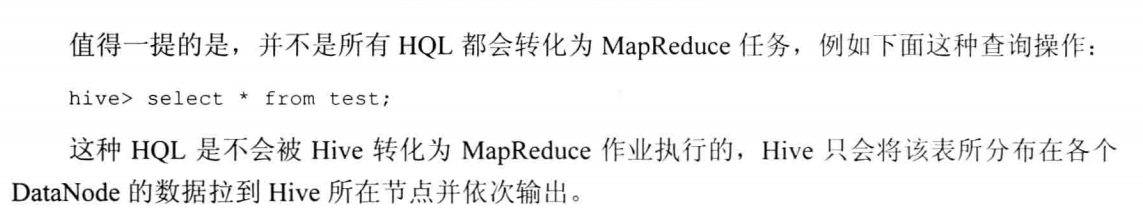
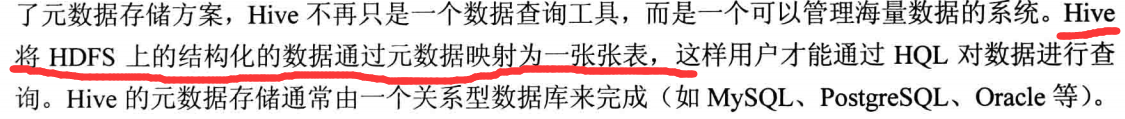
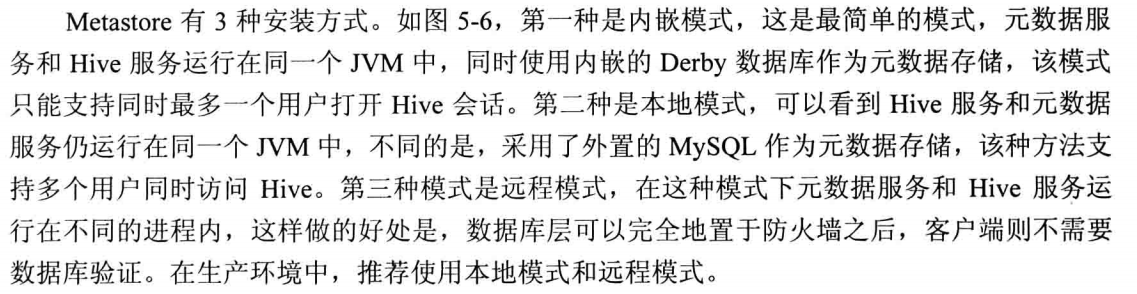
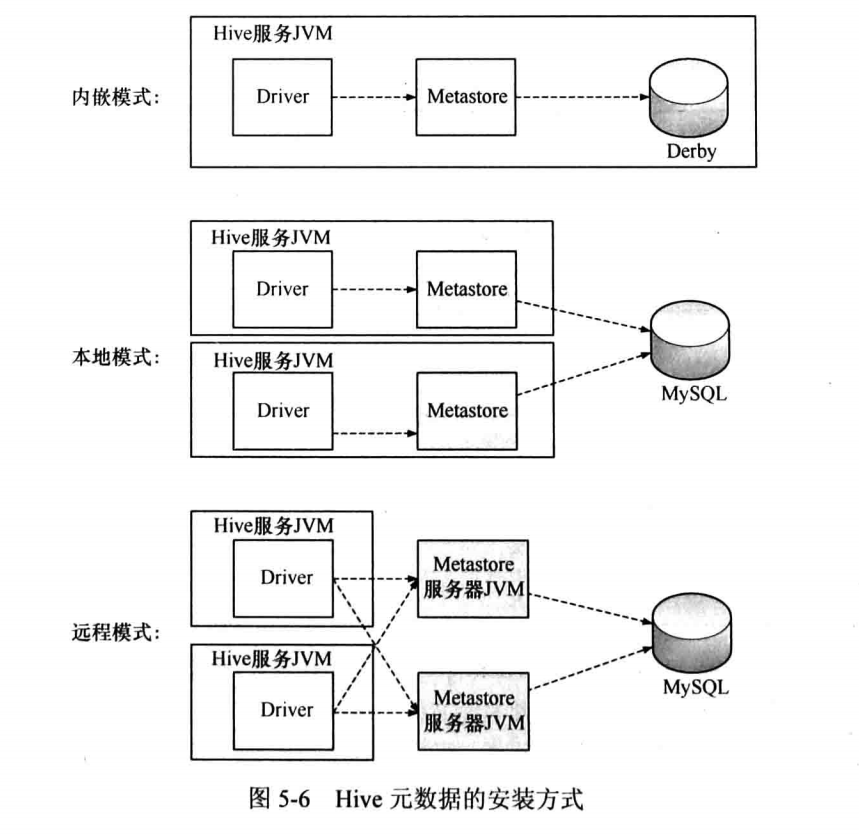
hive与Mapreduce的关系：





