

推荐系统 RES

产品介绍

文档版本 02
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

1 什么是推荐系统.....	1
2 应用场景.....	2
3 基本概念.....	4
4 产品功能.....	6
5 使用限制.....	7
6 计费说明.....	8
7 与其他云服务的关系.....	10
8 权限管理.....	11
9 配额说明.....	14

1 什么是推荐系统

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。

RES 优势

- **开放式推荐**

提供完整的推荐平台和原子推荐算法，不绑定客户的运营场景，客户可以在华为云上根据自己对推荐算法和运营规则的理解，自定义专属的推荐流程。当前只有华为云提供开放式推荐能力，其他友商主打场景式推荐。

- **场景式推荐**

提供多维度的场景推荐，含猜你喜欢、关联推荐、热门推荐，一键式操作，降低客户接入门槛。

- **近线处理能力**

支持实时数据的接入和更新、模型在线学习，近线处理实时训练兴趣模型。

- **全面的推荐实体**

支持以用户推荐物品、以用户推荐用户、以物品推荐物品、以物品推荐用户四种全面的推荐对象，用户根据场景选择不同的推荐实体。

- **独立的排序模块**

独立的基于CTR预估的排序打分模块，支持个性化排序能力。

如何访问 RES

您可以通过以下任何一种方式访问RES。

- **管理控制台**

管理控制台是基于浏览器的可视化界面。通过管理控制台，您可以使用直观的界面进行相应的操作。使用方式请参见《[推荐系统用户指南](#)》。

- **REST API**

RES提供REST形式的访问接口，您能够非常容易地从Web应用中访问RES。具体操作和API详细描述，请参见《[推荐系统API参考](#)》。

2 应用场景

推荐系统支持深度智能挖掘用户和物品的关联关系，将对应场景的推荐结果推送给用户，代替低纬度的人工规则，提升了相关运营指标和用户的体验。包含了互联网信息流，短视频/直播/音乐/阅读，广电媒资，社交，电商等场景。

RES+电商应用场景

场景描述

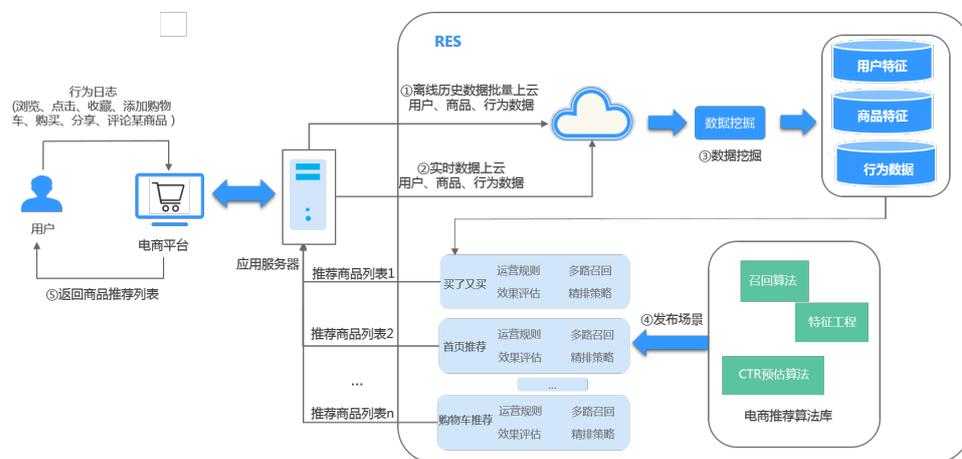
电商场景中，通常涉及首页推荐、购物车推荐、买了又买等推荐场景，但各个子场景的运营规则均不一致。

RES提供一站式电商推荐解决方案，在一套数据源下，支持多种电商推荐场景，提供面向电商推荐场景的多种推荐相关算法和大数据统计分析能力。

场景优势

- 能够精确匹配电商运营规则。
- 最近邻算法与深度学习的结合，挖掘用户高维稀疏特征，匹配最佳推荐结果。
- 融合多种召回策略，网状匹配兴趣标签。
- 改善用户体验，同时降低人工成本。
- 画像与深度模型结合，助力营收收益增长。

图 2-1 RES 电商推荐



RES+媒资应用场景

场景描述

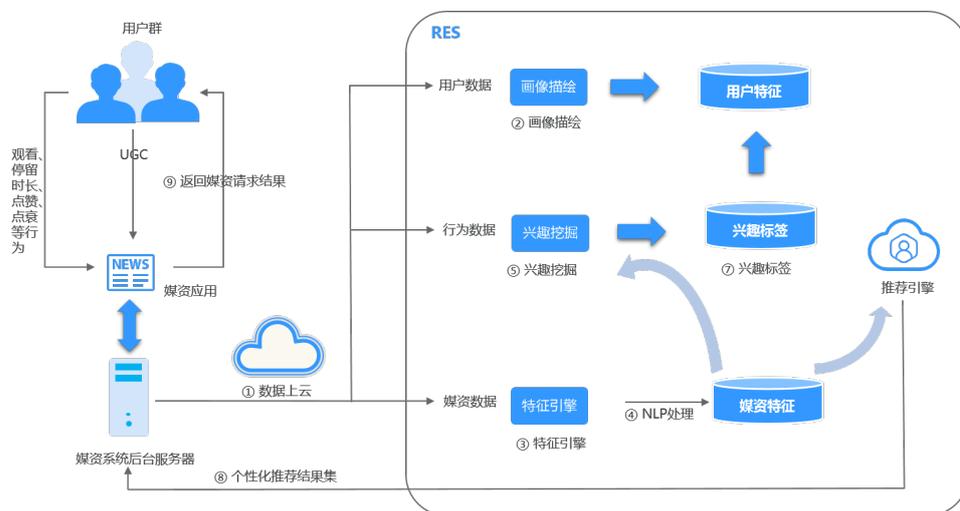
媒资推荐场景中，通常对实时性要求比较高，用户产生的行为需要得到即时的反馈，同时结合用户的长期兴趣和短期兴趣进行个性化推荐。

RES提供一站式媒资推荐解决方案，支持针对行为数据实时生成用户的兴趣标签，提供离线、近线、在线三层计算，完成千人千面的个性化媒资推荐。

场景优势

- 可以实现7*24小时，智能学习用户行为，构建兴趣模型。
- 兴趣文章命中率高，用户粘性增强，PV增幅明显。
- 减少人工运营规则的摄入，减低人力成本。
- 全流程自动化，批/流训练结合，稳定可靠。

图 2-2 RES 媒资推荐



RES+房产应用场景

场景描述

推荐系统助力房产企业APP实现首页推荐、详情推荐和个人中心推荐。

场景优势:

- 支持基于经纬度的向量召回，根据地理位置召回高匹配度的附近房源。
- 特征标签网状匹配。
- 通过卡证识别，快速完成快递录单、手机开户等场景信息录入，实名认证。

3 基本概念

推荐系统

推荐系统，是全力提供媒资、短视频、电商等行业的推荐系统解决方案，帮助互联网企业降低构建推荐应用的技术门槛，提升点击率、留存率和用户体验。

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。

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。

用户

推荐系统被推荐的对象，一般是指使用业务系统的客户。例如，某电商的客户。

物品

被推荐的内容，一般是指业务系统提供的给其用户的商品。例如，某视频网站的视频。

召回策略

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。

过滤规则

过滤规则用于生成推荐的过滤集，包含黑白名单、特征过滤等特性。支持用户在线上推理过程中完成对相关物品的过滤。

特征工程

特征工程常用于对原始数据进行特征挖掘的处理，形成的结果用于排序策略的训练。

排序策略

排序策略利用CTR预估或综合性计算的算法给候选集做打分。

在线服务

在线服务应用于做线上推荐，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。

推荐引擎

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

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。

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。

打散

打散是指推荐的结果集中根据物品表的category字段进行打散，避免推荐结果集过于集中。

4 产品功能

数据源

数据源功能可以在用户上传数据后，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。提供数据源智能检测，输出数据分布和数据质量信息等，智能完成特征工程。

智能场景

根据业务场景选择对应的智能推荐场景，快速搭建专属推荐系统。主要应用为猜你喜欢、关联推荐、热门推荐。

功能优势：

- 多维度管理，支持运营规则设置，一站式推荐平台。
- 自动挖掘特征，采用AUTOML完成特征的自动挖掘和组合，提高特征选择效率。
- 高适用性，多种模板选择，适用多个应用场景。
- 使用便捷，一键式构建推荐系统，提供标准API接口，调用简单，便于被集成。
- 实时更新，具备实时更新能力，更快反馈用户的精准需求。
- 降低成本，减少人力支出投入。

