REPUBLIQUE TUNISIENNE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Carthage
Institut National des Sciences Appliquées et de Technologie

Compte Rendu

Spécialité : Réseau informatique et télécommunications

TP1 - Le traitement Batch avec Hadoop
HDFS et Map Reduce

Présenté par

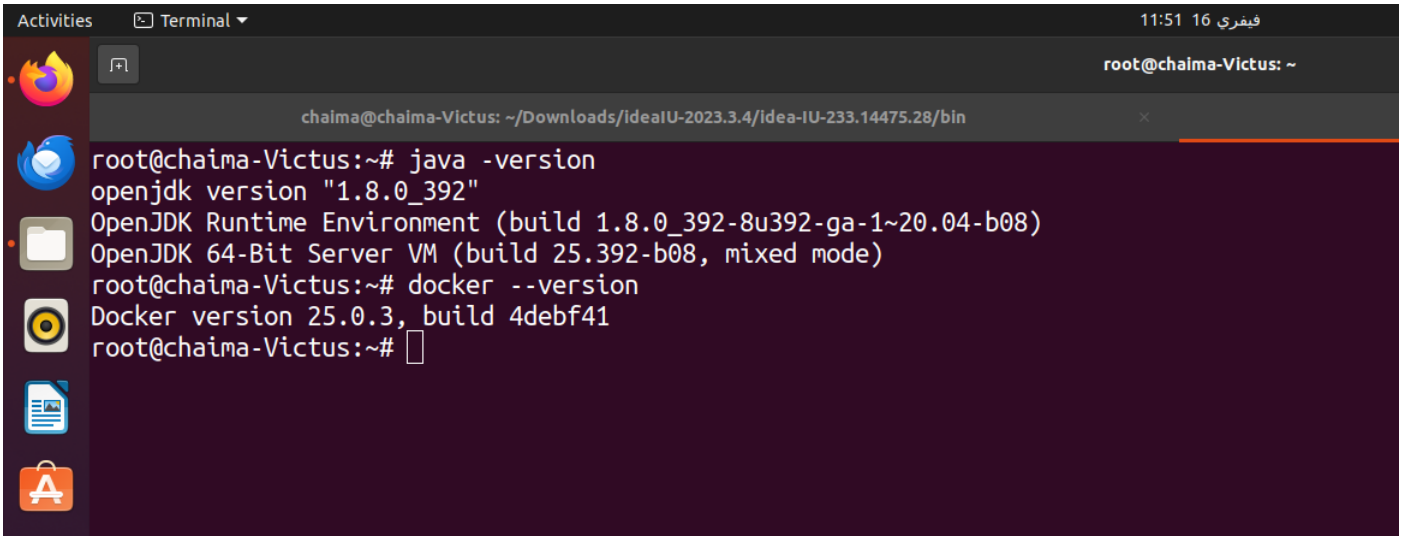
Ghaouari Koussay
Raach Chaima

Année Universitaire : 2023 / 2024

