

御膳房-阿里大数据平台服务

王贲（陈廷梁）

2014-07



自我介绍

淘宝网
Taobao.com



- 陈廷梁，花名：王贲
- 数据平台事业部-商家数据部-架构师

目录

淘宝网
Taobao.com



- 御膳房简介
- 功能介绍
- 技术架构
- 未来规划



- 对外提供一站式大数据服务的平台，让天下没有难玩的数据
- 从提供酒菜到提供厨房
- 从自己玩数据，到提供平台enable大家玩数据
- Enable 数据交换和分享，数据驱动用户创造更大商业价值



- PAAS
 - Data platform as a service
 - Data warehouse as a service
 - Data center as a service



- 阿里数据开放的平台
- 提供大数据存储计算能力
- 一站式数据开发引擎
- 云数据中心（仓库）解决方案
- 数据市场-数据交换分享
- 数据安全服务

御膳房简介-发展历程

淘宝网
Taobao.com



- 2012.9 第一次亲密接触聚石塔商家后台团队
- 2012.10 开始规划聚石塔数据平台-御膳房
- 2012.10~2013.1 前御膳房
- 2013.3 御膳房0.1上线
- 2013.6 ODPS评测
- 2013.7 完成底层云化迁移, HIVE->ODPS, MYSQL->UMP(Unified MySQL Platform)
- 2013.8 M/R开发环境上线
- 2013.8 算法环境上线
- 2014.4 新算法环境上线
- 2014.6 御膳房成为阿里统一的对外数据平台服务, 开始接入非电商市场



- Amazon Redshift
- 京东云海：2014

御膳房功能-数据开发

淘宝网
Taobao.com



APPKEY:999999996

御膳房数据引擎 注销

首页

数据管理

数据开发

生产部署

数据联盟

帮助反馈

数据字典

> 数据查询

基础数据

> 商家主题

> 商品主题

> 客服绩效

> 品牌主题

> 行业主题

> 基础数据

个人数据

> 商家主题

> 商品主题

> 品牌主题

数据管理 > 基础数据 > 商家主题

表名:	sys_seller_dwb_shop_view_d					未授权
等级:	L2	一级主题:	商家主题	二级主题:	商家流量分析	
类型:	卖家表	计算周期:	每天			
描述:	每个店铺的每日流量数据; 统计周期: 每日					

表名:	sys_seller_dwb_shop_trade_d					未授权
等级:	L2	一级主题:	商家主题	二级主题:	商家销售分析	
类型:	卖家表	计算周期:	每天			
描述:	每个店铺的每日销售数据；统计周期：每日					

看数据		一级主题: 商家主题		二级主题: 商家销售分析		三级主题: 商家销售分析	
计算周期: 每天		天级					
描述: 每个店铺的每日销售数据; 统计周期: 每日							

第一步 看数据

御膳房开放商家、商品、品牌、客服相关的淘宝数据，为数据开发提供源动力

御膳房功能-数据开发

淘宝网
Taobao.com



御膳房 阿里云数据引擎
cloudata.taobao.com

数据开发

APPKEY:9999999996

注销

程序 ▾ 运行 ▾ 帮助反馈 ▾

新建 保存 | 新建数据表 | 运行环境 样本数据 运行 停止 | 部署任务

我的程序 基础数据 个人数据

数据开发

演示程序

演示程序2

test

mytest

演示程序 x

```
1 insert overwrite table pri_result.shop_trade_d partition (dt=${date_ymd})
2 select thedate, shop_id, seller_id, alipay_trade_num, gmvt_trade_num
3 from sys_seller_dwb_shop_trade_d
4 where dt=${date_ymd};
```

运行记录

结果数据

运行

停止

第二步 用数据—计算开发

御膳房提供IDE开发工具，通过提交SQL、M/R等程序玩转淘宝数据

御膳房功能-数据开发

淘宝网
Taobao.com



APPKEY:9999999996

御膳房数据引擎 注销

首页

数据管理

数据开发

生产部署

数据联盟

帮助反馈

程序管理

> 程序列表

任务管理

> 任务部署

> 任务列表

> 运行记录

接口发布

> 数据接口发布

> 模型接口发布

接口列表

> 数据接口列表

> 模型接口列表

生产部署 > 任务管理 > 任务部署

任务类型 ☒ 计算 ☐ 导出 ☐ 抽取

任务名称 请填写，不超过50字

任务ID

部署程序

销量预测

部署程序

依赖数据表 暂无数据

产出的结果表 暂无数据

任务描述

商品预测

提交部署

第三步 部署任务

可根据需求对任务进行灵活部署

御膳房功能-数据开发

淘宝网
Taobao.com



APPKEY:999999996

御膳房数据引擎 注销

首页

数据管理

数据开发

生产部署

帮助反馈

程序管理

> 程序列表

任务管理

> 任务部署

> 任务列表

> 运行记录

接口发布

> 数据接口发布

> 模型接口发布

接口列表

> 数据接口列表

> 模型接口列表

生产部署 > 接口发布 > 数据接口发布

接口名称 销售预测

请填写，不通过审核后，系统生成

接口 ID 通过管理员审核后，系统生成

接口描述

销售预测

接口 SQL

接口 SQL

```
select seller_id from sys_s  
where .....
```

```
select seller_id from sys_seller_dwb_shop_view_d  
where .....
```

