Report

Screenshot of running code

BCSF17A505
Mian Muhammad Anwar
Compiler Construction

CS 460: Cloud Computing — Spring 2021
Punjab University College of Information Technology (PUCIT) — University of the Punjab
Assignment 3: Sentiment Analysis using Hadoop — **Instructor:** Waheed Iqbal & Muhammad Abdullah

---

Assignment 03: Sentiment Analysis using Hadoop

Sentiment analysis is a process to identify emotions behind any text. In this assignment, you have to build a simple sentiment analysis MapReduce job. Following are the important files to perform the task:

1. twitter.json.tar.gz: A compressed file containing around 1 million tweets in a JSON format
2. positive_words.txt: A file containing list of positive words
3. negative_words.txt: A file containing list of negative words
4. stop_words.txt: A file containing list of stop words

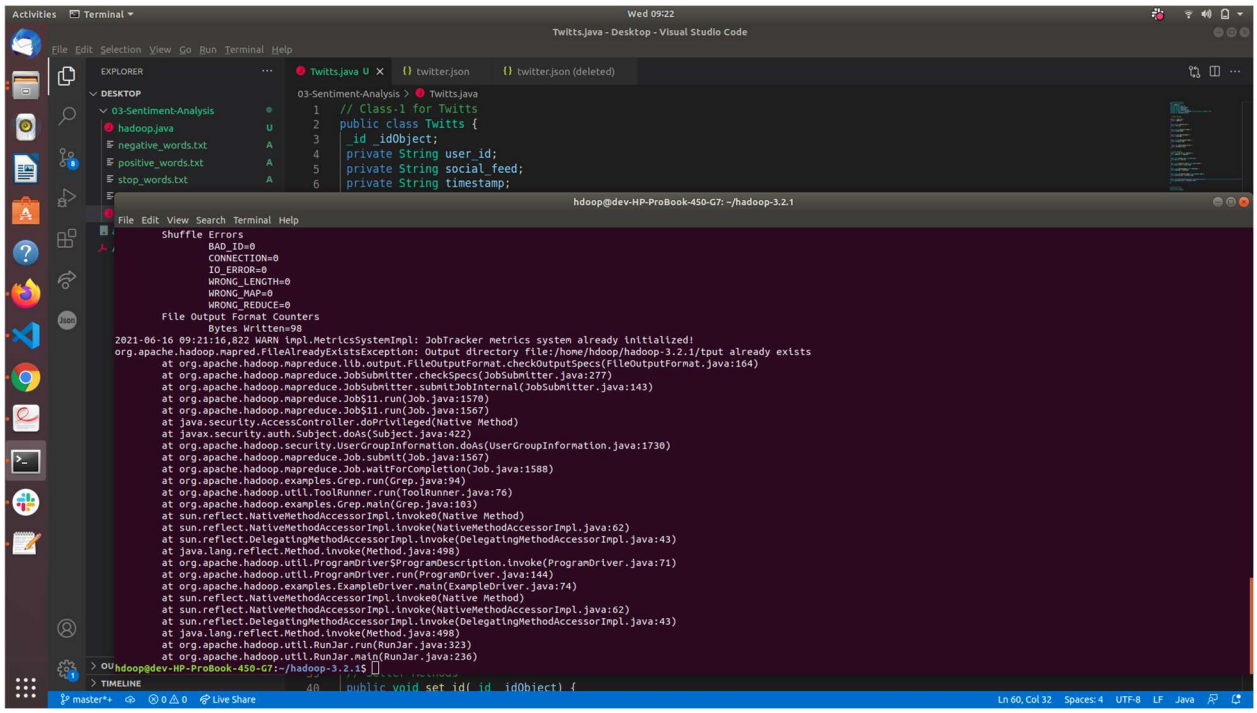
Your MapReduce job should label each tweet either positive, negative, or neutral. For each tweet i, you need to ignore stop words, calculate positive words count p_i, and negative words count n_i. Then use the following formula to calculate the sentiment score s_i:

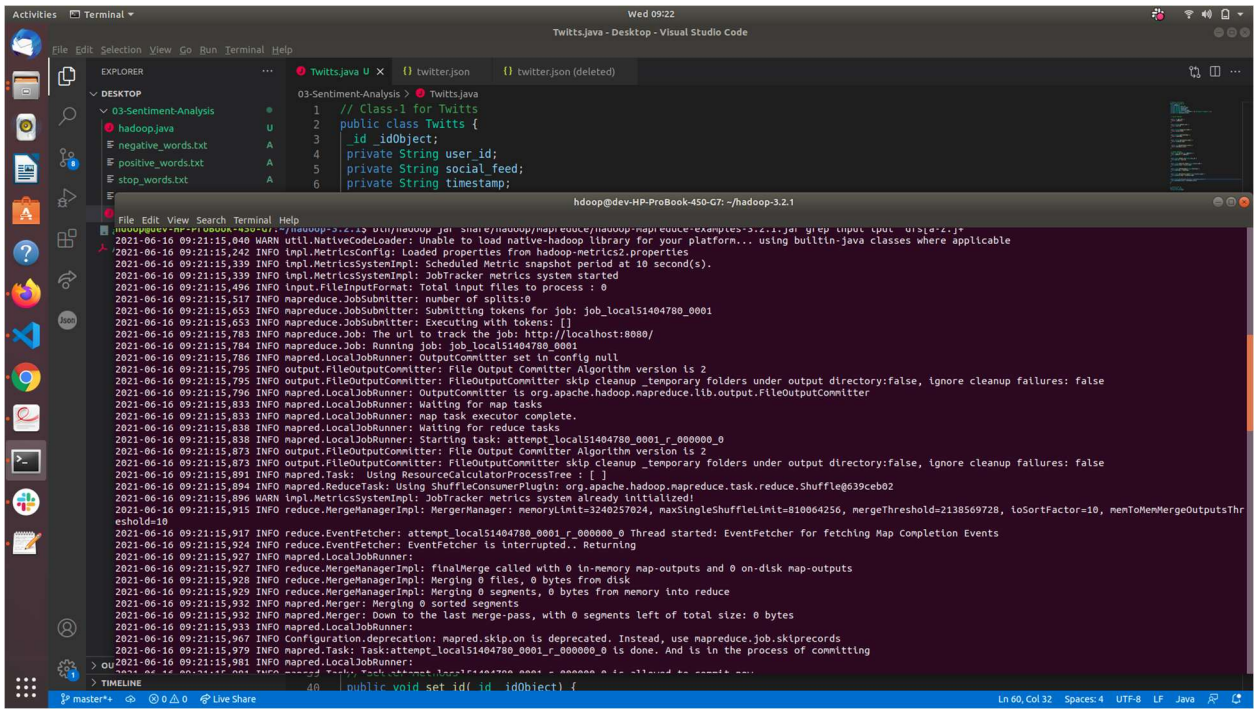
$$s_i = \frac{p_i - n_i}{p_i + n_i + 1} \tag{1}$$

