



大数据生态系统分析

卢亿雷 From AdMaster

johnlya@163.com

@johnlya

目录

1 > 大数据生态系统架构

2 > 大数据生态系统发展

3 > 离线、实时系统分析

4 > 数字营销案例分析

5 > Q & A

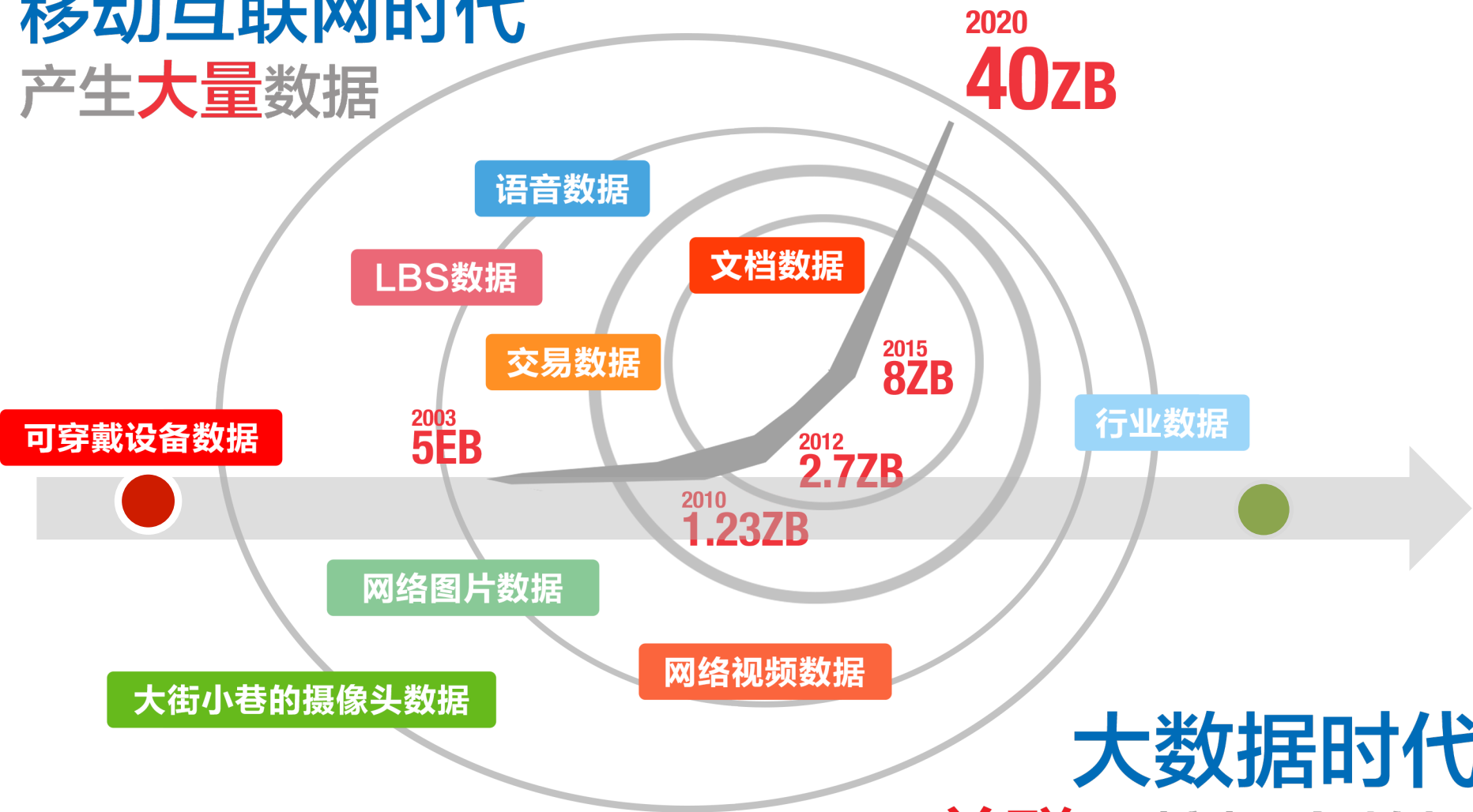
目录

- 1 大数据系统架构
- 2 大数据生态系统发展
- 3 离线、实时系统分析
- 4 数字营销案例分析
- 5 Q & A

数据增长速度

移动互联网时代

产生大量数据



大数据时代

关联、挖掘大数据

大数据生态系统架构

应用服务

Pig

Hive

Druid

Flink

在线计算
(HBase)