自定义场景

基于用户历史行为计算物品相似性，实时更新候选列表，提升用户体验，提高转化率支持多种召回、过滤、排序算子自由组合，训练形式上支持离线批处理、近线流处理、在线实时处理的三种数据处理方式，提供完备的一站式推荐平台，可快速设置运营规则进行AB测试。

功能优势：

- 全开放推荐流程，用户根据业务自定义推荐流程。
- 特征工程，特征处理多样化，支持自定义特征散列等。
- 丰富的推荐策略，提供丰富的召回、过滤、排序算子。
- 运营助手，支持多维度运营，根据业务情况及时调整推荐结果。
- 效果评估，实时效果评估，推荐指标图表展示。

5 使用限制

在使用RES时，需注意以下使用限制。

- 建议使用支持的浏览器登录RES服务。
 - Google Chrome : 43.0及更高版本。
 - Mozilla FireFox : 38.0及更高版本。
 - Internet Explorer : 9.0及更高版本。
- 推荐系统属于高并发低时延场景，建议使用私有网络获取推荐结果。

6 计费说明

计费项

RES服务根据用户使用的不同资源分别进行收费。

📖 说明

- 计费支持区域：华北-北京四。
- 由于RES使用的离线数据需存储在OBS中，数据存储产生的费用，请参见《[OBS价格说明](#)》。

表 6-1 推荐系统计费项说明

计费项	说明
存储资源	应用于物品画像和用户画像的存储计费，对用户和物品的总条目数统计进行收费。 每个数据源默认规格最低是一百万条，如果一个数据源中的用户和物品总条目数不超过100W条，则按照100W条计费，如果超过一百万条，每一百万条为一个计算周期。
计算资源	应用于召回、排序等作业运行时CPU/GPU计算，包含RES的数据源、推荐场景的离线作业计算计费。 如果选取4CU计算规格执行一小时作业运行，则按需消耗为0.49*4=1.96元。

计费项	说明
在线服务	<p>应用于在线服务预测计费。</p> <p>TPS统计规则为每小时平均TPS，例如每秒调用5次，持续调用1小时，TPS即为5，这一小时按需计费消耗为$5 \times 0.95 = 4.95$元。</p> <p>最终纳入计费的有效TPS值由两部分决定：</p> <ul style="list-style-type: none">配置TPS：客户在场景中配置的在线服务TPS。 <p>说明</p> <p>在线服务TPS默认值为5，如果真实调用量不超过5或者没有调用，也会按照基础配置TPS为5来收取基本费用。</p> <ul style="list-style-type: none">真实TPS：按照预测接口在区间内的“平均值”计算得出。 <p>TPS每小时话单计费值由配置TPS和真实TPS共同决定，当配置TPS < 真实TPS时，按照真实TPS计费。当配置TPS > 真实TPS时，按照配置的TPS进行计费，即有效TPS为配置TPS和真实TPS的最大值。</p> <p>例如：当系统DAU为10W，并且推荐接口在登录首页时触发，在线服务一般可配置为10TPS。</p>

计费模式

RES目前提供按需、预付套餐包共2种计费方式。

- 按需购买**：这种购买方式比较灵活，可以即开即停。
- 预付套餐包**：客户预先购买一定的资源使用量配额，在按需使用过程中，系统优先扣减配额，超出配额的使用量才需要额外根据按需费用付费。购买的预付套餐包为在有效期内单位规格的计算时长。例如，您购买了计算型CPU（1U4G）实例20000小时，指单位规格1CU运行20000小时。如果在数据源或者场景中配置计算规格为4CU的话，则可以运行5000小时。同理，存储资源和在线服务也是这种模式。套餐包的有效期为一年，如果在一年内没用完相应的资源，系统会自动清除剩余的资源配额。

续费

为避免产品到期未再次购买及时续费，导致资源数据被删除，您可以在“续费管理”页面中选择“再次购买”，支付完成后再次购买的套餐包将立即生效。您也可以进行续费操作，详情请参见[续费管理](#)。

