

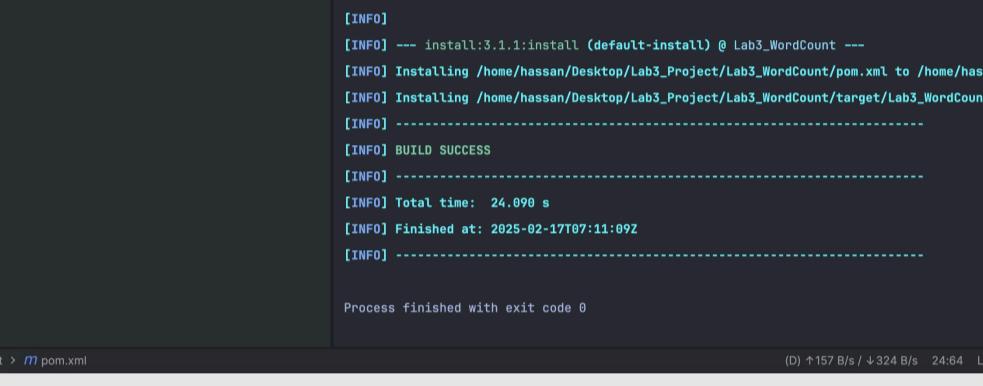
11. *What is the primary purpose of the following statement?*

```
$ java -version  
$ hadoop version  
$jps
```

A screenshot of the Ubuntu 18.04 LTS desktop environment. At the bottom, a dock contains icons for Dash, Home, Applications, and the Dash search bar. The main workspace shows a terminal window with the command 'ssh hassan@172.16.186.132' and its output '— 120x30'. A system status bar at the top displays icons for signal strength, battery level (100%), and system status.

7772 Main
hassan@has

Task 4: Verify the Code JAR build



```
14 JobConf conf = new JobConf(Run > Dependencies
15 conf.setJobName("WordCount");
```

Run **Lab3_WordCount [Install]** x

Lab3_WordCount [Install]: At 2/17/25, 7:11 AM

```
[INFO] --- jar:3.3.0:jar (default-jar) @ Lab3_WordCount ---
[INFO] Building jar: /home/hassan/Desktop/Lab3_Project/Lab3_WordCount/target/Lab3_WordCount-1.0-SNAPSHOT.jar
[INFO]
[INFO] --- install:3.1.1:install (default-install) @ Lab3_WordCount ---
[INFO] Installing /home/hassan/Desktop/Lab3_Project/pom.xml to /home/hassan/.m2/repository/com/hassan/Lab3_WordCount/1.0-SNAPSHOT/Lab3_WordCount-1.0-SNAPSHOT.jar
[INFO] Installing /home/hassan/Desktop/Lab3_Project/Lab3_WordCount/target/Lab3_WordCount-1.0-SNAPSHOT.jar
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 24.090 s
[INFO] Finished at: 2025-02-17T07:11:09Z
[INFO] -----
```

Process finished with exit code 0

(D) ↑157 B/s ↓324 B/s 24:64 LF UTF-8 4 spaces

Task 5: Creating Input Text File (Duplicated Tech Words For example)

algorithm algorithm cloud cloud computing computing data data structure structure database database networking networking artificial artificial intelligence intelligence machine machine learning learning deep deep learning learning cybersecurity cybersecurity encryption encryption authentication authentication API API microservices microservices DevOps DevOps containerization containerization virtualization virtualization blockchain blockchain cryptocurrency cryptocurrency cloud cloud storage storage SaaS SaaS PaaS PaaS IaaS IaaS edge edge computing computing IoT IoT smart smart devices devices automation automation orchestration orchestration CI/CD CI/CD version version control control repository repository GitHub GitHub GitLab GitLab Kubernetes Kubernetes Docker Docker serverless serverless computing computing big big data data analytics analytics NoSQL NoSQL SQL SQL programming programming debugging debugging software software engineering engineering

