

create database miniproject_1;

```
hive> create database miniproject_1;
OK
Time taken: 0.517 seconds
hive> 
```

```
[itv002776@g01 ~]$ hdfs dfs -ls /user/itv002776/warehouse;
Found 11 items
drwxr-xr-x - itv002776 supergroup          0 2022-12-29 10:59 /user/itv002776/warehouse/customers
drwxr-xr-x - itv002776 supergroup          0 2022-12-30 10:27 /user/itv002776/warehouse/dhiraj_itv002776
drwxr-xr-x - itv002776 supergroup          0 2023-03-12 12:35 /user/itv002776/warehouse/dhiraj_itv002776.db
drwxr-xr-x - itv002776 supergroup          0 2023-01-08 02:26 /user/itv002776/warehouse/dhiraj_itv002776_Covi
d_India
drwxr-xr-x - itv002776 supergroup          0 2023-01-08 02:28 /user/itv002776/warehouse/dhiraj_itv002776_Stat
e_Testing
-rw-r--r-- 3 itv002776 supergroup        130 2023-02-05 00:43 /user/itv002776/warehouse/file1.txt
drwxr-xr-x - itv002776 supergroup          0 2023-02-05 04:52 /user/itv002776/warehouse/hive_db_5thfeb.db
drwxr-xr-x - itv002776 supergroup          0 2023-03-22 11:21 /user/itv002776/warehouse/miniproject_1.db
drwxr-xr-x - itv002776 supergroup          0 2022-12-29 10:57 /user/itv002776/warehouse/orders
drwxr-xr-x - itv002776 supergroup          0 2023-03-04 01:28 /user/itv002776/warehouse/temp_files
drwxr-xr-x - itv002776 supergroup          0 2023-01-08 09:10 /user/itv002776/warehouse/week5_assignment.db
[itv002776@g01 ~]$ 
```

1. Create a schema based on the given dataset

```
create table agentloggingreport(SLno int,agent string,date1 date,logintime string,logouttime
string,duration string)row format delimited fields terminated by ','
tblproperties("skip.header.line.count"="1");
```

```
OK
Time taken: 0.032 seconds
hive> create table agentloggingreport(SLno int,agent string,date1 date,logintime string,logouttime string,durat
ion string)row format delimited fields terminated by ',';
OK
Time taken: 0.115 seconds
hive> 
```

```
create table agentperformance(SLno int,date1 date,agentname string,total_chats
int,average_response string,average_resolution string,average_rating float,totalfeedback
int)row format delimited fields terminated by ',' tblproperties("skip.header.line.count"="1")
;
```

```
Time taken: 0.115 seconds
hive> create table agentperformance(SLno int,date1 date,agentname string,total_chats int,average_response strin
g,average_resolution string,average_rating float,totalfeedback int);
OK
Time taken: 0.145 seconds
hive> 
```

2. Dump the data inside the hdfs in the given schema location.

```
hdfs dfs -put AgentLoggingReport.csv /user/itv002776/agentfiles/;
hdfs dfs -put AgentPerformance.csv /user/itv002776/agentfiles/;
```

```

[itv002776@g01 ~]$ hdfs dfs -mkdir /user/itv002776/agentfiles;
[itv002776@g01 ~]$ hdfs dfs -put 'agentloggingreport.csv' /user/itv002776/agentfiles/;
put: `agentloggingreport.csv': No such file or directory
[itv002776@g01 ~]$ hdfs dfs -put agentloggingreport.csv /user/itv002776/agentfiles/;
put: `agentloggingreport.csv': No such file or directory
[itv002776@g01 ~]$ hdfs dfs -put AgentLoggingReport.csv /user/itv002776/agentfiles/;
[itv002776@g01 ~]$ hdfs dfs -put AgentPerformance.csv /user/itv002776/agentfiles/;
[itv002776@g01 ~]$

[itv002776@g01 ~]$ hdfs dfs -ls /user/itv002776/agentfiles/;
Found 2 items
-rw-r--r--  3 itv002776 supergroup      55351 2023-03-22 12:03 /user/itv002776/agentfiles/AgentLoggingReport.csv
-rw-r--r--  3 itv002776 supergroup     109853 2023-03-22 12:05 /user/itv002776/agentfiles/AgentPerformance.csv
[itv002776@g01 ~]$

