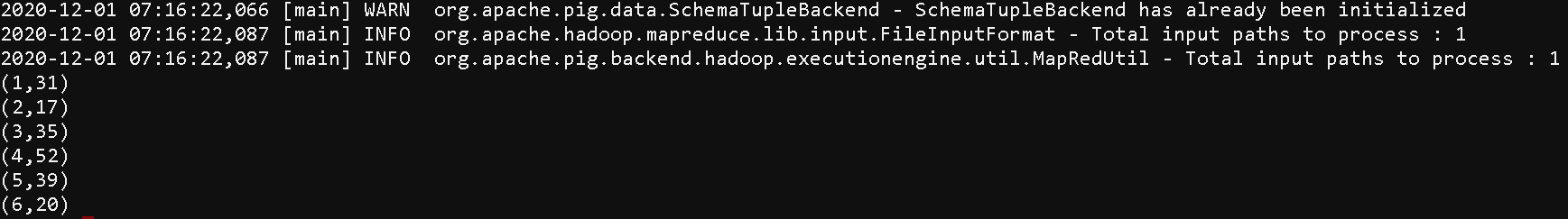
Country Project using PIG

Problem Statement

**A.Count number of countries based on landmass.**

* country\_data = LOAD 'countrydata.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:chararray, population:chararray, lang:int, rel:int, bar:int, stripes:int, col:int, red:int, green:int, blue:int, gold:int, white:int, black:int, orange:int, hue:chararray, circle:int, cross:int, saltire:int, quars:int, suns:int, cres:int, triangle:int, icon:int, animate:int, text:int, left:chararray, right:chararray);
* grouped\_data = GROUP country\_data by landmass;
* count\_data = FOREACH grouped\_data GENERATE group,COUNT(country\_data);
* DUMP count\_data;



**B.Find out top 5 country with Sum of bars and strips in a flag.**

* country\_data = LOAD 'countrydata.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:chararray, population:chararray, lang:int, rel:int, bar:int, stripes:int, col:int, red:int, green:int, blue:int, gold:int, white:int, black:int, orange:int, hue:chararray, circle:int, cross:int, saltire:int, quars:int, suns:int, cres:int, triangle:int, icon:int, animate:int, text:int, left:chararray, right:chararray);
* foreach\_data = FOREACH country\_data GENERATE name,$7+$8;
* grouped\_data = GROUP foreach\_data All;
* bar\_data = FOREACH grouped\_data {

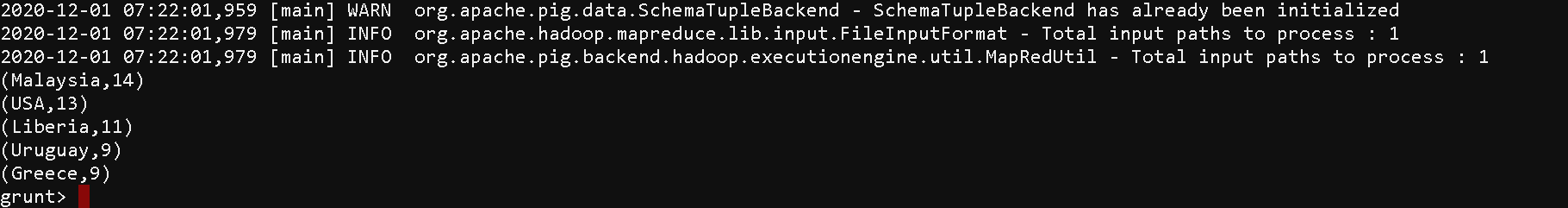
ord = ORDER foreach\_data BY $1 DESC;

top = LIMIT ord 5;

GENERATE FLATTEN(top);

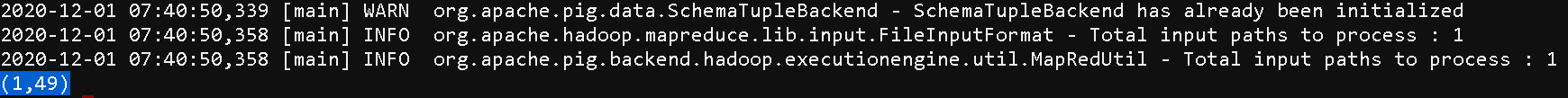
};

* DUMP bar\_data;



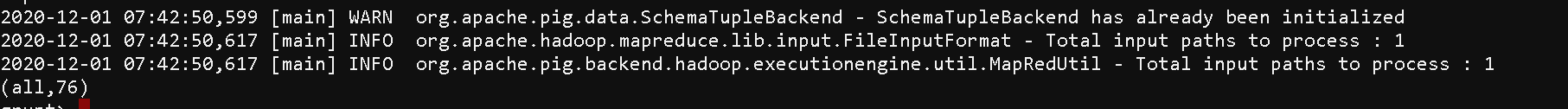
**C.Count of countries with icon**.

