

# 小米程序化广告交易平台 (MAX)的架构实践

欧阳辰

小米MIUI商业产品部



促进软件开发领域知识与创新的传播



关注InfoQ官方信息  
及时获取QCon软件开发者  
大会演讲视频信息



[北京站] 2016年12月2日-3日  
咨询热线: 010-89880682



[北京站] 2017年4月16日-18日  
咨询热线: 010-64738142

# 内容提要

- 小米程序化交易广告介绍
- 架构整体演化过程
- 在线部分的演化
- 数据分析的演化
- 算法分配的演化
- 架构领悟与架构师的KPIs

# 小米MIUI商业产品部

- 日活超过千万的App有21款
- 品牌传播，应用分发，视频广告等
- Intelligent Marketing (IM)  
智能数据，智能科技，智能设备



# 欧阳辰

15年的互联网研发老兵

	开发主管  <b>10<sup>8</sup></b> 	高级开发经理/工程师  Microsoft  搜索和广告	架构师/主管   广告平台大数据
7年	3年	10年	2年

公众号：

互联居



[www.ouyangchen.com](http://www.ouyangchen.com)

# 小米广告系统架构和关键指标



关键指标:



系统:

TPS > 80K/S

可用性 > 99.9%

监控:

200+ 警报

120+ 视图

事故处理:

TTD < 5分钟

TTE < 15分钟

TTM < 45分钟

# 小米广告平台的技术栈

小米开源

开源软件

LVS

Nginx

接入层

Z  
o  
o  
k  
e  
e  
p  
e  
r

Java

Paoding-Rose

Thrift\_RPC

Lucene

Redis

Aerospike

Kafka

MySQL

HBase

Storm

Druid

中间层

HBase

HDFS

Hive/Pig

Hadoop

Impala

Spark

Tensor  
Flow

离线数据

Maven

Git,Nexus

Open-  
Falcon

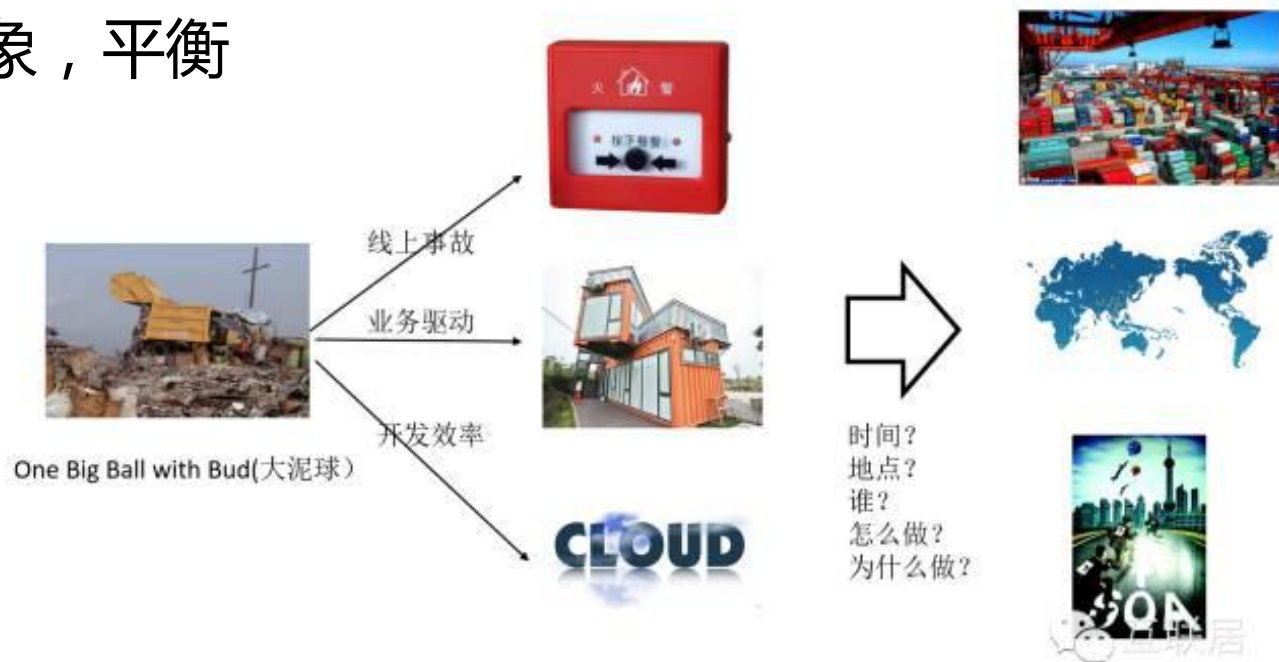
