

随时听讲座
每天看新文
追随技术信仰

火龙果讲堂：

- 一线专家
- 案例回顾
- 经验分享



云之大数据平台架构设计

周小四 / Ray Zhou

目录

- ▶ 云计算与大数据
- ▶ 云上大数据平台的挑战
- ▶ 大数据平台系统架构
- ▶ 数据格式

云计算与大数据 Big Data Requires Elastic, Flexible Infrastructure

▶ 大数据解决方案最大、最重要的决定是平台的选择

Hadoop | Spark | Hbase | GreenPlum ...

▶ 根据实际需求选择平台

实时处理 | 离线处理 | 结构化数据 | 非结构化数据 | CRUD | 事务性 | 数据量大小 ...

▶ 平台的能力

能满足实际需求的平台才是好平台

▶ 业务的不确定性

Relative scale, not absolute scale

The world is changing!

云计算与大数据 Don't try to do it yourself. Let us handle it.

▶ 自动化运维

一键部署 | API 调用 | 定时器 | Auto Scaling

▶ 弹性、敏捷、灵活

纵向、横向伸缩 | 分钟级别

▶ 稳定、高性能与安全

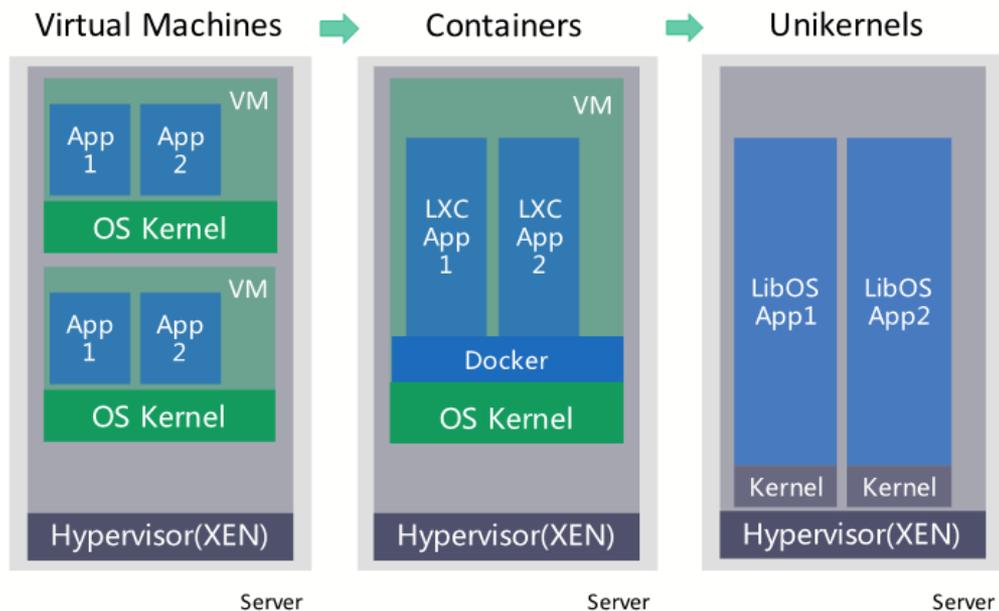
系统优化，平台配置优化 | 私有网络，防火墙 | 高度自动化是一个渐进的过程

目录

- ▶ 云计算与大数据
- ▶ 云上大数据平台的挑战
- ▶ 大数据平台系统架构
- ▶ 数据格式

挑战

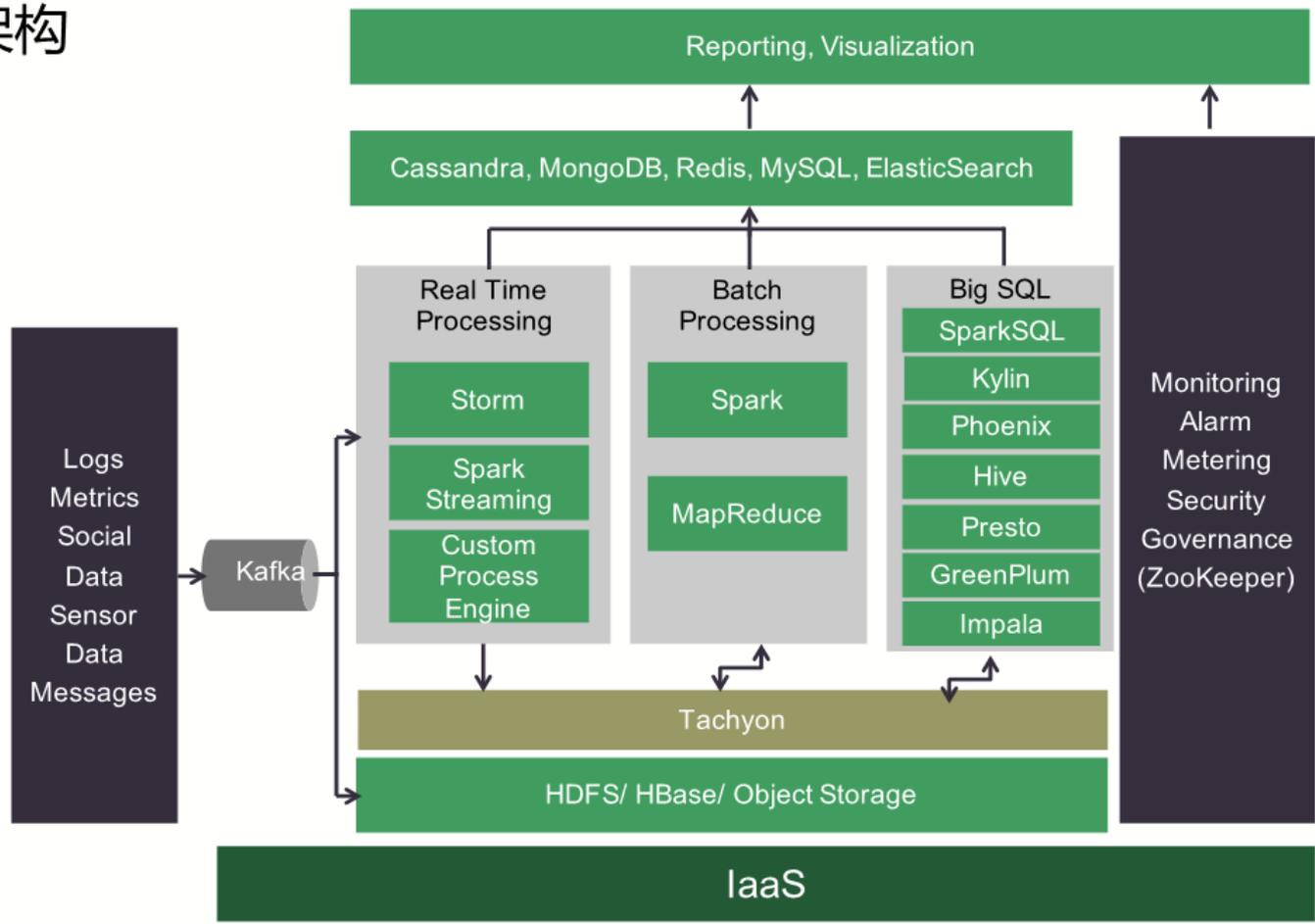
- ▶ 稳定性
- ▶ 性能
 - ❑ 网络 IO – SDN 2.0 (点到点网状结构)
 - ❑ 硬盘 IO – 容器, Unikernel, 定制化IaaS
- ▶ 迁移
 - ❑ Within cloud
 - ❑ Between cloud and on-premise
 - ❑ Tools : Sqoop , Kafka , Flume , hdfs fs -put , distcp , etc.



