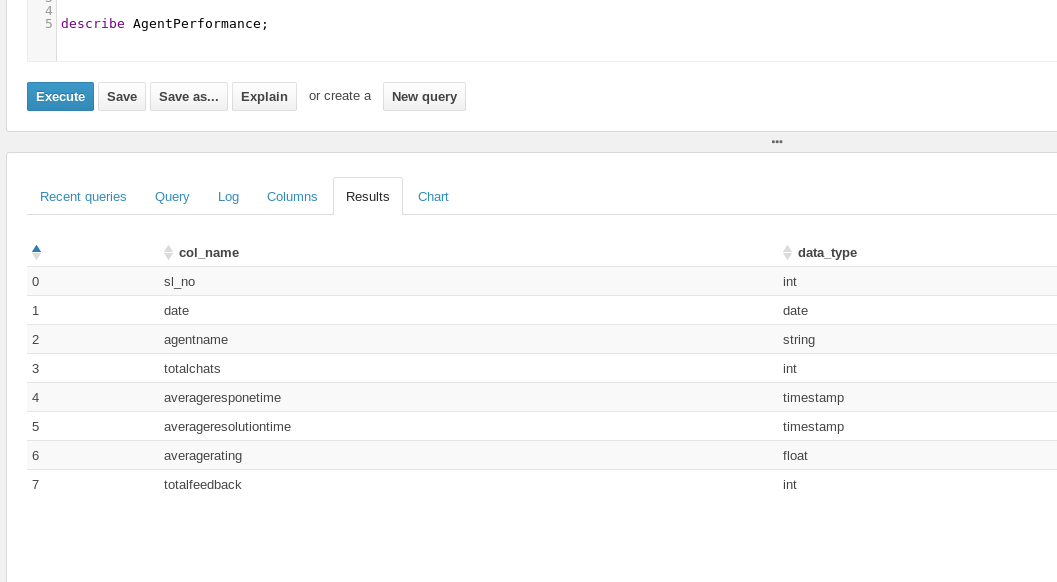
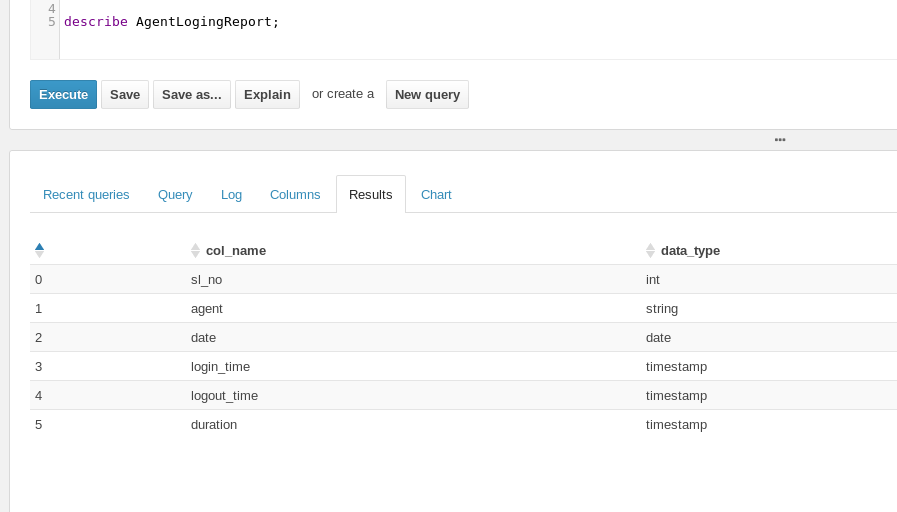
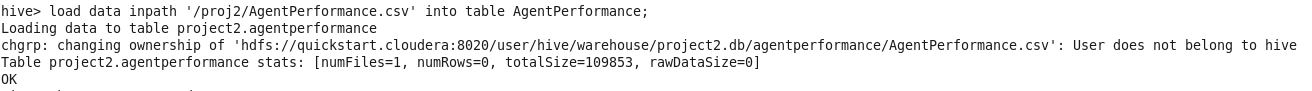
1. **Create a schema based on the given dataset**

