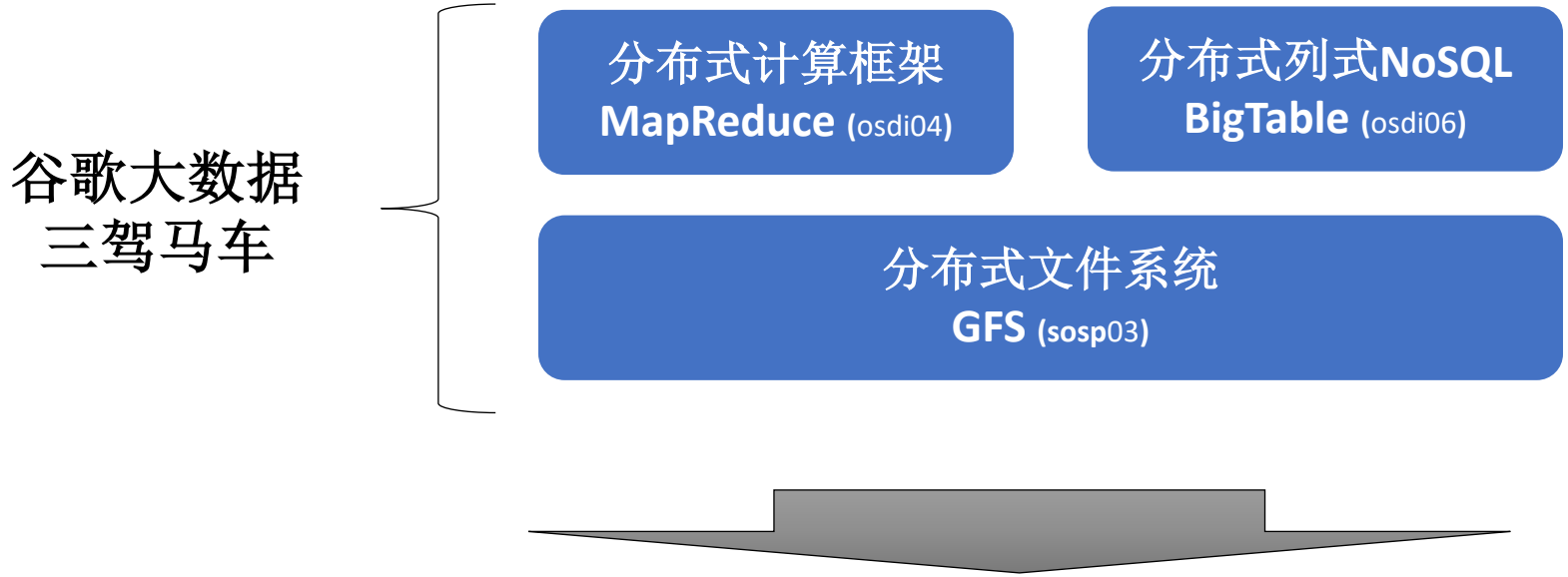
对比MySQL，看HBase的能力及场景

天梧 2019-05

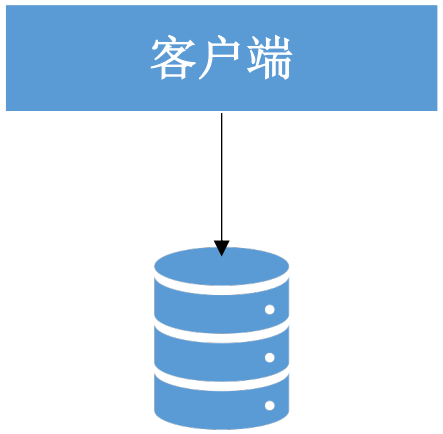
HBase的历史



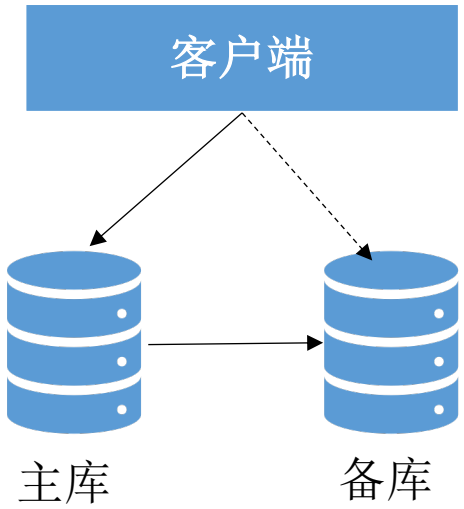
注: Spark是诞生于AMPLab，2010年开源

从架构对比看差异

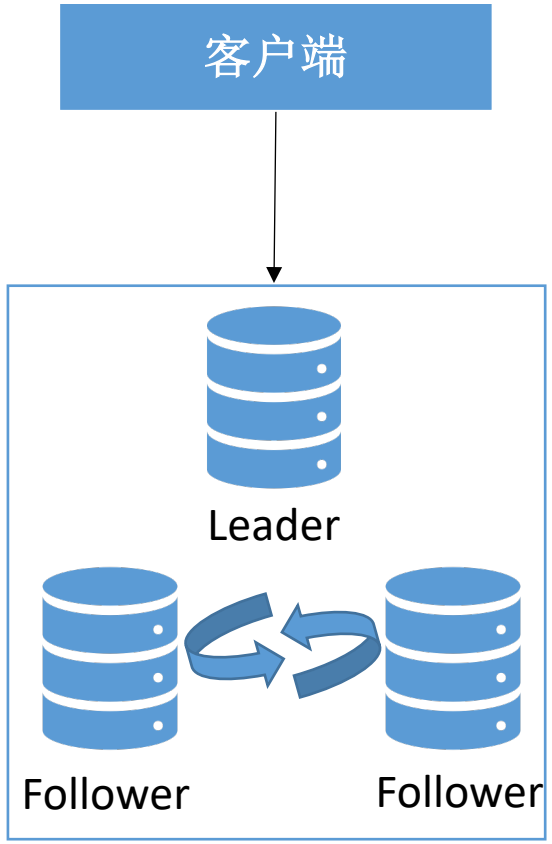
MySQL的常见架构部署



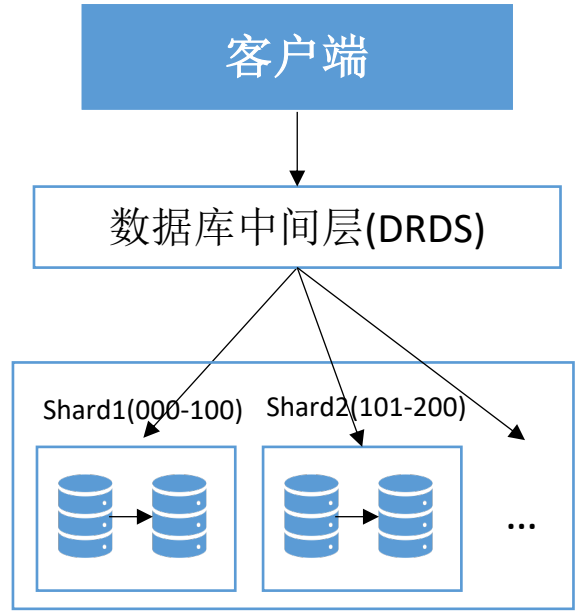
单节点



主备HA

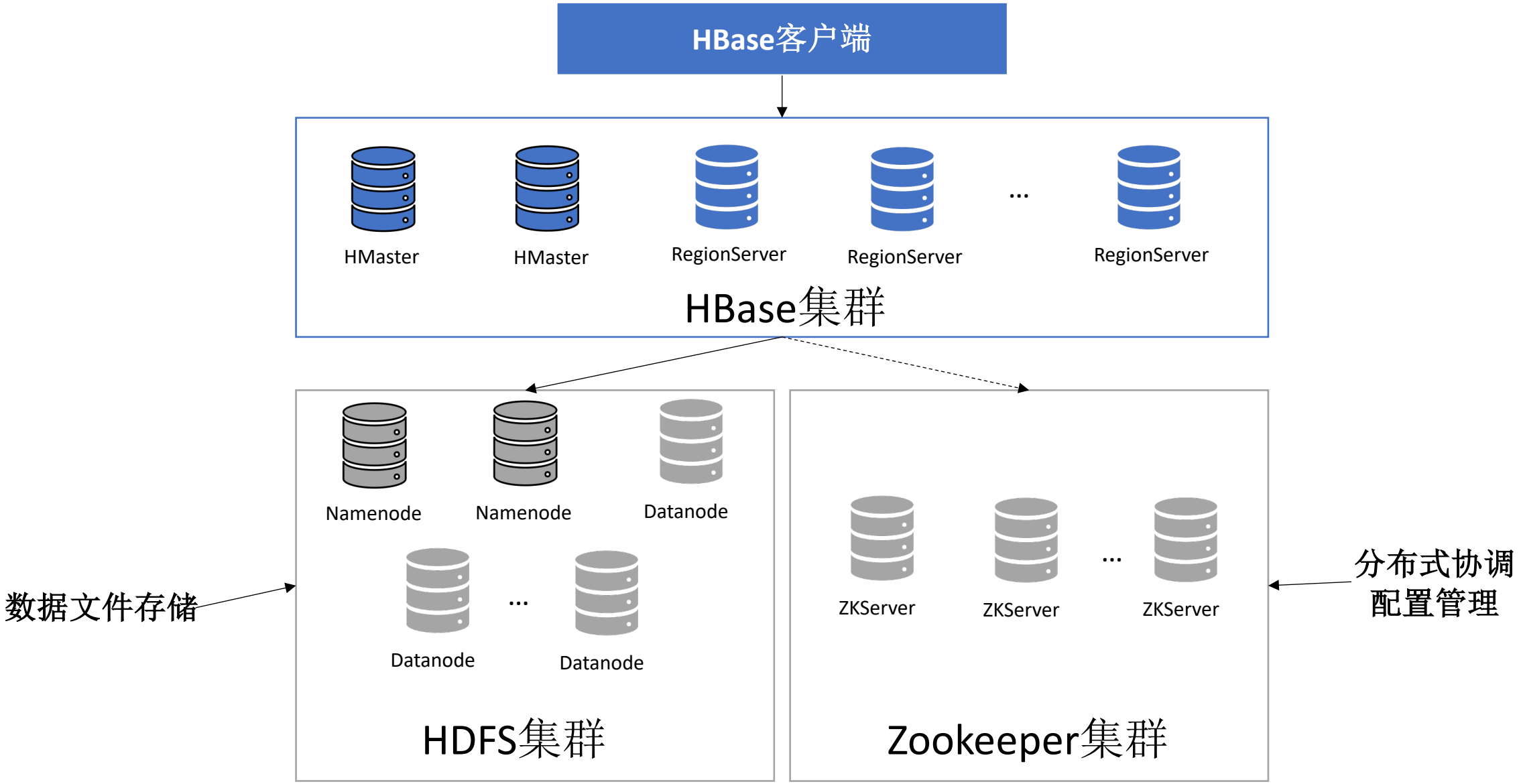


三节点强一致



分库分表
与主备HA

HBase的架构部署


HBase客户端
HMaster
HMaster
RegionServer
RegionServer
...
RegionServer
HBase集群
Namenode
Namenode
Datanode
Datanode
...
Datanode
HDFS集群
ZKServer
ZKServer
...
ZKServer
Zookeeper集群
数据文件存储
分布式协调
配置管理

