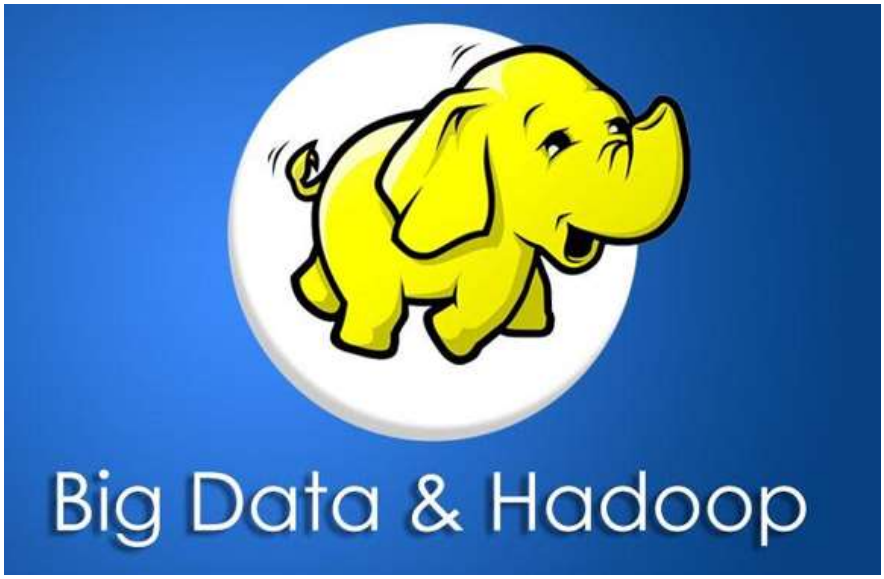
H1B VISA PROJECT DOCUMENT

BY

Gattamaneni pradeepchand chowdary

R171106200132



INDEX

TITLE	PAGE Number
ABSTRACT	2
INTRODUCTION	2
WHAT IS H1B-VISA	2
OBJECTIVES	3
WHY HADOOP?	4
TOOLS USED	4
DATAGIVEN	5
Flow of Project	5
CODING	6
MAPREDUCE	6
HIVE	14
PIG	15
SQOOP	19
OUTPUTS	21
MENU	40

ABSTRACT

The main theme of this project is the implementation of the distributed computing for the large datasets. The project focused on the analysis of H1B -applications data. Using Hadoop, an open source framework for distributed computing, the raw data of the H1B-applications are processed and analysed.

INTRODUCTION

The H1B is an employment-based, non-immigrant visa category for temporary foreign workers in the United States. For a foreign national to apply for H1B visa, an US employer must offer a job and petition for H1B visa with the US immigration department. This is the most common visa status applied for and held by international students once they complete college/ higher education (Masters, Ph.D.) and work in a full-time position.
The system would use Hadoop as its big data framework. Though the analysis can be done using the traditional RDBMS as well, but since the data is big using RDBMS the total time of analysis would be 10 times more than Hadoop. The analysis tool which the system would be using is Map Reduce, Hive, Pig.

WHAT IS H1B-VISA

The US H-1B visa is a non-immigrant visa that allows US companies to employ graduate level workers in specialty occupations that require theoretical or technical expertise in specialized fields such as in IT, finance, accounting, architecture, engineering, mathematics, science, medicine, etc. Any professional level job that usually requires you to have a bachelors degree or higher can come under the H-1B visa for specialty occupations. If you do not have a bachelors degree or higher you may be able to show degree equivalence through work experience and/or other qualifications.

Applying for a non-immigrant visa is generally quicker than applying for a US Green Card, therefore the H-1B visa is popular for companies wishing to bring in staff for long-term assignment in the US. However, because of the lack of available visas employers frequently have to look at applying for other visa categories such as the L-1B for specialized workers, L-1A for managers and executives, E-2 Treaty Investor visa, E-1 Treaty Trader visa, E-3 for Australians etc.

Individuals cannot apply directly for an H-1B visa. Instead the employer must petition for entry of the employee.

Eligibility: specialty occupation

The US H1-B visa is designed to be used for staff in specialty occupations. The job must meet one of the following criteria to qualify as a specialty occupation:

- Have a minimum entry requirement of a Bachelor's or higher degree or its equivalent.
- The degree requirement for the job is common to the industry or the job is so complex or unique that it can be performed only by an individual with a degree.
- The employer normally requires a degree or its equivalent for the position.
- The nature of the specific duties is so specialized and complex that the knowledge required to perform the duties is usually associated with the attainment of a bachelor's or higher degree.

Eligibility: employee qualifications

For you to qualify to accept a job offer in a specialty occupation you must meet one of the following criteria:

- Have completed a US bachelor's or higher degree required by the specific specialty occupation from an accredited college or university.
- Hold a foreign degree that is the equivalent to a U.S. bachelor's or higher degree in the specialty occupation.
- Hold an unrestricted state license, registration, or certification which authorizes you to fully practice the specialty occupation and be engaged in that specialty in the state of intended employment.
- Have education, training, or experience in the specialty that is equivalent to the completion of such a degree, and have recognition of expertise in the specialty through progressively responsible positions directly related to the specialty.

OBJECTIVES

1. Data collection and production of information for various companies and employment authorities, for Analysis purposes.

2. Production of information which serves bodies, organizations and various companies in the fields of employment and career service and analysis the scope and availibility of H1B visa.

4. Processing and analysing large amount of raw data by using map-reduce programming model and distributed computing on HADOOP framework to improve time and complexity.

WHY HADOOP?

As the system involves census data analysis, the amount of raw census data that would be evaluated for insights will be very large. Thus, the involvement of very large data for analysis urges to the use of Hadoop framework for data analysis. Though the data analysis can be done with traditional RDBMS system as well but since the data is very large, the time required for analysis would be very long. The use of Hadoop framework for the data analysis incurs many benefits:

Scalable - New nodes can be added as needed and added without needing to change data formats, how data is loaded, how jobs are written, or the applications on top.

Cost effective - Hadoop brings massively parallel computing to commodity servers. The result is a sizeable decrease in the cost per terabyte of storage, which in turn makes it affordable to model all your data.

Flexible - Hadoop is schema-less, and can absorb any type of data, structured or not, from any number of sources. Data from multiple sources can be joined and aggregated in arbitrary ways enabling deeper analyses than any one system can provide.

Fault tolerant - When you lose a node, the system redirects work to another location of the data and continues processing without missing a beat.

TOOLS USED

1. **MAPREDUCE** - MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.

2. **HIVE** - Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analysing easy. Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. It is used by different companies. For

example, Amazon uses it in Amazon Elastic MapReduce.

3. **PIG** - Apache Pig is an abstraction over MapReduce. It is a tool/platform which is used to analyze larger sets of data representing them as data flows. Pig is generally used with Hadoop; we can perform all the data manipulation operations in Hadoop using Apache Pig.To analyze data using Apache Pig, programmers need to write scripts using Pig Latin language. All these scripts are internally converted to Map and Reduce tasks. Apache Pig has a component known as Pig Engine that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs.

Data Given -

The dataset has nearly 3 million records.

The dataset description is as follows:
The columns in the dataset include:

- CASE_STATUS: Status associated with the last significant event or decision. Valid values include “Certified,” “Certified-Withdrawn,” Denied,” and “Withdrawn”.

  Certified: Employer filed the LCA, which was approved by DOL

  Certified Withdrawn: LCA was approved but later withdrawn by employer

  Withdrawn: LCA was withdrawn by employer before approval

  Denied: LCA was denied by DOL

- EMPLOYER_NAME: Name of employer submitting labour condition application.
- SOC_NAME: The Occupational name associated with the SOC_CODE. SOC_CODE is the occupational code associated with the job being requested for temporary labour condition, as classified by the Standard Occupational Classification (SOC) System.

- JOB_TITLE: Title of the job
- FULL_TIME_POSITION: Y = Full Time Position; N = Part Time Position
- PREVAILING_WAGE: Prevailing Wage for the job being requested for temporary labour condition. The wage is listed at annual scale in USD. The prevailing wage for a job position is defined as the average wage paid to similarly employed workers in the requested occupation in the area of intended employment. The prevailing wage is based on the employer’s minimum requirements for the position.
- YEAR: Year in which the H1B visa petition was filed
- WORKSITE: City and State information of the foreign worker’s intended area of employment
- lon: longitude of the Worksite
- lat: latitude of the Worksite

Flow of Project

1. Menu for select cases.
2. Analysis done using MapReduce, Hive and Pig
3. Require output
4. Graphical Representation of output

CODING

MAPREDUCE:

3)Which industry has the most number of Data Scientist positions?

MAPER:

```
package h1bproject;

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class Mapper3 extends Mapper<LongWritable,Text,Text,IntWritable>{
	public void map(LongWritable key,Text value,Context context) throws IOException, InterruptedExcep-
tion{
			String[] parts = value.toString().split(",(?=([^\"]*\"[^\"]*\")*[^\"]*$)");
			String industry = parts[2];
			String job = parts[4];
			int count = 1;
			if(job.equals("\"DATA SCIENTIST\"")){
				context.write(new Text(industry),new IntWritable(count));
			}
	}
}
```

REDUCER:

```
package h1bproject;

import java.io.IOException;
import java.util.TreeMap;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.NullWritable;
import org.apache.hadoop.io.Text;
```

```
import org.apache.hadoop.mapreduce.Reducer;

public class Reducer3 extends Reducer<Text,IntWritable,NullWritable,Text>{
	private TreeMap<Integer,Text> highest = new TreeMap<>();
	public void reduce(Text key,Iterable<IntWritable> value,Context con-
text) throws IOException, InterruptedException{
		int count = 0;
		for (IntWritable val:value){
			count += val.get();
		}
		String output = key.toString()+","+count;
		highest.put(count,new Text(output));
		if(highest.size()>1){
			highest.remove(highest.firstKey());
		}
	}
	protected void cleanup(Context context) throws IOException, Inter-
ruptedException{
		for(Text t:highest.values()){
			context.write(NullWritable.get(),t);
		}
	}
}
```

DRIVER:

```
package h1bproject;


import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.io.NullWritable;
//import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.mapreduce.Job;
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
//import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;


public class MaxDataScientist {




	public static void main(String[] args) throws IOException, ClassNot-
FoundException, InterruptedException{
		Configuration conf = new Configuration();
		  conf.set("mapreduce.output.textoutputformat.separator",
",");
		  Job job = Job.getInstance(conf, "MaxDataScientist");
		  job.setJarByClass(MaxDataScientist.class);
		  job.setMapperClass(Mapper3.class);
		  job.setReducerClass(Reducer3.class);
		  job.setReducerClass(Reducer3.class);
	   job.setMapOutputKeyClass(Text.class);
```

```
                job.setMapOutputValueClass(IntWritable.class);
                job.setOutputKeyClass(NullWritable.class);
                //job.setNumReduceTasks(0);
                //job.setCombinerClass(ReduceClass.class);
                job.setOutputValueClass(Text.class);
                job.setInputFormatClass(TextInputFormat.class);
                job.setOutputFormatClass(TextOutputFormat.class);
                FileInputFormat.addInputPath(job, new Path(args[0]));
                FileOutputFormat.setOutputPath(job, new Path(args[1]));
                System.exit(job.waitForCompletion(true) ? 0 : 1);
        }

}
```

7) Create a bar graph to depict the number of applications for each year

MAPPER:

```
package h1bproject;

import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class Mapper7 extends Mapper<LongWritable,Text,Text,LongWritable>
{

    public void map(LongWritable key, Text value, Context context) throws
IOException, InterruptedException
    {
         if(key.get()> 0 )

         {

       String [] to-
ken=value.toString().split(",(?=([^\"]*\"[^\"]*\")*[^\"]*$)");

         context.write(new Text(token[7]),new LongWritable(1));
         }
         }
    }
```

REDUCER:

```
package h1bproject;

import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
//import org.apache.hadoop.mapreduce.Reducer.Context;
```

```
public class Reducer7 extends Reducer<Text,LongWritable,Text,LongWritable>
{
public void reduce(Text key,Iterable <LongWritable> values,Context context)
throws IOException, InterruptedException
{
        long sum=0;
for(LongWritable val:values)
        sum=sum+val.get();
context.write(key, new LongWritable(sum));
}

}
```

DRIVER:

```
package h1bproject;

//import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class ApplicationsPerYear
{



        public static void main(String args[]) throws Exception

        {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf,"Applications Per Year");

        job.setJarByClass(ApplicationsPerYear.class);
        job.setMapperClass(Mapper7.class);
        job.setCombinerClass(Reducer7.class);
        job.setReducerClass(Reducer7.class);

        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(LongWritable.class);

        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(LongWritable.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job,new Path(args[1]));

        System.exit(job.waitForCompletion(true)?0:1);

        }
```

```
}
```

9) Which are top ten employers who have the highest success rate in petitions?

