

2017 china Hadoop summit

大数据时代存储-云HBase实践与探索

主讲人：封神

提纲

1. HBase大数据中价值
2. HBase应用案例
3. 云HBase建设
4. 云HBase未来展望

数据库技术发展

HBase支持实时更新、增量导入、多维删除、随机查询、范围查询、高伸缩、高可用、高可靠、高性能、高适应在线分布式NOSQL数据库。

关系型数据库

MySQL

解决中小数据库存储需求！

分库分表

分库分表

解决一定容量需求，实现复杂，业务需感知！

分布式存储

HBase

分布式存储，支持海量存储，高百万级并发！
不支持事物等数据库传统需求！

PB

数据量

GB

传统关系型数据库面对的问题

成本

一般需要高端存储，成本较高！



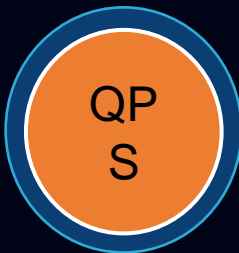
容量

无法满足TB、PB级别的数据存储需求！



QPS

无法满足较高的并发要求，性能不能横向扩展！



分析

计算资源有限，不能满足分析类需求！

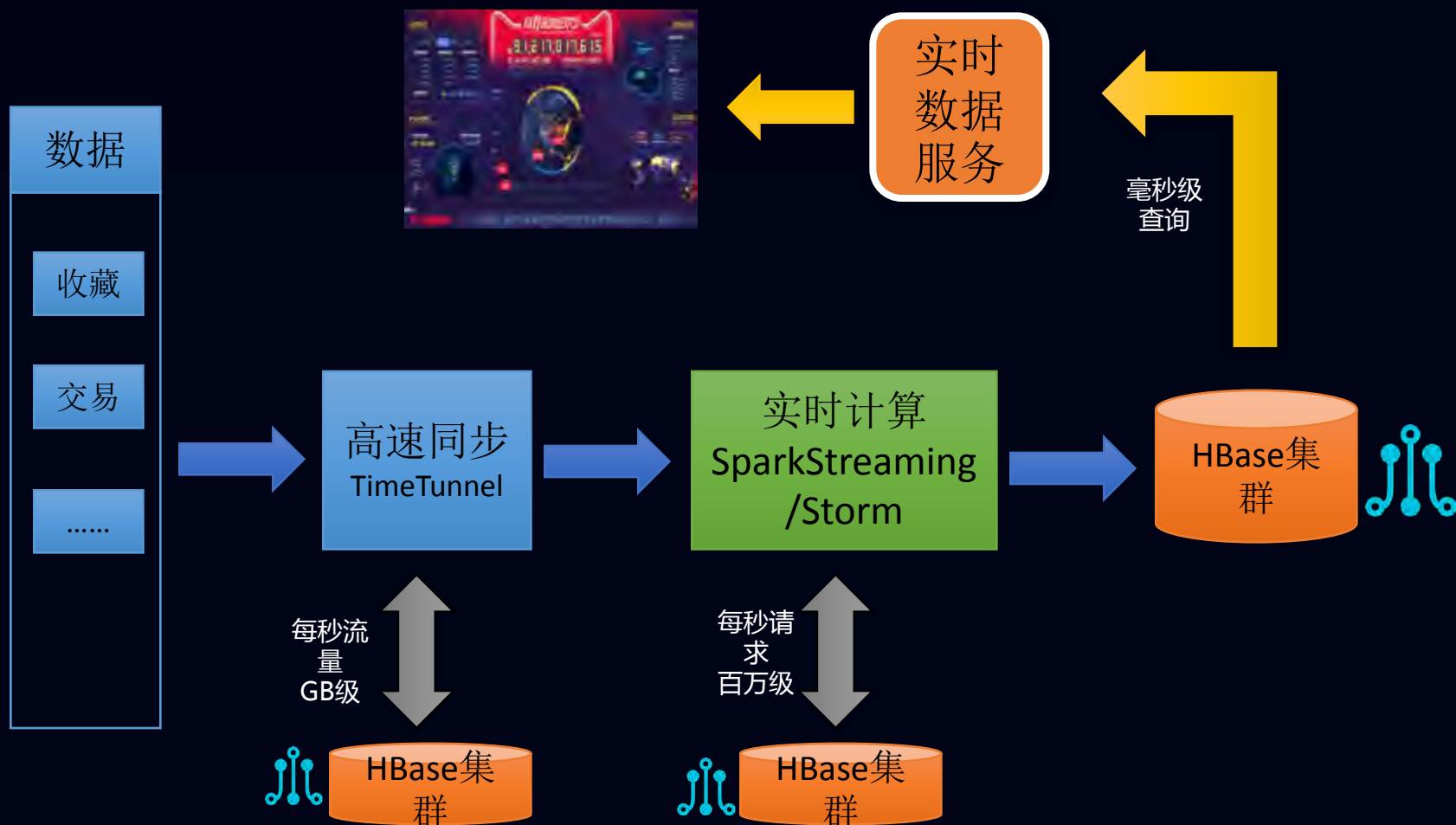
HBase应对办法



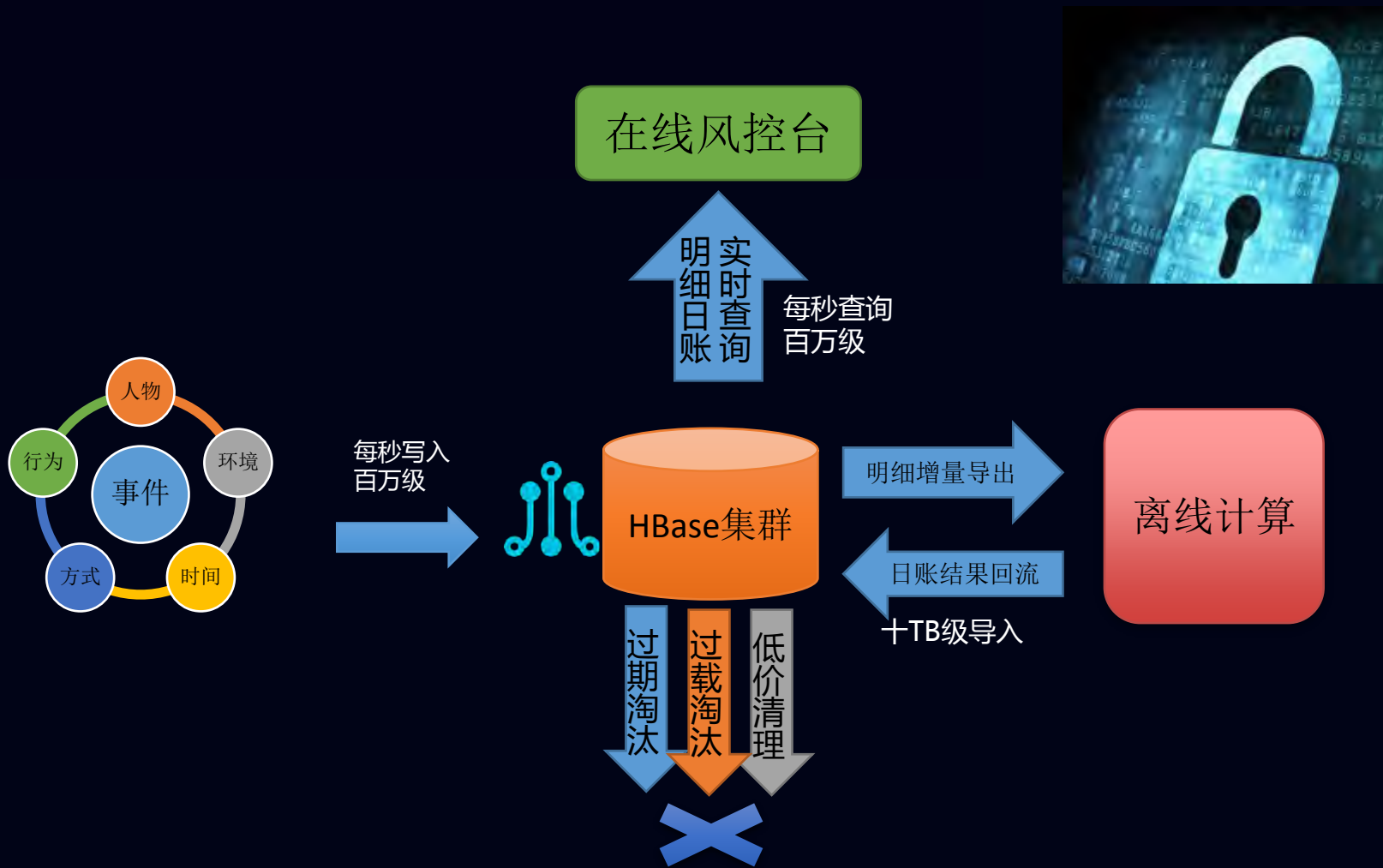
其它问题

业务场景	关系型	HBase
历史数据自动过期，比如删除2年前的日志	业务逻辑	TTL
地址维护：每个客户10个地址	业务逻辑	多版本
业务添加列	DBA添加	动态列
离线数据导入	API写	bulkload
数据分析	无	Spark、Hadoop
冷热数据	无	异构存储

