

ANALYSIS CENSUS DATA

Project Author: Neha Kunwar Solanki

Objective:

In this project we are working on census data to provide guidance and presentation of descriptive data to an organization and policy-makers. And this information makes it possible to plan better services, improve the quality of life and solve existing problems. Census analysis information, which serves as the basis for constructing planning forecasts, procedure of systematically acquiring and recording information about the members of a given population.

Scope:

- I. This census data analysis project provide data collection and production of information for government ministries and local authorities for budgeting purposes.
- II. This production of information which serves bodies, organizations and various other elements in the fields of education, the economy, business, research, etc.

Examples:

- 1. **Government Ministries**
- 2. **Local authorities**
- 3. **Bodies of research**
- 4. **Private and public companies**
- 5. **Journalists**
- 6. **The general public**

Sample Data Set for Analysis: (Census_Records.json)

Age	Education	Marital Status	Gender	Tax filler status	Income	parents	Citizenship	Week Worked

Project Description -

- i) We are provided with certain use-cases to get the required data. For all the use-cases we will be using a Map-reduce approach. The Map Reduce Approach totally works on Key-Value pair as Input and Output. There will be a Driver Class, Mapper Class and a Reducer Class.
- ii) We have used pig Latin language to get required output because Pig tends to create a flow of data: small steps where in each you do some processing. Hive gives you SQL-like language to operate on your data, so transformation from RDBMS is much easier.
- iii) Every use case implemented the Hive Query Language (HiveQL or HQL) for Map Reduce to process structured data using Hive. HiveQL is similar to SQL for querying on schema info on the Metastore. It is one of the replacements of traditional approach for Map Reduce program. Instead of writing Map Reduce program in Java, we can write a query for Map Reduce job and process it.

TECHNOLOGY USED:

1. Apache Hadoop
2. Map-Reduce Programming in java.
3. PIG
4. HIVE
5. Sqoop

SOFTWARE USED:

- 1 Eclipse IDE
- 2 Oracle Virtual Machine
- 3 Cloudera
- 4 JDK 1.7

Secondary Tables:

We have created two or more secondary tables by splitting census record table into sub tables for performing join operation between multiple tables. This functionality can be achieved with the help of sqoop ecosystem tool of hadoop

1. Tax analysis:

In this table we have taken fields like minimum amount, maximum amount, gender, percentage.

2. Pension:

In this table we have taken fields like minimum amount, maximum amount, and pension.

3. Scholarship:

In this table we have taken fields like age, gender, amount, parents for storing scholarship amount according to their age criteria

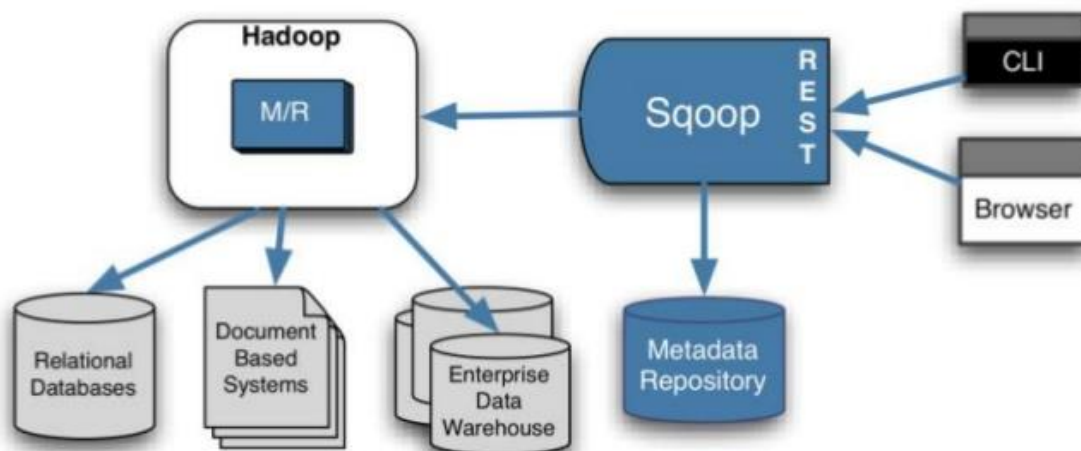
Sqoop:

- I. Apache Sqoop is a tool designed for efficiently transferring bulk data between Apache Hadoop and external data stores such as relational databases, enterprise data warehouses.
- II. Sqoop is used to import data from external data stores into Hadoop Distributed File System or related Hadoop eco-systems like Hive and HBase. Similarly, Sqoop can also be used to extract data from Hadoop or its eco-systems and export it to external data stores such as relational databases, enterprise data warehouses. Sqoop works with relational databases such as Teradata, Netezza, Oracle, MySQL, Postgres etc.

