PIG

**Task 1: Total count of male/female based on education.**

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,Marital:chararray,Gender:chararray,Tax:chararray,Income:float,Parent:chararray,Birth:chararray,Citizen:chararray,Work:chararray');

step2 = foreach step1 generate $1 as Edu,$3 as Gen;

step3 = group step2 by ($0,$1);

step4 = foreach step3 generate group,COUNT(step2.Gen);

dump step4;

**Task 2: Total count of employed/unemployed based on education.**

**Task 2 Employed:**

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,Marital:chararray,Gender:chararray,Tax:chararray,Income:float,Parent:chararray,Birth:chararray,Citizen:chararray,Work:int');

step2 = foreach step1 generate $1 as Edu,$9 as ww;

step3 = filter step2 by $1>0;

step4 = group step3 by $0;

step5 = foreach step4 generate group,COUNT($1);

dump step5;

Task 2 UnEmployed:

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,Marital:chararray,Gender:chararray,Tax:chararray,Income:float,Parent:chararray,Birth:chararray,Citizen:chararray,Work:int');

step2 = foreach step1 generate $1 as Edu,$9 as ww;

step3 = filter step2 by $1=0;

step4 = group step3 by $0;

step5 = foreach step4 generate group,COUNT($1);

dump step5;

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

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,Marital:chararray,Gender:chararray,Tax:chararray,Income:float,Parent:chararray,Birth:chararray,Citizen:chararray,Work:int');

step2 = foreach step1 generate $1 as Edu,$0 as age;

step3 = filter step2 by $1>18 and $1<25;

step4 = group step3 by $0;

step5 = foreach step4 generate group,COUNT($0);

dump step5;

**Task 4: Customer base analysis (Unemployed in Widowed)**

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,Marital:chararray,Gender:chararray,Tax:chararray,Income:float,Parent:chararray,Birth:chararray,Citizen:chararray,Work:int');

step2 = foreach step1 generate Age,Gender,Work,Marital;

step3 = filter step2 by ((Gender==' Female' and work==0 and Marital==' Widowed') and (age>21 and age<60)) ;

step4 = group step3 by age;

step5 = foreach step4 generate group,COUNT(d.age);

dump step5;

**Task 5: For given age range employable female widowed and divorced count(default range 21-60)**

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,Marital:chararray,Gender:chararray,Tax:chararray,Income:float,Parent:chararray,Birth:chararray,Citizen:chararray,Work:int');

step2 = foreach step1 generate Age,Gender,Work,Marital;

step3 = filter step2 by ((Gender==' Female' and work>0) and (Marital==' Widowed' or Marital==' Divorced') and (age>21 and age<60));

step4 = group step3 by age;

step5 = foreach step4 generate group,COUNT(d.age);

dump step5;