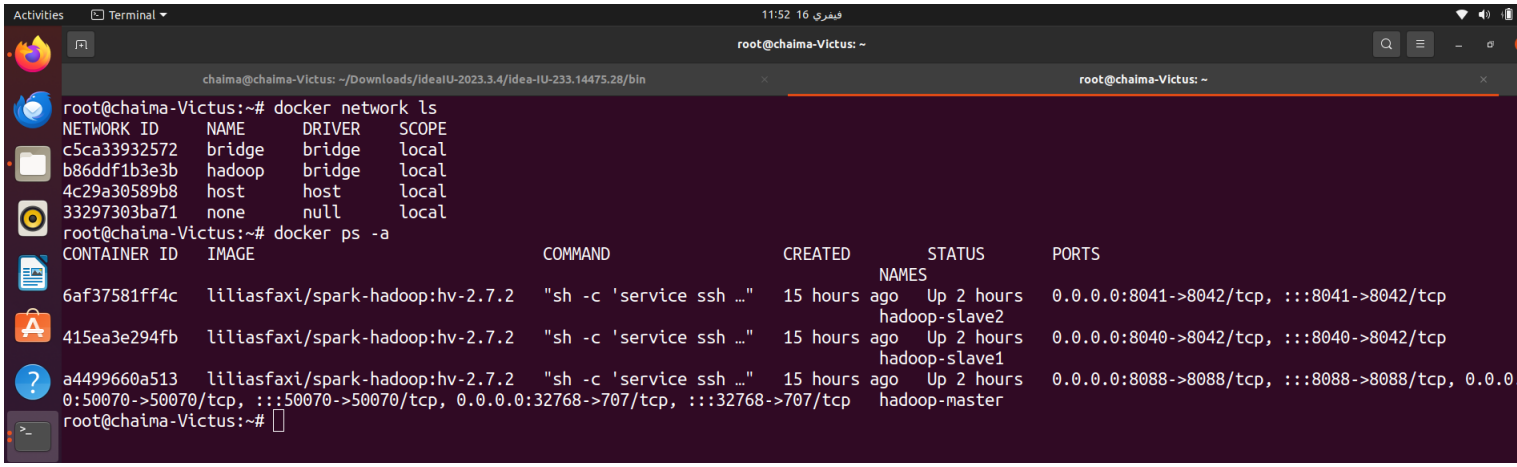
1

Setup et premiers pas

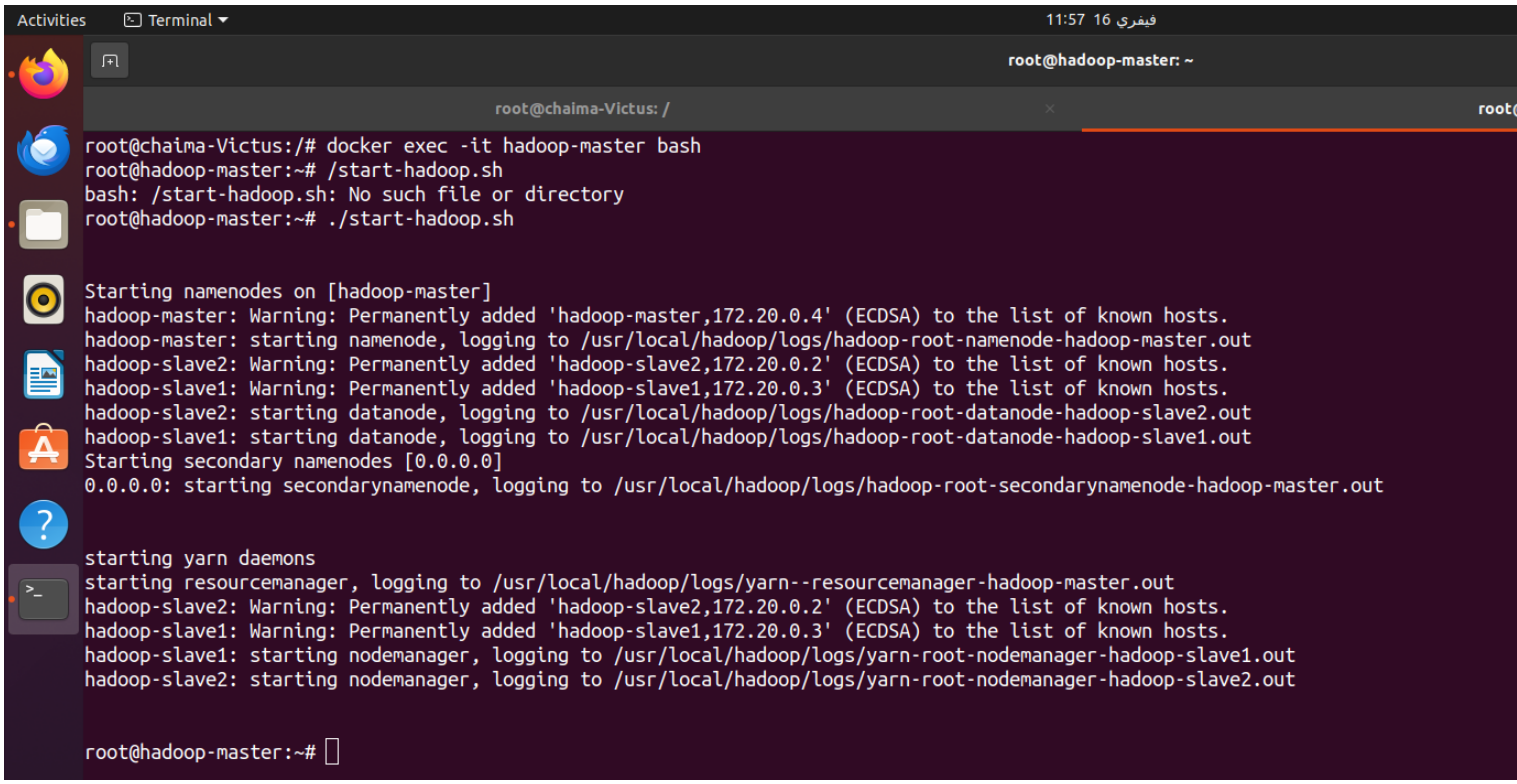
1 Installation des des outils



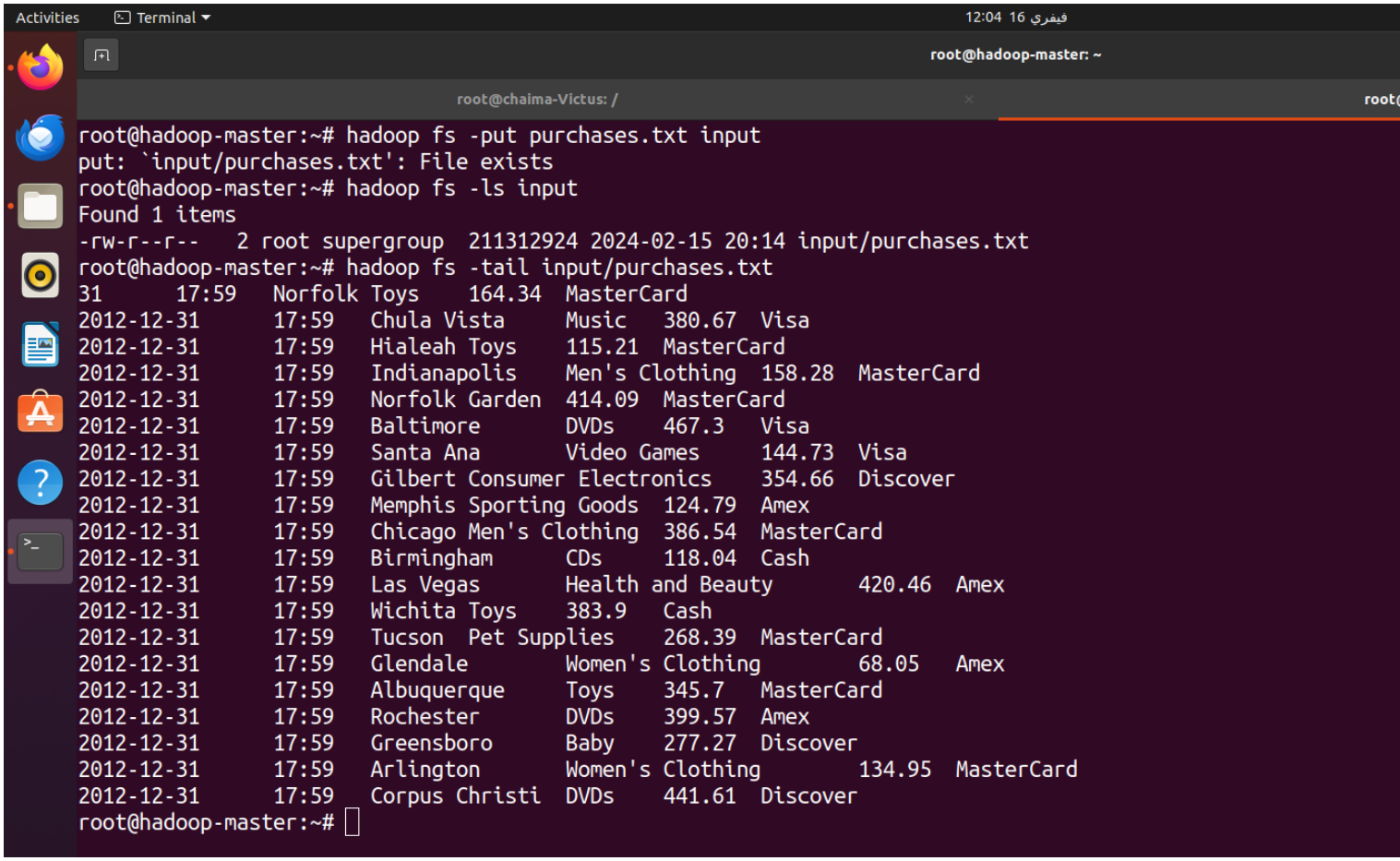
2 Création du réseau et des conteneurs docker



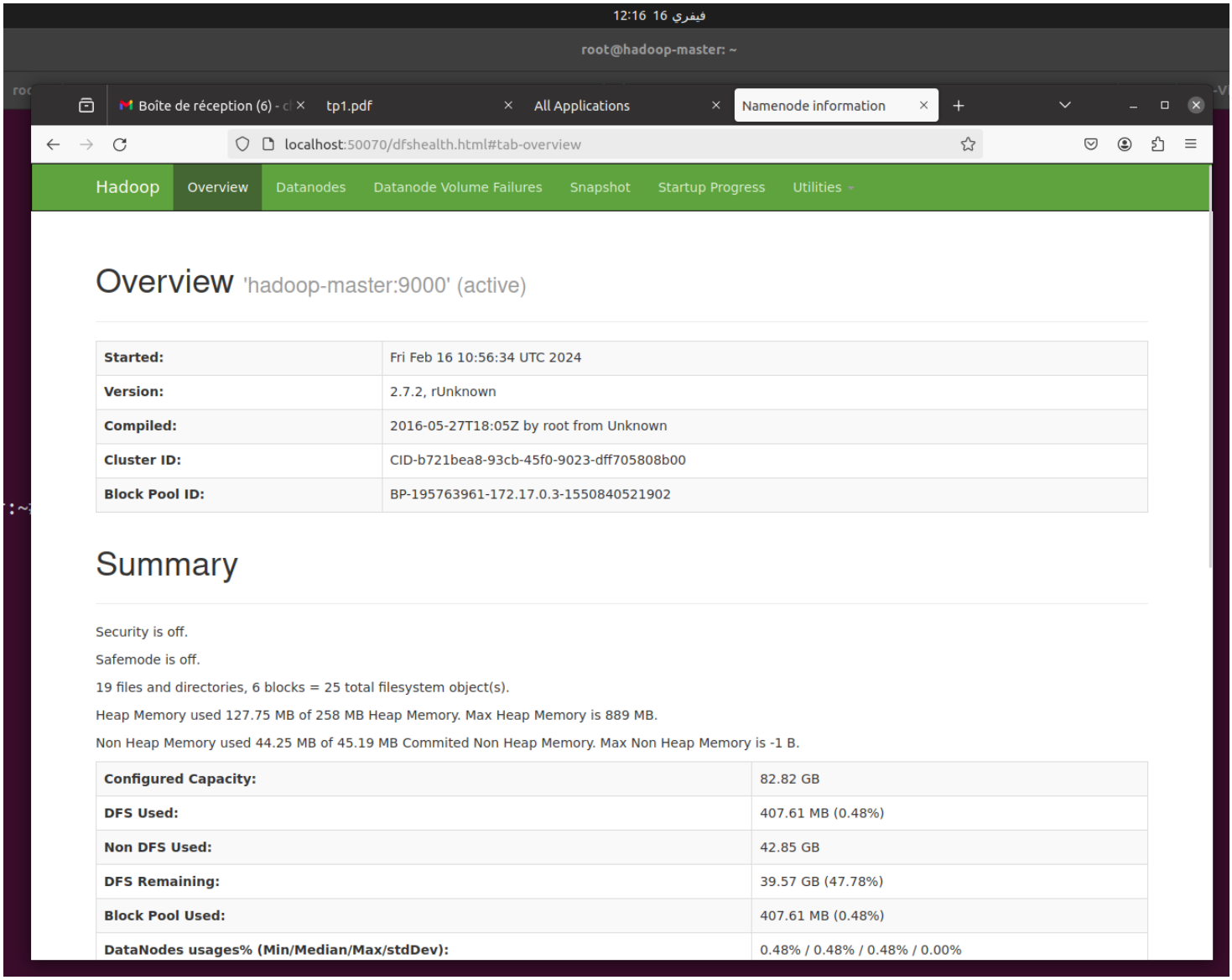
3 Entrer dans le contenaire master pour commencer à l'utiliser et Lancer hadoop et yarn

