

猫眼数据仓库建设分享



李鹏 2017年1月

一、大数据在猫眼的价值落地

- 猫眼有什么数据
- 数据的应用场景

二、猫眼大数据架构体系

- 整体技术架构
- 各方向的核心技术

三、猫眼数据仓库建设

四、猫眼数据平台建设

一、大数据在猫眼的价值落地

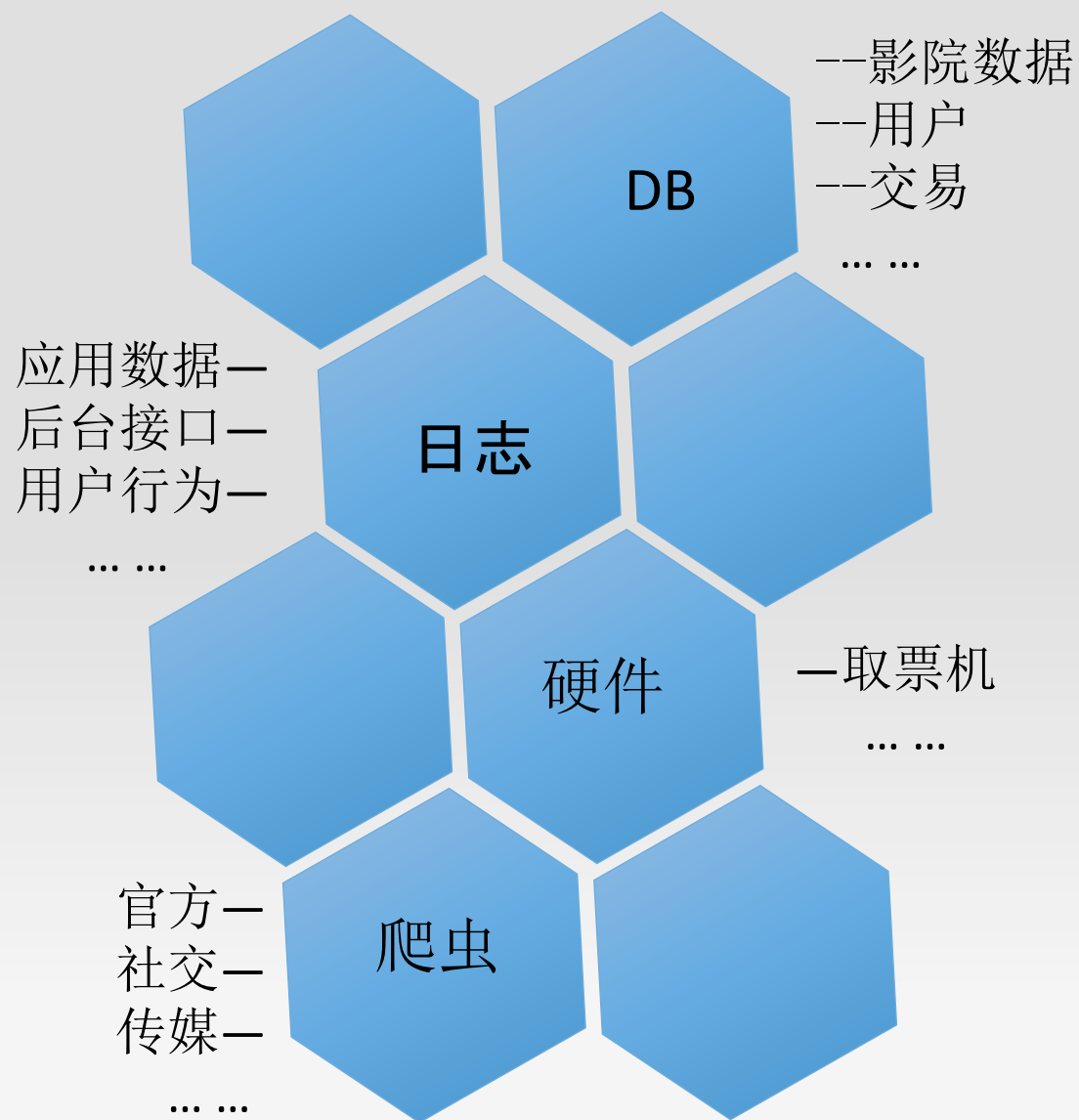
- 猫眼有什么数据
- 数据的应用场景

二、猫眼数据仓库架构体系

- 整体技术架构
- 各方向的核心技术

三、猫眼数据仓库建设

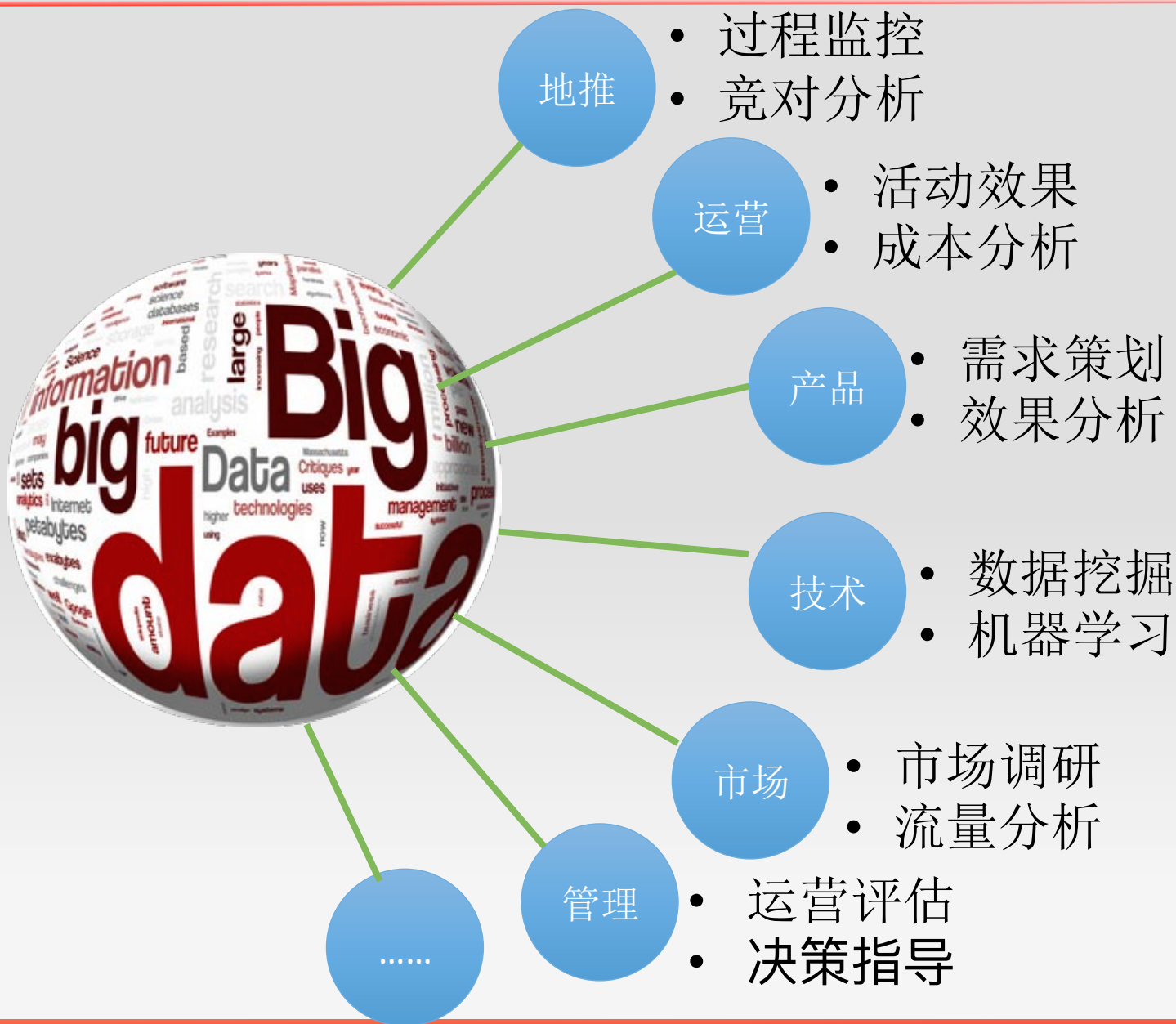
