COUNTRY PROJECT ANALYSIS

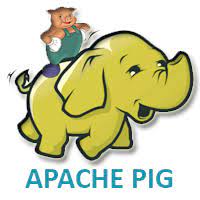
Using Pig Script

Problem Statement

1. Count number of countries based on landmass.
2. Find out top 5 country with Sum of bars and strips in a flag.
3. Count of countries with icon.
4. Count of countries which have same top left and top right color in flag.
5. Count number of countries based on zone.
6. Find out largest county in terms of area in NE zone.
7. Find out least populated country in S.America landmass.
8. Find out largest speaking language among all countries.
9. Find most common colour among flags from all countries.
10. Sum of all circles present in all country flags.
11. Count of countries which have both icon and text in flag.

Big Data Eco-system

Logo

Description automatically generated

Dataset

<http://www.edureka.co/medias/yz5zdt174e/download?media_file_id=171471702>

Dataset Description

<http://www.edureka.co/medias/nsaehvuar2/download?media_file_id=171471703>

1. Graphical user interface, text, application, email

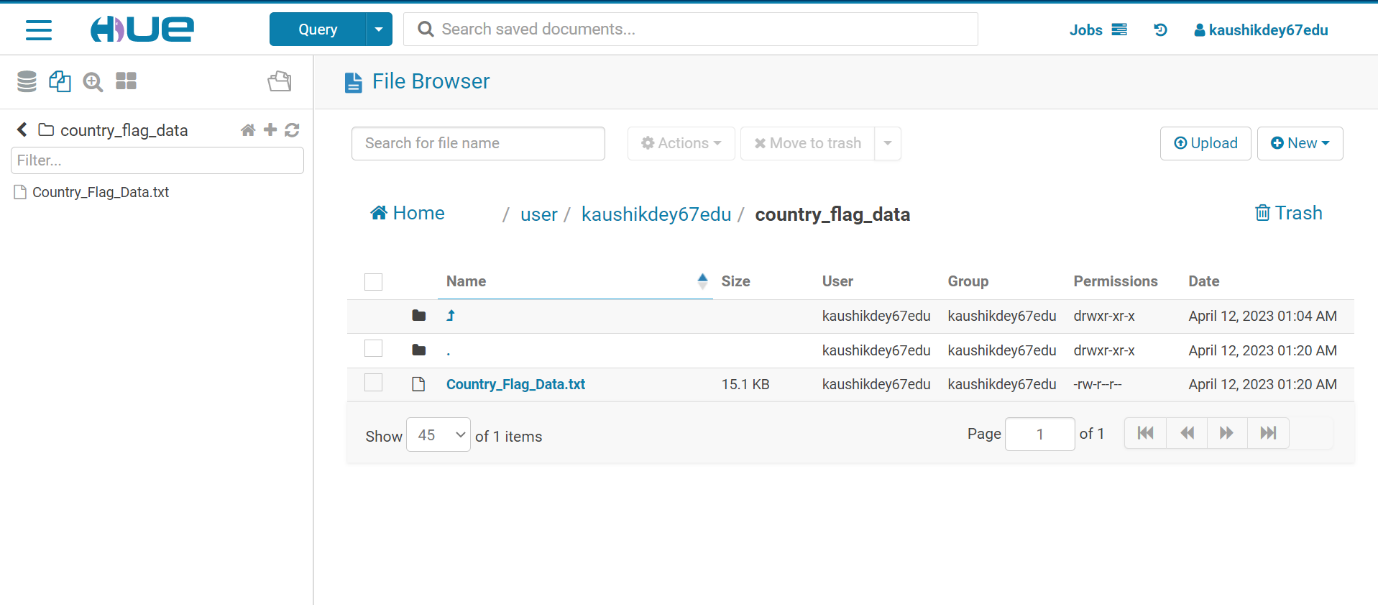
   Description automatically generatedLogin into Big Data Lab

