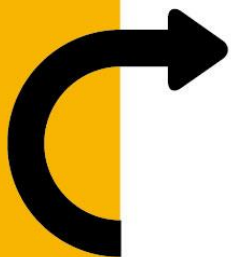




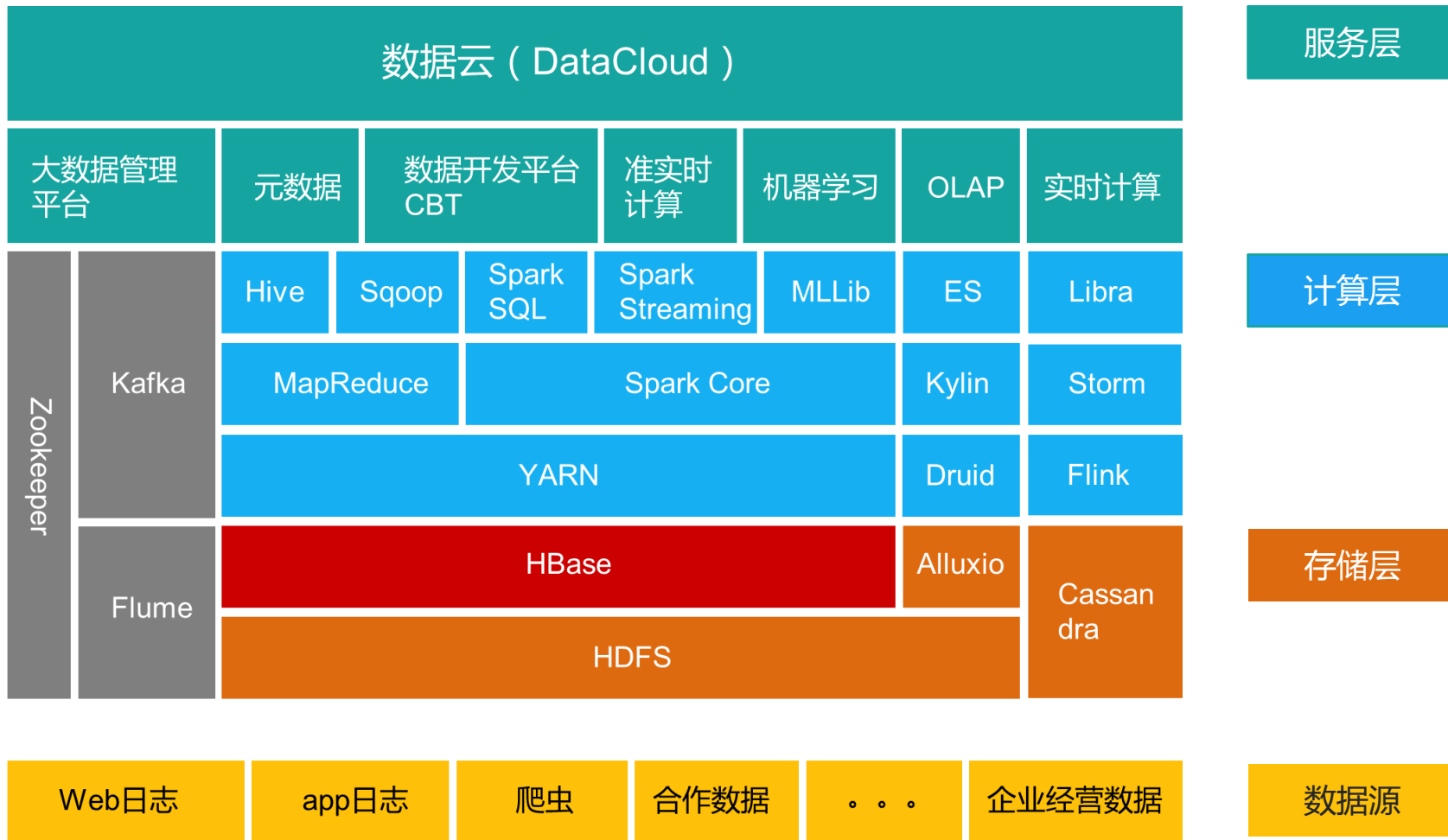
# HBase在苏宁

赵云@苏宁云商-大数据平台



- **苏宁HBase总览**
- **HBase监控**
- **HBase全局流控**
- **HBase多集群高可用**
- **HBase服务化**

# 苏宁大数据平台总览



# HBase在苏宁的现状

## 集群规模

14 线上集群

300 物理服务器

## 业务规模

100+ 业务系统

1000+ Table

5万+ Region

## QPS

日均 400万

峰值 2000万

# 场景-苏宁经营决策平台



实时指标

商品 买家 渠道

销售额: 104,657,265.07 0.00% ↓

退货率: 6.30% 0.00% ↓

付款金额: 104,657,265.07 0.00% ↓

品类	销售量	环比
百货	4,107	0.00% ↓
通讯	3,105	0.00% ↓
家居	1,000	0.00% ↓
小家电	1,104	0.00% ↓
电脑	2,000	0.00% ↓
冰洗	1,000	0.00% ↓
食品酒水	1,000	0.00% ↓
粮油休食	1,000	0.00% ↓
母婴	1,000	0.00% ↓
厨卫	1,000	0.00% ↓
进口保健生鲜	1,000	0.00% ↓
黑电	1,000	0.00% ↓
数码	1,000	0.00% ↓
美妆	1,000	0.00% ↓
空调	1,000	0.00% ↓
医药馆事业部	1,000	0.00% ↓
文化制品	1,000	0.00% ↓
企业定制	1,000	0.00% ↓