四、猫眼数据平台建设



猫眼数据量
日增加：TB级别
总量达：PB级别



猫眼数据应用场景



一、大数据在猫眼的价值落地

- 猫眼有什么数据
- 数据的应用场景

二、猫眼大数据体系架构

- 整体技术架构
- 各方向核心技术

三、猫眼数据仓库建设演进

四、猫眼数据平台建设

机器学习

报表

SQL

竞争情报

Spark

Camus

数据仓库

HBase

Kafka

Hadoop

SQLWeaver

Flume

Presto

BI

Binlog

元数据

Camel

MySQL

调度管理

Hive

爬虫

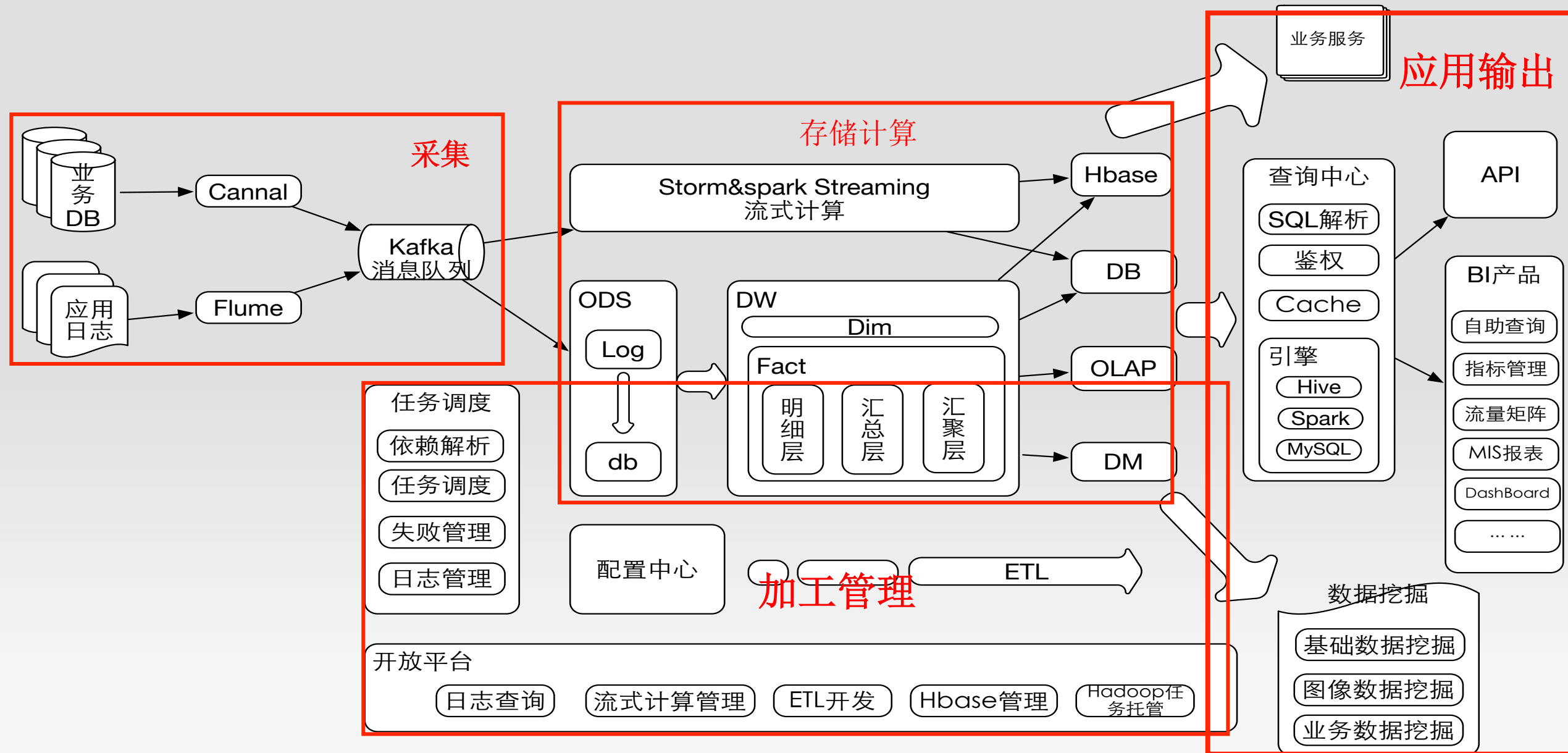
抓取

数据挖掘

NLP

ETL

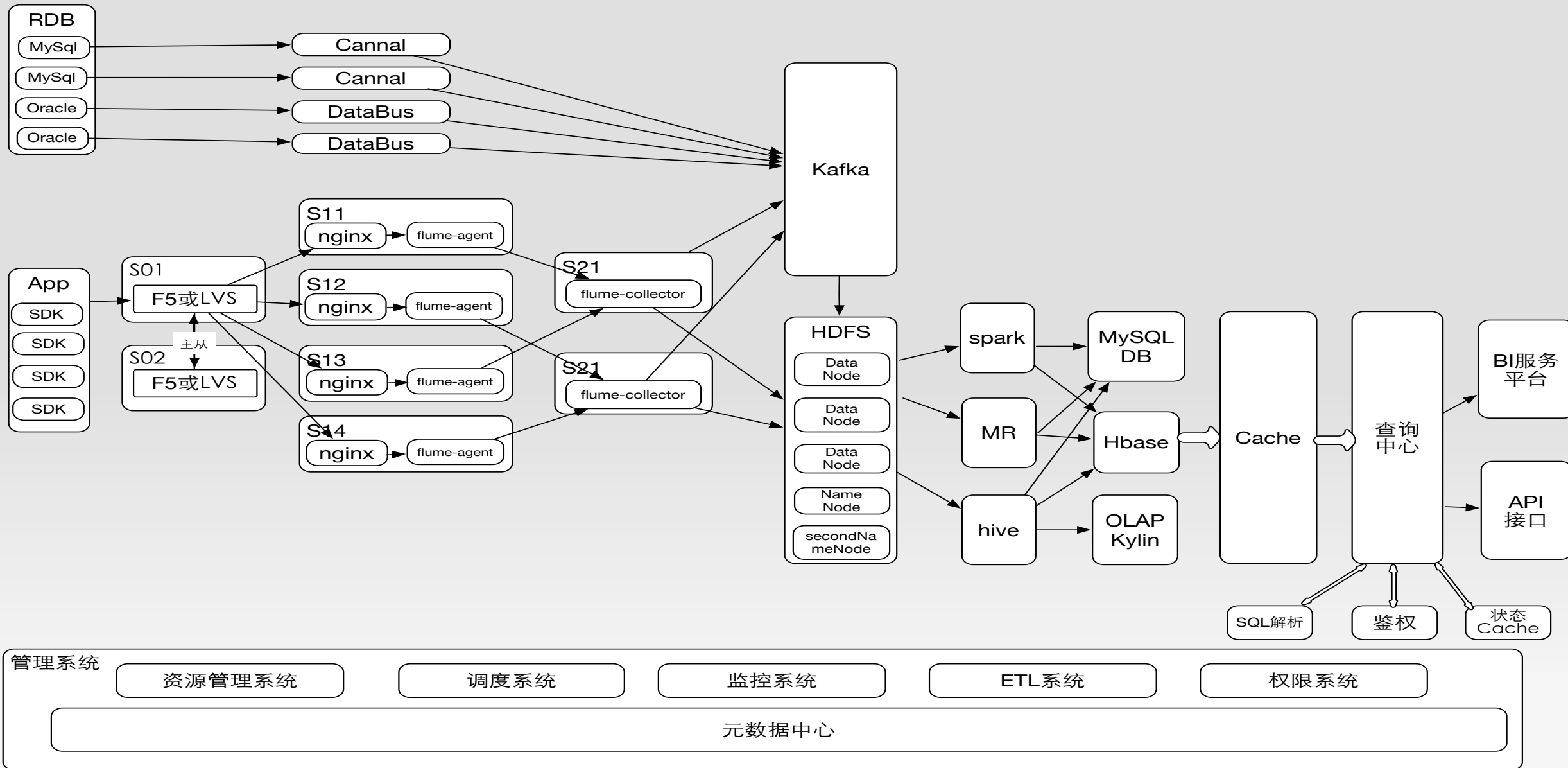
整体技术架构

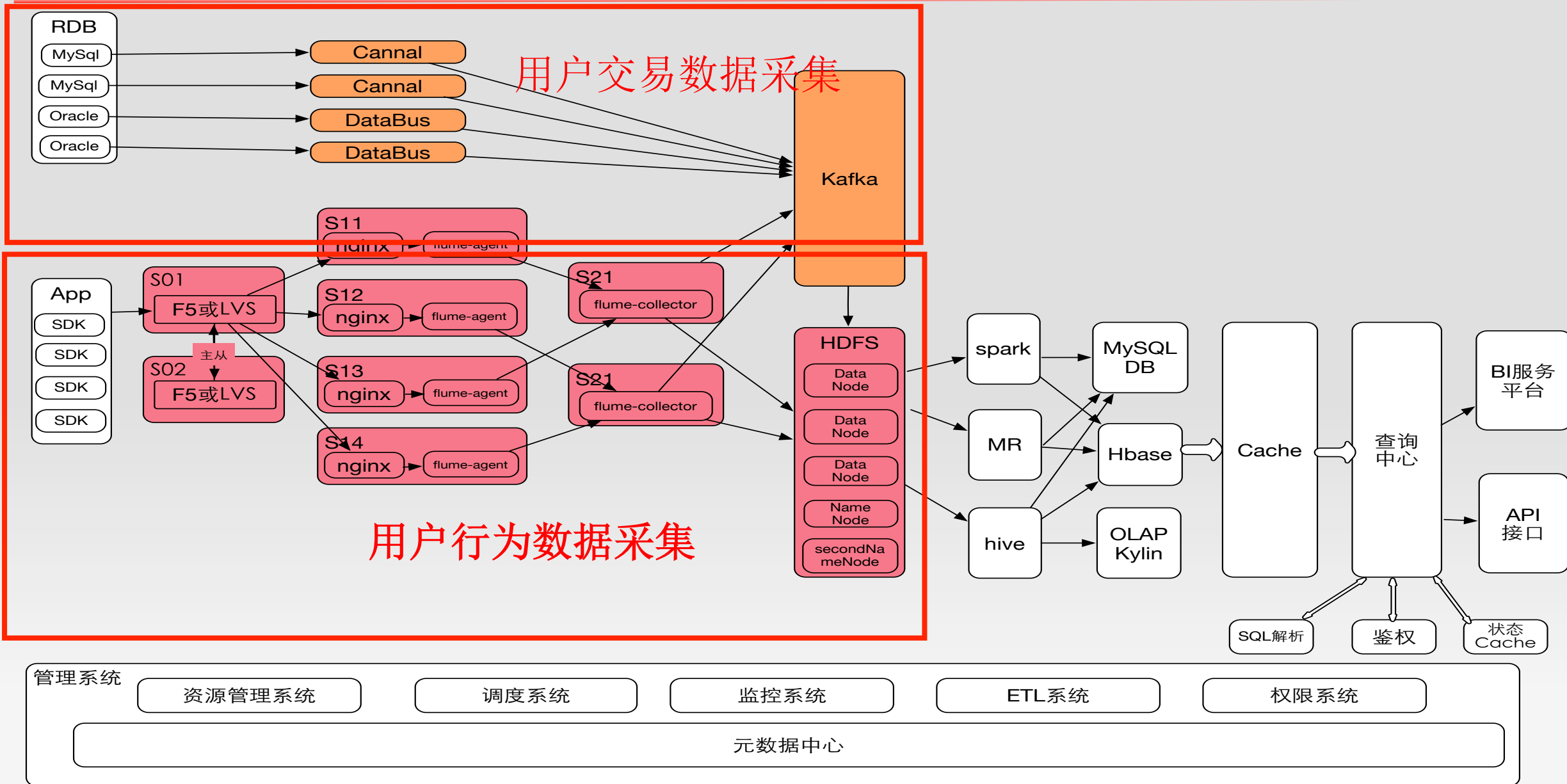




猫眼数据处理流程

猫眼用户的购票行为分析

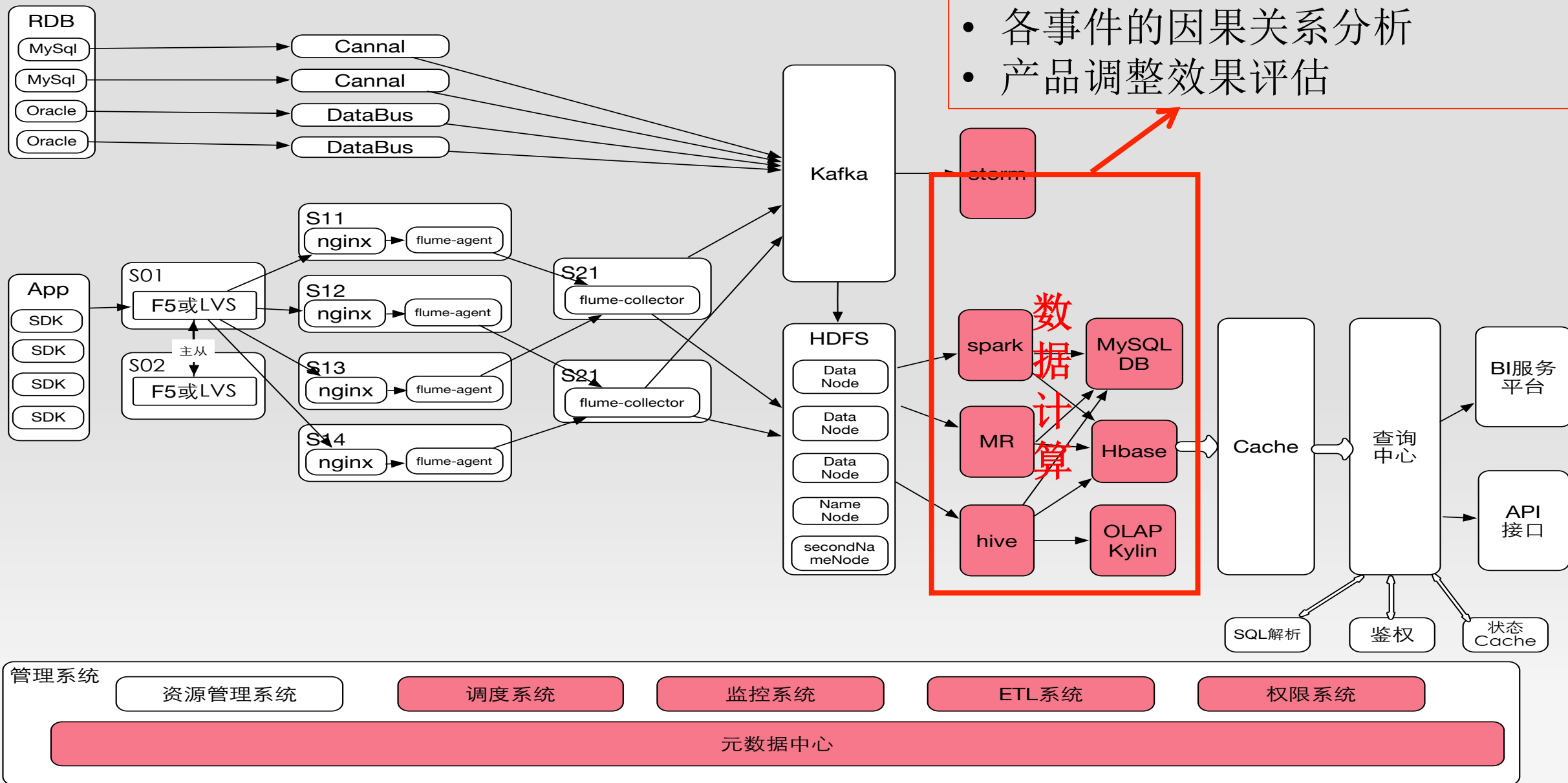




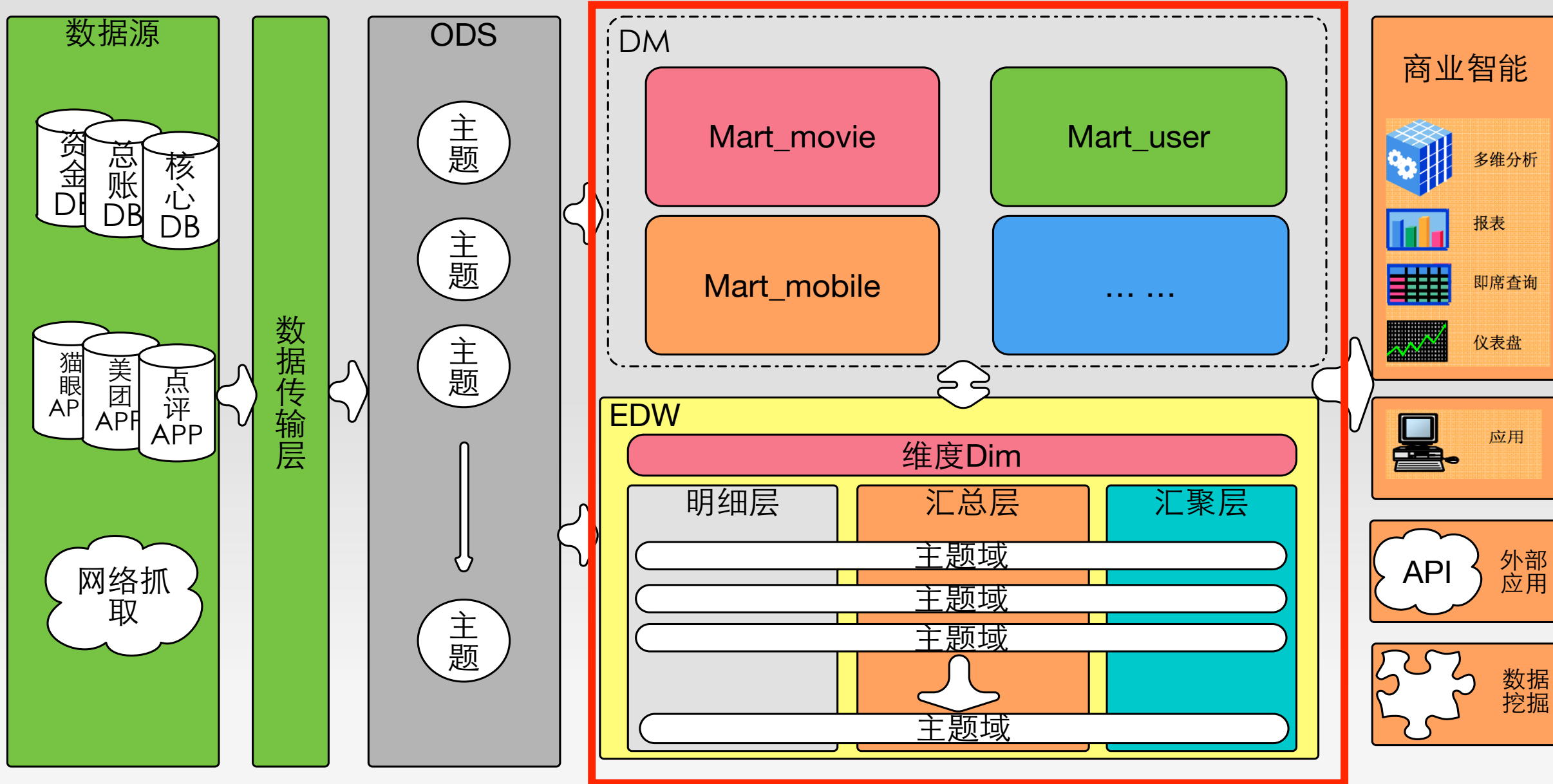