离线计算
(MapReduce)

流式计算
(Storm)

实时计算
(Spark)

Zookeeper

Kafka

MQ

MySQL
Elastic
Search

YARN

HDFS

OS (操作系统)

大数据生态系统架构

Applications

Hive

Cascading

Mahout

Pig

Crunch

Data processing engines

MapReduce



Flink



Spark



Storm



Tez



App and resource management

Yarn

Mesos

Storage, streams

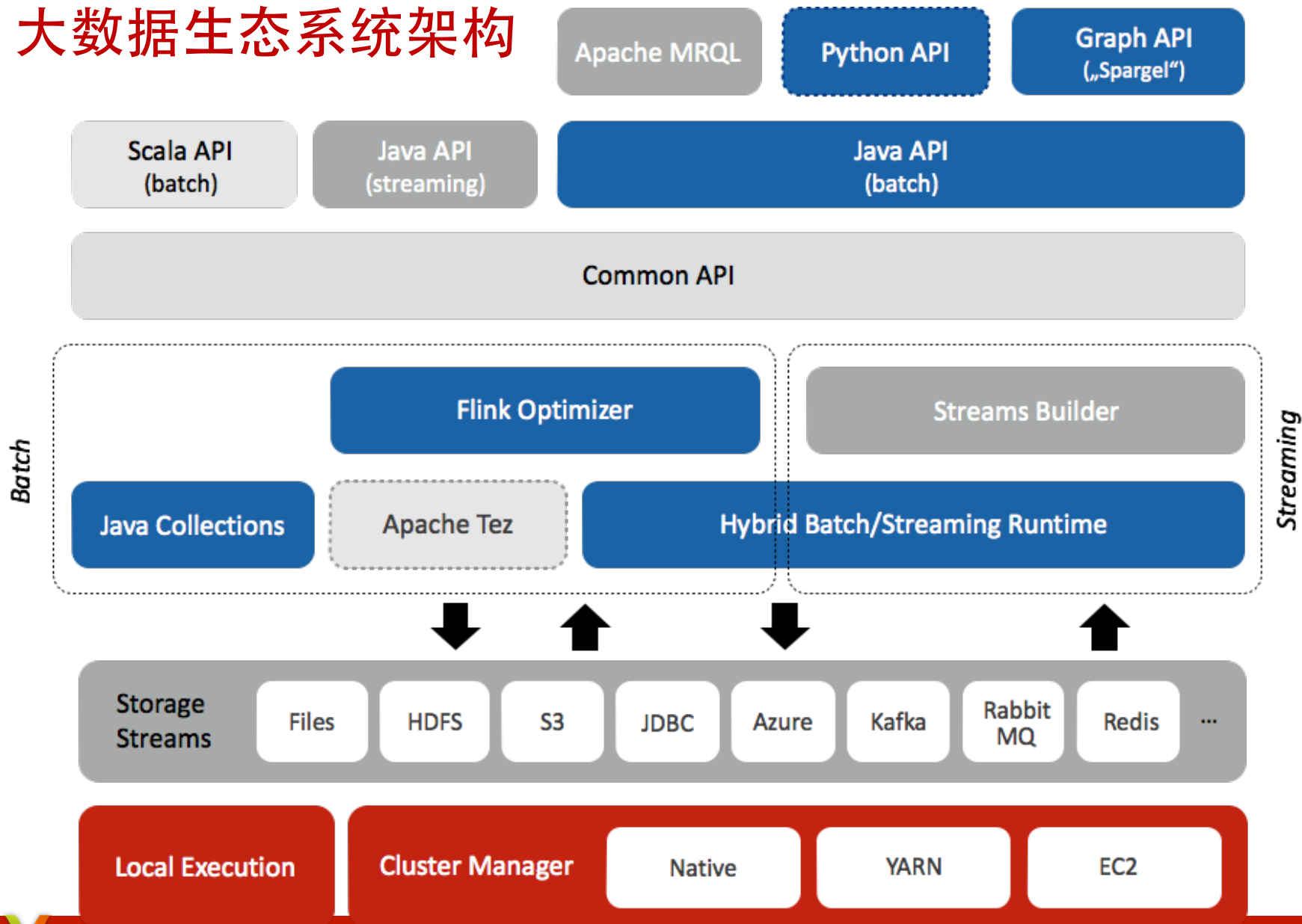
HDFS

HBase

Kafka

...

大数据生态系统架构



目录

- 1 大数据系统架构
- 2 大数据生态系统发展
- 3 离线、实时系统分析
- 4 数字营销案例分析
- 5 Q & A

1. In the beginning the web was full of untapped data.



2. Nutch is built to crawl this web data.



3. All this data had to be stored somewhere, which gave rise to HDFS.



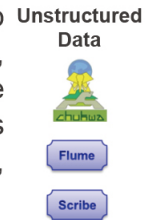
4. Now that we can collect and store this data, how do we use it?



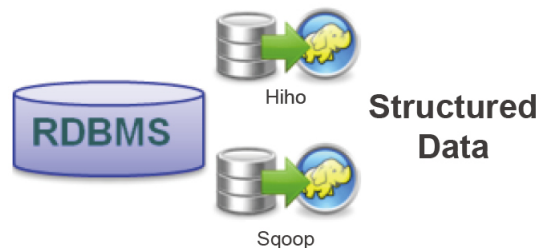
5. Map Reduce framework is built for coding and running analytics.



6. Unstructured Data (e.g. web logs, click streams, Apache logs, server logs, etc.) is able to be incorporated into the mix thanks to tools like Fuse, WebDAV, Chukwa, Flume and Scribe.



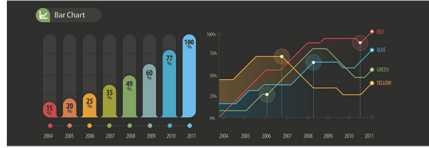
7. Hiho and Sqoop allow for data to be loaded into HDFS; RDBMS joins the Hadoop bandwagon.



8. High-level interfaces such as Pig, Hive and Jaql are required over low-level Map Reduce programming.



9. BI tools like Intellicus with advanced UI reporting (i.e. drilldown) are born.



Intellicus Dashboards

10. Workflow tools over Map Reduce processes and high-level languages are created.

Workflow



Cascading

11. New tools like Hue, Karmasphere, Cacti, Ganglia and the Eclipse plugin arrive, which can monitor/manage Hadoop, run jobs/hive, view HDFS and offer high-level viewing.