全量商品  
数据

PV/UV  
数据

销售数据

商品分析 流量分析 大促直播

## 场景-苏宁经营决策平台

- 实时读写HBase
- 多维数据预计算



# 场景-电商情报分析系统

商品

目录

店铺

品牌

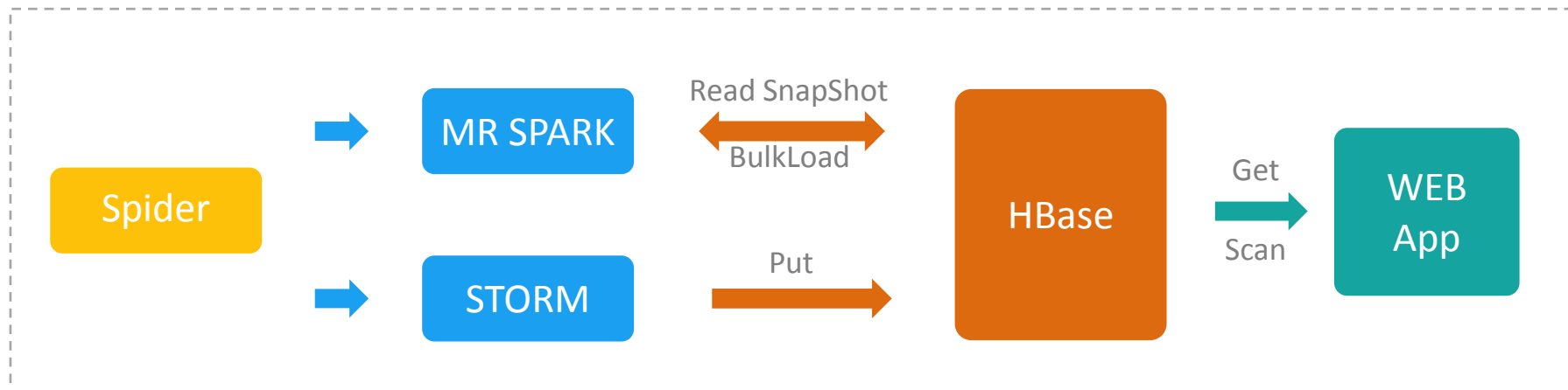
网站: 京东商城 自营平台: 自营 日期: 昨日 或选择其他日期: 点击选择日期

目录	促销商品数	促销占比	降价商品数	提价商品数	下架商品数	新上架商品数	缺货数
合计	5473	71.52%	5473	5727	500	8242	14495
冰洗	99	92.21%	99	121	4	1	4
厨卫	118	84.14%	118	180	6	4	3
黑电	102	87.27%	102	103	2	1	2
空调	65	85.97%	65	36	0	0	1
通讯	196	70.57%	196	102	23	0	1
电脑	920	63.65%	920	450	13	174	1
生活家电	401	90.9%	401	525	17	13	1
百货	1345	66.36%	1345	1792	288	7932	1
数码	112	56.64%	112	380	22	25	1
母婴	403	71.81%	403	480	40	71	1
超市	1712	88.73%	1712	1558	85	21	1

