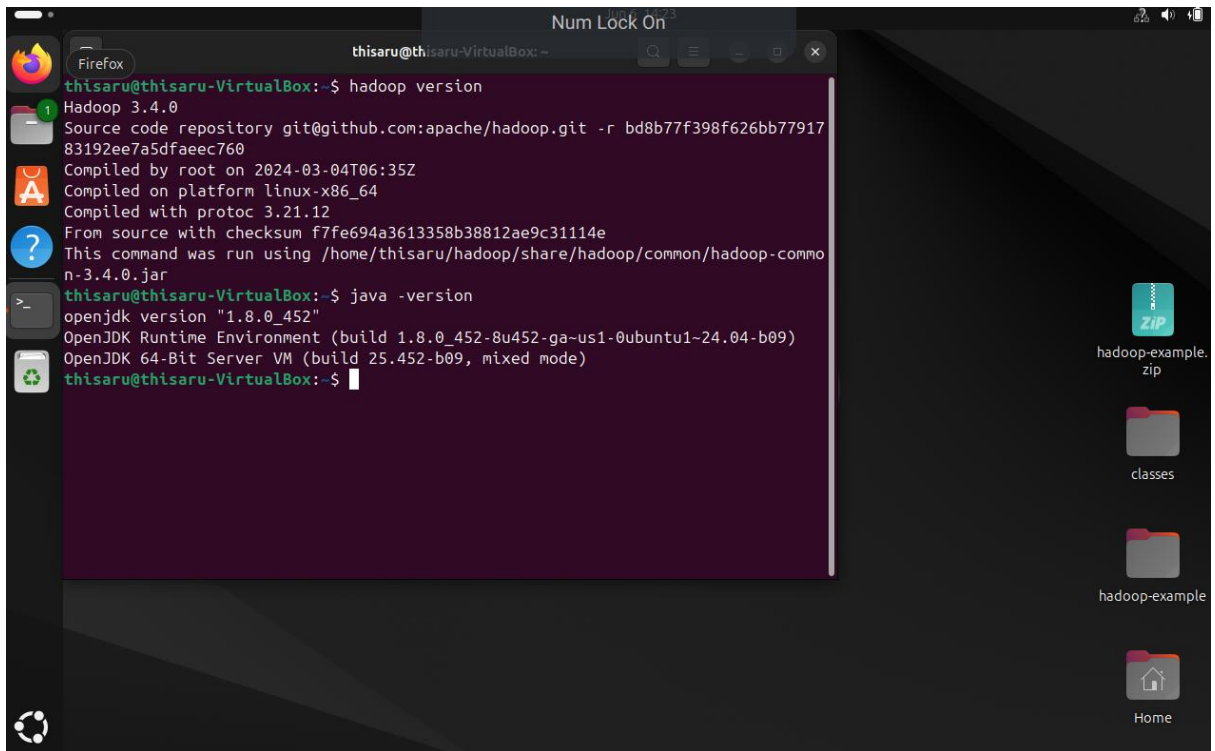


Screenshots For Evidence in Uber Pickup Analysis using Hadoop MapReduce

1 Installation Hadoop and Java Versions



The screenshot shows a terminal window in a virtual machine named 'thisaru@thisaru-VirtualBox'. The user has run the command 'hadoop version', which displays the following information:

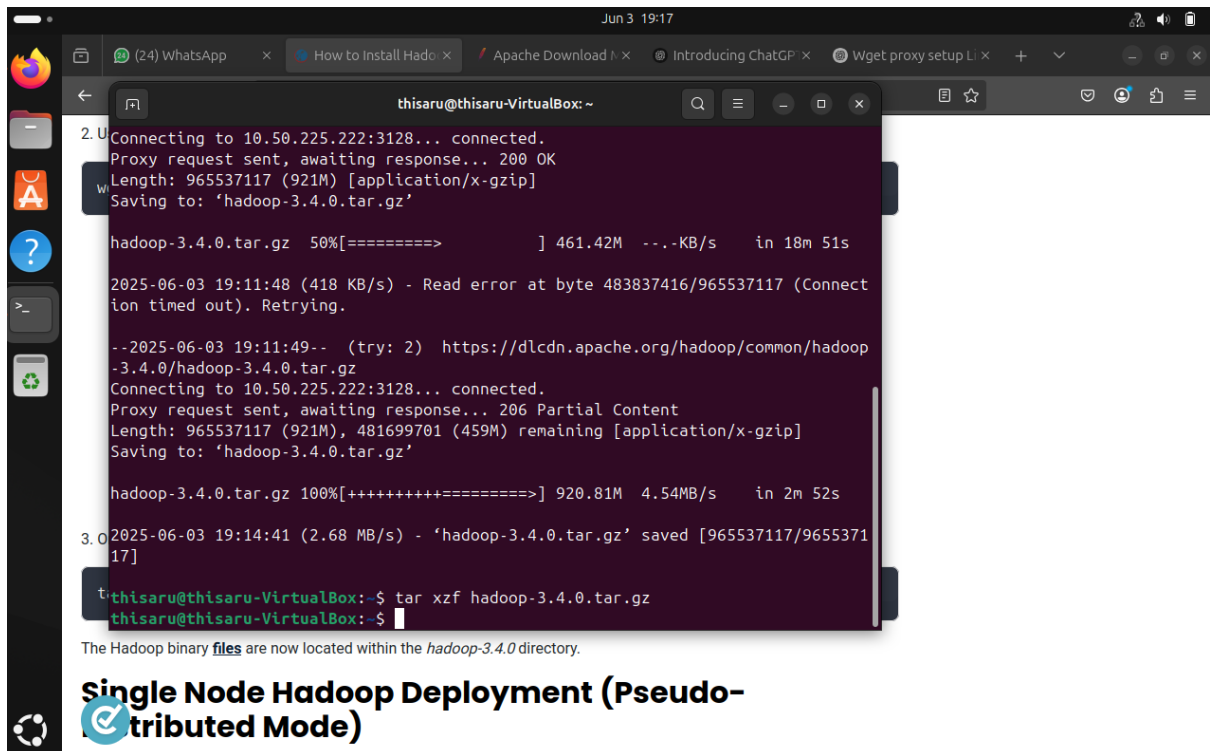
```
thi@thisaru-VirtualBox:~$ hadoop version
Hadoop 3.4.0
Source code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb7791783192ee7a5dfaee760
Compiled by root on 2024-03-04T06:35Z
Compiled on platform linux-x86_64
Compiled with protoc 3.21.12
From source with checksum f7fe694a3613358b38812ae9c31114e
This command was run using /home/thisaru/hadoop/share/hadoop/common/hadoop-common-3.4.0.jar

thi@thisaru-VirtualBox:~$ java -version
openjdk version "1.8.0_452"
OpenJDK Runtime Environment (build 1.8.0_452-8u452-ga-us1-0ubuntu1-24.04-b09)
OpenJDK 64-Bit Server VM (build 25.452-b09, mixed mode)

thi@thisaru-VirtualBox:~$
```

The terminal window is open over a desktop environment with icons for 'hadoop-example.zip', 'classes', 'hadoop-example', and 'Home'.

2 Hadoop Installing



The screenshot shows a terminal window in a virtual machine named 'thisaru@thisaru-VirtualBox'. The user is installing Hadoop 3.4.0 using a proxy. The terminal output is as follows:

```
2. U Connecting to 10.50.225.222:3128... connected.
Proxy request sent, awaiting response... 200 OK
Length: 965537117 (921M) [application/x-gzip]
Saving to: 'hadoop-3.4.0.tar.gz'

hadoop-3.4.0.tar.gz  50%[=====>] 461.42M  ---KB/s   in 18m 51s

2025-06-03 19:11:48 (418 KB/s) - Read error at byte 483837416/965537117 (Connection timed out). Retrying.

--2025-06-03 19:11:49-- (try: 2) https://dldn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz
Connecting to 10.50.225.222:3128... connected.
Proxy request sent, awaiting response... 206 Partial Content
Length: 965537117 (921M), 481699701 (459M) remaining [application/x-gzip]
Saving to: 'hadoop-3.4.0.tar.gz'

hadoop-3.4.0.tar.gz 100%[+++++++>] 920.81M  4.54MB/s   in 2m 52s

3. 2025-06-03 19:14:41 (2.68 MB/s) - 'hadoop-3.4.0.tar.gz' saved [965537117/965537117]

thi@thisaru-VirtualBox:~$ tar xzf hadoop-3.4.0.tar.gz
thi@thisaru-VirtualBox:~$
```

