



# **CSE488: Big Data Analytics [SPRING 2023]**

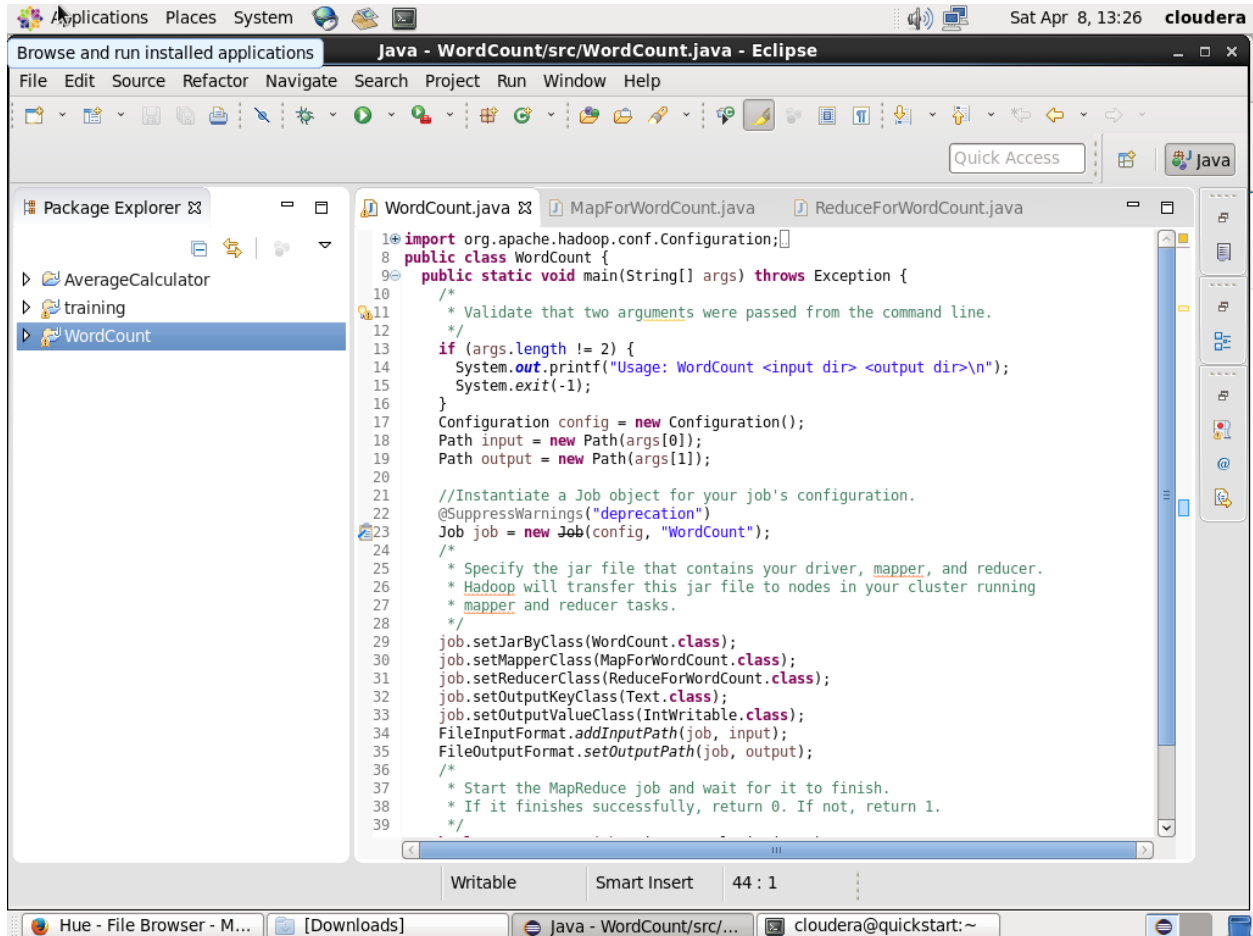
## **Assignment 6**

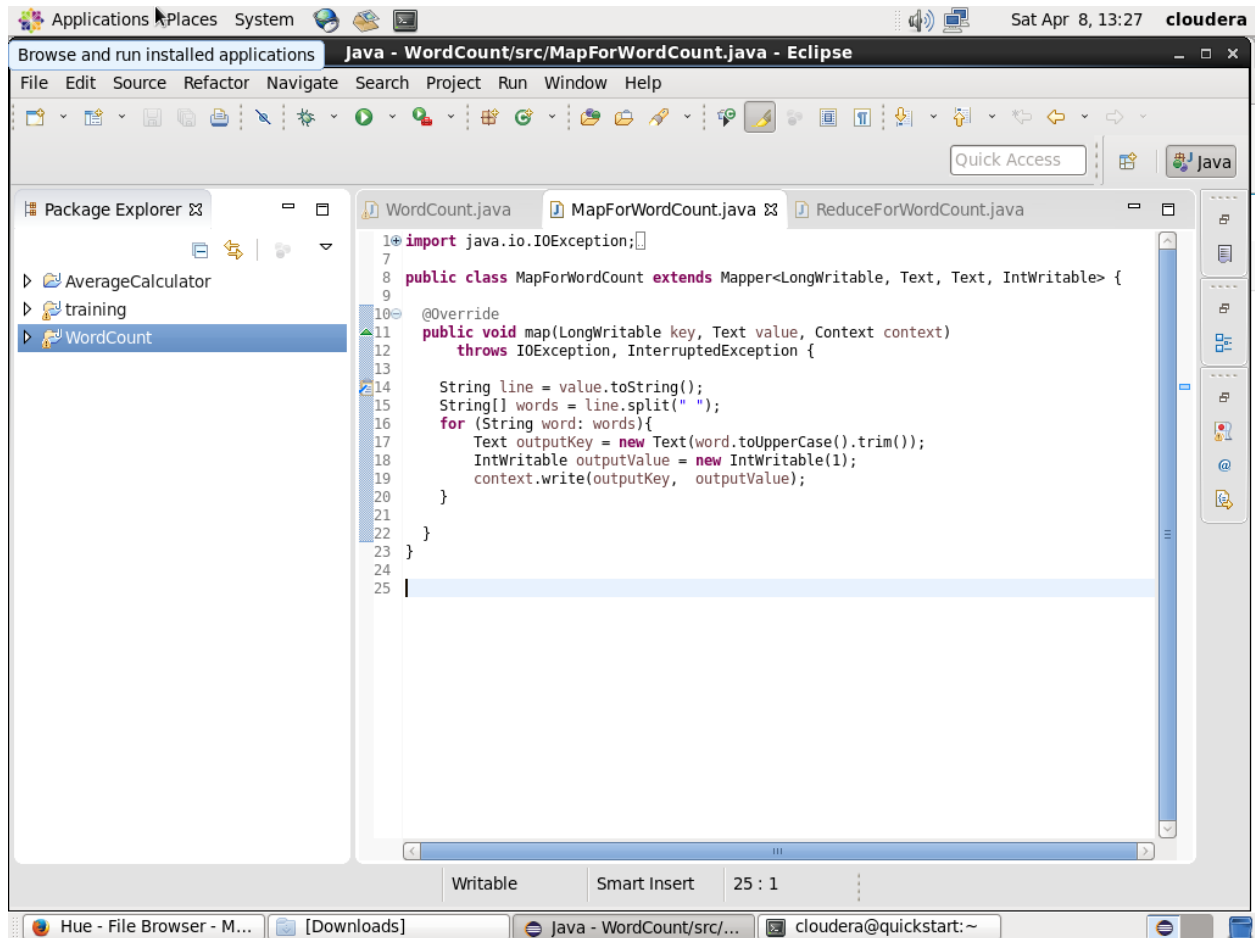
**Submitted by:**

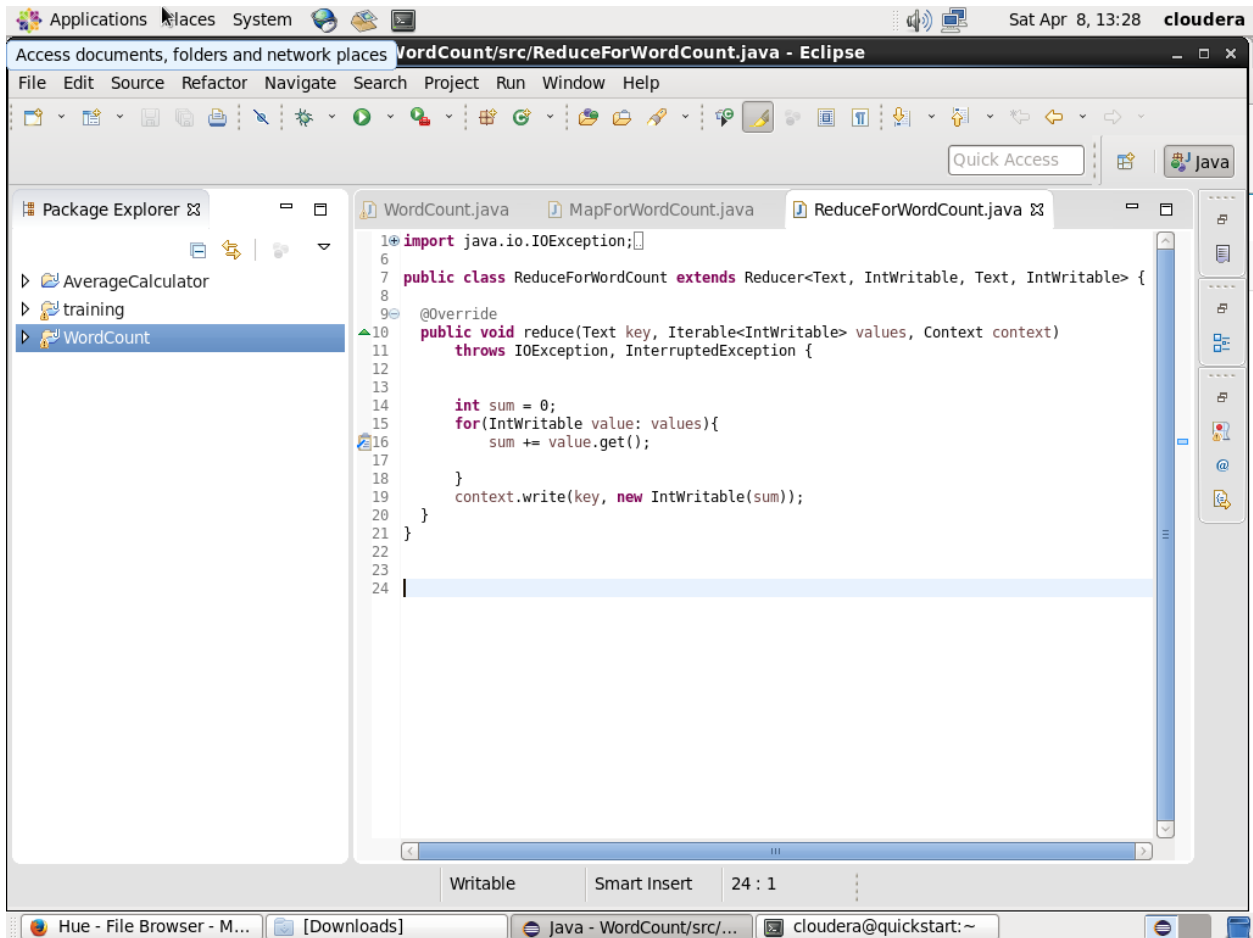
Student ID: 2020-1-60-215

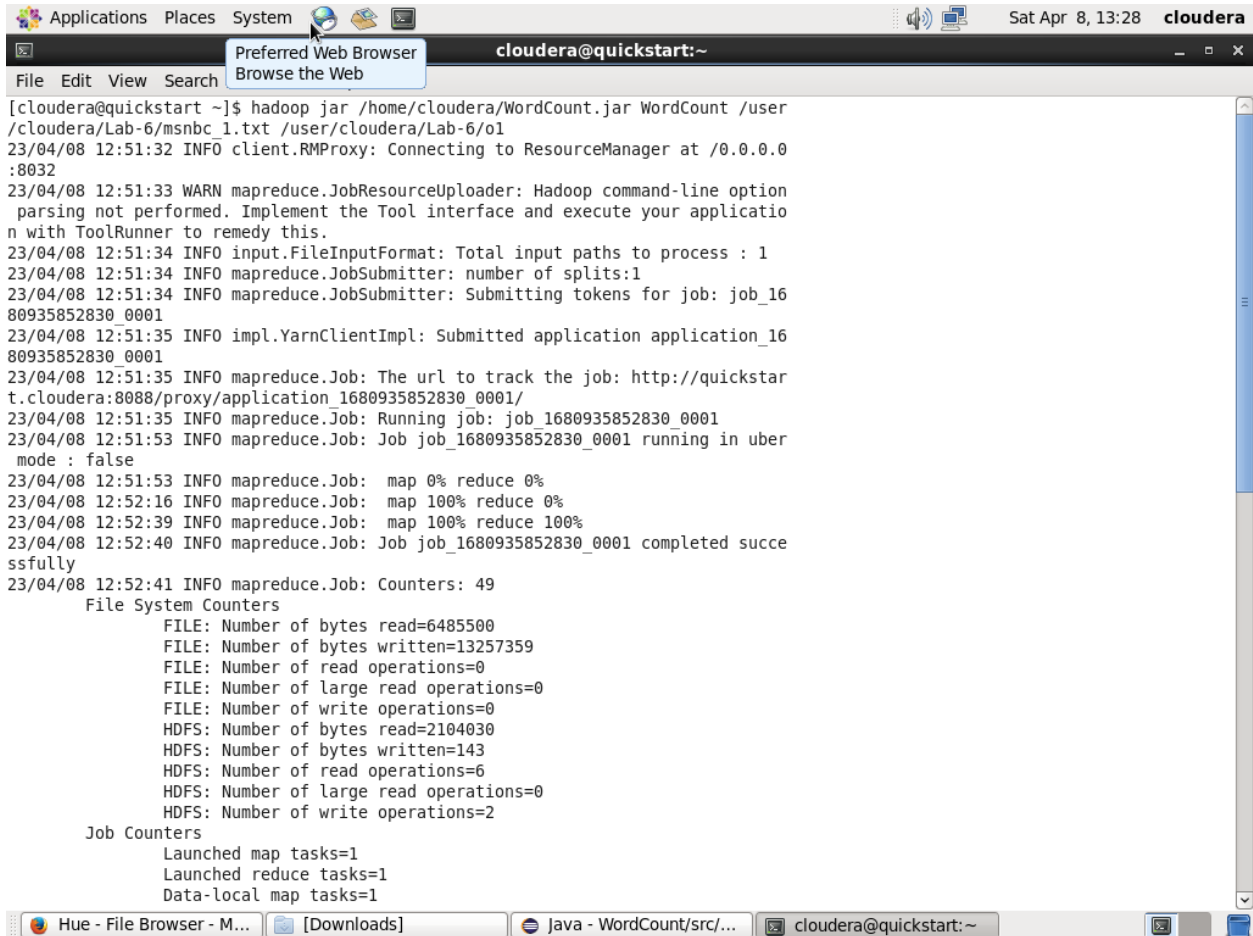
Student Name: Md. Abdul Ahad Rifat

## Problem 1









```
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar WordCount /user/cloudera/Lab-6/msnbc_1.txt /user/cloudera/Lab-6/o1
23/04/08 12:51:32 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/04/08 12:51:33 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
23/04/08 12:51:34 INFO input.FileInputFormat: Total input paths to process : 1
23/04/08 12:51:34 INFO mapreduce.JobSubmitter: number of splits:1
23/04/08 12:51:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1680935852830_0001
23/04/08 12:51:35 INFO impl.YarnClientImpl: Submitted application application_1680935852830_0001
23/04/08 12:51:35 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1680935852830_0001/
23/04/08 12:51:35 INFO mapreduce.Job: Running job: job_1680935852830_0001
23/04/08 12:51:53 INFO mapreduce.Job: Job job_1680935852830_0001 running in uber mode : false
23/04/08 12:51:53 INFO mapreduce.Job: map 0% reduce 0%
23/04/08 12:52:16 INFO mapreduce.Job: map 100% reduce 0%
23/04/08 12:52:39 INFO mapreduce.Job: map 100% reduce 100%
23/04/08 12:52:40 INFO mapreduce.Job: Job job_1680935852830_0001 completed successfully
23/04/08 12:52:41 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=6485500
    FILE: Number of bytes written=13257359
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=2104030
    HDFS: Number of bytes written=143
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
```

Applications Places System Sat Apr 8, 13:29 cloudera

Access documents, folders and network places Hue - File Browser - Mozilla Firefox

Cloudera Live : Welcom... x Hue - File Browser x Online Assignment L... x msnbc\_1.txt - Google... x

quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/Lab-6 Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera

HUE Query Search data and saved documents...