```

Load data into table from hdfs:

load data inpath '/user/itv002776/agentfiles/Agentperformance.csv' into table agentperformance;

load data inpath '/user/itv002776/agentfiles/AgentPerformance.csv' into table agentperformance;

```

Time taken: 0.145 seconds
hive> load data inpath 'AgentLoggingReport.csv' into table agentloggingreport;
FAILED: SemanticException [Error 10001]: Line 1:52 Table not found 'agentloggingreport'
hive> load data inpath 'AgentLoggingReport.csv' into table agentloggingreport;
FAILED: SemanticException Line 1:17 Invalid path 'AgentLoggingReport.csv': No files matching path hdfs://m01.itversity.com:9000/user/itv002776/AgentLoggingReport.csv
hive> load data inpath '/user/itv002776/agentfiles/AgentLoggingReport.csv' into table agentloggingreport;
Loading data to table miniproject_1.agentloggingreport
OK
Time taken: 0.905 seconds
hive> load data inpath '/user/itv002776/agentfiles/Agentperformance.csv' into table agentperformance;
FAILED: SemanticException Line 1:17 Invalid path '/user/itv002776/agentfiles/Agentperformance.csv': No files matching path hdfs://m01.itversity.com:9000/user/itv002776/agentfiles/Agentperformance.csv
hive> load data inpath '/user/itv002776/agentfiles/AgentPerformance.csv' into table agentperformance;
Loading data to table miniproject_1.agentperformance
OK
Time taken: 0.395 seconds
hive>

```

3. List of all agents' names.

Select agentname from agenperformance;

```

Saif Khan
Anurag Tiwari
Maneesh
Harikrishnan Shaji
Ashish
Prateek _iot
Rohan
Saikumarreddy N
Sudhanshu Kumar
Spuri
Suraj S Bilgi
Nandani Gupta
Shivan_S
Mahesh Sarade
Sowmiya Sivakumar
Nitin M
Vivek
Ayushi Mishra
Chaitra K Hiremath
Time taken: 10.816 seconds, Fetched: 2160 row(s)
hive>

```

4. Find out agent average rating.

```

select agentname,avg(average_rating) from agentperformance group by agentname
, totalfeedback)

```

```

hive> select agentname,avg(average_rating) from agentperformance group by agentname;

```

```

Query ID = itv002776_20230322131219_5a3d25e4-5744-402d-90a2-9ca51b5fae27

```

```

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_1675999795986_18794, Tracking URL = http://m02.itversity.com:19088/proxy/application_1675999795986_18794/

```

```

Kill Command = /opt/hadoop/bin/mapred job -kill job_1675999795986_18794

```

```

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

```

```

2023-03-22 13:12:38,829 Stage-1 map = 0%, reduce = 0%

```

```

2023-03-22 13:12:42,912 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.23 sec

```

```

2023-03-22 13:12:46,975 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.22 sec

```

```

MapReduce Total cumulative CPU time: 4 seconds 220 msec

```

Anirudh	0.6449999968210857
Ankit Sharma	0.0
Ankitjha	0.26666666666666666
Anurag Tiwari	0.18333333333333332
Aravind	2.1813333511352537
Ashad Nasim	0.16666666666666666
Ashish	0.0
Ayushi Mishra	3.481999969482422
Bharath	2.9836666584014893
Boktiar Ahmed Bappy	3.567999982833862
Chaitra K Hiremath	0.8646666606267294
Deepranjan Gupta	2.886666695276896
Dibyanshu	0.0
Harikrishnan Shaji	2.6396666526794434
Hitesh Choudhary	0.0
Hrisikesh Neogi	3.1363333304723104
Hyder Abbas	0.0
Ineuron Intelligence	0.0
Ishawant Kumar	3.543333347638448
Jawala Prakash	3.472000018755595
Jayant Kumar	1.068666664759318

5. Total working days for each agents

```
select agent,count(distinct(date1)) from agentloggingreport group by agent;
```

```
e_rating, totalfeedback)
hive> select agent,count(distinct(date1)) from agentloggingreport group by agent;
Query ID = itv002776_20230322132251_4aaa309d-755e-463c-94ab-a9e4d2fba641
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_1675999795986_18797, Tracking URL = http://m02.itversity.com:19088/proxy/application_1675999795986_18797/
Kill Command = /opt/hadoop/bin/mapred job -kill job_1675999795986_18797
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-22 13:23:07,936 Stage-1 map = 0%, reduce = 0%
2023-03-22 13:23:12,023 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.67 sec
2023-03-22 13:23:16,098 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.19 sec
MapReduce Total cumulative CPU time: 3 seconds 190 msec
Ended Job = job_1675999795986_18797
```

Total MapReduce CPU Time Spent: 3 seconds 190 msec

OK

Aditya Shinde	1
Aditya_iot	8
Amersh	2
Ameya Jain	7
Ankitjha	2
Anurag Tiwari	10
Aravind	7
Ayushi Mishra	9
Bharath	8
Boktiar Ahmed Bappy	9
Chaitra K Hiremath	7
Deepranjan Gupta	10
Dibyanshu	9
Harikrishnan Shaji	9
Hrisikesh Neogi	9
Hyder Abbas	2
Ineuron Intelligence	1
Ishawant Kumar	11

6. Total query that each agent have taken

OK

Time taken: 0.316 seconds

hive> select agentname,count(total_chats) from agentperformance group by agentname;

