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>

1. Create a schema based on the given dataset

Table for Dataset 1(AgentLogingReport.csv)

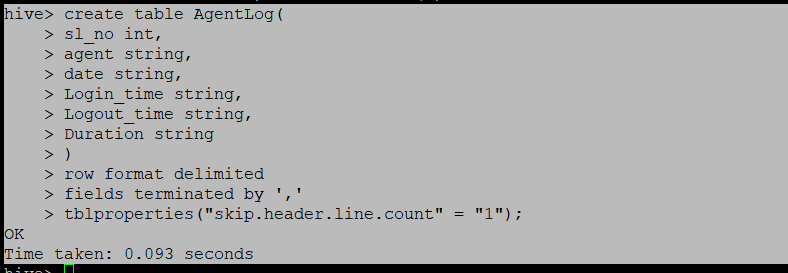
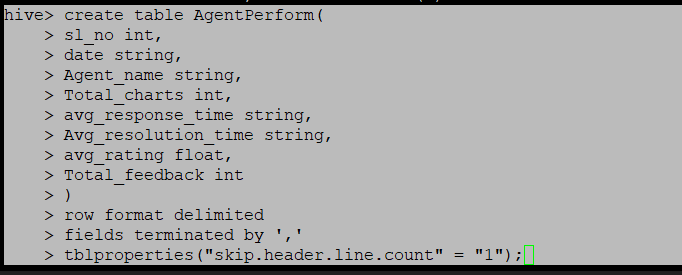
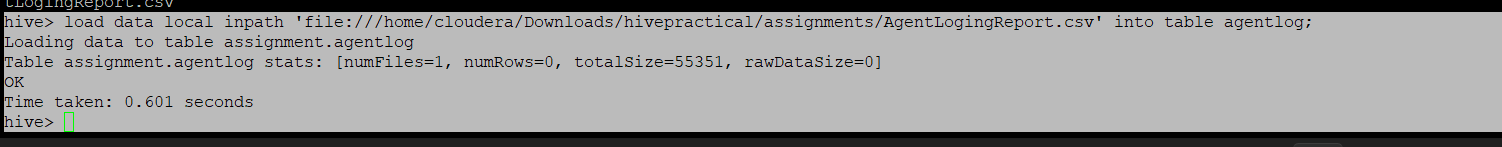


Table for Dataset 2:(AgentPerformance.csv)

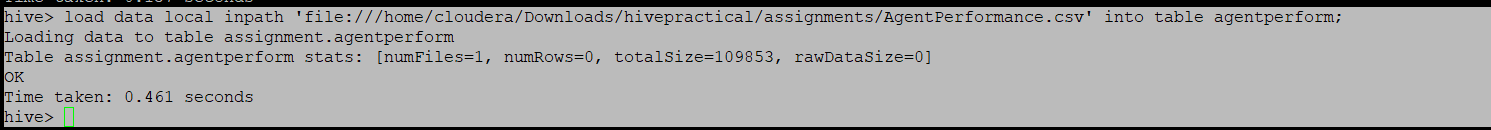


1. Dump the data inside the hdfs in the given schema location.

For Agent Loging Report:

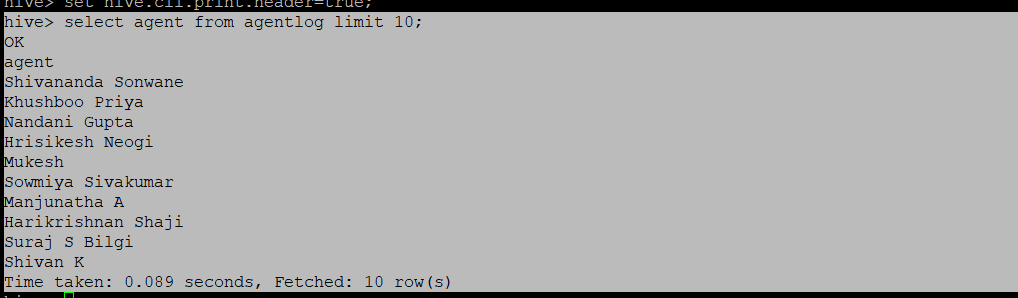


For AgentPerformance:



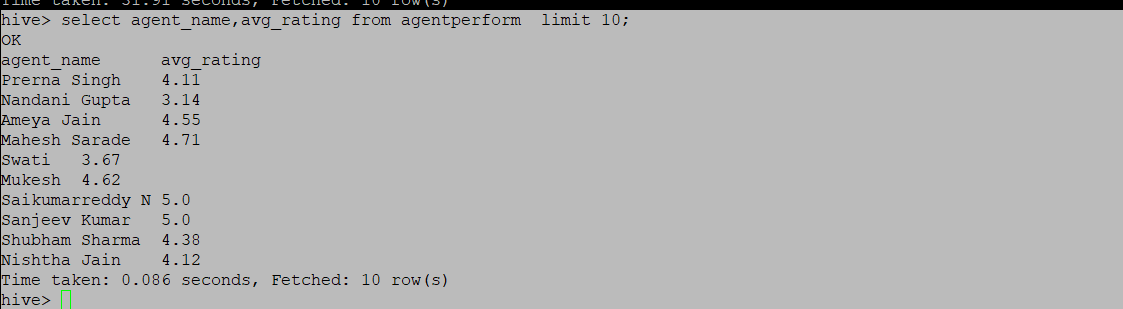
1. List of all agents' names.

