



Compte Rendu

Spécialité : Réseau informatique et télécommunications

TP1 - Le traitement Batch avec Hadoop HDFS et Map Reduce

Présenté par

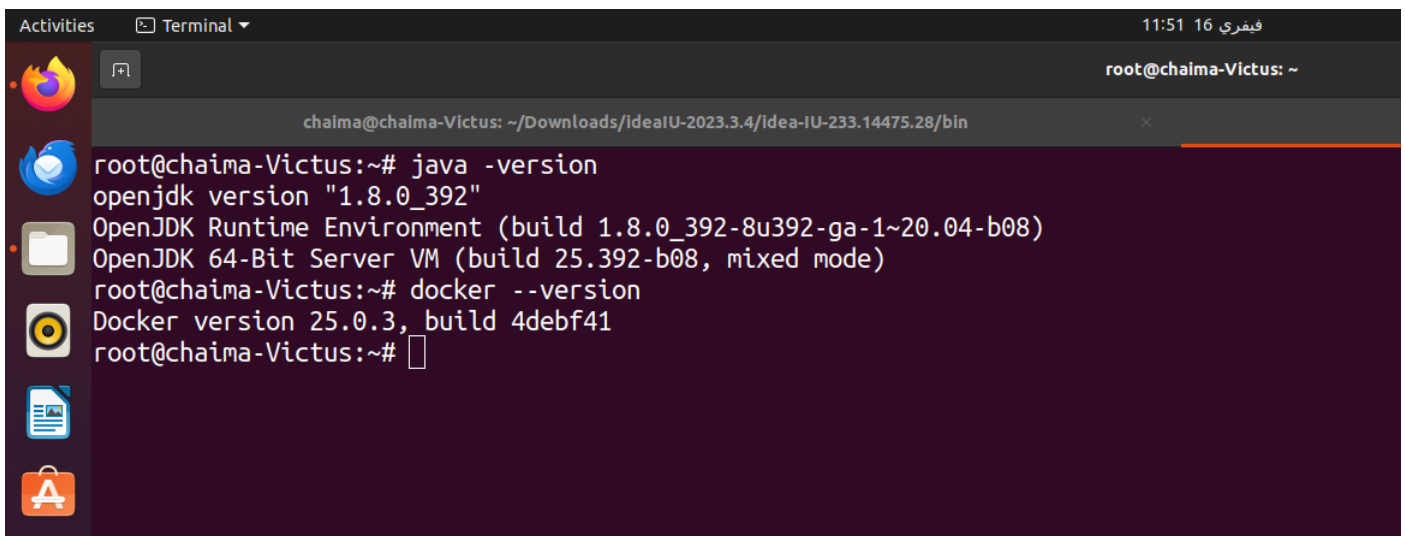
Ghaouari Koussay
Raach Chaima

Année Universitaire : 2023 / 2024

1

Setup et premiers pas

1 Installation des des outils



A terminal window titled "Terminal" with a dark background. The window shows the following commands and output:

```
root@chaima-Victus:~# java -version
openjdk version "1.8.0_392"
OpenJDK Runtime Environment (build 1.8.0_392-8u392-ga-1~20.04-b08)
OpenJDK 64-Bit Server VM (build 25.392-b08, mixed mode)
root@chaima-Victus:~# docker --version
Docker version 25.0.3, build 4debf41
root@chaima-Victus:~#
```

The terminal window has a sidebar on the left with icons for Activities, Terminal, and various applications. The top bar shows the time as 11:51 and the date as 16 فيفري. The window title is "root@chaima-Victus: ~".

2 Création du réseau et des conteneurs docker

```
root@chaima-Victus:~# docker network ls
NETWORK ID      NAME      DRIVER      SCOPE
c5ca33932572    bridge    bridge       local
b86ddf1b3e3b    hadoop    bridge       local
4c29a30589b8    host      host         local
33297303ba71    none      null         local
root@chaima-Victus:~# docker ps -a
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS
6af37581ff4c   liliasfazi/spark-hadoop:hv-2.7.2   "sh -c 'service ssh ..." 15 hours ago   Up 2 hours    0.0.0.0:8041->8042/tcp, :::8041->8042/tcp
415ea3e294fb   liliasfazi/spark-hadoop:hv-2.7.2   "sh -c 'service ssh ..." 15 hours ago   Up 2 hours    0.0.0.0:8040->8042/tcp, :::8040->8042/tcp
a4499660a513   liliasfazi/spark-hadoop:hv-2.7.2   "sh -c 'service ssh ..." 15 hours ago   Up 2 hours    0.0.0.0:8088->8088/tcp, :::8088->8088/tcp, 0.0.0.0:50070->50070/tcp, :::50070->50070/tcp, 0.0.0.0:32768->707/tcp, :::32768->707/tcp
root@chaima-Victus:~#
```

3 Entrer dans le contenaire master pour commencer à l'utiliser et Lancer hadoop et yarn

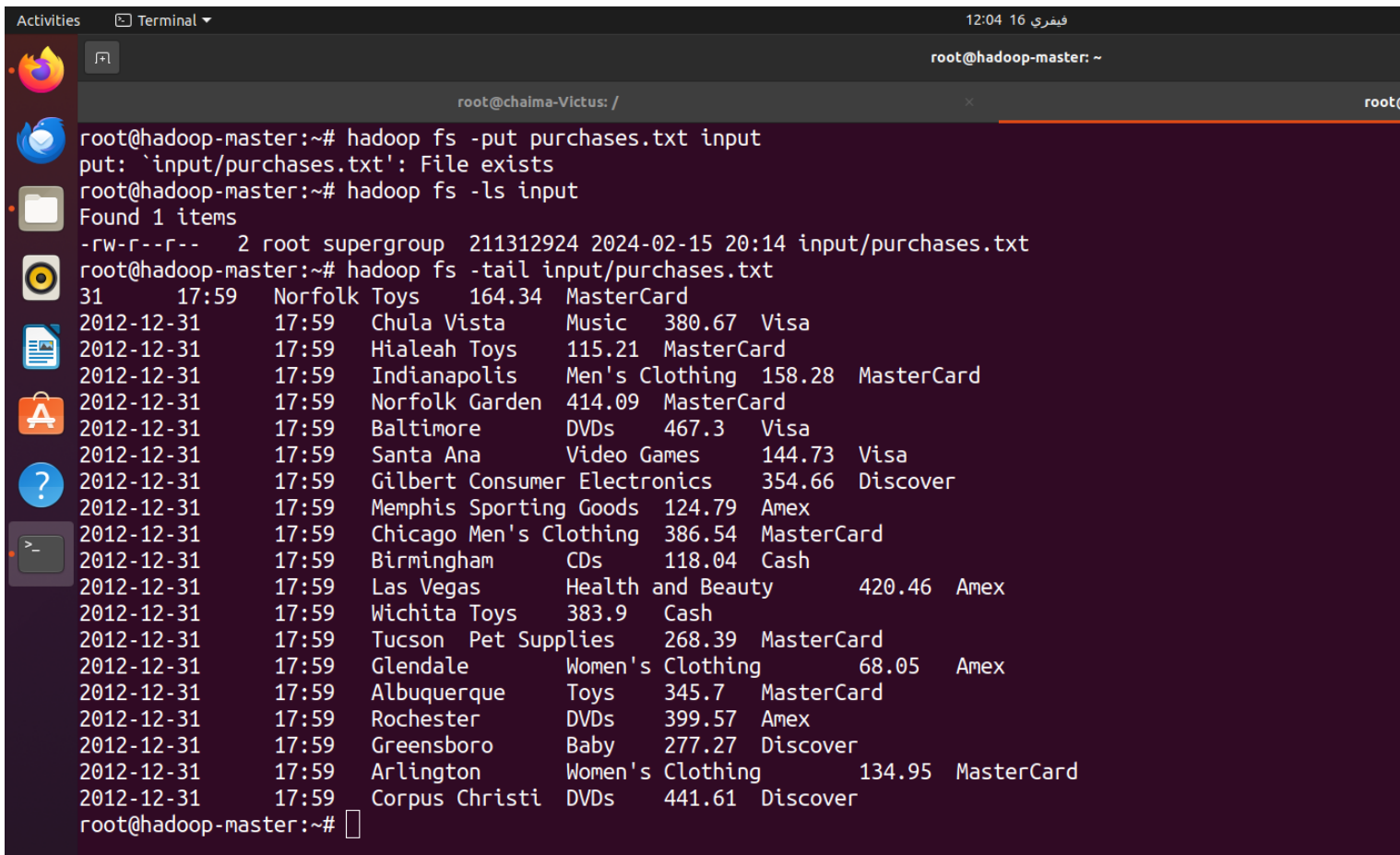
```
root@chaima-Victus:~# docker exec -it hadoop-master bash
root@hadoop-master:~# /start-hadoop.sh
bash: /start-hadoop.sh: No such file or directory
root@hadoop-master:~# ./start-hadoop.sh

Starting namenodes on [hadoop-master]
hadoop-master: Warning: Permanently added 'hadoop-master,172.20.0.4' (ECDSA) to the list of known hosts.
hadoop-master: starting namenode, logging to /usr/local/hadoop/logs/hadoop-root-namenode-hadoop-master.out
hadoop-slave2: Warning: Permanently added 'hadoop-slave2,172.20.0.2' (ECDSA) to the list of known hosts.
hadoop-slave1: Warning: Permanently added 'hadoop-slave1,172.20.0.3' (ECDSA) to the list of known hosts.
hadoop-slave2: starting datanode, logging to /usr/local/hadoop/logs/hadoop-root-datanode-hadoop-slave2.out
hadoop-slave1: starting datanode, logging to /usr/local/hadoop/logs/hadoop-root-datanode-hadoop-slave1.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-root-secondarynamenode-hadoop-master.out

starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn--resourcemanager-hadoop-master.out
hadoop-slave2: Warning: Permanently added 'hadoop-slave2,172.20.0.2' (ECDSA) to the list of known hosts.
hadoop-slave1: Warning: Permanently added 'hadoop-slave1,172.20.0.3' (ECDSA) to the list of known hosts.
hadoop-slave1: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-root-nodemanager-hadoop-slave1.out
hadoop-slave2: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-root-nodemanager-hadoop-slave2.out

root@hadoop-master:~#
```

