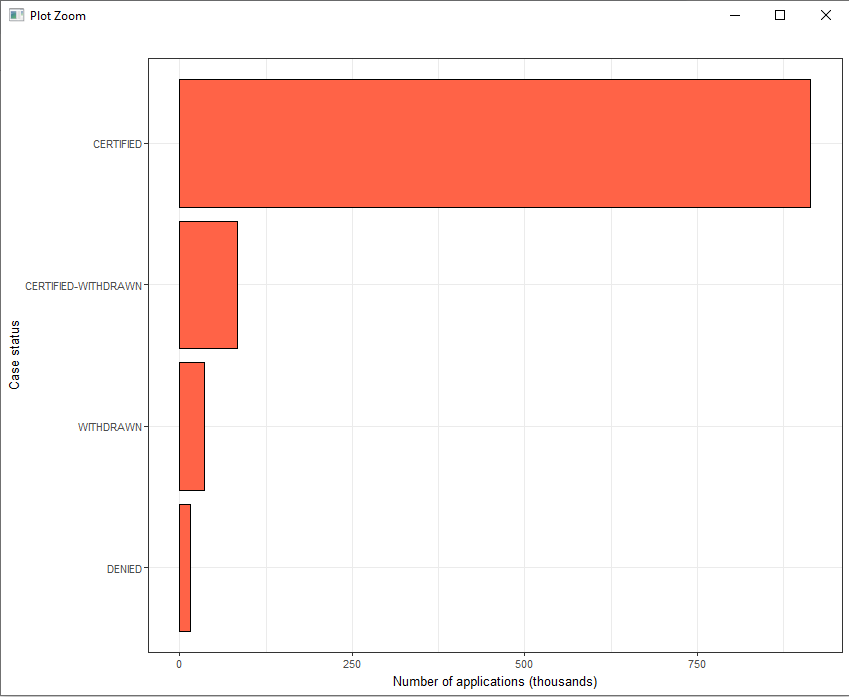
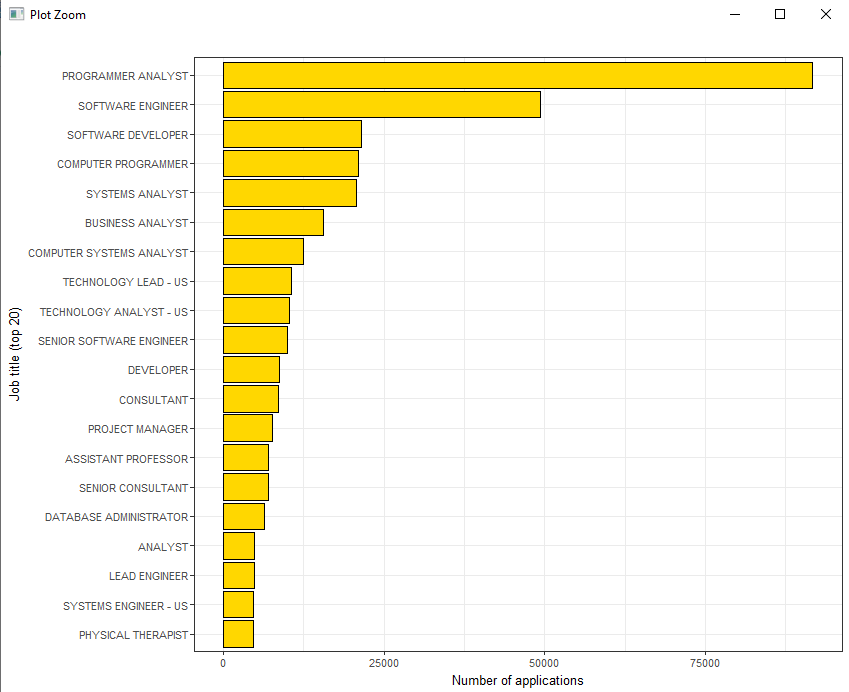
**Understanding dataset**:

* **The meaning of the columns are as follows**:
* Case\_status – Status of the application
* Employer\_name – Name of employer registered in H1b visa application
* Soc\_name – Occupation code for employment
* Job\_title – Job title for employment
* Full\_time\_position – Whether application is for full time position or part-time position
* Prevailing\_wage – The most frequent wage
* Year – The application year
* Worksite – The address of the employer worksite
* **Data Visualization**:

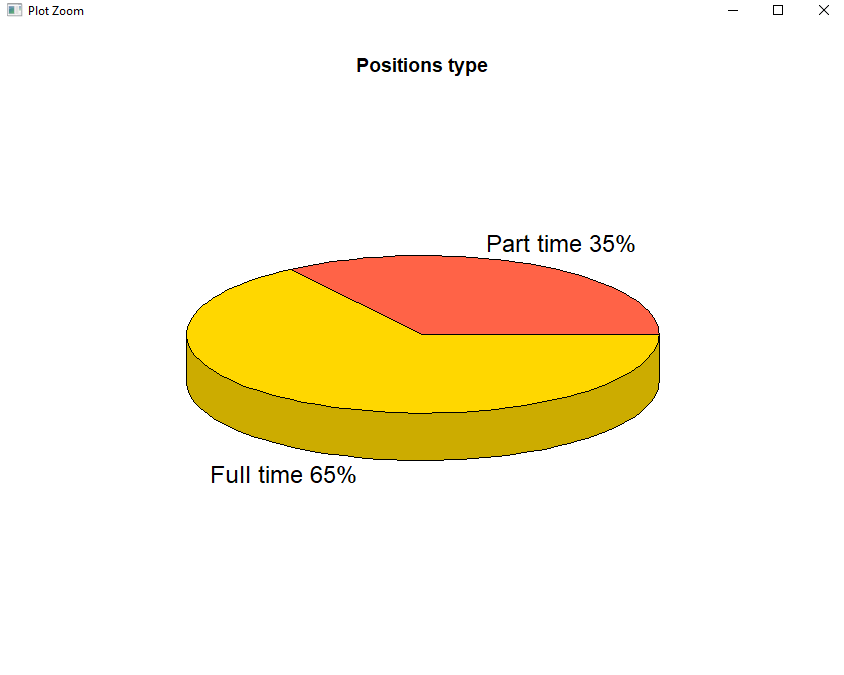
1. Distribution of CASE\_STATUS:



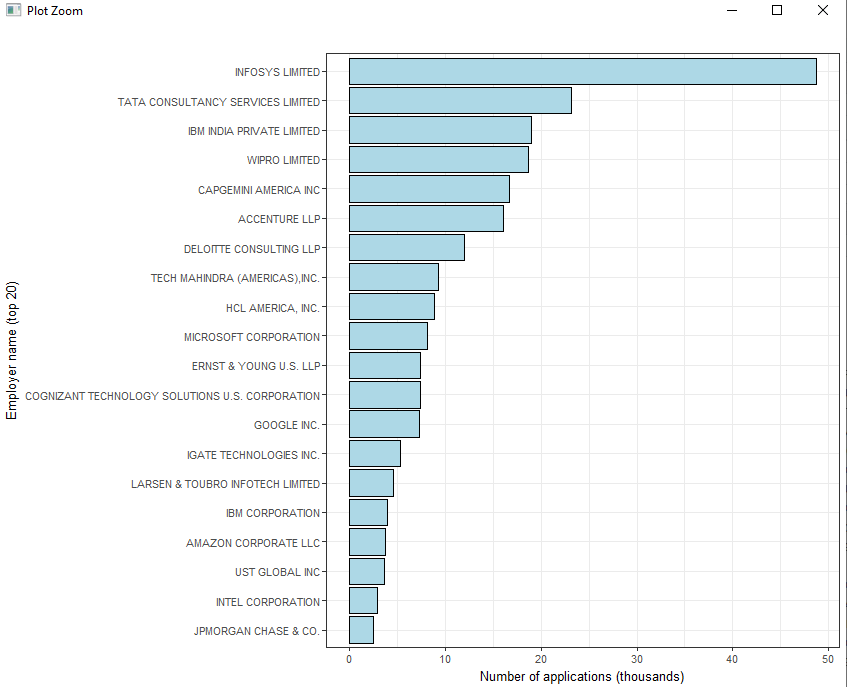
1. Top 20 of most frequent job titles in visa applications:



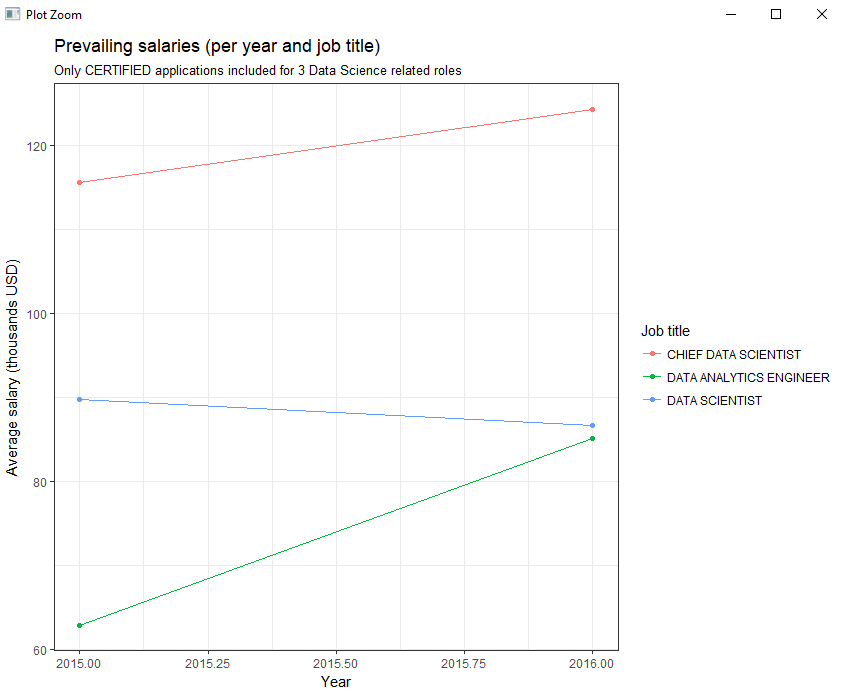
1. Percent of full-time positions from the total number of applications:



1. Top 20 of employers by number of applications:

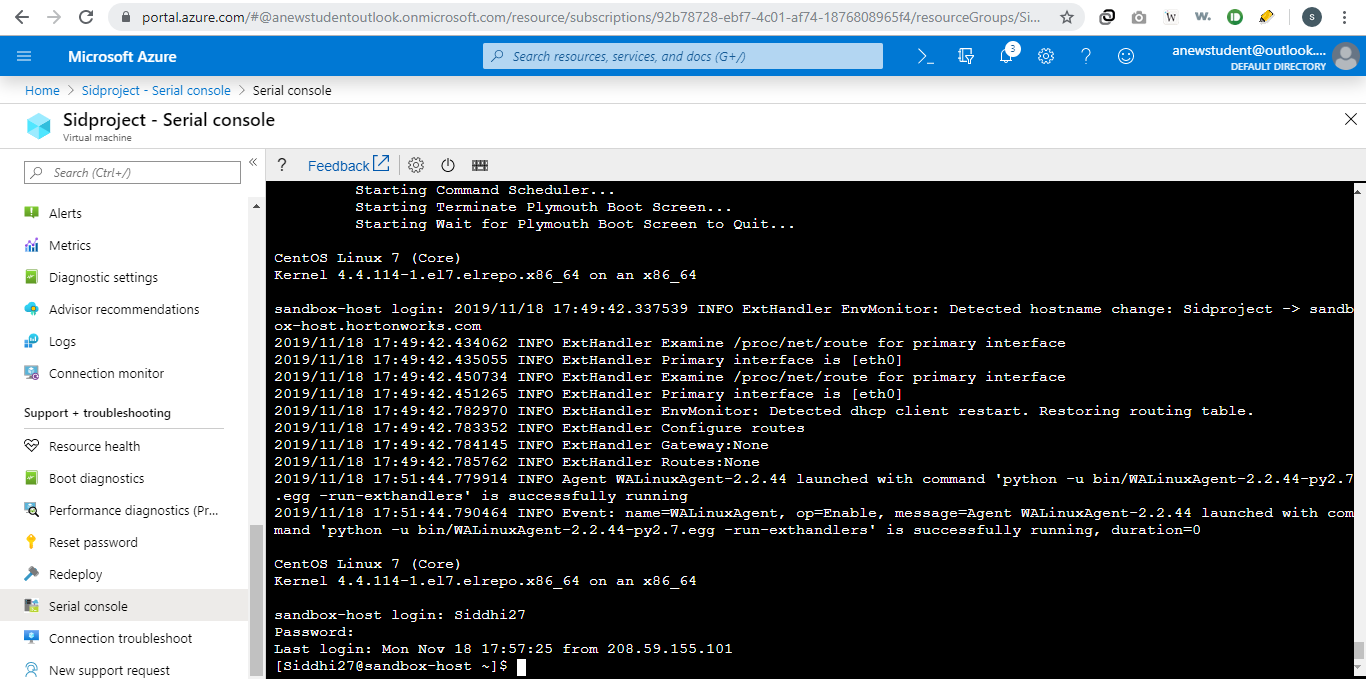


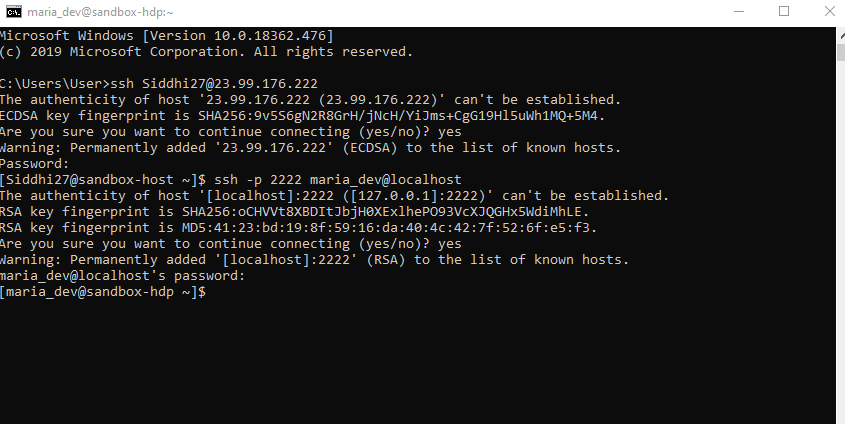
1. Job titles that includes DATA SCIENTIST, CHIEF DATA SCIENTIST and DATA ANALYTICS ENGINEER:

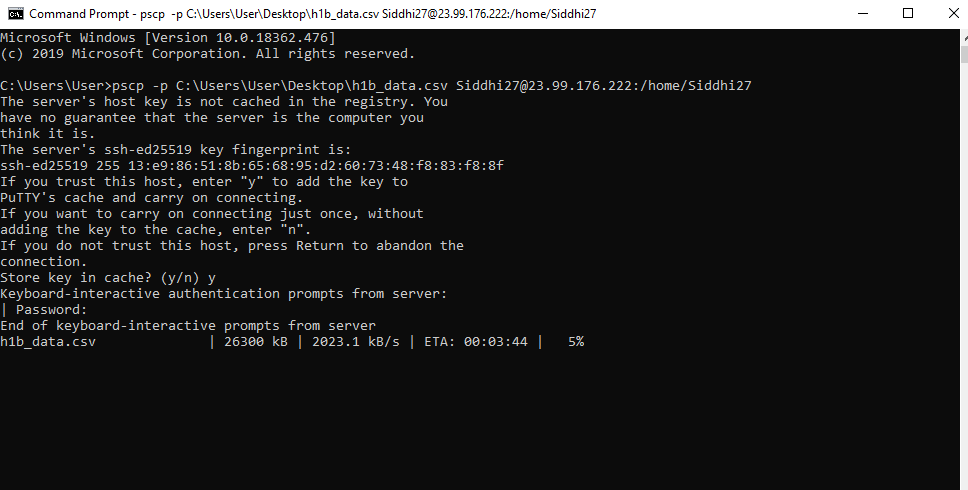


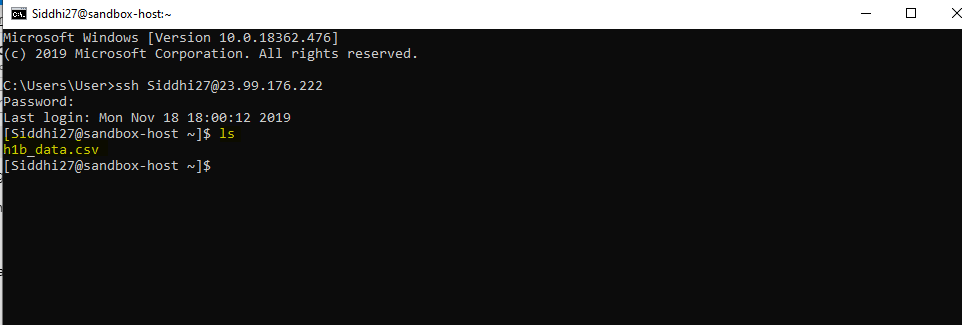
* **Step by step Execution**:

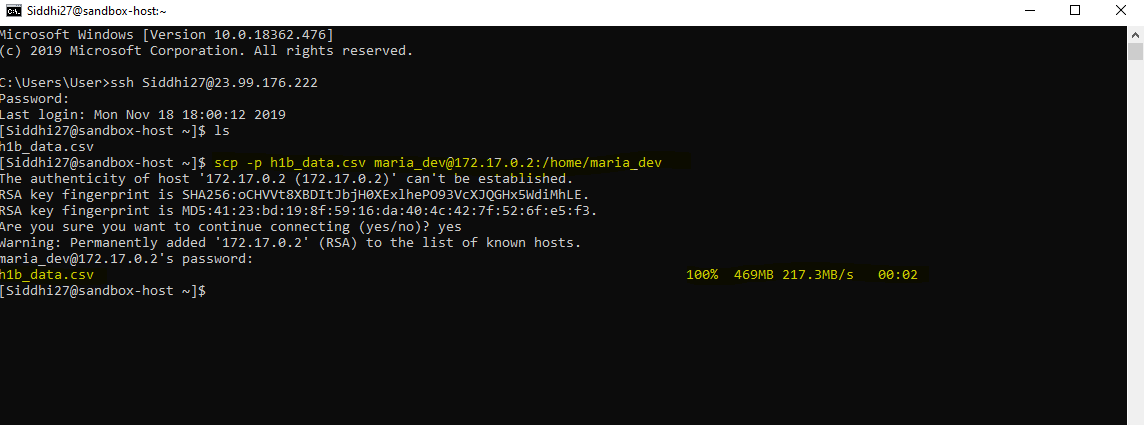
**Starting Virtual Machine:**



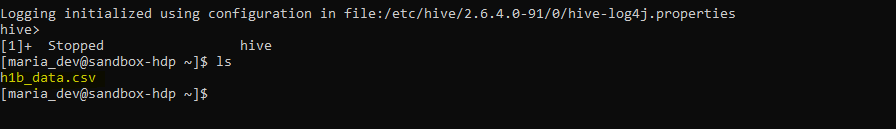






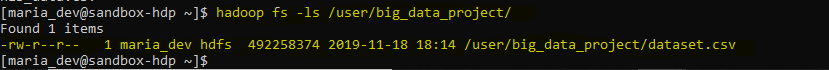


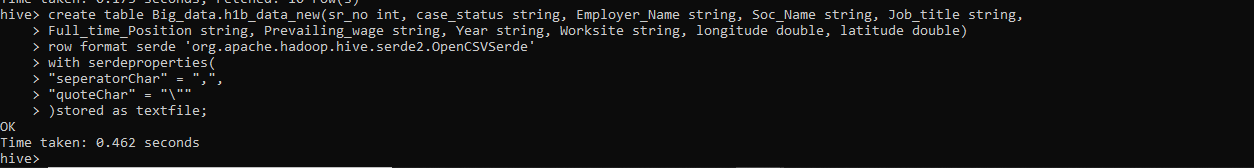
**Starting with Hive:**

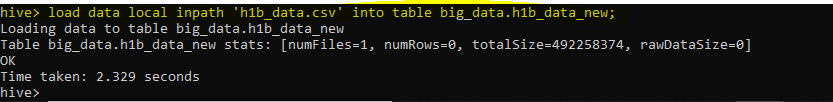




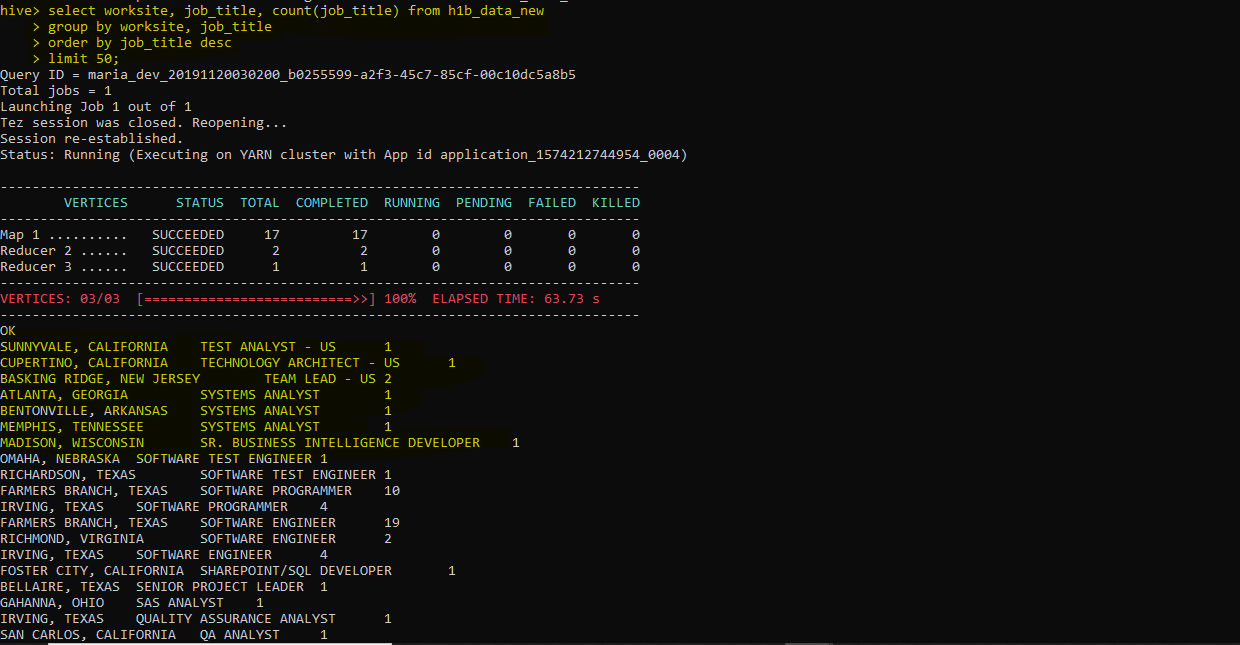




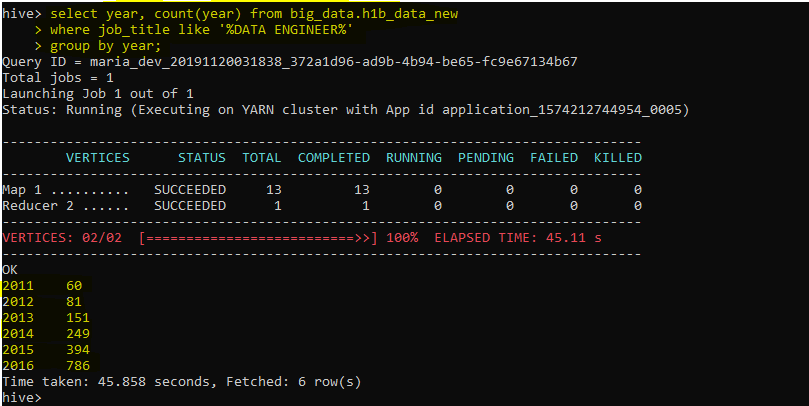
**Loading Data into HIVE:**



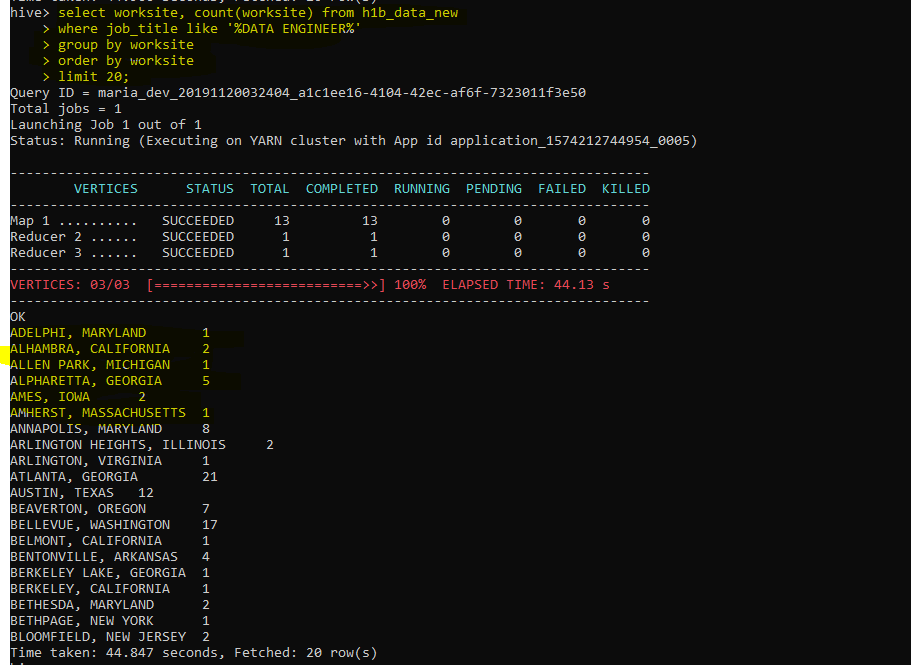
1. Which worksite has popular job title?



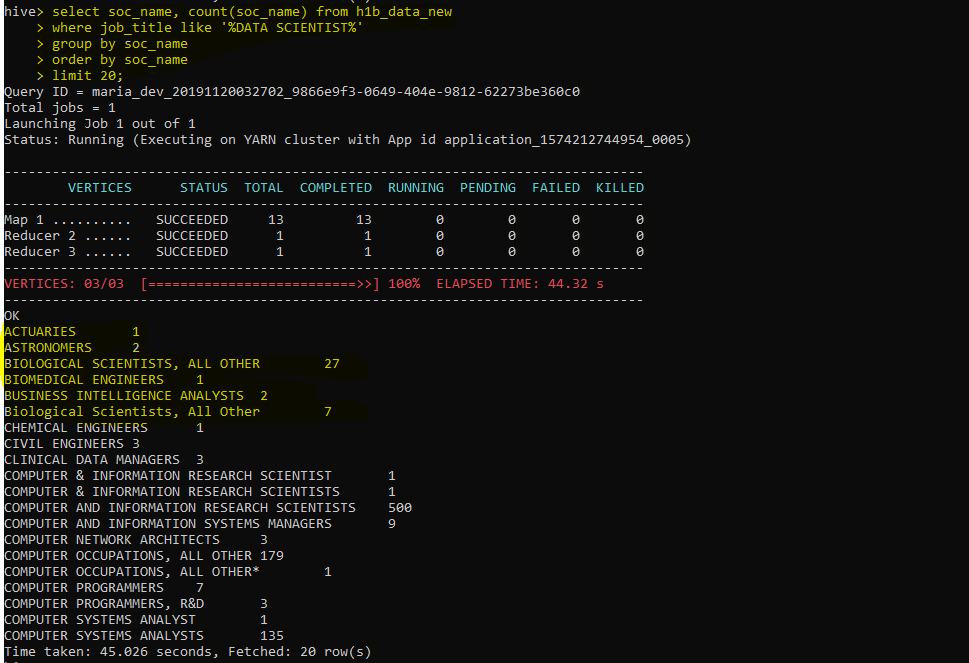
1. Is number of petitions with Data Engineer job title increasing over time?



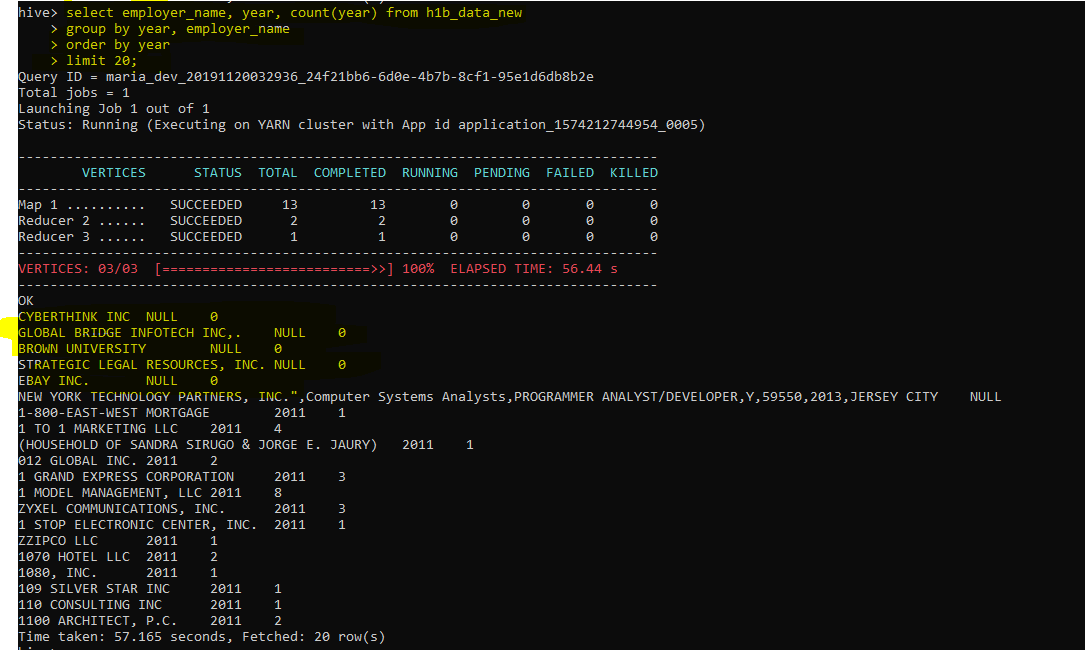
1. Which part of US has most data engineer jobs?