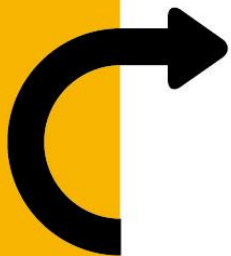
商品分析 竞品分析 差异分析 促销分析

## 场景-电商情报分析系统

- 离线读写HBase
- 实时写HBase





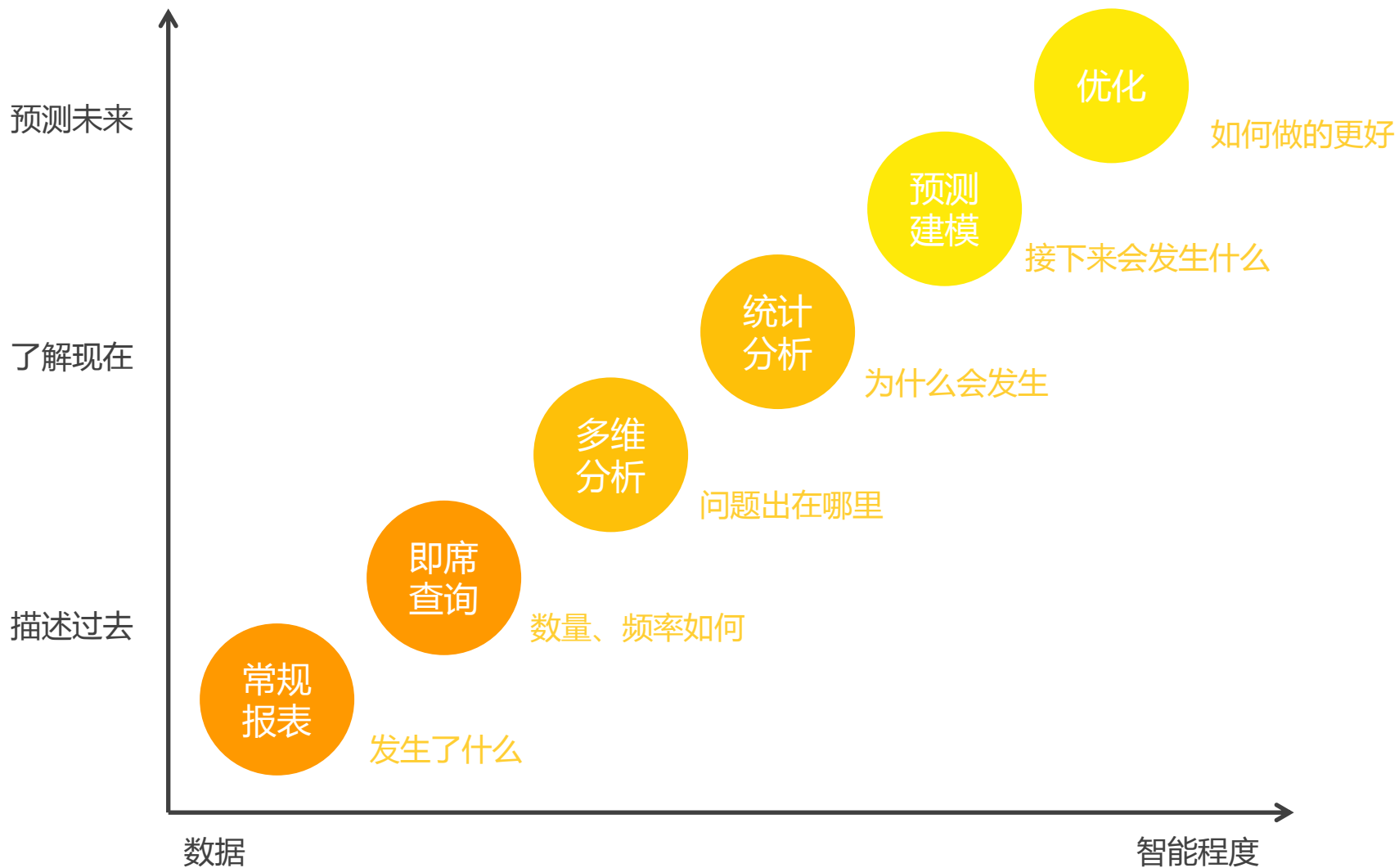


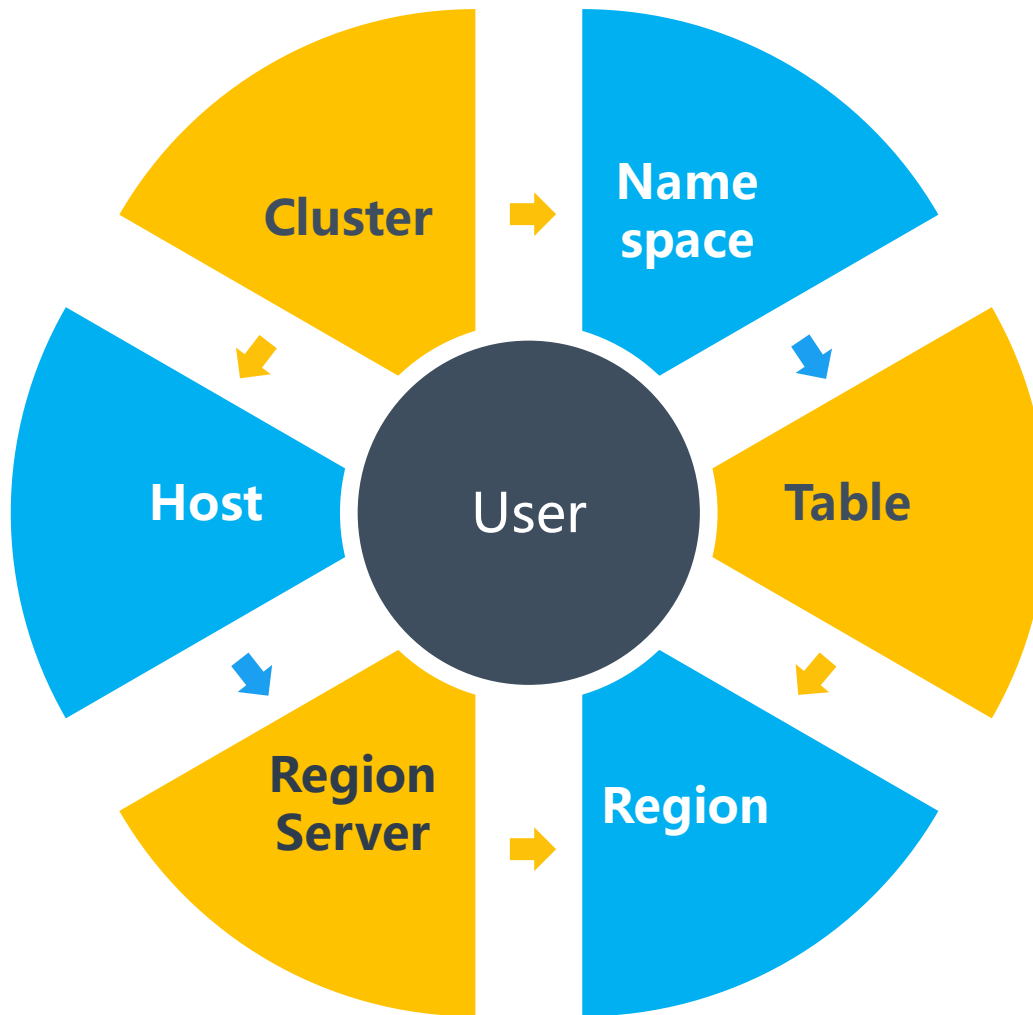
- 苏宁HBase总览
- HBase监控
- HBase全局流控
- HBase多集群高可用
- HBase服务化

