

**Hadoop大數據處理技術及應用基礎**

**實驗報告二**

**專 業：** 數據科學

**授課教師：** 应作斌

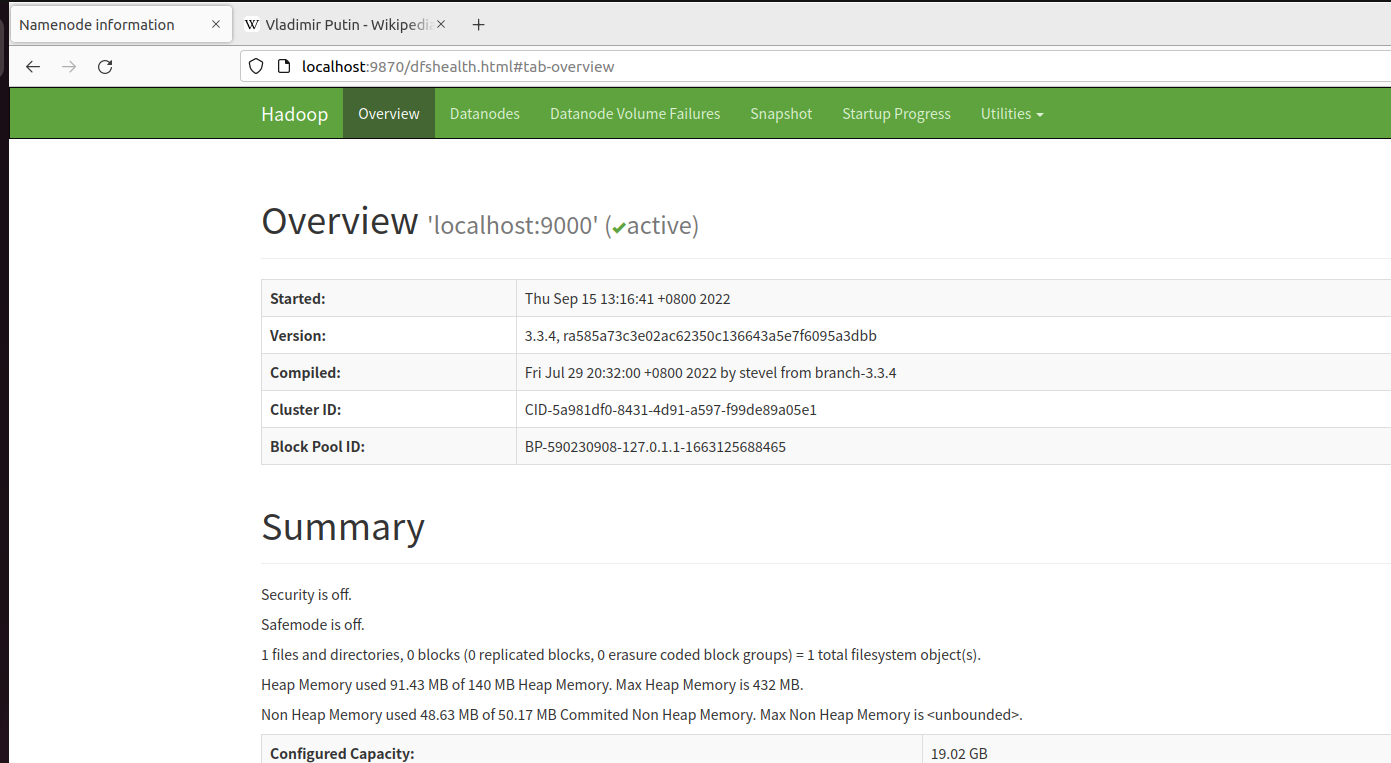
**學生姓名：** 刘传

（一）启动Hadoop

通过运行如下代码：

**./sbin/start-dfs.sh**

开启NameNode和DataNode守护进程，然后访问Web界面http://localhost:9870，如下图所示。

****

**图1 Hadoop的Web界面图**

（二）新建input文件夹

问题①：

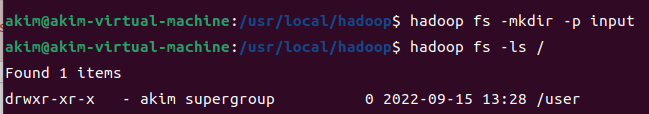
通过CSDN的查阅，了解到实现wordcount的计数需要建立input和output文件夹，但在刚开始并不知道在哪个目录下建立，导致出现如图2的错误。



