统”，进入RES管理控制台，在左侧导航栏选择“数据源”，在页面右上角单击“创建”，尝试创建一个新的数据源，如果无法进行创建（假设当前权限仅包含RES ReadOnlyAccess），表示“RES ReadOnlyAccess”已生效。
- 在“服务列表”中选择除RES外的任一服务（假设当前策略仅包含RES ReadOnlyAccess），如果提示权限不足，表示“RES ReadOnlyAccess”已生效。

## 7.2 RES 自定义策略

如果系统预置的RES权限不满足您的授权要求，可以创建自定义策略。自定义策略中可以添加的授权项（Action）请参考[策略支持的授权项](#)。

目前华为云支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。
- JSON视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写JSON格式的策略内容。

具体创建步骤请参见：[创建自定义策略](#)。本章为您介绍常用的RES自定义策略样例。

## RES 自定义策略样例

- 示例1：拒绝用户删除作业

拒绝策略需要同时配合其他策略使用，否则没有实际作用。用户被授予的策略中，一个授权项的作用如果同时存在Allow和Deny，则遵循Deny优先原则。

如果您给用户授予RES FullAccess的系统策略，但不希望用户拥有RES FullAccess中定义的删除作业权限，您可以创建一条拒绝删除作业的自定义策略，然后同时将RES FullAccess和拒绝策略授予用户，根据Deny优先原则，则用户可以对RES执行除了删除作业外的所有操作。拒绝策略示例如下：

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "res:job:delete"
      ],
      "Effect": "Deny"
    }
  ]
}
```

- 示例2：多个授权项策略，允许查看作业列表和作业详情。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "res:job:get",
        "res:job:list"
      ]
    }
  ]
}
```

# 8 用户指南（旧版）

## 8.1 准备工作

### 8.1.1 注册华为帐号并开通华为云

在使用华为云服务之前您需要申请华为云帐号。通过此帐号，您可以使用所有华为云服务，并且只需为您所使用的服务付费。

#### 操作步骤

1. 进入[华为云](#)首页，单击页面右上角的“注册”。
2. 设置手机号、短信验证码、帐号名、密码并勾选“我已阅读并同意《华为云用户协议》和《隐私政策声明》”，单击“同意协议并注册”。
3. 页面提示注册成功后，使用注册的用户，登录华为云。

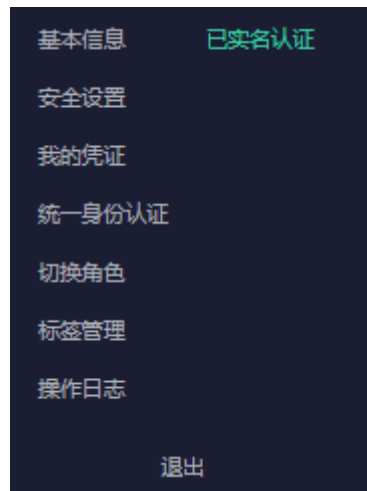
### 8.1.2 获取访问密钥并添加 RES 全局配置

由于使用RES时，用户上传数据、进行离线计算等需要依赖其他服务，如果没有添加访问密钥则无法使用其他服务。因此，在使用RES前，您需要获取访问密钥并在RES管理控制台完成添加访问密钥的配置。

#### 获取访问密钥

1. 登录[华为云](#)，鼠标移动至页面右上方用户名，在下拉列表中选择“帐号中心”。如[图8-1](#)所示。  
如果您当前已登录华为云管理控制台，在帐户名下方选择“基本信息”进入“帐号中心”，或者直接单击“我的凭证”，下一步可直接跳转至步骤3。

图 8-1 账号中心



2. 进入“账号中心”页面，在“基本信息”页签，选择“管理我的凭证”，如图8-2所示。

图 8-2 管理我的凭证

### 基本信息



3. 进入“我的凭证”页面，选择“访问密钥>新增访问密钥”，如图8-3所示。

图 8-3 单击新增访问密钥



4. 进入“新增访问密钥”页面，输入当前用户的登录密码，通过已验证手机或已验证邮箱进行验证，输入对应的验证码，如图8-4所示。

图 8-4 新增访问密钥

新增访问密钥

已验证手机 +86 158\*\*\*\*08      通过已验证邮箱验证

\* 短信验证码

描述

0/255

5. 单击“确定”，根据浏览器提示，保存密钥文件。密钥文件会直接保存到浏览器默认的下载文件夹中。打开名称为“credentials.csv”的文件，即可查看访问密钥（Access Key Id和Secret Access Key）。

## 添加密钥

1. 登录[RES管理控制台](#)，在左侧导航栏单击“全局配置”，进入“全局配置”页面。
2. 单击“添加密钥”，填写获取的访问密钥。
  - 访问密钥（AK）：输入密钥文件中的Access Key Id字段内容。
  - 私有访问密钥（SK）：输入密钥文件中Secret Access Key字段内容。

### 📖 说明

- 在添加AK/SK前，需要在OBS至少创建一个桶，否则会校验失败。具体操作请参见[创建OBS桶](#)。
  - 请确保所填写的AK、SK为当前账号所获取的。
  - 如果您添加访问密钥时，提示“上传的AK/SK不可用”，可能账号状态异常，欠费或被冻结。如有欠费，请您为华为云账户充值。如有其它问题请您[提工单](#)联系工程师解决。
3. 单击“确认”，完成访问密钥的添加。

## 8.1.3 创建 OBS 桶

RES使用对象存储服务（Object Storage Service，简称OBS）进行数据源以及黑白名单和用户数据的存储。因此，在使用RES之前您需要开通OBS服务并创建桶，然后在OBS桶中上传用户数据用于推荐作业的计算。

需要存放在OBS桶中的数据包括：

1. 离线数据源：包含用户类数据，物品类数据，行为数据以及推荐候选列表。
2. 黑名单和白名单：黑名单和白名单的过滤配置，便于用户快速得到有效数据。黑白名单数据都以物品ID的形式在OBS上存储。



## 操作步骤

1. 登录[OBS管理控制台](#)，创建OBS桶。例如，创建名称为“obs-res”的OBS桶。

### 📖 说明

创建桶的区域需要与RES所在的区域一致。例如：当前RES在华北-北京一区域，在对象存储服务创建桶时，请选择华北-北京一。

2. 通过OBS创建用于存放数据的文件夹
3. 推荐的OBS文件夹结构请参见[推荐系统OBS文件夹规范](#)。推荐您按照规范结构创建文件夹并上传至OBS桶中。

## 8.1.4 开通相关资源

在使用RES时需使用到其他的资源，因此需要先开通相关服务才可以正常使用RES。包含服务如下：

- [开通计算引擎DLI、ModelArts](#)
- [存储平台CloudTable](#)
- [（可选）数据接入资源DIS](#)

### 📖 说明

各服务的计费请参见：[产品价格详情](#)。

### 开通计算引擎 DLI、ModelArts

DLI用于推荐系统的离线计算和近线计算。排序策略所使用的计算资源由Modelarts服务提供并收费。在使用推荐系统之前，您需要开通数据湖探索（Data Lake Insight，简称DLI）来创建集群。

- **DLI服务**  
登录[华为云](#)。在华为云“产品”页签，选择“大数据>数据湖探索DLI”，进入DLI产品页。单击“进入控制台”，授权开始使用DLI。
- **ModelArts服务**  
登录[华为云](#)。在华为云“产品”页签，选择“人工智能>AI开发平台ModelArts”，进入ModelArts产品页。单击“立即使用”，开始使用ModelArts。

### 📖 说明

DLI集群建议创建跨源连接，通过RPC方式访问，提高读写性能。

### 开通存储平台 CloudTable

CloudTable作为存储平台，用于用户推荐在线数据和推荐候选集的存储。在使用推荐系统之前，您需要开通表格存储服务（CloudTable Service）来创建集群。

- 登录[华为云](#)。在华为云“产品”页签，选择“大数据>表格存储服务CloudTable”，进入CloudTable产品页。单击“立即使用”，进入CloudTable管理控制台。

### 📖 说明

CloudTable集群需要开启IAM统一身份认证。

## (可选) 开通数据接入资源 DIS

在使用推荐系统之前，数据接入服务 (Data Ingestion Service, 简称DIS) 采集用户实时日志，在计算引擎中消费实时日志并做近线处理。

登录[华为云](#)。在华为云“产品”页签，选择“大数据>数据接入服务DIS”，进入DIS产品页。单击“进入控制台”，开始使用DIS。

## 8.2 RES 操作流程

本章节介绍使用RES，从资源准备到在线服务完成推荐的全流程。RES流程图如[图8-5](#)所示。

图 8-5 RES 操作流程



表 8-1 使用流程说明

流程	子任务	说明	详细指导
准备资源	开通相关资源	基于您的业务需求，您需要开通RES相关服务，包括： <ul style="list-style-type: none"> <li>计算引擎DLI、ModelArts</li> <li>存储平台CloudTable</li> <li>(可选) 数据接入资源DIS</li> </ul>	<a href="#">开通相关资源</a>

流程	子任务	说明	详细指导
	绑定资源	针对您创建的集群等资源，需要完成绑定，才可以在创建作业时可选择绑定的集群进行计算存储等操作。	<a href="#">绑定资源</a>
	创建跨源连接	在使用DLI进行推荐系统的离线和近线计算时，建议创建跨源连接，用于访问CloudTable的数据源，提高读写性能。	<a href="#">创建跨源连接</a>
	开启公共终端节点	推荐系统在需要使用CloudTable集群进行数据存储，需开启公共终端节点之后推荐才能正常使用CloudTable集群。	<a href="#">开启公共终端节点</a>
上传数据	准备离线数据源	需要您准备包含用户类数据，物品类数据，行为数据以及推荐候选列表的离线数据源用于推荐系统的离线计算。	<a href="#">离线数据源</a>
	准备实时日志数据	RES根据实时发送到DIS上的日志，进行数据计算和处理，更新用户的相关数据。	<a href="#">实时日志</a>
运行推荐策略	创建离线作业	创建离线作业包括进行数据质量检测、创建组合作业、特征工程、召回策略、排序策略、过滤规则等作业。	<a href="#">运行推荐作业</a>
创建在线服务	-	创建在线服务用于部署上线服务、更新模型。配置实时计算的逻辑，包括设置在线流量、组装推荐结果和设置排序策略。根据策略做在线推荐结果融合、过滤、重排以及多流程之间的AB，并返回最终结果。	<a href="#">创建在线服务</a>
获取推荐结果	-	您可以通过在线服务预测结果，也可以通过API接口获取最终的推荐结果。	<a href="#">获取推荐结果</a>
（可选）效果评估	-	创建效果评估可以对线上服务设置指标，查看推荐结果的具体衡量指标。	<a href="#">效果评估</a>

## 8.3 管理资源

### 8.3.1 创建资源

用户在使用RES时需要先创建并绑定计算引擎DLI、存储平台CloudTable、数据接入资源DIS相关资源。

#### 背景信息

- 使用RES需要消耗其他服务资源，需要收费。根据您选择的资源不同，收费标准不同，针对不同类型资源的价格，详情请参见[产品价格详情](#)。

- 已开通计算引擎DLI、存储平台CloudTable、数据接入资源DIS相关服务。

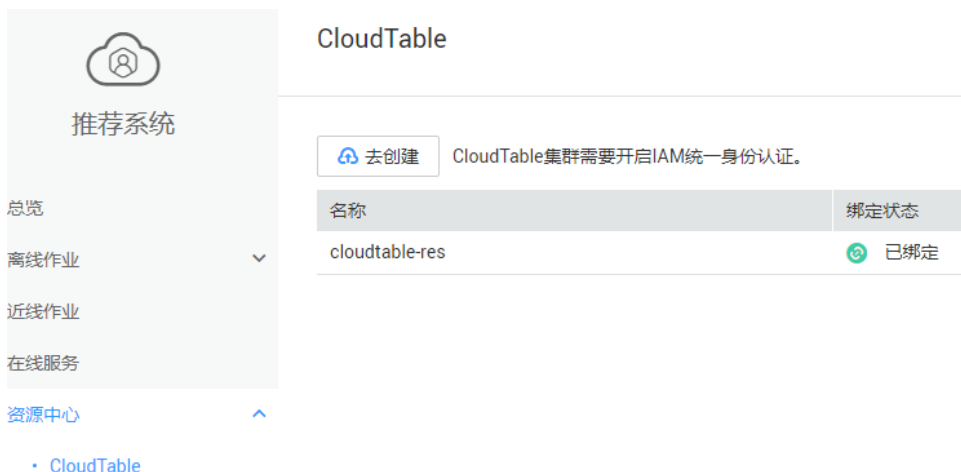
## 创建资源

1. 登录RES管理控制台，在左侧菜单栏中选择“资源中心”下的目标服务，单击页面上方“去创建”，跳转到对应服务管理控制台。
  - DLI创建集群操作详请参见[数据湖探索用户指南>Spark作业控制台>Spark集群管理>创建集群](#)章节。
  - CloudTable创建集群操作详请参见[表格存储服务用户指南> 集群模式> 集群操作指导> 创建集群](#)章节。
  - DIS创建通道操作详请参见[数据接入服务用户指南> 入门> 步骤1：开通DIS通道](#)章节。

### 说明

- CloudTable集群需要开启IAM统一身份认证。
2. 创建资源完成后，创建的资源会展示在“资源中心”列表中，如[图8-6](#)所示。

图 8-6 CloudTable 资源信息



### 8.3.2 绑定或解绑资源

在RES管理控制台界面，资源中心列表中会提供当前账号创建的DLI集群、CloudTable开启IAM认证的集群和DIS通道供用户选择进行绑定或解绑。

## 背景信息

- 绑定资源之后，将该资源应用于RES的作业训练及在线作业获取推荐结果。
- 解绑资源完成资源释放，已经解绑的资源不再应用于RES的相关计算。
- 已开通计算引擎DLI、存储平台CloudTable、数据接入资源DIS相关服务。

## 绑定资源

当用户完成绑定，在创建作业时可选择绑定的集群进行计算存储等操作。具体操作如下：

1. 登录RES管理控制台，在左侧菜单栏单击进入“资源中心”页面，选择计算引擎DLI、存储平台CloudTable、DIS进入列表。

2. 鼠标移动至目标集群所在区域，单击“操作 > 绑定”。
3. 在弹出的确认对话框中，单击“确定”完成资源绑定。

## 解绑资源

当用户需要释放资源或是更换资源绑定，则需要对已经绑定的资源进行解绑。具体操作如下：

1. 登录RES管理控制台，在左侧菜单栏单击进入“资源中心”页面，执行如下操作绑定或解绑资源。选择计算引擎DLI、存储平台CloudTable、DIS进入列表。
2. 鼠标移动至目标集群所在区域，单击“操作 > 解绑”。只有已绑定状态的服务资源支持“解绑”。
3. 在弹出的确认对话框中，单击“确定”完成资源解绑。

### 📖 说明

- 已创建的作业和服务不会因为资源解绑影响计算和推荐结果。
- 直接解绑DLI，不会自动删除创建的DLI跨源连接。您需要在DLI管理控制台进行操作。
- 解绑资源不会删除对应的集群和数据，您需要去服务对应的管理控制台进行删除。

## 8.3.3 创建跨源连接

在使用DLI进行推荐系统的离线和近线计算时，建议创建跨源连接，用于访问CloudTable的数据源，提高读写性能。

### 前提条件

- 已开通计算引擎DLI服务，创建集群并完成资源绑定。

### 创建跨源链接

DLI集群绑定完成后，创建跨源链接。步骤如下：

1. 登录RES管理控制台，在“资源中心”页面，选择计算引擎DLI进入列表。
2. 单击DLI集群名称，进入“跨源连接”页面。
3. 在“跨源连接”页面，单击“创建跨源连接”，输入跨源连接名称、已创建并绑定的DLI资源名称和CloudTable资源名称，如图8-7所示。

### 📖 说明

跨源连接名称只能包含数字、英文字母、下划线和中划线。不能为空。输入长度不能超过64个字符。

图 8-7 创建跨源连接

### 创建跨源连接

\* 名称

\* DLI名称

\* CloudTable名称

- 单击“确定”，完成连接创建。
  - 在“跨源连接”页面，可单击操作列的“删除”，删除不需要的连接。
  - 当连接状态为“创建中”时，连接不可删除。

### 8.3.4 开启公共终端节点

推荐系统在需要使用CloudTable集群进行数据存储，CloudTable集群创建后默认没有打开APIG，需开启之后推荐才能正常使用CloudTable集群。

#### 前提条件

- 已开通CloudTable服务，创建集群并完成资源绑定。
- 公共终端节点默认不开启。

#### 开启公共终端节点

- 登录RES管理控制台，在“资源中心”页面，选择CloudTable进入列表。
- 单击CloudTable集群名称操作列的“开启公共终端节点”，待页面提示“开启公共终端节点成功”即可正常使用集群。

#### 📖 说明

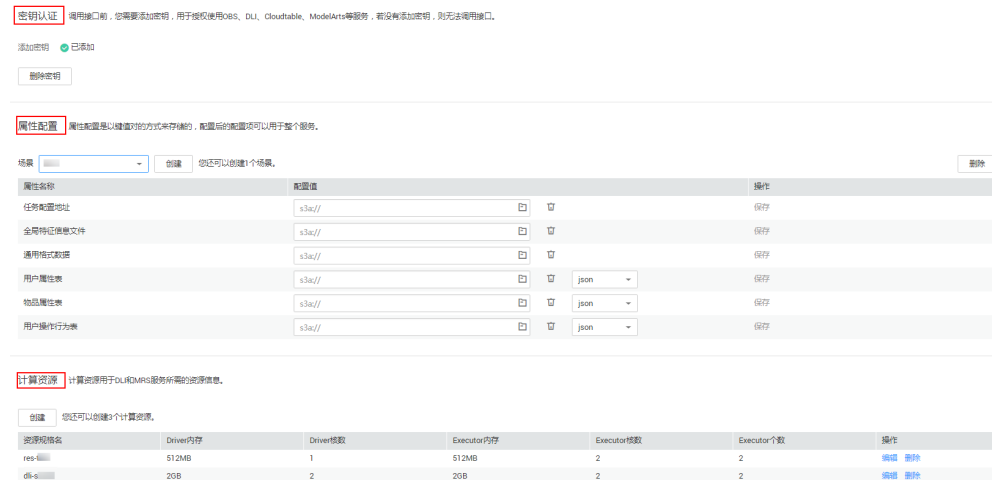
当您不再使用该集群，请先进行解绑，然后在服务对应的管理控制台删除集群。

## 8.4 全局配置

### 8.4.1 全局配置简介

全局配置界面包括密钥认证、属性配置和计算资源三部分，如图8-8所示。

图 8-8 全局配置界面



## 全局配置功能说明

表 8-2 功能说明

功能	说明	详细指导
密钥认证	在使用RES之前，需要获取添加访问密钥，用于授权使用OBS、DLI、Cloudtable、ModelArts等服务，如果没有添加访问密钥，则无法调用接口。	<a href="#">获取访问秘钥并添加RES全局配置</a>
属性配置	以键值对的方式进行存储，配置后的配置项可以用于整个服务。	<a href="#">管理属性配置</a>
计算资源	计算资源包括DLI服务，添加的计算资源需根据业务数据情况进行合理分配。	<a href="#">管理计算资源</a>