****

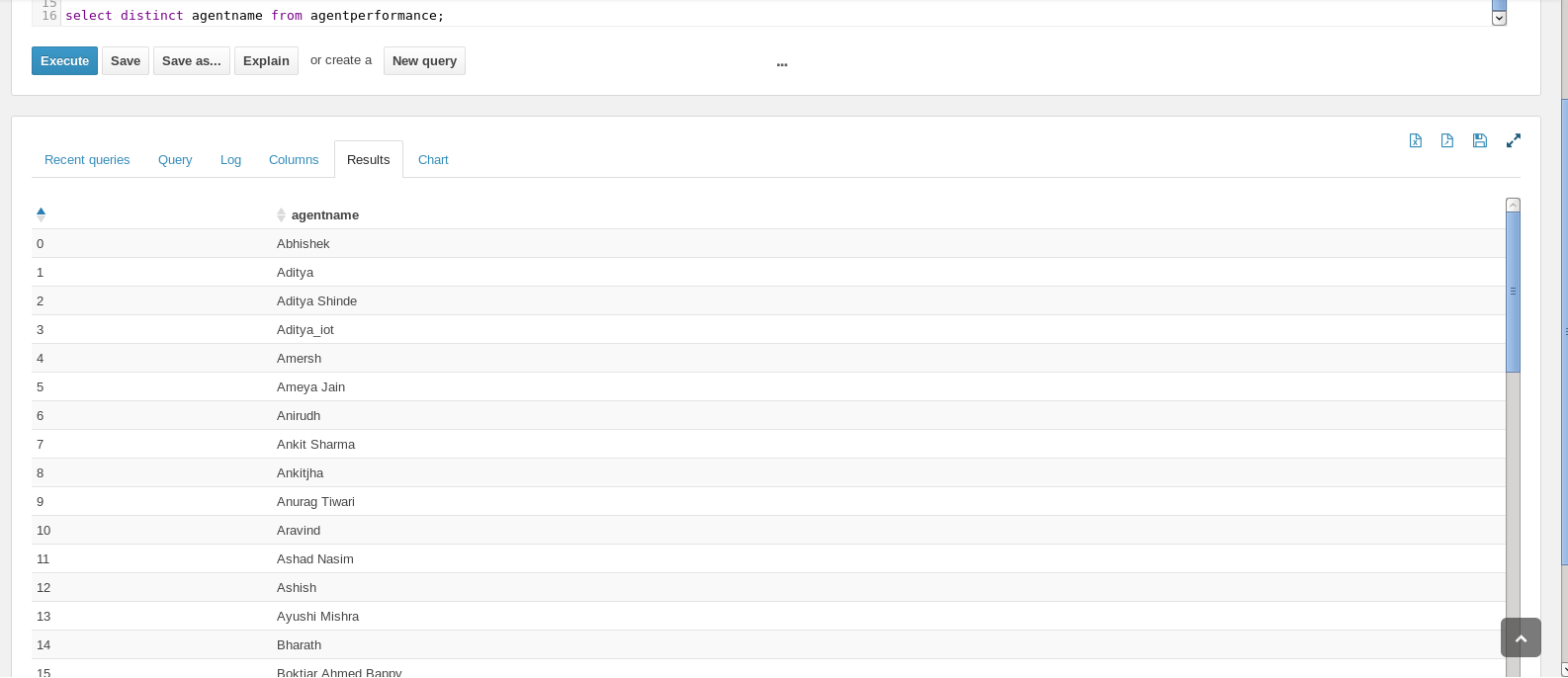
****

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

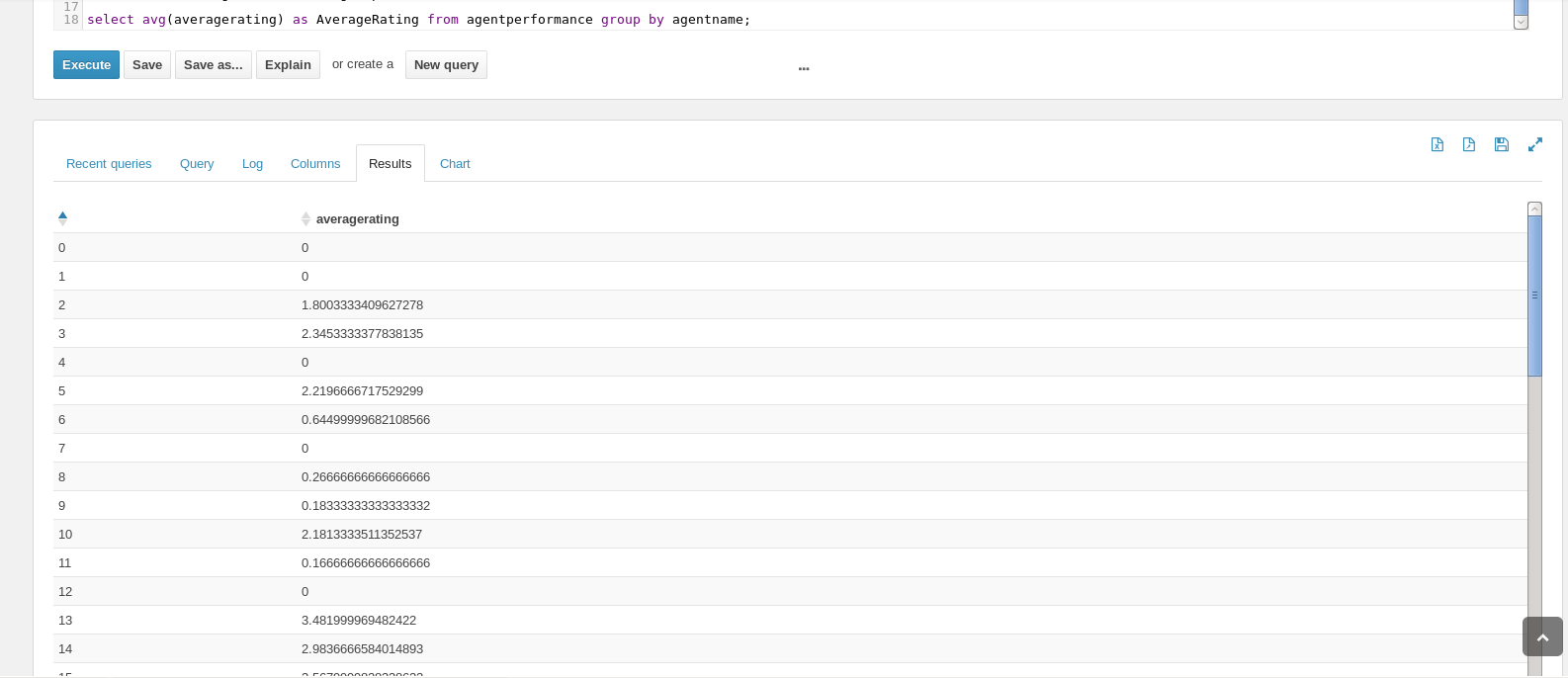
