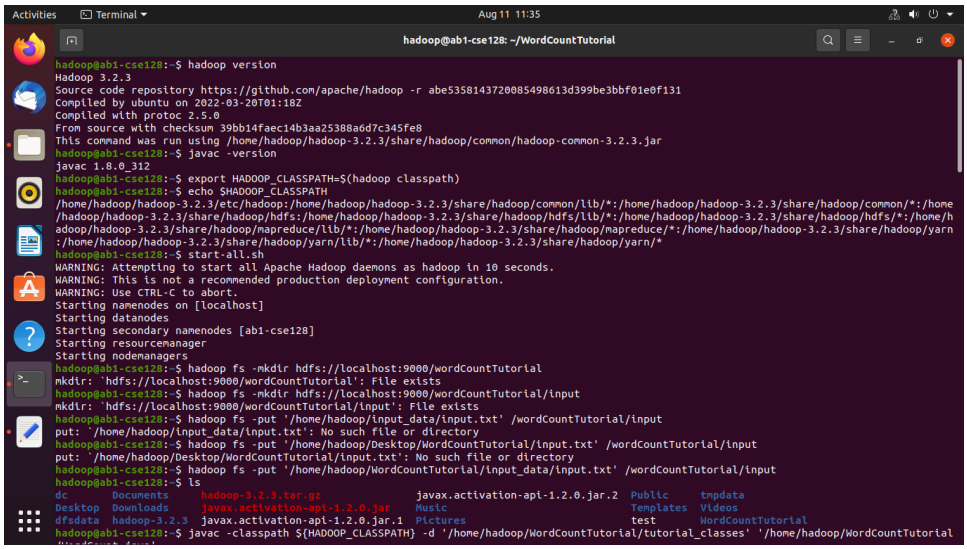
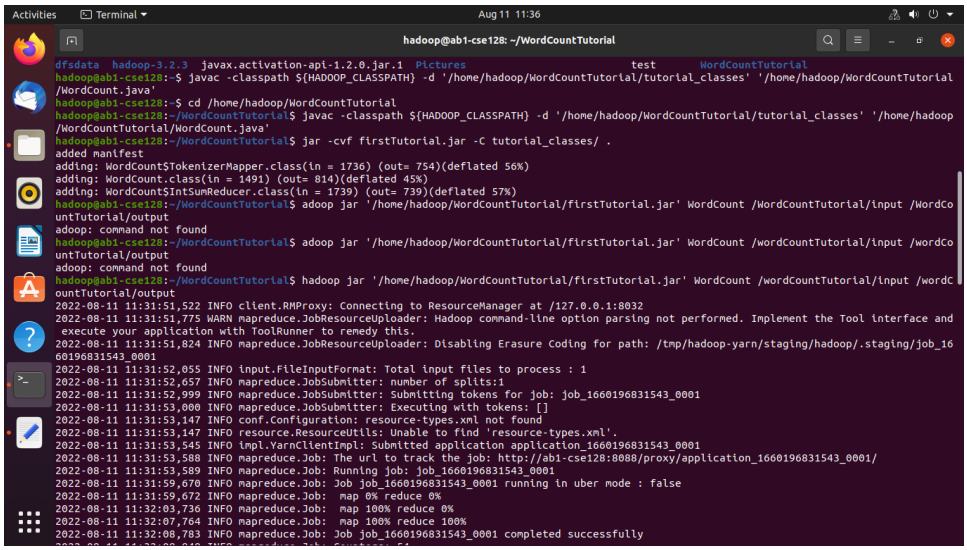
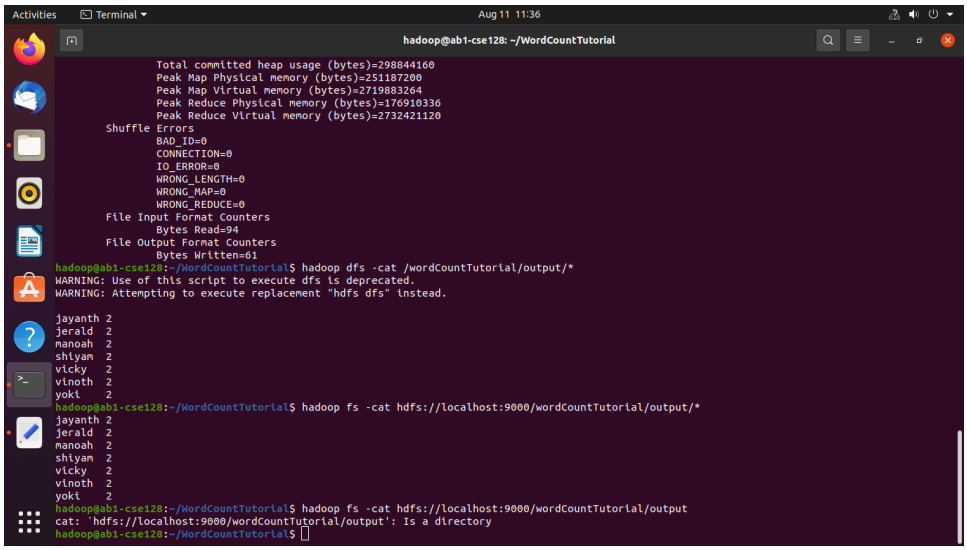