### 8.4.2 管理属性配置

属性配置以键值对的方式进行存储，配置后的配置项可以用于整个服务。该配置项的值为用户指定的OBS路径，OBS相关操作请参见[创建OBS桶](#)。

#### 前提条件

- 已开通OBS服务，并按照[推荐系统OBS文件夹规范](#)建议格式创建文件夹。
- 已上传离线数据。

#### 创建场景

您可以根据自己的业务创建场景进行配置，最多可以创建10个场景。具体操作如下：

1. 登录RES管理控制台，在“全局配置”页面的属性配置区域，您可以直接在默认场景进行配置。也可以单击场景右侧的“创建”，进入“创建场景”页面。
2. 输入场景名称单击“确定”，即完成场景创建。

3. 在场景下拉列表中选择目标场景进行配置，配置完毕单击属性操作列的“保存”。
  - “任务配置地址”：用于存放创建作业时自动生成的JSON格式的配置源文件存储路径。
  - “全局特征信息文件”：根据[全局特征信息文件](#)规范准备并上传的全局特征信息文件路径。
  - “通用格式数据”：经过特征工程处理的宽表路径。
  - “用户属性表”：离线数据用户属性表的OBS路径。
  - “物品属性表”：离线数据物品属性表的OBS路径。
  - “用户操作行为表”：离线数据用户操作行为表的OBS路径。

图 8-9 属性配置



## 删除场景

1. 登录RES管理控制台，在“全局配置”页面的属性配置区域，单击场景右侧的“删除”，进入“场景列表”页面。
2. 单击目标场景操作列的“删除”，即完成场景删除。

### 说明

- 场景删除操作无法恢复，请谨慎操作。场景删除后，OBS路径下的文件依旧保留。
- 默认场景不支持删除。

## 8.4.3 管理计算资源

计算资源DLI服务，添加的计算资源需根据实情情况进行合理分配。开通DLI集群大小就是根据业务需求对集群进行扩容缩容，计算资源最多设置5个。计算资源支持的操作包括：

- [创建计算资源](#)
- [编辑计算资源](#)
- [删除计算资源](#)

### 创建计算资源

1. 登录RES管理控制台，在“全局配置”页面的计算资源区域，单击“创建”添加计算资源。
2. 在“创建计算资源”页面，填写计算资源相关配置，请参见[表8-3](#)。配置资源请勿超过集群资源最大限制，配置过大会导致计算失败。



表 8-3 添加资源参数说明

参数名称	说明
资源名	资源名最大长度为10字符，只能由字母，数字，下划线，中划线组成。
Driver内存	Driver内存大于等于512MB。 ● DLI：默认7GB。
Driver核数	Driver核数，大于等于1。 ● DLI默认2。
Executor内存	Executor内存大于等于1GB。 ● DLI：默认8GB。
Executor核数	Executor核数，大于等于1。 ● DLI默认2。
Executor个数	个数，大于等于1。 ● DLI默认7。

3. 完成该项配置后，单击“确定”，完成创建计算资源。

## 编辑计算资源

1. 登录RES管理控制台，在“全局配置”页面的计算资源区域，单击目标资源规格操作列的“编辑”修改计算资源。
2. 在“编辑计算资源”页面，修改计算资源相关配置，请参见表8-3。计算资源名称不支持修改。
3. 单击“确定”，完成计算资源修改。配置资源不能超过集群资源最大限制，否则会导致计算失败。

## 删除计算资源

1. 登录RES管理控制台，在“全局配置”页面的计算资源区域，单击目标资源规格操作列的“删除”。
2. 在“删除”页面，确认删除信息后，单击“确定”，完成计算资源删除。

### 说明

- 计算资源删除操作无法恢复，请谨慎操作。
- 删除计算资源不影响已经计算成功的作业结果。

## 8.5 离线作业

### 8.5.1 离线作业简介

RES提供了离线作业训练的功能，方便您根据业务需求查看作业结果并不断调整您的作业参数。您还可以基于离线作业得到的推荐候选集，用于在线服务计算得到推荐结果。RES提供了多种推荐离线作业功能，您可以直接使用得到满意的推荐候选集。

用户通过数据质量作业对离线数据进行质量检测，然后将检测合格的数据通过特征工程处理为可用于召回策略、过滤规则、排序策略、近线作业的数据。通过上述离线作业训练出可用于在线服务的推荐候选集。当在线服务运行完成，您可以通过效果评估检测推荐结果。

## 离线作业功能说明

表 8-4 功能说明

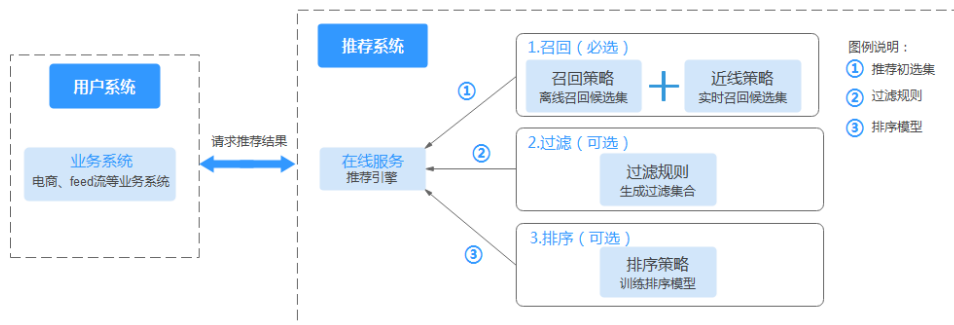
离线作业	说明	详细指导
组合作业	组合作业作为一个包含多个子任务的作业，通常用于多个召回、过滤、排序等任务。	<a href="#">组合作业</a>
数据质量	数据质量是用户在进行离线计算之前使用原始初始格式数据或者通用格式数据检测输入数据是否合法。	<a href="#">数据质量</a>
特征工程	特征工程常用于抽取用户、物品的特征和特定算法的特征生成，一般作为某些算法的前置输入条件。	<a href="#">特征工程</a>
召回策略	召回策略用于生成推荐的候选集，在原始数据中通过算法和规则匹配用户的候选集。	<a href="#">召回策略</a>
排序策略	排序策略根据不同的算法模型对召回策略或者近线策略生成的候选集进行重排序，得到推荐候选集列表。	<a href="#">排序策略</a>
过滤规则	过滤规则用于生成推荐的过滤集，包含黑白名单、历史行为过滤等特性。支持用户在线上推理过程中完成对相关物品的过滤。	<a href="#">过滤规则</a>
效果评估	根据用户行为的数据回流，提供多维度的效果统计。	<a href="#">效果评估</a>

### 8.5.2 组合作业

#### 创建组合作业

通过创建组合作业，用户可以根据配置的策略规则进行离线计算得到不同策略的候选集ID，来进行在线流程计算，得到用户满意的推荐结果。组合作业具体实现请参见图 8-10。

图 8-10 组合作业



创建组合作业主要包括如下设置：

- [基本配置](#)
- [资源选择](#)
- [召回策略](#)
- [过滤规则](#)
- [排序策略](#)
- [预览配置](#)

## 基本配置

基本配置主要包括设置组合作业的名称和描述。

1. 登录RES管理控制台，在左侧导航栏的“离线作业”下，单击“组合作业”进入组合作业页面。
2. 在“组合作业”页面上方，单击“创建”，进入“创建组合作业”页面。
3. 填写“名称”、“场景”和“描述”，其中带\*标志的参数为必填参数。
  - 组合作业名称请以“Standard-”开始，只能由字母、数字、中划线和下划线组成，并且长度小于64个字符。
  - “场景”信息可选择您在全局配置页面创建的场景。
4. 完成该项配置后，单击“下一步”。



## 资源选择

您在使用RES时需要选择计算引擎、存储平台和数据源。计算引擎对数据进行计算，存储平台将处理的数据进行存储。其选择的服务资源即为“资源中心”绑定的资源。

1. 在RES管理控制台完成“基本配置”之后，进入“资源选择”页签。
2. 填写组合作业相关配置参数，请参见[表8-5](#)。

表 8-5 创建组合作业参数说明

模块	参数名称	说明
计算引擎	服务名	计算引擎用于推荐系统的离线计算和近线计算。默认DLI。

模块	参数名称	说明
	集群名称	选择“资源中心”绑定的DLI集群名称。
	任务配置地址	在创建作业时，会自动生成一个JSON格式的配置源文件，该文件存储在指定的OBS路径中，计算引擎可以通过读取配置源文件来进行离线计算。
	资源名	指定DLI运行作业的资源规格。 可选择“全局配置”添加的计算资源或默认值。
存储平台	服务名称	CloudTable作为存储平台，用于用户推荐在线数据和推荐候选集的存储。此处选择已经完成资源绑定的CloudTable。
	集群名称	选择“资源中心”绑定的CloudTable集群名称。
	表名	存储的表格名称。  您可以单击  设置数据版本。RES的数据版本有两种，“V1”版本即数据按照原有格式存储，未做过分区处理。“V2”版本则会依照用户的分区设置做分区处理，当分区合理时，数据将均匀分布在各个节点，有效利用Cloudtable的高并发特性，提升读写效率。其中“预分区数量”和“索引分区数量”可以根据数据量进行设置，如果读写性能达不到要求，可以增加Cloudtable的RS单元数量提升性能。
数据源	初始格式	<ol style="list-style-type: none"> <li>选择提前已经存储在OBS上的如下数据源： <ul style="list-style-type: none"> <li>用户属性表</li> <li>物品属性表</li> <li>用户操作行为表</li> </ul>                     如上数据表的数据格式规范请参见<a href="#">离线数据源</a>。                 </li> <li>在对应表的“数据源”列中，单击  选择数据的数据源OBS存储路径。</li> <li>在对应表的“数据格式”列中，数据格式可选：csv/json。 当选择数据格式为csv时，在弹框中设置数据参数，具体参数如下： <ul style="list-style-type: none"> <li>表头，有或无，根据用户数据格式选取。</li> <li>分隔符，选择逗号(,)、竖线( )、制表符(\t)和自定义。</li> <li>引用字符，单引号(')、双引号(")和自定义。</li> <li>转义字符，反斜杠(\)和自定义。</li> </ul> </li> <li>在对应表的“操作”列中，单击“清除数据”可以删除对应表的数据源。</li> </ol>

模块	参数名称	说明
	通用格式	<p>通用数据由特征工程“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”算子生成。其路径与“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”结果保存路径一致。</p> <p><b>说明</b></p> <p>在使用通用格式数据之前，需要先进行特征工程算子计算。</p> <ol style="list-style-type: none"><li>通用格式数据：从用户属性表、物品属性表和用户操作行为表中提取用户、物品特征和用户行为，并生成JSON数据，即内部通用格式。</li><li>通用格式时间：用户行为数据时间范围，可只有起始时间、结束时间或为空。</li></ol>

- 完成该项配置后，单击“下一步”。

## 召回策略

您可以根据业务需要，选择合适的召回策略。召回策略用于配置离线计算逻辑，通过启动离线计算任务进行候选推荐结果集的生成。

各个召回策略的详细参数设置和输入输出请单击下方链接查看。

- [基于特定行为热度推荐](#)
- [基于综合行为热度推荐](#)
- [基于物品的协同过滤推荐](#)
- [基于用户的协同过滤推荐](#)
- [基于交替最小二乘的矩阵分解推荐](#)
- [业务规则-基于历史行为记忆生成候选集](#)
- [业务规则-人工导入](#)
- [基于属性匹配的召回策略](#)

配置召回策略操作步骤如下：

- 在“创建组合作业”页面，配置完资源选择参数之后，进入“召回策略”页签，单击“添加召回策略”，根据业务需要在下拉框中选择一个合适的策略，如图8-11所示。

图 8-11 召回策略



2. （可选）在目标召回策略右侧，单击“查看输入输出”，可以查看输入数据和输出数据。
3. （可选）在目标召回策略右侧，单击“设置参数”，可以对所选策略进行参数设置。
4. 具体策略的参数说明可单击上方策略名称进行查看，策略设置完成后，单击“下一步”。

## 过滤规则

过滤规则是用于配置候选集的过滤方式，使相关内容不进入候选集。过滤规则生成的候选集UUID用于最终结果推荐。

配置过滤规则操作步骤如下：

1. 在“创建组合作业”页面，配置完召回策略参数之后，进入“过滤规则”页签。
2. 填写业务相关配置，详情请参见[过滤规则](#)。
3. 过滤规则设置完成后，单击“下一步”。

## 排序策略

排序策略是用于对召回策略生成的候选集进行二次排序。如果使用排序策略，需在[特征工程](#)中提交排序样本预处理，创建完成后才可以正常使用排序策略。

各个策略的详细参数设置和输入输出请单击下方链接查看。

- [逻辑斯蒂回归-LR](#)
- [因子分解机-FM](#)
- [域感知因子分解机-FFM](#)
- [深度网络因子分解机-DeepFM](#)
- [核函数特征交互神经网络-PIN](#)

1. 在“创建组合作业”页面，配置完过滤规则参数之后，进入“排序策略”页签，如[图8-12](#)所示。用户可以根据业务需要在“添加排序策略”下拉框中选择一个合适的排序策略，请参见[排序策略](#)对所选排序策略进行参数设置。

图 8-12 排序策略



2. （可选）在目标排序策略，单击“删除”，可以删除对应排序策略。
3. 具体策略的参数说明可单击上方策略名称进行查看，策略设置完成之后，单击“下一步”进入“预览设置”页面。

## 预览配置

在“预览设置”页面，展示了创建作业的名称、资源信息、召回策略信息、过滤规则和排序规则详细信息。

预览结束后，单击“完成”，组合作业创建完成。作业一般需要运行一段时间，根据您的数据量和资源不同，训练时间将耗时几分钟到几十分钟不等。

您可以前往组合作业列表，查看作业的基本情况。在组合作业列表中，刚创建的离线作业“状态”为“计算中”，当离线作业的“状态”变为“计算成功”时，表示作业运行结束，生成的候选集ID将使用于在线服务，为用户生成推荐列表。当离线作业的“状态”变为“计算失败”时，您可以单击作业的名称，进入详情页面，通过查看日志等手段处理问题。

## 8.5.3 数据质量

数据质量检测算子，是用户在进行离线计算之前使用原始初始格式数据（[离线数据源](#)中的离线数据）或者通用格式数据检测输入数据是否合法。包括离线数据中是否包含特殊字符，数据类型是否正确，是否缺少必备信息等。

### 前提条件

- 已将离线数据上传至OBS桶中。

### 创建数据质量作业

创建数据质量操作步骤如下：

1. 在“离线作业”下，单击“数据质量”页签，单击上方“创建”，进入“创建数据质量”页面。
2. 在“创建数据质量”页面，单击“添加数据质量算子”选择“数据质量检测算子”。
  - a. 填写基本信息并选择计算引擎。您可以根据实际情况填写“名称”、“场景”和“描述”信息。其中“场景”信息可选择您在全局配置页面创建的场景。
  - b. 填写参数配置，参数说明请参见[表8-6](#)。

图 8-13 创建数据质量检测作业

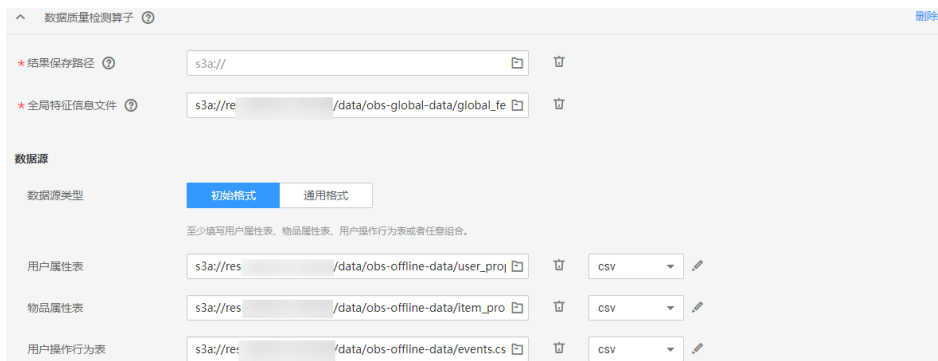


表 8-6 数据质量检测算子参数说明

参数名称	说明
结果保存路径	数据质量检测日志的保存路径。包括错误数据输出及定位等。
全局特征信息文件	<p>用户在使用数据质量检测算子之前，需要提供一份全局的特征信息文件，后续的特征工程、排序算法、在线服务都会用到该文件。全局特征信息文件需要和画像中字段一致，其中 BASIC_INFO 为画像表中定义的基本属性字段，TAGS 为画像表中定义的带权重的标签，Context 为上下文属性。该文件用于说明数据字段信息，以便推荐系统识别用户离线数据，通过特征工程将对应的数据写入到画像中，同时用于排序训练和线上推理服务中使用。</p> <p>文件数据信息请参见<a href="#">全局特征信息文件数据格式</a>。</p> <p>当上传的数据中的特征有变化时，用户需要同步更新该文件。该文件为 JSON 格式，包含特征名、特征大类、特征值类型。</p>
数据源	<p>数据在 OBS 的存放路径。</p> <p>“初始格式”：包括离线数据准备的用户属性表、物品属性表和用户操作行为表。</p> <p>“通用格式”：通用格式数据为经过特征工程作业生成的数据。</p>

- （可选）单击页面右侧“删除”，可以删除该算子。
- 设置完成后，单击“确定”。数据质量检测作业一般需要运行一段时间，请您耐心等待。您可以前往数据质量作业列表，查看作业的基本情况。在作业列表中，刚创建的离线作业“状态”为“计算中”，当离线作业的“状态”变为“计算成功”时，表示作业运行结束，检测完的数据将使用于离线作业。当离线作业的“状态”变为“计算失败”时，您可以单击作业的名称，进入详情页面，通过查看日志等手段处理问题。

### 8.5.4 特征工程

特征工程可对推荐系统的离线数据进行处理，它包含两个功能：

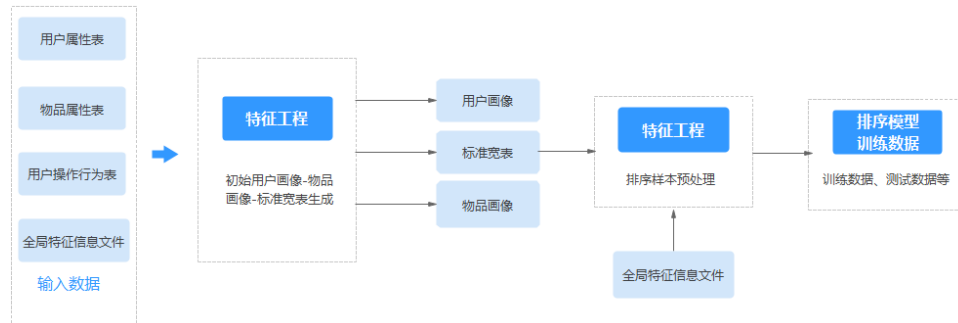


1. 从离线数据中提取用户、物品画像和RES内部通用格式数据；
2. 把RES内部通用格式数据处理成训练排序模型所需的训练数据、测试数据等。

与功能对应，特征工程的两个任务分别是：

- [初始用户画像-物品画像-标准宽表生成](#)
- [排序样本预处理](#)

