

HBase为什么查询快

可回答：1) HBase为什么读快；2) HBase是根据rowkey查询，当数据量相当大的时候，是怎么读的很快的

参考答案：

1、基于LSM树的存储方式

HBase采用基于LSM树的存储方式，这种存储方式将数据分为内存和磁盘两层存储，内存部分称为MemStore，磁盘部分称为HFile。MemStore存储的数据是有序的，并且每次插入数据时会进行排序和合并，因此可以减少数据的查找和排序开销，提高查询效率。

2、分布式存储架构

HBase采用分布式存储架构，数据可以分散存储在集群中的多台机器上。在查询时，HBase可以利用分布式存储架构进行并行查询，从而提高查询效率。

3、列存储方式

HBase采用列存储方式，同一列族的数据被存储在一起，可以避免不必要的IO操作，提高数据读取效率。

4、布隆过滤器

通过Rowkey定位Region，这当中会先经过BlockCache，这边找不到的话，再经过MemStore和Hfile查询，这当中通过布隆过滤器过滤掉一些不需要查询的HFile。

5、基于Region的负载均衡

HBase会将数据划分为多个Region进行存储，每个Region对应着一段连续的行键。HBase会自动对Region进行负载均衡，保证数据分布均匀，避免热点数据和单点故障。

数据量很大的时候，HBase会拆分成多个Region分配到多台RegionServer。客户端通过meta信息定位到某台RegionServer（也可能是多台），通过Rowkey定位Region，这当中会先从BlockCache（缓存）查找，找不到的话，再从MemStore和HFile查询，这当中通过布隆过滤器过滤掉一些不需要查询的HFile，这样就保证HBase的查询速度。



蓦然
送你一张星球优惠券

¥40

可用于

「旧时光大数据」

2023/06/30 12:00 后失效

前 100 名加入可用
长按二维码立抢优惠 ▶



知识星球