Where is Sqoop used in Project?

- I. Relational database systems are widely used to interact with the traditional business applications. So, relational database systems has become one of the sources that generate Big Data.
- II. As we are dealing with Big Data, Hadoop stores and processes the Big Data using different processing frameworks like Map Reduce, Hive, HBase, Cassandra, Pig etc and storage frameworks like HDFS to achieve benefit of distributed computing and distributed storage. In order to store and analyze the Big Data from relational databases, Data need to be transferred between database systems and Hadoop Distributed File System (HDFS). Here, Sqoop comes into picture.
- III. Sqoop acts like a intermediate layer between Hadoop and relational database systems. You can import data and export data between relational database systems and Hadoop and its eco-systems directly using sqoop.

Sqoop 2 Architecture (proposed)



Use Cases: Scenario 1

Objective:

- i) To generate employment opportunities in rural as well as urban areas of the country through setting up of new self-employment ventures / projects / micro enterprises.
- ii) To bring together widely dispersed traditional artisans /rural and urban unemployed youth and give them self-employment opportunities to the extent possible, at their place.

1. Education Statistics Based On Administrative Data:

In this use case, we are fetching records of all the male and female from census data. For example at the state level, some states have consistently even rates across the public and private educational sector.

For example, Rajasthan is the only state with more males than females in its public colleges, and it is very close to a 50-50 split in its private colleges. On the other hand, Rajsamand district typically has among the lowest percentage of males across all types of colleges.

Task1: Total count of male/female based on education

Output: Hive

```
hive-1274 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 610 msec
OK
10th grade      Female 12187
10th grade      Male 10384
11th grade      Female 10815
11th grade      Male 9690
12th grade no diploma  Female 2970
12th grade no diploma  Male 3304
1st 2nd 3rd or 4th grade  Female 2764
1st 2nd 3rd or 4th grade  Male 2591
5th or 6th grade      Female 4992
5th or 6th grade      Male 4761
7th and 8th grade      Female 12609
7th and 8th grade      Male 11518
9th grade          Female 9780
9th grade          Male 8755
Associates degree-academic program      Female 7684
Associates degree-academic program      Male 5266
Associates degree-occup /vocational      Female 9225
Associates degree-occup /vocational      Male 6733
Bachelors degree(BA AB BS)      Female 29557
Bachelors degree(BA AB BS)      Male 29680
Children          Female 69827
Children          Male 71669
Doctorate degree(PhD EdD)      Female 1099
Doctorate degree(PhD EdD)      Male 2714
High school graduate      Female 80977
High school graduate      Male 63857
Less than 1st grade      Female 1279
Less than 1st grade      Male 1133
Masters degree(MA MS MEng MEd MSW MBA)      Female 9493
Masters degree(MA MS MEng MEd MSW MBA)      Male 10150
Prof school degree (MD DDS DVM LLB JD)      Female 1530
Prof school degree (MD DDS DVM LLB JD)      Male 3828
Some college but no degree      Female 45012
Some college but no degree      Male 38690
Time taken: 156.749 seconds
hive>
```

Output: Pig

```
(( Children, Male),71669)
(( Children, Female),69827)
(( 9th grade, Male),8755)
(( 9th grade, Female),9780)
(( 10th grade, Male),10384)
(( 10th grade, Female),12187)
(( 11th grade, Male),9690)
(( 11th grade, Female),10815)
(( 5th or 6th grade, Male),4761)
(( 5th or 6th grade, Female),4992)
(( 7th and 8th grade, Male),11518)
(( 7th and 8th grade, Female),12609)
(( Less than 1st grade, Male),1133)
(( Less than 1st grade, Female),1279)
(( High school graduate, Male),63857)
(( High school graduate, Female),80977)
(( 12th grade no diploma, Male),3304)
(( 12th grade no diploma, Female),2970)
(( 1st 2nd 3rd or 4th grade, Male),2591)
(( 1st 2nd 3rd or 4th grade, Female),2764)
(( Doctorate degree(PhD EdD), Male),2714)
(( Doctorate degree(PhD EdD), Female),1099)
(( Bachelors degree(BA AB BS), Male),29680)
(( Bachelors degree(BA AB BS), Female),29557)
(( Some college but no degree, Male),38690)
(( Some college but no degree, Female),45012)
(( Associates degree-academic program, Male),5266)
(( Associates degree-academic program, Female),7684)
(( Associates degree-occup /vocational, Male),6733)
(( Associates degree-occup /vocational, Female),9225)
(( Masters degree(MA MS MEng MEd MSW MBA), Male),10150)
(( Masters degree(MA MS MEng MEd MSW MBA), Female),9493)
(( Prof school degree (MD DDS DVM LLB JD), Male),3828)
(( Prof school degree (MD DDS DVM LLB JD), Female),1520)
```