## 监控-遇到的问题

- 某个表的QPS的趋势
- 集群的详细运行状况
- ...
- 各Region请求分布是否均匀
- 不能比用户提前发现问题

# 监控-我们需要什么





# 监控-开始有什么

Master: `hbase-hbase01.cnsuning.com`

Region Servers

Base Stats | Memory | Requests | Storefiles | Compactions

ServerName	Start time	Version	Requests Per Second	Num. Regions
<code>hbase-hbase01.cnsuning.com,16020,1507537852877</code>	Mon Oct 09 16:30:52 CST 2017	1.2.6.5	9341	119
<code>hbase-hbase02.cnsuning.com,16020,1507537854001</code>	Mon Oct 09 16:30:54 CST 2017	1.2.6.5	7068	131
<code>hbase-hbase03.cnsuning.com,16020,1507537855076</code>	Mon Oct 09 16:30:55 CST 2017	1.2.6.5	11032	124

```

}, {
  "name" : "Hadoop:service=HBase,name=RegionServer,sub=Server",
  "modelerType" : "RegionServer,sub=Server",
  "tag.zookeeperQuorum" : "hbase-hbase01.cnsuning.com:2181,hbase-hbase02.cnsuning.com:2181,hbase-hbase03.cnsuning.com:2181",
  "tag.serverName" : "hbase-hbase01.cnsuning.com,16020,1507537852877",
  "tag.clusterId" : "e889980a-fba7-4156-8537-6761ced34c6a",
  "tag.Context" : "regionserver",
  "tag.Hostname" : "hbase-hbase01.cnsuning.com",
  "regionCount" : 119,
  "storeCount" : 119,
  "hlogFileCount" : 9,
  "hlogFileSize" : 1020818095,
  "storeFileCount" : 391,
  "memStoreSize" : 8854376,
  "storeFileSize" : 663256005720,
  "regionServerStartTime" : 1507537852877,