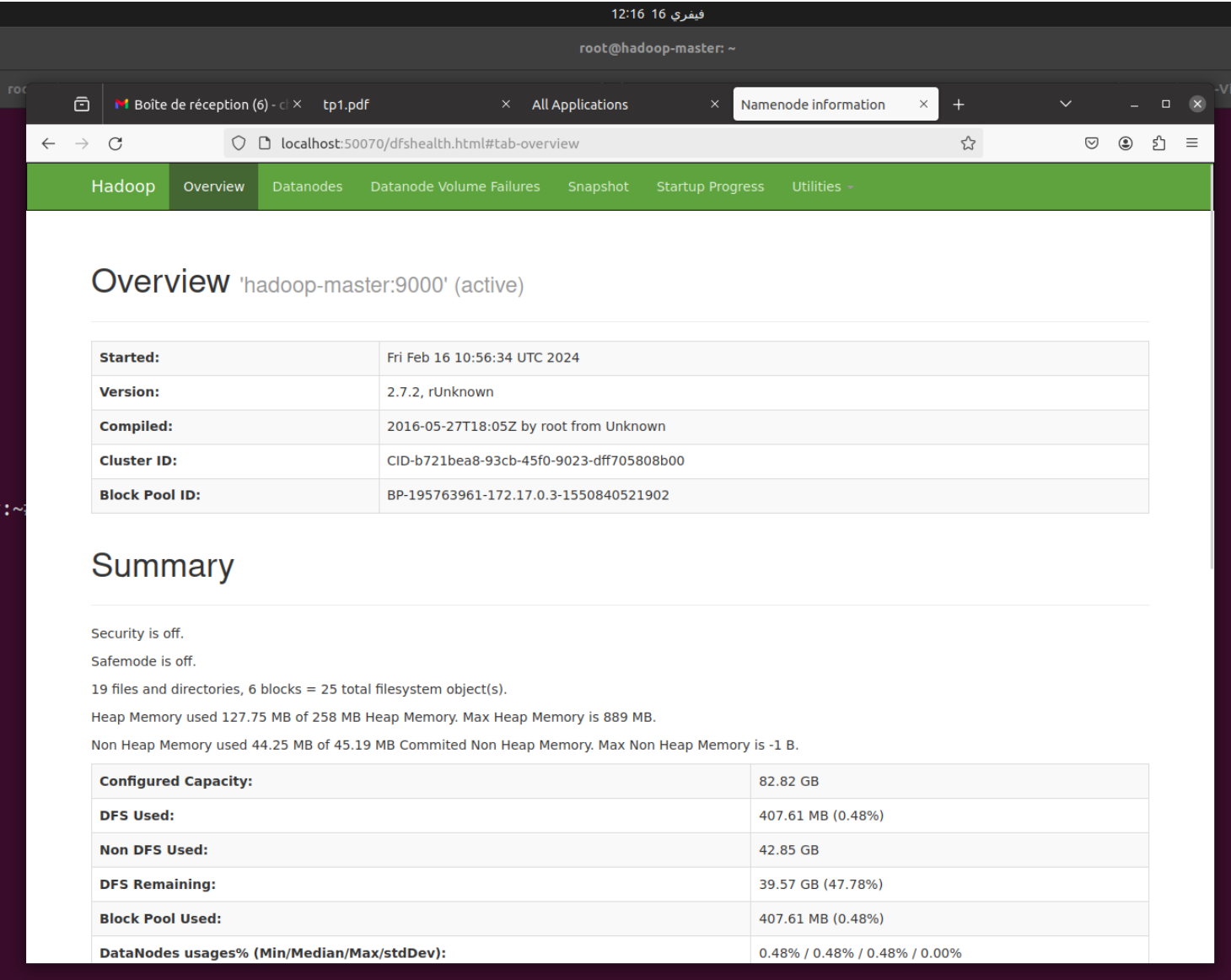
- 4 Créer un répertoire dans HDFS et Charger le fichier purchases dans le répertoire input que vous avez créé, et afficher les dernières lignes du fichier purchases:



The screenshot shows a terminal window with the following commands and output:

```
root@hadoop-master:~# hadoop fs -put purchases.txt input
put: `input/purchases.txt': File exists
root@hadoop-master:~# hadoop fs -ls input
Found 1 items
-rw-r--r--  2 root supergroup 211312924 2024-02-15 20:14 input/purchases.txt
root@hadoop-master:~# hadoop fs -tail input/purchases.txt
31      17:59  Norfolk Toys      164.34  MasterCard
2012-12-31      17:59  Chula Vista      Music  380.67  Visa
2012-12-31      17:59  Hialeah Toys    115.21  MasterCard
2012-12-31      17:59  Indianapolis    Men's Clothing  158.28  MasterCard
2012-12-31      17:59  Norfolk Garden  414.09  MasterCard
2012-12-31      17:59  Baltimore      DVDs    467.3   Visa
2012-12-31      17:59  Santa Ana      Video Games  144.73  Visa
2012-12-31      17:59  Gilbert Consumer Electronics  354.66  Discover
2012-12-31      17:59  Memphis Sporting Goods  124.79  Amex
2012-12-31      17:59  Chicago Men's Clothing  386.54  MasterCard
2012-12-31      17:59  Birmingham     CDs     118.04  Cash
2012-12-31      17:59  Las Vegas      Health and Beauty  420.46  Amex
2012-12-31      17:59  Wichita Toys   383.9   Cash
2012-12-31      17:59  Tucson Pet Supplies  268.39  MasterCard
2012-12-31      17:59  Glendale      Women's Clothing  68.05   Amex
2012-12-31      17:59  Albuquerque    Toys    345.7   MasterCard
2012-12-31      17:59  Rochester     DVDs    399.57  Amex
2012-12-31      17:59  Greensboro     Baby    277.27  Discover
2012-12-31      17:59  Arlington     Women's Clothing  134.95  MasterCard
2012-12-31      17:59  Corpus Christi DVDs    441.61  Discover
root@hadoop-master:~#
```

