HBase与Hive实验

一、小组信息：

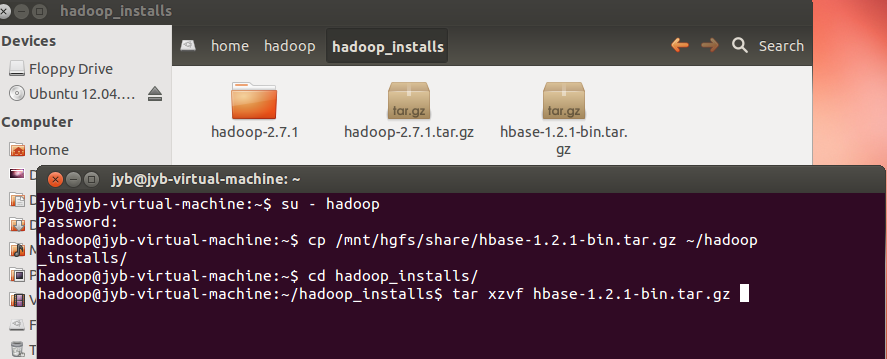
2016st19：张瑞MF1533072、董茂健MG1533007、杨杰MG1533075、金熠波131220168

二、实验环境

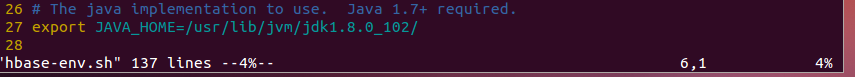
Ubuntu 12.04.2、Java 1.8.0\_102、Maven 3.3.9、IntelliJ 15.06

