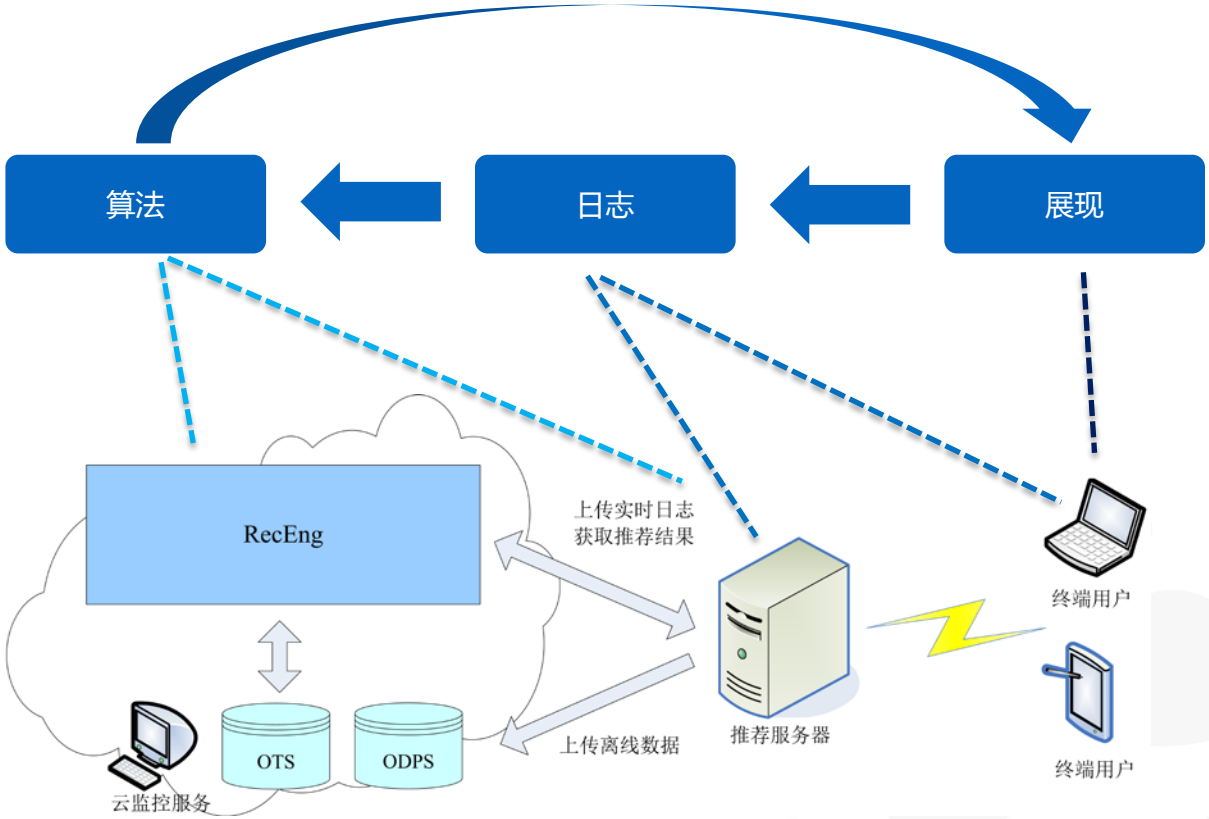
郑重（卢梭），阿里云技术专家

21天搭建推荐系统

目 录
content

阿里云推荐引擎介绍



推荐系统一般包括展现子系统、日志子系统和算法子系统三个部分

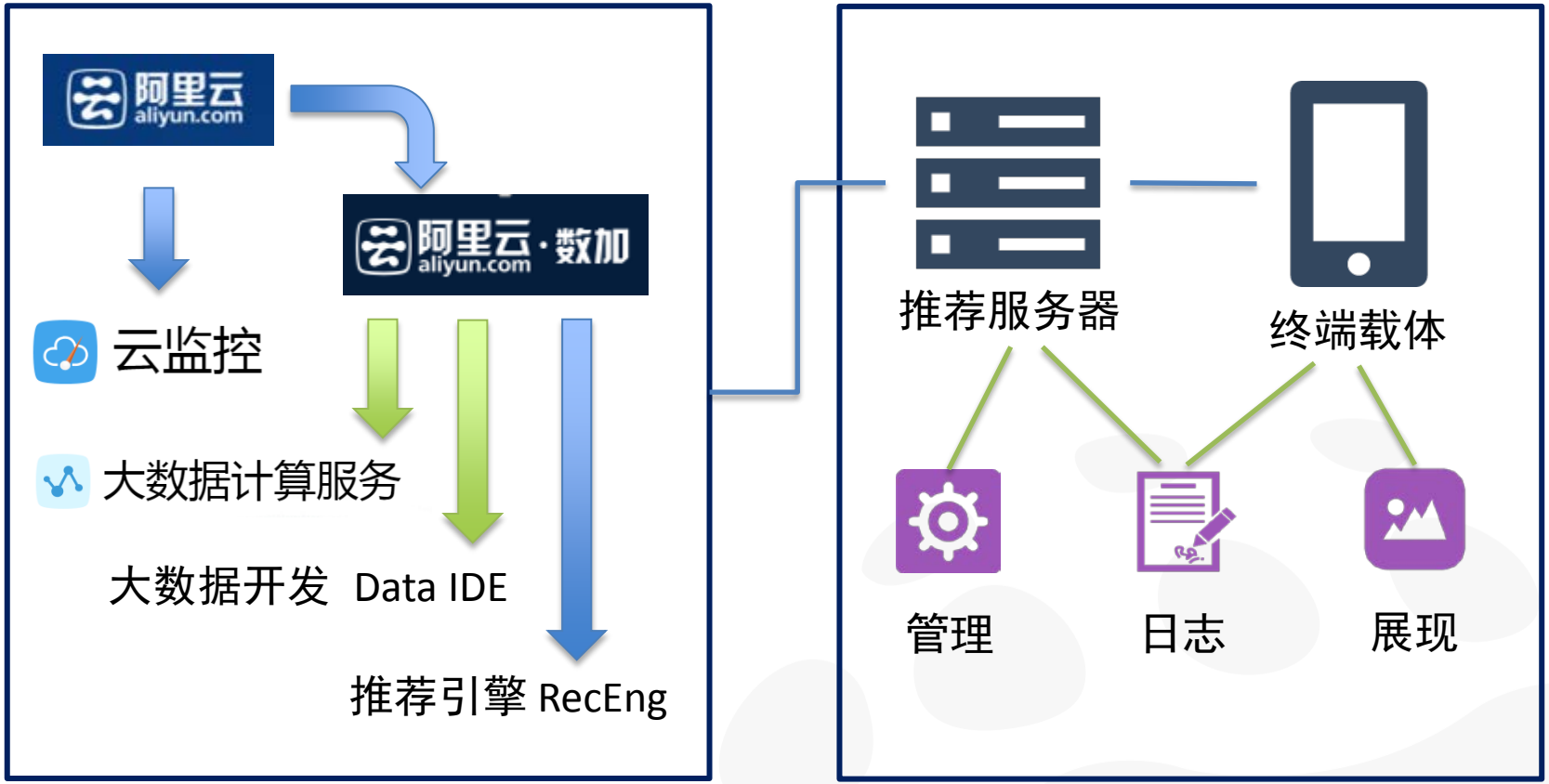
阿里云推荐引擎（RecEng）是推荐系统的一部分，主要实现的是算法子系统，需要和其他子系统配合工作

http://data.aliyun.com/product/re

目 录
content

第一阶段：基本功能

Day1. 环境准备


阿里云 aliyun.com
阿里云·数加 aliyun.com
云监控
大数据计算服务
大数据开发 Data IDE
推荐引擎 RecEng
推荐服务器
终端载体
管理
日志
展现

Day2-3. 数据准备

格式规范

行为表（user_behavior）

字段描述

列名	数据类型	注释
user_id	string	用户ID
item_id	string	物品ID
bhv_type	string	行为类型： view：物品曝光 click：用户点击物品 collect：用户收藏了某个物品 uncollect：用户取消收藏某个物品 search_click：用户点击搜索结果中的物品 comment：用户对物品的评论 share: 分享

Day2-3. 数据准备

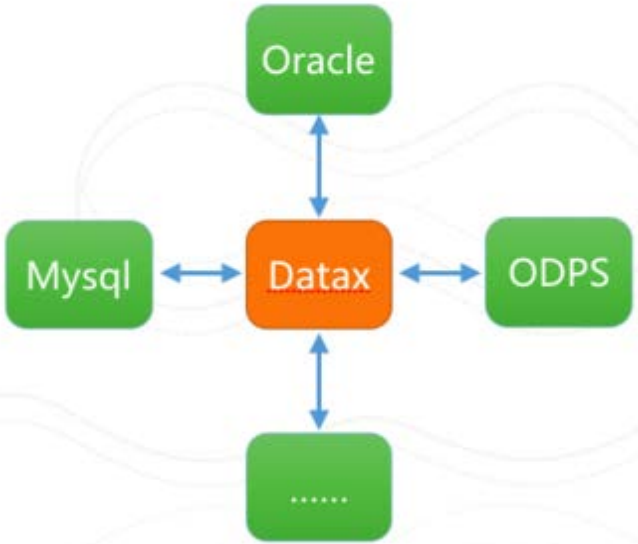
数据上传

利用Tunnel命令上传

1. 集成在ODPS console中
2. tunnel upload log.txt test_project.test_table/p1="b1"

定制DataX上传

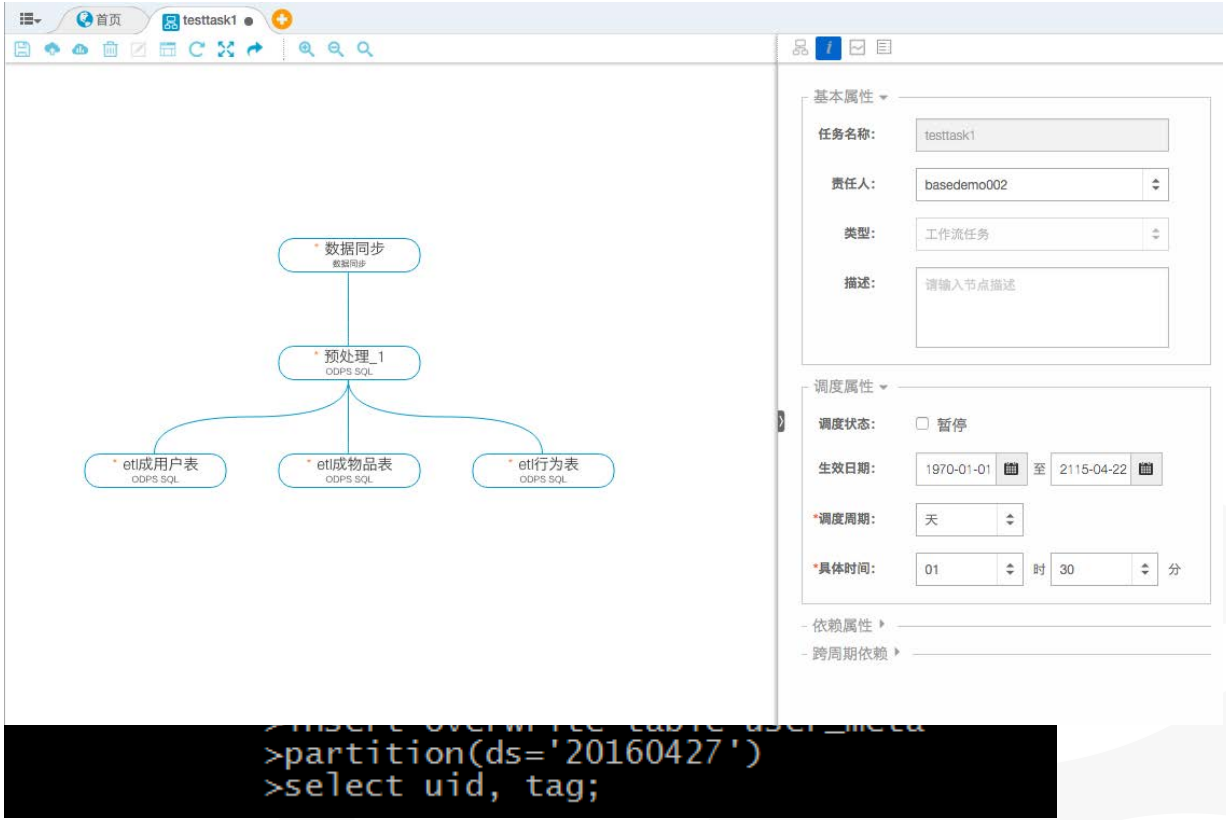
1. 下载DataX
2. 配置DataX的配置文件
3. python datax.py ./mysql2odps.json



Day2-3. 数据准备

格式转换

大数据开发 Data IDE



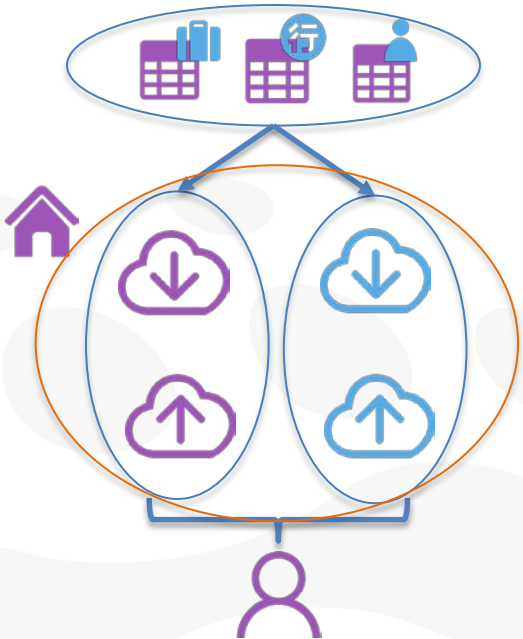
大数据计算服务

Day4-5. 基本配置和离线计算

业务

场景

算法流程



手淘首页

必抢、清单会场首图个性化

人群会场首图个性化

类目会场首图个性化

Day4-5. 基本配置和离线计算

配置业务基本信息 业务Code（即biz_code）唯一标示业务的配置信息

业务名称* 20字以内,支持中文、英文、数字

业务Code* 12个字以内，支持数字和英文字母,业务代码提交后

配置场景和算法 ☑自动配置场景和算法 系统自动为当前业务添加首页推荐和相关推荐两个场景，并为每个场景配置一个默认的推荐算法流程。您可以在场景管理中进行更改。

配置业务依赖云资源 本产品依赖您开通的大数据计算服务(Max Compute,原名ODPS)和表格存储(Table Store)资源，您可以在资源管理中进行添加和管理

大数据计算资源* lusuo_test_odps

表格存储资源* lusuo_test_ots

配置业务数据表 ☐使用本产品日志API接收日志

用户表*	user_table	描述用户元信息的数据表
用户属性维度表	usermeta_config_table	定义用户表中各元信息数据的类型
物品表*	item_table	描述物品元信息的数据表
物品属性维度表	item_config_table	定义物品表中各元信息数据的类型
行为表*	behavior_table	描述用户对物品的行为（包括搜索、点击、浏览、收藏等）信息的数据表
可推荐物品表	rp_table	可被推荐给用户的物品信息数据表，通常是物品表的子集

关闭 提交

新建推荐场景 返回我的推荐

1 配置推荐场景 2 配置API参数 3 配置推荐算法流程

推荐所需的数据来自于业务，业务定义了算法所能使用的数据范围。 了解更多

推荐场景code * ls_scn1

推荐场景名称 * 测试首页推荐

所属业务 * lusuo_biz1 lusuo_test

上一步 下一步

新建推荐场景 返回我的推荐

1 配置推荐场景 2 配置API参数 3 配置推荐算法流程

配置"获取推荐结果API"所需要的输入参数。

场景指的是推荐的上下文，即场景由推荐时可用的参数决定，在推荐引擎中体现为"在线获取推荐结果API"的输入参数。

有两种场景最为常见，分别是首页推荐场景和详情页推荐场景。在执行首页推荐时，可用的参数只有用户信息；而在执行详情页推荐时，可用的参数除了用户信息，还包括当前详情页上所展示的物品信息。