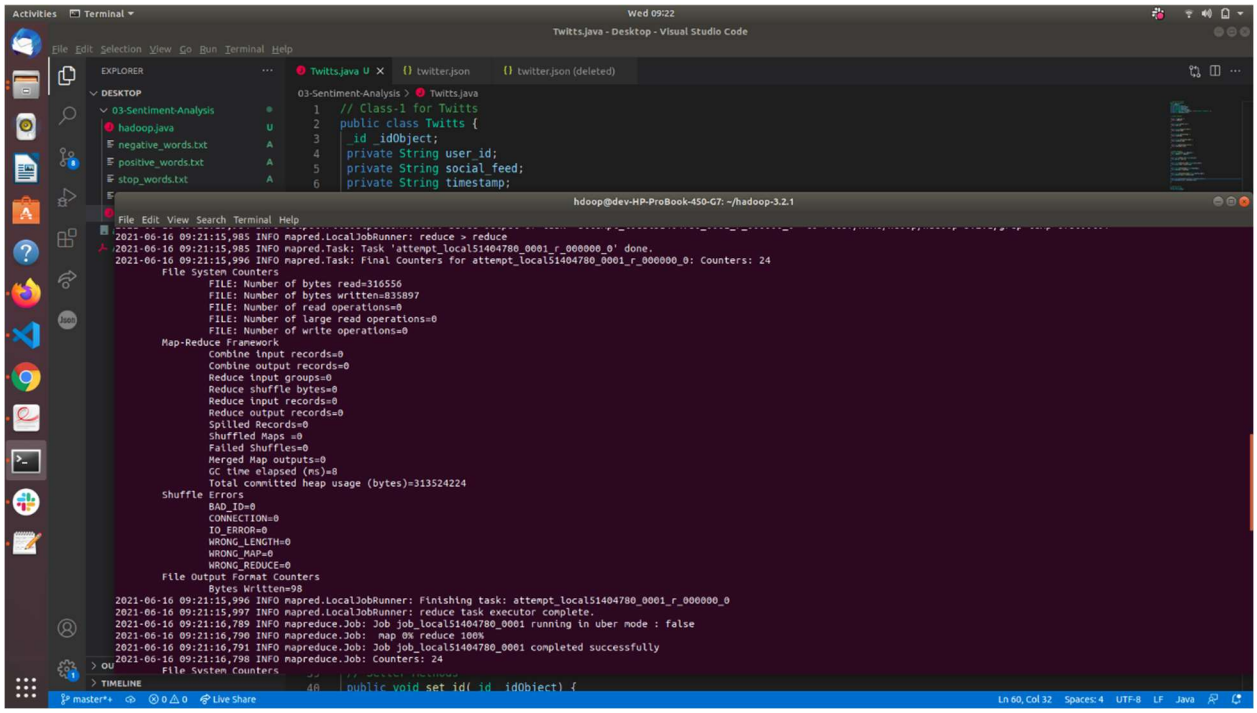
A tweet i will be positive if s_i is greater than 0, negative if s_i is less than 0, and neutral if s_i is equal to 0. You must keep the original JSON document for each tweet and append additional attributed named *sentiment*. You need to produce three separate files for each category in JSON format.

