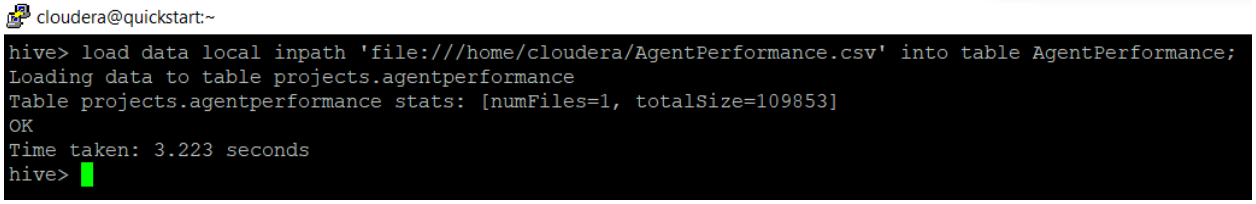
1. Create a schema based on the given dataset

```
hive>
hive> create table AgentPerformance(
    > SL_No int,
    > Date string,
    > Agent_Name string,
    > Total_Chats int,
    > Average_Response_Time string,
    > Average_Resolution_Time string,
    > Average_Rating float,
    > Total_Feedback int
    > )
    > row format delimited
    > fields terminated by ','
    > TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 0.767 seconds
hive>
```

```
hive> Create table AgentLogingReport(
    > SL_No int,
    > Agent_Name string,
    > Date string,
    > Login_Time string,
    > Logout_TIme string,
    > Duration string
    > )
    > row format delimited
    > fields terminated by ','
    > TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 0.324 seconds
hive>
```

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



cloudera@quickstart:~

```
hive> load data local inpath 'file:///home/cloudera/AgentLogingReport.csv' into table AgentLogingReport;
Loading data to table projects.agentlogingreport
Table projects.agentlogingreport stats: [numFiles=1, totalSize=55351]
OK
Time taken: 0.911 seconds
hive>
```

3. List of all agents' names.

cloudera@quickstart:~

```
hive> Select Agent_Name from AgentPerformance;
```

```
Dibyanshu
Uday Mishra
Aditya_iot
Nishtha Jain
Aditya Shinde
Hyder Abbas
Prabir Kumar Satapathy
Bharath
Deepranjan Gupta
Abhishek
Rishav Dash
Ankit Sharma
Samprit
Wasim
Muskan Garg
Aravind
Ashad Nasim
Vasanth P
Amersh
Ineuron Intelligence
Ishawant Kumar
Sandipan Saha
Mahak
Ameya Jain
Mukesh Rao
Zeeshan
Aditya
Ankitjha
Saif Khan
```

4. Find out agent average rating.

```
hive> select agent_name,round(avg(average_rating),2) avg_rating from agentperformance
    > where total_chats>0
    > group by agent_name;
Query ID = cloudera_20220924064343_1b6a3d98-164c-48bb-b6cb-4be3e9d7019b
Total jobs = 1
Launching Job 1 out of 1
```

```
agent_name      avg_rating
Aditya Shinde   4.5
Aditya_iot      4.14
Ameya Jain      4.44
Anirudh         2.76
Ankitjha        2.67
Anurag Tiwari   2.75
Aravind         4.67
Ashad Nasim     2.5
Ayushi Mishra   4.35
Bharath         4.71
Boktiar Ahmed Bappy     4.12
Chaitra K Hiremath      4.32
Deepranjan Gupta        4.12
Dibyanshu       0.0
Harikrishnan Shaji      4.17
Hitesh Choudhary        0.0
Hrisikesh Neogi 4.48
Ishawant Kumar  4.43
Jawala Prakash  4.17
Jayant Kumar    4.01
Jaydeep Dixit   4.52
Khushboo Priya  4.27
Madhulika G     4.37
Mahak   3.0
Mahesh Sarade   4.0
Maitry  4.39
Maneesh         1.67
Manjunatha A    4.31
```

5. Total working days for each agents

cloudera@quickstart:~

```
hive> select Agent_Name,count(distinct date) No_of_Working_days from Agentlogingreport group by Agent_Name;
Query ID = cloudera_20220924035252_86c8d6a3-6473-4d4a-a452-4499815513bf
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:
```

```
agent_name      no_of_working_days
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
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
```

6. Total query that each agent have taken

```
hive> select Agent_Name,sum(total_chats) Total_queries from AgentPerformance group by Agent_Name;
Query ID = cloudera_20220924040505_326cb760-ba82-40af-be99-cc305af33a31
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>
```

```
OK
agent_name      total_queries
Abhishek        0
Aditya  0
Aditya Shinde   277
Aditya_iot      231
Amersh  0
Ameya Jain      322
Anirudh         81
Ankit Sharma    0
Ankitjha        5
Anurag Tiwari   4
Aravind         366
Ashad Nasim     18
Ashish  0
Ayushi Mishra   514
Bharath         369
Boktiar Ahmed Bappy     452
Chaitra K Hiremath      64
Deepranjan Gupta        493
Dibyanshu       1
Harikrishnan Shaji      381
Hitesh Choudhary        1
Hrisikesh Neogi 578
Hyder Abbas     0
Ineuron Intelligence    0
```