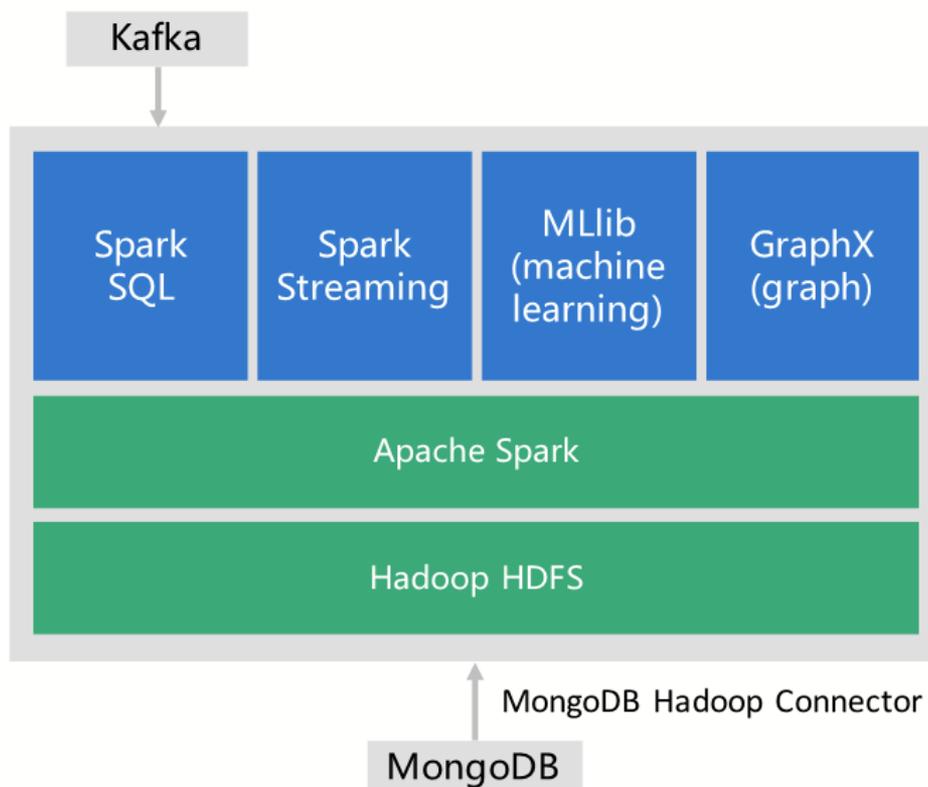
目录

- ▶ 云计算与大数据
- ▶ 云上大数据平台的挑战
- ▶ 大数据平台系统架构
- ▶ 数据格式

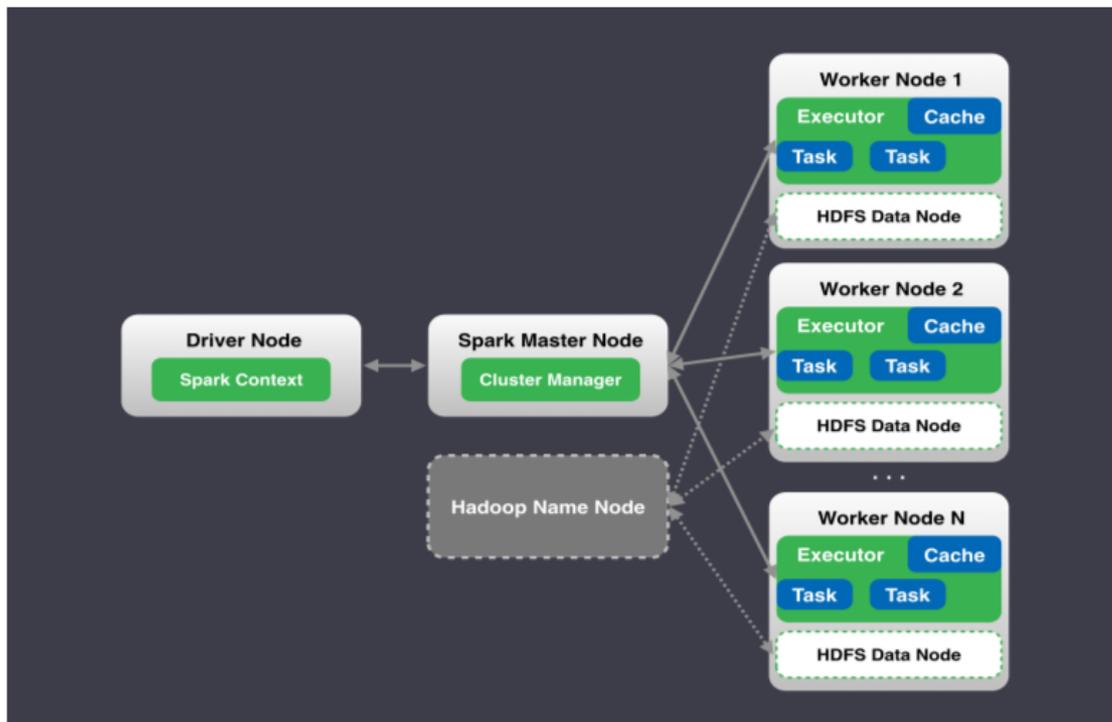
系统架构



计算 – Spark / Hadoop



计算 – Spark / Hadoop



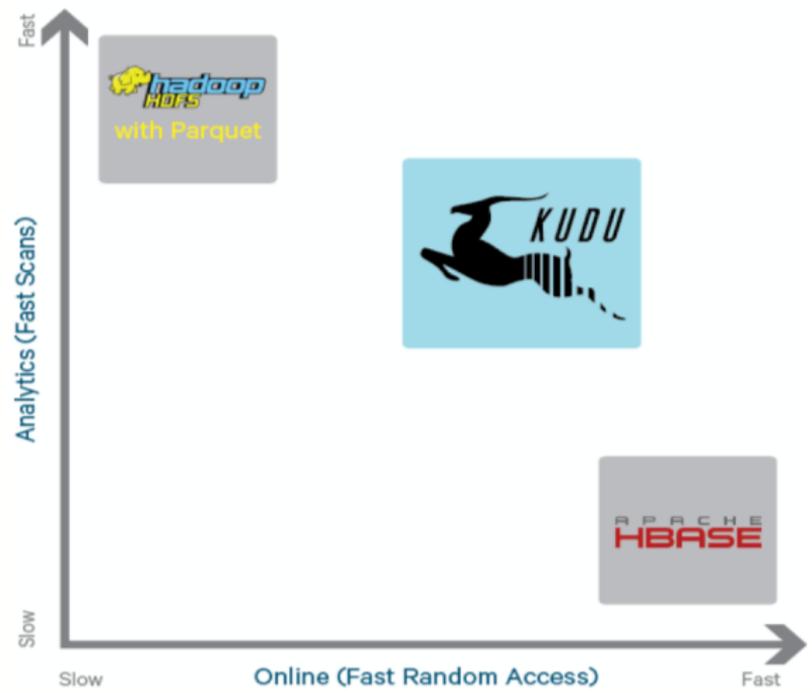
计算平台

	Spark Streaming	Storm*	MapReduce
延时	秒级	毫秒级	较高
吞吐量	大	较大	大
硬盘IO	一般	一般	高
网络	高	高	高
