




2016 杭州·云栖大会
THE COMPUTING CONFERENCE

云栖社区
yq.aliyun.com

基于阿里云数加平台的 大数据Serverless 实践

2016
The Computing
Conference

班输

主办单位:  杭州

 Alibaba Group
阿里巴巴集团

战略合作伙伴: 



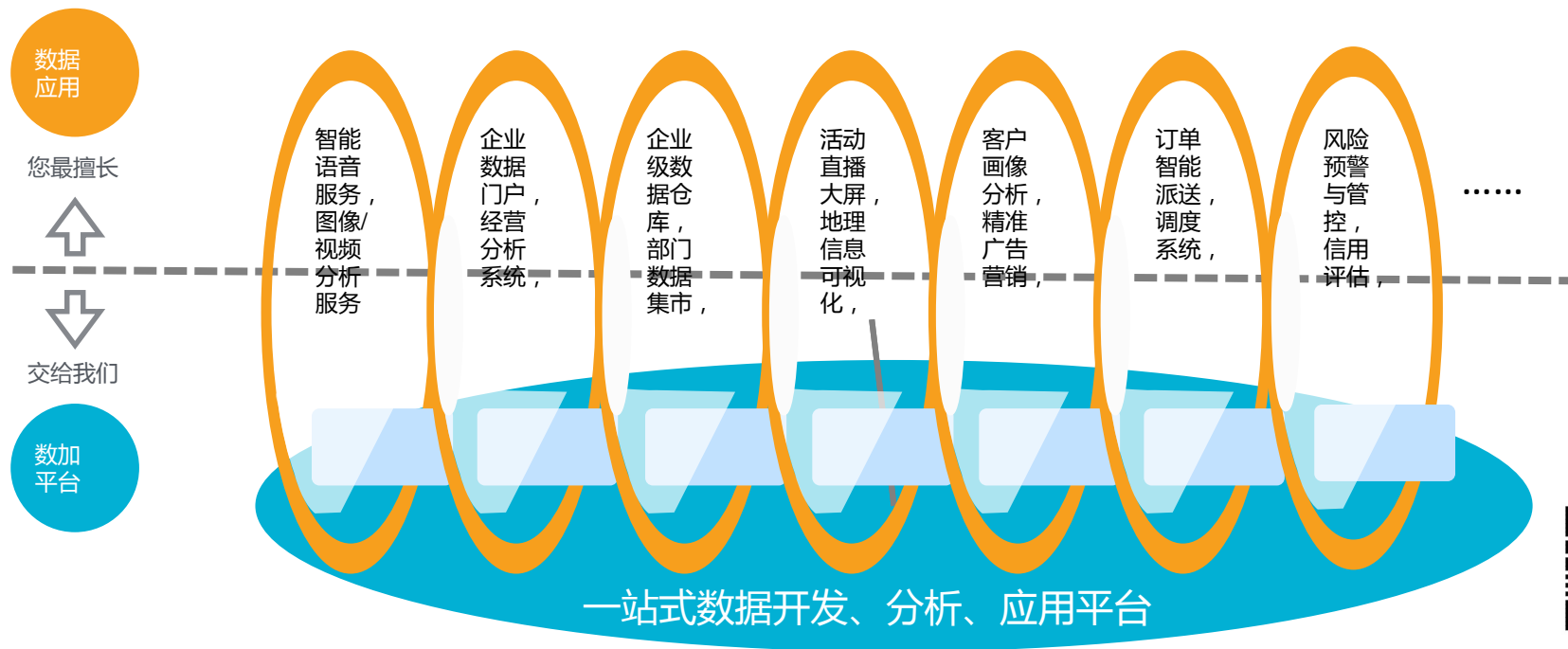
扫码观看大会视频

—————	数加平台简介
目 录	大数据应用的特点
content	数加平台Serverless 架构解析
—————	典型场景实践分享



数加平台简介