图 8-14 特征工程



## 创建特征工程

创建特征工程操作步骤如下：


1. 在“离线作业”下，单击“特征工程”页签，单击上方“创建”，进入“创建特征工程”页面。
2. 在“创建特征工程”页面，填写特征工程“名称”、“场景”和“描述”。
  - 特征工程名称：请以“ETL-”开始，只能由字母、数字、中划线和下划线组成，并且长度小于64个字符。
  - “场景”信息可选择您在全局配置页面创建的场景。
3. 设置计算引擎信息，指定“服务名”、“集群名称”、“任务配置地址”、“资源规格”等信息。
4. 单击“添加特征工程”，根据业务需要在下拉框中选择一个合适的策略。[初始用户画像-物品画像-标准宽表生成](#)和[排序样本预处理](#)策略参数说明请单击策略名称进行查看。
5. （可选）在目标策略右侧，单击“删除”，可以删除该策略。
6. 具体策略的参数说明可单击上方策略名称进行查看，策略设置完成后，单击“确定”。作业一般需要运行一段时间，根据您的数据量和资源不同，将耗时几分钟到几十分钟不等。

您可以前往特征工程列表，查看作业的基本情况。在作业列表中，刚创建的作业“状态”为“计算中”，当作业“状态”变为“计算成功”时，表示作业运行结束，可以将经过处理的数据应用于离线作业。当作业“状态”变为“计算失败”时，您可以单击作业的名称，进入详情页面，通过查看日志等手段处理问题。

## 初始用户画像-物品画像-标准宽表生成

初始用户画像-物品画像-标准宽表生成，是将初始格式数据(离线数据)处理成用户画像、物品画像以及内部通用格式数据。



表 8-7 初始用户画像-物品画像-标准宽表生成参数说明

参数名称	说明
数据源	数据在OBS的存放路径。包括用户属性表、物品属性表、用户操作行为表。
全局特征信息文件	<p>用户在使用特征工程之前，需要提供一份全局的特征信息文件，后续的特征工程、在线模块都会用到该文件。</p> <p>文件数据信息请参见<a href="#">全局特征信息文件</a>。</p> <p>当上传的数据中的特征有变化时，用户需要同步更新该文件。该文件为JSON格式，包含特征名、特征大类、特征值类型。</p>
保留已有宽表	<p>对结果保存路径中已有宽表数据的保留方式：</p> <ul style="list-style-type: none"> <li>否，不保留任何已有的数据。</li> <li>是，保留全部已有的数据。</li> <li>覆盖，将相同日期下的数据覆盖掉，保留不同日期下的数据。</li> </ul>
结果保存路径	<p>行为-用户-物品（通用格式）的保存路径。</p> <p><b>说明</b> 使用初始用户画像-物品画像-标准宽表生成的数据时，其路径具体到文件夹即可。</p>
结果存储平台	<p>存储类型有用户画像表和物品画像表两种。</p> <ul style="list-style-type: none"> <li>用户画像表：用户自定义CloudTable集群名称和表名，用于存储生成的用户数据。</li> <li>物品画像表：用户自定义CloudTable集群名称和表名，用于存储生成的物品数据。</li> <li>设置数据版本：您可以单击  设置数据版本。RES的数据版本有两种，“V1”版本即数据按照原有格式存储，未做过分区处理。“V2”版本则会依照用户的分区设置做分区处理，当分区合理时，数据将均匀分布在各个节点，有效利用Cloudtable的高并发特性，提升读写效率。其中“预分区数量”和“索引分区数量”可以根据数据量进行设置，如果读写性能达不到要求，可以增加Cloudtable的RS单元数量提升性能。</li> </ul>


## 排序样本预处理

将内部通用格式数据处理成排序策略所要求的特定格式数据，同时进行离散化，数据编码等特征工程处理。

表 8-8 排序样本预处理输入参数说明

参数名称	说明
全局特征信息文件	<p>用户在使用特征工程之前，需要提供一份全局的特征信息文件，后续的特征工程、排序算法、在线模块都会用到该文件。文件数据信息请参见<a href="#">全局特征信息文件</a>。</p> <p>当上传的数据中的特征有变化时，用户需要同步更新该文件。该文件为JSON格式，包含特征名、特征大类、特征值类型。</p>
通用格式数据源根路径	<p>通用格式生成所在的根目录，即用户提交初始<a href="#">初始用户画像-物品画像-标准宽表生成</a>或<a href="#">基于行为数据的用户画像更新</a>时所提供的结果保存路径。</p>
行为起止日期	<p>用户行为数据时间范围，可只有起始时间、结束时间或为空。</p>
待提取用户特征	<p>从全局特征信息文件中提取输入的用户特征进行排序模型训练。</p> <p><b>说明</b></p> <p>离散的区间个数不能超过100个，请您根据业务需求合理分配参数值。</p> <p>单击“”，增加用户特征。当“特征值类型”为“单值数值型”时，可选的参数信息如下：</p> <ul style="list-style-type: none"> <li>“等距离散”：根据业务需求限定数值“最小值”、“最大值”和“距离”。例如，根据age进行等距离散，设置年龄最小值为1，最大值为100，离散距离为10。等距离散会按照age将1-10岁，11-20岁等作为一个区间进行离散。</li> <li>“等频离散”：根据业务需求限定数值“最小值”、“最大值”和“频率”。例如，根据weight进行等频离散，设置weight最小值为5，最大值为200，离散频率为200。等频离散会按照weight值的大小进行排序之后，以200个数值为一个区间进行离散。</li> <li>“用户自定义离散”：根据业务需求限定数值“最小值”、“最大值”和“区间名称”。例如，根据age进行离散，设置年龄最小值为1，最大值为3，区间名称为“幼儿”；单击添加区间，最小值为4，最大值为12，区间名称为“少儿”，则按照用户自定义的区间进行离散。</li> <li>“不离散”：（默认）不做归一化，不对数据做处理。归一化，根据业务需求限定数值“最小值”和“最大值”。例如，根据weight进行归一化，设置weight最小值为50，最大值为200。如果给定的数值x在该区间范围内则采用以下计算公式处理：<math>x_{new} = (x - 50) / (200 - 50)</math>，不在区间内的则按异常值处理，如x=80时，<math>x_{new} = 0.2</math>。</li> </ul>

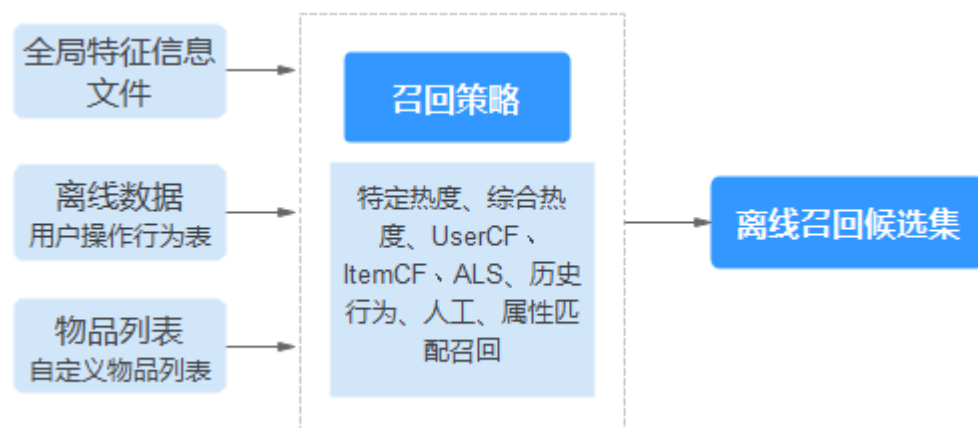
参数名称	说明
待提取物品特征	<p>从全局特征信息文件中提取输入的物品特征进行排序模型训练。</p> <p><b>说明</b> 离散的区间个数不能超过100个，请您根据业务需求合理分配参数值。</p> <p>单击“+”，增加物品特征。当“特征值类型”为“单值数值型”时，可选的参数信息如下：</p> <ul style="list-style-type: none"> <li>“等距离散”：根据业务需求限定数值“最小值”、“最大值”和“距离”。例如，根据order_price进行等距离散，设置价格最小值为1，最大值为100，离散距离为10。那么等距离散会按照价格将1-10元，11-20元等为一个区间进行离散。</li> <li>“等频离散”：根据业务需求限定数值“最小值”、“最大值”和“频率”。例如，根据weight进行等频离散，设置weight最小值为5，最大值为200，离散频率为200。那么等频离散会按照weight值的大小进行排序之后，以200个数值为一个区间进行离散。</li> <li>“用户自定义离散”：根据业务需求限定数值“最小值”、“最大值”和“区间名称”。例如，根据order_price进行离散，设置价格最小值为1，最大值为156，区间名称为“低价区”；单击“+”添加区间，最小值为157，最大值为500，区间名称为“中高价位区”，那么会按照用户自定义的区间进行离散。</li> <li>“不离散”：（默认）不做归一化，不对数据做处理。归一化，根据业务需求限定数值“最小值”和“最大值”。例如，根据weight进行归一化，设置weight最小值为50，最大值为200。如果给定的数值x在该区间范围内则采用以下计算公式处理：<math>x_{new} = (x - 50) / (200 - 50)</math>，不在区间内的则按异常值处理，如x=80时，<math>x_{new} = 0.2</math>。</li> </ul>
正反馈行为类型	<p>用户自定义。行为类型来源于通用格式数据源中“behavior”字段中“actiontype”的值。单击“+”，增加正反馈行为类型。您可以通过+和-来自定义权重。</p>
负反馈行为类型	<p>用户自定义。行为类型来源于通用格式数据源中“behavior”字段中“actiontype”的值。单击“+”，增加负反馈行为类型。您可以通过+和-来自定义权重。</p>
算子类型	<p>排序数据处理算子类型。每一种排序算法都需要进行特定的数据处理，需要根据使用的排序算法来选择排序数据处理类型。LR、FM、FFM、DEEPPM、PIN这五种算法的数据处理互相通用。</p>

参数名称	说明
训练集测试集划分方式	<p>按时间或者比例来划分训练集测试集。</p> <p>“TIME”</p> <ul style="list-style-type: none"> <li>训练数据时间：训练数据起始时间和终止时间，该起始时间和终止时间不得超过行为数据的时间范围。</li> <li>测试数据时间：测试数据起始时间和终止时间，该起始时间和终止时间不得超过行为数据的时间范围。</li> </ul> <p>“RATE”</p> <ul style="list-style-type: none"> <li>训练数据占比：生成的结果中，训练集占整个训练集和测试集的比例，默认0.7。</li> <li>测试数据占比：生成的结果中，训练集占整个训练集和测试集的比例，默认0.3。</li> </ul>
结果保存路径	<p>单击  选择所有输出数据在OBS的保存根路径，会在这个根路径下自动创建feature_map、features_info_online_use、fields_feature_size、test_data、train_data五个文件夹，分别保存特征映射、在线所需特征信息、域特征数量、测试集、训练集这五个文件。</p>

## 8.5.5 召回策略

召回是指对大量的物品做初选，为每一个用户形成个性化候选集。召回策略中内置了多种召回方式，用户可根据自己场景选择。召回策略对应流程请参见图8-15。

图 8-15 召回策略



推荐系统支持的召回方式有：

- 基于特定行为热度推荐
- 基于综合行为热度推荐
- 基于物品的协同过滤推荐
- 基于用户的协同过滤推荐

- [基于交替最小二乘的矩阵分解推荐](#)
- [业务规则-基于历史行为记忆生成候选集](#)
- [业务规则-人工导入](#)
- [基于属性匹配的召回策略](#)
- [基于UCB算法的推荐](#)

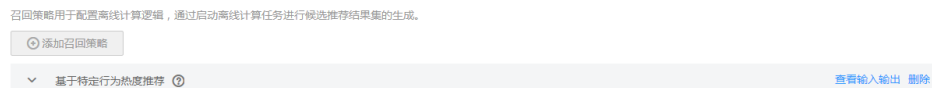
## 创建召回策略

在“创建召回策略”页面，您可以对目标数据选择不同策略进行离线计算，得到合适的候选集。

创建召回策略操作步骤如下：

1. 在“离线作业”下，单击“召回策略”页签，单击上方“创建”，进入“创建召回策略”页面。
2. 在“创建召回策略”页面，填写召回策略“名称”、“场景”和“描述”。
  - 召回策略名称：请以“Retrieval-”开始，只能由字母、数字、中划线和下划线组成，并且长度小于64个字符。
  - “场景”信息可选择您在全局配置页面创建的场景。
3. 设置计算引擎和信息，指定“服务名”、“集群名称”、“任务配置地址”、“资源规格”等信息。
4. 单击“添加召回策略”，根据业务需要在下拉框中选择一个合适的策略。具体策略参数说明请单击上方策略名称查看。

图 8-16 添加召回策略





5. (可选) 在目标召回策略右侧，单击“查看输入输出”，可以查看输入数据和输出数据。
6. (可选) 在目标召回策略右侧，单击“设置参数”，可以对所选策略进行参数设置。
7. (可选) 在目标召回策略右侧，单击“删除”，可以删除该策略。
8. 策略设置完成后，单击“确定”。作业一般需要运行一段时间，根据您的数据量和资源不同，训练时间将耗时几分钟到几十分钟不等，请您耐心等待。

您可以前往召回策略列表，查看作业的基本情况。在作业列表中，刚创建的作业“状态”为“计算中”，当作业“状态”变为“计算成功”时，表示作业运行结束，生成的候选集ID将使用于在线服务，为用户生成推荐列表。当作业“状态”变为“计算失败”时，您可以单击作业的名称，进入详情页面，通过查看日志等手段处理问题。

## 基于特定行为热度推荐

基于特定行为热度推荐通过统计用户操作行为表中物品发生某行为的次数，按照次数从高到低返回物品列表。比如可以统计物品被购买的次数，而返回被购买次数最高的若干个物品。

表 8-9 参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认召回策略的名称
数据源	<p>数据源类型有初始格式和通用格式2种可选。</p> <p>初始格式</p> <ul style="list-style-type: none"> <li>用户操作行为表：单击  选择OBS中存储的用户操作行为表。当选择数据格式为csv时，根据情况单击  设置数据参数。</li> </ul> <p>通用格式</p> <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<b>初始用户画像-物品画像-标准宽表生成</b>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成json数据，即内部通用格式。</li> </ul>	默认选择初始格式
时间选择	<p>时间选择包括数据时间和行为时间跨度。</p> <ul style="list-style-type: none"> <li>数据时间：用于匹配在起始时间和终止时间内的行为数据。</li> <li>行为时间跨度：指定历史行为时间段，选取数据中最靠后的时间往前N天的行为数据计算用户偏好。建议至少设置30天。</li> </ul>	默认选择数据时间的当月所有数据
行为类型	<p>在特定行为热度策略中使用，选择一种行为类型后基于该种行为热度进行推荐。</p> <p>用户行为包含如下几种类型：</p> <ul style="list-style-type: none"> <li>view：物品曝光</li> <li>click：用户点击物品</li> <li>collect：用户收藏了某个物品</li> <li>uncollect：用户取消收藏某个物品</li> <li>search_click：用户点击搜索结果中的物品</li> <li>comment：用户对物品的评论</li> <li>share：分享</li> <li>like：点赞</li> <li>dislike：点衰</li> <li>grade：评分</li> <li>consume：消费</li> <li>use：观看视频/听音乐/阅读。</li> </ul>	物品曝光

参数名称	说明	默认值
行为次数统计方法	统计物品记录数的方式。 <ul style="list-style-type: none"> <li>pv: page view, 即页面浏览量或点击量。</li> <li>uv: unique visitor, 即同一个访客只记录一条数据。</li> </ul>	默认pv
排序算法类型	物品热度是否随着时间衰减。 <ul style="list-style-type: none"> <li>不衰减: normal。</li> <li>衰减: time。</li> </ul>	不衰减
最大推荐结果数	最多生成多少个推荐结果。 该参数会运用在“输出数据”的推荐候选集上。	100

表 8-10 输入数据

数据	是否必须	描述
USER_BEHAVIOR	是	用户操作行为表。

表 8-11 输出数据

名称	类型	描述
HOT_REC_OFFLINE_CANDIDATES	CANDIDATES_SET	特定行为热度推荐算法生成的候选集。





## 基于综合行为热度推荐




基于综合行为热度推荐与基于特定行为热度推荐的召回原理大体相同，不同点是用统计几种行为的加权热度代替了特定行为热度。

表 8-12 参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认召回策略的名称



参数名称	说明	默认值
扩展功能	<p>组合作业的资源选择数据源类型为“通用格式”时，会出现扩展功能选项。</p> <p>创建“基于综合行为热度推荐”的召回策略时，当开启扩展功能，数据源类型仅支持通用格式。</p> <p>该功能表名是否使用分组推荐功能，根据全局特征信息文件勾选的用户特征或物品特征进行分组。</p> <ul style="list-style-type: none"> <li>扩展功能开启: </li> <li>扩展功能关闭: </li> </ul>	默认关闭
数据源	<p>数据源类型有初始格式和通用格式2种可选。</p> <p>初始格式</p> <ul style="list-style-type: none"> <li>用户操作行为表: 初始单击  选择OBS中存储的用户操作行为表。当选择数据格式为csv时，根据情况单击  设置数据参数。</li> </ul> <p>通用格式</p> <ul style="list-style-type: none"> <li>通用格式数据: 特征工程“<b>初始用户画像-物品画像-标准宽表生成</b>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成json数据，即内部通用格式。</li> </ul>	默认选择初始格式
时间选择	<p>时间选择包括数据时间和行为时间跨度。</p> <ul style="list-style-type: none"> <li>数据时间: 用于匹配在起始时间和终止时间内的行为数据。</li> <li>行为时间跨度: 指定历史行为时间段，选取数据中最靠后的时间往前N天的行为数据计算用户偏好。建议至少设置30天。</li> </ul>	默认选择数据时间的当月所有数据

参数名称	说明	默认值
行为权重	<p>当用户行为信息中存在多种行为类型时, 可通过指定行为的权重, 来统一量化行为类型对应的评分。</p> <p>用户行为权重包含:</p> <ul style="list-style-type: none"> <li>• view: 物品曝光</li> <li>• click: 用户点击物品</li> <li>• collect: 用户收藏了某个物品</li> <li>• uncollect: 用户取消收藏某个物品</li> <li>• search_click: 用户点击搜索结果中的物品</li> <li>• comment: 用户对物品的评论</li> <li>• share: 分享</li> <li>• like: 点赞</li> <li>• dislike: 点衰</li> <li>• grade: 评分</li> <li>• consume: 消费</li> <li>• use: 观看视频/听音乐/阅读</li> </ul> <p>您可以单击“增加行为权重”, 新增一个行为权重。通过 <b>+</b> 和 <b>-</b> 来自定义权重大小。</p> <p>单击  可以删除对应行的行为权重。</p>	物品曝光 1.0
全局特征信息文件	<p>开启扩展功能时, 此配置项生效。其中“数据源类型”您可以选择“用户特征”或者“物品特征”进行设置。</p> <ul style="list-style-type: none"> <li>• 全局特征信息文件: 单击  选择该文件的 OBS 存储路径, 当涉及更新时, 单击  更新全局特征信息文件, 并重新设置依赖此文件的配置项。</li> <li>• 用户特征: 从下拉选项中 <input checked="" type="checkbox"/> 勾选全局特征信息文件中提取的用户特征用于进行分组推荐。</li> <li>• 物品特征: 从下拉选项中 <input checked="" type="checkbox"/> 勾选全局特征信息文件中提取的物品特征用于进行分组推荐。</li> </ul>	-
行为次数统计方法	<p>统计物品记录数的方式。</p> <ul style="list-style-type: none"> <li>• pv: page view, 即页面浏览量或点击量。</li> <li>• uv: unique visitor, 即同一个访客只记录一条数据。</li> </ul>	默认pv

参数名称	说明	默认值
排序算法类型	物品热度是否随着时间衰减。 <ul style="list-style-type: none"> <li>不衰减: normal。</li> <li>衰减: time。</li> </ul>	不衰减
最大推荐结果数	最多生成多少个推荐结果。 该参数会运用在“输出数据”的推荐候选集上。	100