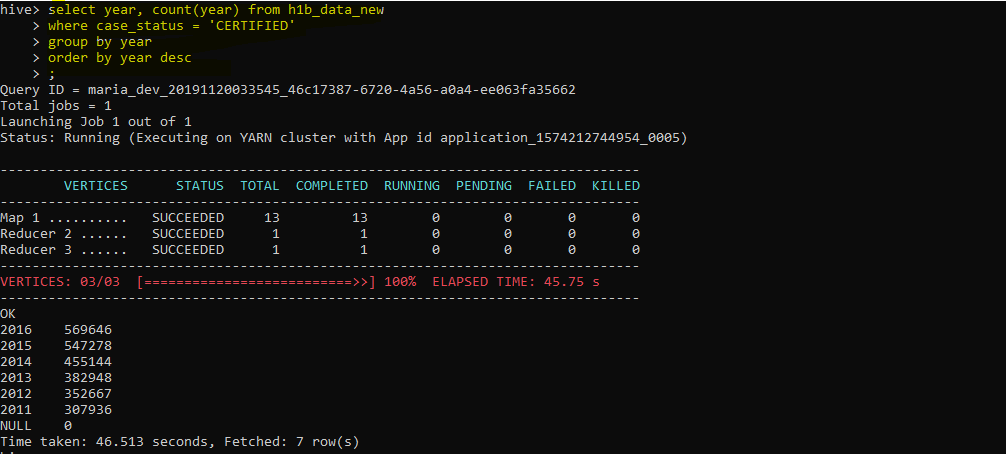
1. Which industry has greatest number of Data scientist position?



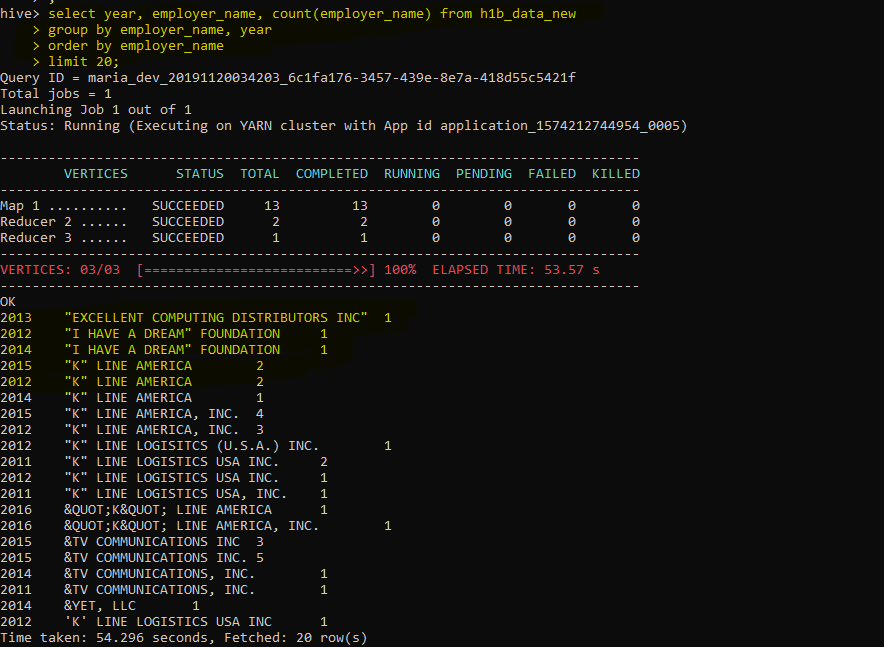
1. Which employer files the most petitions each year?



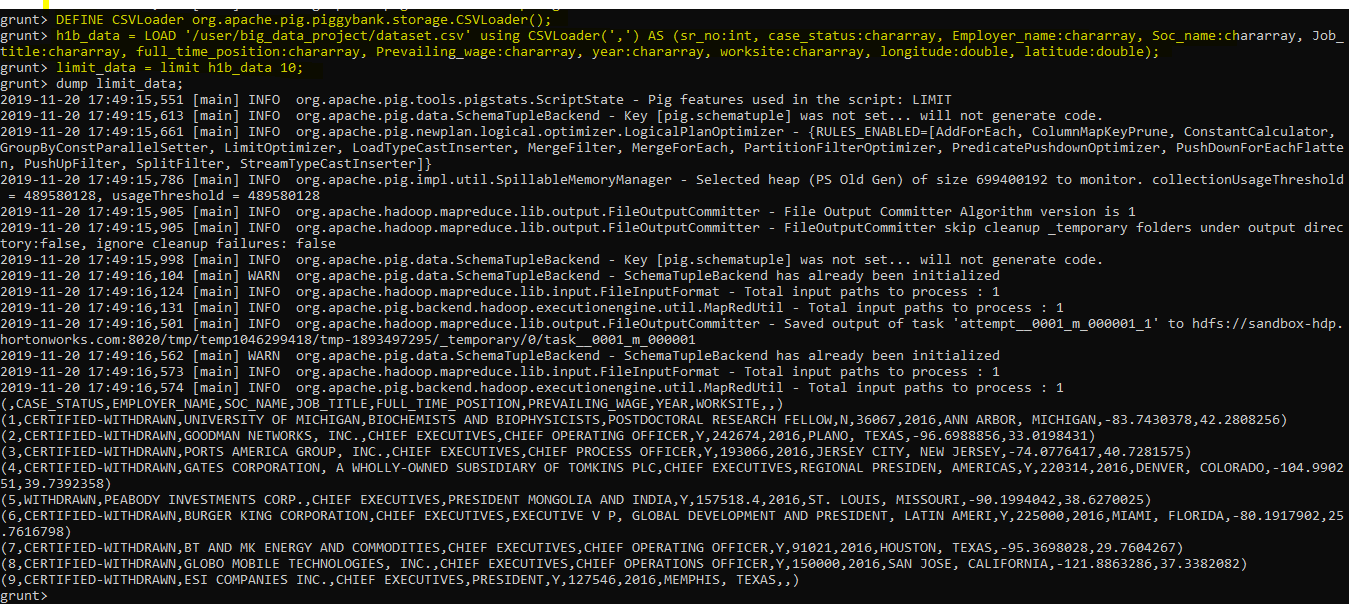
1. What is the general trend of the employer’s H1b status in the period 2011-16?