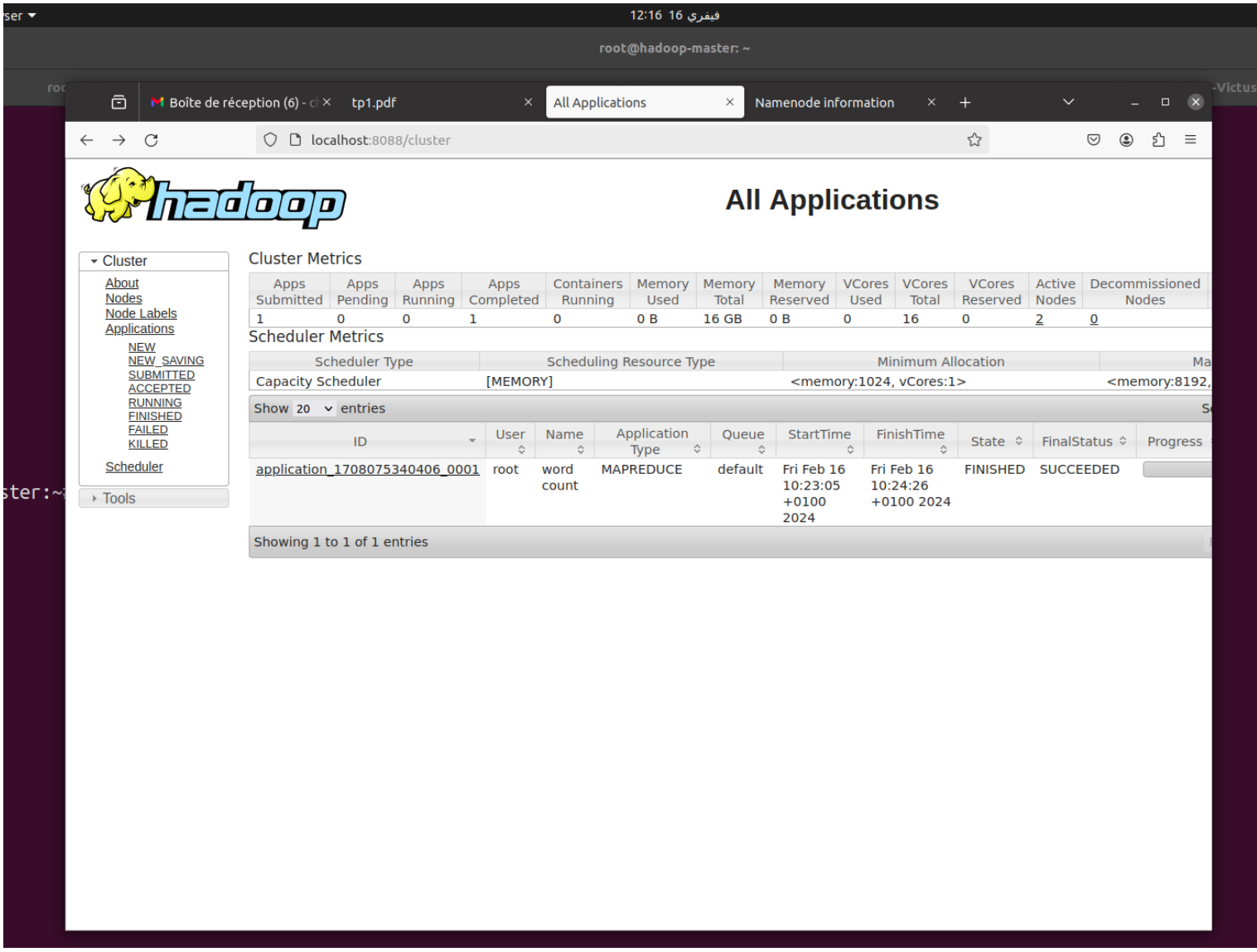


4 Créer un répertoire dans HDFS et Charger le fichier purchases dans le répertoire input que vous avez créé, et afficher les dernières lignes du fichier purchases:



5 Interfaces web pour Hadoop

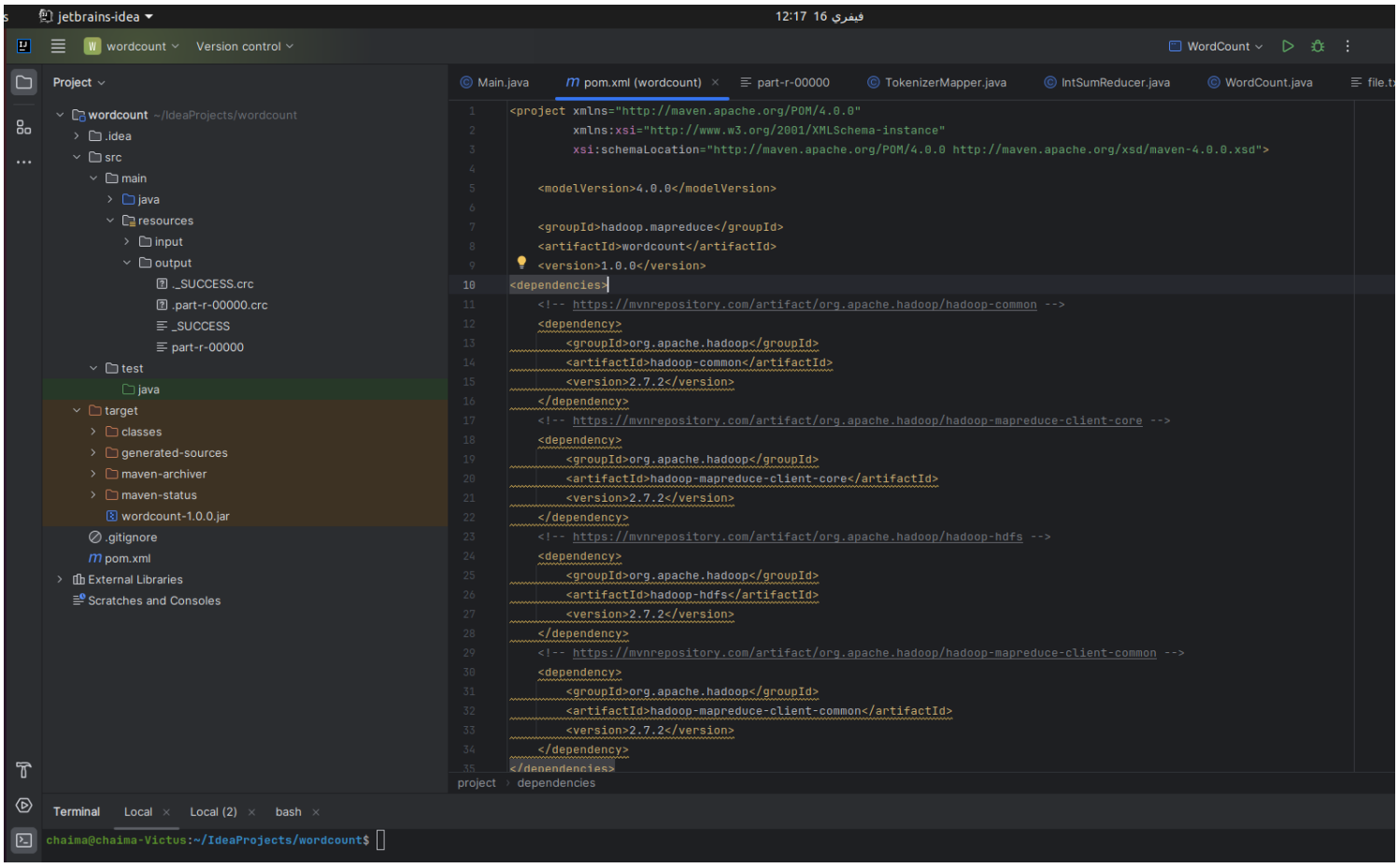



root@hadoop-master: ~
Boîte de réception (6) - c
tp1.pdf
All Applications
Namenode information
localhost:8088/cluster
All Applications
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
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
Capacity Scheduler
[MEMORY]
<memory:1024, 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