应用案例-双十一大屏

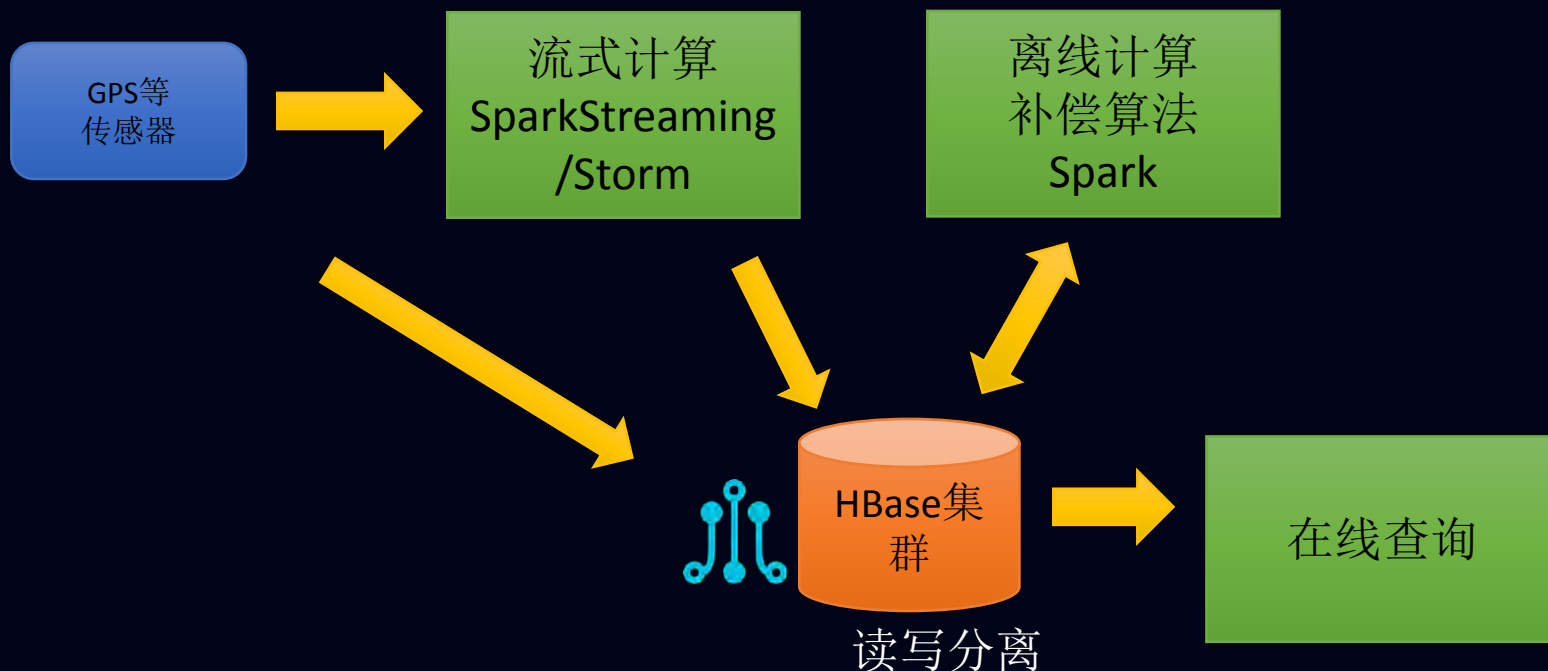


应用案例-蚂蚁风控

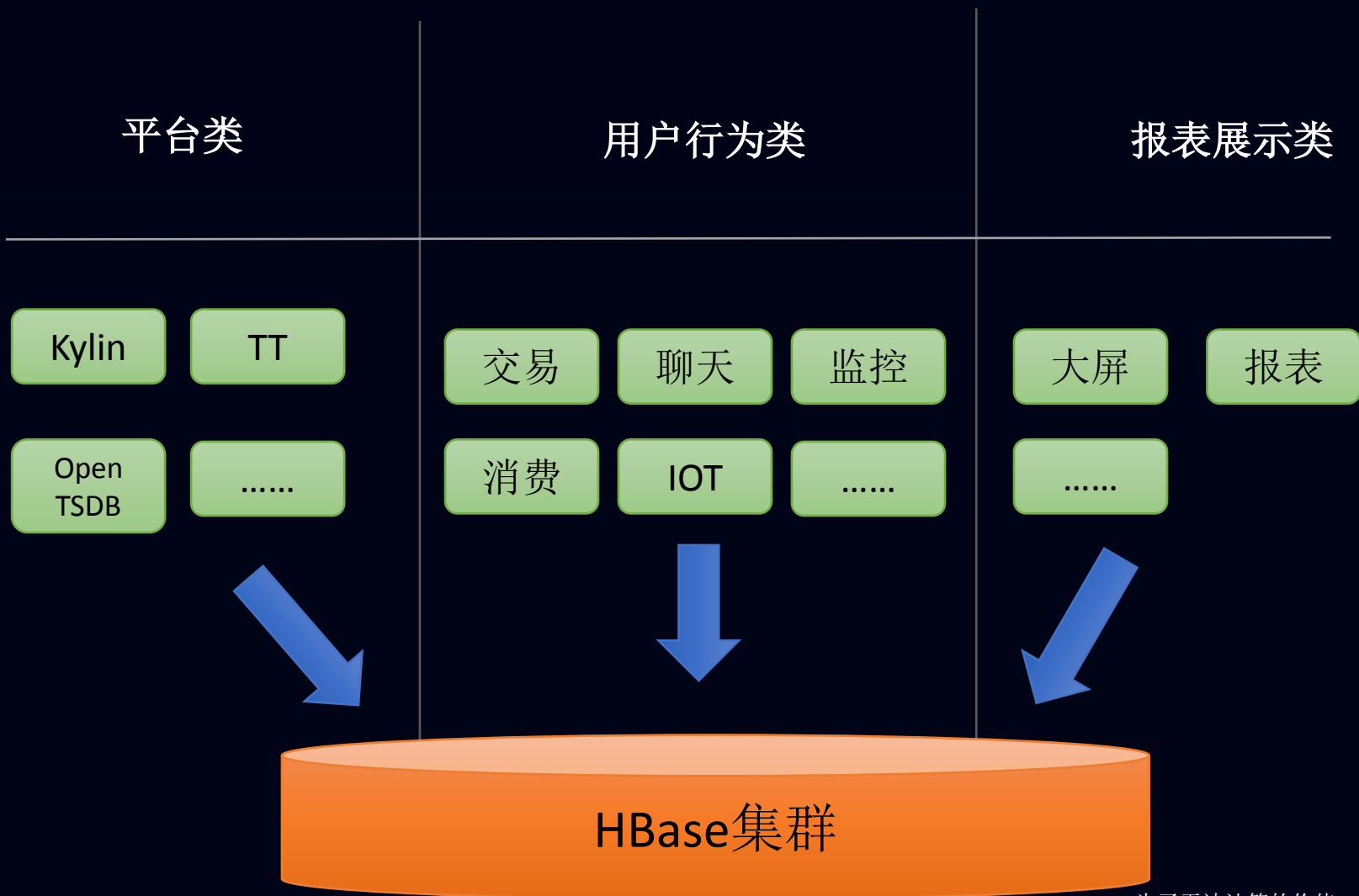