2. Sustainable Employability

In this use case, we are fetching all the records of employed and unemployed people from census data. For example The PMRY has been designed to provide employment to more than a million Person by setting up of 7 lakhs micro enterprises by the educated unemployed youth. It relates to the setting up of the self-employment ventures through industry, service and business routes. The scheme also seeks to associate reputed non-governmental organizations in implementation PMRY scheme especially in the selection, training of entrepreneurs and preparation of project profiles.

Task: Total count of employed/unemployed based on education.

Output: Hive

```
IK
10th grade      EMployed-->    12044.0 UnEMployed-->    10527.0
11th grade      EMployed-->    8798.0 UnEMployed-->    11707.0
12th grade no diploma EMployed-->    2681.0 UnEMployed-->    3593.0
1st 2nd 3rd or 4th grade      EMployed-->    3339.0 UnEMployed-->    2016.0
5th or 6th grade      EMployed-->    5511.0 UnEMployed-->    4242.0
7th and 8th grade      EMployed-->    17234.0 UnEMployed-->    6893.0
9th grade      EMployed-->    11430.0 UnEMployed-->    7105.0
Associates degree-academic program      EMployed-->    2094.0 UnEMployed-->    1
10856.0
Associates degree-occup /vocational      EMployed-->    2820.0 UnEMployed-->    1
1138.0
Bachelors degree(BA AB BS)      EMployed-->    9615.0 UnEMployed-->    49622.0
Children      EMployed-->    141496.0 UnEMployed-->    NULL
Doctorate degree(PhD EdD)      EMployed-->    530.0 UnEMployed-->    3283.0
High school graduate      EMployed-->    44342.0 UnEMployed-->    100492.0
Less than 1st grade      EMployed-->    1678.0 UnEMployed-->    734.0
Masters degree(MA MS MEng MEd MSW MBA)      EMployed-->    2937.0 UnEMployed-->    1
1706.0
Prof school degree (MD DDS DVM LLB JD)      EMployed-->    666.0 UnEMployed-->    4
14692.0
Some college but no degree      EMployed-->    19037.0 UnEMployed-->    64665.0
Time taken: 135.667 seconds
hive>
```

Output: Pig-Emloyed

```
2010-11-20 22:20:21,273 [main] INFO org.apache.pig.backend.hadoop
( 9th grade,7105)
( 10th grade,10527)
( 11th grade,11707)
( 5th or 6th grade,4242)
( 7th and 8th grade,6893)
( Less than 1st grade,734)
( High school graduate,100492)
( 12th grade no diploma,3593)
( 1st 2nd 3rd or 4th grade,2016)
( Doctorate degree(PhD EdD),3283)
( Bachelors degree(BA AB BS),49622)
( Some college but no degree,64665)
( Associates degree-academic program,10856)
( Associates degree-occup /vocational,13138)
( Masters degree(MA MS MEng MEd MSW MBA),16706)
( Prof school degree (MD DDS DVM LLB JD),4692)
[cloudera@localhost ~]$
```


Output: Pig-Unemployed:

```
( Children,141496)
( 9th grade,11430)
( 10th grade,12044)
( 11th grade,8798)
( 5th or 6th grade,5511)
( 7th and 8th grade,17234)
( Less than 1st grade,1678)
( High school graduate,44342)
( 12th grade no diploma,2681)
( 1st 2nd 3rd or 4th grade,3339)
( Doctorate degree(PhD EdD),530)
( Bachelors degree(BA AB BS),9615)
( Some college but no degree,19037)
( Associates degree-academic program,2094)
( Associates degree-occup /vocational,2820)
( Masters degree(MA MS MEng MEd MSW MBA),2937)
( Prof school degree (MD DDS DVM LLB JD),666)
[cloudera@localhost ~]$
```

3. Employability: The Ongoing Process of Finding the Right Job After Graduation

In this use case we are collecting data of people in a range for example A recruiting agency, also referred to as a staffing or temporary agency, searches for, interviews and screens applicants for job placement. The goal of a recruiting agency may be to fill temporary job openings, permanent vacancies or both.

Task 3.Total count for people in age range of 18-25 based on education.

Output: Hive:

```
OK
Education--> 10th grade      Total Count--> 2411
Education--> 11th grade      Total Count--> 5310
Education--> 12th grade no diploma Total Count--> 1824
Education--> 1st 2nd 3rd or 4th grade      Total Count--> 275
Education--> 5th or 6th grade      Total Count--> 871
Education--> 7th and 8th grade      Total Count--> 989
Education--> 9th grade      Total Count--> 1486
Education--> Associates degree-academic program      Total Count--> 1414
Education--> Associates degree-occup /vocational      Total Count--> 1558
Education--> Bachelors degree(BA AB BS)      Total Count--> 5714
Education--> Doctorate degree(PhD EdD)      Total Count--> 15
Education--> High school graduate      Total Count--> 18966
Education--> Less than 1st grade      Total Count--> 187
Education--> Masters degree(MA MS MEng MEd MSW MBA)      Total Count--> 358
Education--> Prof school degree (MD DDS DVM LLB JD)      Total Count--> 27
Education--> Some college but no degree      Total Count--> 20311
Time taken: 29.134 seconds
hive>
```

Use Cases: Scenario 2

Objective:

1. Income tax assessments for New Scheme:

Post Office Saving Account

It is quite easy to open this scheme and caters to the small investors. The account can be opened with only Rs 20. A minimum balance of Rs 50 is required to be maintained in such account. It is however required that the account should have at least one transaction in a period of three financial years. Interest rate of 4% is given in this scheme.

Task1: Tax analysis total and gender wise

Output: Hive

```
Female 1710.1663736369826  
Male 1772.7254616592884  
Time taken: 28.998 seconds  
hive> █
```

5. Per Capita Income at current prices and corresponding percentage change

Monthly Income Deposit

What a monthly income fixed deposit is basically a fixed deposit that can be created and earns an interest in much the same way a regular fixed deposit does with a few differences. The being that with a monthly income plan the investors either have access to the amount deposited or are paid the interest earned on a monthly basis. Some banks even let depositors withdraw the interest every month from regular fixed deposits at discounted interest rates. The monthly income plans are generally linked to a bank account and the interest earned is credited into that account.

Task: Per Capita Income (PCI) analysis consolidated, gender wise and category wise

Output: Hive: Category wise

```
age group--> Teenager      sum of income--> 1689.5446269570016
age group--> adult    sum of income--> 1813.7500828047719
age group--> elderly sum of income--> 1662.5739941670317
age group--> infants sum of income--> 1667.2678898605448
age group--> middle-aged sum of income--> 1737.4900611355397
age group--> senior citizen sum of income--> 1708.379683926455
Time taken: 66.15 seconds
hive> █
```

Output: Hive: Total PCI:

```
TotalPCI--> 1740.0260960934236
Time taken: 29.013 seconds
hive> █
```

Use Cases: Scenario 3

Objective:

Government is providing pension scheme and number of scholarships for students to pursue higher studies. Although Government of India, foreign countries and private institutions offer the international scholarships to academically outstanding students.

1 Amount dispensed on pension

Task: Total amount dispensed on pension in x year(s)

Output: Map Reduce

```
[cloudera@localhost Desktop]$ hadoop jar TotalPension.jar /user/cloudera/CensusData /user/cloudera/outsocials5
Pension in Year : Enter Year
2014
```

```
[cloudera@localhost Desktop]$ hadoop fs -cat /user/cloudera/outsocials5/part-r-00000  
16455420
```

2. Schemes for student scholarships, awards and loans:

The State Governments/Union Territory Administrations shall constitute a committee of the Departments implementing such scholarship schemes to ensure that the student from the minority community, who may also belong to children of those engaged in unclean occupation and OBC do not avail scholarship from other sources for the same purpose and avail only one source.