****

****

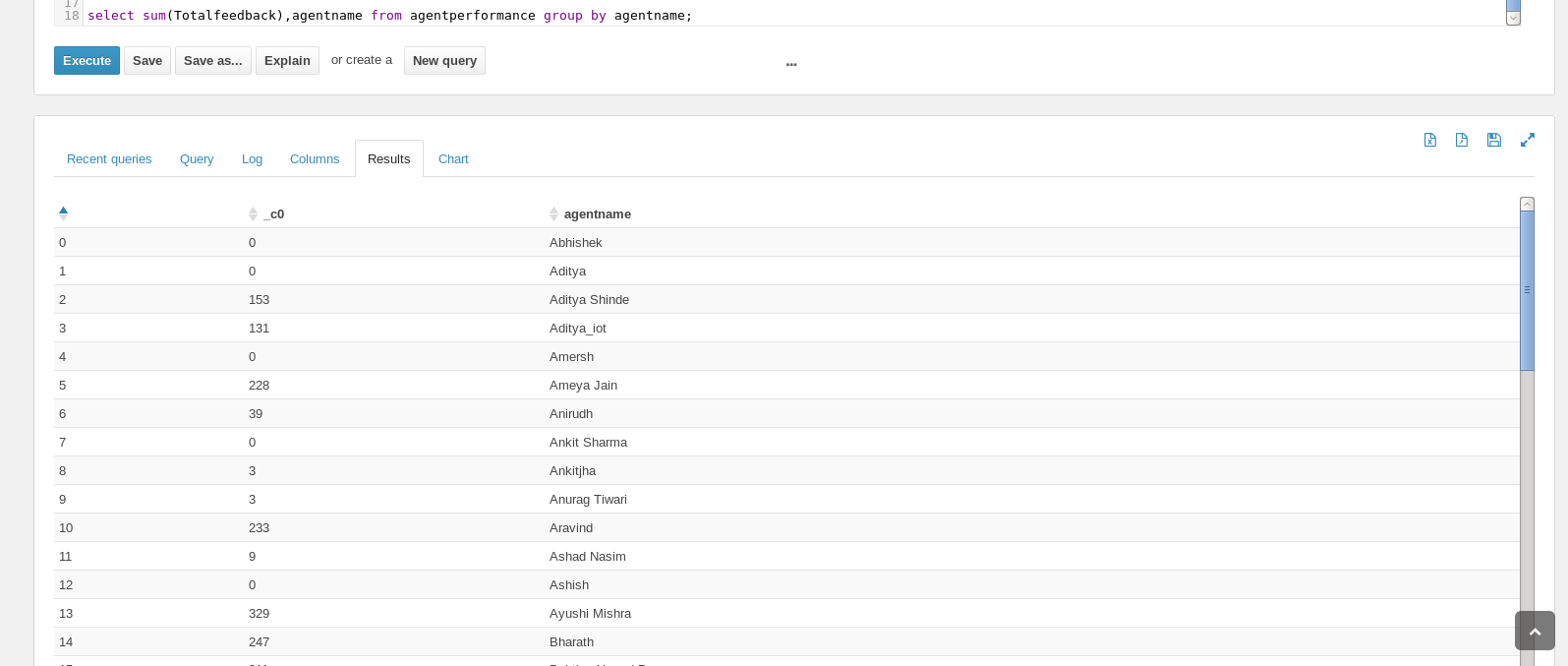
**3. List of all agents' names.**

****

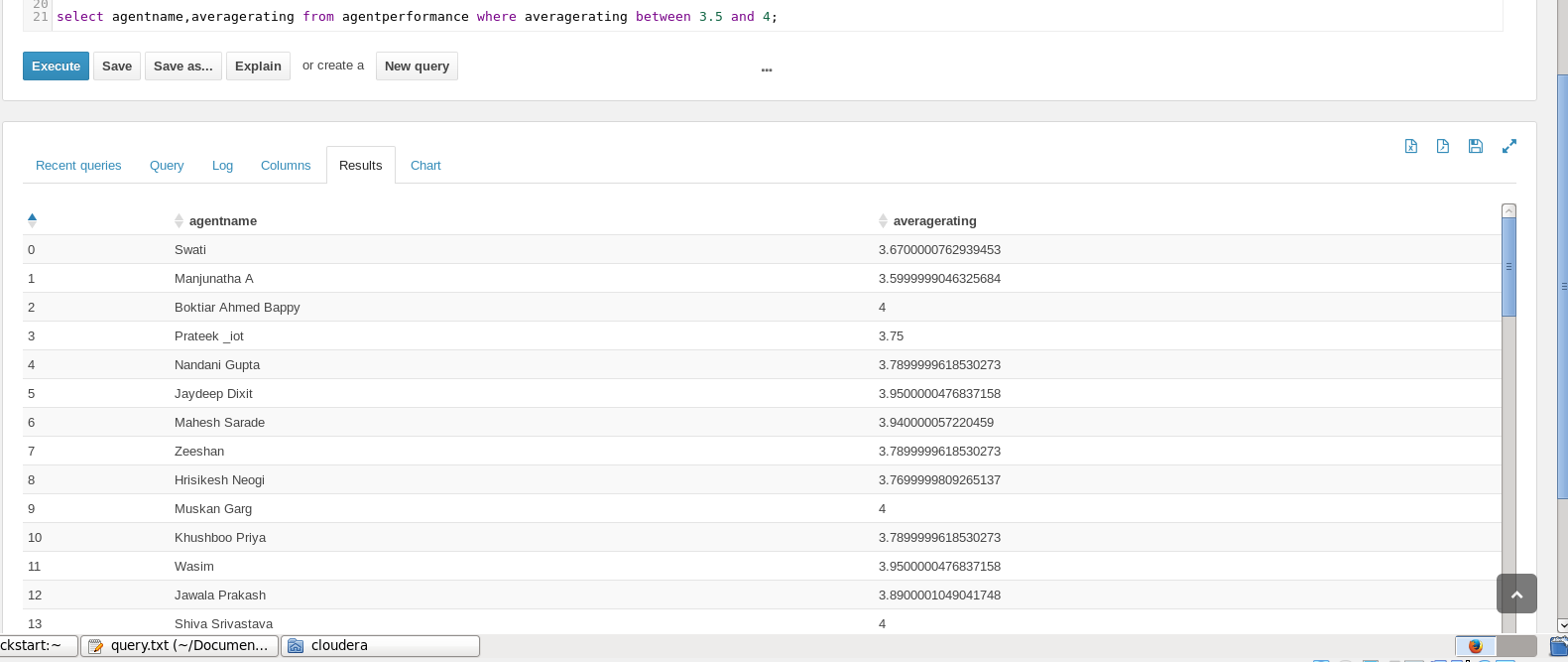