HBase 1.2.1、Hive 2.0.0

三、HBase安装

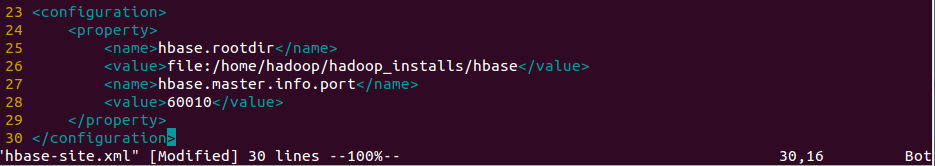


修改hbase-1.2.1/conf/hbase-env.sh：

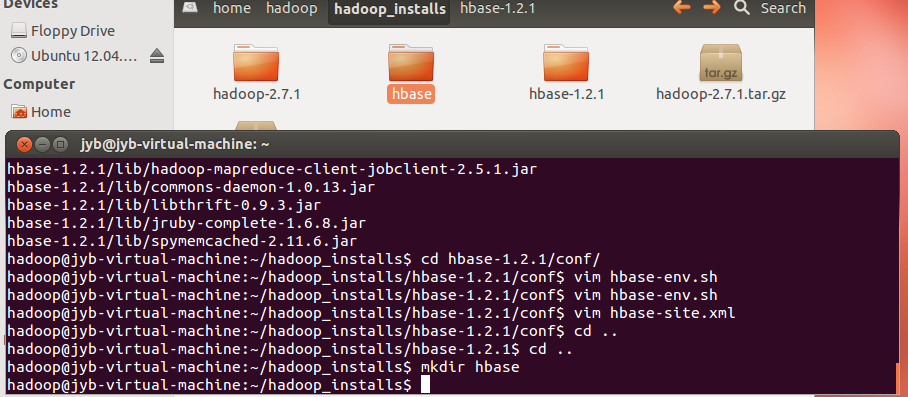




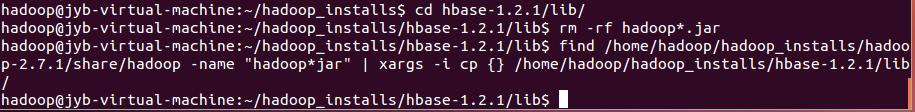
修改hbase-1.2.1/conf/hbase-site.xml：



新建一个用于存放hbase数据的文件夹：

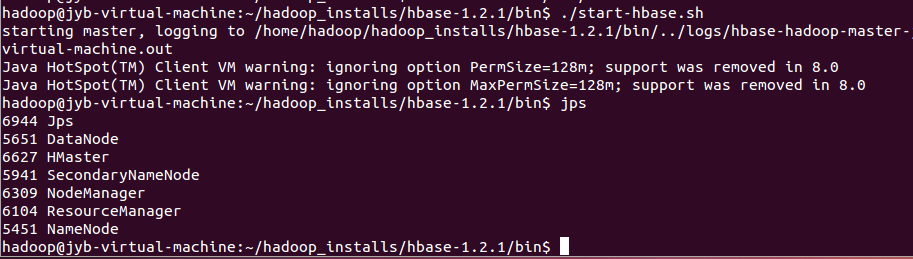


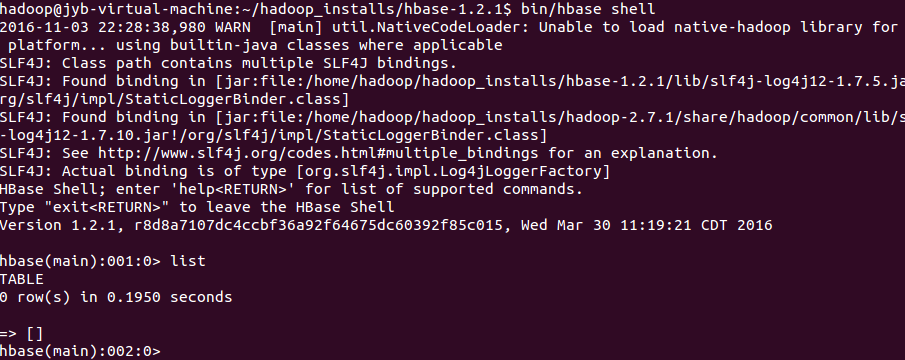
