

推荐系统

# 常见问题

文档版本 01  
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 基础问题</b>	<b>1</b>
1.1 什么是 RES?	1
1.2 与其他云服务的关系	1
1.3 如何开始使用 RES?	2
1.4 获取访问密钥 (AK/SK)	3
1.5 推荐作业有哪几种创建方式?	4
1.6 创建的场景是否会立即发布?	4
1.7 最小在线并发规格支持弹性伸缩, 是否设置最小规格即可?	5
1.8 是否有样例数据支撑我进一步了解 RES?	5
1.9 什么是区域、可用区?	5
1.10 API 查询列表的接口返回结果是否支持分页?	6
<b>2 计费相关</b>	<b>7</b>
2.1 如何查看 RES 正在收费的作业?	7
2.2 如何查看 RES 消费详情?	7
2.3 数据源如何收费?	8
2.4 智能场景和自定义场景如何收费?	8
<b>3 数据源</b>	<b>9</b>
3.1 RES 的离线数据源包括什么?	9
3.2 如何上传数据至 OBS	9
3.3 如何上传实时数据?	9
3.4 离线数据和近线实时数据如何配合使用?	11
3.5 数据探索是什么? 近线实时数据如何在数据探索中的报告体现?	11
3.6 如何确定近线数据源导入实时数据成功?	11
3.7 实时数据能否立即应用到推荐场景?	11
<b>4 智能场景</b>	<b>12</b>
4.1 猜你喜欢的主要应用场景是什么?	12
4.2 关联推荐的主要应用场景是什么?	12
4.3 热门推荐的主要应用场景是什么?	12
<b>5 自定义场景</b>	<b>13</b>
5.1 推荐引擎和排序引擎有什么区别?	13
5.2 RES 支持哪些自定义策略?	13
5.3 重新运行被在线服务所引用的召回策略, 是否需要重新部署在线服务?	13

---

5.4 在线服务获得推荐的调用次数如何计算? .....	14
5.5 自定义场景关闭后, 为什么会自动启动? .....	14

# 1 基础问题

## 1.1 什么是 RES?

推荐系统（Recommender System，简称RES）基于华为大数据和人工智能技术，提供推荐平台和算法服务，并帮助企业构建个性化推荐应用，助力提升网站/APP的点击率、留存率和用户体验。

## 1.2 与其他云服务的关系

表 1-1 RES 与其他服务的关系

相关服务	交互功能
数据湖探索	数据湖探索（Data Lake Insight，简称DLI）用于推荐系统的离线计算和近线计算。DLI的更多信息请参见《 <a href="#">数据湖探索文档</a> 》。
对象存储服务	对象存储服务（Object Storage Service，简称OBS）存储RES的推荐数据源，实现安全、高可靠和低成本存储需求。OBS的更多信息请参见《 <a href="#">对象存储服务文档</a> 》。
数据接入服务	数据接入服务（Data Ingestion Service，简称DIS）提供推荐数据源的实时日志。DIS的更多信息请参见《 <a href="#">数据接入服务文档</a> 》。
统一身份认证服务	统一身份认证服务（Identity and Access Management，简称IAM）为RES提供了华为云统一入口鉴权功能和OBS与DIS的委托授权。IAM的更多信息请参见《 <a href="#">统一身份认证服务文档</a> 》。
ModelArts	ModelArts是面向AI开发者的一站式开发平台，排序策略使用ModelArts的深度学习计算能力训练得到排序模型。ModelArts的更多信息请参见《 <a href="#">ModelArts服务文档</a> 》。

## 1.3 如何开始使用 RES?

使用RES，从资源准备到在线服务完成推荐的全流程，如图1-1所示。

图 1-1 RES 操作流程

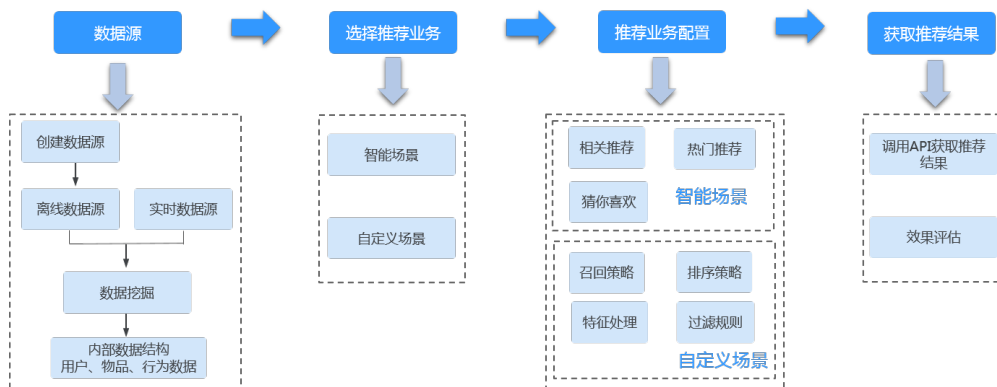


表 1-2 使用流程说明

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>
	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>