**4. Find out agent average rating.**

****

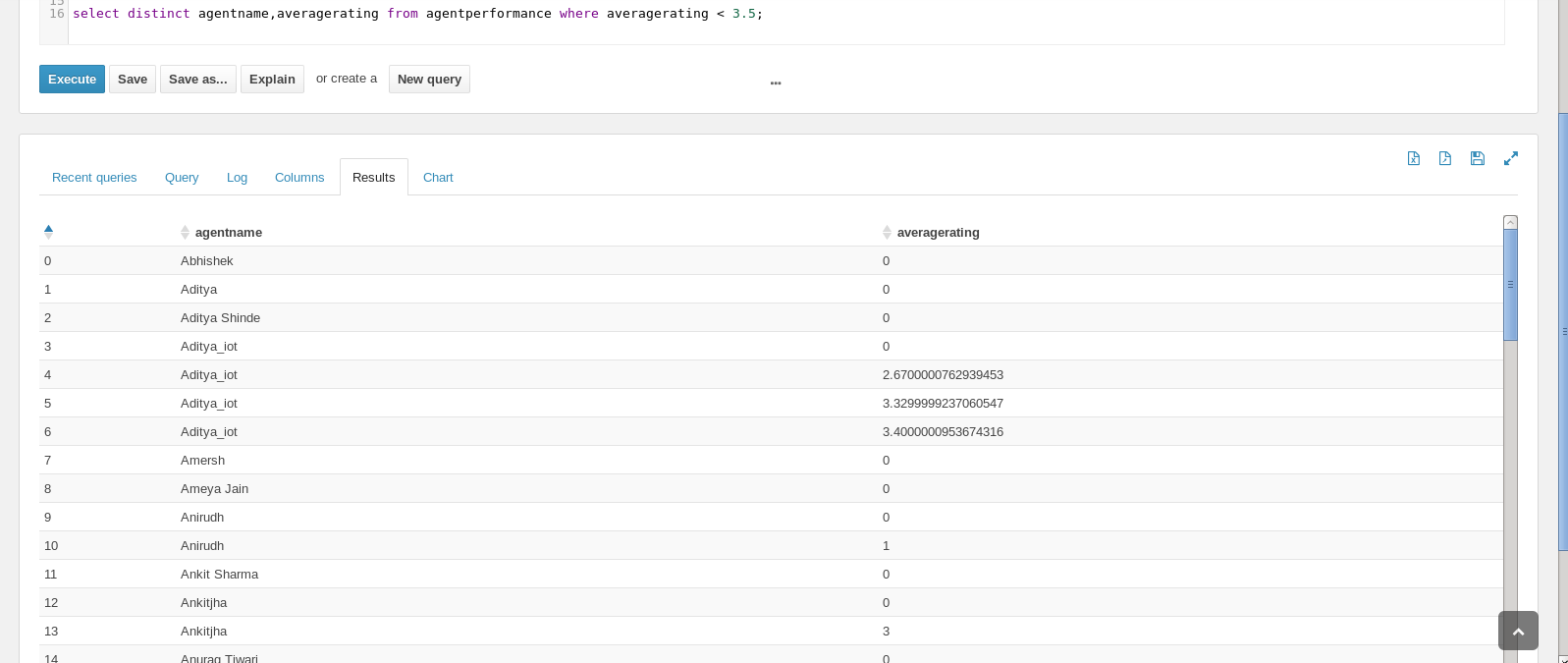
**5. Total Feedback that each agent have received**