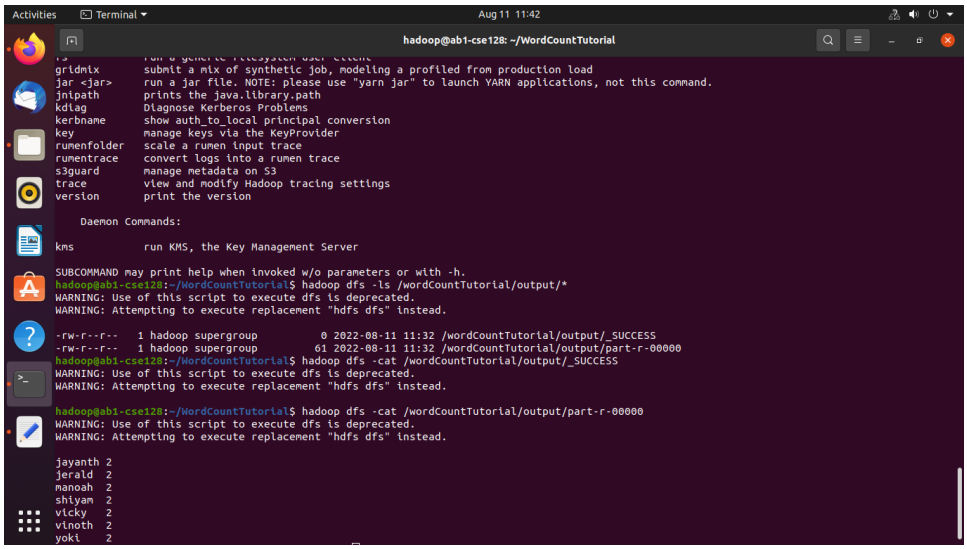
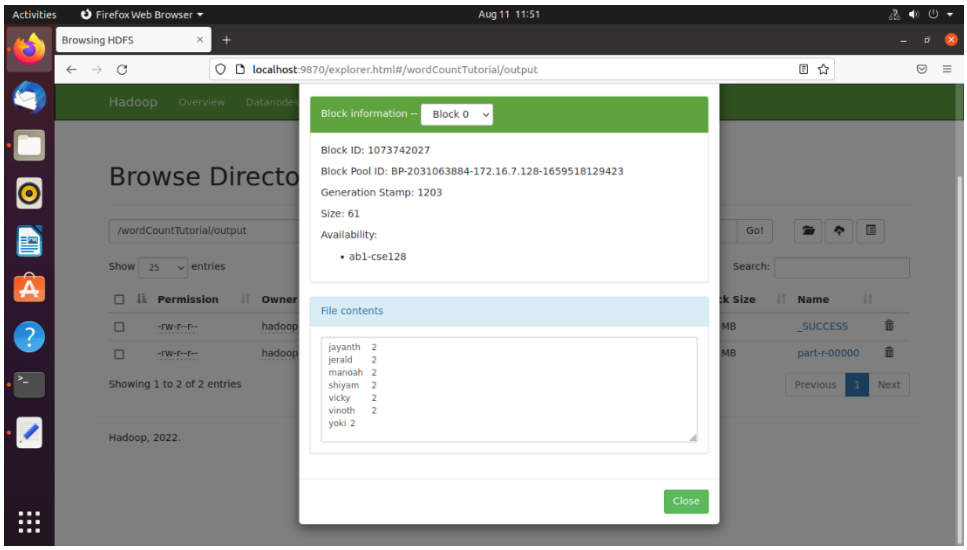
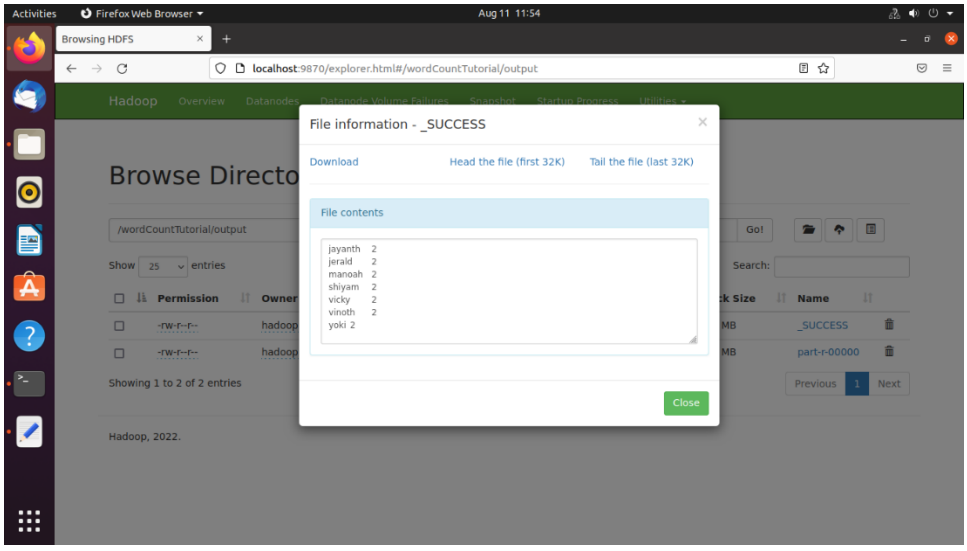