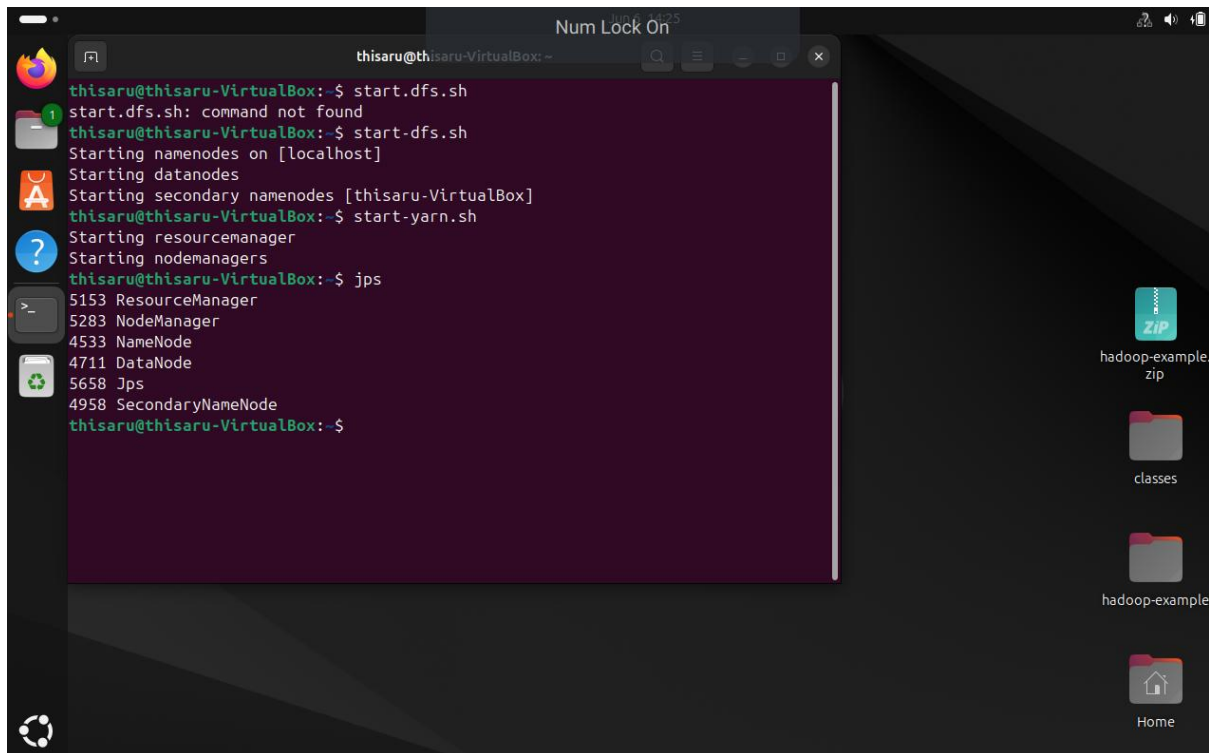
The terminal window is open over a desktop environment with icons for 'hadoop-example.zip', 'classes', 'hadoop-example', and 'Home'.

The Hadoop binary files are now located within the `hadoop-3.4.0` directory.

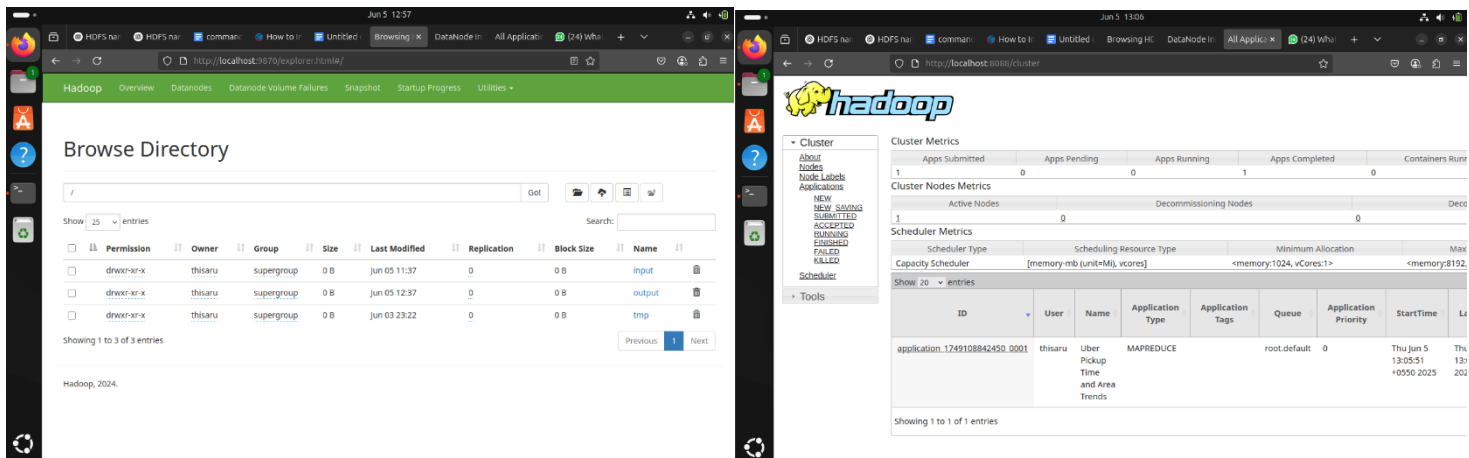
Single Node Hadoop Deployment (Pseudo-Distributed Mode)