表 8-13 输入数据

数据	是否必须	描述
USER_BEHAVIOR	是	用户操作行为表。

表 8-14 输出数据






名称	类型	描述
WEIGHT_BEHAVIORS_REC_OF FLINE_CANDIDATES	CANDIDATES_SET	综合行为热度推荐算法生成的候选集。

## 基于物品的协同过滤推荐

采用经典推荐算法基于物品的协同过滤ItemCF进行召回。

表 8-15 参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认召回策略的名称

参数名称	说明	默认值
数据源	<p>数据源类型有初始格式和通用格式2种可选。</p> <p><b>初始格式</b></p> <ul style="list-style-type: none"> <li>用户操作行为表：单击  选择OBS中存储的用户操作行为表。当选择数据格式为csv时，根据情况单击  设置数据参数。</li> </ul> <p><b>通用格式</b></p> <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成json数据，即内部通用格式。</li> <li>通用格式时间：用户行为数据时间范围，可只有起始时间、结束时间或为空。</li> </ul>	默认选择初始格式
行为时间跨度 (天)	指定历史行为时间段，选取数据中最靠后的时间往前N天的行为数据计算用户偏好。建议至少设置30天。	30天
行为权重	<p>当用户行为信息中存在多种行为类型时，可通过指定行为的权重，来统一量化行为类型对应的评分。</p> <p>用户行为权重项包含如下几种类型：</p> <ul style="list-style-type: none"> <li>view：物品曝光</li> <li>click：用户点击物品</li> <li>collect：用户收藏了某个物品</li> <li>search_click：用户点击搜索结果中的物品</li> <li>comment：用户对物品的评论</li> <li>share：分享</li> <li>like：点赞</li> <li>grade：评分</li> <li>consume：消费</li> <li>use：观看视频/听音乐/阅读。</li> </ul> <p>您可以单击“增加行为权重”，新增一个行为权重。通过  和  来自定义权重大小。</p> <p>单击  可以删除对应行的行为权重。</p>	物品曝光； 1.0
物品最近邻域数	在ItemCF算法中使用，指定某个物品的若干个最近邻物品，构成该物品近邻集合。	50

参数名称	说明	默认值
相似性度量方式	<p>计算用户或物品特征表达相似性的方式。在ItemCF算法中，每个物品会基于对它有过行为的用户表示为特征向量，向量每一维是一个&lt;user_id, score&gt;元组，score表示对应user_id的权重，该参数决定了使用何种函数计算物品向量之间的相似度。</p> <p>相似性度量支持如下方式：</p> <ul style="list-style-type: none"> <li>基于余弦夹角的相似性计算，详细信息请参见 <a href="https://en.wikipedia.org/wiki/Cosine_similarity">https://en.wikipedia.org/wiki/Cosine_similarity</a>。</li> </ul>	基于余弦夹角的相似性计算
行数	用于筛选物品最近邻域的条件组合参数。该值越大，筛选条件越严格。	5
条数	用于筛选物品最近邻域的条件组合参数。该值越小，筛选条件越严格。	4
最大推荐结果数	<p>最多生成多少个推荐结果。</p> <p>该参数会运用在“输出数据”的推荐候选集上。</p>	100

表 8-16 输入数据

数据	是否必须	描述
USER_BEHAVIOR	是	用户操作行为表。

表 8-17 输出数据






名称	类型	描述
ITEMCF_REC_OFFLINE_CANDIDATES	CANDIDATES_SET	ItemCF算法生成的用户-物品列表候选集。
ITEMCF_REC_OFFLINE_CANDIDATES	CANDIDATES_SET	ItemCF算法生成的物品-物品列表候选集。

## 基于用户的协同过滤推荐

采用经典算法基于用户的协同过滤(UserCF)进行召回。

表 8-18 参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认召回策略的名称

参数名称	说明	默认值
数据源	<p>数据源类型有初始格式和通用格式2种可选。</p> <p>初始格式</p> <ul style="list-style-type: none"> <li>用户操作行为表：单击  选择OBS中存储的用户操作行为表。当选择数据格式为csv时，根据情况单击  设置数据参数。</li> </ul> <p>通用格式</p> <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成json数据，即内部通用格式。</li> </ul>	默认选择初始格式
行为时间跨度 (天)	指定历史行为时间段，选取数据中最靠后的时间往前N天的行为数据计算用户偏好。建议至少设置30天。	30天
行为权重	<p>当用户行为信息中存在多种行为类型时，可通过指定行为的权重，来统一量化行为类型对应的评分。</p> <p>用户行为权重项包含如下几种类型：</p> <ul style="list-style-type: none"> <li>view：物品曝光</li> <li>click：用户点击物品</li> <li>collect：用户收藏了某个物品</li> <li>search_click：用户点击搜索结果中的物品</li> <li>comment：用户对物品的评论</li> <li>share：分享</li> <li>like：点赞</li> <li>grade：评分</li> <li>consume：消费</li> <li>use：观看视频/听音乐/阅读。</li> </ul> <p>您可以单击“增加行为权重”，新增一个行为权重。通过  和  来自定义权重大小。</p> <p>单击  可以删除对应的行为权重。</p>	物品曝光 1.0
用户最近邻域数	在UserCF算法中使用，指定某个用户的若干个最近邻用户，构成该用户近邻集合。	50
相似性度量方式	请参见 <a href="#">表8-15</a> 中说明。	基于余弦夹角的相似性计算

参数名称	说明	默认值
行数	用于筛选用户最近邻域的条件组合参数。该值越大，筛选条件越严格。	5
条数	用于筛选用户最近邻域的条件组合参数。该值越小，筛选条件越严格。	4
最大推荐结果数	最多生成多少个推荐结果。 该参数会运用在“输出数据”的推荐候选集上。	100

表 8-19 输入数据

数据	是否必须	描述
USER_BEHAVIOR	是	用户操作行为表。

表 8-20 输出数据






名称	类型	描述
USERCF_REC_OFFLINE_CANDIDATES	CANDIDATES_SET	UserCF算法生成的用户-物品列表候选集。

## 基于交替最小二乘的矩阵分解推荐

基于交替最小二乘的矩阵分解推荐：基于用户-物品的行为信息作为原始矩阵，利用 ALS优化算法对原始矩阵进行矩阵分解，分解之后的用户隐向量矩阵和物品隐向量矩阵可以用来生成预估的新的用户-物品评分矩阵，提取出评分最高的若干个物品作为召回结果。

表 8-21 参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认召回策略的名称

参数名称	说明	默认值
数据源	<p>数据源类型有初始格式和通用格式2种可选。</p> <p><b>初始格式</b></p> <ul style="list-style-type: none"> <li>用户操作行为表：单击  选择OBS中存储的用户操作行为表。当选择数据格式为csv时，根据情况单击  设置数据参数。</li> </ul> <p><b>通用格式</b></p> <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<b>初始用户画像-物品画像-标准宽表生成</b>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成json数据，即内部通用格式。</li> </ul>	默认选择初始格式
行为时间跨度 (天)	指定历史行为时间段，选取数据中最靠后的时间往前N天的行为数据计算用户偏好。建议至少设置30天。	30天
行为权重	<p>当用户行为信息中存在多种行为类型时，可通过指定行为的权重，来统一量化行为类型对应的评分。</p> <p>用户行为权重项包含如下几种类型：</p> <ul style="list-style-type: none"> <li>view：物品曝光</li> <li>click：用户点击物品</li> <li>collect：用户收藏了某个物品</li> <li>search_click：用户点击搜索结果中的物品</li> <li>comment：用户对物品的评论</li> <li>share：分享</li> <li>like：点赞</li> <li>grade：评分</li> <li>consume：消费</li> <li>use：观看视频/听音乐/阅读。</li> </ul> <p>您可以单击“增加行为权重”，新增一个行为权重。通过  和  来自定义权重大小。</p> <p>单击  可以删除对应行的行为权重。</p>	物品曝光 1.0
隐向量维度	在ALS算法中使用，指定用户隐向量、物品隐向量的隐含因子的维度大小。如果离线计算失败，建议调小至10以下。	10
最大迭代次数	指定迭代优化的最大迭代次数。如果离线计算失败，建议调小至10以下。	10
正则化系数	在ALS算法中使用，指定正则化系数，作为优化目标中参数项代价的系数，用于避免过拟合现象发生。	0.01



参数名称	说明	默认值
最大推荐结果数	最多生成多少个推荐结果。 该参数会运用在“输出数据”的推荐候选集上。	100

表 8-22 输入数据

数据	是否必须	描述
USER_BEHAVIOR	是	用户操作行为表。

表 8-23 输出数据

名称	类型	描述
ALSCF_REC_OFFLINE_UIREC_CANDIDATES	CANDIDATES_SET	AlsCF算法生成的用户-物品列表候选集。

## 业务规则-基于历史行为记忆生成候选集

业务规则-基于历史行为记忆生成候选集可以从用户历史行为数据中筛选出发生过某些行为的物品(例如, 筛选出曾经点击过5次的物品)。此策略适用于“看了又看”, “买了又买”等推荐场景。

表 8-24 参数说明

参数名称	说明	默认值
策略别名	策略显示名称, 由中文、英文、数字、下划线、空格或者中划线组成, 并且不能以空格开始和结束, 长度为1~60个字符。	默认召回策略的名称



参数名称	说明	默认值
数据源	<p>数据源类型有初始格式和通用格式2种可选。</p> <p>初始格式</p> <ul style="list-style-type: none"> <li>用户操作行为表：单击  选择 OBS 中存储的用户操作行为表。当选择数据格式为 csv 时，根据情况单击  设置数据参数。</li> </ul> <p>通用格式</p> <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成 json 数据，即内部通用格式。</li> </ul>	默认选择初始格式
行为时间跨度 (天)	指定历史行为时间段，选取数据中最靠后的时间往前 N 天的行为数据计算用户偏好。建议至少设置 30 天。	30 天
行为类型	默认物品曝光，配置参数请参见 <a href="#">表 8-9</a> 中说明。	物品曝光；1 次
最大推荐结果数	最多生成多少个推荐结果。 该参数会运用在“输出数据”的推荐候选集上。	100

表 8-25 输入数据

数据	是否必须	描述
USER_BEHAVIOR	是	用户操作行为表。

表 8-26 输出数据

名称	类型	描述
CUSTOMRULE_REC_OFFLINE_UI_REC_CANDIDATES	CANDIDATES_SET	CustomRule 生成的用户-物品列表候选集。

## 业务规则-人工导入

业务规则-人工导入是指用户自定义物品列表，并将此列表排序后作为候选集，以供在线服务调用。

表 8-27 参数说明

参数名称	说明
策略名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。默认召回策略的名称。
OBS地址	用户可从此OBS地址中选择自定义物品列表。

### 输入数据

数据格式(.csv文件):

itemId1,score1

itemId2,score2

itemId为物品id，score为物品分数。score不是必选，如果不选则算法默认给物品分配分数。算法将物品列表根据分数排序后作为候选集

用户可从OBS中选择保存有人工编辑推荐结果的列表（即物品ID）。

表 8-28 输出数据


名称	类型	描述
MANUALRULE_REC_OFFLINE_UIREC_CANDIDATES	CANDIDATES_SET	人工编辑候选集。 例如，127,1

## 基于属性匹配的召回策略

基于属性匹配的召回策略会用用户画像和物品画像的相关属性进行匹配，为用户召回属性匹配程度高的若干个物品。

表 8-29 参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认召回策略的名称。
全局特征信息文件	用户在使用近线策略之前，需要提供全局特征信息文件，该文件中包含了离线数据中用户、物品的所有特征字段及对应数据类型，并且特征类型要始终保持一致。全局特征信息文件示例请参考 <a href="#">全局特征信息文件</a> 。	-






参数名称	说明	默认值
匹配类型	画像匹配类型，包括： <ul style="list-style-type: none"> <li>• 用户匹配物品</li> <li>• 物品匹配用户</li> <li>• 用户自匹配</li> <li>• 物品自匹配</li> </ul>	用户匹配物品
匹配特征对	用户和物品相关联特征。请根据实际情况配置参数，如果属性匹配特征对相似度较高内存不够时需提升配置。您可以单击  进入“添加匹配特征对”页面进行配置。设置特征对的别名、根据全局特征信息文件匹配用户特征名和物品特征名，设置权重。	-
推荐个数	推荐给用户的物品最大个数。	10
行数	该值越大筛选相似项条件越严格。	5
条数	该值越小筛选相似项条件越严格。	4
用户画像存储	该数据来源于 <b>基于用户数据更新用户画像</b> 更新后的用户画像数据或者 <b>初始用户画像-物品画像-标准宽表生成</b> 的用户画像。	-
物品画像存储	该数据来源于 <b>基于物品数据更新物品画像</b> 更新后的物品画像数据或者 <b>初始用户画像-物品画像-标准宽表生成</b> 的物品画像。	-
最大推荐结果数	最多生成多少个推荐结果。 该参数会运用在“输出数据”的推荐候选集上。	100

## 基于 UCB 算法的推荐

基于UCB算法的推荐综合考虑了用户操作行为表中，物品发生的某几种行为类型及次数，然后给每一个物品都计算一个得分，最终返回得分最高的若干个物品。

表 8-30 参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认召回策略的名称

参数名称	说明	默认值
数据源	<p>数据源类型有初始格式和通用格式2种可选。</p> <p>初始格式</p> <ul style="list-style-type: none"> <li>用户操作行为表：单击  选择OBS中存储的用户操作行为表。当选择数据格式为csv时，根据情况单击  设置数据参数。</li> </ul> <p>通用格式</p> <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成json数据，即内部通用格式。</li> </ul>	默认选择初始格式
行为时间跨度 (天)	指定历史行为时间段，选取数据中最靠后的时间往前N天的行为数据计算用户偏好。建议至少设置30天。	30天
行为权重	<p>当用户行为信息中存在多种行为类型时，可通过指定行为的权重，来统一量化行为类型对应的评分。</p> <p>用户行为权重项包含如下几种类型：</p> <ul style="list-style-type: none"> <li>view：物品曝光</li> <li>click：用户点击物品</li> <li>collect：用户收藏了某个物品</li> <li>search_click：用户点击搜索结果中的物品</li> <li>comment：用户对物品的评论</li> <li>share：分享</li> <li>like：点赞</li> <li>grade：评分</li> <li>consume：消费</li> <li>use：观看视频/听音乐/阅读。</li> </ul> <p>您可以单击“增加行为权重”，新增一个行为权重。通过  和  来自定义权重大小。</p> <p>单击  可以删除对应行的行为权重。</p>	物品曝光 1.0
最小行为次数	在物品上产生过行为的最小用户数，其中一个用户在一个物品上只计算一次行为。	30
折中参数	令alpha为Exploration 和 Exploitation之间的折中参数，其取值范围为[0,1]，alpha越趋近于0，则物品的得分对历史得分高的物品越有利，即 Exploitation。反之，alpha越趋近于1，则物品的得分越倾向于探索新物品，即 Exploration。	0.5

参数名称	说明	默认值
最大推荐结果数	最多生成多少个推荐结果。 该参数会运用在“输出数据”的推荐候选集上。	100

表 8-31 输入数据

数据	是否必须	描述
USER_BEHAVIOR	是	用户操作行为表。

表 8-32 输出数据

名称	类型	描述
UCB_REC_OFFLINE_CANDIDATE_S	CANDIDATES_SET	UCB算法生成的候选集。

## 8.5.6 排序策略

### 排序策略简介

排序策略用于训练排序模型，该模型将被用于对召回策略召回的候选集进行排序，以将推荐物品顺序调整到最优。排序模型可对LR、FM、FFM、DeepFM和PIN等模型进行训练，具体包括如下内容：

- [逻辑斯蒂回归-LR](#)
- [因子分解机-FM](#)
- [域感知因子分解机-FFM](#)
- [深度网络因子分解机-DeepFM](#)
- [核函数特征交互神经网络-PIN](#)

### 创建排序策略

创建排序策略操作步骤详情请参见[排序策略](#)。

作业一般需要运行一段时间，根据您的数据量和资源不同，训练时间将耗时几分钟到几十分钟不等。



您可以前往排序策略列表，查看作业的基本情况。在作业列表中，刚创建的作业“状态”为“计算中”，当作业“状态”变为“计算成功”时，表示作业运行结束，生成的候选集ID将使用于在线服务，为用户生成推荐列表。当作业“状态”变为“计算失败”时，您可以单击作业的名称，进入详情页面，通过查看日志等手段处理问题。


### 逻辑斯蒂回归-LR

逻辑斯蒂回归算法是一种广义的线性回归分析模型，常用于数据挖掘、疾病自动诊断、经济预测等领域。逻辑斯蒂回归算法通过在线性回归的基础上叠加一个sigmoid激

活函数将输出值映射到[0,1]之间，是机器学习领域里常用的二分类算法。单击查看[逻辑斯蒂回归](#)详情信息。

表 8-33 逻辑斯蒂回归参数说明

参数名称	说明
计算节点信息	用户可使用的计算资源种类。包括“8核 16GiB”、“8核 32GiB”、“8核 64GiB”、“16核 128GiB”。
训练数据的obs路径	单击  选择特征工程排序样本预处理生成的训练数据所在的OBS路径。 即特征工程“排序样本预处理”结果保存路径下具体的训练文件路径。
测试数据的obs路径	单击  选择特征工程排序样本预处理生成的测试数据所在的OBS路径。 即特征工程“排序样本预处理”结果保存路径下具体的测试文件路径。
特征值数量统计文件	该文件标识了每一个域下的特征数量，排序数据处理接口会生成这个文件，需要用户提供此文件完整路径。文件路径为 <a href="#">特征工程中排序样本预处理</a> 作业输出数据的结果保存路径的“fields_feature_size”目录下文件名称是part-00000开头的文件，需要用户提供文件的OBS路径。
最大迭代轮数	模型训练的最大迭代轮数，默认50。
提前终止训练轮数	在测试集上连续N轮迭代AUC无提高时，迭代停止，训练提前结束，默认5。
初始化方法	模型参数的初始化方法。 <ul style="list-style-type: none"> <li>• normal: 正态分布 平均值: 默认0 标准差: 0.001</li> <li>• uniform: 均匀分布 最小值: 默认-0.001, 均匀分布的最小值, 必须小于最大值。 最大值: 默认0.001, 均匀分布的最大值, 必须大于最小值。</li> <li>• xavier: 初始化初始值为 均值为0, 方差为 <math>\text{Var}(w_i)=1/n_{in}</math> 的均匀分布 (高斯或者随机分布)。其中 <math>n_{in}</math> 是该神经元的输入数目。</li> </ul>

参数名称	说明
优化器类型	<ul style="list-style-type: none"> <li>grad: 梯度下降算法                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>adam: 自适应矩估计算法 结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认1e-8。</li> </ul> </li> <li>adagrad: 自适应梯度算法 对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>ftl: Follow The Regularized Leader 适用于处理超大规模数据的, 含大量稀疏特征的在线学习的常见优化算法。                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>
L2正则项系数	叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"> <li>full: 指针对全量参数计算。</li> <li>batch: 则仅针对当前批数据中出现的参数计算</li> </ul> 说明 batch模式计算速度快于full模式。
保存根路径	单击  选择训练结果在OBS中的保存根路径, 训练完成后, 会将模型和日志文件保存在该路径下。该路径不能包含中文。