**图2 建立文件夹错误界面图**

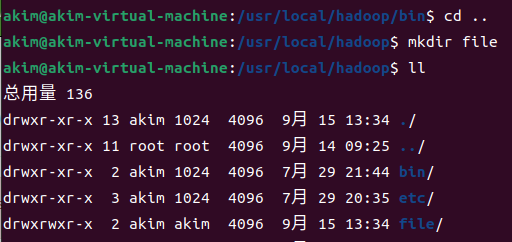
问题②：

输入**$hadoop fs -mkdir input**建立文件夹，却没有反应，通过查阅资料，发现输入**$hadoop fs -mkdir -p input**即可；通过**$hadoop fs -ls /**查看是否建立成功，却遇到了致命的问题，建立的文件夹名并不是**input**，而是一个名为**user**的文件夹，这就为后续出现的问题做了铺垫。

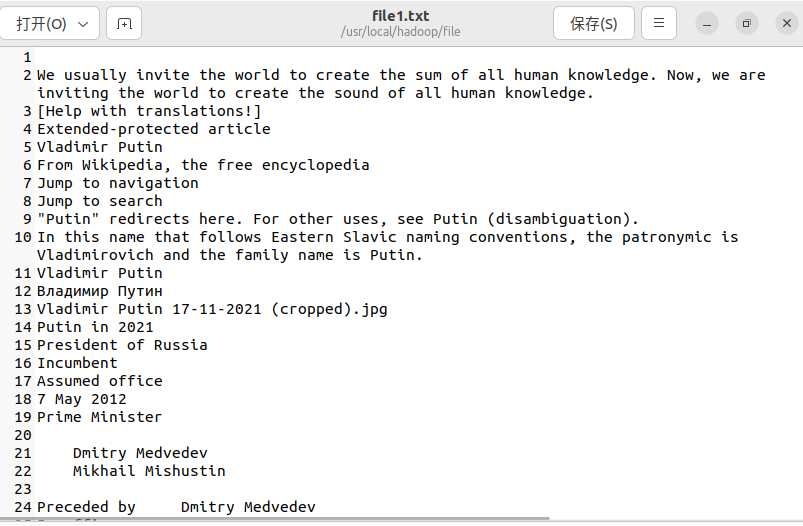


**图3 文件夹建立成功界面图**

（三）新建file文件夹用于建立file.txt，将https://en.wikipedia.org/wiki/Vladimir\_Putin网址中的内容存入file.txt



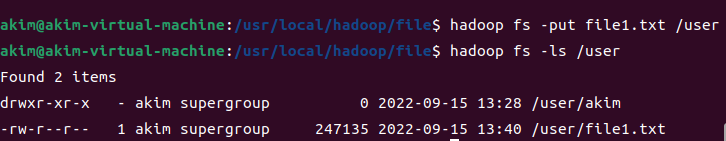
**图4 file文件夹建立界面图**



**图5 将文本写入file.txt界面图**

（四）将文件放入hadoop中的user文件夹中（应为input文件夹）

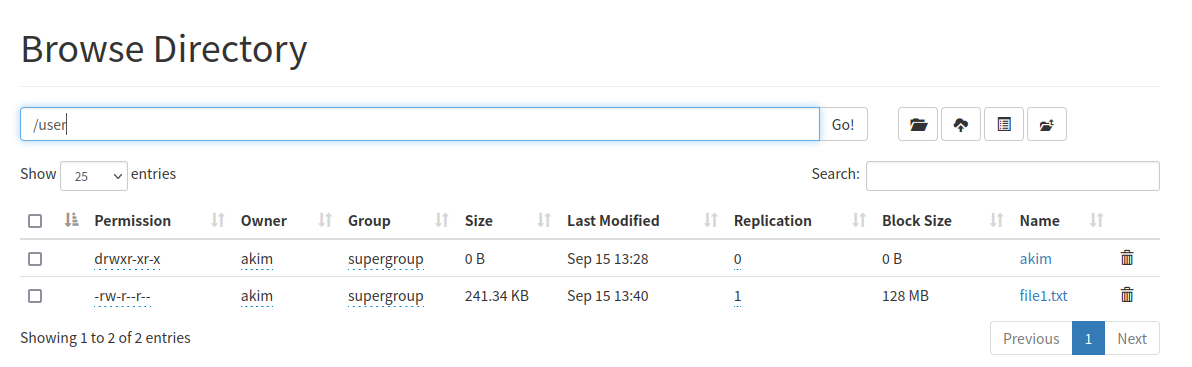
通过代码**$hadoop fs -put file.txt /user**，用**$hadoop fs -ls /user**查看是否推送成功。