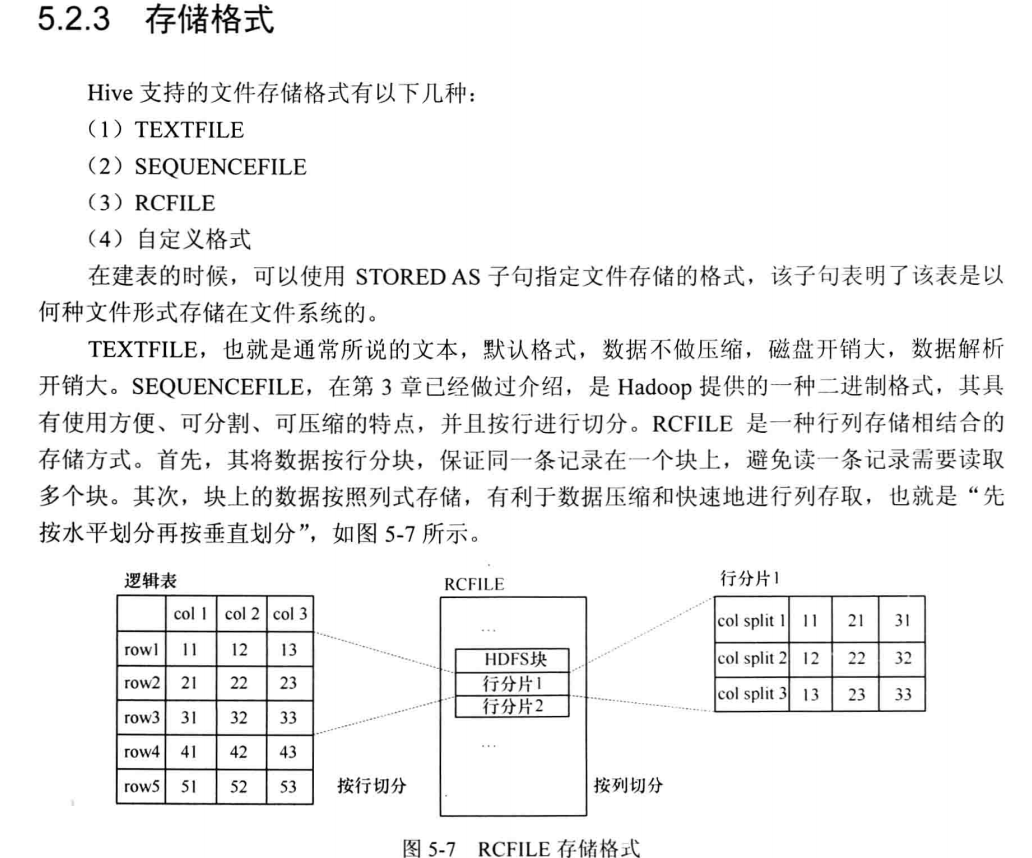
metastore的三种安装模式：



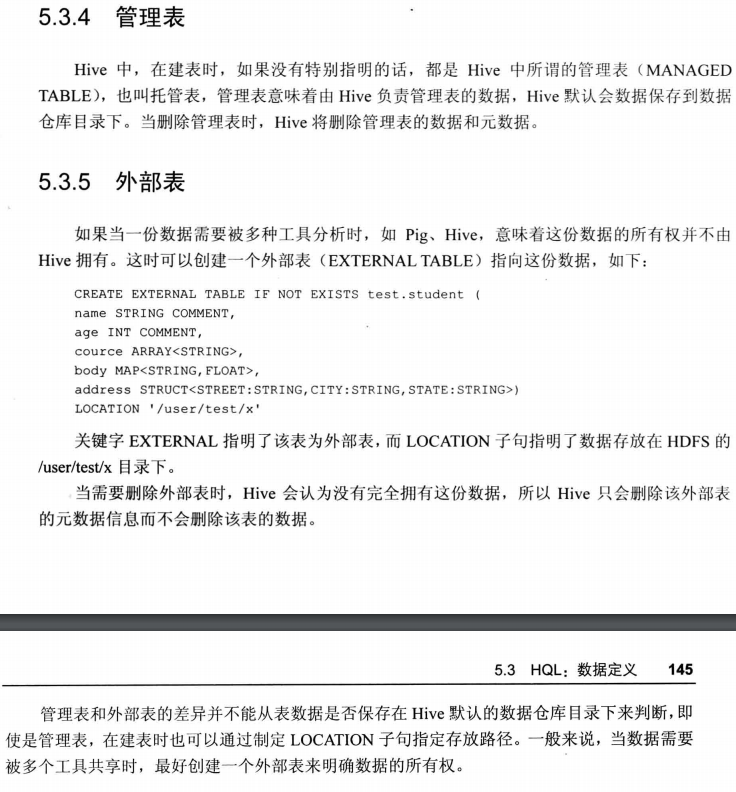


hive与关系型数据库的区别：

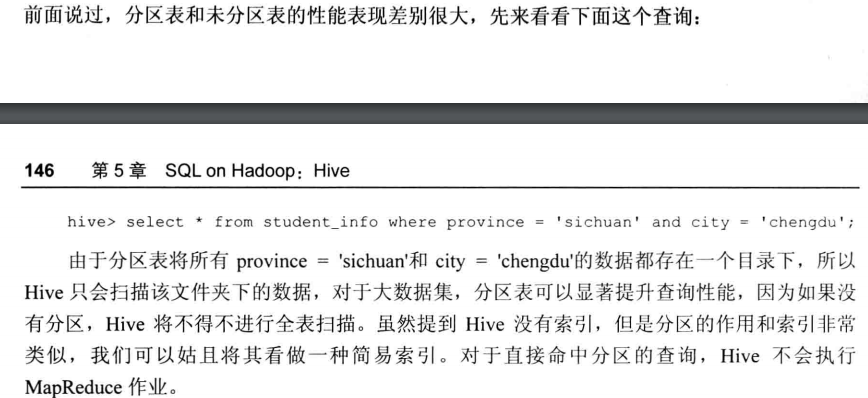


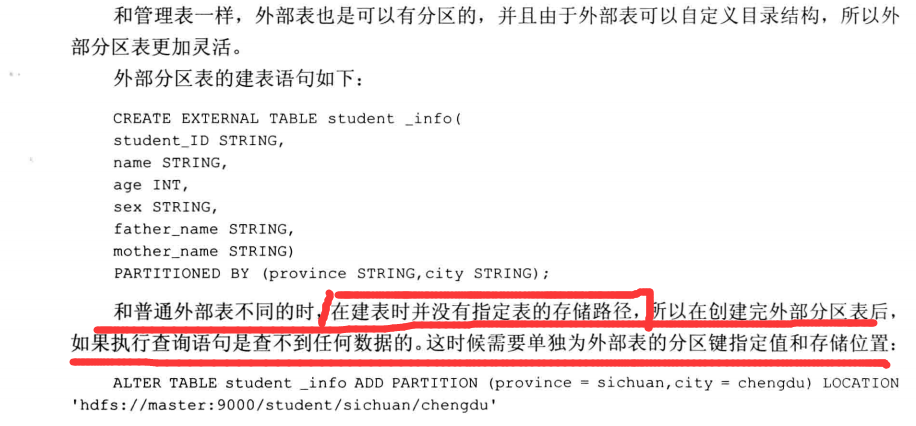