Kaushik Dey

Table of Contents

1. **Count Number of countries based on landmass1 to 3.**
2. **Find out top 5 country with Sum of bars and strips in a flag 4 to 5.**
3. **Count of countries with icon**.  **5 to 6.**
4. **Count of countries which have same top left and top right color in flag**.  **6 to 7.**
5. **Count number of countries based on zone**. ……………………………………………………………….. **8 to 9.**

**Count Number of countries based on Landmass.**

1. **First we have to upload the dataset into hdfs. The following screenshot of Country\_Flag\_Data.text inside hue console.**

Text

Description automatically generated

1. **Now we have to write the pig script in a batch mode, the following code is given below.**

**count\_number\_countries.pig**

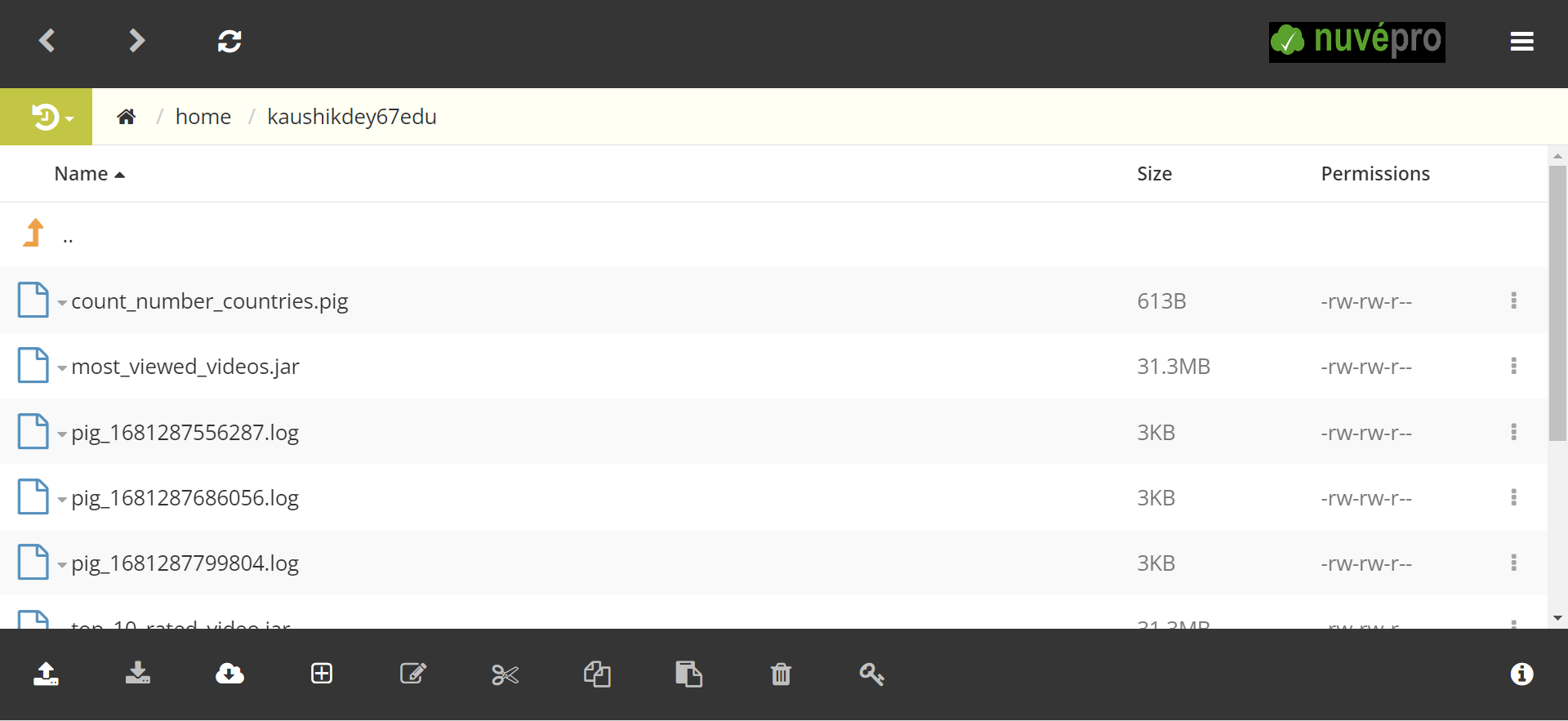
**country\_data = LOAD '/user/kaushikdey67edu/country\_flag\_data/Country\_Flag\_Data.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:chararray, population:chararray, lang:int, rel:int, bar:int, stripes:int, col:int, red:int, green:int, blue:int, gold:int, white:int, black:int, 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;**

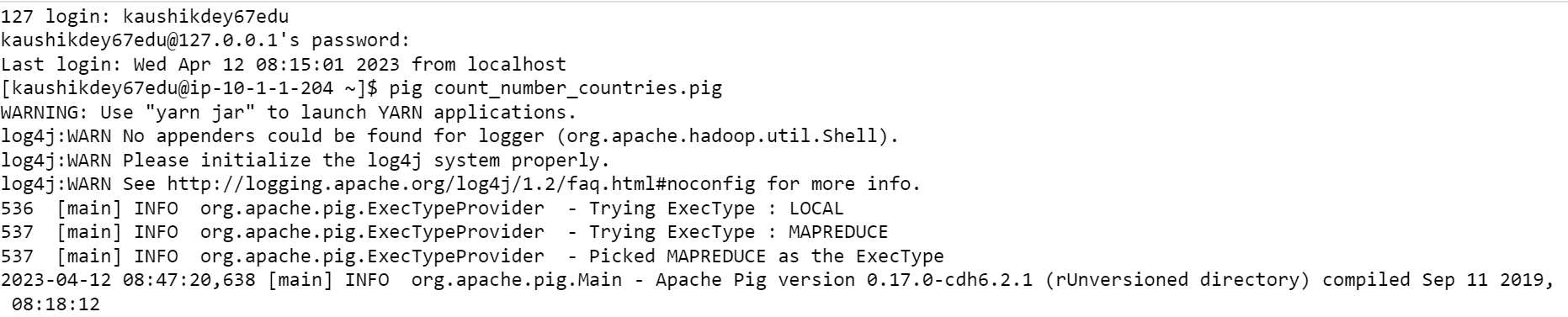
**count\_data = FOREACH grouped\_data GENERATE group,COUNT(country\_data);**

**DUMP count\_data;**

1. **Now we have to upload the script via FTP editor. The following screenshot is given below.**



1. **Now run the pig script via web-shell. The following Screenshot is given below.**



1. **Now we got the map reduce job with following details.**

Text

Description automatically generated

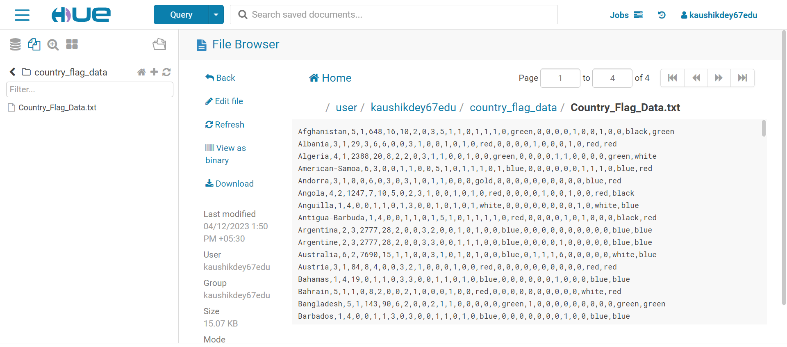
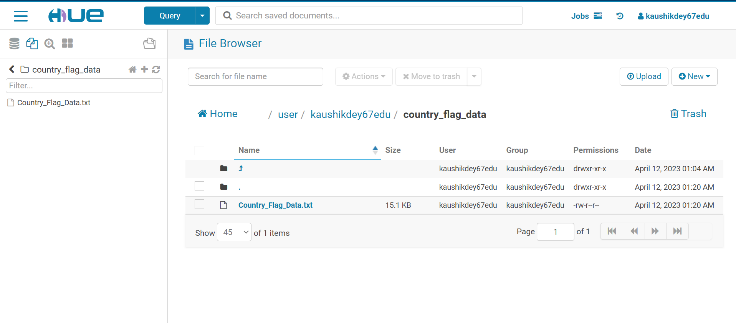
1. **Now we have to check the DAG where Job is executed and is it Succeeded or Failed.**

Graphical user interface, text, application, email

Description automatically generated

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

1. **First we have to upload the dataset into hdfs. The following screenshot of Country\_Flag\_Data.text inside hue console.**



1. **Now we have to write the pig script in a batch mode, the following code is given below.**

**bar\_types\_countries.pig**

bar\_type\_data = LOAD '/user/kaushikdey67edu/country\_flag\_data/Country\_Flag\_Data.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 bar\_type\_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;

1. **Now we have to upload the script via FTP editor. The following screenshot is given below.**
2. **Now run the pig script via web-shell. The following Screenshot is given below.**

Text

Description automatically generated

1. A picture containing graphical user interface

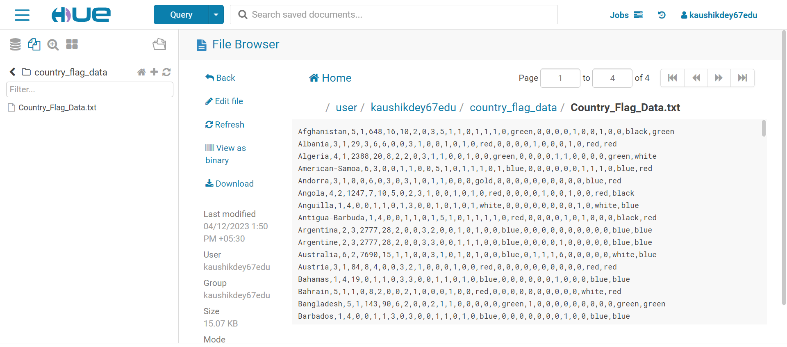
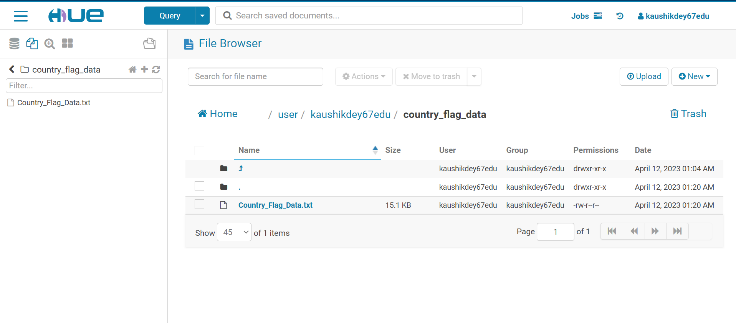
   Description automatically generated**Now we got the map reduce job with following details.**
2. **Now we have to check the DAG where Job is executed and is it Succeeded or Failed.**

Graphical user interface, text

Description automatically generated

**Count of countries with icon**

1. **First we have to upload the dataset into hdfs. The following screenshot of Country\_Flag\_Data.text inside hue console.**



1. **Now we have to write the pig script in a batch mode, the following code is given below.**

**count\_of\_countries.pig**

country\_data = LOAD '/user/kaushikdey67edu/country\_flag\_data/Country\_Flag\_Data.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;

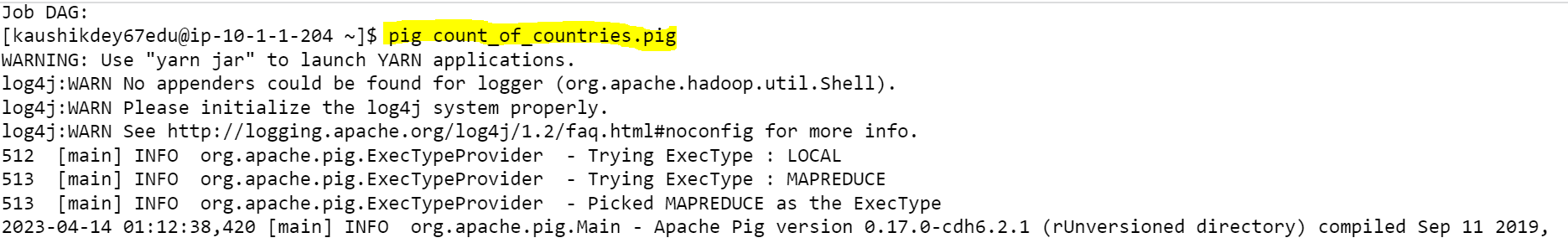
count\_icon\_data = FOREACH grouped\_icon\_data GENERATE group, COUNT(filtered\_icon\_data);

DUMP count\_icon\_data;

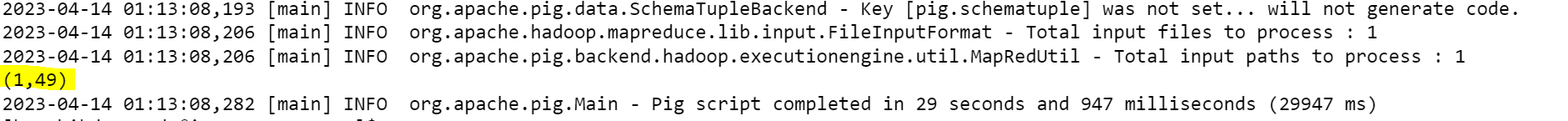
1. **Now we have to upload the script via FTP editor. The following screenshot is given below.**



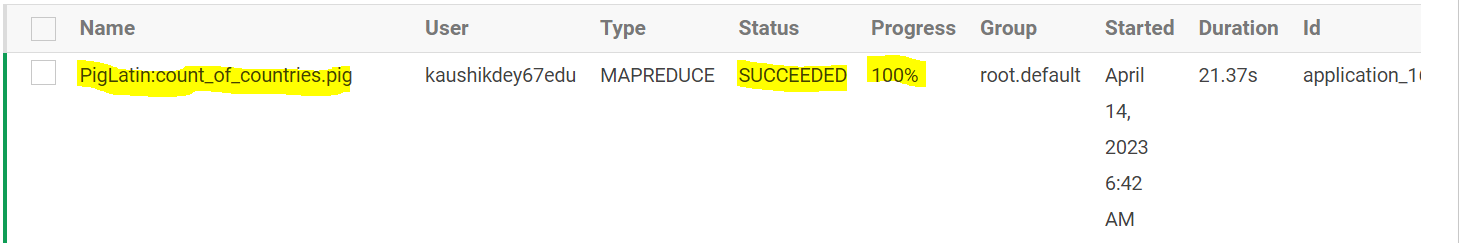
1. **Now run the pig script via web-shell. The following Screenshot is given below.**



1. **Now we got the map reduce job with following details**

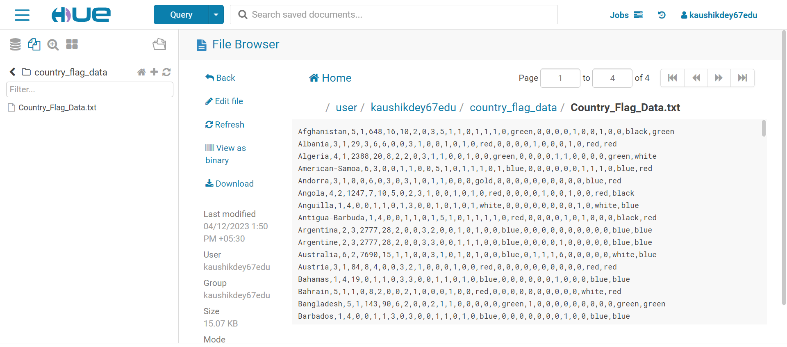
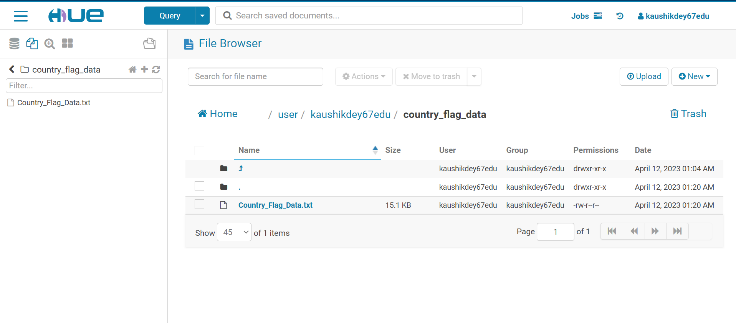


1. **Now we have to check the DAG where Job is executed and is it Succeeded or Failed.**



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

1. **First we have to upload the dataset into hdfs. The following screenshot of Country\_Flag\_Data.text inside hue console**



1. **Now we have to write the pig script in a batch mode, the following code is given below.**

**count\_of\_countries\_top\_left\_right.pig**

country\_data = LOAD '/user/kaushikdey67edu/country\_flag\_data/Country\_Flag\_Data.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;

1. **Now we have to upload the script via FTP editor. The following screenshot is given below.**

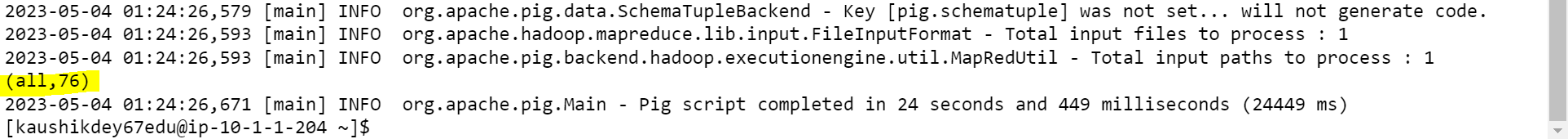


1. **Now run the pig script via web-shell. The following Screenshot is given below.**