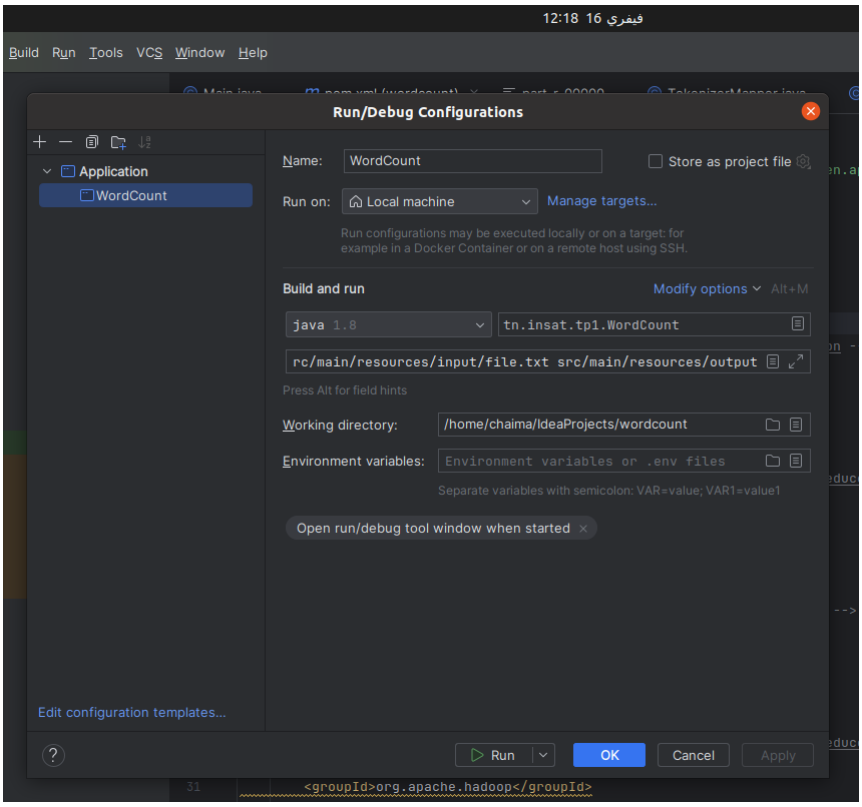
2

Map Reduce

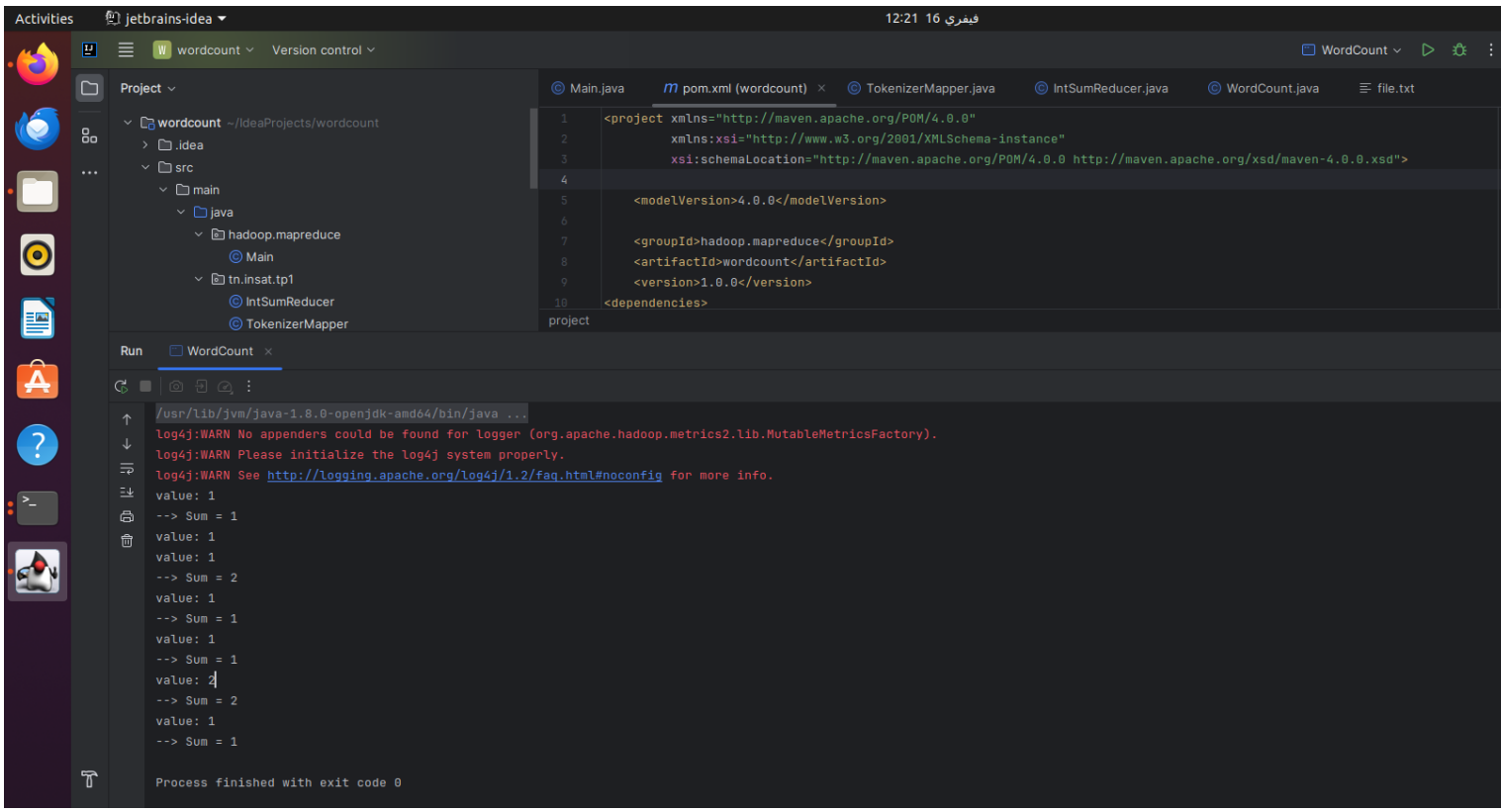
1 Création du fichier pom.xml

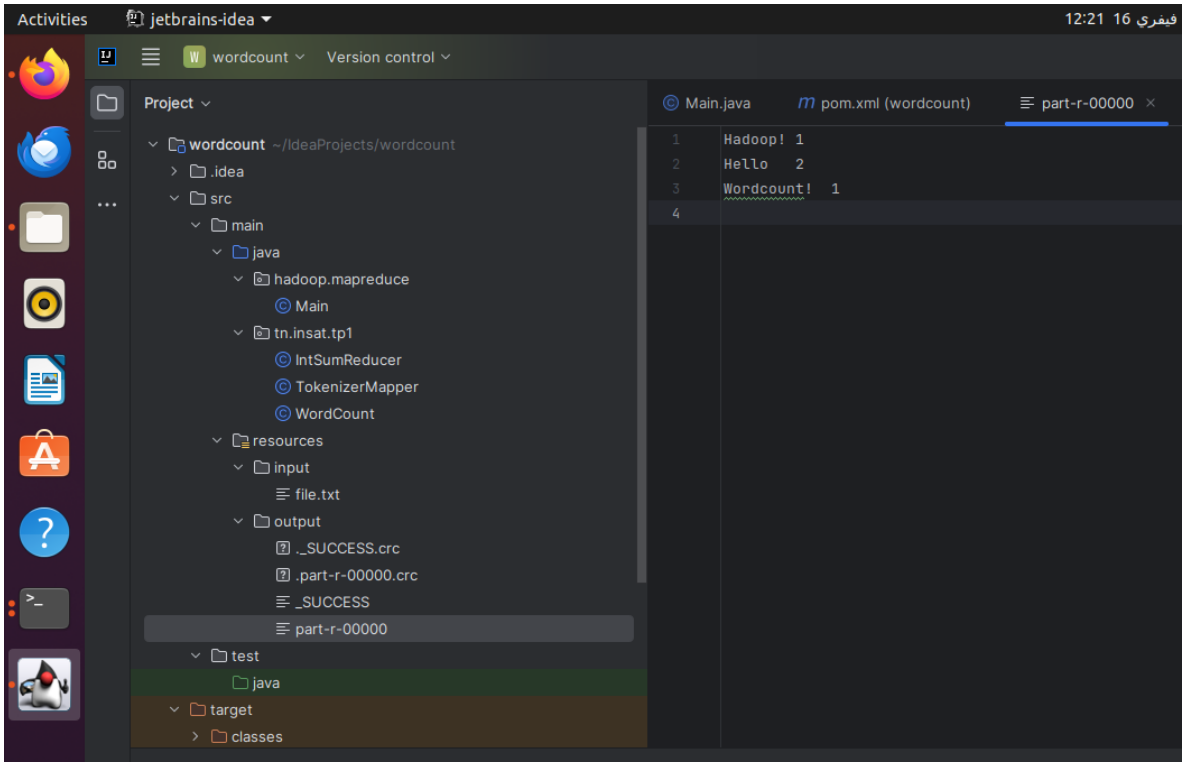


2 Configuration de l'exécution

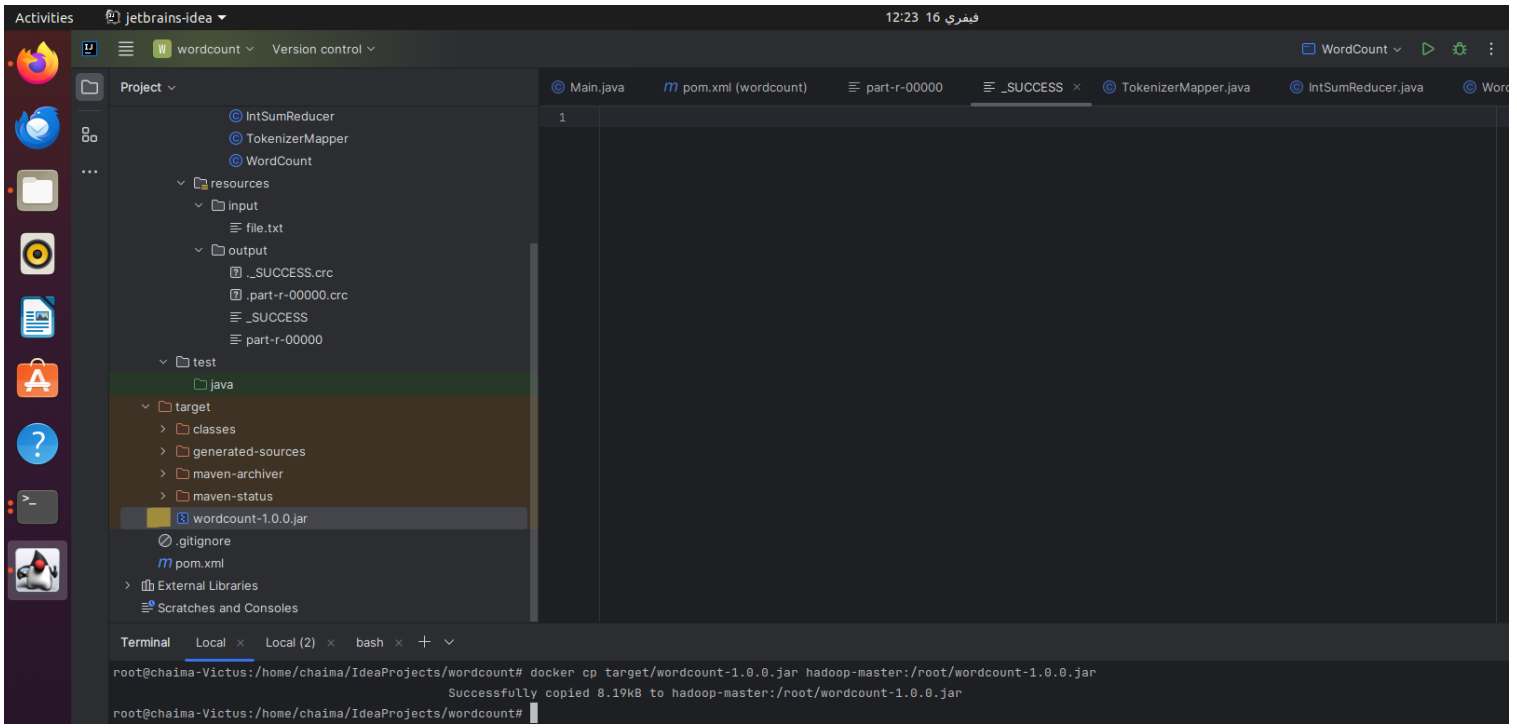


3 Exécution

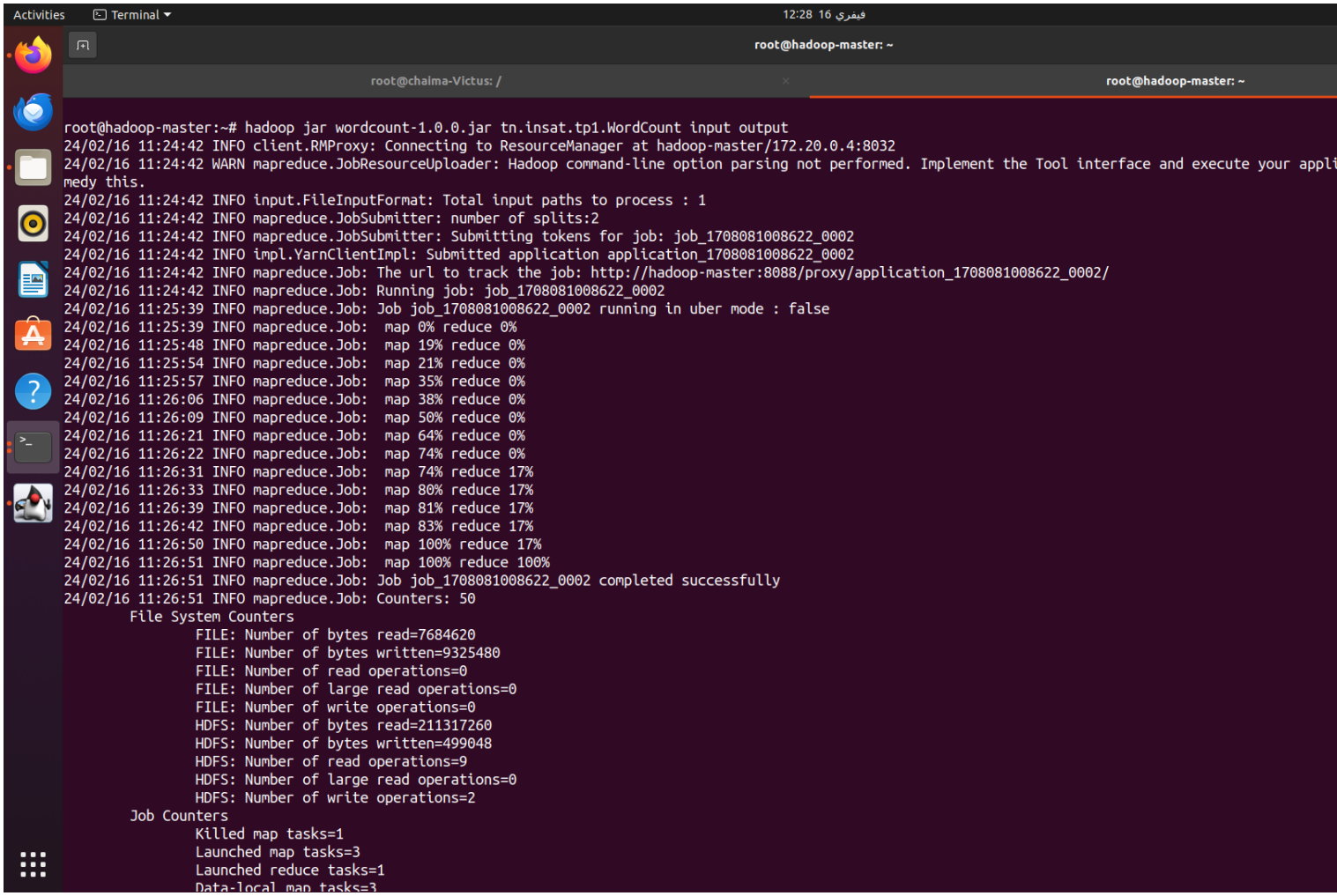


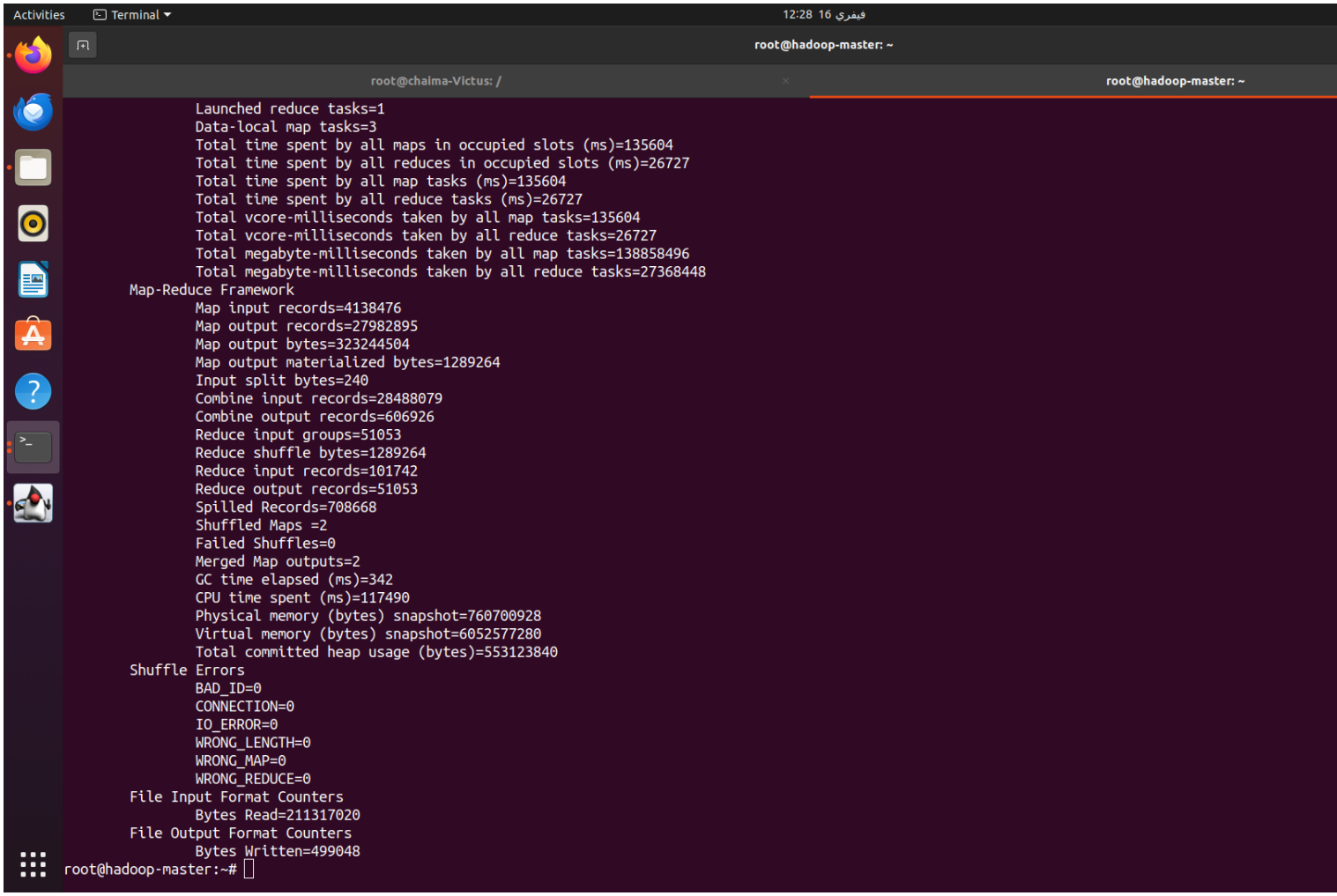