▶ 存储与计算

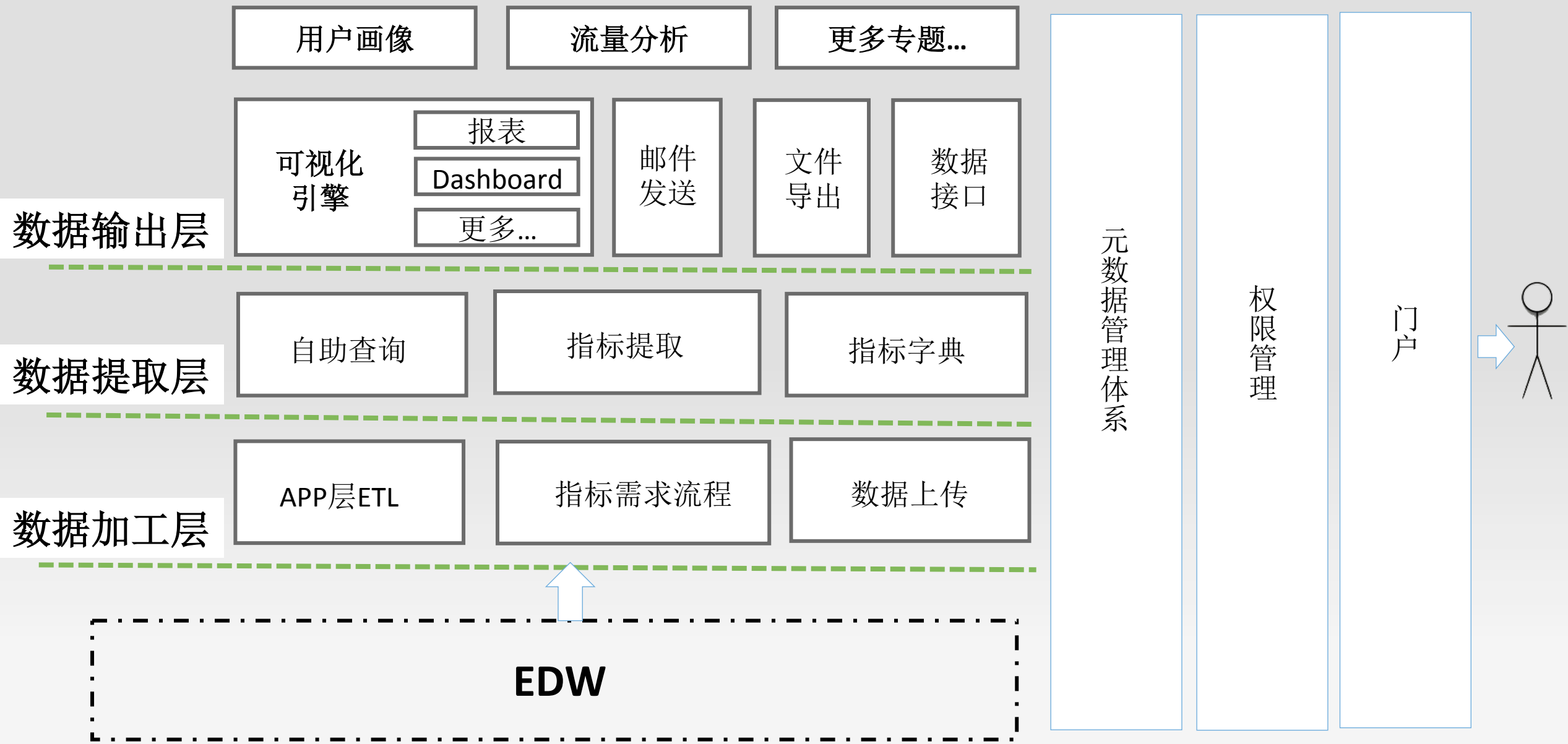


- 用户会话中不同事件的串联
- 各事件的因果关系分析
- 产品调整效果评估



加工与管理







高并发下任务调度和系统可扩展性

大数据量存储和高效查询的设计



一、大数据在猫眼的价值落地

- 猫眼有什么数据
- 数据的应用场景

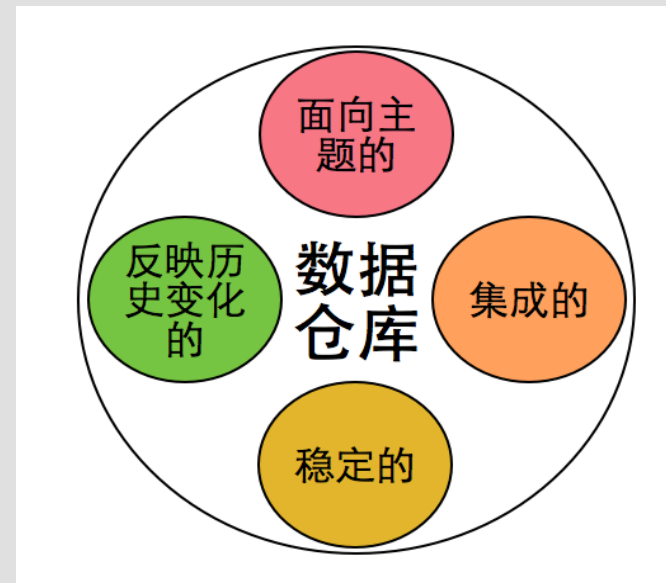
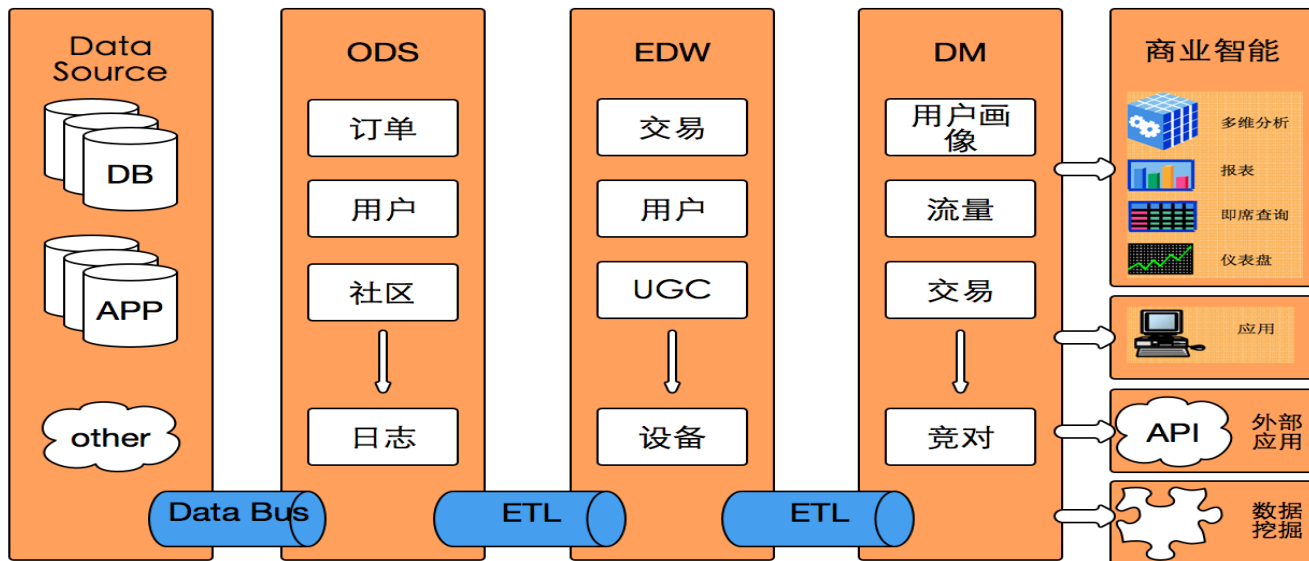
二、猫眼大数据体系架构