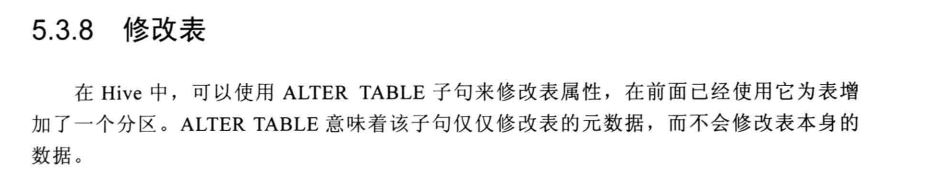


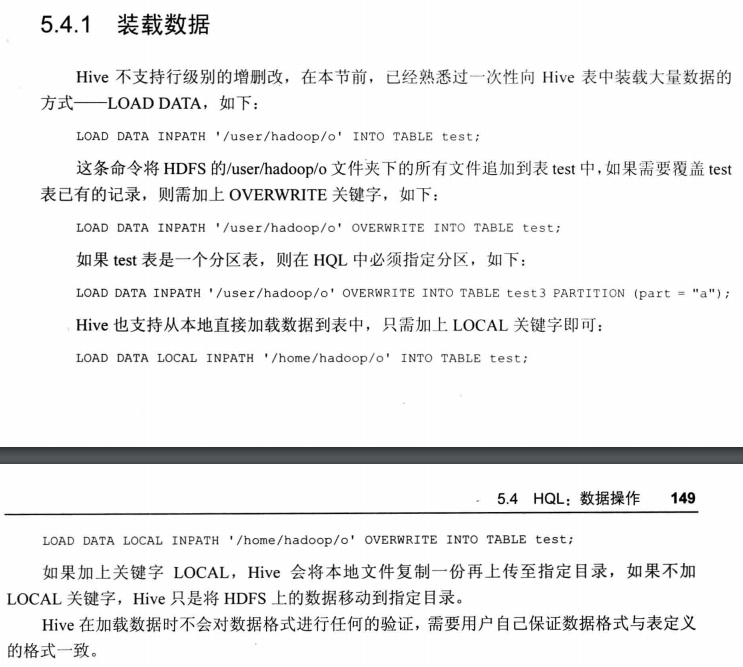


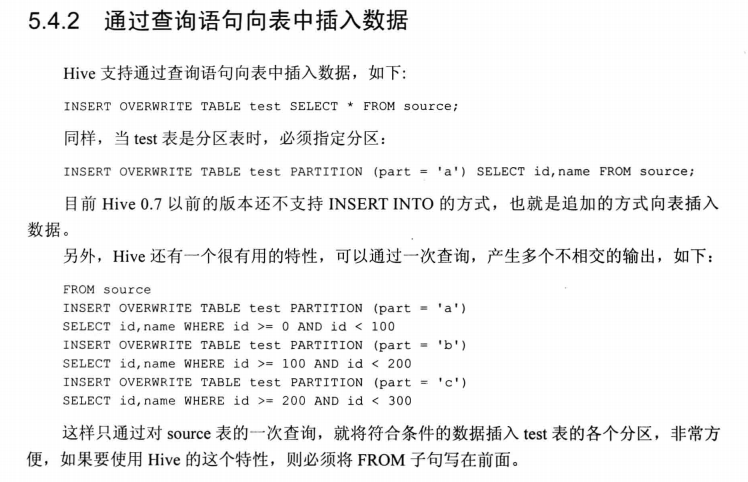
分区：

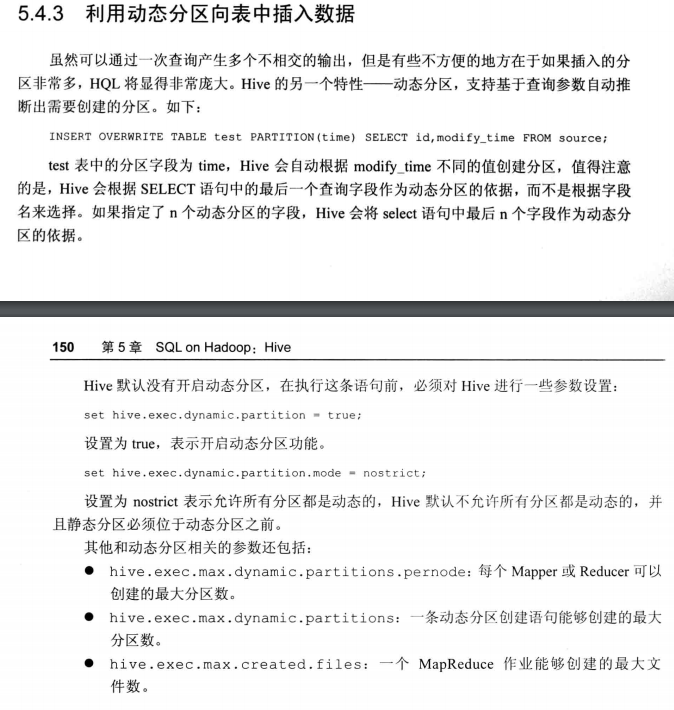