```

一开始我们只有这些

## 监控-通用解决方案

### 大数据管理平台



Apache Ambari

cloudera MANAGER

### 通用监控工具

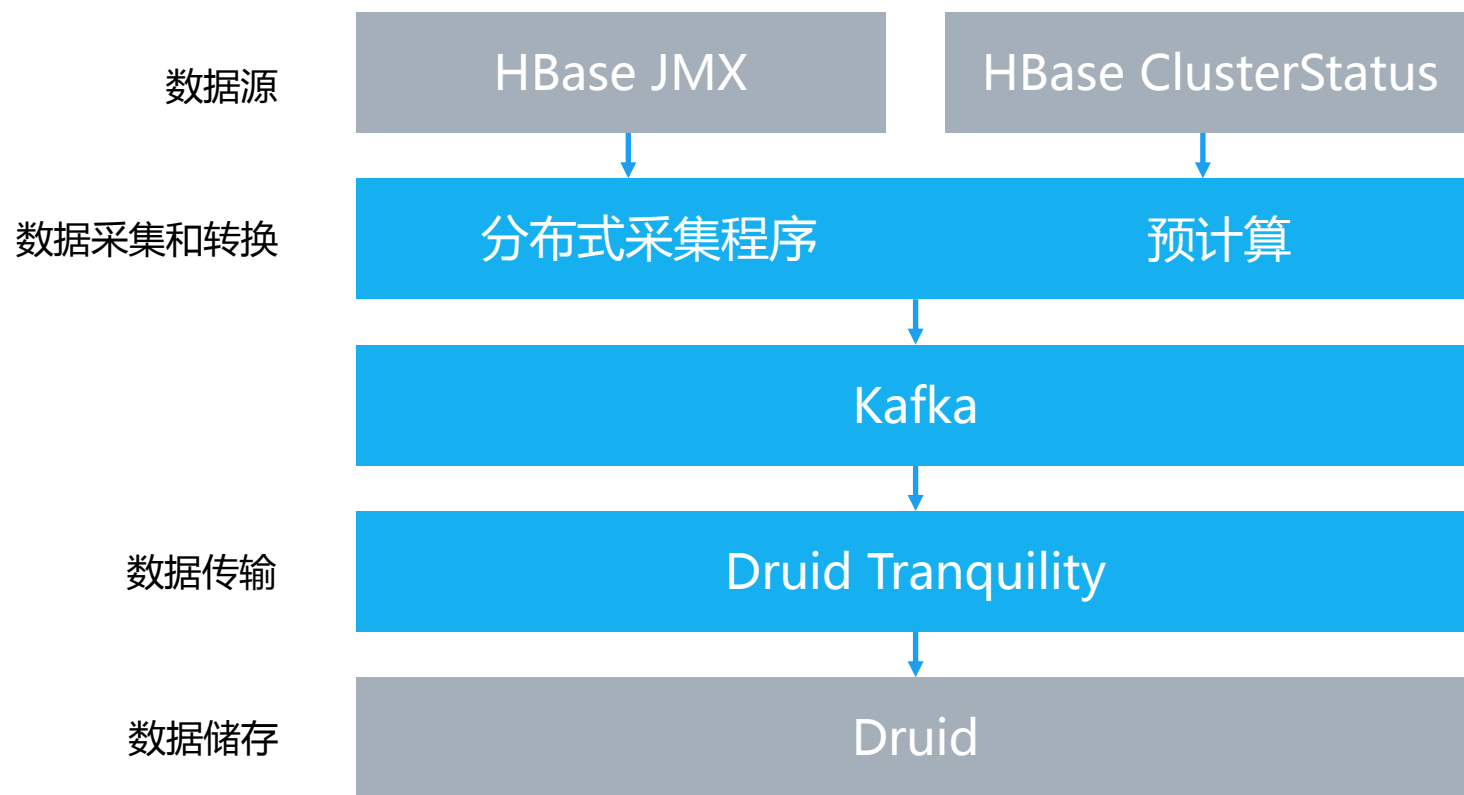


ZABBIX

Grafana



## 监控-苏宁解决方案



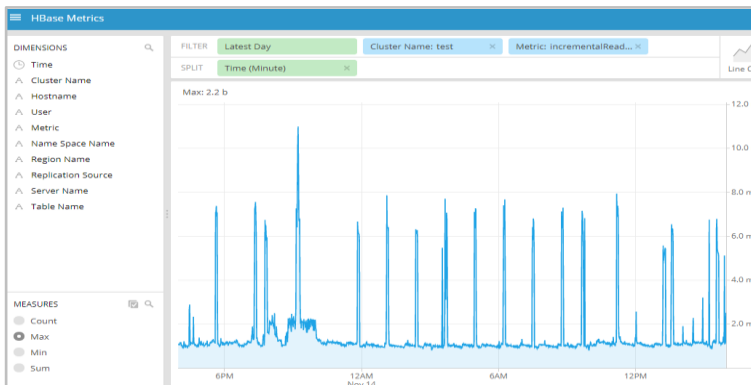
日均 60亿条 100G

# 监控-开源查询视图



Pivot

- 灵活
- 适用于管理员

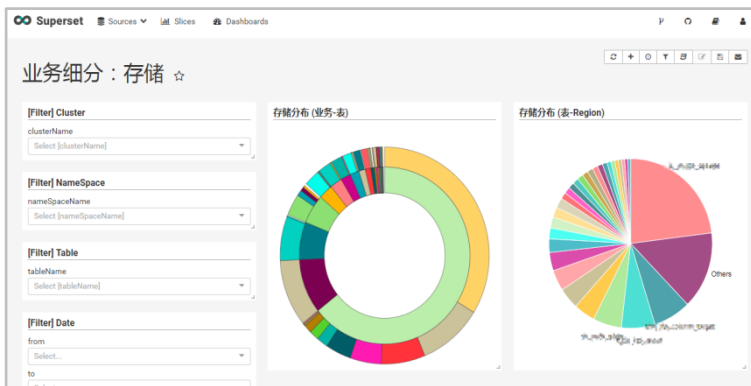


都没有一致的权限管理



Superset

- 预定义
- 适用于普通用户





# 监控-苏宁指标监控



# 监控-苏宁指标监控

Table图表   Table-Top图表   **Table指标**

选择集群名称:    选择Namespace:

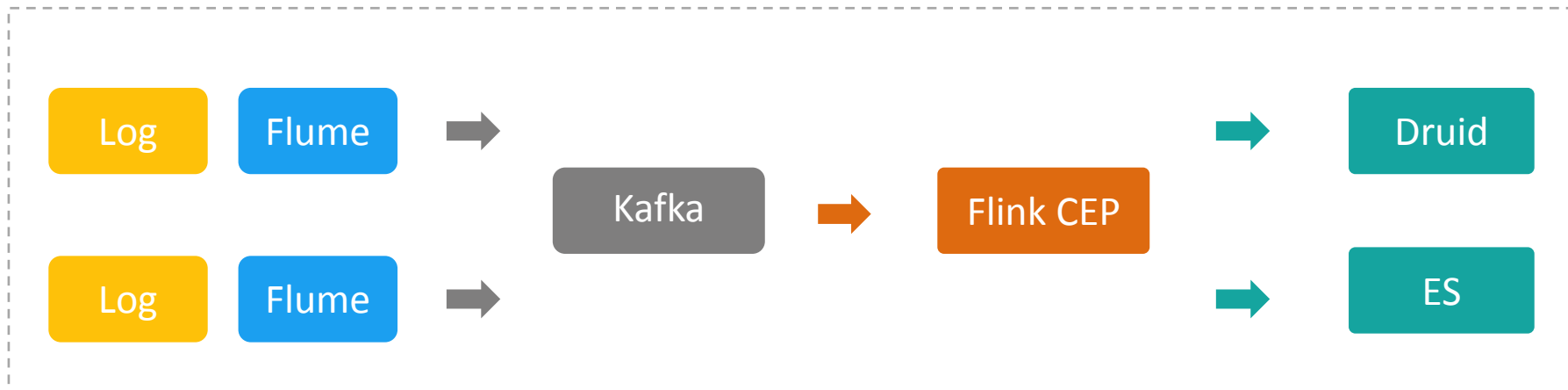
选择TableName:

RegionName	累计总请求数	Storefile(MB)	Memstore(MB)	总存储大小 (MB)	StartKey	EndKey
ns_171001001	14875207	3647	78	3725		020
ns_171001001_980	15183767	3785	6	3791	980	
ns_171001001_040	14761813	3686	26	3712	040	060
ns_171001001_060	14950286	3694	102	3796	060	080
ns_171001001_080	15044672	3752	4	3756	080	100
ns_171001001_100	14800234	3743	1	3744	100	120
ns_171001001_120	14524153	3653	27	3680	120	140
ns_171001001_140	14530104	3678	110	3788	140	160
ns_171001001_160	14681785	3679	61	3740	160	180
ns_171001001_180	15341406	3756	20	3776	180	200
ns_171001001_200	14442653	3694	10	3704	200	220

## 监控-日志监控

仍然有以下问题

- 分布式日志查找和分析
- 事务状态监测
- 异常日志分析
- ...



## 单行日志

- responseTooSlow/Large
- lease expired on region
- wal slow sync
- bad datanode
- Could not complete
- ...

## 多行日志

- Flush
- Compact
- Merge
- Split
- Snapshot
- ...





- 苏宁HBase总览
- HBase监控
- **HBase全局流控**
- HBase多集群高可用
- HBase服务化

## 全局流控-问题提出

- 多租户隔离
- 流量异常
- 非重点业务降级
- 开源社区只支持单个RS的流控
- ...

## 全局流控-实现难点

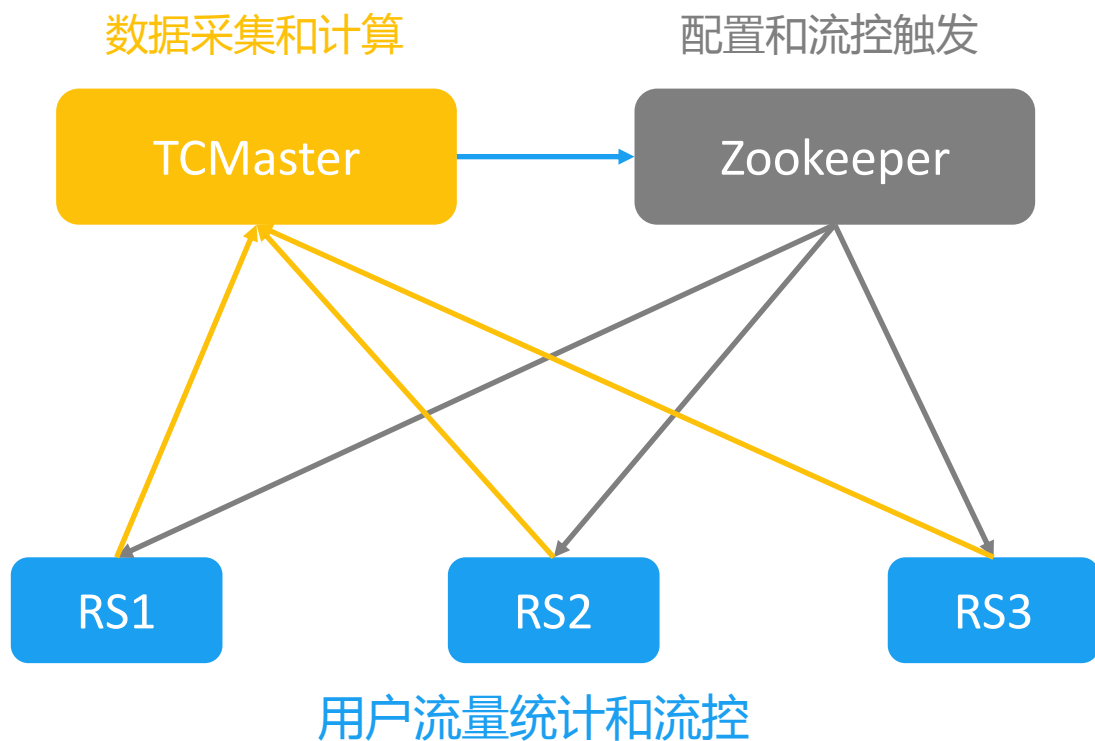
- 不能影响正常的处理效率
- 无法使用全局计数器/加锁方式
- 实时性和吞吐效率无法同时保证