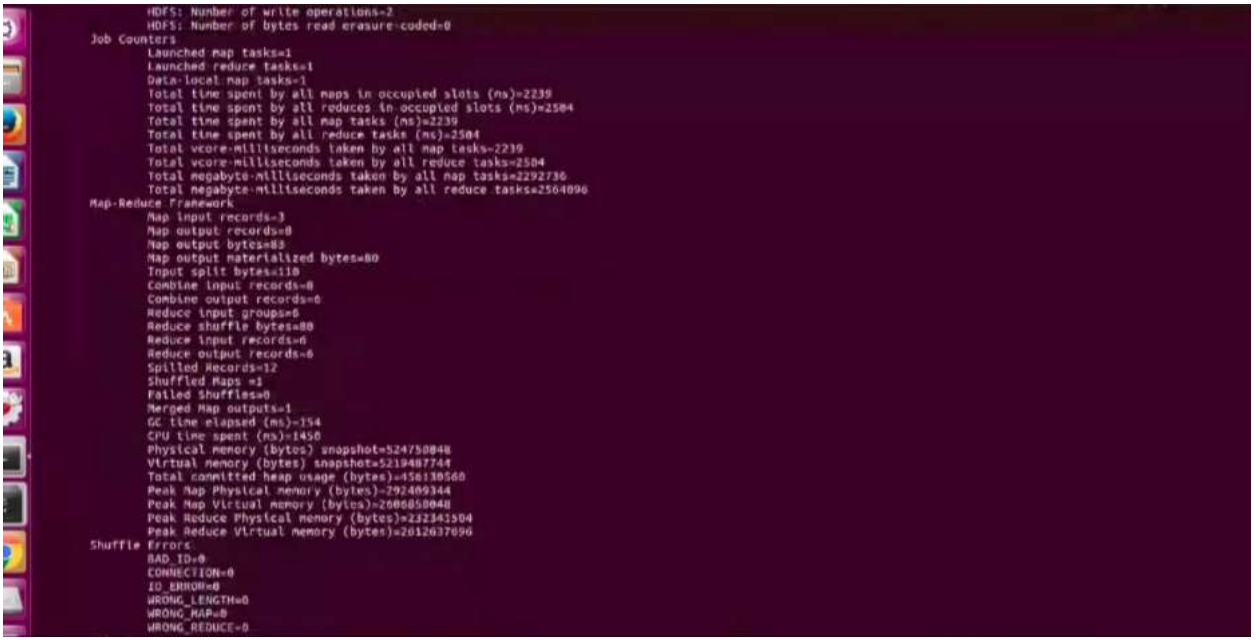
Submission Guidelines

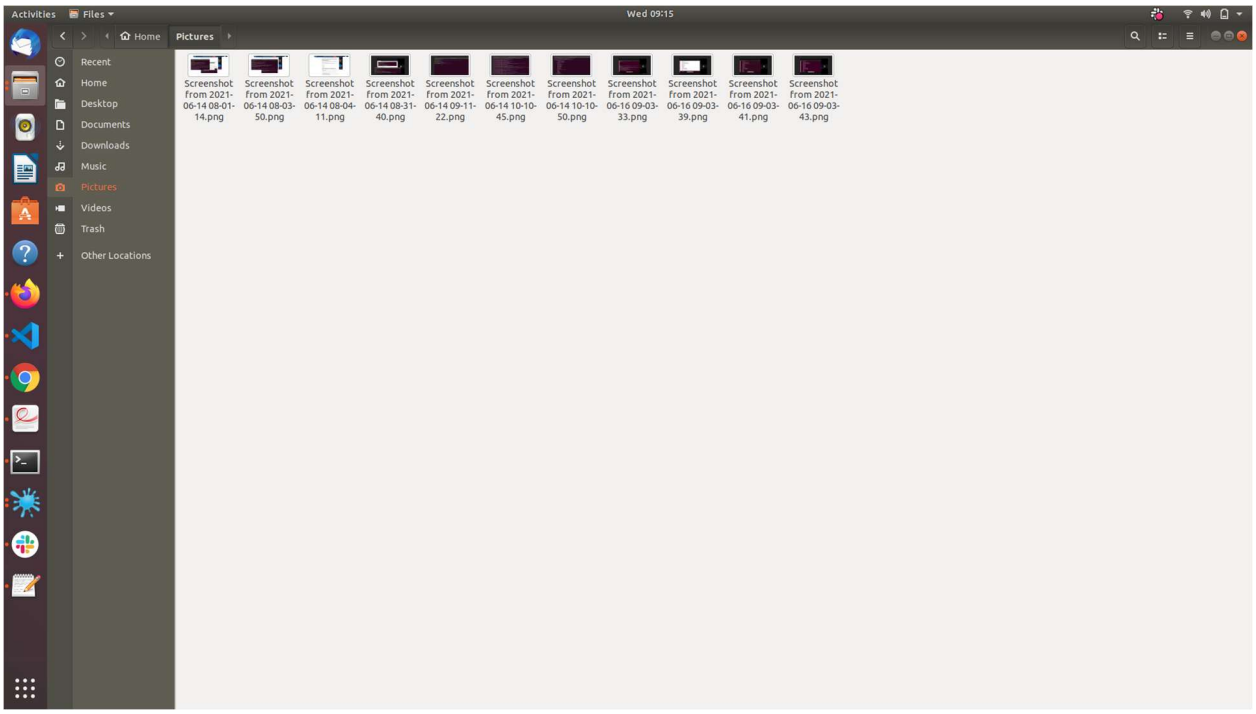
You need to prepare a small report contain screenshot of running your code, attach code in a zip file, and output file before **Tuesday 01 June, 11:59pm** and submit over Google Classroom.