Screenshots For Evidence in Uber Pickup Analysis using Hadoop MapReduce

3 Hadoop Cluster start and verify

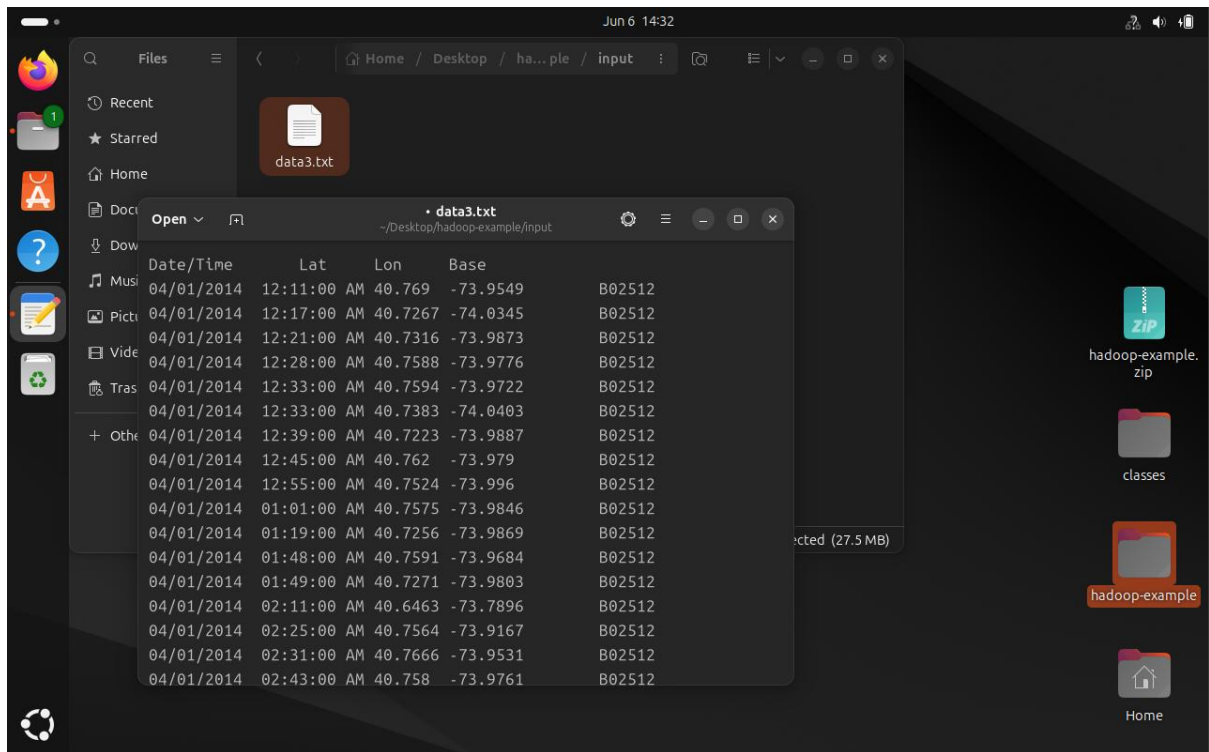


4 Hadoop web UI opened in browser

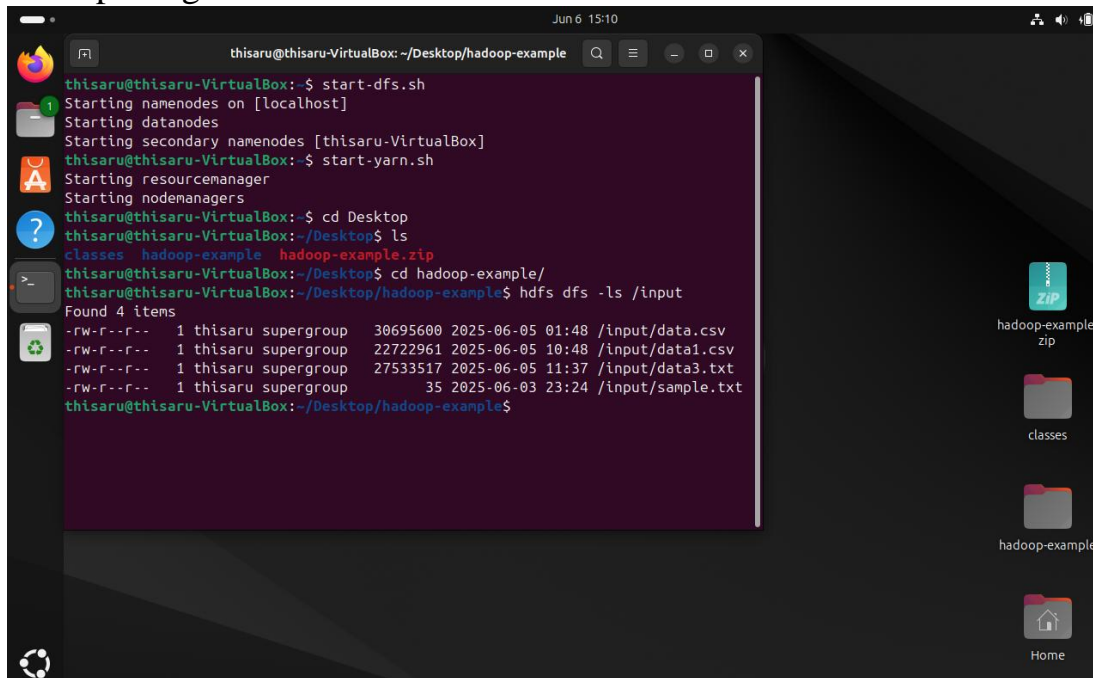


Screenshots For Evidence in Uber Pickup Analysis using Hadoop MapReduce

5 Data Set

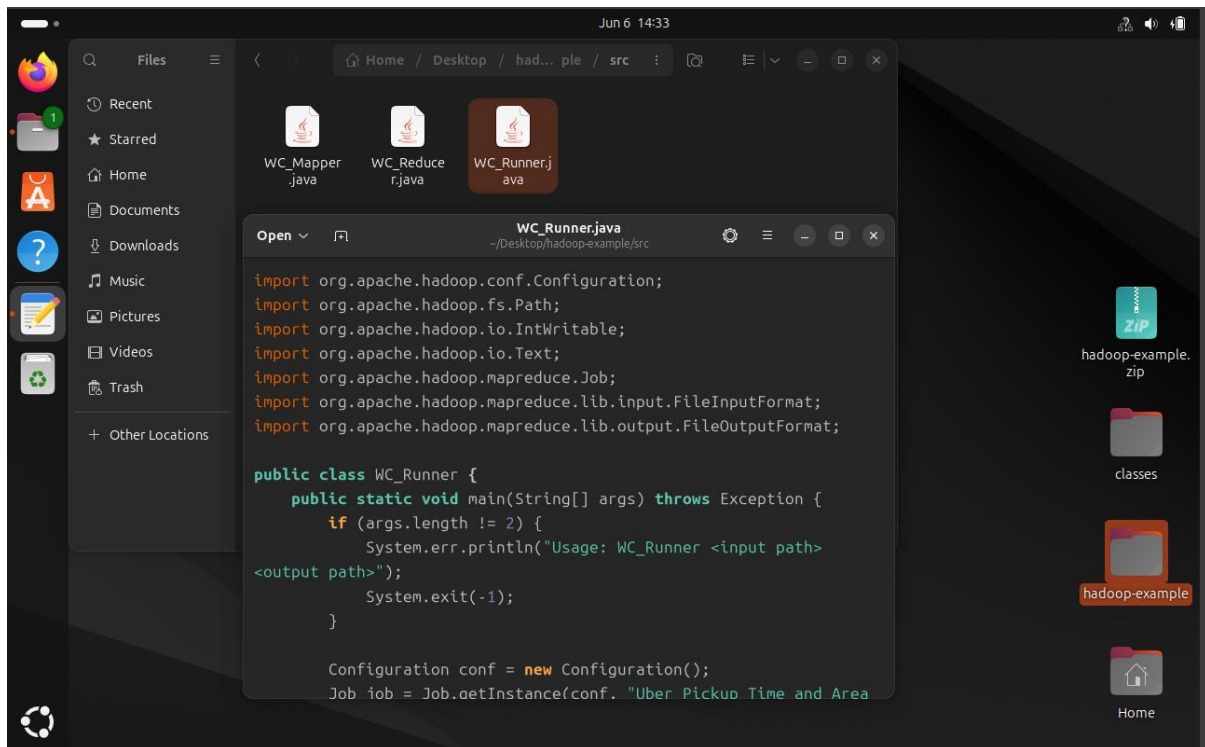


6 Input logs

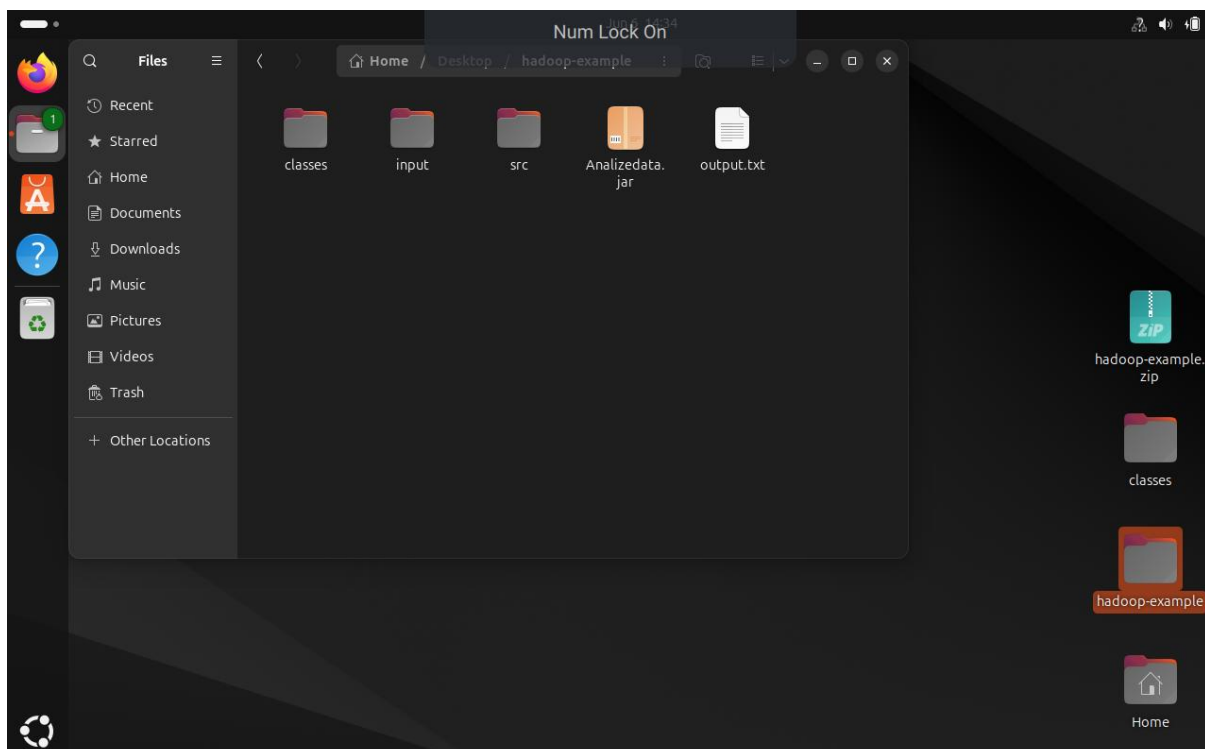


