

Big Data – UE20CS322

Assignment 2

September 07, 2022

Nikhil Raju Mohite

PES1UG20CS667

K section

Computer Science and Engineering

1a.png

```
pes1ug20cs667@pes1ug20cs667:~/A2/UE20CS322-H2$ source install-hive.sh
Downloading Hive
--2022-09-06 21:37:09-- https://downloads.apache.org/hive/hive-3.1.3/apache-hive-3.1.3-bin.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 135.181.214.104, 2a01:4f9:3a:2c57::2, ...
Connecting to downloads.apache.org (downloads.apache.org)[88.99.95.219]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 326940667 (312M) [application/x-gzip]
Saving to: 'apache-hive-3.1.3-bin.tar.gz'

apache-hive-3.1.3-bin.tar.gz          97%[=====] 303.79M  1.01MB/s  eta
apache-hive-3.1.3-bin.tar.gz        100%[=====] 311.79M  556KB/s   tn 6m 46s

2022-09-06 21:43:55 (787 KB/s) - 'apache-hive-3.1.3-bin.tar.gz' saved [326940667/326940667]

Unzipping the downloaded package
Editing bashrc file
HIVE_HOME
/home/pes1ug20cs667/hive/apache_hive
Do you see something like /home/pes1ug20cs999/hive/apache_hive above?[y/n]
y
You're good to go. Run start-hive.sh to start Hive. For now, wait for this process to complete for some post-installation steps.
Finished hive installation. You can use this script again anytime if the installation is broken
```

2a.png

```
mon/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = e2c09735-fd04-43ad-9021-e07f87e8d9d5

Logging initialized using configuration in jar:file:/home/pes1ug20cs667/hive/apache_hive/lib/hive-common-3.1.3.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions.
Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Hive Session ID = d7371397-ffff-479f-8b5d-d0a5c149b5f4
hive> create table netflix(show_id String,type String,title String,director String,
country String,release_year int,primary key (show_id) disable novalidate) row form
t delimited fields terminated by ',';
OK
Time taken: 2.642 seconds
hive> load data local inpath '/home/pes1ug20cs667/A2/UE20CS322-H2/netflix1.csv' int
o table netflix;
Loading data to table default.netflix
OK
Time taken: 3.141 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: 4.796 seconds, Fetched: 3 row(s)
hive>
```

2b.png

```
pes1ug20cs667@pes1ug20cs667: ~/hadoop-3.3.3/sbin

pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$ hdfs dfs -ls /user/hive/warehouse
Found 1 items
drwxr-xr-x - pes1ug20cs667 supergroup 0 2022-09-07 11:48 /user/hive/warehouse/netflix
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$ hdfs dfs -ls /user/hive/warehouse/netflix
Found 1 items
-rw-r--r-- 1 pes1ug20cs667 supergroup 541814 2022-09-07 11:48 /user/hive/warehouse/netflix/netflix1.csv
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$
```

2c1.png

```
hive> insert into table netflix_partition partition(type='Movie') select
> title,director,country,release_year from netflix where type='Movie';
Query ID = pesiug20cs667_20220907164455_d6a9e8d5-621a-40d8-9d52-3499150ecb5a
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_1662548002494_0001, Tracking URL = http://pesiug20cs667:8088/proxy/application_1662548002494_0001/
Kill Command = /home/pesiug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 16:45:39,844 Stage-1 map = 0%, reduce = 0%
2022-09-07 16:46:10,625 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 13.4 sec
2022-09-07 16:46:26,194 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 18.07 sec
MapReduce Total cumulative CPU time: 18 seconds 70 msec
Ended Job = job_1662548002494_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_2022-09-07_16-44-55_931_8883001636153270461-1/-ext-10000
Loading data to table default.netflix_partition partition (type=Movie)
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 18.07 sec HDFS Read: 559899 HDFS Write: 304492 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 70 msec
OK
Time taken: 98.568 seconds
hive>
```

2c2.png

```
hive> insert into table netflix_partition partition(type='TV Show') select
> title,director,country,release_year from netflix where type='TV Show';
Query ID = pesiug20cs667_20220907164820_5ad8bf91-e811-409f-ab75-4ad485f2c5ec
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_1662548002494_0002, Tracking URL = http://pesiug20cs667:8088/proxy/application_1662548002494_0002/
Kill Command = /home/pesiug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 16:48:54,061 Stage-1 map = 0%, reduce = 0%
2022-09-07 16:49:16,923 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.94 sec
2022-09-07 16:49:33,034 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 11.69 sec
MapReduce Total cumulative CPU time: 11 seconds 690 msec
Ended Job = job_1662548002494_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_2022-09-07_16-48-20_114_2287085315104742392-1/-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: 11.69 sec HDFS Read: 559932 HDFS Write: 120389 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 690 msec
OK
Time taken: 76.815 seconds
hive>
```