**Monitor /
Manage
Hadoop
Ecosystem**

12. Support frameworks such as Avro (serialization) and Zookeeper (coordination) see the day.

Support



13. More high-level interfaces that use Mahout and Elastic Map Reduce are created.

More High Level Interfaces



14. OLTP becomes possible thanks to Hbase.

OLTP



15. OLAP tools like Spark and Flink make their debut.

OLTP



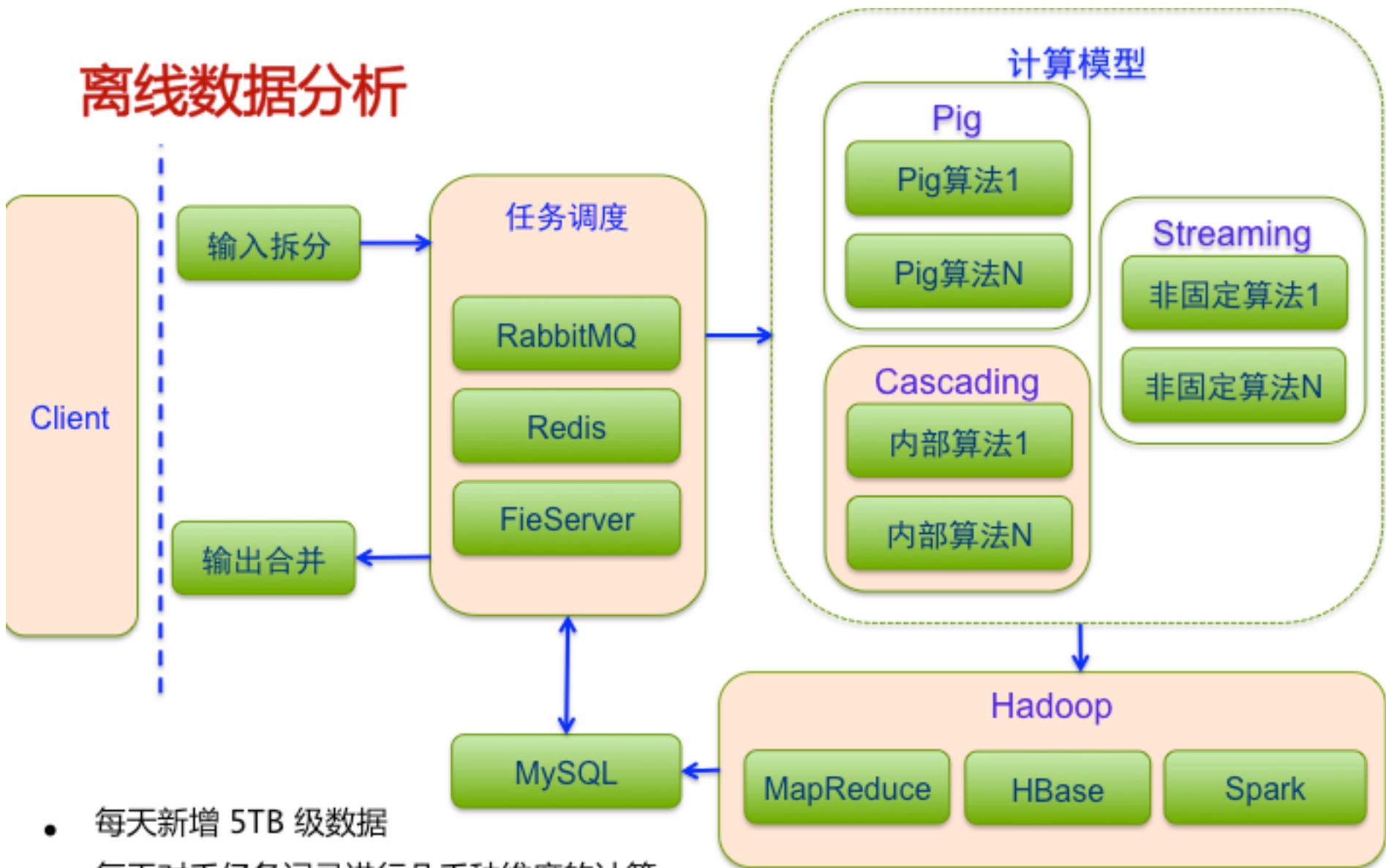
16. Drill supports a variety of NoSQL databases and file systems.