## 因子分解机-FM


因子分解机算法是一种基于矩阵分解的机器学习算法, 能够自动进行二阶特征组合、学习特征之间的关系, 无需人工经验干预, 同时能够解决组合特征稀疏的问题。



表 8-34 因子分解机参数说明

参数名称	说明
计算节点信息	用户可使用的计算资源种类。“8核 16GiB”、“8核 32GiB”、“8核 64GiB”、“16核 128GiB”。
训练数据的obs路径	特征工程排序样本预处理生成的训练数据所在的OBS路径。 即特征工程“排序样本预处理”结果保存路径下具体的训练文件路径。
测试数据的obs路径	特征工程排序样本预处理生成的测试数据所在的OBS路径。 即特征工程“排序样本预处理”结果保存路径下具体的测试文件路径。
特征值数量统计文件	该文件标识了每一个域下的特征数量，排序数据处理接口会生成这个文件，需要用户提供此文件完整路径。文件路径为 <a href="#">特征工程中排序样本预处理</a> 作业输出数据的结果保存路径的“fields_feature_size”目录下文件名称是part-00000开头的文件，需要用户提供文件的OBS路径。
最大迭代轮数	模型训练的最大迭代轮数，默认50。
提前终止训练轮数	在测试集上连续N轮迭代AUC无提高时，迭代停止，训练提前结束，默认5。
初始化方法	模型参数的初始化方法。 <ul style="list-style-type: none"><li>• normal: 正态分布 平均值: 默认0 标准差: 0.001</li><li>• uniform: 均匀分布 最小值: 默认-0.001, 均匀分布的最小值, 必须小于最大值。 最大值: 默认0.001, 均匀分布的最大值, 必须大于最小值。</li><li>• xavier: 初始化初始值为 均值为0, 方差为 <math>\text{Var}(w_i)=1/n_{in}</math> 的均匀分布（高斯或者随机分布）。其中 <math>n_{in}</math> 是该神经元的输入数目。</li></ul>

参数名称	说明
优化器类型	<ul style="list-style-type: none"> <li>● grad: 梯度下降算法                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>● adam: 自适应矩估计算法 结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认1e-8。</li> </ul> </li> <li>● adagrad: 自适应梯度算法 对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>● ftrl: Follow The Regularized Leader 适用于处理超大规模数据的, 含大量稀疏特征的在线学习的常见优化算法。                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>
L2正则项系数	叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"> <li>● full: 指针对全量参数计算。</li> <li>● batch: 则仅针对当前批数据中出现的参数计算</li> </ul> <b>说明</b> batch模式计算速度快于full模式。
隐向量长度	分解后的特征向量的长度。默认10。

参数名称	说明
保存根路径	单击  选择训练结果在OBS中的保存根路径，训练完成后，会将模型和日志文件保存在该路径下。该路径不能包含中文。


## 域感知因子分解机-FFM

域感知因子分解机是因子分解机的改进版本，因子分解机每个特征对其他域的隐向量都一致，而域感知因子分解机每个特征对其他每个域都会学习一个隐向量，能够达到更高的精度，但也更容易出现过拟合。单击查看[域感知因子分解机](#)详细信息。

表 8-35 域感知因子分解机参数说明

参数名称	说明
计算节点信息	用户可使用的计算资源种类。“8核 16GiB”、“8核 32GiB”、“8核 64GiB”、“16核 128GiB”。
训练数据的obs路径	特征工程排序样本预处理生成的训练数据所在的OBS路径。即特征工程“排序样本预处理”结果保存路径下具体的训练文件路径。
测试数据的obs路径	特征工程排序样本预处理生成的测试数据所在的OBS路径。即特征工程“排序样本预处理”结果保存路径下具体的测试文件路径。
特征值数量统计文件	该文件标识了每一个域下的特征数量，排序数据处理接口会生成这个文件，需要用户提供此文件完整路径。文件路径为 <a href="#">特征工程中排序样本预处理</a> 作业输出数据的结果保存路径的“fields_feature_size”目录下文件名称是part-00000开头的文件，需要用户提供文件的OBS路径。
最大迭代轮数	模型训练的最大迭代轮数，默认50。
提前终止训练轮数	在测试集上连续N轮迭代AUC无提高时，迭代停止，训练提前结束，默认5。

参数名称	说明
初始化方法	<p>模型参数的初始化方法。</p> <ul style="list-style-type: none"> <li>● normal: 正态分布 平均值: 默认0 标准差: 0.001</li> <li>● uniform : 均匀分布 最小值: 默认-0.001, 均匀分布的最小值, 必须小于最大值。 最大值: 默认0.001, 均匀分布的最大值, 必须大于最小值。</li> <li>● xavier: 初始化初始值为 均值为0, 方差为 <math>\text{Var}(w_i)=1/n_{in}</math> 的均匀分布 (高斯或者随机分布)。其中 <math>n_{in}</math> 是该神经元的输入数目。</li> </ul>
优化器类型	<ul style="list-style-type: none"> <li>● grad: 梯度下降算法 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>● adam: 自适应矩估计算法 结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认<math>1e-8</math>。</li> </ul> </li> <li>● adagrad: 自适应梯度算法 对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>● ftrl: Follow The Regularized Leader 适用于处理超大规模数据的,含大量稀疏特征的在线学习的常见优化算法。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>

参数名称	说明
L2正则项系数	叠加在模型的2范数之上，用来对模型值进行限制防止过拟合。默认0。
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"> <li>full：指针对全量参数计算。</li> <li>batch：则仅针对当前批数据中出现的参数计算</li> </ul> <b>说明</b> batch模式计算速度快于full模式。
隐向量长度	分解后的表示特征的向量的长度。默认10。
保存根路径	单击  选择训练结果在OBS中的保存根路径，训练完成后，会将模型和日志文件保存在该路径下。该路径不能包含中文。

## 深度网络因子分解机-DeepFM


深度网络因子分解机，结合了因子分解机和深度神经网络对于特征表达的学习，同时学习高阶和低阶特征组合，从而达到准确地特征组合学习，进行精准推荐。单击查看[深度网络因子分解机](#)详细信息。

表 8-36 深度网络因子分解机参数说明

参数名称	说明
计算节点信息	用户可使用的计算资源种类。“8核 16GiB”、“8核 32GiB”、“8核 64GiB”、“16核 128GiB”。
训练数据的obs路径	特征工程排序样本预处理生成的训练数据所在的OBS路径。 即特征工程“排序样本预处理”结果保存路径下具体的训练文件路径。
测试数据的obs路径	特征工程排序样本预处理生成的测试数据所在的OBS路径。 即特征工程“排序样本预处理”结果保存路径下具体的测试文件路径。
特征值数量统计文件	该文件标识了每一个域下的特征数量，排序数据处理接口会生成这个文件，需要用户提供此文件完整路径。文件路径为 <a href="#">特征工程中排序样本预处理</a> 作业输出数据的结果保存路径的“fields_feature_size”目录下文件名称是part-00000开头的文件，需要用户提供文件的OBS路径。
最大迭代轮数	模型训练的最大迭代轮数，默认50。
提前终止训练轮数	在测试集上连续N轮迭代AUC无提高时，迭代停止，训练提前结束，默认5。

参数名称	说明
初始化方法	<p>模型参数的初始化方法。</p> <ul style="list-style-type: none"><li>• normal: 正态分布 平均值: 默认0 标准差: 0.001</li><li>• uniform : 均匀分布 最小值: 默认-0.001, 均匀分布的最小值, 必须小于最大值。 最大值: 默认0.001, 均匀分布的最大值, 必须大于最小值。</li><li>• xavier: 初始化初始值为 均值为0, 方差为 <math>\text{Var}(w_i)=1/n_{in}</math> 的均匀分布 (高斯或者随机分布)。其中 <math>n_{in}</math> 是该神经元的输入数目。</li></ul>

参数名称	说明
优化器类型	<ul style="list-style-type: none"> <li>grad: 梯度下降算法                             <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>adam: 自适应矩估计算法                             <p>结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认1e-8。</li> </ul> </li> <li>adagrad: 自适应梯度算法                             <p>对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>ftrl: Follow The Regularized Leader                             <p>适用于处理超大规模数据的, 含大量稀疏特征的在线学习的常见优化算法。</p> <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>
L2正则项系数	叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"> <li>full: 指针对全量参数计算。</li> <li>batch: 则仅针对当前批数据中出现的参数计算</li> </ul> <b>说明</b> batch模式计算速度快于full模式。
隐向量长度	分解后的表示特征的向量的长度。默认10。
神经网络结构	神经网络的层数与每一层的神经元节点个数。默认400,400,400。

参数名称	说明
激活函数	神经网络中的激活函数，将一个（或一组）神经元的值映射为一个输出值。 <ul style="list-style-type: none"> <li>• relu</li> <li>• tanh</li> <li>• sigmoid</li> </ul>
神经元值保留概率	神经网络前向传播过程中以该概率保留神经元的值。默认 0.8。
保存根路径	单击  选择训练结果在 OBS 中的保存根路径，训练完成后，会将模型和日志文件保存在该路径下。该路径不能包含中文。

## 核函数特征交互神经网络-PIN


核函数特征交互神经网络是深度网络因子分解机的改进版本，深度网络因子分解机通过向量点乘来计算特征之间的关系，而核函数特征交互神经网络使用不同的核（kernel）来对特征交互进行建模，以此来计算两个域中特征的相互关系，其中核的种类包括向量内积外积、矩阵乘法、神经网络等。利用核函数建模特征交互，实现了参数共享，减小了模型复杂度。单击查看[核函数特征交互神经网络](#)详细信息。

表 8-37 核函数特征交互神经网络参数说明

参数名称	说明
计算节点信息	用户可使用的计算资源种类。“8核 16GiB”、“8核 32GiB”、“8核 64GiB”、“16核 128GiB”。
训练数据的obs路径	特征工程排序样本预处理生成的训练数据所在的 OBS 路径。即特征工程“排序样本预处理”结果保存路径下具体的训练文件路径。
测试数据的obs路径	特征工程排序样本预处理生成的测试数据所在的 OBS 路径。即特征工程“排序样本预处理”结果保存路径下具体的测试文件路径。
特征值数量统计文件	该文件标识了每一个域下的特征数量，排序数据处理接口会生成这个文件，需要用户提供此文件完整路径。文件路径为 <a href="#">特征工程中排序样本预处理</a> 作业输出数据的结果保存路径的“fields_feature_size”目录下文件名称是 part-00000 开头的文件，需要用户提供文件的 OBS 路径。
最大迭代轮数	模型训练的最大迭代轮数，默认 50。
提前终止训练轮数	在测试集上连续 N 轮迭代 AUC 无提高时，迭代停止，训练提前结束，默认 5。



参数名称	说明
初始化方法	<p>模型参数的初始化方法。</p> <ul style="list-style-type: none"> <li>• normal: 正态分布 平均值: 默认0 标准差: 0.001</li> <li>• uniform: 均匀分布 最小值: 默认-0.001, 均匀分布的最小值, 必须小于最大值。 最大值: 默认0.001, 均匀分布的最大值, 必须大于最小值。</li> <li>• xavier: 初始化初始值为 均值为0, 方差为 <math>\text{Var}(w_i)=1/n_{in}</math> 的均匀分布 (高斯或者随机分布)。其中 <math>n_{in}</math> 是该神经元的输入数目。</li> </ul>
优化器类型	<ul style="list-style-type: none"> <li>• grad: 梯度下降算法 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> </ul> </li> <li>• adam: 自适应矩估计算法 结合AdaGrad和RMSProp两种优化算法的优点, 对梯度的一阶矩估计 (First Moment Estimation, 即梯度的均值) 和二阶矩估计 (Second Moment Estimation, 即梯度的未中心化的方差) 进行综合考虑, 依次计算出更新步长。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 数值稳定常量: 为保证数值稳定而设置的一个微小常量。默认<math>1e-8</math>。</li> </ul> </li> <li>• adagrad: 自适应梯度算法 对每个不同的参数调整不同的学习率, 对频繁变化的参数以更小的步长进行更新, 而稀疏的参数以更大的步长进行更新。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.001。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> </ul> </li> <li>• ftrl: Follow The Regularized Leader 适用于处理超大规模数据的, 含大量稀疏特征的在线学习的常见优化算法。 <ul style="list-style-type: none"> <li>- 学习率: 优化算法的参数, 决定优化器在最优方向上前进步长的参数。默认0.1。</li> <li>- 初始梯度累加和: 梯度累加和用来调整学习步长。默认0.1。</li> <li>- L1正则项系数: 叠加在模型的1范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> <li>- L2正则项系数: 叠加在模型的2范数之上, 用来对模型值进行限制防止过拟合。默认0。</li> </ul> </li> </ul>

参数名称	说明
L2正则项系数	叠加在模型的2范数之上，用来对模型值进行限制防止过拟合。默认0。
正则损失计算方式	正则损失计算当前有两种方式。 <ul style="list-style-type: none"><li>• full: 指针对全量参数计算。</li><li>• batch: 则仅针对当前批数据中出现的参数计算</li></ul> <b>说明</b> batch模式计算速度快于full模式。
隐向量长度	分解后的表示特征的向量的长度。默认10。
神经网络结构	神经网络的层数与每一层的神经元节点个数。默认400,400,400。
激活函数	神经网络中的激活函数，将一个（或一组）神经元的值映射为一个输出值。 <ul style="list-style-type: none"><li>• relu</li><li>• tanh</li><li>• sigmoid</li></ul>
神经元值保留概率	神经网络前向传播过程中以该概率保留神经元的值。默认0.8。
子网络结构	特征向量之间使用神经网络核来计算相互关系时，该神经网络的结构。默认40,5。
是否移除因子分解机	是否移除模型架构中的因子分解机部分，值为true则蜕变为带有核函数的DNN。默认关闭。
保存根路径	单击  选择训练结果在OBS中的保存根路径，训练完成后，会将模型和日志文件保存在该路径下。该路径不能包含中文。

## 8.5.7 过滤规则

过滤规则用于配置候选集的过滤方式，使之不进入候选集。过滤规则说明请参见图 8-17。

图 8-17 过滤规则



说明：黑名单、白名单、自定义物品列表至少选择一个。

## 创建过滤规则

在“创建过滤规则”页面，用户可以对目标数据选择不同策略进行离线计算，得到合适的候选集。

创建过滤规则操作步骤如下：

1. 在“离线作业”下，单击“过滤规则”页签，单击该页面做上方“创建”，进入“创建过滤规则”页面。
2. 在“创建过滤规则”页面，填写特征工程“名称”、“场景”和“描述”。
  - 特征工程名称：请以“Filter-”开始，只能由字母、数字、中划线和下划线组成，并且长度小于64个字符。
  - “场景”信息可选择您在全局配置页面创建的场景。
3. 根据表8-38设置计算引擎和存储平台相关参数，并根据业务需要在“增加历史行为过滤”，如图8-18所示。

图 8-18 创建过滤规则

过滤规则用于配置候选集的过滤方式。

过滤规则别名

黑名单地址

白名单地址

历史行为过滤  (天)    最小次数    最大次数

增加历史行为过滤



行为过滤逻辑  AND  OR

数据源

数据源类型

\* 用户操作行为表

表 8-38 过滤规则参数说明

参数名称	说明
计算引擎	<ul style="list-style-type: none"> <li>• 服务名，计算引擎DLI用于推荐系统的离线计算和近线计算。</li> <li>• 集群名称，选择“资源中心”绑定的DLI集群名称。</li> <li>• 任务配置地址，在创建作业时，会自动生成一个JSON格式的配置文件，该文件存储在指定的OBS路径中，计算引擎可以通过读取配置文件来进行离线计算。</li> <li>• 资源名，指定DLI运行作业的资源规格。</li> </ul>
存储平台	<ul style="list-style-type: none"> <li>• 服务名称，CloudTable作为存储平台，用于用户推荐在线数据和推荐候选集的存储。</li> <li>• 集群名称，选择“资源中心”绑定的CloudTable集群名称。</li> <li>• 表名，存储的表格名称。</li> </ul>
过滤规则别名	自定义过滤规则名称。由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。
黑名单地址	黑名单所在路径。黑名单中的物品不能出现在最终推荐结果集里面。黑名单内容需要存储在OBS上。
白名单地址	白名单所在的路径。白名单之外的物品不应该出现在最终推荐结果集里面。白名单内容需要存储在OBS上。
历史行为过滤	<p>单击  增加历史行为过滤，单击后方的  删除过滤行为。指定与用户个性化的物品候选集过滤准则。例如对于用户过去3天内有过view行为的物品（如新闻）过滤，使之不进入候选集。</p> <p>行为类型包括。</p> <ul style="list-style-type: none"> <li>• view：物品曝光</li> <li>• click：用户点击物品</li> <li>• collect：用户收藏了某个物品</li> <li>• uncollect：用户取消收藏某个物品</li> <li>• search_click：用户点击搜索结果中的物品</li> <li>• comment：用户对物品的评论</li> <li>• share：分享</li> <li>• like：点赞</li> <li>• dislike：点衰</li> <li>• grade：评分</li> <li>• consume：消费</li> <li>• use：观看视频/听音乐/阅读。</li> </ul>
行为过滤逻辑	各个历史行为过滤规则之间的逻辑。例如，“AND”逻辑为在7天内点赞超过3次且在7天内消费超过3次的物品，“OR”逻辑为在7天内点赞超过3次或在7天内消费超过3次的物品。

参数名称	说明
数据源	<p>数据源可选择初始格式或通用格式。</p> <p><b>说明</b> 过滤规则涉及历史行为过滤，则用到用户操作行为表，需要选取目标数据进行过滤。</p> <p>“初始格式” 用户操作行为表：初始数据中的用户操作行为表。</p> <p>“通用格式”</p> <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成 json 数据，即内部通用格式。</li> <li>通用格式时间：用户行为数据时间范围，可只有起始时间、结束时间或为空。</li> </ul>

- 策略参数设置完成后，单击“确定”。

您可以前往过滤规则列表，查看作业的基本情况。在作业列表中，刚创建的作业“状态”为“计算中”，当作业“状态”变为“计算成功”时，表示作业运行结束，生成的候选集ID将使用于在线服务，为用户生成推荐列表。当作业“状态”变为“计算失败”时，您可以单击作业的名称，进入详情页面，通过查看日志等手段处理问题。

## 8.5.8 效果评估

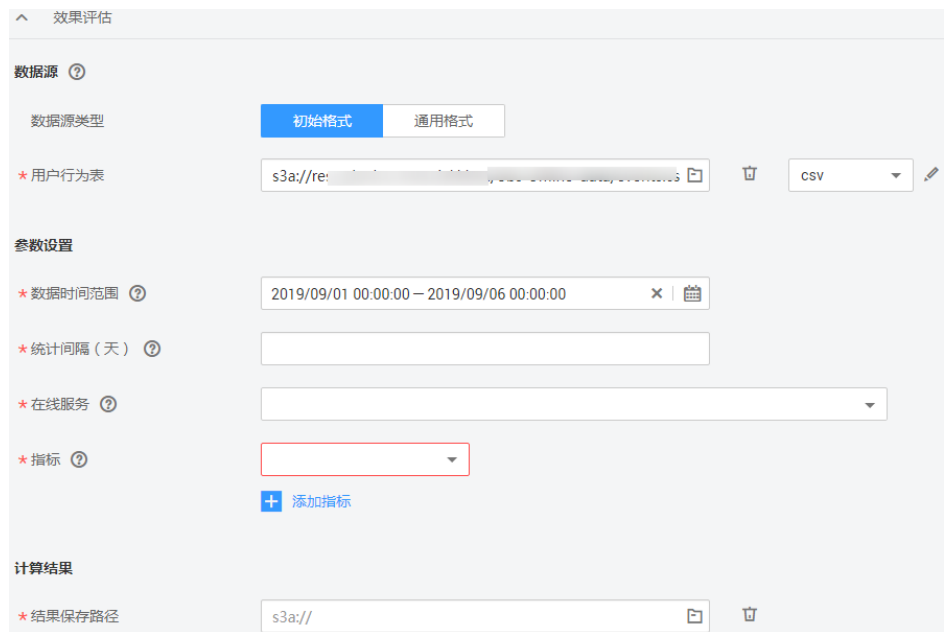
创建效果评估可以对线上服务设置指标，查看推荐效果的反馈，可以根据系统提供的指标添加。

### 创建效果评估

创建效果评估步骤如下：

- 在“离线作业”的“效果评估”页签，单击左上方“创建”，进入“创建效果评估”页面。
- 在“创建效果评估”页面，填写特征工程“名称”、“场景”和“描述”。
  - 效果评估名称：请以“Evaluate-”开始，只能由字母、数字、中划线和下划线组成，并且长度小于64个字符。
  - “场景”信息可选择您在全局配置页面创建的场景。



图 8-19 创建效果评估



3. 配置创建效果评估相关参数，请参见表8-39。

表 8-39 创建效果评估参数说明

参数名称	说明
计算引擎	计算引擎信息，选择指定服务名、集群名称、任务配置地址、资源名等信息。
数据源	数据源类型有初始格式和通用格式可以选择。 “初始格式” <ul style="list-style-type: none"> <li>用户操作行为表：初始数据中的用户操作行为表。</li> </ul> “通用格式” <ul style="list-style-type: none"> <li>通用格式数据：特征工程“<a href="#">初始用户画像-物品画像-标准宽表生成</a>”算子生成的用户推荐系统的数据。从用户特征表、物品特征表以及用户行为表中提取用户、物品特征和用户行为，并生成json数据，即内部通用格式。</li> </ul>
数据时间范围	被统计数据起始时间和终止时间。
统计间隔(天)	统计间隔，以天为单位，每隔多少天计算一次指标，大于0。
在线服务	选择已发布的在线服务进行推荐效果指标计算。
结果保存路径	效果评估结果在OBS的文件输出路径。

参数名称	说明
指标	<p>推荐服务效果评估指标，通过指标后的下拉框选择系统自置指标或自定义指标。您可以单击  添加指标，单击该指标后的  删除指标。</p> <ul style="list-style-type: none"> <li>• 点击PV</li> <li>• 推荐点击PV率</li> <li>• 点击UV</li> <li>• 推荐点击UV率</li> <li>• 自定义指标</li> </ul> <p>自定义指标包含参数设置和指标设置两部分。</p> <ul style="list-style-type: none"> <li>• 参数设置 <ul style="list-style-type: none"> <li>- 参数别名：用户指定参数别名应用于指标公式。</li> <li>- 行为类型：选择需要进行评估的行为类型，如物品曝光。</li> <li>- 阈值：阈值是用来衡量用户行为有效性的标准，当数据源的actionMeasure的值大于阈值时，当前用户行为有效。</li> <li>- 去重：您可以单击勾选 <input checked="" type="checkbox"/>，根据用户对行为记录去重。</li> </ul> </li> <li>• 指标设置 <ul style="list-style-type: none"> <li>- 指标名称：请您定义评估的指标名称。</li> <li>- 指标公式：用户指定自定义指标公式，如：<math>A/(A+B)</math>，参数A、B代表自定义参数的参数别名。只支持+、-、*、/。</li> </ul> </li> </ul>

4. 选择计算引擎，完成配置后单击“确定”。在“效果评估”页面可以单击“指标图”查看效果评估结果。

## 8.5.9 管理离线作业

您可以对各离线作业进行“复制”、“编辑”、“终止”、“重新执行”、“删除”等操作。您也可以通过查看服务的详细信息判读作业训练状态和查询训练结果。

### 复制离线作业

用户可以通过复制组合作业再次创建新的作业进行离线计算。生成的数据和原来的作业生成的数据相互独立，复制的离线作业会生成新的线上指定的UUID。

操作步骤如下：

1. 登录RES管理控制台，在左侧菜单栏中选择“离线作业”下的目标推荐作业，进入作业列表。
2. 在作业列表页面，您可以单击目标作业“操作”列的“复制”，进入“创建离线作业”页面。

3. 离线作业配置方法, 请参见[组合作业](#), 配置结束单击“完成”。

## 编辑离线作业

用户可以通过编辑离线作业修改该作业的参数信息进行离线计算。生成的数据会覆盖原来的离线作业生成的数据, 编辑并运行之后的离线作业会生成新的线上指定的UUID。

“启动中”、“计算中”状态的作业不支持编辑。

1. 登录RES管理控制台, 在左侧菜单栏中选择“离线作业”下的目标推荐作业, 进入作业列表。
2. 在作业列表页面, 您可以单击目标作业“操作”列的“编辑”, 进入“编辑离线作业”页面。
3. 离线作业配置方法, 请参见[组合作业](#), 配置结束单击“完成”。

## 终止离线作业

您可以对处于“启动中”、“计算中”状态的作业进行“终止”操作, “计算成功”和“计算失败”状态的作业无法终止。

1. 登录RES管理控制台, 在左侧菜单栏中选择“离线作业”下的目标推荐作业, 进入作业列表。
2. 在作业列表页面, 您可以单击目标作业“操作”列的“终止”结束离线作业的启动或计算。

## 重新执行离线作业

重新执行离线作业的目的是将本次执行生成的数据和之前的作业数据融合在一起, 无需重新创建作业就能更新数据。

重新执行不按作业不会生成新的线上指定的UUID。

“启动中”、“计算中”状态的作业不支持重新执行。

1. 登录RES管理控制台, 在左侧菜单栏中选择“离线作业”下的目标推荐作业, 进入作业列表。
2. 在作业列表页面, 您可以单击目标作业“操作”列的“重新执行”运行离线作业。

## 查看离线作业详情

在创建离线作业完成后, 您可以通过单击离线作业名称查看作业详情。组合作业包含的信息如[表8-40](#)所示。

在RES管理控制台, 选择“离线作业>组合作业”, 进入组合作业列表页面。在组合作业列表中, 您可以单击作业名称, 查看该作业的详情。



表 8-40 组合作业详情

参数名称	详情信息
资源信息	资源信息包括： <ul style="list-style-type: none"> <li>• 计算引擎：服务名、集群名称、任务配置地址、资源名信息、</li> <li>• 存储平台：服务名、集群名称、表名。</li> <li>• 数据源：数据表类型、数据源、数据格式。</li> </ul>
召回策略信息	该离线作业设置的召回策略详情，召回策略信息包括： <ul style="list-style-type: none"> <li>• 策略别名，策略显示的名称。</li> <li>• OBS地址，推荐业务人员可从OBS中选择保存有人工编辑推荐结果的列表，进行人工推荐便于进行ABTest。</li> <li>• 该策略生成的候选集ID。</li> </ul>
过滤规则信息	过滤规则用于配置候选集的过滤方式，过滤规则信息包括： <ul style="list-style-type: none"> <li>• 过滤规则别名。</li> <li>• 黑名单地址，黑名单所在路径。黑名单中的物品不能出现在最终推荐结果集里面。</li> <li>• 白名单地址，白名单所在的路径。白名单之外的物品不应该出现在最终推荐结果集里。</li> <li>• 历史行为过滤，指定与用户个性化的物品候选集过滤准则。例如对于用户过去3天内有过view行为的物品（如新闻）过滤，使之不进入候选集。</li> <li>• 过滤ID。</li> </ul>
排序规则信息	该离线作业设置的排序规则详情。
日志信息	日志信息包括： <ul style="list-style-type: none"> <li>• 策略类型。</li> <li>• 日志起始行。</li> <li>• 日志总行数。</li> <li>• 日志类型。</li> <li>• 日志详情。</li> </ul>

## 8.5.10 删除离线作业

删除离线作业即清理离线作业产生的数据，数据清理完成后删除元数据。操作必须确保流程中没有使用该作业产生的UUID，可以参考以下操作完成。

### 删除离线作业

您可以对“计算成功”、“计算失败”等状态的作业进行删除，“启动中”、“计算中”状态的作业不支持删除操作。

1. 登录RES管理控制台。在左侧菜单栏中选择“离线作业”下的目标推荐作业，进入作业列表。
2. 在作业列表页面，您可以单击目标作业“操作”列的“删除”。
3. 在删除页面，您可以直接单击“是”只删除该作业。也可以勾选下方的“同时删除作业产生的CloudTable数据”，选择该作业对应的数据进行离线删除。

图 8-20 删除作业



### 说明

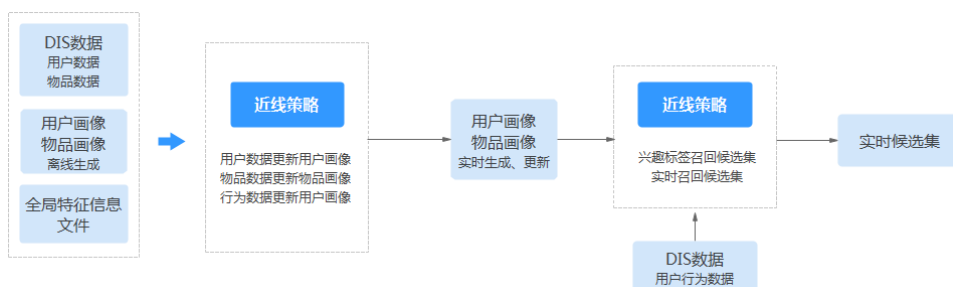
离线作业删除后，将不可恢复，请谨慎操作。

## 8.6 近线作业

### 近线作业简介

近线作业为推荐系统提供实时计算能力。近线作业以数据接入服务DIS中的数据为数据源，实时计算并更新用户画像、物品画像和推荐候选集等数据。使用近线作业，用户需先将业务系统埋点日志转换成**实时日志**指定格式，并实时写入DIS相应通道。近线作业具体实现请参见图8-21。

图 8-21 近线策略



近线策略的具体功能如下：

- [基于行为数据的用户画像更新](#)
- [基于用户的标签搜索候选集](#)
- [基于用户数据更新用户画像](#)
- [基于物品数据更新物品画像](#)
- [流式训练](#)

可上传3份数据至DIS：实时行为日志、实时用户日志和实时物品日志，如[实时日志](#)所示。

实时行为日志会被“基于行为数据的用户画像更新”和“基于用户的标签搜索候选集”使用；

实时用户日志会被“基于用户数据更新用户画像”使用；

实时物品日志会被“基于物品数据更新物品画像”使用。

## 创建近线作业

创建近线作业操作步骤如下：

1. 登录RES管理控制台，在左侧导航栏单击进入“近线作业”页面。
2. 在页面上方单击“创建”，进入“创建近线作业”页面。
3. 填写近线作业“名称”、“场景”和“描述”。
  - 近线策略名称：请以“Nearline-”开始，只能由字母、数字、中划线和下划线组成，并且长度小于64个字符。
  - “场景”信息可选择您在全局配置页面创建的场景。
4. 设置计算引擎信息，指定“服务名”、“集群名称”、“任务配置地址”、“资源规格”等信息。
5. 单击“添加近线策略”，根据业务需要在下拉框中选择一个合适的策略。各策略参数说明请单击策略名称进行查看。
  - [基于行为数据的用户画像更新](#)
  - [基于用户的标签搜索候选集](#)
  - [基于用户数据更新用户画像](#)
  - [基于物品数据更新物品画像](#)
  - [流式训练](#)
6. (可选) 在目标策略右侧，可单击“删除”，删除该策略。
7. 策略设置完成后，单击“确定”。作业一般需要运行一段时间，请您耐心等待。您可以前往近线作业列表，查看作业的基本情况。在作业列表中，刚创建的作业“状态”为“计算中”时，表示作业运行结束。当作业的“状态”变为“计算失败”时，您可以单击作业的名称，进入详情页面，通过查看日志等手段处理问题。


## 基于行为数据的用户画像更新

基于行为数据的用户画像更新可以对用户画像进行持续更新，更新频率可达秒级。以DIS中的实时行为日志为数据源，从中分析出用户画像更新信息，并实时更新用户画像，使用户画像随着用户行为的发生而不断变化。

通过此任务可以更新的用户画像内容有：用户标签等动态信息，以及用户年龄、位置、联系方式等上下文信息。

表 8-41 基于行为数据的用户画像更新参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认基于行为数据的用户画像更新。
服务名	DIS。将实时近线任务需要的 <b>实时行为日志</b> 数据添加到DIS中，推荐系统通过读取该数据进行近线计算。	DIS。
通道名称	数据通道名称，不同类型的数据需要创建不同的通道。	无。
起始位置	读取DIS数据的起始位置，latest表示从最新的数据开始读取，earliest表示从通道中现存最旧的数据开始读取。	默认为latest。
行为类型	用户行为包含如下几种类型： <ul style="list-style-type: none"> <li>• view：物品曝光</li> <li>• click：用户点击物品</li> <li>• collect：用户收藏了某个物品</li> <li>• uncollect：用户取消收藏某个物品</li> <li>• search_click：用户点击搜索结果中的物品</li> <li>• comment：用户对物品的评论</li> <li>• share：分享</li> <li>• like：点赞</li> <li>• dislike：点衰</li> <li>• grade：评分</li> <li>• consume：消费</li> <li>• use：观看视频/听音乐/阅读。</li> </ul>	默认选取所有行为。
窗口间隔 (秒)	近线策略处理的窗口间隔，单位为秒，10代表每隔10s进行一次算。	10
结果保存路径	根据行为日志，及cloudtable中已存在的用户画像和物品画像实时构造出样本通用数据，以供其他功能使用。	-
更新上下文	用户操作行为表中的“context”字段代表动作发生的上下文信息。更新上下文就是将这些信息写入到Cloudtable中进行实时的更新。	默认关闭。
是否更新物品的热度值	根据行为日志，计算物品的热度值。近线召回策略中（如利用兴趣标签搜索候选集），召回的候选集按何种顺序排列，如要选择根据热度排序，则需要打开该功能。	默认关闭。

参数名称	说明	默认值
是否保存历史记录	根据行为日志，统计用户的历史记录。近线召回策略中（如利用兴趣标签搜索候选集），召回的候选集中过滤掉用户的历史记录依赖于此项，保存了历史记录才可使用。	默认关闭。
兴趣标签衰减参数	兴趣标签的衰减参数，数值越小，衰减能力越强。数值越大，衰减能力越弱。如果值为0，则代表不衰减。	0.99
兴趣标签维护长度	各标签体系下，兴趣标签的最大长度。	20
全局特征信息文件	用户在使用近线策略之前，需要提供全局特征信息文件，该文件为JSON格式，包含特征名、特征大类、特征值类型。当上传的数据中的特征有变化时，需要同步更新该文件。全局特征信息文件示例请参考 <a href="#">全局特征信息文件</a> 。	-
异常数据输出路径	单击右侧的  按钮，选择数据在OBS中的存放路径，此路径下会记录不符合任务要求的输入数据。	-
用户画像存储	该数据来源于 <a href="#">基于用户数据更新用户画像</a> 更新后的用户画像数据。服务名默认为绑定的CloudTable资源信息。指定集群名称和表名用于存储更新后的用户画像。	不涉及。
物品画像存储	该数据来源于 <a href="#">基于物品数据更新物品画像</a> 更新后的物品画像数据。服务名默认为绑定的CloudTable资源信息。	不涉及。
过滤存储	过滤出用户发生过某行为的物品，并保存在此表中。例如，过滤出用户看过的物品，并存储在此表中。指定集群名称和表名用于存储更新后的用户画像。	不涉及。



## 基于用户的标签搜索候选集

基于用户的标签搜索候选集可以持续为用户召回新的候选集，召回频率可达秒级。以DIS中的实时行为日志为数据源，分析用户喜好，并以此为依据为用户召回候选集。此任务包含两种召回方式：

- 根据兴趣标签召回候选集
- 根据实时标签召回候选集

表 8-42 基于用户的兴趣标签搜索候选集参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认基于用户的标签搜索候选集。
数据源	<ul style="list-style-type: none"> <li>● 服务名：DIS。将实时近线任务需要的<b>实时行为日志</b>数据添加到DIS中，推荐系统通过读取该数据进行近线计算。</li> <li>● 通道名称：数据通道名称，不同类型的数据需要创建不同的通道。</li> <li>● 起始位置：读取DIS数据的起始位置，latest表示从最新的数据开始读取，earliest表示从最旧的数据开始读取。</li> </ul>	服务名默认DIS。 通道名无。 起始位置默认为latest。
行为类型	用户行为包含如下几种类型： <ul style="list-style-type: none"> <li>● view：物品曝光</li> <li>● click：用户点击物品</li> <li>● collect：用户收藏了某个物品</li> <li>● uncollect：用户取消收藏某个物品</li> <li>● search_click：用户点击搜索结果中的物品</li> <li>● comment：用户对物品的评论</li> <li>● share：分享</li> <li>● like：点赞</li> <li>● dislike：点衰</li> <li>● grade：评分</li> <li>● consume：消费</li> <li>● use：观看视频/听音乐/阅读。</li> </ul>	默认选取所有行为。
窗口间隔 (秒)	近线策略处理的窗口间隔，单位为秒，10代表每隔10s进行一次算。	10
过滤历史记录	生成的候选集中是否过滤历史记录，如需要过滤历史记录，则开启此项。	默认关闭。

参数名称	说明	默认值
候选集的排序策略	<p>生成候选集的排列规则，其排列顺序包括：</p> <ul style="list-style-type: none"> <li>● 随机排序：不对候选集排序。</li> <li>● 热度排序：根据物品属性表里的物品热度排序，由基于行为数据的用户画像更新生成，热度越大排序越靠前。</li> <li>● 时间排序：根据时间对候选集进行排序。时间排序需要指定特征名称和推荐天数。 <ul style="list-style-type: none"> <li>- 特征名称：值为时间戳(10位)的特征的名称，任务会根据此特征对候选集进行排序。</li> <li>- 推荐天数：推荐数据的时间段，该时间段从当前开始往前推N天，默认15天。</li> </ul> </li> </ul>	默认热度排序。
候选集最大长度	生成候选集的最大长度，每次计算更新的候选集中的个数不会超过最大值。	默认50。
候选集的召回策略	<p>召回候选集的策略。</p> <ul style="list-style-type: none"> <li>● 兴趣标签召回候选集：根据用户画像的兴趣标签召回候选集。</li> <li>● 实时标签召回候选集：根据用户实时操作的物品的标签召回候选集。</li> </ul>	默认兴趣标签召回候选集。
兴趣宽度	<p>生成候选集中的兴趣宽度，值越小候选集中的类型越少。</p> <p><b>说明</b> 选择兴趣宽度数量对应的，权重值最高的兴趣标签个数进行检索，得到与标签匹配的物品候选集。</p>	默认3。
全局特征信息文件	用户在使用近线策略之前，需要提供全局特征信息文件，该文件为JSON格式，包含特征名、特征大类、特征值类型。当上传的数据中的特征有变化时，需要同步更新该文件。全局特征信息文件示例请参考 <a href="#">全局特征信息文件</a> 。	-
异常数据输出路径	单击右侧的  按钮，选择数据在OBS中的存放路径，此路径下会记录不符合任务要求的输入数据。	-
自定义搜索	<p>改变默认搜索条件，按照所选自定义搜索规则进行搜索。自定义搜索的内容来自于全局特征信息文件。您可以单击添加自定义搜索，在下拉选项中选择用户的兴趣标签，单击转换为另外的标签体系进行召回。</p> <p><b>说明</b> 存在不同标签体系时可使用此功能，自定义搜索条件。一般情况下不建议使用。</p>	-

