Task 1 : Total count of male/female based on education. :if it bachelor, how many male/female

PIG :

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,MartialStatus:chararray,Gender:chararray,TaxFilerStatus:chararray,Income:float,Parents:chararray,CountryOfBirth:chararray,Citizenship:chararray,WeeksWorked: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;

Employed Counts :

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,MartialStatus:chararray,Gender:chararray,TaxFilerStatus:chararray,Income:float,Parents:chararray,CountryOfBirth:chararray,Citizenship:chararray,WeeksWorked: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;

UnEmployed:

step1 = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,MartialStatus:chararray,Gender:chararray,TaxFilerStatus:chararray,Income:float,Parents:chararray,CountryOfBirth:chararray,Citizenship:chararray,WeeksWorked: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;

2. Total amount dispensed on scholarship in current year:

PIG:

PIG File: t1.txt

a = load '/user/cloudera/Census\_Records.json' using JsonLoader('Age:int,Education:chararray,MartialStatus:chararray,Gender:chararray,TaxFilerStatus:chararray,Income:float,Parents:chararray,CountryOfBirth:chararray,Citizenship:chararray,WeeksWorked:chararray');

b = load '/user/cloudera/scholar1' using PigStorage(',') as (status:chararray,schamt:int);

c = join a by Parents,b by status;

d = foreach c generate $6 as parent,$11 as Schamt;

e = group d by $0;

f = foreach e generate group,SUM(d.Schamt);

dump f;

Secondary table: scholar1:

Father only present, 2000

Mother only present, 4000

Neither parent present, 7000

Not in universe, 10000