适用场景	近实时较大数据块 分析计算	实时小数据块 分析计算	离线高吞吐量 分析计算
当前稳定版本	1.5.2	0.9.5	2.6.2
起源	UC Berkeley AMPLab	Twitter	Google

*Heron – claimed better than Storm by Twitter

存储 - Hadoop

Hadoop Storage Engines



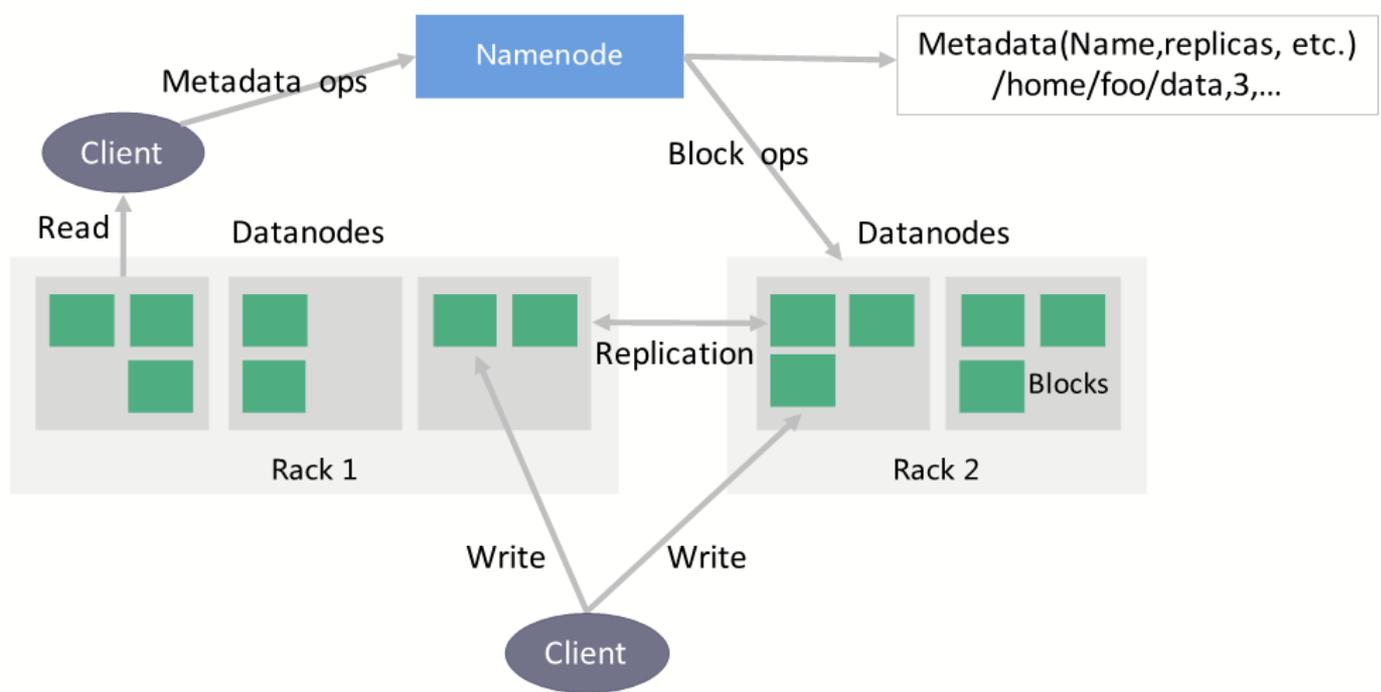
<https://vision.cloudera.com/introducing-kudu-the-new-hadoop-storage-engine-for-fast-analytics-on-fast-data/>