替换hbase中hadoop2.5.1版本的jar包，进行版本统一：

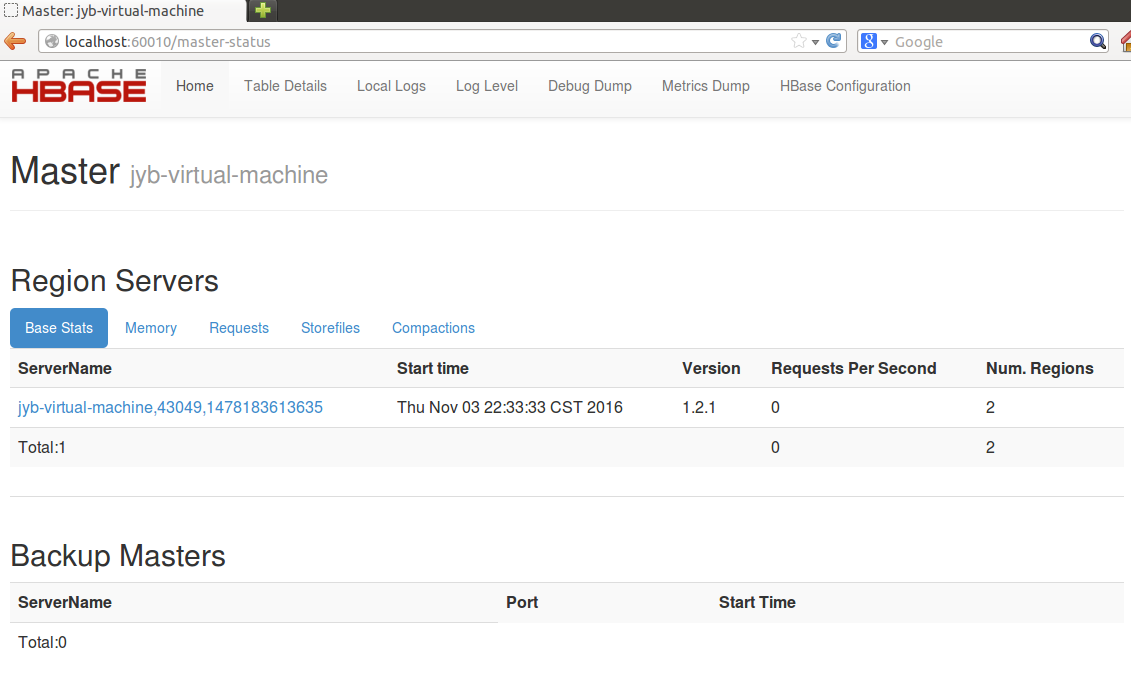




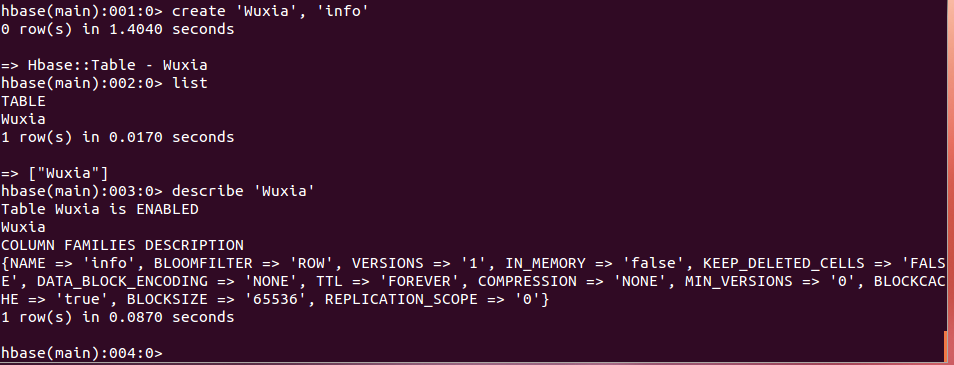
启动hbase：





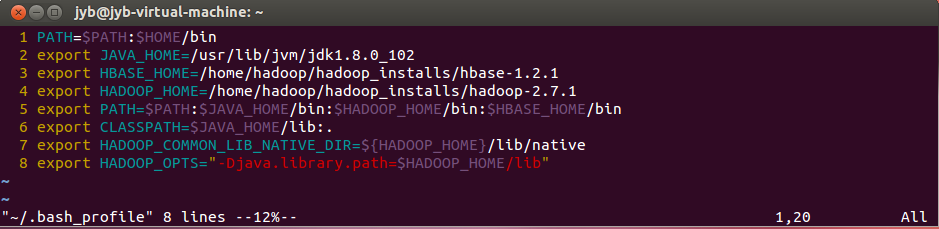


在HBase中创建Wuxia表：

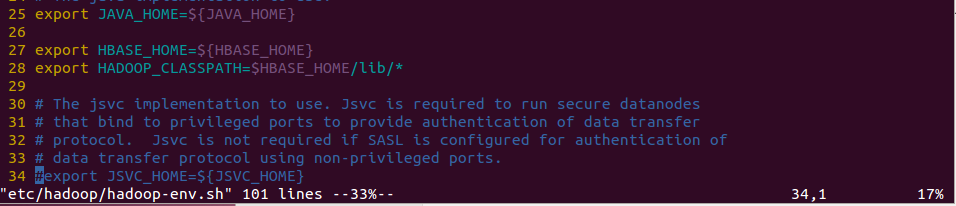


四、通过MapReduce写入HBase（HBaseWuxia工程）

修改.bash\_profile



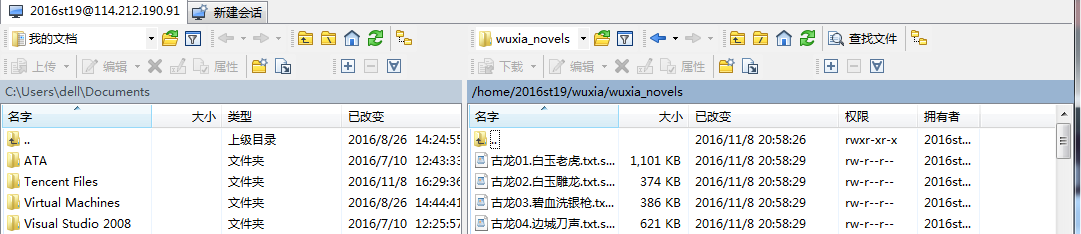
