WORDCOUNT PROBLEM ON HADOOP REPORT

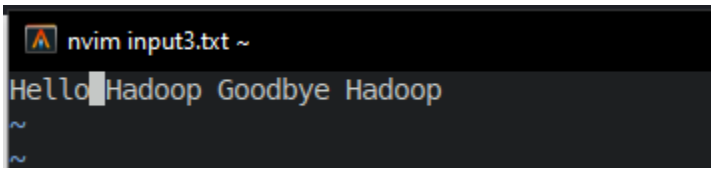
1. Set up:

```
tml_21127642@  > (dragonfarm) ~/hadoop/hadoop-3.3.6) nvim WordCount.java
tml_21127642@  > (dragonfarm) ~/hadoop/hadoop-3.3.6) jar cf wc.jar WordCount*.class
tml_21127642@  > (dragonfarm) ~/hadoop/hadoop-3.3.6) ls
 LICENSE-binary   README.txt                      WordCount.java   lib/              sbin/
 LICENSE.txt     'WordCount$IntSumReducer.class'  bin/             libexec/          share/
 NOTICE-binary   'WordCount$TokenizerMapper.class' etc/            licenses-binary/  wc.jar
 NOTICE.txt       WordCount.class                 include/         logs/
tml_21127642@  > (dragonfarm) ~/hadoop/hadoop-3.3.6) hdfs dfs -mkdir /input
tml_21127642@  > (dragonfarm) ~/hadoop/hadoop-3.3.6) cd
tml_21127642@  > (dragonfarm) ~) nvim input3.txt
```

- **Step 1:** Create a file named “WordCount.java”

Note: the nvim command will both create a file and open a code editor neovim

- **Step 2:** Copy code from the tutorial into the file
- **Step 3:** Compile the java file into jar file (wc.jar)
- **Step 4:** Create a input folder for our input file
- **Step 5:** Create a input file (input3.txt)



The content of the input file (input3.txt)

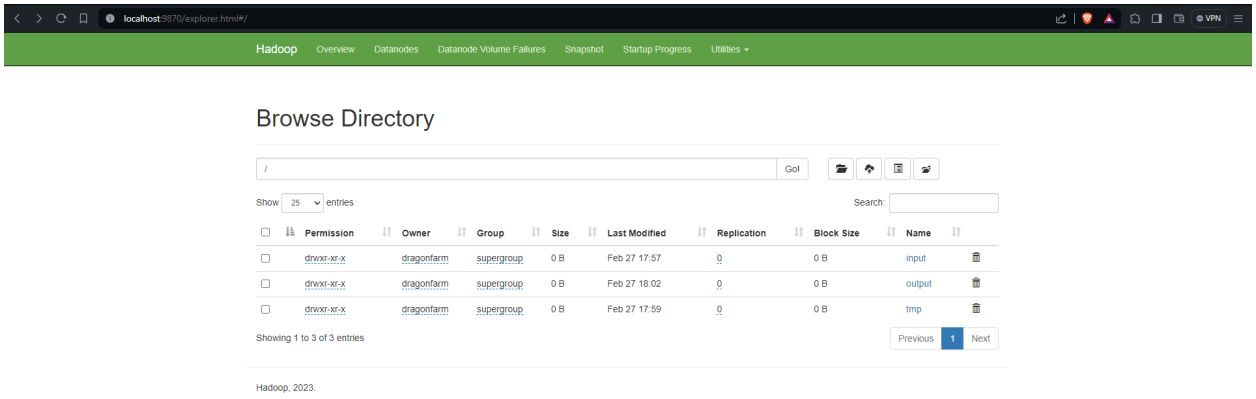
2. Run the application:

```
tml_21127642@  > (dragonfarm) ~/hadoop/hadoop-3.3.6) cd
tml_21127642@  > (dragonfarm) ~) nvim input3.txt
tml_21127642@  > (dragonfarm) ~) hdfs dfs -put input3.txt /input
tml_21127642@  > (dragonfarm) ~) hdfs dfs -cat /input/input3.txt
Hello Hadoop Goodbye Hadoop
tml_21127642@  > (dragonfarm) ~)
```