Screenshots For Evidence in Uber Pickup Analysis using Hadoop MapReduce

7 Code Implementation

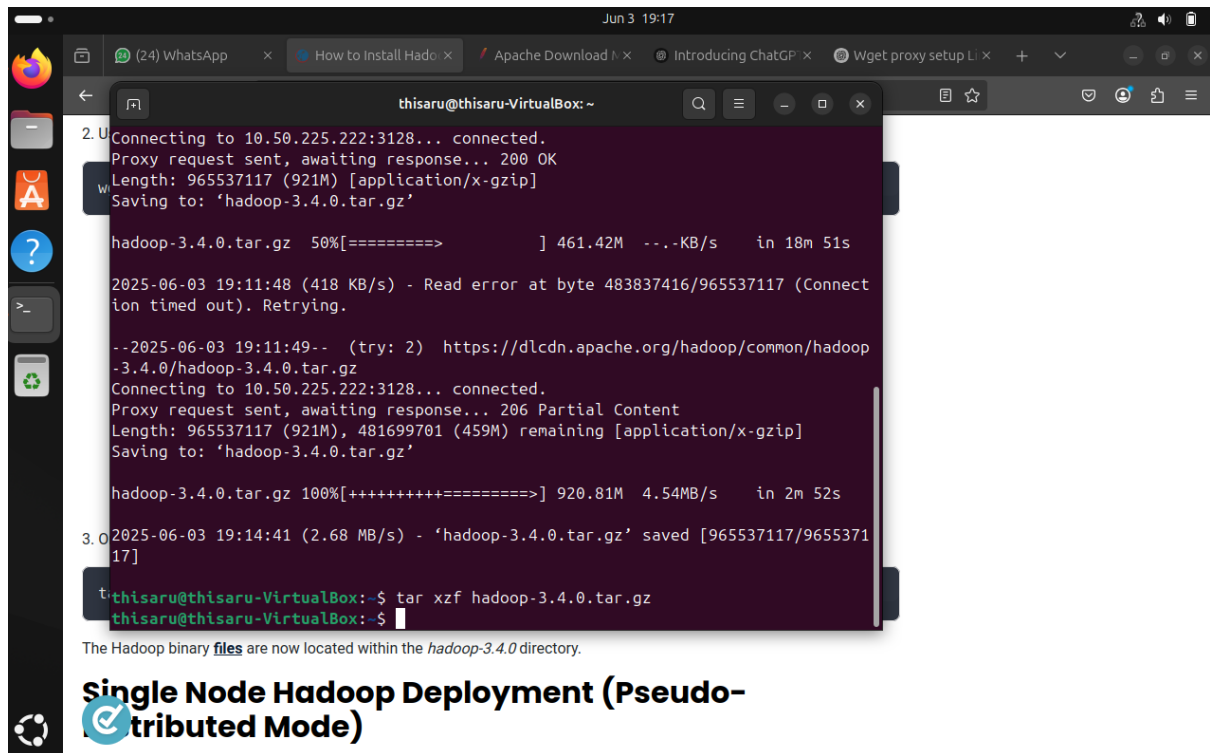


8 JAR Creation

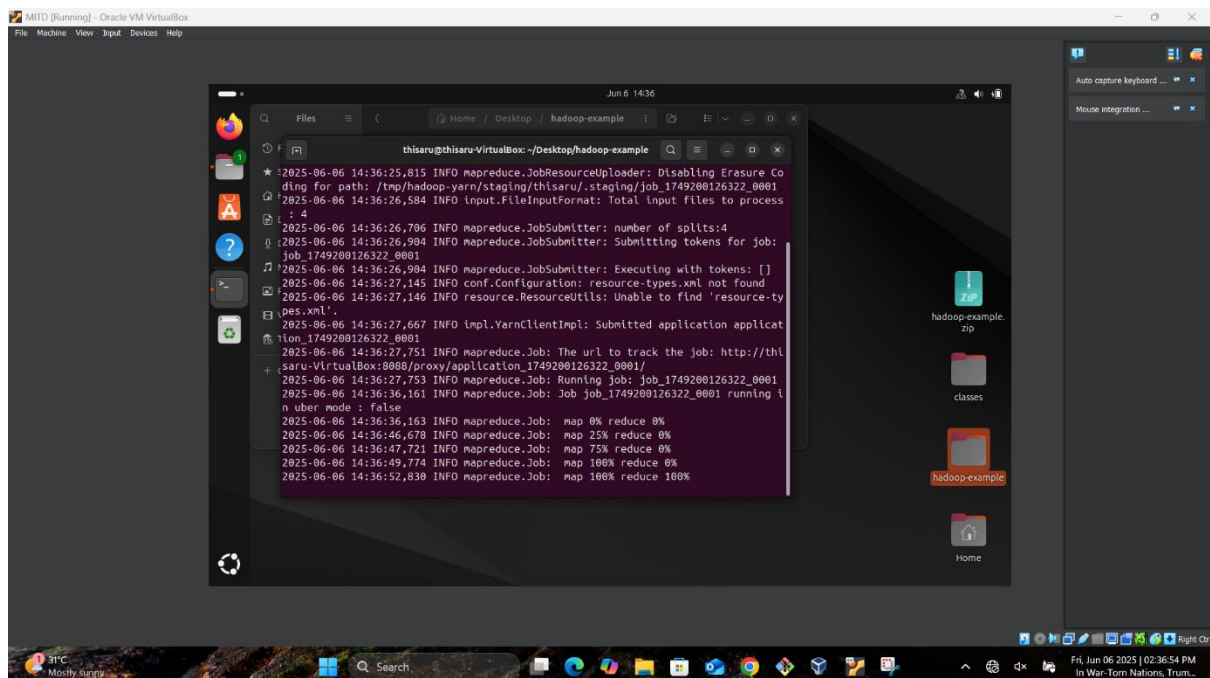


Screenshots For Evidence in Uber Pickup Analysis using Hadoop MapReduce

9 Running the Hadoop Job

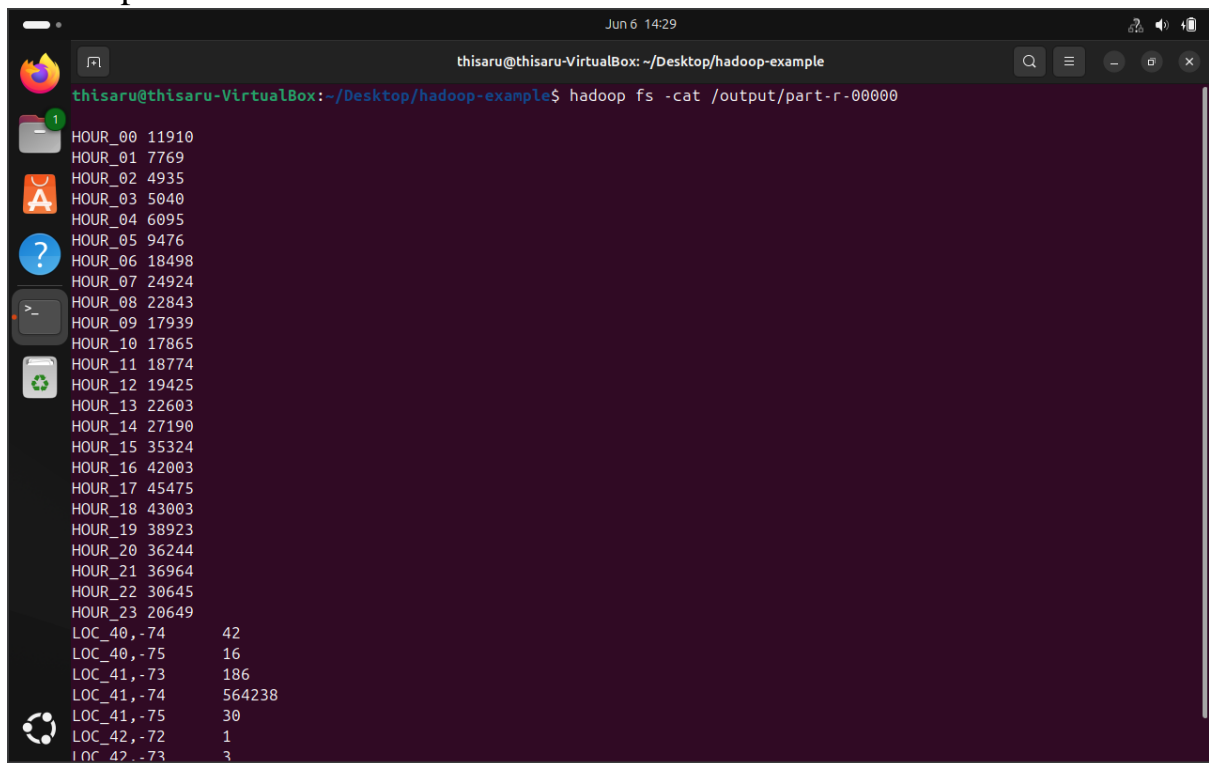


10 Hadoop Job Complete



Screenshots For Evidence in Uber Pickup Analysis using Hadoop MapReduce

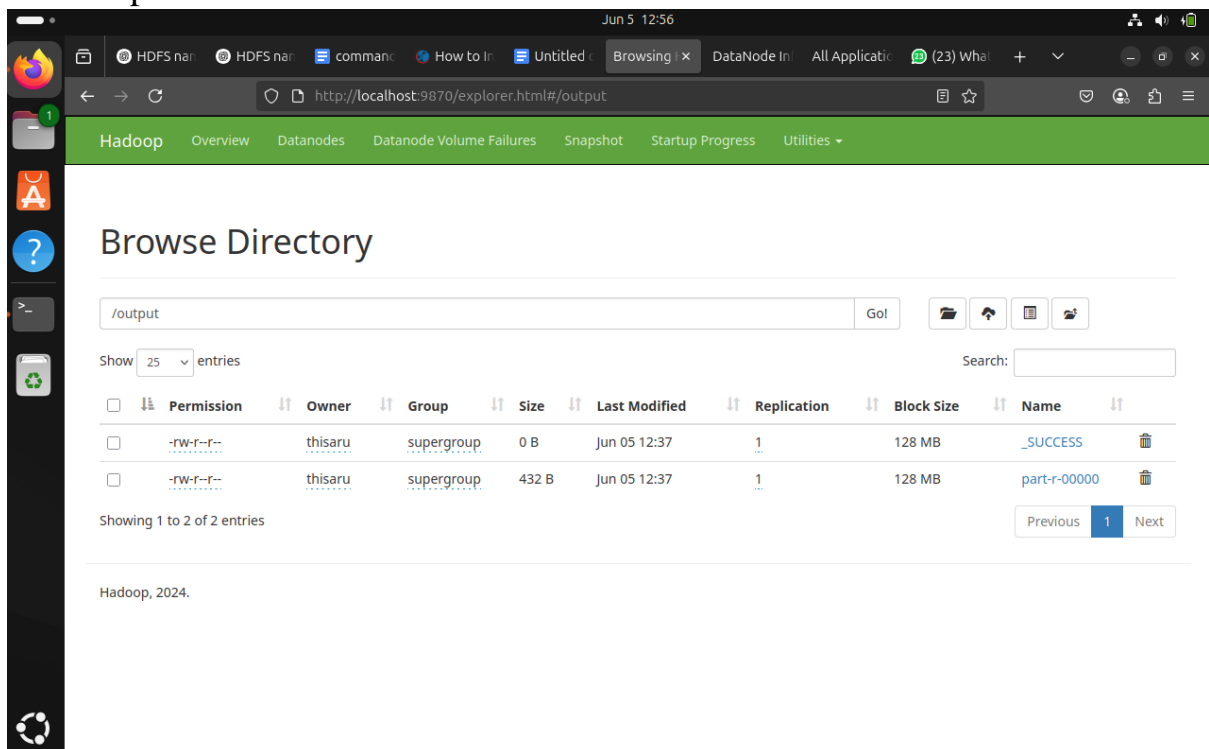