* filtered\_icon\_data = FILTER country\_data BY icon == 1;
* grouped\_icon\_data = GROUP filtered\_icon\_data BY icon;
* count\_icon\_data = FOREACH grouped\_icon\_data GENERATE group, COUNT(filtered\_icon\_data);
* DUMP count\_icon\_data;



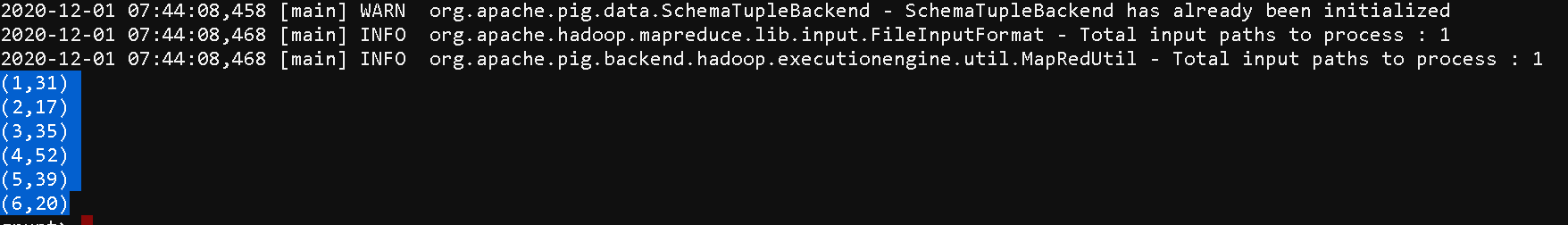
**D.Count of countries which have same top left and top right color in flag.**

* filtered\_flag\_data = FILTER country\_data BY $28 == $29;
* grouped\_flag\_data = GROUP filtered\_flag\_data All;
* count\_flag\_data = FOREACH grouped\_flag\_data GENERATE group, COUNT(filtered\_flag\_data);
* DUMP count\_flag\_data;



**E.Count number of countries based on zone.**

* grouped\_zone\_data = GROUP country\_data by landmass;
* count\_zone\_data = FOREACH grouped\_zone\_data GENERATE group,COUNT(country\_data);
* DUMP count\_zone\_data;



**F.Find out largest county in terms of area in NE zone**.

* filtered\_NE\_data = FILTER country\_data BY zone == 1;
* grouped\_NE\_data = GROUP filtered\_NE\_data All;
* max\_area\_data = FOREACH grouped\_NE\_data {

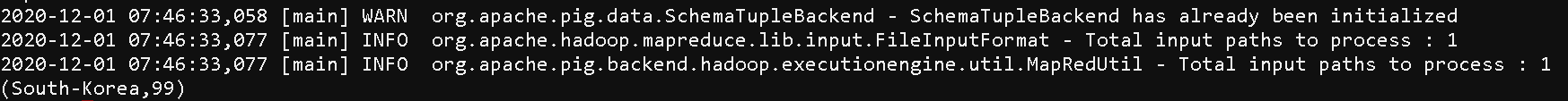
ord = ORDER filtered\_NE\_data BY $3 DESC;

top = LIMIT ord 1;

GENERATE FLATTEN(top);

};

* max\_area\_country = FOREACH max\_area\_data GENERATE name,area;
* DUMP max\_area\_country;