Query ID = itv002776_20230323141457_19ade6f9-4691-44d0-b547-0500e06f3aad

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_1675999795986_19023, Tracking URL = http://m02.itversity.com:19088/proxy/application_1675999795986_19023/

Kill Command = /opt/hadoop/bin/mapred job -kill job_1675999795986_19023

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2023-03-23 14:15:23,124 Stage-1 map = 0%, reduce = 0%

2023-03-23 14:15:27,249 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.03 sec

2023-03-23 14:15:32,369 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.48 sec

MapReduce Total cumulative CPU time: 3 seconds 480 msec

```

MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.48 sec HDFS Read: 124856 HDFS Write: 2004 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 480 msec
OK
Abhishek      30
Aditya      30
Aditya Shinde  30
Aditya_iot    30
Amersh      30
Ameya Jain    30
Anirudh      30
Ankit Sharma  30
Ankitjha     30
Anurag Tiwari 30
Aravind      30
Ashad Nasim   30
Ashish      60
Ayushi Mishra 30
Bharath      30
Boktiar Ahmed Bappy 30
Chaitra K Hiremath 30

```

itv002776@n01

7. Total Feedback that each agent have received

```
select agentname,sum(totalfeedback) from agentperformance group by agentname;
```

```

hive> select agentname,sum(totalfeedback) from agentperformance group by agentname;
Query ID = itv002776_20230323141829_alf3eefd-74b2-434a-ae61-aecb69f6271a
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_1675999795986_19024, Tracking URL = http://m02.itversity.com:19088/proxy/application_1675999795986_19024/
Kill Command = /opt/hadoop/bin/mapred job -kill job_1675999795986_19024
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-23 14:18:46,845 Stage-1 map = 0%, reduce = 0%
2023-03-23 14:18:50,942 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.28 sec
2023-03-23 14:18:56,043 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.05 sec
MapReduce Total cumulative CPU time: 4 seconds 50 msec

```

```

Ankit Sharma      0
Ankitjha         3
Anurag Tiwari     3
Aravind          233
Ashad Nasim       9
Ashish           0
Ayushi Mishra    329
Bharath          247
Boktiar Ahmed Bappy 311
Chaitra K Hiremath 37
Deepranjan Gupta 312
Dibyanshu        0
Harikrishnan Shaji 231
Hitesh Choudhary  0
Hrisikesh Neogi  367
Hyder Abbas      0
Ineuron Intelligence 0
Ishawant Kumar   202
Jawala Prakash   250
Jayant Kumar     70
Jaydeep Dixit    305

```

itv002776@n01

8. Agent name who have average rating between 3.5 to 4

Select agentname from agentperformance where average_rating>3.5 and average_rating<4

```
Time taken: 0.071 seconds
hive> Select agentname from agentperformance where average_rating>3.5 and average_rating<4
> ;
OK
Swati
Manjunatha A
Prateek _iot
Nandani Gupta
Jaydeep Dixit
Mahesh Sarade
Zeeshan
Hrisikesh Neogi
Khushboo Priya
Wasim
Jawala Prakash
Nishtha Jain
Nandani Gupta
Prerna Singh
```

9. Agent name who have average rating less than 3.5

Select agentname from agentperformance where average_rating<3.5

```
Sandipan Saha
Mukesh Rao
Saif Khan
Anurag Tiwari
Ankitjha
Ashish
Maneesh
Sudhanshu Kumar
Suraj S Bilgi
Rohan
Spuri
Nandani Gupta
Shivan_S
Vivek
Nitin M
Shivan K
Mahesh Sarade
Madhulika G
Chaitra K Hiremath
Sowmiya Sivakumar
Boktiar Ahmed Bappy
```

10. Agent name who have rating more than 4.5

Select agentname from agentperformance where average_rating>4.5

=====

Time taken: 9.312 seconds, Fetched: 1474 row(s)

hive> Select agentname from agentperformance where average_rating>4.5;

OK

Ameya Jain
Mahesh Sarade
Mukesh
Saikumarreddy N
Sanjeev Kumar
Harikrishnan Shaji
Sowmiya Sivakumar
Boktiar Ahmed Bappy
Shivananda Sonwane
Ishawant Kumar
Deepranjan Gupta
Shivananda Sonwane
Muskan Garg

11. How many feedback agents have received more than 4.5 average

ack)

hive> Select agentname,totalfeedback from agentperformance where average_rating>4.5;

OK

Ameya Jain	11
Mahesh Sarade	7
Mukesh	8
Saikumarreddy N	12
Sanjeev Kumar	6
Harikrishnan Shaji	7
Sowmiya Sivakumar	16
Boktiar Ahmed Bappy	4
Shivananda Sonwane	1
Ishawant Kumar	3
Deepranjan Gupta	5
Shivananda Sonwane	9
Muskan Garg	4
