RDD assignments

1. Create RDDs in three different ways.
2. Read a text file and count the number of words in the file using RDD operations.
3. Write a program to find the word frequency in a given file.
4. Write a program to convert all words in a file to uppercase.
5. Write a program to convert all words in a file to lowercase.
6. Write a program to capitalize first letter of each words in file (use string `capitalize()` method).
7. Find the longest length of word from given set of words.
8. Map the Registration numbers to corresponding branch. 6000 series BDA, 9000 series HDA, 1000 series ML, 2000 series VLSI, 3000 series ES, 4000 series MSc, 5000 series CC. Given registration number, generate a key-value pair of Registration Number and Corresponding Branch.
9. Text file contain numbers. Numbers are separated by one white space. There is no order to store the numbers. One line may contain one or more numbers. Find the maximum, minimum, sum and mean of numbers.
10. A text file (citizen.txt) contains data about citizens of country. Fields (information in file) are Name, dob, Phone, email and state name. Another file contains mapping of state names to state code like Karnataka is codes as KA, TamilNadu as TN, Kerala KL etc. Compress the citizen.txt file by changing full state name to state code. Write compressed data back into new file.

Create dataset (text file) with fields like 'Student Name', 'Institute', 'Program Name', and 'Gender' and solve following questions.

1. Compute number of students from each Institute.
2. Number of students enrolled to any program.
3. Number of 'boy' and 'girl' students.
4. Number of 'boy' and 'girl' students from selected Institute.

Dataset: Temperature of Indian Cities. Fields of dataset are Date, Average Temperature, City, Country, Latitude and Longitude (Use dataset attached to MapReduce assignment). Solve following questions

1. Find maximum and minimum temperature of all cities from the given dataset

2. Count number of data point for each city.

3. Find the maximum and minimum temperature for city **Bangalore** from the given dataset.

4. Find the maximum and minimum temperature for any given city from the given dataset. City name should be passed through command line argument.

Create dataset (text file) of bank transactions. Fields in file are 'Bank ID', 'Account Number', 'Transaction Date', 'Transaction Type' (credit or debit), 'Transaction Amount'. Date format is dd-mm-yyyy.

1. Count unique number of customers

2. Count unique number of Bank ID

3. Count unique number of customers per Bank ID

4. Number of transactions for given Account Number

5. Number of credit transactions for given Account Number in a given year