# Big Data H – 2

**NAME:** Vijay J

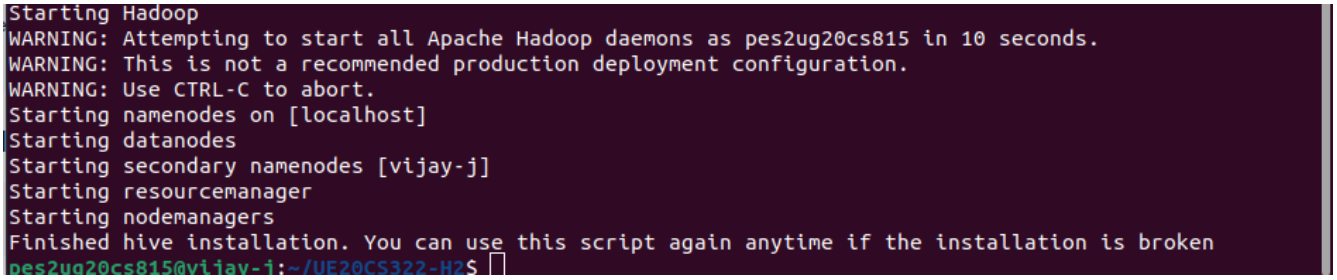
**SRN:** PES2UG20CS815

**SECTION:** J

## ScreenShots:

## 1a:

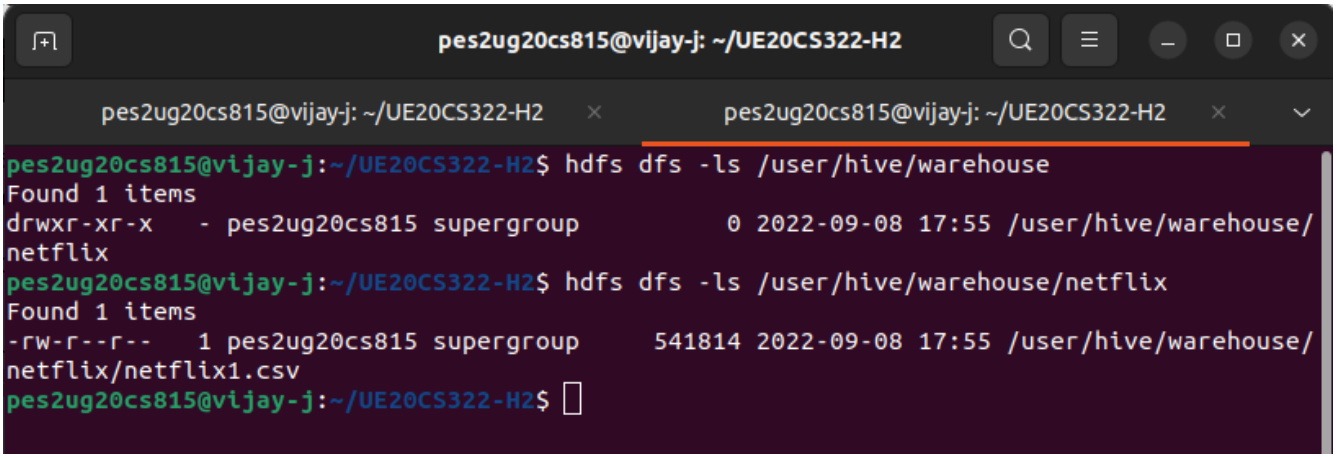

## 1a1:



## 2a:

```
hive> create table netflix(show_id String,type String,title String,director String,country String,release_year int,primary key (show_id) disa
ble novalidate) row format delimited fields terminated by ',';
OK
Time taken: 4.11 seconds
hive> load data local inpath '/home/pes2ug20cs815/UE20CS322-H2/netflix1.csv' into table netflix;
Loading data to table default.netflix
OK
Time taken: 4.633 seconds
hive> select * from netflix limit 3;
OK
s1      Movie   Dick Johnson Is Dead    Kirsten Johnson United States   2020
s3      TV Show Ganglands       Julien Leclercq France  2021
s6      TV Show Midnight Mass   Mike Flanagan   United States   2021
Time taken: 5.972 seconds, Fetched: 3 row(s)
hive>
```