Activities
Terminal
Aug 11 11:35
hadoop@ab1-cse128: ~/WordCountTutorial
hadoop@ab1-cse128:~$ hadoop version
Hadoop 3.2.3
Source code repository https://github.com/apache/hadoop -r abe5358143720085498613d399be3bbf01e0f131
Compiled by ubuntu on 2022-03-20T01:18Z
Compiled with protoc 2.5.0
From source with checksum 39bb14faec14b3aa25388a6d7c345fe8
This command was run using /home/hadoop/hadoop-3.2.3/share/hadoop/common/hadoop-common-3.2.3.jar
hadoop@ab1-cse128:~$ javac -version
javac 1.8.0_312
hadoop@ab1-cse128:~$ export HADOOP_CLASSPATH=$(hadoop classpath)
hadoop@ab1-cse128:~$ echo $HADOOP_CLASSPATH
/home/hadoop/hadoop-3.2.3/etc/hadoop:/home/hadoop/hadoop-3.2.3/share/hadoop/common/lib/*:/home/hadoop/hadoop-3.2.3/share/hadoop/common/*:/home/hadoop/hadoop-3.2.3/share/hadoop/hdfs:/home/hadoop/hadoop-3.2.3/share/hadoop/hdfs/lib/*:/home/hadoop/hadoop-3.2.3/share/hadoop/hdfs/*:/home/hadoop/hadoop-3.2.3/share/hadoop/mapreduce/lib/*:/home/hadoop/hadoop-3.2.3/share/hadoop/mapreduce/*:/home/hadoop/hadoop-3.2.3/share/hadoop/yarn:/home/hadoop/hadoop-3.2.3/share/hadoop/yarn/lib/*:/home/hadoop/hadoop-3.2.3/share/hadoop/yarn/*
hadoop@ab1-cse128:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ab1-cse128]
Starting resourcemanager
Starting nodemanagers
hadoop@ab1-cse128:~$ hadoop fs -mkdir hdfs://localhost:9000/wordCountTutorial
mkdir: `hdfs://localhost:9000/wordCountTutorial': File exists
hadoop@ab1-cse128:~$ hadoop fs -mkdir hdfs://localhost:9000/wordCountTutorial/input
mkdir: `hdfs://localhost:9000/wordCountTutorial/input': File exists
hadoop@ab1-cse128:~$ hadoop fs -put '/home/hadoop/input_data/input.txt' /wordCountTutorial/input
put: `/home/hadoop/input_data/input.txt': No such file or directory
hadoop@ab1-cse128:~$ hadoop fs -put '/home/hadoop/Desktop/WordCountTutorial/input.txt' /wordCountTutorial/input
put: `/home/hadoop/Desktop/WordCountTutorial/input.txt': No such file or directory
hadoop@ab1-cse128:~$ hadoop fs -put '/home/hadoop/WordCountTutorial/input_data/input.txt' /wordCountTutorial/input
hadoop@ab1-cse128:~$ ls
dc Documents hadoop-3.2.3.tar.gz javax.activation-api-1.2.0.jar.2 Public tmpdata
Desktop Downloads javax.activation-api-1.2.0.jar Music Templates Videos
dfsdata hadoop-3.2.3 javax.activation-api-1.2.0.jar.1 Pictures test WordCountTutorial
hadoop@ab1-cse128:~$ javac -classpath ${HADOOP_CLASSPATH} -d '/home/hadoop/WordCountTutorial/tutorial_classes' '/home/hadoop/WordCountTutorial