存储 – Greenplum Database

- ▶ GPDB：为大数据存储、计算、挖掘而设计
 - ❑ 标准 SQL 数据库：ANSI SQL 2008 标准，OLAP，JDBC/ODBC
 - ❑ 支持ACID、分布式事务
 - ❑ 分布式数据库：线性扩展，支持上百物理节点
 - ❑ 企业级数据库：全球大客户超过 1000+ 安装集群
 - ❑ 百万行源代码，超过10年的全球研发投入
 - ❑ 开源数据库(greenplum.org)，良性生态系统

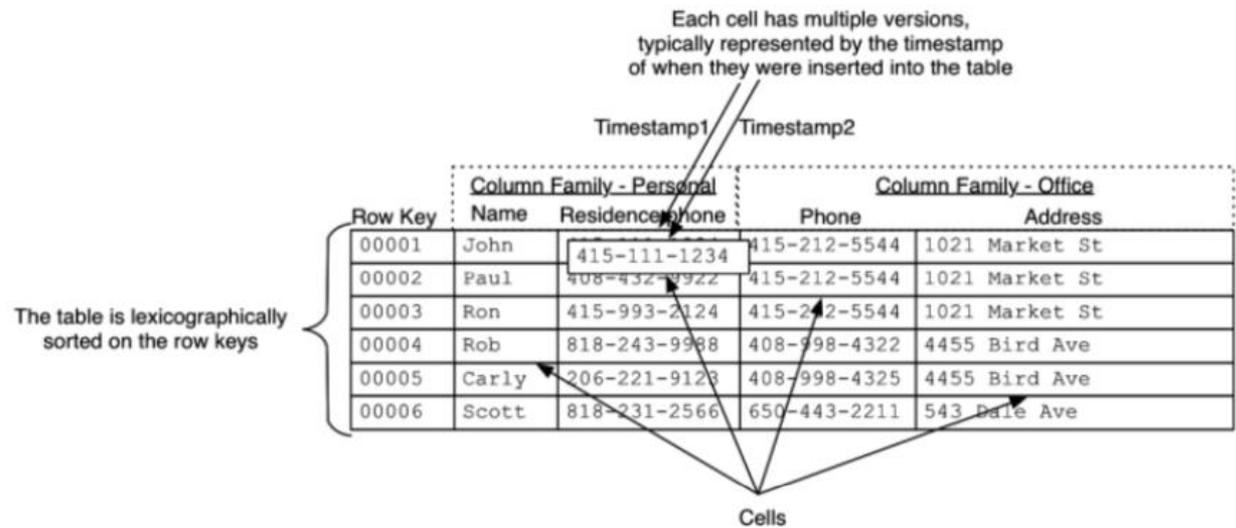
存储 - HDFS

HDFS Architecture

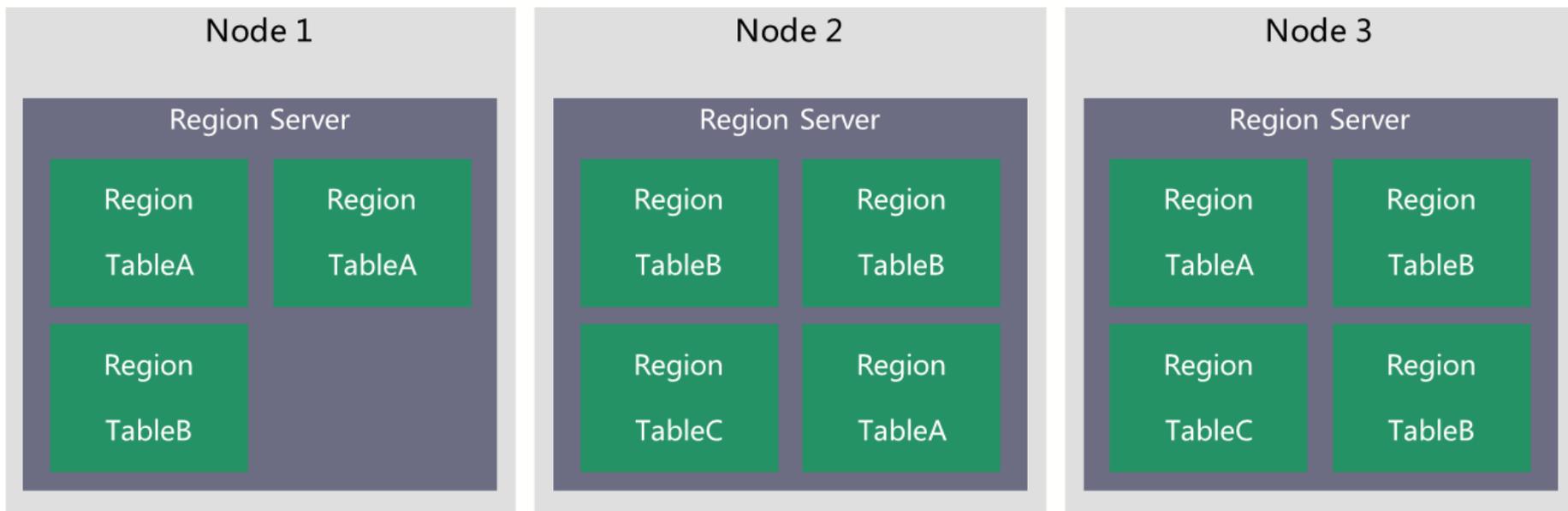


存储 – HBase

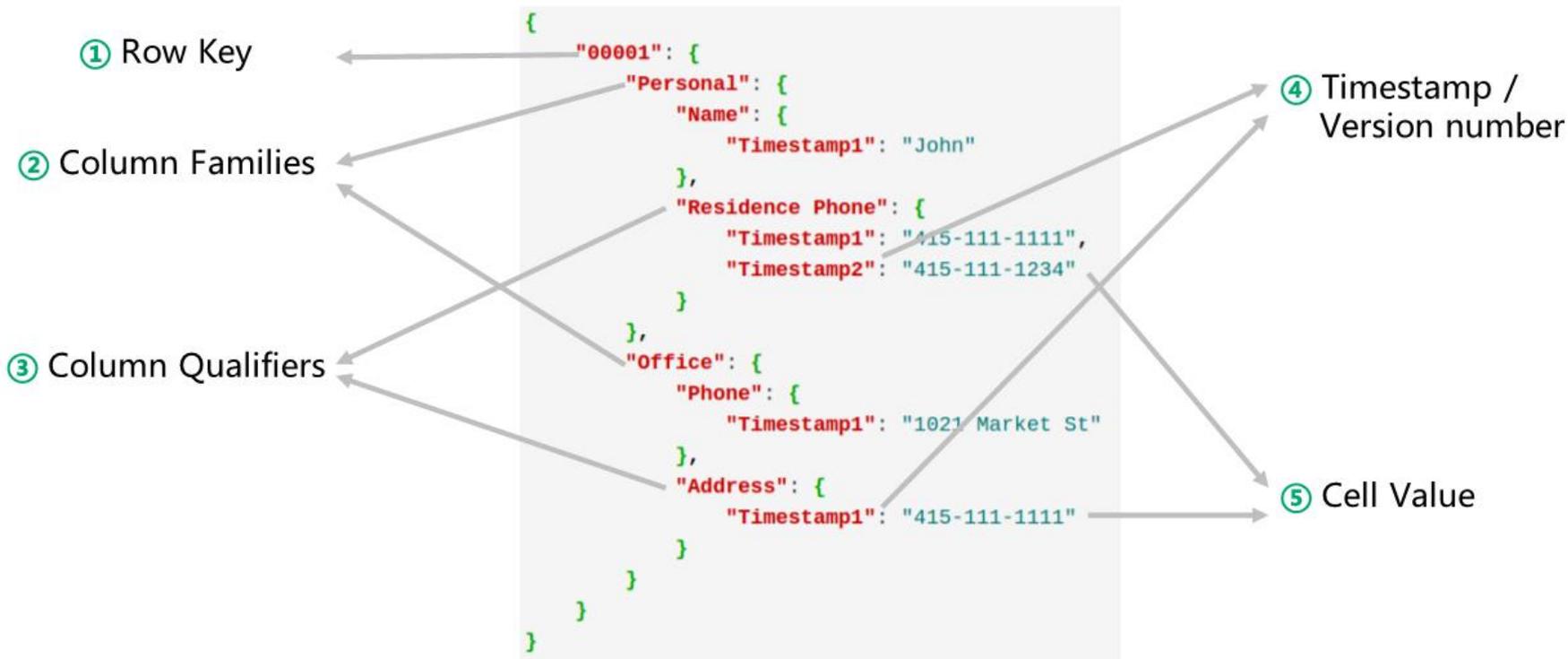
A sparse, distributed, persistent multidimensional sorted map



存储 - HBase



存储 - HBase