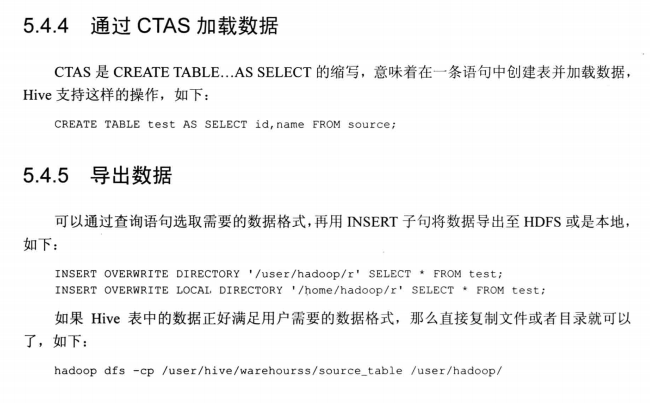




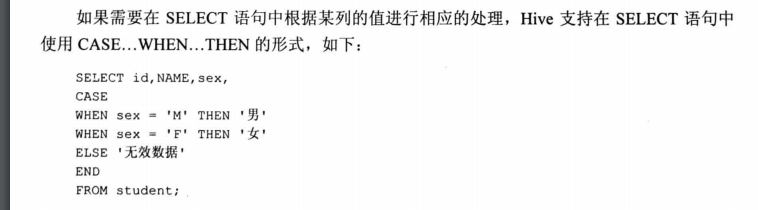


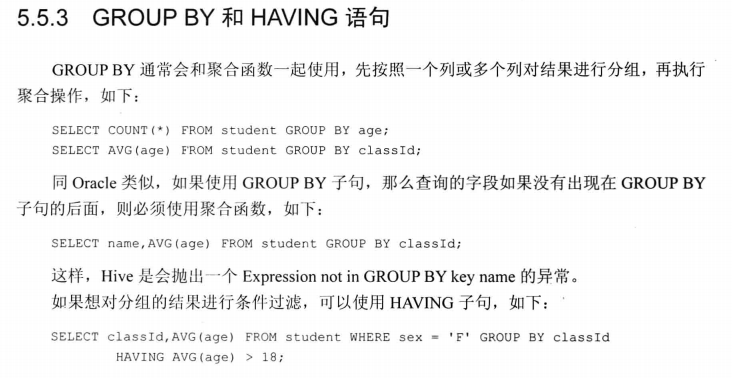


导出数据：

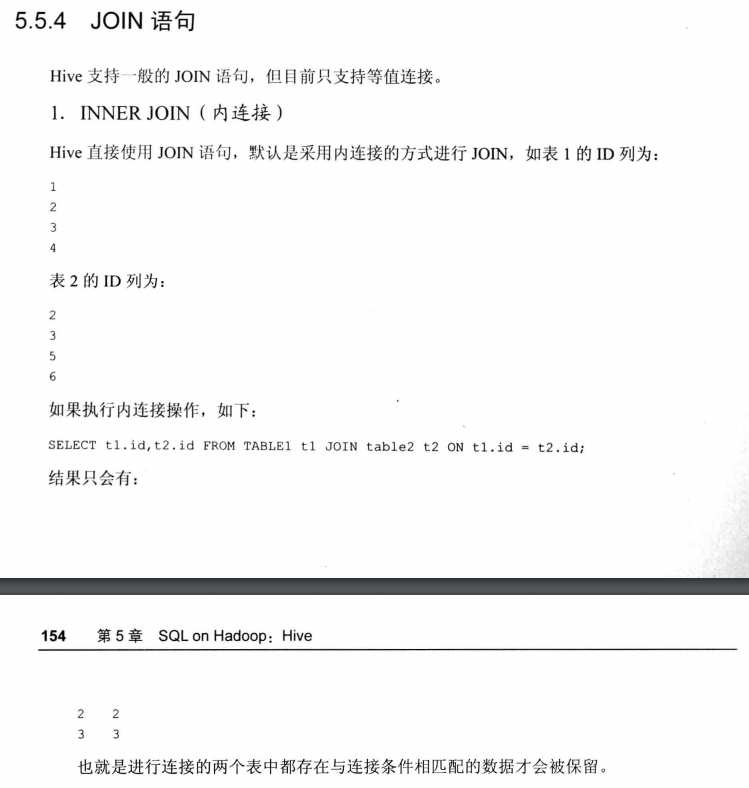