X-Box

MARATHON  
/ Docker

基础服务

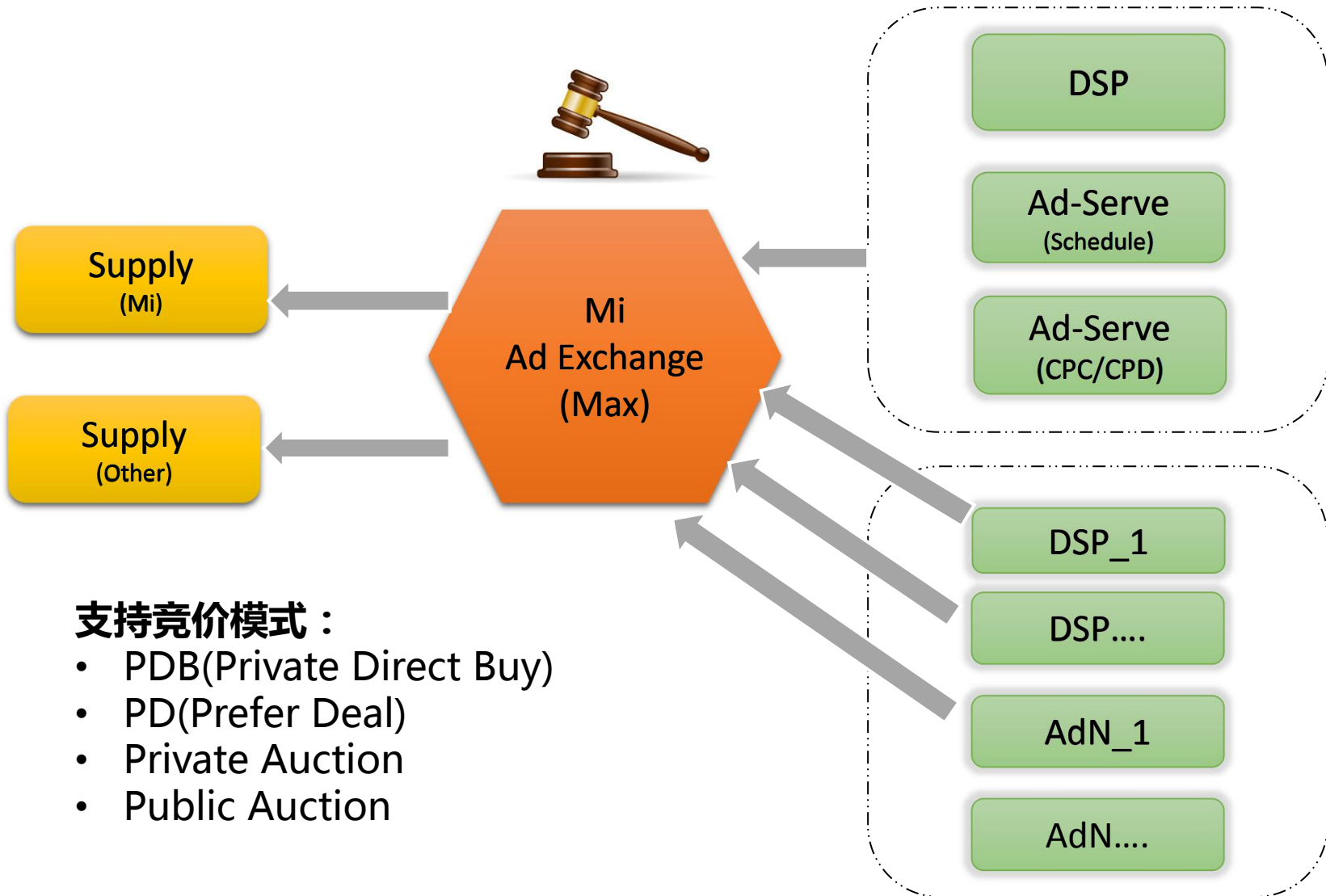
# 广告平台的架构演化(10X)

1. “加”：新业务疯狂上线，耦合
2. “减”：服务化，解耦
3. “乘”：微创新，引入新技术
4. “除”：抽象，平衡





# 小米广告交易平台(MAX)



# 小米MAX大计算挑战-大量优化工作

## 请求规模

- 近百亿日请求
- 对接数十个DSP/Ad Network
- 支持多模式：PDB,PD,RTB等

## MAX的性能

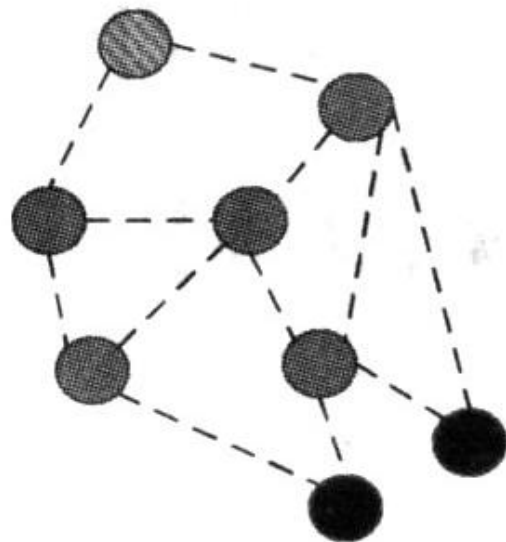
- DSP 100毫秒返回结果
- 99P <200ms
- 单机最大支持5000 QPS