MAPPER:

```
package h1b9;

import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class Mapper9 extends Mapper<LongWritable,Text,Text,Text>{

               public void map(LongWritable key,Text value,Context context)
throws IOException, InterruptedException{
                      String[] parts =
value.toString().split(",(?=([^\"]*\"[^\"]*\")*[^\"]*$)");
                      String status = parts[1];
                      String employer = parts[2];
                      context.write(new Text(employer), new Text(status));

       }

}
```

REDUCER:

```
package h1b9;

import java.io.IOException;
import java.util.TreeMap;

import org.apache.hadoop.io.NullWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class Reduce9 extends Reducer <Text,Text,NullWritable,Text>{

               private TreeMap<Double,String> top10 = new TreeMap<>();
               public void reduce(Text key,Iterable<Text> value,Context con-
text){
                      double total = 0;
                      double sucess = 0;
                      for (Text val:value){
                             String status = val.toString();
                             if(status.equals("\"CERTIFIED\"") || sta-
tus.equals("\"CERTIFIED-WITHDRAWN\"")){
```

```
                                        total++;
                                        sucess++;
                                }
                                else total++;
                        }
                        double sucess_rate = (sucess/total)*100;
                        if (sucess_rate >= 70 && total >=1000){
                                String output = key.toString() +","+String.for-
mat("%.0f",total)+","+String.format("%.2f %%",sucess_rate);
                                top10.put(sucess_rate, out-
put);
                        }
                }
                protected void cleanup(Context context) throws IOException,
InterruptedException{
                        for(String val : top10.descendingMap().values()){
                                context.write(NullWritable.get(),new
Text(val));
                        }
                }
        }
```

DRIVER:

```
package h1b9;

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;



public class H1B9 {




        public static void main(String[] args) throws IOException, ClassNot-
FoundException, InterruptedException{
                Configuration conf = new Configuration();
                  conf.set("mapreduce.output.textoutputformat.separator",
",");
                  Job job = Job.getInstance(conf, "H1B9");
                  job.setJarByClass(H1B9.class);
                  job.setMapperClass(Mapper9.class);
                  job.setReducerClass(Reduce9.class);
               job.setReducerClass(Reduce9.class);
                  job.setMapOutputKeyClass(Text.class);
                  job.setMapOutputValueClass(Text.class);
                  job.setOutputKeyClass(NullWritable.class);
                  //job.setNumReduceTasks(0);
                  //job.setCombinerClass(ReduceClass.class);
                  job.setOutputValueClass(Text.class);
```

```
                job.setInputFormatClass(TextInputFormat.class);
                job.setOutputFormatClass(TextOutputFormat.class);
                FileInputFormat.addInputPath(job, new Path(args[0]));
                FileOutputFormat.setOutputPath(job, new Path(args[1]));
                System.exit(job.waitForCompletion(true) ? 0 : 1);
        }

}
```

10) Which are the top 10 job positions which have the highest success rate in petitions?

MAPPER:

```
package h1b10;

import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
//import org.apache.hadoop.mapreduce.Mapper.Context;

public  class MapClass10 extends Mapper<LongWritable,Text,Text,Text>{
        public void map(LongWritable key,Text value,Context context) throws
IOException, InterruptedException{
                String[] parts =
value.toString().split(",(?=([^\"]*\"[^\"]*\")*[^\"]*$)");//From statckover-
flow.com
                String status = parts[1];
                String job = parts[4].replaceAll("\"", "");
                context.write(new Text(job), new Text(status));
        }
}
```

REDUCER:

```
package h1b10;

import java.io.IOException;
import java.util.TreeMap;

import org.apache.hadoop.io.NullWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
//import org.apache.hadoop.mapreduce.Reducer.Context;

public class ReduceClass10 extends Reducer <Text,Text,NullWritable,Text>{
        private TreeMap<Double,String> top10 = new TreeMap<>();
        public void reduce(Text key,Iterable<Text> value,Context context){
                double total = 0;
                double sucess = 0;
                for (Text val:value){
                        String status = val.toString();
                        if(status.equals("\"CERTIFIED\"") || sta-
tus.equals("\"CERTIFIED-WITHDRAWN\"")){
                                total++;
                                sucess++;
                        }
                        else total++;
```

```
                }              System.out.printf("%f,%f",sucess,total);
                double sucess_rate = (sucess/total)*100;
                if (sucess_rate >= 70 && total >=1000){
                        String output = key.toString() +"@"+String.for-
mat("%.0f",total)+"@"+String.format("%.2f ",sucess_rate);
                        top10.put(sucess_rate, out-
put);
                }
        }
        protected void cleanup(Context context) throws IOException, Inter-
ruptedException{
                for(String val : top10.descendingMap().values()){
                        context.write(NullWritable.get(),new Text(val));
                }
        }//
}
```

DRIVER:

```
package h1b10;

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
//import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;


public class H1B10 {

        public static void main(String[] args) throws IOException, ClassNot-
FoundException, InterruptedException{
                Configuration conf = new Configuration();
                  conf.set("mapreduce.output.textoutputformat.separator",
",");
                  Job job = Job.getInstance(conf, "H1B10");
                  job.setJarByClass(H1B10.class);
                  job.setMapperClass(MapClass10.class);
                  job.setReducerClass(ReduceClass10.class);
                  job.setReducerClass(ReduceClass10.class);
                  job.setMapOutputKeyClass(Text.class);
                  job.setMapOutputValueClass(Text.class);
                  job.setOutputKeyClass(NullWritable.class);
                  //job.setNumReduceTasks(0);
                  //job.setCombinerClass(ReduceClass.class);
                  job.setOutputValueClass(Text.class);
                  job.setInputFormatClass(TextInputFormat.class);
                  job.setOutputFormatClass(TextOutputFormat.class);
                  FileInputFormat.addInputPath(job, new Path(args[0]));
                  FileOutputFormat.setOutputPath(job, new Path(args[1]));
                  System.exit(job.waitForCompletion(true) ? 0 : 1);
        }

}
```

HIVE:

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

```
select worksite,Tjob,year from(Select rank() over(partition by year order by Tjob desc) as
rank_1,Tjob,worksite,year from(Select count(job_title)as Tjob,worksite,year from h1b_final where job_ti-
tle="DATA ENGINEER" group by worksite,year order by Tjob) a)b where rank_1=1 and year is not null;
```

b) find top 5 locations in the US who have got certified visa for each year.

```
select worksite,Tcity,year from(Select rank() over(partition by year order by Tcity desc)as
rank_1,worksite,Tcity,year from(Select count(worksite)as Tcity,worksite,year from h1b_final where
case_status="CERTIFIED" or case_status="CERTIFIED-WITHDRAWN" group by worksite,year) a)b where
rank_1<6 and year is not null;
```

5) Find the most popular top 10 job positions for H1B visa applications for each year?

```
select job_title,year from(select rank() over(partition by year order by TJob desc)as rank_1,job_title,year
from(select count(job_title)as TJob,job_title,year from h1b_final group by job_title,year)a)b where
rank_1<=10 and year!= 'YEAR' and year is not null;
```

8) Find the average Prevailing Wage for each Job for each Year (take part time and full time separate)

full time:

```
select job_title,TJob,year from(select rank() over(partition by year order by TJob desc)as
rank_1,TJob,job_title,year from(select AVG(prevailing_wage)as TJob,job_title,year from h1b_final where
full_time_position ='Y' and prevailing_wage is not null group by job_title,year)a)b;
```

part time:

```
select job_title,TJob,year from(select rank() over(partition by year order by TJob desc)as
rank_1,TJob,job_title,year from(select AVG(prevailing_wage)as TJob,job_title,year from h1b_final where
full_time_position ='N' and prevailing_wage is not null group by job_title,year)a)b;
```

PIG:

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

```
REGISTER /home/hduser/Downloads/piggybank.jar;
DEFINE pigloader org.apache.pig.piggybank.storage.CSVExcelStorage();
data = load '/home/hduser/Downloads/h1b_kaggle.csv' using pigloader() as
(f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12);
noheader = filter data by $0 > 1;
data = order noheader by $0;
data1 = filter data by ($4 matches 'DATA ENGINEER');
--data1 = foreach data1 generate
data2 = group data1 by $7;
data3 = foreach data2 generate FLATTEN(group),COUNT(data1.$4);
data2011 = filter data3 by ($0 matches '2011');
data2012 = filter data3 by ($0 matches '2012');
data2013 = filter data3 by ($0 matches '2013');
data2014 = filter data3 by ($0 matches '2014');
data2015 = filter data3 by ($0 matches '2015');
data2016 = filter data3 by ($0 matches '2016');
data4 =foreach data2011 generate
data2011.$1,data2012.$1,data2013.$1,data2014.$1,data2015.$1,data2016.$1;
--dump data4;
data5 = foreach data4 generate  $0,$1,$2,$3,$4,$5,CONCAT((chararray)ROUND_TO((float)((($1-
$0)*100)/$0),2),'%'),CONCAT((chararray)ROUND_TO((float)((($2-$1)*100)/$1),2),'%'),CONCAT((cha-
rarray)ROUND_TO((float)((($3-$2)*100)/$2),2),'%'),CONCAT((chararray)ROUND_TO((float)((($4-
$3)*100)/$3),2),'%'),CONCAT((chararray)ROUND_TO((float)((($5-$4)*100)/$4),2),'%');
--dump data5;
data6 = foreach data5 generate $6,$7,$8,$9,$10;
perbyyear = foreach data6 generate FLATTEN(TOBAG(*));
dump perbyyear
```

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

```
REGISTER /home/hduser/Downloads/piggybank.jar;
DEFINE pigloader org.apache.pig.piggybank.storage.CSVExcelStorage();
data = load '/home/hduser/Downloads/h1b_kaggle.csv' using pigloader() as
(f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12);
noheader = filter data by $0 > 1;
data = order noheader by $0;
data2011 = filter data by ($7 matches '2011');
data2012 = filter data by ($7 matches '2012');
data2013 = filter data by ($7 matches '2013');
data2014 = filter data by ($7 matches '2014');
data2015 = filter data by ($7 matches '2015');
data2016 = filter data by ($7 matches '2016');
group2011 = group data2011 by $4;
count2011 = foreach group2011 generate group,COUNT(data2011.$1);
--dump count2011;
group2012 = group data2012 by $4;
count2012 = foreach group2012 generate group,COUNT(data2012.$1);
group2013 = group data2013 by $4;
```

```
count2013 = foreach group2013 generate group,COUNT(data2013.$1);
group2014 = group data2014 by $4;
count2014 = foreach group2014 generate group,COUNT(data2014.$1);
group2015 = group data2015 by $4;
count2015 = foreach group2015 generate group,COUNT(data2015.$1);
group2016 = group data2016 by $4;
count2016 = foreach group2016 generate group,COUNT(data2016.$1);


joindata = join count2011 by $0,count2012 by $0,count2013 by $0,count2014 by $0,count2015 by $0,count2016 by $0;

l1 = limit joindata 5;
--dump l1;
--dataclear = foreach joindata generate $0,$1,$11;
dataclear = foreach joindata generate $0,$1,$3,$5,$7,$9,$11;
percentage = foreach dataclear generate $0,ROUND_TO((float)((($2-$1)*100)/$1),2),ROUND_TO((float)((($3-$2)*100)/$2),2),ROUND_TO((float)((($4-$3)*100)/$3),2),ROUND_TO((float)((($5-$4)*100)/$4),2),ROUND_TO((float)((($6-$5)*100)/$5),2);

top = foreach percentage generate $0,($1+$2+$3+$4+$5)/5;
top = order top by $1 desc;
top5 = limit top 5;
top5 = foreach top5 generate $0,CONCAT((chararray) $1,'%');
dump top5;
```