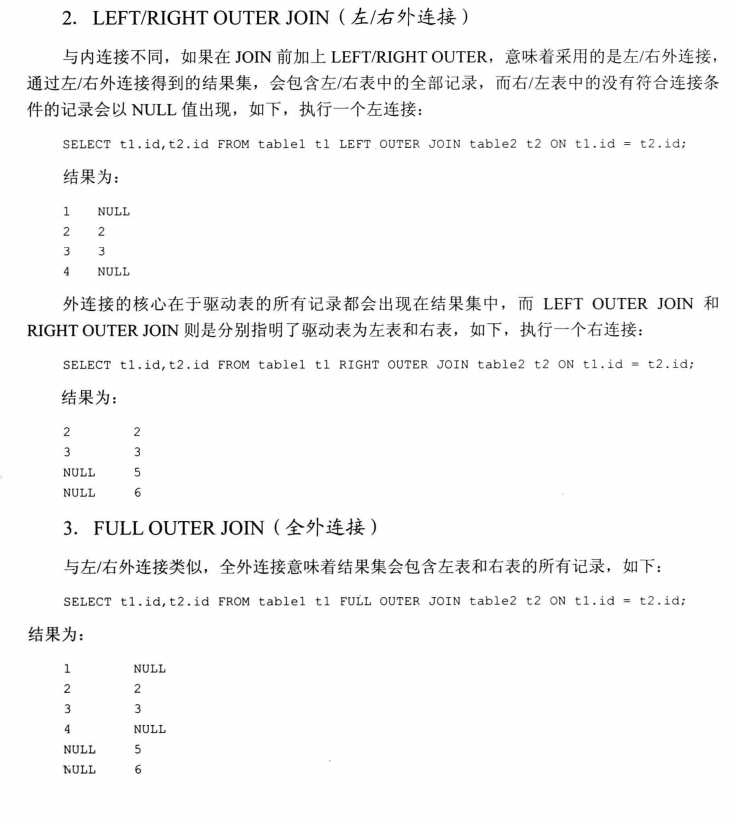
查询：

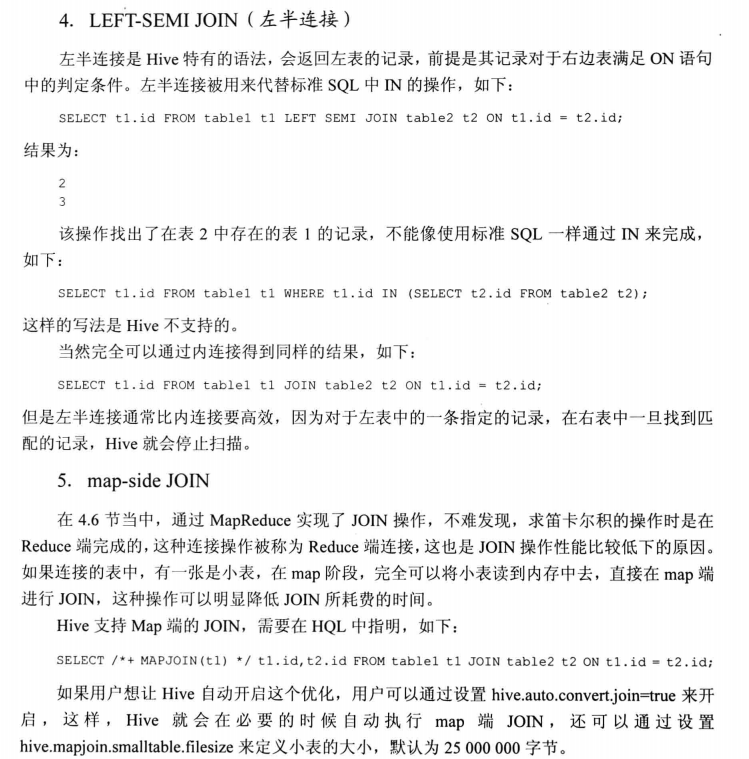


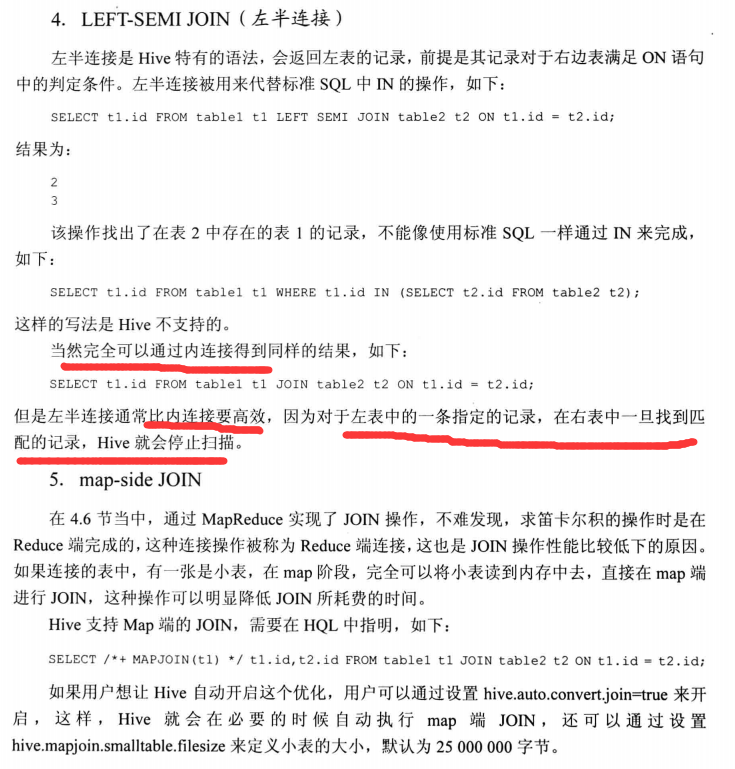


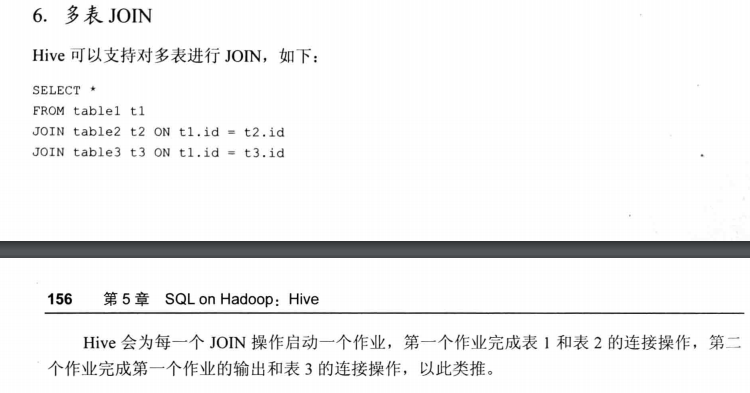
连接：

