

BIG DATA FUNDAMENTALS ASSIGNMENT

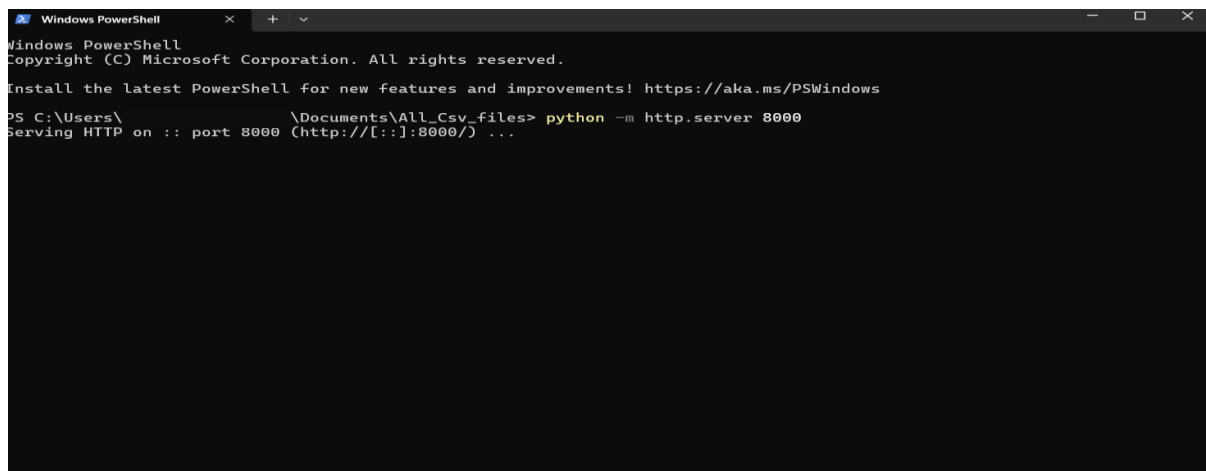
Name : Ganesh Agrahari

Section: Bca ds&ai 23

Roll No. :24

Hive

1. First of all I downloaded the dataset of **merged_dataset.csv** from Kaggle dataset
2. Created a folder in my local machine (windows) **All_Csv_File** and in that folder I move the **merged_dataset.csv** file



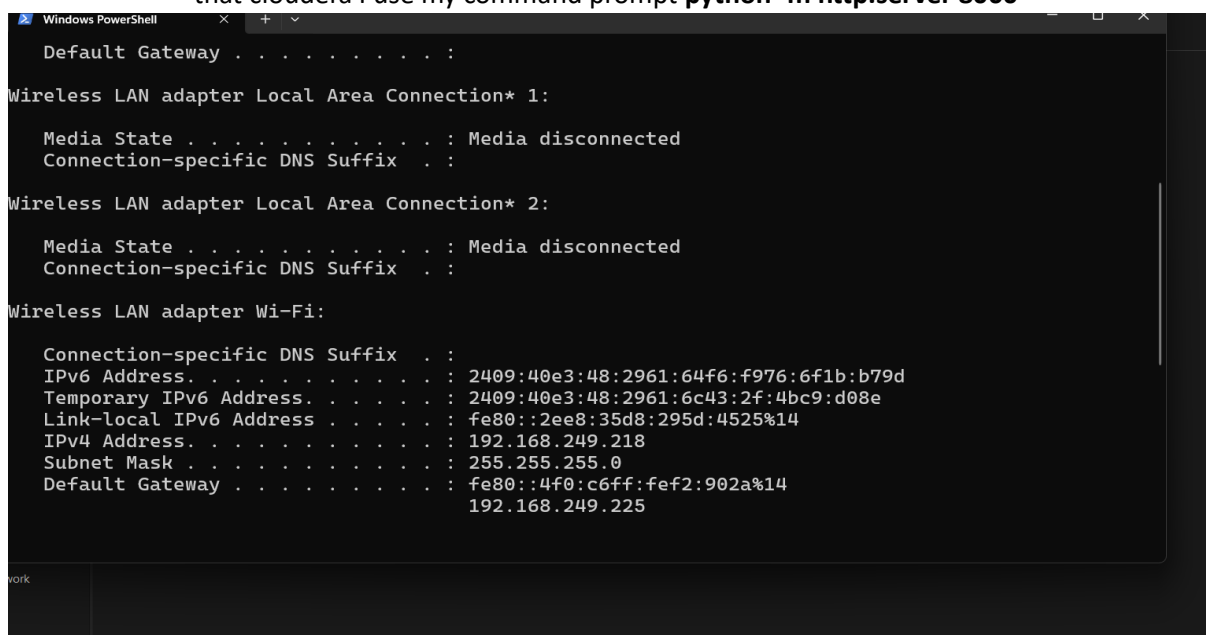
```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\...\Documents> python -m http.server 8000
Serving HTTP on :: port 8000 (http://[::]:8000/) ...
```