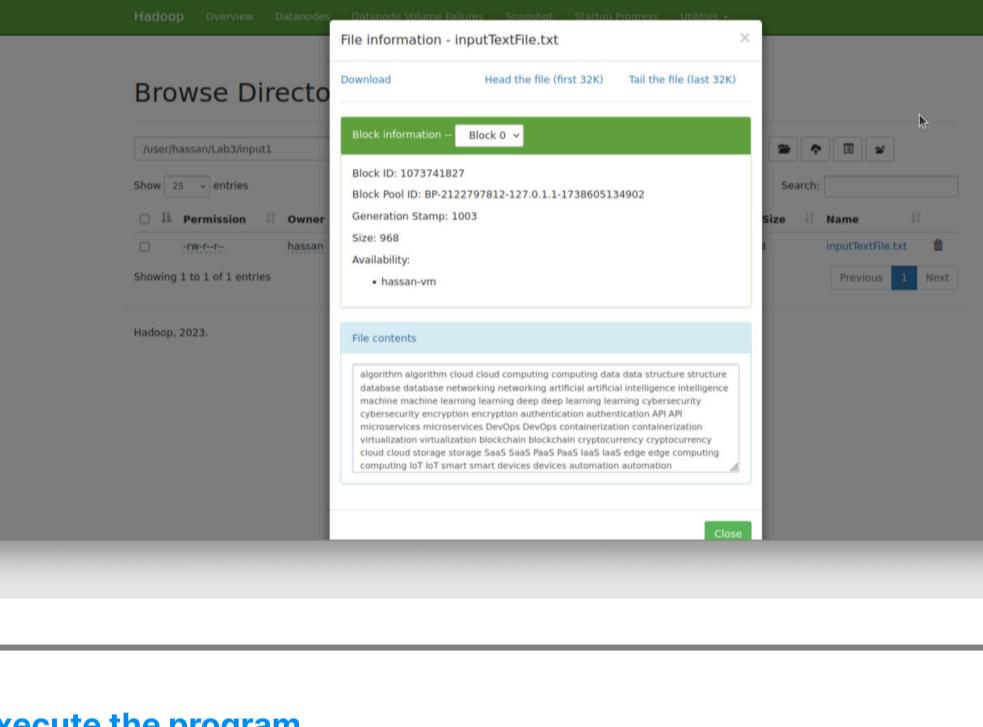
```
dit View Window Help
drn - hassan@hassan-vm: ~/Desktop/Lab3_Project - ssh hassan@172.16.186.132 - 120x30
hassan@hassan-vm:~/Desktop/Lab3_Project$ nano inputTextFile.txt
hassan@hassan-vm:~/Desktop/Lab3_Project$ cat inputTextFile.txt
algorithm algorithm cloud cloud computing computing data data structure structure database database networking networking
artificial artificial intelligence intelligence machine machine learning learning deep deep learning learning cybersecurity
cybersecurity encryption encryption authentication authentication API API microservices microservices DevOps DevOps
containerization containerization virtualization virtualization blockchain blockchain cryptocurrency cryptocurrency
cryptocurrency cloud cloud storage storage SaaS SaaS PaaS PaaS IaaS IaaS edge edge computing computing IoT IoT smart
smart devices devices automation automation orchestration orchestration CI/CD CI/CD version version control control
repository repository GitHub GitHub GitLab GitLab Kubernetes Kubernetes Docker Docker serverless serverless computing
computing big big data data analytics analytics NoSQL NoSQL SQL SQL programming programming debugging debugging
software software engineering engineering
hassan@hassan-vm:~/Desktop/Lab3_Project$
```

Task 6: Cre

```
$ hadoop fs -mkdir -p /user/hassan/Lab3/input
$ hadoop fs -ls /user/hassan/
$ hadoop fs -put ~/Desktop/Lab3_Project/inputTextFile.txt /user/hassan/Lab3/input/
$ hadoop fs -ls /user/hassan/Lab3/input/
$ hadoop fs -cat /user/hassan/Lab3/input/inputTextFile.txt

Terminal
dit View Window Help drn — hassan@hassan-vm: ~ — ssh hassan@172.16.186.132 — 120x30
[hassan@hassan-vm:~$ hadoop fs -mkdir -p /user/hassan/Lab3/input
[hassan@hassan-vm:~$ hadoop fs -ls /user/hassan/
Found 2 items
drwxr-xr-x  - hassan supergroup          0 2025-02-04 08:45 /user/hassan/Folder1
drwxr-xr-x  - hassan supergroup          0 2025-02-17 07:22 /user/hassan/Lab3
[hassan@hassan-vm:~$ hadoop fs -put ~/Desktop/Lab3_Project/inputTextFile.txt /user/hassan/Lab3/input/
[hassan@hassan-vm:~$ hadoop fs -ls /user/hassan/Lab3/input/
Found 1 items
-rw-r--r--  1 hassan supergroup      968 2025-02-17 07:27 /user/hassan/Lab3/input/inputTextFile.txt
[hassan@hassan-vm:~$ hadoop fs -cat /user/hassan/Lab3/input/inputTextFile.txt
algorithm algorithm cloud cloud computing computing data data structure structure database database networking networking
artificial artificial intelligence intelligence machine machine learning learning deep deep learning learning cybersecurity
cybersecurity encryption encryption authentication authentication API API microservices microservices DevOps DevOps
containerization containerization virtualization virtualization blockchain blockchain cryptocurrency cryptocurrency
cloud cloud storage storage SaaS SaaS PaaS PaaS IaaS IaaS edge edge computing computing IoT IoT smart smart devices
devices automation automation orchestration orchestration CI/CD CI/CD version version control control repository repository
GitHub GitHub GitLab GitLab Kubernetes Kubernetes Docker Docker serverless serverless computing computing big big data
data analytics analytics NoSQL NoSQL SQL SQL programming programming debugging debugging software software engineering
engineering
hassan@hassan-vm:~$ 
```