## 我们做的优化工作：

1. 线程池的调优
2. 独立线程清理器
3. HTTP Keep Alive优化
4. 快速JSON解析
  - Netty JSON
  - Fast JSON
  - Gson ✓
1. 协程(Coroutine)
2. 智能流量分配
3. 多实例部署
4. 多机房优化
5. ....

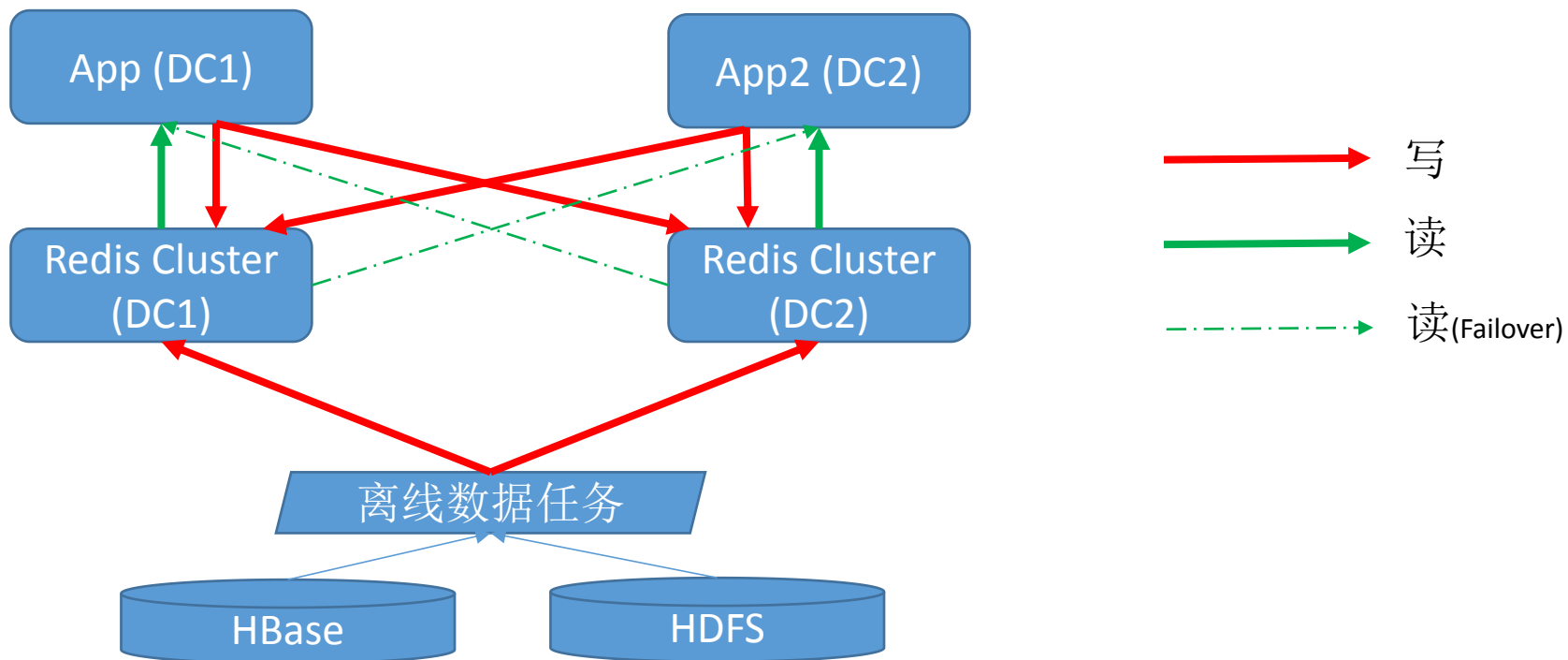
# 小米MAX - 雪崩效应的应对

- 服务的业务隔离
- 尽早失败，避免等待
- 连接池共享和独享
- 服务无状态设计
- 负载均衡和熔断
- 降级服务
- 灰度发布和主动监控



# Redis 的实践之旅 (1/3)

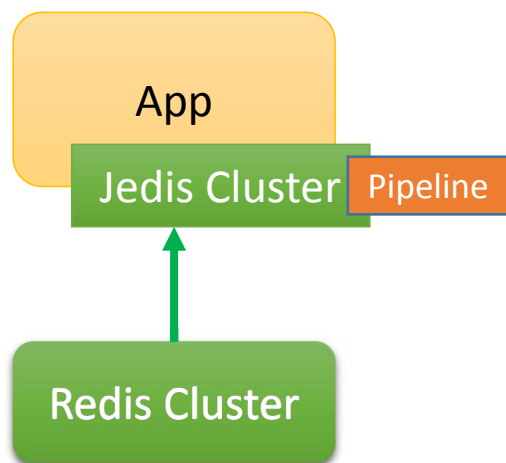
- Redis单机-》Redis Cluster(3.0Beta1-3.2)
- Redis Cluster -》Cross DC Cluster (双写)



“单个集群内存容量1TB+，15亿+的Key，百万级QPS吞吐量”

# Redis 的实践之旅 (2/3)

- Jedis Cluster + Pipeline ( 已经开源 )
- K多V小 : String -> Hash ( 80% Space saved, ziplist)
- 结合Bloom filter



Key->String



Sub(Key) →

<Key1, Value>

<Key2, Value>

.....