数加是阿里云大数据的品牌名，旗下包含一系列的大数据产品及服务。



扫码观看大会视频

大数据产品概览



大数据应用的特点

1

流程长，业务逻辑复杂

从数据的采集，存储，分析，挖掘到最终提供数据服务
需要把多种数据源融合，关联分析

2

场景多样化，不确定性强

大数据相关的应用要产生价值，需要业务紧密结合，很多场景都是探索性质，并且要随着业务变化和反馈持续的调整

3

技术门槛高

多种计算引擎，离线，在线，流式，多种计算模型，
SQL,MR，大规模机器学习的组合使用

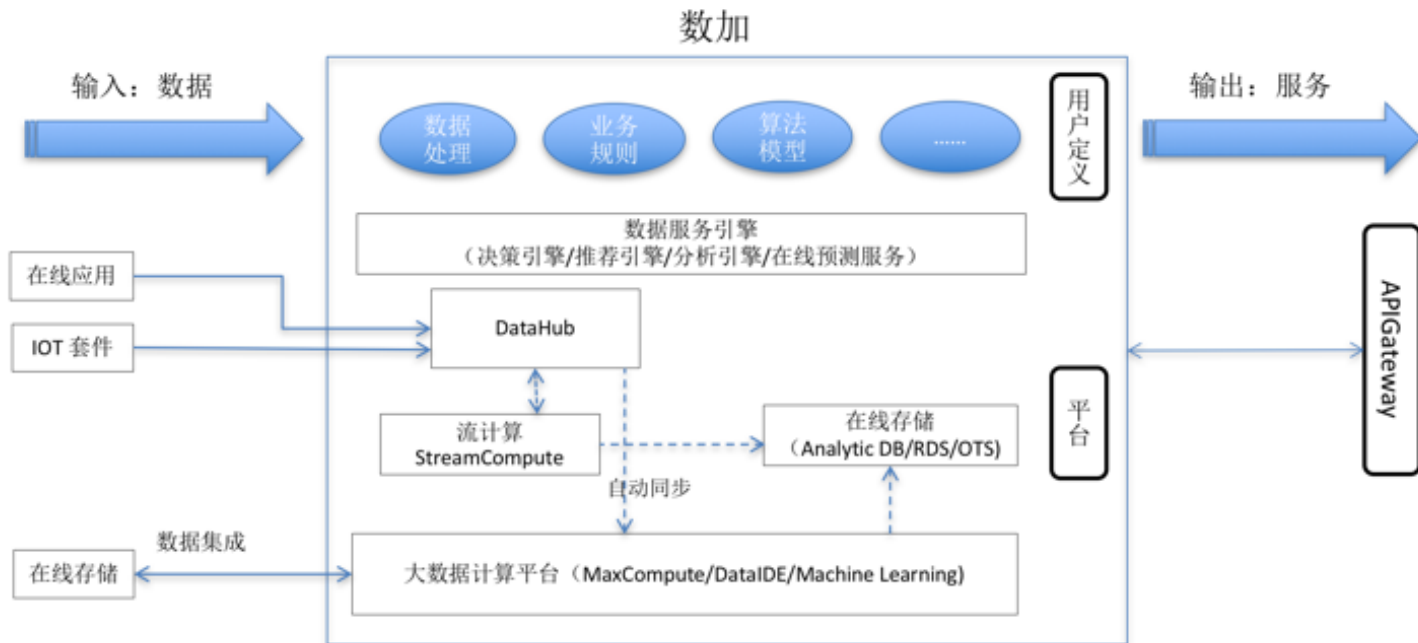


如何把数据科学家从基础设施运维的复杂性中解放出来？

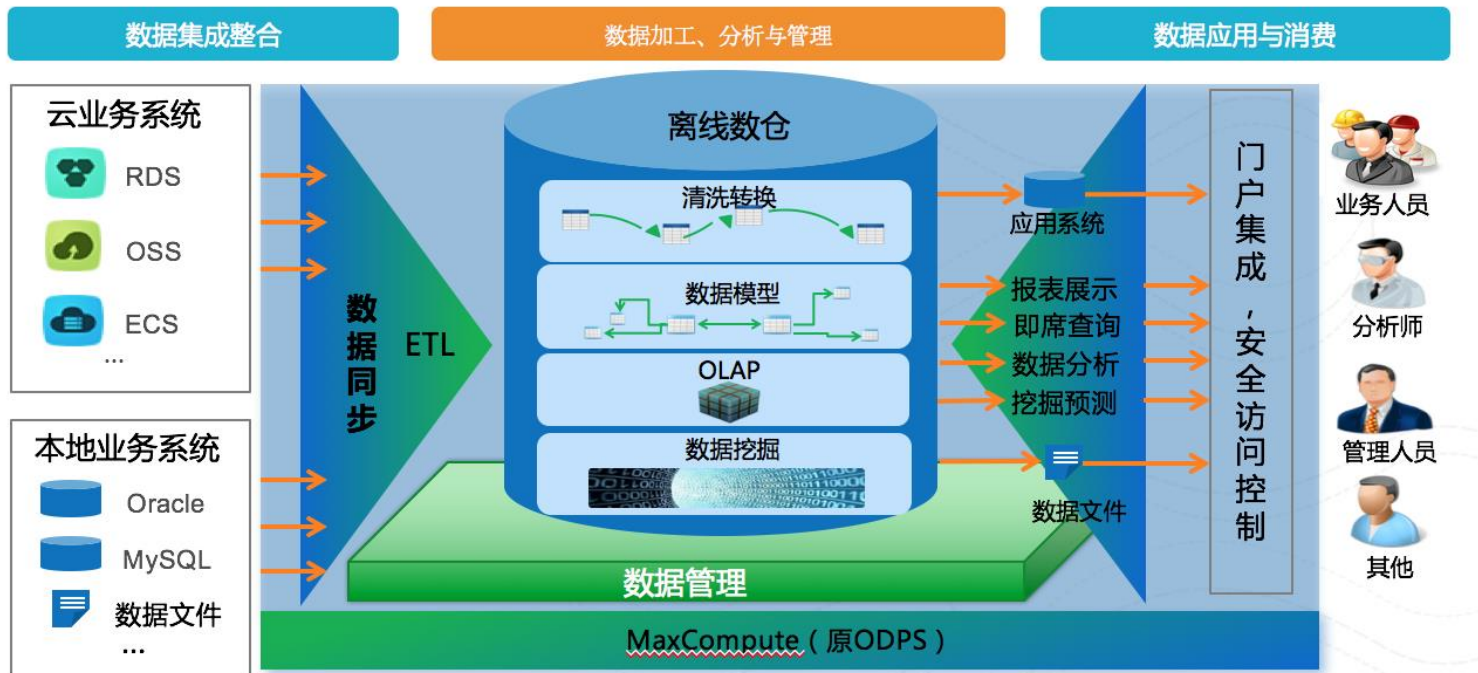
数加平台给出的答案：Serverless 架构



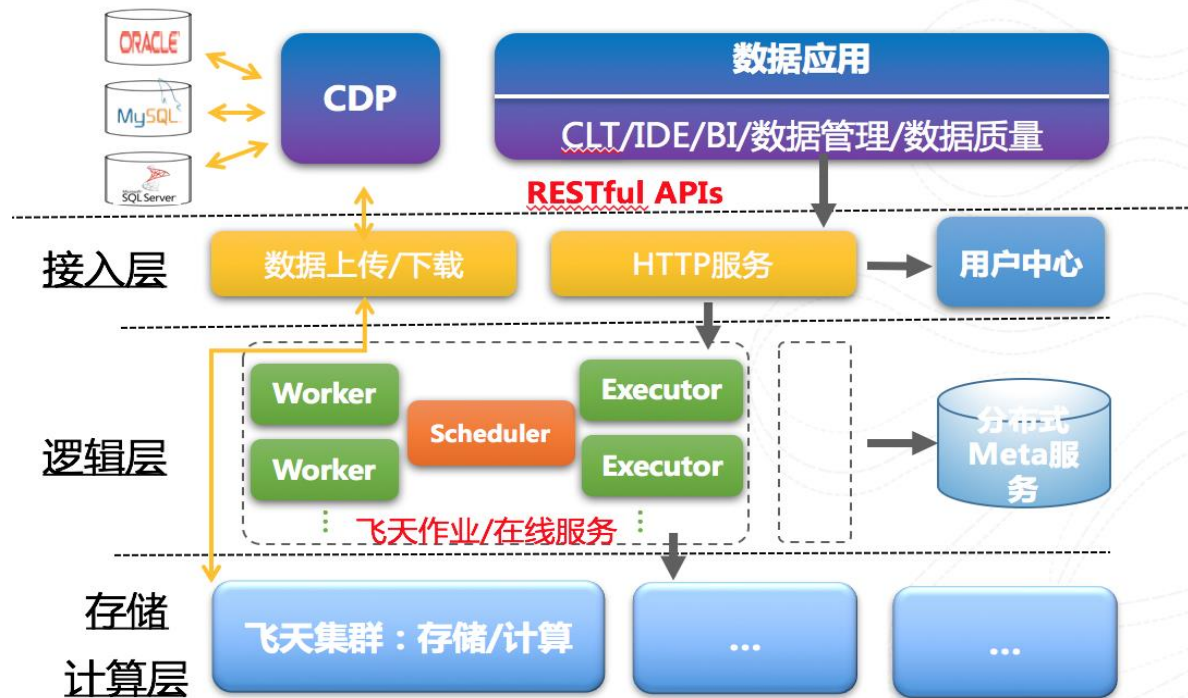
数加Serverless 架构解析



Serverless 架构产品典范：MaxCompute



MaxCompute 架构



MaxCompute 服务化的架构，提供的是大数据计算的能力，而不是计算的工具或者环境



典型场景实践分享

1

物联网大数据应用

组合使用数加平台提供的Datahub, StreamCompute, MaxCompute, DataV 等服务实现物联网数据的分析应用

2

预测即服务

组合使用数加平台提供的MaxCompute, 机器学习, 在线预测等服务实现孩子成绩的预测

3

个性化推荐

使用数加平台的推荐引擎快速搭建个性化推荐系统

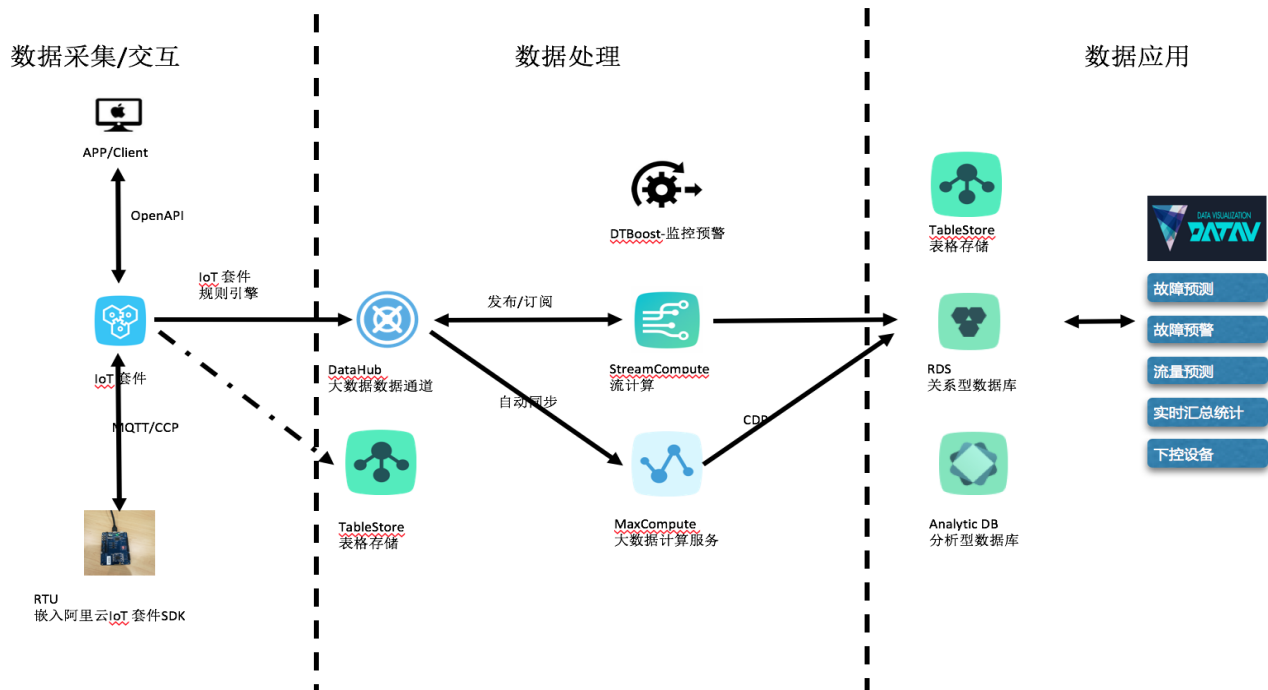


物联网案例：智慧水务



扫码观看大会视频

物联网大数据应用参考架构



扫码观看大会视频

在线预测部署

当前模型状态: 部署成功 当前版本: 1 部署时间: 2016-08-30 11:50:38 请查看页面下方信息进行

行接口调用。如需更新, 请点击: [重新部署](#)

查看历史版本信息, 请点击版本进行切换, 重新部署新增预测服务, 不会覆盖原有服务

接口模式

SDK模式

返回样例

预测服务endpoint: <http://prediction.odps.aliyun.com>

部署project: garvin_test

在线模型名称: xlab_m_logisticregress_191457_v1

接口方式: Restful Api支持Json和Protobuf

返回格式: JSON/XML

接口样例:

POST

http://prediction.odps.aliyun.com/projects/garvin_test/onlinemodels/xlab_m_logisticregress_191457_v1

1 HTTP/1.1

Authorization: ODPS

AccessId:AccessKey

Date: Tue, 31 Mar 2015 06:32:27 GMT

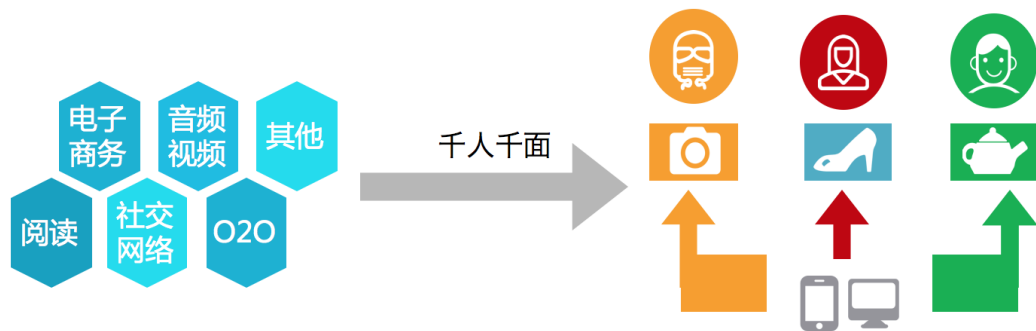
Content-Type: application/json

Host: 10.97.180.57



扫码观看大会视频

个性化推荐案例



典型客户:



扫码观看大会视频

五个步骤，轻松实现个性化推荐

1 添加资源



2 添加业务
配置数据



3 配置推荐
场景算法



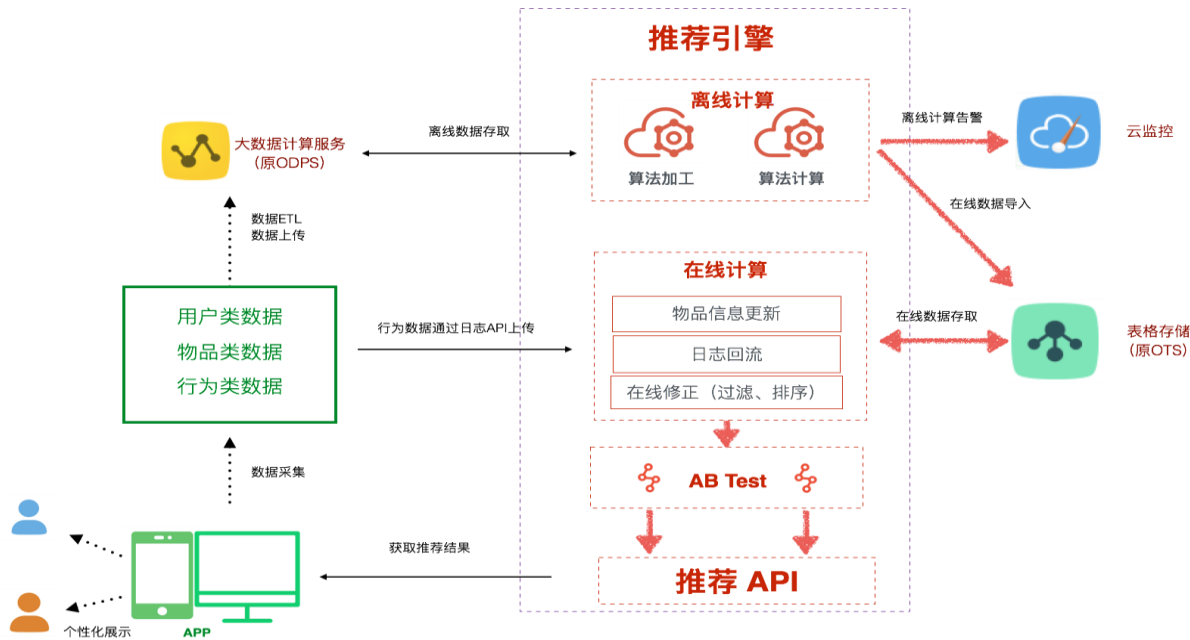
4 API对接



5 查看效果
报表



推荐引擎系统架构



通过场景管理配置推荐算法

推荐场景是指在客户的APP中使用推荐功能的模块名称，场景隶属于业务，使用到的数据就是业务中配置的数据。

场景包含一个或多个算法流程，支持AB Test测试。每一个算法流程代表一种推荐物品的逻辑，由多个算法拼装组成。

提供多种算法模板供客户选择

添加计算路径

路径代码* 12个字符以内，英文

计算路径类型 普通计算路径

离线计算模版 test1

在线计算模版 spl_on_tpl simple_online_tpl

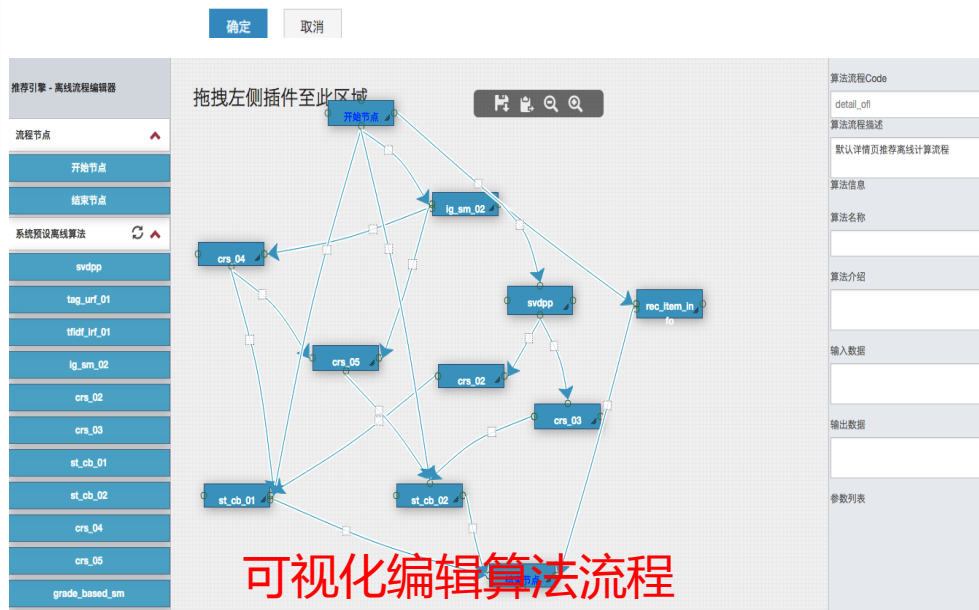
Close 提交

支持同一个推荐位同时测试多种推荐算法

计算路径管理

添加计算路径

路径代码	路径类型	离线计算流程	在线计算流程	流量占比	操作
path1	普通计算流程	修改	修改	10%	删除
path_2	普通计算流程	修改	修改	90%	删除
path_3	普通计算流程	修改	修改	0%	删除

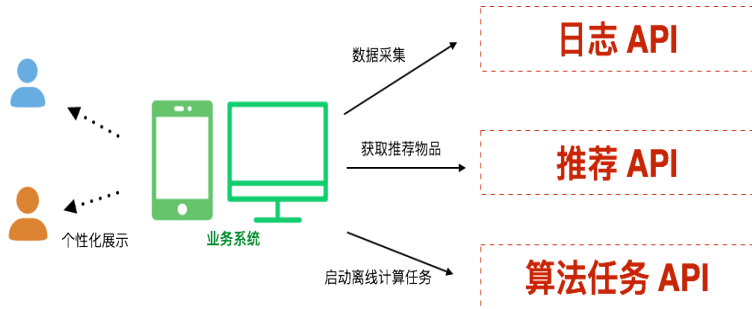


系统之间的对接通过API方式实现，API包括日志API、推荐API、算法任务API三类。

日志API用于接收您的业务系统采集的数据，以行为类数据为主。

推荐API提供推荐的物品列表，用于在您的业务系统中展示给消费者进行推荐。

算法任务API用于启动离线计算流程的算法任务、查看任务状态等。



API

请选择接口:

Admin.Admin.addNodeCategory post /admin/cat

请求方法:

post

请求地址:

/ars/admin/cat

请求参数:

json格式

调试接口

返回结果:



扫码观看大会视频

2016 The
Computing
Conference
THANKS

