GitHub URL:

https://github.com/illinoistech-itm/nshah90

Run the acite75_99.txt file with PatentCitation.java file.

The output will be the counts of cited Patents.

Then convert the output as a comma separated output with the following command.

➔ sed "s/\s/,/g" citedOutput.txt > citedOutput_comma.txt

Run the citedOutput_comma.txt with PatentCitation.java file

The Output will give us the count of count.

Created a hive environment with help of following commands.

- Download the binary file "apache-hive-2.1.0-bin.tar" and untar it then moves it to "/usr/lib/hive".
- export HIVE_HOME and path in .bashrc file.

Run this command to setup schema:

➔ schematool -initSchema -dbType derby

Create a database in the hive environment.

Create a table in the hive shell using following command:

create table PatentCitedTable (key string, value int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' stored as textfile;

1st column key- this has all the values of patent citations

2nd column value – this has all the values of no of counts.

Load the data into the table with the help of following command:

load data inpath '/user/vagrant/data/citedOutput_comma.txt' into table PatentCitedTable;

To find the Top 10 Number of Citations we will run the following query which will retrieve top 10 citation in descending order.

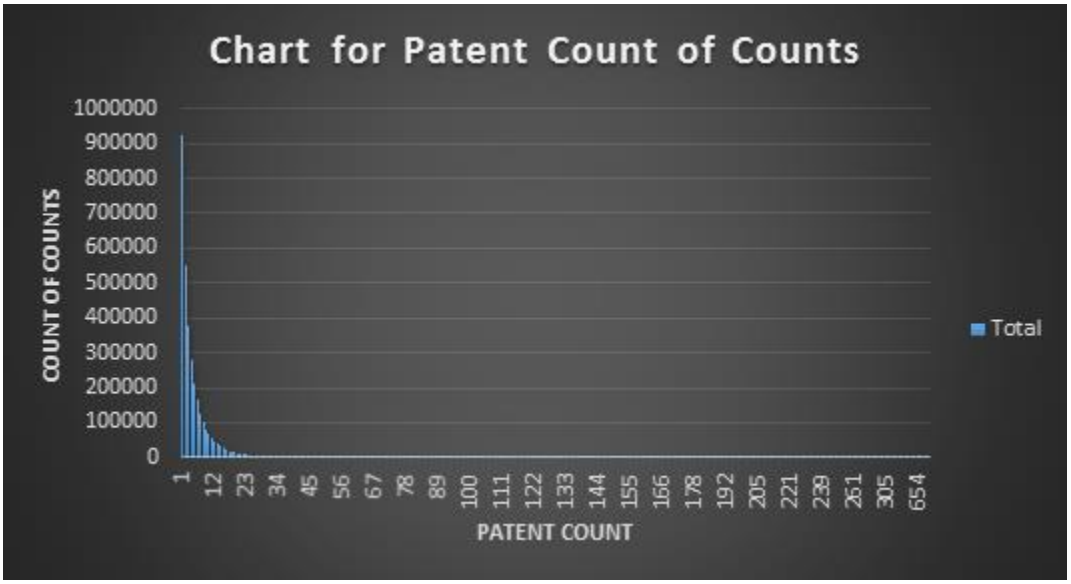
➔ Select * from PatentCitedTable order by value DESC limit 10;