相比MySQL， HBase的架构特点：

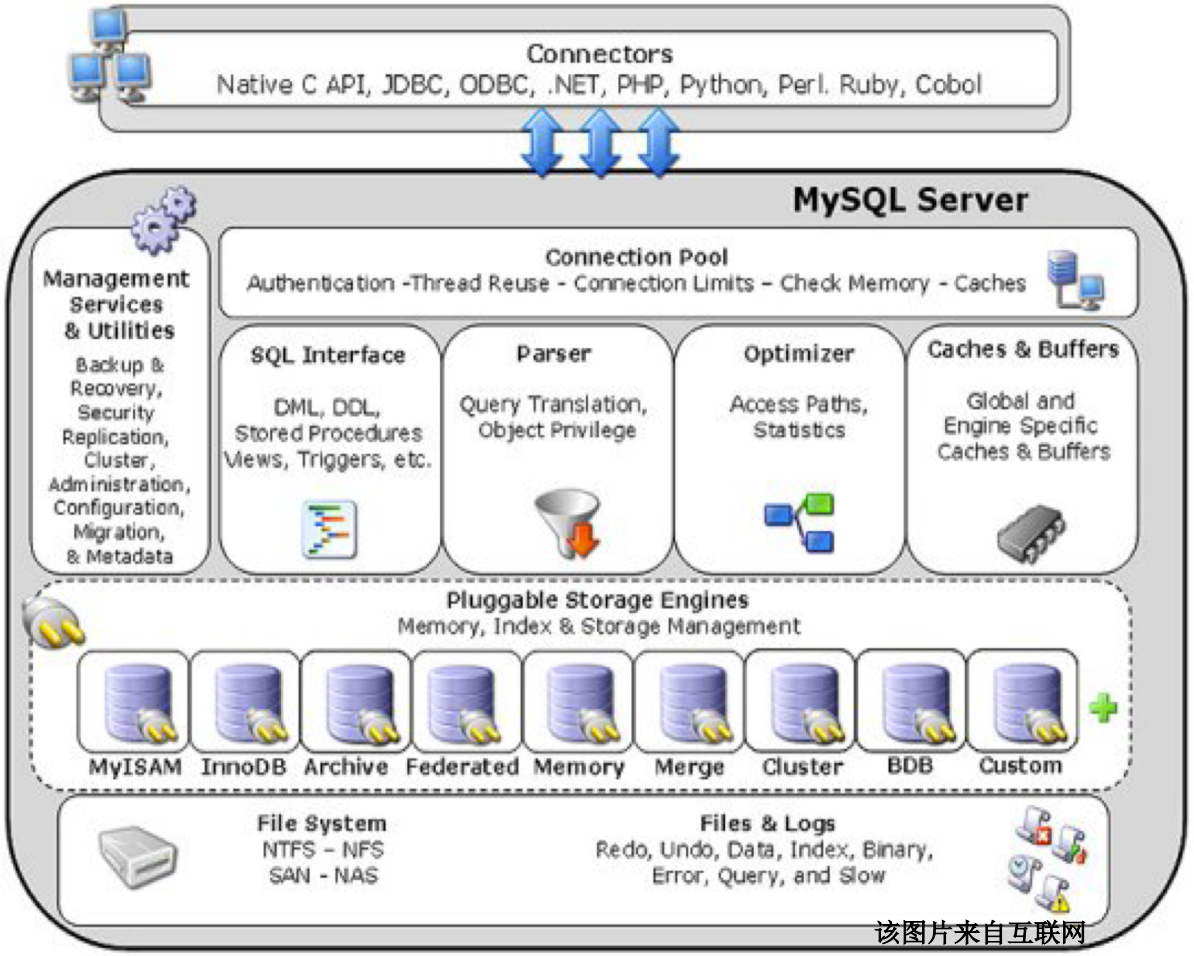
1. 完全分布式 (数据分片、故障自恢复)
2. 底层使用HDFS(存储计算分离)

由架构看到的能力差异：

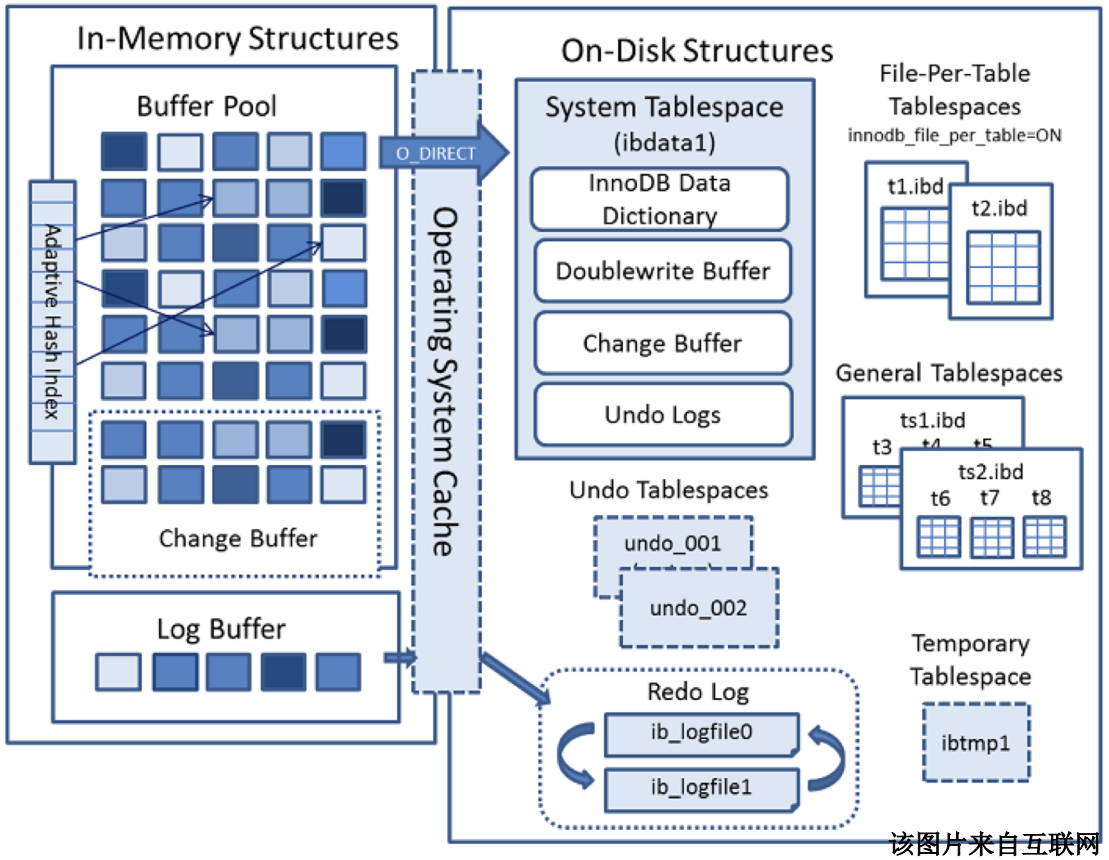
1. MySQL：运维简单(组件少)、延时低(访问路径短)
2. HBase：**扩展性好**、内置**容错恢复**与**数据冗余**

