

Term Work 5

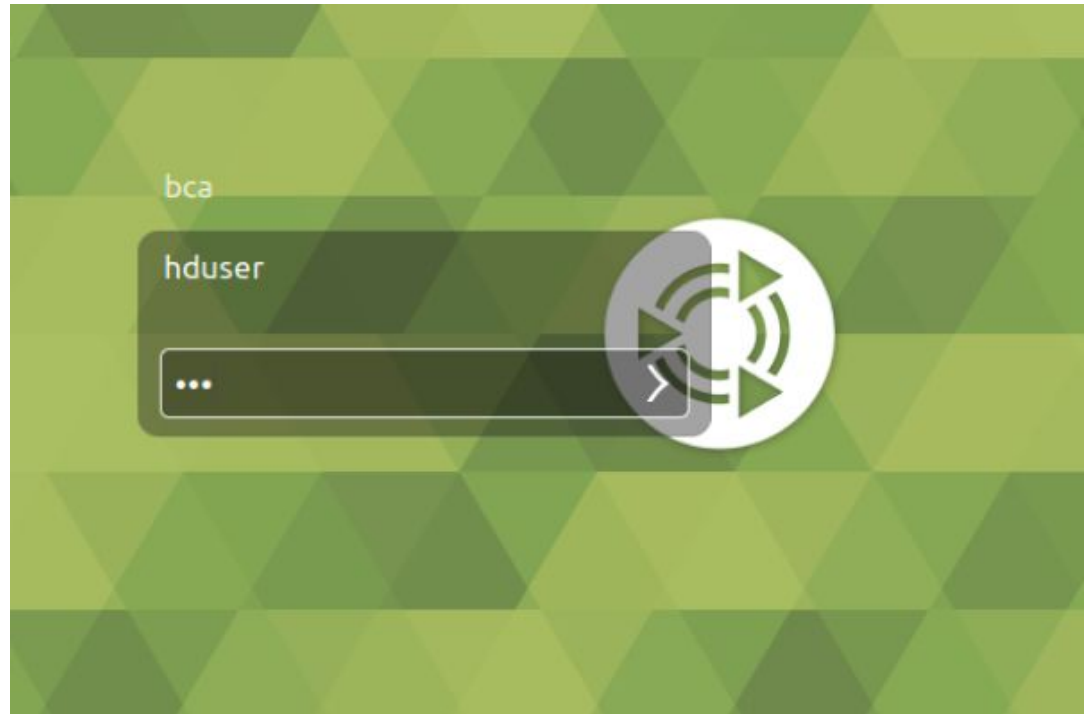


Password

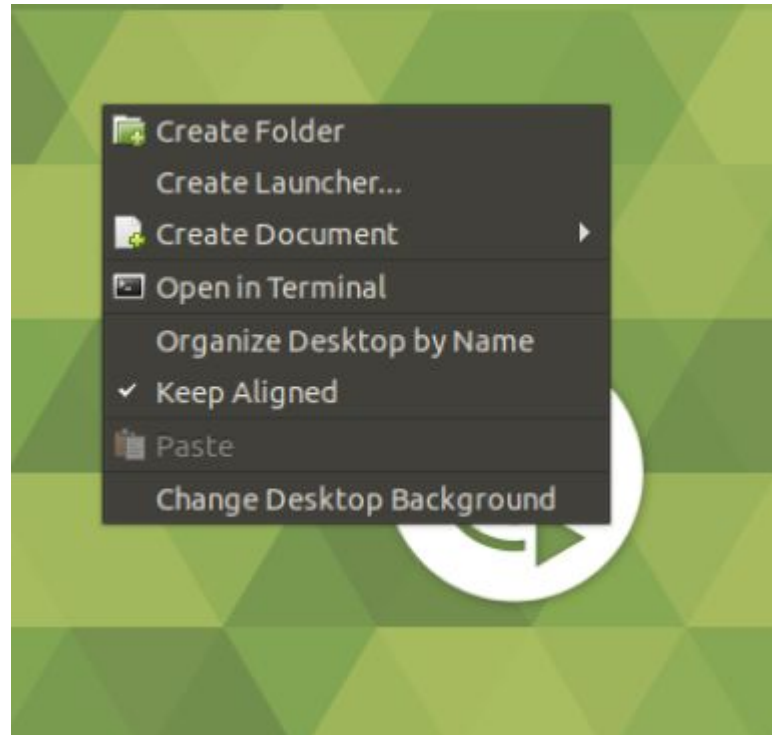
dv

or

bca



Right Click & select
Open in Terminal



Type jps &



```
● hduser@ubuntu: ~  
File Edit View Search Terminal Help  
hduser@ubuntu:~$ jps  
2472 Jps  
hduser@ubuntu:~$ █
```

Type & press



Blue Colour - Directory

Password bca or dv

```
File Edit View Search Terminal Help
hduser@ubuntu:~$ cd Documents
hduser@ubuntu:~/Documents$ sudo mkdir termwork5
[sudo] password for hduser:
hduser@ubuntu:~/Documents$ ls
bgm20          map90.txt      map.py         -reducer       shop90.txt     termwork4      words.txt
bgm27          -mapper       -output        reducer45.py   termwork1     termwork5      word.txt
flipkart.txt  mapper45.py   red90.py      reducer.py     termwork2     word13.txt
map90.py      mapper.py     red.py        shop1.txt     termwork3     word23.txt
hduser@ubuntu:~/Documents$
```

Type `cd termwork5`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents$ cd termwork5
hduser@ubuntu:~/Documents/termwork5$
```

Type `sudo nano salary.csv`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ sudo nano salary.csv
```

Type the text
& press Ctrl x y & then press



```
File Edit View Search Terminal Help
GNU nano 2.9.3 salary.csv

E001,Sunita,Accounts,15000
E002,Harsh,IT,50000
E003,Ragini,IT,75000
E004,Mithun,Accounts,20000
E005,Pruthavi,Marketing,45000
E006,Anjali,IT,70000
E007,Kunal,Marketing,60000
E008,Mitali,Accounts,55000
E009,Roopa,IT,70000
E010,Deepthi,Accounts,30000
E011,Janavi,Marketing,25000
E012,Lata,Accounts,30000
E013,Brijmohan,Marketing,45000