流程	子任务	说明	详细指导
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>智能场景 <a href="#">获取推荐结果</a></li> <li>自定义场景 <a href="#">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>智能场景 <a href="#">效果评估</a></li> <li>自定义场景 <a href="#">效果评估</a></li> </ul>

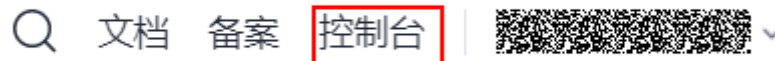
## 1.4 获取访问密钥（AK/SK）

由于使用RES时，上传数据、进行计算等需要依赖其他服务，若没有添加访问密钥则无法使用其他服务。

### 获取访问密钥

1. 登录[华为云](#)，在页面右上方单击“控制台”，进入华为云管理控制台。

图 1-2 控制台入口



2. 在控制台右上角的帐户名下方，单击“我的凭证”，进入“我的凭证”页面。

图 1-3 我的凭证



3. 在“我的凭证”页面，选择“访问密钥>新增访问密钥”，如[图1-4](#)所示。

图 1-4 单击新增访问密钥



4. 填写该密钥的描述说明，单击“确定”。根据提示单击“立即下载”，下载密钥。

图 1-5 新增访问密钥



5. 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）。

## 1.5 推荐作业有哪几种创建方式？

推荐系统支持如下几种作业创建方式：

- 通过RES管理控制台创建作业、查看推荐和效果评估结果。详情参见《[推荐系统用户指南](#)》。
- 通过API提交任务并获取结果。详情请参见《[推荐系统API参考](#)》。

## 1.6 创建的场景是否会立即发布？

新创建的智能场景和自定义场景不会立即发布。

在新创建的场景中，需要完成相关推荐配置才能完成整体场景的创建。配置完成的场景处于待发布状态，需要在场景列表页面对目标场景执行“发布”操作，才能运行场景中涉及的作业，完成场景的发布。



## 1.7 最小在线并发规格支持弹性伸缩，是否设置最小规格即可？

目前推荐系统支持的默认在线并发规格为5。由于伸缩资源有一定的时延导致预测接口出现变慢的情况，因此在可预见的并发规格中，建议填写实际并发的规格，推荐系统后台会根据并发规格提前预留好资源，避免峰值的到来。