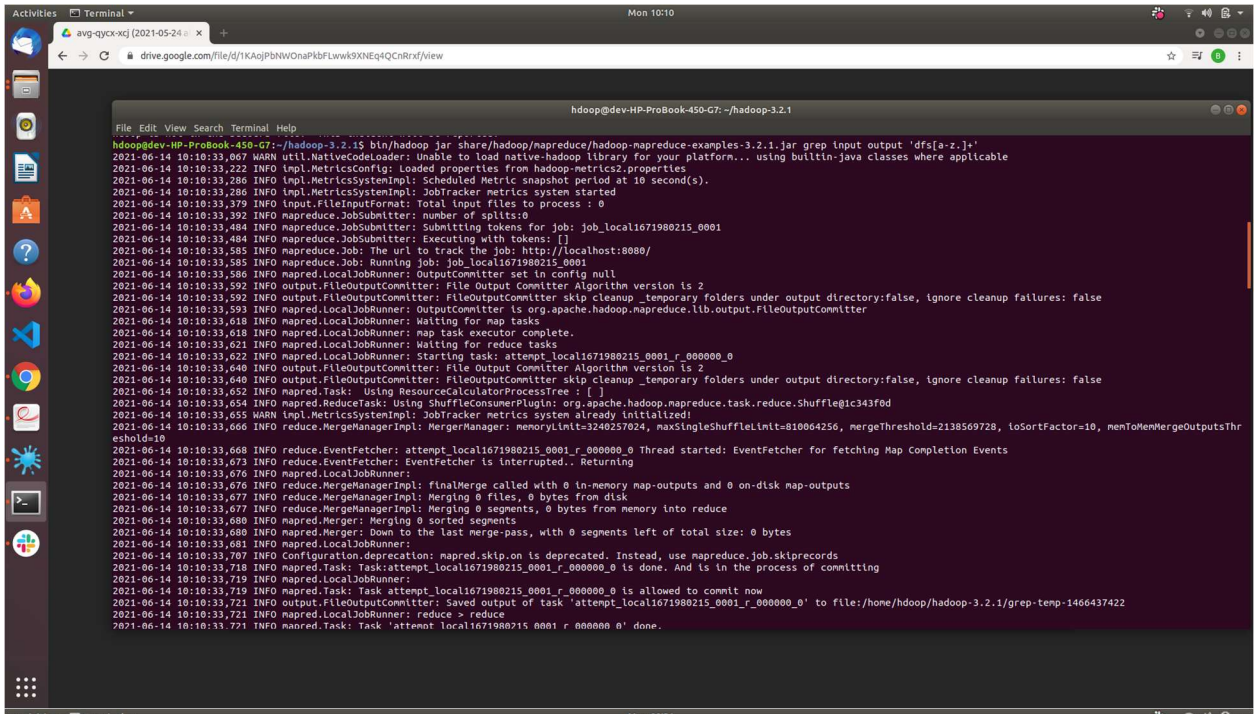

Activities
Terminal
Wed 09:22
Twitts.java - Desktop - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
DESKTOP
03-Sentiment-Analysis
hadoop.java
negative_words.txt
positive_words.txt
stop_words.txt
Twitts.java U
twitter.json
twitter.json (deleted)
03-Sentiment-Analysis > Twitts.java
1 // Class-1 for Twitts
2 public class Twitts {
3 _id _idObject;
4 private String user_id;
5 private String social_feed;
6 private String timestamp;
hdoop@dev-HP-ProBook-450-G7: ~/hadoop-3.2.1
File Edit View Search Terminal Help
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Output Format Counters
Bytes Written=98
2021-06-16 09:21:16,822 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory file:/home/hdoop/hadoop-3.2.1/tput already exists
at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:164)
at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:277)
at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1730)
at org.apache.hadoop.mapreduce.Job.submit(Job.java:1567)
at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1588)
at org.apache.hadoop.examples.Grep.run(Grep.java:94)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
at org.apache.hadoop.examples.Grep.main(Grep.java:103)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:71)
at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hdoop@dev-HP-ProBook-450-G7:~/hadoop-3.2.1$
OUTLINE
TIMELINE
40 public void set_id(_id _idObject) {
master*+
0 0
Live Share
Ln 60, Col 32
Spaces: 4
UTF-8
LF
Java