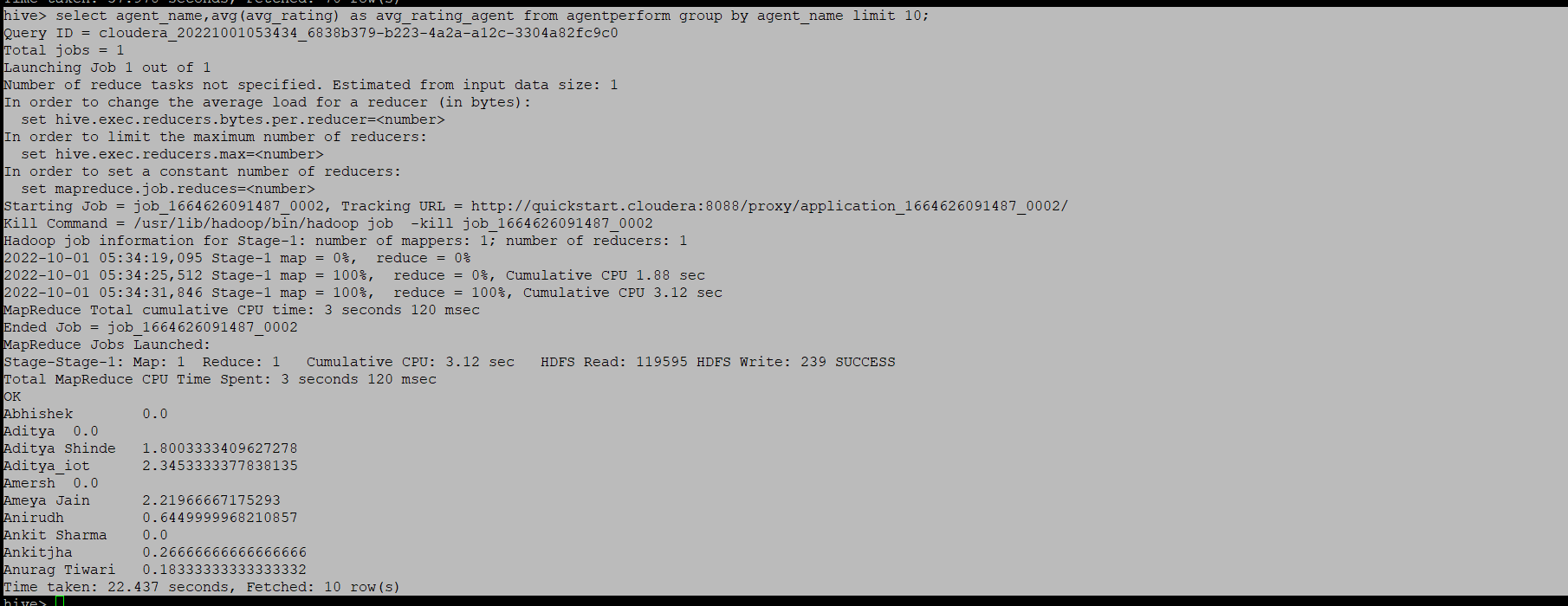
For Agent Loging Report:



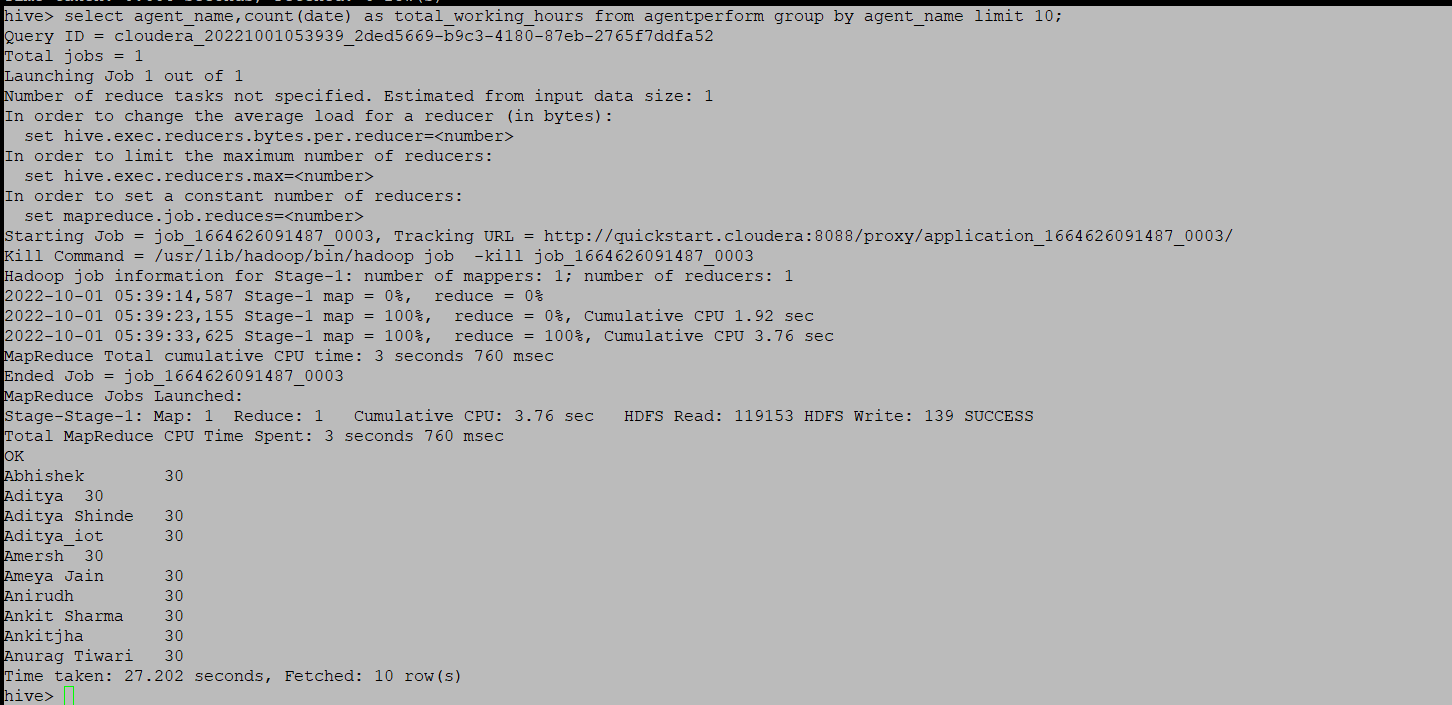
1. Find out agent average rating.

For Agent Performance:

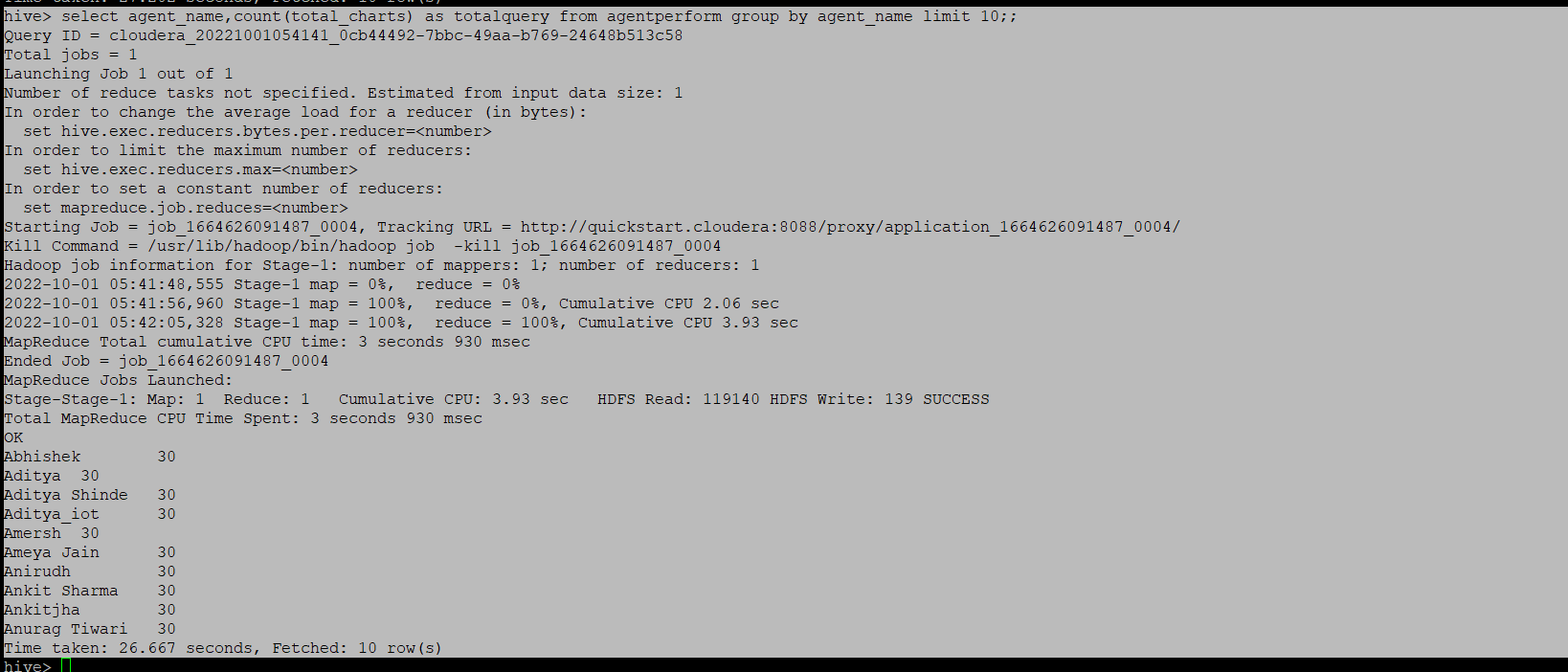




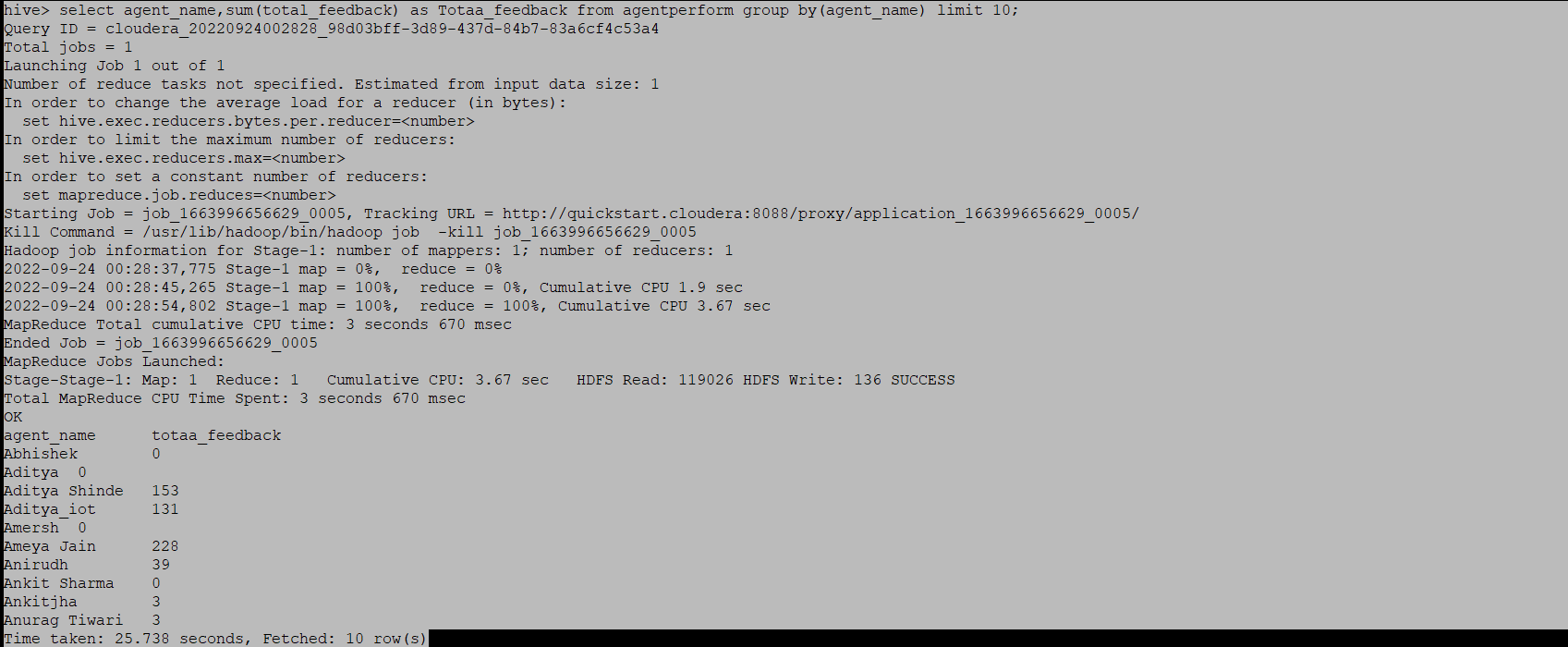
5. Total working days for each agents



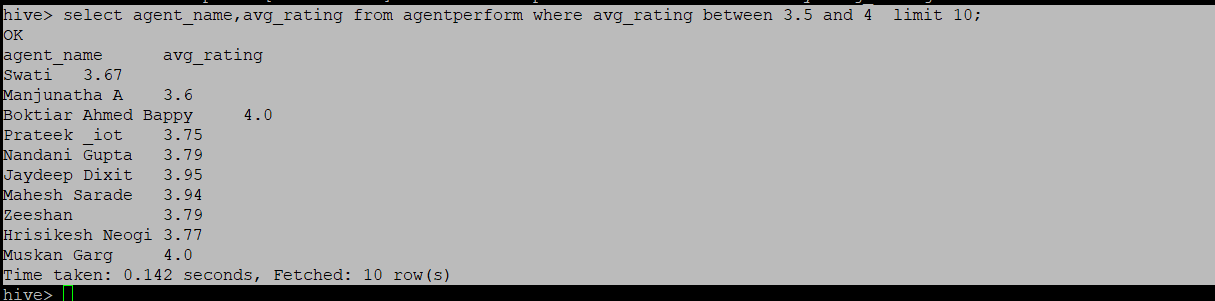
6.Total query that each agent have taken



