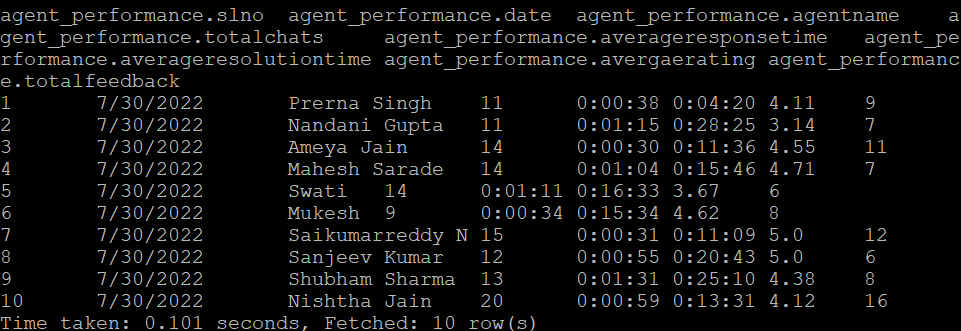
1. Create a schema based on the given dataset

create table agent\_performance (slno int, date string,agentname string, totalchats int, averageresponsetime string, averageresolutiontime string, avergaerating float, totalfeedback int ) row format delimited fields terminated by ',' lines terminated by '\n' tblproperties ("skip.header.line.count"="1");

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

load data inpath 'file:///tmp/hive\_classes/AgentPerformance.csv' into table agent\_perfromance;



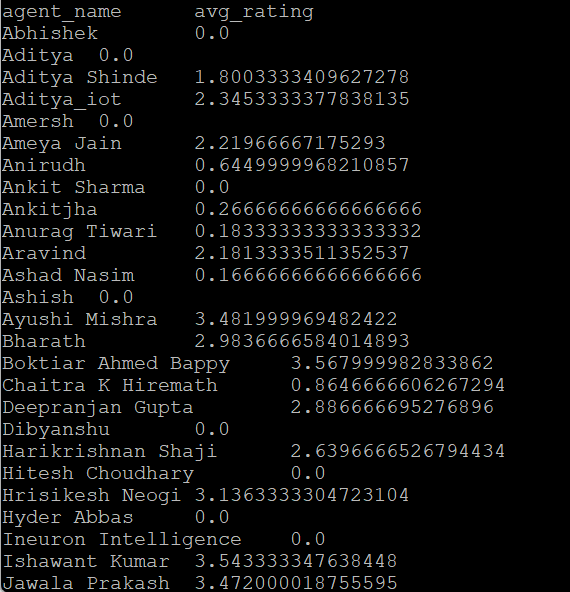
1. List of all agents' names.

select distinct agentname from agent\_performance;



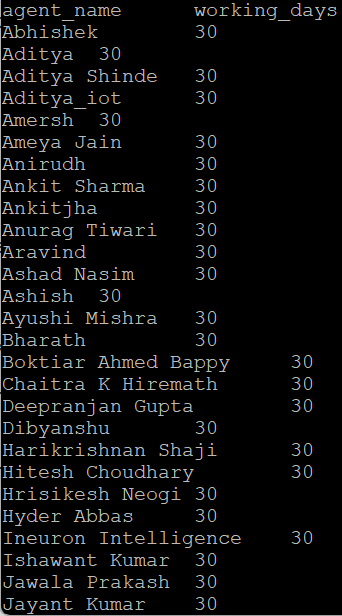
1. Find out agent average rating.

select agentname as Agent\_Name,avg(avergaerating) as Avg\_Rating from agent\_performance group by agentname;



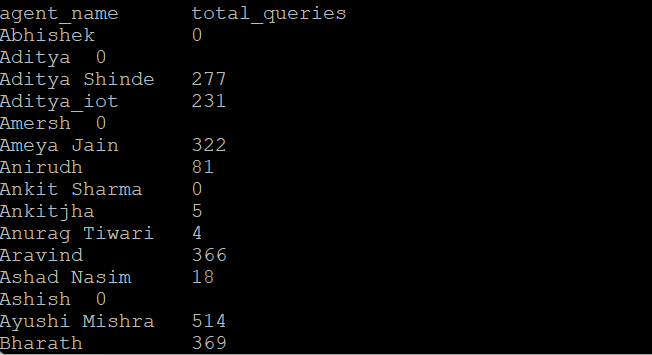
1. Total working days for each agents

select agentname as Agent\_Name, count(distinct(date)) as Working\_Days from agent\_performance group by agentname;



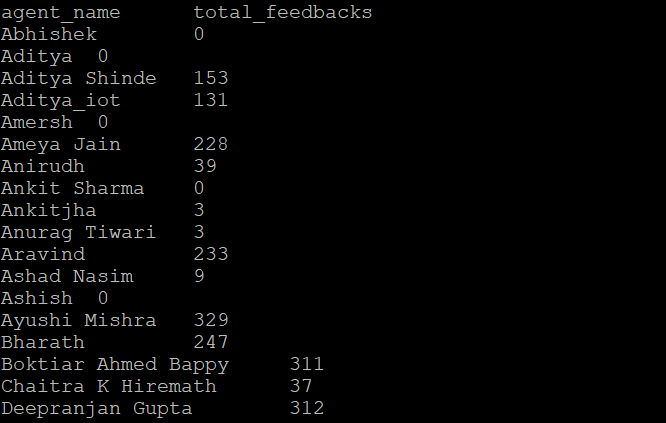
1. Total query that each agent have taken

select agentname as Agent\_Name, sum(totalchats) as Total\_Queries from agent\_performance group by agentname;