2 Housing scheme for widows, divorcees

- i) The Department of Minority Welfare is constructing 250 houses across the State for widows belonging to minority religions such as Muslim, Christian, Sikh, Buddha, and Parsi, and divorcee women who have not remarried.
- ii) The department proposes to give Rs.2 lakh a house as financial assistance, which need not be repaid.
- iii) Applicants should not have remarried within three years of submitting the application form and should have a minimum of three cents of litigation-free and liability-free land in their own name. Those who have received housing aid from the government or similar agencies earlier are not eligible to apply.

Task: For given age range employable female widowed and divorced count

Output: Map Reduce

```
[cloudera@localhost ~]$ hadoop jar FemaleDivorceWidow.jar /user/cloudera/CensusData /user/cloudera/results  
Enter Minimum Age  
34  
Enter Maximum Age  
45
```

```
Divorced      7134
Widowed 580
```

3. Use of vote counting for meta-analysis:

- i) In this use case we are fetching total vote count for literature research. As Vote counting is a simple but limited method for synthesizing evidence from multiple evaluations, which involves simply comparing the number of positive studies (studies showing benefit) with the number of negative studies (studies showing harm).
- ii) It does not take into account the quality of the studies, the size of the samples, or the size of the effect.

Task: Voter(s) count in x year(s)

Output: Hive

```
OK
Total_Voters_Count--> 437549
Time taken: 31.156 seconds
hive> █
```

10. Senior citizen Free Computer Classes:

Many local and state governments offer computer classes to senior citizens through libraries and community centers. These classes can range from essential skills to advanced skills, such as coding. Libraries in larger municipalities will often have more extensive programming.

Community colleges offer a variety of computer classes, from beginning programming to learning how to use software products such as Word and Excel. Learning annexes -- educational facilities often attached to high schools, community colleges or libraries -- also offer computer classes.

Task 2. Senior Citizen(s) count in x year(s)

Output: Hive

```
OK
Total_Senior_Citizen in given year--> 100079
Time taken: 30.949 seconds
hive> █
```

3. Pradhan Mantri Jeevan Jyoti Bima Yojana (PMJJBY) planning

These schemes are designed to be pro poor and promise to provide protection against the risks of dying too early (Pradhan Mantri Jeevan Jyoti Bima Yojana) or living too long (Atal Pension Yojana) or unable to work & earn due to partial or full disability (Pradhan Mantri Suraksha Bima Yojana).

Task: Total number of Male/Female

Output:Hive

```
OK
gender--> Female Total count--> 311800
gender--> Male Total count--> 284723
Time taken: 29.985 seconds
hive> █
```

4.

Task 4.Citizens and immigrants count for employed lot

Output: Hive

```
OK
CitizenShip--> Immigrants Total Count--> 67265
CitizenShip--> Native Born United States Total Count--> 529258
Time taken: 26.96 seconds
hive> █
```

13. Option with your Foundation Degree:

Task1: Degree wise count for Employability

Output:MapReduce:

```
hduser@ubuntu64server:~$ hadoop fs -cat /kkl/p*
10th grade      10527
11th grade      11707
12th grade no diploma  3593
1st 2nd 3rd or 4th grade  2016
5th or 6th grade  4242
7th and 8th grade  6893
9th grade       7105
Associates degree-academic program  10856
Associates degree-occup /vocational  13138
Bachelors degree (BA AB BS)  49622
Children        0
Doctorate degree (PhD EdD)  3283
High school graduate  100492
Less than 1st grade  734
Masters degree (MA MS MEng MEd MSW MBA)  16706
Prof school degree (MD DDS DVM LLB JD)  4692
Some college but no degree  64665
hduser@ubuntu64server:~$
```

```
hive>
> select 'Education-->',edu,'Total Count-->',count(*) from final_census where ww=0 group by edu;
```

Output: Hive

```

OK
Education--> 10th grade      Total Count--> 12044
Education--> 11th grade      Total Count--> 8798
Education--> 12th grade no diploma Total Count--> 2681
Education--> 1st 2nd 3rd or 4th grade      Total Count--> 3339
Education--> 5th or 6th grade      Total Count--> 5511
Education--> 7th and 8th grade      Total Count--> 17234
Education--> 9th grade      Total Count--> 11430
Education--> Associates degree-academic program      Total Count--> 2094
Education--> Associates degree-occup /vocational      Total Count--> 2820
Education--> Bachelors degree(BA AB BS)      Total Count--> 9615
Education--> Children      Total Count--> 141496
Education--> Doctorate degree(PhD EdD)      Total Count--> 530
Education--> High school graduate      Total Count--> 44342
Education--> Less than 1st grade      Total Count--> 1678
Education--> Masters degree(MA MS MEng MEd MSW MBA) Total Count--> 2937
Education--> Prof school degree (MD DDS DVM LLB JD) Total Count--> 666
Education--> Some college but no degree      Total Count--> 19037
Time taken: 28.947 seconds
hive>