File Browser

default (0) Tables The database has no tables

View as binary Home Page 1 to

Edit file / user / cloudera / Lab-6

Download	1	157802
	10	22559
	11	16380
View file location	12	45173
	13	35768
	14	66621
	15	10999
Refresh	16	3122
	17	2832
Last modified	2	75956
04/08/2023 6:52 AM	3	33785
	4	63607
	5	24607
	6	69522
User	7	50531
cloudera	8	73214
Group	9	32777
cloudera		
Size		

New Window New Private Window Save Page

Print History Full Screen

Find Preferences Add-ons

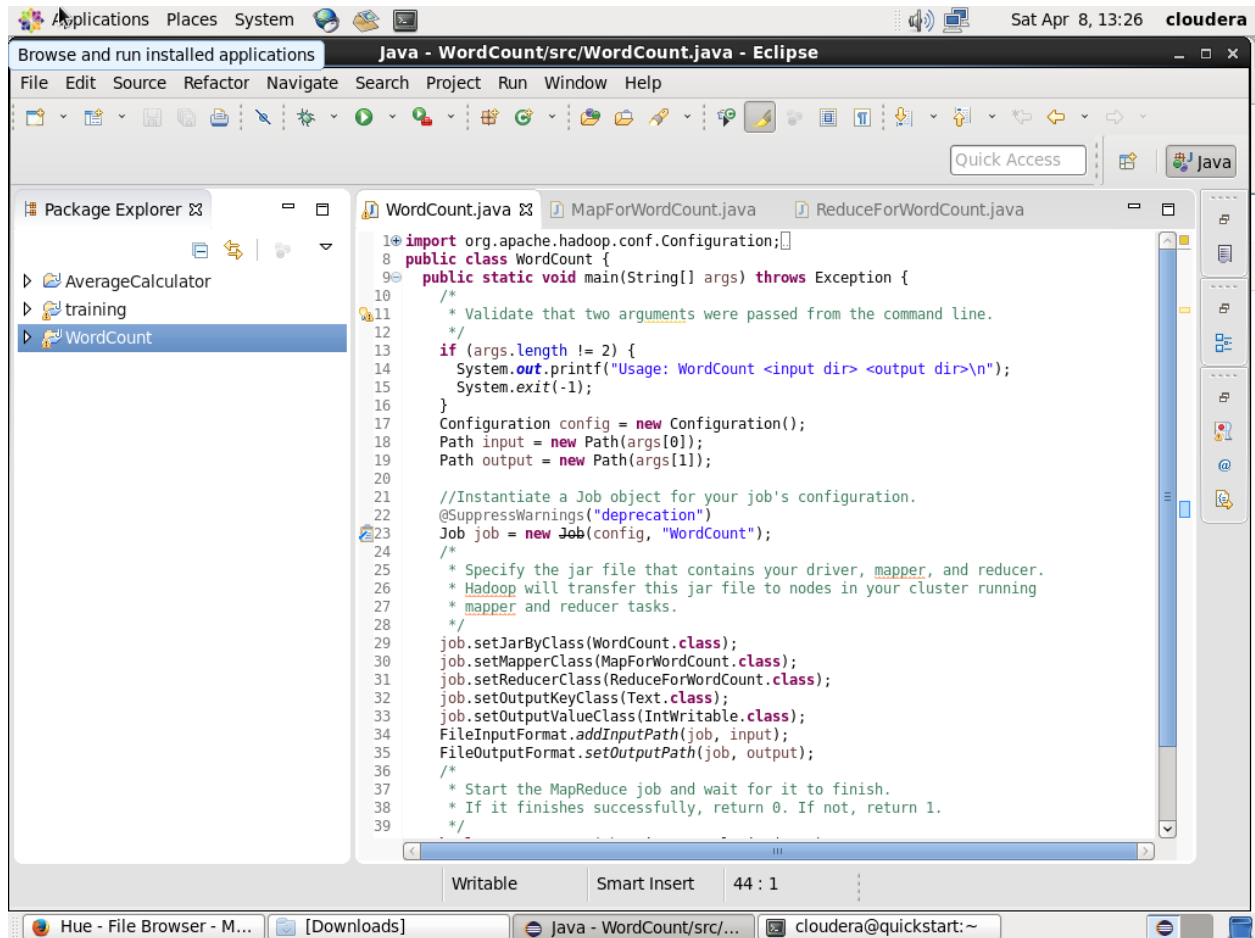
Developer

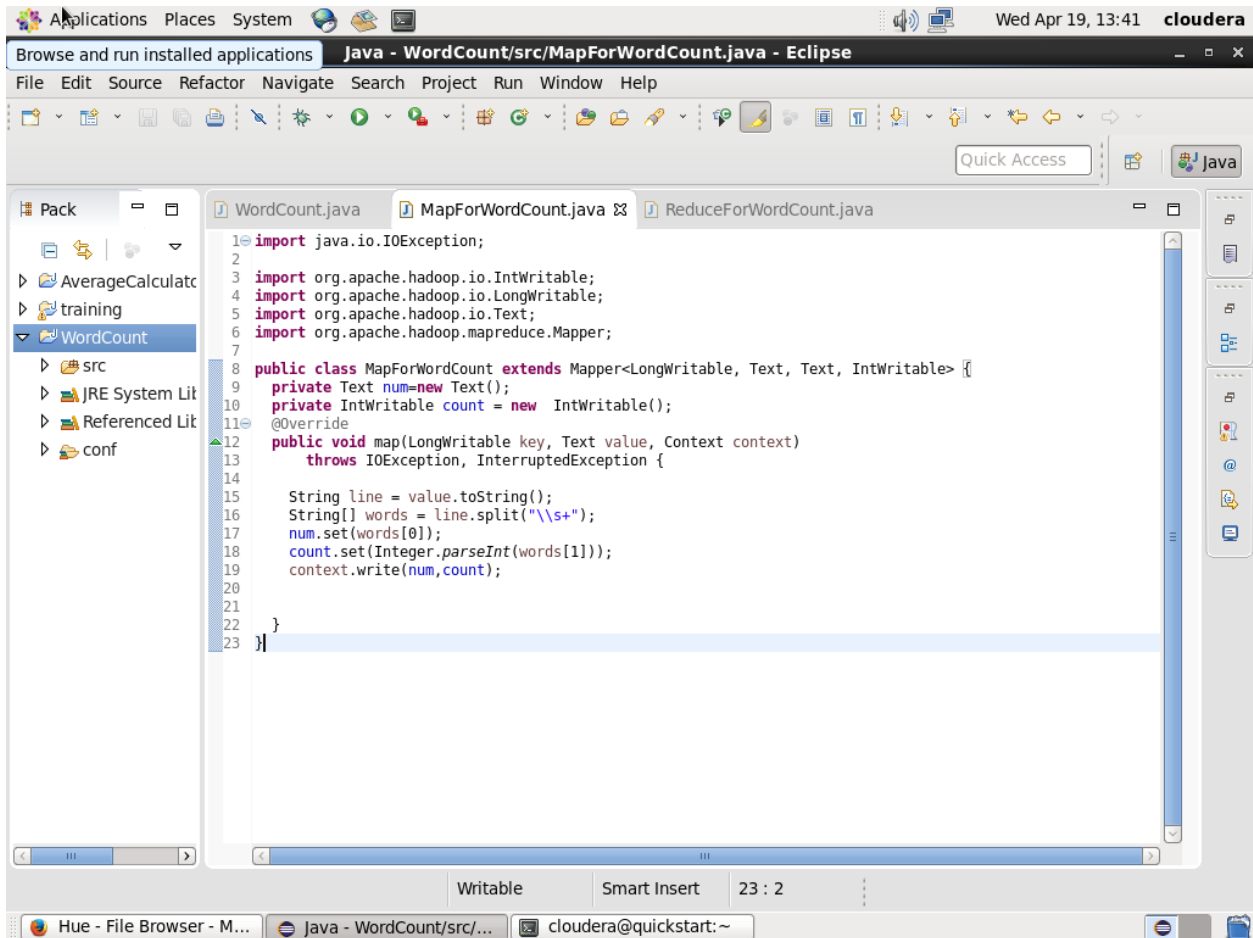
Sign in to Sync

Customize

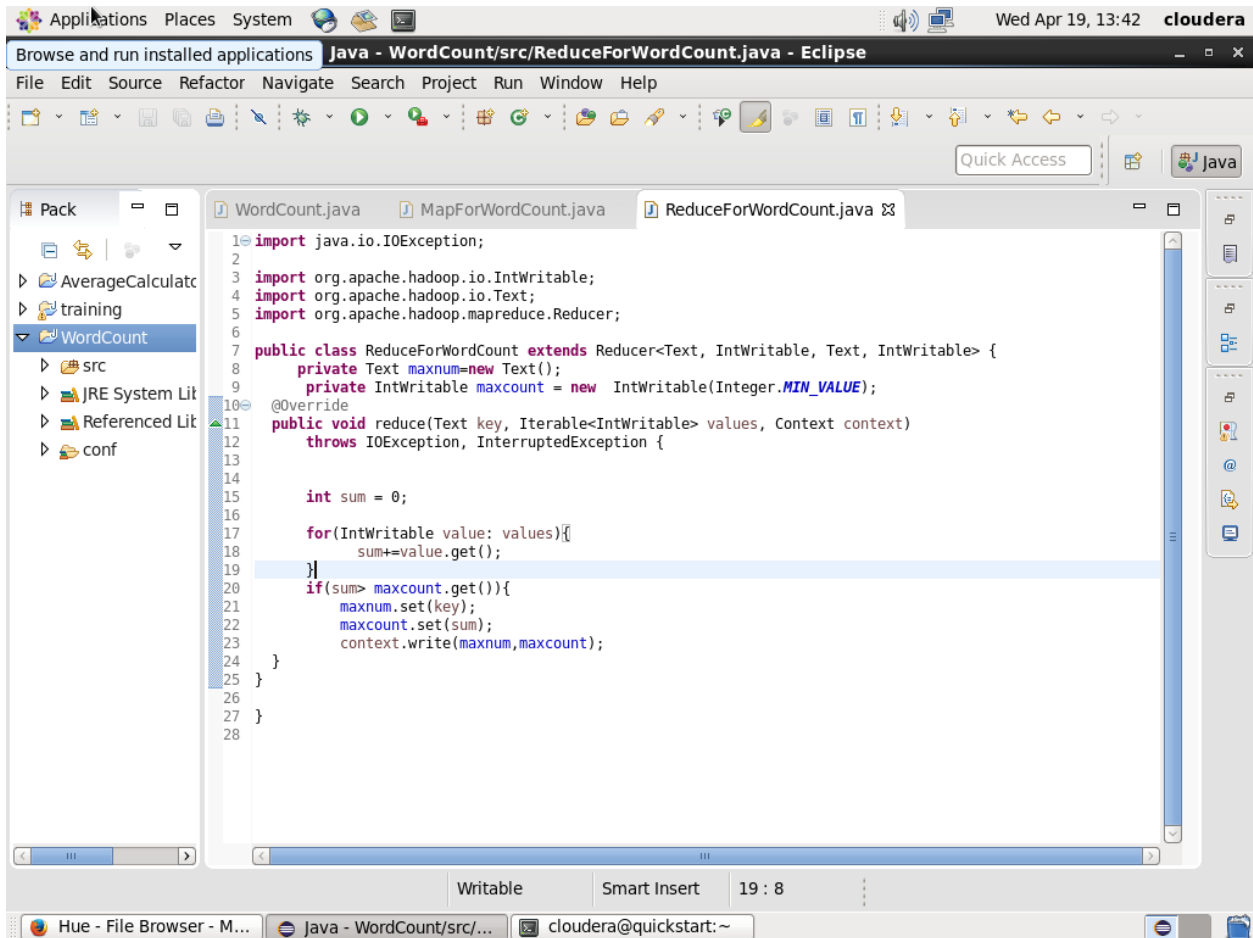
Hue - File Browser - M... [Downloads] Java - WordCount/src/... cloudera@quickstart:~