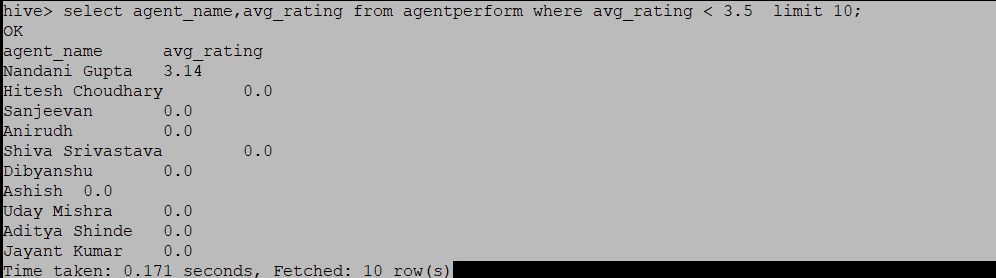
7. Total Feedback that each agent have received



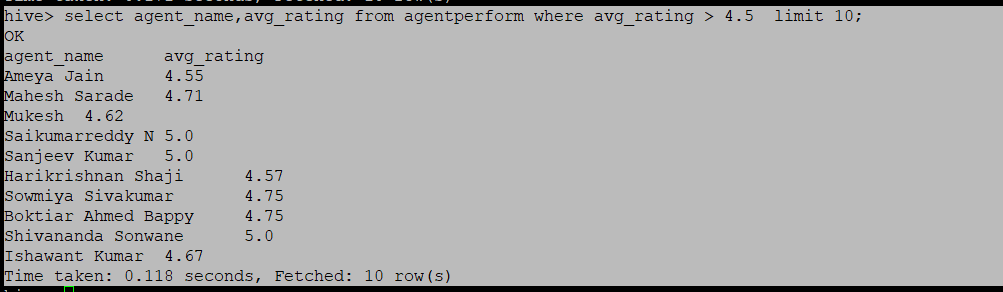
8. Agent name who have average rating between 3.5 to 4



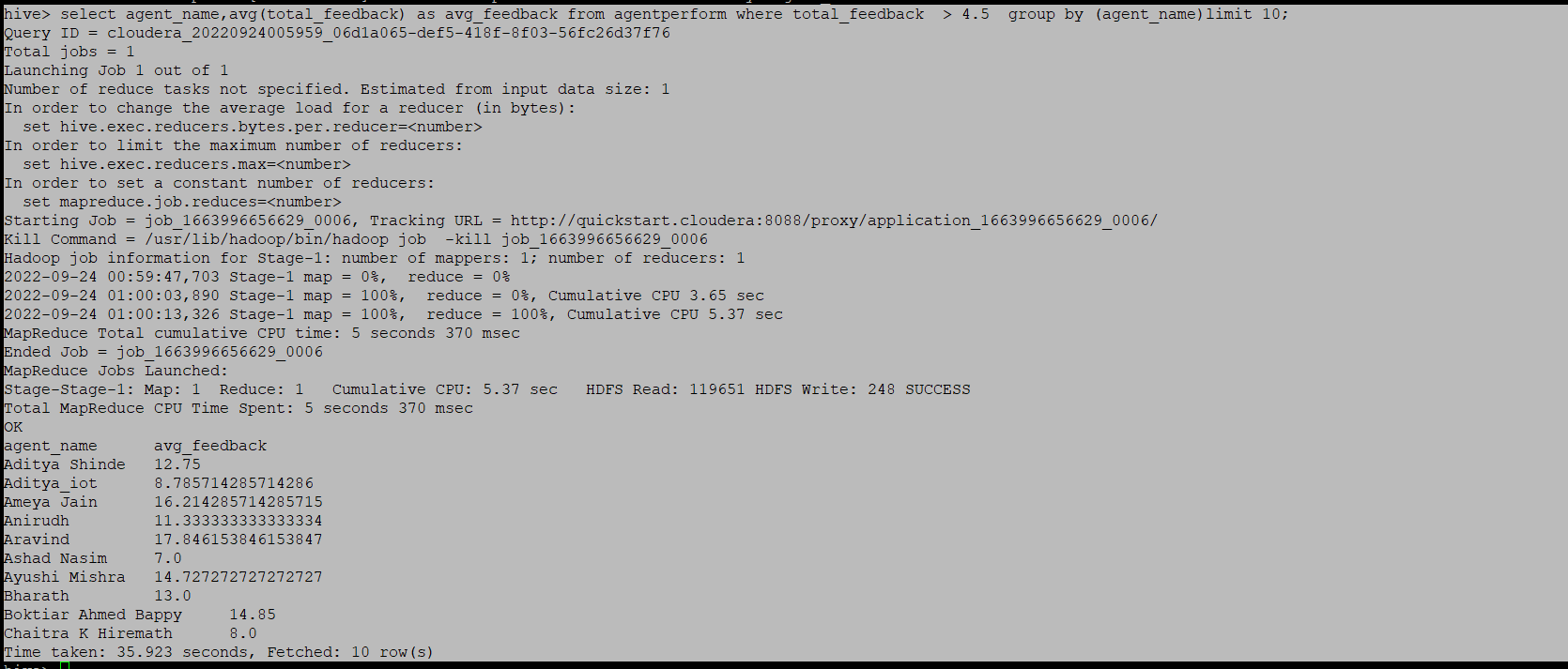
9. Agent name who have rating less than 3.5



