hive_project

hive project by ineuron

This is a real time dataset of the ineuron technical consultant team. You have to perform hive analysis on this given dataset.

Download Dataset 1 - https://drive.google.com/file/d/1WrG-9qv6atP-W3P_-gYln1hHyFKRKMHP/view

Download Dataset 2 - https://drive.google.com/file/d/1-JIPCZ34dyN6k9CqJa-Y8yxIGq6vTVXU/view

Note: both files are csv files.

show databases;

1. Create a schema based on the given dataset
2. Dump the data inside the hdfs in the given schema location.

```
create database hive_proj;

use hive_proj;

create table LoginReport
(
SL_No int,
Agent String,
date_dd date,
Login_time string,
Logout_time string,
duration string
)
row format delimited
```

```
fields terminated by ',';

load data local inpath 'file:///config/workspace/AgentLogingReport (1).csv'into table LoginReport;

select * from LoginReport limit 10;

create table AgentPerfromance
(
Sl_no int,
Date_d string,
Agent_name string,
Total_chats int,
Avg_responce_time string,
Avg_resolution string,
Avg_rating float,
total_feedback int
)
row format delimited
fields terminated by ',';

load data local inpath 'file:///config/workspace/AgentPerformance (3).csv'into table
AgentPerfromance;

select * from AgentPerfromance limit 10;
```

3. List of all agents' names.

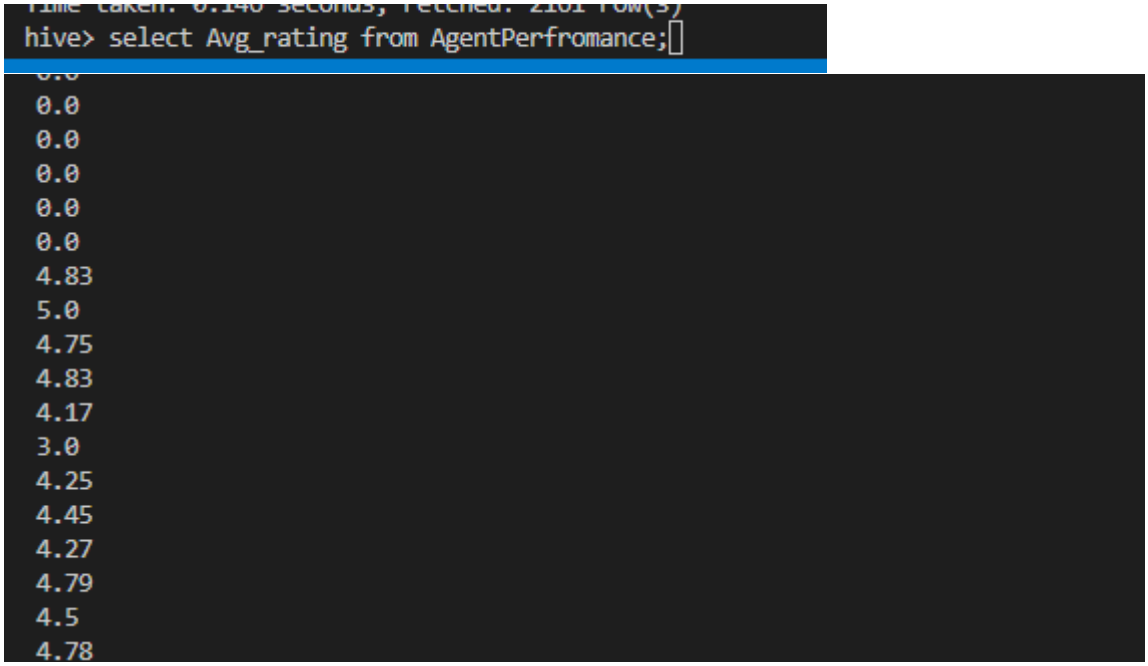
select Agent from loginreport;

```
hive> select Agent from loginreport;
OK
Agent
Shivananda Sonwane
Khushboo Priya
Nandani Gupta
Hrisikesh Neogi
Mukesh
Sowmiya Sivakumar
Manjunatha A
Harikrishnan Shaji
Suraj S Bilgi
Shivan K
Anurag Tiwari
Ishawant Kumar
Shivan K
Shubham Sharma
Shivan K
Prerna Singh
Shivan K
Shivan K
Hrisikesh Neogi
Khushboo Priya
```

```
Prateek _iot
Suraj S Bilgi
Aditya_iot
Shivananda Sonwane
Prerna Singh
Mahesh Sarade
Prateek _iot
Dibyanshu
Dibyanshu
Dibyanshu
Dibyanshu
Ishawant Kumar
Prabir Kumar Satapathy
Mithun S
Nandani Gupta
Dibyanshu
Dibyanshu
Deepranjan Gupta
Chaitra K Hiremath
Dibyanshu
Prerna Singh
Jaydeep Dixit
Nandani Gupta
Sanjeev Kumar
Mahesh Sarade
Dibyanshu
Zeeshan
Ishawant Kumar
Dibyanshu
Mukesh
Shivan K
```