1. Total Feedback that each agent have received

select agentname as Agent\_Name, sum(totalfeedback) as Total\_Feedbacks from agent\_performance group by agentname;



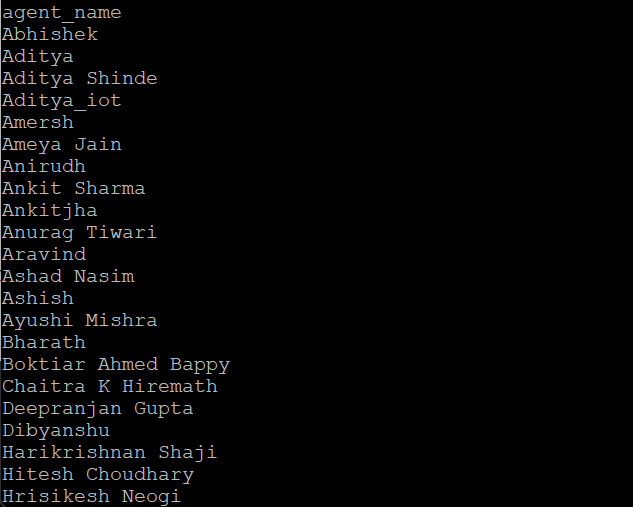
1. Agent name who have average rating between 3.5 to 4

select distinct agentname as Agent\_Name from agent\_performance where avergaerating between 3.5 and 4;



1. Agent name who have rating less than 3.5

select distinct agentname as Agent\_Name from agent\_performance where avergaerating < 3.5;



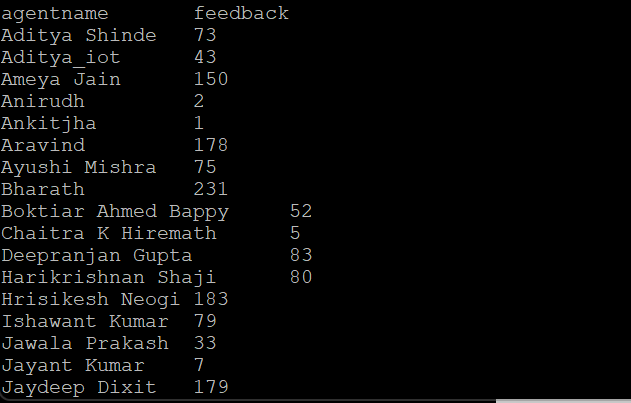
1. Agent name who have rating more than 4.5

select distinct agentname as Agent\_Name from agent\_performance where avergaerating > 4.5;



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

select agentname, sum(totalfeedback) as feedback from agent\_performance where avergaerating > 4.5 group by agentname;



1. average weekly response time for each agent

select agentname, sum(cast(substr(averageresponsetime,3,2) as decimal(20,10)))+(sum(cast(substr(averageresponsetime,6,2) as decimal(20,10)))/60) rt,

date\_format(date\_trunc('month', a.event\_date), '%m/%d/%Y') as date

from agent\_performance group by agentname;

1. average weekly resolution time for each agents

select agentname, sum(cast(substr(averageresultiontime,3,2) as decimal(20,10)))+(sum(cast(substr(averageresulutiontime,6,2) as decimal(20,10)))/60) rt, date\_format(date\_trunc('month', a.event\_date), '%m/%d/%Y') as date

from agent\_performance group by agentname;

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

select agentname,sum(totalchats) as numberofchats from agent\_performance where totalchats > 0 group by agentname;

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

select agentname, (sum(cast(substr(averageresultiontime,3,2) as decimal(20,10)))+(sum(cast(substr(averageresulutiontime,6,2) as decimal(20,10)))/60))/60 hour, date\_format(date\_trunc('month', a.event\_date), '%m/%d/%Y') as date

from agent\_performance group by agentname;

1. 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.

Select a.\*, b.\* from agent\_performance a inner join agent\_login\_report b on a.agentname = b.agentname;

Select a.\*, b.\* from agent\_performance a left join agent\_login\_report b on a.agentname = b.agentname;

Select a.\*, b.\* from agent\_performance a right join agent\_login\_report b on a.agentname = b.agentname;

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

create table agentperformance\_part1(slno int, date string,totalchats int,averageresponsetime string,averageresulutiontime string,averagerating float,totalfeedback int) partitioned by (agentname string);

insert overwrite table agentperformance\_part1 partition(agentname) select slno,date,totalchats,averageresponsetime,averageresolutiontime,avergaerating,totalfeedback from agent\_performance;

create table agentperformance\_buket(slno int, agentname string,date string,totalchats int,averageresponsetime string,averageresulutiontime string,averagerating float,totalfeedback int) clustered by (agentname) into 4 buckets row format delimited fields terminated by ‘,’;

from agent\_performance insert overwrite table agentperformance\_bucket select slno,date,totalchats,averageresponsetime,averageresolutiontime,avergaerating,totalfeedback from agent\_performance;