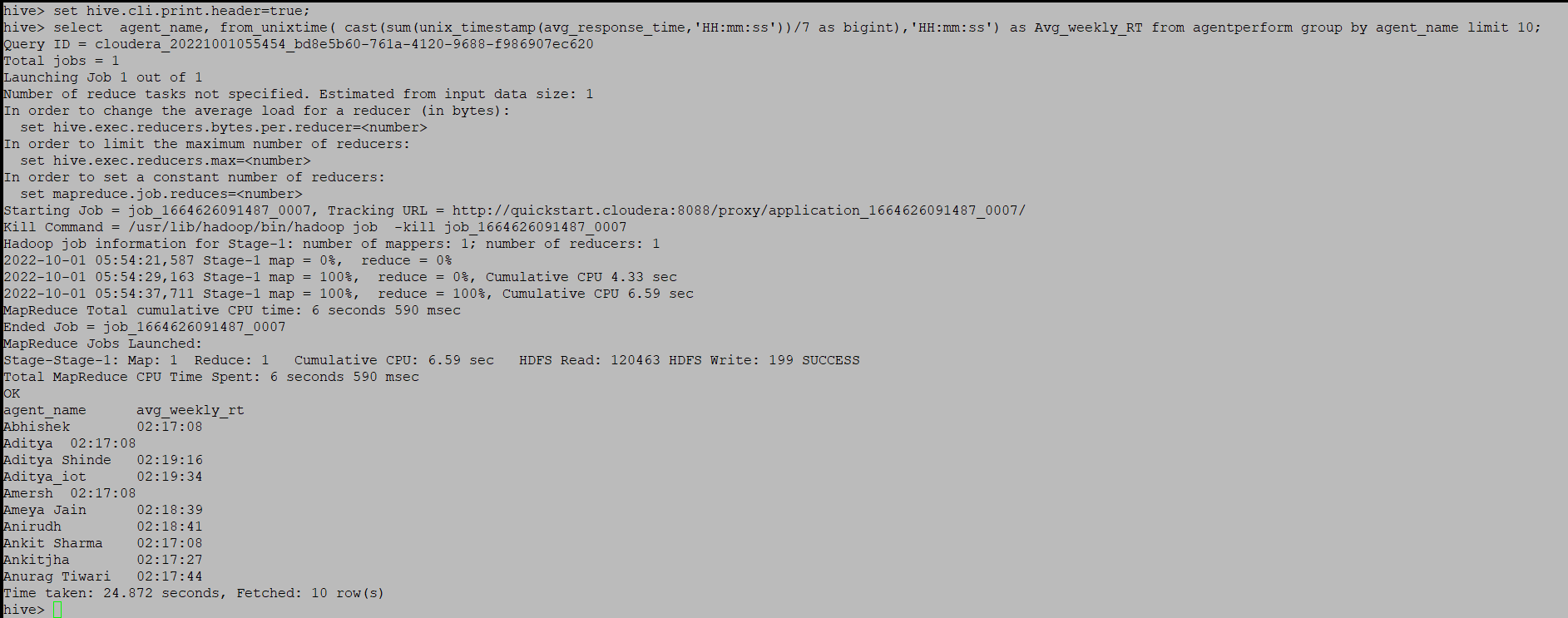
10. Agent name who have rating more than 4.5



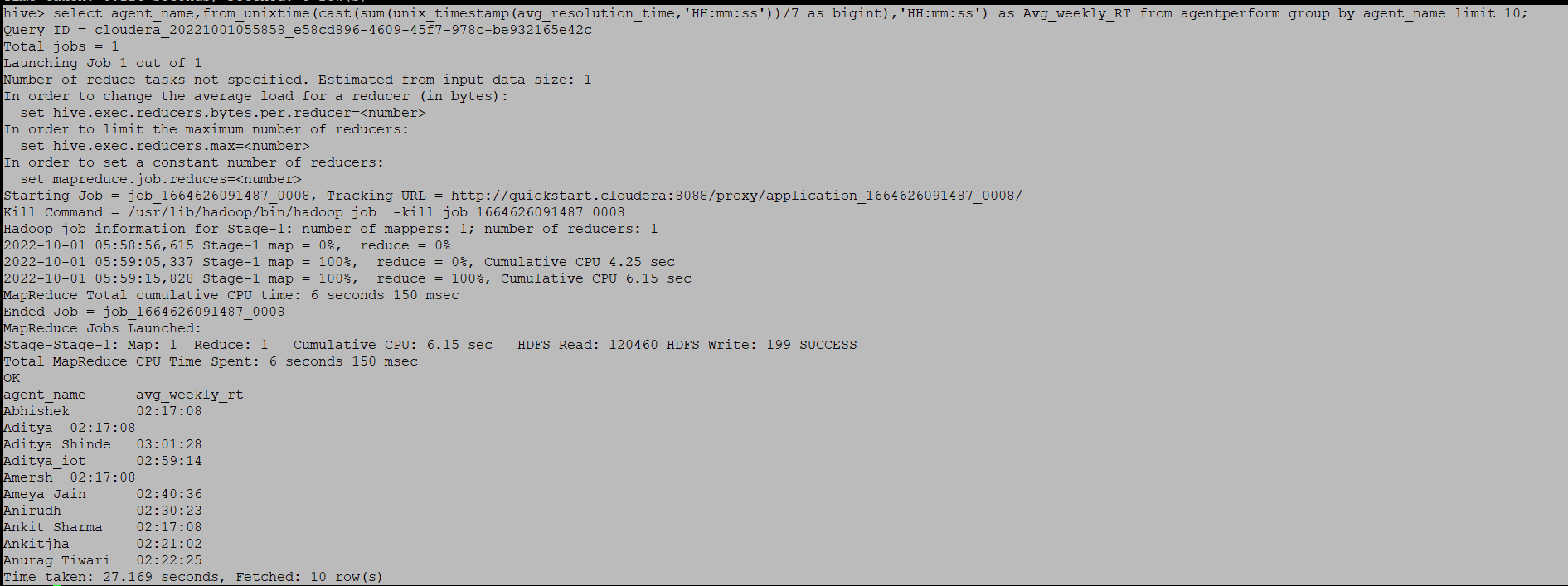
11. How many feedback agents have received more than 4.5 average



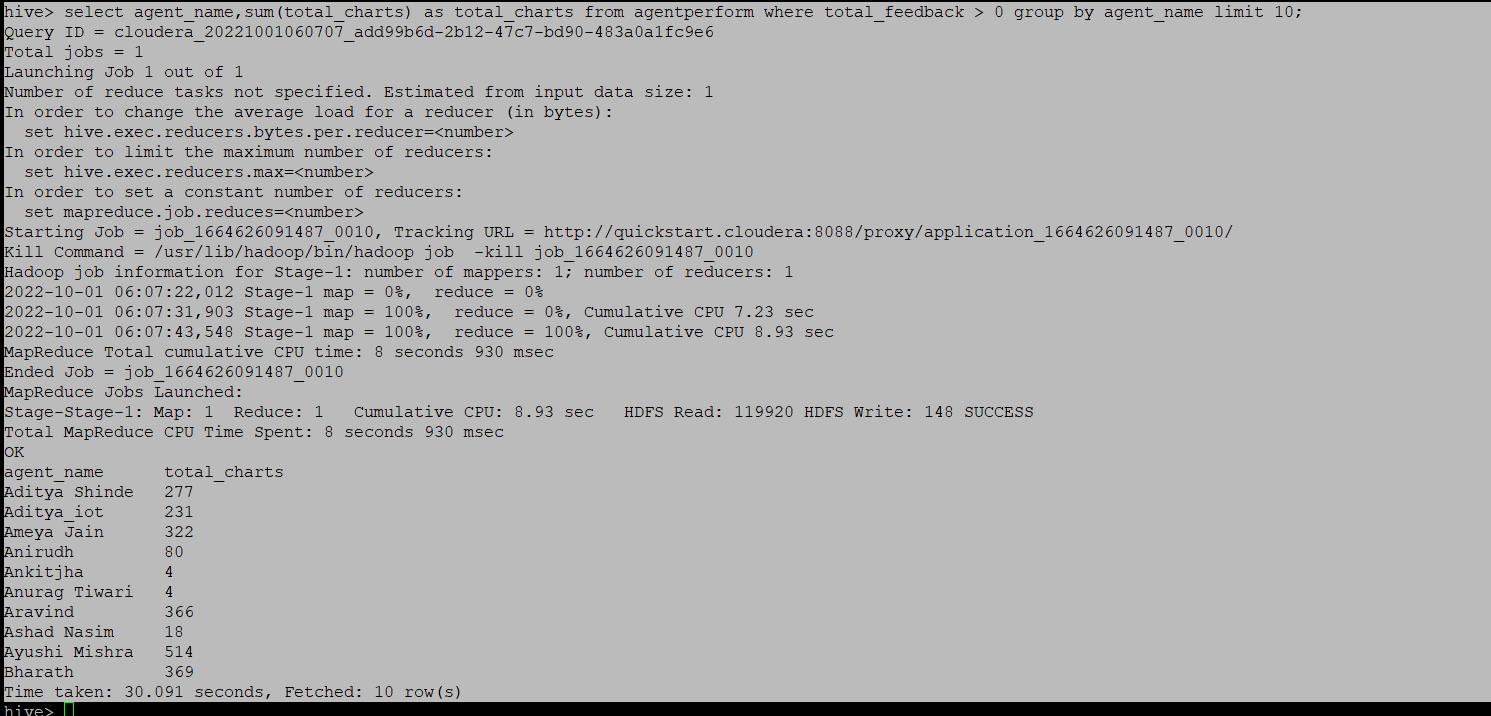
12. average weekly response time for each agent.