您完全可以根据自己的需求建立全新的场景，比如针对搜索关键词的推荐场景，这时可用的参数除了用户信息，还有用户所输入的关键词。

设置OTS预留读吞吐量: 0

请选择推荐API所用的参数: +新建自定义参数

☑biz_code	表示业务code
☑scn_code	表示场景code
☑user_id	表示用户的id
☐item_id	表示物品的id

上一步 下一步

Day4-5. 基本配置和离线计算

1 配置推荐场景 > 2 配置API参数 > 3 配置推荐算法流程

推荐算法流程指数据端到端的处理流程，即从客户的原始输入数据开始（如用户数据、物品数据、行为数据等），一直到产出能够被客户产品在线获取的推荐结果为止。

算法流程由离线流程和在线流程两部分组成，分别完成离线计算和在线推荐。 了解更多

+新建算法流程

开发测试环境 | 线上生产环境

发布

算法流程code	离线流程	在线流程	操作
test_path	修改	修改	删除 回滚 在线流程调试

添加算法流程

上一步 完成

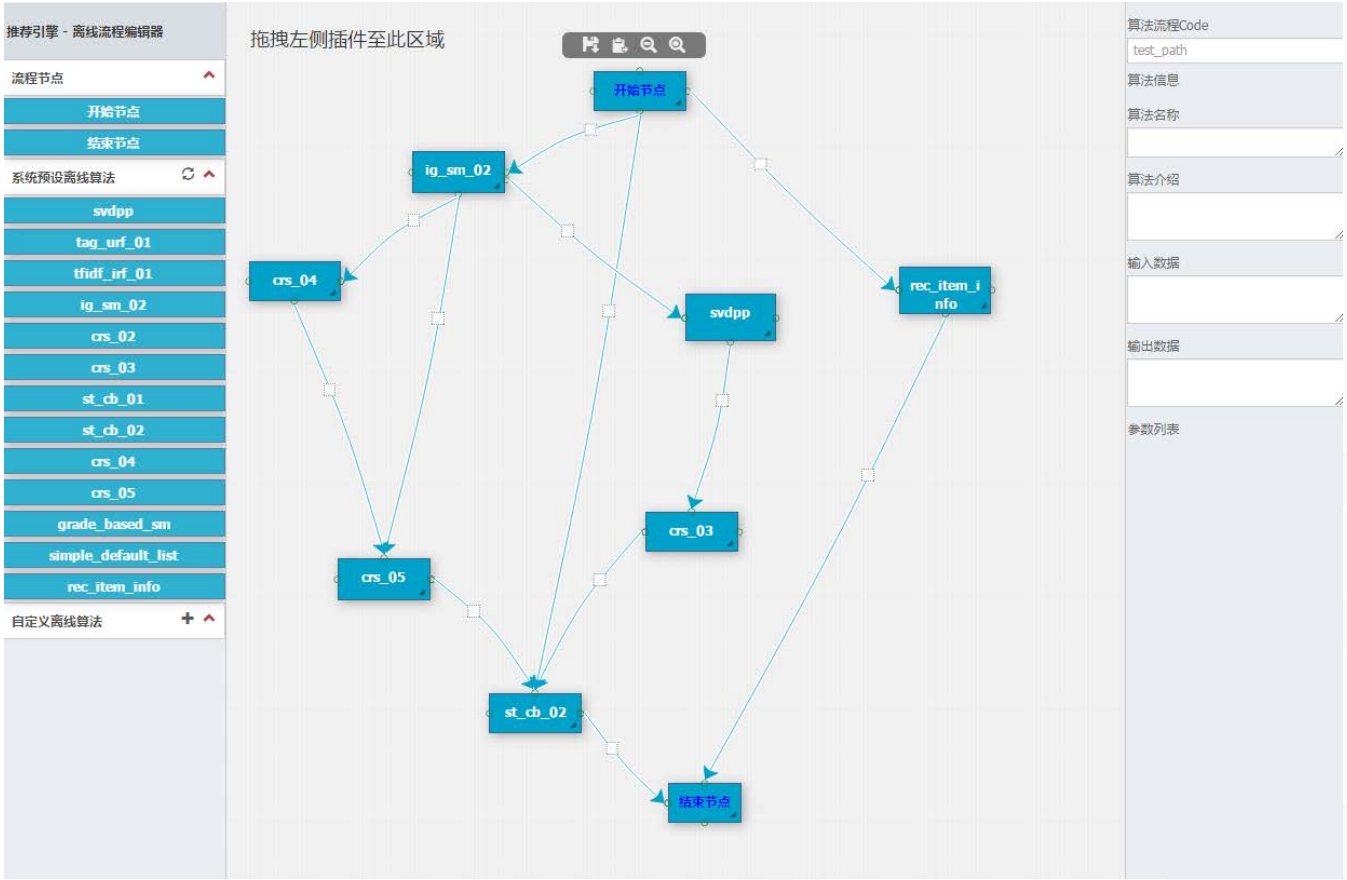
算法流程代码* test_path

离线计算模版 main_ofl 默认首页推荐离线计算流程

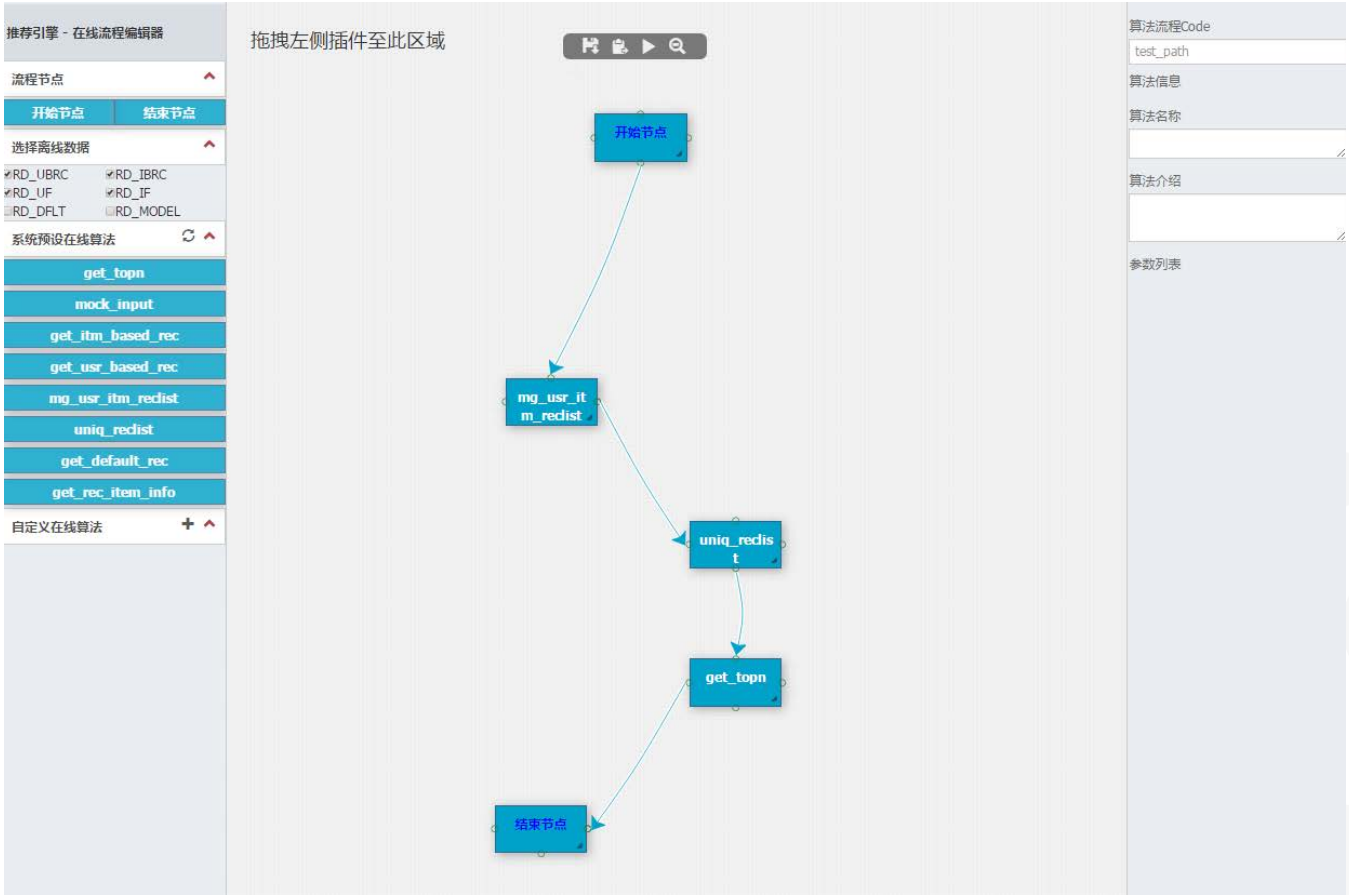
在线计算模版 detail_ol 默认详情页推荐在线计算流程

关闭 提交

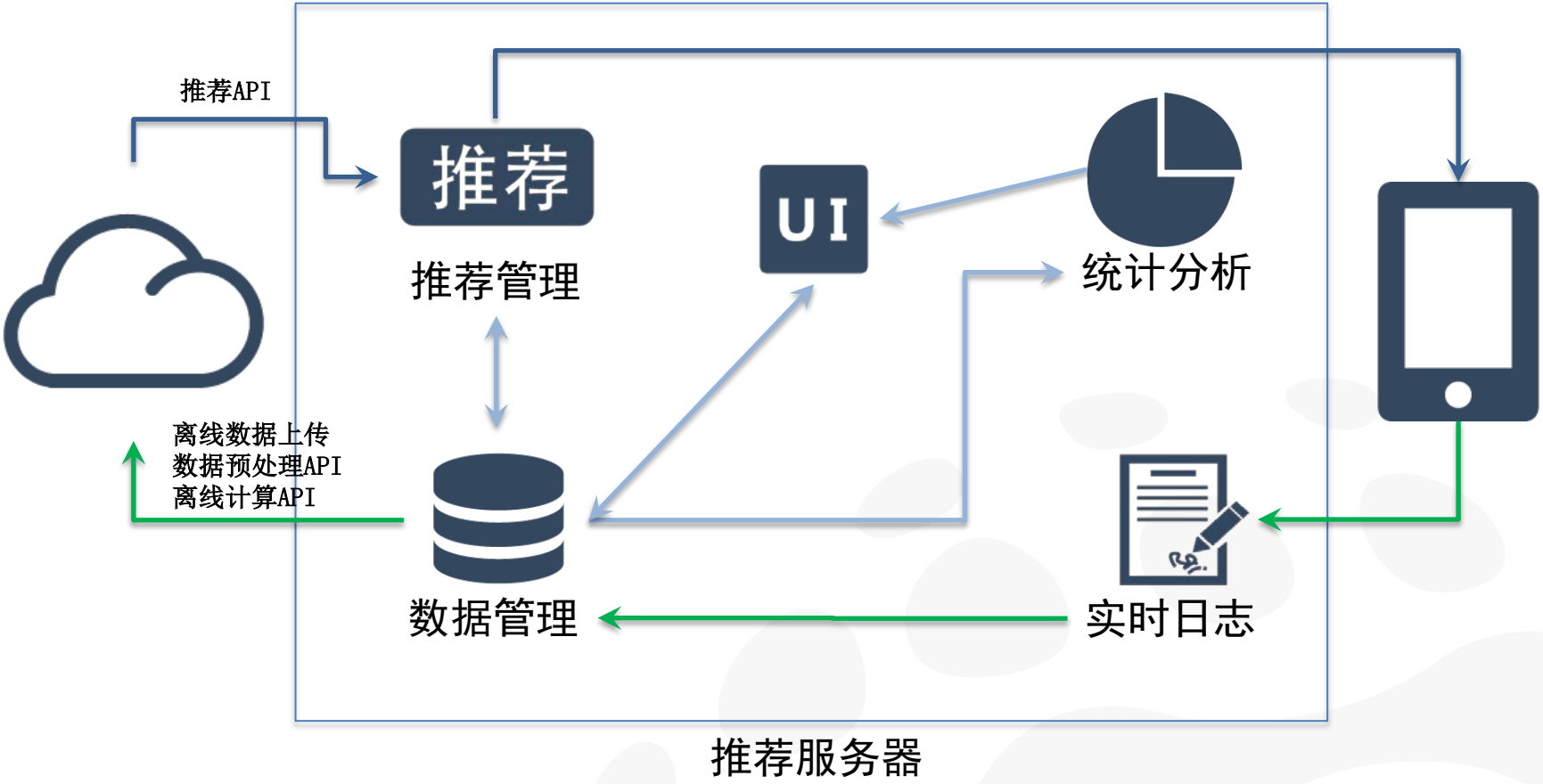
Day4-5. 基本配置和离线计算



Day4-5. 基本配置和离线计算



Day6-8. 推荐API集成

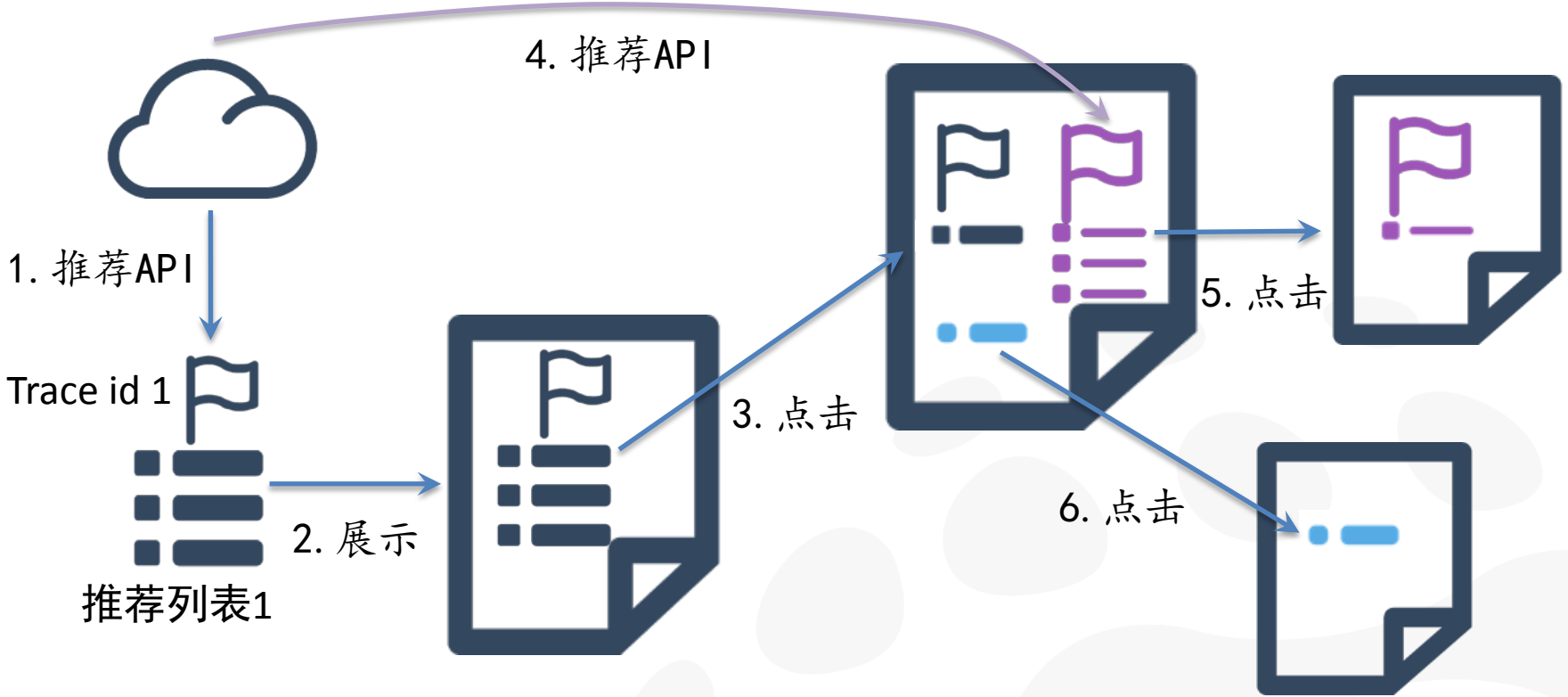