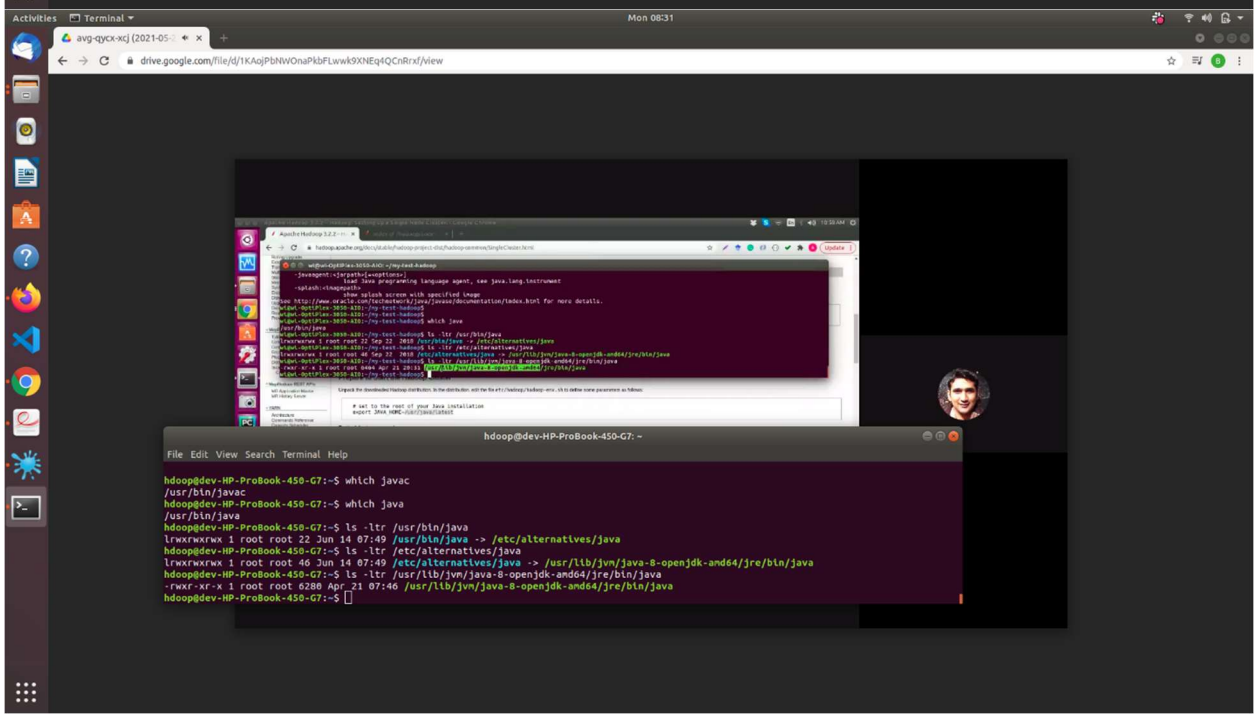

Activities
Terminal
Wed 09:22
Twitts.java - Desktop - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
DESKTOP
03-Sentiment-Analysis
hadoop.java
negative_words.txt
positive_words.txt
stop_words.txt
Twitts.java U
twitter.json
twitter.json (deleted)
03-Sentiment-Analysis > Twitts.java
1 // Class-1 for Twitts
2 public class Twitts {
3 _id _idObject;
4 private String user_id;
5 private String social_feed;
6 private String timestamp;
hdoop@dev-HP-ProBook-450-G7: ~/hadoop-3.2.1
File Edit View Search Terminal Help
2021-06-16 09:21:15,040 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-06-16 09:21:15,242 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2021-06-16 09:21:15,339 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2021-06-16 09:21:15,339 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2021-06-16 09:21:15,496 INFO input.FileInputFormat: Total input files to process : 0
2021-06-16 09:21:15,517 INFO mapreduce.JobSubmitter: number of splits:0
2021-06-16 09:21:15,653 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local51404780_0001
2021-06-16 09:21:15,653 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-06-16 09:21:15,783 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2021-06-16 09:21:15,784 INFO mapreduce.Job: Running job: job_local51404780_0001
2021-06-16 09:21:15,786 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2021-06-16 09:21:15,795 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-06-16 09:21:15,795 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2021-06-16 09:21:15,796 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2021-06-16 09:21:15,833 INFO mapred.LocalJobRunner: Waiting for map tasks
2021-06-16 09:21:15,833 INFO mapred.LocalJobRunner: map task executor complete.
2021-06-16 09:21:15,838 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2021-06-16 09:21:15,838 INFO mapred.LocalJobRunner: Starting task: attempt_local51404780_0001_r_000000_0
2021-06-16 09:21:15,873 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-06-16 09:21:15,873 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2021-06-16 09:21:15,891 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2021-06-16 09:21:15,894 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@639ceb02
2021-06-16 09:21:15,896 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2021-06-16 09:21:15,915 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=3240257024, maxSingleShuffleLimit=810064256, mergeThreshold=2138569728, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2021-06-16 09:21:15,917 INFO reduce.EventFetcher: attempt_local51404780_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2021-06-16 09:21:15,924 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2021-06-16 09:21:15,927 INFO mapred.LocalJobRunner:
2021-06-16 09:21:15,927 INFO reduce.MergeManagerImpl: finalMerge called with 0 in-memory map-outputs and 0 on-disk map-outputs
2021-06-16 09:21:15,928 INFO reduce.MergeManagerImpl: Merging 0 files, 0 bytes from disk
2021-06-16 09:21:15,929 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2021-06-16 09:21:15,932 INFO mapred.Merger: Merging 0 sorted segments
2021-06-16 09:21:15,932 INFO mapred.Merger: Down to the last merge-pass, with 0 segments left of total size: 0 bytes
2021-06-16 09:21:15,933 INFO mapred.LocalJobRunner:
2021-06-16 09:21:15,967 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2021-06-16 09:21:15,979 INFO mapred.Task: Task:attempt_local51404780_0001_r_000000_0 is done. And is in the process of committing
2021-06-16 09:21:15,981 INFO mapred.LocalJobRunner:
OUTLINE
TIMELINE
40 public void set_id(_id _idObject) {
master*+
0 0
Live Share
Ln 60, Col 32
Spaces: 4
UTF-8
LF
Java