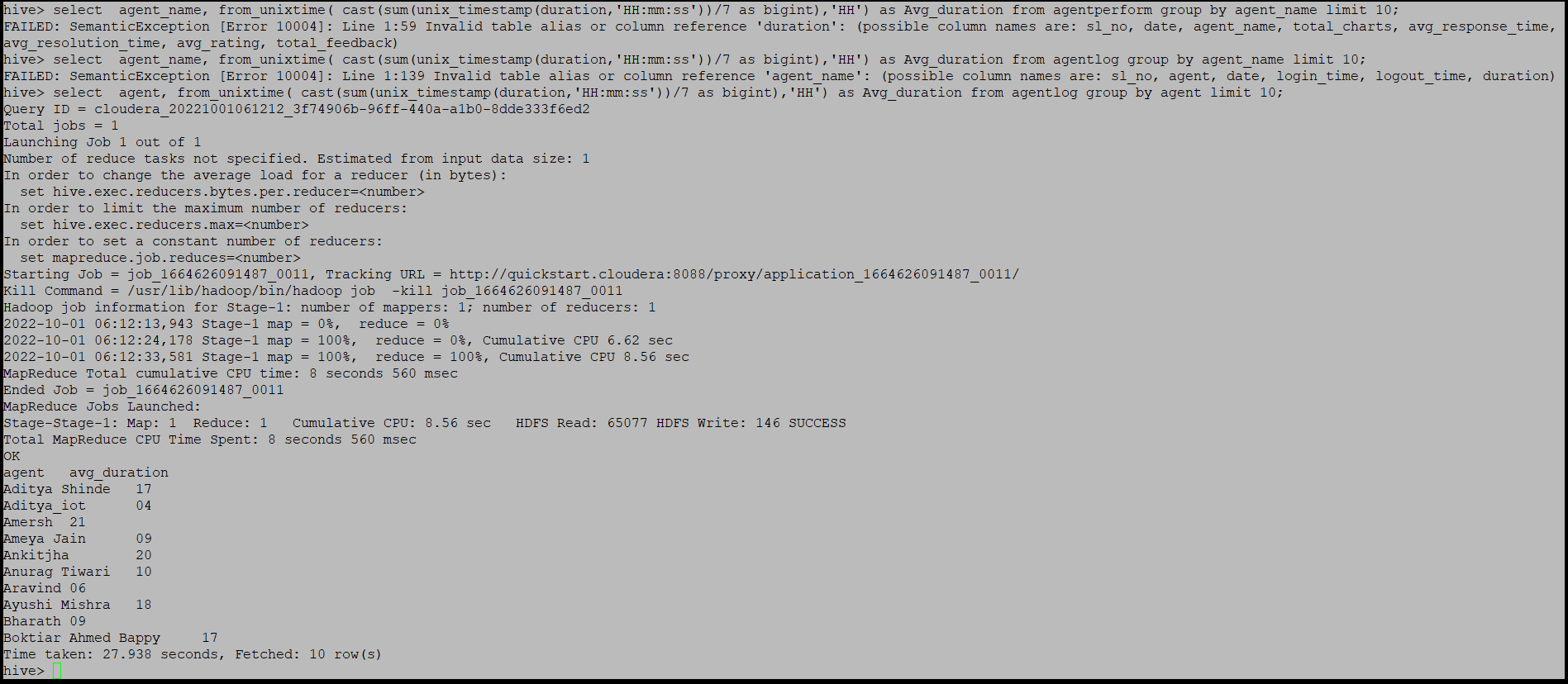
13. average weekly resolution time for each agents



