Query 3

Which industry has the most number of Data Scientist positions?

Platform used: HIVE

Query for above question:

hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/cloudera/project\_op/h1b\_query3' row format delimited fields terminated by ','

select soc\_name,job\_title,count(\*) as cnt from h1b\_app2 where job\_title like 'DATA SCIENTIST%' group by soc\_name,job\_title order by cnt desc limit 1;

Explanation:

In the above query we select the count and group it by soc\_name and job\_title and filter the job\_title by value ‘DATA SCIENTIS%’ because there is another value which has ‘DATA SCIENTIST II’ in the job\_title column. We limit sort in descending order based on the count and limit the number of records to 1.

Output file is below:

