**Lab 7: Pig**

1. Work the “max-temperature” example. Write down the commands in each step and print out a screenshot for the result of the example in CSUEB Hadoop.

records = LOAD 'pig\_1/sample.txt'

AS (year:chararray, temperature:int, quality:int);

-- ^^ max\_temp\_load

-- vv max\_temp\_dump\_records

DUMP records;

-- ^^ max\_temp\_dump\_records

-- vv max\_temp\_describe\_records

DESCRIBE records;

-- ^^ max\_temp\_describe\_records

-- vv max\_temp\_filter\_records

filtered\_records = FILTER records BY temperature != 9999 AND

quality IN (0, 1, 4, 5, 9);

DUMP filtered\_records;

-- ^^ max\_temp\_filter\_records

-- vv max\_temp\_dump\_grouped\_records

grouped\_records = GROUP filtered\_records BY year;

DUMP grouped\_records;

-- ^^ max\_temp\_dump\_grouped\_records

DESCRIBE grouped\_records;

-- vv max\_temp\_max\_temp

max\_temp = FOREACH grouped\_records GENERATE group,

MAX(filtered\_records.temperature);

-- ^^ max\_temp\_max\_temp

-- vv max\_temp\_result

DUMP max\_temp;



2. Work the “Macros” example. Write down the commands in each step and print out screenshots for the result of the example in CSUEB Hadoop.

DEFINE max\_by\_group(X, group\_key, max\_field) RETURNS Y {

A = GROUP $X by $group\_key;

$Y = FOREACH A GENERATE group, MAX($X.$max\_field);

};

records = LOAD ‘/home/22student22/inputpig/sample.txt ‘ AS (year:chararray, temperature:int, quality:int);

filtered\_records = FILTER records BY temperature != 9999 AND

quality IN (0, 1, 4, 5, 9);

max\_temp = max\_by\_group(filtered\_records, year, temperature);

DUMP max\_temp

Text

Description automatically generated

3. Work “Trim” JAVA based UDF example. Type out all the commands in each step of the process and print out a screenshot of the final results in CSUEB Hadoop.

A = LOAD '/home/student22/A' AS (fruit:chararray);

DUMP A;

DESCRIBE A;

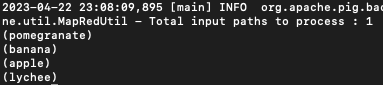
sh javac -classpath /home/student22/hadoop-common-2.6.1.jar:/home/student22/hadoop-mapreduce-client-core-2.6.1.jar:/home/student22/commons-cli-2.0.jar:/home/student22/pig-0.11.0.jar -d . Trim.java

sh jar -cvf pig-trim.jar com/hadoopbook/pig/Trim.class

REGISTER pig-trim.jar;

B = FOREACH A GENERATE com.hadoopbook.pig.Trim(fruit);

DUMP B;



4. Create and Work “Strip” python based UDF example. Type out all the commands in each step of the process and print out a screenshot of the final results in CSUEB Hadoop.

pig -x local

A = LOAD '/home/student22/A' AS (fruit:chararray);

REGISTER 'PythonStrip.py' USING streaming\_python AS my\_strip;

relation = FOREACH A GENERATE my\_strip.strip\_function(fruit);

dump relation;

Text

Description automatically generated