## Problem 2









```
Applications Places System cloudera Wed Apr 19, 13:42 cloudera
Browse and run installed applications cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/max.jar WordCount /user/cloudera/Lab-6/ol/part-r-00000 /user/cloudera/Lab-6/02
23/04/19 13:38:49 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/04/19 13:38:50 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
23/04/19 13:38:50 INFO input.FileInputFormat: Total input paths to process : 1
23/04/19 13:38:50 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedExcep
tion
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
23/04/19 13:38:50 INFO mapreduce.JobSubmitter: number of splits:1
23/04/19 13:38:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1680954430650_0011
23/04/19 13:38:51 INFO impl.YarnClientImpl: Submitted application application_1680954430650_0011
23/04/19 13:38:51 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_16809544306
50_0011/
23/04/19 13:38:51 INFO mapreduce.Job: Running job: job_1680954430650_0011
23/04/19 13:39:02 INFO mapreduce.Job: Job job_1680954430650_0011 running in uber mode : false
23/04/19 13:39:02 INFO mapreduce.Job: map 0% reduce 0%
23/04/19 13:39:09 INFO mapreduce.Job: map 100% reduce 0%
23/04/19 13:39:17 INFO mapreduce.Job: map 100% reduce 100%
23/04/19 13:39:18 INFO mapreduce.Job: Job job_1680954430650_0011 completed successfully
23/04/19 13:39:18 INFO mapreduce.Job: Counters: 49
File System Counters
    FILE: Number of bytes read=150
    FILE: Number of bytes written=286667
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=275
    HDFS: Number of bytes written=9
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
Job Counters
    Launched map tasks=1
```


Applications Places System Wed Apr 19, 13:42 cloudera






Change desktop appearance and behavior, get help, or log out e Browser - Mozilla Firefox

Cloudera Live : Welco... x Hue - File Browser x OFFLINE ASSIGNM... x msnbc\_1.txt - Goo... x Cloudera Live : Welco... x




quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/Lab-6 Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

 Query Search data and sav Jobs cloudera

default

Tables (0)   