sys_seller_id: seller id
sys_user_nick: 卖家nick
sys_appkey: 发出TOP API调用的appkey

查看示例

变量	默认值	描述	操作
seller_id	100	卖家id	删除

输入变量名

输入默认值

输入描述

+ 添加变量

快速测试SQL

点击后自动保存，并进入提交进行测试，返回结果到下面

测试结果 sql解析错误

第四步 发布接口

保存

提交发布

简单的接口发布，通过TOP接口，拿到数据计算结果

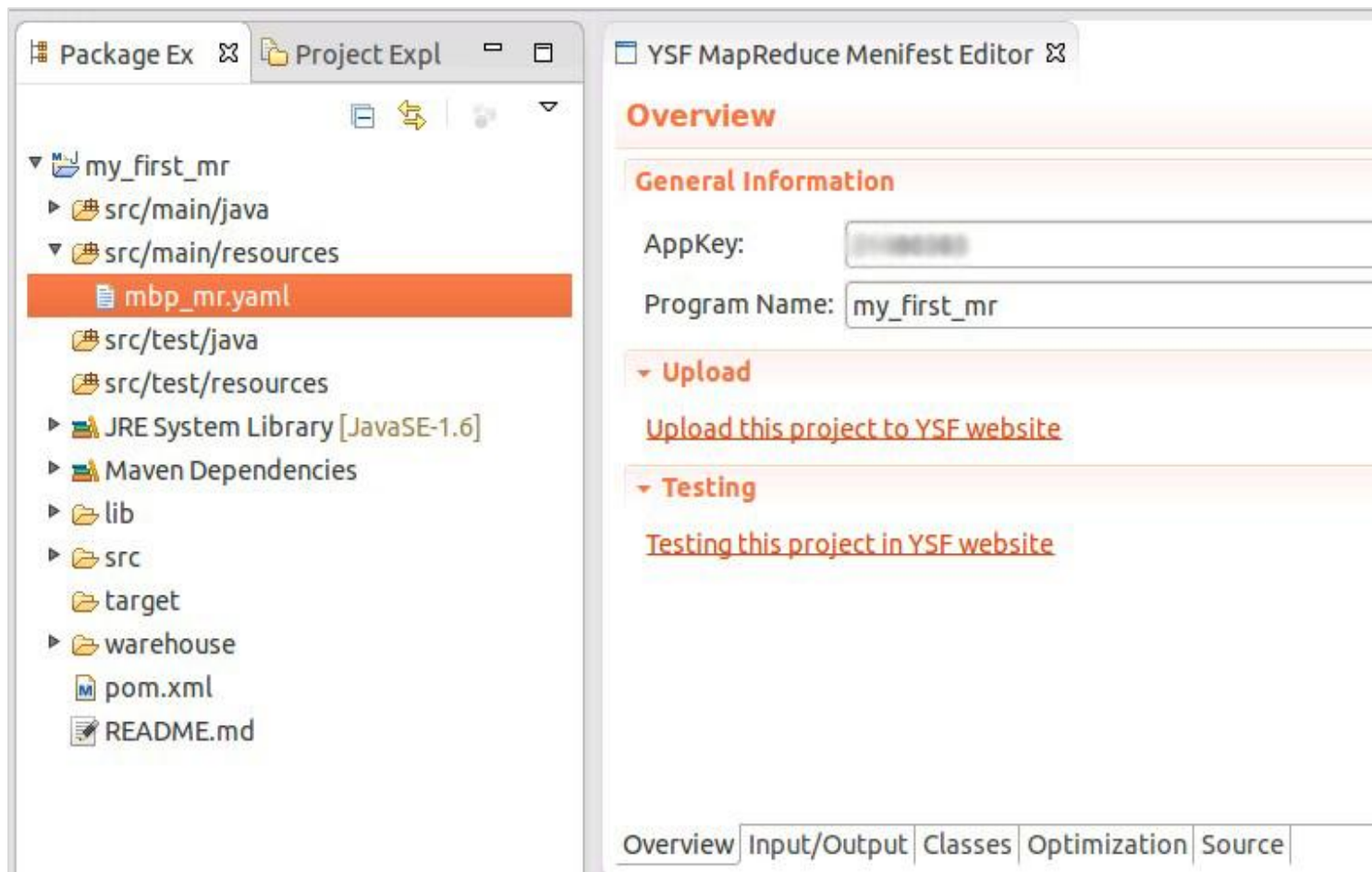
御膳房功能-数据开发

淘宝网
Taobao.com



MapReduce

MapReduce简称MR
平台也提供了
Eclipse开发插
件来辅助
MapReduce开
发与调试。





UDF

User Defined Function, 即用户自定义函数。

平台提供了Eclipse开发插件来辅助UDF开发与调试。

The screenshot displays the 'YSF UDF Manifest Editor' interface. On the left, the 'Project Explorer' shows a project named 'my_first_udf' with a structure including 'src/main/java' (containing 'udf' and 'udf.examples' sub-packages with various Java files like 'MyUDAF.java', 'MyUDF.java', 'MyUDTF.java', 'GcdUDF.java', 'NormalizeUDTF.java', and 'VarianceUDAF.java'), 'src/main/resources' (containing 'mbp_udf.yaml'), 'src/test/java', 'src/test/resources', 'JRE System Library [java-6-op', 'Maven Dependencies', and 'lib'. The 'mbp_udf.yaml' file is selected and highlighted. On the right, the 'Overview' tab is active, showing 'General Information' with fields for 'AppKey' (21480388), 'UDF Name' (my_first_udf), 'UDF Class' (udf.MyUDF), 'Type' (UDF), 'Category' (字符串处理函数), and 'Description' (This is my first UDF.). Below this, there is an 'Upload' section with a link to 'Upload this project to YSF website'. At the bottom, there are tabs for 'Overview' and 'Source'.

御膳房功能-算法开发

淘宝网
Taobao.com



御膳房-大数据挖掘平台

11:14 AM



数据获取



数据挖掘环境 by
Python



数据挖掘环境 by R



御膳房集成开发环境



XLab



MySQL Workbench



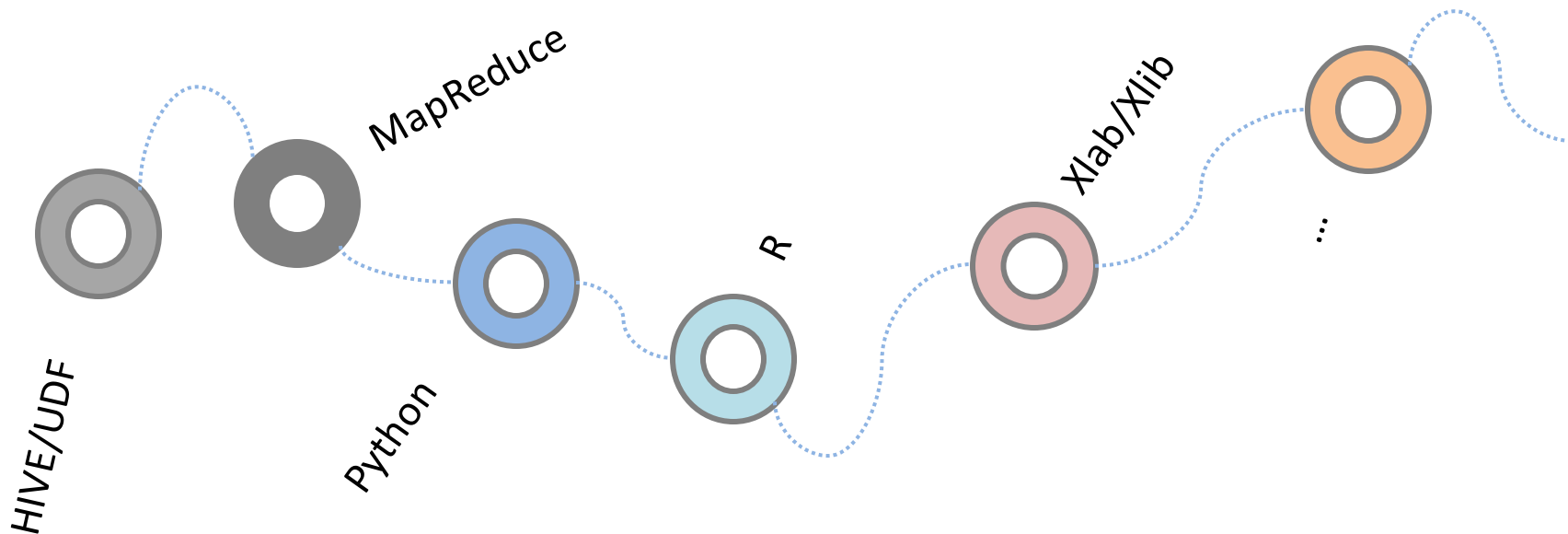
御膳房

聚石塔数据引擎
clouddata.taobao.com





基于 **工具链** 的一站式数据挖掘平台



* 虚线表示使用过程中的链式关系。工具可自由选择，两两之间，都可协同工作。

御膳房功能-算法开发

淘宝网
Taobao.com



XLib

XLib是ODPS的分布式算法库，支持分类预测、回归、聚类、关联分析、矩阵计算等。

The screenshot displays the XLib interface with several components:

- Odps Table:** A list of tables including 'iris', 'iris555', 'iris666', 'iris_3', 'iris_bayes_mdlist_predict_out', 'iris_bayes_model', 'iris_bayes_model_111', and 'iris_bayes_predicted'.
- 窗口管理器 (Window Manager):** A sidebar on the left showing a tree view of the workspace, including 'iris - 数据表浏览', '统计图表', '系统输出: stdout2014_04_16_11_26_50_719.log', and '脚本运行窗口 E:\work\数据平台-产品与平台服务\naive_bayes_test.py'.
- log查看 (Log Viewer):** A window showing the contents of 'logstdout2014_04_16_11_26_50_719.log', displaying a list of log entries with timestamps and status (Running).
- 脚本运行 (Script Execution):** A window titled '脚本运行窗口 E:\work\数据平台-产品与平台服务\naive_bayes_test.py' showing the execution of a Python script. The script includes commands for training and predicting using Naive Bayes.
- iris - 数据表浏览 (Data Table Viewer):** A window displaying a table with columns: 'sepal_length', 'sepal_width', 'petal_length', 'petal_width', and 'category'. The table contains 24 rows of data.
- 统计图表 (Statistical Chart):** A window titled '统计图表' showing a bar chart with the legend: 'sepal_length' (red), 'sepal_width' (blue), 'petal_length' (green), and 'petal_width' (yellow). The chart displays the distribution of these features across the data.

Red arrows point to specific elements: 'log查看' points to the log viewer window, '脚本运行' points to the script execution window, '表' points to the 'iris' table in the Odps Table list, and '表操作窗口' points to the 'iris - 数据表浏览' window.