****

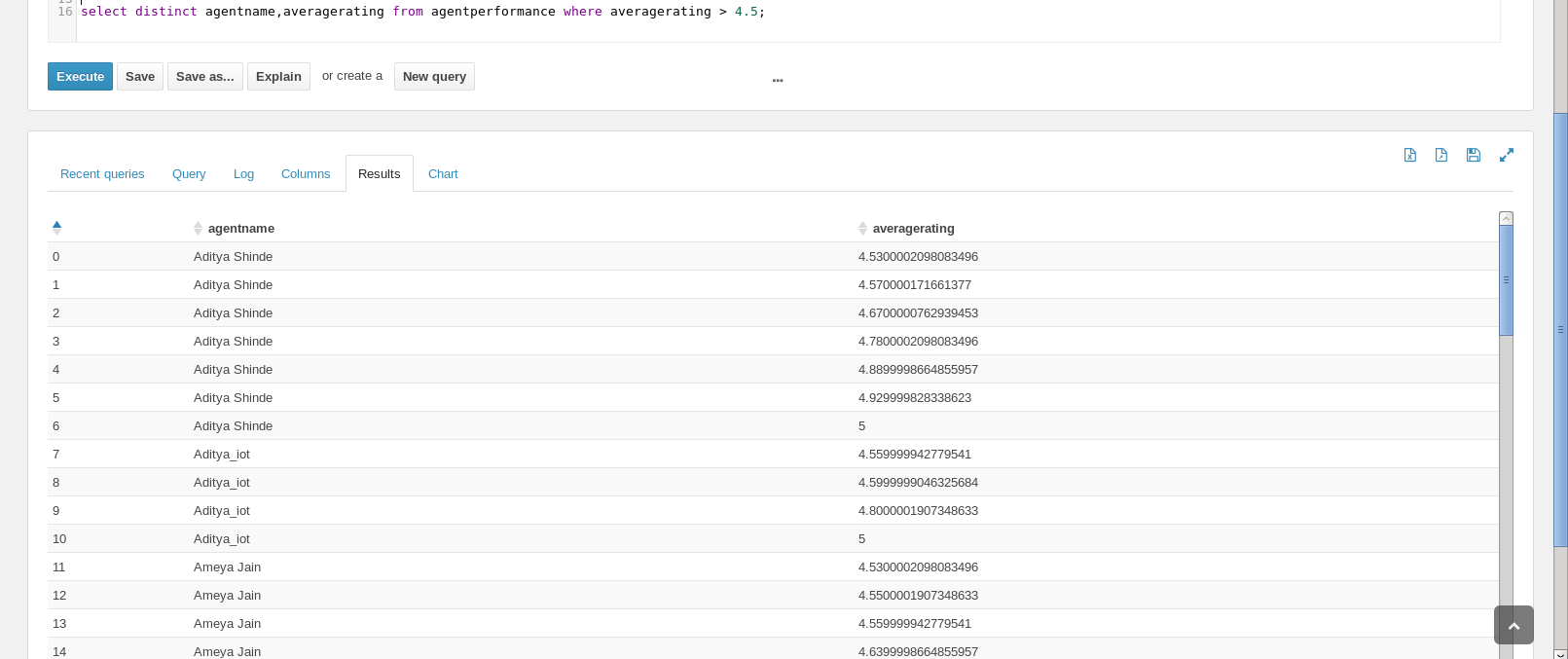
**6. Agent name who have average rating between 3.5 to 4**

****

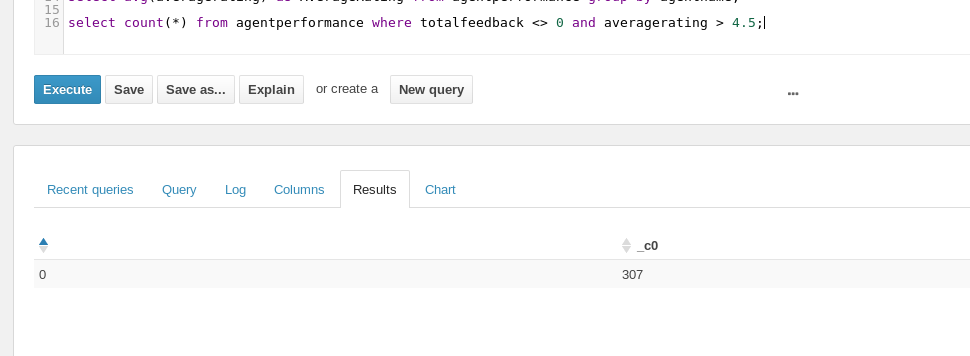
**7. Agent name who have rating less than 3.5**