**图 6 检查推送结果界面图**

由图6可知，file1.txt文件已放入user文件夹。

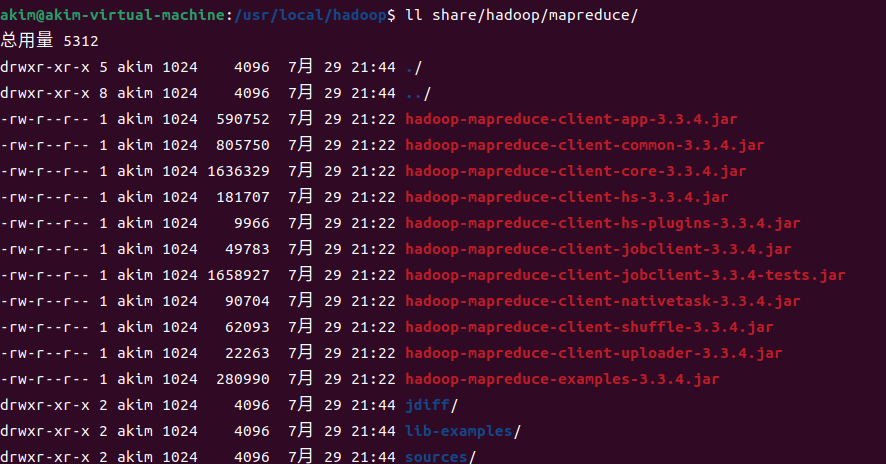
通过Web查看，如图7所示。



**图7 Web界面查看图**

（五）查看并使用hadoop-mapreduce-example.jar包

通过代码**$ ll share/hadoop/mapreduce**查看到，存在**hadoop-mapreduce-example-3.3.4.jar**包。



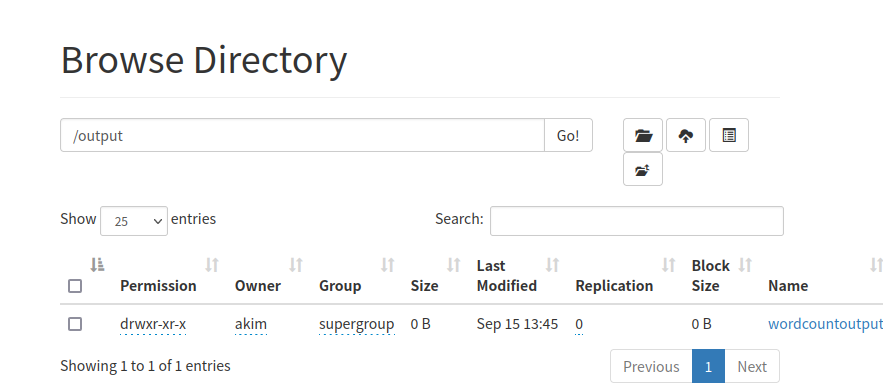
**图8 查看jar包界面图**

使用这个包执行wordcount，通过如下代码：

**$hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-example-3.3.4.jar wordcount /user/ /output/wordcountoutput**（这行命令的意思是调用**hadoop-mapreduce-example.jar**这个jar包并使用wordcount方法，使用hadoop中的user文件中的内容，将结果输出到output中的wordcountoutput文件中）

问题①：

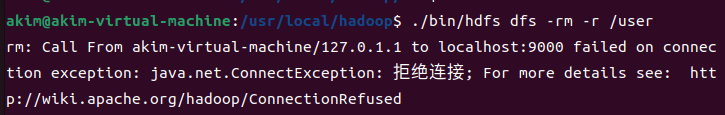
通过Web查看，并未在wordcountoutput文件夹中出现\_SUCCESS和part-r-0000文件，这就说明前面某一环节导致user中的文件并没有写入output中，于是我将user文件夹和output文件夹全部删除并重新试验。



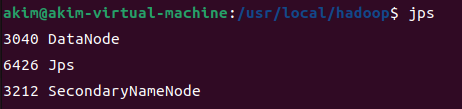
**图9 wordcountoutput为空界面图**

问题②：

通过**$ ./bin/hdfs dfs -rm -r /user**，将user文件夹删除时，出现如下图的错误，网上并未查到原因，后来通过jps查看到，NameNode进程被不知为何杀掉了，导致Web界面也无法打开，进而导致出现该问题，于是我重新启动守护进程，可重新开启Web界面。

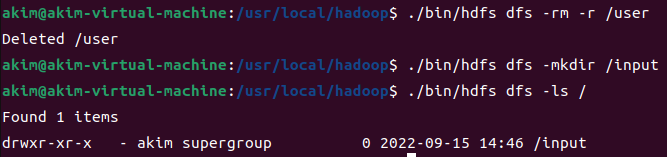


**图10 错误界面图**

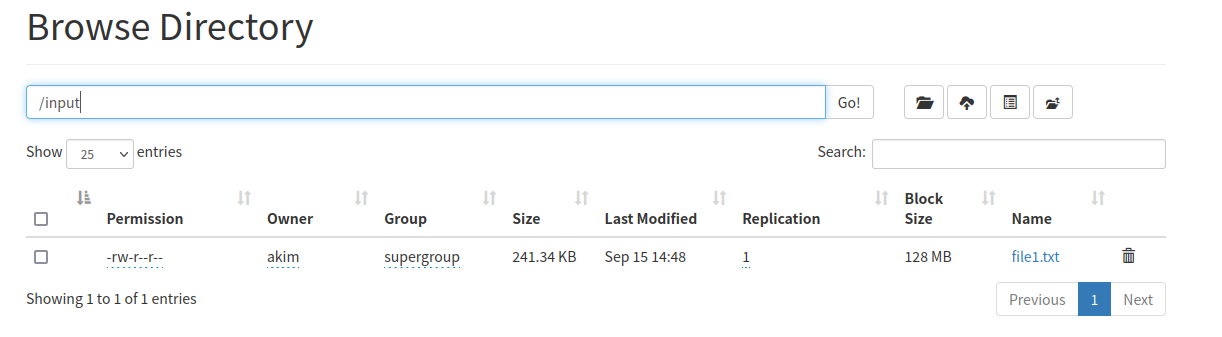


**图11 jps检测进程界面图**

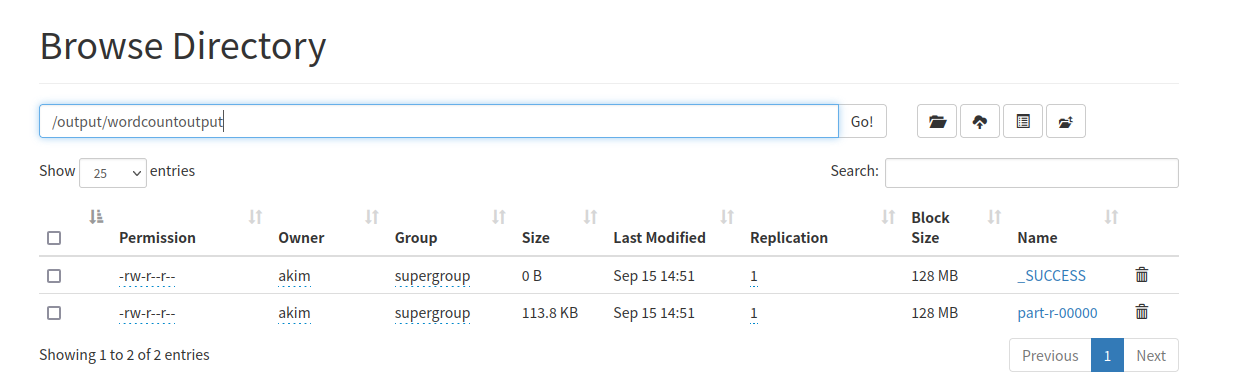
重新删除user文件后，通过查阅资料，选择另一种代码进行文件夹的建立，即**$ ./bin/hdfs dfs -mkdir /input**，并用**$ ./bin/hdfs dfs -ls** /查看到input文件终于成功建立。并且在Web界面成功看到\_SUCCESS和part-r-0000文件，这表明input中的文件成功写入output文件中。



**图12 重新建立input文件夹界面图**



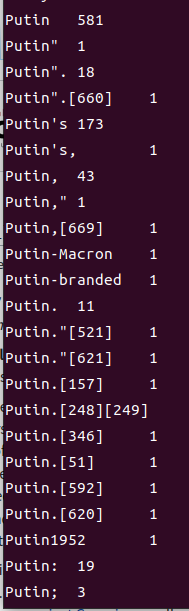
**图13 成功界面图1**



**图14 成功界面图2**

（六）查看统计结果

通过**$hadoop fs -cat /output/wordcountoutput/\***，查看到结果，并找到Putin的次数。



**图15 Putin次数统计界面图**