[uploading the file on server 8000]

After that I need to upload this file on **Hadoop Cloudera Manager** so to upload or paste this file on that cloudera I use my command prompt **python -m http.server 8000**



```
Windows PowerShell
Default Gateway . . . . . :
Wireless LAN adapter Local Area Connection* 1:
Media State . . . . . : Media disconnected
Connection-specific DNS Suffix . :
Wireless LAN adapter Local Area Connection* 2:
Media State . . . . . : Media disconnected
Connection-specific DNS Suffix . :
Wireless LAN adapter Wi-Fi:
Connection-specific DNS Suffix . :
IPv6 Address. . . . . : 2409:40e3:48:2961:64f6:f976:6f1b:b79d
Temporary IPv6 Address. . . . . : 2409:40e3:48:2961:6c43:2f:4bc9:d08e
Link-local IPv6 Address . . . . . : fe80::2ee8:35d8:295d:4525%14
IPv4 Address. . . . . : 192.168.249.218
Subnet Mask . . . . . : 255.255.255.0
Default Gateway . . . . . : fe80::4f0:c6ff:fef2:902a%14
192.168.249.225
```

1. After successfully we need to get that files on cloudera so on cloudera terminal we run the command
2. **wget http:// 192.168.249.218/merged_dataset.csv**

192.168.249.218:- is a IPV4 address of my computer

Note:- I have run this command two times that's why there is a two file of csv are present in my cloudera second name file is **movies**

```
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ wget http://192.168.249.218/netflix_titles.csv
--2025-04-24 22:15:17-- http://192.168.249.218/netflix_titles.csv
Connecting to 192.168.249.218:80... connected.
HTTP request sent, awaiting response... 404 File not found
2025-04-24 22:15:17 ERROR 404: File not found.

[cloudera@quickstart ~]$ http://192.168.249.218:8000/netflix_titles.csv
bash: http://192.168.249.218:8000/netflix_titles.csv: No such file or directory
[cloudera@quickstart ~]$ wget http://192.168.249.218:8000/netflix_titles.csv
--2025-04-24 22:16:59-- http://192.168.249.218:8000/netflix_titles.csv
Connecting to 192.168.249.218:8000... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2129404 (2.0M) [application/vnd.ms-excel]
Saving to: "netflix_titles.csv"

100%[=====] 2,129,404 4.06M/s in 0.5s

2025-04-24 22:16:59 (4.06 MB/s) - "netflix_titles.csv" saved [2129404/2129404]

[cloudera@quickstart ~]$
```

[verifying the folder has import or not]

ON cloudera terminal by doing **ls cmd** I verify the **merged_dataset.csv** has import or not

```
[cloudera@quickstart ~]$ ls
cloudera-manager Desktop Downloads enterprise-deployment.json input.txt lib Music n
m_api.py Documents eclipse express-deployment.json kerberos mapper.py netflix_titles.csv p
[cloudera@quickstart ~]$ -rm -r netflix_titles.csv
bash: -rm: command not found
[cloudera@quickstart ~]$
```

hdfs dfs -ls /user/cloudera

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera
Found 6 items
-rw-r--r-- 1 cloudera cloudera 157 2025-04-15 11:41 /user/cloudera/Student.txt
drwxr-xr-x - cloudera cloudera 0 2025-04-20 01:38 /user/cloudera/input
drwxr-xr-x - cloudera cloudera 0 2025-04-25 00:48 /user/cloudera/movies
drwxr-xr-x - cloudera cloudera 0 2025-04-24 23:43 /user/cloudera/netflix_titles1
drwxr-xr-x - cloudera cloudera 0 2025-04-23 23:14 /user/cloudera/statewisepopl
drwxr-xr-x - cloudera cloudera 0 2025-04-20 02:49 /user/cloudera/user
```

Upload the **merged_dataset.csv** to the Hadoop so, first I've made a folder Hadoop fs -mkdir /user/cloudera/moviesAnd put the **merged_dataset.csv** file in that folder by Hadoop fs -put **merged_dataset.csv** /user/cloudera/movies

[now go to hive]

By just clicking hive we can go to hive (sometimes it take time)

```

[cloudera@quickstart ~]$ wget http://192.168.249.218/netflix_titles.csv
--2025-04-24 22:15:17-- http://192.168.249.218/netflix_titles.csv
Connecting to 192.168.249.218:80... connected.
HTTP request sent, awaiting response... 404 File not found
2025-04-24 22:15:17 ERROR 404: File not found.

