

planning

```
1. select COUNT(*) as Total_Voters from final_census where age+($hiveconf:year)-YEAR(from_unixtime(unix_timestamp()))>=18;
```

```
2. select COUNT(*) as Total_Senior_Citizen from final_census where age+($hiveconf:year)-YEAR(from_unixtime(unix_timestamp()))>=60;
```

```
3. select gender, COUNT(*) as Total from final_census group by gender;
```

```
4. select citizenship, COUNT(*) from ( select CASE citizenship when 'Native- Born in the United States' then 'Native Born United States' else 'Immigrants' END citizenship from final_census) a group by citizenship;
```

Education

```
1. select education ,gender, COUNT(*) Total from final_census group by education, gender;
```

```
2. select education , SUM(CASE when weeks_worked <=0 then '1' else null END) as Employed , SUM(CASE when weeks_worked >0 then '1' else null END) as Unemployed from final_census group by education;
```

```
3.select education ,COUNT(*) as Total_Peoples from final_census where age between 18 and 25;
```

PIG(planing)

1.

```
step1 = LOAD '/user/cloudera/CensusData' using PigStorage(',') as (age : int , education , marital_status , gender , tax_fil_status , income: double , parents , country_birth , citizenship , weeks_worked );
```

```
step2 = FILTER step1 by age + ($YEAR-GetYear(CurrentTime()))>=18;
```

```
step3 = FOREACH step2 GENERATE 1 as one, age;
```

```
step4 = GROUP step3 by one;
```

```
step5 = FOREACH step4 GENERATE COUNT(step3.age) as TOTAL_VOTERS;
```

```
DUMP step5;
```

2.

```
step1 = LOAD '/user/cloudera/CensusData' using PigStorage(',') as (age : int , education , marital_status , gender , tax_fil_status , income: double , parents , country_birth , citizenship , weeks_worked );
```

```
step2 = FILTER step1 by age + ($YEAR-GetYear(CurrentTime()))>=$SENIOR_AGE;
```

```
step3 = FOREACH step2 GENERATE 1 as one, age;
```

```
step4 = GROUP step3 by one;
```

```
step5 = FOREACH step4 GENERATE COUNT(step3.age) as TOTAL_SENIOR_CITIZEN;
```

```
DUMP step5;
```

3.

```
step1      =      LOAD      '/user/cloudera/Census_Records.json'      using
JsonLoader('Age:chararray,      Education:chararray,
MaritalStatus:chararray, Gender:chararray,
TaxFilerStatus:chararray,      Income:      double,      Parents:chararray,
CountryOfBirth:chararray, Citizenship:chararray, WeeksWorked:chararray');
step2 = GROUP step1 by Gender;
step3 = FOREACH step2 GENERATE group, COUNT(step1.Gender) as Total_Num;
DUMP step3;
```

Financial

1.

```
select SUM(income*tax_pct) as Total_tax ,SUM(CASE f.gender WHEN ' Male'
then Income END) as Tax_Male, SUM(CASE f.gender WHEN ' Female' then
Income END) as Tax_Female from fina_census f join gen_wise_tax t
on(f.gender=t.gender) where f.income between t.minamount and t.maxamount;
```

Finance

```
2.select Education,SUM(income)/COUNT(*), SUM(CASE gender when ' Male'
then Income END)/COUNT(CASE gender when ' Male' then 1 END) as For_Male,
SUM(CASE gender when ' Female' then Income END)/COUNT(CASE gender when '
Female' then 1 END) as For_Female from final_census group by Education
```

PIG (Miscellaneous)

1.

```
step1      =      LOAD      '/user/cloudera/Census_Records.json'      using
JsonLoader('Age:chararray,      Education:chararray,
MaritalStatus:chararray,
Gender:chararray,      TaxFilerStatus:chararray,      Income:      double,
Parents:chararray, CountryOfBirth:chararray, Citizenship:chararray,
WeeksWorked:double');
step2 = FILTER step1 by WeeksWorked==0;
step3 = GROUP step2 by Education;
step4 = FOREACH step3 GENERATE group, COUNT(step2.Age) as Total;
DUMP step4;
```

2.

```
step1      =      LOAD      '/user/cloudera/Census_Records.json'      using
JsonLoader('Age:chararray,      Education:chararray,
MaritalStatus:chararray,
Gender:chararray,      TaxFilerStatus:chararray,      Income:      double,
Parents:chararray, CountryOfBirth:chararray,
Citizenship:chararray, WeeksWorked:double');
step2 = FILTER step1 by Education matches '.*grade.*';
step3 = FILTER step2 by not(Parents matches ' Both parents present');
step4 = FOREACH step3 GENERATE 1 as one, Education;
step5 = GROUP step4 by one;
```

```
step6 = FOREACH step5 GENERATE COUNT(step4.Education) as Edcnt;  
DUMP step6;
```

3.

```
step1 = LOAD '/user/cloudera/Census_Records.json' using  
JsonLoader('Age:chararray, Education:chararray,  
MaritalStatus:chararray,  
Gender:chararray, TaxFilerStatus:chararray, Income: double,  
Parents:chararray, CountryOfBirth:chararray, Citizenship:chararray,  
WeeksWorked:double');  
step2 = FILTER step1 by not (Citizenship matches ' Native- Born in the  
United States');  
step3 = FOREACH step2 GENERATE Age, Education, TaxFilerStatus, Gender,  
Citizenship;  
DUMP step3;
```

4.

```
step1 = LOAD '/user/cloudera/Census_Records.json' using  
JsonLoader('Age:chararray, Education:chararray,  
MaritalStatus:chararray,  
Gender:chararray, TaxFilerStatus:chararray, Income: double,  
Parents:chararray, CountryOfBirth:chararray, Citizenship:chararray,  
WeeksWorked:double');  
step2 = FILTER step1 by Citizenship matches ' Foreign born- U S citizen  
by naturalization';  
step3 = GROUP step2 by CountryOfBirth;  
step4 = FOREACH step3 GENERATE group, COUNT(step2.Age);  
DUMP step4;
```

social

2.

```
step1 = LOAD '/user/cloudera/Census_Records.json' using  
JsonLoader('Age:chararray, Education:chararray,  
MaritalStatus:chararray, Gender:chararray, TaxFilerStatus:chararray,  
Income: double, Parents:chararray, CountryOfBirth:chararray,  
Citizenship:chararray, WeeksWorked:double');  
step2 = LOAD '/user/cloudera/scholarship/part-m-00000' using  
PigStorage(',') as (Parents , scholarship:int);  
step3 = JOIN step1 by Parents , step2 by Parents;  
step4 = FOREACH step3 GENERATE step1.Parents , scholarship;  
step5 = GROUP step4 by Parents;  
step6 = FOREACH step5 GENERATE group, SUM(step4.scholarship);  
DUMP step6;
```