目录

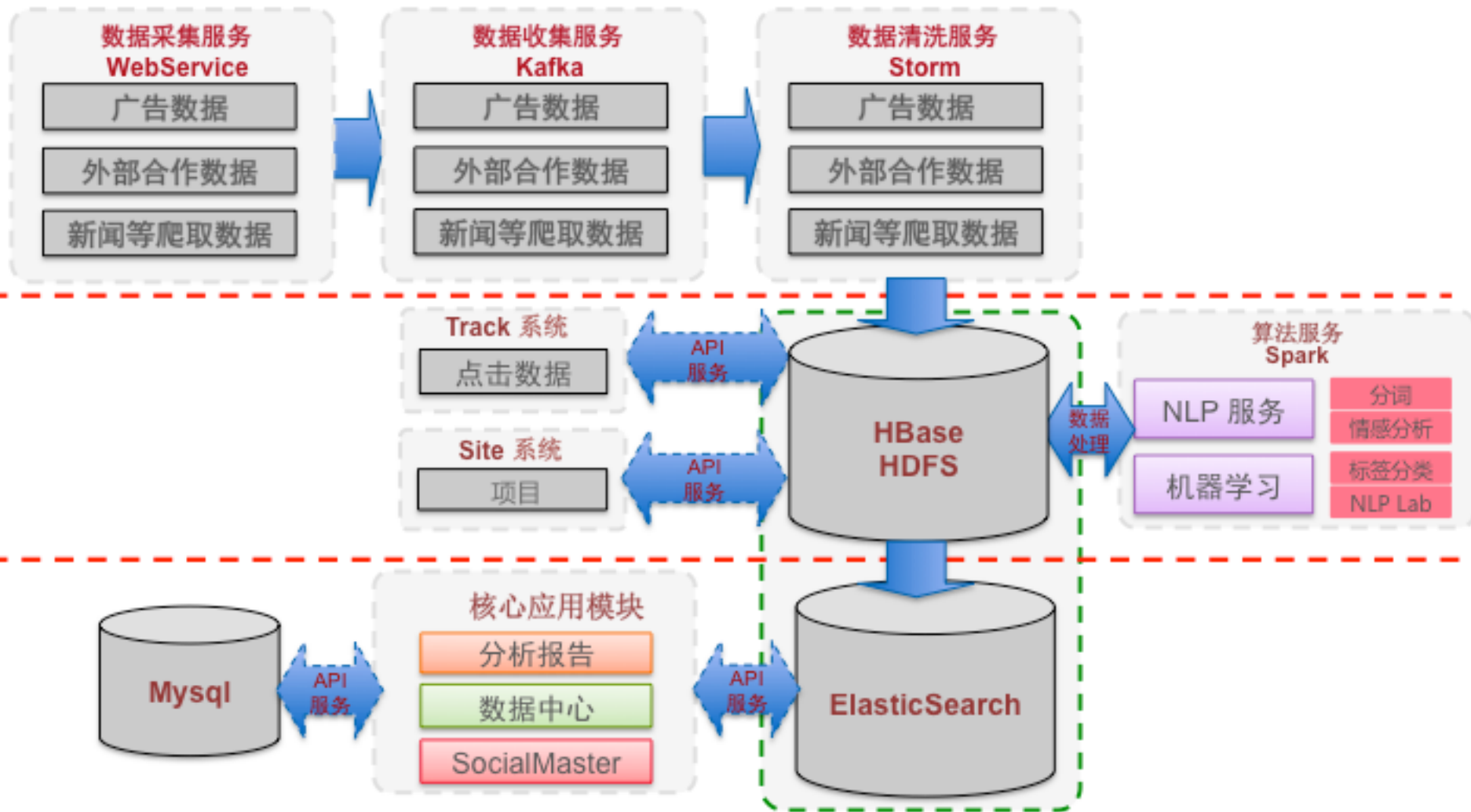
- 1 大数据系统架构
- 2 大数据生态系统发展
- 3 离线、实时系统分析
- 4 数字营销案例分析
- 5 Q & A

离线数据分析



- 每天新增 5TB 级数据
- 每天对千亿条记录进行几千种维度的计算

在线数据分析



系统对比

◆ NFS & HDFS

Storage

◆ Pig & Hive

◆ MapReduce & Spark

◆ Spark & Flink

◆ Storm & Spark Streaming

Computing

◆ MySQL & HBase

◆ MongoDB & HBase

◆ HBase & Cassandra

Database

◆ ElasticSearch & Druid(Pinot)