4 Création de l'artifact et copie dans le contenaire master

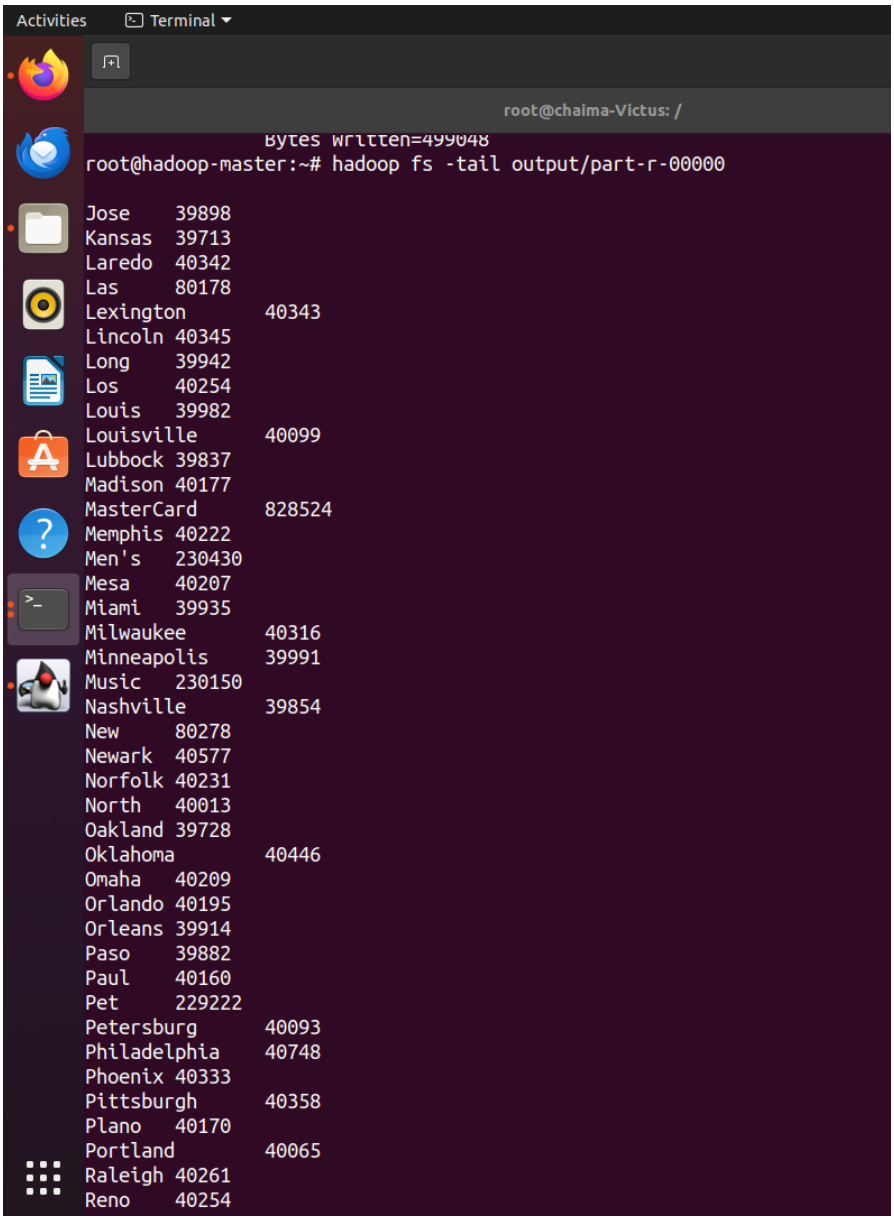


5 Lancer le job map reduce

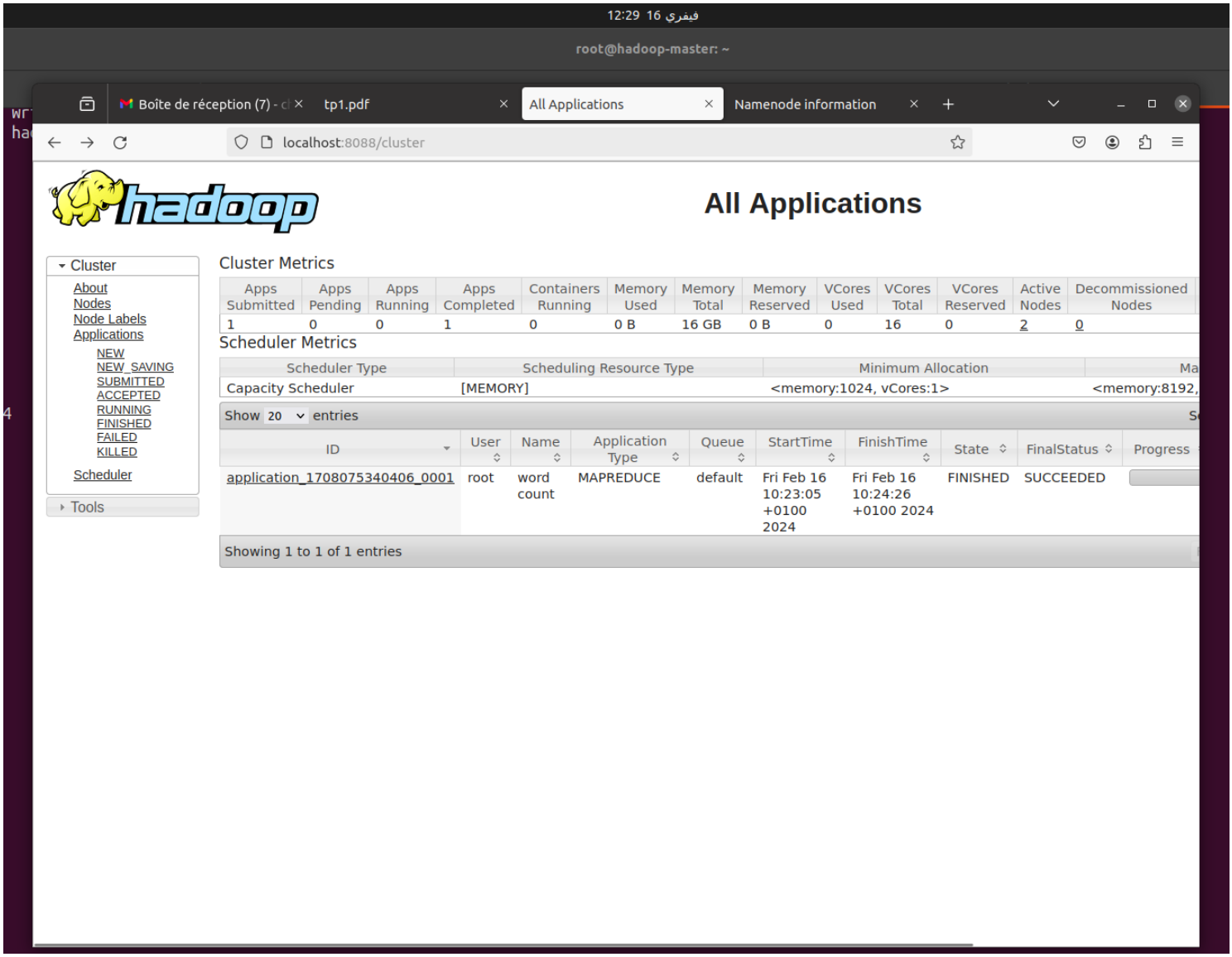


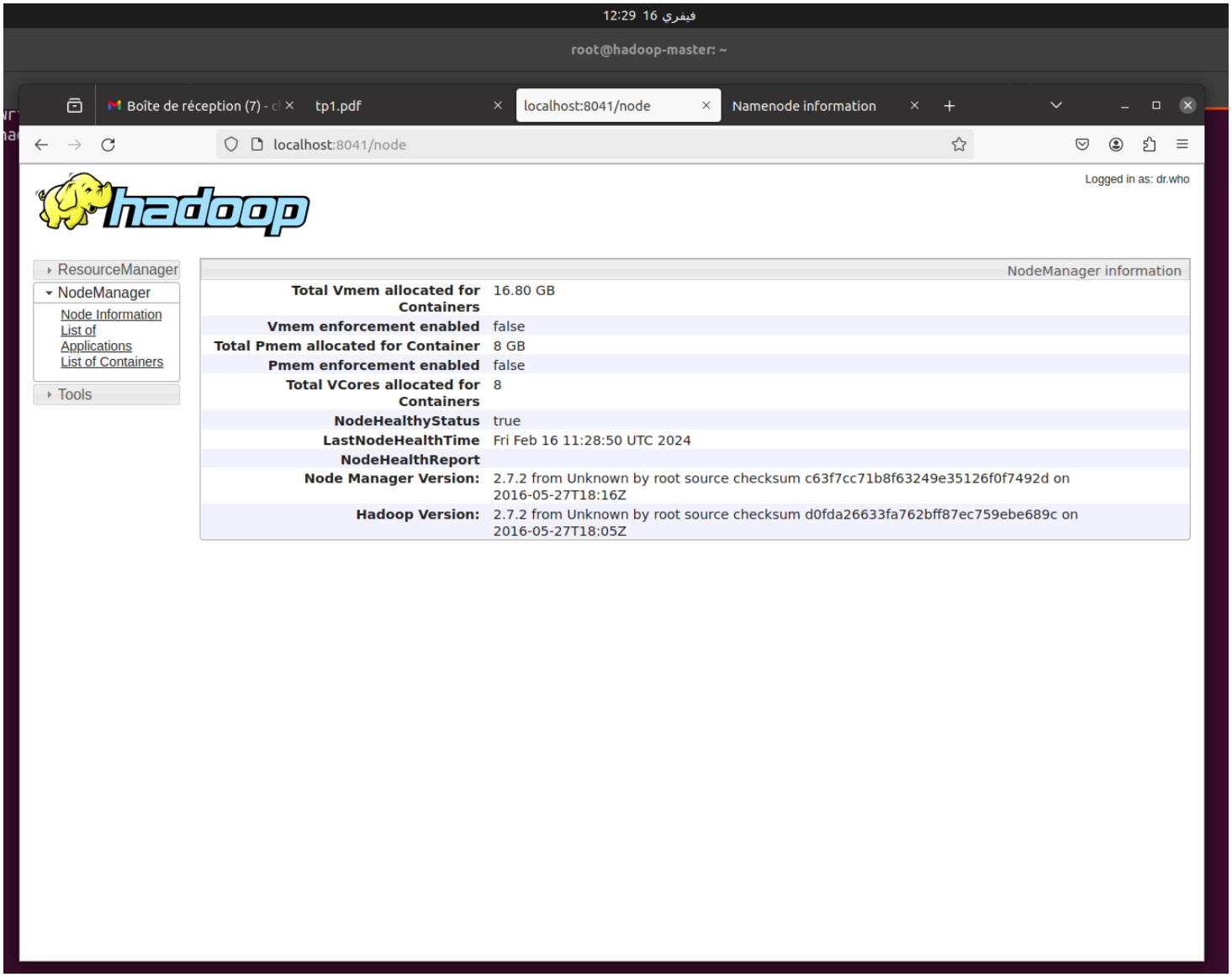

Activities
Terminal
12:28 16 فيفري
root@hadoop-master: ~
root@chaima-Victus: /
root@hadoop-master: ~
Launched reduce tasks=1