应用案例-汽车物联网

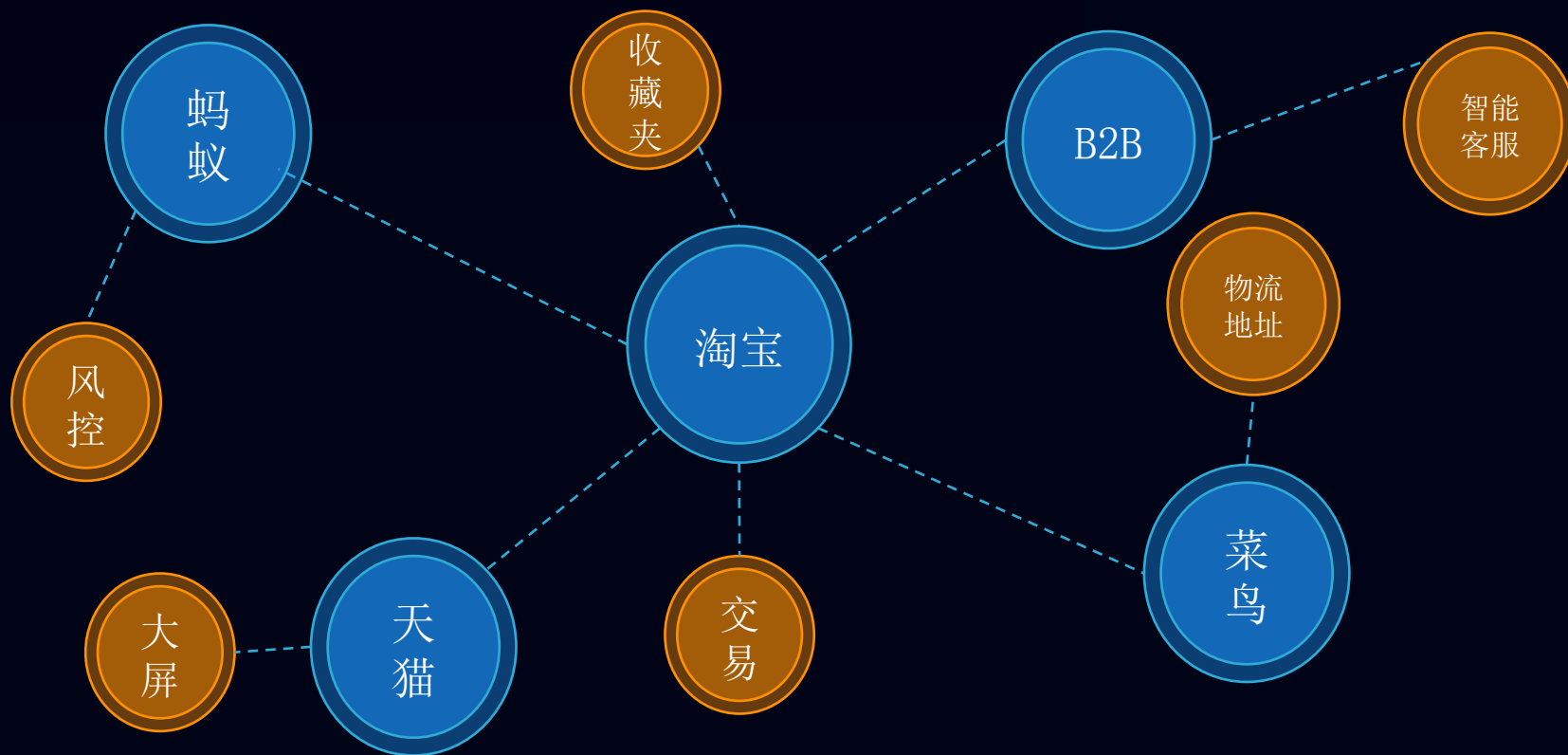
传感器会收集每辆货车的定位数据上传到MQ中，其中包括经度、纬度、设备上报时间、车辆状态等