5 Interfaces web pour Hadoop



ser

12:16 16 فيفري


root@hadoop-master: ~

Boîte de réception (6) - cl x tp1.pdf

All Applications x

Namenode information x +

localhost:8088/cluster



All Applications

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes
1	0	0	1	0	0 B	16 GB	0 B	0	16	0	2	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:1>

Show 20 entries

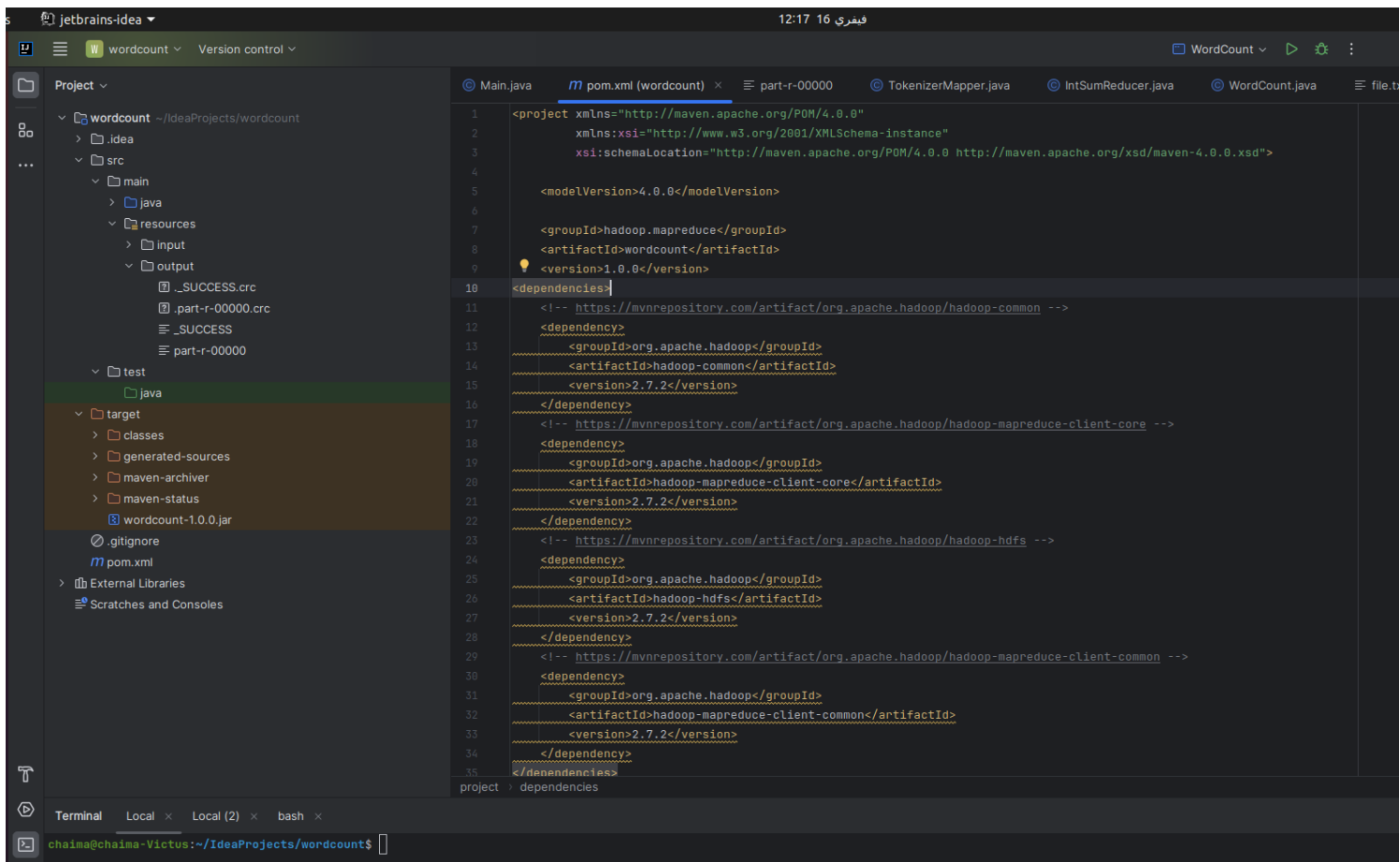
ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress
application_1708075340406_0001	root	word count	MAPREDUCE	default	Fri Feb 16 10:23:05 +0100 2024	Fri Feb 16 10:24:26 +0100 2024	FINISHED	SUCCEEDED	

