**Bulkload相关：**

1.为什么要Bulkload？Bulkload的本质是什么？

各个阶段的目的

Bulkload是批量入库，本质上他是ETL，直接把source文件变成HBase底层的Hfile；

1. Bulkload有几个要点？

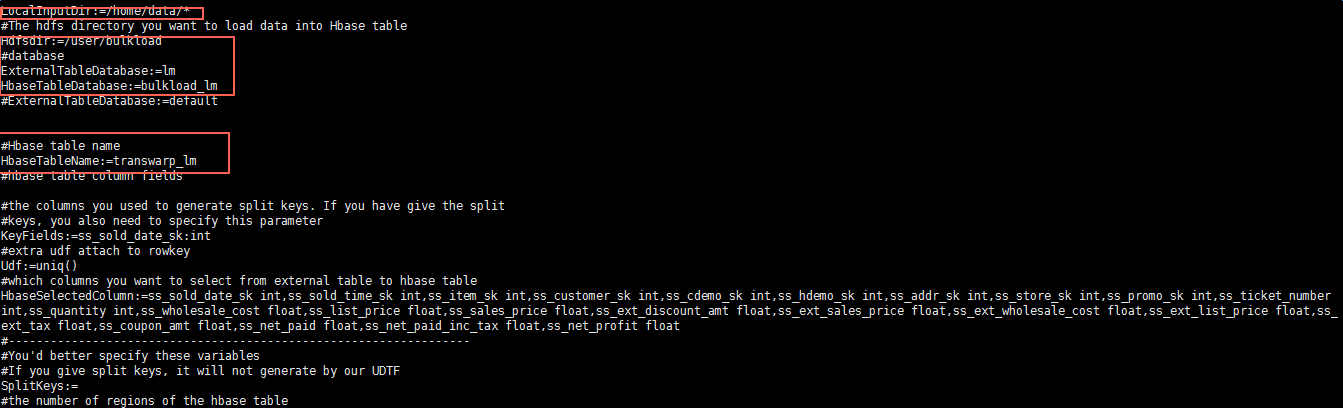
两个要点，一是mapper映射关系，二是reducer对结果进行全排列

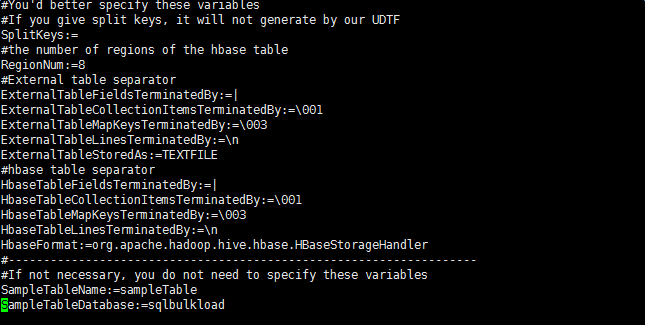
1. 什么是SQLbulkload？

SQLbulkload是一种入库方式，它是利用hbase的数据信息按照特定格式存储在hdfs内这一原理，直接在HDFS中生成持久化的HFile数据格式文件，然后上传至合适位置，即完成巨量数据快速入库的办法。配合mapreduce完成，高效便捷，而且不占用region资源，增添负载，在大数据量写入时能极大的提高写入效率，并降低对HBase节点的写入压力。

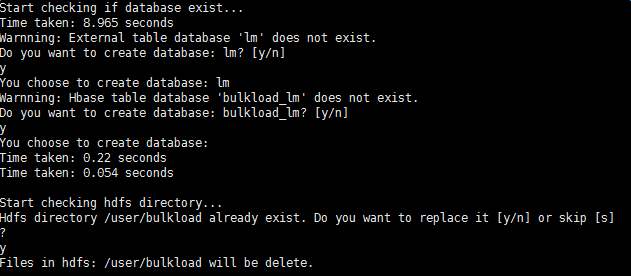
**SQLBulkload操作（TPCDS一张表的数据）**

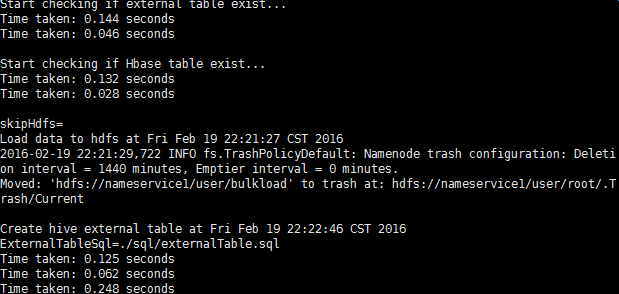
第一步，修改bulkload里面的config.txt文件

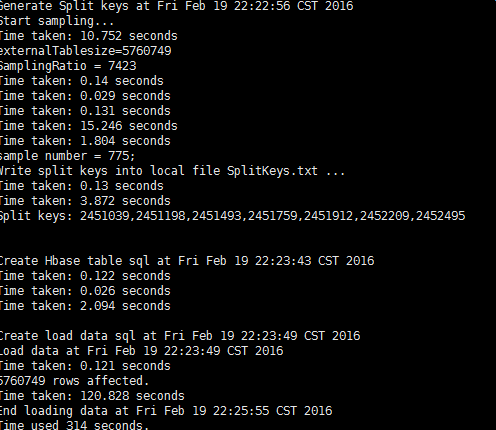




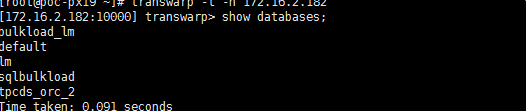
第二步，执行run.sh文件

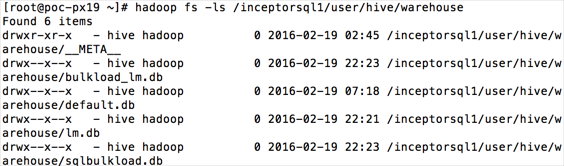




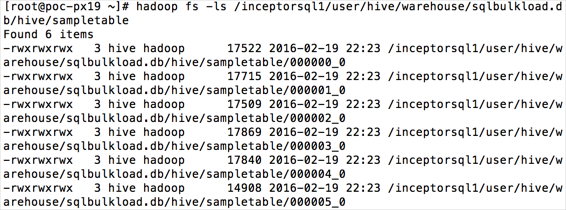


验证是否创建好的数据库以及各个表

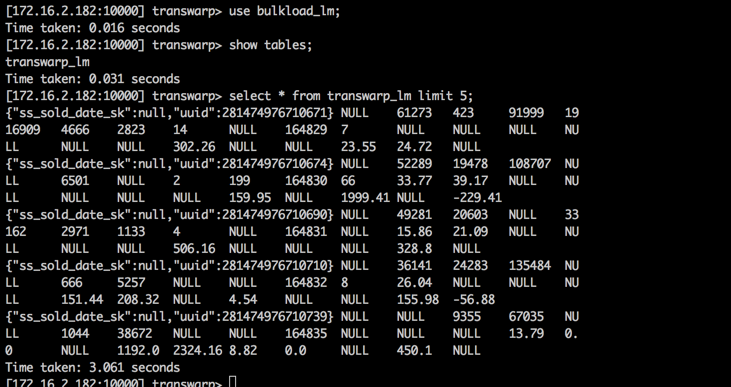




Simple抽样表



验证创建好的Hbase表是否成功



进入60010端口查看新建成的Hbase表的信息