**G.Find out least populated country in S.America Landmass**

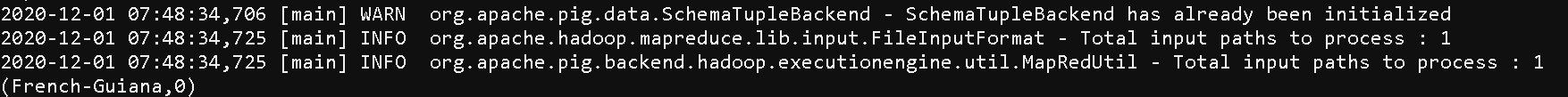
* filtered\_SA\_data = FILTER country\_data BY landmass == 2;
* grouped\_SA\_data = GROUP filtered\_SA\_data All;
* min\_population\_data = FOREACH grouped\_SA\_data {

ord = ORDER filtered\_SA\_data BY $4 ASC;

top = LIMIT ord 1;

GENERATE FLATTEN(top);

* };
* min\_population\_country = FOREACH min\_population\_data GENERATE name,population;
* DUMP min\_population\_country



**H.Find out largest speaking language among all countries.**

* foreach\_data = FOREACH country\_data GENERATE name,lang;
* grouped\_data = GROUP foreach\_data BY lang;
* count\_data = FOREACH grouped\_data GENERATE group,(int)COUNT(foreach\_data);
* grouped\_count\_data = GROUP count\_data All;
* max\_lang\_data = FOREACH grouped\_count\_data {

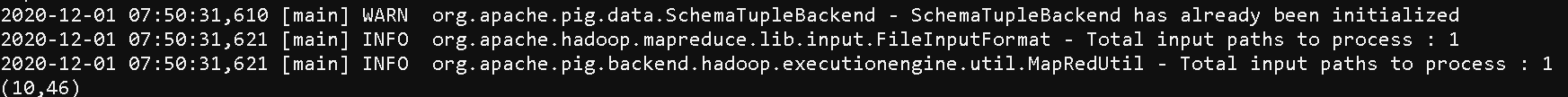
ord = ORDER count\_data BY $1 DESC;

top = LIMIT ord 1;

GENERATE FLATTEN(top);

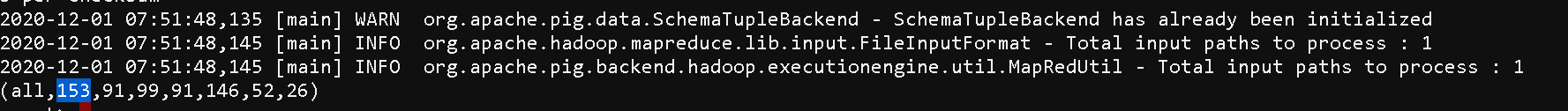
};

* DUMP max\_lang\_data;



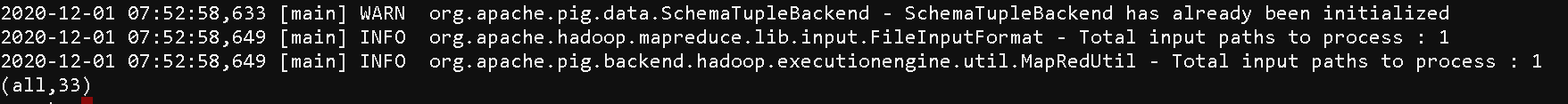
**I.Find most common colour among flags from all countries.**

* grouped\_data = GROUP country\_data All;
* final\_color\_data = FOREACH grouped\_data GENERATE group,SUM(country\_data.$10),SUM(country\_data.$11),SUM(country\_data.$12),SUM(country\_data.$13),SUM(country\_data.$14),SUM(country\_data.$15),SUM(country\_data.$16);
* dump final\_color\_data;



**J.Sum of all circles present in all country flags.**

* grouped\_circle\_data = GROUP country\_data All;
* sum\_circle = FOREACH grouped\_circle\_data GENERATE group,SUM(country\_data.circle);
* dump sum\_circle;



**K.Count of countries which have both icon and text in flag.**

* filtered\_country\_data= FILTER country\_data BY ($25 == 1) | ($29!=0);
* coun = FOREACH country\_data GENERATE (country\_data.text is NULL ? 0 : 1);
* filtered\_country\_data2= FILTER filtered\_country\_data1 BY ($29 == 1);
* grouped\_text\_data = GROUP filtered\_country\_data All;
* count\_text\_data = FOREACH grouped\_text\_data GENERATE group, COUNT(filtered\_country\_data);
* DUMP count\_text\_data;