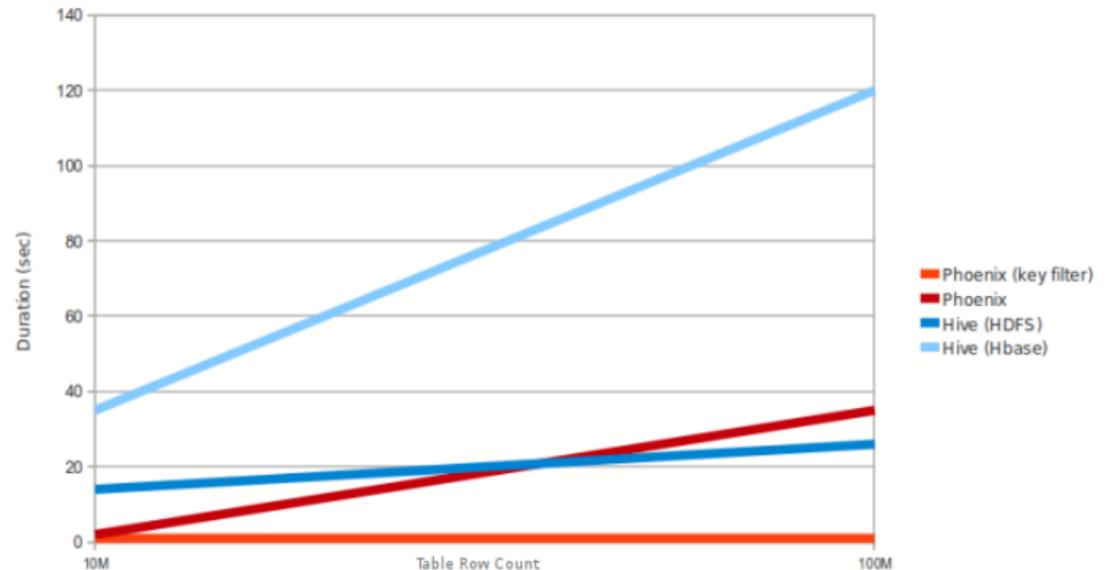


存储 – HBase

- ▶ Row keys - the most important aspect
 - ❑ Sorted lexicographically
 - ❑ Hotspotting
 - ❑ Monotonically Increasing Row Keys/Timeseries Data - TSDB
 - ❑ Rowkey Length – hash
 - ❑ Minimize row and column (family, qualifier) sizes
- ▶ Can store anything in the form of byte[]
- ▶ Similar access patterns in the same column family
- ▶ Denormalized
- ▶ Tall tables v.s. Wide tables
- ▶ Single API calls rather than multiple API calls

