HIVE QUERIES

**1 ) Is the number of petitions with Data Engineer job title increasing over time?**

**select year , lag(count(\*)) over(order by year),count(\*) from h1b\_final where job\_title like'DATA ENGINEER%' group by year;**

**select (curr\_year-prev\_year)/prev\_year\*100 from (select year , lag(count(\*)) over(order by year) prev\_year , count(\*) curr\_year from h1b\_final where job\_title like 'DATA ENGINEER%' group by year) p;**

**select avg ((curr\_year-prev\_year)/prev\_year\*100) growth from (select year,coalesce(lag(count(\*)) over(order by year),0) prev\_year,count(\*) curr\_year from h1b\_final where job\_title like'DATA ENGINEER%' group by year) p**;

**2) Which part of the US has the most Data Engineer jobs for each year**

**create table yearsite as select year,worksite,count(\*) year\_wise\_count from h1b\_applications where job\_title like '%DATA ENGINEER%'group by year,worksite;**

**select \* from yearsite a where year\_wise\_count in (select max(year\_wise\_count) from yearsite b where b.year=a.year);**

**3) Which top 5 employers file the most petitions each year**

**Create table emp\_top as select COUNT(\*) year , year\_count , employer\_name from emp\_top group by year\_count**

**select year,employer\_name, year\_count from (select rank() over(partition by year order by year\_count desc) nrow,\* from emp\_top) where nrow<6**

**4) Find the percentage and the count of each case status on total applications for each year. Create a graph depicting the pattern of All the cases over the period of time**

**select a.year ,case\_status,case\_count, case\_count/year\_case\*100 case\_percent from (select year,case\_status,count(\*) case\_count from h1b\_final2 group by year,case\_status) a join year\_wise\_case b on a.year=b.year order by year**;

**5) Find top 5 job titles who are having highest growth in applications**

**create table job\_growth as select year,job\_title,coalesce(lag(count(\*)) over(partition by job\_title order by year),0) prev\_year,count(\*) curr\_year from h1b\_app group by year,job\_title order by job\_title,year;  
  
select year,job\_title,avg((curr\_year-prev\_year)/curr\_year\*100) growth from job\_growth group by year,job\_title order by growth desc limit 5**

**6) Which are employers along with the number of petitions who have the success rate more than 70% in petitions and total petitions filed more than 1000**

**select employer\_name,success\_rate from (select employer\_name,count(case case\_status when 'CERTIFIED' then 1 when 'CERTIFIED-WITHDRAWN' then 1 else null end) /count(case\_status)\*100 success\_rate from h1b\_final group by employer\_name having count(\*)>1000) success\_tab where success\_rate>70**