HOMEWORK ASSIGNMENT: WEEK 3

Assignment 01:

1. Create a simple text file having 5 fields per records with ',' as delimiter(You are free to choose your data in those fields) and save it in /inputs/mr_inputs hdfs directory, also

   Write

   - A Pig Script to Load this file
   - A Pig Script to Select first and 4th field
   - An UDF to create another field which is the concatenation of first and 4th field

2. Write a Pig Program to count the number of lines in your input file (again, you are free to choose your input file).

3. For the given file (exercise_3_data) as Input, write a Filter UDF which removes a record if the second field or third field is NOT a number.

4. For the given data set of exercise 4 (exercise_4_data):

   Load the data

   Create three data sets from this original data set, with first dataset having second field as even number and the second data set having the second field as odd numbers and the third data set having the second field as non-number.