SQL on Hadoop

- ▶ SparkSQL, Phoenix, Hive ,
Kylin, Presto, HAWQ
- ▶ SparkSQL支持交互式查询

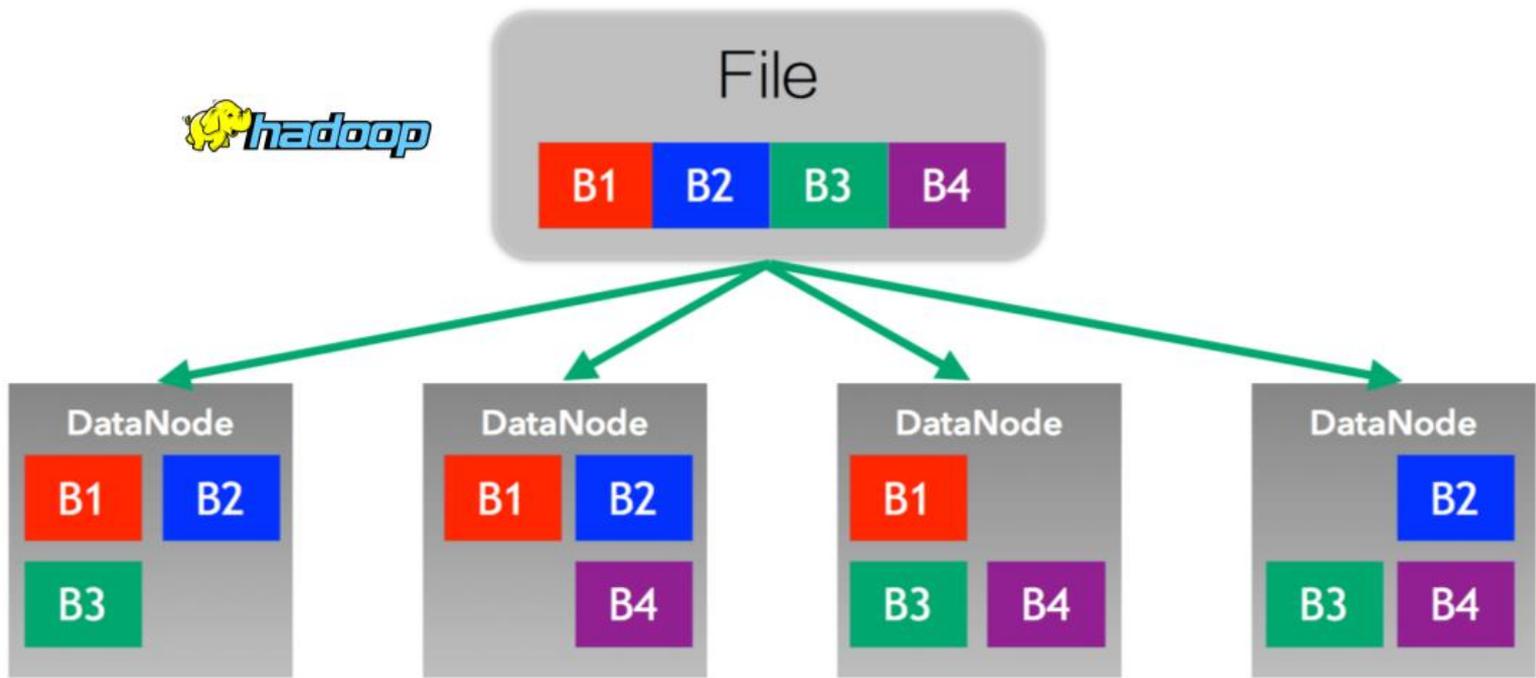


Query: select count(1) from table over 10M and 100M rows. Data is 5 narrow columns. Number of Region Servers: 4 (HBase heap: 10GB, Processor: 6 cores @ 3.3GHz Xeon)
<https://phoenix.apache.org/performance.html>

