Module 5: PIG

**Assignment** – Find out successful students.

1. Student Results :
2. Create a pig script file name it as student\_results.pig :

student = LOAD 'student' using PigStorage('\t') AS (student\_name:chararray, student\_roll:int);

results = LOAD 'results' using PigStorage('\t') AS (student\_roll:int, student\_result:chararray);

joined\_data = join student by student\_roll , results by student\_roll;

filtered\_results = FILTER joined\_data by student\_result matches 'pass';

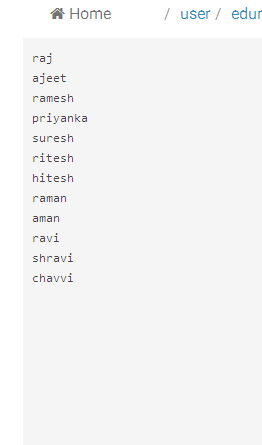
final\_results = FOREACH filtered\_results GENERATE student\_name;

store final\_results into 'outputResults1' using PigStorage(',');

1. Upload pig script file in FTP server.
2. Upload data set on HUE.
3. Log in to Webconsole, enter into pig environment and execute the following command :

Exec student\_results.pig

1. Output :



1. Execute Data-set and Pig Script for Health Care Use-Case :
2. Write java code for processing
3. Export the code into a jar file decript.jar
4. Create a pig script file name it as deidentify\_script.pig

REGISTER decript.jar;

A = LOAD 'healthcare\_Sample\_dataset2.csv' using PigStorage(',') AS (PatientID: int, Name: chararray, DOB: chararray, PhoneNumber: chararray, EmailAddress: chararray, SSN: chararray, Gender: chararray, Disease: chararray, weight: float);

B = LOAD 'healthcare\_Sample\_dataset1.csv' using PigStorage(',') AS (PatientID: int, Name: chararray, DOB: chararray, PhoneNumber: chararray, EmailAddress: chararray, SSN: chararray, Gender: chararray, Disease: chararray, weight: float);

C = UNION A, B;

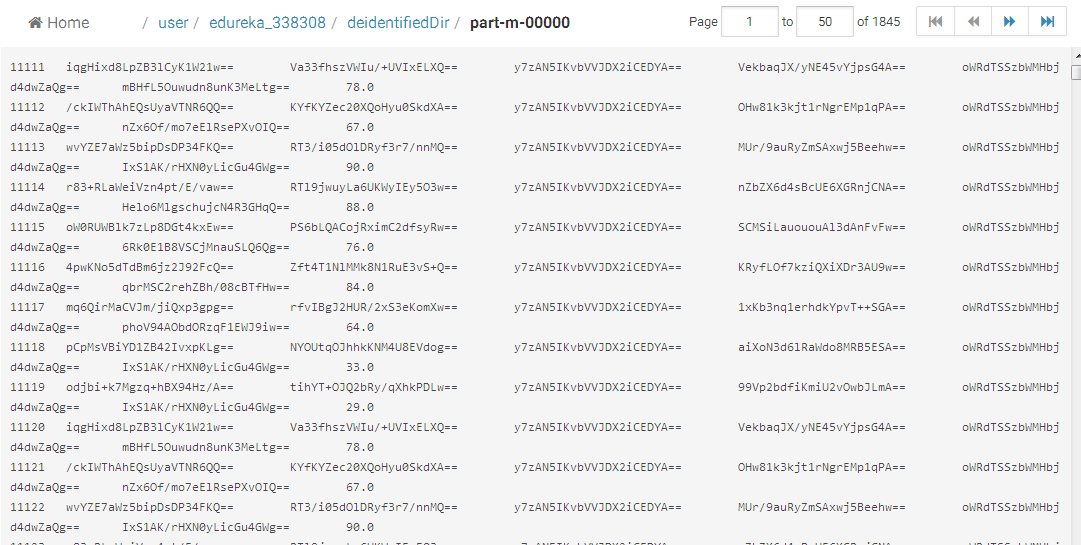
D = FOREACH C GENERATE PatientID, DeIdentifyUDF(Name,'12345678abcdefgh'), DeIdentifyUDF(DOB,'12345678abcdefgh'), DeIdentifyUDF(PhoneNumber,'12345678abcdefgh'), DeIdentifyUDF(EmailAddress,'12345678abcdefgh'),DeIdentifyUDF(SSN,'12345678abcdefgh'), DeIdentifyUDF(Disease,'12345678abcdefgh'),weight;

STORE D into 'deidentifiedDir';

1. Upload jar file, pig script file in FTP server.
2. Upload data set on HUE.
3. Log in to Webconsole, enter into pig environment and execute the following command :

Exec decript.jar

1. Output will give encrypted version of data :



1. Execute Data-set and Pig Script for Weather Use Case :
2. Write java code for processing
3. Export the code into a jar file udf\_corrupt.jar
4. Create a pig script file name it as weatherPIG\_script.pig :

Register udf\_corrupt.jar;

A = load 'weatherPIG.txt' using TextLoader as (data:chararray);

AF = foreach A generate TRIM(SUBSTRING(data, 6, 14)), IfCorrupted(TRIM(SUBSTRING(data, 46, 53))), IfCorrupted(TRIM(SUBSTRING(data, 38, 45)));

store AF into 'outputweatherData' using PigStorage(',');

1. Upload jar file, pig script file in FTP server.
2. Upload data set on HUE.
3. Log in to Webconsole, enter into pig environment and execute the following command :

Exec weatherPIG\_script.pig.pig

1. Output :