配置hadoop-env.sh



登录集群，将集群上HDFS的文件下到账户的本地文件



用Winscp下载到本地：



将Windows下下载的武侠小说复制到Linux中：



复制到HDFS中：



将hbase下的jar包复制到hadoop的lib下：



MapReduce任务的实现：

这里只列出与倒排索引不同的部分。

Reduce部分：

Reduce继承自TableReducer



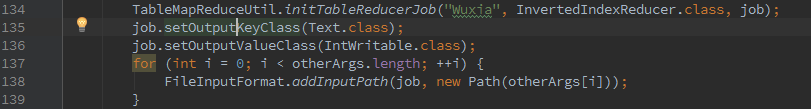
使用put写入表中：



Main函数中的配置：



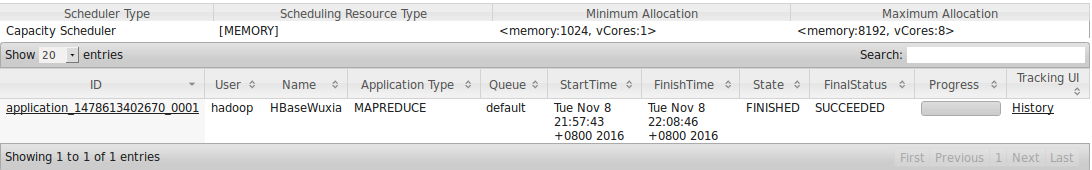
执行写入HBase的MapReduce Job：

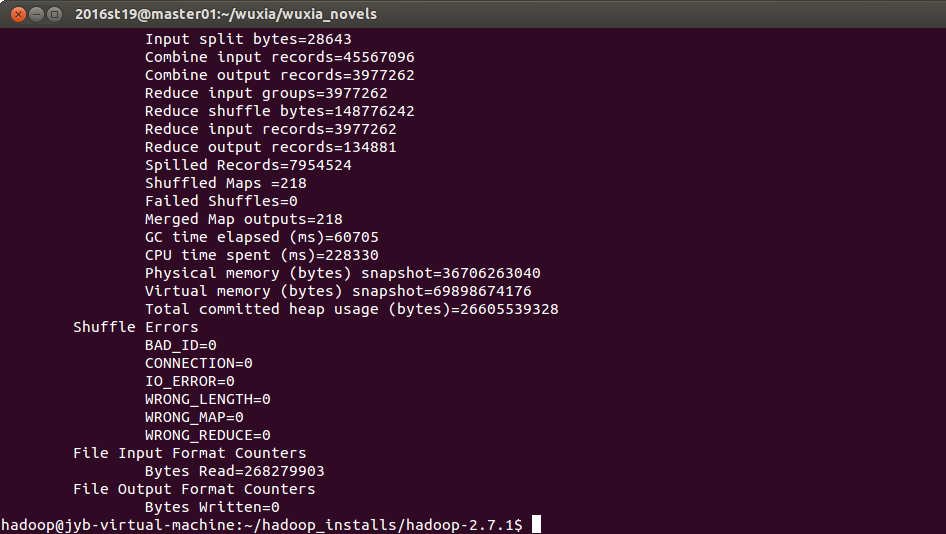


任务的执行：



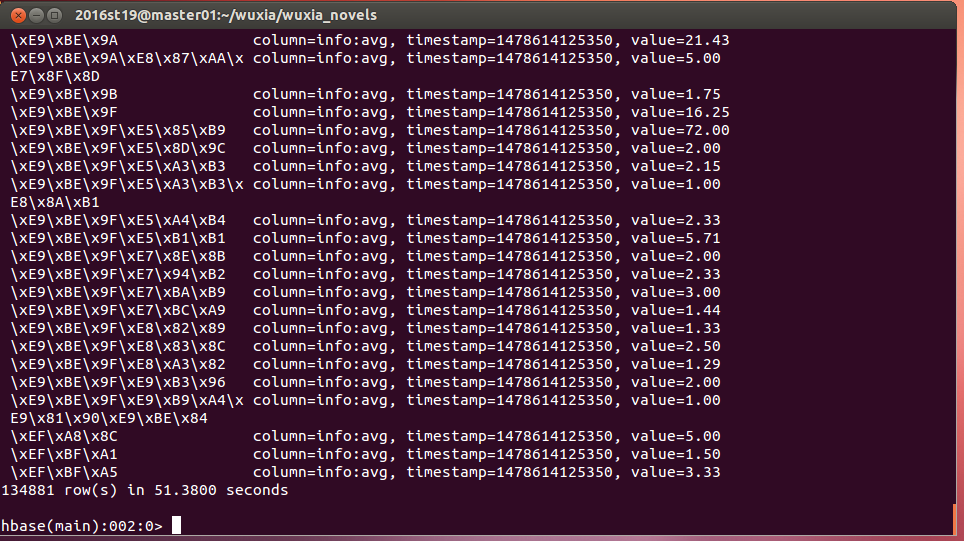
任务执行结果：





检查是否已将结果写入Wuxia：





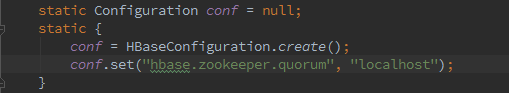
五、从HBase读出并写入文件（HBaseToFile工程）

实现细节：

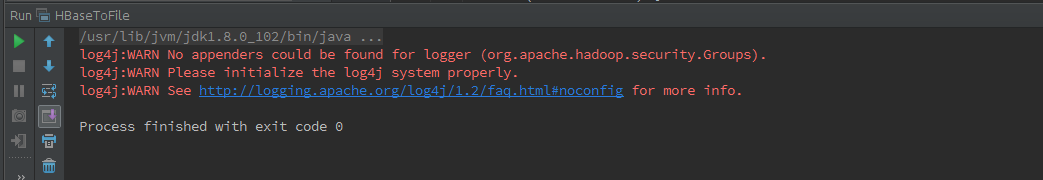
用Scan不断从HBase的指定表中获取结果：



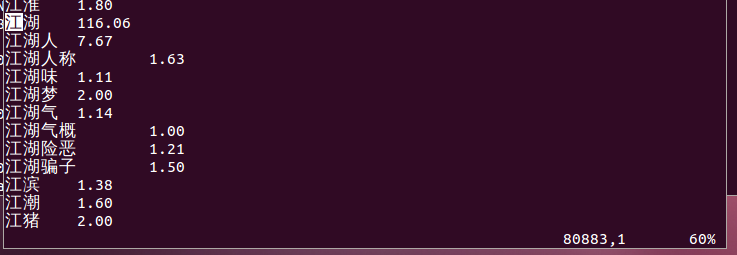
Conf配置：

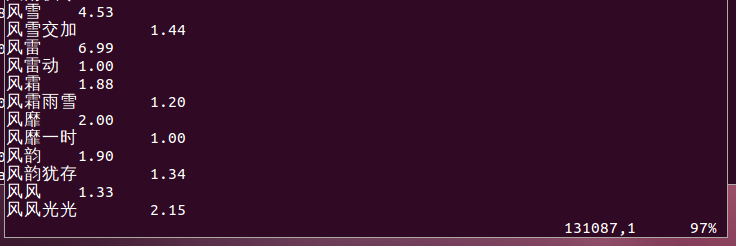


运行HBaseToFile Java程序：

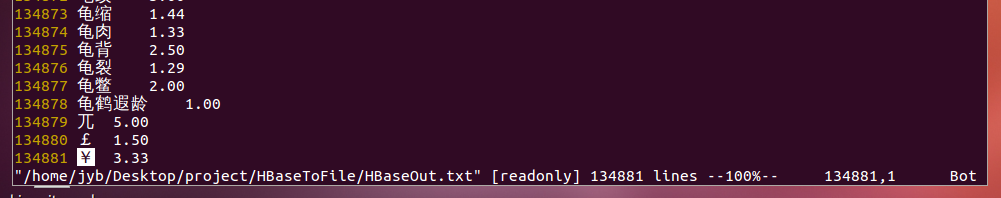


下列出江湖和风雪的结果：



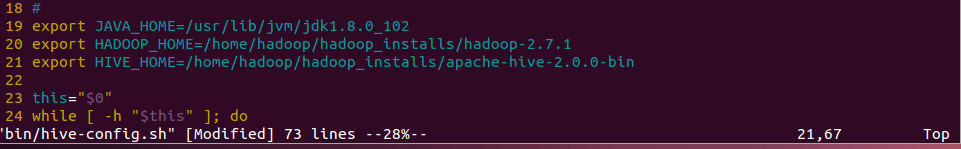


总数也与HBase中scan的一致：

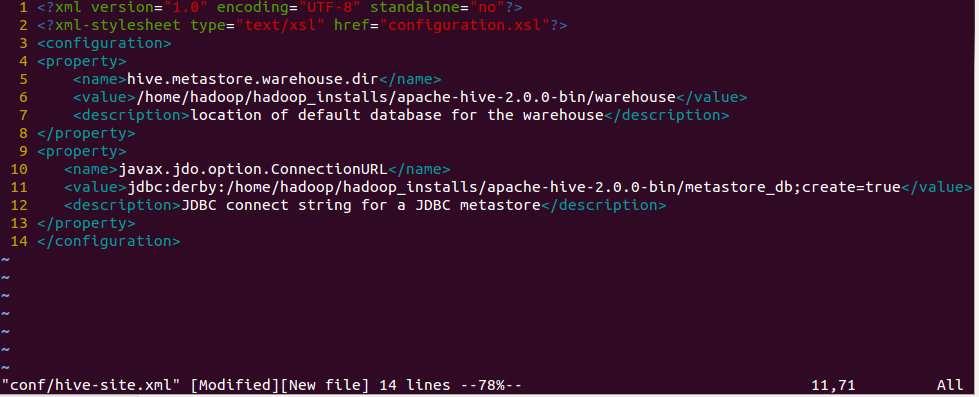


六、Hive安装

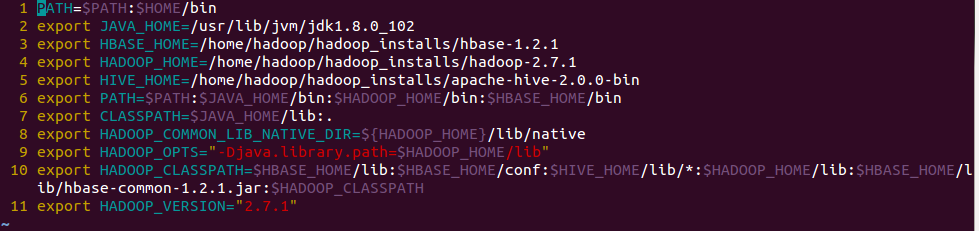
配置bin/hive-config.sh:



创建并修改conf/hive-site.xml:

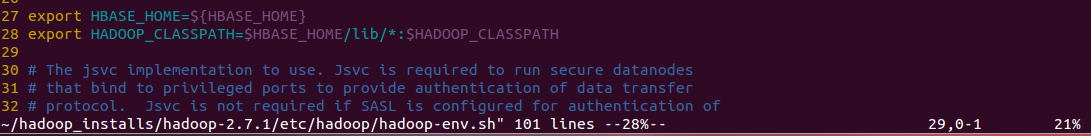


修改~/.bash\_profile并应用：

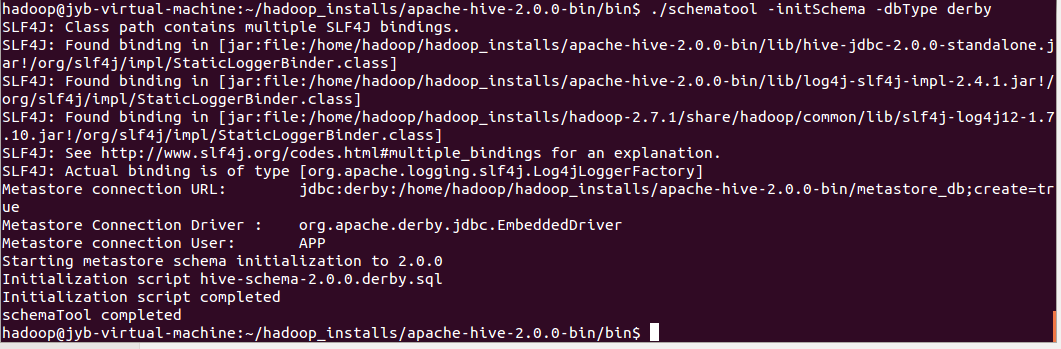




修改Hadoop中etc/hadoop/hadoop-env.sh:



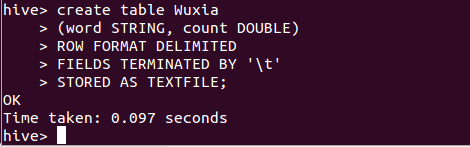
初始化数据库：



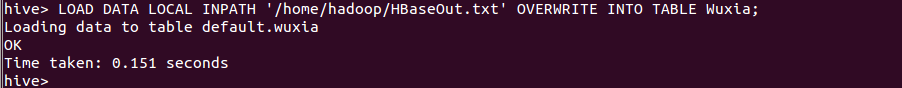
创建并升级数据文件夹：



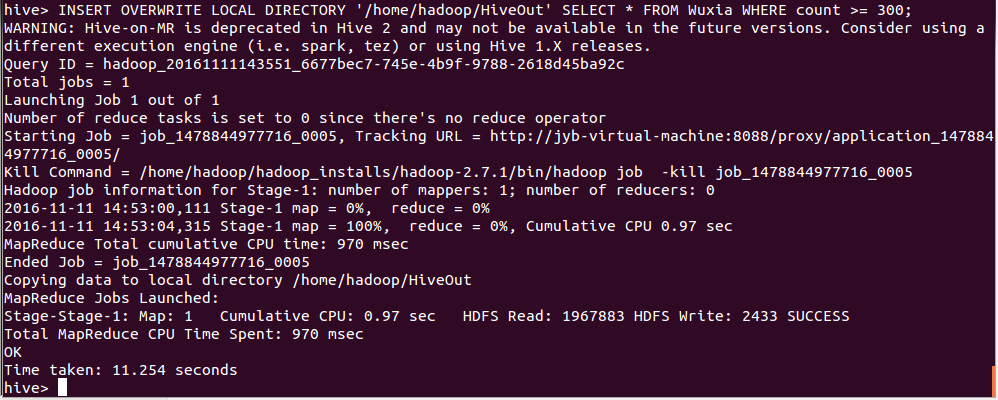
创建Wuxia表：

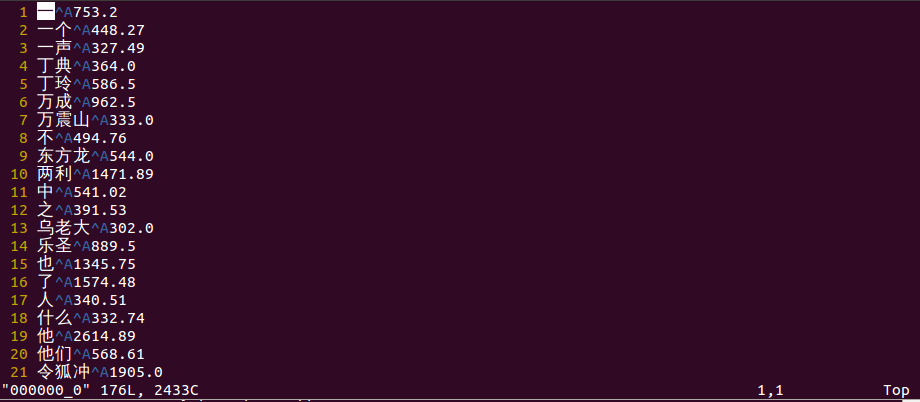


将从HBase中导出的数据导入Wuxia表：

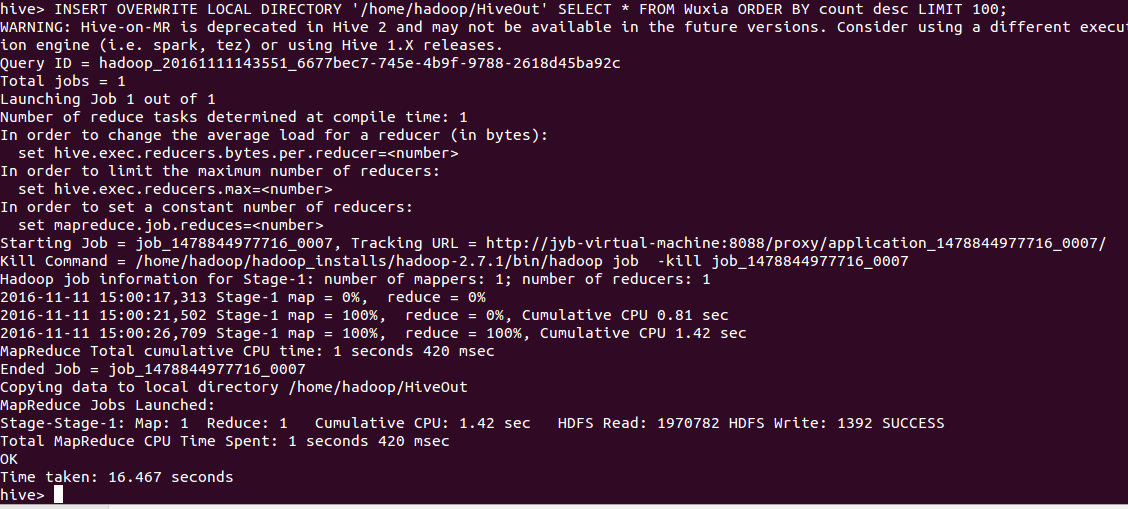


查询次数大于300的单词：

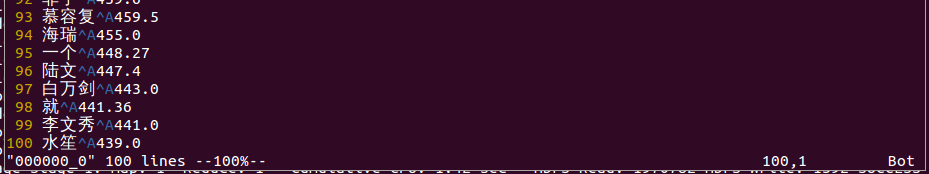




查询前100次数最多的单词：







六、附件及说明

随附件包括本实验报告、ReadMe、HBaseWuxia工程（删去out生成目录）、HBaseWuxia.jar、HBaseToFile工程、HBaseToFile的中间生成文件、Hive查询生成的两个查询文件。