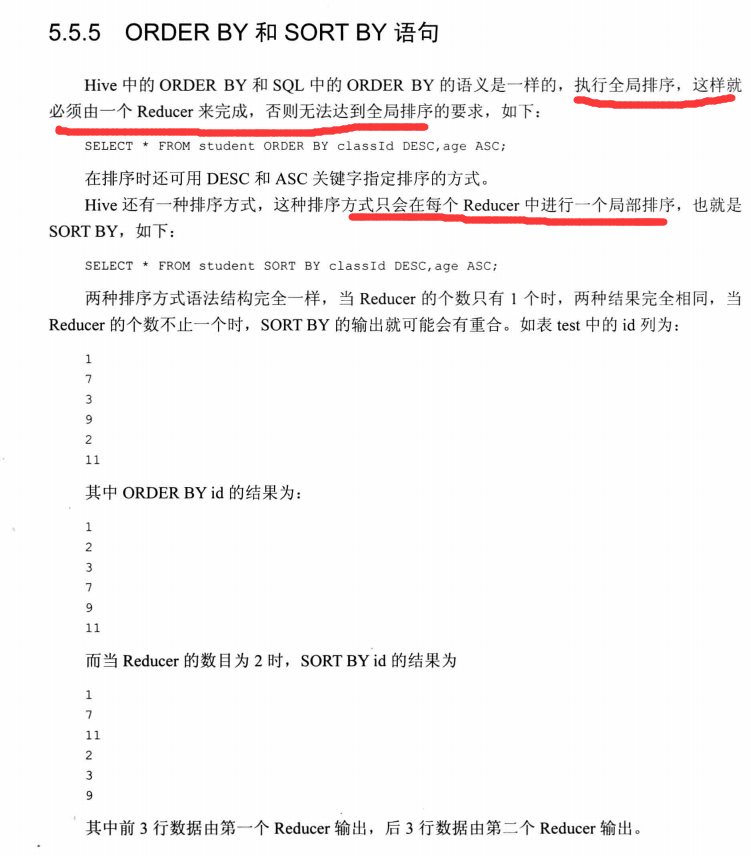


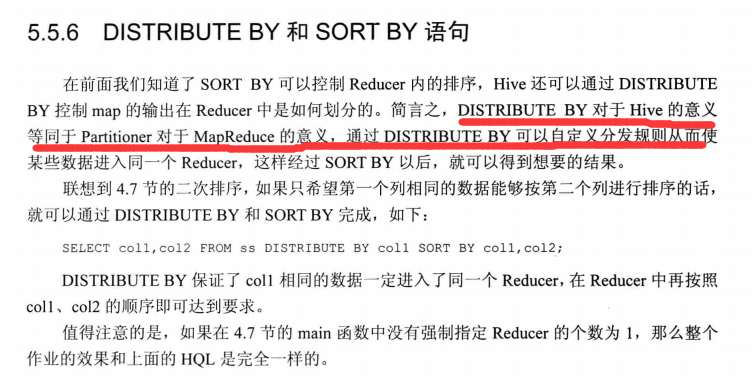




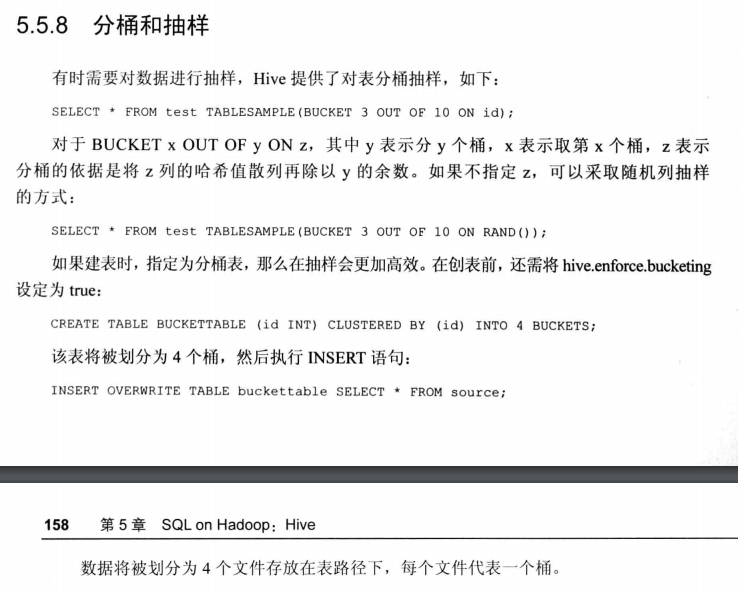


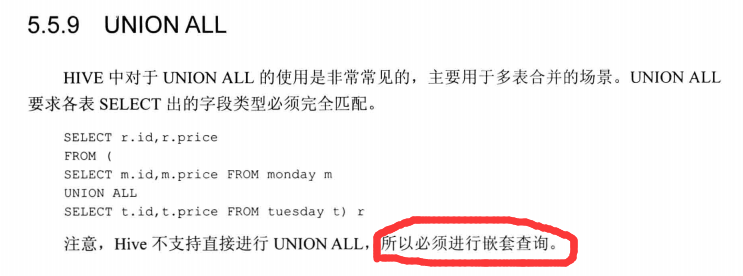








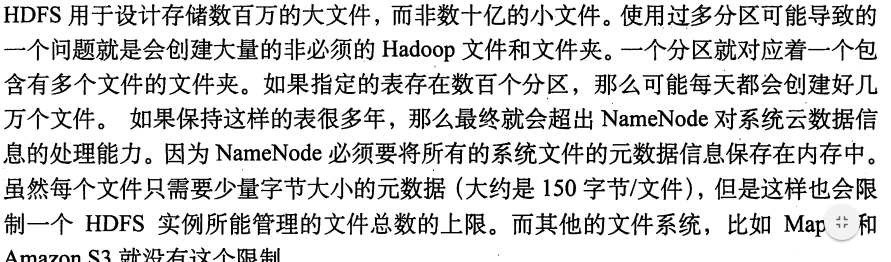


