BIG DATA LAB HIVE ASSIGNMENT 1

Name : Asha Venugopal Madguni

USN : 1NT18IS033

Sem and Sec : 6-'A' (A2 Batch)

- **Create database called customer**

```
hive> create database Customer;
OK
Time taken: 5.329 seconds
hive> use Customer;
OK
Time taken: 0.13 seconds
hive> create table bank(tid int,bname string,cname string,amount float)row format delimited fields terminated by ",";
OK
Time taken: 3.487 seconds
hive> insert into bank values(101,"SBI","Asha",50000.00),(102,"Axis","Ananya",30000.00),(103,"BOB","Goutam",80000.00),(104,"Canara","Nandita",60000.00),(105,"KDCC","Pavitra",40000.00);
Query ID = hdoop_20210707123059_e8e8be0a-ed84-4746-8264-0dfa82d52639
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):
```

- **Create table bank (tid, bname, cname, amount)**

```
hive> create table bank(tid int,bname string,cname string,amount float)row format delimited fields terminated by ",";
OK
Time taken: 3.487 seconds
hive> insert into bank values(101,"SBI","Asha",50000.00),(102,"Axis","Ananya",30000.00),(103,"BOB","Goutam",80000.00),(104,"Canara","Nandita",60000.00),(105,"KDCC","Pavitra",40000.00);
Query ID = hdoop_20210707123059_e8e8be0a-ed84-4746-8264-0dfa82d52639
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):
```
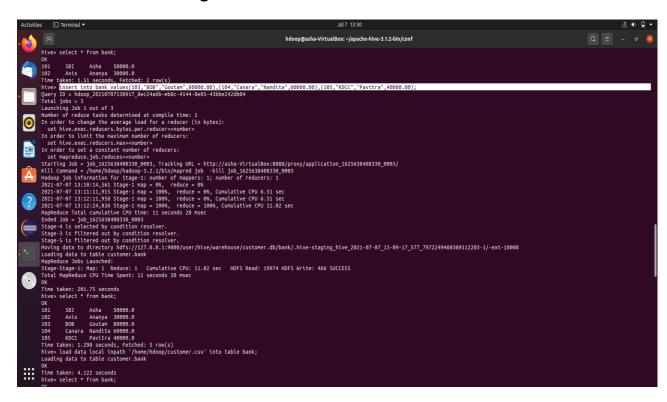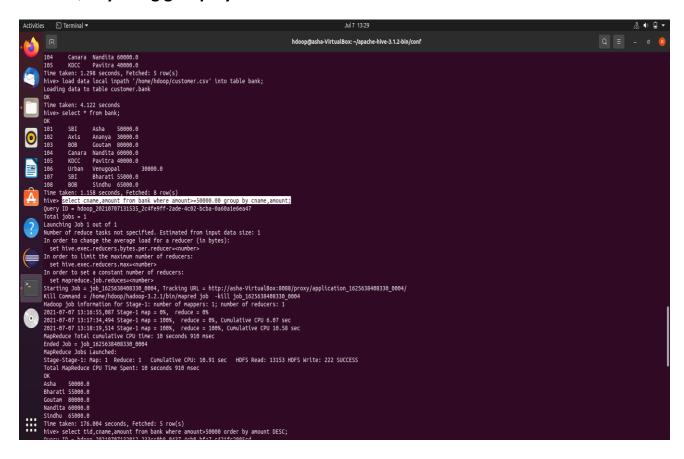
- **Insert values using insert statement**

- **Import the values using csv**

- **Query using group by clause**


```
104    Canara  Nandita 60000.0
105    KDCC    Pavitra 40000.0
Time taken: 1.298 seconds, Fetched: 5 row(s)
hive> load data local inpath '/home/hdoop/customer.csv' into table bank;
Loading data to table customer.bank
OK
Time taken: 4.122 seconds
hive> select * from bank;
OK
101    SBI     Asha    50000.0
102    Axis    Ananya  30000.0
103    BOB     Goutam  80000.0
104    Canara  Nandita 60000.0
105    KDCC    Pavitra 40000.0
106    Urban   Venugopal       30000.0
107    SBI     Bharati 55000.0
108    BOB     Sindhu  65000.0
Time taken: 1.158 seconds, Fetched: 8 row(s)
hive> select cname,amount from bank where amount>=50000.00 group by cname,amount;
Query ID = hdoop_20210707131535_2c4fe9ff-2ade-4c02-bcba-0a60a1e6ea47
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_1625638408330_0004, Tracking URL = http://asha-VirtualBox:8088/proxy/application_1625638408330_0004/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1625638408330_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-07 13:16:55,087 Stage-1 map = 0%,  reduce = 0%
2021-07-07 13:17:34,494 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.07 sec
2021-07-07 13:18:19,514 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.58 sec
MapReduce Total cumulative CPU time: 10 seconds 910 msec
Ended Job = job_1625638408330_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.91 sec   HDFS Read: 13153 HDFS Write: 222 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 910 msec
OK
Asha    50000.0
Bharati 55000.0
Goutam  80000.0
Nandita 60000.0
Sindhu  65000.0
Time taken: 176.004 seconds, Fetched: 5 row(s)
hive> select tid,cname,amount from bank where amount>50000 order by amount DESC;
Query ID = hdoop_20210707132012_233cc0b8_0437_4cb0_bfa7_c421fc2005cd
```