2c3.png

```
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: 0.657 seconds, Fetched: 3 row(s)
hive>
```

2d.png

```
pesiug20cs667@pesiug20cs667: ~/hadoop-3.3.3/sbin

pesiug20cs667@pesiug20cs667:~/hadoop-3.3.3/sbin$ hdfs dfs -ls /user/hive/warehouse/netflix_partition/type=Movie
Found 1 items
-rw-r--r-- 1 pesiug20cs667 supergroup 301387 2022-09-07 16:46 /user/hive/warehouse/netflix_partition/type=Movie/000000_0
pesiug20cs667@pesiug20cs667:~/hadoop-3.3.3/sbin$ hdfs dfs -ls /user/hive/warehouse/netflix_partition/type=TV\ Show
Found 1 items
-rw-r--r-- 1 pesiug20cs667 supergroup 117456 2022-09-07 16:49 /user/hive/warehouse/netflix_partition/type=TV Show/000000_0
pesiug20cs667@pesiug20cs667:~/hadoop-3.3.3/sbin$ hdfs dfs -ls /user/hive/warehouse/netflix_partition/
Found 2 items
drwxr-xr-x - pesiug20cs667 supergroup 0 2022-09-07 16:46 /user/hive/warehouse/netflix_partition/type=Movie
drwxr-xr-x - pesiug20cs667 supergroup 0 2022-09-07 16:49 /user/hive/warehouse/netflix_partition/type=TV Show
pesiug20cs667@pesiug20cs667:~/hadoop-3.3.3/sbin$
```

2e1.png

```
hive> set hive.enforce.bucketing=True;
hive> CREATE TABLE netflix_bucket(title String,director String,country String) PARTITIONED BY(type String) CLUSTERED BY (country) INTO 10 BUCKETS;
OK
Time taken: 0.401 seconds
```

2e2.png

```
hive> insert into table netflix_bucket partition(type='Movie') select
> title,director,country from netflix where type='Movie';
Query ID = pes1ug20cs667_20220907165339_009b37af-c658-4556-97ee-d86654c3e321
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_1662548002494_0003, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0003/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 10
2022-09-07 16:54:09,496 Stage-1 map = 0%, reduce = 0%
2022-09-07 16:54:21,397 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.28 sec
2022-09-07 16:54:50,274 Stage-1 map = 100%, reduce = 7%, Cumulative CPU 8.48 sec
2022-09-07 16:55:01,453 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 11.32 sec
2022-09-07 16:55:07,646 Stage-1 map = 100%, reduce = 20%, Cumulative CPU 11.32 sec
2022-09-07 16:55:14,884 Stage-1 map = 100%, reduce = 27%, Cumulative CPU 13.73 sec
2022-09-07 16:55:23,648 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 19.88 sec
2022-09-07 16:55:24,968 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 22.29 sec
2022-09-07 16:55:32,558 Stage-1 map = 100%, reduce = 43%, Cumulative CPU 28.57 sec
2022-09-07 16:55:33,682 Stage-1 map = 100%, reduce = 50%, Cumulative CPU 34.29 sec
2022-09-07 16:55:35,036 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 41.92 sec
2022-09-07 16:55:43,434 Stage-1 map = 100%, reduce = 60%, Cumulative CPU 48.67 sec
2022-09-07 16:56:39,396 Stage-1 map = 100%, reduce = 73%, Cumulative CPU 54.92 sec
2022-09-07 16:56:40,683 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 61.65 sec
2022-09-07 16:56:52,252 Stage-1 map = 100%, reduce = 97%, Cumulative CPU 64.92 sec
2022-09-07 16:56:53,438 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 66.23 sec
MapReduce Total cumulative CPU time: 1 minutes 7 seconds 550 msec
Ended Job = job_1662548002494_0003
Loading data to table default.netflix_bucket partition (type=Movie)
```

2e3.png

```
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_1662548002494_0004, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0004/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0004
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-09-07 16:58:22,648 Stage-3 map = 0%, reduce = 0%
2022-09-07 16:58:33,182 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 3.56 sec
2022-09-07 16:58:44,434 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 6.61 sec
MapReduce Total cumulative CPU time: 6 seconds 610 msec
Ended Job = job_1662548002494_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 10 Cumulative CPU: 67.55 sec HDFS Read: 631549 HDFS Write: 289418 SUCCESS
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 6.61 sec HDFS Read: 31049 HDFS Write: 2464 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 14 seconds 160 msec
OK
Time taken: 308.474 seconds
hive>
```

2f.png

```
pes1ug20cs667@pes1ug20cs667: ~/hadoop-3.3.3/sbin
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$ hdfs dfs -ls /user/hive/warehouse/netflix_bucket/type=Movie
Found 10 items
-rw-r--r-- 1 pes1ug20cs667 supergroup 14237 2022-09-07 16:55 /user/hive/warehouse/netflix_bucket/type=Movie/000000_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 113867 2022-09-07 16:55 /user/hive/warehouse/netflix_bucket/type=Movie/000001_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 23287 2022-09-07 16:55 /user/hive/warehouse/netflix_bucket/type=Movie/000002_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 10311 2022-09-07 16:55 /user/hive/warehouse/netflix_bucket/type=Movie/000003_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 5485 2022-09-07 16:55 /user/hive/warehouse/netflix_bucket/type=Movie/000004_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 46603 2022-09-07 16:55 /user/hive/warehouse/netflix_bucket/type=Movie/000005_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 8528 2022-09-07 16:56 /user/hive/warehouse/netflix_bucket/type=Movie/000006_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 9748 2022-09-07 16:57 /user/hive/warehouse/netflix_bucket/type=Movie/000007_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 22963 2022-09-07 16:56 /user/hive/warehouse/netflix_bucket/type=Movie/000008_0
-rw-r--r-- 1 pes1ug20cs667 supergroup 17074 2022-09-07 16:56 /user/hive/warehouse/netflix_bucket/type=Movie/000009_0
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$
```

TASK2

3a1.png

```
hive> create table customers(customer_id int,initials String,street
> String,country String);
OK
Time taken: 0.76 seconds
hive> create table orders(customer_id int,order_id String,order_date
> date,total_cost int);
OK
Time taken: 0.22 seconds
hive> insert into customers values
> (1,"GM","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 = pes1ug20cs667_20220907170144_0ae2f78d-fa74-4865-be24-c28612bc0fce
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_1662548002494_0005, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0005/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 17:02:07,404 Stage-1 map = 0%, reduce = 0%
2022-09-07 17:02:17,982 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.16 sec
2022-09-07 17:02:27,456 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.74 sec
MapReduce Total cumulative CPU time: 7 seconds 740 msec
Ended Job = job_1662548002494_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_2022-09-07_17-01-44_780_4980136662364088513-1/-ext-10000
Loading data to table default.customers
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.74 sec HDFS Read: 18510 HDFS Write: 476 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 740 msec
OK
Time taken: 48.355 seconds
hive>
```

3a2.png

```
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 = pes1ug20cs667_20220907170357_2f409cab-e3b6-4145-84af-bdb4deef4f91
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_1662548002494_0006, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0006/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-09-07 17:04:14,456 Stage-1 map = 0%, reduce = 0%
2022-09-07 17:04:23,363 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.7 sec
MapReduce Total cumulative CPU time: 2 seconds 700 msec
Ended Job = job_1662548002494_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_2022-09-07_17-03-57_684_3178791715676340764-1/-ext-10000
Loading data to table default.orders
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 2.7 sec HDFS Read: 6126 HDFS Write: 144 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 700 msec
OK
Time taken: 28.559 seconds
hive>
```

3b.png

```
hive> select customers.initials,orders.order_id,orders.total_cost from
> customers join orders on customers.customer_id=orders.customer_id;
Query ID = pesiug20cs667_20220907170535_794d6298-ca17-4001-8d96-6d394fc7e120
Total jobs = 1
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hadoop-3.3.3/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
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_1662548002494_0007, Tracking URL = http://pesiug20cs667:8088/proxy/application_1662548002494_0007/
Kill Command = /home/pesiug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0007
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-07 17:06:45,736 Stage-3 map = 0%, reduce = 0%
2022-09-07 17:06:55,084 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 4.65 sec
MapReduce Total cumulative CPU time: 4 seconds 650 msec
Ended Job = Job_1662548002494_0007
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 4.65 sec HDFS Read: 9621 HDFS Write: 169 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 650 msec
OK
GH 1 100
JK 4 20
NL 2 60
NL 3 150
Time taken: 82.73 seconds, Fetched: 4 row(s)
hive>
```

3c.png

```
hive> set hive.auto.convert.join=true;
hive> SELECT /*+ MAPJOIN(orders) */
> customers.initials,orders.order_id,orders.total_cost from customers join
> orders on customers.customer_id=orders.customer_id;
Query ID = pesiug20cs667_20220907170859_43349bc2-b0c1-4696-b281-7a5a2c8a5861
Total jobs = 1
2022-09-07 17:09:13 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/pesiug20cs667/54010fd7-0ffe-4e5f-a4e9-756dd5895a5f/hive_2022-09-07_17-09-00_132_50730889920234602528-1/-locat
l-100004/HashTable-Stage-3/MapJoin-mapfile11...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_1662548002494_0008, Tracking URL = http://pesiug20cs667:8088/proxy/application_1662548002494_0008/
Kill Command = /home/pesiug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-07 17:09:27,179 Stage-3 map = 0%, reduce = 0%
2022-09-07 17:09:33,594 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.92 sec
MapReduce Total cumulative CPU time: 2 seconds 920 msec
Ended Job = Job_1662548002494_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.92 sec HDFS Read: 9638 HDFS Write: 169 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 920 msec
OK
GH 1 100
JK 4 20
NL 2 60
NL 3 150
Time taken: 36.819 seconds, Fetched: 4 row(s)
hive>
```

TASK 3

4a1.png

```
hive> select * from costs;
OK
1      chocolate      100.0
2      grape          50.0
3      chips          10.0
4      oranges        80.0
5      apples         90.0
6      chips          20.0
7      chocolate      90.0
8      grape          100.0
9      chips          40.0
10     oranges        70.0
11     apples         90.0
12     chips          20.0
Time taken: 0.368 seconds, Fetched: 12 row(s)
hive> UPDATE costs
> SET item_cost=30
> WHERE item_name="chips";
Query ID = pes1ug20cs667_20220907180026_604cfa55-0c47-4f02-bb8e-9b9afeea6d95
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_1662548002494_0020, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0020/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 18:00:37,011 Stage-1 map = 0%, reduce = 0%
2022-09-07 18:00:45,911 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.74 sec
2022-09-07 18:00:54,772 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.65 sec
MapReduce Total cumulative CPU time: 7 seconds 650 msec
Ended Job = job_1662548002494_0020
Loading data to table default.costs
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.65 sec HDFS Read: 14375 HDFS Write: 1675 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 650 msec
OK
Time taken: 30.454 seconds
```

4a2.png

```
hive> UPDATE costs
> SET item_cost=30
> WHERE item_name="chips";
Query ID = pes1ug20cs667_20220907180026_604cfa55-0c47-4f02-bb8e-9b9afeea6d95
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_1662548002494_0020, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0020/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 18:00:37,011 Stage-1 map = 0%, reduce = 0%
2022-09-07 18:00:45,911 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.74 sec
2022-09-07 18:00:54,772 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.65 sec
MapReduce Total cumulative CPU time: 7 seconds 650 msec
Ended Job = job_1662548002494_0020
Loading data to table default.costs
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.65 sec HDFS Read: 14375 HDFS Write: 1675 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 650 msec
OK
Time taken: 30.454 seconds
hive> select * from costs;
OK
1      chocolate      100.0
2      grape          50.0
4      oranges        80.0
5      apples         90.0
7      chocolate      90.0
8      grape          100.0
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: 18.892 seconds, Fetched: 12 row(s)
hive> █
```


4b1.png

```
hive> select * from costs;
OK
1      chocolate      100.0
2      grape      50.0
4      oranges      80.0
5      apples      90.0
7      chocolate      90.0
8      grape      100.0
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: 0.212 seconds, Fetched: 12 row(s)
hive> DELETE FROM costs WHERE item_cost IN (select max(item_cost) from costs);
Query ID = pes1ug20cs667_20220907180348_b276dc9c-2f6f-4596-ac6a-c86d7fcb9769
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_1662548002494_0021, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0021/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0021
Hadoop job information for Stage-4: number of mappers: 2; number of reducers: 1
2022-09-07 18:03:58,633 Stage-4 map = 0%, reduce = 0%
2022-09-07 18:04:15,927 Stage-4 map = 50%, reduce = 0%, Cumulative CPU 5.27 sec
2022-09-07 18:04:16,976 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 12.75 sec
2022-09-07 18:04:24,336 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 15.74 sec
MapReduce Total cumulative CPU time: 15 seconds 740 msec
Ended Job = job_1662548002494_0021
Stage-7 is selected by condition resolver.
Stage-1 is filtered out by condition resolver.
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_1662548002494_0022, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0022/
```

4b2.png

```
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0022
Hadoop job information for Stage-5: number of mappers: 2; number of reducers: 0
2022-09-07 18:05:01,948 Stage-5 map = 0%, reduce = 0%
2022-09-07 18:05:14,047 Stage-5 map = 50%, reduce = 0%, Cumulative CPU 4.0 sec
2022-09-07 18:05:15,087 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 7.51 sec
MapReduce Total cumulative CPU time: 7 seconds 510 msec
Ended Job = job_1662548002494_0022
Launching Job 4 out of 4
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_1662548002494_0023, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0023/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0023
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-07 18:05:31,798 Stage-2 map = 0%, reduce = 0%
2022-09-07 18:05:39,386 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.39 sec
2022-09-07 18:05:45,771 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.8 sec
MapReduce Total cumulative CPU time: 4 seconds 800 msec
Ended Job = job_1662548002494_0023
Loading data to table default.costs
MapReduce Jobs Launched:
Stage-Stage-4: Map: 2 Reduce: 1 Cumulative CPU: 15.74 sec HDFS Read: 23229 HDFS Write: 121 SUCCESS
Stage-Stage-5: Map: 2 Cumulative CPU: 7.51 sec HDFS Read: 19969 HDFS Write: 250 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.8 sec HDFS Read: 9590 HDFS Write: 834 SUCCESS
Total MapReduce CPU Time Spent: 28 seconds 50 msec
OK
Time taken: 119.566 seconds
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: 0.861 seconds, Fetched: 10 row(s)
hive>
```


4c.png

```
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: 0.377 seconds, Fetched: 10 row(s)
hive> select item_name, count(*) as getcount FROM costs GROUP BY item_name;
Query ID = pes1ug20cs667_20220907182404_47069537-59c6-4b3f-875e-48402a163ce1
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_1662548002494_0026, Tracking URL = http://pes1ug20cs667:8088/proxy/application_1662548002494_0026/
Kill Command = /home/pes1ug20cs667/hadoop-3.3.3/bin/mapred job -kill job_1662548002494_0026
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-09-07 18:24:14,336 Stage-1 map = 0%, reduce = 0%
2022-09-07 18:24:21,634 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.8 sec
2022-09-07 18:24:26,825 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.42 sec
MapReduce Total cumulative CPU time: 5 seconds 420 msec
Ended Job = job_1662548002494_0026
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 5.42 sec HDFS Read: 26100 HDFS Write: 194 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 420 msec
OK
apples 2
chips 4
chocolate 1
grape 1
oranges 2
Time taken: 25.637 seconds, Fetched: 5 row(s)
hive>
```

Number of mappers = 2

Number of reducers = 1

Output:

5a1.png

```
pesiug20cs667@pesiug20cs667:~/A2/UE20CS322-H2$ python3 eval-rr.pyc
Starting H2 evaluation...
Setting up...
Verifying output...

Verifying Task 1...
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hadoop-3.3.3/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = f59c0bb1-33cc-4b9d-bd21-80411cba54b1
Hive Session ID = 29b2aacf-f5e4-4e5f-b4cb-f9a1dc05bdc9

Task 1 completed successfully

Verifying Task 2...
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hadoop-3.3.3/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = f9843bc0-b5af-447f-9e4f-82d965d43b87
Hive Session ID = 70f90c06-8c0c-4f59-85bd-f131ee578b0c

Task 2 completed successfully
```

5a2.png

```
Verifying Task 3...
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hadoop-3.3.3/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = df4df532-4fd9-421c-9b1f-103a526ed5e1
Hive Session ID = b9046ac2-01d2-4d53-9304-d492cia63568
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pesiug20cs667/hadoop-3.3.3/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = 5320f484-3ee2-420b-b85f-44cad9e049ce
Hive Session ID = 12e89c6f-4eed-450b-be05-dd4ad8604edc

Task 3 completed successfully

Sending result to server...
Server response: Your Hive evaluation is complete. You scored 4 marks out of 4. You can run the pyc file again to get a better score.

Auto-evaluation concluded.
pesiug20cs667@pesiug20cs667:~/A2/UE20CS322-H2$
```