Activities
Terminal
Aug 11 11:36
hadoop@ab1-cse128: ~/WordCountTutorial
dfsdata hadoop-3.2.3 javax.activation-api-1.2.0.jar.1 Pictures test WordCountTutorial
hadoop@ab1-cse128:~$ javac -classpath ${HADOOP_CLASSPATH} -d '/home/hadoop/WordCountTutorial/tutorial_classes' '/home/hadoop/WordCountTutorial/WordCount.java'
hadoop@ab1-cse128:~$ cd /home/hadoop/WordCountTutorial
hadoop@ab1-cse128:~/WordCountTutorial$ javac -classpath ${HADOOP_CLASSPATH} -d '/home/hadoop/WordCountTutorial/tutorial_classes' '/home/hadoop/WordCountTutorial/WordCount.java'
hadoop@ab1-cse128:~/WordCountTutorial$ jar -cvf firstTutorial.jar -C tutorial_classes/ .
added manifest
adding: WordCount$TokenizerMapper.class(in = 1736) (out= 754)(deflated 56%)
adding: WordCount.class(in = 1491) (out= 814)(deflated 45%)
adding: WordCount$IntSumReducer.class(in = 1739) (out= 739)(deflated 57%)
hadoop@ab1-cse128:~/WordCountTutorial$ adoop jar '/home/hadoop/WordCountTutorial/firstTutorial.jar' WordCount /WordCountTutorial/input /WordCountTutorial/output
adoop: command not found
hadoop@ab1-cse128:~/WordCountTutorial$ adoop jar '/home/hadoop/WordCountTutorial/firstTutorial.jar' WordCount /wordCountTutorial/input /wordCountTutorial/output
adoop: command not found
hadoop@ab1-cse128:~/WordCountTutorial$ hadoop jar '/home/hadoop/WordCountTutorial/firstTutorial.jar' WordCount /wordCountTutorial/input /wordCountTutorial/output
2022-08-11 11:31:51,522 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-08-11 11:31:51,775 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2022-08-11 11:31:51,824 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1660196831543_0001
2022-08-11 11:31:52,055 INFO input.FileInputFormat: Total input files to process : 1
2022-08-11 11:31:52,057 INFO mapreduce.JobSubmitter: number of splits:1
2022-08-11 11:31:52,999 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1660196831543_0001
2022-08-11 11:31:53,000 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-08-11 11:31:53,147 INFO conf.Configuration: resource-types.xml not found
2022-08-11 11:31:53,147 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-08-11 11:31:53,545 INFO impl.YarnClientImpl: Submitted application application_1660196831543_0001
2022-08-11 11:31:53,588 INFO mapreduce.Job: The url to track the job: http://ab1-cse128:8088/proxy/application_1660196831543_0001/
2022-08-11 11:31:53,589 INFO mapreduce.Job: Running job: job_1660196831543_0001
2022-08-11 11:31:59,670 INFO mapreduce.Job: Job job_1660196831543_0001 running in uber mode : false
2022-08-11 11:31:59,672 INFO mapreduce.Job: map 0% reduce 0%
2022-08-11 11:32:03,736 INFO mapreduce.Job: map 100% reduce 0%
2022-08-11 11:32:07,764 INFO mapreduce.Job: map 100% reduce 100%
2022-08-11 11:32:08,783 INFO mapreduce.Job: Job job_1660196831543_0001 completed successfully