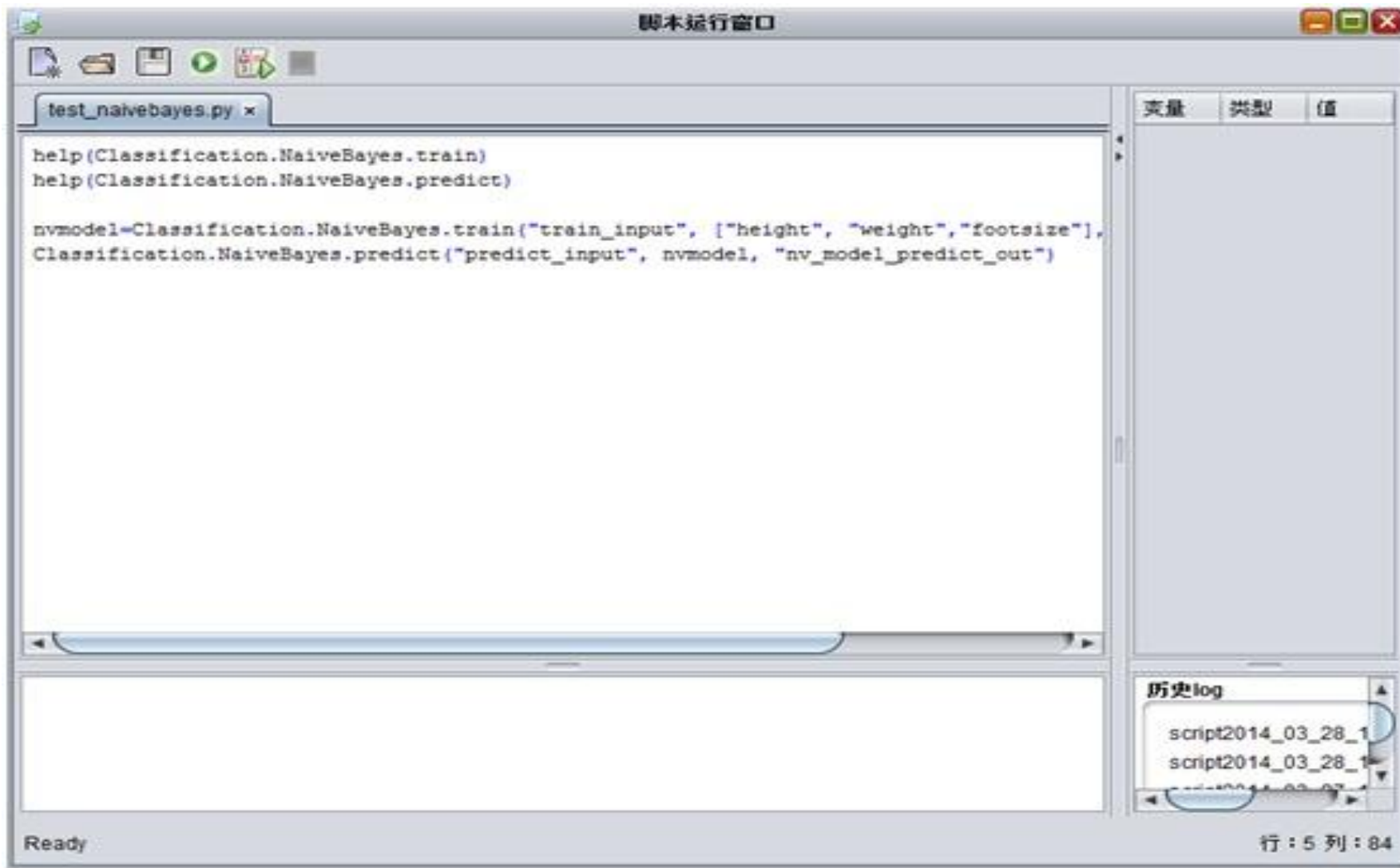
御膳房功能-算法开发

淘宝网
Taobao.com



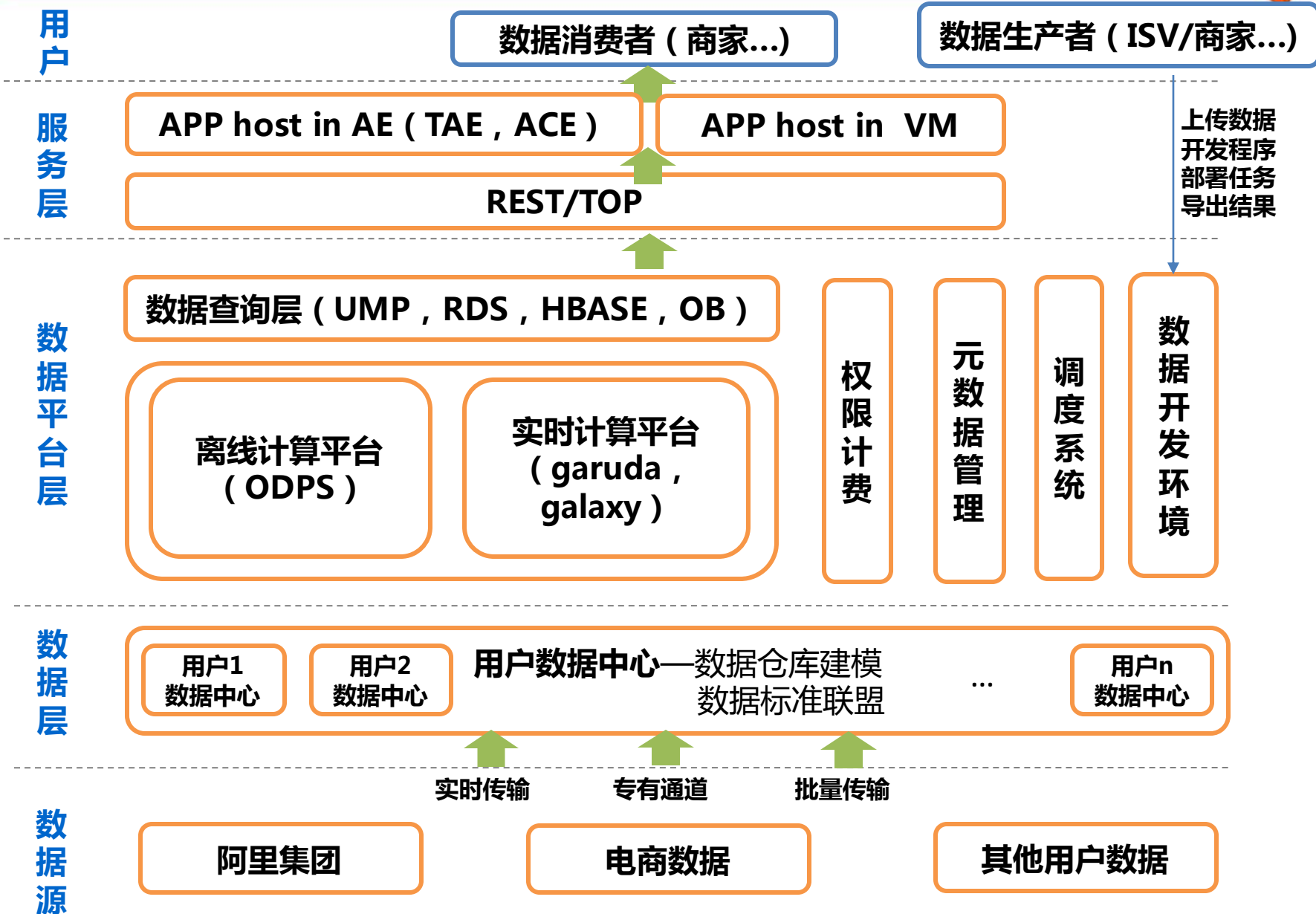
XLib

XLib是ODPS的分布式算法库，支持分类预测、回归、聚类、关联分析、矩阵计算等。



御膳房架构-按数据流分层

