西安科技大学

**计算机学院实验报告**

班 级 网络工程1801 姓 名 吴斌 学 号 18408020129

课程名称 大数据技术原理与应用 实验（五）

**1实验名称：**

大数据技术原理与应用综合实验

**2实验目的：**

MapReduce基础实验

**3实验原理：**

实验Part1：直接使用example里给出的jar包运行词频统计应用

实验Part2：下载java源代码自己编译运行词频统计应用

**4实验内容与结果；**

4.1：

启动hdfs和mapreduce相关组件

4.2：

在hdfs上准备好待词频统计的文本文件

4.3：

向mapreduce提交wordcount应用

4.4：

查看词频统计结果

4.5：

尝试下载java源代码自己编译运行

**5实验截图及简单总结：**

**第一步：启动hdfs和mapreduce相关组件**

若已完成过hdfs的格式化，不再需要进行格式化；若未进行格式化，则首先在hadoop-3.3.1的bin目录下执行格式化操作，命令如下：./hadoop namenode –format。

进入hadoop-3.3.1/sbin目录，执行如下命令：./start-all.sh。通过jps命令，确认如下五个进程成功运行：NameNode、DataNode、SecondaryNameNode、NodeManager、ResourceManager。

**第二步：在hdfs上准备好待词频统计的文本文件**

在本地创建一个文本文件，例如在/home/lxyseu/Downloads目录下创建一个名为text的文件，内容是fox wolf dog cat chicken duck wolf tiger fox。

在hdfs中创建input文件夹，需要在hadoop-3.3.1/bin目录下运行如下命令：./hdfs dfs –mkdir /input。

将本地文件上传至hdfs的/input目录，即在hadoop-3.3.1/bin目录下运行./hdfs dfs –copyFromLocal /home/lxyseu/Downloads/text /input。

**第三步：向mapreduce提交wordcount应用**

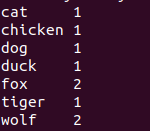
在hadoop-3.3.1/bin下执行如下命令：

./hadoop jar ../share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar wordcount /input /output。

**第四步：查看词频统计结果**

在hadoop-3.3.1/bin下执行命令./hdfs dfs –ls /output命令，可以查看到结果文件part-r-00000。

在hadoop-3.3.1/bin下继续执行./hdfs dfs –cat /output/part-r-00000命令，可以看到统计结果：



**实验Part2：下载java源代码自己编译运行词频统计应用**

**第一步：通过官网下载src，找到wordcount源代码**

**第二步：编译源代码，并生成jar包**

通过命令vi ~/.bashrc，增加如下内容：

export HADOOP\_HOME=/home/lxyse/hadoop-3.3.1（根据你的hadoop文件夹实际位置填写）

export CLASSPATH=$($HADOOP\_HOME/bin/hadoop classpath):$CLASSPATH

通过命令source ~/.bashrc生效。

通过命令javac /home/lxyseu/Downloads/wordcount.java进行编译（根据实际下载目录）

通过jar -cvf 命令完成jar包生成，jar -cvf WordCount.jar ./WordCount\*.class

**第三步：通过hadoop jar命令运行**

在hadoop-3.3.1./bin目录下运行如下命令：

/hadoop jar WordCount.jar org/apache/hadoop/examples/WordCount /input /output2 （填写WordCount.jar文件的实际地址）（备注：实验Part1的第一步、第二步、第四步仍然必要）