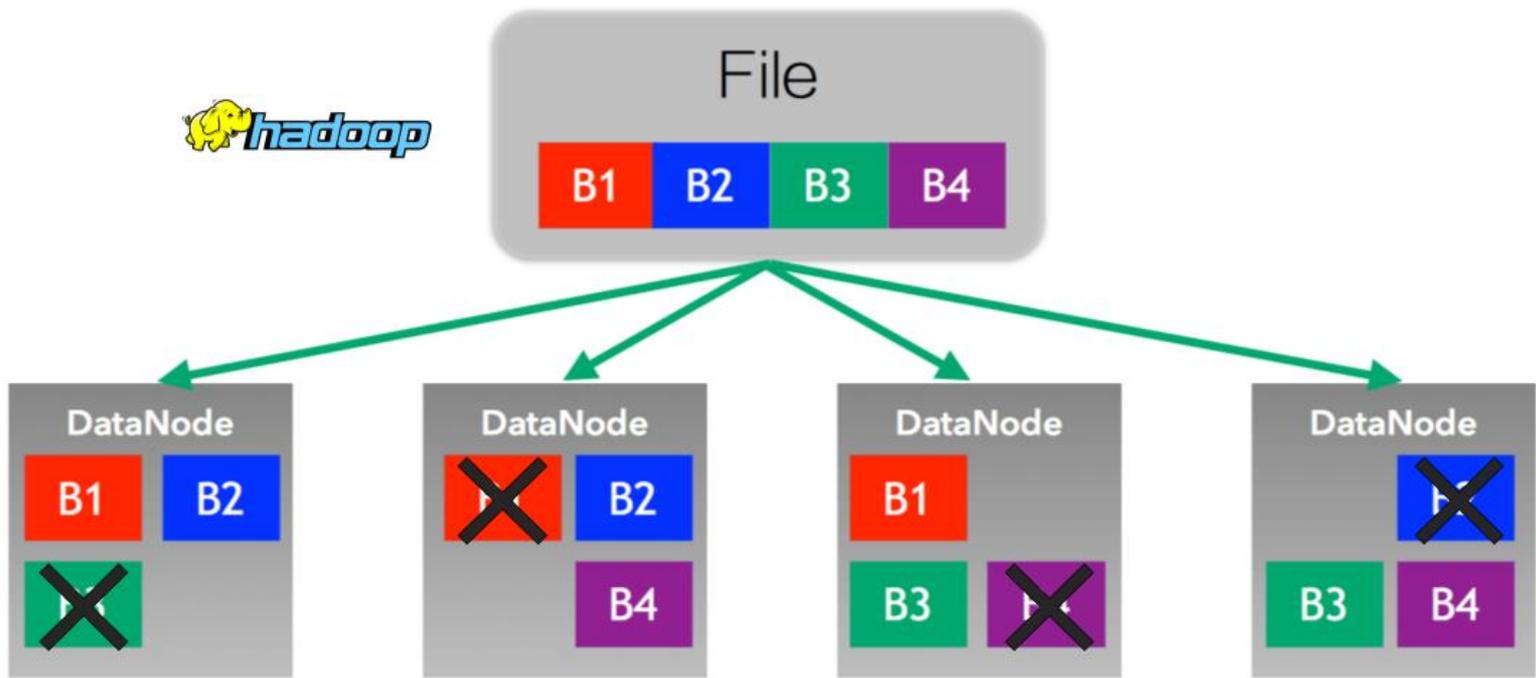
存储 – Hadoop on Cloud

- ▶ 默认3个副本因子
- ▶ 云上真的需要3个吗？
- ▶ 怎么做最优？

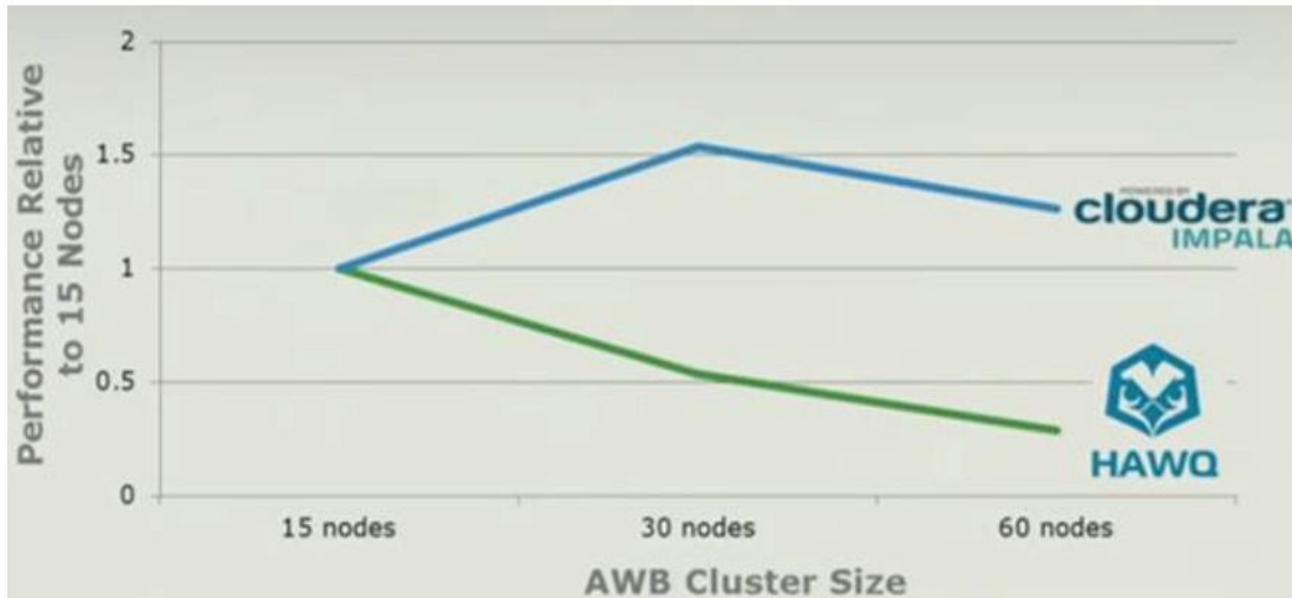
存储 - Hadoop on Cloud



存储 - Hadoop on Cloud



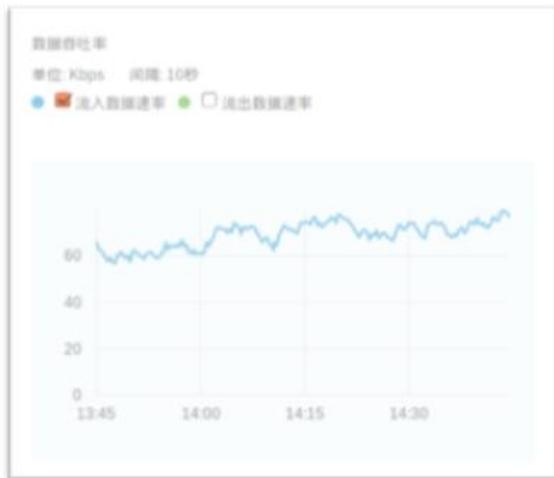
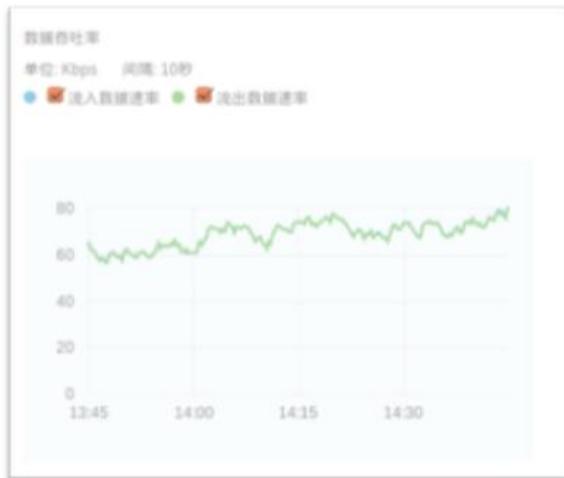
SQL on Hadoop