## 2b:




## 2c1:

```
hive> set hive.exec.dynamic.partition=True;
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> create table netflix_partition(title String,director String,country String,release_
year int) partitioned by (type String);
OK
Time taken: 0.595 seconds
hive>
```

nshot of the terminal output and name it 2c.png(If you have multiple screenshots

## 2c2:

```
hive> insert into table netflix_partition partition(type='Movie') select title,director,country,release_year from netflix where type='Movie';
Query ID = pes2ug20cs815_20220908175932_26af45eb-ec86-4e9b-b341-e6d8c755317a
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 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_1662639493345_0001, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0001/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-08 18:00:20,625 Stage-1 map = 0%,  reduce = 0%
2022-09-08 18:01:00,040 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.72 sec
2022-09-08 18:01:12,829 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.17 sec
MapReduce Total cumulative CPU time: 10 seconds 170 msec
Ended Job = job_1662639493345_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/netflix_partition/type=Movie/.hive-staging_hive_2022-09-08_17-59-32_209_59
46220248287289730-1/-ext-10000
Loading data to table default.netflix_partition partition (type=Movie)
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.17 sec   HDFS Read: 559899 HDFS Write: 304492 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 170 msec
OK
Time taken: 109.288 seconds
hive>
```

## 2c3:

```
hive> insert into table netflix_partition partition(type='TV Show') select title,director,country,release_year from netflix where type='TV Sh
ow';
Query ID = pes2ug20cs815_20220908180232_af14dfc2-e80e-4b0f-995b-f3e508d7dc97
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 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_1662639493345_0002, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0002/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-08 18:02:53,430 Stage-1 map = 0%,  reduce = 0%
2022-09-08 18:03:35,838 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 7.05 sec
2022-09-08 18:03:54,837 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.06 sec
MapReduce Total cumulative CPU time: 11 seconds 60 msec
Ended Job = job_1662639493345_0002
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/netflix_partition/type=TV Show/.hive-staging_hive_2022-09-08_18-02-32_343_
6693865936898144663-1/-ext-10000
Loading data to table default.netflix_partition partition (type=TV Show)
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.06 sec   HDFS Read: 559939 HDFS Write: 120389 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 60 msec
OK
Time taken: 87.649 seconds
hive>
```

## 2c4:

```
hive> select * from netflix_partition limit 3;
OK
Dick Johnson Is Dead    Kirsten Johnson United States   2020    Movie
Confessions of an Invisible Girl        Bruno Garotti   Brazil  2021    Movie
Sankofa Haile Gerima    United States   1993    Movie
Time taken: 0.675 seconds, Fetched: 3 row(s)
hive>
```

## 2d:

```
pes2ug20cs815@vijay-j: ~/UE20CS322-H2
pes2ug20cs815@vijay-j: ~/UE20CS322-H2   pes2ug20cs815@vijay-j: ~/UE20CS322-H2
pes2ug20cs815@vijay-j:~/UE20CS322-H2$ hdfs dfs -ls /user/hive/warehouse/netflix_partition/type=Movie
Found 1 items
-rw-r--r--   1 pes2ug20cs815 supergroup     301387 2022-09-08 18:00 /user/hive/warehouse/netflix_partition/type=Movie/000000_0
pes2ug20cs815@vijay-j:~/UE20CS322-H2$ hdfs dfs -ls /user/hive/warehouse/netflix_partition/type=TV\ Show
Found 1 items
-rw-r--r--   1 pes2ug20cs815 supergroup     117456 2022-09-08 18:03 /user/hive/warehouse/netflix_partition/type=TV Show/000000_0
pes2ug20cs815@vijay-j:~/UE20CS322-H2$
```

## 2e1:

```
hive> CREATE TABLE netflix_bucket(title String,director String,country String) PARTITIONED BY(type String) CLUSTERED BY (country) INTO 10 BUC
KETS;
OK
Time taken: 0.511 seconds
hive>
```