```

t1 (~/Desktop/mydata) - gedit

Output: Adv Map Reduce

```

hduser@ubuntu64server:~$ hadoop fs -cat /kk1/p*
10th grade      10527
11th grade      11707
12th grade no diploma      3593
1st 2nd 3rd or 4th grade      2016
5th or 6th grade      4242
7th and 8th grade      6893
9th grade      7105
Associates degree-academic program      10856
Associates degree-occup /vocational      13138
Bachelors degree(BA AB BS)      49622
Children      0
Doctorate degree(PhD EdD)      3283
High school graduate      100492
Less than 1st grade      734
Masters degree(MA MS MEng MEd MSW MBA)      16706
Prof school degree (MD DDS DVM LLB JD)      4692
Some college but no degree      64665
hduser@ubuntu64server:~$

```

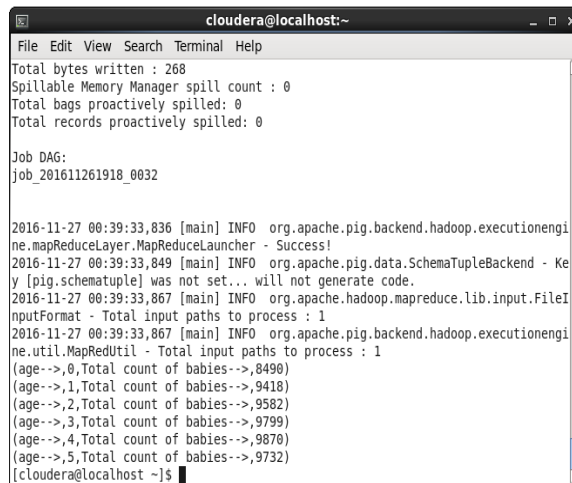

14. Customer base analysis

Task .Customer base analysis

Analyzing number of babies for vaccination whose age <=5:

Input: Output: Pig

```
a = load '/user/cloudera/Census_Records.json' using JsonLoader
('age:int,edu:chararray,mar:chararray,gen:chararray,tax:chararray,income:long,parent:chararray,country:chararray,citizen:chararray,ww:int');
b = foreach a generate age,gen,income;
c = filter b by age <=5;
d = group c by age;
e = foreach d generate 'age-->', group,'Total count of babies-->',COUNT(c.age);
dump e;
```



The screenshot shows a terminal window titled 'cloudera@localhost:~'. The output of the Pig job is as follows:

```
File Edit View Search Terminal Help
Total bytes written : 268
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_201611261918_0032

2016-11-27 00:39:33,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2016-11-27 00:39:33,849 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2016-11-27 00:39:33,867 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2016-11-27 00:39:33,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(age-->,0,Total count of babies-->,8490)
(age-->,1,Total count of babies-->,9418)
(age-->,2,Total count of babies-->,9582)
(age-->,3,Total count of babies-->,9799)
(age-->,4,Total count of babies-->,9870)
(age-->,5,Total count of babies-->,9732)
[cloudera@localhost ~]$
```

. Income tax assessments

Task 3.Non-US citizen(s) tax filer status

```
hive> select age,edu,gen,'TaxFilerStatus-->', tax,'Citizenship-->',citizen ,income,ww from final_census where citizen not in(' Native- Born in the United States');
total MapReduce jobs = 1
```