Activities
Terminal
Aug 11 11:36
hadoop@ab1-cse128: ~/WordCountTutorial
Total committed heap usage (bytes)=298844160
Peak Map Physical memory (bytes)=251187200
Peak Map Virtual memory (bytes)=2719883264
Peak Reduce Physical memory (bytes)=176910336
Peak Reduce Virtual memory (bytes)=2732421120
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=94
File Output Format Counters
Bytes Written=61
hadoop@ab1-cse128:~/WordCountTutorial$ hadoop dfs -cat /wordCountTutorial/output/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
jayanth 2
jerald 2
manoah 2
shiyam 2
vicky 2
vinoth 2
yoki 2
hadoop@ab1-cse128:~/WordCountTutorial$ hadoop fs -cat hdfs://localhost:9000/wordCountTutorial/output/*
jayanth 2
jerald 2
manoah 2
shiyam 2
vicky 2
vinoth 2
yoki 2
hadoop@ab1-cse128:~/WordCountTutorial$ hadoop fs -cat hdfs://localhost:9000/wordCountTutorial/output
cat: `hdfs://localhost:9000/wordCountTutorial/output': Is a directory
hadoop@ab1-cse128:~/WordCountTutorial$


Activities
Terminal
Aug 11 11:42
hadoop@ab1-cse128: ~/WordCountTutorial
gridmix submit a mix of synthetic job, modeling a profiled from production load
jar <jar> run a jar file. NOTE: please use "yarn jar" to launch YARN applications, not this command.
jnipath prints the java.library.path
kdiag Diagnose Kerberos Problems
kerbname show auth_to_local principal conversion
key manage keys via the KeyProvider
rumenfolder scale a rumen input trace
rumentrace convert logs into a rumen trace
s3guard manage metadata on S3
trace view and modify Hadoop tracing settings
version print the version
Daemon Commands:
kms run KMS, the Key Management Server
SUBCOMMAND may print help when invoked w/o parameters or with -h.
hadoop@ab1-cse128:~/WordCountTutorial$ hadoop dfs -ls /wordCountTutorial/output/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
-rw-r--r-- 1 hadoop supergroup 0 2022-08-11 11:32 /wordCountTutorial/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 61 2022-08-11 11:32 /wordCountTutorial/output/part-r-00000
hadoop@ab1-cse128:~/WordCountTutorial$ hadoop dfs -cat /wordCountTutorial/output/_SUCCESS
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
hadoop@ab1-cse128:~/WordCountTutorial$ hadoop dfs -cat /wordCountTutorial/output/part-r-00000
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
jayanth 2
jerald 2
manoah 2
shiyam 2
vicky 2
vinoth 2
yoki 2