E014,Nina,Accounts,50000
E015,Pallavi,Marketing,25000
```


Type `cat salary.csv`
& press



This shows your typed text

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ cat salary.csv
E001,Sunita,Accounts,15000
E002,Harsh,IT,50000
E003,Ragini,IT,75000
E004,Mithun,Accounts,20000
E005,Pruthavi,Marketing,45000
E006,Anjali,IT,70000
E007,Kunal,Marketing,60000
E008,Mitali,Accounts,55000
E009,Roopa,IT,70000
E010,Deepti,Accounts,30000
E011,Janavi,Marketing,25000
E012,Lata,Accounts,30000
E013,Brijmohan,Marketing,45000
E014,Nina,Accounts,50000
E015,Pallavi,Marketing,25000
hduser@ubuntu:~/Documents/termwork5$
```

Type `ls`
& press



This shows your file

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ ls
salary.csv
hduser@ubuntu:~/Documents/termwork5$
```

Type `sudo nano mapper.py`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ sudo nano mapper.py
```

Type the code
& press Ctrl x y & then press



```
File Edit View Search Terminal Help
GNU nano 2.9.3 mapper.py

#!/usr/bin/env python
import sys
# input comes from STDIN (standard input)
for line in sys.stdin:
    line = line.strip()
    line = line.split(",")
    if len(line) >= 2:
        dept = line[2]
        sal = line[3]
        print '%s\t%s' % (dept, sal)
```

Type `cat mapper.py`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ cat mapper.py
#!/usr/bin/env python
import sys
# input comes from STDIN (standard input)
for line in sys.stdin:
    line = line.strip()
    line = line.split(",")
    if len(line) >= 2:
        dept = line[2]
        sal = line[3]
        print '%s\t%s' % (dept, sal)
hduser@ubuntu:~/Documents/termwork5$
```

Type `ls`
& press



This shows your files

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ ls
mapper.py  salary.csv
hduser@ubuntu:~/Documents/termwork5$
```

Type `cat salary.csv | python mapper.py`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ cat salary.csv | python mapper.py
Accounts      15000
IT      50000
IT      75000
Accounts      20000
Marketing     45000
IT      70000
Marketing     60000
Accounts      55000
IT      70000
Accounts      30000
Marketing     25000
Accounts      30000
Marketing     45000
Accounts      50000
Marketing     25000
hduser@ubuntu:~/Documents/termwork5$
```

Type `sudo nano reducer.py`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ sudo nano reducer.py
```


Type the code
& press Ctrl x y & then press



```
File Edit View Search Terminal Help
GNU nano 2.9.3 reducer.py

#!/usr/bin/env python
import sys
deptdic = {}
for line in sys.stdin:
    line = line.strip()
    dept,sal = line.split('\t')
    if dept in deptdic:
        deptdic[dept].append(int(sal))
    else:
        deptdic[dept] = []
        deptdic[dept].append(int(sal))
for dept in deptdic.keys():
    sum_sal=sum(deptdic[dept])*1.0
    print '%s\t%s'% (dept,sum_sal)
```

Type `cat reducer.py`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ cat reducer.py
#!/usr/bin/env python
import sys
deptdic = {}
for line in sys.stdin:
    line = line.strip()
    dept,sal = line.split('\t')
    if dept in deptdic:
        deptdic[dept].append(int(sal))
    else:
        deptdic[dept] = []
        deptdic[dept].append(int(sal))
for dept in deptdic.keys():
    sum_sal=sum(deptdic[dept])*1.0
    print '%s\t%s'% (dept,sum_sal)
hduser@ubuntu:~/Documents/termwork5$
```

Type `ls`
& press



This shows your files

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ ls
mapper.py reducer.py salary.csv
hduser@ubuntu:~/Documents/termwork5$
```

Type `cat salary.csv | python mapper.py | python reducer.py`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ cat salary.csv | python mapper.py | python reducer
.py
Marketing      200000.0
Accounts       200000.0
IT             265000.0
hduser@ubuntu:~/Documents/termwork5$
```

Type jps
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ jps
2718 Jps
hduser@ubuntu:~/Documents/termwork5$
```

Type start-all.sh
& press



Provide password 4 times (it will ask)

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
```

Type `jps`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ jps
3776 NodeManager
3409 SecondaryNameNode
3601 ResourceManager
3126 DataNode
4102 Jps
2954 NameNode
hduser@ubuntu:~/Documents/termwork5$ █
```

Type `hdfs dfs -mkdir /tw5`
& press



If it goes into safe mode then type
`hdfs dfsadmin -safemode leave`
After this type
`hdfs dfs -mkdir /tw5`

This creates a directory in cluster

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ hdfs dfs -mkdir /tw5
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
24/07/26 23:15:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@ubuntu:~/Documents/termwork5$
```


Type `hdfs dfs -ls /`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ hdfs dfs -ls /
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
24/07/26 23:16:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 26 items
drwxr-xr-x - hduser supergroup 0 2024-07-09 04:50 /90
drwxr-xr-x - hduser supergroup 0 2024-06-25 03:02 /TERM2
drwxr-xr-x - hduser supergroup 0 2024-07-02 22:35 /TERM3
drwxr-xr-x - hduser supergroup 0 2024-07-02 22:35 /TERM4
drwxr-xr-x - hduser supergroup 0 2024-07-02 22:35 /TERM5
```

Type `hdfs dfs -copyFromLocal`
`/home/hduser/Documents/termwork5/salary.csv /tw5`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ hdfs dfs -copyFromLocal /home/hduser/Documents/termwork5/salary.csv /tw5
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
24/07/26 23:19:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@ubuntu:~/Documents/termwork5$
```

Type `hdfs dfs -ls /tw5`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ hdfs dfs -ls /tw5
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
24/07/26 23:21:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--  1 hduser supergroup      400 2024-07-26 23:19 /tw5/salary.csv
hduser@ubuntu:~/Documents/termwork5$
```

Type `sudo chmod 777 mapper.py reducer.py`
& press



Provide password dv or bca

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ sudo chmod 777 mapper.py reducer.py
[sudo] password for hduser:
hduser@ubuntu:~/Documents/termwork5$
```

Type `ls -l`
& press



```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ ls -l
total 12
-rwxrwxrwx 1 root root 253 Jul 26 23:03 mapper.py
-rwxrwxrwx 1 root root 360 Jul 26 23:08 reducer.py
-rw-r--r-- 1 root root 400 Jul 26 22:59 salary.csv
hduser@ubuntu:~/Documents/termwork5$
```


Type

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ hadoop jar /home/hduser/Downloads/hadoop-streaming-2.7.3.jar \
> -input /tw5/salary.csv \
> -output /tw5/output \
> -mapper /home/hduser/Documents/termwork5/mapper.py \
> -reducer /home/hduser/Documents/termwork5/reducer.py
```

& press



Everything working fine, it will show the output directory
& press



```
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=600
File Output Format Counters
      Bytes Written=49
24/07/26 23:28:43 INFO streaming.StreamJob: Output directory: /tw5/output
hduser@ubuntu:~/Documents/termwork5$
```

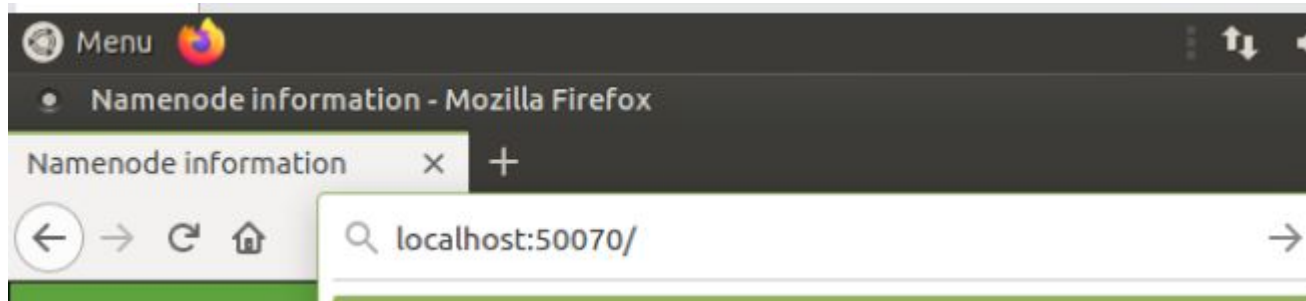
Type `hdfs dfs -cat /tw5/output/part-00000`
& press



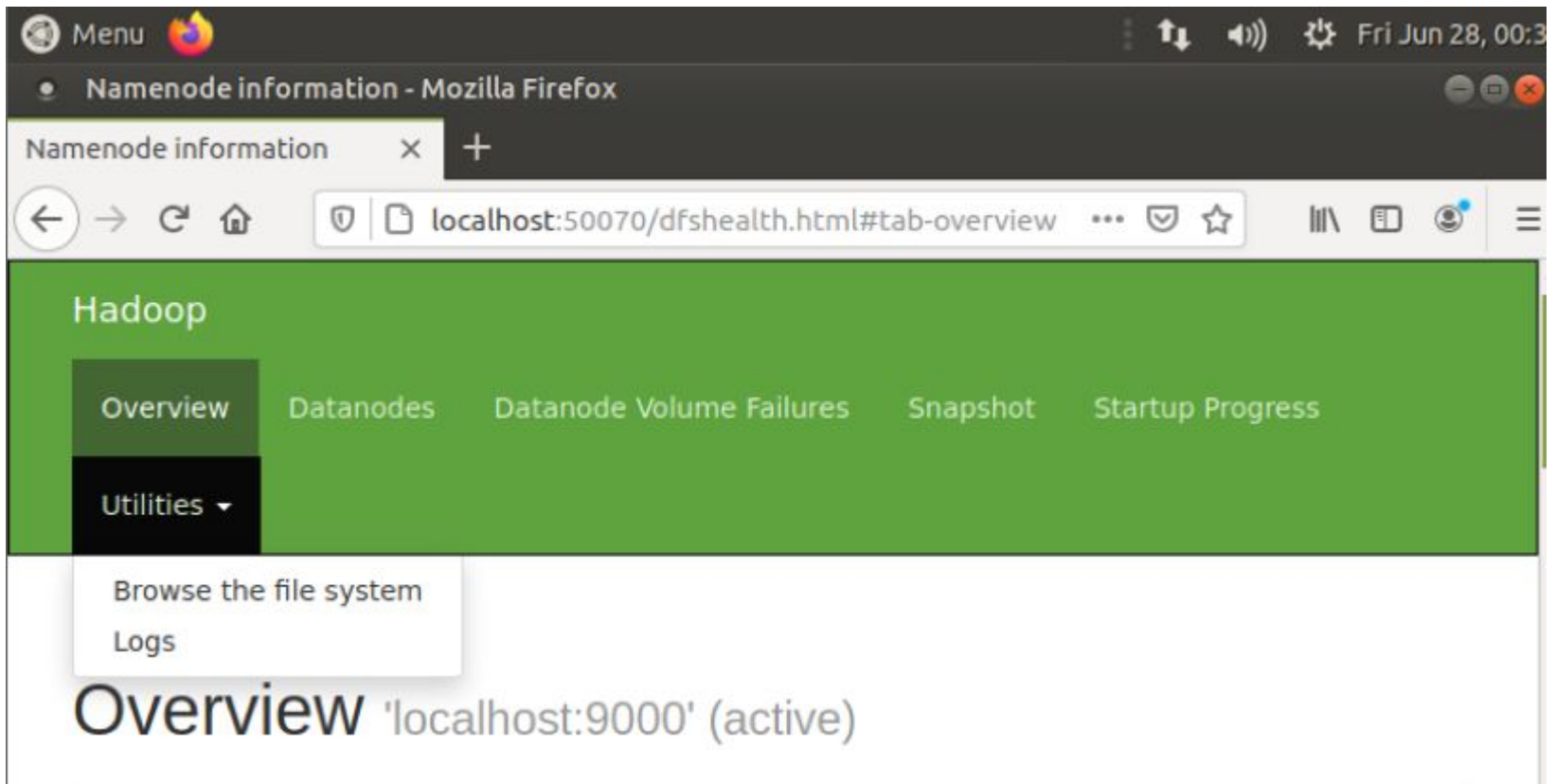
This shows the final output

```
File Edit View Search Terminal Help
hduser@ubuntu:~/Documents/termwork5$ hdfs dfs -cat /tw5/output/part-00000
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
24/07/26 23:30:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Marketing      200000.0
Accounts       200000.0
IT             265000.0
hduser@ubuntu:~/Documents/termwork5$
```


Now open firefox and type localhost:50070/
& press



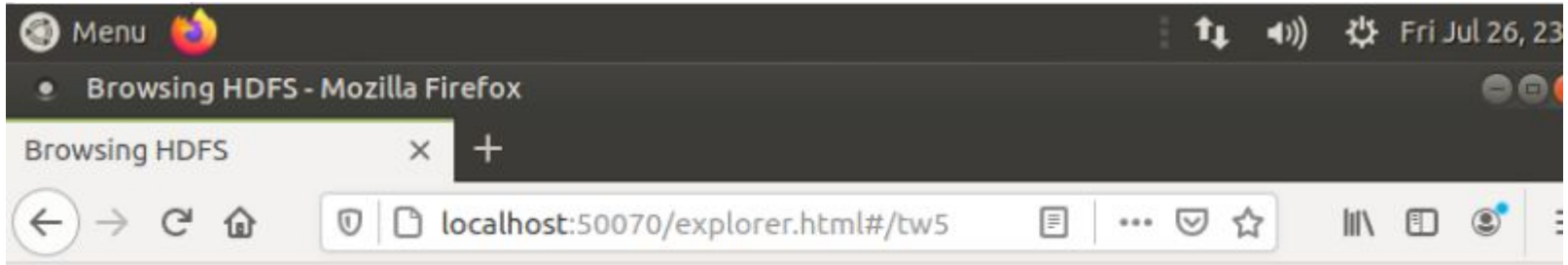
Click utilities & click Browse the file system



Click tw5

drwxr-xr-x	hduser	supergroup	0 B	6/28/2024, 12:30:12 AM	0	0 B	tw1
drwxr-xr-x	hduser	supergroup	0 B	7/10/2024, 12:39:46 AM	0	0 B	tw2
drwxr-xr-x	hduser	supergroup	0 B	7/12/2024, 1:05:10 AM	0	0 B	tw3
drwxr-xr-x	hduser	supergroup	0 B	7/25/2024, 2:51:50 AM	0	0 B	tw4
drwxr-xr-x	hduser	supergroup	0 B	7/26/2024, 11:28:00 PM	0	0 B	tw5

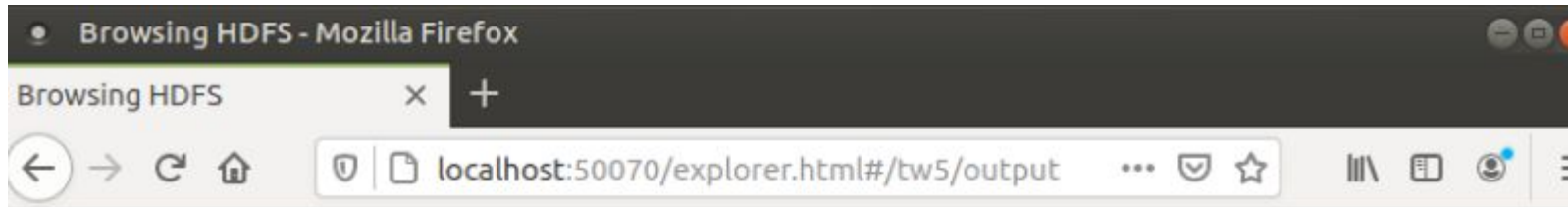
Click output



Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hduser	supergroup	0 B	7/26/2024, 11:28:40 PM	0	0 B	output
-rw-r--r--	hduser	supergroup	400 B	7/26/2024, 11:19:57 PM	1	128 MB	salary.csv

Click part-00000



Browse Directory

/tw5/output

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	hduser	supergroup	0 B	7/26/2024, 11:28:40 PM	1	128 MB	_SUCCESS
-rw-r--r--	hduser	supergroup	49 B	7/26/2024, 11:28:39 PM	1	128 MB	part-00000

Click Download

Browsing HDFS - Mozilla Firefox

Browsing HDFS x +

localhost:50070/explorer.html#/tw5/output

File information - part-00000

[Download](#)

Block information -- Block 0

Block ID: 1073742088

Block Pool ID: BP-1279745871-127.0.1.1-1717407862132

Generation Stamp: 1264

Size: 49

Availability:

- ubuntu

Hadoop,

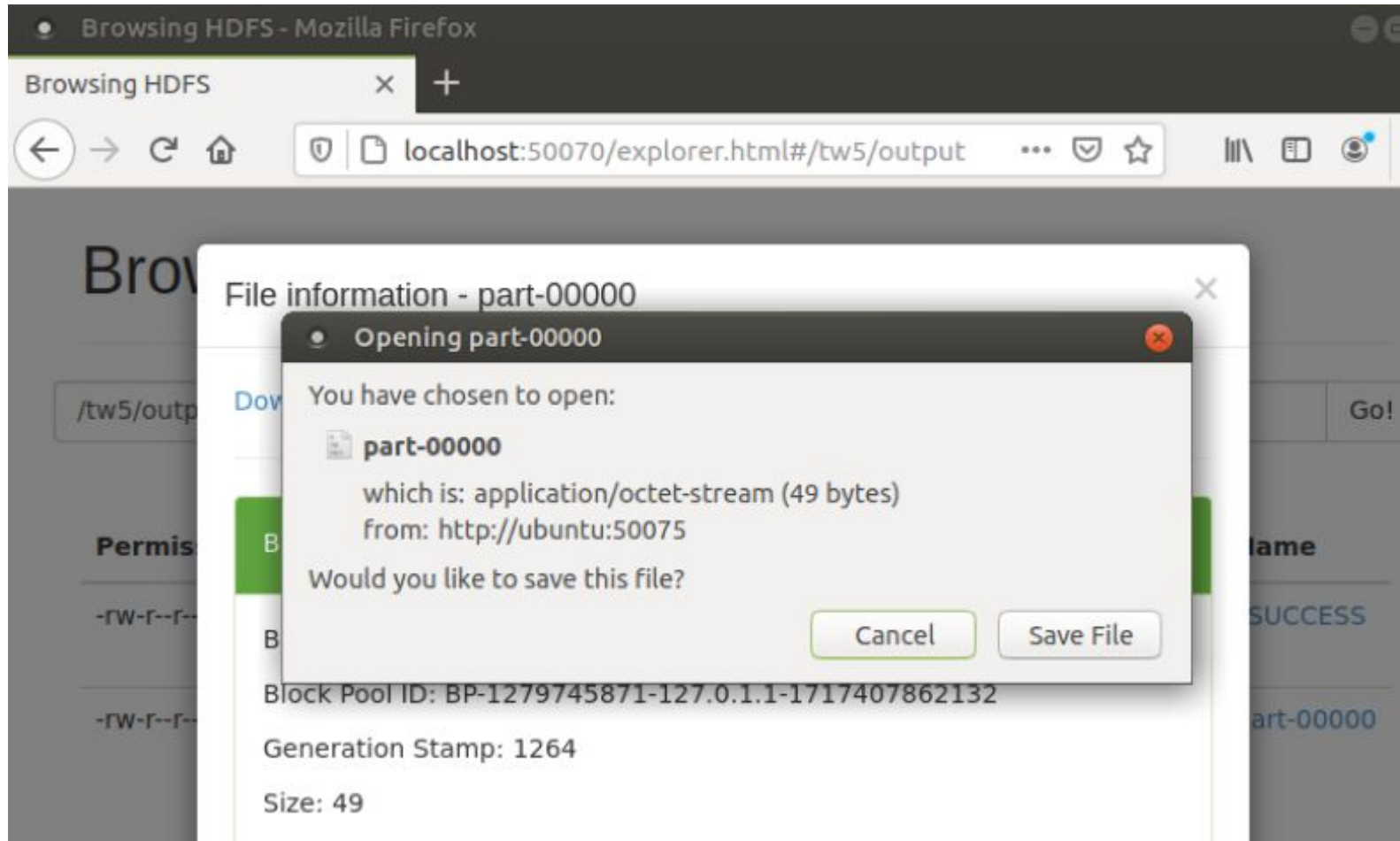
Go!

name

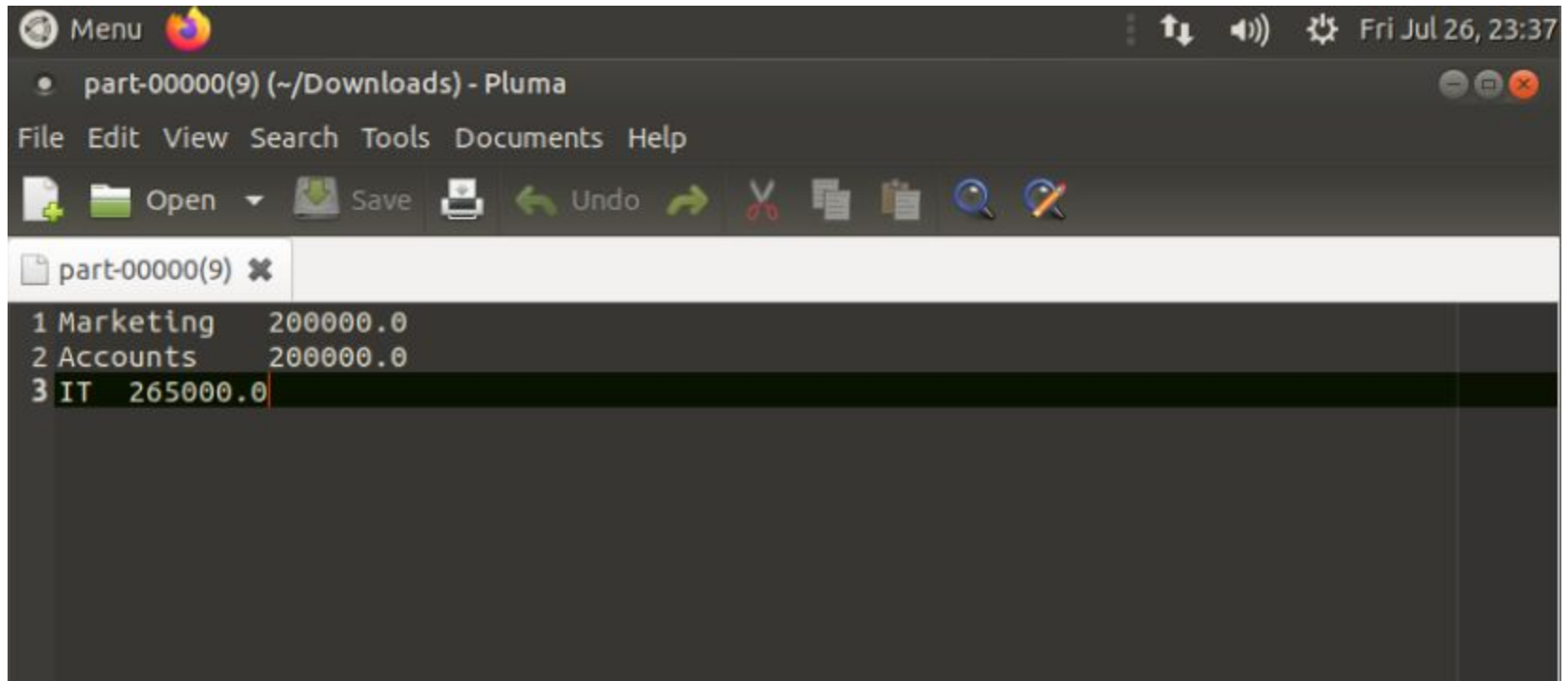
SUCCESS

part-00000

Click savefile



Go to download and open the file part-00000 to view the output



The screenshot shows a Pluma text editor window titled 'part-00000(9) (~/.Downloads) - Pluma'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Tools', 'Documents', and 'Help'. Below the menu bar is a toolbar with icons for 'Open', 'Save', 'Print', 'Undo', 'Redo', 'Cut', 'Copy', 'Paste', 'Find', and 'Replace'. The main text area contains a table with three rows of data. The first row is '1 Marketing 200000.0', the second row is '2 Accounts 200000.0', and the third row is '3 IT 265000.0'. The third row is highlighted with a red background.

1	Marketing	200000.0
2	Accounts	200000.0
3	IT	265000.0