## 全局流控-基本思路

- 牺牲实时性，保证吞吐效率
- 按表对用户进行流控
- 独立服务进行流控的管理

# 全局流控-总体方案

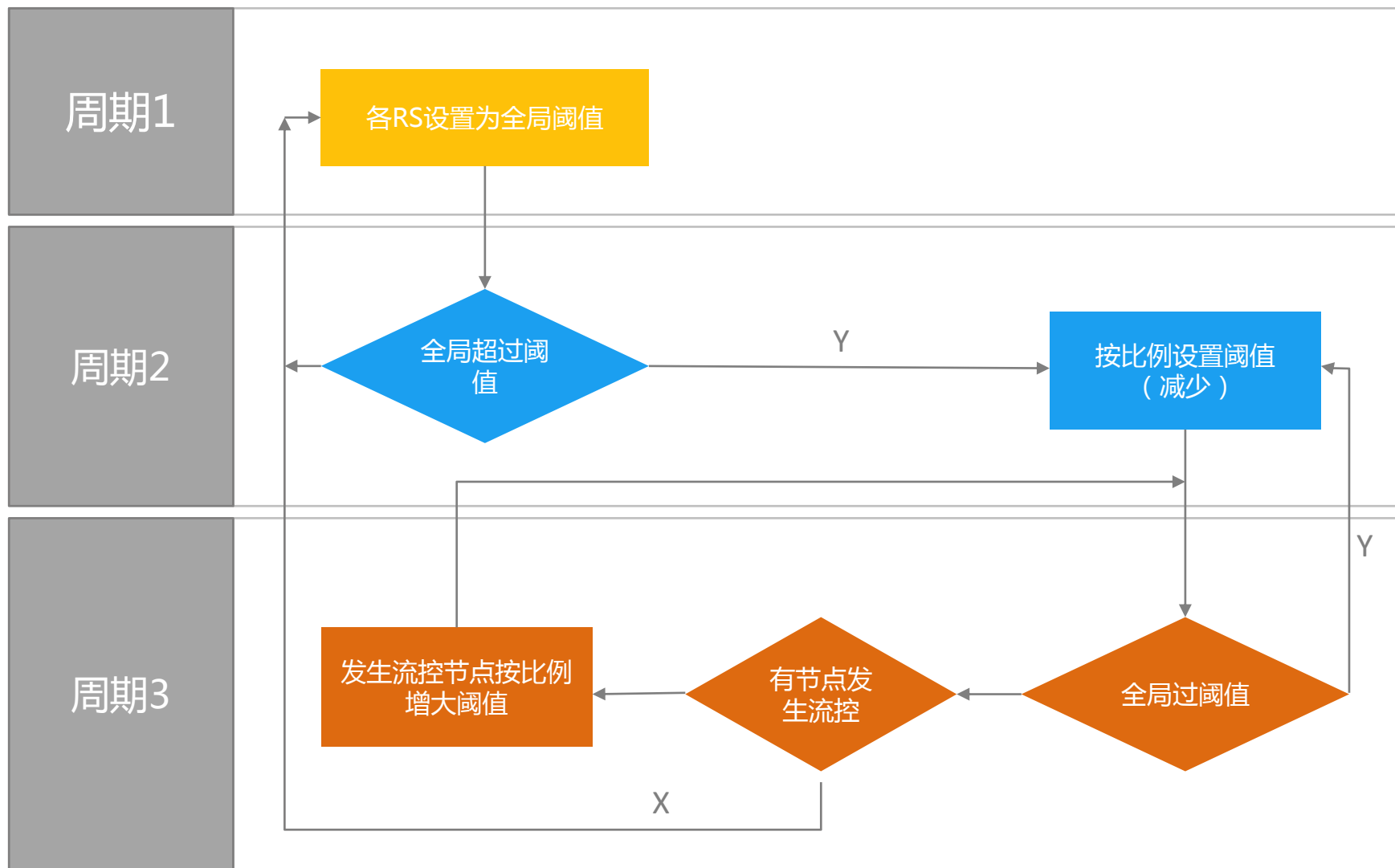


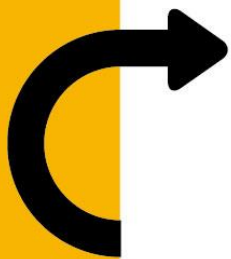
→ 周期性收集各RS的用户流量数据

→ Watch流控通知，针对下个周期采取限制

→ 流控通知

# 全局流控-流控逻辑





- 苏宁HBase总览
- HBase监控
- HBase全局流控
- HBase多集群高可用
- HBase服务化

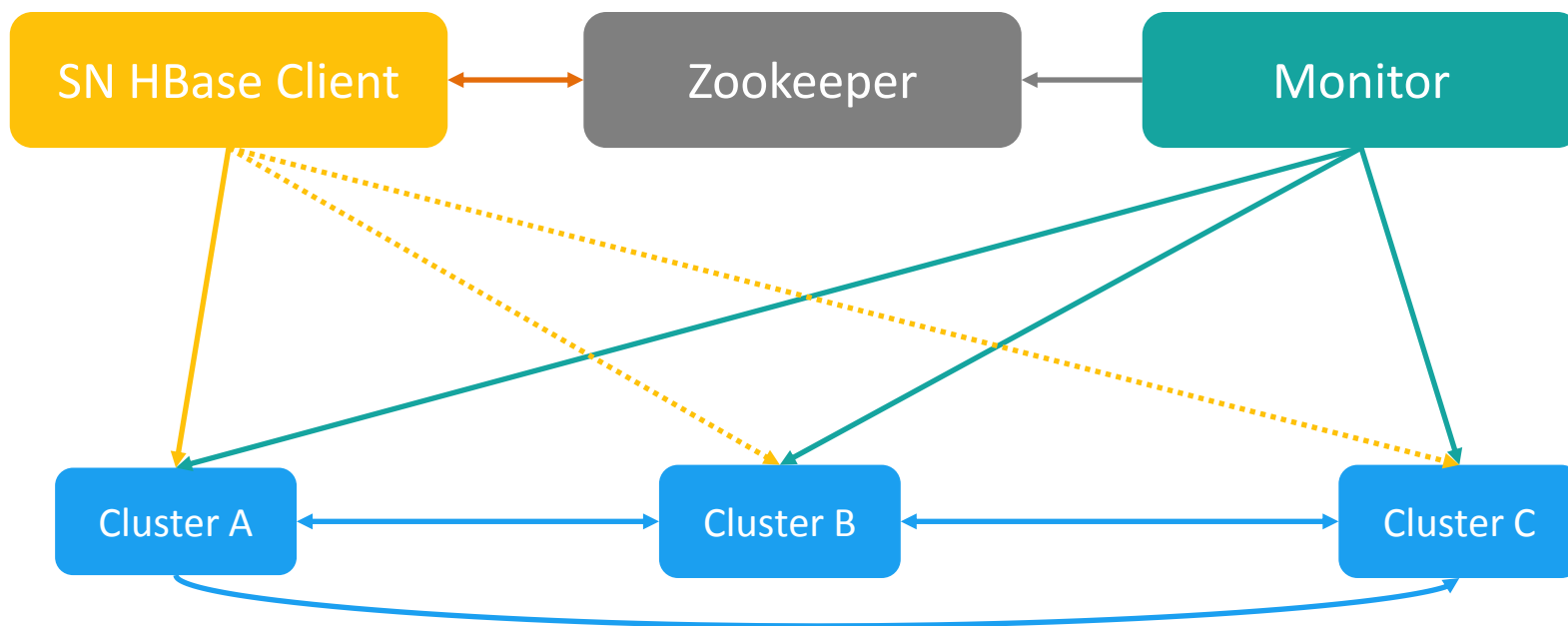
## 多集群高可用-问题提出

- 7×24在线服务
- 防止单集群故障
- 版本热升级
- 上帝说要有
- ...

## 多集群高可用-基本思路

- 多集群间复制
- 切换通知
- ...
- 集群状态监控
- 客户端封装

