Hadoop学习之Hive编程指南学习(六)

HiveQL 索引

Hive没有RDB中键的概念，而且索引处理模块被设计成可以定制java编码的插件。

当逻辑分区实际上太多而几乎无法使用时，建立索引成为分区的另一个选择。可以帮助裁剪掉一张表的一些数据块，可以减少MapReduce的输入数据量。通过explain可以查看某个查询语句是否用到索引。

* 1. 创建索引
  2. 方法
  3. Dd
  4. Dd
  5. Dd
  6. Dd
  7. D
  8. D
  9. 总是使用压缩

压缩可以使磁盘上存储的数据量变小，这样可以降低I/O来提高查询执行速度。

压缩和解压缩会消耗CPU资源，但MapReduce任务是I/O密集型，非cpu密集型。因此CPU开销不是问题。