从引擎结构看差异

MySQL的引擎结构



该图片来自互联网

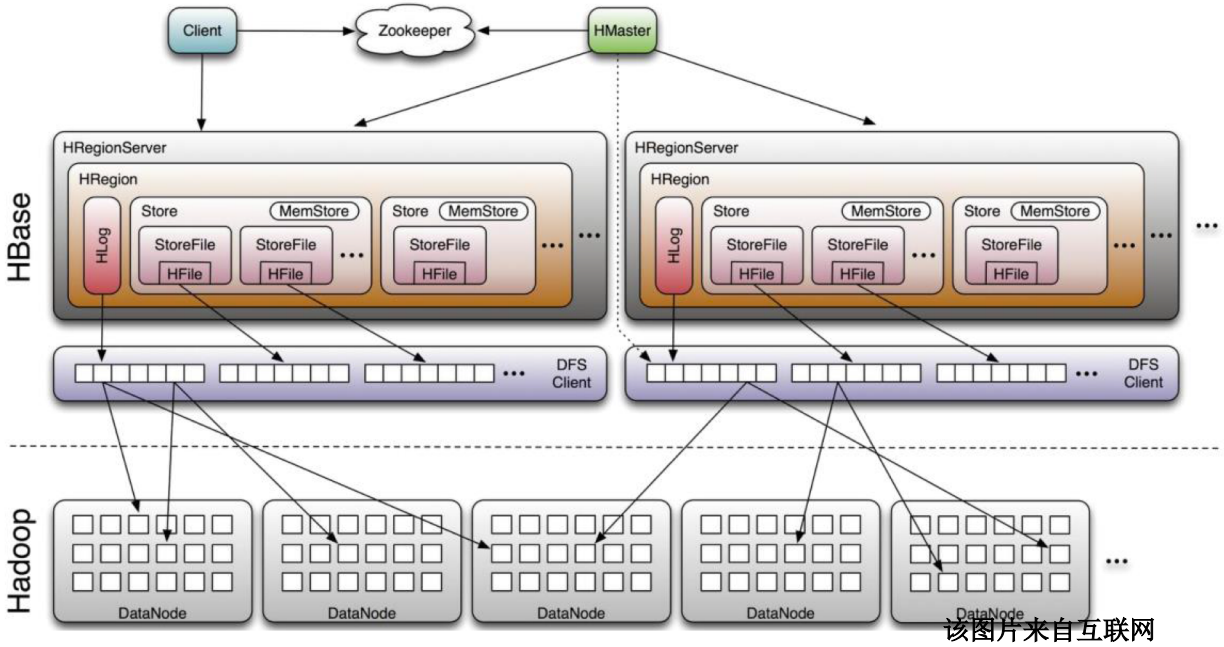


该图片来自互联网

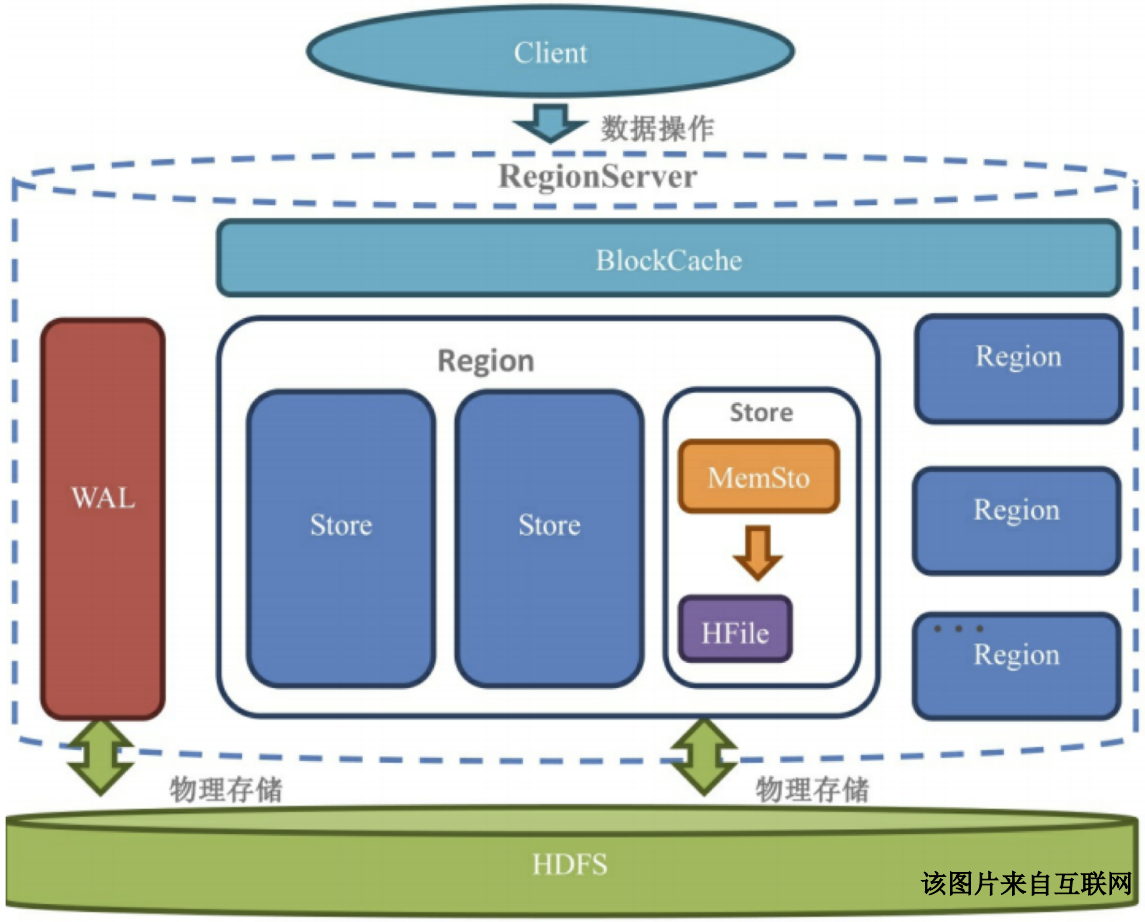
SQL引擎 + 可插拔存储引擎

MySQL的默认存储引擎InnoDB
(B+ Tree)

HBase的引擎结构



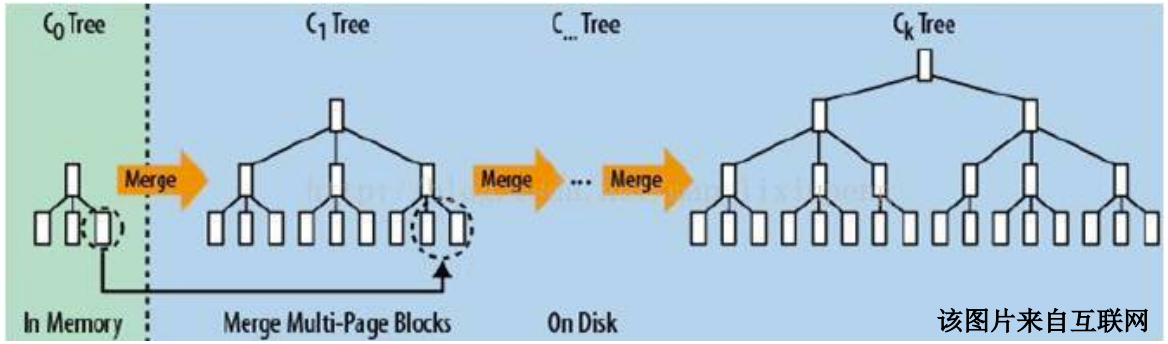
RegionServer (负责数据的组织和查询)
HDFS(负责文件存储，类比Disk)
一个Table会拆成多个分片(Region)分配到多个Regionserver



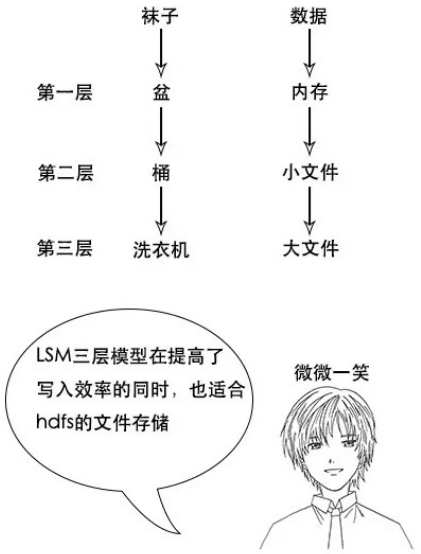
Regionserver的内部结构
(LSM Tree)

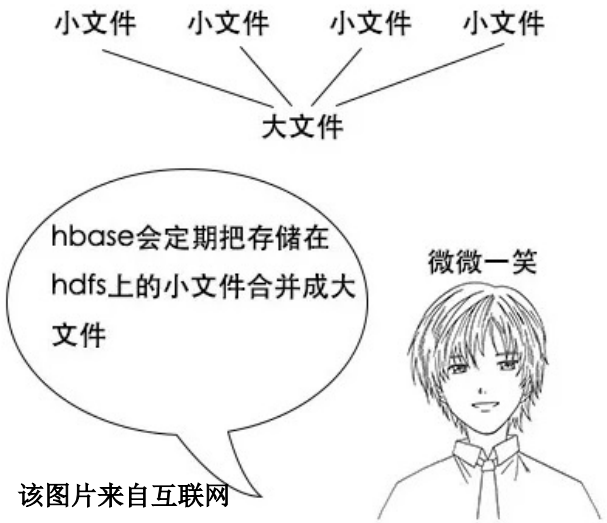
相比MySQL， HBase的内部引擎特点：

1. HBase原生没有SQL引擎 (无法使用SQL访问，使用API)， 云HBase增强版(Lindorm)及开源Phoenix均提供SQL能力
2. HBase使用LSM(Log-Structure Merge)树， Innodb使用B+树



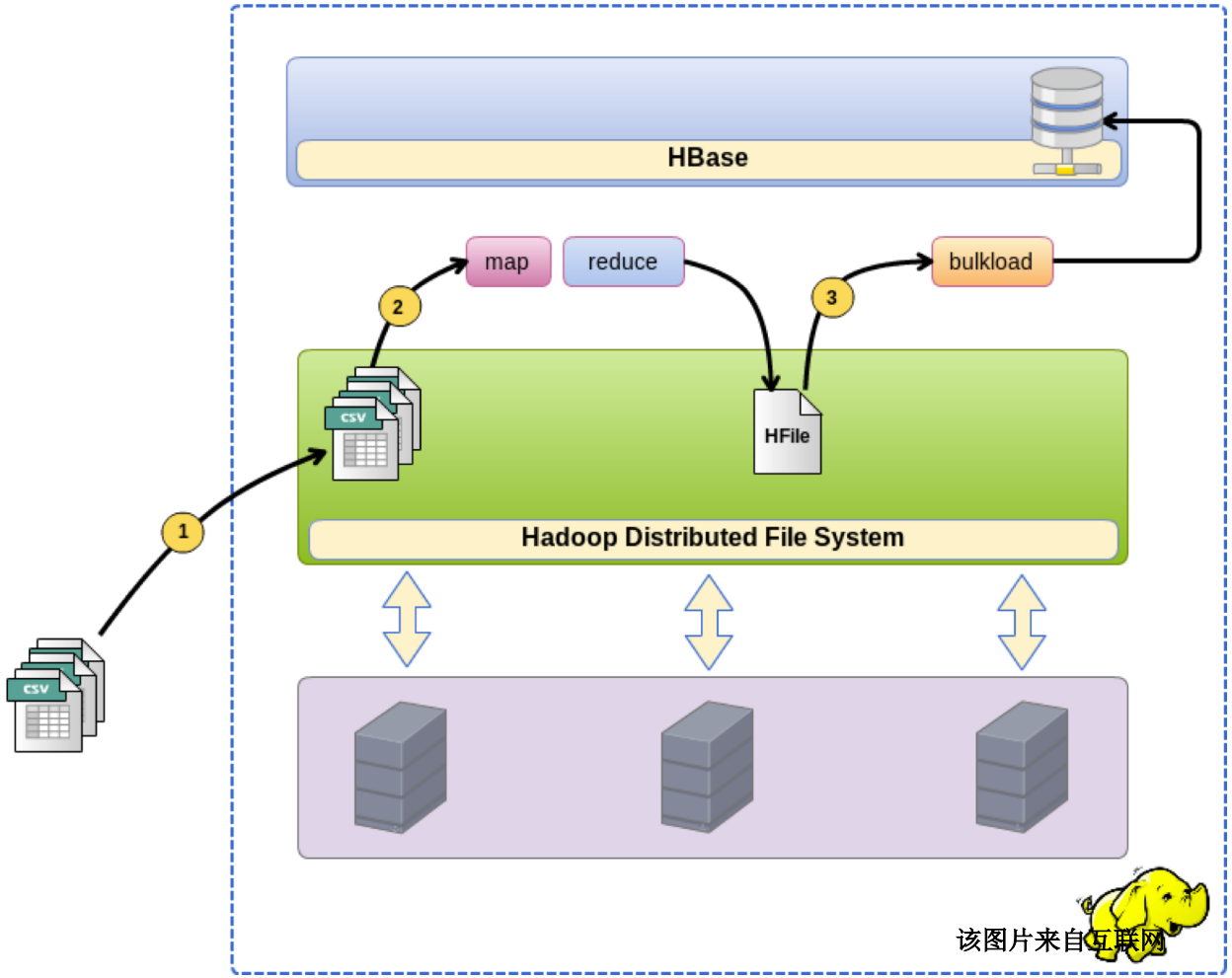
该图片来自互联网





该图片来自互联网

- LSM特点
  - 侧重写（只需写memory）
  - 读需要访问多颗树，IO次数相比B+树，波动大
  - 写都是顺序IO，随机读是随机IO，顺序读是顺序IO
- 优点
  - **写性能大幅度提高**
  - 不受**SSD随机写入放大**干扰
  - 不受**空间放大**干扰
- 缺点
  - **读性能有牺牲**，IO次数比B+树大
  - 需要定期compaction，对整体**网络/磁盘IO存在放大**



HBase BulkLoad 过程

(MapReduce可以使用Hadoop、Spark等，也可以用本地单机方式)

HBase BulkLoad简介

将需导入的数据，按照HFile格式(HBase中数据文件格式)存储在HDFS，然后以**文件方式加载**到HBase的Regionserver中，对外提供访问。

HBase BulkLoad优点

1. **高效的导入效率**，是普通API的**十倍至百倍**以上
2. 导入过程，对HBase**在线服务**几乎**无影响**
3. 数据可以在很小时间(秒级)窗口，**批量生效**（同时对外可读）

