

1a1.

```
pes2ug20cs389@pes2ug20cs389-VirtualBox:~/UE20CS322-H2$ source install-hive.sh
Downloading Hive
--2022-09-06 08:34:10-- 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, 2a01:4f8:10a:201a::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. 100%[=====] 311.79M 13.1MB/s in 2m 26s

2022-09-06 08:36:37 (2.13 MB/s) - 'apache-hive-3.1.3-bin.tar.gz' saved [326940667/326940667]

Unzipping the downloaded package
Editing bashrc file
HIVE_HOME
/home/pes2ug20cs389/hive/apache_hive
Do you see something like /home/peslug20cs999/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.
Starting Hadoop
WARNING: Attempting to start all Apache Hadoop daemons as pes2ug20cs389 in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
```

1a2.

```
Starting datanodes
Starting secondary namenodes [pes2ug20cs389-VirtualBox]
Starting resourcemanager
Starting nodemanagers
mkdir: Cannot create directory /root/hive/warehouse. Name node is in safe mode.
Finished hive installation. You can use this script again anytime if the installation is broken
```

2a1.

```
Initialization script completed
schemaTool completed
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/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 = 8e64e46b-c957-4e49-a10e-3a3a93707bea

Logging initialized using configuration in jar:file:/home/pes2ug20cs389/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 = ba36e248-50ef-4e48-975a-c5ed8c9444d5
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 format delimited fields terminated by ',';
OK
Time taken: 1.263 seconds
hive> load data local inpath '/home/pes2ug20cs389/UE20CS322-H2/netflix1.csv' into table netflix;
Loading data to table default.netflix
OK
Time taken: 1.391 seconds
hive> select * from netflix limit 3;
```

2a2.

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

2b.

```

pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ hdfs dfs -ls /user/hive/warehouse
Found 1 items
drwxr-xr-x - pes2ug20cs389 supergroup          0 2022-09-07 12:28 /user/hive/warehouse/netflix
pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ hdfs dfs -ls /user/hive/warehouse/netflix
Found 1 items
-rw-r--r-- 1 pes2ug20cs389 supergroup      541814 2022-09-07 12:28 /user/hive/warehouse/netflix/netflix1.csv
pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ 

```

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: 0.107 seconds
hive> insert into table netflix_partition partition(type='Movie') select title,director,country,release_year from netflix where type='Movie';
Query ID = pes2ug20cs389_20220907123649_fd8cb993-b50e-47f8-83d5-0419dfb9040
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_1662533681004_0001, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0001/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 12:37:05,009 Stage-1 map = 0%,  reduce = 0%
2022-09-07 12:37:13,512 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.0 sec
2022-09-07 12:37:19,845 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.78 sec
MapReduce Total cumulative CPU time: 5 seconds 780 msec

```

2c2.

```

Ended Job = job_1662533681004_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_12-36-49_958_8345390872263355214-1/-ext-10000
Loading data to table default.netflix_partition partition (type=Movie)
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 5.78 sec   HDFS Read: 559899 HDFS Write: 304492 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 780 msec
OK
Time taken: 31.736 seconds
hive> insert into table netflix_partition partition(type='TV Show') select title,director,country,release_year from netflix where type='TV Show';
Query ID = pes2ug20cs389_20220907123741_e659ae53-9577-4905-bd71-f08d3ec5a34d
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_1662533681004_0002, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0002/

```

2c3.

```

Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 12:37:51,023 Stage-1 map = 0%, reduce = 0%
2022-09-07 12:37:57,655 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.78 sec
2022-09-07 12:38:04,662 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.53 sec
MapReduce Total cumulative CPU time: 4 seconds 530 msec
Ended Job = job_1662533681004_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_12-37-41_615_571782965565249777-1/-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: 4.53 sec HDFS Read: 559933 HDFS Write: 120389 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 530 msec
OK
Time taken: 24.648 seconds
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: 0.273 seconds, Fetched: 3 row(s)
hive> █

```

2d.