参数名称	说明	默认值
用户画像存储	指定存储用户画像的存储平台信息。服务名默认 CloudTable，指定集群名称和表名用于存储更新后的用户画像。	不涉及。
物品画像存储	指定存储物品画像的存储平台信息。服务名默认 CloudTable，指定集群名称和表名用于存储更新后的用户画像。	不涉及。
过滤存储	指定存储过滤集合的存储平台信息。服务名默认 CloudTable，指定集群名称和表名用于存储更新后的用户画像。	不涉及
候选集存储	指定存储候选集的存储平台信息。服务名默认 CloudTable，指定集群名称和表名用于存储更新后的用户画像。	不涉及。



## 基于用户数据更新用户画像

基于用户数据更新候选集可以持续更新用户画像，更新频率可达秒级。以DIS中的实时用户数据为数据源，实时捕捉用户数据近线，持续更新或添加用户画像数据，使用户画像处于最新状态。

表 8-43 基于用户数据更新用户画像参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认基于用户数据更新用户画像。
数据源	<ul style="list-style-type: none"> <li>服务名：DIS。将实时近线任务需要的<b>实时用户日志</b>数据添加到DIS中，推荐系统通过读取该数据进行近线计算。</li> <li>通道名称：数据通道名称，不同类型的数据需要创建不同的通道。</li> <li>起始位置：读取DIS数据的起始位置，latest表示从最新的数据开始读取，earliest表示从最旧的数据开始读取。</li> </ul>	服务名默认 DIS。 通道名无。 起始位置默认为 latest。
窗口间隔 (秒)	近线策略处理的窗口间隔，单位为秒，10代表每隔10s进行一次算。	10s。
全局特征信息文件	用户在使用近线策略之前，需要提供全局特征信息文件，该文件为JSON格式，包含特征名、特征大类、特征值类型。当上传的数据中的特征有变化时，需要同步更新该文件。全局特征信息文件示例请参考 <b>全局特征信息文件</b> 。	-





参数名称	说明	默认值
异常数据输出路径	单击右侧的  按钮，选择数据在OBS中的存放路径，此路径下会记录不符合任务要求的输入数据。	-
用户画像存储	<p>指定存储用户画像的存储平台信息。</p> <ul style="list-style-type: none"> <li>• 服务名：默认CloudTable，用于存储更新后的用户画像。</li> <li>• 集群名称：指定对应的集群名称。</li> <li>• 表名：选择对应表名称。您可以单击  设置数据版本。RES的数据版本有两种，“V1”版本即数据按照原有格式存储，未做过分区处理。“V2”版本则会依照用户的分区设置做分区处理，当分区合理时，数据将均匀分布在各个节点，有效利用Cloudtable的高并发特性，提升读写效率。其中“预分区数量”和“索引分区数量”可以根据数据量进行设置，如果读写性能达不到要求，可以增加Cloudtable的RS单元数量提升性能。</li> </ul>	-

## 基于物品数据更新物品画像

基于物品数据更新物品画像可持续更新物品画像，更新频率可达秒级。以DIS中的实时物品日志为数据源，持续更新或添加物品画像数据，使物品画像处于最新的状态。

表 8-44 基于物品数据更新物品画像参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认基于物品数据更新物品画像。
数据源	<ul style="list-style-type: none"> <li>• 服务名：DIS。将实时近线任务需要的<b>实时物品日志</b>数据添加到DIS中，推荐系统通过读取该数据进行近线计算。</li> <li>• 通道名称：数据通道名称，不同类型的数据需要创建不同的通道。</li> <li>• 起始位置：读取DIS数据的起始位置，latest表示从最新的数据开始读取，earliest表示从最旧的数据开始读取。</li> </ul>	<p>服务名默认DIS。 通道名无。 起始位置默认为latest。</p>
窗口间隔 (秒)	近线策略处理的窗口间隔，单位为秒，10代表每隔10s进行一次算。	10s。

参数名称	说明	默认值
全局特征信息文件	用户在使用近线策略之前，需要提供全局特征信息文件，该文件为JSON格式，包含特征名、特征大类、特征值类型。当上传的数据中的特征有变化时，需要同步更新该文件。全局特征信息文件示例请参考 <a href="#">全局特征信息文件</a> 。	-
异常数据输出路径	单击右侧的  按钮，选择数据在OBS中的存放路径，此路径下会记录不符合任务要求的输入数据。	-
候选集	是否维护维新入库物品候选集。 <ul style="list-style-type: none"> <li>时间特征名称：特征值为10位时间戳的特征名称。</li> <li>候选集最大长度：维护的候选集的最大长度。</li> <li>保留天数：保留时间为最近N天的候选集。</li> </ul>	false
物品画像存储	指定存储物品画像的存储平台信息： <ul style="list-style-type: none"> <li>服务名：默认CloudTable，用于存储更新后的物品画像。</li> <li>集群名称：指定对应的集群名称。</li> <li>表名：选择对应表名称。您可以单击设置数据版本。RES的数据版本有两种，“V1”版本即数据按照原有格式存储，未做过分区处理。“V2”版本则会依照用户的分区设置做分区处理，当分区合理时，数据将均匀分布在各个节点，有效利用Cloudtable的高并发特性，提升读写效率。其中“预分区数量”和“索引分区数量”可以根据数据量进行设置，如果读写性能达不到要求，可以增加Cloudtable的RS单元数量提升性能。</li> </ul>	-

## 流式训练

流式训练接入用户行为数据，进行在线训练，以分钟级更新在线服务依赖的排序模型，减少排序模型的效果衰减。


### 前提条件

1. 已存在运行成功的排序策略。建议“优化器类型”选择“ftrl”。
2. 已存在运行成功的在线服务。该在线服务“排序方式”为“点击率预估”并且“模型文件路径”为1排序策略生成的模型存储路径。

 说明

- 流式训练仅支持2019年12月4日0点之后创建的排序策略和在线服务。
- 流式训练使用到的排序策略不支持“核函数特征交互神经网络-PIN”算法。
- 建议您定期“重新执行”条件1对应的离线排序任务。重新执行排序策略后，在线服务和流式训练作业将自动同步新产生的模型，无需重新配置或重新执行在线服务和流式训练作业。
- 在线服务可以减少排序模型的效果衰减，但仍需定期重新执行离线任务。

表 8-45 流式训练参数说明

参数名称	说明	默认值
策略别名	策略显示名称，由中文、英文、数字、下划线、空格或者中划线组成，并且不能以空格开始和结束，长度为1~60个字符。	默认流式训练。
数据源	<ul style="list-style-type: none"> <li>• 服务名：添加已经进行资源绑定的DIS资源。</li> <li>• 输入流位置：行为数据流的输入配置。将实时近线任务需要的<b>实时物品日志</b>数据添加到输入流指定的DIS通道中，推荐系统通过读取该数据进行近线计算。                             <ul style="list-style-type: none"> <li>- 通道名称：DIS的数据通道名称，不同类型的数据需要创建不同的通道。</li> <li>- 起始位置：读取DIS数据的起始位置，latest表示从最新的数据开始读取，earliest表示从最旧的数据开始读取。</li> </ul> </li> <li>• 输出流位置：行为数据流处理后的输出配置。选择对应的DIS通道名称进行存放。输出流存放的数据属于流式训练作业的中间数据，用户无需获取或发送数据到输出流绑定的数据通道，只需确保该通道仅供本作业作为输出流使用。</li> </ul>	服务名默认DIS。 通道名无。 起始位置默认为latest。
窗口间隔 (秒)	近线策略处理的窗口间隔，单位为秒，10代表每隔10s进行一次算。	10s。
算子参数	<ul style="list-style-type: none"> <li>• 在线服务名：使用的在线服务的名称。该在线服务需满足<b>前提条件</b>的要求。</li> <li>• 流程名：在线服务对应的在线流程名称。</li> <li>• 异常数据输出路径：单击右侧的按钮，选择数据在OBS中的存放路径，此路径下会记录不符合任务要求的输入数据。</li> </ul>	-
训练设置	<ul style="list-style-type: none"> <li>• 计算节点规格：您可以根据业务需求选择节点规格，“8核 16GiB”、“8核 32GiB”、“8核 64GiB”、“16核 128GiB”。</li> <li>• 模型更新间隔：模型更新时间间隔，更新模型文件。</li> </ul>	默认“8核 16GiB” 时间间隔10min

参数名称	说明	默认值
优化策略相关参数	<ul style="list-style-type: none"> <li>• 优化器类型：ftrl。适用于处理超大规模数据的,含大量稀疏特征的在线学习的常见优化算法</li> <li>• 学习率：优化算法的参数，决定优化器在最优方向上前进长的参数。默认0.1。</li> <li>• 初始梯度累加和：梯度累加和用来调整学习步长。默认0.1。</li> <li>• L1正则项系数：叠加在模型的1范数之上，用来对模型值进行限制防止过拟合。默认0。</li> <li>• L2正则项系数：叠加在模型的2范数之上，用来对模型值进行限制防止过拟合。默认0。</li> </ul>	-

## 8.7 在线服务

### 8.7.1 创建在线服务

在线服务负责分析结果的应用过程，泛指部署线上服务后提供的推理服务，对外提供API接口。在推荐系统中，包含推荐引擎、文本标签、排序三种在线服务，具体说明如下：

- **推荐引擎**  
推荐引擎用于对RES召回策略跑出来的候选集结果进行融合过滤和排序。
- **文本标签**  
文本标签服务为用户提供自然语言处理工具，可用于关键词提取和命名实体识别。
- **排序**  
排序服务允许用户提供自己的候选集，使用RES的排序策略进行排序。

#### 前提条件

- 已经有计算成功的离线作业并且已经生成候选集UUID。
- 您可以部署10个在线服务，每个在线服务可以创建2个在线流程，如需增加配额请您提[工单](#)处理。
- 由于在线运行需消耗资源，确保账户未欠费。
- 如果账户欠费，账号将被冻结，服务会被停止运行，您无法对服务执行调用、启动或者编辑操作。

#### 操作步骤

1. 登录RES管理控制台，在左侧导航栏选择进入“在线服务”页面。
2. 在页面上方单击“创建”进入“创建在线服务”，在页面填写相关参数。
  - a. 填写基本信息并选择服务类型。您可以根据实际情况填写“名称”、“场景”和“描述”信息。其中“场景”信息可选择您在全局配置页面创建的[场景](#)。

图 8-22 基本信息

基本信息

\* 名称

场景

描述

0/100

- b. 根据业务需求选择服务类型。包括“推荐引擎”、“排序”和“文本标签”。
- c. 单击“添加在线流程”，并进行命名，您最多可以部署5个在线流程。根据选择的服务类型配置在线流程，包括融合、过滤、排序和模型及配置等关键信息。具体参数信息参见表8-46。服务类型选择推荐引擎。

图 8-23 创建在线服务

融合

\* 推荐结果设置

任务别名	UUID	优先级(P1..)	同优先级数据占比	操作
在线候选召回	online-recall	P1	- 100 + (%)	设置参数

+ 添加推荐候选集 + 添加在线候选集

容错   选择 清空

离线过滤

过滤   选择 清空

在线过滤

属性过滤规则 + 增加属性过滤规则

[去重] 物品属性 + 增加物品属性

[去重] 忽略长度

排序

排序方式  点击率预估  属性权重

特征工程   选择 清空

\* 模型文件路径

公共配置

\* 全局特征信息文件


\* 画像   选择 清空

模型及配置

模型名称	模型版本	计算节点规格	计算节点个数	分流 (%)
res-infer-def...	1.0.0	2核8GiB	- 2 +	- 100 +

表 8-46 创建在线服务参数说明

参数名称	子参数	说明
融合	“推荐结果设置”	<p>“添加推荐候选集”（选择离线或近线任务所生成的推荐候选集进行排序）</p> <ul style="list-style-type: none"> <li>任务别名和UUID：单击操作列表的“选择”添加离线或近线的任务名称和候选集ID。</li> <li>优先级：优先级高的推荐结果将确保展示在优先级低的之前。</li> <li>同优先级数据占比：优先级相同的推荐候选集，该占比展示推荐数量，同优先级下的数据占比之和需要等于100%。</li> </ul> <p>“添加在线候选集”（根据设置的参数在线进行召回，必须添加全局特征信息文件才可设置参数）</p> <ul style="list-style-type: none"> <li>任务别名和UUID：分别默认为“在线候选集召回”和“online-recall”，无需改动。</li> <li>优先级、同优先级数据占比：同添加推荐候选集。</li> <li>设置参数：单击操作列的“设置参数”进行召回策略参数的配置。根据业务需求选择召回类型为物品或者用户，配置在线召回特征。在线召回的特征属性来自于公共配置的全局特征信息文件。</li> </ul> <p>可单击“添加推荐候选集、添加在线候选集”配置多个候选集，作为当前在线流程的排序候选集。</p> <p><b>说明</b> 在线候选集的延迟较推荐候选集较高，如无特殊需求，建议选择推荐候选集。</p>
	容错	容错用于数据请求异常时的固定人工输出召回集。需要在离线作业创建召回策略进行人工策略的计算。单击“选择”获取人工导入策略的任务别名和UUID。
离线过滤	过滤	过滤数据来源于过滤规则产生的候选集，单击“选择”获取过滤的任务别名和UUID。
在线过滤	【去重】物品属性	属性名从画像算子生成，如“product_color”，则对产品颜色相同的物品进行去重。
	【去重】忽略长度	截断物品ID末尾指定长度后的字符串进行去重，如指定长度为2，则SKU_A1234和SKU_A1244只会保留其中一个。

参数名称	子参数	说明
	属性过滤规则	<p>指定定制化用户属性以及物品属性过滤规则，属性过滤规则用于过滤最终用户的推荐结果。例如，对于一线城市用户过滤敏感信息物品，使之不进入候选集。单击  增加属性过滤规则。</p> <ul style="list-style-type: none"> <li>用户属性：指定在用户属性中需要过滤的字段，包含属性名和属性值。来源于画像数据，即<b>特征工程</b>中<b>初始用户画像-物品画像-标准宽表生成</b>算子作业输出的数据，用户属性来自于公共配置的全局特征信息文件。如过滤籍贯是广东且性别为男性的用户。</li> <li>物品属性：指定在物品属性中需要过滤的字段，包含属性名和属性值。来源于画像数据，即<b>特征工程</b>中<b>初始用户画像-物品画像-标准宽表生成</b>算子作业输出的数据，物品属性的名称来自于公共配置的全局特征信息文件。如过滤产品颜色为红色且产品品牌为华为的物品。</li> </ul>
排序方式	“点击率预估”	<ul style="list-style-type: none"> <li>特征工程：排序数据来源于排序算子作业产生的候选集。单击“选择”获取排序策略的任务别名和UUID。</li> <li>模型文件路径：排序策略生成的模型存储路径。</li> </ul>
	“属性权重”	属性权重：输入属性，权重默认1.0。也可单击“增加属性权重”。
公共配置	“全局特征信息文件”	为json格式文件。指定在去重过滤即属性过滤中物品属性和用户属性的特征信息文件，如物品属性需要“feature_name”、“feature_type”、“feature_value_type”字段来表示特征名，特征类型以及特征值类型。需预先存储在OBS中。
	“画像”	画像为 <b>特征工程</b> 中 <b>初始用户画像-物品画像-标准宽表生成</b> 算子的结果。且此画像会用于去重过滤、属性过滤以及排序的计算中。单击“选择”获取特征工程作业产生的UUID。
模型及配置	-	<p>设置模型名称、模型版本、计算节点规格、计算节点个数和分流（%）。</p> <p>模型名称和模型版本选择调用API接口的模型名称和模型版本。“计算节点规格”默认2核 8GiB，“计算节点个数”默认为2，“分流”之和必须是100%。</p>

- 完成该项配置后，单击“下一步”进入规格确认页面，核对无误后，单击“完成”。当您通过在线服务列表查看在线服务的状态由“初始化”、“部署中”变更为“运行中”时即完成在线服务的部署。在线服务一般需要运行一段时间，根据您的数据量和资源不同，运行时间将耗时几分钟到几十分钟不等，请您耐心等待。

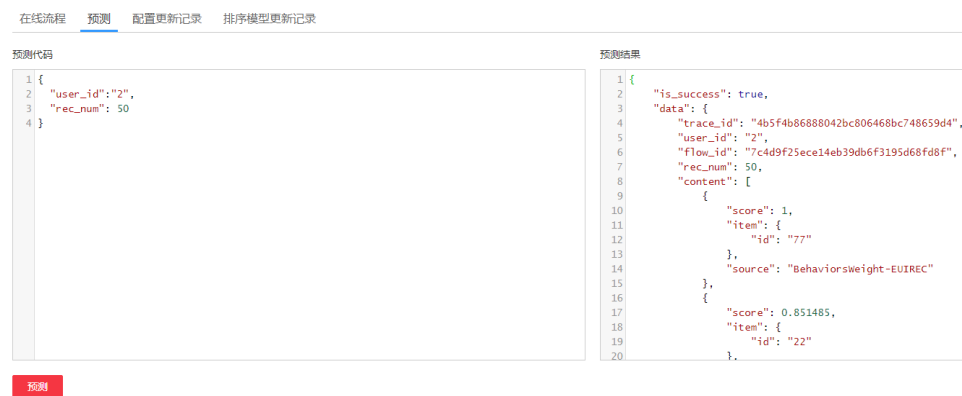
## 8.7.2 获取推荐结果

在线服务创建完成，部署成功后，当服务状态会显示“运行中”，表示服务状态正常。您可以通过在线预测功能测试推荐结果进一步调整作业参数，也可以通过预测接口来调用API，获取推荐结果。

### 在线预测

1. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表页面。
2. 单击目标服务名称进入服务详情页面，单击下方的“预测”页签，输入预测代码，单击“预测”后显示预测结果，如图8-24所示。输入用户ID和推荐数量，根据您的召回策略等返回用户的预测结果。

图 8-24 在线预测



### 获取预测接口



- 通过在线服务页面获取接口
  - a. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表页面。
  - b. 单击目标服务名称前方的  查看预测接口，通过单击预测接口右侧的 ，复制接口地址，调用服务。

图 8-25 获取预测接口




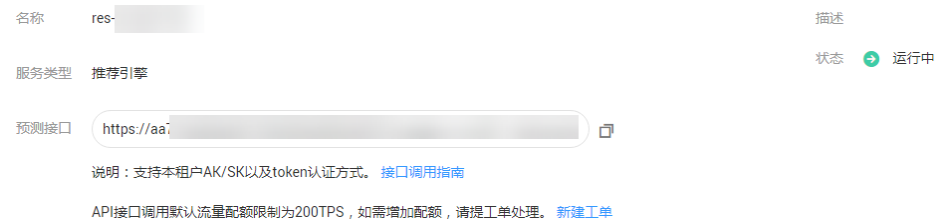
- 通过在线服务详情页面获取接口
  - a. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表页面。
  - b. 单击目标服务名称进入服务详情页面，通过单击预测接口右侧的 ，复制接口地址，调用服务。