****

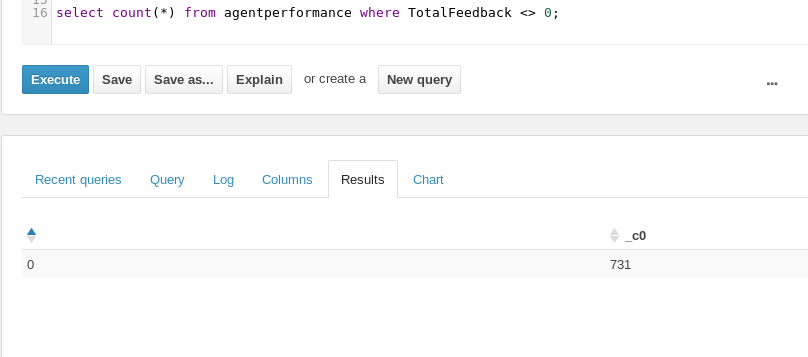
**8. Agent name who have rating more than 4.5**

****

**9. How many feedback agents have received more than 4.5 average**

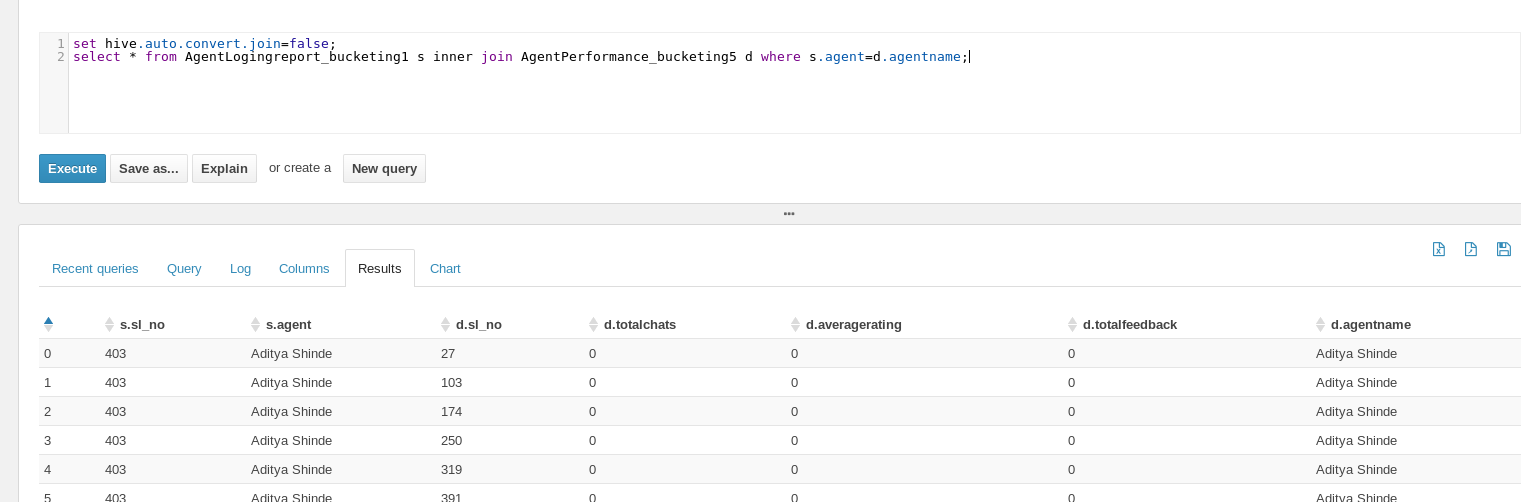
