

I have completed the project as per below:

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-JPCZ34dyN6k9CqJa-Y8yxIGq6vTVXU/view>

Note: both files are csv files.

1. Create a schema based on the given dataset
2. Dump the data inside the hdfs in the given schema location.
3. List of all agents' names.
4. Find out agent average rating.
5. Total working days for each agents
6. Total query that each agent have taken
7. Total Feedback that each agent have received
8. Agent name who have average rating between 3.5 to 4
9. Agent name who have rating less than 3.5
10. Agent name who have rating more than 4.5
11. How many feedback agents have received more than 4.5 average
12. average weekly response time for each agent
13. average weekly resolution time for each agents

14. Find the number of chat on which they have received a feedback
15. Total contribution hour for each and every agents weekly basis
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.
17. Perform partitioning on top of the agent column and then on top of that perform bucketing for each partitioning.