The database has no tables





View as binary

Edit file

Download

View file location

Refresh

Page 1 to 1 of 1    

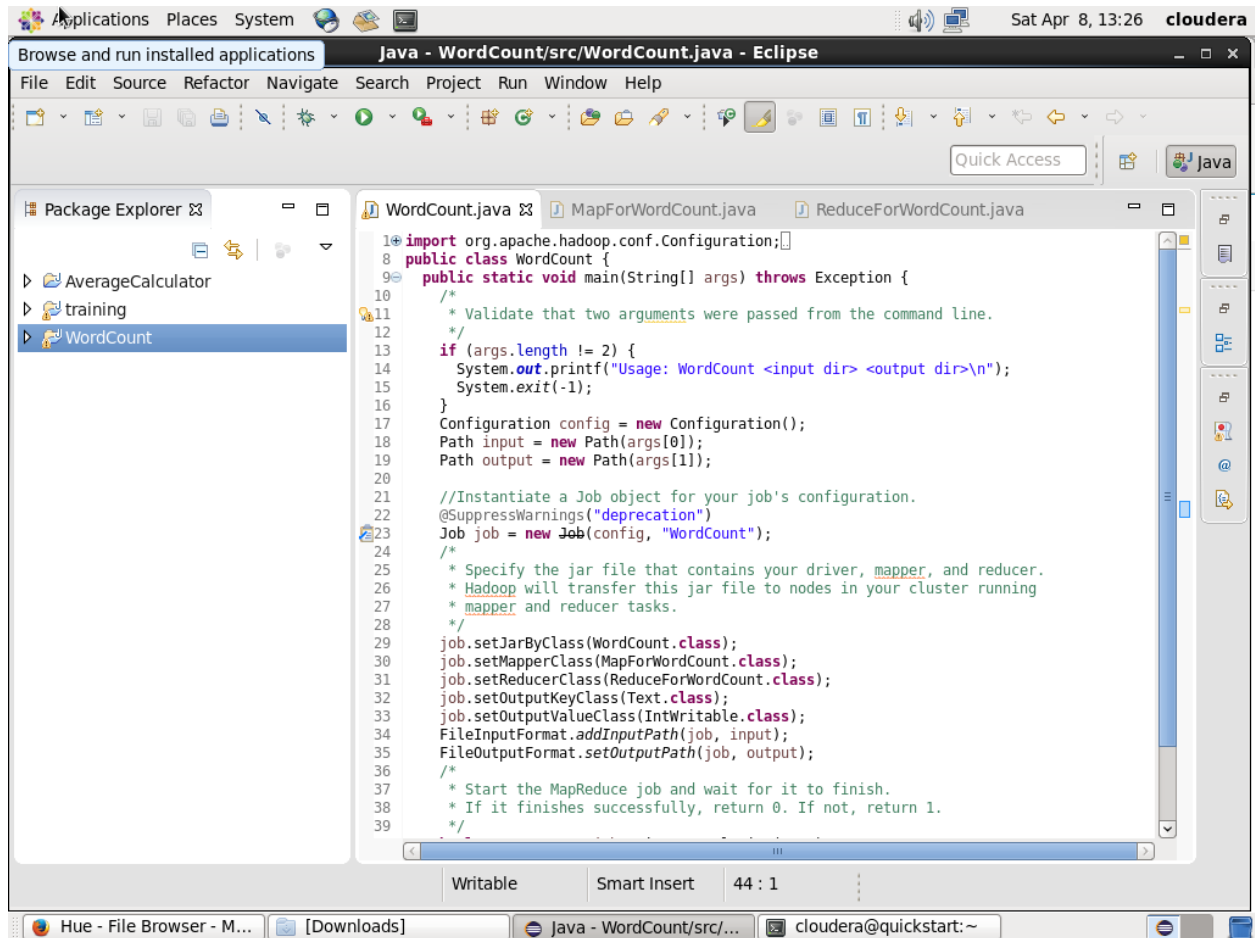
Home

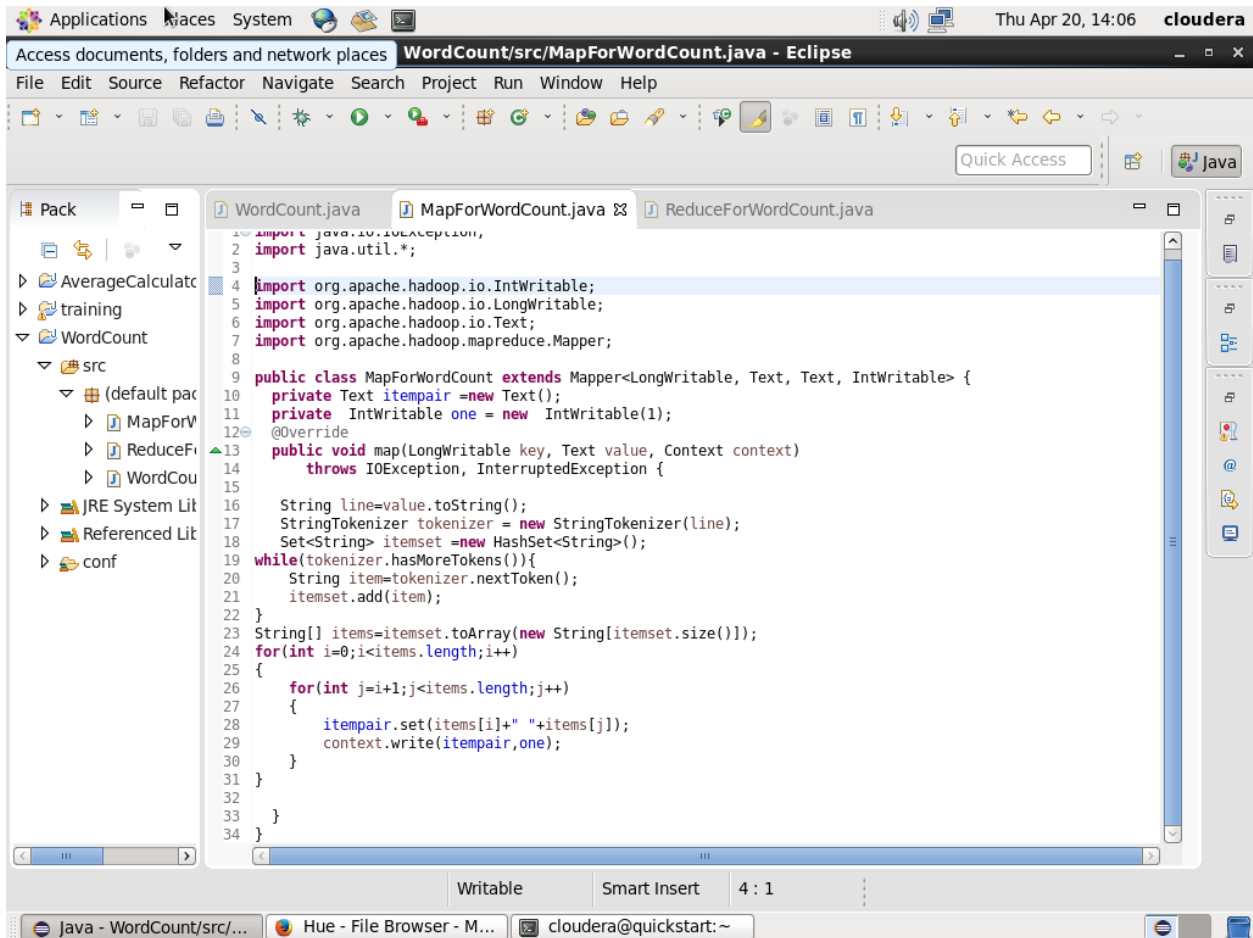
/ user / cloudera / Lab-6 / o2 /

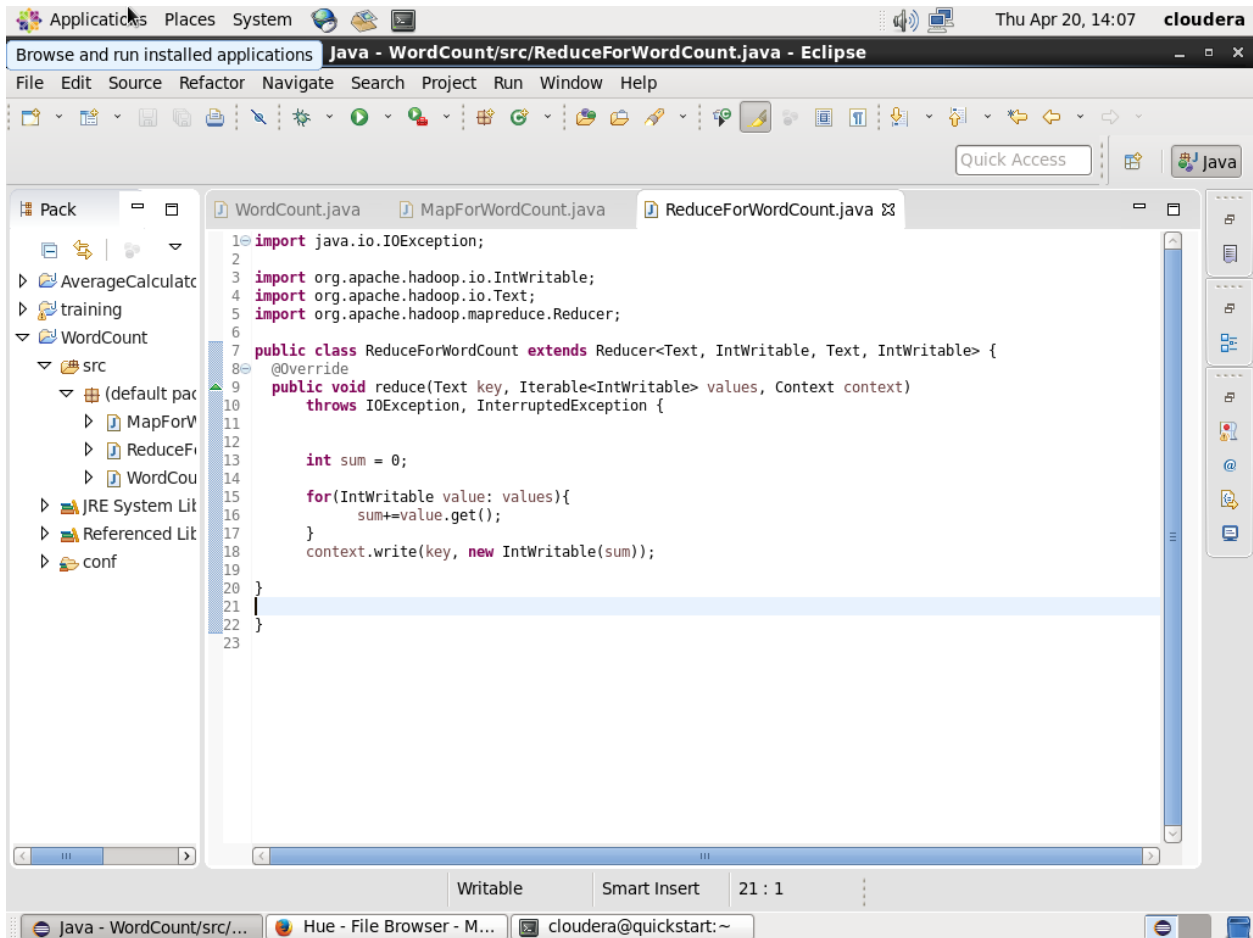
part-r-00000

1	157802
---	--------

## Problem 4







Applications Places System Thu Apr 20, 14:07 cloudera

Change desktop appearance and behavior, get help, or log out cloudera@quickstart:~

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/itmpair.jar WordCount /user/cloudera/Lab-6/msnbc_1.txt /user/cloudera/Lab-6/o3
23/04/20 14:02:32 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/04/20 14:02:33 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
23/04/20 14:02:33 INFO input.FileInputFormat: Total input paths to process : 1
23/04/20 14:02:33 INFO mapreduce.JobSubmitter: number of splits:1
23/04/20 14:02:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1680954430650_0025
23/04/20 14:02:34 INFO impl.YarnClientImpl: Submitted application application_1680954430650_0025
23/04/20 14:02:34 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_16809544306
50_0025/
23/04/20 14:02:34 INFO mapreduce.Job: Running job: job_1680954430650_0025
23/04/20 14:02:44 INFO mapreduce.Job: Job job_1680954430650_0025 running in uber mode : false
23/04/20 14:02:44 INFO mapreduce.Job: map 0% reduce 0%
23/04/20 14:02:54 INFO mapreduce.Job: map 100% reduce 0%
23/04/20 14:03:04 INFO mapreduce.Job: map 100% reduce 100%