Showing 1 to 1 of 1 entries

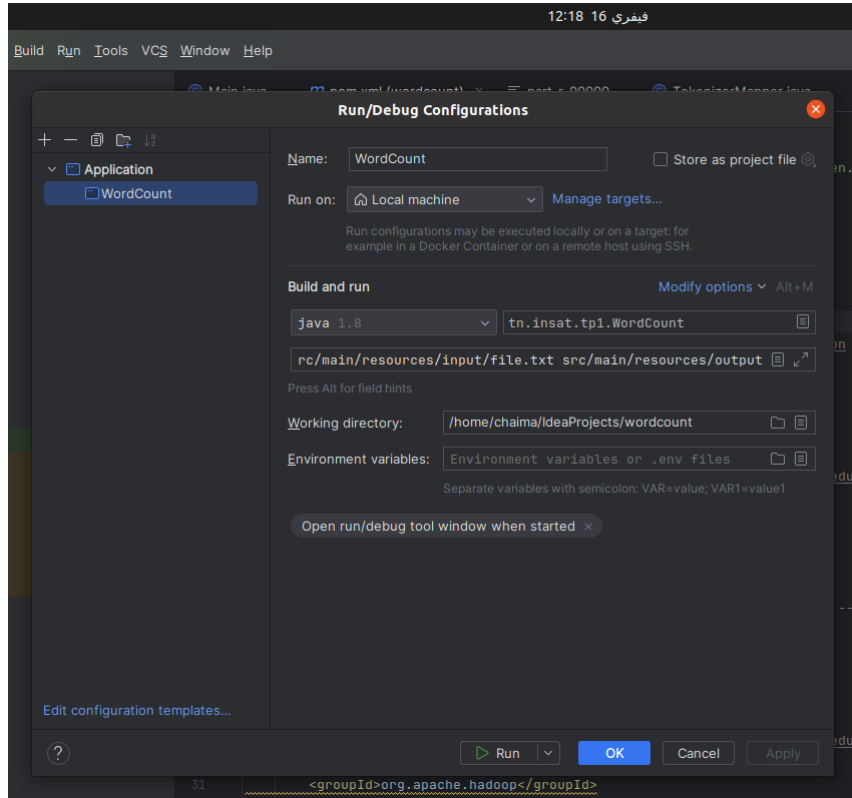
2

Map Reduce

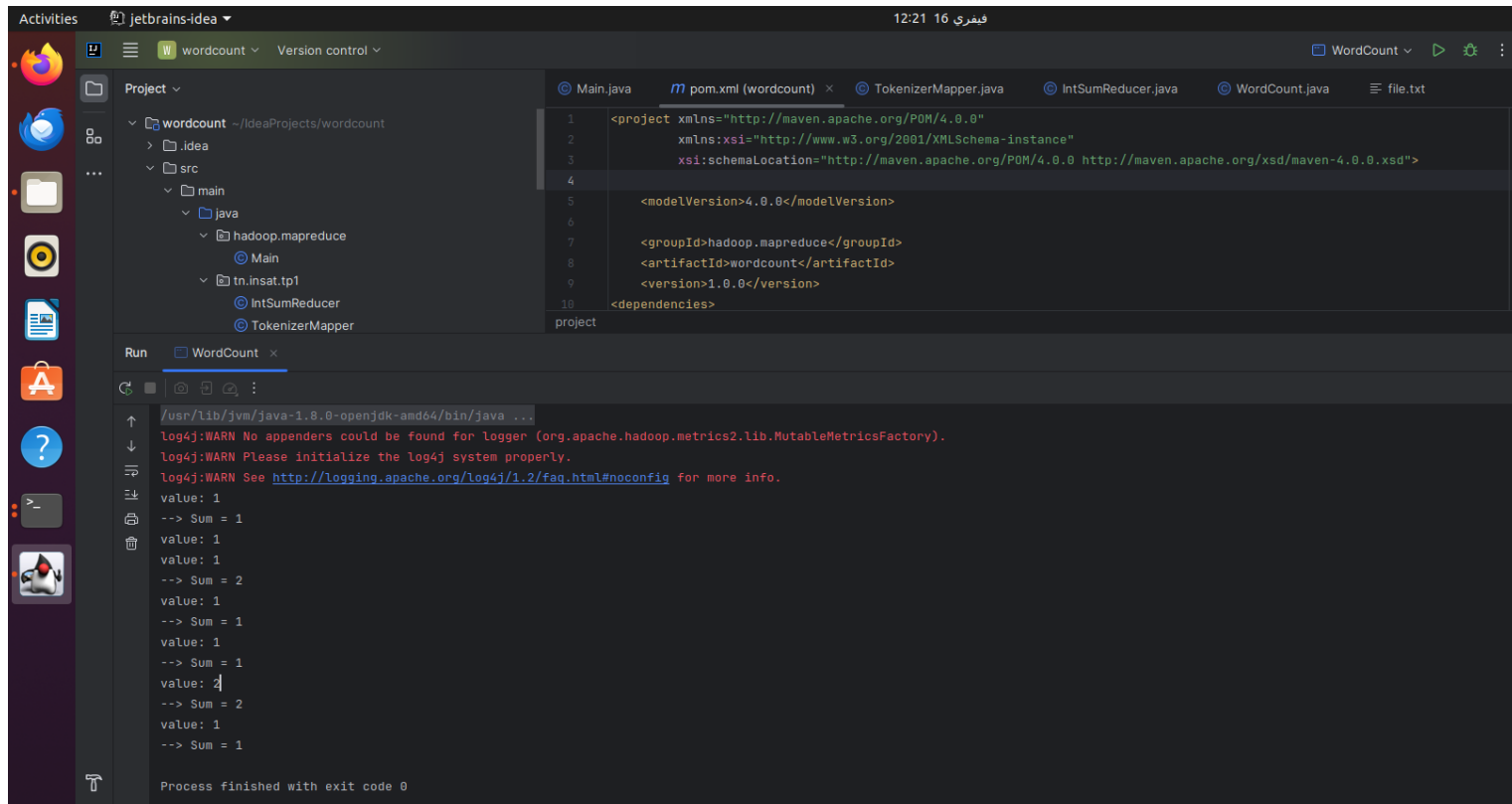
1 Création du fichier pom.xml

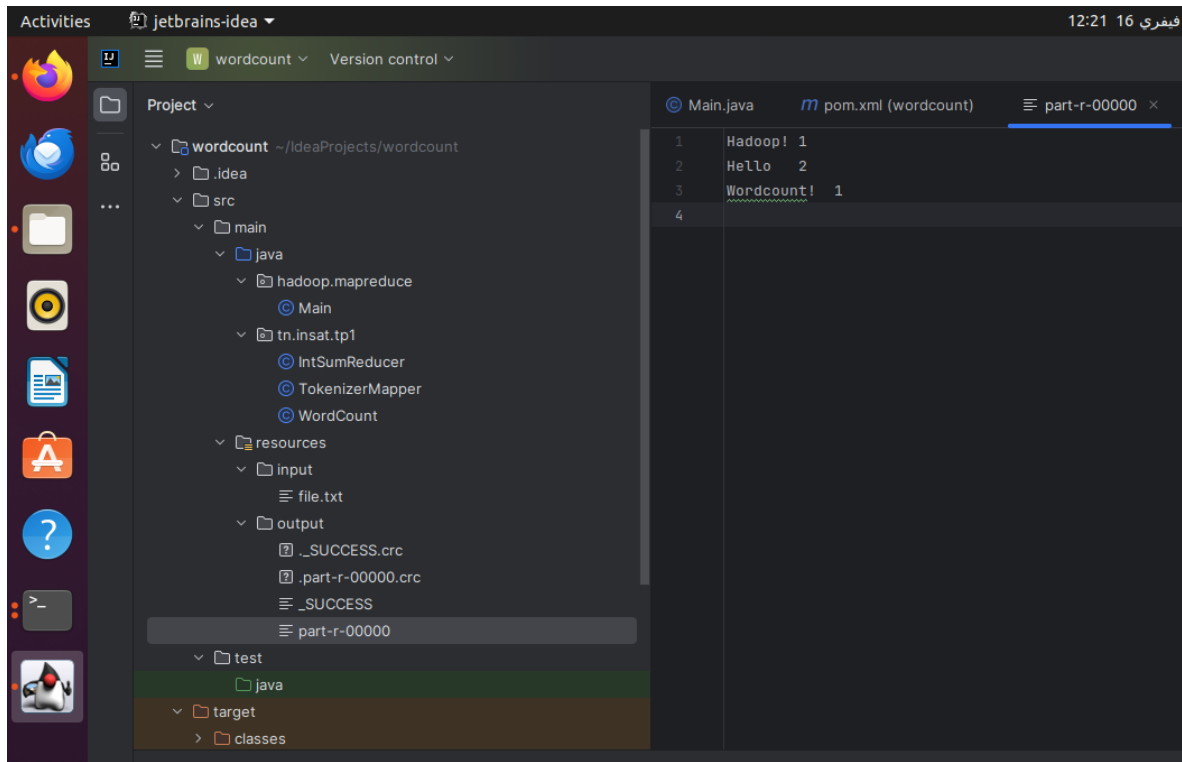


2 Configuration de l'exécution

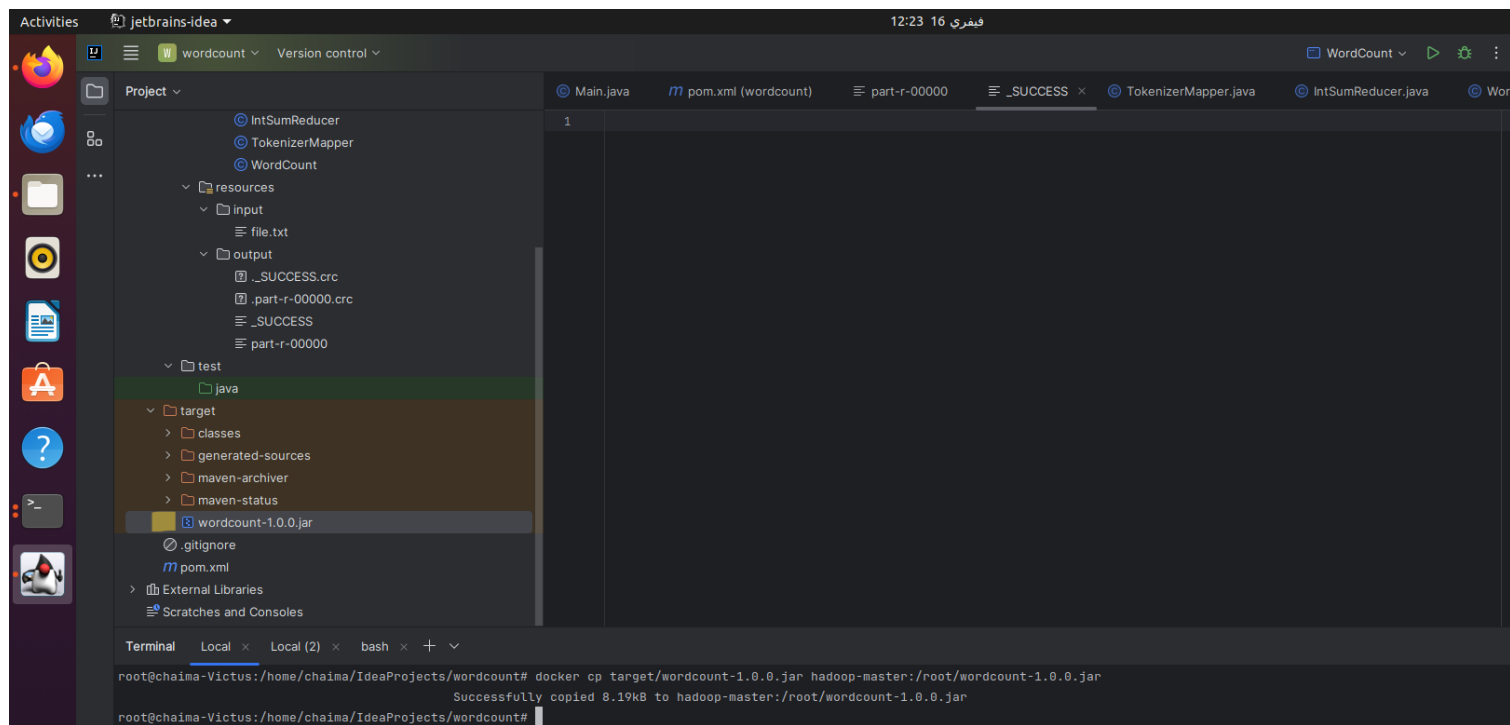


3 Exécution

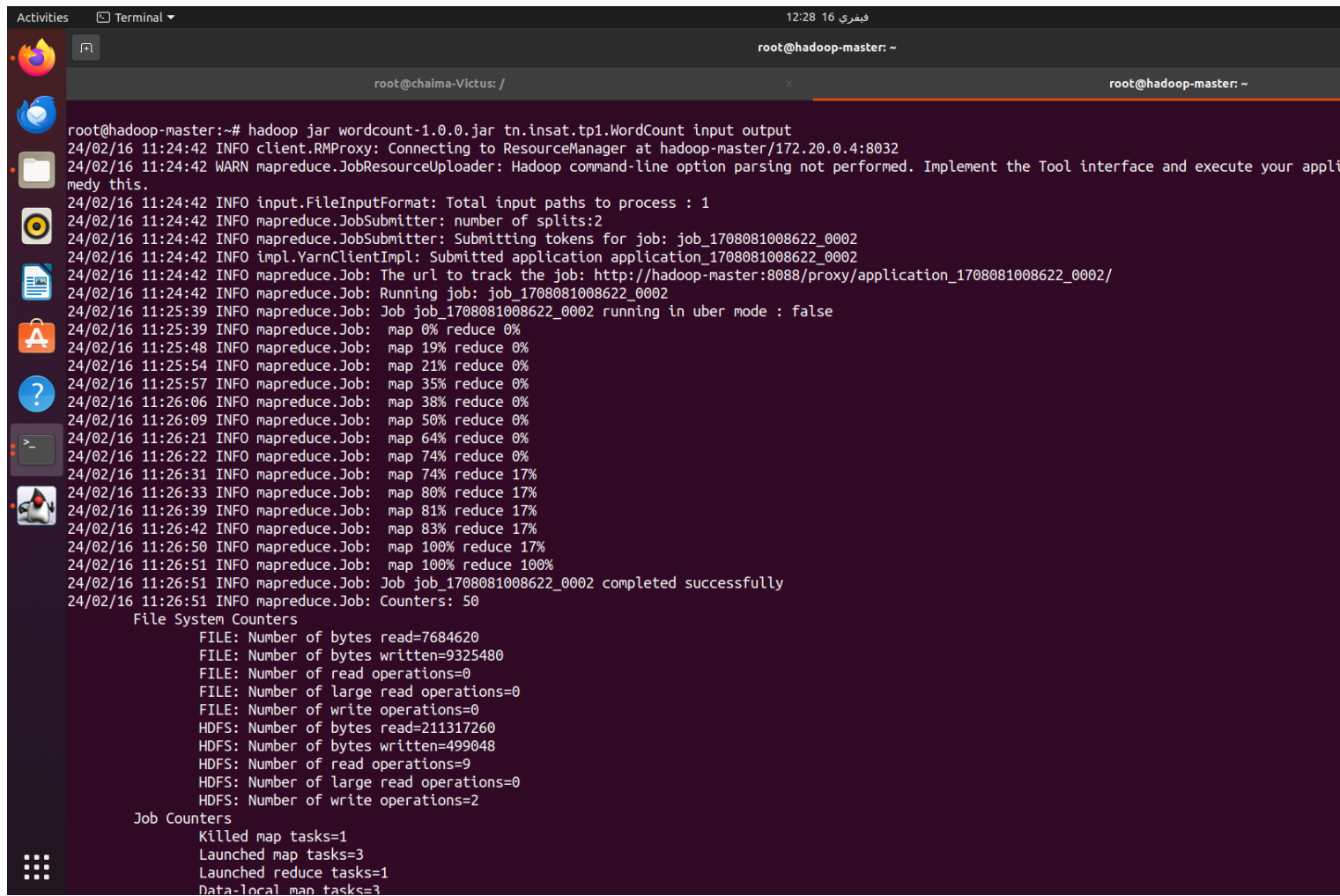




4 Création de l'artifact et copie dans le conteneur master



5 Lancer le job map reduce

A terminal window titled 'Terminal' with a dark background and light text. The window shows the execution of a Hadoop MapReduce job. The user is at the prompt 'root@hadoop-master: ~' and has entered the command 'hadoop jar wordcount-1.0.0.jar tn.insat.tp1.WordCount input output'. The logs show the job submission process, including connecting to the ResourceManager, submitting tokens, and the job running in uber mode. Progress updates show the map and reduce percentages increasing from 0% to 100%. The job completes successfully, and the final counters are displayed, including File System Counters and Job Counters.

```
root@hadoop-master:~# hadoop jar wordcount-1.0.0.jar tn.insat.tp1.WordCount input output
24/02/16 11:24:42 INFO client.RMPProxy: Connecting to ResourceManager at hadoop-master/172.20.0.4:8032
24/02/16 11:24:42 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with this.
24/02/16 11:24:42 INFO input.FileInputFormat: Total input paths to process : 1
24/02/16 11:24:42 INFO mapreduce.JobSubmitter: number of splits:2
24/02/16 11:24:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1708081008622_0002
24/02/16 11:24:42 INFO impl.YarnClientImpl: Submitted application application_1708081008622_0002
24/02/16 11:24:42 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1708081008622_0002/
24/02/16 11:24:42 INFO mapreduce.Job: Running job: job_1708081008622_0002
24/02/16 11:25:39 INFO mapreduce.Job: Job job_1708081008622_0002 running in uber mode : false
24/02/16 11:25:39 INFO mapreduce.Job: map 0% reduce 0%
24/02/16 11:25:48 INFO mapreduce.Job: map 19% reduce 0%
24/02/16 11:25:54 INFO mapreduce.Job: map 21% reduce 0%
24/02/16 11:25:57 INFO mapreduce.Job: map 35% reduce 0%
24/02/16 11:26:06 INFO mapreduce.Job: map 38% reduce 0%
24/02/16 11:26:09 INFO mapreduce.Job: map 50% reduce 0%
