

蚂蜂窝大数据平台架构及Druid引擎实战

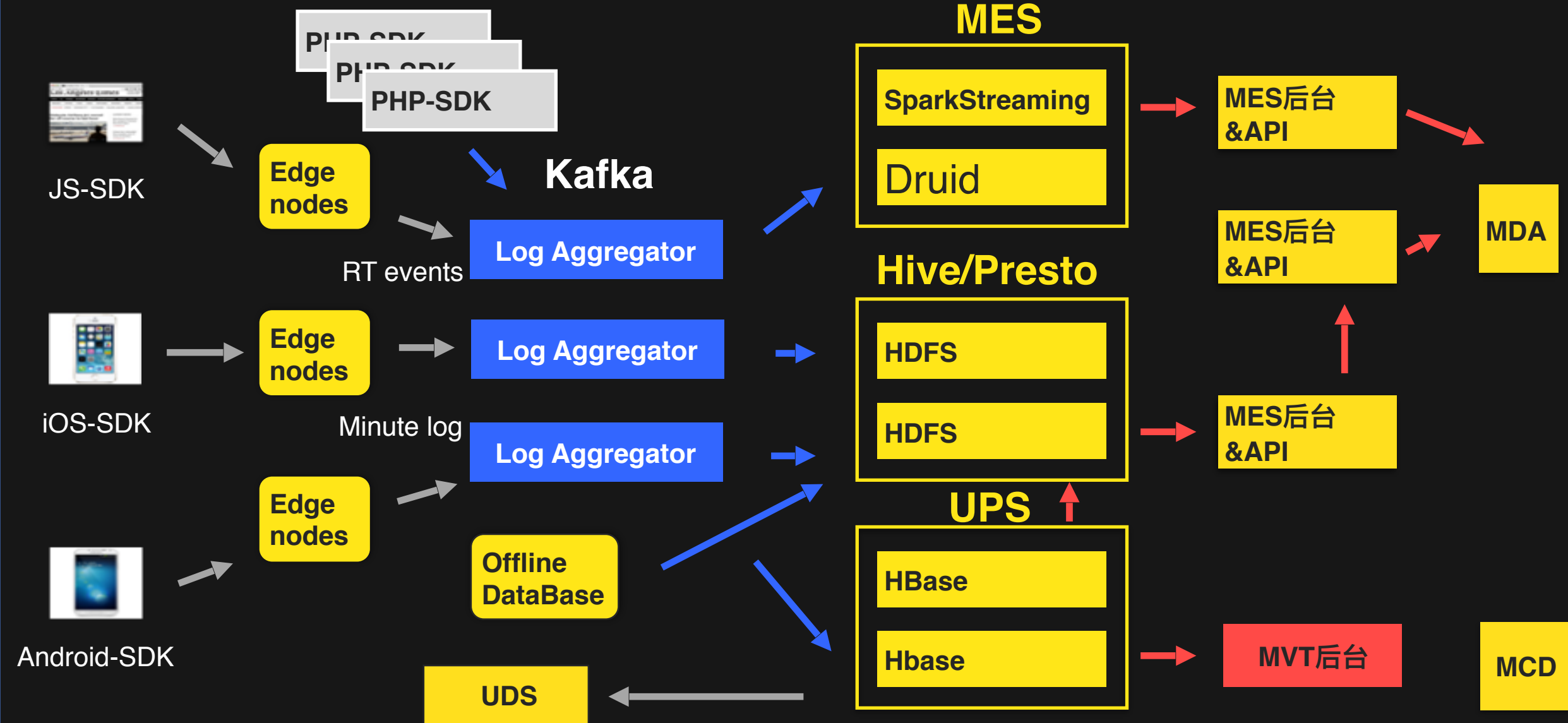
2017.08.19

汪木铃

自我介绍:

曾就职于360，14年加入蚂蜂窝，目前负责
蚂蜂窝大数据平台

数据流





K-V型事件模型

basic

attr

```
{  
  'dt': '20170804'  
  'app_code': 'xxx',  
  'event_code': 'xxx',  
  'event_guid': 'xxxx',  
  
  'attr': {  
    '自定义属性1': 'xxxx',  
    '自定义属性2': 'xxxx',  
    '自定义属性3': 'xxxx'  
  }  
}
```

实时数据探索

- dashboard
- server_event
- mobile_event
- page_event

频次统计

显示代码 显示名称 事件配置

前两天 前一天 今天 2017-08-1

当前规则: default

basic.app_code=default -> basic.event_code=h5ToApp -> attr.os -> attr.type -> attr.closeTime

当日汇总

生成MQL

系统的事件数	2200000	加入MDA	独立ip	2200000	加入MDA
系统的不同取值数	5	加入MDA	独立uuid	2200000	加入MDA
			独立uid	2200000	加入MDA

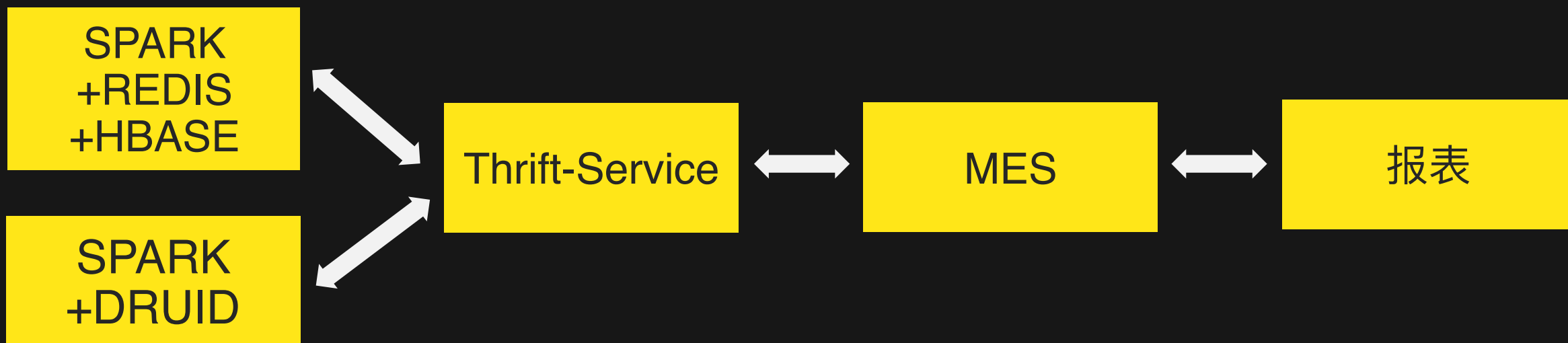
当日排行

生成MQL 下载

搜索

序号	系统	事件数	占比	独立ip	独立uuid	独立uid	
1	android	2200000	100.00%	2200000	2200000	2200000	历史趋势 加入MDA
2	ios10	0	0.00%	0	0	0	历史趋势 加入MDA

0 生成MDA报表 清空数据源



引擎	UV	虚拟key	一对多
spark+reids+hbase	精确	支持	支持
spark+druid	误差(2%左右) datasketch	不支持	不支持
	解决UV带来的计算压力		