HBase BulkLoad原理

这是由LSM + 存储计算分离(使用HDFS)设计催生的独有功能，前者(LSM)可以保证数据能够以文件的方式灌入数据库，后者(HDFS)可以提供数据文件的共享存储。该特性在大数据量导入场景拥有非常大的优势

由引擎结构(B+ Tree vs LSM Tree)看到的能力差异：

1. MySQL：读写均衡、存在空间碎片
2. HBase：**侧重于写**、**存储紧凑**无浪费、IO放大、数据导入能力强

功能差异

数据访问

姓名	年龄	职业	性别	城市	年消费	支付宝等级
小王	20			杭州		
小张		程序员	女		10w	钻石
小李				上海		

相同之处：

数据以表的模型进行逻辑组织，应用对数据进行增删改查

不同之处：

MySQL的SQL功能更丰富；事务能力更强

HBase既可以用API进行更灵活、性能更好的访问，也可以借助Phoenix使用标准SQL访问；只支持单行事务

HBase的特色功能--TTL

数据的生存时间（TTL）：

1. 对于每一行中的每一列数据，系统都保存了其最后更新时间
2. 以秒为单位来设置 TTL（Time To Live）长度，每一行中的每一列**数据**一旦达到**到期**时间，将被**自动删除**

Metric Name	Date	Value
Cpu_user	2019-5-1 16:00:00	0.05
Cpu_user	2019-5-1 16:10:00	0.1
Load_one	2019-5-2 15:00:00	1.1
Load_one	2019-5-3 16:00:00	1.3
Load_one	2019-5-3 16:10:00	1.5

若TTL为一个月

2019-6-1 16:00:00 过期

2019-6-1 16:10:00 过期

2019-6-2 15:00:00 过期

2019-6-3 16:00:00 过期

2019-6-3 16:10:00 过期

适用于无需永久保存的数据，如：日志、监控、轨迹、浏览记录、费用详单等
可提升性能、易于开发

HBase的特色功能—动态列

动态列(Dynamic Column)：

1. **无需定义**表中的**字段**，可以直接往表中插入新字段
2. **单表**的**字段**数可以**达到百万**个
3. **空字段不占用存储**空间

姓名	年龄	职业	性别	城市	年消费	支付宝等级
小王	20			杭州		
小张		程序员	女		10w	钻石
小李				上海		

如： 若要新增字段“优酷会员级别”， 无需任何操作，新插入的行数据直接带该字段的值就行。 对于存量数据，该字段为空。

该图片来自互联网

刻画用户的上千维度

适用于**表结构经常调整或字段数非常多的数据**，如**用户画像**、**安全风控**场景，需要**维护**对人、设备的静动态、行为统计等**大量维度**，并且需要**频繁变更字段**来准确衡量各个维度的数据价值

HBase的特色功能—多版本

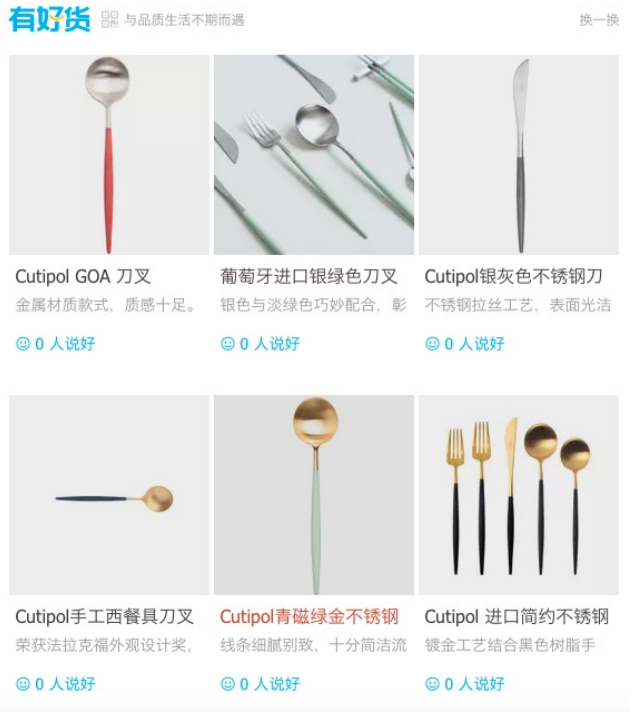
多版本(Multi Versions)：

1. 对于任何一个字段，当数据更新后，其**旧版本**仍然可以**被访问**
2. 可被访问的**旧版本数量**可设定，**可多达上百万**
3. 旧版本数量超过限制后，**自动删除**多余的旧版本

设备ID	上次登陆时间	上次登陆城市
1001	2019-5-9 16:00	杭州
	2019-5-1 11:00	北京
	2019-4-9 12:00	杭州
	…	…

如： 若保留版本数为100，则系统中会为设备1001的”上次登陆时间”、”上次登陆城市”字段最多保留100条记录，超出部分自动删除，等价于**维护最近100次的登陆时间和城市**

适用于需要**维护最近N次变更值的数据**，如浏览记录、轨迹记录、登陆记录、交易记录等，这些记录常用于**实时推荐**、**安全风控**、**营销圈人**等场景



基于商品浏览
推荐好货

HBase的特色功能—多列簇

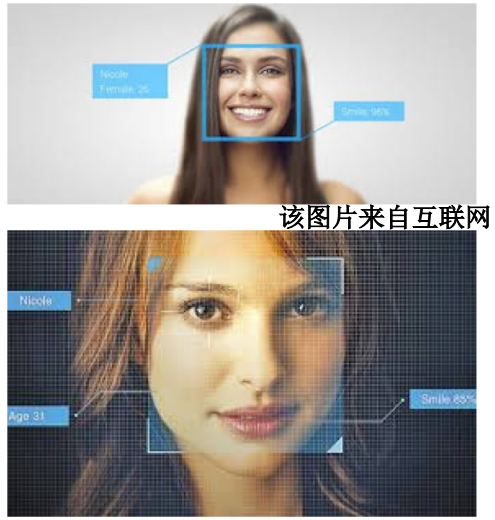
多列簇(Column Family)：

1. 列簇即把一些列在物理存储上放在一起
2. 不同列簇的数据一定存储在不同的物理文件
3. 可减少非相关列在查询时的性能影响
4. 可增加类似列在存储时的压缩率

网页域名	列簇(Content)	列簇(Metadata)		
	网页内容	大小	类别	更新时间
com.cnn.www	<html …>…</html>	20480	新闻	2019-05-01
com.taobao.www	<html…淘！我喜欢….	147456	购物	2019-05-09

如：网页内容一个列簇，其Value值一般较大10KB+; 网页的元数据在另一个列簇; 不同列簇，既能在逻辑上保持统一访问，又能在性能/存储上保持高效

适用于**不同列的大小、访问频率有较明显差异的数据**，如网页、图片、音频、文档等数据的内容与元数据，常用于**搜索**、**人脸识别**、**多媒体**等场景



该图片来自互联网

人脸图片及其元数据统一存储

HBase的特色功能—MOB

MOB(Medium Object)：

1. 适用于存储中等大小(100KB-10MB)数据的特性，如图片、文档、PDF、小视频等
2. 使用透明，与普通访问方式一致，支持所有

使用HBase统一存储各个大小的数据

<100KB

默认

姓名	年龄	职业
小王	20	
小张		程序员
小李		教师

100KB~10MB

MOB

> 10MB

存到HDFS

与传统方案的对比

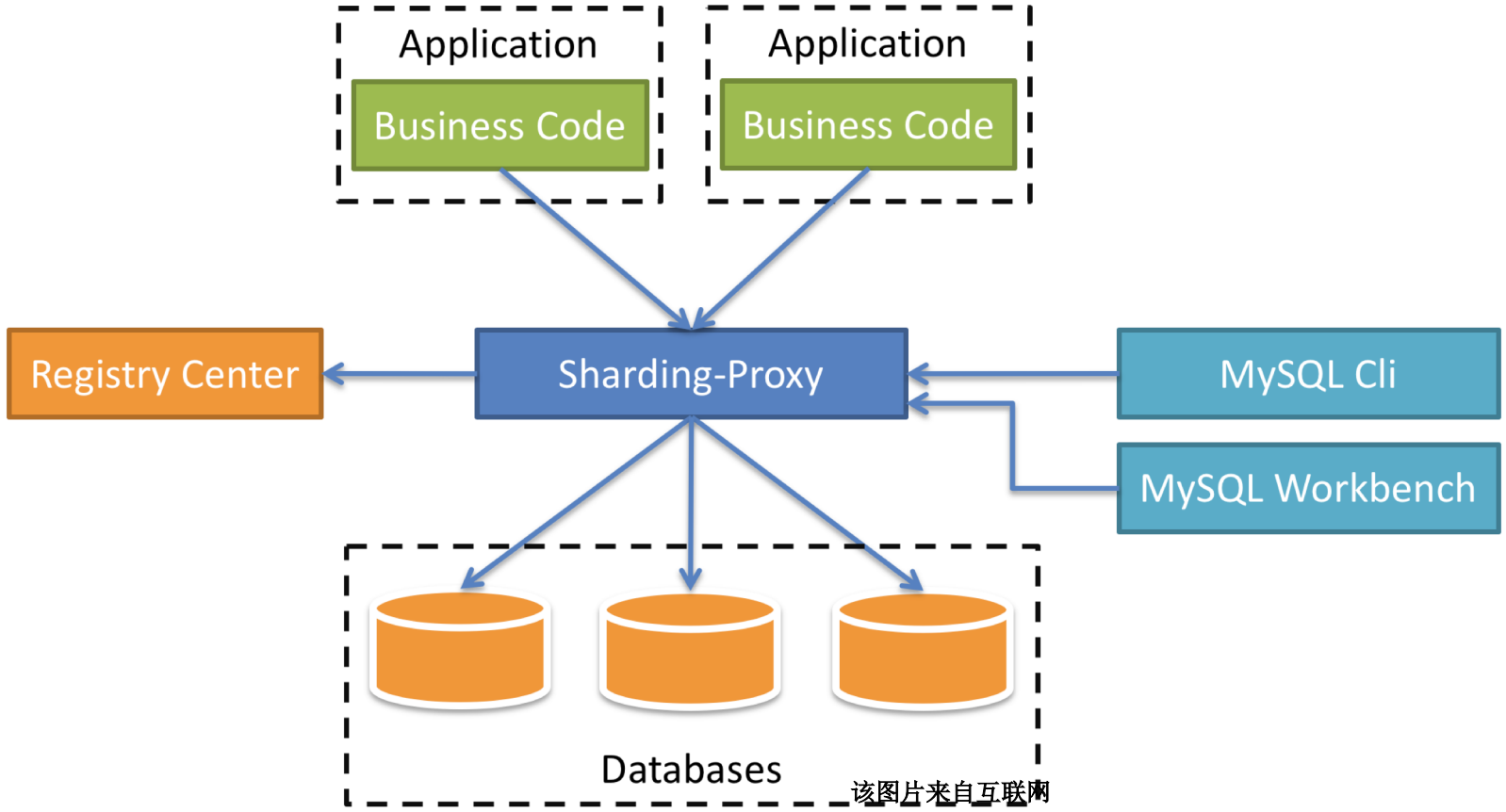
(文件元数据存到MySQL+文件内容存到对象存储)