- 整体技术架构
- 各方向核心技术

三、猫眼数据仓库建设

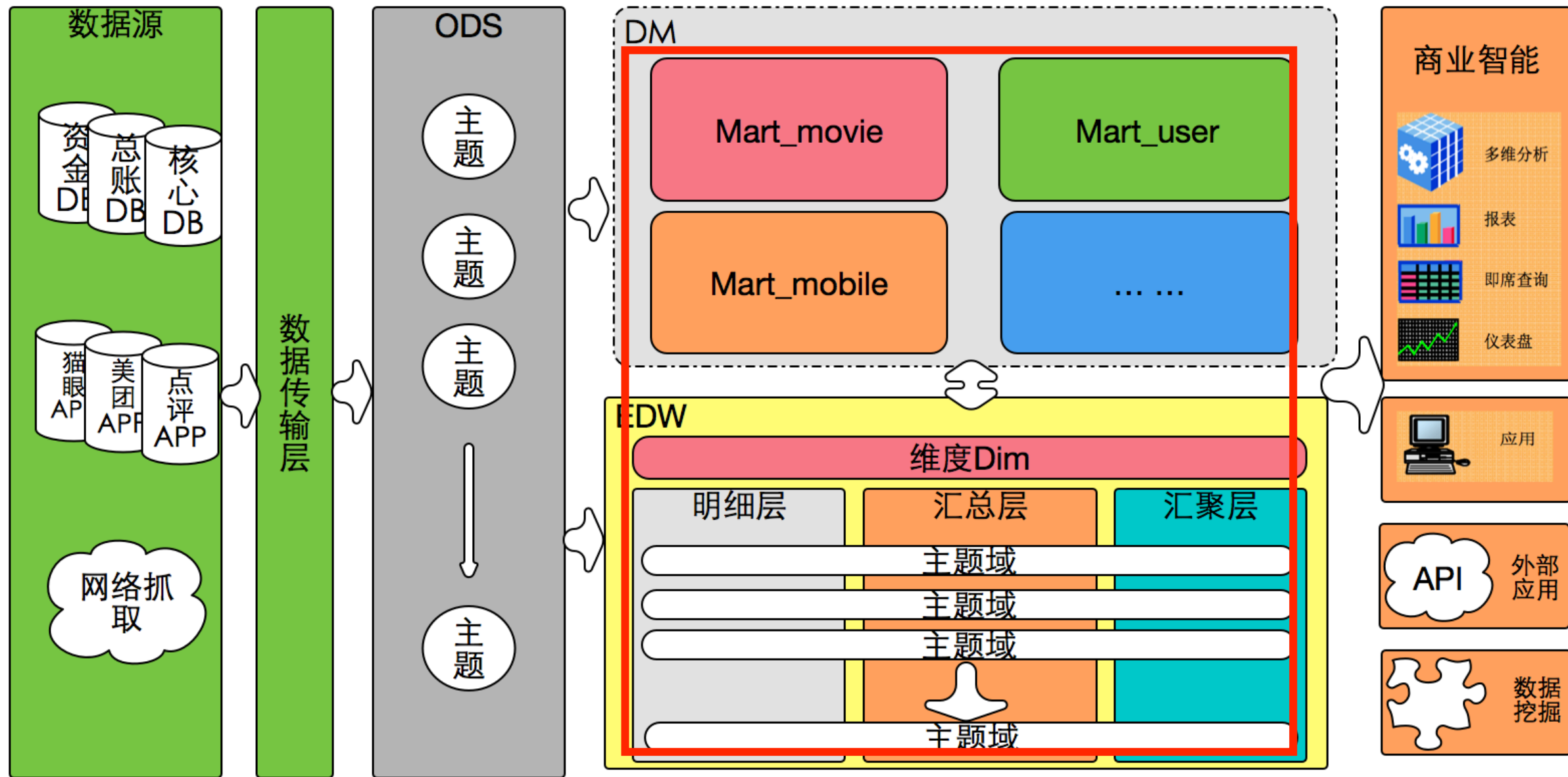
四、猫眼数据平台建设

数据仓库定义

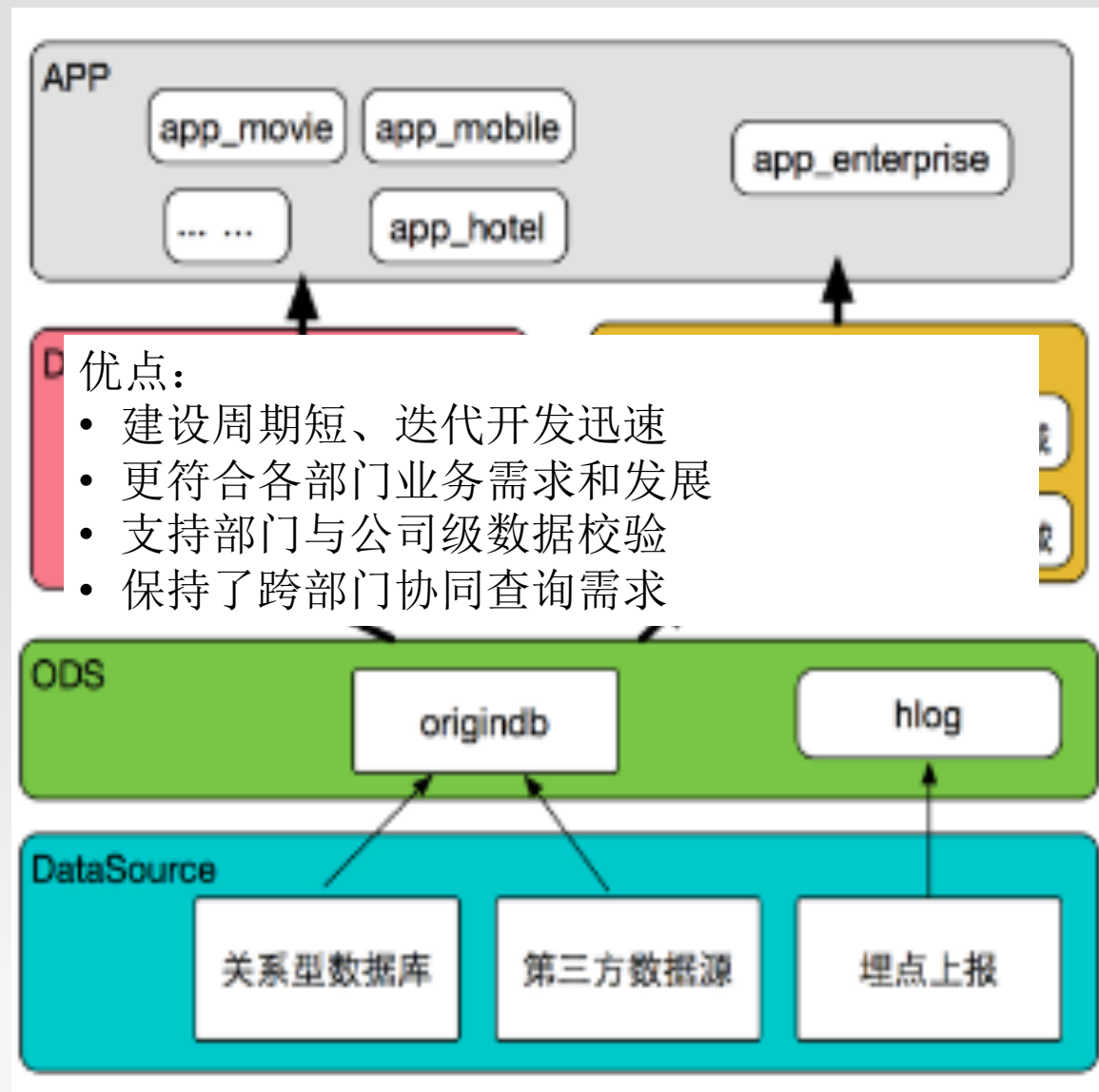
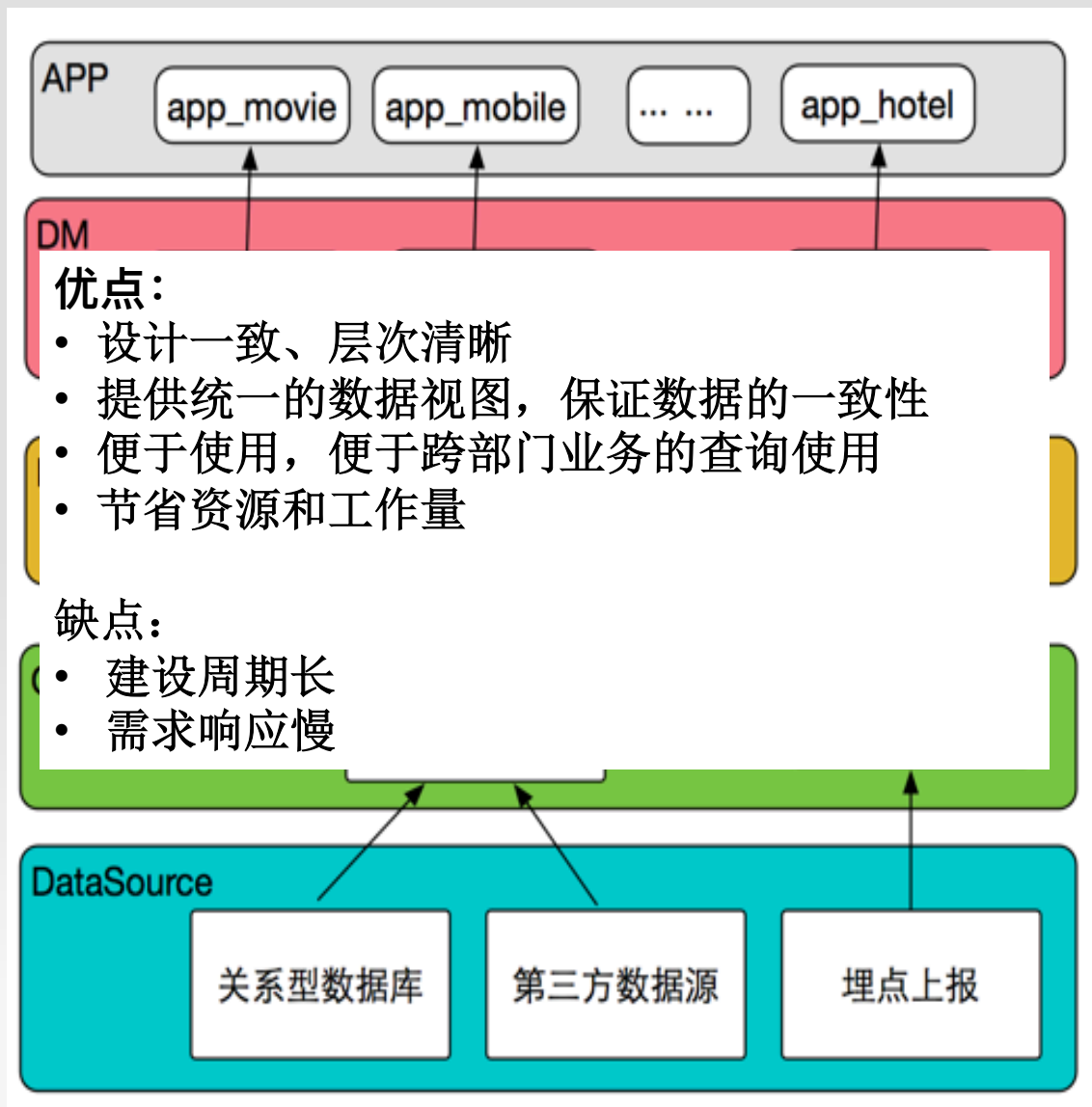