MES第一代:Spark+Redis+Hbase实时数据探 索

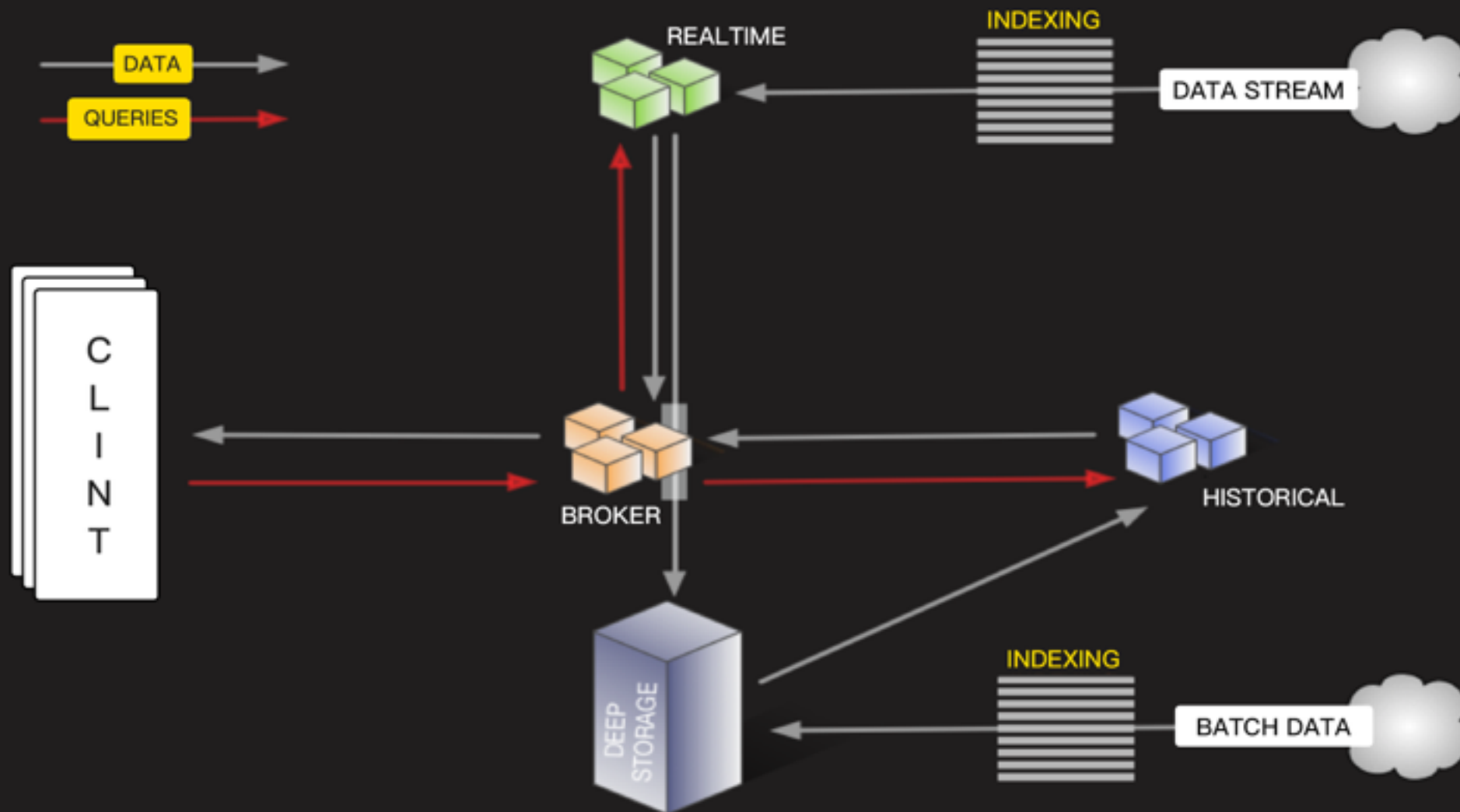
流量高峰期处理延迟

维度交叉分析，不灵活

消耗资源大

系统故障,重算慢

MES第二代:Druid实时数据探索



3 nodes

overlord

coordinator

3 nodes

middleManager

8 nodes

2-Hot-historical

6-Cold-historical

3 nodes

1-hot-broker

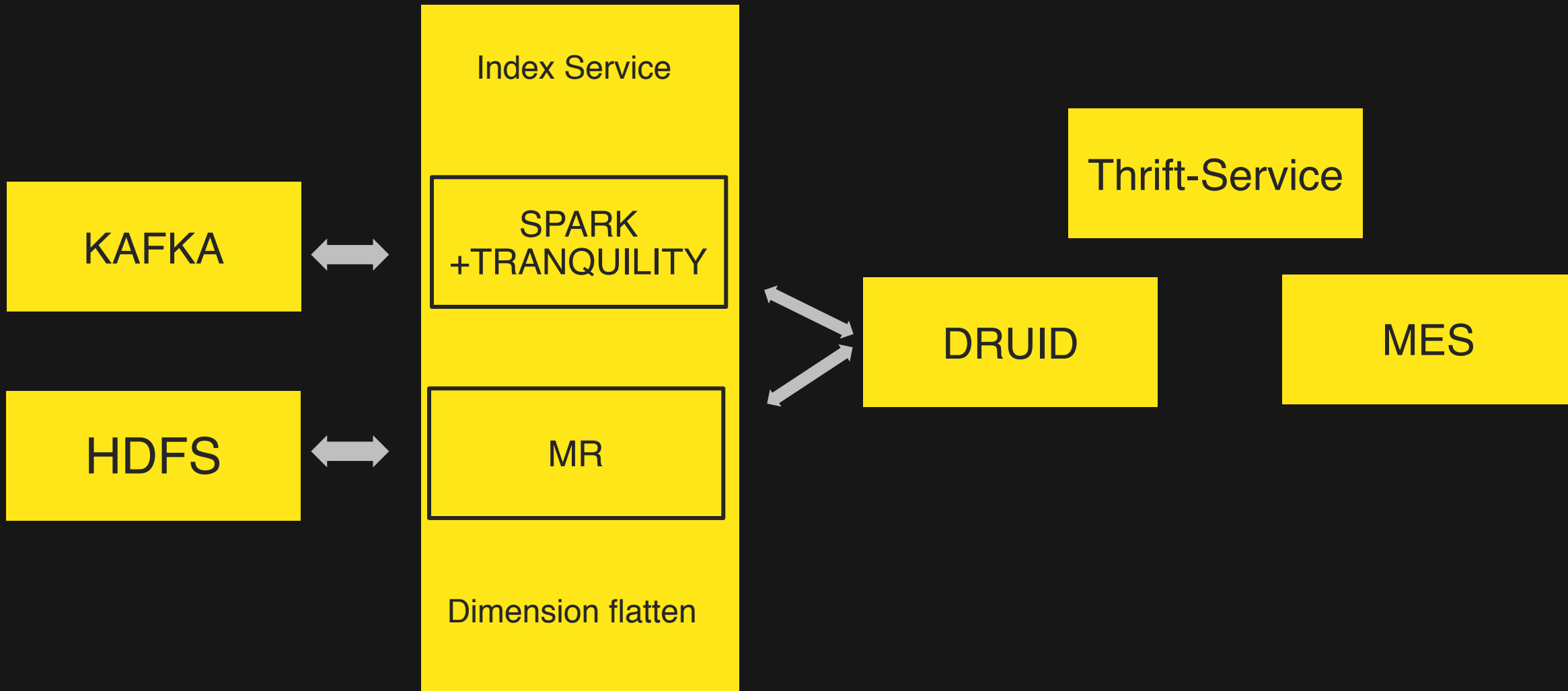
