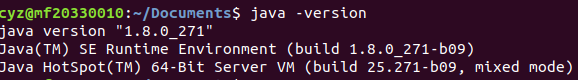
实验一：安装单机Hadoop系统与WordCount程序实验

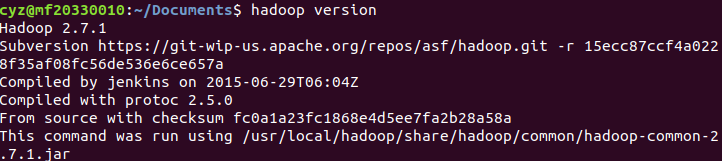
#### 系统安装与运行

Mac下的运行的Linux虚拟机：Ubuntu18.04

Java版本：1.8

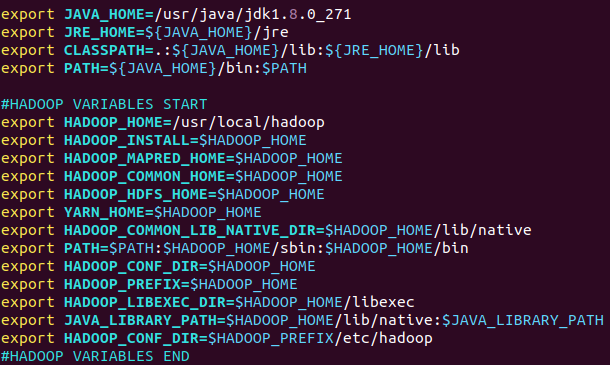


单机伪分布式Hadoop：2.7.1



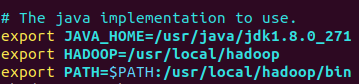
相关文件配置：

环境变量：~/.bashrc

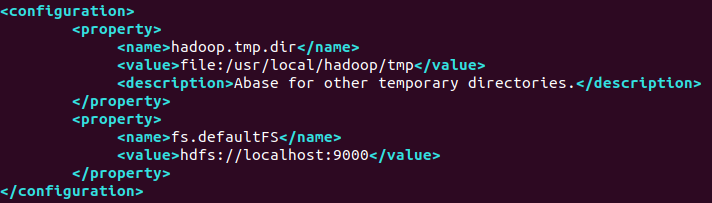


Hadoop相关配置：修改$HADOOP\_HOME/etc/hadoop目录下配置文件

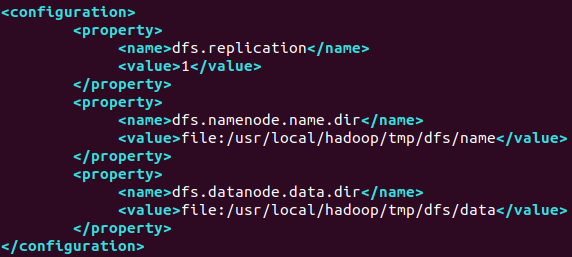
hadoop-env.sh:



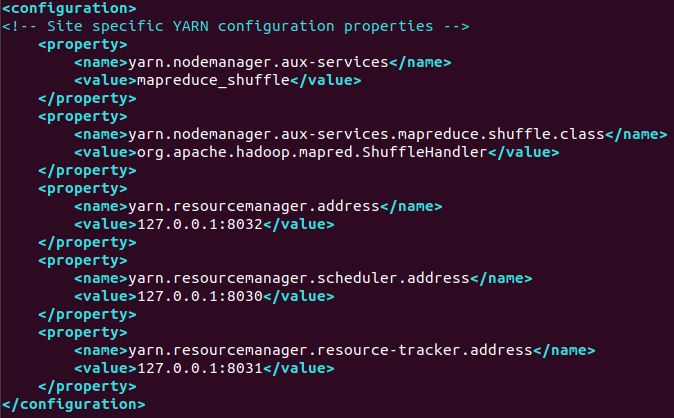
core-site.xml:



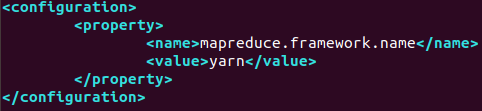
hdfs-site.xml:



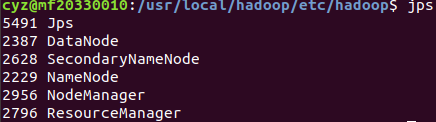
yarn-site.xml:



mapred-site.xml:



启动HDFS和MapReduce后打印各进程信息：jps



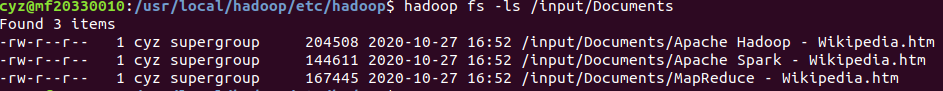
#### 实验数据说明

对维基百科中[Apache Spark](https://en.wikipedia.org/wiki/Apache_Spark), [Apache Hadoop](https://en.wikipedia.org/wiki/Apache_Hadoop), [MapReduce](https://en.wikipedia.org/wiki/MapReduce)这三个词条的页面保存为html文件作为本次WordCount实验的输入，其存储在~/Documents目录下。

在HDFS文件系统中创建input文件夹：hdfs dfs -mkdir /input

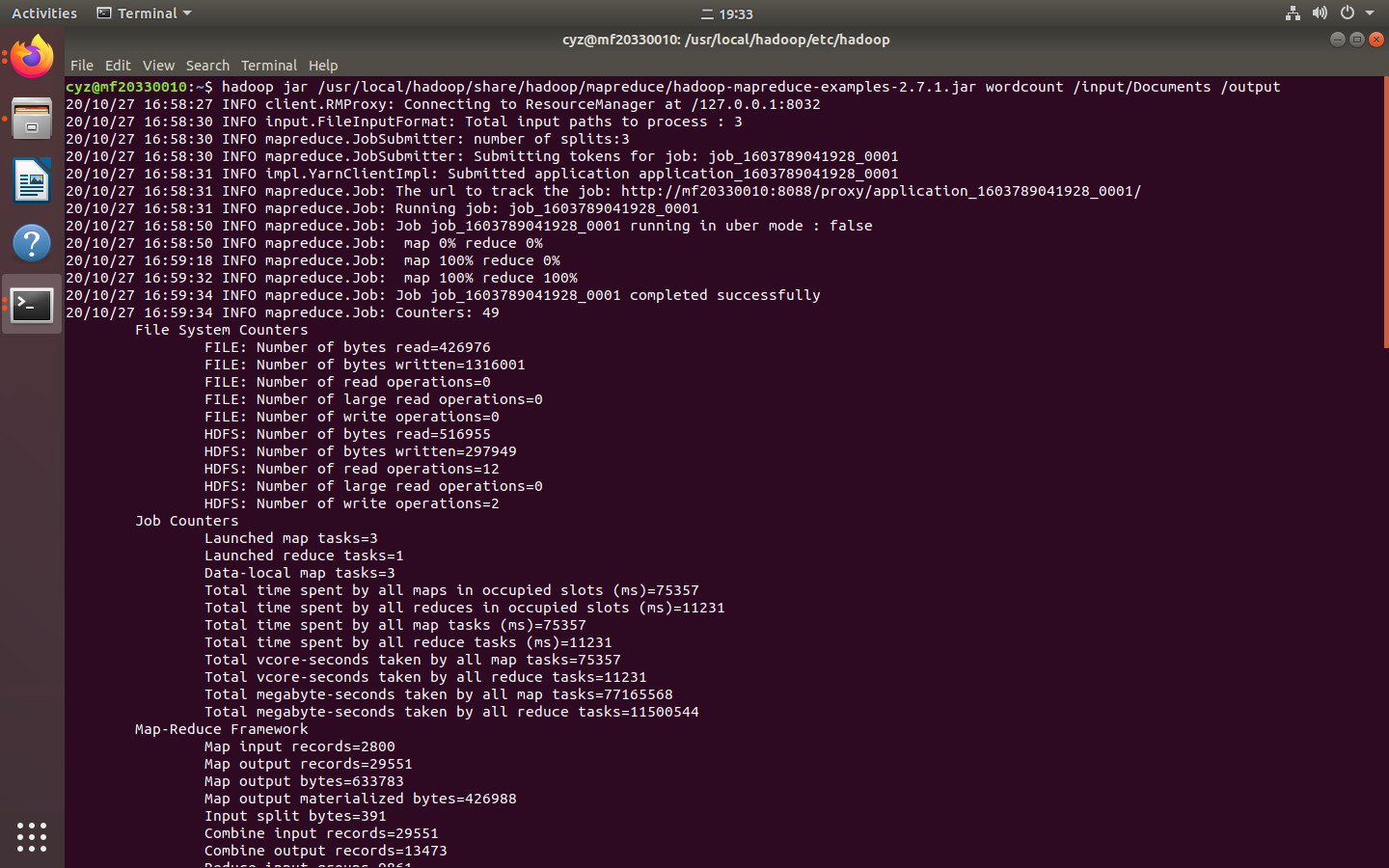
将该文件夹从本地复制到HDFS文件系统下：hdfs fs –put ~/Documents /input