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

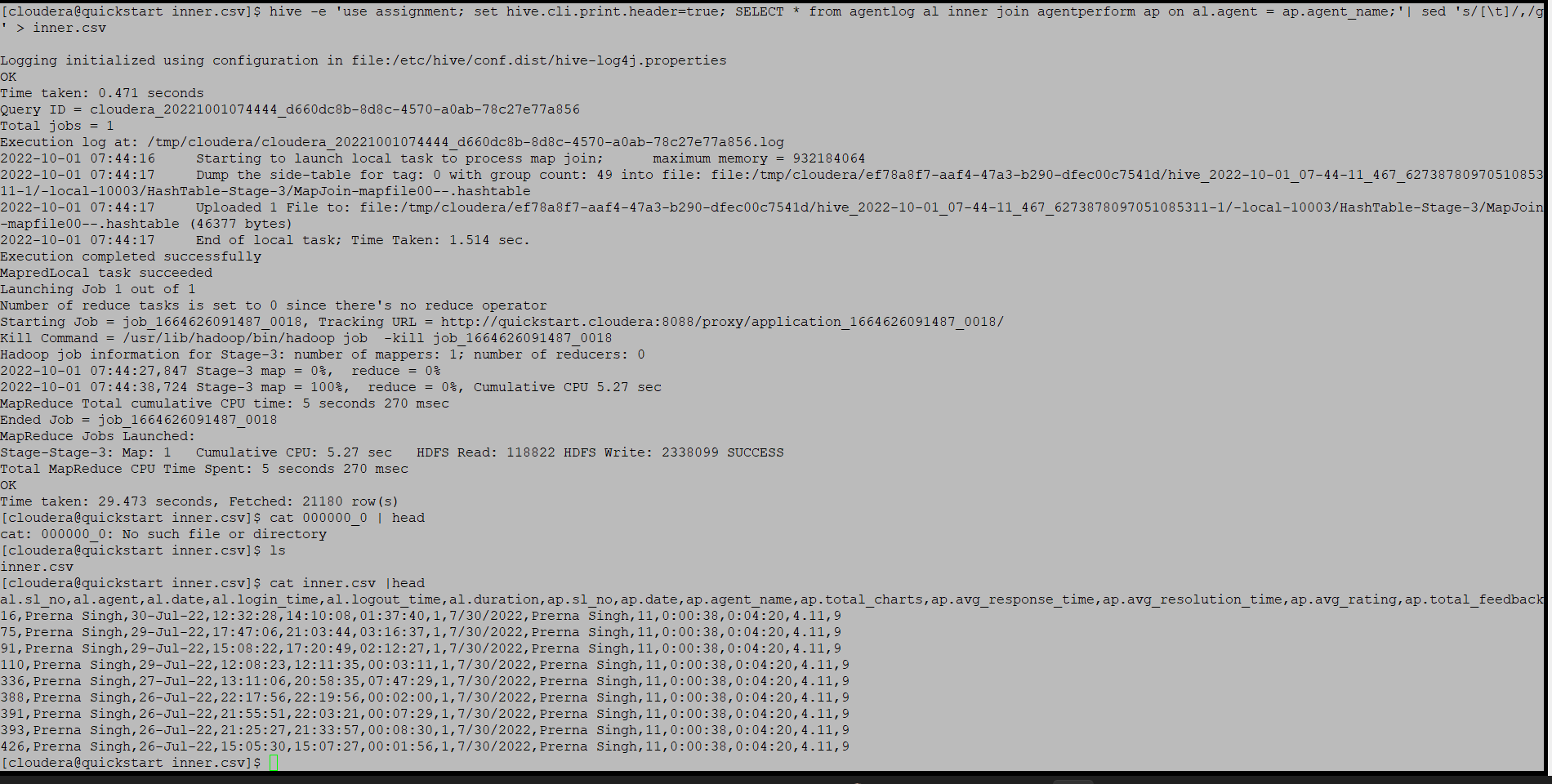


15. Total contribution hour for each and every agents weekly basis

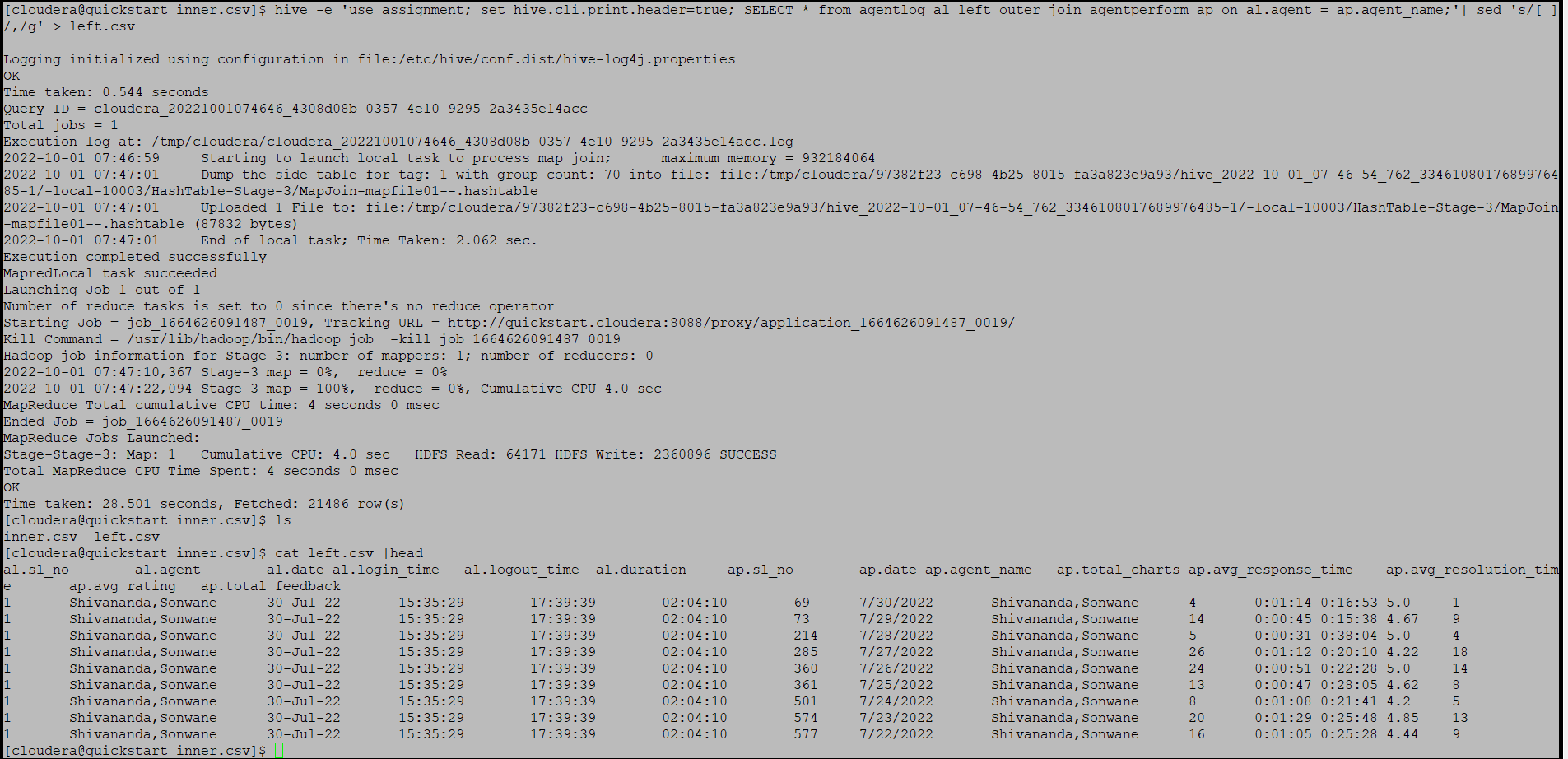


16. Perform inner join, left join and right join based on the agent column and after joining the table export that data into your local system.

