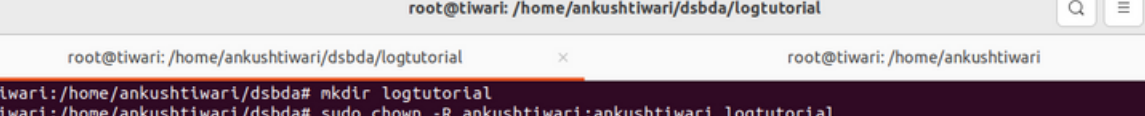


AIM: Design a distributed application using MapReduce which processes a log file of a system.



The terminal window shows the following commands and output:

```
root@tiwari: /home/ankushtiwari/dsbdalognutorial
root@tiwari:/home/ankushtiwari/dsbdalognutorial# mkdir logtutorial
root@tiwari:/home/ankushtiwari/dsbdalognutorial# sudo chown -R ankushtiwari:ankushtiwari logtutorial
root@tiwari:/home/ankushtiwari/dsbdalognutorial# start-all.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [tiwari]
Starting resourcemanager
Starting nodemanagers
root@tiwari:/home/ankushtiwari/dsbdalognutorial# cd logtutorial/
root@tiwari:/home/ankushtiwari/dsbdalognutorial/logtutorial# touch LogFileMapper.java
root@tiwari:/home/ankushtiwari/dsbdalognutorial/logtutorial# touch LogFileReducer.java
root@tiwari:/home/ankushtiwari/dsbdalognutorial/logtutorial# touch LogFileCountryDriver.java
root@tiwari:/home/ankushtiwari/dsbdalognutorial/logtutorial# touch log_file.txt
root@tiwari:/home/ankushtiwari/dsbdalognutorial/logtutorial# cd ..
root@tiwari:/home/ankushtiwari/dsbdalognutorial# sudo chown -R ankushtiwari:ankushtiwari logtutorial
root@tiwari:/home/ankushtiwari/dsbdalognutorial# cd logtutorial/
root@tiwari:/home/ankushtiwari/dsbdalognutorial/logtutorial#
```

Activities LibreOffice Calc Apr 16 00:45

Untitled 1 - LibreOffice Calc

File Edit View Insert Format

Liberation Sans

A1

You are running version 7.3 of LibreOffice

Text Import - [log_file.txt]

Import

Character set: Unicode (UTF-8)

Language: Default - English (India)

From row: 2

Separator Options

☐ Fixed width ☒ Separated by

☐ Tab ☐ Comma ☐ Semicolon ☐ Space ☒ Other {

☐ Merge delimiters ☐ Trim spaces String delimiter: "