查看HDFS文件系统下的Documents文件夹:Hadoop fs -ls /input/Documents

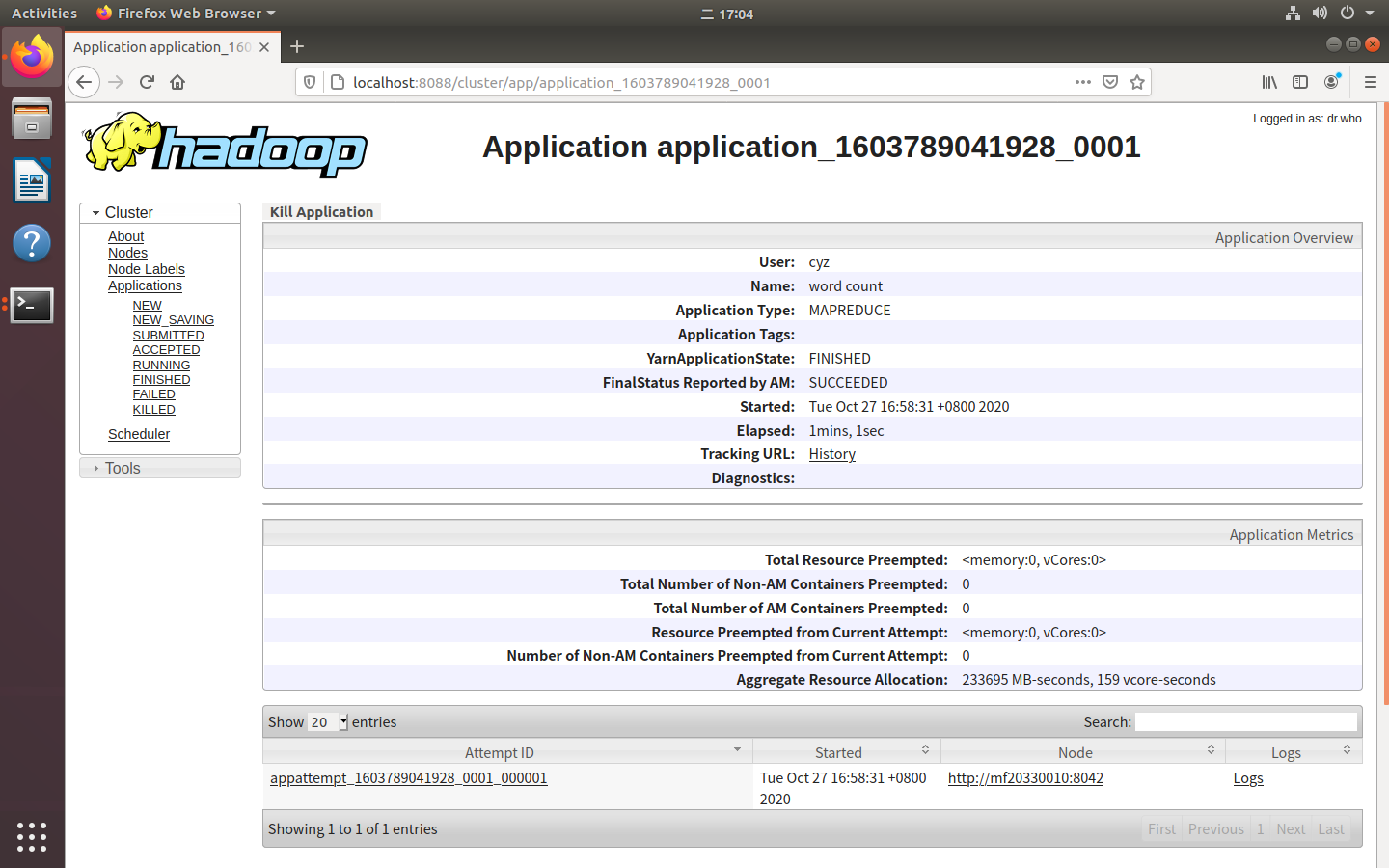


#### 作业运行状态

hadoop jar /usr/local/Hadoop/share/hadoop/mapreduce/Hadoop-mapreduce-examples-2.7.1.jar wordcount /input/Documents /output