7. Total Feedback that each agent have received

```
hive> select agent_name,sum(total_feedback) total_feedback
    > from agentperformance
    > group by agent_name;
Query ID = cloudera_20220924070202_b09f31ca-2111-446e-b027-f745ac8af2d2
Total jobs = 1
Launching Job 1 out of 1
```

```
agent_name      total_feedback
Abhishek        0
Aditya  0
Aditya Shinde   153
Aditya_iot      131
Amersh  0
Ameya Jain      228
Anirudh         39
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
```

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

```
hive> select agent_name,round(avg(average_rating),2) avg_rating
    > from agentperformance
    > where total_chats>0
    > group by agent_name
    > having avg(average_rating) between 3.5 and 4;
Query ID = cloudera_20220924070707_ea797c6b-829f-403a-8b9d-e1aee4141fbb
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_1663936619422_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663936619422_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663936619422_0017
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-24 07:07:27,036 Stage-1 map = 0%,  reduce = 0%
2022-09-24 07:07:35,771 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.27 sec
2022-09-24 07:07:45,145 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.36 sec
MapReduce Total cumulative CPU time: 11 seconds 360 msec
Ended Job = job_1663936619422_0017
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.36 sec   HDFS Read: 120803 HDFS Write: 28 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 360 msec
OK
Mithun S        3.93
Shivan K        3.87
Time taken: 29.208 seconds, Fetched: 2 row(s)
```

9. Agent name who have rating less than 3.5

cloudera@quickstart:~

```
hive> select agent_name,round(avg(average_rating),2) avg_rating
    > from agentperformance
    > where total_chats>0
    > group by agent_name
    > having avg(average_rating)<3.5;
Query ID = cloudera_20220924071313_9284f3cd-c8e2-4455-8746-2d210140f26c
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_1663936619422_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663936619422_0019
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-24 07:13:51,526 Stage-1 map = 0%,  reduce = 0%
2022-09-24 07:14:00,217 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.3 sec
2022-09-24 07:14:08,579 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.12 sec
MapReduce Total cumulative CPU time: 10 seconds 120 msec
Ended Job = job_1663936619422_0019
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.12 sec   HDFS Read: 120734 HDFS Write: 177 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 120 msec
OK
Anirudh         2.76
Ankitjha        2.67
Anurag Tiwari   2.75
Ashad Nasim     2.5
Dibyanshu       0.0
Hitesh Choudhary        0.0
Mahak   3.0
Maneesh         1.67
Mukesh Rao      2.56
Samprit         0.0
Tarun   1.5
Vivek   3.0
Time taken: 25.821 seconds, Fetched: 12 row(s)
```

10. Agent name who have rating more than 4.5

```
hive> select agent_name,round(avg(average_rating),2) avg_rating
    > from agentperformance
    > where total_chats>0
    > group by agent_name
    > having avg(average_rating)>4.5;
Query ID = cloudera_20220924071010_5aedb302-2899-48c8-ab8a-3735d6c0f0a0
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_1663936619422_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663936619422_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-24 07:10:39,253 Stage-1 map = 0%,  reduce = 0%
2022-09-24 07:10:46,824 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 7.01 sec
2022-09-24 07:10:56,229 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.94 sec
MapReduce Total cumulative CPU time: 8 seconds 940 msec
Ended Job = job_1663936619422_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.94 sec   HDFS Read: 120658 HDFS Write: 193 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 940 msec
OK
Aditya Shinde   4.5
Aravind         4.67
Bharath         4.71
Jaydeep Dixit   4.52
Mukesh  4.64
Saikumarreddy N 4.57
Shivananda Sonwane      4.53
Shubham Sharma  4.61
Sudhanshu Kumar 5.0
Suraj S Bilgi   4.68
Wasim   4.5
Time taken: 25.228 seconds, Fetched: 11 row(s)
```

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

```
hive> select agent_name,sum(total_feedback) total_feedback
    > from agentperformance
    > where total_chats>0
    > group by agent_name
    > having avg(average_rating)>4.5;
Query ID = cloudera_20220924071818_7559d0fd-1a8d-4d9b-8369-3997a47cfa69
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_1663936619422_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663936619422_0021
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-24 07:18:53,833 Stage-1 map = 0%,  reduce = 0%
2022-09-24 07:19:00,832 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.46 sec
2022-09-24 07:19:10,781 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.21 sec
MapReduce Total cumulative CPU time: 8 seconds 210 msec
Ended Job = job_1663936619422_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.21 sec   HDFS Read: 121233 HDFS Write: 181 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 210 msec
OK
agent_name      total_feedback
Aditya Shinde   153
Aravind         233
Bharath         247
Jaydeep Dixit   305
Mukesh  17
Saikumarreddy N 290
Shivananda Sonwane      263
Shubham Sharma  300
Sudhanshu Kumar 2
Suraj S Bilgi   15
Wasim   284
Time taken: 26.468 seconds, Fetched: 11 row(s)
```