如果需要更多的规格，请您[提工单](#)联系工程师解决。

## 1.8 是否有样例数据支撑我进一步了解 RES？

RES提供了可用来测试的全量数据，包括智能场景和自定义场景的样例测试。

- 智能场景的样例测试，可参见[智能场景（猜你喜欢）](#)。
- 自定义场景的样例测试，可参见[自定义场景（热度推荐）](#)。

## 1.9 什么是区域、可用区？

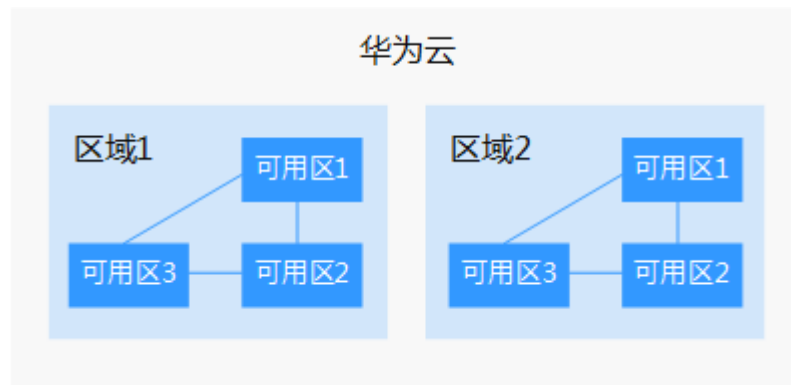
### 什么是区域、可用区？

使用区域和可用区来描述数据中心的位置，您可以在特定的区域、可用区创建资源。

- 区域（Region）：从地理位置和网络时延维度划分，同一个Region内共享弹性计算、块存储、对象存储、VPC网络、弹性公网IP、镜像等公共服务。Region分为通用Region和专属Region，通用Region指面向公共租户提供通用云服务的Region；专属Region指只承载同一类业务或只面向特定租户提供业务服务的专用Region。
- 可用区（AZ，Availability Zone）：一个AZ是一个或多个物理数据中心的集合，有独立的风火水电，AZ内逻辑上再将计算、网络、存储等资源划分成多个集群。一个Region中的多个AZ间通过高速光纤相连，以满足用户跨AZ构建高可用性系统的需求。

[图1](#)阐明了区域和可用区之间的关系。

图 1-6 区域和可用区



目前，华为云已在全球多个地域开放云服务，您可以根据需求选择适合自己的区域和可用区。更多信息请参见[华为云全球站点](#)。

## 如何选择区域？

选择区域时，您需要考虑以下几个因素：

- **地理位置**

一般情况下，建议就近选择靠近您或者您的目标用户的区域，这样可以减少网络时延，提高访问速度。

- 在除中国大陆以外的亚太地区有业务的用户，可以选择“中国-香港”、“亚太-曼谷”或“亚太-新加坡”区域。
- 在非洲地区有业务的用户，可以选择“非洲-约翰内斯堡”区域。
- 在欧洲地区有业务的用户，可以选择“欧洲-巴黎”区域。
- 在拉丁美洲地区有业务的用户，可以选择“拉美-圣地亚哥”区域。

### 说明

“拉美-圣地亚哥”区域位于智利。

- **资源的价格**

不同区域的资源价格可能有差异，请参见[华为云服务价格详情](#)。

## 如何选择可用区？

是否将资源放在同一可用区内，主要取决于您对容灾能力和网络时延的要求。

- 如果您的应用需要较高的容灾能力，建议您将资源部署在同一区域的不同可用区内。
- 如果您的应用要求实例之间的网络延时较低，则建议您将资源创建在同一可用区内。

## 区域和终端节点

当您通过API使用资源时，您必须指定其区域终端节点。有关华为云的区域和终端节点的更多信息，请参阅[地区和终端节点](#)。

## 1.10 API 查询列表的接口返回结果是否支持分页？

API查询列表的接口不支持分页。

# 2 计费相关

## 2.1 如何查看 RES 正在收费的作业？

在RES管理控制台，单击左侧菜单栏的“总览”，您可以在“总览”区域查看智能场景和自定义场景中处于“运行中”状态的作业。再根据实际情况进入对应管理页面，“终止”或“删除”对应场景停止收费。

## 2.2 如何查看 RES 消费详情？

在“费用中心”，您可以根据需求按照账期、产品类型等查询RES的消费详情。本章节以查询“账单详情”为例指导您查看计费情况，如需了解更多的账单情况，请参见[查看费用账单](#)。

查询方法：

1. 在“费用中心”中，选择“账单管理>流水和明细账单”，在流水和明细账单页面，切换“流水账单”和“明细账单”页签。
2. 在“流水账单”列表页，罗列该帐号下各种产品类型，每个任务产生的费用详细。您可以单击“操作 > 详情”，查看使用量详情。可拖动详情下方的进度条，查看“使用量”、“应付金额”等信息。

图 2-1 流水账单



交易时间	订单号/交易号	账单类型	企业项目	账号	产品类型	产品	计费模式	操作
2023/02/06 10:23:44 ...		消费-使用			--	数据库库服...	按需	详情
2023/02/06 10:23:44 ...		消费-使用			--	数据库库服...	按需	详情

3. 在“明细账单”列表页，罗列了该帐号下各种的资源的计费模式、使用量和单价等信息。可以按账期、统计维度和统计周期筛选查看明细账单。

图 2-2 明细账单

账期	企业项目	账号	产品类型	产品	计费模式	账单类型	资源名称/ID	资源标签	规格
2023/02			数据湖探索...	DLI存储空间	按需	消费-使用		--	DLI存储数据量
2023/02			对象存储服务...	云存储	按需	消费-使用		--	OBS对象存...