目 录
content

第二阶段：高级功能

Day9-11. 效果报表

Trace ID的生命周期



Day9-11. 效果报表

效果计算

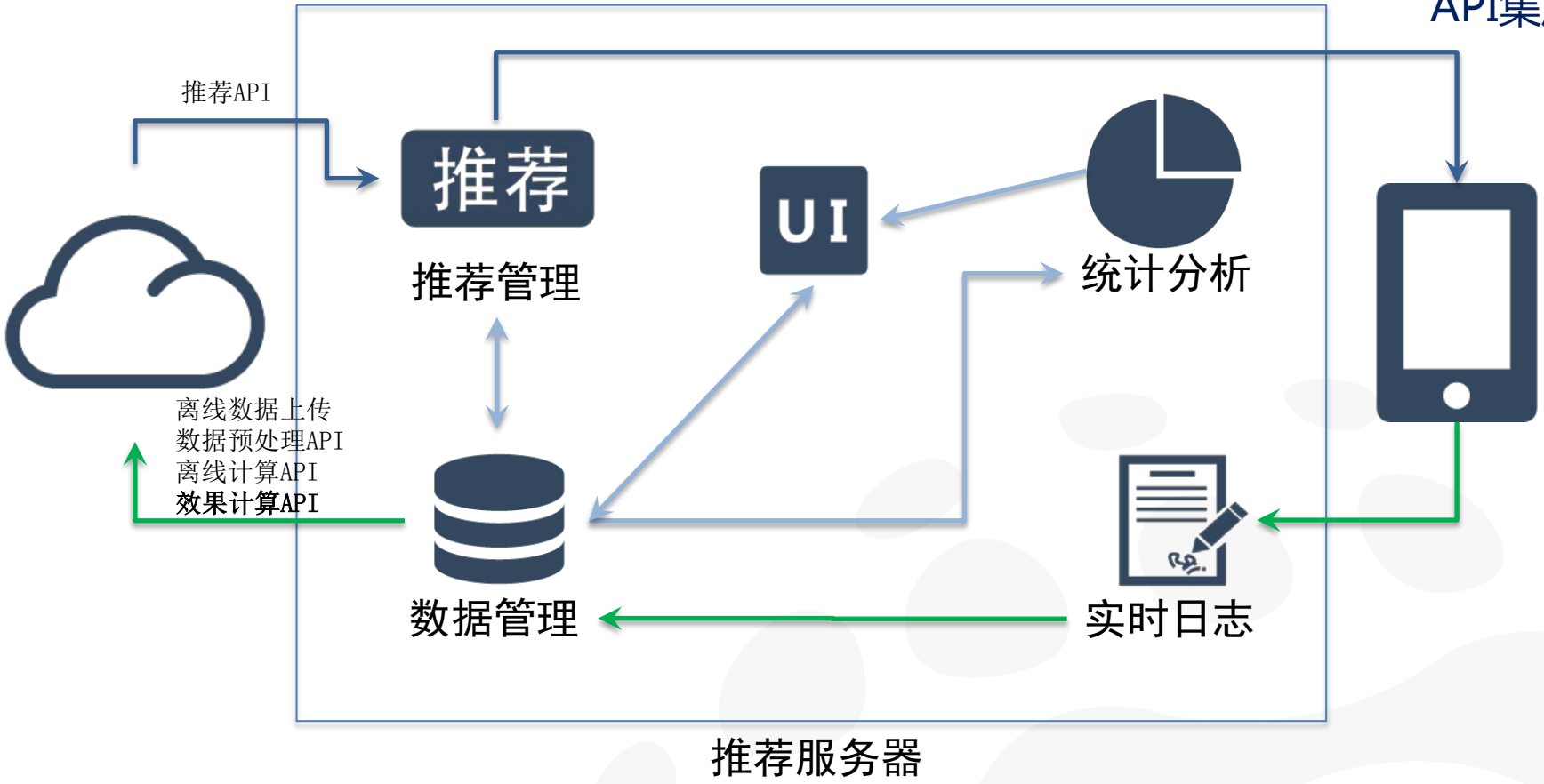
效果算法：以行为类型为参数，根据行为数据计算指标的算法

效果指标：明确行为类型、时间跨度、统计口径的效果算法实例。选择效果指标后即生成计算任务

效果报表：根据算好的效果指标，配置显示图表后的可视化展示

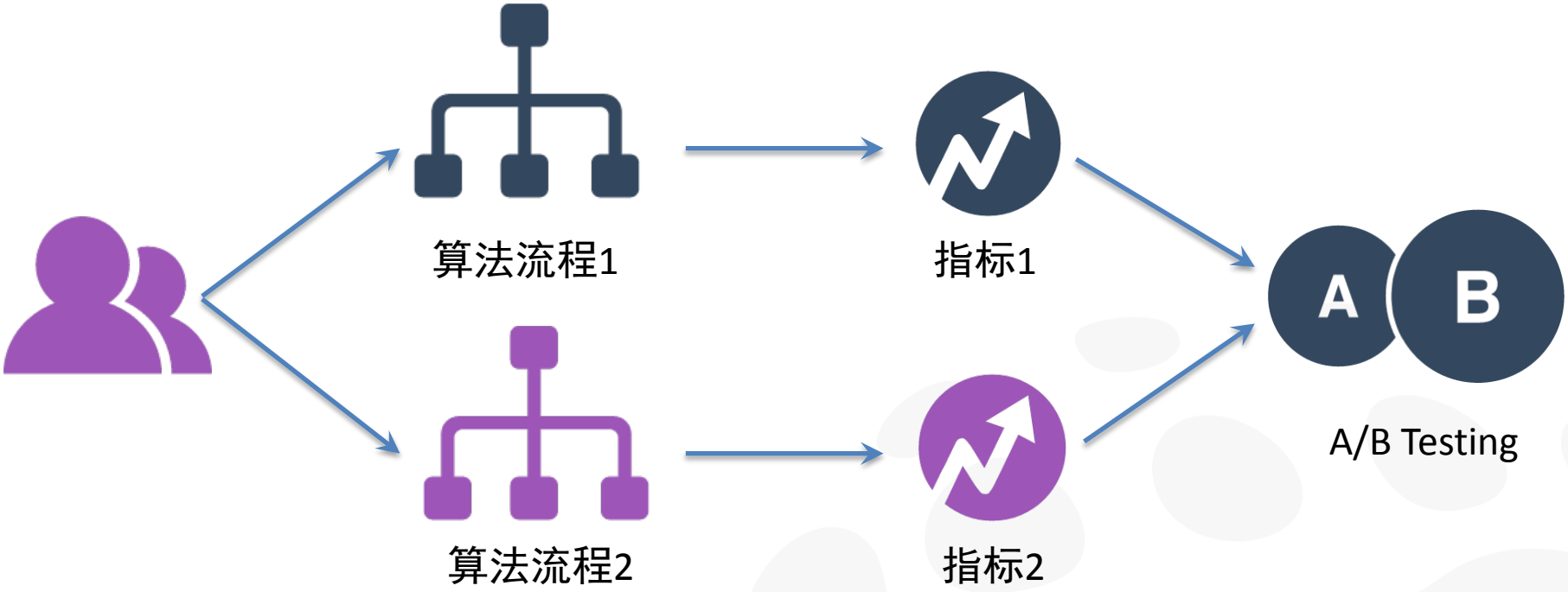
Day9-11. 效果报表

API集成



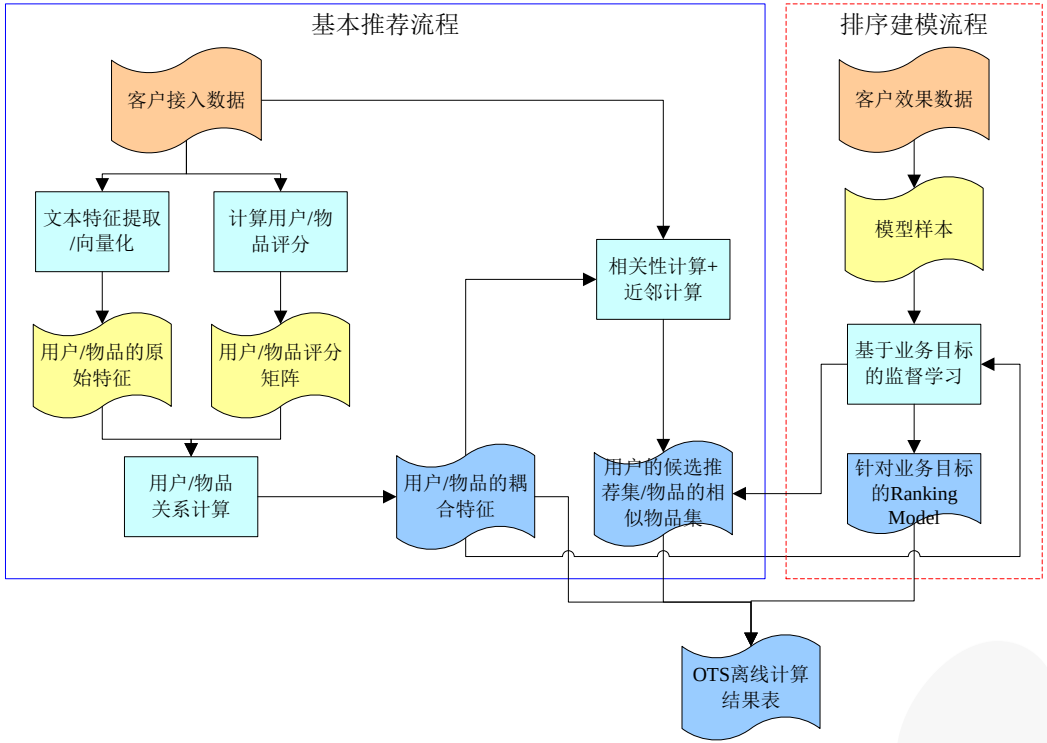
Day12-15. 优化

A/B Testing

