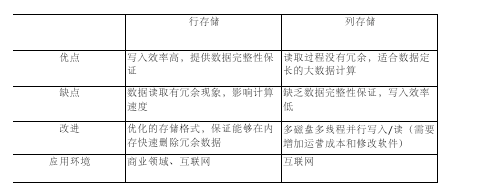
# 列存储与行存储的区别

2018年03月14日 15:24:30

阅读数：527



写入：

行存储的写入是一次完成，数据的完整性因此可以确定。

列存储需要把一行记录拆分成单列保存，写入次数明显比行存储多。

行存储在写入上占有很大的优势

数据修改：

行存储是在指定位置写入一次，列存储是将磁盘定位到多个列上分别写入。

行存储在数据修改也是占优的

数据读取：

行存储通常将一行数据完全读出，如果只需要其中几列数据，就会存在冗余列

列存储每次读取的数据是集合中的一段或者全部。

由于列储存的数据是同质的，这种情况使数据解析变得容易。行存储则复杂的多，因为在一行记录中保存了多种类型的数据，数据解析需要在多种数据类型之间频繁转换，这个操作很消耗cpu

所以列存储的解析过程中更有利于分析大数据

显而易见，两种存储格式都有各自的优缺点：行存储的写入是一次性完成，消耗的时间比列存储少，并且能够保证数据的完整性，缺点是数据读取过程中会产生冗余数据，如果只有少量数据，此影响可以忽略；数量大可能会影响到数据的处理效率。列存储在写入效率、保证数据完整性上都不如行存储，它的优势是在读取过程，不会产生冗余数据，这对数据完整性要求不高的大数据处理领域，比如互联网，犹为重要。

什么时候应该使用行式存储？什么时候应该使用列式存储呢？

如果你大部分时间都是关注整张表的内容，而不是单独某几列，并且所关注的内容是不需要通过任何聚集运算的，那么推荐使用行式存储。原因是重构每一行数据（即解压缩过程）对于HANA来说，是一个不小的负担。

列式存储的话，比如你比较关注的都是某几列的内容，或者有频繁聚集需要的，通过聚集之后进行数据分析的表。