Activities
Firefox Web Browser
Aug 11 11:51
Browsing HDFS
localhost:9870/explorer.html#/wordCountTutorial/output
Hadoop
Overview
Block information -- Block 0
Block ID: 1073742027
Block Pool ID: BP-2031063884-172.16.7.128-1659518129423
Generation Stamp: 1203
Size: 61
Availability:
ab1-cse128
File contents
jayanth 2
jerald 2
manoah 2
shiyam 2
vicky 2
vinoth 2
yoki 2
Close
Browse Directo
/wordCountTutorial/output
Go!
Show 25 entries
Search:
Permission
Owner
Name
-rw-r--r--
hadoop
_SUCCESS
part-r-00000
Showing 1 to 2 of 2 entries
Previous
1
Next
Hadoop, 2022.


Activities
Firefox Web Browser
Aug 11 11:54
Browsing HDFS
localhost:9870/explorer.html#/wordCountTutorial/output
Hadoop
Overview
File information - _SUCCESS
Download
Head the file (first 32K)
Tail the file (last 32K)
File contents
jayanth 2
jerald 2
manoah 2
shiyam 2
vicky 2
vinoth 2
yoki 2
Close
Browse Directo
/wordCountTutorial/output
Go!
Show 25 entries
Search:
Permission
Owner
Name
-rw-r--r--
hadoop
_SUCCESS
part-r-00000
Showing 1 to 2 of 2 entries
Previous
1
Next
Hadoop, 2022.

```
import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {
    public static class TokenizerMapper extends Mapper<Object, Text, Text,
IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        public void map(Object key, Text value, Context context) throws
IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }
    public static class IntSumReducer extends
Reducer<Text,IntWritable,Text,IntWritable> {
        private IntWritable result = new IntWritable();
        public void reduce(Text key, Iterable<IntWritable> values,Context
context) throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "word count");
        job.setJarByClass(WordCount.class);
        job.setMapperClass(TokenizerMapper.class);
```

```
        job.setCombinerClass(IntSumReducer.class);
        job.setReducerClass(IntSumReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}
```

COMMANDS

hadoop version

javac -version

export HADOOP_CLASSPATH=$(hadoop classpath)

echo $HADOOP_CLASSPATH

hadoop fs -mkdir /wordCountTutorial

start-all.sh

hadoop fs -mkdir hdfs://localhost:9000/wordCountTutorial

hadoop fs -mkdir hdfs://localhost:9000/wordCountTutorial/input

hadoop fs -put '/home/hadoop/input_data/input.txt' /wordCountTutorial/input

ls

javac -classpath ${HADOOP_CLASSPATH} -d '/home/hadoop/tutorial_classes' '/home/hadoop/WordCount.java'

jar -cvf firstTutorial.jar -C tutorial_classes/ .

hadoop jar '/home/hadoop/firstTutorial.jar' WordCount /wordCountTutorial/input /wordCountTutorial/output

hadoop dfs -cat /wordCountTutorial/output/*

hadoop fs -cat hdfs://localhost:9000/wordCountTutorial/output/*

hadoop fs -cat hdfs://localhost:9000/wordCountTutorial/output

hadoop dfs -cat /wordCountTutorial/output/*

hadoop hdfs -cat /wordCountTutorial/output/*

hadoop dfs -ls /wordCountTutorial/output/*

hadoop dfs -cat /wordCountTutorial/output/_SUCCESS

hadoop dfs -cat /wordCountTutorial/output/part-r-00000