HBase三大类应用



HBase阿里业务



云HBase技术大图

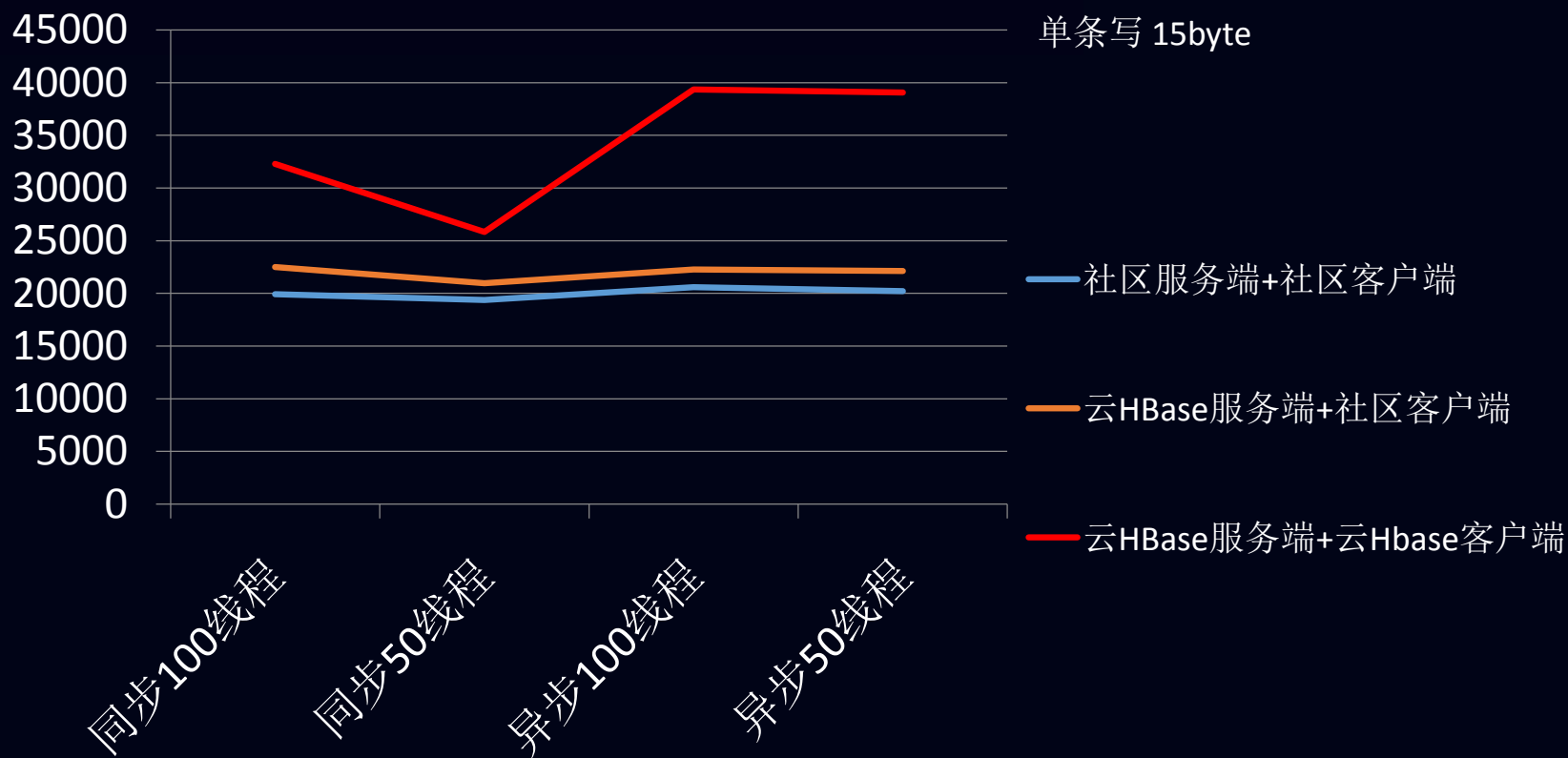


云HBase内核改造

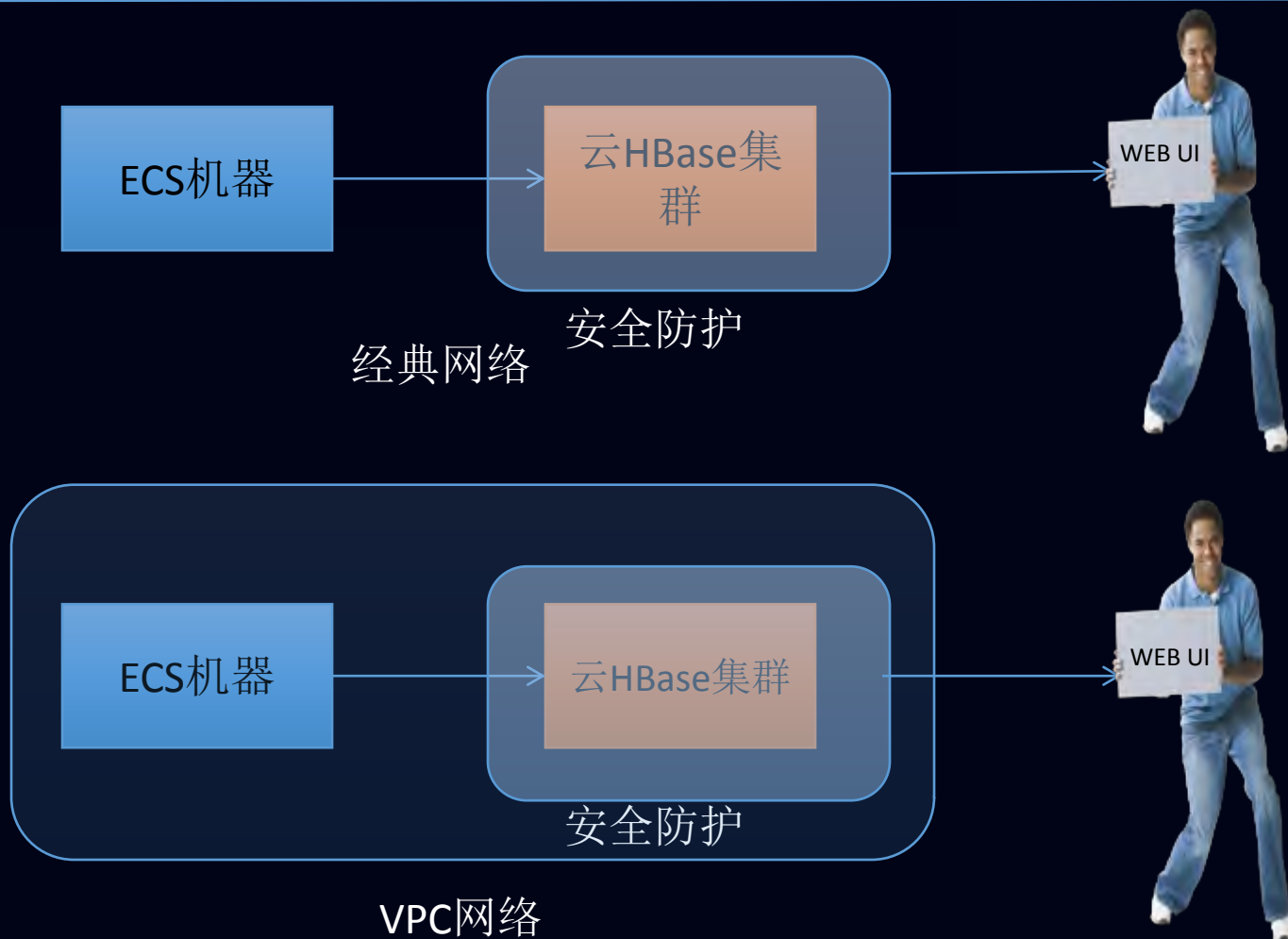