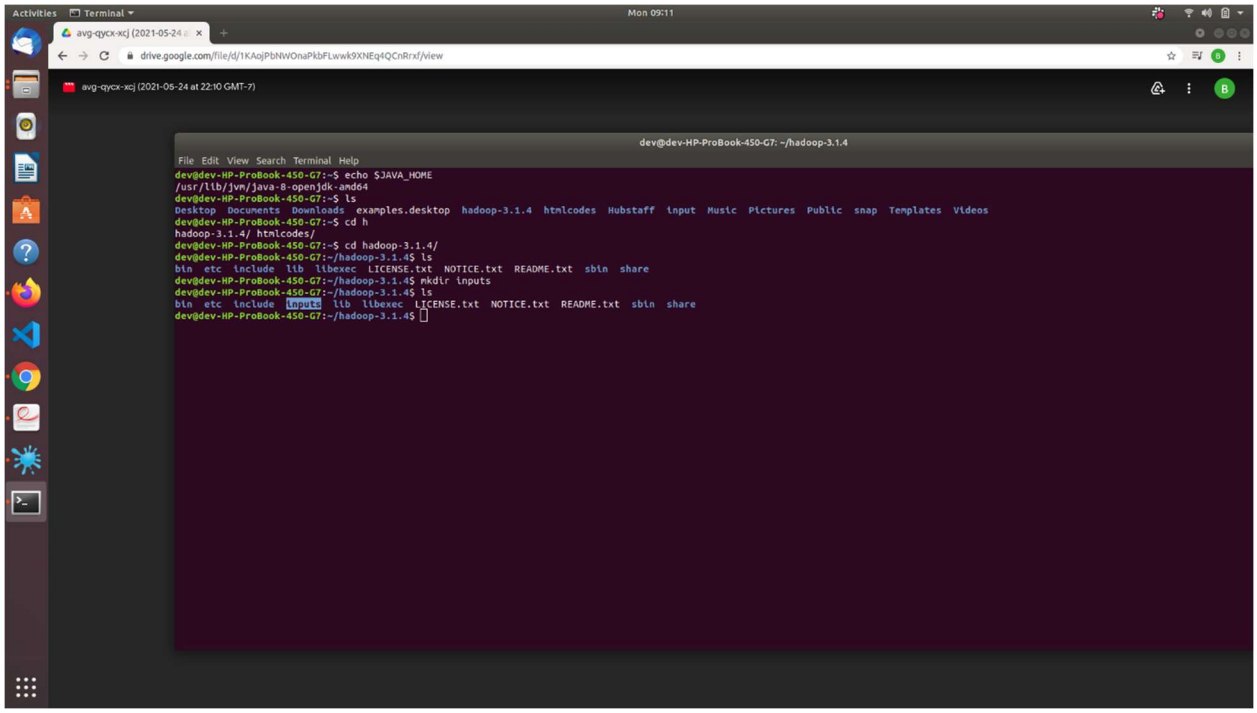

Activities
Terminal
Wed 09:22
Twitts.java - Desktop - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
DESKTOP
03-Sentiment-Analysis
hadoop.java
negative_words.txt
positive_words.txt
stop_words.txt
Twitts.java U
twitter.json
twitter.json (deleted)
03-Sentiment-Analysis > Twitts.java
1 // Class-1 for Twitts
2 public class Twitts {
3 _id _idObject;
4 private String user_id;
5 private String social_feed;
6 private String timestamp;
hdoop@dev-HP-ProBook-450-G7: ~/hadoop-3.2.1
File Edit View Search Terminal Help
2021-06-16 09:21:15,985 INFO mapred.LocalJobRunner: reduce > reduce
2021-06-16 09:21:15,985 INFO mapred.Task: Task 'attempt_local51404780_0001_r_000000_0' done.
2021-06-16 09:21:15,996 INFO mapred.Task: Final Counters for attempt_local51404780_0001_r_000000_0: Counters: 24
File System Counters
FILE: Number of bytes read=316556
FILE: Number of bytes written=835897
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
Map-Reduce Framework
Combine input records=0
Combine output records=0
Reduce input groups=0
Reduce shuffle bytes=0
Reduce input records=0
Reduce output records=0
Spilled Records=0
Shuffled Maps =0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=8
Total committed heap usage (bytes)=313524224
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Output Format Counters
Bytes Written=98
2021-06-16 09:21:15,996 INFO mapred.LocalJobRunner: Finishing task: attempt_local51404780_0001_r_000000_0
2021-06-16 09:21:15,997 INFO mapred.LocalJobRunner: reduce task executor complete.
2021-06-16 09:21:16,789 INFO mapreduce.Job: Job job_local51404780_0001 running in uber mode : false
2021-06-16 09:21:16,790 INFO mapreduce.Job: map 0% reduce 100%
2021-06-16 09:21:16,791 INFO mapreduce.Job: Job job_local51404780_0001 completed successfully
2021-06-16 09:21:16,798 INFO mapreduce.Job: Counters: 24
File System Counters
40 public void set_id(_id _idObject) {
TIMELINE
master*+
0 0
Live Share
Ln 60, Col 32
Spaces: 4
UTF-8
LF
Java

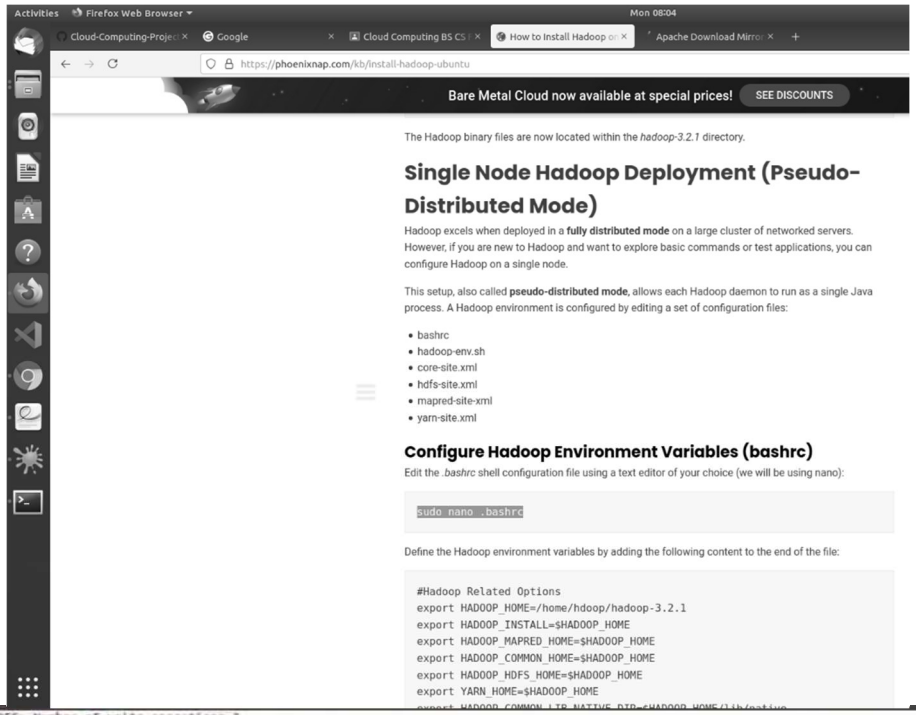

HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=2239
Total time spent by all reduces in occupied slots (ms)=2504
Total time spent by all map tasks (ms)=2239
Total time spent by all reduce tasks (ms)=2504
Total vcore-milliseconds taken by all map tasks=2239
Total vcore-milliseconds taken by all reduce tasks=2504
Total megabyte-milliseconds taken by all map tasks=2292736
Total megabyte-milliseconds taken by all reduce tasks=2564096
Map-Reduce Framework
Map input records=3
Map output records=8
Map output bytes=83
Map output materialized bytes=80
Input split bytes=110
Combine input records=8
Combine output records=6
Reduce input groups=6
Reduce shuffle bytes=80
Reduce input records=6
Reduce output records=6
Spilled Records=12
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=154
CPU time spent (ms)=1450
Physical memory (bytes) snapshot=524750848
Virtual memory (bytes) snapshot=5219487744
Total committed heap usage (bytes)=456130560
Peak Map Physical memory (bytes)=292409344
Peak Map Virtual memory (bytes)=2606850048
Peak Reduce Physical memory (bytes)=232341504
Peak Reduce Virtual memory (bytes)=2612637696
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