12. average weekly response time for each agent

```
hive>    select agent_name,
    >    date_format(from_unixtime(unix_timestamp(date,'MM/dd/yyyy')),'W') week_of_month,
    >    round(avg(3600*hour(average_response_time)+60*minute(average_response_time)+second(average_response_time)),2) avg_response_time_in_seconds
    >    from agentperformance
    >    where total_chats>0
    >    group by agent_name, date_format(from_unixtime(unix_timestamp(date,'MM/dd/yyyy')),'W');
Query ID = cloudera_20220926054343_c318cb34-7b21-466e-9753-79c6720ec3fe
```

OK

agent_name	week_of_month	avg_response_time_in_seconds
Aditya Shinde	1	86.0
Aditya Shinde	2	68.5
Aditya Shinde	3	79.0
Aditya_iot	2	70.5
Aditya_iot	3	72.8
Aditya_iot	4	42.2
Aditya_iot	5	60.6
Ameya Jain	2	0.0
Ameya Jain	3	40.75
Ameya Jain	4	46.2
Ameya Jain	5	48.0
Anirudh	2	135.5
Anirudh	3	37.0
Ankitjha	2	0.0
Ankitjha	5	66.5
Anurag Tiwari	2	126.5
Aravind	2	34.0

13. average weekly resolution time for each agents

```
hive>   select agent_name,
    >    date_format(from_unixtime(unix_timestamp(date,'MM/dd/yyyy')),'W') week_of_month,
    >    round(avg(60*hour(average_resolution_time)+minute(average_resolution_time)+second(average_resolution_time)/60),2) avg_resolution_time_in_minutes
    >    from agentperformance
    >    where total_chats>0
    >    group by agent_name, date_format(from_unixtime(unix_timestamp(date,'MM/dd/yyyy')),'W')
    > ;
```

agent_name	week_of_month	avg_resolution_time_in_minutes
Aditya Shinde	1	19.87
Aditya Shinde	2	21.99
Aditya Shinde	3	31.7
Aditya_iot	2	18.08
Aditya_iot	3	21.26
Aditya_iot	4	15.68
Aditya_iot	5	14.76
Ameya Jain	2	1.78
Ameya Jain	3	11.13
Ameya Jain	4	11.68
Ameya Jain	5	11.9
Anirudh	2	14.01
Anirudh	3	12.23
Ankitjha	2	21.37
Ankitjha	5	2.93
Anurag Tiwari	2	18.48
Aravind	2	17.87
Aravind	3	20.6
Aravind	4	18.81
Aravind	5	17.13
Ashad Nasim	2	5.23
Ayushi Mishra	1	30.25
Ayushi Mishra	2	18.07
Ayushi Mishra	3	20.58
Ayushi Mishra	4	17.23
Ayushi Mishra	5	19.53

14. Find the number of chat on which they have received a feedback

cloudera@quickstart:~

```
hive> Select agent_name,sum(total_chats) total_chats,sum(total_feedback) total_feedback
    > from agentperformance
    > group by agent_name;
Query ID = cloudera_20220926073232_e161b441-7db1-44a8-ac3f-fdf58610f276
```

agent_name	total_chats	total_feedback
Abhishek	0	0
Aditya 0	0	
Aditya Shinde	277	153
Aditya_iot	231	131
Amersh 0	0	
Ameya Jain	322	228
Anirudh	81	39
Ankit Sharma	0	0
Ankitjha	5	3
Anurag Tiwari	4	3
Aravind	366	233
Ashad Nasim	18	9
Ashish 0	0	
Ayushi Mishra	514	329
Bharath	369	247
Boktiar Ahmed Bappy	452	311
Chaitra K Hiremath	64	37
Deepranjan Gupta	493	312
Dibyanshu 1	0	
Harikrishnan Shaji	381	231
Hitesh Choudhary	1	0
Hrisikesh Neogi 578	367	