****

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

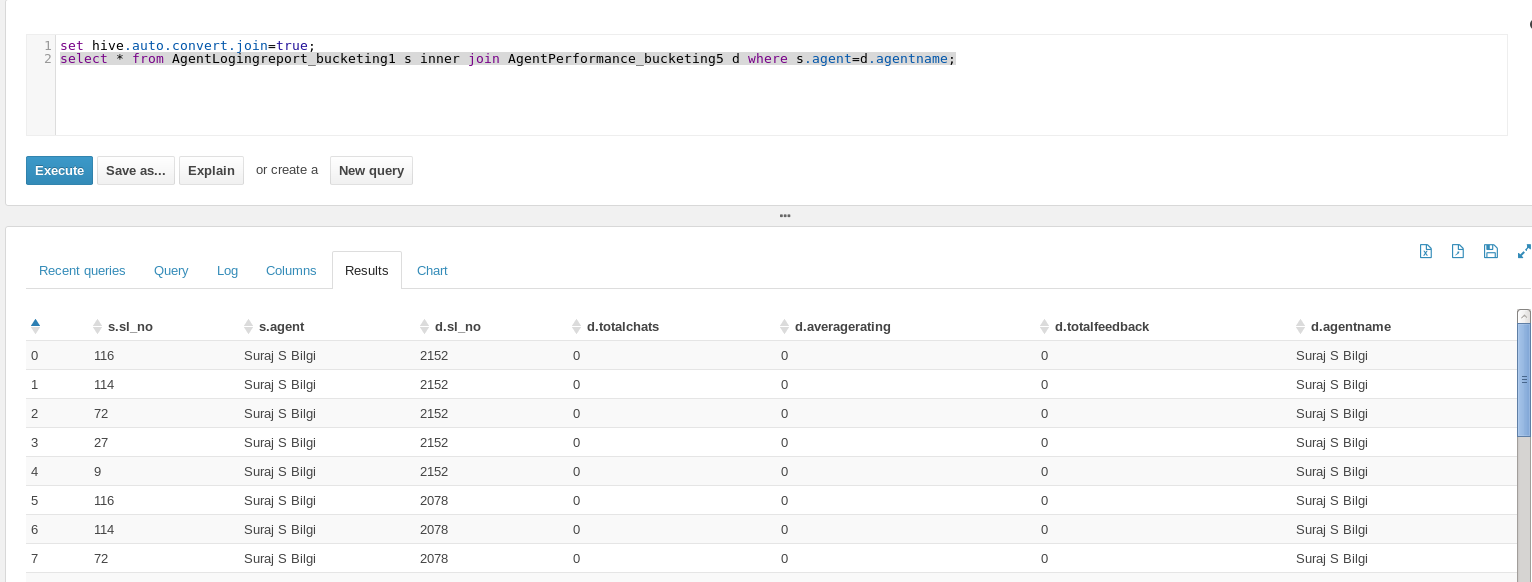
****

**11. 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.**

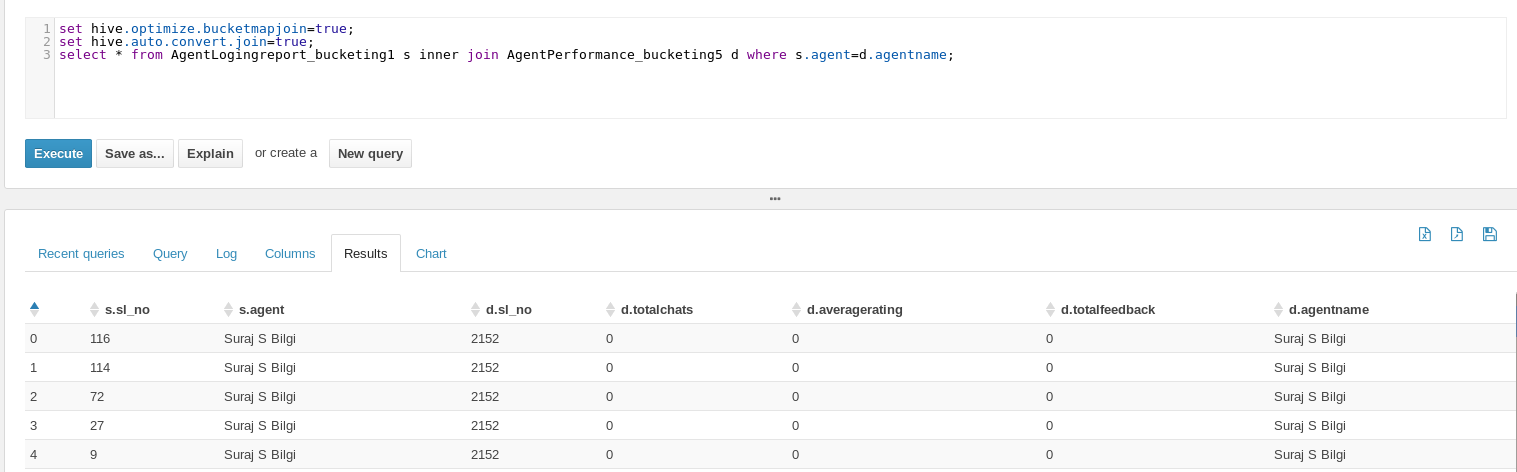
Reduce-Side Join



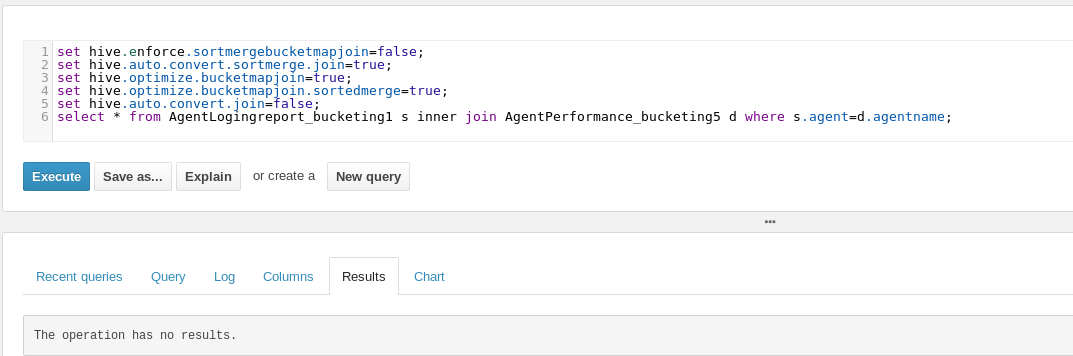