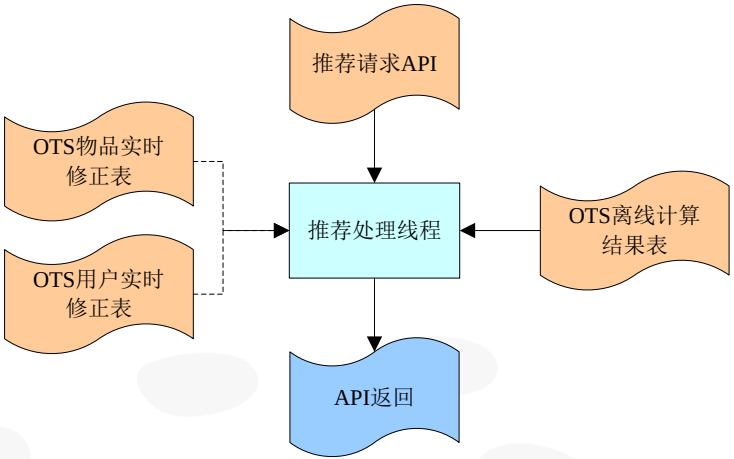


Day12-15. 优化

算法流程



离线算法流程



在线算法流程

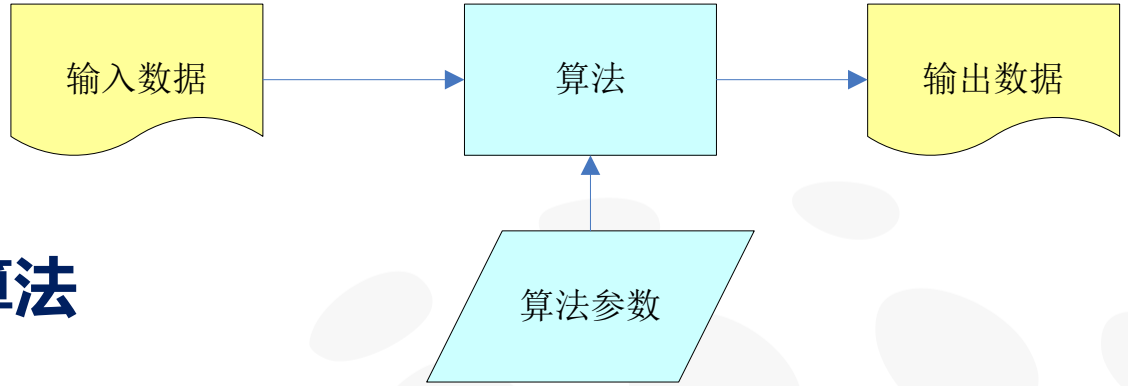
Day12-15. 优化

自定义算法

1

自定义离线算法

算法开发
单元测试
算法注册
流程测试



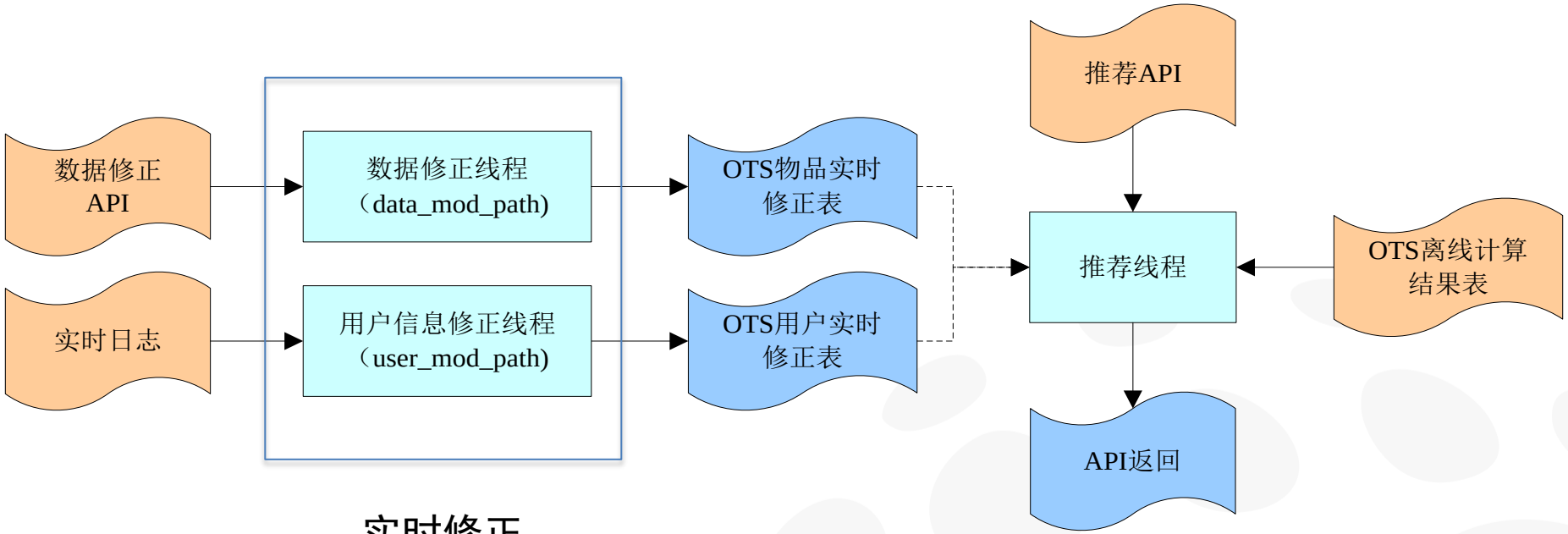
2

自定义在线算法

算法开发
单元测试
算法注册
流程测试

Day16-20. 实时修正

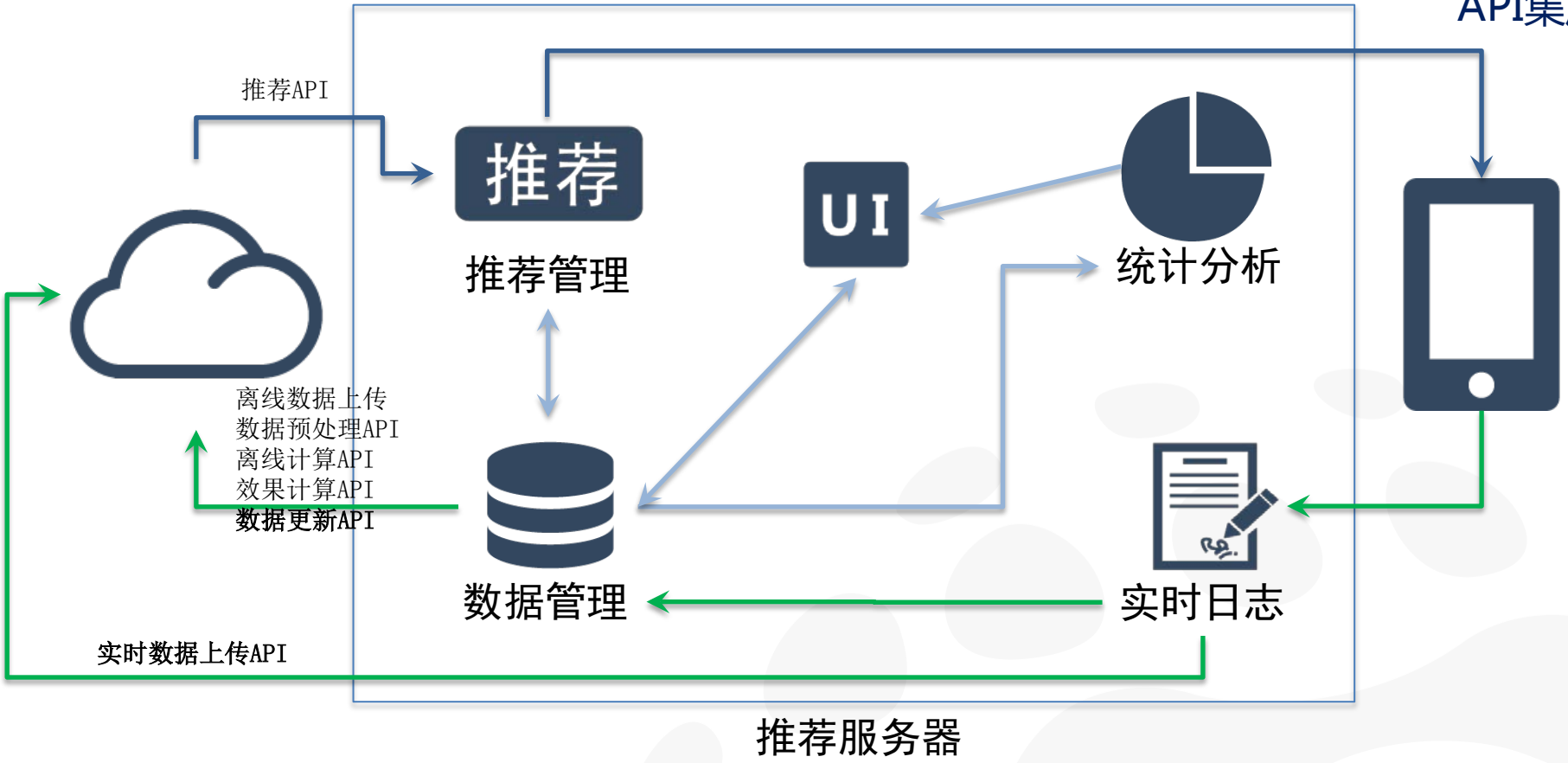
整体流程



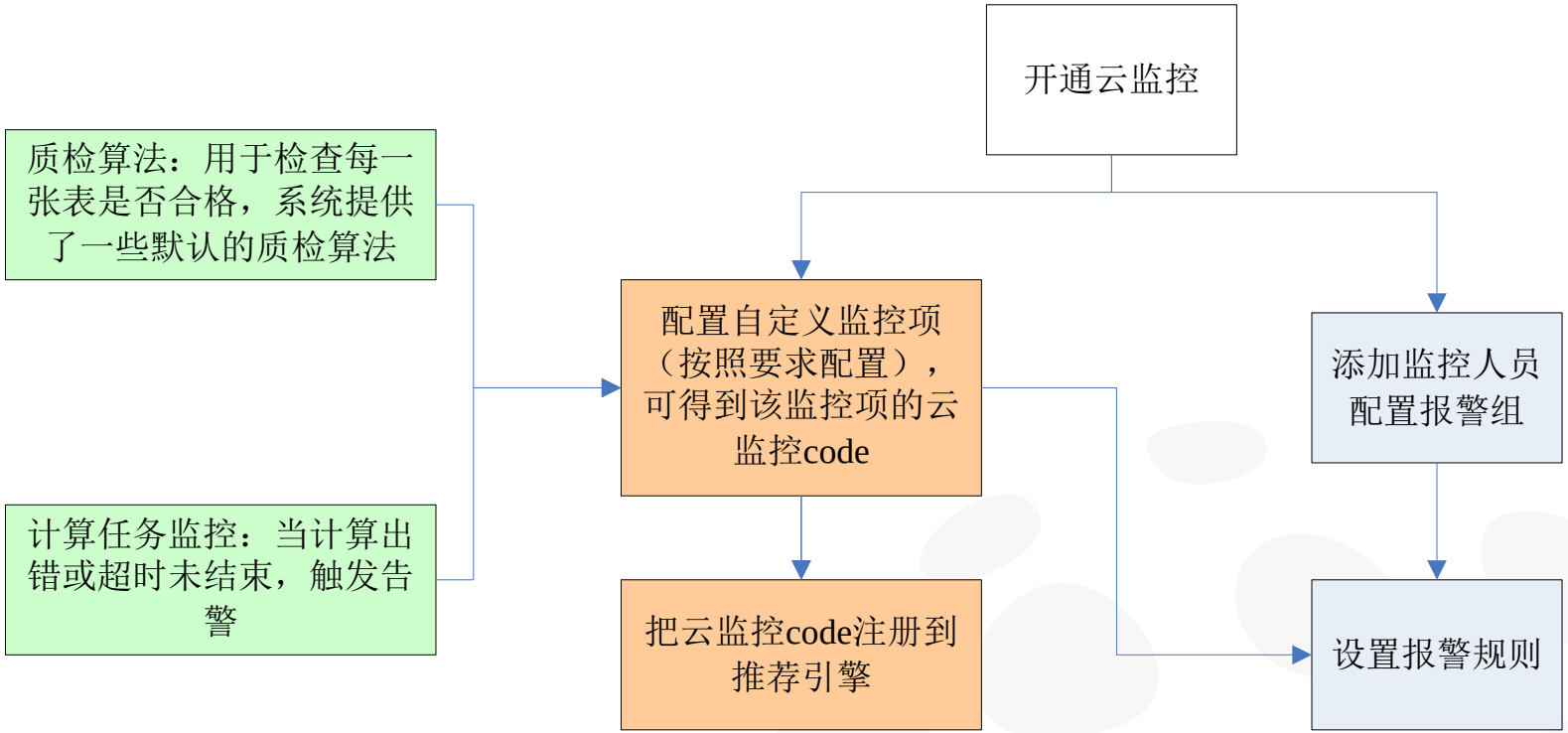
实时修正

Day16-20. 实时修正

API集成



Day21. 监控和告警


开通云监控
质检算法：用于检查每一张表是否合格，系统提供了一些默认的质检算法
计算任务监控：当计算出错或超时未结束，触发告警
配置自定义监控项（按照要求配置），可得到该监控项的云监控code
把云监控code注册到推荐引擎
添加监控人员配置报警组
设置报警规则

Future Work

简化基于规则的推荐

更好的集成业务和运营积累的知识

提供更多的算法

实现更多的推荐算法
针对行业的算法模板

推荐算法大赛

计划在下半年启动，敬请关注！

Conclusion

目标：让客户专注于推荐业务，不再被系统问题困扰

方法：通用的推荐引擎，集中实现与业务无关的内容

效果：推荐是个系统工程，算法很重要，但不是全部

关注我们

个性化推荐
钉钉交流群

该钉钉群二维码将在2017-06-13失效

关注云栖社区
微信公众号