11 Output



A terminal window titled 'thisaru@thisaru-VirtualBox: ~/Desktop/hadoop-example' showing the command 'hadoop fs -cat /output/part-r-00000'. The output lists hourly pickup counts (HOUR_00 to HOUR_23) and location counts (LOC_40 to LOC_47).

```
thisaru@thisaru-VirtualBox: ~/Desktop/hadoop-example$ hadoop fs -cat /output/part-r-00000
HOUR_00 11910
HOUR_01 7769
HOUR_02 4935
HOUR_03 5040
HOUR_04 6095
HOUR_05 9476
HOUR_06 18498
HOUR_07 24924
HOUR_08 22843
HOUR_09 17939
HOUR_10 17865
HOUR_11 18774
HOUR_12 19425
HOUR_13 22603
HOUR_14 27190
HOUR_15 35324
HOUR_16 42003
HOUR_17 45475
HOUR_18 43003
HOUR_19 38923
HOUR_20 36244
HOUR_21 36964
HOUR_22 30645
HOUR_23 20649
LOC_40,-74 42
LOC_40,-75 16
LOC_41,-73 186
LOC_41,-74 564238
LOC_41,-75 30
LOC_42,-72 1
LOC_42,-73 3
```

12 Output create



A screenshot of the Hadoop web interface (http://localhost:9870/explorer.html#/output) showing the 'Browse Directory' view for the '/output' path. The interface displays a table of files with columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, and Name. Two files are listed: '_SUCCESS' and 'part-r-00000'.

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	thisaru	supergroup	0 B	Jun 05 12:37	1	128 MB	_SUCCESS
-rw-r--r--	thisaru	supergroup	432 B	Jun 05 12:37	1	128 MB	part-r-00000

Showing 1 to 2 of 2 entries

Hadoop, 2024.