4. Find out agent average rating

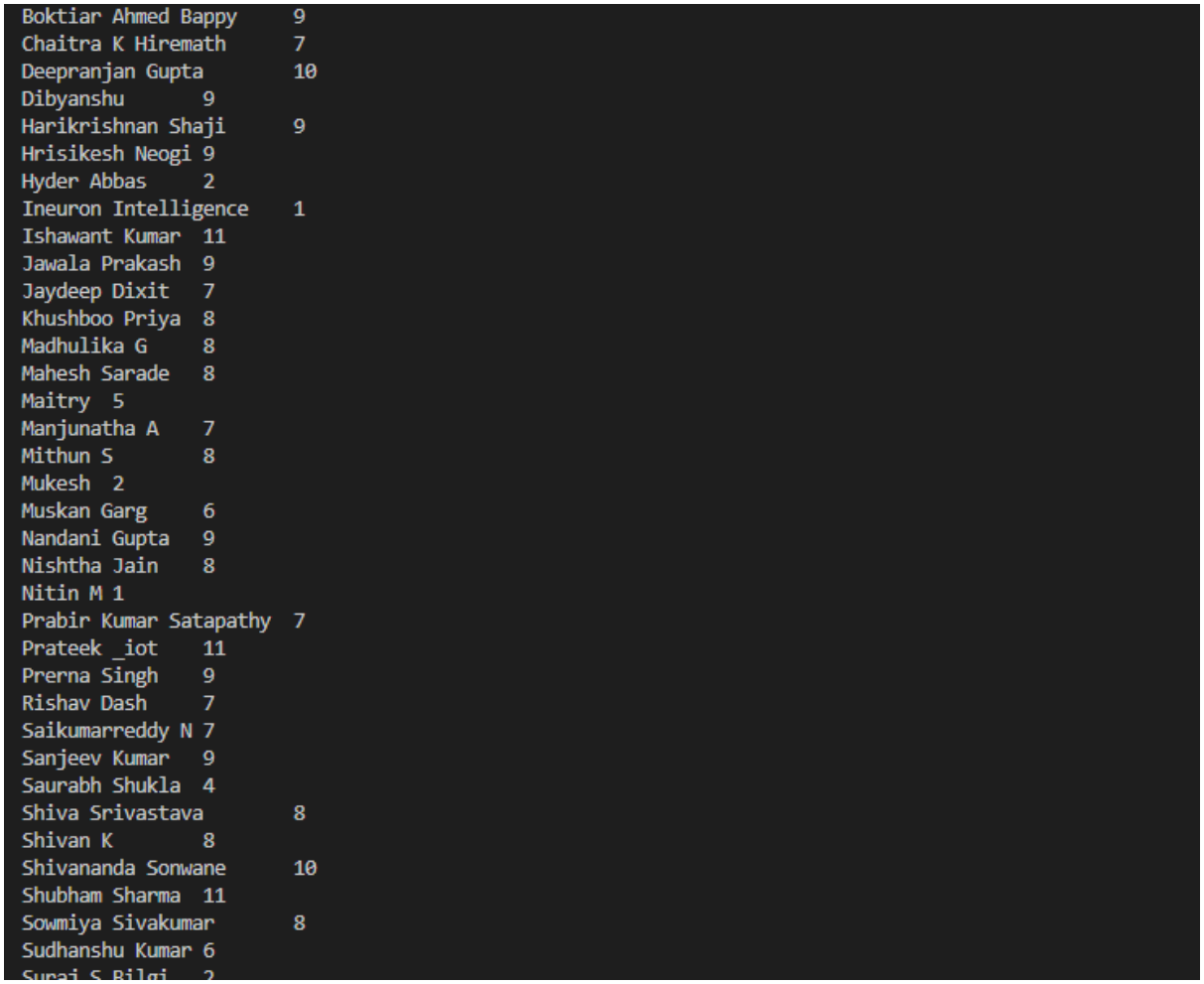
select Avg_rating from AgentPerfromance;