Map Side Join

****

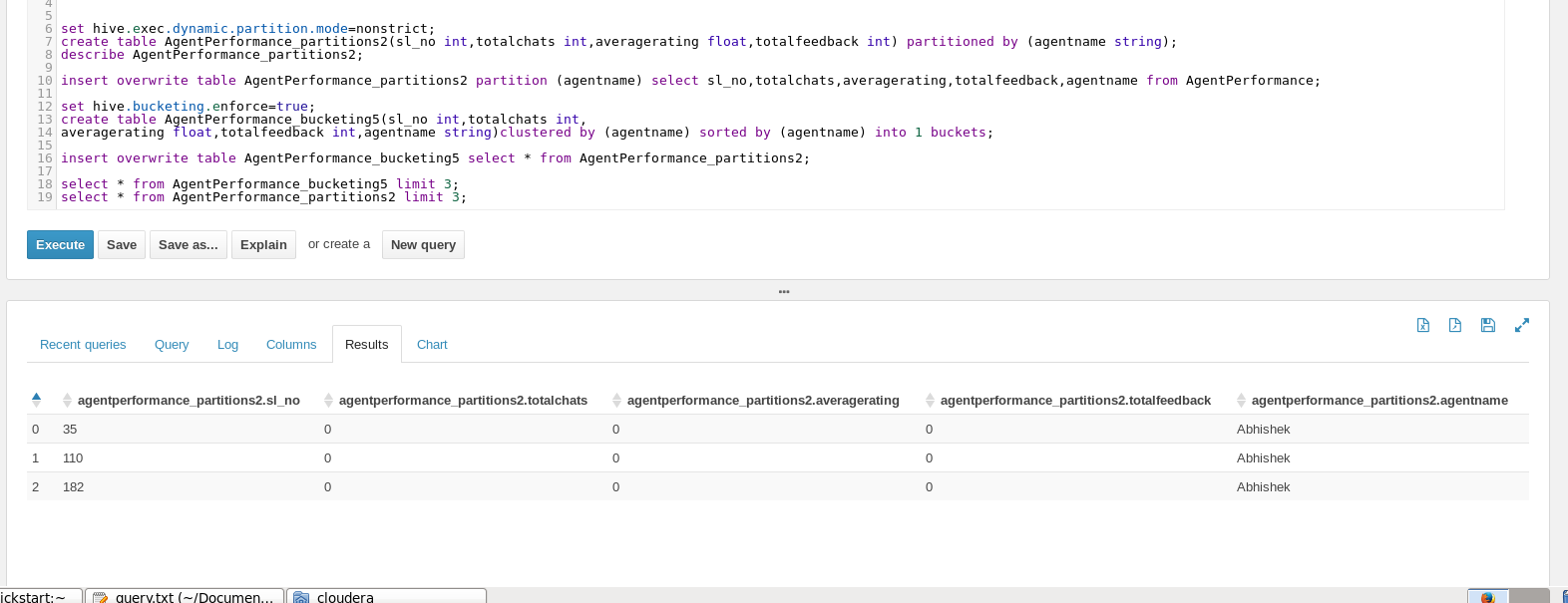
Bucket Map Join

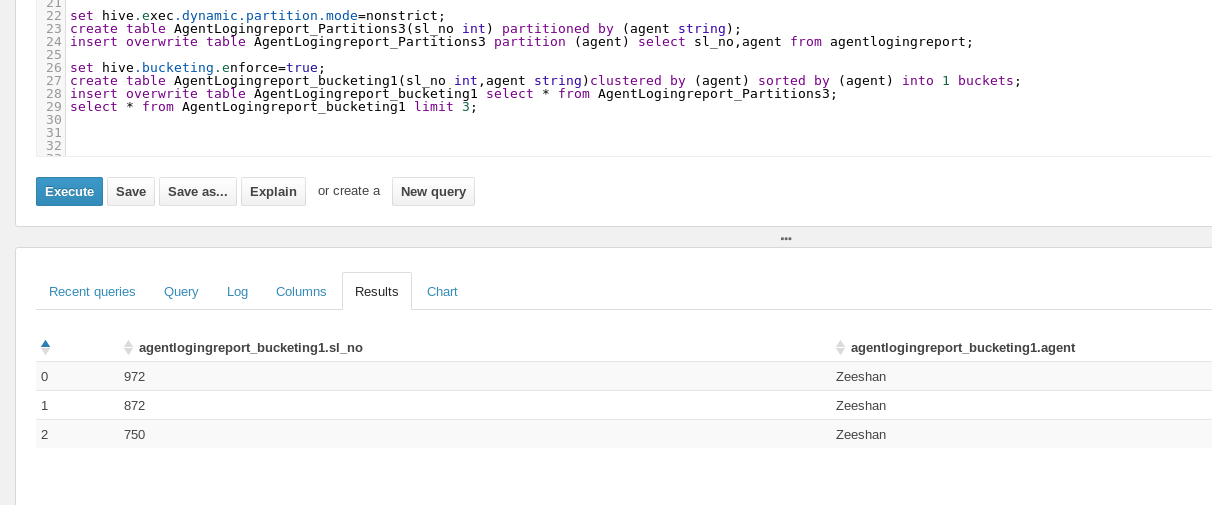


Sorted Merge Bucket Map Join

****

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

****

****