

大数据时代结构化存储云HBase技术架构及最佳实践

原创



VIP文章 普通网友 于 2017-09-23 21:19:28 发布



21



收藏

文章标签: [数据库](#) [大数据](#) [分布式](#) [hadoop](#) [编程语言](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: <https://blog.csdn.net/lveil9911108/article/details/115754576>

版权

在10年, 阿里研究HBase, 是为了解决阿里容量及并发的实际问题, 按照数据库要求, 阿里深入HBase技术, 并致力于保障稳定性和性能, 目前已经有10000台规模, 数百个集群, 大约1亿的QPS, 服务整个集团的业务。17年, 把这部分能力也开放给公有云客户。本文中, 阿里云高级专家封神带来了主题演讲《大数据时代结构化存储云HBase技术架构及最佳实践》, 介绍HBase的应用选择、实战案例、技术平台解读以及后续的规划。

为什么应用HBase

一般而言, 传统关系型数据库面临着成本、容量、QPS、分析等多方面的问题: 存储成本较高; 无法满足TB、PB级别的数量存储需求; QPS无法满足较高的并发要求, 性能不能横向扩展; 数据隔离, 从而不能满足分析类的需求。

通过关系型数据库MySQL, 可以解决中小数据库存储需求; 通过分库分表, 能够解决一定容量及并发的需求, 但是其实现复杂, 需要业务感知; 通过以HBase为代表的分布式数据库, 可以支持高到千万的并发, 满足海量数据的存储。

那么怎么解决传统数据库这些问题呢? HBase给出了相应的应对方法:

LSM-Tree: 写吞吐高, 离线导入效率高;

存储无限扩容: 计算存储分离, 分布式存储可以无限扩容;

自动分区: 分区自动分裂, 分区自动Merge;

Hadoop生态: Phoenix满足查询需求, Spark接HBase, 可以满足分析类需求。