89e6d2b383471fc370d828e552c19e65      v

Sub(Key)      Key1

<https://github.com/cityonsky/jedis-cluster-ext/tree/master/redis>

# Redis的实践之旅(3/3)

容量和速度

监控和告警-Open-Falcon  
---没有指标，无法改进

Redis

DRAM

Aerospike

DRAM/SSD

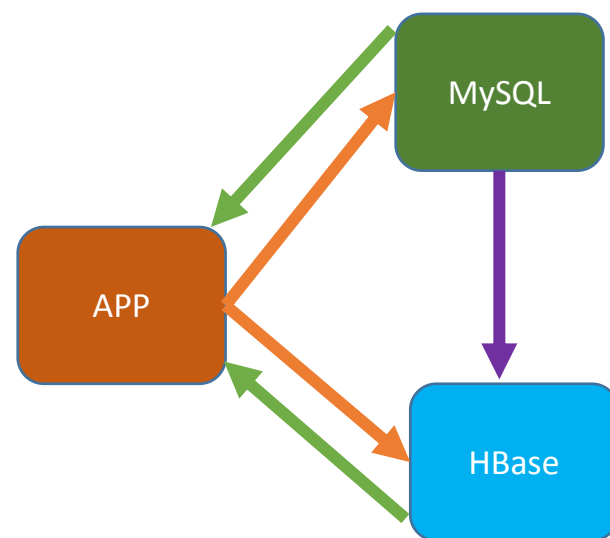
HBase

SSD/HDD

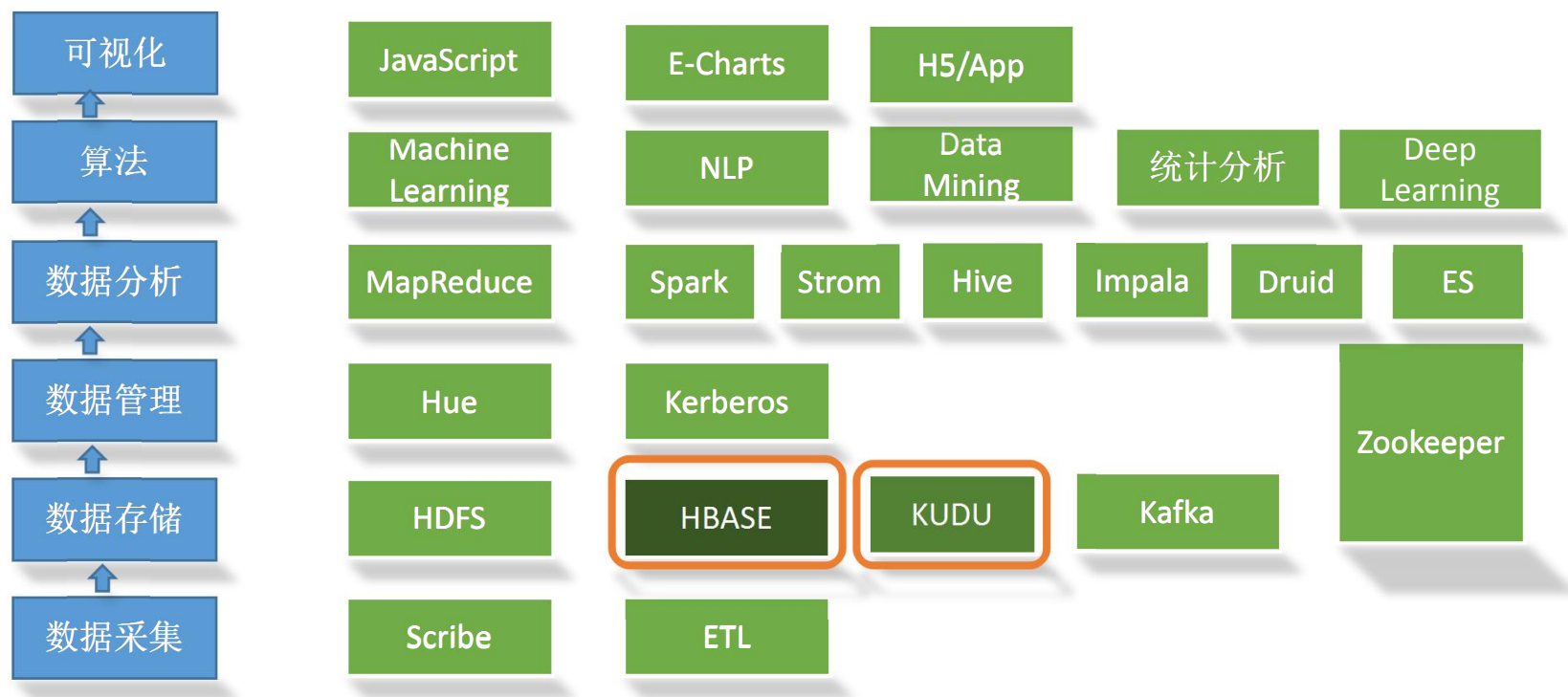
- ```
Redis Server 监控
```
- Redis local alive
  - Redis Connection
    - connected\_clients
    - connected\_clients\_pct( $\text{connected\_clients}/\text{maxclients}$ )
    - rejected\_connections
    - total\_connections\_received
    - client\_longest\_output\_time
    - client\_biggest\_input\_length
  - Redis Memory
    - used\_memory
    - redis\_memory\_usage\_peak
    - used\_memory\_rss
    - mem\_fragmentation\_ratio
  - Redis Cluster
    - cluster\_state
    - cluster\_slots\_assigned
    - cluster\_slots\_fail
    - cluster\_size
  - Redis Replication
    - master\_link\_status
    - master\_last\_io\_seconds\_ago
    - slave\_lag
    - connected\_slave
    - repl\_backlog\_size
  - Redis Performance
    - qps
    - cmdstat\_xxx
    - keys
    - lasttest\_fork\_usec
    - slowlog\_len
    - slowlog\_max\_time
    - total\_net\_output\_bytes
    - keyspace\_hit\_ratio
    - keyspace\_misses
  - Redis Persistence(rdb)
    - rdb\_last\_bgsave\_status
    - rdb\_last\_bgsave\_time\_sec
    - rdb\_last\_save\_time
    - rdb\_changes\_since\_last\_save
  - Redis Respond time
    - respond\_time\_max
    - respond\_time\_99\_max
    - respond\_time\_99\_avg