Data-local map tasks=3
Total time spent by all maps in occupied slots (ms)=135604
Total time spent by all reduces in occupied slots (ms)=26727
Total time spent by all map tasks (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



7 Monitorer les Jobs Map Reduce

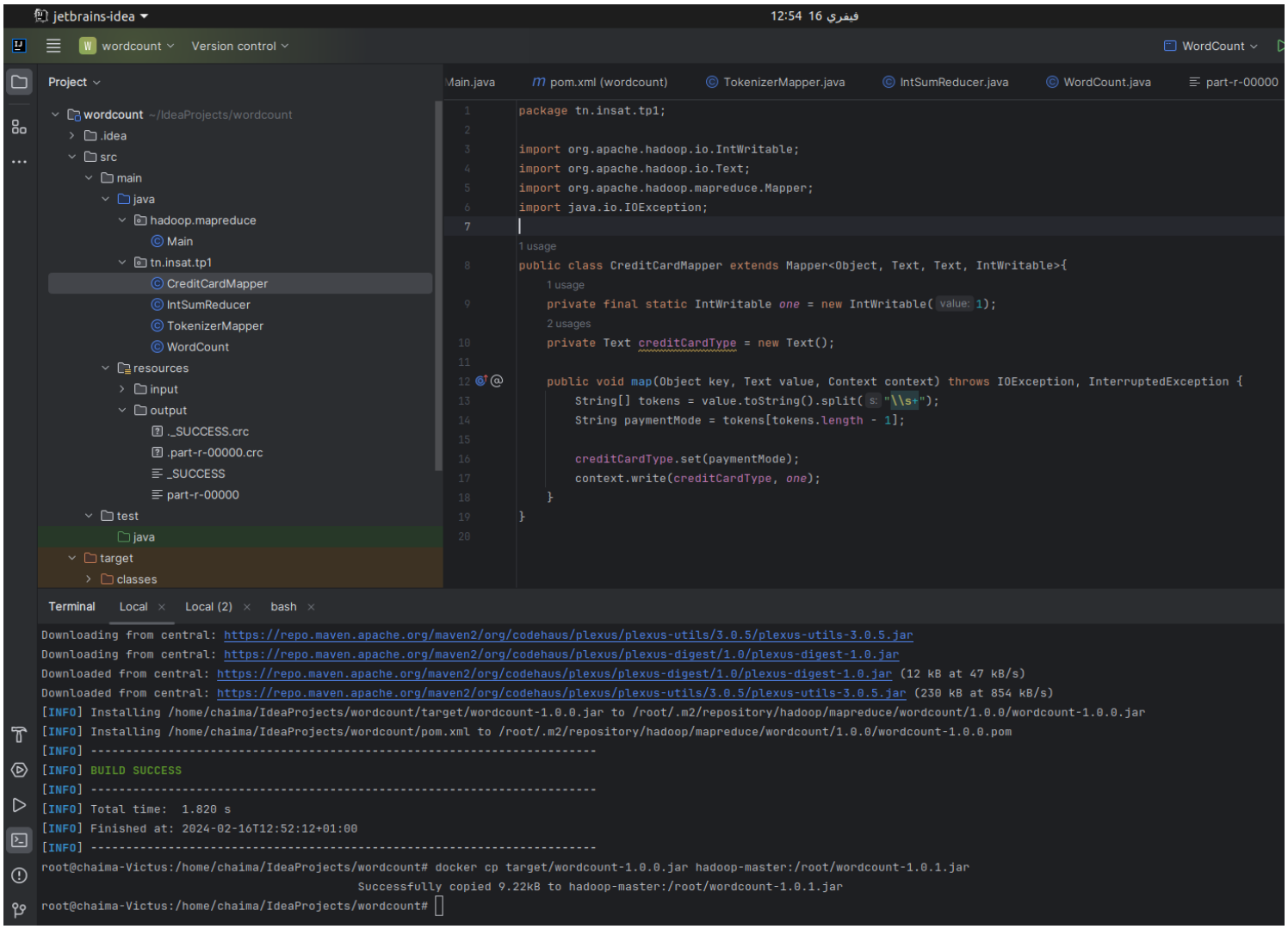



12:29 16 فيفري
root@hadoop-master: ~
Boîte de réception (7) - ch
tp1.pdf
localhost:8041/node
Namenode information
localhost:8041/node
Logged in as: dr.who
hadoop
ResourceManager
NodeManager
Node Information
List of Applications
List of Containers
Tools
NodeManager information
Total Vmem allocated for 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