[cloudera@quickstart ~]$ http://192.168.249.218:8000/netflix_titles.csv
ash: http://192.168.249.218:8000/netflix_titles.csv: No such file or directory
[cloudera@quickstart ~]$ wget http://192.168.249.218:8000/netflix_titles.csv
--2025-04-24 22:16:59-- http://192.168.249.218:8000/netflix_titles.csv
Connecting to 192.168.249.218:8000... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2129404 (2.0M) [application/vnd.ms-excel]
Saving to: "netflix_titles.csv"

100%[=====>] 2,129,404 4.06M/s in 0.5s

2025-04-24 22:16:59 (4.06 MB/s) - "netflix_titles.csv" saved [2129404/2129404]

[cloudera@quickstart ~]$ ls
cloudera-manager  enterprise-deployment.json  Music  Public  Videos
m_api.py          express-deployment.json    netflix_titles.csv  reducer.py  WordCount.jar
desktop          input.txt                  parcels           sample.csv  WordCount.java
documents        kerberos                  Pictures         statewisepop.csv  workspace
downloads        lib                      Placement Data.csv  statewisepop.csv.1
eclipse          mapper.py                 Projectsfolder  Templates

[cloudera@quickstart ~]$ hdfs dfs -mkdir/user/cloudera/netflix_titles.csv
mkdir/user/cloudera/netflix_titles.csv: Unknown command
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/cloudera/netflix_titles
[cloudera@quickstart ~]$ hdfs dfs -mkdir ls
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 6 items
-rw-r--r-- 1 cloudera cloudera 157 2025-04-15 11:41 Student.txt
drwxr-xr-x - cloudera cloudera 0 2025-04-20 01:38 input
drwxr-xr-x - cloudera cloudera 0 2025-04-24 22:21 ls
drwxr-xr-x - cloudera cloudera 0 2025-04-24 22:20 netflix_titles
drwxr-xr-x - cloudera cloudera 0 2025-04-23 23:14 statewisepop1
drwxr-xr-x - cloudera cloudera 0 2025-04-20 02:49 user
[cloudera@quickstart ~]$ hdfs dfs -rm -r ls
Deleted ls
[cloudera@quickstart ~]$ hdfs dfs -put netflix_titles.csv /user/cloudera/netflix_titles
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive>

```

[query commands]

```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Edit View Search Terminal Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera
Found 5 items
-rw-r--r-- 1 cloudera cloudera 157 2025-04-15 11:41 /user/cloudera/Student.txt
drwxr-xr-x - cloudera cloudera 0 2025-04-20 01:38 /user/cloudera/input
drwxr-xr-x - cloudera cloudera 0 2025-04-24 07:42 /user/cloudera/statewisepop
drwxr-xr-x - cloudera cloudera 0 2025-04-23 23:14 /user/cloudera/statewisepop1
drwxr-xr-x - cloudera cloudera 0 2025-04-20 02:49 /user/cloudera/user
[cloudera@quickstart ~]$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> [cloudera@quickstart]hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> DROP TABLE state_population;
OK
Time taken: 2.026 seconds
hive> CREATE TABLE state_population(states uts STRING, population 2024 INT, male literacy rate FLOAT, female literacy
_avg rate FLOAT, average literacy rate FLOAT, area sq km INT, gdp FLOAT, per capita income FLOAT, sex ratio INT, hindu
FLOAT, muslim FLOAT, christian FLOAT, sikhs FLOAT, buddhist FLOAT, others FLOAT, majority STRING) ROW FORMAT DELIMIT
ED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/cloudera/statewisepop1';
OK
Time taken: 4.749 seconds
hive> SELECT * FROM state_population;
OK
States/Uts NULL NULL NULL NULL NULL NULL NULL NULL NULL NULL NULL NULL NULL
ULL Majority
Andaman and Nicobar Islands 404000 90.11 81.84 86.27 8249 0.89 229000.0 1045 0.6945 0.085
2 0.2128 0.0034 9.0E-4 0.0032 Hindu
73.4 59.5 66.4 162970 14.5 219518.0 963 0.9087 0.0733 0.013
8 2.0E-4 1.0E-4 0.0039 Hindu
Arunachal Pradesh 150000 73.69 59.57 66.95 83743 0.38 256410.0 997 0.2904 0.0195 0.302
6 0.0024 0.1177 0.2674 Christian
Assam 36159000 90.1 81.2 85.9 78438 5.67 118504.0 1012 0.6147 0.3422 0.0374 7.0E-
4 0.0018 0.0032 Hindu
Bihar 129205000 79.7 60.5 70.9 94163 8.59 54000.0 1090 0.8269 0.1687 0.0012 2.0E-4 2.0E-
4 0.0028 Hindu
Chandigarh 1247000 90.54 81.38 86.43 114 0.49 333000.0 917 0.8078 0.0487 0.0011 0.131
1 NULL
Chhattisgarh 30638000 85.4 68.7 77.3 135192 5.09 152348.0 1015 0.9325 0.0262 0.019
2 0.0027 0.0028 0.0226 Hindu
Dadra & Nagar Haveli and Diu 1387000 89.5 76.2 80.5 603 NULL NULL 827 0.9393 0.037
6 0.0149 6.0E-4 0.0018 0.0058 Hindu
Delhi 21684000 93.7 82.4 88.7 1483 10.4 444768.0 913 0.8168 0.1286 0.0087 0.034
0.0011 0.0108 Hindu
Goa 1585000 92.81 81.84 87.4 3702 1.0 527146.0 1027 0.6608 0.0833 0.251 0.001 8.0E-
4 0.0031 Hindu

```