Other Options

☐ Format quoted field as text ☐ Detect special numbers

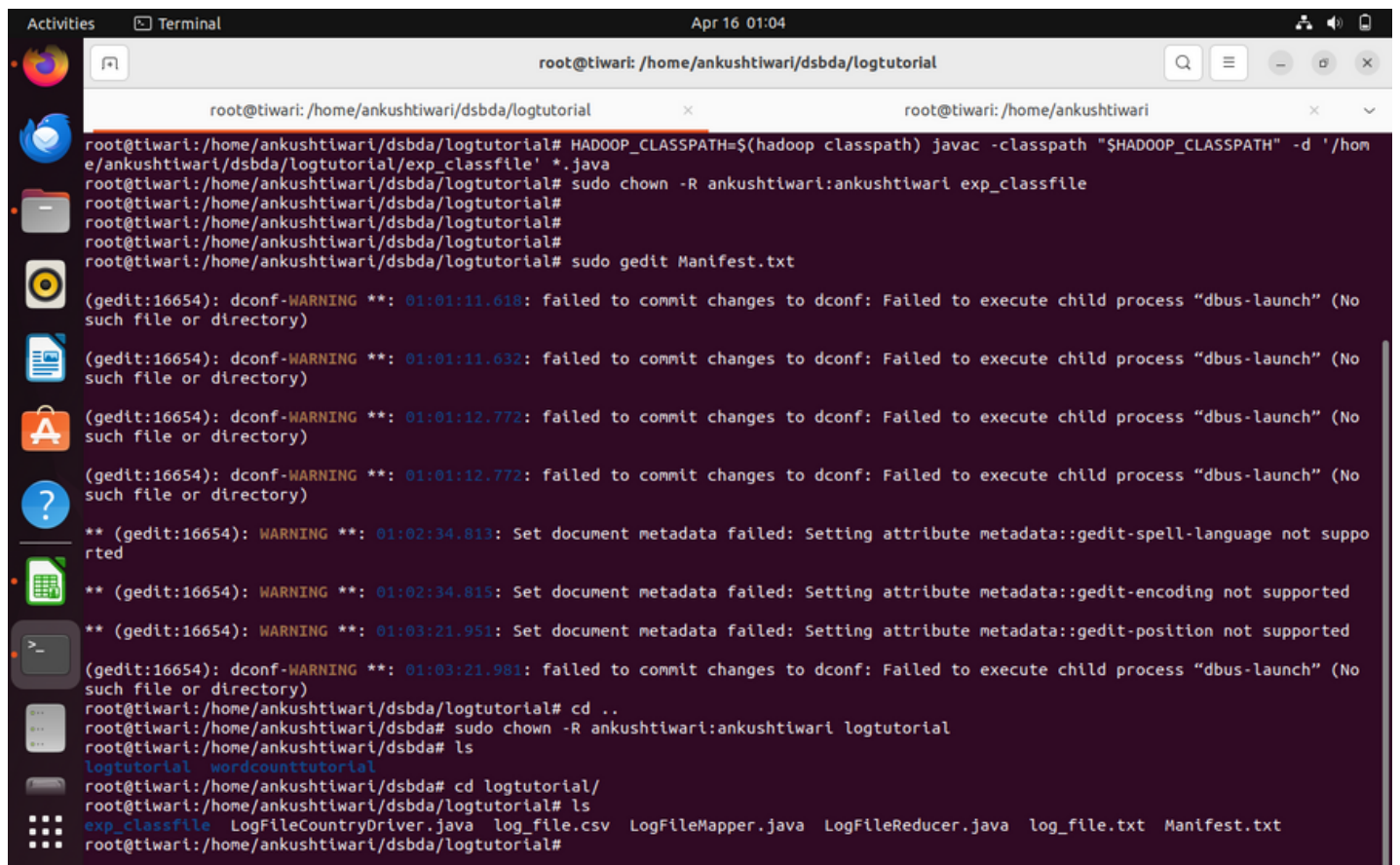
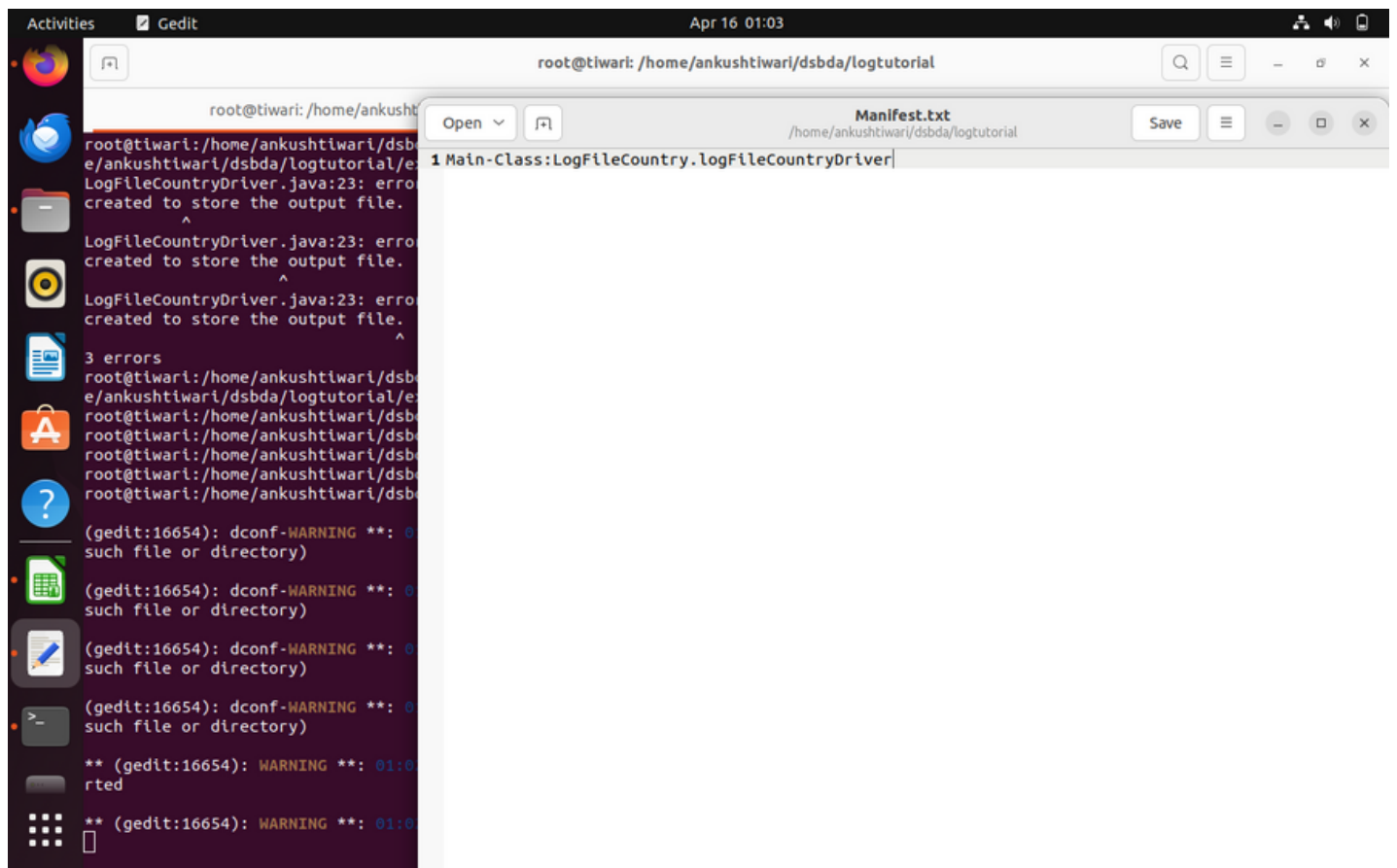
☐ Evaluate formulas