4)Which top 5 employers file the most petitions each year?

```
REGISTER /home/hduser/Downloads/piggybank.jar;
DEFINE pigloader org.apache.pig.piggybank.storage.CSVExcelStorage();
data = load '/home/hduser/Downloads/h1b_kaggle.csv' using pigloader() as (f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12);
noheader = filter data by $0 > 1;
data = order noheader by $0;
data = foreach data generate $1,$2,$7;
data2011 = filter data by ($2 matches '2011');
data2012 = filter data by ($2 matches '2012');
data2013 = filter data by ($2 matches '2013');
data2014 = filter data by ($2 matches '2014');
data2015 = filter data by ($2 matches '2015');
data2016 = filter data by ($2 matches '2016');

groupdata2011 = group data2011 by ($1,$2);
groupdata2012 = group data2012 by ($1,$2);
groupdata2013 = group data2013 by ($1,$2);
groupdata2014 = group data2014 by ($1,$2);
groupdata2015 = group data2015 by ($1,$2);
groupdata2016 = group data2016 by ($1,$2);
```

```
data2011 = foreach groupdata2011 generate Flatten(group),COUNT(data2011.$0);
data2012 = foreach groupdata2012 generate FLATTEN(group),COUNT(data2012.$0);
data2013 = foreach groupdata2013 generate FLATTEN(group),COUNT(data2013.$0);
data2014 = foreach groupdata2014 generate FLATTEN(group),COUNT(data2014.$0);
data2015 = foreach groupdata2015 generate FLATTEN(group),COUNT(data2015.$0);
data2016 = foreach groupdata2016 generate FLATTEN(group),COUNT(data2016.$0);

dataorderd2011 = order data2011 by $2 desc;
dataorderd2012 = order data2012 by $2 desc;
dataorderd2013 = order data2013 by $2 desc;
dataorderd2014 = order data2014 by $2 desc;
dataorderd2015 = order data2015 by $2 desc;
dataorderd2016 = order data2016 by $2 desc;

top5_2011 = limit dataorderd2011 5;
top5_2012 = limit dataorderd2012 5;
top5_2013 = limit dataorderd2013 5;
top5_2014 = limit dataorderd2014 5;
top5_2015 = limit dataorderd2015 5;
top5_2016 = limit dataorderd2016 5;

uniondata = union top5_2011,top5_2012,top5_2013,top5_2014,top5_2015,top5_2016;
uniondata = order uniondata by $1;

dump uniondata;
```

6) 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.

```
REGISTER /home/hduser/Downloads/piggybank.jar;
DEFINE pigloader org.apache.pig.piggybank.storage.CSVExcelStorage();
data = load '/home/hduser/Downloads/h1b_kaggle.csv' using pigloader() as
(f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12);
noheader = filter data by $0 >= 1;
data = order noheader by $0;
data = foreach data generate $1,$7;
data2011 = filter data by ($1 matches '2011');
data2012 = filter data by ($1 matches '2012');
data2013 = filter data by ($1 matches '2013');
data2014 = filter data by ($1 matches '2014');
data2015 = filter data by ($1 matches '2015');
data2016 = filter data by ($1 matches '2016');

group2011 = group data2011 by $1;
group2012 = group data2012 by $1;
group2013 = group data2013 by $1;
group2014 = group data2014 by $1;
group2015 = group data2015 by $1;
group2016 = group data2016 by $1;
```

```
count2011 = foreach group2011 generate COUNT(data2011.$0);
count2012 = foreach group2012 generate COUNT(data2012.$0);
count2013 = foreach group2013 generate COUNT(data2013.$0);
count2014 = foreach group2014 generate COUNT(data2014.$0);
count2015 = foreach group2015 generate COUNT(data2015.$0);
count2016 = foreach group2016 generate COUNT(data2016.$0);


groupdata2011 = group data2011 by ($1,$0);
groupdata2012 = group data2012 by ($1,$0);
groupdata2013 = group data2013 by ($1,$0);
groupdata2014 = group data2014 by ($1,$0);
groupdata2015 = group data2015 by ($1,$0);
groupdata2016 = group data2016 by ($1,$0);

data2011 = foreach groupdata2011 generate group,COUNT(data2011.$0);
data2012 = foreach groupdata2012 generate group,COUNT(data2012.$0);
data2013 = foreach groupdata2013 generate group,COUNT(data2013.$0) ;
data2014 = foreach groupdata2014 generate group,COUNT(data2014.$0) ;
data2015 = foreach groupdata2015 generate group,COUNT(data2015.$0) ;
data2016 = foreach groupdata2016 generate group,COUNT(data2016.$0) ;


percentage2011 = foreach data2011 generate FLAT-
TEN(group),ROUND_TO((float)(($1)*100)/count2011.$0),2),$1;
percentage2012 = foreach data2012 generate FLAT-
TEN(group),ROUND_TO((float)(($1)*100)/count2012.$0),2),$1;
percentage2013 = foreach data2013 generate FLAT-
TEN(group),ROUND_TO((float)(($1)*100)/count2013.$0),2),$1;
percentage2014 = foreach data2014 generate FLAT-
TEN(group),ROUND_TO((float)(($1)*100)/count2014.$0),2),$1;
percentage2015 = foreach data2015 generate FLAT-
TEN(group),ROUND_TO((float)(($1)*100)/count2015.$0),2),$1;
percentage2016 = foreach data2016 generate FLAT-
TEN(group),ROUND_TO((float)(($1)*100)/count2016.$0),2),$1;


store percentage2011 into '/home/hduser/project/projectoutputs/project62011' using PigStorage(',');
store percentage2012 into '/home/hduser/project/projectoutputs/project62012' using PigStorage(',');
store percentage2013 into '/home/hduser/project/projectoutputs/project62013' using PigStorage(',');
store percentage2014 into '/home/hduser/project/projectoutputs/project62014' using PigStorage(',');
store percentage2015 into '/home/hduser/project/projectoutputs/project62015' using PigStorage(',');
store percentage2016 into '/home/hduser/project/projectoutputs/project62016' using PigStorage(',');

--dump percentage2011
--dump percentage2012
--dump percentage2013
--dump percentage2014
--dump percentage2015
--dump percentage2016
```

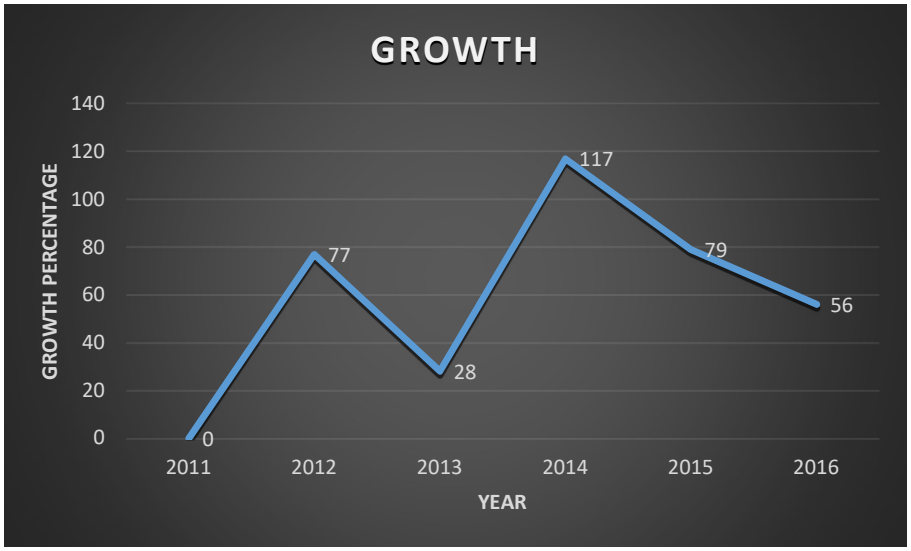
SQOOP:

```
mysql -u root -p'hadoop' -e 'drop database h1b11;
create database if not exists h1b11;
use h1b11;
create table h1b11(job_title varchar(100),petitions int,success_rate float);

sqoop export --connect jdbc:mysql://localhost/h1b11 --username root --password 'hadoop' --table
h1b11 --update-mode allowinsert  --export-dir /projectoutput/q10/p* --input-fields-terminated-by
'@' ;

echo -e '\n\nDisplay contents from MySQL Database.\n\n'

echo -e '\n  Which are the top 10 job positions that have  success rate more than 70% in petitions
and total petitions filed more than 1000?\n\n'

mysql -u root -p'hadoop' -e 'select * from h1b11.h1b11';
```

OUTPUTS:

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

2011-2012,(77.0%)
2012-2013,(28.0%)
2013-2014,(117.0%)
2014-2015,(79.0%)
2015-2016,(56.0%)



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

(SENIOR SYSTEMS ANALYST JC60,4255.4%)
(SOFTWARE DEVELOPER 2,3480.8%)
(PROJECT MANAGER 3,3233.4%)
(SYSTEMS ANALYST JC65,2985.0%)
(MODULE LEAD,2917.2%)

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

SAN FRANCISCO, CALIFORNIA	3	2011
SAN FRANCISCO, CALIFORNIA	7	2012
MENLO PARK, CALIFORNIA	10	2013
MENLO PARK, CALIFORNIA	13	2014
SAN FRANCISCO, CALIFORNIA	33	2015
MENLO PARK, CALIFORNIA	35	2016

b) find top 5 locations in the US who have got certified visa for each year.

NEW YORK, NEW YORK	23172	2011
HOUSTON, TEXAS	8184	2011
CHICAGO, ILLINOIS	5188	2011
SAN JOSE, CALIFORNIA	4713	2011
SAN FRANCISCO, CALIFORNIA	4711	2011
NEW YORK, NEW YORK	23736	2012

HOUSTON, TEXAS 9963 2012
SAN FRANCISCO, CALIFORNIA 6116 2012
CHICAGO, ILLINOIS 5671 2012
ATLANTA, GEORGIA 5565 2012

NEW YORK, NEW YORK 23537 2013
HOUSTON, TEXAS 11136 2013
SAN FRANCISCO, CALIFORNIA 7281 2013
SAN JOSE, CALIFORNIA 6722 2013
ATLANTA, GEORGIA 6377 2013

NEW YORK, NEW YORK 27634 2014
HOUSTON, TEXAS 13360 2014
SAN FRANCISCO, CALIFORNIA 9798 2014
SAN JOSE, CALIFORNIA 8223 2014
ATLANTA, GEORGIA 8213 2014

NEW YORK, NEW YORK 31266 2015
HOUSTON, TEXAS 15242 2015
SAN FRANCISCO, CALIFORNIA 12594 2015
ATLANTA, GEORGIA 10500 2015
SAN JOSE, CALIFORNIA 9589 2015

NEW YORK, NEW YORK 34639 2016
SAN FRANCISCO, CALIFORNIA 13836 2016
HOUSTON, TEXAS 13655 2016
ATLANTA, GEORGIA 11678 2016
CHICAGO, ILLINOIS 11064 2016

3)Which industry has the most number of Data Scientist positions?

Launched map tasks=5
Launched reduce tasks=1
Data-local map tasks=5

"MICROSOFT CORPORATION",139

4)Which top 5 employers file the most petitions each year?

