Assignment 6:

1.

assignment6 = LOAD '/inputs/mr\_inputs/assignment6' using PigStorage(',') AS (productid: long, name, category, price: double, location);

dump assignment6;

(101,comb,commercial,12.0,chennai)

(102,powder,commercial,34.0,chennai)

(103,watermelon,fruits,29.0,chennai)

(104,apple,fruits,35.0,chennai)

(105,carrot,veg,32.0,chennai)

transform = FOREACH assignment6 GENERATE $0,$3

dump transform;

2017-05-14 02:09:19,197 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2017-05-14 02:09:19,200 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

(101,12.0)

(102,34.0)

(103,29.0)

(104,35.0)

(105,32.0)

Please find the jar attached for the UDF

Steps:

REGISTER /home/hduser/PigUDF.jar

DEFINE CONCAT UDF.PigUDF;

A = LOAD '/inputs/mr\_inputs/assignment6' Using PigStorage(',');

dump A;

(101,comb,commercial,12,chennai)

(102,powder,commercial,34,chennai)

(103,watermelon,fruits,29,chennai)

(104,apple,fruits,35,chennai)

(105,carrot,veg,32,chennai)

B = FOREACH A GENERATE $1,$2,CONCAT(\*),$4;

dump B;

(comb,commercial,10112,chennai)

(powder,commercial,10234,chennai)

(watermelon,fruits,10329,chennai)

(apple,fruits,10435,chennai)

(carrot,veg,10532,chennai)

2.

counter = group assignment6 BY $0;

counter\_result = foreach counter generate COUNT(assignment6);

c1 = group counter\_result ALL;

c2 = foreach c1 generate SUM(counter\_result.$0);

dump c2

(5)

3. Please find attached code Pig-UDF1

4.

A = LOAD '/inputs/mr\_inputs/exercise\_4\_data.txt' Using PigStorage(',');

B = FILTER A BY ($1%2)==0;

C = FILTER A BY ($1%2)!=0;

STORE B INTO 'hdfs://localhost:9000/input/mr\_output6' Using PigStorage(',');

STORE C INTO 'hdfs://localhost:9000/input/mr\_output7' Using PigStorage(',');

D = FILTER A BY org.apache.pig.piggybank.evaluation.IsNumeric($1);

STORE C INTO 'hdfs://localhost:9000/input/mr\_output8' Using PigStorage(',');