	MySQL+对象存储	HBase MOB
读写强一致	NO	Y
查询能力	强	强
查询响应时间	高	低
运维成本	高	低
水平扩展	Y	Y

MOB适用于网页、图片、音频、文档等数据的高效存储，与多列簇特性配合，常用于**搜索**、**人脸识别**、**多媒体**等场景

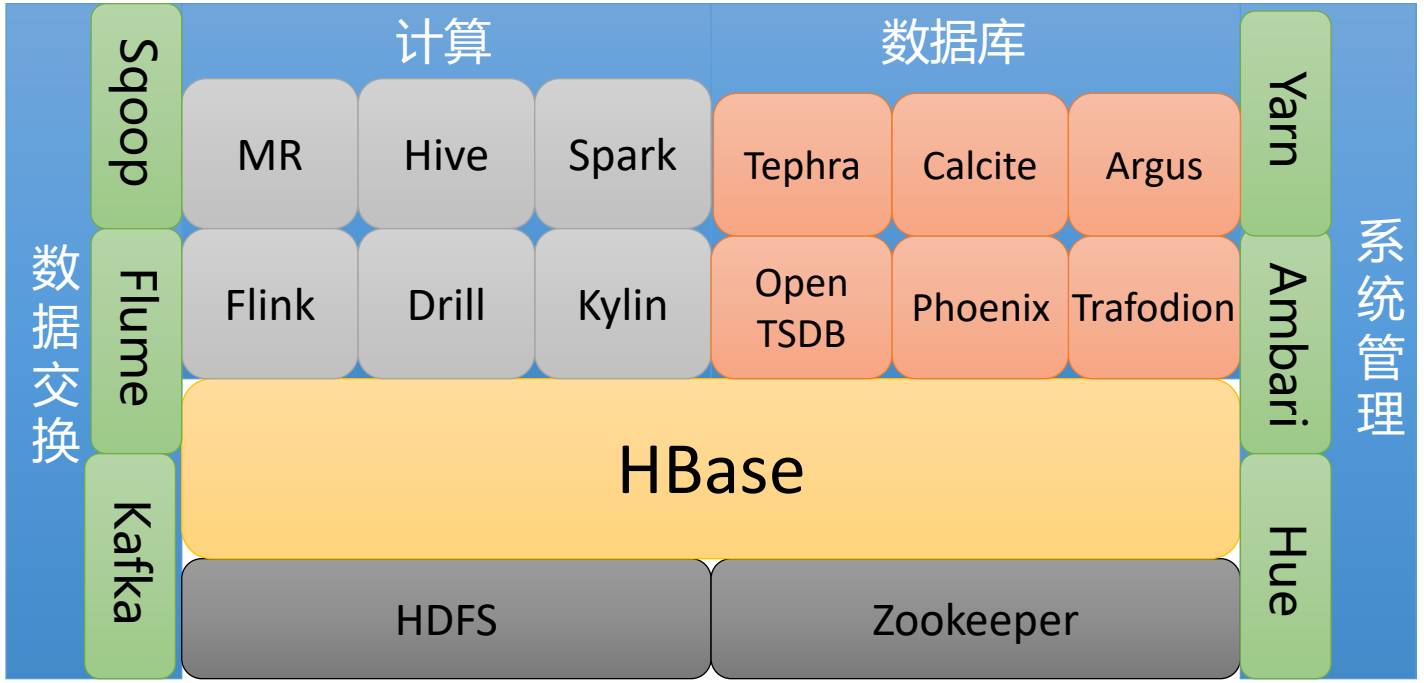
从生态看差异

MySQL的生态



MySQL： 满足APP的在线数据库存储，一般有我足矣

HBase的生态



大数据圈：应用于大数据场景的存储、计算及管理组件

MySQL：

一般可独立满足在线应用的数据存储需求，或者与少量组件配合(如缓存、分库中间件)

HBase：

一般需要和较多大数据组件一起配合完成应用场景，场景架构的设计、实施存在较大的挑战

总结

侧重于在线应用

事务

丰富SQL

运维简单

MySQL

4个9 SLA

集中式架构

B+ Tree

低延迟

读写均衡

侧重于大数据量场景

存储成本低

Schema Free

写能力突出

扩展性好

APACHE HBASE

TTL

存储计算分离

分布式架构

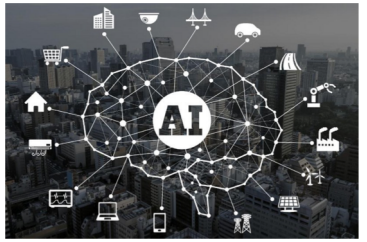
多版本

LSM Tree

容错恢复

数据冗余

数据导入强



哪些场景的存储适合HBase？

用户画像

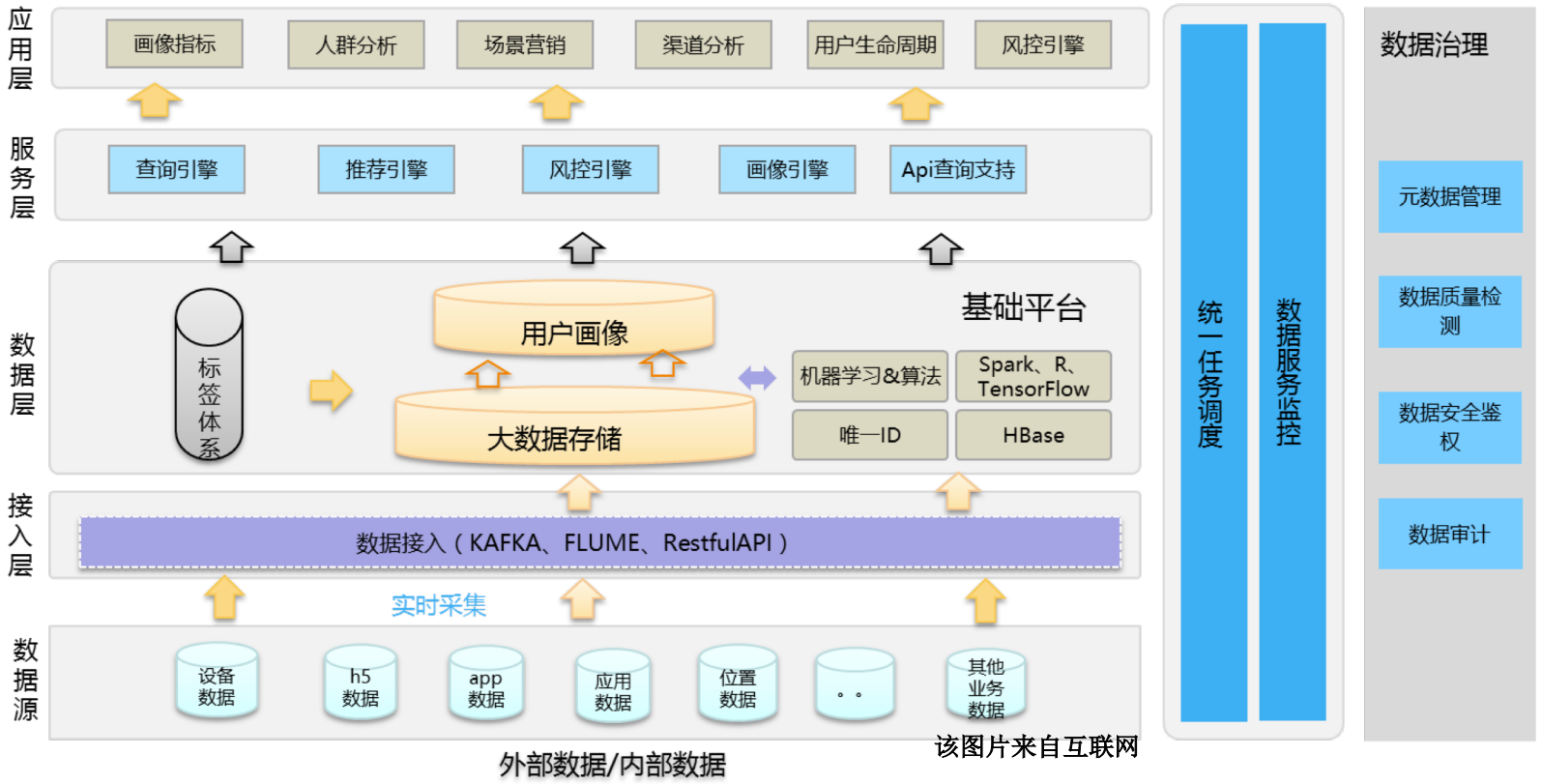
用户行为数据

数据挖掘

该图片来自互联网

用户画像是很多以数据为驱动应用的基础，
如推荐、风控、搜索、智能客服等等

用户画像的参考架构



HBase适合画像存储的几个优势：

① **动态列**
- 满足画像中的巨多、稀疏、灵活的维度

② **BulkLoad导入**
- 满足离线产出的海量画像数据**高效导入**
- 满足导入过程对**在线服务的零影响**
- 数据在短时间(秒) 内，**批量生效**(同时可读)

③ **低存储成本**
- 支持HDD、OSS等便宜的存储介质
- 压缩率高

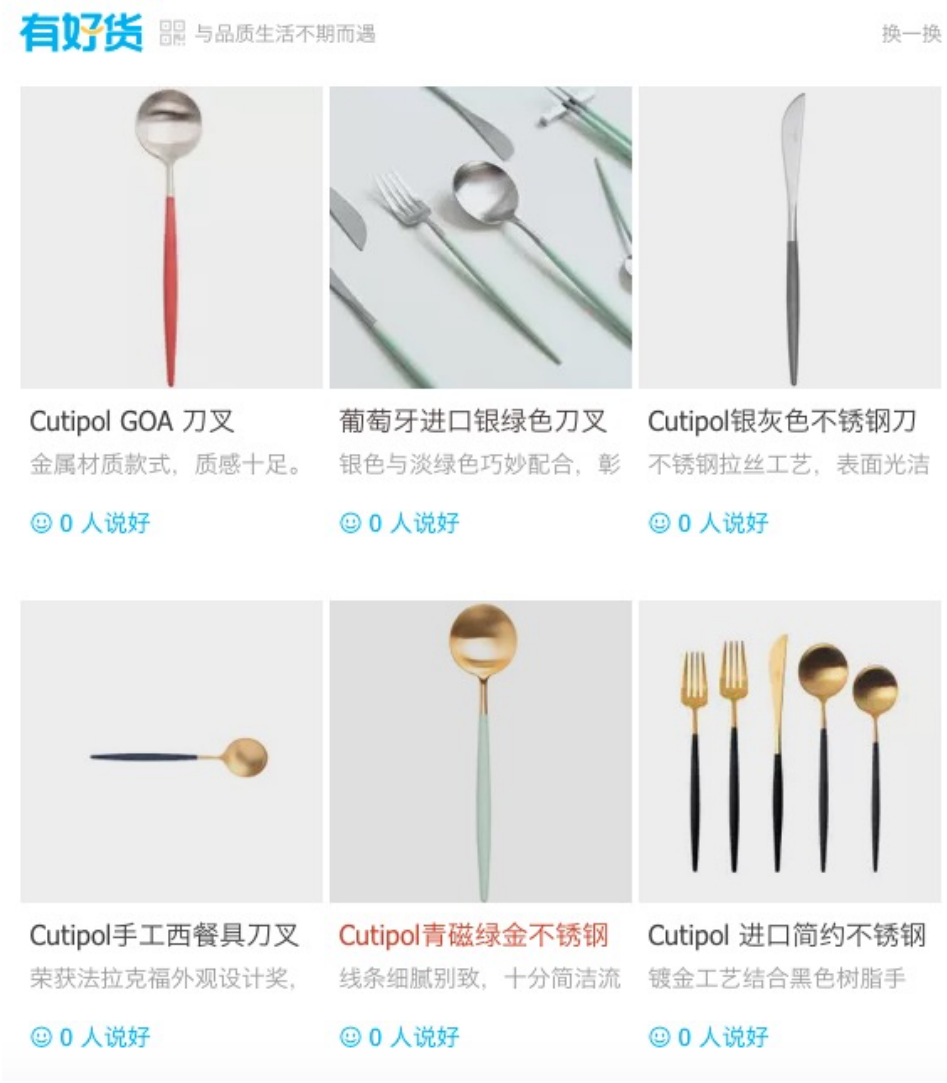
④ **高扩展性**
- 满足数据的大规模增长，提升画像数据的价值

⑤ **在线响应**
- 访问延迟在ms级，满足应用的在线调用

⑥ **与Spark、Hadoop的集成性好**

1. 使用HBase**存储**用户的**行为数据**，服务于Spark/Hadoop的计算
2. 使用HBase**存储**用户的**画像数据**，服务于在线业务查询及分析

实时推荐



B端服务与内容推荐

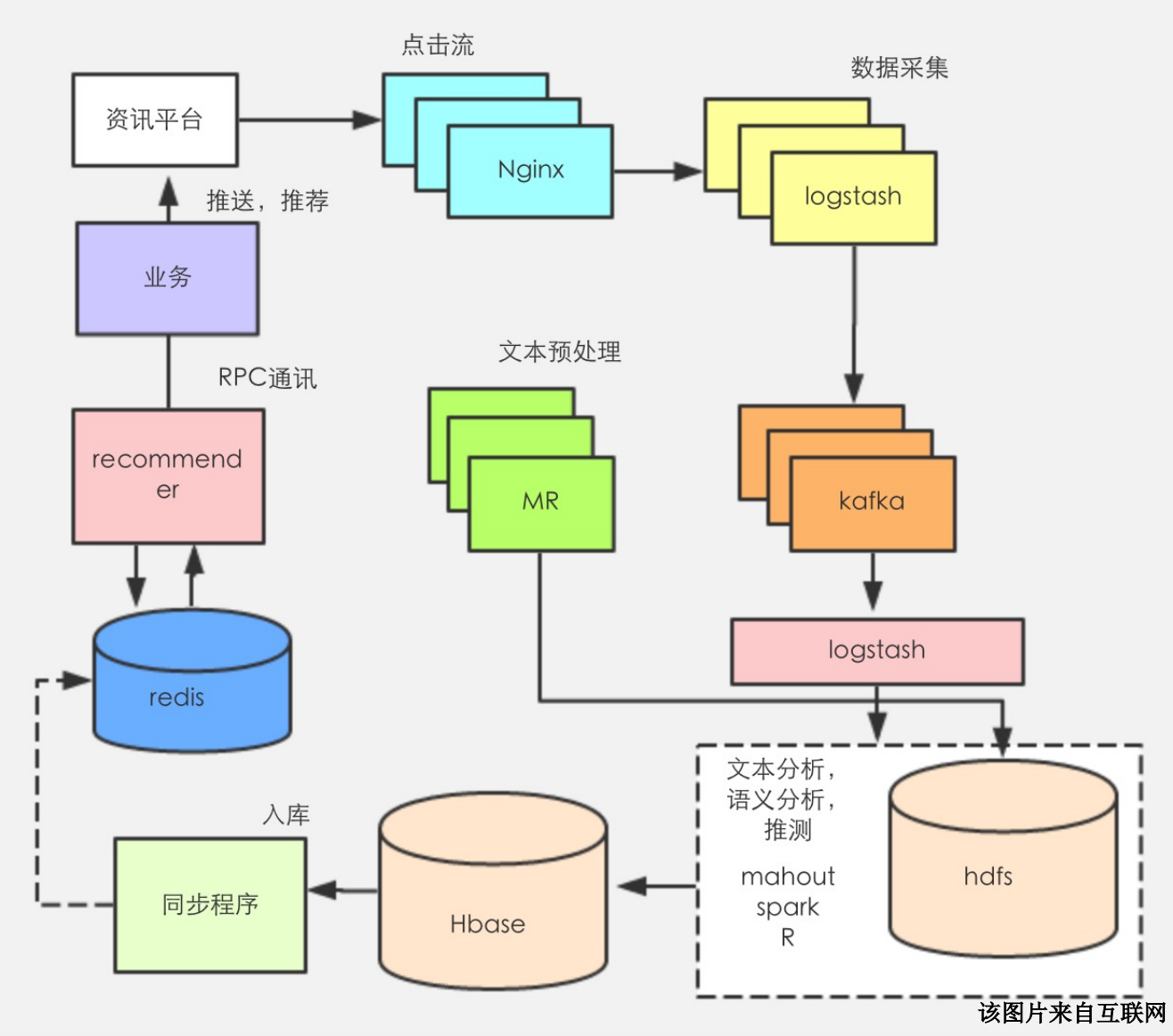
C端服务推荐：应用中心-限时推广与城市服务

C端权益推荐：惠支付

C端内容推荐：汇生活

推荐带动业务增长

实时推荐的参考架构



该图片来自互联网

1. 将Spark/Hadoop**离线计算的推荐结果，导入到HBase**，服务于业务应用
2. 将Spark Streaming/Flink**实时计算的推荐结果，写入到HBase**，从而对推荐效果进行实时调整，服务于业务应用
3. 将实时**行为数据/维表数据写入到HBase存储**，供计算使用

HBase适合推荐场景的存储的几个优势：

① **BulkLoad导入**
- 满足离线产出的推荐结果数据**高效导入**
- 满足导入过程对**在线服务的零影响**
- 数据在短时间(秒) 内，**批量生效**(同时可读)

② **低存储成本**
- 支持HDD、OSS等便宜的存储介质
- 压缩率高

③ **数据生命周期**
- 满足实时行为数据、推荐结果数据的**过期自动淘汰**

④ **高吞吐**
- 满足实时数据及计算结果的大量写入

⑤ **在线响应**
- 访问延迟在ms级，满足应用的在线调用

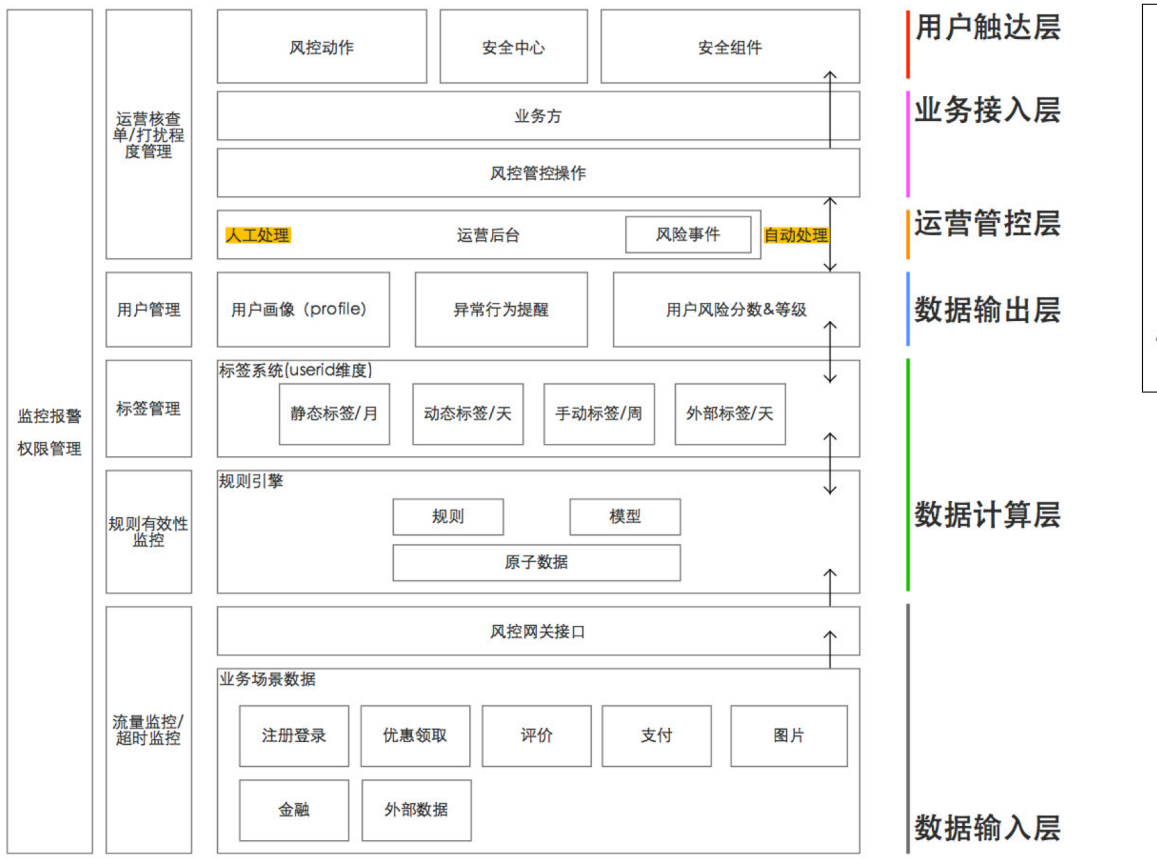
⑥ **与Spark、Hadoop的集成性好**

实时风控

本页图片来自互联网

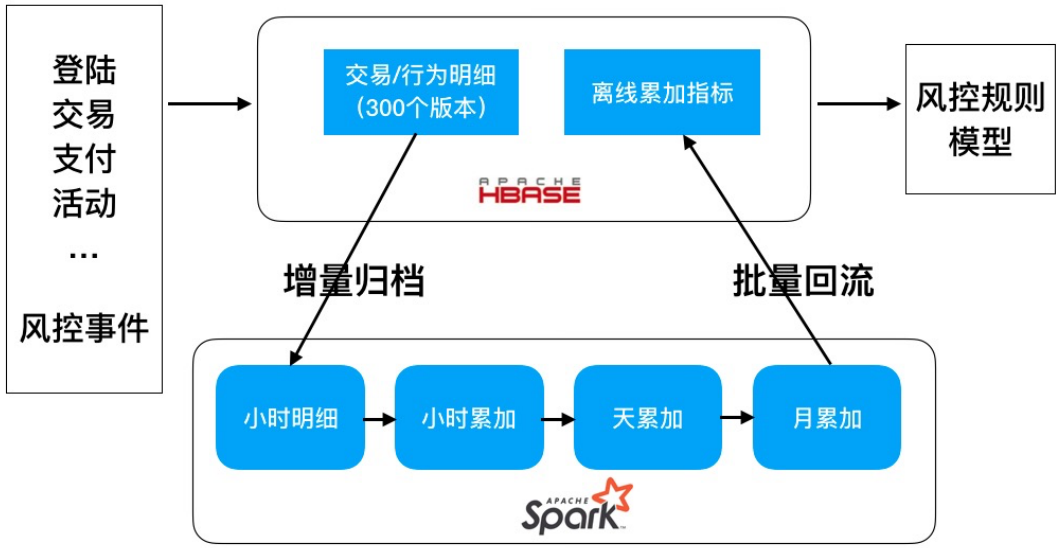
看不见的安全保护

实时风控的参考架构



该图片来自互联网

1. 用HBase存储风控的事件数据，提供给上游计算层
2. 用HBase存储设备、指纹、画像等信息，提供给计算层、应用层
3. 用HBase存储各维度的**指标统计数据**，由计算层写入，供应用层实时调用，做风险决策



HBase适合实时风控存储的几个优势：

① 多版本
- 高效维护最近N次登陆、支付、浏览等事件数据

② BulkLoad导入
- 满足离线产出的指标统计数据**高效导入**
- 满足导入过程对**在线服务的零影响**
- 数据在短时间(秒) 内，**批量生效**(同时可读)

③ 动态列
- 满足风控指标的巨多、稀疏、灵活的特点

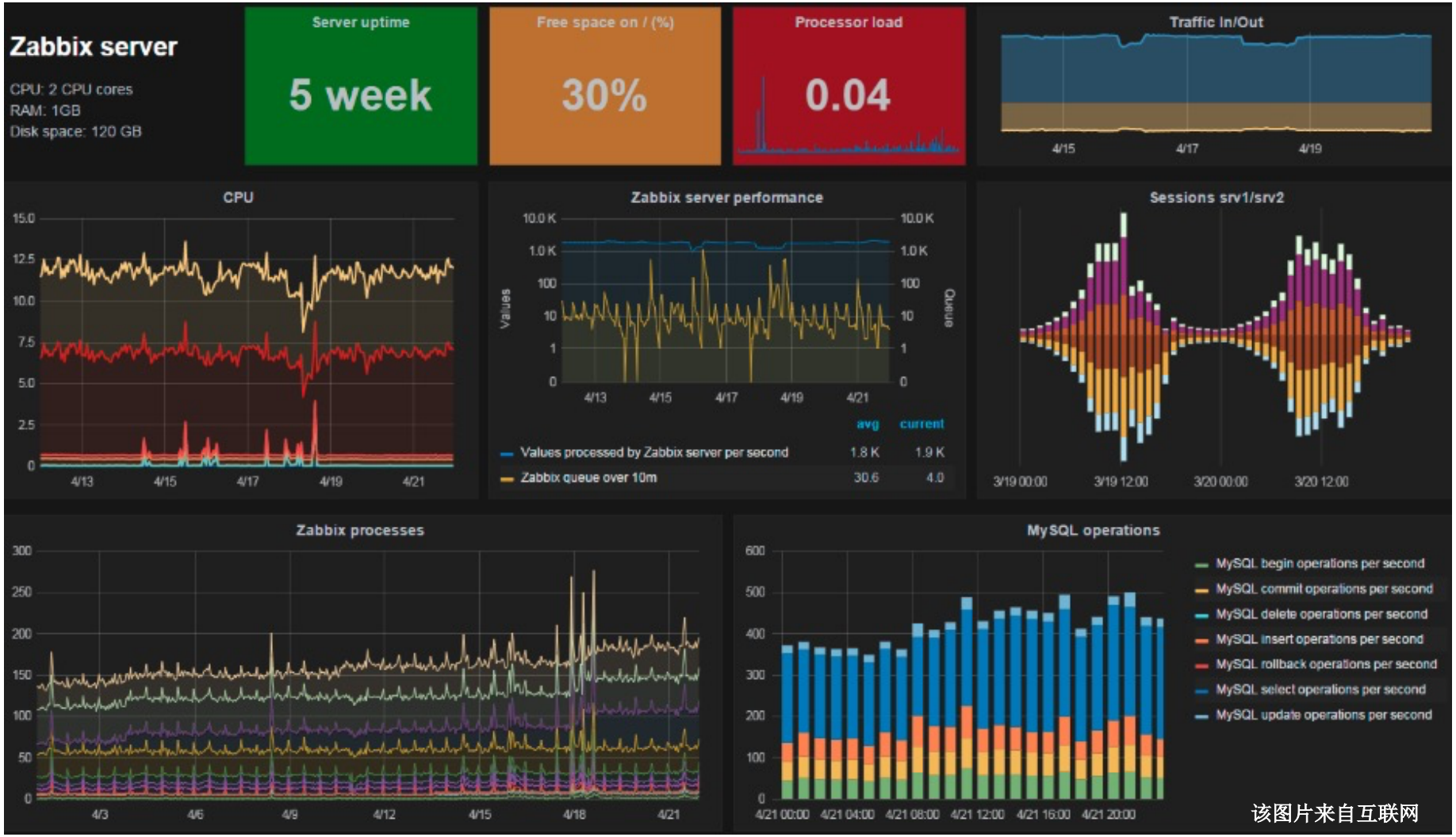
④ 数据生命周期
- 满足事件数据、指标统计数据的**过期自动淘汰**

⑤ 低存储成本、高吞吐

⑥ 在线响应

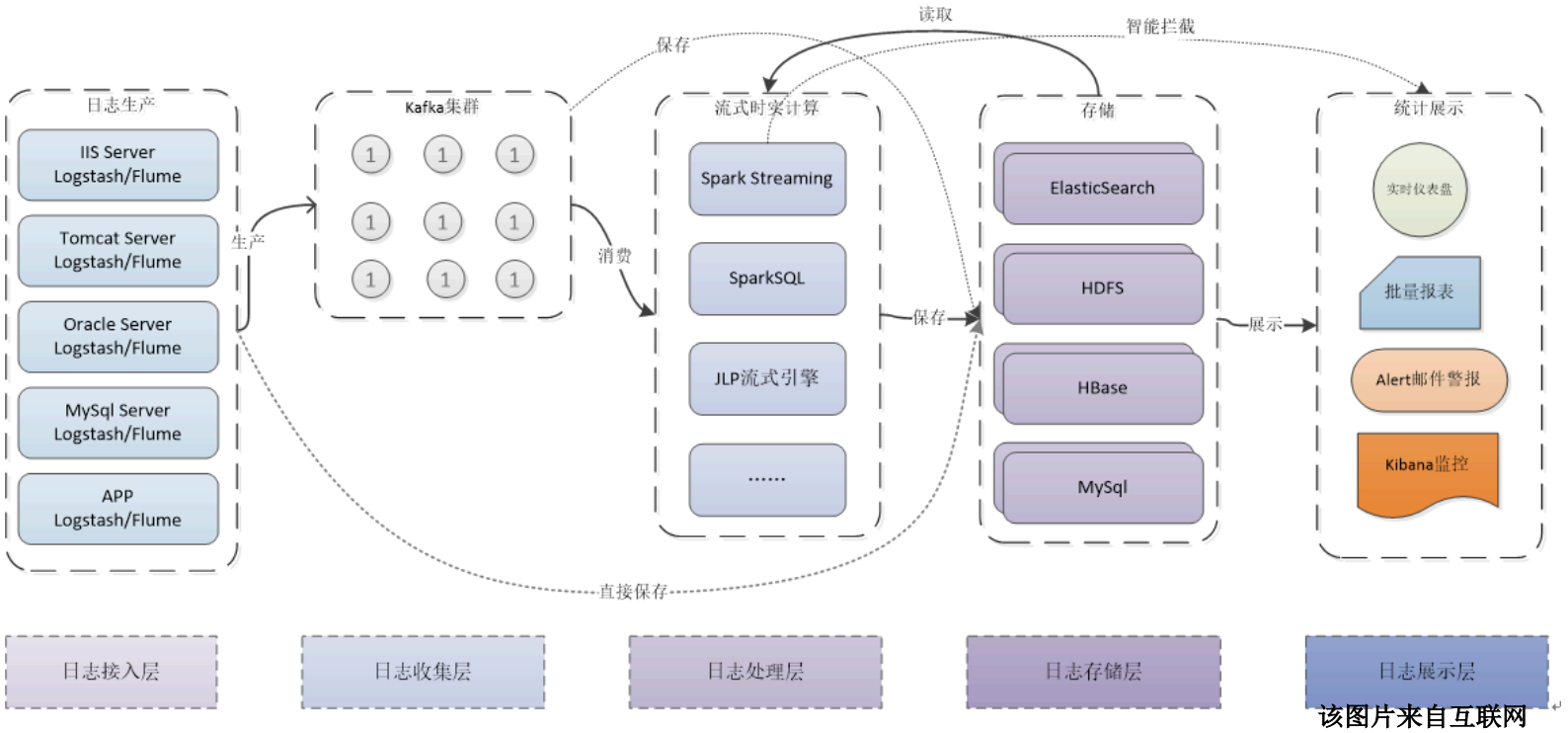
⑦ 与Spark、Hadoop的集成性好

监控系统



该图片来自互联网

监控系统的参考架构



该图片来自互联网

HBase适合监控系统存储的几个优势：

① **数据生命周期(TTL)**
- 满足原始Metrics数据的**过期自动淘汰**

② **高并发写入**
- LSM结构大大优化随机写性能

③ **高效范围查询**
- LSM结构下近期、历史数据存在分层，满足近期数据的高效范围查询

④ **低存储成本**

⑤ **在线响应**

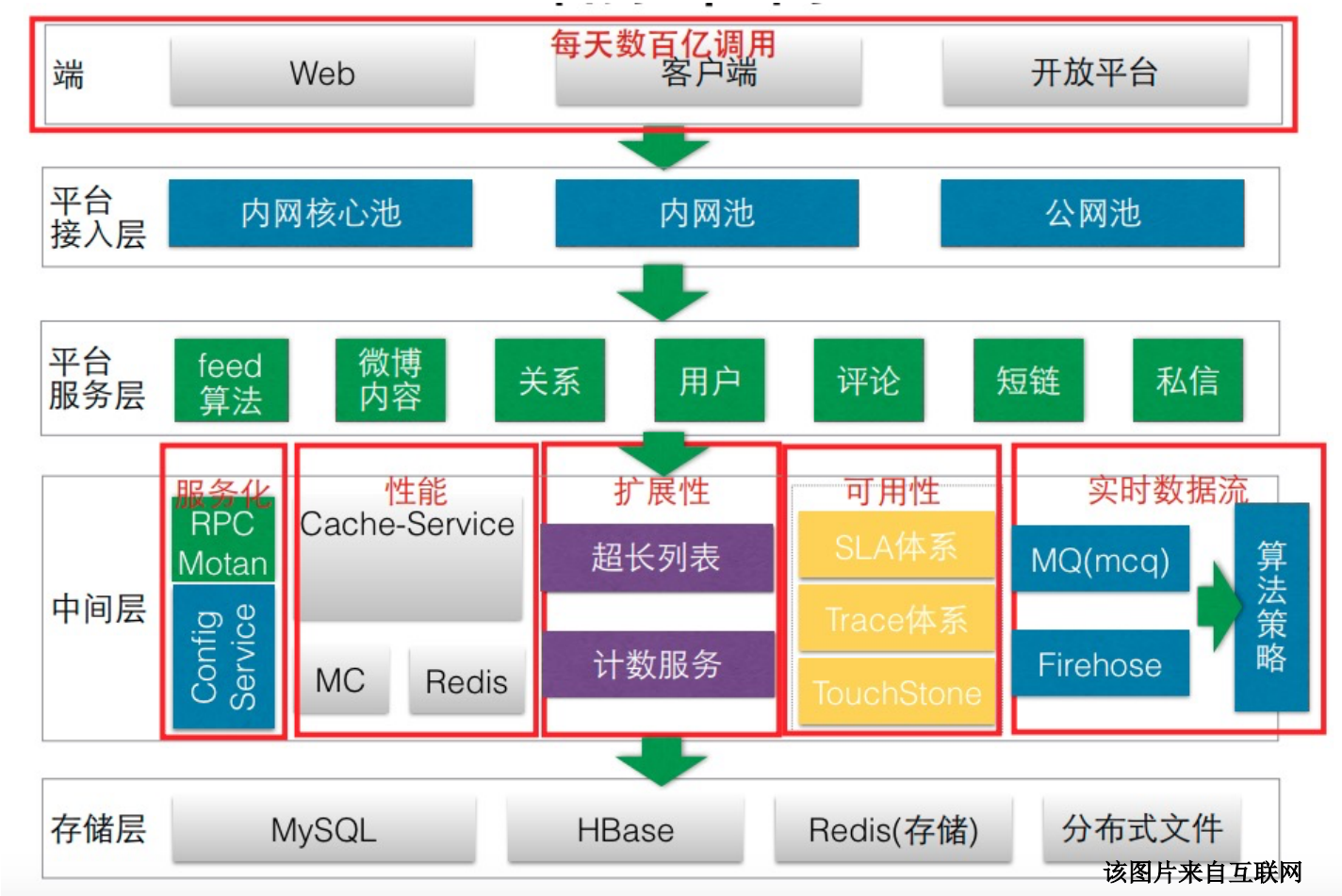
⑥ **与Spark、Hadoop的集成性好**

1. 用HBase**存储**监控的原始Metrics数据，提供给应用展示、计算分析
2. 用HBase**存储**监控的聚合结果数据，由离线计算或实时计算写入

社交Feed流

本页图片来自互联网

社交Feed流的参考架构



HBase适合社交Feed流存储的几个优势：

① **高并发写入**
- Feed流推模式(写扩散)，写入放大；LSM结构、batch操作使得HBase的写能力非常优秀

② **MOB特性**
- 满足长文本、图片、小视频等几百KB至几MB的中等对象存储

③ **在线响应**
- 访问延迟在ms级，满足应用的在线调用

④ **低存储成本**

1. 用HBase存储社交中的短文本数据
2. 用HBase存储Feed流中的列表数据
3. 用HBase存储社交中的长文本、图片、小视频等数据

The More

汽车轨迹

BI报表

搜索系统

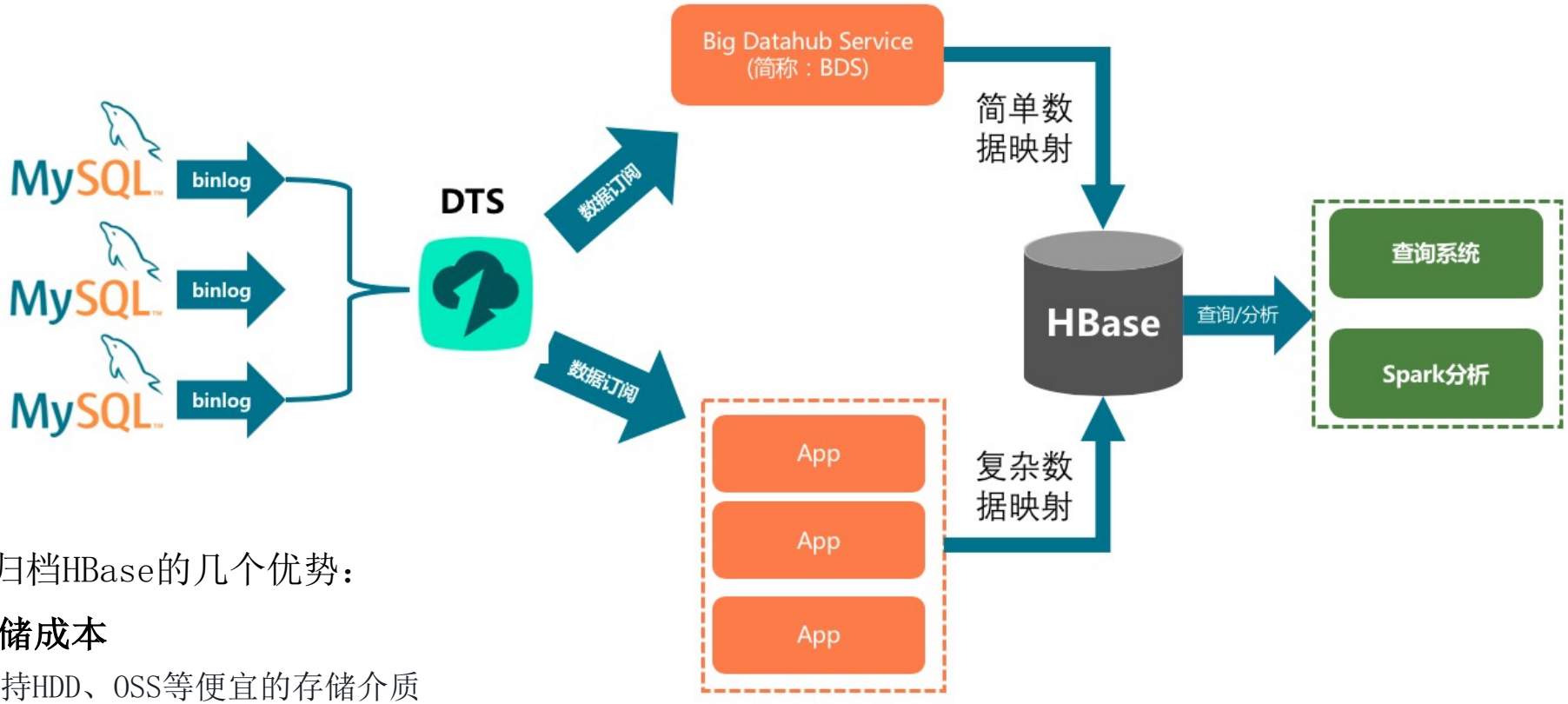
IM系统

以上场景均适合使用HBase存储

HBase不是MySQL的替换

HBase是业务规模及场景扩张后，对MySQL的自然延伸

MySQL数据归档HBase



MySQL数据归档HBase的几个优势：

- **降低存储成本**
  - 支持HDD、OSS等便宜的存储介质
  - 压缩率高
- **提升在线库性能及变更效率**
  - 更少的数据，更短的DDL时间，更低的业务影响
  - 更低的索引层级，更高的读写性能，更好的用户体验
- **充分利用全局/全量历史数据进行数据挖掘**
  - 例如：分析交易流水，作为会员运营的输入，是否场景活跃，打新&防止流失。

HBase助您畅游大数据
欢迎您选择阿里云HBase

云数据库 HBase 版介绍