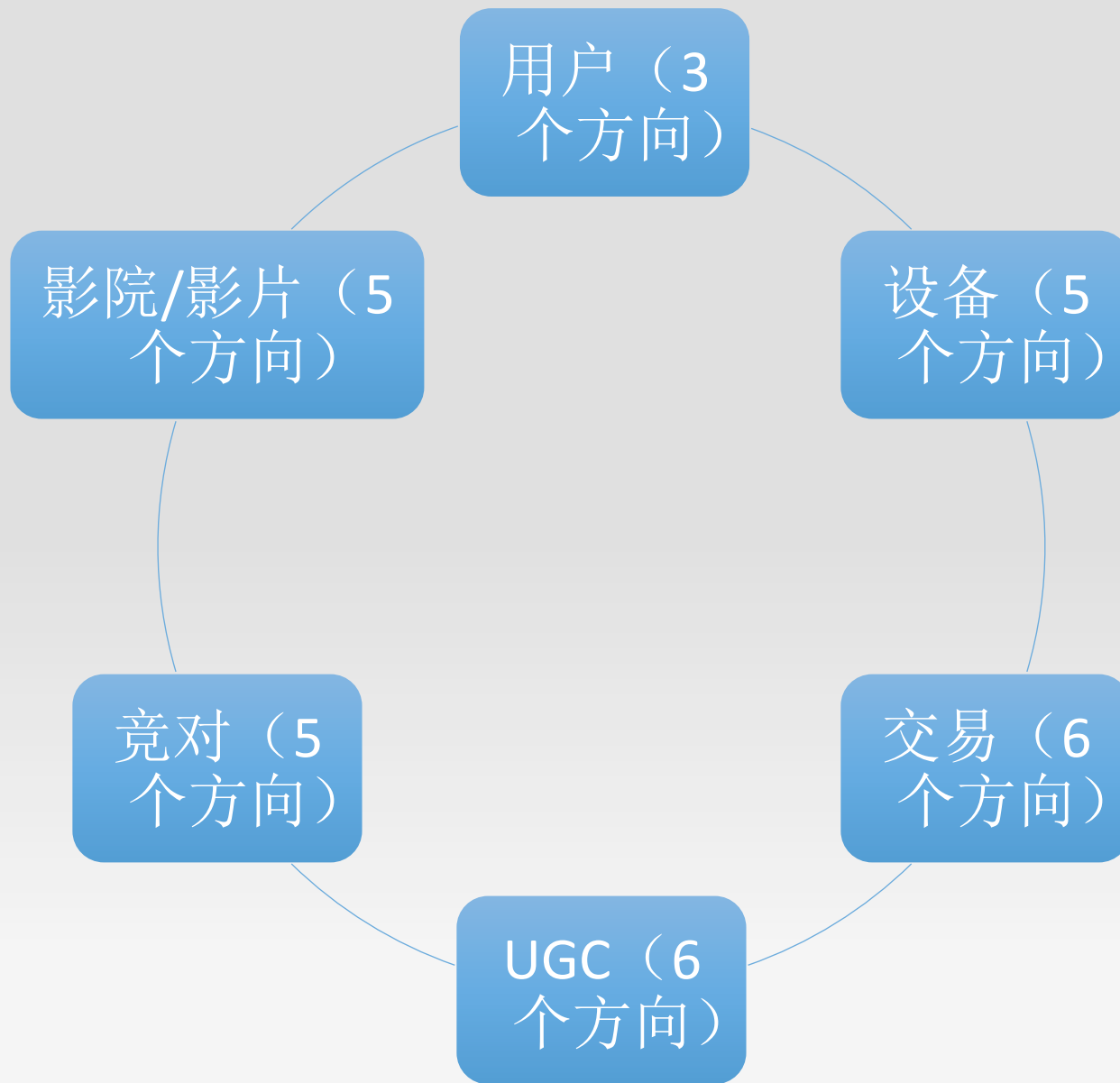
DB VS DW	ODS VS DW	DM VS DW
<ul style="list-style-type: none">• 面向应用的• 事务型的• 高实时性的• 少数据、高访问的	<ul style="list-style-type: none">• 面向主题的• 集成的• 可更新的• 当前或接近当前的	一组特定的、针对某个主题域、某个部门或者某些特殊用户而进行分类的数据集合
<ul style="list-style-type: none">• 面向主题的• 分析型的• 低实时性的• 数据量大、访问频次低的	<ul style="list-style-type: none">• 面向主题的• 集成的• 稳定的• 反映历史变化的	面向企业的、完整主题域的数据集合