24/02/16 11:26:21 INFO mapreduce.Job: map 64% reduce 0%
24/02/16 11:26:22 INFO mapreduce.Job: map 74% reduce 0%
24/02/16 11:26:31 INFO mapreduce.Job: map 74% reduce 17%
24/02/16 11:26:33 INFO mapreduce.Job: map 80% reduce 17%
24/02/16 11:26:39 INFO mapreduce.Job: map 81% reduce 17%
24/02/16 11:26:42 INFO mapreduce.Job: map 83% reduce 17%
24/02/16 11:26:50 INFO mapreduce.Job: map 100% reduce 17%
24/02/16 11:26:51 INFO mapreduce.Job: map 100% reduce 100%
24/02/16 11:26:51 INFO mapreduce.Job: Job job_1708081008622_0002 completed successfully
24/02/16 11:26:51 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=7684620
    FILE: Number of bytes written=9325480
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=211317260
    HDFS: Number of bytes written=499048
    HDFS: Number of read operations=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=3
    Launched reduce tasks=1
    Data-local map tasks=3
```

```
Activities Terminal 12:28 16 فيفري
root@hadoop-master: ~

root@chalma-Victus: /
x
root@hadoop-master: ~

Launched reduce tasks=1
Data-local map tasks=3
Total time spent by all maps in occupied slots (ms)=135604
Total time spent by all reduces in occupied slots (ms)=26727
Total time spent by all map tasks (ms)=135604
Total time spent by all reduce tasks (ms)=26727
Total vcore-milliseconds taken by all map tasks=135604
Total vcore-milliseconds taken by all reduce tasks=26727
Total megabyte-milliseconds taken by all map tasks=138858496
Total megabyte-milliseconds taken by all reduce tasks=27368448

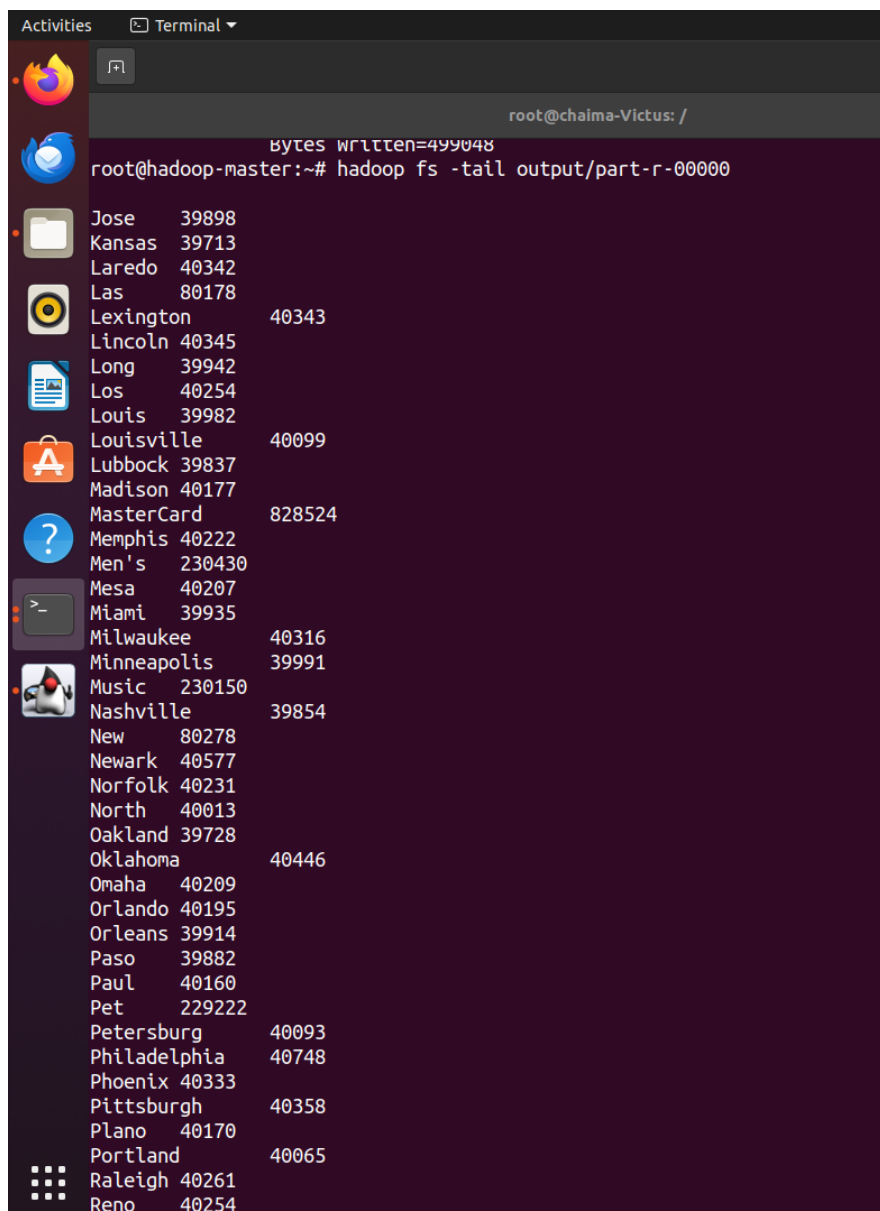
Map-Reduce Framework
Map input records=4138476
Map output records=27982895
Map output bytes=323244504
Map output materialized bytes=1289264
Input split bytes=240
Combine input records=28488079
Combine output records=606926
Reduce input groups=51053
Reduce shuffle bytes=1289264
Reduce input records=101742
Reduce output records=51053
Spilled Records=708668
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=342
CPU time spent (ms)=117490
Physical memory (bytes) snapshot=760700928
Virtual memory (bytes) snapshot=6052577280
Total committed heap usage (bytes)=553123840

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=211317020