```

pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ hdfs dfs -ls /user/hive/warehouse/netflix_partition/type=Movie
Found 1 items
-rw-r--r-- 1 pes2ug20cs389 supergroup 301387 2022-09-07 12:37 /user/hive/warehouse/netflix_partition/type=Movie/00000_0
pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ hdfs dfs -ls /user/hive/warehouse/netflix_partition/type=TV\ Show
Found 1 items
-rw-r--r-- 1 pes2ug20cs389 supergroup 117456 2022-09-07 12:37 /user/hive/warehouse/netflix_partition/type=TV Show/00000_0
pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ hdfs dfs -ls /user/hive/warehouse/netflix_partition
Found 2 items
drwxr-xr-x - pes2ug20cs389 supergroup 0 2022-09-07 12:37 /user/hive/warehouse/netflix_partition/type=Movie
drwxr-xr-x - pes2ug20cs389 supergroup 0 2022-09-07 12:38 /user/hive/warehouse/netflix_partition/type=TV Show
pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ █

```

2e1.

```

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: 0.131 seconds
hive> insert into table netflix_bucket partition(type='Movie') select title,director,country from netflix where type='Movie';
Query ID = pes2ug20cs389_20220907124939_595fe0ca-ccc0-46b9-9bf3-179550bd9f0e
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_1662533681004_0003, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0003/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 10
2022-09-07 12:49:48,585 Stage-1 map = 0%, reduce = 0%
2022-09-07 12:49:55,948 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.05 sec
2022-09-07 12:50:10,242 Stage-1 map = 100%, reduce = 10%, Cumulative CPU 5.46 sec
2022-09-07 12:50:14,801 Stage-1 map = 100%, reduce = 20%, Cumulative CPU 8.61 sec
2022-09-07 12:50:19,570 Stage-1 map = 100%, reduce = 30%, Cumulative CPU 11.04 sec
2022-09-07 12:50:23,250 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 15.04 sec

```

2e2.

```

2022-09-07 12:50:14,801 Stage-1 map = 100%, reduce = 20%, Cumulative CPU 8.61 sec
2022-09-07 12:50:19,570 Stage-1 map = 100%, reduce = 30%, Cumulative CPU 11.04 sec
2022-09-07 12:50:21,958 Stage-1 map = 100%, reduce = 50%, Cumulative CPU 15.04 sec
2022-09-07 12:50:23,252 Stage-1 map = 100%, reduce = 60%, Cumulative CPU 17.46 sec
2022-09-07 12:50:28,165 Stage-1 map = 100%, reduce = 70%, Cumulative CPU 19.75 sec
2022-09-07 12:50:30,412 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 21.69 sec
2022-09-07 12:50:32,533 Stage-1 map = 100%, reduce = 90%, Cumulative CPU 24.1 sec
2022-09-07 12:50:33,570 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 26.52 sec
MapReduce Total cumulative CPU time: 26 seconds 520 msec
Ended Job = job_1662533681004_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=<nnumber>
Starting Job = job_1662533681004_0004, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0004/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0004
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-09-07 12:50:47,113 Stage-3 map = 0%, reduce = 0%
2022-09-07 12:50:53,396 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.23 sec
2022-09-07 12:50:58,672 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 3.07 sec

```

2e3.

```

2022-09-07 12:50:58,672 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 3.07 sec
MapReduce Total cumulative CPU time: 3 seconds 70 msec
Ended Job = job_1662533681004_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 10 Cumulative CPU: 26.52 sec HDFS Read: 631549 HDFS Write: 289418 SUCCESS
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 3.07 sec HDFS Read: 31049 HDFS Write: 2464 SUCCESS
Total MapReduce CPU Time Spent: 29 seconds 590 msec
OK
Time taken: 80.297 seconds
hive> 
```

2f.

```

pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ hdfs dfs -ls /user/hive/warehouse/netflix_bucket/type=Movie
Found 10 items
-rw-r--r-- 1 pes2ug20cs389 supergroup 14237 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000000
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 113867 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000001
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 23287 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000002
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 10311 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000003
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 5485 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000004
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 46603 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000005
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 8528 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000006
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 9748 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000007
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 22963 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000008
  0
-rw-r--r-- 1 pes2ug20cs389 supergroup 17074 2022-09-07 12:50 /user/hive/warehouse/netflix_bucket/type=Movie/000009
  0
pes2ug20cs389@pes2ug20cs389-VirtualBox:~$ 
```

3a1.

```

hive> create table customers(customer_id int,initials String,street String,country String);
OK
Time taken: 0.087 seconds
hive> create table orders(customer_id int,order_id String,order_date date,total_cost int);
OK
Time taken: 0.093 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 = pes2ug20cs389_20220907125528_31dd17ab-466e-459d-8eee-ddb2ee288c47
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_1662533681004_0005, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0005/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

```

3a2 .