# 多集群高可用-总体方案



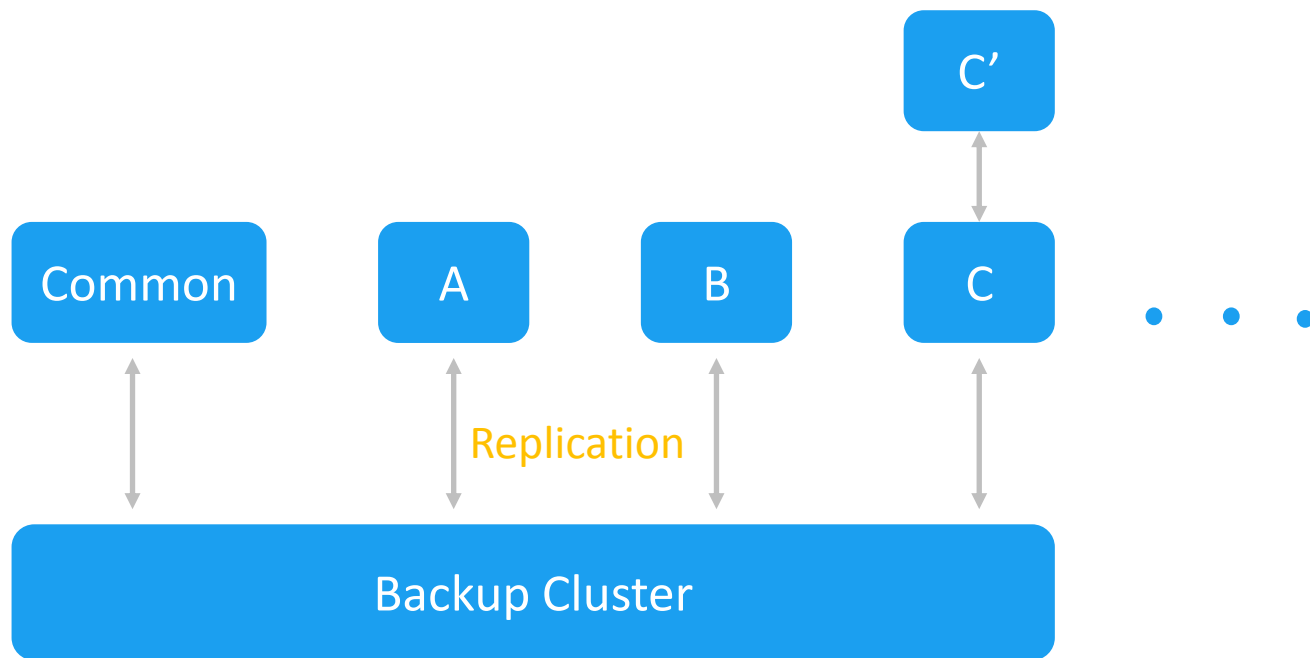
- ↔ 集群间Replication
- 集群状态更新
- 客户端请求
- 集群监控
- ↔ 集群获取和watch
- 客户端读 ( hedged )



- 苏宁HBase总览
- HBase监控
- HBase全局流控
- HBase多集群高可用
- HBase服务化



# 服务化-集群



## 服务化-能力

- 集群
- 资源
- 元数据
- 流程
- Replication
- Snapshot

管理能力

工具能力

HBase  
服务化

监控能力

- 数据迁移
- 自动合并
- 批量删除
- 重新分区
- 数据本地化

- 服务端
- 客户端
- 告警
- 用户日报

## 未来规划

- 容器化部署
- 更多的社区贡献
- 自动扩缩容
- 拥抱HBase 2.0
- 实时Replication

# Thanks!