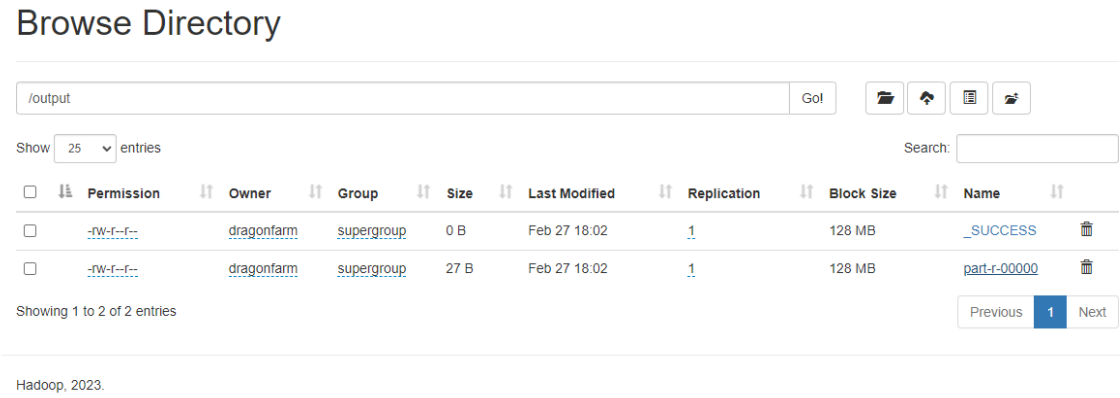
```
tml_21127642@  > (dragonfarm) ~/hadoop/hadoop-3.3.6) bin/hadoop jar wc.jar WordCount /input /output
2024-02-27 18:02:18,283 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-02-27 18:02:18,531 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
 interface and execute your application with ToolRunner to remedy this.
2024-02-27 18:02:18,552 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dragonfa
rm/.staging/job_1709031098048_0002
2024-02-27 18:02:18,756 INFO input.FileInputFormat: Total input files to process : 1
2024-02-27 18:02:19,613 INFO mapreduce.JobSubmitter: number of splits:1
2024-02-27 18:02:19,698 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1709031098048_0002
2024-02-27 18:02:19,699 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-02-27 18:02:19,820 INFO conf.Configuration: resource-types.xml not found
2024-02-27 18:02:19,821 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-02-27 18:02:19,863 INFO impl.YarnClientImpl: Submitted application application_1709031098048_0002
2024-02-27 18:02:19,888 INFO mapreduce.Job: The url to track the job: http://DESKTOP-DLGRKJK.:8088/proxy/application_17090310980
48_0002/
2024-02-27 18:02:19,888 INFO mapreduce.Job: Running job: job_1709031098048_0002
```

- **Step 6**: Put the input3.txt file into the input folder
- **Step 7:** Go into the hadoop folder and run the application

3. Checking the result:

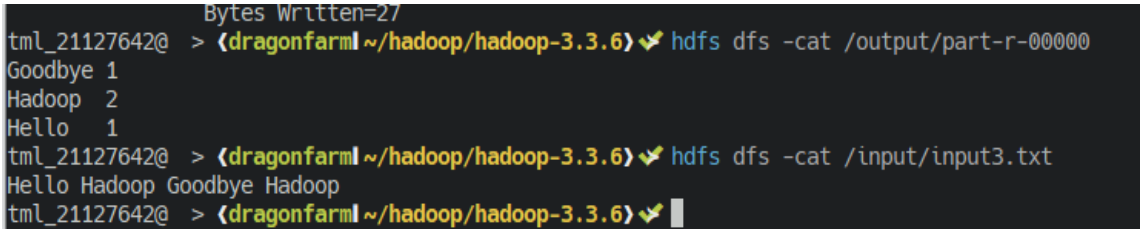


Website should have both input folder and output folder



Output folder content

- **Step 8:** Go into your terminal.
- **Step 9:** Enter command **"hdfs dfs -cat /output/part-r-00000".**
- **Step 10:** Now it should show the result like the image below.



The result

4. Source:

- https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html#Example%3A_WordCount_v1.0 – ("Apache Hadoop 3.3.6 – MapReduce Tutorial")