1-cold-broker

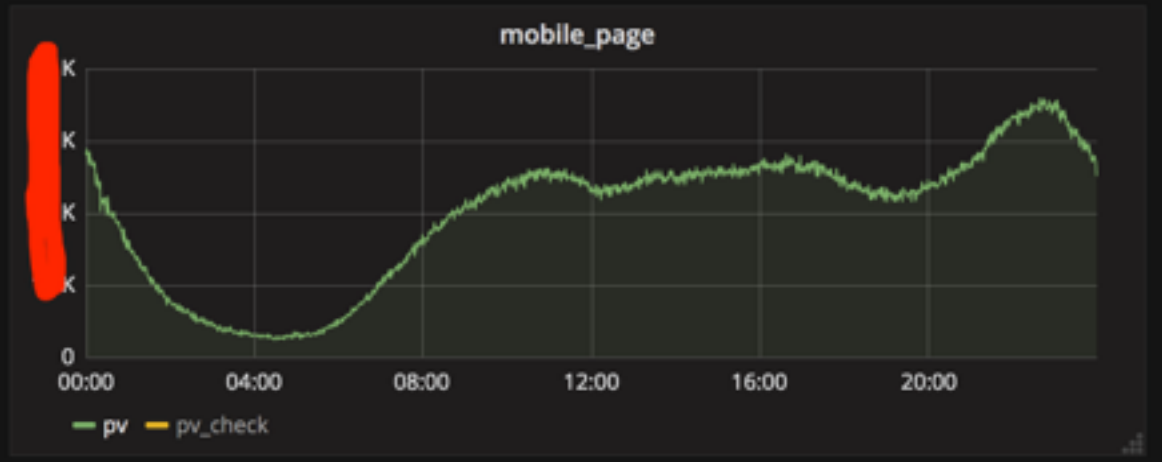
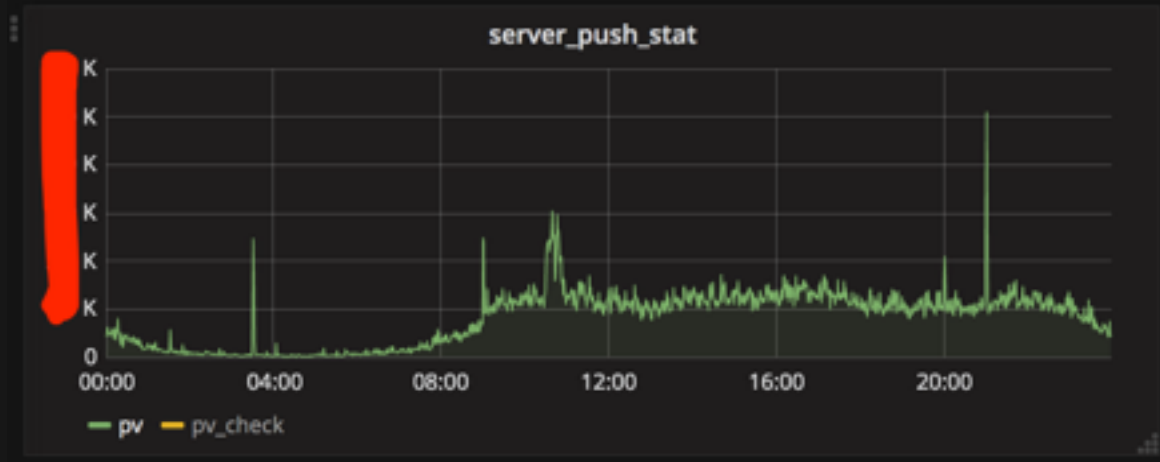
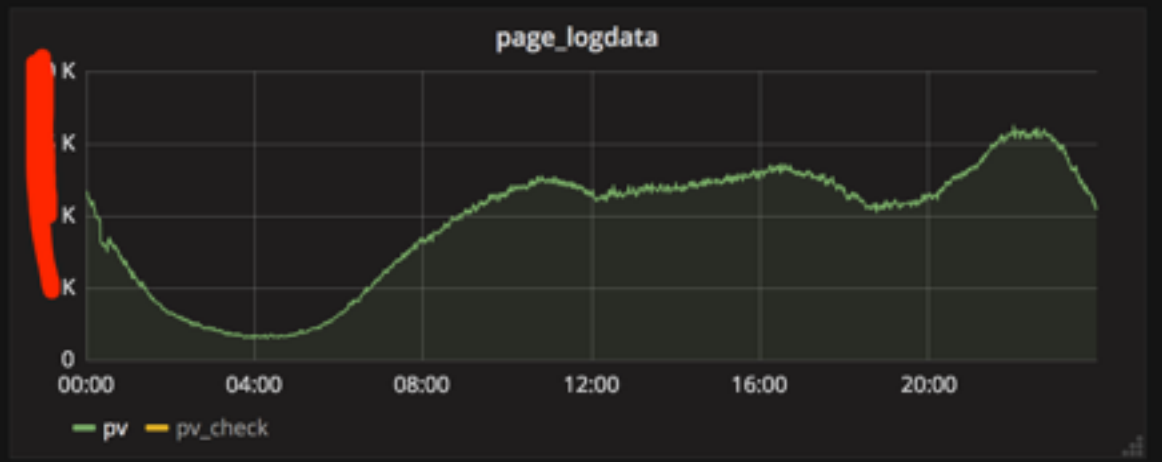
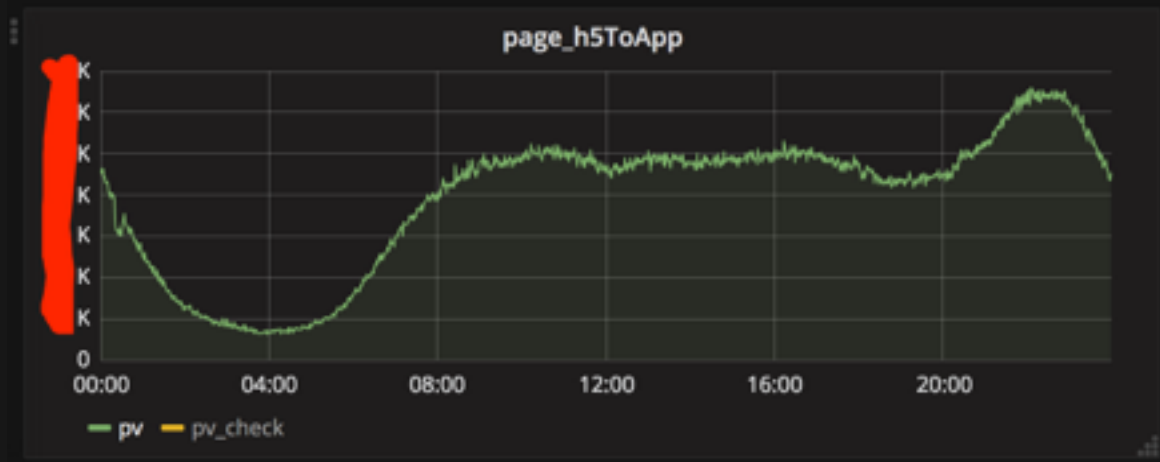
1-router-broker

日志源	维度	指标	保留周期	查询耗时
server_event	700+	5	15天	分钟级别
mobile_event	700+	5	15天	分钟级别
page_event	300+	5	15天	分级级别

日志源	维度	指标	保留周期	查询耗时
logdata	45	5	永久	秒级别
mobile_page	80	5	永久	秒级别
Page_h5ToApp	28	5	永久	秒级别
Push_stat	23	7	永久	秒级别










粒度	15天内	历史数据	
segment	HOUR	HOUR/DAY	
query	MINUTE	HOUR	





+ ADD ROW

DATA CUBES

 Page Event No description	 Server Event No description	 Monitor Event No description
 Mysql Binlog No description	 Druid Cluster Monitor No description	 Logdata No description
 Server Push Stat No description	 Page H5 To App No description	 Mobile Page No description

Druid集群注意事项

维度不要太多且维度大小不要太大

维度不能含逗号,中文等

合理配置historical磁盘 内存比例

middleManager尽可能单独物理运行

合理设置segment大小(300-700M)

优先升级Druid版本

谢谢IQ&A

2017.08.19