Output:Hive

37	Bachelors degree(BA AB BS)	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	624.68	0
3	Children	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	1205.3	0
25	Some college but no degree	Male	TaxFilerStatus-->	Single CitizenShip-->	Foreign born- Not a citizen of U S	3442.4	52	
28	Bachelors degree(BA AB BS)	Female	TaxFilerStatus-->	Single CitizenShip-->	Foreign born- U S citizen by naturalization	1741.48	52	
53	1st 2nd 3rd or 4th grade	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	1370.33	28
38	7th and 8th grade	Female	TaxFilerStatus-->	Head of household	CitizenShip-->	Foreign born- Not a citizen of U S	1219.11	0
14	Children	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	877.03	0
2	Children	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	862.44	0
3	Children	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	1804.74	0
46	5th or 6th grade	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- U S citizen by naturalization	2688.61	52
43	Some college but no degree	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	2239.83	52
41	5th or 6th grade	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	938.99	52
26	11th grade	Male	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	1199.34	52
22	Some college but no degree	Male	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Native- Born abroad of American Parent(s)	1900.14	52
12	Children	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- U S citizen by naturalization	1532.61	0
52	12th grade no diploma	Male	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Native- Born in Puerto Rico or U S Outlying	1140.64	0
25	Some college but no degree	Male	TaxFilerStatus-->	Single CitizenShip-->	Foreign born- Not a citizen of U S	1740.9	52	
46	Some college but no degree	Male	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	1031.19	52
48	High school graduate	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- U S citizen by naturalization	740	52
35	High school graduate	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	1584.92	0
26	9th grade	Male	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	1171.52	52
28	12th grade no diploma	Male	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	1140.07	52
43	Some college but no degree	Male	TaxFilerStatus-->	Single CitizenShip-->	Native- Born abroad of American Parent(s)	1019.25	36	
24	High school graduate	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- U S citizen by naturalization	852.49	52
31	High school graduate	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- U S citizen by naturalization	648.07	26
39	12th grade no diploma	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	1432.86	0
53	High school graduate	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- U S citizen by naturalization	2590.42	26
19	5th or 6th grade	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	1329.61	0
49	High school graduate	Female	TaxFilerStatus-->	Single CitizenShip-->	Native- Born in Puerto Rico or U S Outlying	1198.34	52	
23	High school graduate	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- Not a citizen of U S	2632.78	52
38	Some college but no degree	Female	TaxFilerStatus-->	Joint both under 65	CitizenShip-->	Foreign born- U S citizen by naturalization	1386.91	52
32	Some college but no degree	Male	TaxFilerStatus-->	Single CitizenShip-->	Foreign born- Not a citizen of U S	1230.37	0	
46	1st 2nd 3rd or 4th grade	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	762.63	0
37	7th and 8th grade	Male	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	1195.92	0
24	High school graduate	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	1761.07	0
24	7th and 8th grade	Male	TaxFilerStatus-->	Single CitizenShip-->	Foreign born- Not a citizen of U S	2938.54	52	
51	Masters degree(MA MS Meng MED MSW MBA)	Male	TaxFilerStatus-->	Single CitizenShip-->	Foreign born- U S citizen by naturalization	672.59	52	
5	Children	Male	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	649.07	0
26	5th or 6th grade	Female	TaxFilerStatus-->	Nonfiler	CitizenShip-->	Foreign born- Not a citizen of U S	1287.85	0

Time taken: 38.064 seconds
vive>

16. Naturalization Trends:

- i) In this use case we want to examine the roles of origin, mode of entry, and immigrant visa class in immigrants' transition to naturalization over the first decade or more of lawful permanent residence.
- ii) The next section reviews prior research and presents hypotheses as to origin, immigrant visa class, and mode of entry in the process of naturalizing. Following that review,
- iii) the Immigration-to-Naturalization Project (INP) dataset is described. In this use case, we are fetching all the records from "country of birth" field we are finding person who are not born in the United States.

Task .Country of birth wise count for US citizenship by naturalization

```
Time taken: 0.1200 seconds
vive> select cntry,count(citizen) from final_census where citizen
      >=' Foreign born- U S citizen by naturalization' group by cntry;
```

Output:Hive:

Country of birth wise count for US citizenship by naturalization

Hive

```
OK
?      3113
Cambodia      75
Canada 770
China  430
Columbia      397
Cuba    1251
Dominican-Republic  379
Ecuador      192
El-Salvador  227
England      496
France  87
Germany      1054
Greece 300
Guatemala      98
Haiti  144
Holand-Netherlands  28
Honduras      87
Hong Kong      99
Hungary      187
India  384
Iran   141
Ireland      206
Italy  793
Jamaica      342
Japan  152
Laos   82
Mexico 2218
Nicaragua      110
Panama  38
Peru   202
Philippines  1220
Poland  577
Portugal      248
Scotland      106
South Korea   472
Taiwan  283
Thailand      53
Trinidad&Tobago  62
Vietnam      371
Yugoslavia    141
Time taken: 27.363 seconds
hive> █
```