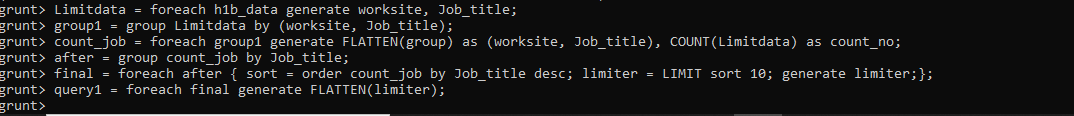
1. What is general trend of the employer name with respect to the year?

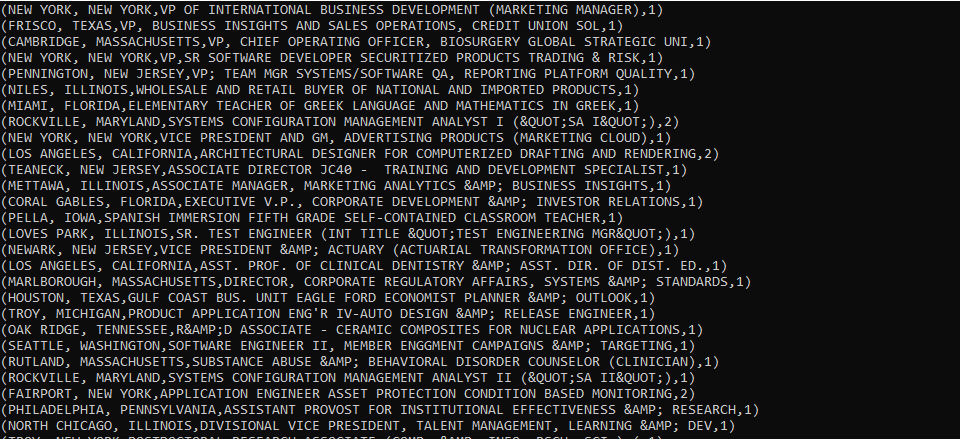


**Starting with Pig:**

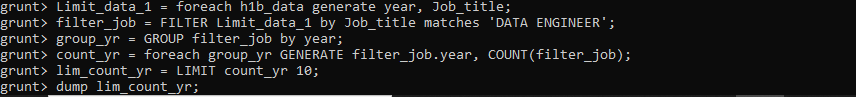


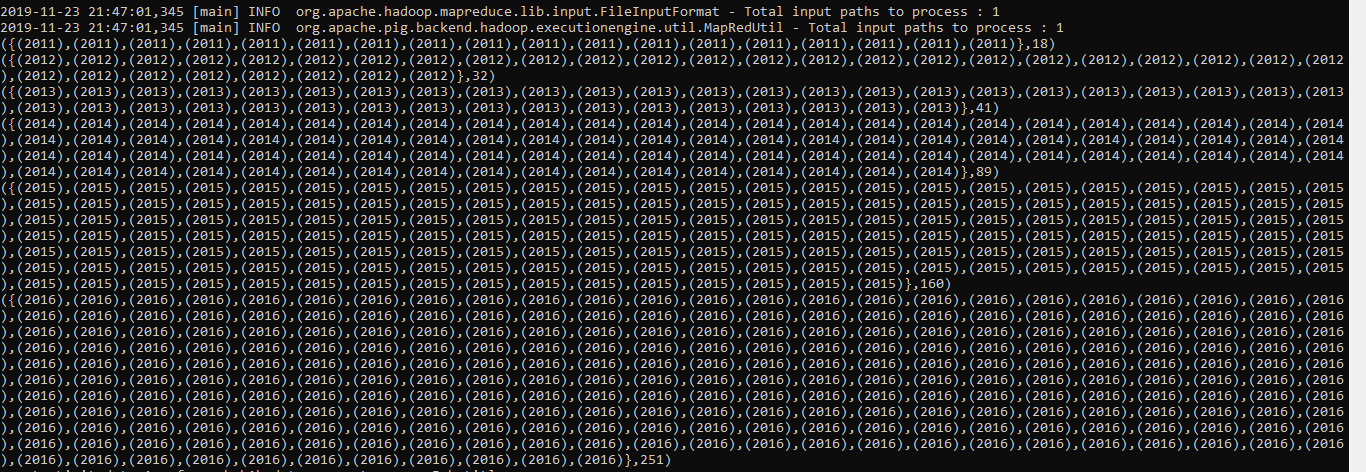
1. Which worksite has popular job title?



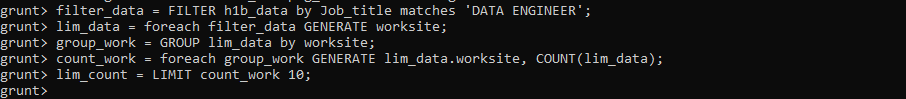


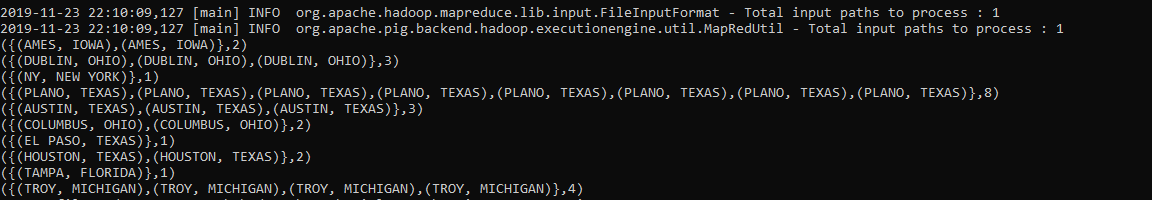
2. Is number of petitions with Data Engineer job title increasing over time?



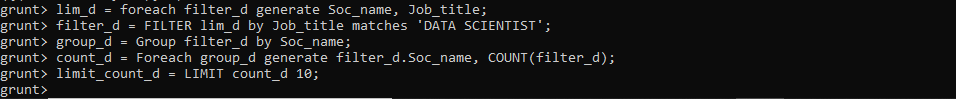


3.Which part of US has most data engineer jobs?

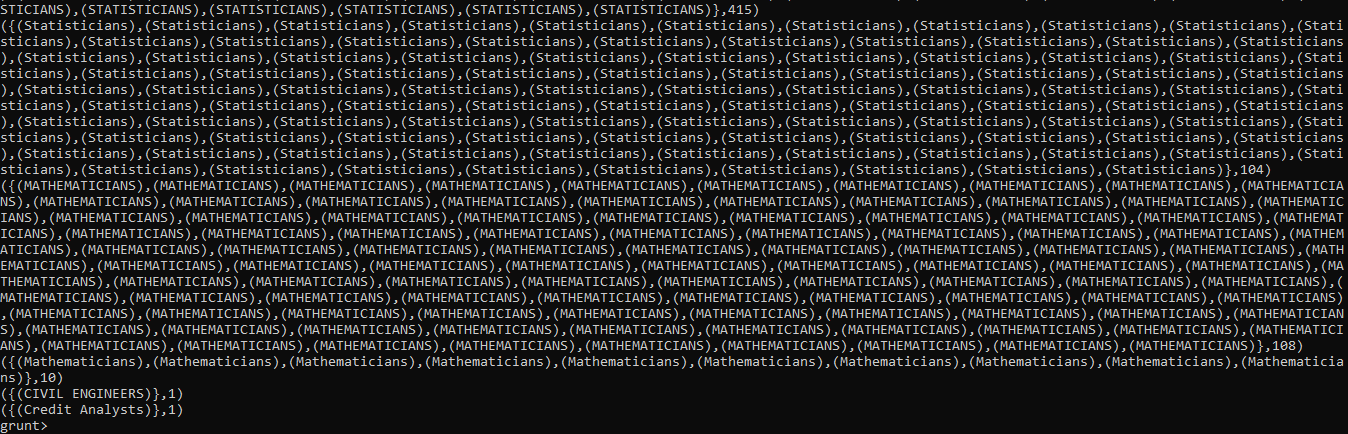




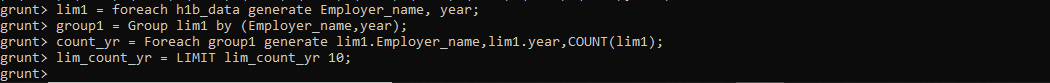
4. Which industry has greatest number of Data scientist position?