猫眼数据仓库架构

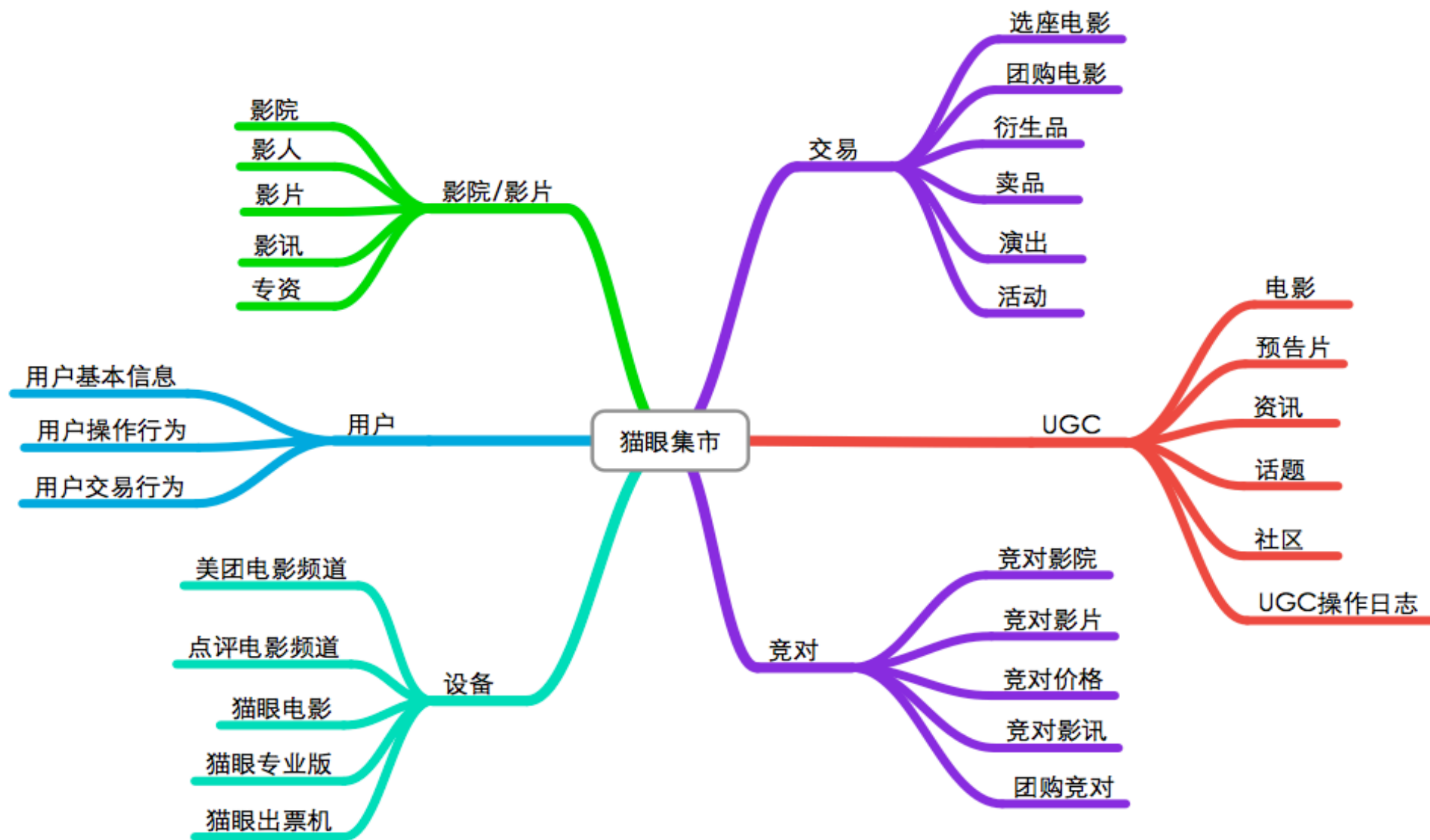


猫眼数据仓库建设变更历程





主题域内容组成



一、大数据在猫眼的价值落地

- 猫眼有什么数据
- 数据的应用场景

二、猫眼大数据体系架构

- 整体技术架构
- 各方向核心技术

三、猫眼数据仓库建设

四、猫眼数据平台建设

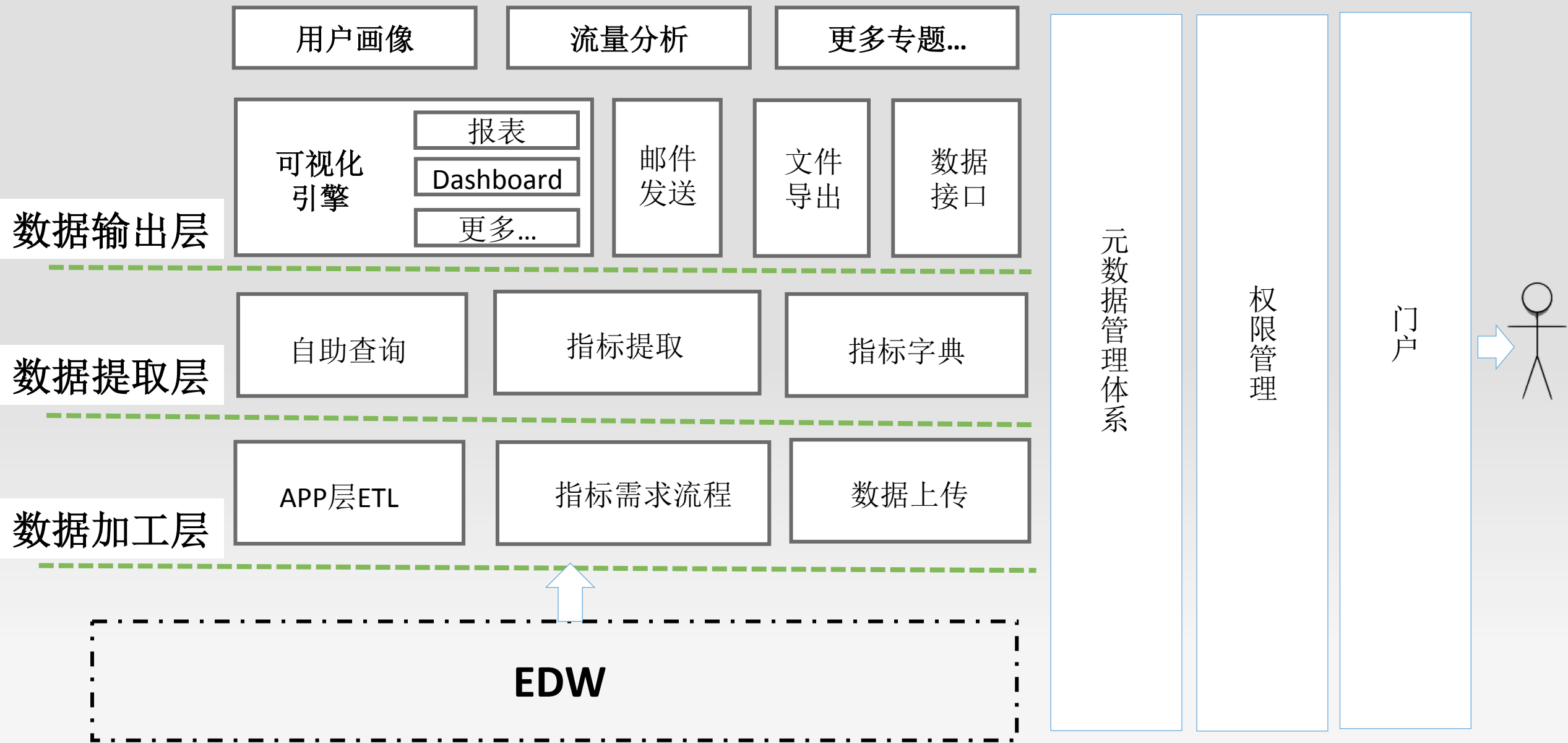
权限系统
数据源系统
报表系统
数据魔方系统

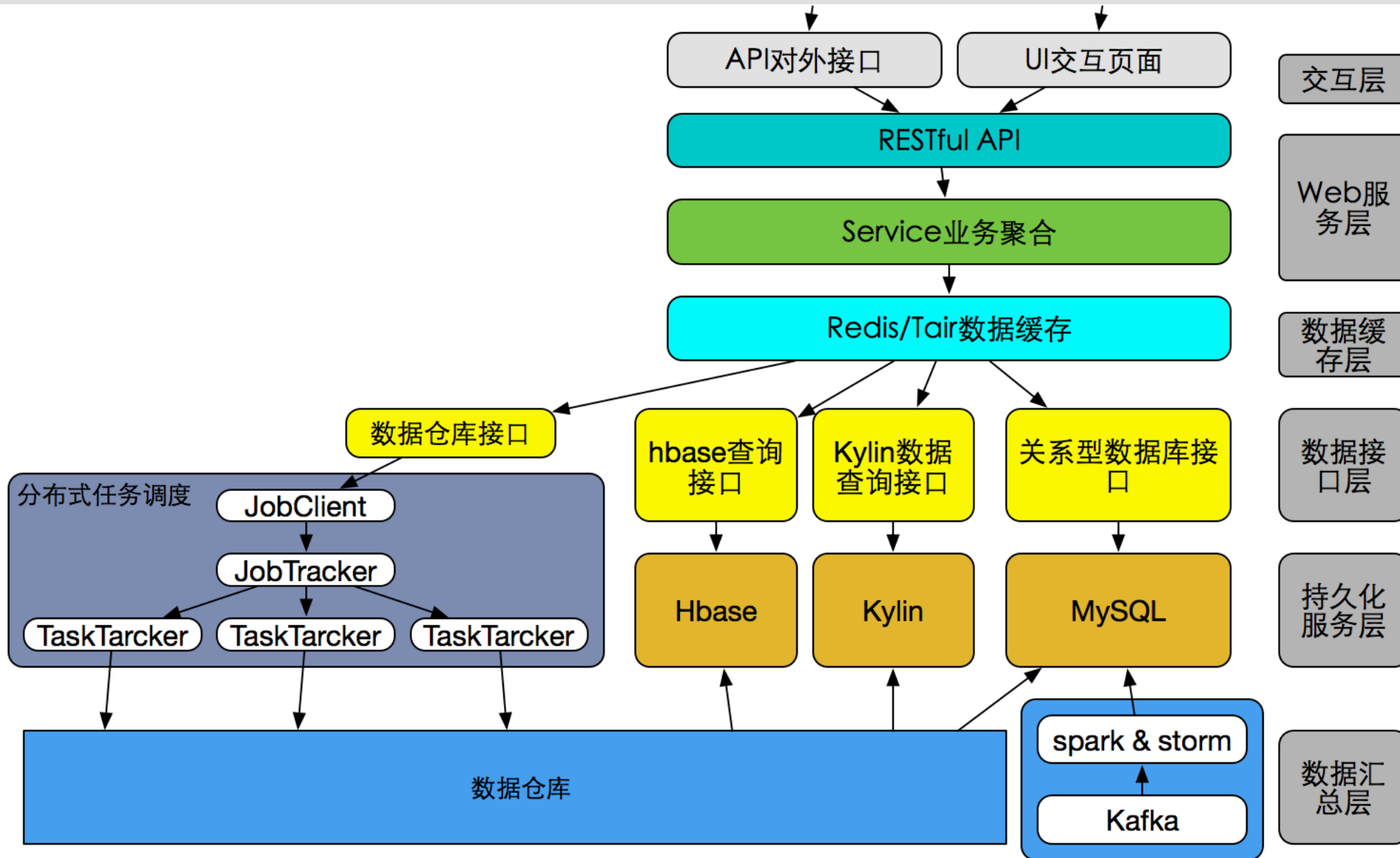
流量分析系统
数据直播间
BI系统监控

指标字典
集群监控管理

用户画像系统
任务调度系统

数据源与数据
质量监控系统





大盘概览 2017/01/08

12,747.5

大盘票房 (万)

日环比
-19.1%

周同比
-49.5%

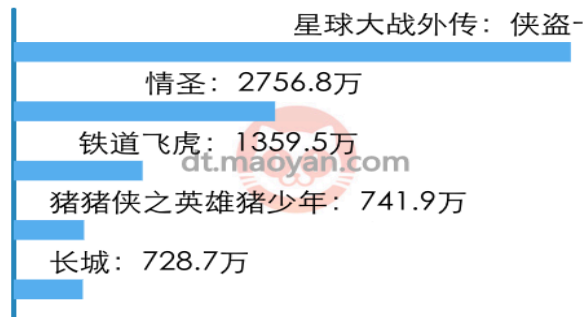
387.4

大盘人次 (万)

日环比
-18.5%

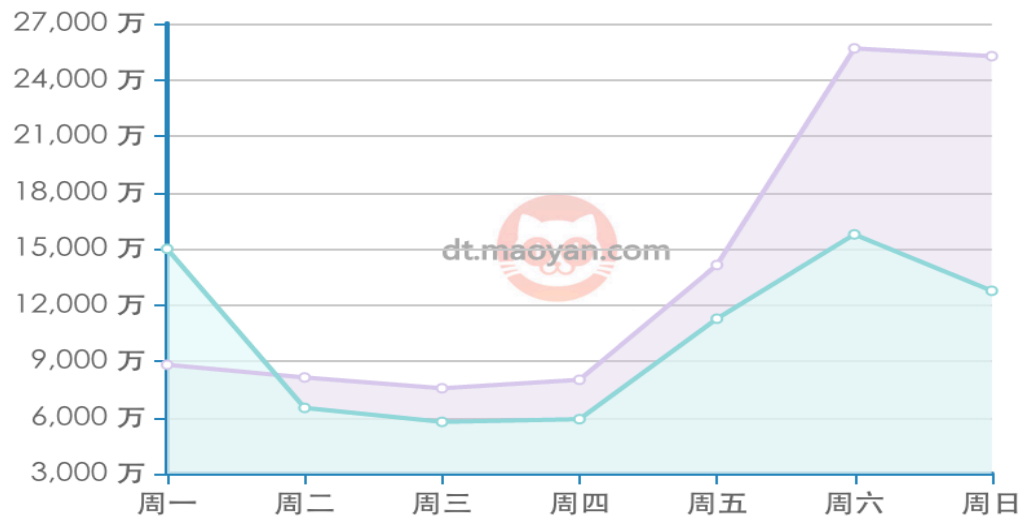
周同比
-50.2%

影片票房排名 2017/01/08