## 2e2:

```
pes2ug20cs815@vijay-j: ~/UE20CS322-H2
hive> insert into table netflix_bucket partition(type='Movie') select title,director,country from netflix where type='Movie';
Query ID = pes2ug20cs815_20220908181041_f3c33f9b-0957-4d0b-ba01-3dc9072437f0
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks determined at compile time: 10
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_1662639493345_0003, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0003/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 10
2022-09-08 18:11:04,572 Stage-1 map = 0%,  reduce = 0%
2022-09-08 18:11:14,122 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.6 sec
2022-09-08 18:11:50,124 Stage-1 map = 100%,  reduce = 7%, Cumulative CPU 5.29 sec
2022-09-08 18:11:55,539 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 6.65 sec
2022-09-08 18:12:00,738 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 11.03 sec
2022-09-08 18:12:56,325 Stage-1 map = 100%,  reduce = 60%, Cumulative CPU 36.09 sec
2022-09-08 18:13:41,109 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 42.98 sec
2022-09-08 18:13:42,219 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 50.28 sec
2022-09-08 18:13:52,137 Stage-1 map = 100%,  reduce = 97%, Cumulative CPU 52.76 sec
2022-09-08 18:13:57,901 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 52.76 sec
MapReduce Total cumulative CPU time: 54 seconds 40 msec
Ended Job = job_1662639493345_0003
Loading data to table default.netflix_bucket partition (type=Movie)
Launching Job 2 out of 2
Number of reduce tasks not specified. Estimated from input data size: 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_1662639493345_0004, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0004/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0004
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
```

## 2e3:

```
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-09-08 18:15:17,223 Stage-3 map = 0%,  reduce = 0%
2022-09-08 18:15:27,828 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 2.92 sec
2022-09-08 18:15:36,278 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 6.5 sec
MapReduce Total cumulative CPU time: 6 seconds 500 msec
Ended Job = job_1662639493345_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 10   Cumulative CPU: 54.04 sec   HDFS Read: 631549 HDFS Write: 289418 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 6.5 sec   HDFS Read: 31049 HDFS Write: 2464 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 0 seconds 540 msec
OK
Time taken: 304.627 seconds
hive>
```

## 2f:

```
pes2ug20cs815@vijay-j: ~/UE20CS322-H2
pes2ug20cs815@vijay-j: ~/UE20CS322-H2   pes2ug20cs815@vijay-j: ~/UE20CS322-H2
pes2ug20cs815@vijay-j:~/UE20CS322-H2$ hdfs dfs -ls /user/hive/warehouse/netflix_bucket/type=Movie
Found 10 items
-rw-r--r--   1 pes2ug20cs815 supergroup      14237 2022-09-08 18:12 /user/hive/warehouse/netflix_bucket/type=Movie/000000_0
-rw-r--r--   1 pes2ug20cs815 supergroup     113867 2022-09-08 18:12 /user/hive/warehouse/netflix_bucket/type=Movie/000001_0
-rw-r--r--   1 pes2ug20cs815 supergroup      23287 2022-09-08 18:12 /user/hive/warehouse/netflix_bucket/type=Movie/000002_0
-rw-r--r--   1 pes2ug20cs815 supergroup      10311 2022-09-08 18:12 /user/hive/warehouse/netflix_bucket/type=Movie/000003_0
-rw-r--r--   1 pes2ug20cs815 supergroup       5485 2022-09-08 18:12 /user/hive/warehouse/netflix_bucket/type=Movie/000004_0
-rw-r--r--   1 pes2ug20cs815 supergroup      46603 2022-09-08 18:12 /user/hive/warehouse/netflix_bucket/type=Movie/000005_0
-rw-r--r--   1 pes2ug20cs815 supergroup       8528 2022-09-08 18:13 /user/hive/warehouse/netflix_bucket/type=Movie/000006_0
-rw-r--r--   1 pes2ug20cs815 supergroup       9748 2022-09-08 18:13 /user/hive/warehouse/netflix_bucket/type=Movie/000007_0
-rw-r--r--   1 pes2ug20cs815 supergroup      22963 2022-09-08 18:13 /user/hive/warehouse/netflix_bucket/type=Movie/000008_0
-rw-r--r--   1 pes2ug20cs815 supergroup      17074 2022-09-08 18:13 /user/hive/warehouse/netflix_bucket/type=Movie/000009_0
pes2ug20cs815@vijay-j:~/UE20CS322-H2$
```

## 3a1:

```
pes2ug20cs815@vijay-j: ~/UE20CS322-H2
Total MapReduce CPU Time Spent: 1 minutes 0 seconds 540 msec
OK
Time taken: 304.627 seconds
hive> create table customers(customer_id int,initals String,street String,country String);
OK
Time taken: 0.745 seconds
hive> create table orders(customer_id int,order_id String,order_date date,total_cost int);
OK
Time taken: 0.379 seconds
hive> insert into customers values
    > (1,"GH","123 road","UK"),
    > (3,"JK","456 road","SP"),
    > (2,"NL","789 road","BZ"),
    > (4,"AJ","1011 road","AU"),
    > (5,"PK","1213 road","IN");
Query ID = pes2ug20cs815_20220908182159_a90ac7a2-871d-43f6-85bd-366549dd03cc
Total jobs = 3
Launching Job 1 out of 3
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_1662639493345_0005, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0005/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-08 18:22:23,441 Stage-1 map = 0%,  reduce = 0%
2022-09-08 18:22:40,635 Stage-1 map = 100%,  reduce = 0%
2022-09-08 18:22:53,313 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.83 sec
MapReduce Total cumulative CPU time: 9 seconds 830 msec
Ended Job = job_1662639493345_0005
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/customers/.hive-staging_hive_2022-09-08_18-21-59_942_8308967754115814012-1
/-ext-10000
```

# 3a2:

```
Total MapReduce CPU Time Spent: 9 seconds 830 msec
OK
Time taken: 60.497 seconds
hive> insert into orders values
    > (1,1,"2022-01-04",100),
    > (3,4,"2022-03-07",20),
    > (2,2,"2022-01-02",60),
    > (2,3,"2022-02-01",150);
Query ID = pes2ug20cs815_20220908182319_b72987d8-b88a-4da2-823a-0e4e002ae61b
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1662639493345_0006, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0006/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-09-08 18:23:44,679 Stage-1 map = 0%,  reduce = 0%
2022-09-08 18:23:54,460 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.7 sec
MapReduce Total cumulative CPU time: 3 seconds 700 msec
Ended Job = job_1662639493345_0006
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/orders/.hive-staging_hive_2022-09-08_18-23-19_644_520159898455446929 5-1/-e
xt-10000
Loading data to table default.orders
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.7 sec   HDFS Read: 6106 HDFS Write: 144 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 700 msec
OK
Time taken: 40.336 seconds
hive>
```

# 3b1:

```
hive> select customers.initals,orders.order_id,orders.total_cost from customers join orders on customers.customer_id=orders.customer_id;
Query ID = pes2ug20cs815_20220908182556_0b90c865-611b-4c23-88e8-979cdd3c5ec9
Total jobs = 1

SLF4J: Found binding in [jar:file:/home/pes2ug20cs815/hadoop-3.3.3/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLo
ggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.SLF4J: Actual binding is of type [org.apache.logging.slf4j.Lo
g4jLoggerFactory]2022-09-08 18:26:44    Starting to launch local task to process map join;      maximum memory = 239075328
2022-09-08 18:26:56     Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/pes2ug20cs815/ad8d73a3-bb6a-43e8-be78-94f6be5
bd797/hive_2022-09-08_18-25-56_091_3583215019619232418-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1662639493345_0007, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0007/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0007
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-08 18:27:21,762 Stage-3 map = 0%,  reduce = 0%
2022-09-08 18:27:56,749 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 4.29 sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_1662639493345_0007
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 4.29 sec   HDFS Read: 9621 HDFS Write: 169 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 290 msec
OK
GH      1       100
JK      4       20
NL      2       60
NL      3       150
Time taken: 127.425 seconds, Fetched: 4 row(s)
hive>
```

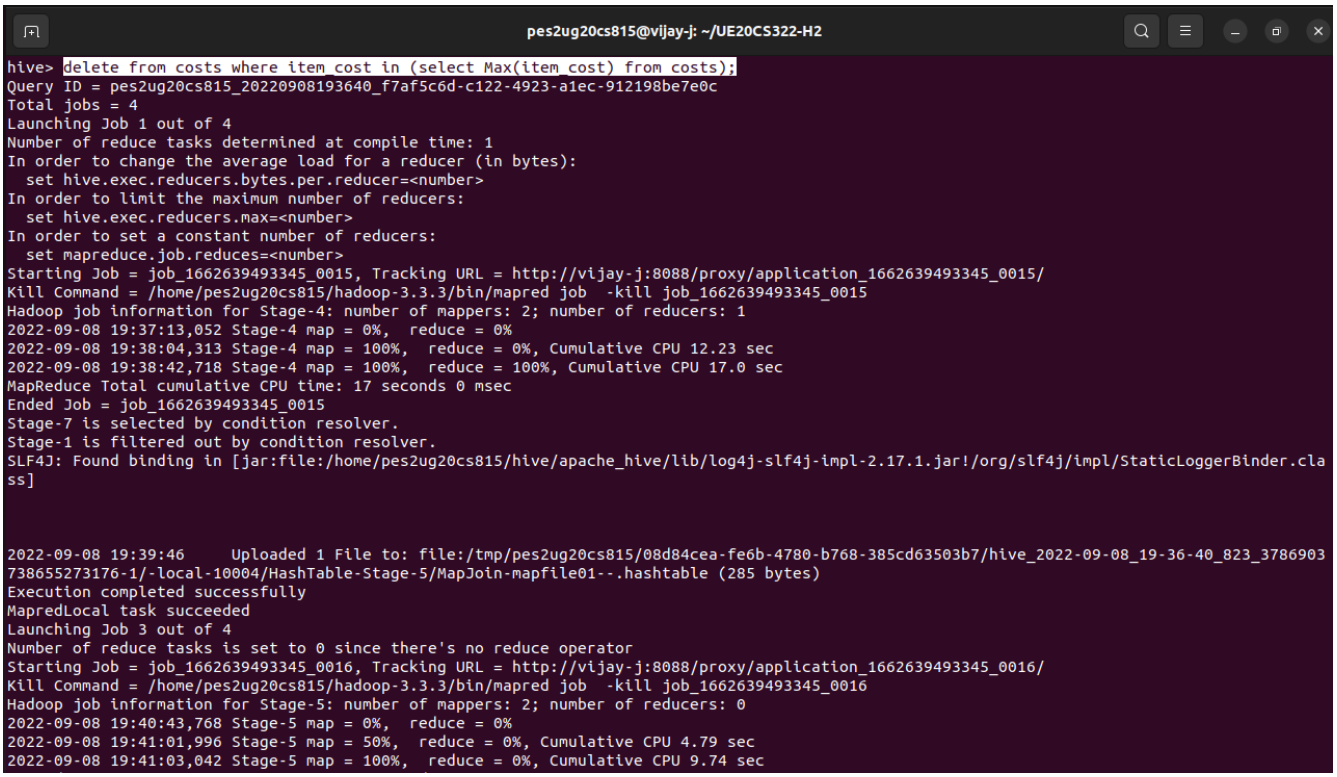
# 3c:

```
hive> set hive.auto.convert.join=true;
hive> SELECT /*+ MAPJOIN(orders) */ customers.initals,orders.order_id,orders.total_cost from customers join orders on customers.customer_id=o
rders.customer_id;
Query ID = pes2ug20cs815_20220908183145_1c178759-76d6-46f6-9a63-06a8c8567b00
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1662639493345_0008, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0008/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-08 18:32:37,224 Stage-3 map = 0%,  reduce = 0%
2022-09-08 18:32:59,652 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 3.8 sec
MapReduce Total cumulative CPU time: 3 seconds 800 msec
Ended Job = job_1662639493345_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 3.8 sec   HDFS Read: 9638 HDFS Write: 169 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 800 msec
OK
GH      1       100
JK      4       20
NL      2       60
NL      3       150
Time taken: 78.693 seconds, Fetched: 4 row(s)
hive>
```

# 4a:

```
hive> UPDATE costs SET item_cost = 30 WHERE item_name="chips";
Query ID = pes2ug20cs815_20220908193312_df14bf5f-df01-41b5-81d4-baa01ae48828
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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_1662639493345_0014, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0014/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-08 19:33:47,224 Stage-1 map = 0%,  reduce = 0%
2022-09-08 19:34:16,482 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.99 sec
2022-09-08 19:34:35,850 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.83 sec
MapReduce Total cumulative CPU time: 11 seconds 830 msec
Ended Job = job_1662639493345_0014
Loading data to table default.costs
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.83 sec   HDFS Read: 14372 HDFS Write: 1675 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 830 msec
OK
Time taken: 93.138 seconds
hive>
```

# 4b1:


pes2ug20cs815@vijay-j: ~/UE20CS322-H2
hive> delete from costs where item_cost in (select Max(item_cost) from costs);
Query ID = pes2ug20cs815_20220908193640_f7af5c6d-c122-4923-a1ec-912198be7e0c
Total jobs = 4
Launching Job 1 out of 4
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_1662639493345_0015, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0015/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job -kill job_1662639493345_0015
Hadoop job information for Stage-4: number of mappers: 2; number of reducers: 1
2022-09-08 19:37:13,052 Stage-4 map = 0%, reduce = 0%
2022-09-08 19:38:04,313 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 12.23 sec
2022-09-08 19:38:42,718 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 17.0 sec
MapReduce Total cumulative CPU time: 17 seconds 0 msec
Ended Job = job_1662639493345_0015
Stage-7 is selected by condition resolver.
Stage-1 is filtered out by condition resolver.
SLF4J: Found binding in [jar:file:/home/pes2ug20cs815/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
2022-09-08 19:39:46 Uploaded 1 File to: file:/tmp/pes2ug20cs815/08d84cea-fe6b-4780-b768-385cd63503b7/hive_2022-09-08_19-36-40_823_3786903738655273176-1/-local-10004/HashTable-Stage-5/MapJoin-mapfile01--.hashtable (285 bytes)
Execution completed successfully
MapredLocal task succeeded
Launching Job 3 out of 4
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1662639493345_0016, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0016/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job -kill job_1662639493345_0016
Hadoop job information for Stage-5: number of mappers: 2; number of reducers: 0
2022-09-08 19:40:43,768 Stage-5 map = 0%, reduce = 0%
2022-09-08 19:41:01,996 Stage-5 map = 50%, reduce = 0%, Cumulative CPU 4.79 sec
2022-09-08 19:41:03,042 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 9.74 sec

# 4b2:


hive> select * from costs;
OK
2 grape 50.0
4 oranges 80.0
5 apples 90.0
7 chocolate 90.0
10 oranges 70.0
11 apples 90.0
3 chips 30.0
6 chips 30.0
9 chips 30.0
12 chips 30.0
Time taken: 2.295 seconds, Fetched: 10 row(s)
hive>

# 4c:

```
hive> select item_name,count(*) from costs group by item_name;
Query ID = pes2ug20cs815_20220908194521_58951ffd-fdf2-4d1a-9d96-065bfea9e7d3
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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_1662639493345_0018, Tracking URL = http://vijay-j:8088/proxy/application_1662639493345_0018/
Kill Command = /home/pes2ug20cs815/hadoop-3.3.3/bin/mapred job  -kill job_1662639493345_0018
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-09-08 19:45:46,297 Stage-1 map = 0%,  reduce = 0%
2022-09-08 19:46:08,838 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.73 sec
2022-09-08 19:46:20,471 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 13.05 sec
MapReduce Total cumulative CPU time: 13 seconds 50 msec
Ended Job = job_1662639493345_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1   Cumulative CPU: 13.05 sec   HDFS Read: 25860 HDFS Write: 194 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 50 msec
OK
apples  2
chips   4
chocolate       1
grape   1
oranges 2
Time taken: 64.715 seconds, Fetched: 5 row(s)
hive>
```