(TATA CONSULTANCY SERVICES LIMITED,2011,5416)
(MICROSOFT CORPORATION,2011,4253)
(DELOITTE CONSULTING LLP,2011,3621)
(WIPRO LIMITED,2011,3028)
(COGNIZANT TECHNOLOGY SOLUTIONS U.S. CORPORATION,2011,2721)

(WIPRO LIMITED,2012,7182)
(TATA CONSULTANCY SERVICES LIMITED,2012,6735)
(DELOITTE CONSULTING LLP,2012,4727)
(IBM INDIA PRIVATE LIMITED,2012,4074)
(INFOSYS LIMITED,2012,15818)

(ACCENTURE LLP,2013,4994)
(DELOITTE CONSULTING LLP,2013,6124)

(WIPRO LIMITED,2013,6734)
(TATA CONSULTANCY SERVICES LIMITED,2013,8790)
(INFOSYS LIMITED,2013,32223)

(ACCENTURE LLP,2014,5498)
(DELOITTE CONSULTING LLP,2014,7017)
(WIPRO LIMITED,2014,8365)
(TATA CONSULTANCY SERVICES LIMITED,2014,14098)
(INFOSYS LIMITED,2014,23759)

(INFOSYS LIMITED,2015,33245)
(TATA CONSULTANCY SERVICES LIMITED,2015,16553)
(WIPRO LIMITED,2015,12201)
(IBM INDIA PRIVATE LIMITED,2015,10693)
(ACCENTURE LLP,2015,9605)

(INFOSYS LIMITED,2016,25352)
(CAPGEMINI AMERICA INC,2016,16725)
(TATA CONSULTANCY SERVICES LIMITED,2016,13134)
(WIPRO LIMITED,2016,10607)
(IBM INDIA PRIVATE LIMITED,2016,9787)

5) Find the most popular top 10 job positions for H1B visa applications for each year?

PROGRAMMER ANALYST	31799	2011
SOFTWARE ENGINEER	12763	2011
COMPUTER PROGRAMMER	8998	2011
SYSTEMS ANALYST	8644	2011
BUSINESS ANALYST	3891	2011
COMPUTER SYSTEMS ANALYST	3698	2011
ASSISTANT PROFESSOR	3467	2011
PHYSICAL THERAPIST	3377	2011
SENIOR SOFTWARE ENGINEER	2935	2011
SENIOR CONSULTANT	2798	2011

PROGRAMMER ANALYST	33066	2012
SOFTWARE ENGINEER	14437	2012
COMPUTER PROGRAMMER	9629	2012
SYSTEMS ANALYST	9296	2012
BUSINESS ANALYST	4752	2012
COMPUTER SYSTEMS ANALYST	4706	2012
SOFTWARE DEVELOPER	3895	2012
PHYSICAL THERAPIST	3871	2012
ASSISTANT PROFESSOR	3801	2012
SENIOR CONSULTANT	3737	2012

PROGRAMMER ANALYST	33880	2013
SOFTWARE ENGINEER	15680	2013
COMPUTER PROGRAMMER	11271	2013
SYSTEMS ANALYST	8714	2013
TECHNOLOGY LEAD - US	7853	2013
TECHNOLOGY ANALYST - US	7683	2013

BUSINESS ANALYST 5716 2013
COMPUTER SYSTEMS ANALYST 5043 2013
SOFTWARE DEVELOPER 5026 2013
SENIOR CONSULTANT 4326 2013

PROGRAMMER ANALYST 43114 2014
SOFTWARE ENGINEER 20500 2014
COMPUTER PROGRAMMER 14950 2014
SYSTEMS ANALYST 10194 2014
SOFTWARE DEVELOPER 7337 2014
BUSINESS ANALYST 7302 2014
COMPUTER SYSTEMS ANALYST 6821 2014
TECHNOLOGY LEAD - US 5057 2014
TECHNOLOGY ANALYST - US 4913 2014
SENIOR CONSULTANT 4898 2014

PROGRAMMER ANALYST 53436 2015
SOFTWARE ENGINEER 27259 2015
COMPUTER PROGRAMMER 14054 2015
SYSTEMS ANALYST 12803 2015
SOFTWARE DEVELOPER 10441 2015
BUSINESS ANALYST 8853 2015
TECHNOLOGY LEAD - US 8242 2015
COMPUTER SYSTEMS ANALYST 7918 2015
TECHNOLOGY ANALYST - US 7014 2015
SENIOR SOFTWARE ENGINEER 6013 2015

PROGRAMMER ANALYST 53743 2016
SOFTWARE ENGINEER 30668 2016
SOFTWARE DEVELOPER 14041 2016
SYSTEMS ANALYST 12314 2016
COMPUTER PROGRAMMER 11668 2016
BUSINESS ANALYST 9167 2016
COMPUTER SYSTEMS ANALYST 6900 2016
SENIOR SOFTWARE ENGINEER 6439 2016
DEVELOPER6084 2016
TECHNOLOGY LEAD - US 5410 2016

6) 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.

2011,DENIED,8.0%,29130
2011,CERTIFIED,85.0%,307936
2011,WITHDRAWN,2.0%,10105
2011,CERTIFIED-WITHDRAWN,3.0%,11596

2012,DENIED,5.0%,21096
2012,CERTIFIED,84.0%,352668
2012,WITHDRAWN,2.0%,10725
2012,CERTIFIED-WITHDRAWN,7.0%,31118

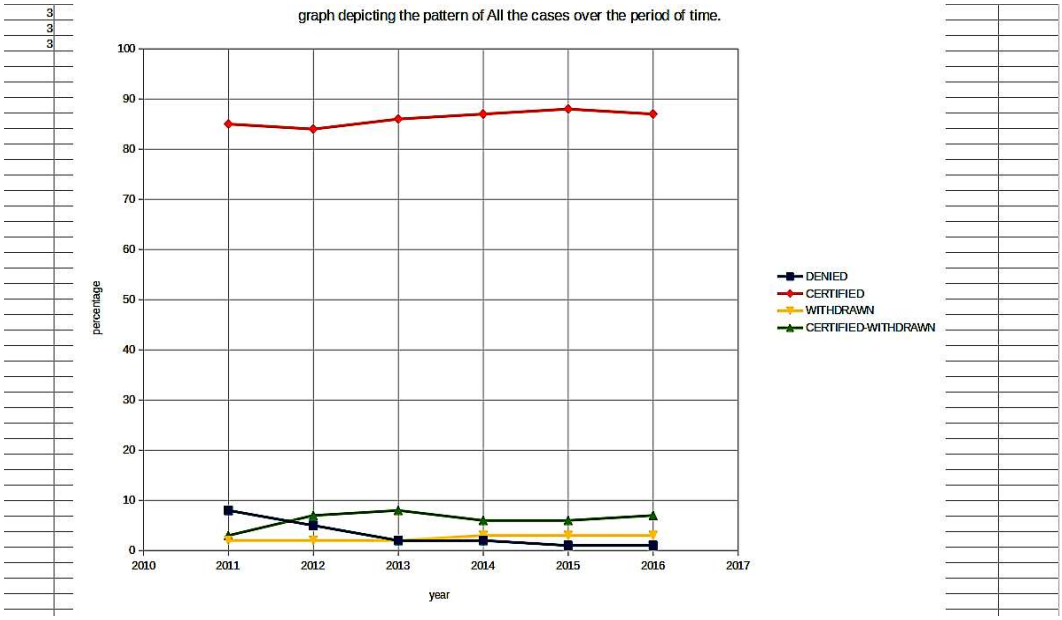
2013,DENIED,2.0%,12126
2013,CERTIFIED,86.0%,382951

2013,WITHDRAWN,2.0%,11590
2013,CERTIFIED-WITHDRAWN,8.0%,35432
2013,PENDING QUALITY AND COMPLIANCE REVIEW – UNASSIGNED,0.0%,15

2014,DENIED,2.0%,11896
2014,REJECTED,0.0%,2
2014,CERTIFIED,87.0%,455144
2014,WITHDRAWN,3.0%,16034
2014,INVALIDATED,0.0%,1
2014,CERTIFIED-WITHDRAWN,6.0%,36350

2015,DENIED,1.0%,10923
2015,CERTIFIED,88.0%,547278
2015,WITHDRAWN,3.0%,19455
2015,CERTIFIED-WITHDRAWN,6.0%,41071

2016,DENIED,1.0%,9175
2016,CERTIFIED,87.0%,569646
2016,WITHDRAWN,3.0%,21890
2016,CERTIFIED-WITHDRAWN,7.0%,47092

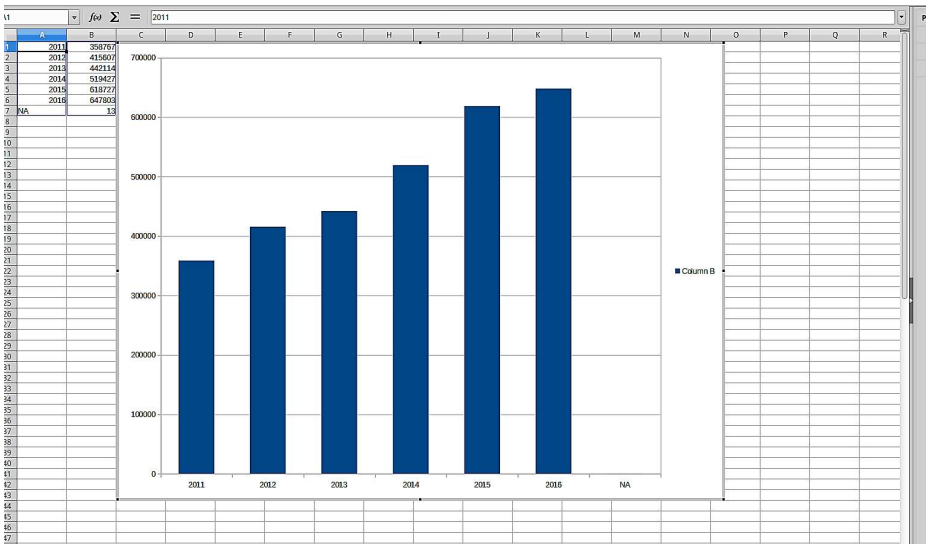


7) Create a bar graph to depict the number of applications for each year

Launched map tasks=5
Launched reduce tasks=1
Data-local map tasks=5

2011 358767
2012 415607
2013 442114
2014 519427
2015 618727

2016 647803
NA 13



8) Find the average Prevailing Wage for each Job for each Year (take part time and full time separate)

full time:

PROGRAM MANAGER (POSITIVE PATHWAYS) 3.289741175E8 2011
DIRECTOR OF SRCH ENGINE OPTMZTN/SRCH ENGINE MKTNG 2.4593712E8 2011
ASSOCIATE DIRECTOR, HEALTH ECONOMICS AND OUTCOMES 2.2142432E8 2011
AREA MANAGER, PHARMACEUTICAL PACKAGING 2.1298784E8 2011
SR. WEB INFRASTRUCTURE CONSULT 2.0801248E8 2011
DEVELOPER (SOFTWARE SYSTEMS APPLICATIONS) 1.765608E8 2011
PHYSICIAN, UROLOGIST 1.731392E8 2011
ELECTRICAL COMPLETIONS ENGINEER - LEVEL 4 1.70261177E8 2011
ASSISTANT PROFESSOR/HEMATOLOGY, BONE MARROW TRANSP 1.60738521E8 2011
VICE PRESIDENT, DIRECTOR OF STRATEGIC PLANNING 1.53941975E8 2011

HOSPITALIST (INTERNAL MEDICING 2.980016E8 2012
PHYSICIAN (GENERAL INVASIVE CARDIOLOGY/ELECTROPHYS 1.731392E8 2012
DIRECTOR, STRATEGIC ANALYSIS & IMPLEMENTATION 1.45805265E8 2012
PHYSICIAN, NEUROLOGIST 1.43553623E8 2012
DIRECTOR, STRATEGIC TALENT PLANNING 1.407619615E8 2012
SHAREPOINTDEVELOPER 1.3117728E8 2012
BUSINESS/COMPUTER SYSTEMS ANALYST 1.2797408E8 2012
DIRECTOR, INFORMATION TECHNOLOGIES 1.233398295E8 2012
SOFTWARE ENGINEER(.NET DEVELOPER) 1.208792E8 2012
DIRECTOR FINANCE & CONTROLLING 1.18383928E8 2012

ASSOCIATE MED. DIRECTOR & HOSPICE/PALLIATIVE CARE 2.093179796E8 2013
PROGRAMMER ANALYST/SAP SRM ANALYST 2.071056E8 2013
AREA MANAGER RAN ENGINEERING 1.8698784E8 2013
PROGRAM DIRECTOR, STORE ORDER AND INVENTORY OPTIMI 1.7638816E8 2013
STAFF CONSULTANT - MICRO 1.6950752E8 2013

QAULITY TEST ENGINEER 1.5587936E8 2013
SLAMBERTPA@AOL.COM 1.4458912E8 2013
QUALITY ASSURANCE ANALYST/ COMPUTER PROGRAMMER 1.40088E8 2013
PROJECT GEOSTATISTICIAN 1.3943904E8 2013
DIRECTOR, GLOBAL OPERATIONS 1.39096121E8 2013

CHIEF FINANCIAL OFFICER, HUNTSWORTH HEALTH GLOBAL 1.80605828E8 2014
PHYSICIAL THERAPIST 1.487408E8 2014
SENIOR FLOW ASSURANCE ENGINEER 1.379921505E8 2014
AREA RETAIL SALES AND OPERATIONS MANAGER 1.3052832E8 2014
SENIOR R&D EMI/SI ENGINEER 1.29053215E8 2014
DEPUTY PROGRAM DIRECTOR 1.22928832E8 2014
TECHNICAL MANAGER - PORTALS & COLLABORATION 1.159814135E8 2014
PRESIDENT, AMERICAS 1.134280265E8 2014
ASSOCIATE DIRECTOR, SAP FICO 1.06761543E8 2014
TECHNICAL MANAGER, APPLICATION IMPLEMENTATION & IN 1.037534575E8 2014

MANAGER, GEORGIAN, CAUCASUS, AND EASTERN EUROPE REGIONAL MAN 1.2308608E8 2015
STAFF DIGITAL VERIFICATION ENGINEER 1.108309385E8 2015
ENGINEERING QUALITY ANALYST (15-1199.0) 9.993984E7 2015
SR. MANAGER, SOX & INTERNAL AUDIT GROUP 9.06598055E7 2015
PEOPLESOFT HCM ANALYST 8.652141675E7 2015
REGISTERED PHARMACIST 8.61050652E7 2015
TECHNICAL SUPPORT SPECIALIST (ACCOUNTING) 6.632288E7 2015
COMPUTER SOFTWARE TECHNICAL SUPPORT SPECIALIST 6.632288E7 2015
MUNICIPAL FIXED-INCOME RESEARCH ASSOCIATE 5.2049972E7 2015
SENIOR CONSULTING GEOLOGIST 4.18909008E7 2015

DIRECTOR, SOCIAL AND DIGITAL MEDIA 3.291392E8 2016
CONSTRUCTION TECHNICAL SERVICES SPECIALIST II 1.3827216E8 2016
LAB ENGINEER 1.2992096E8 2016
GLOBAL BUSINESS MANAGER 1.25962065E8 2016
ERP BI DEVELOPER 3 1.152112E8 2016
ARCHITECTURAL DESIGNER (INTERN ARCHITECT) 1.1499488E8 2016
ELEMENTARY TEACHER 1.147224E8 2016
SENIOR ANALYST, IT ANALYSIS 1.1214112E8 2016
RESEARCHER- NANOSCIENCE/NANOENGINEERING 9.621872E7 2016
WORKDAY SOLUTIONS CONSULTANT 9.57894705E7 2016

part time:

OPERATIONS/CLIENT SERVICE MANAGER 1.0673312E8 2011
STRATEGIC MANAGEMENT ANALYST 3.5611915333333336E7 2011
DIRECTOR 2.8270314888888888E7 2011
REGGAE DANCEHALL ARTIST 1.311648E7 2011
HEADLINE ENTERTAINER 1.04E7 2011
DATA ADMINISTRATOR 8976796.8 2011
INTERNAL AUDITOR 5770078.933333334 2011
PMO MANAGER 4186894.0 2011
INTERNATIONAL MERCHANDISE SALES MANAGER 2768604.6666666665 2011
PROJECT BUDGET ANALYST 2115412.0 2011

TEST ANALYST - US 9.215232E7 2012
ADJUNCT INSTRUCTOR FASHION DESIGN 3.25364E7 2012
PURCHASING MANAGERS 2.1827256333333332E7 2012
SPANISH TEACHER 9069436.3 2012
ADJUNCT ASSISTANT PROFESSOR CLINICAL 4376320.0 2012
FAMILY MEDICINE PHYSICIAN/HOSPITALIST 3457916.5 2012
PASTORAL ADMINISTRATOR 2256800.0 2012
ENGINEERING MANAGER 1623987.5 2012
MECHANICAL ENGINEER 1605184.2467532468 2012
VIDEO EDITOR, DIRECTOR OF PHOTOGRAPHY AND VIDEOGRA 1239721.3333333333 2012
MANAGEMENT ANALYST, OPERATIONS 1.2351872E8 2013
BIOMEDICAL ENGINEERING 1.1711648E8 2013
BIZTALK PROGRAMMER 1.1252696E8 2013
BRAND SPECIALIST 1.0854896E8 2013
PROCESS ENGINEER 6.66571565E7 2013
HUMAN RESOURCES SPECIALISTS 6.35997335E7 2013
GEOSCIENTIST 3.7202429E7 2013
MANAGING DIRECTOR 3.1738961E7 2013
MATHEMATICS INSTRUCTOR 2.7937426E7 2013
SALES ENGINEER 6656320.2272727275 2013

GYMNASTICS TRAINING CENTER OPERATIONS MANAGER 6.5684684E7 2014
WEB DEVELOPERS 5.58810925E7 2014
MINISTER, PART TIME 2.2117295E7 2014
NETWORK ENGINEER 1.3616277625E7 2014
INDUSTRIAL ENGINEER 1.2808594708333334E7 2014
LAWYER 3694982.9696969697 2014
LASER MICROMACHINING ENGINEER 2115557.5 2014
COMPUTER PROGRAMMER 800276.1384615385 2014
STAFF ACCOUNTANT 527876.4495412844 2014
VISITING ASSOCIATE PROFESSOR 429000.0 2014

VASCULAR SURGEON & SPECIALITY MEDICINE CONSULTANT 324688.0 2015
PSYCHIATRIST II 291200.0 2015
ASSISTANT PROFESSOR ADJUNCT OF ARCHITECTURE 284377.0 2015
ADJUNCT ASSOCIATE PROFESSOR 257920.0 2015
CARDIOLOGIST 246176.0 2015
SPECIAL LEGAL COUNSEL (CORPORATE) 245856.0 2015
PHYSICIAN - EMERGENCY MEDICIN 242860.0 2015
SOCIAL ENTREPRENEUR IN RESIDENCE 239200.0 2015
PRESIDENT AND CEO 233272.0 2015
PEDIATRIC CRITICAL CARE PHYSICIAN 229444.0 2015

CLINICAL PSYCHOLOGIST - TESTING COORDINATOR 70000.0 2016
TESTING COORDINATOR - CLINICAL PSYCHOLOGIST 70000.0 2016
ELECTRICAL ENGINEER R&D 70000.0 2016
VP, DEVELOPMENT AND PUBLIC AFFAIRS 70000.0 2016
DEMAND GENERATION MANAGER 70000.0 2016
SOFTWARE DEVELOPER, HIGH PERFORMANCE COMPUTING APP DEVELOPER 70000.0 2016
RESEARCH ENGINEER, ASSOCIATE 70000.0 2016
COMMERCIALIZATION SERVICES MANAGER 70000.0 2016
ASSISTANT PROFESSOR OF COMPUTER AND INFORMATION SCIENCES 70000.0 2016

NETWORK ARCHITECT & SECURITY SPECIALISTS 70000.0 2016
CHIEF FINANACIAL OFFICER 70000.0 2016
SYSTEM PERFORMANCE ENGINEER 70000.0 2016
IMMIGRATION ATTORNEY 70000.0 2016
OUTREACH AND RESEARCH COORDINATOR 70000.0 2016

9) Which are top ten employers who have the highest success rate in petitions?

Launched map tasks=5
Launched reduce tasks=1
Data-local map tasks=5

"HTC GLOBAL SERVICES, INC.",1164,100.00 %
"INFOSYS LIMITED",130592,99.54 %
"DIASPARK, INC.",1419,99.51 %
"ACCENTURE LLP",33447,99.39 %
"TECH MAHINDRA (AMERICAS),INC.",10732,99.34 %
"TATA CONSULTANCY SERVICES LIMITED",64726,99.34 %
"YASH TECHNOLOGIES, INC.",2214,99.28 %
"YASH & LUJAN CONSULTING, INC.",1372,99.27 %
"HCL AMERICA, INC.",22678,99.27 %
"RELIABLE SOFTWARE RESOURCES, INC.",1992,99.15 %
"NTT DATA, INC.",4611,99.13 %
"ERP ANALYSTS, INC.",1785,99.10 %
"PATNI AMERICAS INC.",3149,99.08 %
"MINDTREE LIMITED",4067,99.07 %
"KFORCE INC.",1596,99.06 %
"TECH MAHINDRA (AMERICAS), INC",1170,99.06 %
"GRANDISON MANAGEMENT, INC.",1386,98.99 %
"GENPACT LLC",1046,98.85 %
"SMARTPLAY, INC.",1377,98.84 %
"SYNTEL CONSULTING INC.",3167,98.83 %
"CREDIT SUISSE SECURITIES (USA) LLC",2546,98.82 %
"MASTECH, INC., A MASTECH HOLDINGS, INC. COMPANY",5228,98.81 %
"GENESIS ELDERCARE REHABILITATION SERVICES, INC.",1320,98.79 %
"HORIZON TECHNOLOGIES INC",1731,98.79 %
"SYNTEL INC",1946,98.77 %
"THE BOSTON CONSULTING GROUP, INC.",1352,98.74 %
"AMDOCS INC.",1023,98.73 %
"SAP AMERICA, INC.",1456,98.70 %
"DELOITTE TAX LLP",2501,98.64 %
"MPHASIS CORPORATION",5199,98.63 %
"3I INFOTECH, INC.",2041,98.58 %
"COMPUNNEL SOFTWARE GROUP, INC.",3378,98.58 %
"THE MATHWORKS, INC.",2020,98.47 %
"PERFICIENT, INC.",1366,98.46 %
"DALLAS INDEPENDENT SCHOOL DISTRICT",1729,98.44 %
"CGI TECHNOLOGIES AND SOLUTIONS INC.",1995,98.40 %
"VEDICSOFT",1169,98.37 %
"UNIVERSITY OF PITTSBURGH",1632,98.35 %
"DELOITTE CONSULTING LLP",36742,98.33 %
"BLOOMBERG, LP",2352,98.30 %
"WIPRO LIMITED",48117,98.29 %

"MCKINSEY & COMPANY, INC. UNITED STATES",1097,98.27 %
"AVCO CONSULTING INC",1424,98.24 %
"MICHIGAN STATE UNIVERSITY",1191,98.24 %
"NVIDIA CORPORATION",1182,98.22 %
"ALINDUS, INC.",1046,98.18 %
"PRICEWATERHOUSECOOPERS, LLP",2529,98.14 %
"CVS RX SERVICES, INC.",2735,98.14 %
"PYRAMID TECHNOLOGY SOLUTIONS, INC",1056,98.11 %
"MICROSOFT CORPORATION",25576,98.09 %
"CYBERTHINK INC",1618,98.08 %
"BARCLAYS SERVICES CORP.",1605,98.07 %
"HARVARD UNIVERSITY",1966,98.07 %
"INOVANT, LLC",1086,98.07 %
"ERNST & YOUNG U.S. LLP",18232,98.05 %
"DELOITTE & TOUCHE LLP",9642,98.00 %
"CAPGEMINI AMERICA INC",16725,97.96 %
"TECH MAHINDRA (AMERICAS) INC.",2102,97.95 %
"SUNERA TECHNOLOGIES, INC",1440,97.92 %
"SRS CONSULTING INC.",1150,97.91 %
"AKVARR INC",1372,97.89 %
"LARSEN & TOUBRO TECHNOLOGY SERVICES LTD",1385,97.83 %
"MARLABS, INC",2626,97.83 %
"CYIENT, INC.",1281,97.81 %
"MERRILL LYNCH",1873,97.81 %
"NATSOFT CORPORATION",1137,97.80 %
"WASHINGTON UNIVERSITY IN ST. LOUIS",1576,97.78 %
"KPMG LLP",4629,97.77 %
"THE UNIV. OF ALA. AT BIRMINGHAM (UAB)",1288,97.75 %
"APEX TECHNOLOGY SYSTEMS, INC",1060,97.74 %
"LINKEDIN CORPORATION",2194,97.72 %
"HEADSTRONG SERVICES LLC",2587,97.72 %
"SATYAM COMPUTER SERVICES LTD",1622,97.72 %
"BIRLASOFT INC",2370,97.68 %
"ERICSSON INC.",3359,97.65 %
"UBER TECHNOLOGIES, INC.",1006,97.61 %
"INFOSYS TECHNOLOGIES LIMITED",1336,97.60 %
"DOTCOM TEAM, LLC",1125,97.60 %
"APPLE INC.",7317,97.59 %
"PHOTON INFOTECH, INC.",1235,97.57 %
"UNIVERSITY OF UTAH",1069,97.57 %
"UNIVERSITY OF MINNESOTA",1353,97.56 %
"TEXAS INSTRUMENTS INCORPORATED",1780,97.53 %
"COMPUTER SCIENCES CORPORATION",1089,97.52 %
"PRICEWATERHOUSECOOPERS ADVISORY SERVICES LLC",1724,97.51 %
"MEMORIAL SLOAN-KETTERING CANCER CENTER",1080,97.50 %
"CAPITAL ONE SERVICES, LLC",2796,97.50 %
"ORACLE AMERICA, INC.",7684,97.49 %
"CSC COVANSYS CORPORATION",2251,97.47 %
"CITIGROUP GLOBAL MARKETS INC.",1435,97.42 %
"MICROEXCEL, INC",1159,97.41 %
"SCHLUMBERGER TECHNOLOGY CORPORATION",2310,97.40 %
"RITE AID CORP.",1577,97.40 %
"ASTIR IT SOLUTIONS INC.",1955,97.34 %

"SATYAM COMPUTER SERVICES LIMITED",2403,97.34 %
"INTONE NETWORKS INC.",1575,97.21 %
"COMCAST CABLE COMMUNICATIONS, LLC",1214,97.20 %
"BANK OF AMERICA N.A.",4282,97.20 %
"RJT COMPUQUEST, INC.",1662,97.17 %
"CHILDREN'S HOSPITAL CORPORATION",1017,97.15 %
"UNIVERSITY OF CALIFORNIA, SAN FRANCISCO",1348,97.11 %
"VMWARE, INC.",2617,97.02 %
"TESLA MOTORS, INC.",1441,97.02 %
"PRICEWATERHOUSECOOPERS LLP",2719,96.98 %
"HP ENTERPRISE SERVICES, LLC",1149,96.95 %
"AMERICAN EXPRESS TRAVEL RELATED SERVICES COMPANY, INC.",1045,96.94 %
"UNIVERSITY OF MICHIGAN",2893,96.92 %
"TWITTER, INC.",1328,96.91 %
"JPMORGAN CHASE & CO.",7035,96.87 %
"HCL GLOBAL SYSTEMS INC",3677,96.87 %
"IDHASOFT, INC.",1423,96.84 %
"CAPGEMINI FINANCIAL SERVICES USA INC",4426,96.81 %
"CHARTER GLOBAL, INC.",1188,96.80 %
"RANDSTAD TECHNOLOGIES, LP",3419,96.78 %
"SMARTSOFT INTERNATIONAL, INC.",1212,96.78 %
"VEDICSOFT SOLUTIONS LLC",1274,96.78 %
"V-SOFT CONSULTING GROUP, INC",4283,96.75 %
"UNIVERSITY OF ILLINOIS",1196,96.74 %
"THE UNIVERSITY OF CHICAGO",1277,96.71 %
"TECHDEMOCRACY LLC",1027,96.69 %
"VERIZON DATA SERVICES LLC",1635,96.64 %
"EMC CORPORATION",4467,96.62 %
"IDEXCEL, INC.",1360,96.62 %
"GOOGLE INC.",16473,96.59 %
"BLACKROCK FINANCIAL MANAGEMENT, INC.",1048,96.47 %
"VERINON TECHNOLOGY SOLUTIONS LTD.",1245,96.47 %
"HEWLETT-PACKARD COMPANY",1639,96.46 %
"EBAY INC.",3464,96.45 %
"IGATE TECHNOLOGIES INC.",12564,96.44 %
"LARSEN & TOUBRO INFOTECH LIMITED",17457,96.43 %
"T-MOBILE USA, INC.",1457,96.43 %
"UNIVERSITY OF WISCONSIN-MADISON",1115,96.41 %
"GENERAL HOSPITAL CORPORATION",1429,96.36 %
"SYMANTEC CORPORATION",2290,96.33 %
"PAYPAL, INC.",2830,96.33 %
"FACEBOOK, INC.",4145,96.31 %
"COLUMBIA UNIVERSITY",1841,96.31 %
"SAGARSOFT, INC",1082,96.30 %
"PURDUE UNIVERSITY",1076,96.28 %
"SALESFORCE.COM, INC.",2245,96.26 %
"COLLABORATE SOLUTIONS INC",1209,96.20 %
"NIIT TECHNOLOGIES LIMITED",1339,96.19 %
"PROKARMA, INC.",1333,96.17 %
"TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA",2258,96.10 %
"EXPERIS US, INC.",1641,96.10 %
"NEW YORK UNIVERSITY SCHOOL OF MEDICINE",1126,96.09 %
"CERNER CORPORATION",2268,96.08 %

"MASSACHUSETTS INSTITUTE OF TECHNOLOGY",1347,96.07 %
"JOHNS HOPKINS UNIVERSITY",1823,96.05 %
"LEAD IT CORPORATION",1720,96.05 %
"WAL-MART ASSOCIATES, INC.",3670,96.02 %
"UNIVERSITY OF CALIFORNIA, LOS ANGELES",1172,95.99 %
"UST GLOBAL INC.",6363,95.93 %
"THE UNIVERSITY OF IOWA",1569,95.92 %
"DEUTSCHE BANK SECURITIES INC.",1170,95.90 %
"POLARIS SOFTWARE LAB (INDIA) LTD.",1326,95.85 %
"A2Z DEVELOPMENT CENTER, INC.",1025,95.80 %
"EXILANT TECHNOLOGIES PRIVATE LIMITED",1572,95.80 %
"NATIONAL INSTITUTES OF HEALTH, HHS",2327,95.75 %
"THE OHIO STATE UNIVERSITY",1587,95.72 %
"TECHNOSOFT CORPORATION",1625,95.69 %
"SYNECHRON, INC.",3802,95.69 %
"SEARS HOLDINGS MANAGEMENT CORPORATION",1105,95.66 %
"INTRAEDGE, INC.",1254,95.61 %
"ADVENT GLOBAL SOLUTIONS INC.",1048,95.52 %
"HEXAWARE TECHNOLOGIES, INC.",5466,95.50 %
"YAHOO! INC.",3348,95.46 %
"SATYAM COMPUTER SERVICES LTD.",1694,95.34 %
"CMC AMERICAS, INC.",1157,95.33 %
"AMAZON CORPORATE LLC",9026,95.29 %
"DELL MARKETING L.P.",1532,95.23 %
"EXPEDIA, INC.",1311,95.19 %
"CYMA SYSTEMS INC",1269,95.19 %
"MULTIVISION INC.",1502,95.01 %
"UNIVERSITY OF CALIFORNIA, DAVIS",1334,94.98 %
"MOUNT SINAI MEDICAL CENTER",1114,94.97 %
"SYSTEM SOFT TECHNOLOGIES LLC",3102,94.97 %
"CIBER, INC.",2097,94.95 %
"CISCO SYSTEMS, INC.",3140,94.90 %
"THE UNIVERSITY OF TEXAS AT AUSTIN",1274,94.90 %
"HOWARD HUGHES MEDICAL INSTITUTE",1135,94.89 %
"QUALCOMM TECHNOLOGIES, INC.",6113,94.88 %
"BRIGHAM AND WOMEN'S HOSPITAL",1117,94.81 %
"ADOBE SYSTEMS INCORPORATED",1167,94.77 %
"AVANT HEALTHCARE PROFESSIONALS",1006,94.73 %
"UST GLOBAL INC",6355,94.73 %
"POPULUS GROUP",2635,94.72 %
"ORACLE FINANCIAL SERVICES SOFTWARE, INC.",1532,94.65 %
"BROOKHAVEN NATIONAL LABORATORY",1023,94.62 %
"SAPIENT CORPORATION",2237,94.59 %
"AT&T SERVICES, INC.",1201,94.59 %
"LARSEN & TOUBRO LIMITED",3066,94.59 %
"UNIVERSITY OF FLORIDA",1429,94.54 %
"HITACHI CONSULTING CORPORATION",2854,94.50 %
"MAYO CLINIC",1772,94.47 %
"KPIT INFOSYSTEMS, INC.",3114,94.44 %
"QUALCOMM INCORPORATED",3965,94.43 %
"FUJITSU AMERICA, INC.",5309,94.35 %
"ITECH US, INC.",2476,94.35 %
"YALE UNIVERSITY",1852,94.33 %

"UNIVERSITY OF ILLINOIS AT CHICAGO",1131,94.25 %
"UNIVERSITY OF COLORADO",1599,94.12 %
"MANAGEMENT HEALTH SYSTEMS, INC.",2000,94.10 %
"UNIVERSITY OF WASHINGTON",1187,94.02 %
"ECLINICALWORKS, LLC",1547,93.92 %
"GLOBALFOUNDRIES U.S. INC.",1391,93.82 %
"TECH MAHINDRA (AMERICAS), INC.",7019,93.79 %
"NORTHWESTERN UNIVERSITY",1439,93.47 %
"CAPGEMINI U.S. LLC",3712,93.35 %
"ORION SYSTEMS INTEGRATORS, INC",1160,93.28 %
"UNIVERSITY OF CALIFORNIA, SAN DIEGO",1202,93.26 %
"SOFTWARE PARADIGMS INTERNATIONAL GROUP, LLC",1034,93.23 %
"DUKE UNIVERSITY AND MEDICAL CENTER",1330,93.16 %
"GOLDMAN, SACHS & CO.",3713,93.11 %
"EMORY UNIVERSITY",1680,93.10 %
"BAYLOR COLLEGE OF MEDICINE",1666,93.04 %
"MICRON TECHNOLOGY, INC.",1934,93.02 %
"QUALCOMM ATHEROS, INC.",1274,93.01 %
"AKAMAI TECHNOLOGIES, INC.",1092,92.95 %
"MORGAN STANLEY & CO. LLC",1669,92.93 %
"UNIVERSITY OF MARYLAND COLLEGE PARK",1354,92.61 %
"INTEL CORPORATION",11415,92.56 %
"BATTELLE",1052,92.49 %
"COGNIZANT TECHNOLOGY SOLUTIONS U.S. CORPORATION",17528,92.49 %
"INTUIT INC.",1404,92.45 %
"PEOPLE TECH GROUP INC.",1124,92.26 %
"ITC INFOTECH (USA), INC.",1859,91.93 %
"MEDTRONIC, INC.",1050,91.90 %
"BALTIMORE CITY PUBLIC SCHOOLS",1014,91.52 %
"SIRI INFOSOLUTIONS INC.",1039,91.34 %
"THE PENNSYLVANIA STATE UNIVERSITY",1042,91.17 %
"ADVANCED MICRO DEVICES, INC.",1512,91.14 %
"HSBC BANK USA, N.A.",1110,90.99 %
"L&T TECHNOLOGY SERVICES LIMITED",3722,90.57 %
"MARVELL SEMICONDUCTOR, INC.",1631,89.94 %
"CUMMINS INC.",4737,89.82 %
"IBM INDIA PVT. LTD.",1284,89.56 %
"VIRTUSA CORPORATION",2217,88.77 %
"INDIANA UNIV. PURDUE UNIV. INDIANAPOLIS",1007,88.58 %
"THE UNIVERSITY OF ARIZONA",1037,88.33 %
"IBM CORPORATION",13276,88.20 %
"BROADCOM CORPORATION",2862,88.12 %
"IBM INDIA PRIVATE LIMITED",34219,87.69 %
"PERSISTENT SYSTEMS, INC.",3225,87.07 %
"AMERICAN INFORMATION TECHNOLOGY CORPORATION",1358,86.67 %
"CITRIX SYSTEMS, INC.",1044,85.34 %
"NETAPP, INC.",1870,84.55 %

10) Which are the top 10 job positions which have the highest success rate in petitions?

Launched map tasks=5
Launched reduce tasks=1
Data-local map tasks=5

PRODUCTION SUPPORT LEAD - US@1301@100.00
ASSOCIATE CONSULTANT - US@4393@99.93
SYSTEMS ENGINEER - US@10036@99.90
TEST ENGINEER - US@2198@99.86
PRODUCTION SUPPORT ANALYST - US@1451@99.86
TEST ANALYST - US@4958@99.82
CONSULTANT - US@7426@99.81
TECHNOLOGY LEAD - US@28350@99.80
TECHNICAL TEST LEAD - US@5374@99.80
SENIOR TECHNOLOGY ARCHITECT - US@1417@99.79
TECHNOLOGY ARCHITECT - US@4707@99.77
TECHNOLOGY ANALYST - US@26055@99.76
SENIOR PROJECT MANAGER - US@2774@99.75
DEVELOPER USER INTERFACE@5247@99.71
COMPUTER SYSTEMS ANALYST 2@4031@99.70
SYSTEMS ANALYST - II@1339@99.70
PROJECT MANAGER - III@1651@99.70
PROJECT MANAGER - US@7046@99.69
PROGRAMMER ANALYST - II@3588@99.67
LEAD CONSULTANT - US@3402@99.65
COMPUTER SYSTEMS ANALYST 3@2170@99.59
COMPUTER PROGRAMMER/CONFIGURER 2@6729@99.57
PROGRAMMER ANALYST - I@1432@99.51
SYSTEMS ANALYST - III@1006@99.50
PRINCIPAL CONSULTANT - US@1352@99.48
COMPUTER SPECIALIST/TESTING AND QUALITY ANALYST 2@3998@99.42
COMPUTER PROGRAMMER/CONFIGURER 3@1145@99.39
COMPUTER SPECIALIST/SYSTEM SUPPORT AND DEVELOPMENT@1339@99.33
COMPUTER SPECIALIST/SYSTEM SUPPORT AND DEVELOPMENT ADMIN 2@1085@99.26
DATA WAREHOUSE SPECIALIST@1631@99.20
SPECIALIST MASTER@1119@99.20
ASSURANCE STAFF@2334@99.06
COMPUTER SYSTEMS ENGINEER/ARCHITECT@2067@98.79
SOFTWARE QUALITY ASSURANCE ENGINEER AND TESTER@1568@98.66
ADVISORY SENIOR@5416@98.65
AUDIT SENIOR@1070@98.60
TEST CONSULTANT@1454@98.56
SOFTWARE ENGINEER AND TESTER@1216@98.52
ARCHITECT LEVEL 2@2892@98.51
PROGRAMMER/DEVELOPER@1560@98.46
TEST ENGINEER LEVEL 2@2372@98.44
MODULE LEAD@2226@98.34
ADVISORY MANAGER@3255@98.31
AUDIT ASSISTANT@1205@98.26
LEAD ENGINEER@11157@98.23
COMPUTER SPECIALIST@2175@98.21
SPECIALIST SENIOR@1447@98.20
CONSULTANT LEVEL 3@1171@98.12
ADVISORY STAFF@2413@98.01
DEVELOPER@12909@98.01
ERS SENIOR CONSULTANT@2249@97.95
TAX SENIOR@1838@97.93

TEST ENGINEER LEVEL 1@1036@97.88
PROGRAMMER ANALYST LEVEL 1@2395@97.87
BUSINESS TECHNOLOGY ANALYST@2005@97.86
TECHNICAL ANALYST@2932@97.78
SOFTWARE DEVELOPMENT ENGINEER IN TEST@4258@97.65
ADVISORY SENIOR ASSOCIATE@1332@97.52
SOFTWARE ENGINEER 2@4166@97.17
ERS CONSULTANT@2170@97.14
FUNCTIONAL CONSULTANT@1115@97.04
QA TESTER@1170@96.92
SOFTWARE DEVELOPMENT ENGINEER@7284@96.86
PROGRAMMER ANALYSTS@1133@96.82
COMPUTER SYSTEMS ENGINEER@11090@96.70
BUSINESS SYSTEM ANALYST@4436@96.69
LEAD CONSULTANT@2169@96.63
ELEMENTARY BILINGUAL TEACHER@2088@96.60
ASSISTANT VICE PRESIDENT@2132@96.58
CONSULTANT@23081@96.57
PRINCIPAL CONSULTANT@1836@96.57
SYSTEMS ANALYSTS@1252@96.57
SYSTEM ADMINISTRATOR@5048@96.55
SOLUTIONS ARCHITECT@1915@96.50
APPLICATIONS CONSULTANT@1180@96.44
ORACLE DATABASE ADMINISTRATOR@1527@96.40
SOFTWARE DESIGN ENGINEER@1080@96.39
PROGRAMMER/ANALYST@9375@96.37
ASSOCIATE RESEARCH SCIENTIST@1400@96.36
RESEARCH FELLOW@5981@96.36
APPLICATIONS DEVELOPER@3366@96.35
SENIOR SOFTWARE DEVELOPER@10208@96.32
PROGRAMMER ANALYST@249038@96.13
SENIOR PROGRAMMER ANALYST@5810@96.13
ASSISTANT RESEARCH SCIENTIST@1103@96.10
SENIOR ASSOCIATE@3540@96.02
SOFTWARE QUALITY ASSURANCE ENGINEER@4920@95.98
SENIOR MANAGER@1439@95.97
ASSURANCE SENIOR@1607@95.96
PHYSICIAN IN A POST GRADUATE TRAINING PROGRAM@2421@95.95
SYSTEMS ANALYST@61965@95.95
QUALITY ASSURANCE ANALYST@7325@95.95
TECHNICAL ARCHITECT@2908@95.94
PROJECT LEAD@2363@95.94
SOFTWARE ENGINEER III@1328@95.93
SOFTWARE ANALYST@1072@95.90
SR. SYSTEMS ANALYST@1151@95.83
SOFTWARE ENGINEER 3@1891@95.82
LEAD DEVELOPER@1049@95.81
QUALITY ANALYST@2616@95.80
SENIOR SOFTWARE DEVELOPMENT ENGINEER@1399@95.78
SOFTWARE QA ANALYST@1112@95.77
TEST ANALYST@1419@95.77
SENIOR CONSULTANT@24904@95.76
TECHNICAL SPECIALIST@1295@95.75

SENIOR TECHNICAL CONSULTANT@1882@95.75
SR. PROGRAMMER ANALYST@3716@95.75
DATA SCIENTIST@1932@95.70
SAP CONSULTANT@3023@95.70
COMPUTER PROGRAMMER@70570@95.63
PROGRAM MANAGER@3920@95.59
BUSINESS SYSTEMS ANALYST@10110@95.57
PROGRAMMER / ANALYST@1173@95.57
MEMBER OF TECHNICAL STAFF@1774@95.55
ETL DEVELOPER@1841@95.55
SOFTWARE PROGRAMMER@3577@95.53
PRINCIPAL SOFTWARE ENGINEER@2257@95.53
POSTDOCTORAL SCHOLAR@3186@95.48
SOFTWARE ENGINEER & TESTER@1538@95.45
MANAGER@8561@95.42
JAVA DEVELOPER@7596@95.39
SOFTWARE DEVELOPMENT ENGINEER II@3274@95.36
POSTDOCTORAL RESEARCH ASSOCIATE@6039@95.31
SCIENCE TEACHER@1127@95.30
POSTDOCTORAL ASSOCIATE@5145@95.30
SENIOR DATABASE ADMINISTRATOR@1229@95.28
SYSTEMS ANALYST II@1036@95.27
BUSINESS ANALYST@39681@95.20
ANALYST@11751@95.19
QA ANALYST@6871@95.18
DATABASE ADMINISTRATOR@16665@95.18
ASSOCIATE@12502@95.18
SENIOR BUSINESS ANALYST@3402@95.15
ASSOCIATE SOFTWARE ENGINEER@1215@95.14
SENIOR SYSTEMS ANALYST@5353@95.14
SOFTWARE ENGINEER@121307@95.12
SOFTWARE APPLICATION ENGINEER@1126@95.12
COMPUTER PROGRAMMER/ANALYST@1122@95.10
SYSTEMS ENGINEER@8078@95.10
COMPUTER PROGRAMMER ANALYST@13634@95.07
DATABASE ANALYST@1050@95.05
SOFTWARE QA ENGINEER@1169@95.04
.NET DEVELOPER@2921@95.04
COMPUTER SYSTEMS ANALYSTS@4728@95.03
WEB DEVELOPER@8024@95.01
ARCHITECT@4982@95.00
HARDWARE ENGINEER@2556@94.99
RESEARCH ASSISTANT PROFESSOR@1973@94.98
QUALITY ASSURANCE ENGINEER@3647@94.98
COMPUTER SOFTWARE ENGINEER, APPLICATIONS@4426@94.92
VALIDATION ENGINEER@1159@94.91
DATABASE DEVELOPER@1155@94.89
COMPUTER SYSTEMS ANALYST@35086@94.86
POSTDOCTORAL FELLOW@7857@94.86
SENIOR HARDWARE ENGINEER@1653@94.86
QA ENGINEER@2224@94.83
ENGINEER II@1249@94.72
SR. SOFTWARE ENGINEER@4863@94.72

STAFF SCIENTIST@1242@94.69
BUSINESS INTELLIGENCE ANALYST@1972@94.68
PRINCIPAL ENGINEER@1066@94.65
RF ENGINEER@2794@94.63
SYSTEMS ADMINISTRATOR@6659@94.59
SENIOR SOFTWARE ENGINEER@27133@94.56
SENIOR JAVA DEVELOPER@1395@94.55
SOFTWARE TEST ENGINEER@3591@94.54
VISITING ASSISTANT PROFESSOR@1311@94.51
TECHNICAL CONSULTANT@3420@94.50
PROGRAMMER/ ANALYST@1000@94.50
SOFTWARE DEVELOPER@42907@94.50
STAFF ENGINEER@1869@94.49
NETWORK ENGINEER@5422@94.47
PROJECT MANAGER@20172@94.44
IT PROJECT MANAGER@2473@94.42
APPLICATION ENGINEER@1458@94.38
SOFTWARE ARCHITECT@1878@94.36
PROGRAMMER@6011@94.33
VICE PRESIDENT@3159@94.30
CLINICAL ASSISTANT PROFESSOR@1281@94.30
ASSISTANT PROFESSOR@25265@94.30
ASSOCIATE CONSULTANT@1350@94.30
SENIOR SYSTEMS ENGINEER@2030@94.29
TECHNICAL RECRUITER@1364@94.28
RESEARCHER@1031@94.28
SOFTWARE DEVELOPER, APPLICATIONS@1830@94.26
STAFF SOFTWARE ENGINEER@2976@94.25
TEST ENGINEER@3936@94.23
RESEARCH ASSOCIATE@13623@94.20
SENIOR PRODUCT MANAGER@1085@94.19
DATA ANALYST@3805@94.17
TECHNICAL SUPPORT ENGINEER@1230@94.15
PROGRAMMER ANALYST II@1059@94.15
SYSTEM ANALYST@4684@94.13
POSTDOCTORAL RESEARCHER@2130@94.08
LECTURER@2257@94.06
SR. SOFTWARE DEVELOPER@1161@94.06
SOFTWARE ENGINEER II@2051@94.05
PRODUCT MANAGER@3367@94.03
IT CONSULTANT@3497@93.99
QUALITY ENGINEER@2381@93.99
SENIOR RESEARCH ASSOCIATE@1015@93.99
SPEECH LANGUAGE PATHOLOGIST@1381@93.99
SENIOR ANALYST@1646@93.99
SYSTEMS ANALYST JC65@3321@93.98
RESEARCH ENGINEER@1338@93.95
DESIGN ENGINEER@6325@93.93
POSTDOCTORAL RESEARCH FELLOW@2346@93.90
MANUFACTURING ENGINEER@1906@93.81
SENIOR DESIGN ENGINEER@1209@93.80
COMPUTER PROGRAMMERS@4963@93.77
TECHNICAL MANAGER@1060@93.77

COMPONENT DESIGN ENGINEER@2851@93.76
NETWORK ADMINISTRATOR@2624@93.71
SOFTWARE DEVELOPMENT ENGINEER I@2128@93.70
RESEARCH ASSISTANT@1777@93.70
QUANTITATIVE ANALYST@1293@93.66
TECHNICAL LEAD@5218@93.66
POST DOCTORAL FELLOW@1507@93.56
ASSOCIATE PROFESSOR@1441@93.55
RESEARCH SCIENTIST@5154@93.46
SENIOR DEVELOPER@2994@93.45
PHARMACIST@5864@93.40
INSTRUCTOR@3014@93.40
SENIOR SYSTEMS ANALYST JC60@3069@93.39
SENIOR ENGINEER@3773@93.37
PRODUCT ENGINEER@2634@93.24
ENGINEER@4941@93.06
NETWORK AND COMPUTER SYSTEMS ADMINISTRATOR@1928@93.05
LEAD SOFTWARE ENGINEER@1572@93.00
SOFTWARE DEVELOPERS, APPLICATIONS@1195@92.97
COMPUTER SYSTEM ANALYST@3753@92.75
OCCUPATIONAL THERAPIST@4437@92.72
SENIOR APPLICATION DEVELOPER@1965@92.62
SENIOR PROJECT MANAGER@1015@92.61
CLINICAL FELLOW@1146@92.58
INDUSTRIAL DESIGNER@3619@92.57
PROJECT ENGINEER@6439@92.56
PHYSICAL THERAPIST@20207@92.54
APPLICATIONS ENGINEER@1688@92.54
SOLUTION ARCHITECT@1994@92.53
HOSPITALIST@4387@92.50
TEST LEAD@1726@92.41
DIRECTOR@1333@92.35
HOSPITALIST PHYSICIAN@4067@92.33
SENIOR FINANCIAL ANALYST@1196@92.22
SYSTEM ENGINEER@2145@92.21
MANAGER JC50@1874@91.89
MEDICAL RESIDENT@2336@91.78
PSYCHIATRIST@1289@91.54
COMPUTER SOFTWARE ENGINEER@2684@91.51
SCIENTIST@1340@91.49
RESIDENT@1245@91.49
APPLICATION DEVELOPER@7692@91.46
TECHNICAL PROJECT MANAGER@1052@91.25
PROCESS ENGINEER@4377@91.23
PHYSICIAN@4417@91.06
ENGINEERING MANAGER@1199@90.91
MECHANICAL ENGINEER@7301@90.84
RESEARCH ANALYST@1869@90.74
SPECIAL EDUCATION TEACHER@1721@90.47
MANAGING CONSULTANT@3873@90.34
TEST SPECIALIST@1011@90.31
STRUCTURAL ENGINEER@1094@90.22
ACCOUNT MANAGER@1066@90.15

SENIOR SYSTEM ENGINEER@1408@90.13
APPLICATION PROGRAMMER@1686@90.09
NEPHROLOGIST@1027@89.97
ARCHITECTURAL DESIGNER@2334@89.76
MEDICAL TECHNOLOGIST@3125@89.57
RESIDENT PHYSICIAN@2119@89.48
SALES ENGINEER@2167@89.39
PEDIATRICIAN@1214@89.37
DESIGNER@1992@89.31
INDUSTRIAL ENGINEER@2093@89.20
OPERATIONS RESEARCH ANALYST@1946@89.16
IT SPECIALIST@2585@89.13
ASSOCIATE ATTORNEY@1533@89.04
ELECTRONICS ENGINEER@1060@88.87
CHEMIST@1380@88.84
MARKETING ANALYST@1573@88.68
ELECTRICAL ENGINEER@4174@88.43
MANAGEMENT ANALYST@5386@88.04
BUSINESS DEVELOPMENT MANAGER@2345@88.02
BUSINESS DEVELOPMENT ANALYST@1148@87.98
ELEMENTARY SCHOOL TEACHER@1304@87.35
STAFF ACCOUNTANT@4491@87.26
FINANCIAL ANALYST@8515@87.15
BUDGET ANALYST@1687@87.14
BUSINESS DEVELOPMENT SPECIALIST@1482@87.11
SYSTEM ANALYST JC65@1419@86.82
DENTIST@3250@86.74
TEACHER@3576@86.72
MARKETING SPECIALIST@2150@86.56
BUSINESS OPERATIONS SPECIALIST@1034@85.98
ATTORNEY@1050@85.90
INTERIOR DESIGNER@1361@85.45
CIVIL ENGINEER@2257@84.98
MARKET RESEARCH ANALYST@8934@84.45
LAW CLERK@1709@83.91
SALES MANAGER@1233@83.62
OPERATIONS MANAGER@1785@83.47
ACCOUNTANT@14048@83.47
GRAPHIC DESIGNER@5020@83.15
PUBLIC RELATIONS SPECIALIST@1931@82.70
FINANCIAL MANAGER@1080@82.41
MARKETING MANAGER@2230@80.81
CHIEF EXECUTIVE OFFICER@1095@80.64
GENERAL MANAGER@1348@78.49

MENU:

```
#!/bin/bash
show_menu()
{
   NORMAL=`echo "\033[m"`
   MENU=`echo "\033[36m"` #Blue
   NUMBER=`echo "\033[33m"` #yellow
   FGRED=`echo "\033[41m"`
   RED_TEXT=`echo "\033[31m"`
   ENTER_LINE=`echo "\033[33m"`
   echo -e "${MENU}**********H1B
APPLICATIONS**********${NORMAL}"
   echo -e "${MENU}**${NUMBER} 1) ${MENU} Is the number of petitions with Data Engineer
job title increasing over time?${NORMAL}"
   echo -e "${MENU}**${NUMBER} 2) ${MENU} Find top 5 job titles who are having highest
growth in applications. ${NORMAL}"
   echo -e "${MENU}**${NUMBER} 3) ${MENU} Which part of the US has the most Data
Engineer jobs for each year? ${NORMAL}"
   echo -e "${MENU}**${NUMBER} 4) ${MENU} find top 5 locations in the US who have got
certified visa for each year.${NORMAL}"
   echo -e "${MENU}**${NUMBER} 5) ${MENU} Which industry has the most number of Data
Scientist positions?${NORMAL}"
   echo -e "${MENU}**${NUMBER} 6) ${MENU} Which top 5 employers file the most petitions
each year? ${NORMAL}"
   echo -e "${MENU}**${NUMBER} 7) ${MENU} Find the most popular top 10 job positions for
H1B visa applications for each year?${NORMAL}"
   echo -e "${MENU}**${NUMBER} 8) ${MENU} Find the percentage and the count of each
case status on total applications for each year. Create a graph depicting the pattern of All the cases
over the period of time.${NORMAL}"
   echo -e "${MENU}**${NUMBER} 9) ${MENU} Create a bar graph to depict the number of
applications for each year${NORMAL}"
   echo -e "${MENU}**${NUMBER} 10) ${MENU}Find the average Prevailing Wage for each
Job for each Year (take part time and full time separate) arrange output in descending
order${NORMAL}"
   echo -e "${MENU}**${NUMBER} 11) ${MENU} Which are employers who have the highest
success rate in petitions more than 70% in petitions and total petions filed more than
1000?${NORMAL}"
   echo -e "${MENU}**${NUMBER} 12) ${MENU} Which are the top 10 job positions which
have the  success rate more than 70% in petitions and total petitions filed more than 1000?
${NORMAL}"
   echo -e "${MENU}**${NUMBER} 13) ${MENU}Export result for option no 12 to MySQL
database.${NORMAL}"
   echo -e "${MENU}**********************************************${NORMAL}"
   echo -e "${ENTER_LINE}Please enter a menu option and enter or ${RED_TEXT}enter to exit.
${NORMAL}"
   read opt
}
function option_picked()
{
   COLOR='\033[01;31m' # bold red
   RESET='\033[00;00m' # normal white
```

```
   MESSAGE="$1"  #modified to post the correct option selected
   echo -e "${COLOR}${MESSAGE}${RESET}"
}
clear
show_menu
        while [ opt != '' ]
   do
   if [[ $opt = "" ]]; then
         exit;
   else
      case $opt in
      1) clear;
         stop-all.sh
      option_picked "1 a) Is the number of petitions with Data Engineer job title increasing over
time?";
                  pig -x local /home/hduser/project/project1a.pig
      show_menu;
      ;;
      2) clear;
         stop-all.sh
      option_picked "1 b) Find top 5 job titles who are having highest growth in applications. ";

                  pig -x local /home/hduser/project/project1b.pig
      show_menu;
      ;;
      3) clear;
      option_picked "2 a) Which part of the US has the most Data Engineer jobs for each year?";
                  start-all.sh
                  hive -f /home/hduser/project/project2a.sql

      show_menu;
      ;;
            4) clear;

      option_picked "2 b) find top 5 locations in the US who have got certified visa for each year.";

                  start-all.sh
                  hive -f /home/hduser/project/project2b.sql

      show_menu;
      ;;
            5) clear;
      option_picked "3) Which industry has the most number of Data Scientist positions?";
      start-all.sh
                  hadoop fs -rm -r -f /projectoutput/q3
            hadoop jar /home/hduser/project/project3.jar h1bproject.MaxDataScientist /project
/projectoutput/q3
                  hadoop fs -cat /projectoutput/q3/p*
      show_menu;
      ;;
      6) clear;
         stop-all.sh
```

```
option_picked "4)Which top 5 employers file the most petitions each year?";

        pig -x local /home/hduser/project/project4.pig
show_menu;
;;
7) clear;
   stop-all.sh
option_picked "5) Find the most popular top 10 job positions for H1B visa applications for
each year?";
   start-all.sh
        hive -f /home/hduser/project/project5.sql
   show_menu;
;;
8) clear;
   stop-all.sh
        option_picked "6) 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.";

        pig -x local /home/hduser/project/project6.pig
        cat /home/hduser/project/projectoutputs/project6

show_menu;
;;
        9) clear;
option_picked "7) Create a bar graph to depict the number of applications for each year";
        start-all.sh
        hadoop fs -rm -r -f /projectoutput/q7
        hadoop jar /home/hduser/project/project7.jar h1bproject.ApplicationsPerYear /project
/projectoutput/q7
        hadoop fs -cat /projectoutput/q7/p*
show_menu;
;;
        10) clear;
option_picked "8) Find the average Prevailing Wage for each Job for each Year (take part time
and full time separate) arrange output in descending order";

        echo -e "Enter the choice Full time/ Part time.(Y/N)"
        read var
hive -e "select job_title,TJob,year from(select rank() over(partition by year order by TJob
desc)as rank_1,TJob,job_title,year from(select AVG(prevailing_wage)as TJob,job_title,year from
h1b_final where full_time_position ='$var' and prevailing_wage is not null group by
job_title,year)a)b ;
"
show_menu;
;;
        11) clear;

        option_picked "9) Which are   employers who have the highest success rate in
petitions more than 70% in petitions and total petions filed more than 1000?"
   start-all.sh
        hadoop fs -rm -r -f /projectoutput/q9
```

```
                hadoop jar /home/hduser/project/project9.jar h1b9.H1B9 /project /projectoutput/q9
                hadoop fs -cat /projectoutput/q9/p*
        show_menu;
        ;;
                12) clear;

                option_picked "10) Which are the top 10 job positions which have the  success rate
more than 70% in petitions and total petitions filed more than 1000?"
          start-all.sh
                hadoop fs -rm -r -f /projectoutput/q10
                hadoop jar /home/hduser/project/project10.jar h1b10.H1B10 /project
/projectoutput/q10
                hadoop fs -cat /projectoutput/q10/p*
        show_menu;
        ;;
                13) clear;
                option_picked "11) Export result for question no 10 to MySql database."
                start-all.sh

                bash /home/hduser/project/sqoop.sh
        show_menu;
        ;;
                \n) exit;
                ;;
        *) clear;
        option_picked "Pick an option from the menu";
        show_menu;
        ;;
    esac
fi
done
```