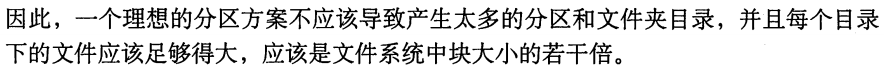


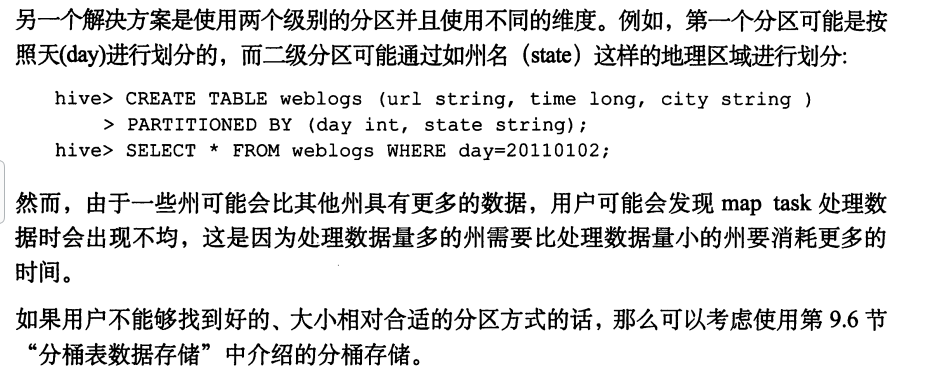
hive中分区和分桶的异同及联系

首先来说，分桶和分区都是为了缩小查找的范围加快hive或者底层mapreduce程序执行的速度。