传输

► Why Kafka

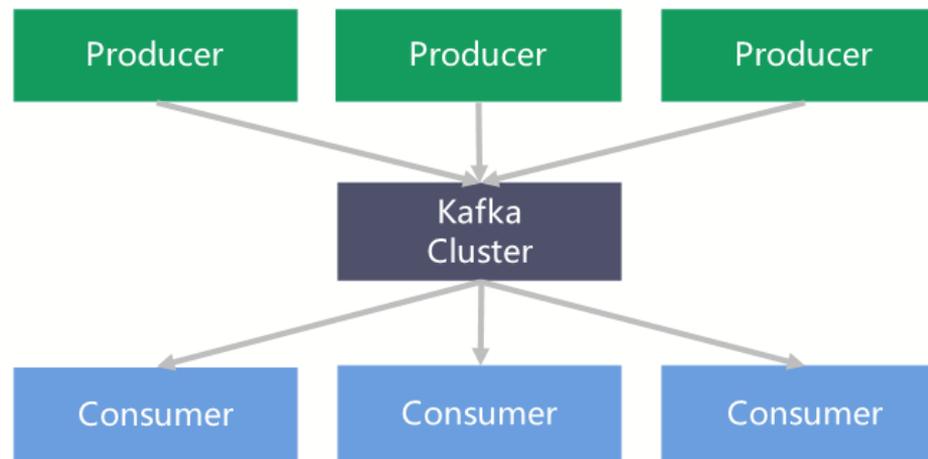
- ❑ 分布式、可分区、多副本的高吞吐量低延迟消息系统
- ❑ 3 Kafka servers, 1 producer -- 821,557 records/sec(78.3 MB/sec)*
- ❑ 适用于处理活跃的流式数据，比如网页的访问量、日志等



* <https://engineering.linkedin.com/kafka/benchmarking-apache-kafka-2-million-writes-second-three-cheap-machines>

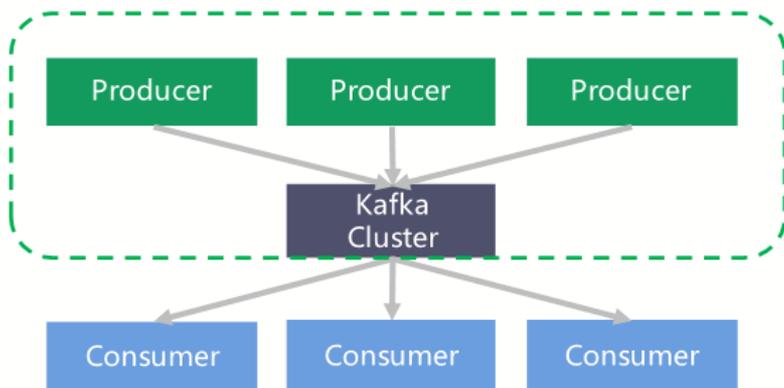
传输 - Kafka

- ▶ Producer, consumer, broker



传输 - Kafka

- ▶ Producer, consumer, broker
- ▶ Topic, partition, replicate

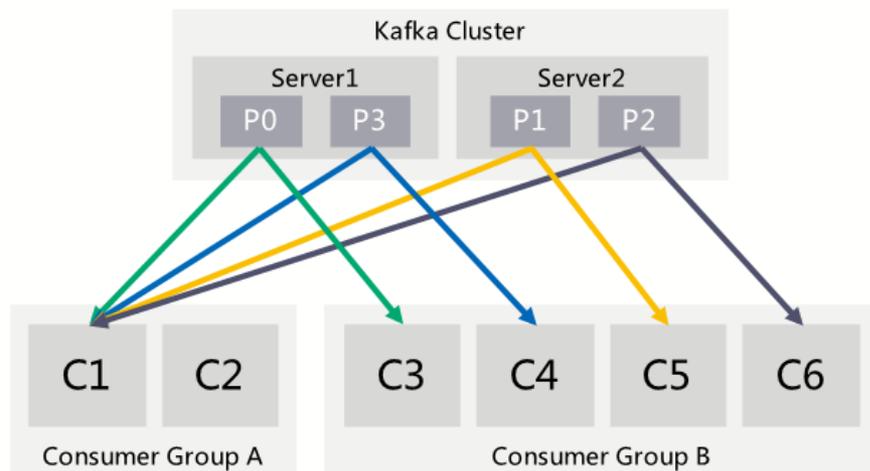
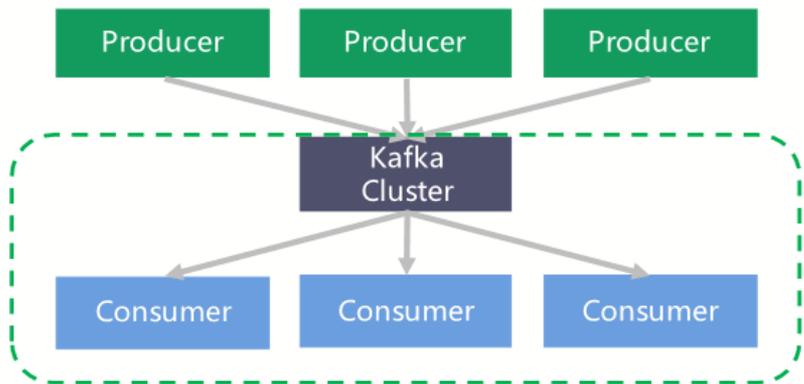


Anatomy of Topic



传输 - Kafka

- ▶ Producer, consumer, broker
- ▶ Topic, partition, replicate



目录

- ▶ 云计算与大数据
- ▶ 云上大数据平台的挑战
- ▶ 大数据平台系统架构
- ▶ 数据格式

数据格式

▶ 错误的格式

- ❑ 性能成倍下降
- ❑ 空间成倍上升

▶ 可分割的(Splittable)

- ❑ 支持：CSV, JSON记录, Avro, Parquet
- ❑ 不支持：XML, JSON文件

▶ 可块压缩的(Block Compressible)

- ❑ 支持：Avro, Parquet
- ❑ 不支持：CSV, JSON记录



Thank you.

交流时间



讲座 5月14日 如何塑造敏捷团队的价值观与文化



随时听讲座
每天看新文
——
追随技术信仰