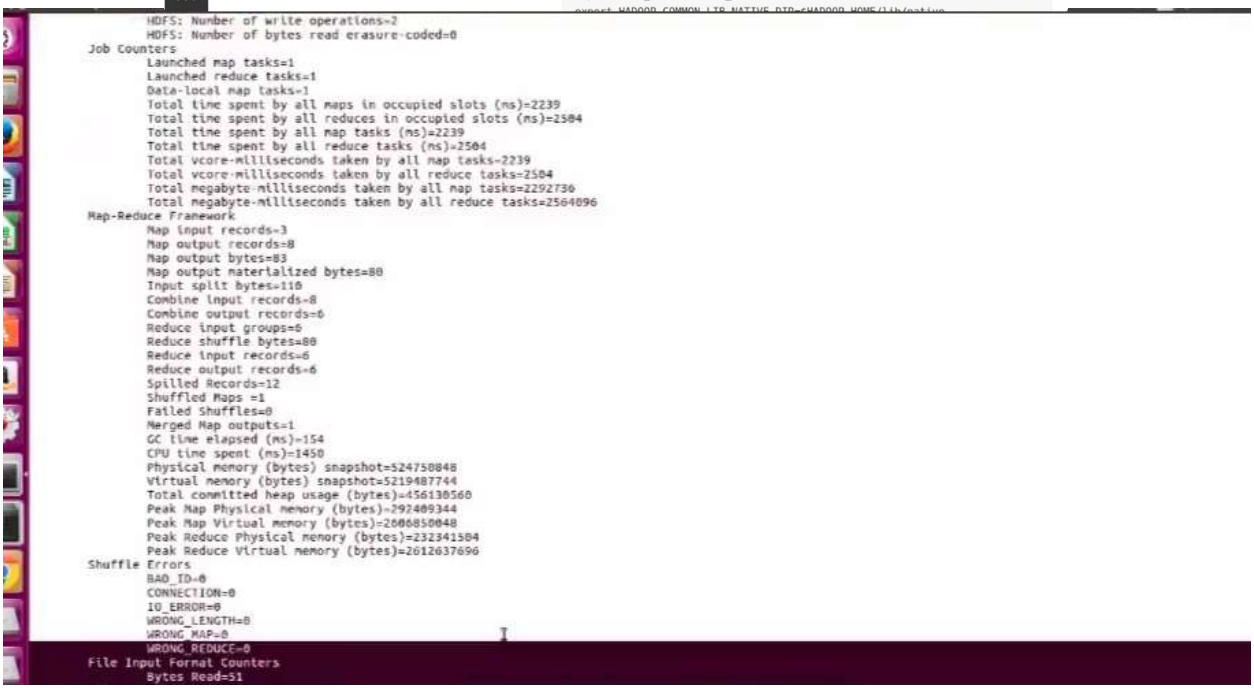

Activities
Files
Wed 09:15
Home
Pictures
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
Screenshot from 2021-06-14 08-01-14.png
Screenshot from 2021-06-14 08-03-50.png
Screenshot from 2021-06-14 08-04-11.png
Screenshot from 2021-06-14 08-31-40.png
Screenshot from 2021-06-14 09-11-22.png
Screenshot from 2021-06-14 10-10-45.png
Screenshot from 2021-06-14 10-10-50.png
Screenshot from 2021-06-16 09-03-33.png
Screenshot from 2021-06-16 09-03-39.png
Screenshot from 2021-06-16 09-03-41.png
Screenshot from 2021-06-16 09-03-43.png


Activities
Terminal
Mon 10:10
drive.google.com/file/d/1KAojPbNWOnaPkbFLwwk9XNEq4QCnRrxf/view
hdoop@dev-HP-ProBook-450-G7: ~/hadoop-3.2.1
File Edit View Search Terminal Help
hdoop@dev-HP-ProBook-450-G7:~/hadoop-3.2.1$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar grep input output 'dfs[a-z.]+'
2021-06-14 10:10:33,067 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-06-14 10:10:33,222 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2021-06-14 10:10:33,286 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2021-06-14 10:10:33,286 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2021-06-14 10:10:33,379 INFO input.FileInputFormat: Total input files to process : 0
2021-06-14 10:10:33,392 INFO mapreduce.JobSubmitter: number of splits:0
2021-06-14 10:10:33,484 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1671980215_0001
2021-06-14 10:10:33,484 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-06-14 10:10:33,585 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2021-06-14 10:10:33,585 INFO mapreduce.Job: Running job: job_local1671980215_0001
2021-06-14 10:10:33,586 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2021-06-14 10:10:33,592 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-06-14 10:10:33,592 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2021-06-14 10:10:33,593 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2021-06-14 10:10:33,618 INFO mapred.LocalJobRunner: Waiting for map tasks
2021-06-14 10:10:33,618 INFO mapred.LocalJobRunner: map task executor complete.
2021-06-14 10:10:33,621 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2021-06-14 10:10:33,622 INFO mapred.LocalJobRunner: Starting task: attempt_local1671980215_0001_r_000000_0
2021-06-14 10:10:33,640 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2021-06-14 10:10:33,640 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2021-06-14 10:10:33,652 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2021-06-14 10:10:33,654 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@1c343f0d
2021-06-14 10:10:33,655 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2021-06-14 10:10:33,666 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=3240257024, maxSingleShuffleLimit=810064256, mergeThreshold=2138569728, ioSortFactor=10, memToMemMergeOutputsThr
eshold=10
2021-06-14 10:10:33,668 INFO reduce.EventFetcher: attempt_local1671980215_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2021-06-14 10:10:33,673 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2021-06-14 10:10:33,676 INFO mapred.LocalJobRunner:
2021-06-14 10:10:33,676 INFO reduce.MergeManagerImpl: finalMerge called with 0 in-memory map-outputs and 0 on-disk map-outputs
2021-06-14 10:10:33,677 INFO reduce.MergeManagerImpl: Merging 0 files, 0 bytes from disk
2021-06-14 10:10:33,677 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2021-06-14 10:10:33,680 INFO mapred.Merger: Merging 0 sorted segments
2021-06-14 10:10:33,680 INFO mapred.Merger: Down to the last merge-pass, with 0 segments left of total size: 0 bytes
2021-06-14 10:10:33,681 INFO mapred.LocalJobRunner:
2021-06-14 10:10:33,707 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2021-06-14 10:10:33,718 INFO mapred.Task: Task:attempt_local1671980215_0001_r_000000_0 is done. And is in the process of commiting
2021-06-14 10:10:33,719 INFO mapred.LocalJobRunner:
2021-06-14 10:10:33,719 INFO mapred.Task: Task attempt_local1671980215_0001_r_000000_0 is allowed to commit now
2021-06-14 10:10:33,721 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1671980215_0001_r_000000_0' to file:/home/hdoop/hadoop-3.2.1/grep-temp-1466437422
2021-06-14 10:10:33,721 INFO mapred.LocalJobRunner: reduce > reduce
2021-06-14 10:10:33,721 INFO mapred.Task: Task 'attempt_local1671980215_0001_r_000000_0' done.