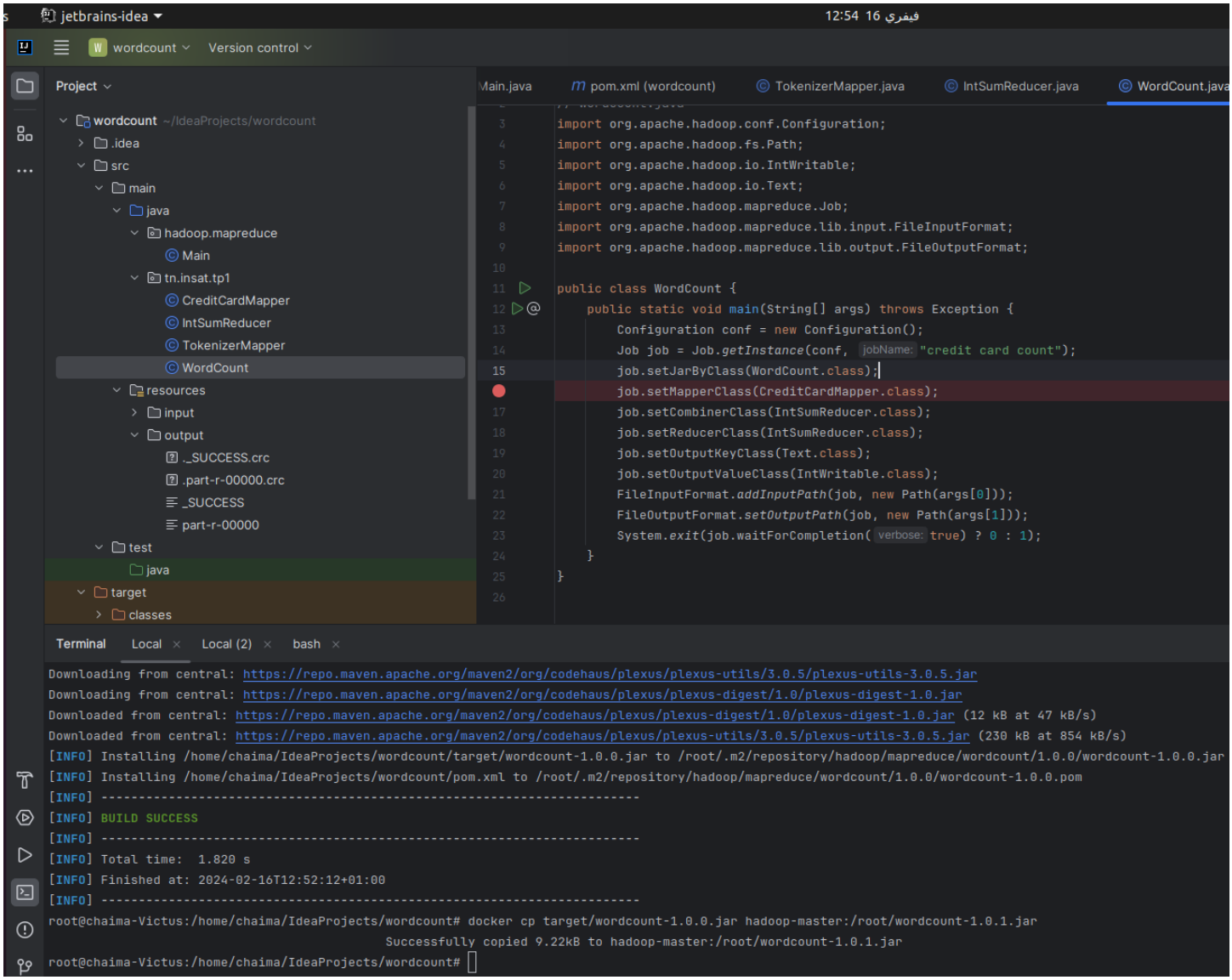
3

Bonus

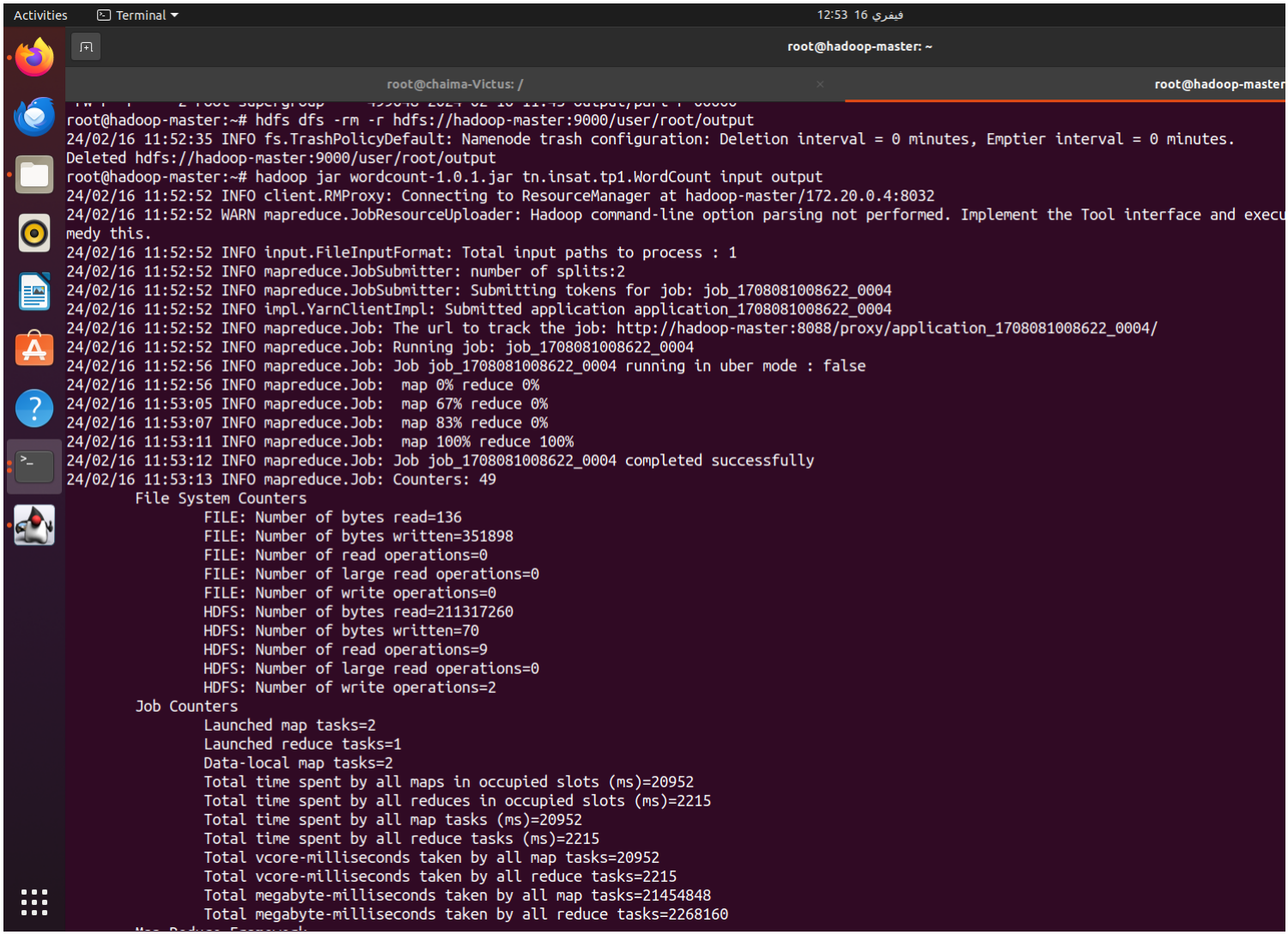
1 Ajout d'une classe "CreditCardMapper"



2 Modification dans la classe main



3 Exécution du nouveau artifact dans le master



4 Affichage du résultat