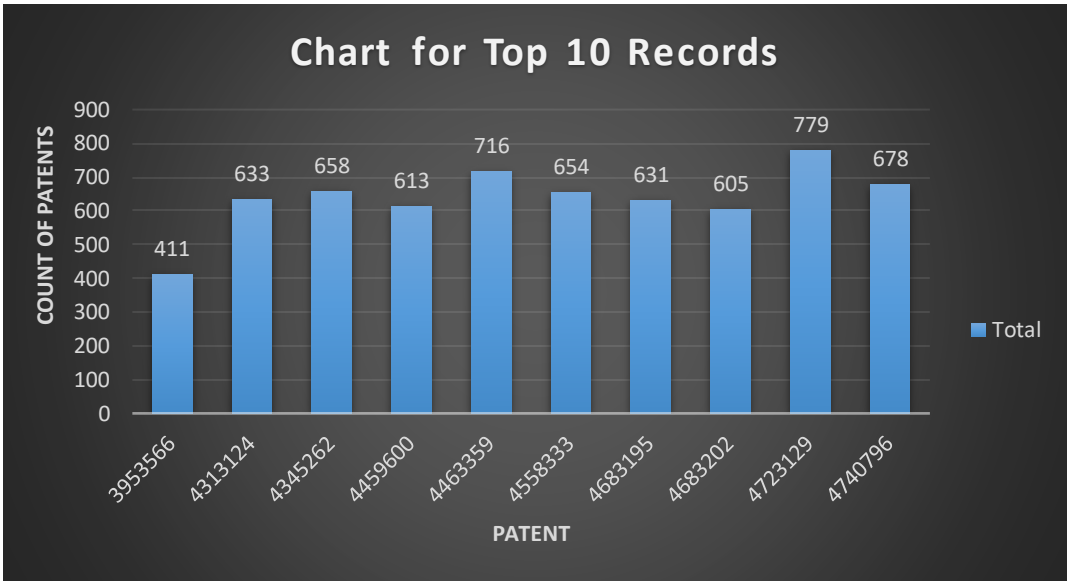
To find the Top 1000 Number of Citations we will run the following query which will retrieve top 1000 citation in descending order.

➔ Select * from PatentCitedTable order by value DESC limit 1000;

Histogram Analysis for all the graphs are as follows:

Charts:





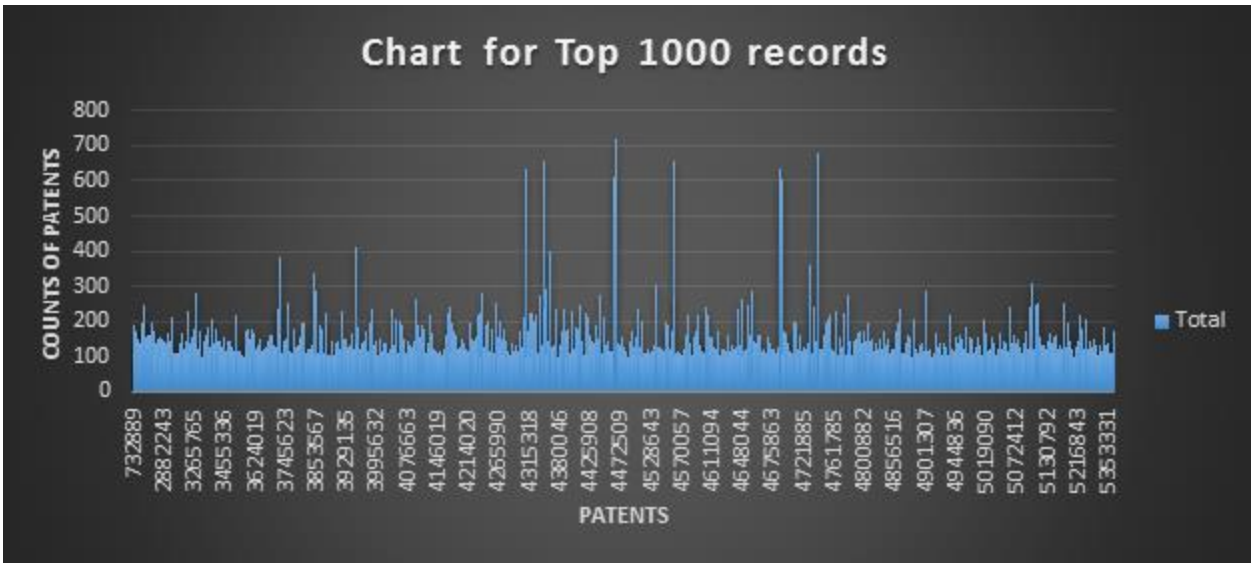


Chart for Top 1000 records
COUNTS OF PATENTS
800
700
600
500
400
300
200
100
0
732889
882243
3265765
3455336
3624019
3745623
3853567
3929135
3995632
4076663
4146019
4214020
4265990
4315318
4380046
4425908
4472509
4528643
4570057
4611094
4648044
4675863
4721885
4761785
4800882
4856516
4901307
4944835
5019090
5072412
5130792
5216843
5353331
PATENTS
Total