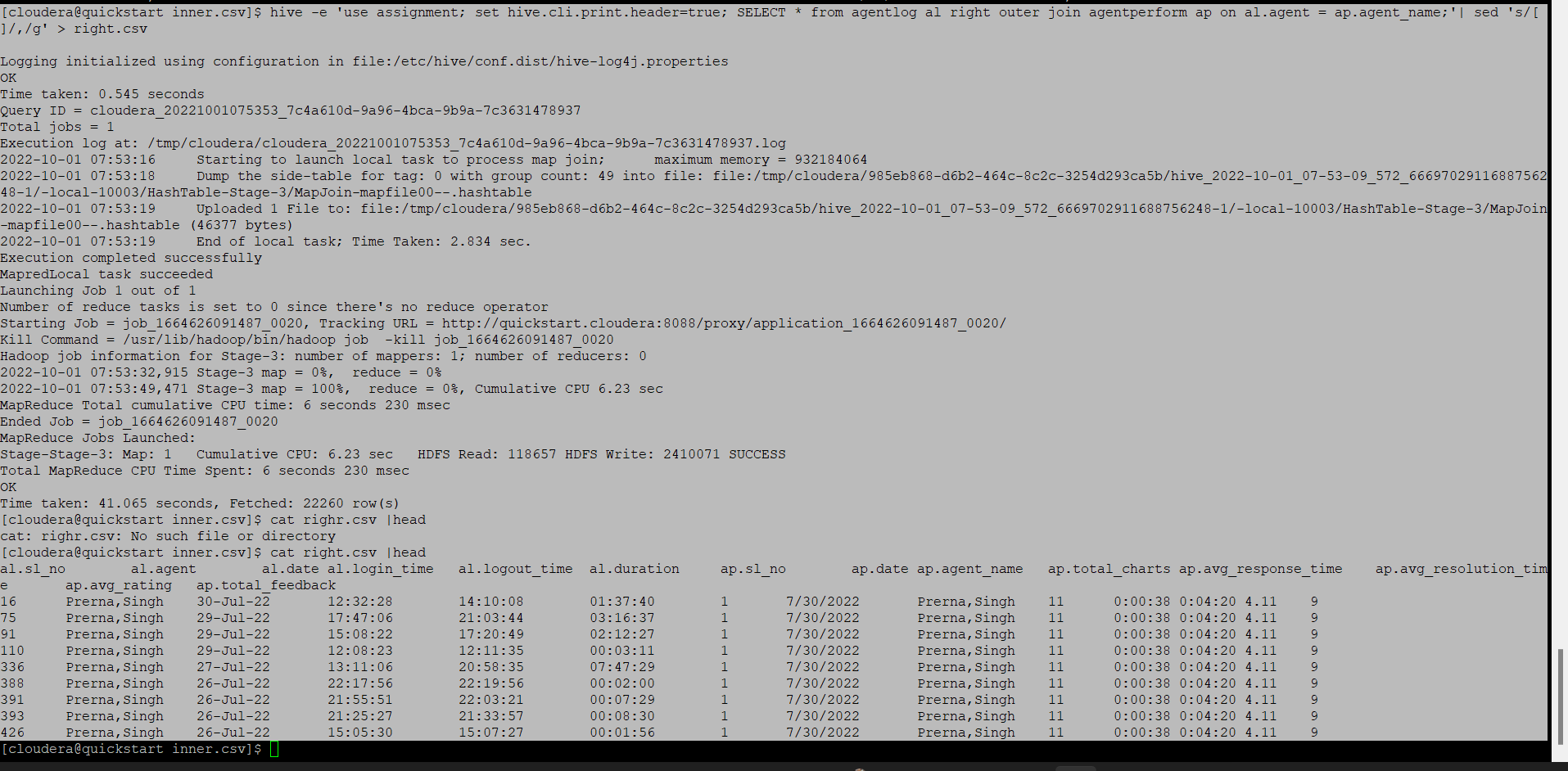
Inner join:



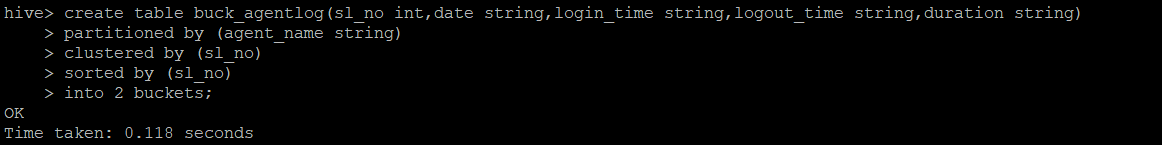
Left Outer Join:

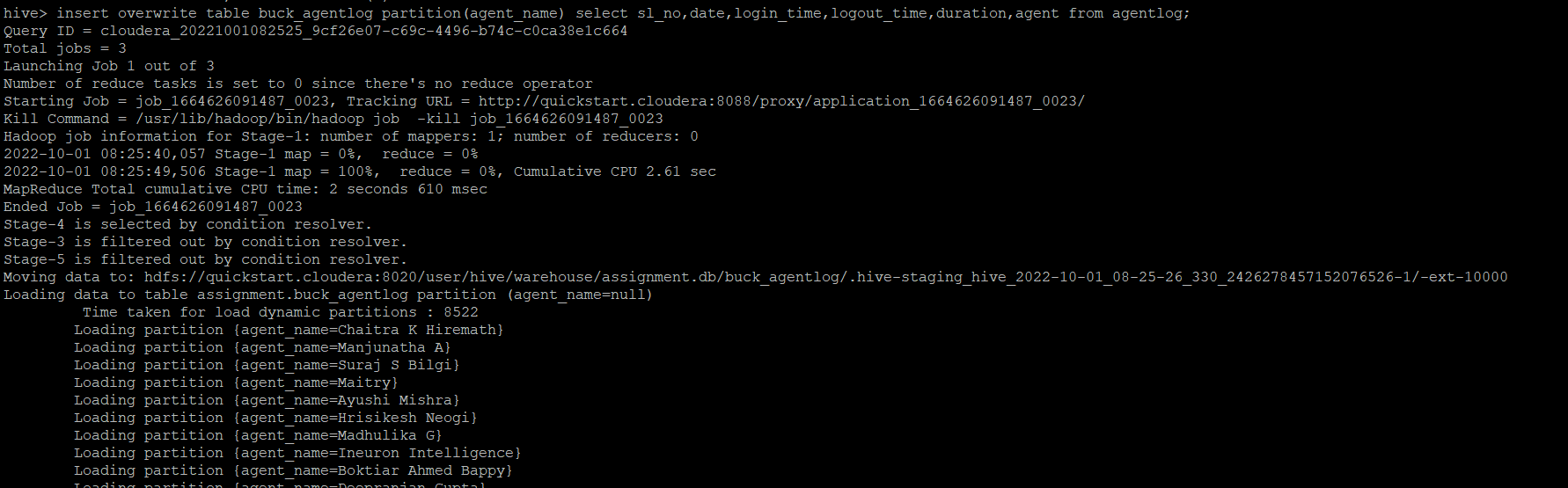


Right Outer Join:



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





For agent performance:

