A. Count number of countries based on landmass.

Command:

country_data = LOAD 'flag.data_dataset.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;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:chara
rray, 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:i
nt, animate:int, text:int, left:chararray, right:chararray);  grouped_data = GROUP country_data by landmass;  count_data = FORE
ACH grouped_data GENERATE group,COUNT(country_data); DUMP count_data;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime     MinMapTIme      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgRedu
ceTime  MedianReducetime        Alias   Feature Outputs
job_1616262377730_6523  1       1       4       4       4       4       3       3       3       3       count_data,country_data
,grouped_data   GROUP_BY,COMBINER       hdfs://nameservice1/tmp/temp1114119728/tmp-471805958,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/edureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 6 records (47 bytes) in: "hdfs://nameservice1/tmp/temp1114119728/tmp-471805958"

Counters:
Total records written : 6
Total bytes written : 47
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6523
```

```
2021-04-25 19:40:05,799 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 19:40:05,800 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead,
 use fs.defaultFS
2021-04-25 19:40:05,800 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not gen
erate code.
2021-04-25 19:40:05,812 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-04-25 19:40:05,812 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to proce
ss : 1
(1,31)
(2,17)
(3,35)
(4,52)
(5,39)
(6,20)
grunt>
```

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

Command:
country_data = LOAD 'flag.data_dataset.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;

Screenshot:


```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:chararray, population:chararray, l
ang: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, anima
te:int, text:int, left:chararray, right:chararray); foreach_data = FOREACH cou
ntry_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);};
grunt> DUMP bar_data;
```

Results:

```
Counters:
Total records written : 5
Total bytes written : 76
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6524


2021-04-25 19:44:17,295 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 19:44:17,295 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 19:44:17,296 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 19:44:17,299 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 19:44:17,299 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(Malaysia,14)
(USA,13)
(Liberia,11)
(Uruguay,9)
(Greece,9)
grunt>
```

C. Count of countries with icon.

Command:
country_data = LOAD 'flag.data_dataset.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);
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;

Screenshots:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:chararray, population:chararray, l
ang: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, anima
te:int, text:int, left:chararray, right:chararray);
grunt> filtered_icon_data = FILTER country_data BY icon == 1;
grunt> grouped_icon_data = GROUP filtered_icon_data BY icon;
grunt> count_icon_data = FOREACH grouped_icon_data GENERATE group, COUNT(filte
red_icon_data);
grunt> DUMP count_icon_data;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps     Reduces MaxMapTime      MinMapTIme      AvgMapTime      Median
MapTime MaxReduceTime    MinReduceTime    AvgReduceTime    MedianReducetime     A
lias     Feature Outputs
job_1616262377730_6525  1       1       14      14      14      14      2     2
2       2               count_icon_data,country_data,filtered_icon_data,grouped_icon_d
ata      GROUP_BY,COMBINER        hdfs://nameservice1/tmp/temp1114119728/tmp3075
37119,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (7 bytes) in: "hdfs://nameservice1/tmp/temp11141
19728/tmp307537119"

Counters:
Total records written : 1
Total bytes written : 7
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6525


2021-04-25 19:49:01,081 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 19:49:01,081 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 19:49:01,082 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 19:49:01,084 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 19:49:01,084 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(1,49)
```

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

Command:
country_data = LOAD 'flag.data_dataset.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);
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;

Screenshots:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:chararray, population:chararray, l
ang: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, anima
te:int, text:int, left:chararray, right:chararray);
grunt> filtered_flag_data = FILTER country_data BY $28 == $29;
grunt> grouped_flag_data = GROUP filtered_flag_data All;
grunt> count_flag_data = FOREACH grouped_flag_data GENERATE group, COUNT(filte
red_flag_data);
grunt> DUMP count_flag_data;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps      Reduces MaxMapTime      MinMapTIme      AvgMapTime      Median
MapTime MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime     A
lias     Feature Outputs
job_1616262377730_6526  1       1       3       3       3       3       3       3
3       3       count_flag_data,country_data,filtered_flag_data,grouped_flag_d
ata     GROUP_BY,COMBINER       hdfs://nameservice1/tmp/temp1114119728/tmp-126
353399,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (12 bytes) in: "hdfs://nameservice1/tmp/temp1114
119728/tmp-126353399"

Counters:
Total records written : 1
Total bytes written : 12
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6526


2021-04-25 19:51:41,983 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 19:51:41,984 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 19:51:41,984 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 19:51:41,987 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 19:51:41,987 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(all,76)
```

E. Count number of countries based on zone.

Command:

country_data = LOAD 'flag.data_dataset.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_zone_data = GROUP country_data by landmass;
count_zone_data = FOREACH grouped_zone_data GENERATE group,COUNT(country_data);
DUMP count_zone_data;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:chararray, population:chararray, l
ang: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, anima
te:int, text:int, left:chararray, right:chararray);
grunt> grouped_zone_data = GROUP country_data by landmass;
grunt> count_zone_data = FOREACH grouped_zone_data GENERATE group,COUNT(countr
y_data);
grunt> DUMP count_zone_data;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps     Reduces MaxMapTime      MinMapTIme      AvgMapTime      Median
MapTime MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime     A
lias     Feature Outputs
job_1616262377730_6527  1       1       3       3       3       3       14    1
4       14      14      count_zone_data,country_data,grouped_zone_data   GROUP_
BY,COMBINER     hdfs://nameservice1/tmp/temp1114119728/tmp329597333,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 6 records (47 bytes) in: "hdfs://nameservice1/tmp/temp1114
119728/tmp329597333"

Counters:
Total records written : 6
Total bytes written : 47
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6527
```

```
2021-04-25 19:56:01,981 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 19:56:01,981 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 19:56:01,982 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 19:56:01,985 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 19:56:01,985 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(1,31)
(2,17)
(3,35)
(4,52)
(5,39)
(6,20)
grunt>
```

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

Command:
country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:int, 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);
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;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:int, population:chararray, lang:in
t, 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, s
altire:int, quars:int, suns:int, cres:int, triangle:int, icon:int, animate:int
, text:int, left:chararray, right:chararray);
2021-04-25 19:58:25,817 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> filtered_NE_data = FILTER country_data BY zone == 1;
grunt> grouped_NE_data = GROUP filtered_NE_data All;
grunt> max_area_data = FOREACH grouped_NE_data {
>> ord = ORDER filtered_NE_data BY $3 DESC;
>> top = LIMIT ord 1;
>> GENERATE FLATTEN(top); };
grunt> max_area_country = FOREACH max_area_data GENERATE name,area;
grunt> DUMP max_area_country;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps    Reduces MaxMapTime     MinMapTIme     AvgMapTime     Median
MapTime MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime    A
lias    Feature Outputs
job_1616262377730_6529  1       1       5       5       5       5       2       2
2       2       country_data,filtered_NE_data,grouped_NE_data,max_area_country
,max_area_data,top      GROUP_BY        hdfs://nameservice1/tmp/temp1114119728
/tmp204194171,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (14 bytes) in: "hdfs://nameservice1/tmp/temp1114
119728/tmp204194171"

Counters:
Total records written : 1
Total bytes written : 14
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6529


2021-04-25 19:59:05,661 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 19:59:05,662 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 19:59:05,662 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 19:59:05,665 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 19:59:05,665 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(USSR,22402)
```

G. Find out least populated country in S. America landmass.

Command:

country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:int, 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);
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;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:int, population:chararray, lang:in
t, 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, s
altire:int, quars:int, suns:int, cres:int, triangle:int, icon:int, animate:int
, text:int, left:chararray, right:chararray);
grunt> filtered_SA_data = FILTER country_data BY landmass == 2;
grunt> grouped_SA_data = GROUP filtered_SA_data All;
grunt> min_population_data = FOREACH grouped_SA_data {
>> ord = ORDER filtered_SA_data BY $4 ASC;
>> top = LIMIT ord 1;
>> GENERATE FLATTEN(top); };
grunt> min_population_country = FOREACH min_population_data GENERATE name,popu
lation;
grunt> DUMP min_population_country;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps     Reduces MaxMapTime      MinMapTIme      AvgMapTime      Median
MapTime MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime    A
lias    Feature Outputs
job_1616262377730_6530  1       1       2       2       2       2       8       8
8       8       country_data,filtered_SA_data,grouped_SA_data,min_population_c
ountry,min_population_data,top  GROUP_BY        hdfs://nameservice1/tmp/temp11
14119728/tmp1433800941,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (24 bytes) in: "hdfs://nameservice1/tmp/temp1114
119728/tmp1433800941"

Counters:
Total records written : 1
Total bytes written : 24
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
Job DAG:
job_1616262377730_6530


2021-04-25 20:01:00,716 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 20:01:00,717 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 20:01:00,717 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 20:01:00,720 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 20:01:00,720 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(French-Guiana,0)
grunt>
```

H. Find out largest speaking language among all countries.

Command:
country_data = LOAD 'flag.data_dataset.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,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;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:chararray, population:chararray, l
ang: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, anima
te:int, text:int, left:chararray, right:chararray);
2021-04-25 20:03:29,428 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> foreach_data = FOREACH country_data GENERATE name,lang;
grunt> grouped_data = GROUP foreach_data BY lang;
grunt> count_data = FOREACH grouped_data GENERATE group,(int)COUNT(foreach_dat
a);
grunt> grouped_count_data = GROUP count_data All;
grunt> max_lang_data = FOREACH grouped_count_data {
>> ord = ORDER count_data BY $1 DESC;
>> top = LIMIT ord 1;
>> GENERATE FLATTEN(top);
>> };
grunt> DUMP max_lang_data;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps     Reduces MaxMapTime      MinMapTIme      AvgMapTime      Median
MapTime MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime    A
lias     Feature Outputs
job_1616262377730_6532  1        1       3       3       3       3       3    3
3       3       count_data,country_data,foreach_data,grouped_data       GROUP_
BY,COMBINER
job_1616262377730_6533  1        1       2       2       2       2       14   1
4       14      14      grouped_count_data,max_lang_data,ord,top       GROUP_
BY      hdfs://nameservice1/tmp/temp1114119728/tmp1079149918,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (8 bytes) in: "hdfs://nameservice1/tmp/temp11141
19728/tmp1079149918"

Counters:
Total records written : 1
Total bytes written : 8
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
Job DAG:
job_1616262377730_6532  ->      job_1616262377730_6533,
job_1616262377730_6533


2021-04-25 20:04:29,167 [main] INFO  org.apache.hadoop.ipc.Client - Retrying c
onnect to server: ip-20-0-31-221.ec2.internal/20.0.31.221:38969. Already tried
 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3,
sleepTime=1000 MILLISECONDS)
2021-04-25 20:04:30,168 [main] INFO  org.apache.hadoop.ipc.Client - Retrying c
onnect to server: ip-20-0-31-221.ec2.internal/20.0.31.221:38969. Already tried
 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3,
sleepTime=1000 MILLISECONDS)
2021-04-25 20:04:31,169 [main] INFO  org.apache.hadoop.ipc.Client - Retrying c
onnect to server: ip-20-0-31-221.ec2.internal/20.0.31.221:38969. Already tried
 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3,
sleepTime=1000 MILLISECONDS)
2021-04-25 20:04:31,272 [main] INFO  org.apache.hadoop.mapred.ClientServiceDel
egate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redi
recting to job history server
2021-04-25 20:04:31,370 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 20:04:31,371 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 20:04:31,371 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 20:04:31,378 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 20:04:31,378 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(10,46)
```

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

Command:
country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:chararray, population:chararray, lang:chararray, 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 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;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:chararray, population:chararray, l
ang:chararray, rel:int, bar:int, stripes:int, col:int, red:int, green:int, blu
e: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);
grunt> grouped_data = GROUP country_data All;
grunt> 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(co
untry_data.$14),SUM(country_data.$15),SUM(country_data.$16);
grunt> dump final_color_data;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps     Reduces MaxMapTime      MinMapTIme       AvgMapTime        Median
MapTime MaxReduceTime   MinReduceTime    AvgReduceTime    MedianReducetime     A
lias     Feature Outputs
job_1616262377730_6534  1        1        2       2        2       2       2    2
2       2         country_data,final_color_data,grouped_data       GROUP_BY,COMBI
NER     hdfs://nameservice1/tmp/temp1114119728/tmp302683494,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (26 bytes) in: "hdfs://nameservice1/tmp/temp1114
119728/tmp302683494"

Counters:
Total records written : 1
Total bytes written : 26
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
Job DAG:
job_1616262377730_6534


2021-04-25 20:07:30,232 [main] INFO  org.apache.hadoop.mapred.ClientServiceDel
egate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redi
recting to job history server
2021-04-25 20:07:30,275 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 20:07:30,275 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 20:07:30,275 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 20:07:30,278 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 20:07:30,278 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(all,153,91,99,91,146,52,26)
```

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

Command:

country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:chararray, population:chararray, lang:chararray, 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_circle_data = GROUP country_data All;
sum_circle = FOREACH grouped_circle_data GENERATE group,SUM(country_data.circle);
dump sum_circle;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (n
ame:chararray, landmass:int, zone:int, area:chararray, population:chararray, l
ang:chararray, rel:int, bar:int, stripes:int, col:int, red:int, green:int, blu
e: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);
grunt> grouped_circle_data = GROUP country_data All;
grunt> sum_circle = FOREACH grouped_circle_data GENERATE group,SUM(country_dat
a.circle);
grunt> dump sum_circle;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps     Reduces MaxMapTime      MinMapTIme      AvgMapTime      Median
MapTime MaxReduceTime    MinReduceTime    AvgReduceTime    MedianReducetime    A
lias     Feature Outputs
job_1616262377730_6535  1       1       3       3       3       3       2       2
2       2           country_data,grouped_circle_data,sum_circle     GROUP_BY,COMBI
NER     hdfs://nameservice1/tmp/temp1114119728/tmp466570499,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/ed
ureka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (12 bytes) in: "hdfs://nameservice1/tmp/temp1114
119728/tmp466570499"

Counters:
Total records written : 1
Total bytes written : 12
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6535

Job DAG:
job_1616262377730_6535


2021-04-25 20:09:13,267 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 20:09:13,267 [main] INFO  org.apache.hadoop.conf.Configuration.depr
ecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 20:09:13,268 [main] INFO  org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2021-04-25 20:09:13,271 [main] INFO  org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2021-04-25 20:09:13,271 [main] INFO  org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(all,33)
```

K. Count of countries which have both icon and text in flag.
Command:
country_data = LOAD 'flag.data_dataset.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);
filtered_country_data = FILTER country_data BY ($25 == 1) AND ($27 == 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;

Screenshot:

```
grunt> country_data = LOAD 'flag.data_dataset.txt' USING PigStorage(',') as (nam
e: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, sal
tire:int, quars:int, suns:int, cres:int, triangle:int, icon:int, animate:int, te
xt:int, left:chararray, right:chararray);
2021-04-25 20:44:34,501 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> filtered_country_data = FILTER country_data BY ($25 == 1) AND ($27 == 1);
grunt> grouped_text_data = GROUP filtered_country_data All;
grunt> count_text_data = FOREACH grouped_text_data GENERATE group, COUNT(filtere
d_country_data);
grunt> DUMP count_text_data;
```

Results:

```
Success!

Job Stats (time in seconds):
JobId    Maps     Reduces MaxMapTime       MinMapTIme       AvgMapTime       MedianMa
pTime    MaxReduceTime   MinReduceTime    AvgReduceTime    MedianReducetime      A
lias     Feature Outputs
job_1616262377730_6539  1       1       2       2       2       2       2       2
2       2        count_text_data,country_data,filtered_country_data,grouped_text_
data    GROUP_BY,COMBINER       hdfs://nameservice1/tmp/temp-1096107230/tmp74089
314,

Input(s):
Successfully read 194 records (15815 bytes) from: "hdfs://nameservice1/user/edur
eka_1472935/flag.data_dataset.txt"

Output(s):
Successfully stored 1 records (12 bytes) in: "hdfs://nameservice1/tmp/temp-10961
07230/tmp74089314"

Counters:
Total records written : 1
Total bytes written : 12
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
Job DAG:
job_1616262377730_6539


2021-04-25 20:45:09,098 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2021-04-25 20:45:09,099 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-25 20:45:09,099 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2021-04-25 20:45:09,102 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2021-04-25 20:45:09,102 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(all,13)
grunt>
```