HDFS GUI



File information - inputTextFile.txt

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

Block information - Block 0

Block ID: 1073741827
Block Pool ID: BP-2122797812-127.0.1.1-1738605134902
Generation Stamp: 1003
Size: 968
Availability:
• hassan-vm

File contents

algorithm algorithm cloud cloud computing computing data data structure structure database database networking networking
artificial artificial intelligence intelligence machine machine learning learning deep deep learning learning cybersecurity
cybersecurity encryption encryption authentication authentication API API microservices microservices DevOps DevOps
containerization containerization virtualization virtualization blockchain blockchain cryptocurrency cryptocurrency
cloud cloud storage storage SaaS SaaS PaaS PaaS IaaS IaaS edge edge computing computing IoT IoT smart smart devices
devices automation automation orchestration orchestration CI/CD CI/CD version version control control repository repository
GitHub GitHub GitLab GitLab Kubernetes Kubernetes Docker Docker serverless serverless computing computing big big data
data analytics analytics NoSQL NoSQL SQL SQL programming programming debugging debugging software software engineering
engineering

Close

Task 7: Execute the program

```
$ hadoop jar ~/Desktop/Lab3_Project/Lab3_WordCount-1.0-SNAPSHOT.jar edu.hassan.RunnerClass
/user/hassan/Lab3/input/inputTextFile.txt /user/hassan/Lab3/output/
```

(1)

```
Edit View Window Help drn — hassan@hassan-vm: ~ — ssh hassan@172.16.186.132 — 117x28
[hassan@hassan-vm:~$ hadoop jar ~/Desktop/Lab3_Project/Lab3_WordCount-1.0-SNAPSHOT.jar edu.hassan.RunnerClass /user/hassan/Lab3/input/inputTextFile.txt /user/hassan/Lab3/output/
2025-02-17 07:41:58,761 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8080
2
2025-02-17 07:41:59,035 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8080
2
2025-02-17 07:42:00,466 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-02-17 07:42:00,499 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hassan/.staging/job_1739764183578_0001
2025-02-17 07:42:02,129 INFO mapred.FileInputFormat: Total input files to process : 1
2025-02-17 07:42:02,141 INFO mapred.FileInputFormat: Input file: /user/hassan/Lab3/input/inputTextFile.txt
```

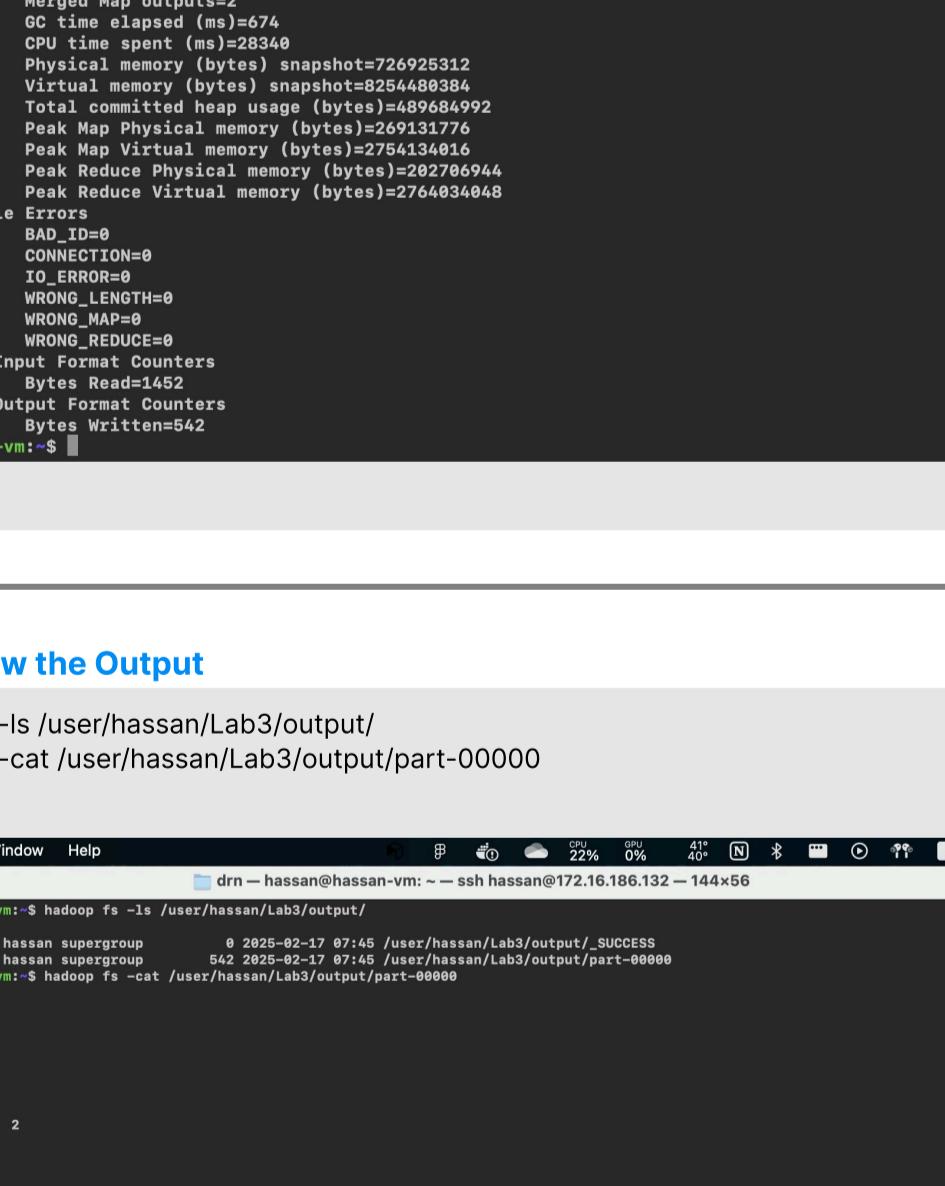
```
Total megabyte-milliseconds taken by all reduce tasks=9018368
Map-Reduce Framework
  Map input records=1
  Map output records=198
  Map output bytes=1499
```

```
Input split bytes=228
Combine input records=108
Combine output records=49
Reduce input groups=49
Reduce shuffle bytes=750
Reduce input records=49
Reduce output records=49
Spilled Records=98
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=674
CPU time spent (ms)=28340
Physical memory (bytes) snapshot=726925312
Virtual memory (bytes) snapshot=8254480384
Total committed heap usage (bytes)=489684992
Peak Map Physical memory (bytes)=269131776
Peak Map Virtual memory (bytes)=2754134016
Peak Reduce Physical memory (bytes)=202706944
Peak Reduce Virtual memory (bytes)=2764034048
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=1452
File Output Format Counters
    Bytes Written=542
hassan@hassan-vm:~$
```

Task 8: Show the Output

```
$ hadoop fs -ls /user/hassan/Lab3/output/
$ hadoop fs -cat /user/hassan/Lab3/output/part-00000
```

Terminal



```
drn — hassan@hassan-vm: ~ — ssh hassan@172.16.186.132 — 144x56
[hassan@hassan-vm:~$ hadoop fs -ls /user/hassan/Lab3/output/
Found 2 items
-rw-r--r-- 1 hassan supergroup 0 2025-02-17 07:45 /user/hassan/Lab3/output/_SUCCESS
-rw-r--r-- 1 hassan supergroup 542 2025-02-17 07:45 /user/hassan/Lab3/output/part-00000
[hassan@hassan-vm:~$ hadoop fs -cat /user/hassan/Lab3/output/part-00000
API 2
CI/CD 2
DevOps 2
Docker 2
GitHub 2
GitLab 2
IaaS 2
IoT 2
Kubernetes 2
NoSQL 2
PaaS 2
SQL 2
SaaS 2
algorithm 2
analytics 2
artificial 2
authentication 2
automation 2
big 2
blockchain 2
cloud 4
computing 6
containerization 2
control 2
cryptocurrency 2
```

```
    security 2
    data 4
    database 2
    debugging 2
    deep 2
    devices 2
    edge 2
    encryption 2
    engineering 2
    intelligence 2
```

```
learning      4
machine 2
microservices 2
networking 2
orchestration 2
programming 2
repository 2
serverless 2
smart 2
software 2
storage 2
structure 2
version 2
virtualization 2
hassan@hassan-vm:~$
```

HDFS GUI

Explanation of the Implementation

- This lab discuss about the Word Count in a text file by using the MapReduce mechanism and the Hadoop Distributed File System.
- First, I created a Java program that consists of three classes: one for Mapping, one for Reducing, and one for Running the job of word counting.
- Each class implements a function for the MapReduce mechanism.
- After that, I built the program into a JAR file and created an input text file containing many duplicated tech related words (just for example).
- I then uploaded this input text file to the Hadoop File System on a new directory “/Lab3/input” inside an old directory I created before in Lab 2.

- And then I ran the `count` program which
- Hadoop distributed the word counts along with its duplicate times in the

- About the system I'm using:
 - Macbook Air M1

- VMware Fusion “Virtual Machine Hypervisor”
 - Ubuntu for Servers 22.04 “With ubuntu-desktop package”
 - Hadoop version 3.3.6
 - JDK version 11
 - IntelliJ IDEA with SSH feature “for remote programming”

- https://github.com/drni/DS_Lab3