```

Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 12:55:38,577 Stage-1 map = 0%, reduce = 0%
2022-09-07 12:55:45,101 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.35 sec
2022-09-07 12:55:51,354 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.22 sec
MapReduce Total cumulative CPU time: 4 seconds 220 msec
Ended Job = job_1662533681004_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_12-55-28_938_7
794748012251821186-1/-ext-10000
Loading data to table default.customers
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.22 sec HDFS Read: 18510 HDFS Write: 476 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 220 msec
OK
Time taken: 23.961 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 = pes2ug20cs389_20220907125556_bf7cf030-5003-48ff-bd21-d628255f3e03
Total jobs = 3
Launching Job 1 out of 3

```

3a3 .

```

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_1662533681004_0006, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0006/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-09-07 12:56:04,874 Stage-1 map = 0%, reduce = 0%
2022-09-07 12:56:11,144 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.71 sec
MapReduce Total cumulative CPU time: 1 seconds 710 msec
Ended Job = job_1662533681004_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_12-55-56_499_8567
420446078416402-1/-ext-10000
Loading data to table default.orders
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 1.71 sec HDFS Read: 6106 HDFS Write: 144 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 710 msec
OK
Time taken: 15.93 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 = pes2ug20cs389_20220907125759_c5f2014c-b4f8-4193-9ab9-a1cb3feb49c5
Total jobs = 1
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/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_1662533681004_0007, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0007/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0007
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-07 12:58:15,775 Stage-3 map = 0%, reduce = 0%
2022-09-07 12:58:22,012 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.21 sec
MapReduce Total cumulative CPU time: 2 seconds 210 msec
Ended Job = job_1662533681004_0007
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.21 sec HDFS Read: 9621 HDFS Write: 169 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 210 msec

```

3b2 .

```

MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.21 sec HDFS Read: 9621 HDFS Write: 169 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 210 msec
OK
GH      1      100
JK      4      20
NL      2      60
NL      3     150
Time taken: 24.023 seconds, Fetched: 4 row(s)
hive> █

```

3c .

```

hive> SELECT /*+ MAPJOIN(orders) */ customers.initals,orders.order_id,orders.total_cost from customers join orders on customers.customer_id=orders.customer_id;
Query ID = pes2ug20cs389_20220907130005_d31bc927-8cff-40ab-9d52-8760451f161b
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_1662533681004_0008, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0008/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-09-07 13:00:21,815 Stage-3 map = 0%, reduce = 0%
2022-09-07 13:00:28,164 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.97 sec
MapReduce Total cumulative CPU time: 1 seconds 970 msec
Ended Job = job_1662533681004_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.97 sec HDFS Read: 9638 HDFS Write: 169 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 970 msec
OK
GH      1      100
JK      4      20
NL      2      60
NL      3     150
Time taken: 23.571 seconds, Fetched: 4 row(s)
hive> █

```

4a .

```

hive> UPDATE costs SET item_cost = 30 WHERE item_name = "chips";
Query ID = pes2ug20cs389_20220907133242_8390a7d7-5bf0-449d-9f97-f7061475a461
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_1662533681004_0010, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0010/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-07 13:32:49,375 Stage-1 map = 0%, reduce = 0%
2022-09-07 13:32:54,555 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.91 sec
2022-09-07 13:33:00,759 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.43 sec
MapReduce Total cumulative CPU time: 3 seconds 430 msec
Ended Job = job_1662533681004_0010
Loading data to table default.costs
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 3.43 sec  HDFS Read: 14375 HDFS Write: 1675 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 430 msec
OK
Time taken: 19.228 seconds

```

4b1.

```

hive> DELETE FROM costs WHERE item_cost IN (SELECT MAX(item_cost) FROM costs);
Query ID = pes2ug20cs389_20220907155657_f000a3da-6e23-4dd1-81cf-df7c1890bcb9
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_1662533681004_0014, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0014/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0014
Hadoop job information for Stage-4: number of mappers: 2; number of reducers: 1
2022-09-07 15:57:05,025 Stage-4 map = 0%, reduce = 0%
2022-09-07 15:57:12,703 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 3.62 sec
2022-09-07 15:57:17,836 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 5.14 sec
MapReduce Total cumulative CPU time: 5 seconds 140 msec
Ended Job = job_1662533681004_0014
Stage-7 is selected by condition resolver.
Stage-1 is filtered out by condition resolver.
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
Execution completed successfully
MapredLocal task succeeded

```

4b2.