Text

Description automatically generated with medium confidence

1. **Now we got the map reduce job with following details.**

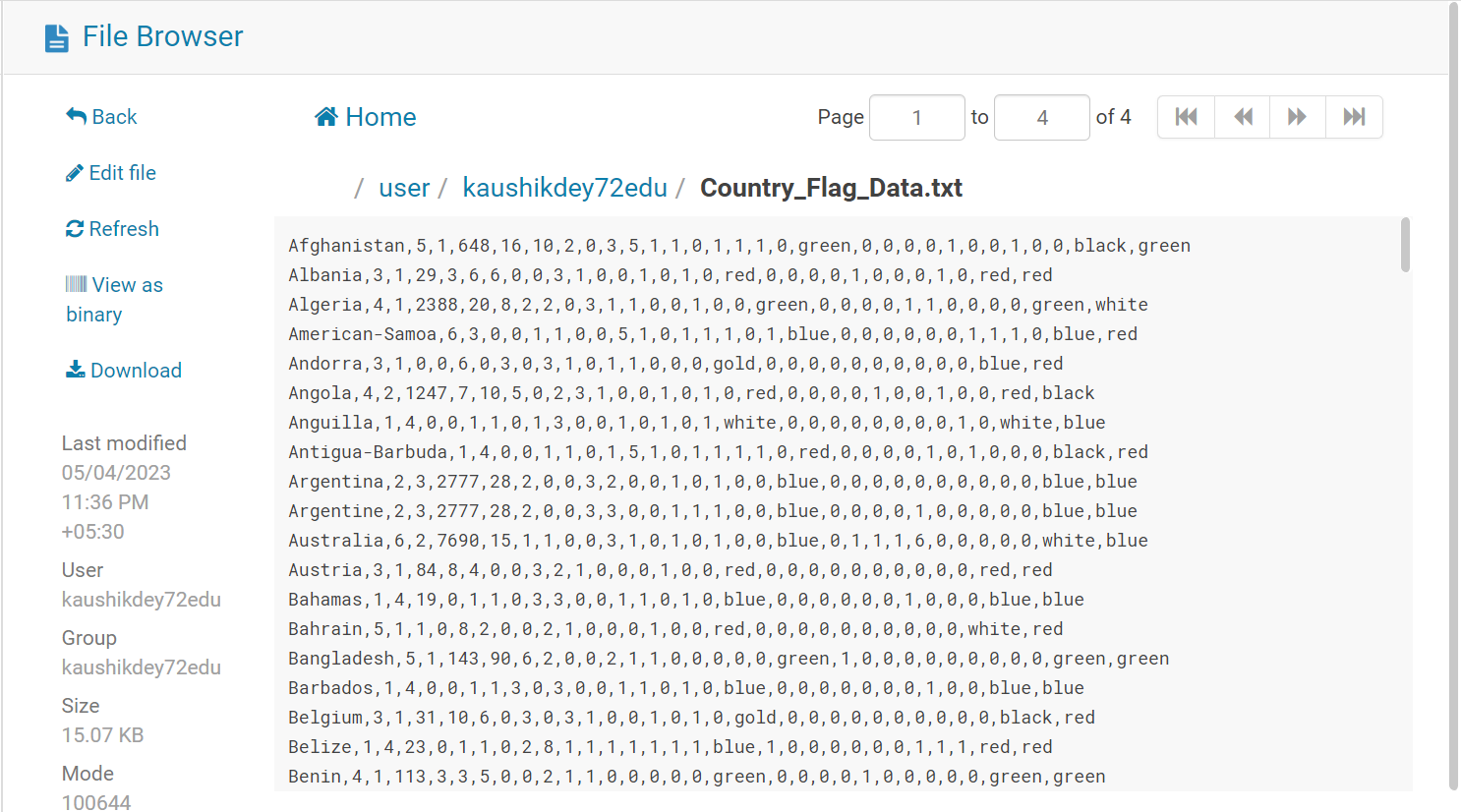


1. **Now we have to check the DAG where Job is executed and is it Succeeded or Failed.**

Graphical user interface, text, email

Description automatically generated

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

1. **First we have to upload the dataset into hdfs. The following screenshot of Country\_Flag\_Data.text inside hue console.**
2. **Now we have to write the pig script in a batch mode, the following code is given below.**

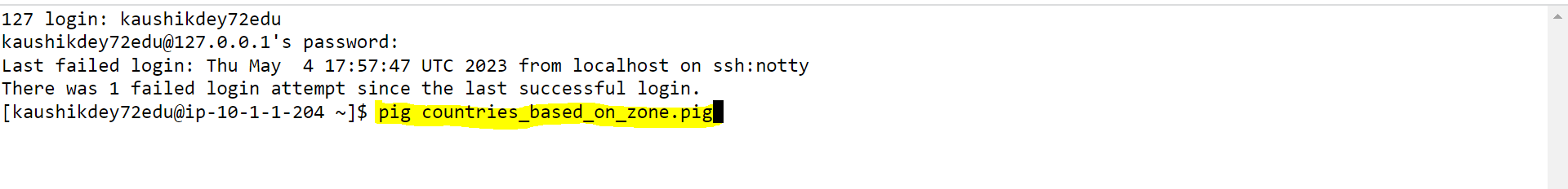
country\_data = LOAD '/user/kaushikdey72edu/Country\_Flag\_Data.txt' USING PigStorage(',') as (name:chararray, landmass:int, zone:int, area:chararray, population:chararray, lang:int, rel:int, bar:int, stripes:int, col:int, red:int, green:int, blue:int, gold:int, white:int, black:int, orange:int, hue:chararray, 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;

1. Graphical user interface

   Description automatically generated with medium confidence**Now we have to upload the script via FTP editor. The following screenshot is given below.**
2. **Now run the pig script via web-shell. The following Screenshot is given below.**
3. Text

   Description automatically generated**Now we got the map reduce job with following details.**

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