影片	票房	票房占比	线上化率
星球大战外传...	5,875.6万	46.0%	34.2%
情圣	2,756.8万	21.6%	19.9%
铁道飞虎	1,359.5万	10.6%	7.5%
猪猪侠之英雄...	741.9万	5.8%	4.7%
长城	728.7万	5.7%	4.1%

影市大盘 2016/12/26 - 2017/01/01 2017/01/02 - 2017/01/08

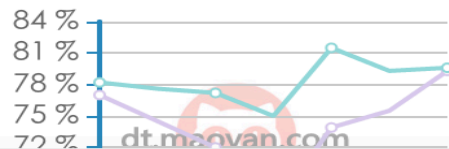


其他指标 2016/12/26 - 2017/01/01 2017/01/02 - 2017/01/08

大盘人次



线上化率

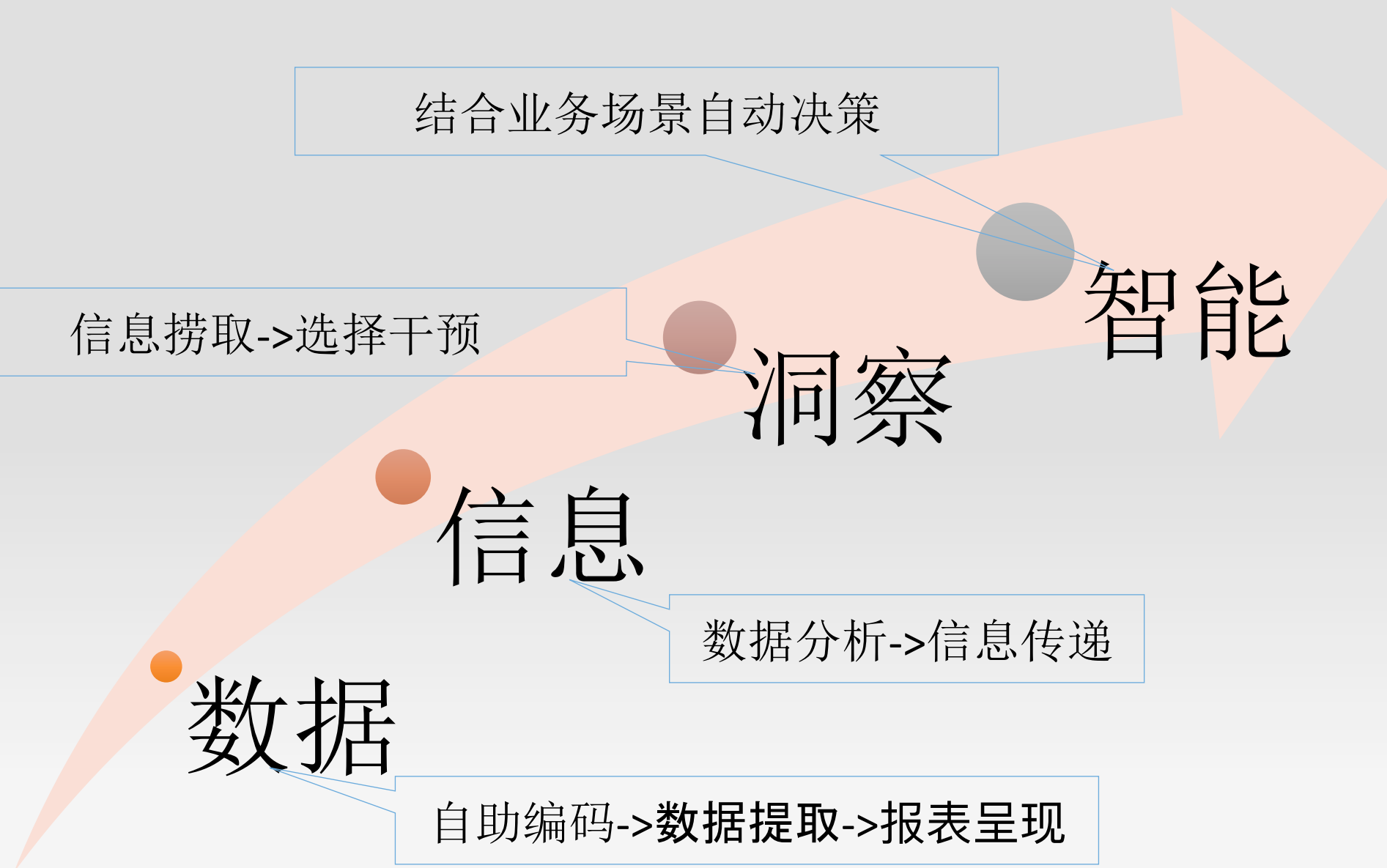


一周热映影片2017/01/02-2017/01/08

周五	上映3天
星球大战外传：侠盗一号	2.1亿
玩命直播	718.2万
一万公里的约定	248.3万
一路逆风	234万

一周待映影片2017/01/09-2017/01/15

周一	今天上映
青海湖畔	180人想看
周五	4天后上映
太空旅客	39,748人想看
魔弦传说	10,365人想看



Thank **You**!