云HBase正在免费试用，欢迎选购 >>

欢迎加入HBase+Spark技术交流钉钉群：
https://dwz.cn/Fvqv066s

阿里云HBase对比开源自建的八大优势

5 倍

综合生产性能

1/3

平均存储成本

稳

九年双十一考验

全

运维效能提升

多重

企业级安全保护

丝滑

生态体验

强

技术专家团

1 小时

在线透明搬迁

阿里云HBase X-Pack

低成本、云原生、一站式的数据处理平台

HBase X-Pack基于Apache HBase + Spark 深度扩展，融合Phoenix、Solr等技术，支持海量数据的一站式存储、检索、分析，历经阿里巴巴近十年的大规模锤炼，被广泛用于风控、推荐、搜索、画像、社交、物联网、人工智能、对象存储等场景，助力企业数据智能化

低成本

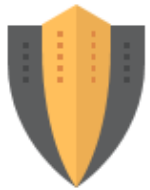
内置自主可控的**高性能引擎**，最高达开源版的**5X吞吐**，拥有冷热分层、异构存储、高压缩算法、共享文件系统等**多项成本节省的核心技术**，结合按需收费、弹性伸缩的云原生能力，大幅优化您的总体拥有成本。

弹性扩展

支持GB至PB、数百至千万并发的**平滑弹性伸缩**，系统采用**完全分布式**架构，具备数据表自动分区、在线分裂、秒级扩容等能力，并拥有热点识别及加速、级联分裂、透明加盐等**多项去倾斜的核心技术**，让系统永无单点瓶颈。

低延迟

千万级并发下保持请求的**单个毫秒响应**，在RPC、内存管理、缓存结构、日志写入等方面深度优化，保障读写路径的无锁实现。同时，基于PACELC理论构建的多副本冗余并发能力，让系统的毛刺率减少一个数量级。



高可用

内置多副本复制和自动容灾技术，面对节点或机房的故障，无需应用任何改造，服务**秒级自动恢复**。同时，支持同城多活、异地多活、三机房强一致容灾等高级功能。

安全

系统支持**网络隔离、白名单、身份认证、权限控制、访问审计**等丰富安全功能，来满足您不同层次的需求

丰富检索

内置的多维索引模块通过LSM、倒排、BKD树等技术支持**主键查询、多字段组合查询、地理位置查询、全文检索、模糊查询**等能力

灵活分析

深度集成Spark、 Spark Streaming/SQL、Phoenix 等大数据技术，一体化**体验交互式、流、批分析**

易开发

支持开发者灵活选择KV、SQL、时序、时空、图、FeedStream等**多模型**，并通过**Native OSS API**开发，如HBase、Phoenix、Gremlin等，提供100%的开源兼容性

轻松运维

一体化的**运维管理平台**，帮助您可视化完成监控报警、容量规划、升级配置、备份迁移、数据查询、问题诊断、性能优化等一系列繁琐操作，让运维更高效、智能与轻松，更有可选的Serverless服务让您可完全免运维。

经过验证

系统经阿里经济体内上万节点的**大规模长久锤炼**，拥有国内在HBase、Spark技术领域最强大的技术团队，保障您在使用过程的**持续稳定、可靠**

© 2019 Alibaba Group. All rights reserved.

谢谢指导

THANKS!