Fields

Column type:

	Standard	Standard	Standard	Standard
1	0.223.157.186		[15/Jul/2009:20:50:32 0700]	"GET /assets/js/the
2	10.223.157.186		[15/Jul/2009:20:50:33 0700]	"GET /assets/img/h
3	10.223.157.186		[15/Jul/2009:20:50:33 0700]	"GET /assets/img/du
4	10.223.157.186		[15/Jul/2009:20:50:33 0700]	"GET /assets/img/du
5	10.223.157.186		[15/Jul/2009:20:50:33 0700]	"GET /assets/img/h
6				
7	10.223.157.186		[15/Jul/2009:20:50:33 0700]	"GET /assets/img/du
8	10.223.157.186		[15/Jul/2009:20:50:33 0700]	"GET /assets/img/lo

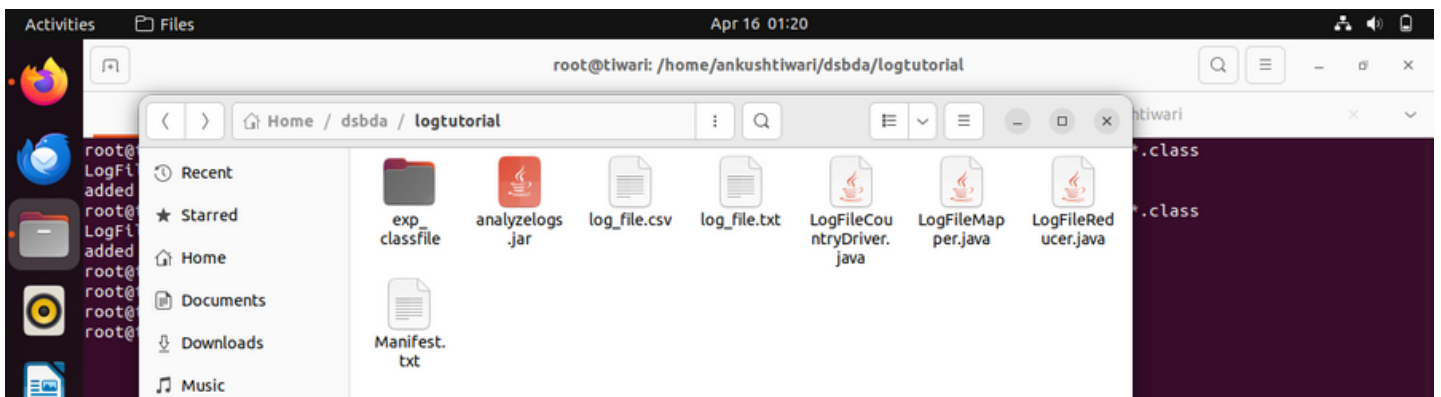
Help Cancel OK



Activities Terminal Apr 16 01:19

root@tiwari: /home/ankushtiwari/dsbda/logtutorial

```
root@tiwari: /home/ankushtiwari/dsbda/logtutorial# jar -cvfm analyzelogs.jar Manifest.txt LogFileCountry/*.class
LogFileCountry/*.class : no such file or directory
added manifest
root@tiwari: /home/ankushtiwari/dsbda/logtutorial# jar -cvfm analyzelogs.jar Manifest.txt LogFileCountry/*.class
LogFileCountry/*.class : no such file or directory
added manifest
root@tiwari: /home/ankushtiwari/dsbda/logtutorial# cd ..
root@tiwari: /home/ankushtiwari/dsbda# sudo chown -R ankushtiwari:ankushtiwari logtutorial
root@tiwari: /home/ankushtiwari/dsbda# cd logtutorial/
root@tiwari: /home/ankushtiwari/dsbda/logtutorial#
```