File Output Format Counters
Bytes Written=499048

root@hadoop-master:~#
```

6 Afficher les dernières lignes du fichier généré output/part-r-00000

A terminal window titled 'Terminal' with a dark background. The prompt is 'root@chaima-Victus: /'. The command 'root@hadoop-master:~# hadoop fs -tail output/part-r-00000' has been executed. The output shows 'bytes written=499048' followed by a list of city names and their corresponding values. The cities listed are Jose, Kansas, Laredo, Las, Lexington, Lincoln, Long, Los, Louis, Louisville, Lubbock, Madison, MasterCard, Memphis, Men's, Mesa, Miami, Milwaukee, Minneapolis, Music, Nashville, New, Newark, Norfolk, North, Oakland, Oklahoma, Omaha, Orlando, Orleans, Paso, Paul, Pet, Petersburg, Philadelphia, Phoenix, Pittsburgh, Plano, Portland, Raleigh, and Reno. The values range from 229222 to 828524.

```
root@chaima-Victus: /
root@hadoop-master:~# hadoop fs -tail output/part-r-00000
bytes written=499048
Jose      39898
Kansas    39713
Laredo    40342
Las        80178
Lexington 40343
Lincoln   40345
Long      39942
Los       40254
Louis     39982
Louisville 40099
Lubbock   39837
Madison   40177
MasterCard 828524
Memphis   40222
Men's     230430
Mesa       40207
Miami     39935
Milwaukee 40316
Minneapolis 39991
Music     230150
Nashville 39854
New        80278
Newark    40577
Norfolk   40231
North     40013
Oakland   39728
Oklahoma  40446
Omaha     40209
Orlando   40195
Orleans   39914
Paso      39882
Paul      40160
Pet       229222
Petersburg 40093
Philadelphia 40748
Phoenix   40333
Pittsburgh 40358
Plano     40170
Portland  40065
Raleigh   40261
Reno      40254
```


7 Monitorer les Jobs Map Reduce

12:29 16 فيفري

root@hadoop-master: ~

Boîte de réception (7) - cl × tp1.pdf × All Applications × Namenode information × +

localhost:8088/cluster



Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes
1	0	0	1	0	0 B	16 GB	0 B	0	16	0	2	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Ma
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192,

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress
application_1708075340406_0001	root	word count	MAPREDUCE	default	Fri Feb 16 10:23:05 +0100 2024	Fri Feb 16 10:24:26 +0100 2024	FINISHED	SUCCEEDED	


Showing 1 to 1 of 1 entries

12:29 16 فيفري

root@hadoop-master: ~

Boîte de réception (7) - cl × tp1.pdf × localhost:8041/node × Namenode information × +

localhost:8041/node

 Logged in as: dr.who

ResourceManager

NodeManager

Node Information

List of Applications

List of Containers

Tools

NodeManager information

Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	false
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	false
Total VCores allocated for Containers	8
NodeHealthyStatus	true
LastNodeHealthTime	Fri Feb 16 11:28:50 UTC 2024
NodeHealthReport	
Node Manager Version:	2.7.2 from Unknown by root source checksum c63f7cc71b8f63249e35126f0f7492d on 2016-05-27T18:16Z
Hadoop Version:	2.7.2 from Unknown by root source checksum d0fda26633fa762bff87ec759ebe689c on 2016-05-27T18:05Z

3

Bonus

1 Ajout d'une classe "CreditCardMapper"

The screenshot displays the IntelliJ IDEA IDE interface. The top bar shows 'jetbrains-idea' and the time '12:54 16 فيفري'. The project 'wordcount' is open, and the 'CreditCardMapper' class is being implemented. The code uses Hadoop MapReduce APIs. The terminal shows the successful build and execution of the application.

```
1 package tn.insat.tp1;
2
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapreduce.Mapper;
6 import java.io.IOException;
7
8 public class CreditCardMapper extends Mapper<Object, Text, Text, IntWritable>{
9     1 usage
10     private final static IntWritable one = new IntWritable( value: 1);
11     2 usages
12     private Text creditCardType = new Text();
13
14     public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
15         String[] tokens = value.toString().split( "\\s+");
16         String paymentMode = tokens[tokens.length - 1];
17
18         creditCardType.set(paymentMode);
19         context.write(creditCardType, one);
20     }
21 }
```

Terminal output:

```
Downloading from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-utils/3.0.5/plexus-utils-3.0.5.jar
Downloading from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-digest/1.0/plexus-digest-1.0.jar
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-digest/1.0/plexus-digest-1.0.jar (12 kB at 47 kB/s)
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-utils/3.0.5/plexus-utils-3.0.5.jar (230 kB at 854 kB/s)
[INFO] Installing /home/chaïma/IdeaProjects/wordcount/target/wordcount-1.0.0.jar to /root/.m2/repository/hadoop/mapreduce/wordcount/1.0.0/wordcount-1.0.0.jar
[INFO] Installing /home/chaïma/IdeaProjects/wordcount/pom.xml to /root/.m2/repository/hadoop/mapreduce/wordcount/1.0.0/wordcount-1.0.0.pom
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 1.820 s
[INFO] Finished at: 2024-02-16T12:52:12+01:00
[INFO] -----
root@chaïma-Victus:/home/chaïma/IdeaProjects/wordcount# docker cp target/wordcount-1.0.0.jar hadoop-master:/root/wordcount-1.0.1.jar
Successfully copied 9.22kB to hadoop-master:/root/wordcount-1.0.1.jar
root@chaïma-Victus:/home/chaïma/IdeaProjects/wordcount#
```

2 Modification dans la classe main

The screenshot displays the IntelliJ IDEA IDE interface. The left sidebar shows the project structure for 'wordcount', with the 'Main' class in the 'main' package selected. The main editor window shows the code for 'Main.java', which is a Hadoop MapReduce application. The code includes imports for Hadoop classes and a 'main' method that configures a Hadoop job named 'credit card count'. The job is configured with 'CreditCardMapper' as the mapper, 'IntSumReducer' as the combiner and reducer, and 'Text' as the output key class. The input path is set to 'args[0]' and the output path is set to 'args[1]'. The terminal at the bottom shows the build process, including downloading dependencies from Maven Central and copying the resulting JAR file to the Hadoop master node.

```

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.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {

    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "credit card count");
        job.setJarByClass(WordCount.class);
        job.setMapperClass(CreditCardMapper.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(verbose: true) ? 0 : 1);
    }
}

```

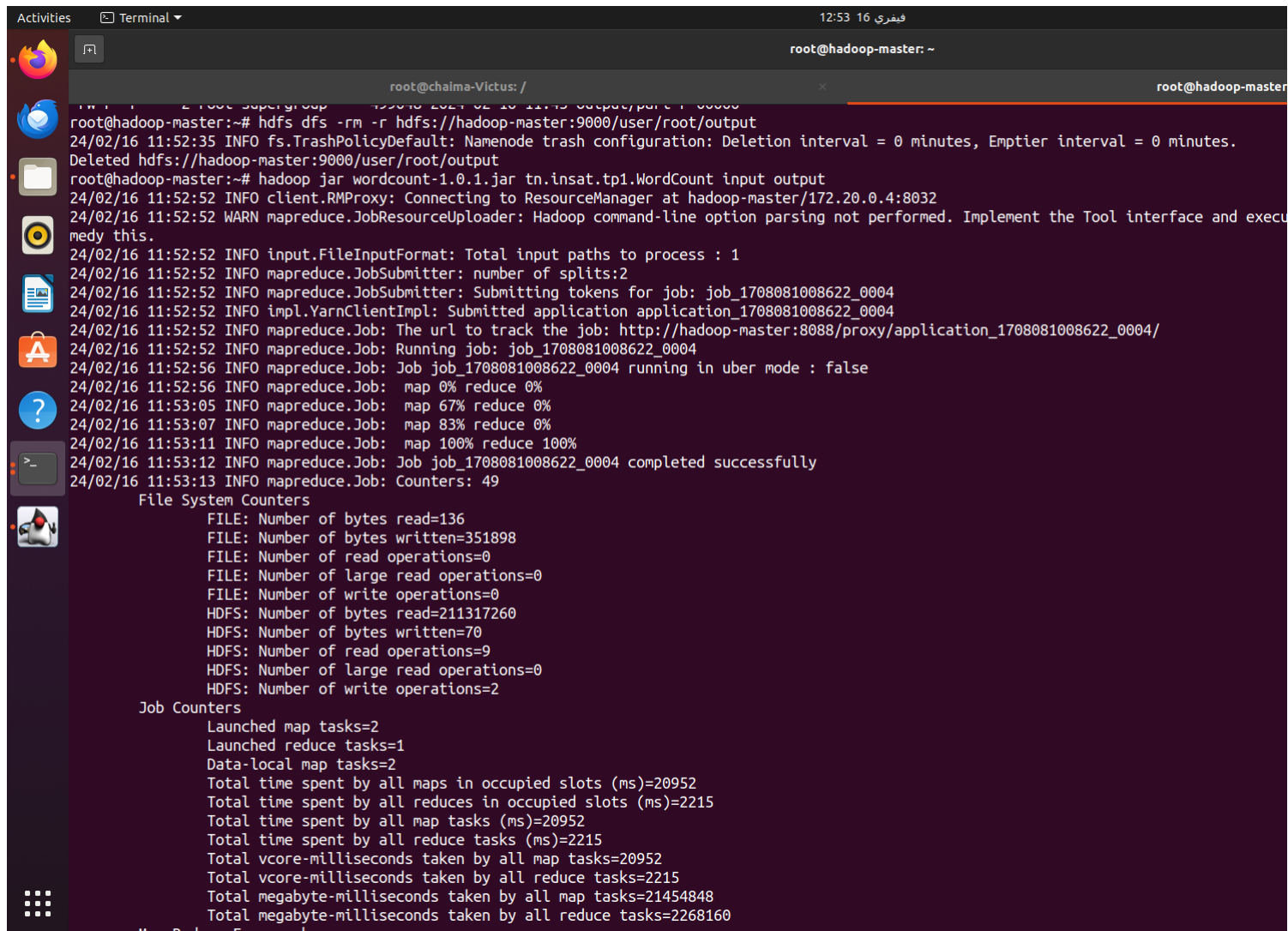
Terminal output:

```

Downloading from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-utils/3.0.5/plexus-utils-3.0.5.jar
Downloading from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-digest/1.0/plexus-digest-1.0.jar
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-digest/1.0/plexus-digest-1.0.jar (12 kB at 47 kB/s)
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-utils/3.0.5/plexus-utils-3.0.5.jar (230 kB at 854 kB/s)
[INFO] Installing /home/chaima/IdeaProjects/wordcount/target/wordcount-1.0.0.jar to /root/.m2/repository/hadoop/mapreduce/wordcount/1.0.0/wordcount-1.0.0.jar
[INFO] Installing /home/chaima/IdeaProjects/wordcount/pom.xml to /root/.m2/repository/hadoop/mapreduce/wordcount/1.0.0/wordcount-1.0.0.pom
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 1.820 s
[INFO] Finished at: 2024-02-16T12:52:12+01:00
[INFO] -----
root@chaima-Victus:/home/chaima/IdeaProjects/wordcount# docker cp target/wordcount-1.0.0.jar hadoop-master:/root/wordcount-1.0.1.jar
Successfully copied 9.22kB to hadoop-master:/root/wordcount-1.0.1.jar
root@chaima-Victus:/home/chaima/IdeaProjects/wordcount#

```


3 Exécution du nouveau artifice dans le master

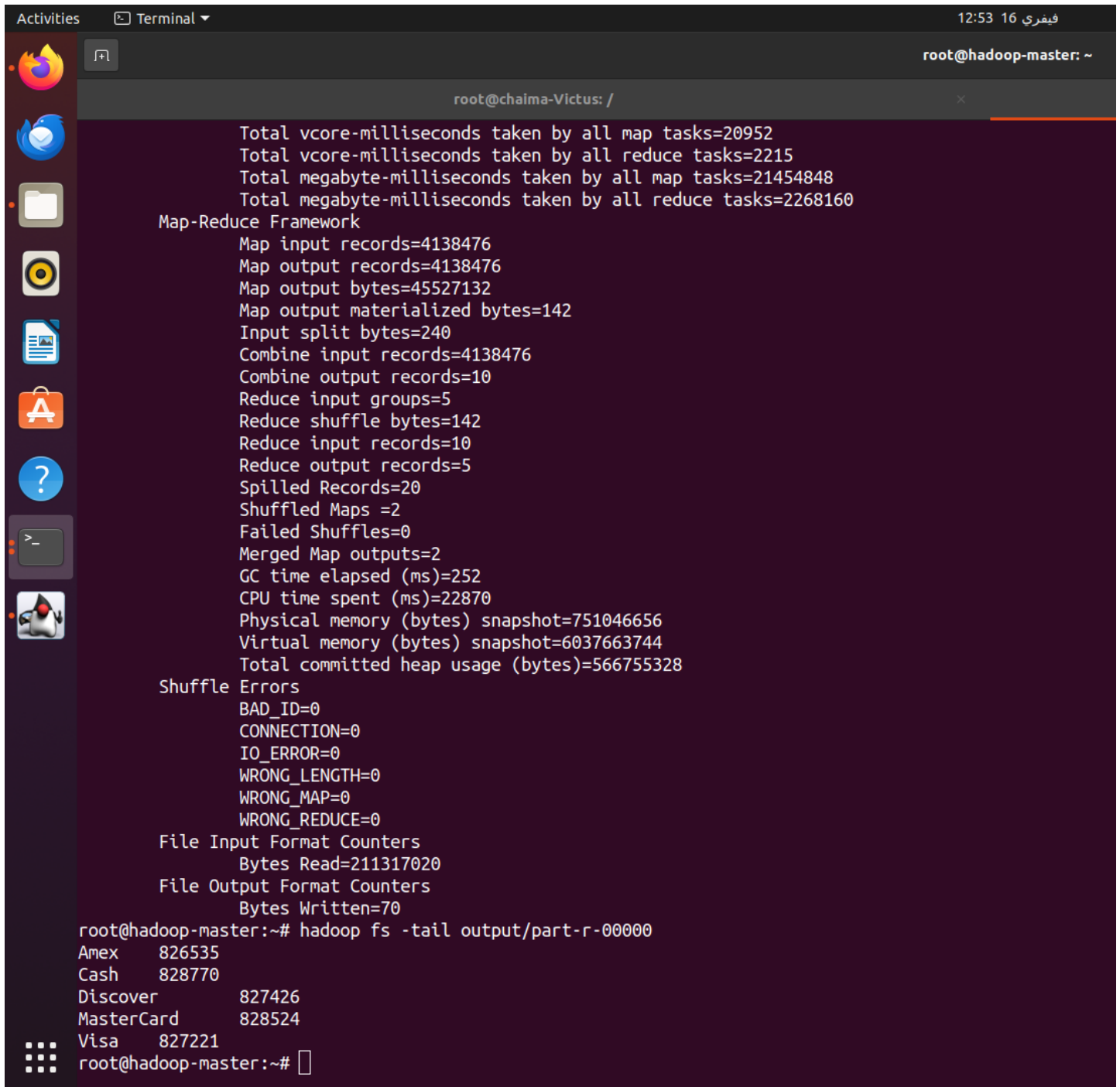


```

root@hadoop-master:~# hdfs dfs -rm -r hdfs://hadoop-master:9000/user/root/output
24/02/16 11:52:35 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted hdfs://hadoop-master:9000/user/root/output
root@hadoop-master:~# hadoop jar wordcount-1.0.1.jar tn.insat.tp1.WordCount input output
24/02/16 11:52:52 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.20.0.4:8032
24/02/16 11:52:52 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute
medy this.
24/02/16 11:52:52 INFO input.FileInputFormat: Total input paths to process : 1
24/02/16 11:52:52 INFO mapreduce.JobSubmitter: number of splits:2
24/02/16 11:52:52 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1708081008622_0004
24/02/16 11:52:52 INFO impl.YarnClientImpl: Submitted application application_1708081008622_0004
24/02/16 11:52:52 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1708081008622_0004/
24/02/16 11:52:52 INFO mapreduce.Job: Running job: job_1708081008622_0004
24/02/16 11:52:56 INFO mapreduce.Job: Job job_1708081008622_0004 running in uber mode : false
24/02/16 11:52:56 INFO mapreduce.Job: map 0% reduce 0%
24/02/16 11:53:05 INFO mapreduce.Job: map 67% reduce 0%
24/02/16 11:53:07 INFO mapreduce.Job: map 83% reduce 0%
24/02/16 11:53:11 INFO mapreduce.Job: map 100% reduce 100%
24/02/16 11:53:12 INFO mapreduce.Job: Job job_1708081008622_0004 completed successfully
24/02/16 11:53:13 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=136
  FILE: Number of bytes written=351898
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=211317260
  HDFS: Number of bytes written=70
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=2
  Launched reduce tasks=1
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=20952
  Total time spent by all reduces in occupied slots (ms)=2215
  Total time spent by all map tasks (ms)=20952
  Total time spent by all reduce tasks (ms)=2215
  Total vcore-milliseconds taken by all map tasks=20952
  Total vcore-milliseconds taken by all reduce tasks=2215
  Total megabyte-milliseconds taken by all map tasks=21454848
  Total megabyte-milliseconds taken by all reduce tasks=2268160

```

4 Affichage du résultat



```

Activities Terminal 12:53 16 فيفري
root@hadoop-master: ~

root@chaima-Victus: /

Total vcore-milliseconds taken by all map tasks=20952
Total vcore-milliseconds taken by all reduce tasks=2215
Total megabyte-milliseconds taken by all map tasks=21454848
Total megabyte-milliseconds taken by all reduce tasks=2268160
Map-Reduce Framework
  Map input records=4138476
  Map output records=4138476
  Map output bytes=45527132
  Map output materialized bytes=142
  Input split bytes=240
  Combine input records=4138476
  Combine output records=10
  Reduce input groups=5
  Reduce shuffle bytes=142
  Reduce input records=10
  Reduce output records=5
  Spilled Records=20
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=252
  CPU time spent (ms)=22870
  Physical memory (bytes) snapshot=751046656
  Virtual memory (bytes) snapshot=6037663744
  Total committed heap usage (bytes)=566755328
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=211317020
File Output Format Counters
  Bytes Written=70
root@hadoop-master:~# hadoop fs -tail output/part-r-00000
Amex      826535
Cash      828770
Discover   827426
MasterCard 828524
Visa      827221
root@hadoop-master:~#

```