Homework Assignment 4

1. Work the Pig examples “Cross”, “Group” and “Sort” from Chapter 11 using data “hadoop-book-master/input/pig/join/A”, “hadoop-book-master/input/pig/group/A”, and “hadoop-book-master/input/pig/sort/A” respectively. Type out all the commands in each step of the process and **Print** out a couple of screenshots for the results of these examples in CSUEB Hadoop.
2. Create the “isGoodQuanlity.py” python file demonstrated in class. Work this python UDF on the sample file “sample\_badrecords.txt’ in Pig. Type out all the commands in each step of the process and print out a screenshot of the final results in CSUEB Hadoop.
3. Work Java based UDF “isGoodQuality” (/hadoop-book-master/ch16-pig/src/main/java/com/hadoopbook/pig/IsGoodQuality.java) from example 11-1 in Pig. Type out all the commands in each step of the process and print out a screenshot of the final results in CSUEB Hadoop.
4. Work the Hive example “Multitable insert”. Type out all the commands in each step of the process and print out a screenshot of the final results in CSUEB Hadoop.

The codes to create table records2 are as follows:

DROP TABLE IF exists records2;

CREATE TABLE records2 (station STRING, year STRING, temperature INT, quality INT)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t';

LOAD DATA LOCAL INPATH '/your/path/to /sample2.txt'

OVERWRITE INTO TABLE records2;

The sample2.txt can be found in Hadoop-Book-Master/input/ncdc/micro-tab/sample2.txt.

1. Work the “hiveudf.py” on “hivesample2.txt”. Type out all the commands in each step of the process and print out a screenshot of the final results in CSUEB Hadoop.