```
SELECT * FROM merged_dataset LIMIT 5;
```

```
hive> SELECT * FROM merged_dataset LIMIT 5;
OK
name      NULL      movie_rated  run_length  genres  release_date  NULL      NULL      NULL
Inception  2010      PG-13      2h 28min    Action; Adventure; Sci-Fi;  16 July 2010 (USA)      8.8      19816
75      3820
The Lord of the Rings: The Fellowship of the Ring  2001      PG-13      2h 58min    Action; Adventure; Drama;  1
9 December 2001 (USA)      8.8      1609165 5365
The Lord of the Rings: The Return of the King  2003      PG-13      3h 21min    Adventure; Drama; Fantasy;  17 De
cember 2003 (USA)      8.9      1593859 3681
The Dark Knight Rises  2012      PG-13      2h 44min    Action; Adventure;  20 July 2012 (USA)      8.4      14703
29      2979
Time taken: 0.133 seconds, Fetched: 5 row(s)
hive>
```

```
SELECT COUNT(*) FROM merged_dataset;
```

```
Time taken: 0.133 seconds, Fetched: 5 row(s)
hive> SELECT COUNT (*) FROM merged_dataset;
Query ID = cloudera_20250425013232_e15fcd02-c66a-40c8-ac21-127fb701ba9b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1745503643740_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1745503643740_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1745503643740_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-04-25 01:32:42,427 Stage-1 map = 0%, reduce = 0%
2025-04-25 01:32:50,996 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.97 sec
2025-04-25 01:33:09,405 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.07 sec
MapReduce Total cumulative CPU time: 3 seconds 70 msec
Ended Job = job_1745503643740_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.07 sec HDFS Read: 149321 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 70 msec
OK
1501
Time taken: 42.861 seconds, Fetched: 1 row(s)
hive>
```