- **Query using order by clause**

```
Time taken: 176.004 seconds, Fetched: 5 row(s)
hive> select tid,cname,amount from bank where amount>50000 order by amount DESC;
Query ID = hdoop_20210707132012_233cc0b8-0437-4cb0-bfa7-c421fc2005cd
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_1625638408330_0005, Tracking URL = http://asha-VirtualBox:8088/proxy/application_1625638408330_0005/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1625638408330_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-07 13:21:14,254 Stage-1 map = 0%,  reduce = 0%
2021-07-07 13:22:03,577 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.88 sec
2021-07-07 13:23:04,296 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.88 sec
2021-07-07 13:23:08,538 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.68 sec
MapReduce Total cumulative CPU time: 9 seconds 680 msec
Ended Job = job_1625638408330_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.68 sec   HDFS Read: 12597 HDFS Write: 213 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 680 msec
OK
103     Goutam  80000.0
108     Sindhu  65000.0
104     Nandita 60000.0
107     Bharati 55000.0
Time taken: 185.854 seconds, Fetched: 4 row(s)
```
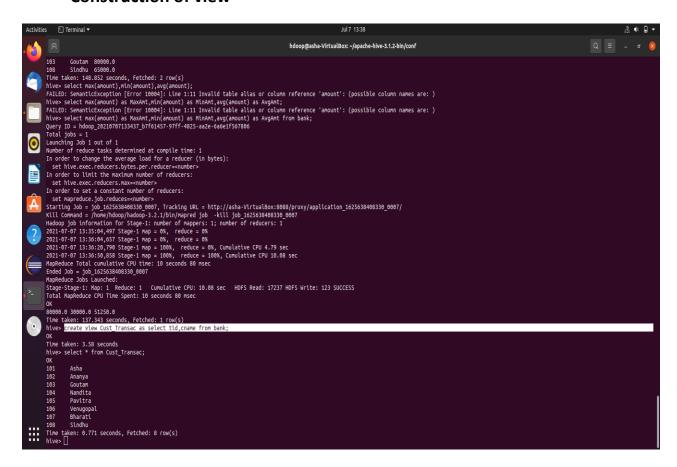
- **Query using having clause**

```
Time taken: 185.854 seconds, Fetched: 4 row(s)
hive> select tid,cname,sum(amount) from bank group by tid,cname having sum(amount)>60000;
Query ID = hdoop_20210707132606_5ec869a4-a0cd-4317-bf6b-667e54bd131b
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_1625638408330_0006, Tracking URL = http://asha-VirtualBox:8088/proxy/application_1625638408330_0006/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1625638408330_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-07 13:26:53,668 Stage-1 map = 0%,  reduce = 0%
2021-07-07 13:27:30,953 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.59 sec
2021-07-07 13:28:30,035 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.23 sec
MapReduce Total cumulative CPU time: 10 seconds 230 msec
Ended Job = job_1625638408330_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.23 sec   HDFS Read: 14852 HDFS Write: 149 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 230 msec
OK
103     Goutam  80000.0
108     Sindhu  65000.0
Time taken: 148.852 seconds, Fetched: 2 row(s)
hive>
```

- **Query on aggregate functions**

```
hive> select max(amount) as MaxAmt,min(amount) as MinAmt,avg(amount) as AvgAmt;
FAILED: SemanticException [Error 10004]: Line 1:11 Invalid table alias or column reference 'amount': (possible column names are: )
hive> select max(amount) as MaxAmt,min(amount) as MinAmt,avg(amount) as AvgAmt from bank;
Query ID = hdoop_20210707133437_b7f61457-97ff-4825-aa2e-6a6e1f567806
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_1625638408330_0007, Tracking URL = http://asha-VirtualBox:8088/proxy/application_1625638408330_0007/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1625638408330_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-07 13:35:04,497 Stage-1 map = 0%,  reduce = 0%
2021-07-07 13:36:04,657 Stage-1 map = 0%,  reduce = 0%
2021-07-07 13:36:20,790 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.79 sec
2021-07-07 13:36:50,858 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.08 sec
MapReduce Total cumulative CPU time: 10 seconds 80 msec
Ended Job = job_1625638408330_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.08 sec   HDFS Read: 17237 HDFS Write: 123 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 80 msec
OK
80000.0 30000.0 51250.0
Time taken: 137.343 seconds, Fetched: 1 row(s)
hive>
```

- **Construction of view**

- **Creating one more table called customer(cust_id int,tid int,loc string) And importing values from csv file**

- **Experimenting on join operation**

- **Experimenting on alter operation**