1、 系统安装运行情况

etc/hadoop/core-site.xml:

```xml
<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>file:/usr/local/hadoop/tmp</value>
                <description>Abase for other temporary directories.</description>
        </property>
        <property>
                <name>fs.defaultFS</name>
                <value>hdfs://localhost:9000</value>
        </property>
        <property>
          <name>hadoop.native.lib</name>
          <value>true</value>
          <description>Should native hadoop libraries, if present, be used.</description>
        </property>
</configuration>
```
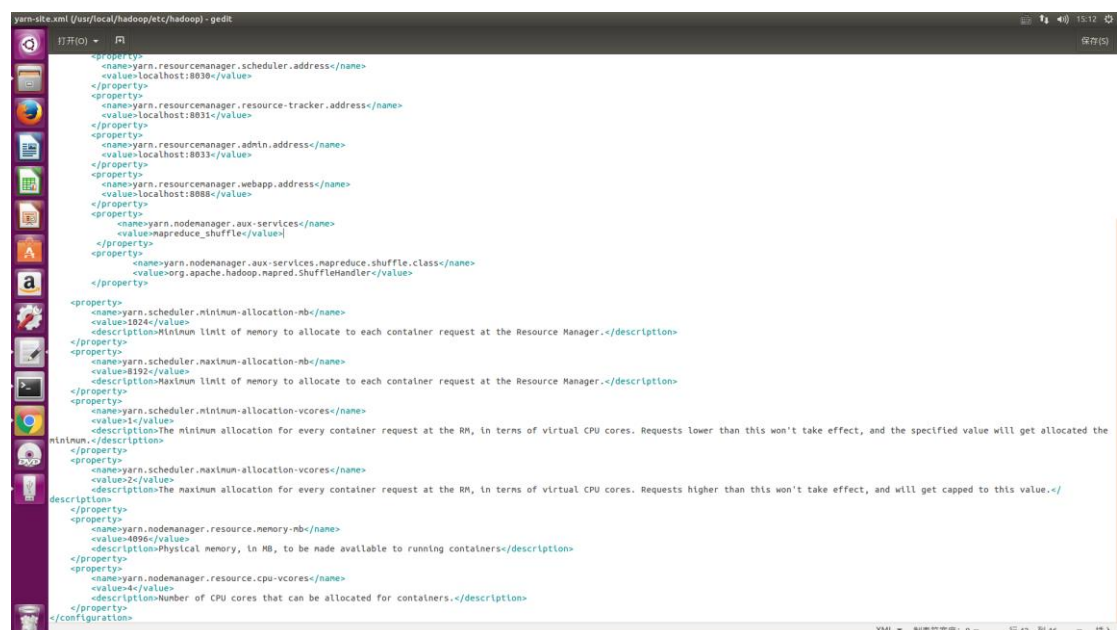
etc/hadoop/hdfs-site.xml:

```xml
<configuration>
        <property>
                <name>dfs.replication</name>
                <value>1</value>
        </property>
        <property>
                <name>dfs.namenode.name.dir</name>
                <value>file:/usr/local/hadoop/tmp/dfs/name</value>
        </property>
        <property>
                <name>dfs.datanode.data.dir</name>
                <value>file:/usr/local/hadoop/tmp/dfs/data</value>
        </property>
</configuration>
```

etc/hadoop/yarn-site.xml:
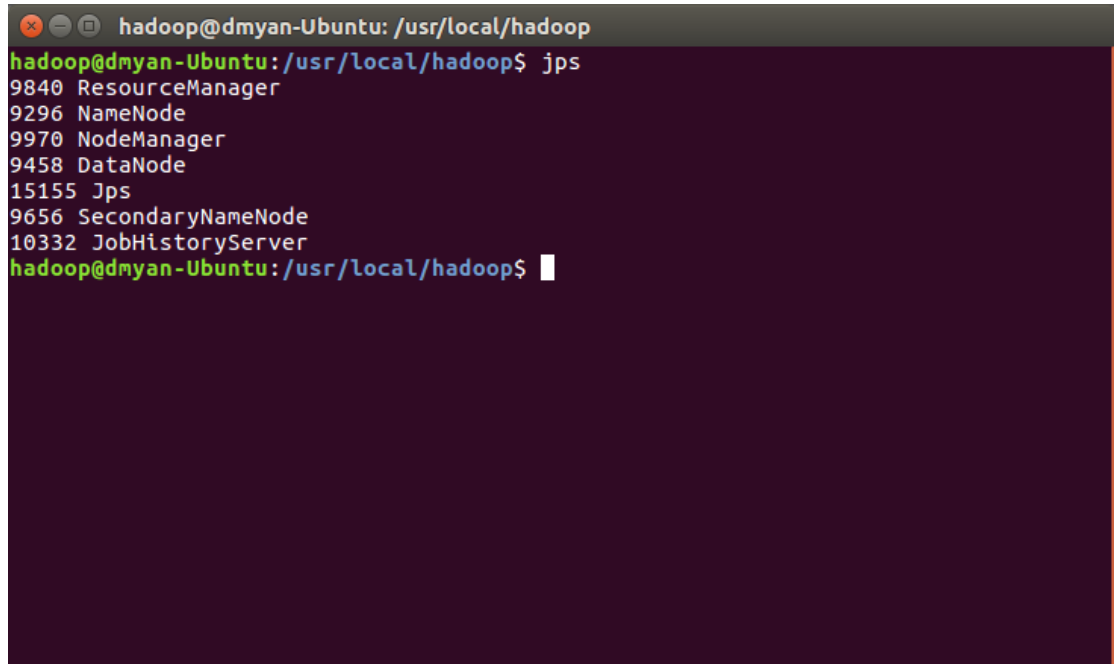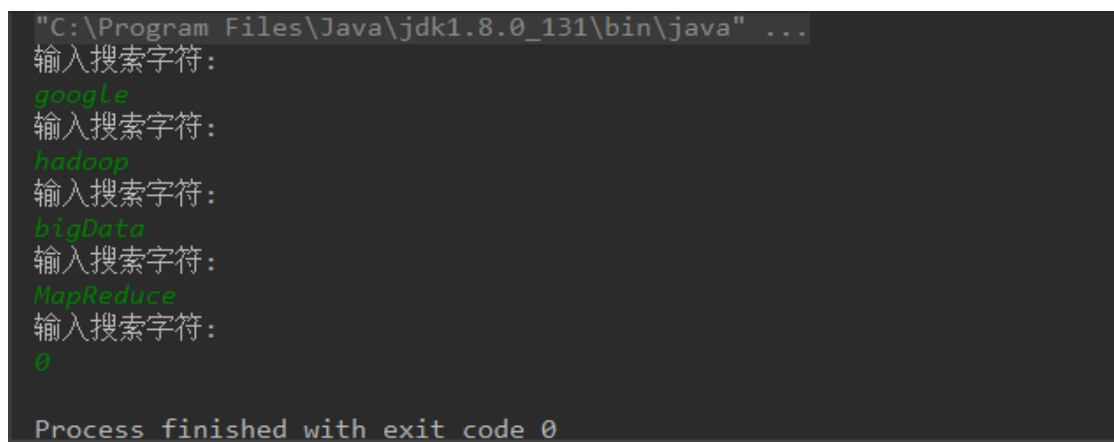

