

HBase的一个region由哪些东西组成?

问过的一些公司：快手

参考答案：

Region是HBase集群分布数据的**最小单位**。

Region 是部分数据，所以是所有数据的一个自己，但Region包括完整的行，所以Region 是行为单位表的一个子集。

每个Region 有三个主要要素：

- 它所属于哪张表
- 它所包含的第一行(第一个region 没有首行)
- 他所包含的最后一行(末一个region 没有末行)

当表初写数据时，此时表只有一个Region，当随着数据的增多，Region开始变大，等到它达到限定的阈值大小时，变化把Region分裂为两个大小基本相同的Region，而这个阈值就是StoreFile的设定大小（参数：hbase.hRegion.max.filesize），在第一次分裂Region之前，所有加载的数据都放在原始区域的那台服务器上，随着表的变大，Region的个数也会相应的增加，而Region是HBase集群分布数据的最小单位。

但Region也是由block组成，Region是属于单一的RegionServer，除非这个RegionServer 宕机，或者其它方式挂掉，再或者执行balance时，才可能会将这部分Region的信息转移到其它机器上。

这也就是为什么 Region比较少的时候，导致Region 分配不均，总是分派到少数的节点上，读写并发效果不显著，这就是HBase 读写效率比较低的原因。

欢迎加入知识星球，获取《大数据面试题 V4.0》以及更多大数据开发内容



知识星球

长按扫码领取优惠 ▶