5. Total working days for each agents

select Agent as name, count(distinct date_dd) as total_working_Days from LoginReport group by Agent;

```
hive> select Agent as name, count(distinct date_dd) as total_working_Days from LoginReport group by Agent;
Query ID = abc_20230728120011_5f60c354-bb3b-410a-88aa-e8bc8d557b4c
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_1690525332945_0001, Tracking URL = http://7c4a1be4cbc3:8088/proxy/application_1690525332945_0001/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1690525332945_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-28 12:00:25,125 Stage-1 map = 0%,  reduce = 0%
2023-07-28 12:00:33,391 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.02 sec
2023-07-28 12:00:39,575 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.08 sec
MapReduce Total cumulative CPU time: 6 seconds 80 msec
Ended Job = job_1690525332945_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.08 sec   HDFS Read: 65886 HDFS Write: 1437 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 80 msec
OK
Aditya Shinde   1
Aditya_iot      8
Agent   0
Amersh  2
Ameya Jain      7
Ankitjha        2
Anurag Tiwari   10
Aravind 7
Ayushi Mishra   9
Bharath 8
Boktiar Ahmed Bappy     9
```



```
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
Jawala Prakash  9
Jaydeep Dixit   7
Khushboo Priya  8
Madhulika G     8
Mahesh Sarade   8
Maitry  5
Manjunatha A    7
Mithun S        8
Mukesh  2
Muskan Garg     6
Nandani Gupta   9
Nishtha Jain    8
Nitin M 1
Prabir Kumar Satapathy  7
Prateek _iot    11
Prerna Singh    9
Rishav Dash     7
Saikumarreddy N 7
Sanjeev Kumar   9
Saurabh Shukla  4
Shiva Srivastava        8
Shivan K        8
Shivananda Sonwane      10
Shubham Sharma  11
Sowmiya Sivakumar       8
Sudhanshu Kumar 6
Suraj S Bilgi   2
```

6. Total query that each agent have taken

select Agent_name as name, sum(Total_chats) as total_query_take from AgentPerfromance group by Agent_name order by total_query_take desc;

```
hive> select Agent_name as name, sum(Total_chats) as total_query_take from AgentPerfromance group by Agent_name order by total_query_take desc;
Query ID = abc_20230728120711_dadb3a55-708b-4715-8d6a-341e6636c0d2
Total jobs = 2
Launching Job 1 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_1690525332945_0002, Tracking URL = http://7c4a1be4cbc3:8088/proxy/application_1690525332945_0002/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1690525332945_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-28 12:07:21,713 Stage-1 map = 0%,  reduce = 0%
2023-07-28 12:07:30,964 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.39 sec
2023-07-28 12:07:37,118 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.27 sec
MapReduce Total cumulative CPU time: 6 seconds 270 msec
Ended Job = job_1690525332945_0002
Launching Job 2 out of 2
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_1690525332945_0003, Tracking URL = http://7c4a1be4cbc3:8088/proxy/application_1690525332945_0003/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1690525332945_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-07-28 12:07:52,427 Stage-2 map = 0%,  reduce = 0%
```

```
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1690525332945_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-07-28 12:07:52,427 Stage-2 map = 0%,  reduce = 0%
2023-07-28 12:08:00,630 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.41 sec
2023-07-28 12:08:06,781 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 5.03 sec
MapReduce Total cumulative CPU time: 5 seconds 30 msec
Ended Job = job_1690525332945_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.27 sec   HDFS Read: 123539 HDFS Write: 2316 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.03 sec   HDFS Read: 9832 HDFS Write: 2041 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 300 msec
OK
Hrisikesh Neogi 578
Nandani Gupta   560
Zeeshan         542
Maitry  542
Swati   524
Ayushi Mishra   514
Jaydeep Dixit   512
Shubham Sharma  510
Sanjeev Kumar   507
Mithun S        503
Deepranjan Gupta        493
Madhulika G     469
Boktiar Ahmed Bappy     452
Khushboo Priya  446
Shivananda Sonwane      441
Jawala Prakash  439
Wasim   433
Manjunatha A    413
Rishav Dash     409
Prerna Singh    401
```

```
Moskan Garg     36
Shiva Srivastava        53
Vivek   44
Sandipan Saha   30
Suraj S Bilgi   28
Tarun   22
Mukesh  19
Ashad Nasim     18
Saurabh Shukla  16
Mahak   7
Shivan_S        7
Ankitjha        5
Mukesh Rao      5
Maneesh         4
Anurag Tiwari   4
Sudhanshu Kumar 2
Samprit         1
Hitesh Choudhary        1
Dibyanshu       1
Sanjeevan       0
Saif Khan       0
Rohan   0
Nitin M 0
Spuri   0
Ineuron Intelligence    0
Hyder Abbas     0
Uday Mishra     0
Vasanth P       0
Ashish  0
Ankit Sharma    0
Amersh  0
Aditya  0
Abhishek        0
Agent Name      NULL
Time taken: 56.336 seconds, Fetched: 71 row(s)
```

7. Total Feedback that each agent have received

select Agent_name as name, sum(total_feedback) as total_feedback_taken from AgentPerfromance group by Agent_name order by total_feedback_taken desc;

```
Time taken: 56.336 seconds, Fetched: 71 row(s)
hive> select Agent_name as name, sum(total_feedback) as total_feedback_taken from AgentPerfromance group by Agent_name order by total_feedback_taken desc;
Query ID = abc_20230728121255_50f3adfd-a57d-4aff-b493-55da25eb6ad0
Total jobs = 2
Launching Job 1 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_1690525332945_0004, Tracking URL = http://7c4a1be4cbc3:8088/proxy/application_1690525332945_0004/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1690525332945_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-28 12:13:05,213 Stage-1 map = 0%,  reduce = 0%
2023-07-28 12:13:13,432 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.79 sec
2023-07-28 12:13:19,582 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.2 sec
MapReduce Total cumulative CPU time: 5 seconds 200 msec
Ended Job = job_1690525332945_0004
Launching Job 2 out of 2
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_1690525332945_0005, Tracking URL = http://7c4a1be4cbc3:8088/proxy/application_1690525332945_0005/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1690525332945_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-07-28 12:13:34,634 Stage-2 map = 0%,  reduce = 0%
```

```
2023-07-28 12:13:34,634 Stage-2 map = 0%,  reduce = 0%
2023-07-28 12:13:42,825 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.32 sec
2023-07-28 12:13:48,971 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.89 sec
MapReduce Total cumulative CPU time: 4 seconds 890 msec
Ended Job = job_1690525332945_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.2 sec   HDFS Read: 123638 HDFS Write: 2303 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.89 sec   HDFS Read: 9829 HDFS Write: 2037 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 90 msec
OK
Hrisikesh Neogi 367
Mithun S        364
Maitry  347
Zeeshan         335
Ayushi Mishra   329
Deepranjan Gupta        312
Boktiar Ahmed Bappy     311
Sanjeev Kumar   311
Nandani Gupta   308
Jaydeep Dixit   305
Swati   302
Shubham Sharma  300
Saikumarreddy N 290
Khushboo Priya  289
Wasim   284
Madhulika G     281
Rishav Dash     264
Shivananda Sonwane      263
Nishtha Jain    257
Manjunatha A    254
Jawala Prakash  250
Bharath         247
Shivan K        243
Prerna Singh    235
Aravind         233
Harikrishnan Shaji      231
```

```
Chaitra K Hiremath      37
Muskan Garg     37
Vivek   20
Sandipan Saha   18
Mukesh  17
Suraj S Bilgi   15
Ashad Nasim     9
Saurabh Shukla  8
Tarun   6
Mahak   5
Mukesh Rao      5
Shivan_S        4
Anurag Tiwari   3
Ankitjha        3
Maneesh         3
Sudhanshu Kumar 2
Samprit         0
Saif Khan       0
Rohan   0
Sanjeevan       0
Nitin M 0
Spuri   0
Ineuron Intelligence    0
Hyder Abbas     0
Hitesh Choudhary        0
Dibyanshu       0
Uday Mishra     0
Vasanth P       0
Ashish  0
Ankit Sharma    0
Amersh  0
Aditya  0
Abhishek        0
Agent Name      NULL
Time taken: 55.557 seconds, Fetched: 71 row(s)
```

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

```
select Agent_name from AgentPerfromance where Avg_rating BETWEEn 3.5 and 4;
```