OLAP

目录

- 1 大数据系统架构
- 2 大数据生态系统发展
- 3 离线、实时系统分析
- 4 数字营销案例分析
- 5 Q & A

数字营销精准投放（3W原则）



最合适的时间
When



合适的地点
Where



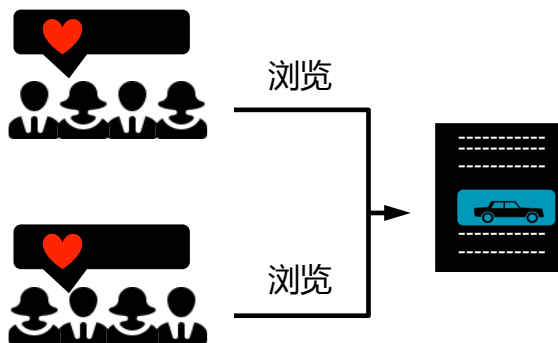
合适的商品
What



合适的人
User

传统与DSP广告对比

传统广告



目标用户成本 ¥ 50

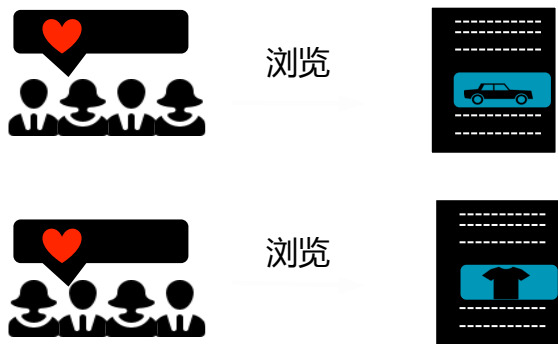


非目标用户成本 ¥ 50



汽车广告主
总成本 ¥ 100

DSP广告



目标用户成本 ¥ 50



目标用户成本 ¥ 50



汽车广告主
总成本 ¥ 50

服装广告主
总成本 ¥ 50

数字营销用户网络流程



- 基于cookie技术的在线消费者和广告效果调研。



- 多维度跨屏广告监测和评估（网络，手机，pad，电视）



- 跳转后行为监测，了解消费者的行为习惯，优化网站设计

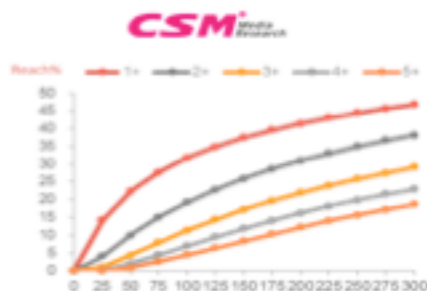


- 跨电商平台的综合分析，有助于制定销售策略，优化销售渠道