Activities Terminal Apr 16 01:23

root@tiwari: /home/ankushtiwari/dsbda/logtutorial

```
root@tiwari: /home/ankushtiwari/dsbda/logtutorial# hdfs dfs -mkdir /LogFileExp
root@tiwari: /home/ankushtiwari/dsbda/logtutorial# hdfs dfs -mkdir /LogFileExp/input
root@tiwari: /home/ankushtiwari/dsbda/logtutorial# hdfs dfs -mkdir /LogFileExp/output
root@tiwari: /home/ankushtiwari/dsbda/logtutorial#
```

Activities Firefox Web Browser Apr 16 01:23

(no subject) - creativeins APznzaYECuQf4sjHF3E91bT Ubuntu Command Summ Browsing HDFS

localhost:9870/explorer.html#/LogFileExp

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/LogFileExp Go!

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Apr 16 01:22	0	0 B	input	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Apr 16 01:22	0	0 B	output	

Showing 1 to 2 of 2 entries Previous 1 Next

Hadoop, 2023.

Activities Terminal Apr 16 01:44

root@tiwari: /home/ankushtiwari/dsdba/logtutorial

```
root@tiwari: /home/ankushtiwari/dsdba/logtutorial# jar -cvf analyzeLogs1.jar -C '/home/ankushtiwari/dsdba/logtutorial/exp_classfile' LogFileCountry/*.class
/home/ankushtiwari/dsdba/logtutorial/exp_classfile/LogFileCountry/*.class : no such file or directory
added manifest
root@tiwari: /home/ankushtiwari/dsdba/logtutorial# jar -cvf analyzeLogs1.jar -C '/home/ankushtiwari/dsdba/logtutorial/exp_classfile/L ogFileCountry' .
added manifest
adding: LogFileCountryDriver.class(in = 1677) (out= 826)(deflated 50%)
adding: LogFileReducer.class(in = 1580) (out= 636)(deflated 59%)
adding: LogFileMapper.class(in = 1713) (out= 646)(deflated 62%)
root@tiwari: /home/ankushtiwari/dsdba/logtutorial# jar -tf analyzeLogs.jar
java.io.FileNotFoundException: analyzeLogs.jar (No such file or directory)
    at java.util.zip.ZipFile.open(Native Method)
    at java.util.zip.ZipFile.<init>(ZipFile.java:231)
    at java.util.zip.ZipFile.<init>(ZipFile.java:157)
    at java.util.zip.ZipFile.<init>(ZipFile.java:128)
    at sun.tools.jar.Main.list(Main.java:1115)
    at sun.tools.jar.Main.run(Main.java:293)
    at sun.tools.jar.Main.main(Main.java:1288)
root@tiwari: /home/ankushtiwari/dsdba/logtutorial# jar -tf analyzeLogs1.jar
META-INF/
META-INF/MANIFEST.MF
LogFileCountryDriver.class
LogFileReducer.class
LogFileMapper.class
root@tiwari: /home/ankushtiwari/dsdba/logtutorial#
```



```

Map-Reduce Framework
  Map input records=8
  Map output records=139
  Map output bytes=1355
  Map output materialized bytes=1265
  Input split bytes=120
  Combine input records=139
  Combine output records=102
  Reduce input groups=102
  Reduce shuffle bytes=1265
  Reduce input records=102
  Reduce output records=102
  Spilled Records=204
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=244
  CPU time spent (ms)=1950
  Physical memory (bytes) snapshot=536207360
  Virtual memory (bytes) snapshot=5133156352
  Total committed heap usage (bytes)=401080320
  Peak Map Physical memory (bytes)=320638976
  Peak Map Virtual memory (bytes)=2560835584
  Peak Reduce Physical memory (bytes)=215568384
  Peak Reduce Virtual memory (bytes)=2572320768
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=799
File Output Format Counters
  Bytes Written=851

```

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/LogFileExp/Output

Show 25 entries

Permission	Owner
-rw-r--r--	huser
-rw-r--r--	huser

Showing 1 to 2 of 2 entries

Hadoop, 2021.

File information - part-00000

Download Head the file (first 32K) Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741897
 Block Pool ID: BP-1388353168-127.0.1.1-1647528100285
 Generation Stamp: 1073
 Size: 3838
 Availability:
 • ubuntu-college

File contents

10.240.170.50	1
10.241.107.75	1
10.241.9.187	1
10.243.51.109	5
10.244.166.195	5
10.245.208.15	20
10.246.151.162	3
10.247.111.104	9

Close