15. Total contribution hour for each and every agents weekly basis

```
hive>    Select agent_name,
    >    date_format(from_unixtime(unix_timestamp(date,'dd-MMM-yy')),'W') week_of_month,
    >    round(sum(hour(duration)+minute(duration)/60+second(duration)/3600),2) total_contribution_hrs
    >    from agentlogingreport
    >    group by agent_name,date_format(from_unixtime(unix_timestamp(date,'dd-MMM-yy')),'W');
Query ID = cloudera_20220926061919_f98f5eea-480c-4317-a46c-4cbef5d5849e
```

agent_name	week_of_month	total_contribution_hrs
Aditya Shinde	5	0.04
Aditya_iot	4	6.1
Aditya_iot	5	9.64
Amersh 5	3.06	
Ameya Jain	4	17.23
Ameya Jain	5	24.85
Ankitjha	5	2.27
Anurag Tiwari	4	0.22
Anurag Tiwari	5	2.56
Aravind 4	18.21	
Aravind 5	6.09	
Ayushi Mishra	4	11.83
Ayushi Mishra	5	26.29
Bharath 4	17.79	
Bharath 5	30.29	
Boktiar Ahmed Bappy	4	17.42
Boktiar Ahmed Bappy	5	22.85
Chaitra K Hiremath	4	2.23
Chaitra K Hiremath	5	32.09
Deepranjan Gupta	4	39.73
Deepranjan Gupta	5	66.55
Dibyanshu	4	27.04
Dibyanshu	5	25.56
Harikrishnan Shaji	4	21.44
Harikrishnan Shaji	5	32.28
Hrisikesh Neogi	4	20.53
Hrisikesh Neogi	5	37.04
Hyder Abbas	4	0.34

16. Perform inner join, left join and right join based on the agent column and after joining the table export that data into your local system.

Inner join:

```
hive> insert overwrite LOCAL DIRECTORY '/home/cloudera/hive_data'
    > ROW FORMAT DELIMITED
    > Fields terminated by ','
    > Select p.date,p.agent_name,p.total_chats,p.average_response_time,
    > p.average_resolution_time,p.average_rating,p.total_feedback,
    > l.Login_Time,l.Logout_Time,L.duration
    > from agentperformance p
    > inner join
    > agentlogingreport l
    > on p.agent_name = l.agent_name and
    > from_unixtime(unix_timestamp(p.date,'MM/dd/yyyy'))=from_unixtime(unix_timestamp(l.date,'dd-MMM-yy'));
Query ID = cloudera_20220926071616_1d169501-fbfc-488c-97db-a19bcb99a425
```

Left join:

```
hive> insert overwrite LOCAL DIRECTORY '/home/cloudera/hive_data/left_join'
    > ROW FORMAT DELIMITED
    > Fields terminated by ','
    > Select p.date,p.agent_name,p.total_chats,p.average_response_time,
    > p.average_resolution_time,p.average_rating,p.total_feedback,
    > l.Login_Time,l.Logout_Time,L.duration
    > from agentperformance p
    > Left join
    > agentlogingreport l
    > on p.agent_name = l.agent_name and
    > from_unixtime(unix_timestamp(p.date,'MM/dd/yyyy'))=from_unixtime(unix_timestamp(l.date,'dd-MMM-yy'));
Query ID = cloudera_20220926072626_091d6a8d-8450-4577-8c2d-6babf3dd6cf1
```

Right Join:

cloudera@quickstart:~

```
hive> insert overwrite LOCAL DIRECTORY '/home/cloudera/hive_data/right_join'
    > ROW FORMAT DELIMITED
    > Fields terminated by ','
    > Select p.date,p.agent_name,p.total_chats,p.average_response_time,
    > p.average_resolution_time,p.average_rating,p.total_feedback,
    > l.Login_Time,l.Logout_Time,L.duration
    > from agentperformance p
    > right join
    > agentlogingreport l
    > on p.agent_name = l.agent_name and
    > from_unixtime(unix_timestamp(p.date,'MM/dd/yyyy'))=from_unixtime(unix_timestamp(l.date,'dd-MMM-yy'));
Query ID = cloudera_20220926072727_21e01172-a8bd-45e4-ab13-a2f8db9fd4f3
```

17. Perform partitioning on top of the agent column and then on top of that perform bucketing for each partitioning.

```
hive> CREATE TABLE agentperformance_par(
    >   sl_no int,
    >   date date,
    >   total_chats int,
    >   average_response_time string,
    >   average_resolution_time string,
    >   average_rating float,
    >   total_feedback int)
    >   partitioned by(agent_name string)
    >   clustered by (date)
    >   sorted by (date)
    >   into 10 buckets
    >   ;
OK
Time taken: 0.697 seconds
```

cloudera@quickstart:~

```
hive>    insert overwrite table agentperformance_part partition(agent_name)
    >    select SL_No,to_date(from_unixtime(unix_timestamp(date,'MM/dd/yyyy'))),
    >    total_chats,average_response_time,average_resolution_time,
    >    average_rating,total_feedback,agent_name
    >    from agentperformance;
Query ID = cloudera_20220926093434_3e861360-f8c7-4f18-8055-ba6b6a9b69bf
Total jobs = 3
```