Activities
Terminal
Mon 08:31
drive.google.com/file/d/1KAojPbNWOnaPkbFLwwk9XNEq4QCnRrxf/view
hdoop@dev-HP-ProBook-450-G7: ~
File Edit View Search Terminal Help
hdoop@dev-HP-ProBook-450-G7:~$ which javac
/usr/bin/javac
hdoop@dev-HP-ProBook-450-G7:~$ which java
/usr/bin/java
hdoop@dev-HP-ProBook-450-G7:~$ ls -ltr /usr/bin/java
lrwxrwxrwx 1 root root 22 Jun 14 07:49 /usr/bin/java -> /etc/alternatives/java
hdoop@dev-HP-ProBook-450-G7:~$ ls -ltr /etc/alternatives/java
lrwxrwxrwx 1 root root 46 Jun 14 07:49 /etc/alternatives/java -> /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java
hdoop@dev-HP-ProBook-450-G7:~$ ls -ltr /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java
-rwxr-xr-x 1 root root 6280 Apr 21 07:46 /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java
hdoop@dev-HP-ProBook-450-G7:~$


Activities
Terminal
Mon 09:11
avg-qycx-xcj (2021-05-24
drive.google.com/file/d/1KAojPbNWOnaPkbFLwwk9XNEq4QCnRrxf/view
avg-qycx-xcj (2021-05-24 at 22:10 GMT-7)
dev@dev-HP-ProBook-450-G7: ~/hadoop-3.1.4
File Edit View Search Terminal Help
dev@dev-HP-ProBook-450-G7:~$ echo $JAVA_HOME
/usr/lib/jvm/java-8-openjdk-amd64
dev@dev-HP-ProBook-450-G7:~$ ls
Desktop Documents Downloads examples.desktop hadoop-3.1.4 htmlcodes Hubstaff input Music Pictures Public snap Templates Videos
dev@dev-HP-ProBook-450-G7:~$ cd h
hadoop-3.1.4/ htmlcodes/
dev@dev-HP-ProBook-450-G7:~$ cd hadoop-3.1.4/
dev@dev-HP-ProBook-450-G7:~/hadoop-3.1.4$ ls
bin etc include lib libexec LICENSE.txt NOTICE.txt README.txt sbin share
dev@dev-HP-ProBook-450-G7:~/hadoop-3.1.4$ mkdir inputs
dev@dev-HP-ProBook-450-G7:~/hadoop-3.1.4$ ls
bin etc include inputs lib libexec LICENSE.txt NOTICE.txt README.txt sbin share
dev@dev-HP-ProBook-450-G7:~/hadoop-3.1.4$


Activities
Firefox Web Browser
Mon 08:04
Cloud-Computing-Projec
Google
Cloud Computing BS CS I
How to Install Hadoop on
Apache Download Mirror
https://phoenixnap.com/kb/install-hadoop-ubuntu
Bare Metal Cloud now available at special prices!
SEE DISCOUNTS
The Hadoop binary files are now located within the hadoop-3.2.1 directory.
Single Node Hadoop Deployment (Pseudo-Distributed Mode)
Hadoop excels when deployed in a fully distributed mode on a large cluster of networked servers. However, if you are new to Hadoop and want to explore basic commands or test applications, you can configure Hadoop on a single node.
This setup, also called pseudo-distributed mode, allows each Hadoop daemon to run as a single Java process. A Hadoop environment is configured by editing a set of configuration files:
bashrc
hadoop-env.sh
core-site.xml
hdfs-site.xml
mapred-site-xml
yarn-site.xml
Configure Hadoop Environment Variables (bashrc)
Edit the .bashrc shell configuration file using a text editor of your choice (we will be using nano):
sudo nano .bashrc
Define the Hadoop environment variables by adding the following content to the end of the file:
#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.2.1
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME


HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=2239
Total time spent by all reduces in occupied slots (ms)=2504
Total time spent by all map tasks (ms)=2239
Total time spent by all reduce tasks (ms)=2504
Total vcore-milliseconds taken by all map tasks=2239
Total vcore-milliseconds taken by all reduce tasks=2504
Total megabyte-milliseconds taken by all map tasks=2292736
Total megabyte-milliseconds taken by all reduce tasks=2564096
Map-Reduce Framework
Map input records=3
Map output records=8
Map output bytes=83
Map output materialized bytes=80
Input split bytes=110
Combine input records=8
Combine output records=6
Reduce input groups=6
Reduce shuffle bytes=80
Reduce input records=6
Reduce output records=6
Spilled Records=12
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=154
CPU time spent (ms)=1450
Physical memory (bytes) snapshot=524750848
Virtual memory (bytes) snapshot=5219487744
Total committed heap usage (bytes)=456130560
Peak Map Physical memory (bytes)=292409344
Peak Map Virtual memory (bytes)=2606850048
Peak Reduce Physical memory (bytes)=232341504
Peak Reduce Virtual memory (bytes)=2612637696
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=51