- 通过社会化媒体聆听，了解消费需求及市场营销效果，了解消费者眼中的“我们”，了解我们的竞争对手。

广告主整合跨屏投放策略



电视
投放
优化

PC端
视频
优化

Mobile
视频
优化



输入媒体重叠率的调
研数据



Ipsos MAC 益普索中
国媒体地图辑



OPTIMIZATION
测试不同的组合方案

输入媒体价格
TV : CPRP
OTV : CPM

OUTPUT
优化方案输出

广告反作弊分析

作弊识别模型

广告投放作弊

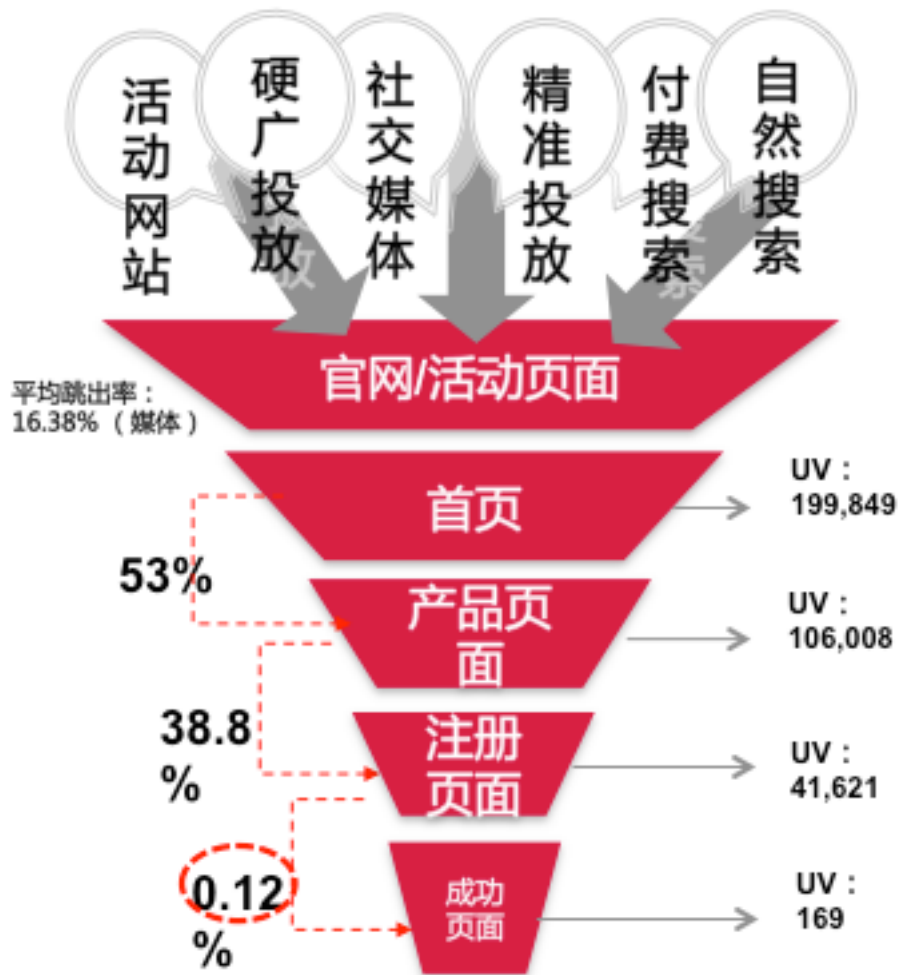
监播
实录
Snap
实时
展现

定投
识别
智能
监督

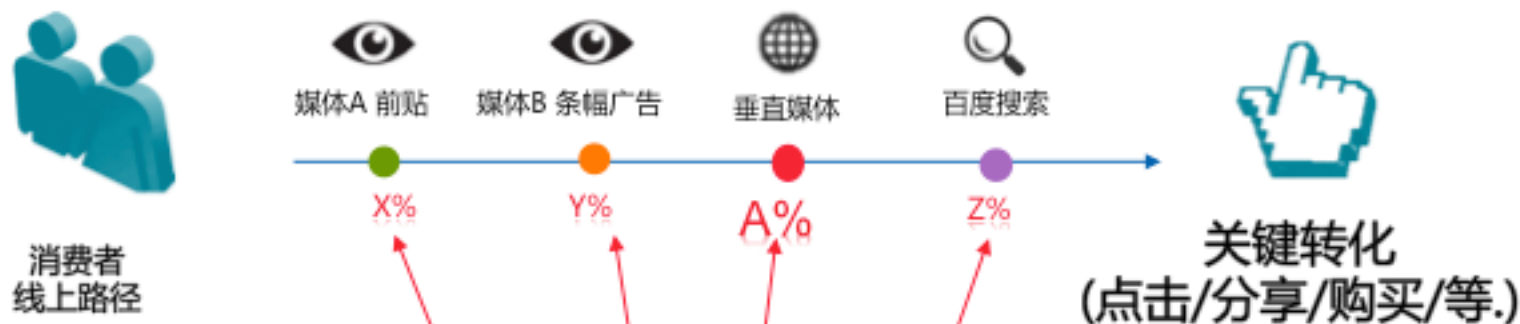
多种
维度
佐证
判断



全流量分析，发掘营销效果



移动端消费分析-归因模型



AdMaster归因模型

输入变量：分媒体曝光数据

输出变量：关键转化（点击、分享或购买等行为）

可口可乐分析：“昵称瓶”



- 用户男女性别
- 用户年龄分段
- 用户收入
- 用户爱好
- 用户情感



- 声量
- 互动性
- 发帖率



Q & A

Email:johnlya@163.com
微信:johnlya