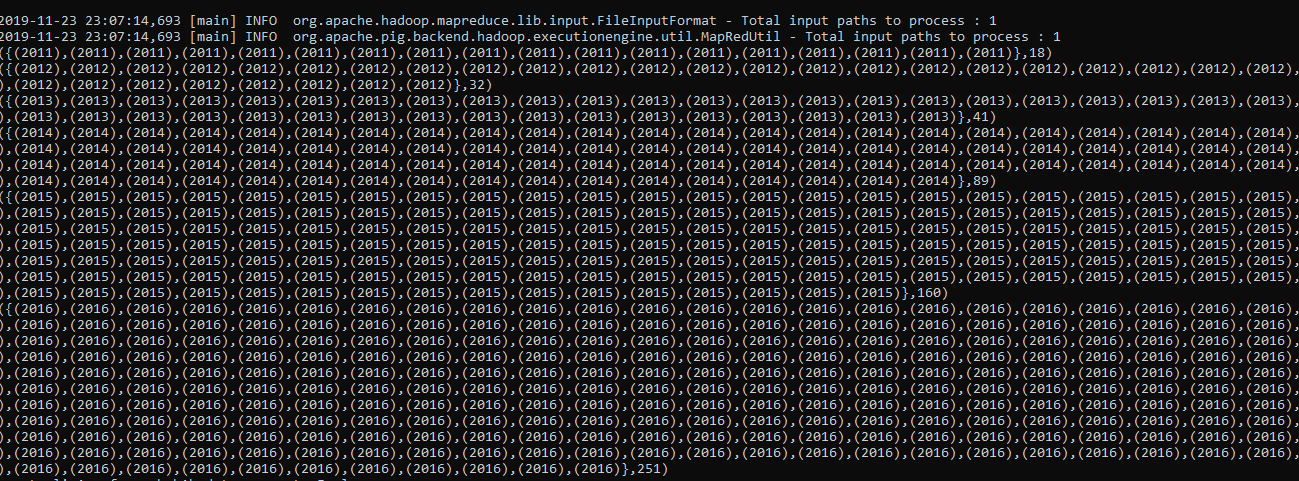






5. Which employer files the most petitions each year?

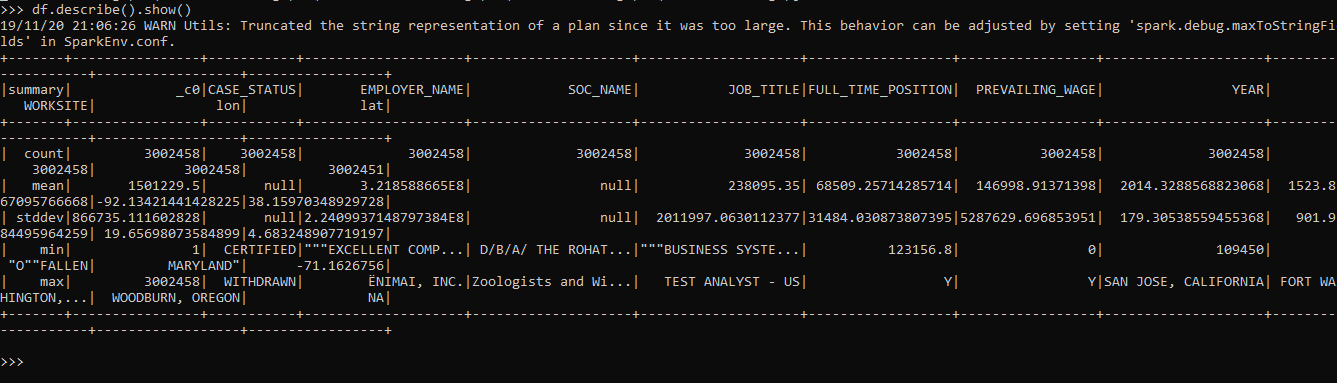




**Starting with SPARK**:

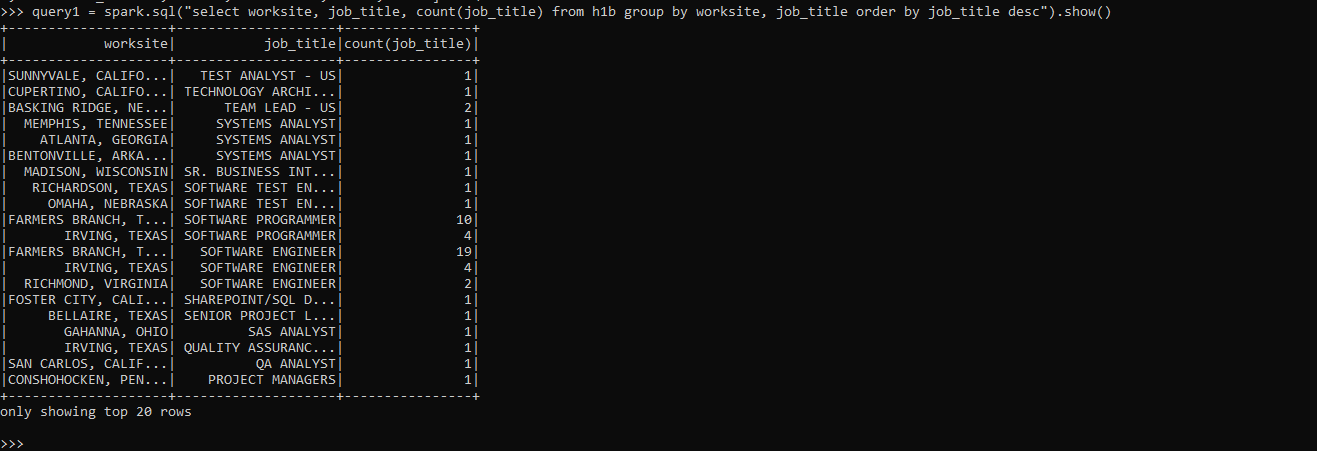




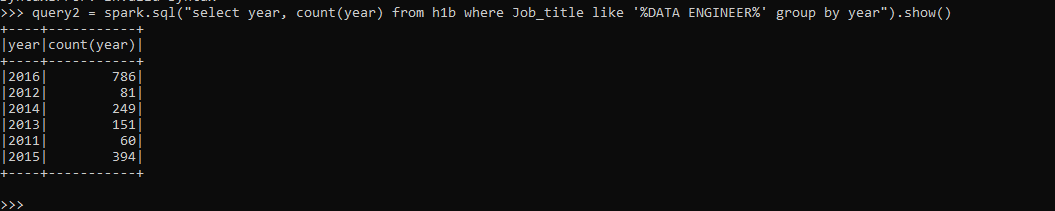




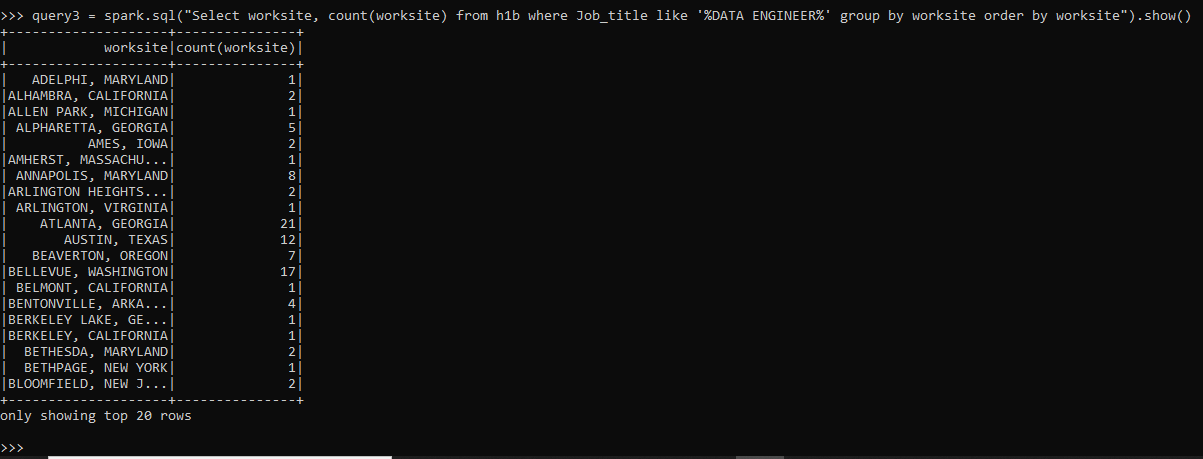
Query 1: Which worksite has popular job title?



Query 2: Is number of petitions with Data Engineer job title increasing over time?



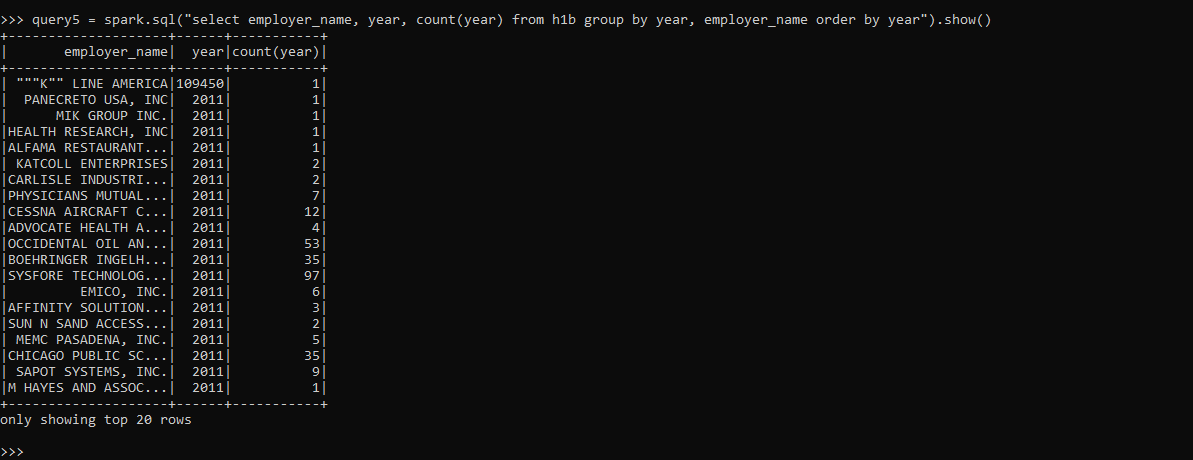
Query 3: Which part of US has most data engineer jobs?



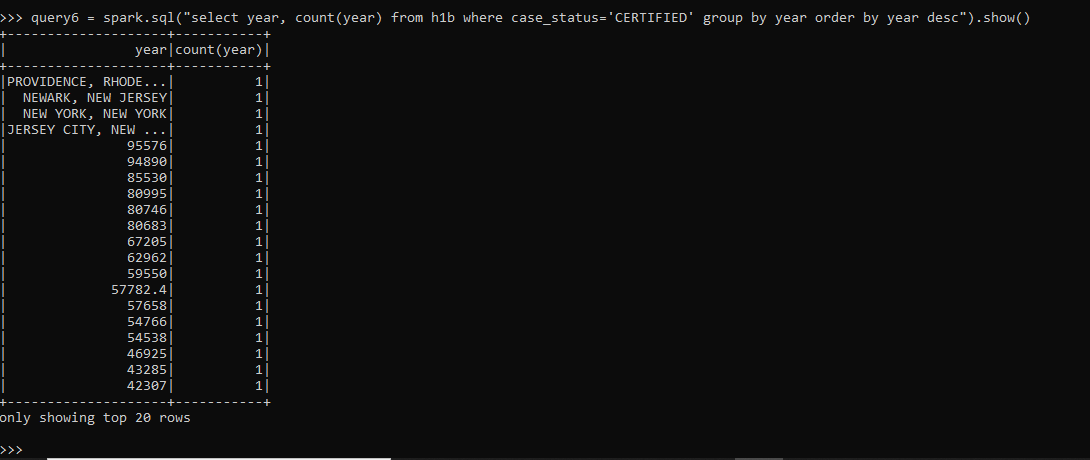
Query 4: Which industry has greatest number of Data scientist position?



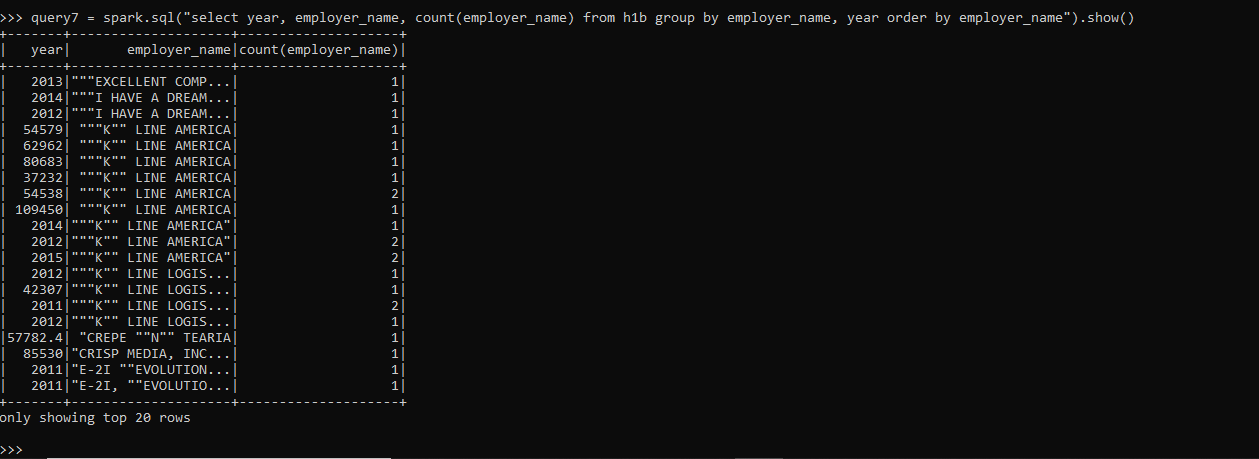
Query 5: Which employer files the most petitions each year?



Query 6: What is the general trend of the employer H1b status in the period 2011-16?



Query 7: What is general trend of the employer name with respect to the year?



**Comparison of the query execution time of Hive, Pig and Spark:**