```
SELECT name, COUNT(*) AS total
```

```
FROM merged_dataset
```

```
GROUP BY year;
```

```
hive> SELECT type , COUNT(*) AS total FROM merged_dataset GROUP BY type ;
FAILED: SemanticException [Error 10004]: Line 1:61 Invalid table alias or column reference 'type': (possible column names are: name, year, movie_rated, run_length, genres, release_date, rating, num_raters, num_reviews)
hive>
```

merged_dataset to create the following output .

Output of the command **SELECT * FROM merged_dataset ;**

Don Jon 2013	R	1h 30min	Comedy; Romance;	27 September 2013 (USA)	6.3	224890	338	
The Fountain 2006	PG-13	1h 37min	Romance;	22 November 2006 (USA)	7.2	224185	985	
Marriage Story 2019	R	2h 17min	Comedy; Romance;	6 December 2019 (USA)	8.0	219894	1178	
Bridget Jones's Diary 5	2001	R	1h 37min	Comedy; Romance;	13 April 2001 (USA)	6.7	21871	
The Beach 2000	R	1h 59min	Romance;	11 February 2000 (USA)	6.7	218197	587	
Warm Bodies 2013	PG-13	1h 38min	Comedy; Romance;	1 February 2013 (USA)	6.9	215679	378	
Just Go with It 2011	PG-13	1h 57min	Comedy; Romance;	11 February 2011 (USA)	6.4	215475	271	
Singin' in the Rain 9	1952	G	1h 43min	Comedy; Romance;	11 April 1952 (USA)	8.3	21137	
Mr. Nobody 2009	R	2h 21min	Romance;	26 September 2013 (USA)	7.8	210352	393	
Meet Joe Black 1998	PG-13	2h 58min	Romance;	13 November 1998 (USA)	7.2	208791	799	
Mamma Mia! 2008	PG-13	1h 48min	Comedy; Romance;	18 July 2008 (USA)	6.4	208596	752	
No Strings Attached 9	2011	R	1h 48min	Comedy; Romance;	21 January 2011 (USA)	6.2	20828	
The Lobster 4	2015	R	1h 59min	Comedy; Romance;	16 October 2015 (Ireland)	7.2	20745	
Romeo + Juliet 1996	PG-13	2h	Romance;	1 November 1996 (USA)	6.7	205296	576	
Garden State 2004	R	1h 42min	Comedy; Romance;	20 August 2004 (USA)	7.4	204812	973	
How to Lose a Guy in 10 Days 93365	2003	PG-13	1h 56min	Comedy; Romance;	7 February 2003 (USA)	6.4	2	
Match Point 2005	R	2h 4min	Romance;	20 January 2006 (USA)	7.6	201777	786	
True Romance 1993	R	1h 59min	Romance;	10 September 1993 (USA)	7.9	200842	537	
The Ugly Truth 2009	R	1h 36min	Comedy; Romance;	24 July 2009 (USA)	6.4	199488	195	
Closer 2004	R	1h 44min	Romance;	3 December 2004 (USA)	7.2	198838	952	
The Girl Next Door 9	2004	R	1h 49min	Comedy; Romance;	9 April 2004 (USA)	6.7	19748	
Me Before You 2016	PG-13	1h 46min	Romance;	3 June 2016 (USA)	7.4	197474	457	
P.S. I Love You 2007	PG-13	2h 6min	Comedy; Romance;	21 December 2007 (USA)	7.0	197346	276	
Call Me by Your Name 2017	R	2h 12min	Romance;	19 January 2018 (USA)	7.9	195583	951	
A Walk to Remember 2002	PG	1h 41min	Romance;	25 January 2002 (USA)	7.4	195515	1010	
Dirty Dancing 1987	PG-13	1h 40min	Romance;	21 August 1987 (USA)	7.0	194838	462	
"I Love You" NULL 2009	R	1h 45min	Comedy; Romance;	NULL	7	191014		
Ghost 1990	PG-13	2h 7min	Romance;	13 July 1990 (USA)	7.0	190746	301	
Revolutionary Road 2008	R	1h 59min	Romance;	23 January 2009 (USA)	7.3	190069	486	
When Harry Met Sally... 9	1989	R	1h 35min	Comedy; Romance;	21 July 1989 (USA)	7.6	18907	
Bad Teacher 2011	R	1h 32min	Comedy; Romance;	24 June 2011 (USA)	5.6	189027	325	
Inception 88	2010	PG-13	2h 28min	Action; Adventure; Sci-Fi;	16 July 2010 (USA)	8.8	19816	
The Matrix 3820	1999	R	2h 16min	Action; Sci-Fi;	31 March 1999 (USA)	8.7	1619772	4281
Interstellar 10	2014	PG-13	2h 49min	Adventure; Drama; Sci-Fi;	7 November 2014 (USA)	8.6	14308	
The Avengers 53	2012	PG-13	2h 23min	Action; Adventure; Sci-Fi;	4 May 2012 (USA)	8.0	12360	
The Prestige 21	2006	PG-13	2h 10min	Drama; Mystery; Sci-Fi;	20 October 2006 (USA)	8.5	11445	
Back to the Future 1394	1985	PG	1h 56min	Adventure; Comedy; Sci-Fi;	3 July 1985 (USA)	8.5	1	

MapReduce

Use the **ls-l** command to show the list of the files

Create the directory DS22 by **mkdir**.

Change the directory to DS22 by **cd** command

Use **cat >input.txt** to create the file named **input.txt** by the **cat** command and enter the data.