etc/hadoop/mapred-site.xml:

```
<configuration>
        <property>
                <name>mapreduce.framework.name</name>
                <value>yarn</value>
        </property>
</configuration>
```

运行情况：

```
hadoop@dmyan-Ubuntu: /usr/local/hadoop
hadoop@dmyan-Ubuntu:/usr/local/hadoop$ jps
9840 ResourceManager
9296 NameNode
9970 NodeManager
9458 DataNode
15155 Jps
9656 SecondaryNameNode
10332 JobHistoryServer
hadoop@dmyan-Ubuntu:/usr/local/hadoop$
```
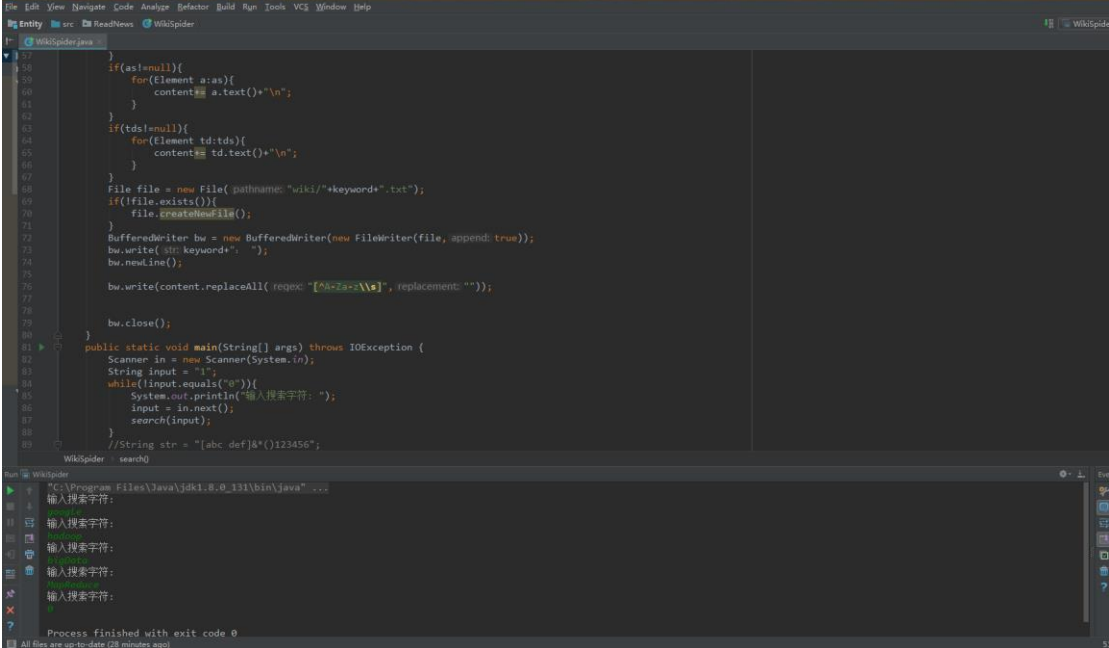
2、 实验数据说明：

从 wikipedia 网页中抓取了关于 bigData、google、hadoop、

MapReduce 等关键词的英文网页中的单词数据

```
"C:\Program Files\Java\jdk1.8.0_131\bin\java" ...
输入搜索字符：
google
输入搜索字符：
hadoop
输入搜索字符：
bigData
输入搜索字符：
MapReduce
输入搜索字符：
0

Process finished with exit code 0
```

下面是关键词为 google 的 wikipedia 上词汇抓取结果（关键词 google 的结果一共 3425 行单词数据）：

3、程序运行后在 Hadoop Web 作业状态查看界面上的作业运行状态

4、实验输出结果开头部分

5、实验体会

本次实验经过一步步的实践学会了安装 hadoop 平台，简单的使用 Hadoop，Yarn 进行 Hadoop 集群的资源管理，运行 WordCount 程序并进行词频统计。