```

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_1662533681004_0015, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0015/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0015
Hadoop job information for Stage-5: number of mappers: 2; number of reducers: 0
2022-09-07 15:57:31,276 Stage-5 map = 0%, reduce = 0%
2022-09-07 15:57:38,884 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 4.03 sec
MapReduce Total cumulative CPU time: 4 seconds 30 msec
Ended Job = job_1662533681004_0015
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_1662533681004_0016, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0016/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0016

```

4b3.

```

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-07 15:57:51,488 Stage-2 map = 0%,  reduce = 0%
2022-09-07 15:57:56,609 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.37 sec
2022-09-07 15:58:01,781 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.89 sec
MapReduce Total cumulative CPU time: 2 seconds 890 msec
Ended Job = job_1662533681004_0016
Loading data to table default.costs
MapReduce Jobs Launched:
Stage-Stage-4: Map: 2  Reduce: 1  Cumulative CPU: 5.14 sec  HDFS Read: 23230 HDFS Write: 121 SUCCESS
Stage-Stage-5: Map: 2  Cumulative CPU: 4.03 sec  HDFS Read: 19977 HDFS Write: 250 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 2.89 sec  HDFS Read: 9597 HDFS Write: 834 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 60 msec
OK
Time taken: 65.292 seconds
hive> 
```

4c1.

```

hive> SELECT item_name, COUNT(*) FROM costs GROUP BY item_name;
Query ID = pes2ug20cs389_20220907161804_8a9b480d-b874-4ca8-8683-684d7ac28a3d
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_1662533681004_0018, Tracking URL = http://pes2ug20cs389-VirtualBox:8088/proxy/application_1662533681004_0018/
Kill Command = /home/pes2ug20cs389/hadoop-3.3.3/bin/mapred job -kill job_1662533681004_0018
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2022-09-07 16:18:10,628 Stage-1 map = 0%,  reduce = 0%
2022-09-07 16:18:16,993 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.17 sec
2022-09-07 16:18:18,025 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.88 sec
2022-09-07 16:18:22,141 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.19 sec
MapReduce Total cumulative CPU time: 4 seconds 190 msec
Ended Job = job_1662533681004_0018
hive> 
```

4c2.

```

MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 4.19 sec  HDFS Read: 26100 HDFS Write: 194 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 190 msec
OK
apples 2
chips 4
chocolate      1
grape 1
oranges 2
Time taken: 18.546 seconds, Fetched: 5 row(s)
hive> 
```

5a.

```

pes2ug20cs389@pes2ug20cs389-VirtualBox:~/UE20CS322-H2$ python3 eval-ec.py
Starting H2 evaluation...
Setting up...
Verifying output...

Verifying Task 1...
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/:org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/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 = 6ad72ec8-1505-4628-8cc3-e0b77df5ca17
Hive Session ID = b86193e7-10de-4330-9cd5-9430ca4790dd

Task 1 completed successfully

Verifying Task 2...
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/:org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hadoop-3.3.3/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/:org/slf4j/impl/StaticLoggerBinder.class] 
```

```
aticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/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.
```

```
Task 2 completed successfully
```

```
Verifying Task 3...
```

```
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/*org/slf4j/impl/St
aticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/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 = 14b1bff7-6da9-4063-898b-faa62f664238
Hive Session ID = e20d1f98-548e-4d67-9103-38a4d1a04731
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/*org/slf4j/impl/St
aticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/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 = 14b1bff7-6da9-4063-898b-faa62f664238
```

```
Hive Session ID = e20d1f98-548e-4d67-9103-38a4d1a04731
```

```
SLF4J: Class path contains multiple SLF4J bindings.
```

```
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/hive/apache_hive/lib/log4j-slf4j-impl-2.17.1.jar!/*org/slf4j/impl/St
aticLoggerBinder.class]
```

```
SLF4J: Found binding in [jar:file:/home/pes2ug20cs389/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 = 25a78ff9-0d5b-431a-a90e-8b363469750d
```

```
Hive Session ID = af4bbba4-e98b-485f-9b12-f8bf6c36eb9d
```

```
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.
```