hive的分区是非常有用的，因为hive通常要对输入进行全盘扫描来满足查询条件（这里忽略hive的索引功能）。

分区：



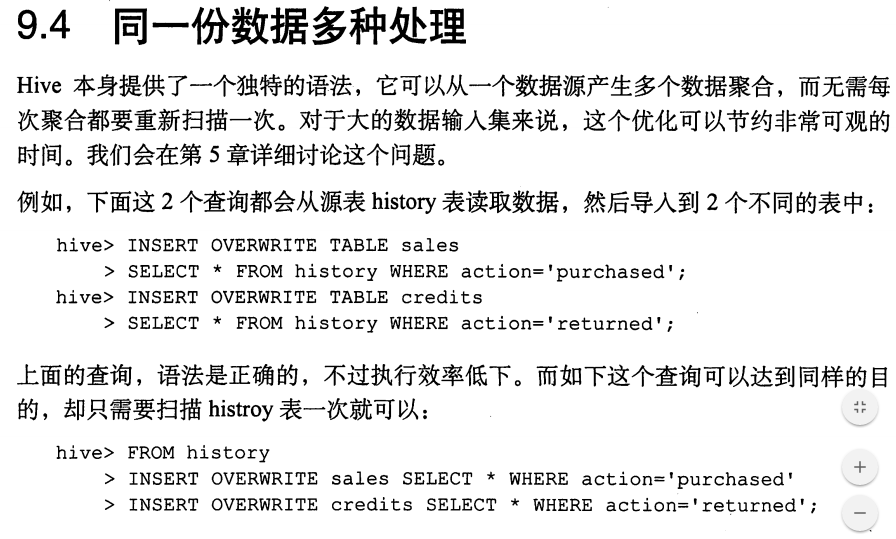




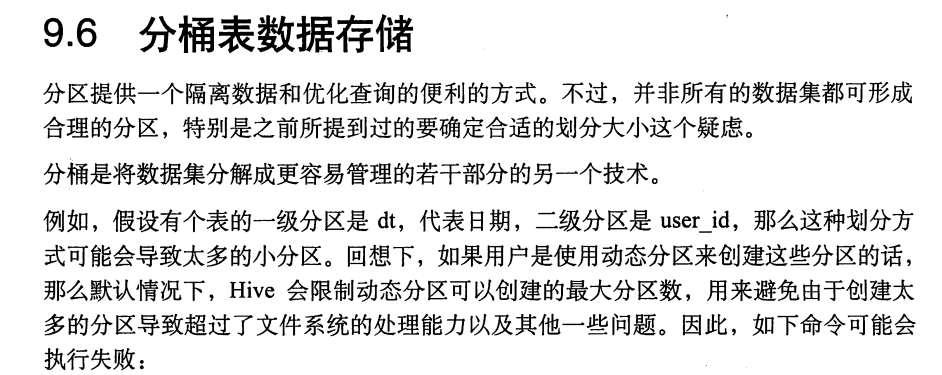
所以分区的时候要考虑两个主要问题：

1.分区的数量不宜过多

2.每个分区中的数据尽量保持均衡



分桶：





调优：