图 8-26 获取预测接口



## 调用接口

Mozilla、Google都为REST提供了图形化的浏览器插件，发送处理请求消息。

预测接口的调用和接口参数请参见[预测接口](#)，更多接口信息请参见《[推荐系统API参考](#)》。

### 8.7.3 查看在线服务详情

当在线服务创建成功后，您可以进入“在线服务”页面，来查看服务详情。

1. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入“在线服务”管理页面。
2. 您可以在服务列表直接查看服务的“名称”、“状态”、“调用总次数”、“开通状态”和“服务类型”等。
  - 当服务部署成功：您可以单击服务名称左侧的 查看预测接口、ID、在线流程名称、在线流程ID进行调用。
  - 当服务部署失败：您可以单击服务名称左侧的 查看失败原因，根据失败原因对服务配置进行调整，重新部署。常见的可能失败原因及解决方法请参见[表8-47](#)。

表 8-47 部署失败解决方法

报错	失败原因	解决方法
Failed to deploy service(A maximum of (5) real-time service instances are allowed.	一个服务允许选择的节点个数不能超过5个。	减少使用的节点个数或者申请增加节点个数。 如有更多问题请 <a href="#">提工单</a> 咨询技术人员为您解答。

3. 单击目标服务名称，进入服务详情页面查看服务具体信息。
  - 您可以查看服务的“名称”、“状态”、“描述”、“服务类型”、和“预测接口”。
  - 您也可以通过单击预测接口右侧的 ，复制接口地址，调用服务。

图 8-27 在线服务详情



4. 您可以在如下页面查看服务相关参数信息，包括“在线流程”、“预测”、“配置更新记录”、“排序模型更新记录”。

表 8-48 在线服务详情

参数	说明
在线流程	展示在线流程名称、在线流程ID以及在线流程关于融合、过滤、排序、公共配置和模型相关的参数信息。
预测	可对运行中状态的在线服务进行预测，输入预测代码，单击“预测”后显示预测结果。
配置更新记录	展示配置相关记录。 单击配置记录前方的▾可查看该记录下详细的模型信息，包括模型名称、版本、分流、计算节点规格和计算节点个数。
排序模型更新记录	展示模型的更新记录。包括模型名称、状态、更新时间和消息。

## 8.7.4 管理在线服务

您可以对在线作业进行“编辑”、“启动”、“开通”、“停止”、“删除”等操作。您可以通过单击在线服务名称查看在线服务的详细信息。

### 编辑服务

用户可以通过“编辑”在线服务修改该参数信息进行计算。生成的数据会覆盖原来的在线服务计算生成的数据。“部署中”的在线服务不支持编辑。操作步骤如下：

1. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表。
2. 在“在线服务”页面，您可以单击目标服务操作列的“编辑”，进入“编辑在线服务”页面。
3. 在线服务配置方法，请参见[创建在线服务](#)，配置结束单击“立即修改”。

### 启动服务

当在线服务创建完成后，用户可以根据业务需求启动服务运行，当服务状态为“运行中”时服务可用。启动服务之后会生成预测接口用于调用服务接口，查看推荐结果。

您可以对“已停止”和“部署失败”的在线服务进行“启动”操作。“初始化”和“部署中”状态的在线服务不支持启动。

1. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表。
2. 在“在线服务”页面，单击目标服务操作列的“启动”完成操作。

## 开通服务

在线服务创建完成后，默认为开通状态，您可以对“运行中”状态的服务进行正常调用。当您在调用过程中提示“开通”时，需要开通才能正常使用。

您可以对处于“运行中”、“已停止”状态的服务进行“开通”操作。

1. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表。
2. 在“在线服务”页面，单击目标服务操作列的“开通”完成操作。

## 停止服务

由于运行中的在线服务将一直收费，您可以通过停止操作，停止收费。停止服务即释放当前该在线服务占用的容器资源，停止状态的服务不可用，会调用失败。您可以对处于“运行中”状态的服务进行“停止”操作。

1. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表。
2. 在“在线服务”页面，单击目标服务操作列的“停止”结束作业的启动或计算。

## 删除服务

删除服务即清理服务产生的数据，完成资源释放。

1. 登录RES管理控制台，在左侧菜单栏中选择“在线服务”，进入服务列表。
2. 在“在线服务”页面，单击目标服务操作列的“删除”，在“删除”页面确认在线服务信息后单击“确定”进行删除。

### 说明

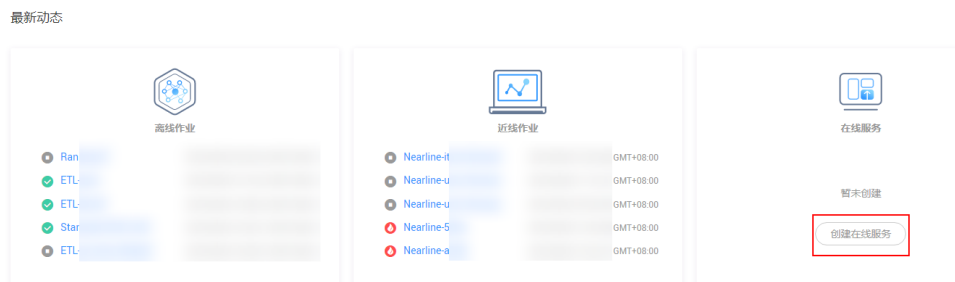
服务删除后，将不可恢复，请谨慎操作。存储在OBS中的数据不会被删除。

## 8.8 服务总览信息

您可以在RES管理控制台的总览页查看服务的最新动态、了解作业状态、快捷创建服务。

1. 登录RES管理控制台，单击左侧导航栏上的“总览”，进入总览页面。
2. 总览界面包括“最新动态”、“视频教程”和“常用链接”，如图8-28所示。其中，在最新动态可查看所创建“离线作业”、“近线作业”和“在线服务”的名称、状态和创建时间。

图 8-28 总览



3. 单击作业或服务名称，进入该详情页面，可以查看资源信息、策略信息和日志信息。您也可以通过页面快速创建作业或服务。单击“创建在线服务”，快速进入创建在线服务页面。

## 8.9 数据格式规范

### 8.9.1 推荐系统 OBS 文件夹规范

使用推荐系统时，需要在OBS创建桶并导入离线数据，同时作业所产生的数据也会保存在OBS中。为了方便您快速定位文件路径，建议您按照如下结构准备数据创建文件夹，并上传至OBS桶。

#### OBS 文件夹示例

自定义OBS桶名

- obs-offline-data 用于存放用户行为表、用户属性表和物品属性表。
- obs-customize-data 用于存放用户自定义列表“customize.csv”，该列表应用于召回策略的人工导入策略。
- obs-filter-data 用于存放用户黑白名单，应用于过滤策略。
- obs-global-data 用于存放全局特征文件，例如放入名为“global\_feature\_info.json”的文件。
- obs-general-data 用于存放经过特征工程处理的宽表，又名通用格式数据。
- obs-pretrain-data 用于存放排序预处理任务生成的预训练数据。
- obs-error-data 用于存放异常数据输出路径，来源于近线策略。
- obs-nearline-data 用于存放近线任务结果数据，宽表形式。
- obs-model-data 排序模型保存路径。
- obs-config-data 计算引擎配置地址，用来记录计算引擎的日志。

### 8.9.2 离线数据源

调用RES之前，您需要准备3种基础数据包并上传至OBS，离线数据源目前支持CSV和JSON。具体数据包请参见表1 基础数据表。

表 8-49 基础数据表

数据类型	表名
用户类数据	用户属性表
物品类数据	物品属性表
行为类数据	用户操作行为表

### 说明

1. 用户需要自己手工创建整理这些表并存储到OBS上。
2. 每张表的表结构必须符合推荐系统的要求，列名和字段类型需要和规范中保持一致（参考下面的表结构说明）。
3. 每张表中填充的数据，必须符合推荐引擎的要求。
4. 对于业务数据中无法提供的字段可以填NULL。

## 用户属性表

用户属性表记录用户的属性信息，例如地域、爱好等，属性名和属性值成对出现。

表 8-50 字段描述

字段名	类型	描述	是否必选
userId	String	被推荐用户的唯一标识符。例如，用户的登录账号、imei号等。	是
userProperties	String	<p>用户属性kv串。不同属性之间用\002（ASCII值）分隔，属性的key和value之间用\003（ASCII值）分隔，多值枚举型属性的不同值之间以\004（ASCII值）分隔，kv数值型属性的不同键值对同样以\004（ASCII值）分隔，每一个键值对内部用冒号（:）将键和值分隔。</p> <p>基本格式： k1\003v1\002k2\003v2\k3\003v1\004v2\004v3\002k4\003\k1:v1\004k2:v2\004k3:v3</p> <p>例子：比如用户有四个标签，分别是年龄、性别、爱好、画像，取值分别为18、1、[read,gym,music]、{morality:0.8,iq:0.7,art:0.9}，则这条记录可编码为age\00318\002gender\0031\002hobbies\003read\004gym\004music\002profile\003morality:0.8\004iq:0.7\004art:0.9。</p> <p>userProperties字段每个KV组中value的取值类型信息由user_meta_config给出。</p>	否

## 用户属性 JSON 格式示例

```
{
  "userId": "user1",
  "userProperties": "provinceId\u00035147\u0002cityId\u00035148\u0002districtId\u00035154\u0002streetId\u0003130533\u0002payment_type\u00031\u0002payment_method\u0003CMBLIFE_CMBLIFE\u0002payment_channel\u000312"
}
```

## 物品属性表

物品属性表记录物品的属性信息，例如类别、长度等，属性名和属性值成对出现。

表 8-51 字段描述

字段名	类型	描述	是否必选
itemId	String	物品ID, 唯一标识。	是
itemProperties	String	描述物品的key-value信息, key标识描述名称, value标识值。基本格式同用户属性表中userProperties。	否

## 物品属性示例

```
{
  "itemId":"item1",
  "itemProperties":"product_name\u0003wyf-name\u0002order_price\u0003500.00\u0002weight\u0003130.00\u0002volume\u0003v1"
}
```

## 用户操作行为表

用户-物品行为表, 每行记录用户的单次行为信息, 包含用户标识符、行为对象标识符、行为类型和行为时间等信息。

表 8-52 字段描述

字段名	类型	描述	是否必选
userId	String	用户ID。	是
objectType	String	发生行为对象的名称。 item: 是用户和物品本身发生的行为。	是
actionObject	String	对应行为发生对象的值, 如果是和物品发生关系, 则是物品的id(itemId)的值。	是

字段名	类型	描述	是否必选
actionType	String	行为类型： <ul style="list-style-type: none"> <li>view: 物品曝光</li> <li>click: 用户点击物品</li> <li>collect: 用户收藏了某个物品</li> <li>uncollect: 用户取消收藏某个物品</li> <li>search_click: 用户点击搜索结果中的物品</li> <li>comment: 用户对物品的评论</li> <li>share: 分享</li> <li>like: 点赞</li> <li>dislike: 点衰</li> <li>grade: 评分</li> <li>consume: 消费</li> <li>use: 观看视频/听音乐/阅读。</li> </ul>	是
actionMeasure	String	发生行为的度量，金额，评分，次数（整数）等。 默认值： <ul style="list-style-type: none"> <li>评分和观看视频/听音乐/阅读由用户根据实际情况自行定义，如观看两分钟，值为“0.5”，观看5分钟，值为“0.8”。</li> <li>其余行为的默认值为1.0。</li> </ul>	是
actionContent	String	发生行为为评论时，需要记录评论内容。 当是search_click时，需要记录搜索的关键词。	否
actionDateTime	String	行为发生的时间，采用UTC标准时间，单位以秒计。	否
actionLocation	String	行为发生的位置和精度维度。 格式：[latitude, longitude]	否
context	String	动作发生的上下文信息，内容为json格式字符串，也可作为行为表的扩展字段。 例如，用户当前的设备id, ip地址等信息。	否
subSite	String	行为发生的位置ID。例如，在首页推荐里面点击，在详情页里面浏览。	否
traceld	String	用于追踪每个被推荐物品的唯一ID。用于效果的计算。	否
flowId	String	用于计算每一个在线服务的效果。flowId由推荐系统的API返回给用户，用户需把flowId写到用户行为日志中。	否

## 用户操作行为示例

```
{  
  "userId":"user1",  
  "objectType":"item",  
  "actionObject":"item1",  
  "actionType":"collect",  
  "actionMeasure":"1.0",  
  "actionDateTime":"1512689700"  
}
```

### 8.9.3 实时日志

RES根据实时发送到DIS上的日志，进行数据计算和处理，更新用户的相关数据。用户发送到DIS上的数据具体如下：

#### 实时行为日志

实时行为日志的作用包括：

- 更新用户的兴趣标签。
- 记录所选行为类型的历史记录。
- 更新用户的上下文信息。
- 召回候选集。

表 8-53 实时行为日志字段描述

字段名	类型	描述	是否必选
userId	String	用户ID。	是
objectType	String	发生行为对象的类型，item是用户和商品本身发生的行为。	是
actionObject	String	对应行为发生的对象的值，如果是和商品发生关系：则是商品的id(itemId)的值。	是
actionType	String	行为类型： <ul style="list-style-type: none"><li>• 物品曝光</li><li>• 用户点击物品</li><li>• 用户收藏了某个物品</li><li>• 用户取消收藏某个物品</li><li>• 用户点击搜索结果中的物品</li><li>• 用户对物品的评论</li><li>• 分享</li><li>• 点赞</li><li>• 点衰</li><li>• 评分</li><li>• 消费</li><li>• 观看视频/听音乐/阅读</li></ul>	是



字段名	类型	描述	是否必选
actionMeasure	String	发生行为的度量，金额，评分，次数（整数）等。 默认值： <ul style="list-style-type: none"><li>评分和观看视频/听音乐/阅读可由用户根据实际情况自行定义，如观看两分钟，值为“1.5”，观看5分钟，值为“1.8”。</li><li>其余行为的默认值为1.0。</li></ul>	是
actionContent	String	发生行为为评论时，需要记录评论内容。 行为为搜索时，需要记录搜索的关键词。	否
actionDateTime	String	行为发生的时间，采用UTC标准时间，单位以秒计。	否
actionLocation	String	行为发生的位置，精度维度格式：[latitude, longitude]。	否
context	Json	动作发生的上下文信息，内容为json对象，也可作为行为表的扩展字段。例如，用户当前的设备id，ip地址等信息。	否
subSite	String	行为发生的位置ID，比如，在首页推荐里面点击，在详情页里面浏览。	否
traceId	String	用于追踪每个被推荐物品的唯一ID。用于推荐效果的计算。	否
flowId	String	用于计算每一个在线服务的效果。flowId由推荐系统的API返回给用户，用户需把flowId写到用户行为日志中。	否

## 实时行为日志示例

```
{
  "userId": "user1",
  "objectType": "item",
  "actionObject": "item1",
  "actionType": "collect",
  "actionMeasure": "1.0",
  "actionDateTime": "1512689700"
}
```

## 实时用户日志

实时用户日志用于在实时流中对用户画像表进行写入。

表 8-54 用户实时日志字段描述

字段名	类型	描述	是否必选
userId	String	全局唯一用户ID。	是
BASICINFO	Json	用户基本属性值。其中字段的值只能是数值型，字符串或字符串数组。	否
TAGS	Json	用户的兴趣标签。其中字段的值只能是Map类型。 可传入不同类型的标签信息（如人工标签，关键词标签等）。 此标签体系需与物品标签体系相同。若无历史兴趣标签，则无需传入此字段。推荐系统将会根据特定行为匹配的标签进行计算并完成更新。 <b>说明</b> 传入TAGS字段会直接覆盖原程序中已计算出的兴趣标签权重。	否
CONTEXT	Json	用户上下文信息。其中字段的值只能是数值型，字符串或字符串数组。	否

### 实时用户日志示例

```
{
  "userId":"user123",
  "BASICINFO":{
    "age":15,
    "gender":"male",
    "salary":31693.84172,
    "hobbies":[
      "体育",
      "美术",
      "音乐",
      "格斗",
      "书法"
    ]
  },
  "TAGS":{
    "artificialLabel":{
      "tag1":0.84,
      "tag2":0.22,
      "tag3":0.88
    },
    "keywordLabel":{
      "tag4":0.22,
      "tag5":0.99
    }
  },
  "CONTEXT":{
    "provinceld":"3853",
    "cityld":"3868",
    "districtld":"3873"
  }
}
```

## 实时物品日志

实时物品日志用于在实时流中对物品画像表进行写入。

表 8-55 物品实时日志字段描述

字段名	类型	描述	是否必选
itemId	String	全局唯一物品ID。	是
BASICINFO	Json	物品基本属性值。其中字段的值只能是数值型，字符串或字符串数组。	否
TAGS	Json	物品标签。其中字段的值只能是Map类型。可传入不同类型的标签信息（如人工标签，关键词标签等）。 <b>说明</b> 如果此字段为空，则RES无法根据行为计算用户的兴趣标签。	否

表 8-56 物品实时日志示例表

itemId	BASICINFO			TAGS	
item123	weight	brand	price	artificialLabel	keywordLabel
	12	"male"	333	{"tag1":0.84,"tag2":0.22,"tag3":0.88}	{"tag4":0.22,"tag5":0.99}

## 实时物品日志示例

```
{
  "itemId":"item123",
  "BASICINFO":{
    "weight":12,
    "brand":"male",
    "price":333
  },
  "TAGS":{
    "artificialLabel":{
      "tag1":0.84,
      "tag2":0.22,
      "tag3":0.88
    },
    "keywordLabel":{
      "tag4":0.22,
      "tag5":0.99
    }
  }
}
```

## 8.9.4 全局特征信息文件

在特征工程、在线模块、近线模块时都会用到该全局的特征信息文件。当上传的数据中的特征有变化时，用户需要同步更新该文件。该文件为JSON格式，包含特征名、特征大类、特征值类型。

表 8-57 全局特征信息文件字段描述

字段名	类型	描述	是否必选
user_features	Array[Object]	用户特征，值为数组，其元素为json对象，参见表8-58。	是
item_features	Array[Object]	物品特征，值为数组，其元素为json对象，参见表8-58。	是

表 8-58 全局特征信息文件示例表

字段名	类型	描述	是否必选
feature_name	String	特征名称	是
feature_type	String	特征类型，包含BASIC_INFO、TAGS和CONTEXT。 <b>说明</b> CONTEXT只有user_features下有。	是
feature_value_type	String	特征值类型，包含string、numerical、strArray、map。 <b>说明</b> <ul style="list-style-type: none"> <li>当feature_type为BASIC_INFO或CONTEXT时可选string,numerical和strArray。</li> <li>feature_type为TAGS时只可选map。</li> </ul>	是

### 示例

```
{
  "user_features":[
    {
      "feature_name":"provinceld",
      "feature_type":"BASIC_INFO",
      "feature_value_type":"string"
    },
    {
      "feature_name":"salary",
      "feature_type":"BASIC_INFO",
      "feature_value_type":"numerical"
    },
    {
      "feature_name":"hobbies",
      "feature_type":"BASIC_INFO",
      "feature_value_type":"strArray"
    }
  ],
}
```

```
{
  "feature_name":"user_tags",
  "feature_type":"TAGS",
  "feature_value_type":"map"
},
{
  "feature_name":"payment_type",
  "feature_type":"CONTEXT",
  "feature_value_type":"string"
}
],
"item_features":[
  {
    "feature_name":"product_name",
    "feature_type":"BASIC_INFO",
    "feature_value_type":"string"
  },
  {
    "feature_name":"order_price",
    "feature_type":"BASIC_INFO",
    "feature_value_type":"numerical"
  },
  {
    "feature_name":"categories",
    "feature_type":"BASIC_INFO",
    "feature_value_type":"strArray"
  },
  {
    "feature_name":"item_tags",
    "feature_type":"TAGS",
    "feature_value_type":"map"
  }
]
}
```