```

cloudera@quickstart:~/DS22
File Edit View Search Terminal Help

[cloudera@quickstart ~]$ ls -l
total 188
-rwxrwxr-x 1 cloudera cloudera 5387 Oct 23 2017 cloudera-manager
-rwxrwxr-x 1 cloudera cloudera 9964 Oct 23 2017 cm_api.py
drwxrwxr-x 2 cloudera cloudera 4896 Apr 23 02:34 Desktop
drwxrwxr-x 2 cloudera cloudera 4896 Oct 23 2017 Documents
drwxrwxr-x 9 cloudera cloudera 4896 Apr 16 21:54 Downloads
drwxrwxr-x 1 cloudera cloudera 4896 Apr 23 02:13 eclipse
-rw-rw-r-- 1 cloudera cloudera 53655 Oct 23 2017 enterprise-deployment.json
-rw-rw-r-- 1 cloudera cloudera 50515 Oct 23 2017 express-deployment.json
-rwxrwxr-x 1 cloudera cloudera 5887 Oct 23 2017 kerberos
drwxrwxr-x 2 cloudera cloudera 4896 Oct 23 2017 lib
drwxrwxr-x 2 cloudera cloudera 4896 Apr 16 21:54 Music
-rwxrwxr-x 1 cloudera cloudera 4228 Oct 23 2017 parcels
drwxrwxr-x 2 cloudera cloudera 4896 Apr 16 21:54 Pictures
drwxrwxr-x 2 cloudera cloudera 4896 Apr 16 21:54 Public
drwxrwxr-x 2 cloudera cloudera 4896 Apr 16 21:54 Templates
drwxrwxr-x 2 cloudera cloudera 4896 Apr 16 21:54 Videos
drwxrwxr-x 5 cloudera cloudera 4896 Apr 23 02:14 workspace
[cloudera@quickstart ~]$ mkdir DS22
[cloudera@quickstart ~]$ cd DS22
[cloudera@quickstart DS22]$ cat >input.txt
Hello class
This is our map reduce practical
Hello map reduce
map reduce is important to learn
[cloudera@quickstart DS22]$ pwd
/home/cloudera/DS22
[cloudera@quickstart DS22]$
cloudera@quickstart DS22$

```

use **put** cmd to put data of input.txt to DS22.

Then check the the data by **ls**.

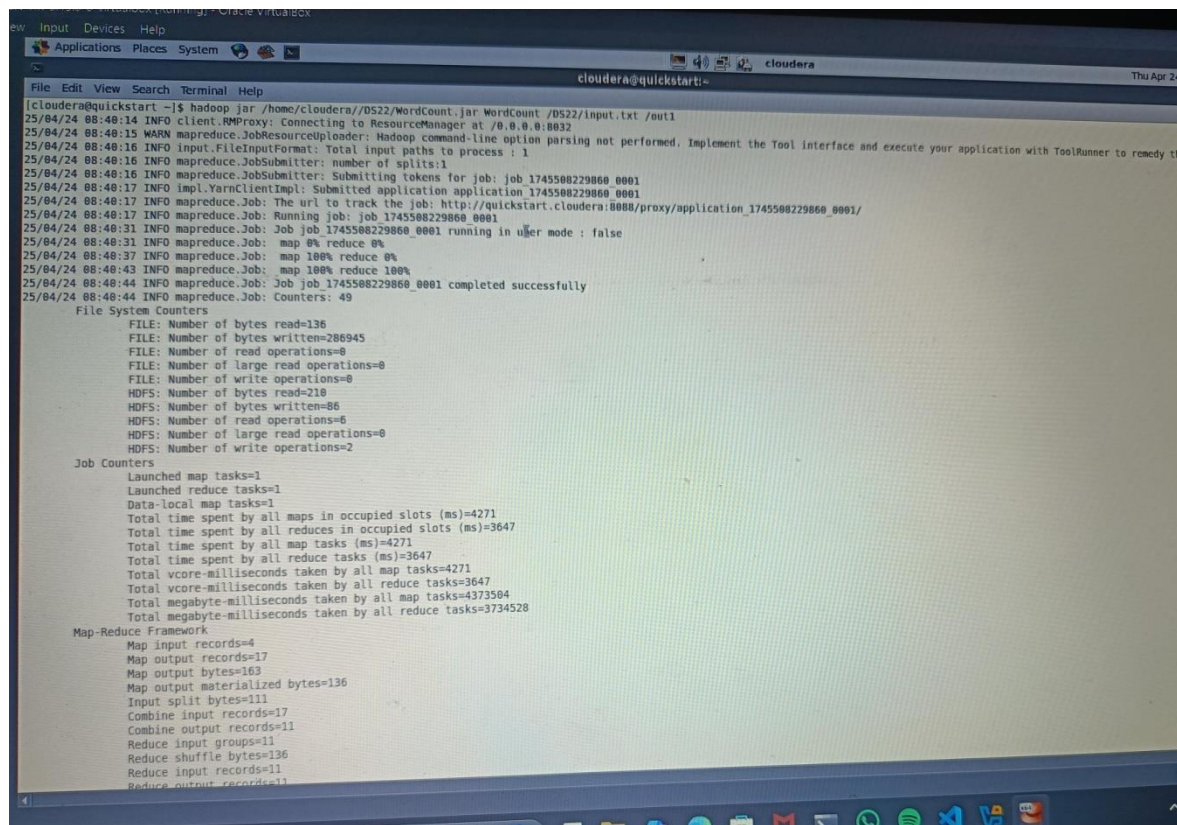
```

cloudera@quickstart:~
File Edit View Search Terminal Help

[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxr-xr-x - cloudera supergroup 0 2025-04-23 09:17 /DS22
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - cloudera supergroup 0 2025-04-23 10:09 /ds22
drwxr-xr-x - hbase supergroup 0 2025-04-23 21:40 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2025-04-16 21:53 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -ls /ds22
Found 1 items
-rw-r--r-- 1 cloudera supergroup 32 2025-04-23 10:09 /ds22/input.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /DS22
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/DS22/input.txt/DS22
put: '/home/cloudera/DS22/input.txt/DS22': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/DS22/input.txt /DS22/
[cloudera@quickstart ~]$ hdfs dfs -ls /DS22/
Found 1 items
-rw-r--r-- 1 cloudera supergroup 99 2025-04-23 22:01 /DS22/input.txt
[cloudera@quickstart ~]$

```

Imported the **mapReduce** framework in to perform the the the task of mapReduce work.



```

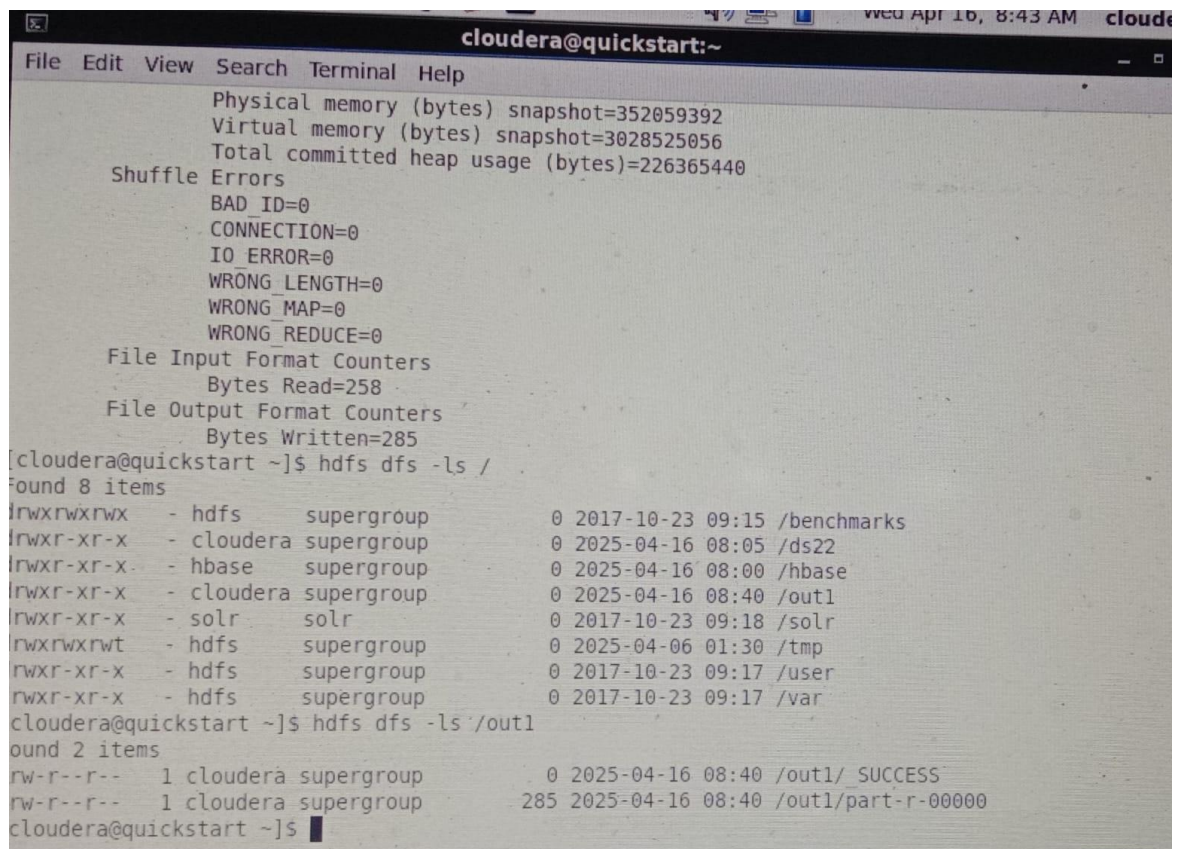
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/DS22/WordCount.jar WordCount /DS22/input.txt /out1
25/04/24 08:48:14 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/04/24 08:48:15 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy th
25/04/24 08:48:16 INFO input.FileInputFormat: Total input paths to process : 1
25/04/24 08:48:16 INFO mapreduce.JobSubmitter: number of splits:1
25/04/24 08:48:16 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1745508229860_0001
25/04/24 08:48:17 INFO impl.YarnClientImpl: Submitted application application_1745508229860_0001
25/04/24 08:48:17 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1745508229860_0001/
25/04/24 08:48:17 INFO mapreduce.Job: Running job: job_1745508229860_0001
25/04/24 08:48:31 INFO mapreduce.Job: Job job_1745508229860_0001 running in user mode : false
25/04/24 08:48:31 INFO mapreduce.Job: map 0% reduce 0%
25/04/24 08:48:37 INFO mapreduce.Job: map 100% reduce 0%
25/04/24 08:48:43 INFO mapreduce.Job: map 100% reduce 100%
25/04/24 08:48:44 INFO mapreduce.Job: Job job_1745508229860_0001 completed successfully
25/04/24 08:48:44 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=136
  FILE: Number of bytes written=286945
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=218
  HDFS: Number of bytes written=86
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4271
  Total time spent by all reduces in occupied slots (ms)=3647
  Total time spent by all map tasks (ms)=4271
  Total time spent by all reduce tasks (ms)=3647
  Total vcore-milliseonds taken by all map tasks=4271
  Total vcore-milliseonds taken by all reduce tasks=3647
  Total megabyte-milliseonds taken by all map tasks=4373504
  Total megabyte-milliseonds taken by all reduce tasks=3734528

Map-Reduce Framework
  Map input records=4
  Map output records=17
  Map output bytes=163
  Map output materialized bytes=136
  Input split bytes=111
  Combine input records=17
  Combine output records=11
  Reduce input groups=11
  Reduce shuffle bytes=136
  Reduce input records=11
  Reduce output records=11
  
```

check the directory by cmd **hdfs dfs -ls /out1** (created a out1 to store the output).

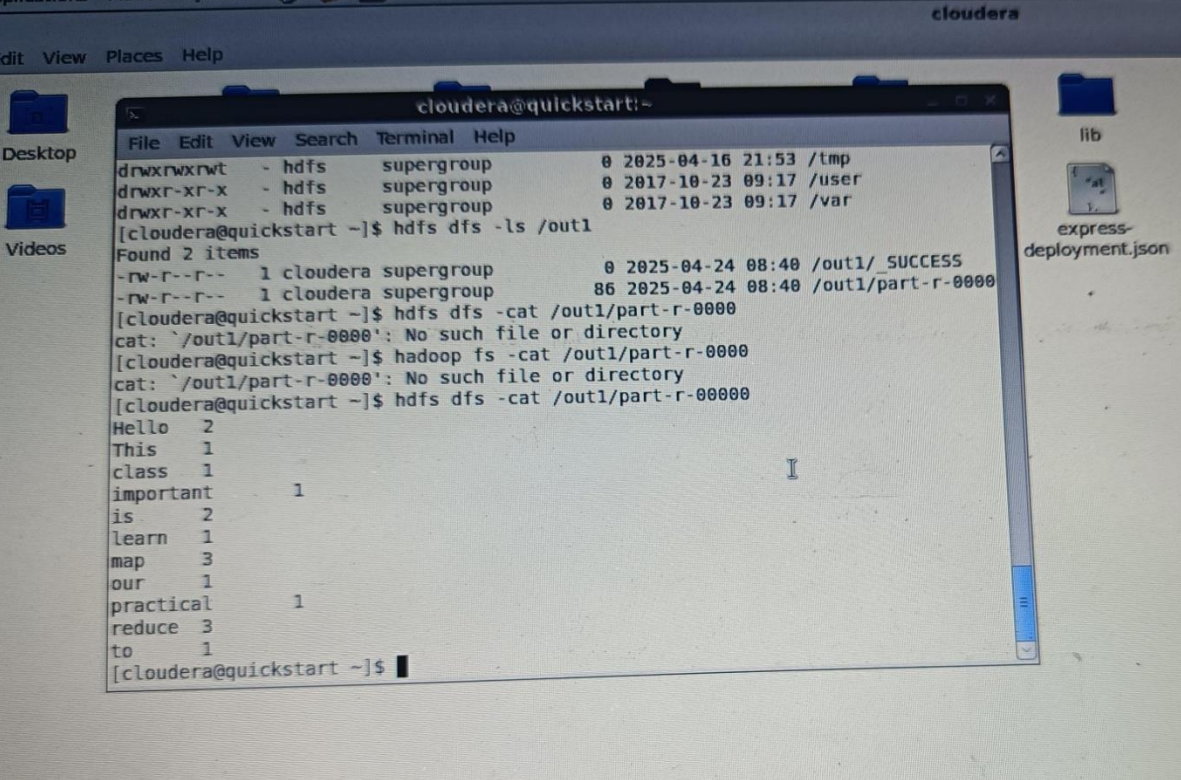


```

[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - cloudera supergroup 0 2025-04-16 08:05 /ds22
drwxr-xr-x - hbase supergroup 0 2025-04-16 08:00 /hbase
drwxr-xr-x - cloudera supergroup 0 2025-04-16 08:40 /out1
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
rwxrwxrwt - hdfs supergroup 0 2025-04-06 01:30 /tmp
rwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
rwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart ~]$ hdfs dfs -ls /out1
Found 2 items
rw-r--r-- 1 cloudera supergroup 0 2025-04-16 08:40 /out1/ SUCCESS
rw-r--r-- 1 cloudera supergroup 285 2025-04-16 08:40 /out1/part-r-00000
[cloudera@quickstart ~]$
  
```


Use the cmd **hdfs dfs -cat /out1/part-r-00000** to get the output of map_reduce .



The screenshot shows a terminal window titled "cloudera@quickstart:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the following commands and output:

```
cloudera@quickstart:~$ hdfs dfs -ls /out1
Found 2 items
-rw-r--r-- 1 cloudera supergroup 0 2025-04-24 08:40 /out1/_SUCCESS
-rw-r--r-- 1 cloudera supergroup 86 2025-04-24 08:40 /out1/part-r-00000
cloudera@quickstart:~$ hdfs dfs -cat /out1/part-r-00000
cat: '/out1/part-r-00000': No such file or directory
cloudera@quickstart:~$ hadoop fs -cat /out1/part-r-00000
cat: '/out1/part-r-00000': No such file or directory
cloudera@quickstart:~$ hdfs dfs -cat /out1/part-r-00000
Hello 2
This 1
class 1
important 1
is 2
learn 1
map 3
our 1
practical 1
reduce 3
to 1
cloudera@quickstart:~$
```

The background desktop shows icons for Desktop, Videos, lib, and express-deployment.json.