23/04/20 14:03:04 INFO mapreduce.Job: Job job_1680954430650_0025 completed successfully
23/04/20 14:03:04 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=2296065
    FILE: Number of bytes written=4878489
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=2104030
    HDFS: Number of bytes written=1814
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  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)=7976
    Total time spent by all reduces in occupied slots (ms)=7500
    Total time spent by all map tasks (ms)=7976
    Total time spent by all reduce tasks (ms)=7500
    Total vcore-milliseconds taken by all map tasks=7976
    Total vcore-milliseconds taken by all reduce tasks=7500
```

[Java - WordCount/src/...] [Hue - File Browser - ...] cloudera@quickstart:~

Application Places System

Thu Apr 20, 14:07 cloudera

Browse and run installed applications

Hue - File Browser - Mozilla Firefox

Hue - File Browser

quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/Lab-6

Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE

Query

Search data and sav

Jobs

cloudera

default

Tables

The database has no tables

(0)

as binary

Home

/ user / cloudera / Lab-6 / o3 /

part-r-00000

1 10 243

1 11 5472

1 12 7116

1 13 651

1 14 6504

1 15 1492

1 16 178

1 17 1252

1 4 6578

1 5 1985

Download

View file location

Refresh

Last modified

04/20/2023

8:03 AM

[Java - WordCount/src/...

Hue - File Browser - M...

cloudera@quickstart: ~