## 2.3 数据源如何收费？

数据源的收费包括三项。

- “OBS存储”：将离线数据源上传至OBS进行存储，根据存储空间大小进行计费。详见[OBS计费详情](#)。
- “画像数据”：在数据源导入数据成功后，按每小时每百万条画像进行计费，在删除数据源后，将停止计费。
- “训练作业”：在创建离线作业和导入实时数据均需要选取训练规格，根据训练规格大小进行计费。

### 📖 说明

如果需要数据源停止计费，请先终止近线数据源的导入，删除对应数据源和对应OBS中的数据存储。

## 2.4 智能场景和自定义场景如何收费？

各场景是按需计费，根据您选择的不同规格，价格不同。训练作业运行一次，根据此次运行时耗费的资源进行计费。当该场景作业处于结束状态，如“运行失败”状态，将停止计费。

每“执行”一次训练作业，都会产生计费，“运行中”的场景均处于计费中。

# 3 数据源

## 3.1 RES 的离线数据源包括什么？

离线数据包括如下几张表：

- 用户属性表
- 物品属性表
- 用户操作行为表

每张表的字段描述和规范详情请参见《推荐系统用户指南》中[准备离线数据源](#)章节。

## 3.2 如何上传数据至 OBS

使用RES进行推荐时，您需要将数据上传至对象存储服务（OBS）桶中。您可以登录[OBS管理控制台](#)创建OBS桶，并在您创建的OBS桶中创建文件夹，然后再进行数据的上传。创建OBS桶和上传文件夹的操作指导请参见[创建桶](#)和[上传对象](#)。

### 说明

您在创建OBS桶时，需保证您的OBS桶与RES在同一个区域。

## 3.3 如何上传实时数据？

推荐系统支持您通过SDK上传实时数据，具体操作方法如下。

### 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 mrr

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();
```

### 3.4 离线数据和近线实时数据如何配合使用？

在推荐系统初始化阶段，需要用户提供批量的离线数据源并按照推荐系统要求的数据格式上传至OBS，完成数据的检测和导入。

近线实时数据源推荐使用RES SDK上传，此操作所有的数据更新都是实时生效的。

### 3.5 数据探索是什么？近线实时数据如何在数据探索中的报告体现？

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，由于近线实时数据会实时入库，因此可以通过定时执行数据探索任务来覆盖增量数据。

### 3.6 如何确定近线数据源导入实时数据成功？

近线数据源上传成功后，可在推荐系统控制台的“数据源”直接查看。

### 3.7 实时数据能否立即应用到推荐场景？

需要确认关联的召回策略，近线召回可以，离线召回不可以。离线召回需要重新执行。

# 4 智能场景

---

## 4.1 猜你喜欢主要应用场景是什么？

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。

## 4.2 关联推荐的主要应用场景是什么？

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。

## 4.3 热门推荐的主要应用场景是什么？

热门推荐只要适用于首页、热点类场景，满足流行度统计，有效吸引新用户。



# 5 自定义场景

## 5.1 推荐引擎和排序引擎有什么区别？

### 推荐引擎

推荐引擎是以推荐为业务逻辑的引擎，即系统根据配置生成召回集作为起点，输出推荐结果集为终点的引擎。

### 排序引擎

排序引擎是以排序为业务逻辑的引擎，即用户提供排序集为输入，系统根据排序算法输出排序结果的引擎。

## 5.2 RES 支持哪些自定义策略？

目前RES所支持的推荐策略如下所示：

- 召回策略
- 过滤规则
- 特征工程
- 排序策略
- 近线策略
- 效果评估

推荐策略详细信息请参见《推荐系统用户指南》[算法介绍及其参数说明](#)章节。

## 5.3 重新运行被在线服务所引用的召回策略，是否需要重新部署在线服务？

不需要。重新执行召回策略，产生新的候选集会被在线服务直接引用，无需重启在线服务。

## 5.4 在线服务获得推荐的调用次数如何计算？

RES从全局角度计算在线服务获得推荐的调用次数，不区分每次调用的用户。例如A用户调用请求推荐接口是每秒5次，B用户调用请求推荐接口每秒5次，当A用户和B用户同时调用此接口时，总的获得推荐的调用请求为A用户和B用户之和，即 $5+5=10$ 。

## 5.5 自定义场景关闭后，为什么会自动启动？

在创建自定义场景时，如果设置了自动召回策略，且此召回策略关联了在线服务，就会自动运行场景实例。用户可关闭召回策略，或者在在线服务中删除依赖的这个策略。