# 如何从MySQL 平滑迁移到HBASE?

1. 双写HBase和mysql
2. 迁移历史数据 ( 使用旧时间戳)
3. 双读HBase和mysql , 验证数据一一一致性
4. 灰度返回HBase结果



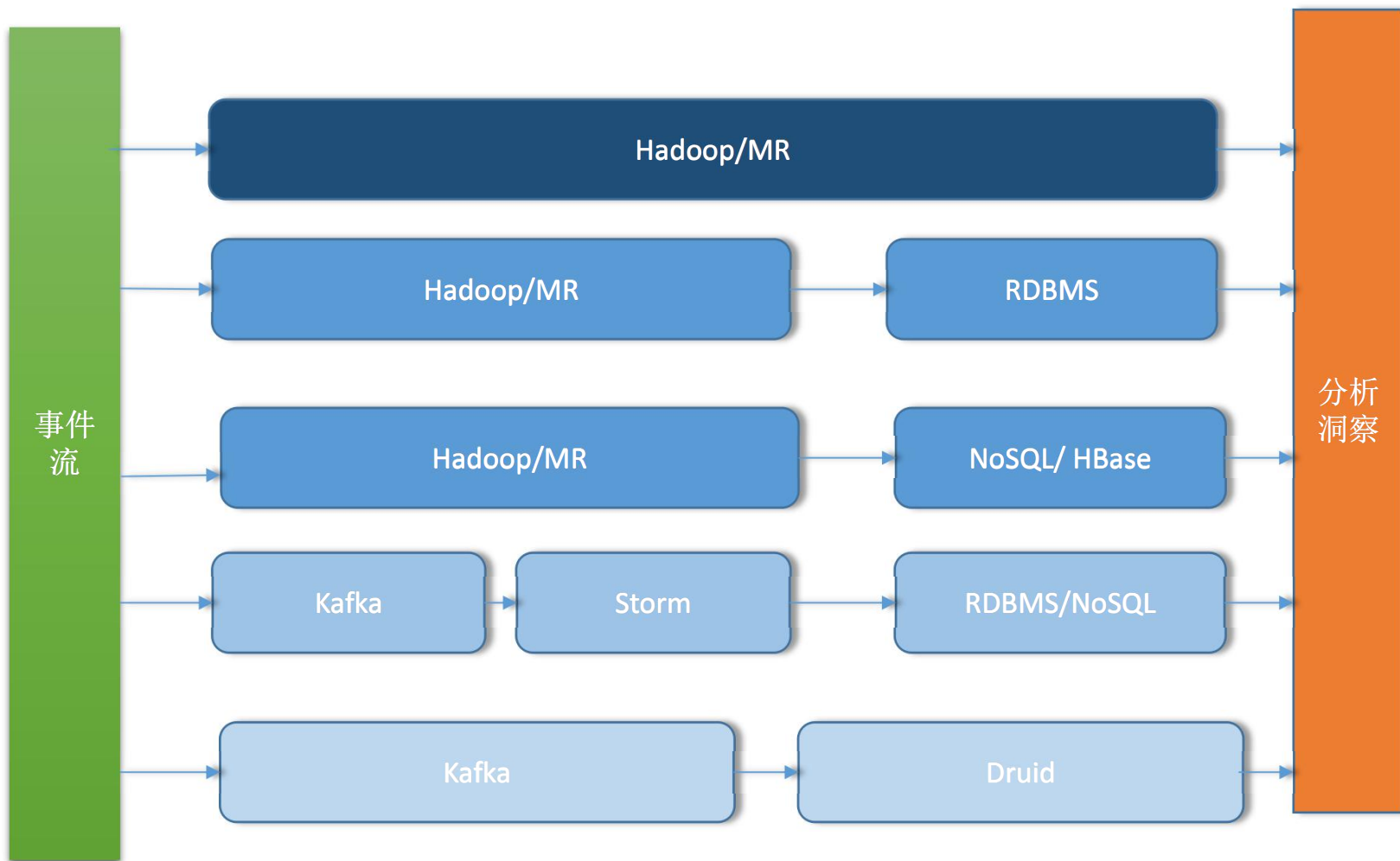
# 小米的大数据技术框架



—— 小米积极参与的项目，并且有Committee



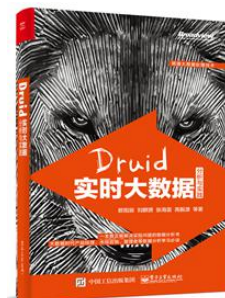
# 小米广告大数据分析架构的演化



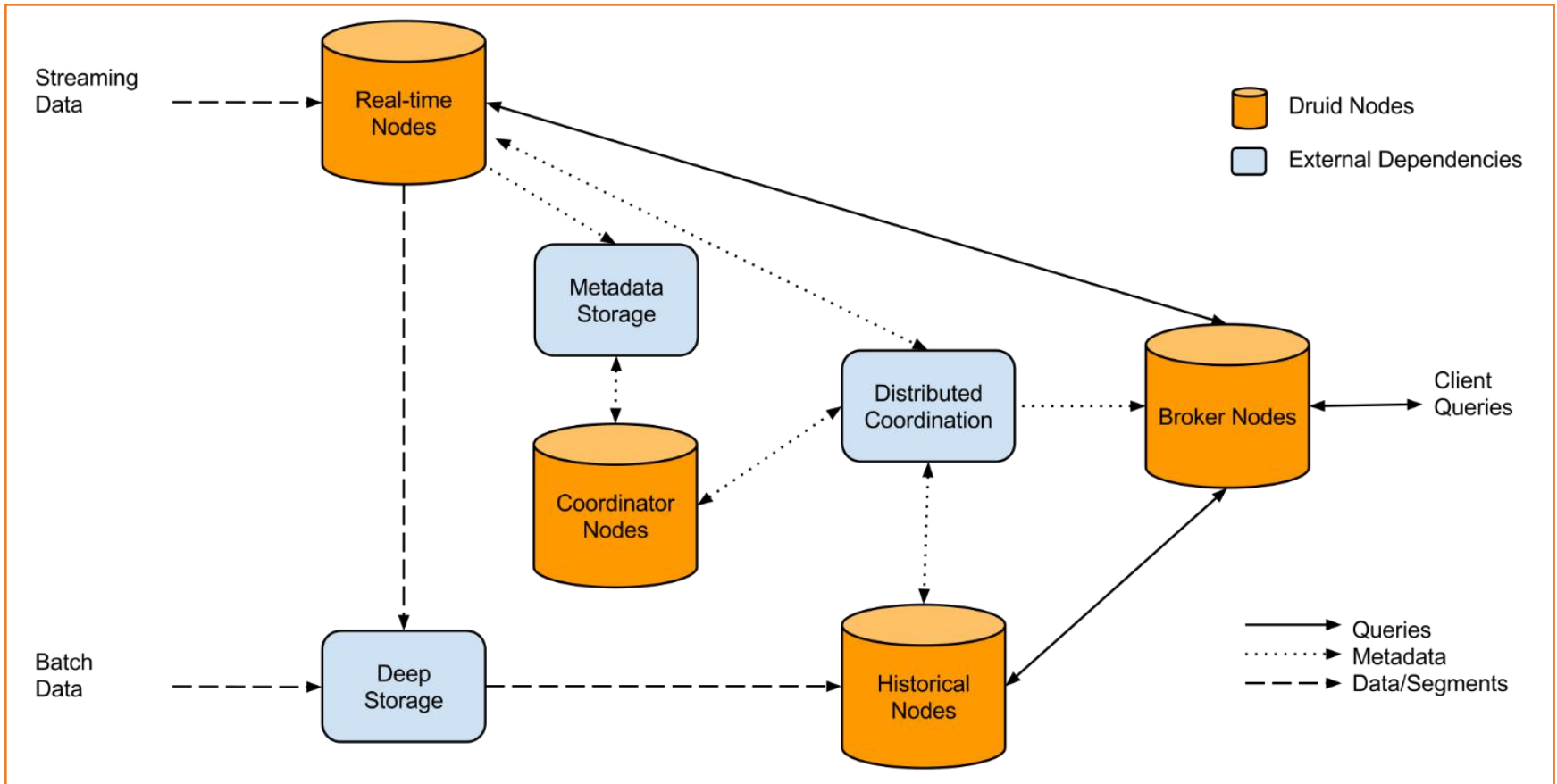
# 几种实时大数据分析工具的比较

|         | Druid                    | Pinot    | KYLIN     |
|---------|--------------------------|----------|-----------|
| 使用场景    | 实时处理分析                   | 实时处理分析   | OLAP分析引擎  |
| 开发语言    | JAVA                     | JAVA     | JAVA      |
| 接口协议    | JSON                     | JSON     | OLAP/JDBC |
| 发布时间    | 2013                     | 2015     | 2015      |
| Sponsor | Imply.io/<br>MetaMarkets | LinkedIn | eBay      |
| 技术      | 实时聚合                     | 实时聚合     | 预处理，Cache |