淘宝网
Taobao.com



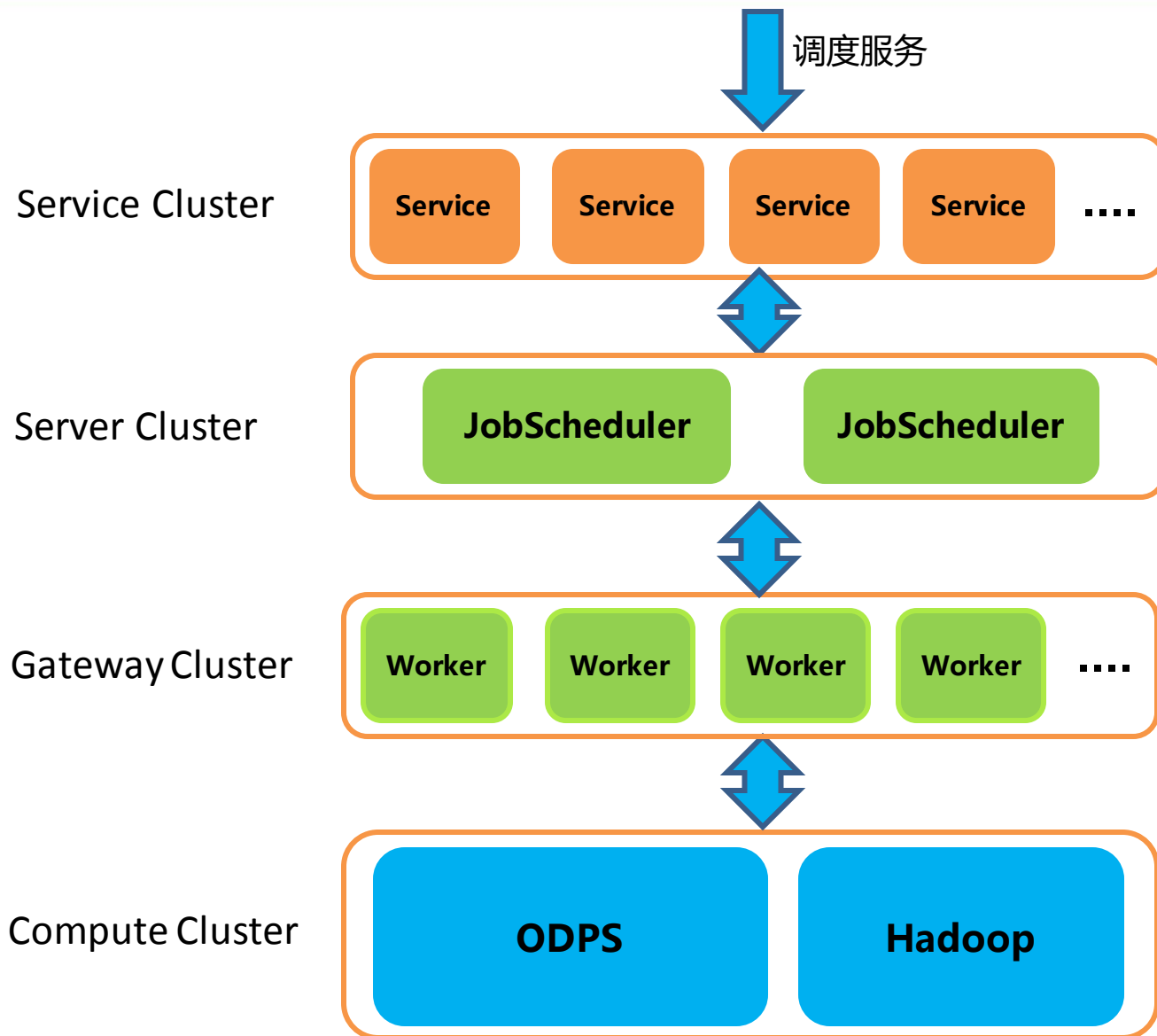


- 起
- 停
- 重跑
- 补跑
- 支持即时调度和周期调度
- 最小调度周期1分钟
- 支持优先级
- 可扩展，LB，HA



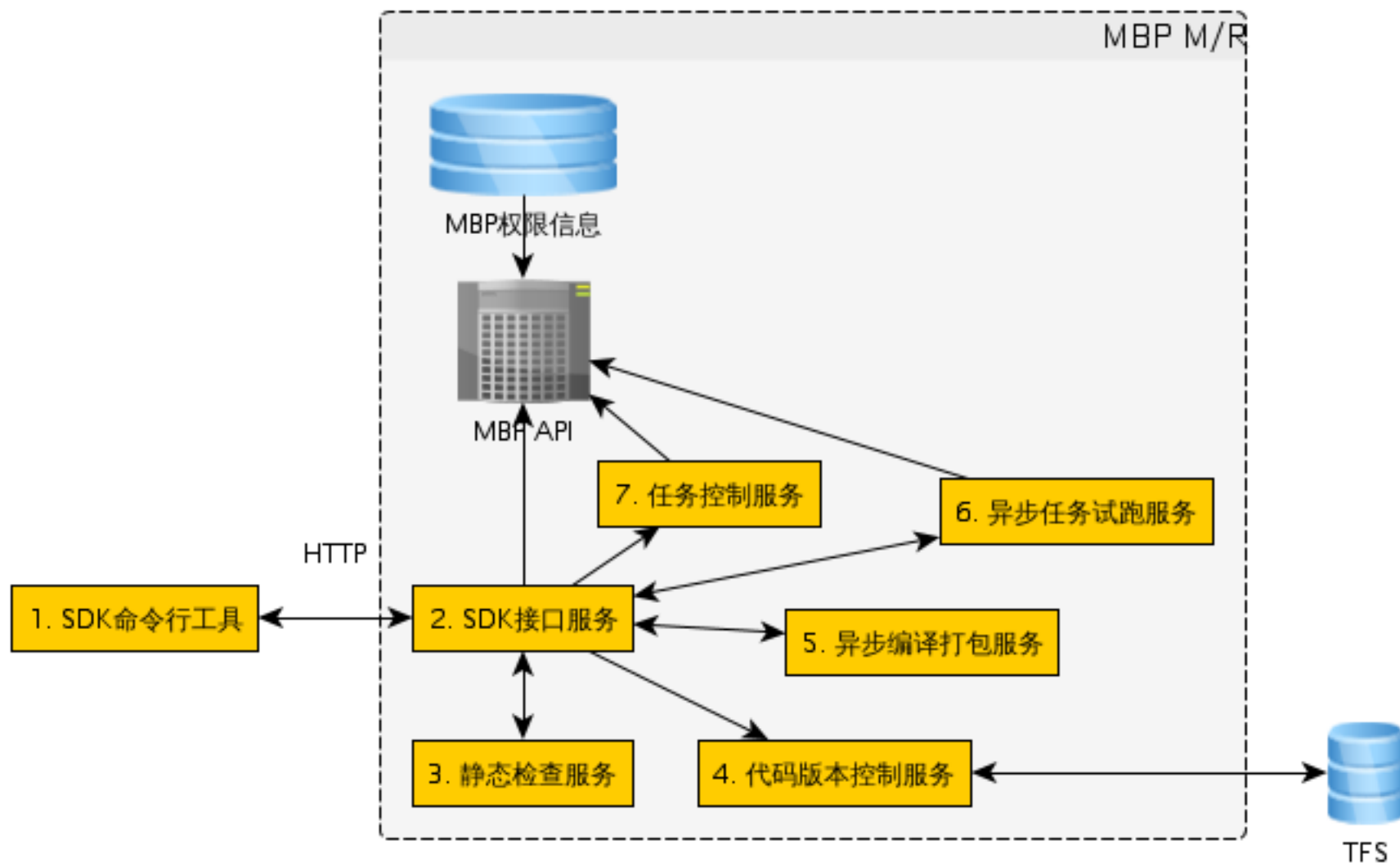