云HBase内核性能

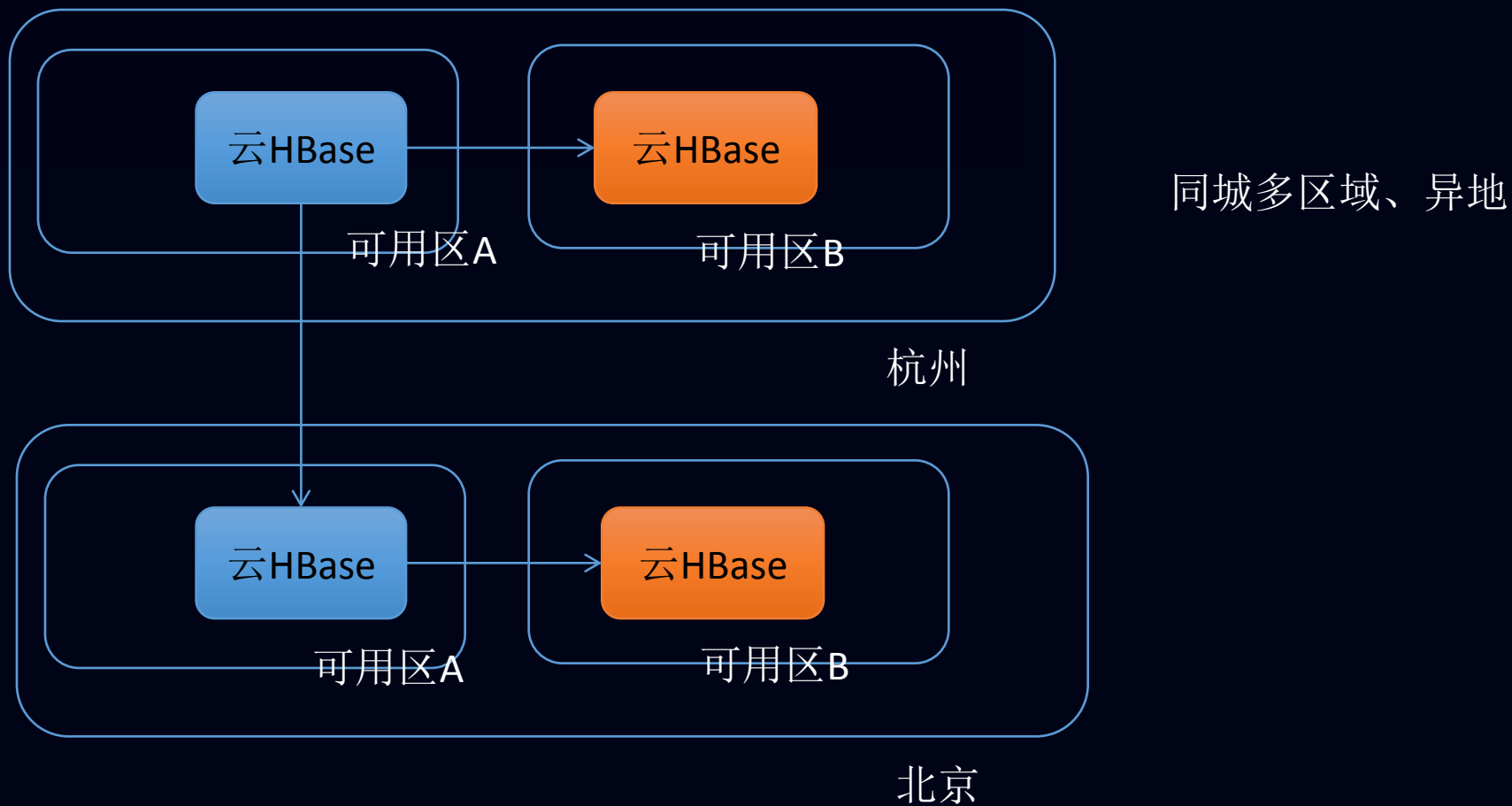
社区版本：1.1.1 VS 阿里云云HBase版本
2个slave 4cpu16g 启动单个RegionServer
单条写 15byte



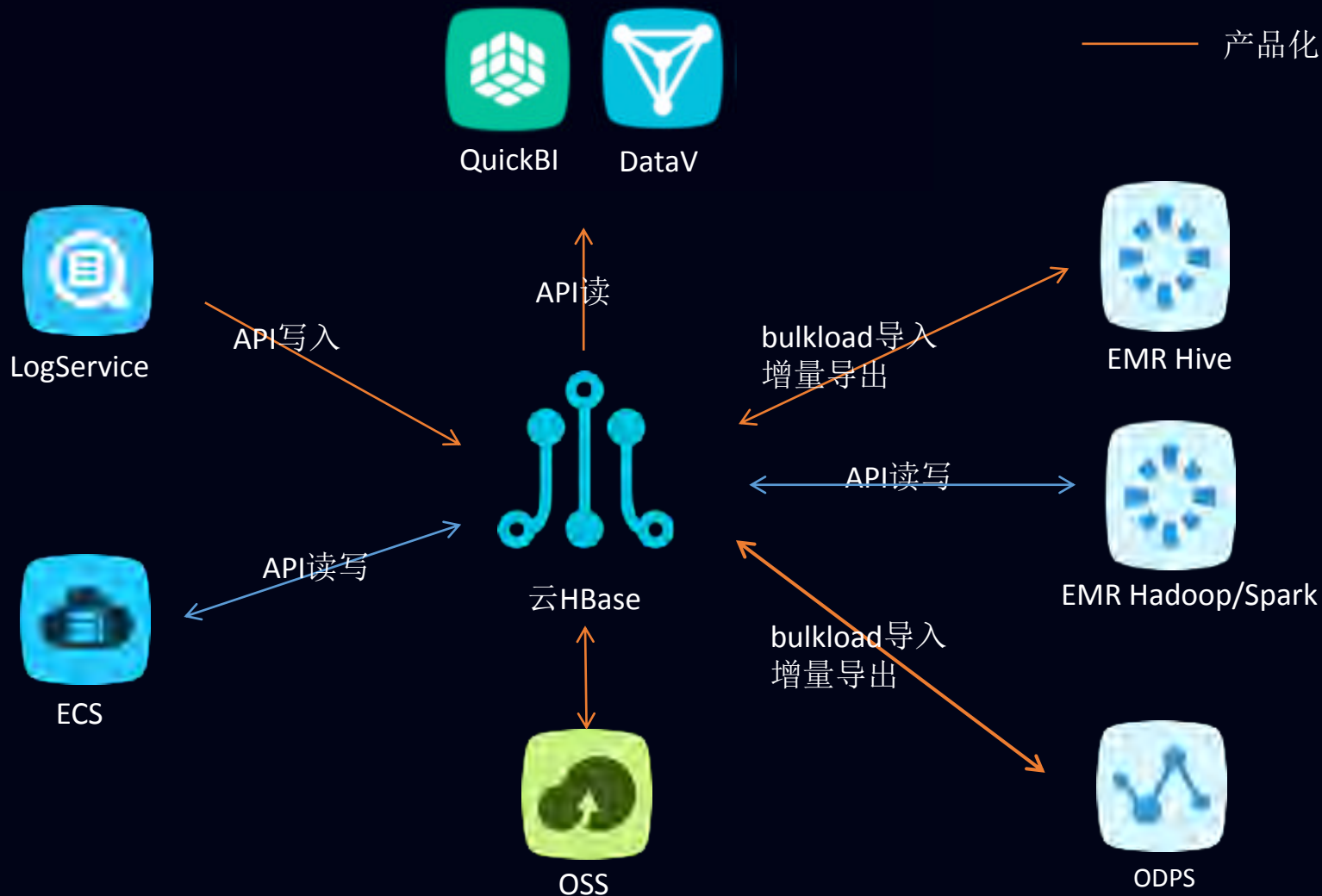
云HBase访问控制



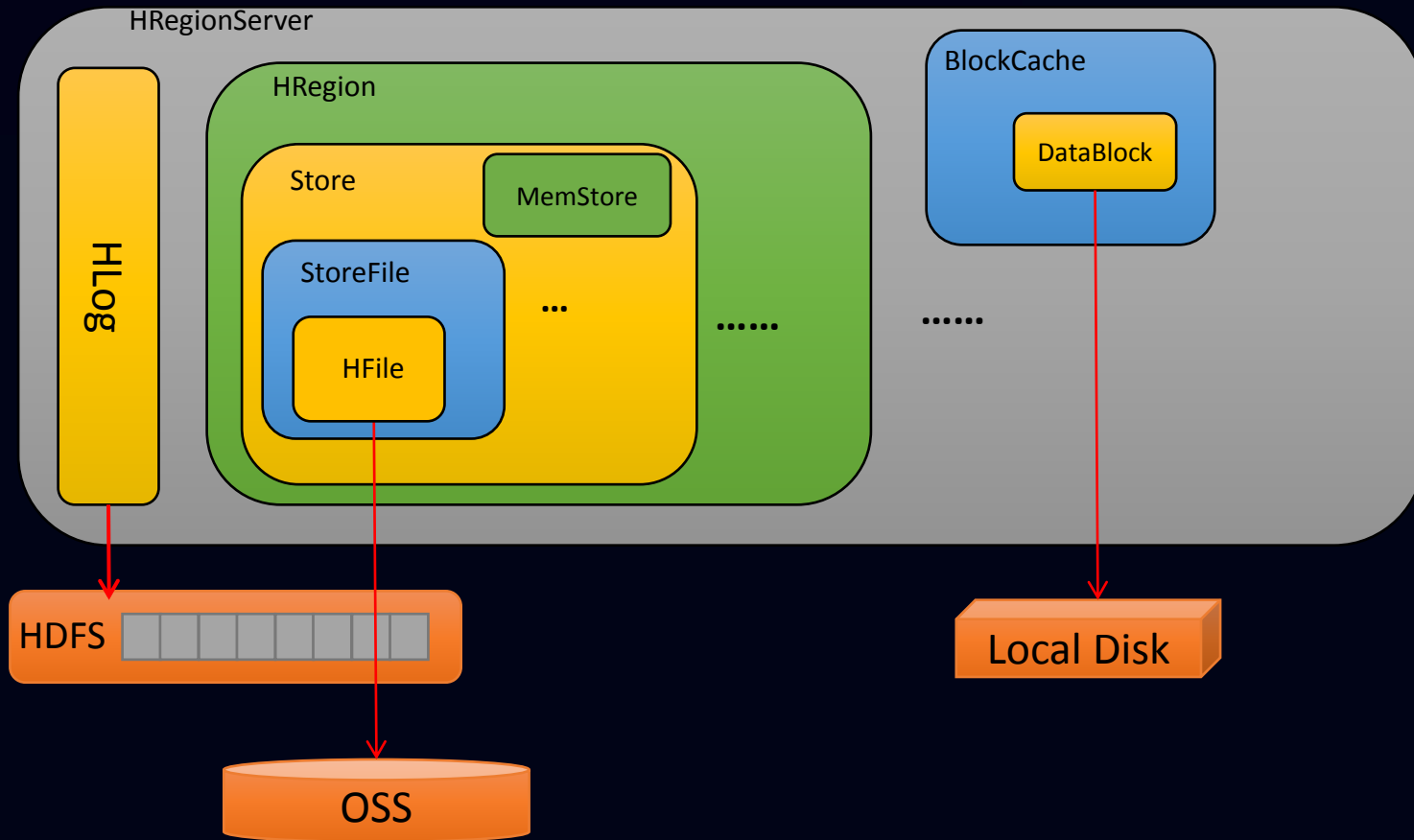
云HBase高可用-容灾



云HBase数据流



云HBase On OSS



新内核改进

- HBase2.0
 - Offheaping
 - Spark Integration
 - In-Memory Flush/Compaction
 - Async WAL
- 硬件发展带来的变化
 - 万兆网络普及，存储计算分离更加彻底
 - 固态存储容量更大、更廉价，IO不再成为瓶颈
 - 内存快速增大、可持久化，数据存储启用内存介质
- 云
 - 存储存放在OSS中

回顾

1. HBase在大数据中的价值
2. HBase应用案例
3. 云HBase建设
4. 云HBase未来展望

2017 china Hadoop summit

欢迎使用



云HBase产品首页



云HBase钉钉技术交流群