欠费

客户欠费后，可以查看欠费详情。为防止相关资源被停止或者释放，需要及时进行充值。具体欠费策略请参见[欠费还款](#)。

7 与其他云服务的关系

表 7-1 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《 数据湖探索文档 》。
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《 对象存储服务文档 》。
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《 数据接入服务文档 》。
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《 统一身份认证服务文档 》。
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《 ModelArts服务文档 》。

8 权限管理

如果您需要对华为云上购买的RES资源，给企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制华为云资源的访问。

通过IAM，您可以在华为云账号中给员工创建IAM用户，并使用策略来控制他们对华为云资源的访问范围。例如您的员工中有负责软件开发的人员，您希望他们拥有RES的使用权限，但是不希望他们拥有删除RES任务等高危操作的权限，那么您可以使用IAM为开发人员创建用户，通过授予仅能使用RES，但是不允许删除RES的权限策略，控制他们对RES资源的使用范围。

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

IAM是华为云提供权限管理的基础服务，无需付费即可使用，您只需要为您账号中的资源进行付费。关于IAM的详细介绍，请参见《[IAM产品介绍](#)》。

RES 权限

默认情况下，管理员创建的IAM用户没有任何权限，需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的用户获得策略定义的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对云服务进行操作。

RES部署时通过物理区域划分，为项目级服务。授权时，“作用范围”需要选择“区域级项目”，然后在指定区域（如华北-北京1）对应的项目（cn-north-1）中设置相关权限，并且该权限仅对此项目生效；如果在“所有项目”中设置权限，则该权限在所有区域项目中都生效。访问RES时，需要先切换至授权区域。

根据授权精细程度分为角色和策略。

- **角色**：IAM最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度，提供有限的服务相关角色用于授权。由于华为云各服务之间存在业务依赖关系，因此给用户授予角色时，可能需要一并授予依赖的其他角色，才能正确完成业务。角色并不能满足用户对精细化授权的要求，无法完全达到企业对权限最小化的安全管控要求。
- **策略**：IAM最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式，能够满足企业对权限最小化的安全管控要求。例如：针对ECS服务，管理员能够控制IAM用户仅能对某一类云服务器资源进行指定的管理操作。多数细粒度策略以API接口为粒度进行权限拆分，RES支持的API授权项请参见《[权限策略和授权项](#)》。

如表8-1所示，包括了RES的所有系统权限。

表 8-1 RES 系统策略

系统策略名称	描述	类别
RES FullAccess	推荐系统服务管理员权限，拥有该权限的用户可以操作并使用所有推荐系统服务。	系统策略
RES ReadOnlyAccess	推荐系统服务只读权限，拥有该权限的用户仅能查看推荐系统服务数据。	系统策略

RES常用操作与系统策略的授权关系如表8-2所示，您可以参照该表选择合适的系统策略。

表 8-2 常用操作与系统策略的关系

操作	RES FullAccess	RES ReadOnlyAccess
新增工作空间	√	x
查询工作空间详情	√	√
查询工作空间列表	√	√
修改工作空间	√	x
删除工作空间	√	x
新增数据源	√	x
查询数据源详情	√	√
查询数据源列表	√	√
修改数据源	√	x
删除数据源	√	x
新增场景	√	x
查询场景	√	√
查询场景列表	√	√
修改场景	√	x
删除场景	√	x
运行场景	√	x
新增作业	√	x
查询作业详情	√	√
查询作业列表	√	√

操作	RES FullAccess	RES ReadOnlyAccess
修改作业算子	√	x
删除作业	√	x
运行作业	√	x

相关链接

- [IAM产品介绍](#)
- [创建用户组、用户并授予RES权限](#)
- [策略支持的授权项](#)

9 配额说明

为防止资源滥用，平台限定了各服务资源的配额，对用户的资源数量和容量做了限制。

表 9-1 RES 服务配额

资源	限制条件	建议
推荐引擎预测接口中最多请求结果数量	20	可 提工单 支持更高规格。
单份画像数据中最多支持的特征数量	30	
单场景在线服务最多支持每秒请求的次数(TPS)	200	
数据源个数	5	
场景个数	10	
单场景下推荐预测返回的结果集个数	20	

如果当前资源配额限制无法满足使用需要，您可以申请扩大配额。配额查看及修改请参见[关于配额](#)。