调度系统Octopus架构

淘宝网
Taobao.com



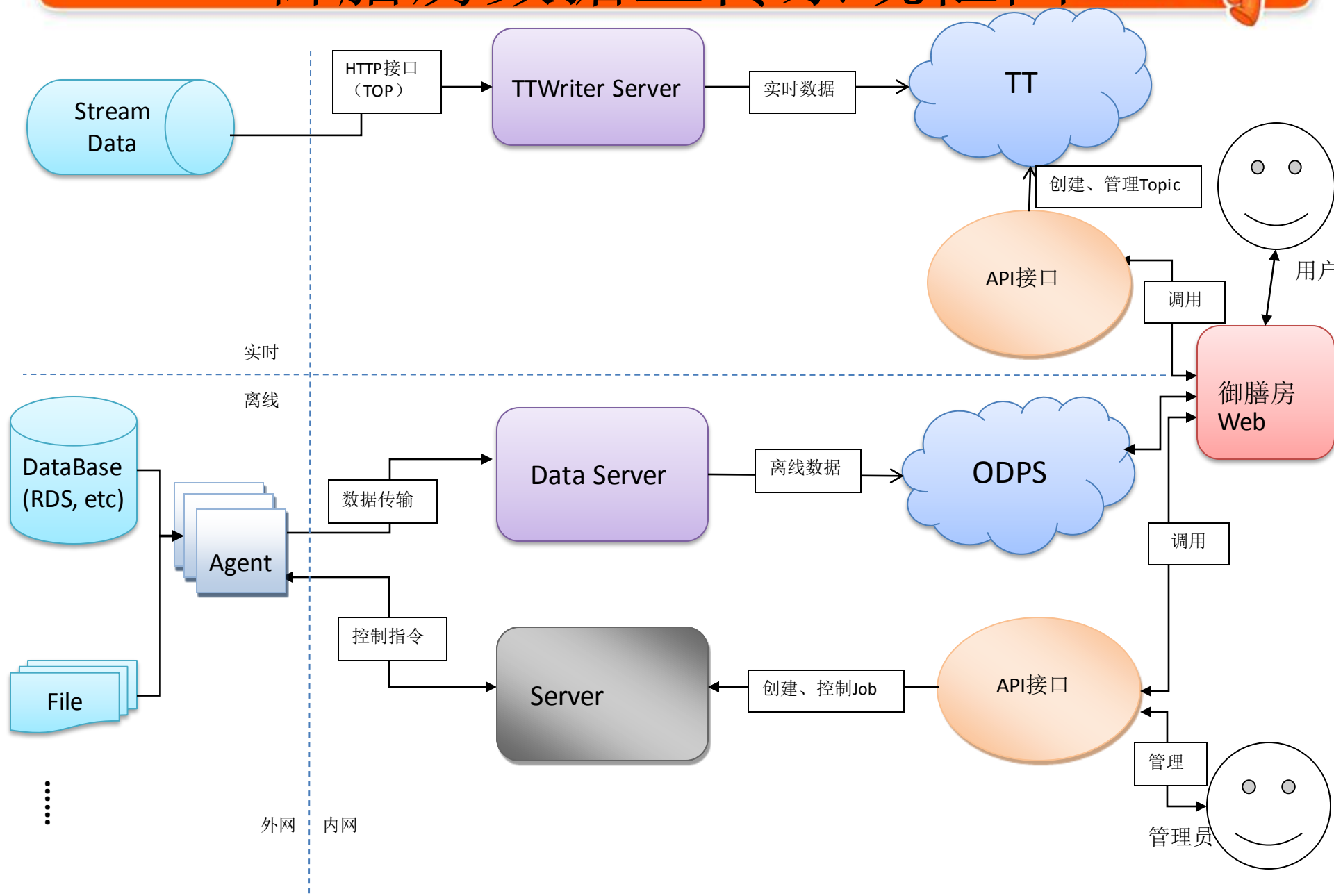
M/R计算环境系统框图

淘宝网
Taobao.com



御膳房数据上传系统框图

淘宝网
Taobao.com

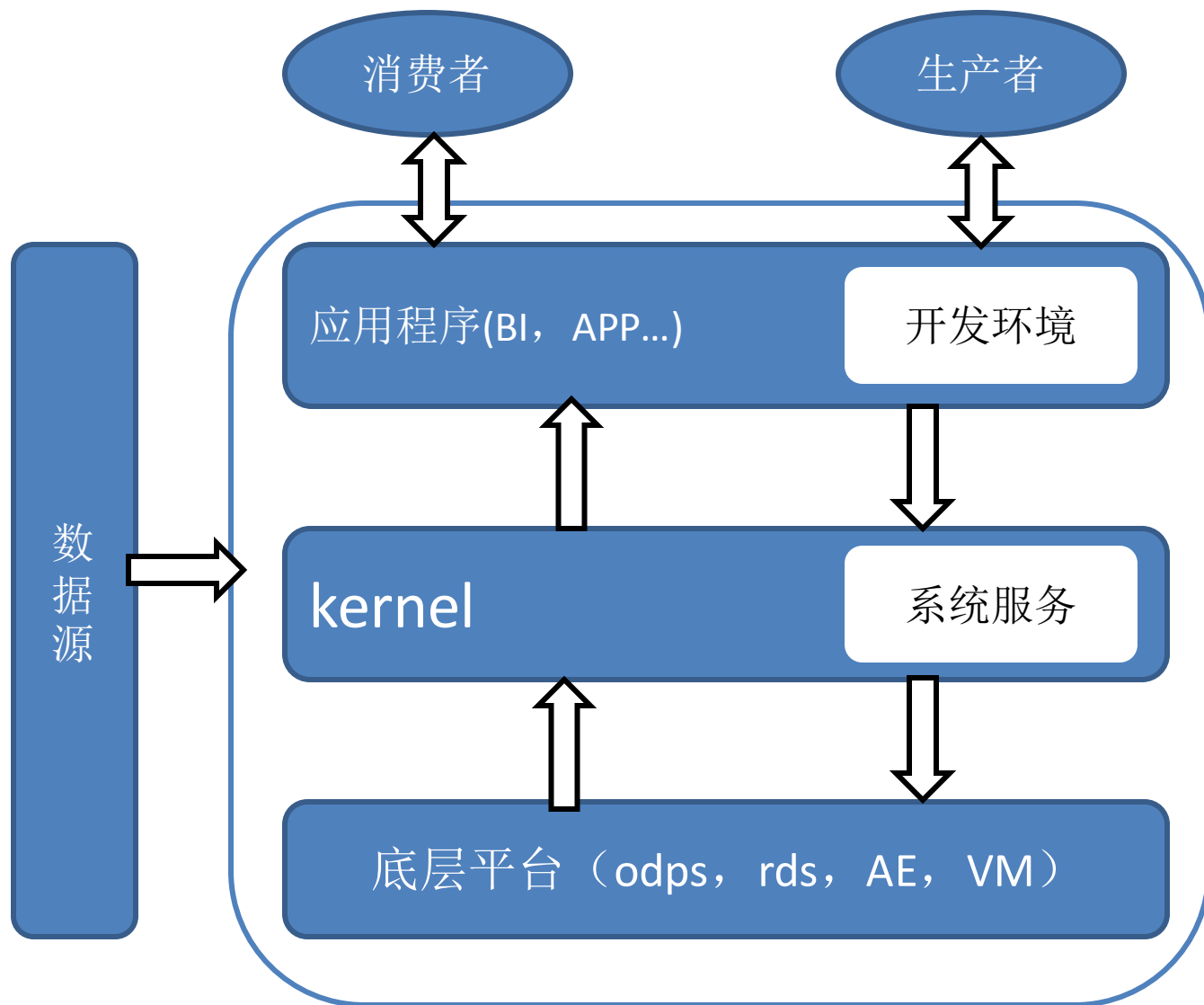




- 更大用户群
 - 电商：商家，ISV，品牌商
 - 广告
 - 其他：气象局，交通局
- 建立用户数据中心以及长在上面的app生态
 - BI 工具：MSTR，Cognos，数云，Tableau
 - 自用或第三方app
- 促进数据交换分享

御膳房未来规划-架构

淘宝网
Taobao.com



谢谢！