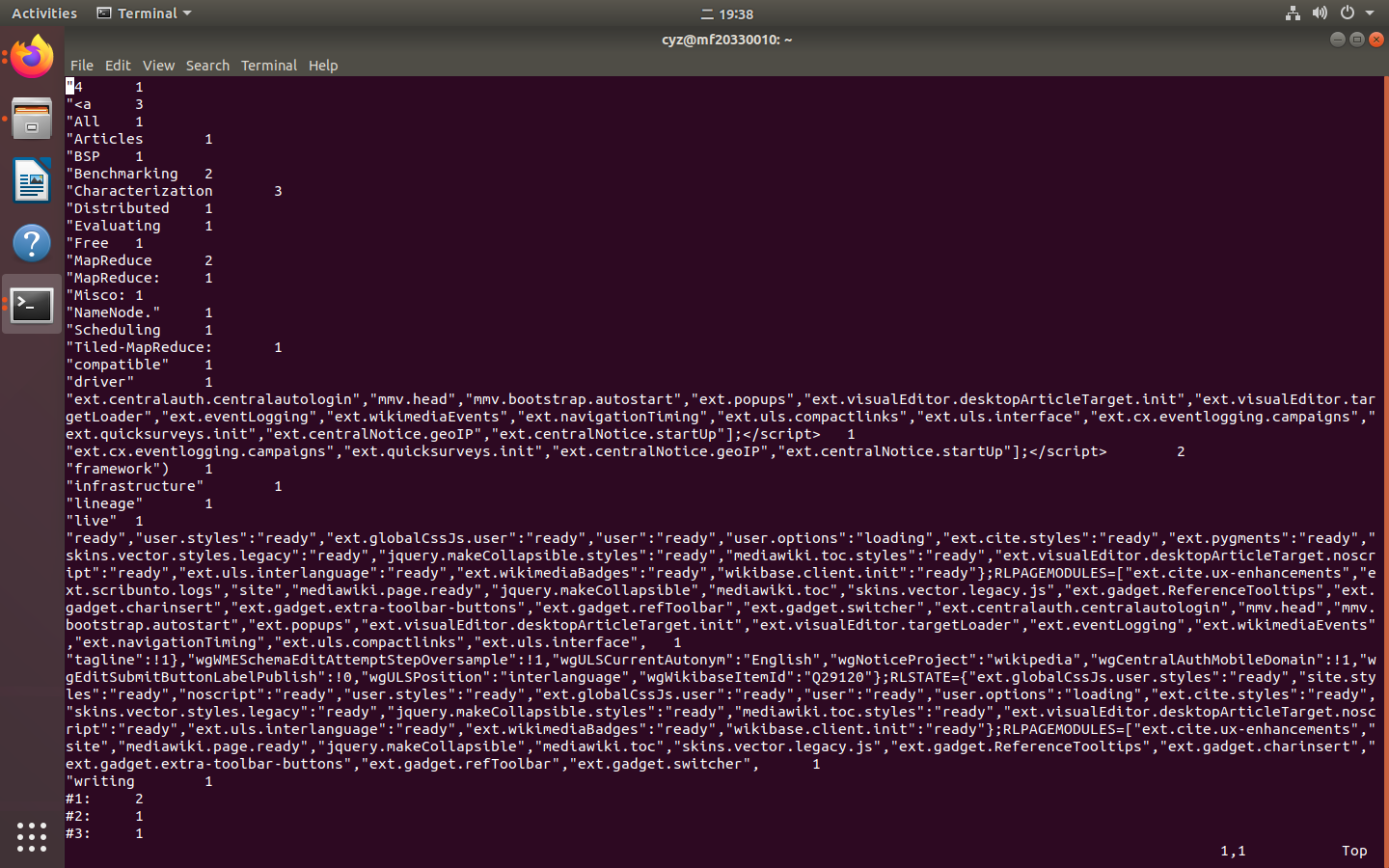
在localhost:8088观察作业运行状态



#### 4.实验结果

将输出结果从HDFS文件系统中复制到本地：hadoop fs -get /output/part-r-00000 ~/Documents

查看结果：



#### 5.实验体会

通过本次实验，我从实践中认识了Hadoop架构与HDFS文件系统，并且对其有了更深刻的认识。