《Druid实时大数据分析原理与实践》即将出版

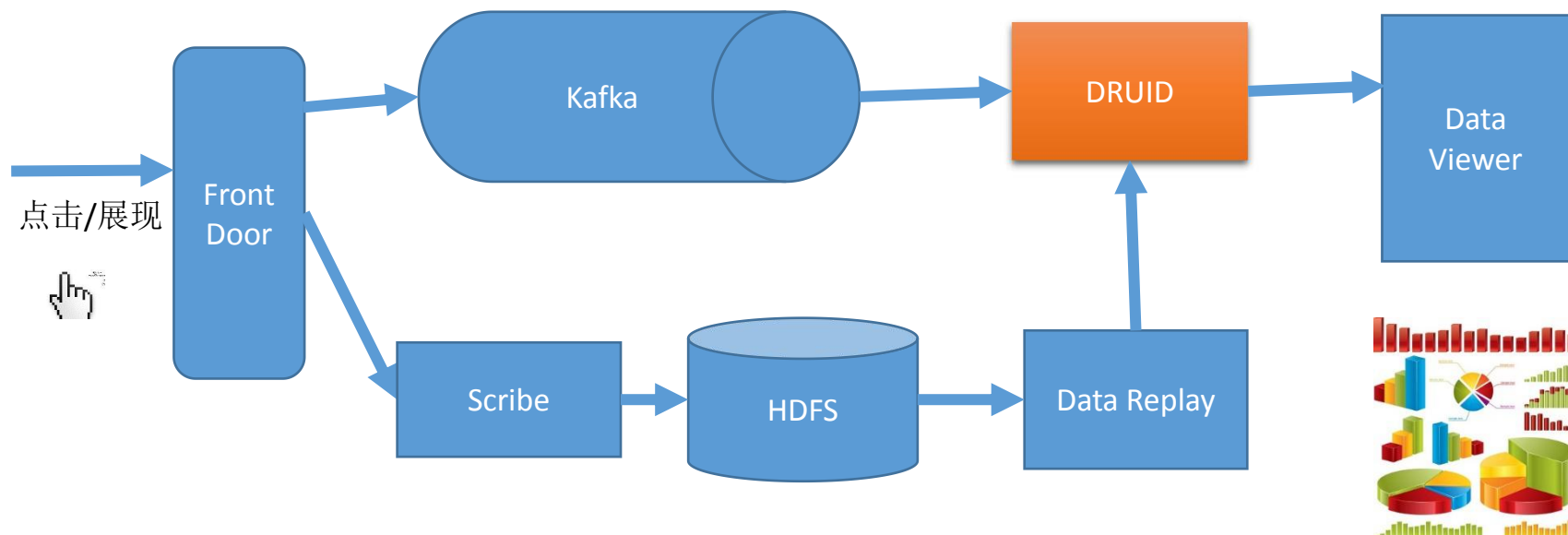


# DRUID Architecture :



<http://druid.io>

# DRUID使用场景：广告实时统计分析架构图 (非计费部分)



实时:秒级别更新

# MAX 流量智能分配

- 第一阶段：
  - 方法：初期流量分配给全部Demand
  - 问题：很多DSP的响应时间有问题，浪费带宽
- 第二阶段：
  - 业务模式PDB, Prefer Deal, Private Auction, Public Auction.
  - 问题：流量控制和分配解耦，流量服从业务/订单。
- 第三阶段：
  - 智能流量分配，优化收入，性能体验
    - 基于历史数据学习，智能判断，逻辑回归(LR)

架构的感悟！

# 什么是架构？

- 所有架构都是设计，不是每个设计都是架构。  
架构代表着发展一个系统的重要设计决策，这个重要性是通过变化引入成本来衡量的(Grady Booch, 06)
- “一切圣贤，皆以无为法而有差别”（金刚经）
- 架构是学习和演化，不是蓝图(Chen)



# 架构演化：Stack Overflow-Scale Up



(stackexchange.com/performance)



Alexa Traffic Ranks  
How is this site ranked relative to other sites?



Global Rank

57

Rank in India

14

1.这个软件架构是很“烂”(Boring)的

2.保持一个很“烂”的架构是非常有趣的



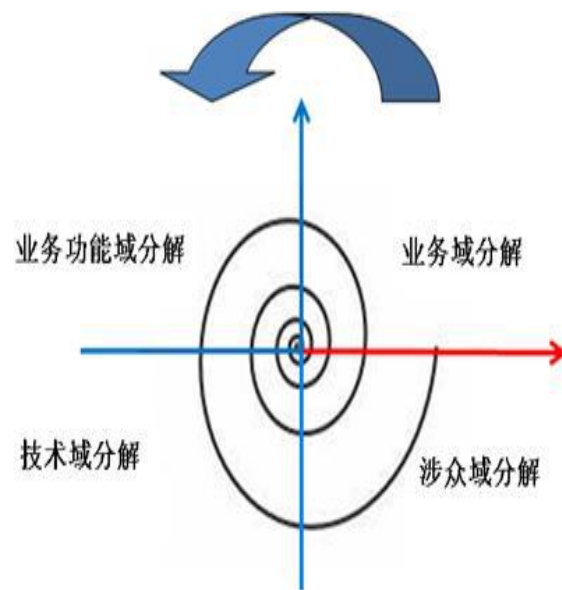
# 我的解耦4大基本原则和技术

## 原则

- 先业务，后技术；先逻辑，后物理
- 奥卡姆剃刀：如无必须，勿增实体
- 正交性：分解出模块无职责的重复
- 稳定性原则：稳定和易变的分解

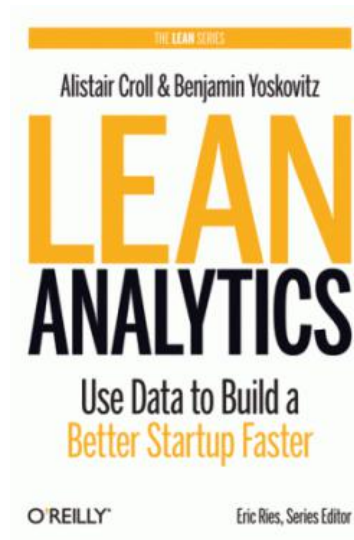
## 技术

- 接口
- 消息队列
- 模块化，服务化
- 异步化



# 架构师的OKR或KPIs

- 关键指标 (One Metric That Matters)
- 向业务负责
- 帮助团队获得满足感
- 随时回答团队的问题
- 保持谦逊和诚实



OKR  
Objective, Key Results





道阻且长，行则将至

谢谢！