Assignment 2

Distributed Systems

February 13, 2023
Due: 11:59 PM, March 7, 2023

1 Introduction

We will be using the Map-Reduce framework using Hadoop streaming. You are expected to implement Mapper and Runner components in any language, with the runner script that calls these in a language you would prefer (Python, C++, Java, or a bash script).

- **Map Reduce:** link
- **Hadoop Documentation :**link
- **References :** link
- **Installation:** link
- We recommend to use **docker** to install and use Hadoop and its components. This tutorial might be of help in setting up that.

You are required to do **any TWO out of THREE** given questions

2 Problems

You will be given an input file which you can change in your runner script the way your mapper needs it. Runner script would be taking five arguments that are Hadoop-stream jar file address, input file address (local one), input directory address(where to copy local input files in HDFS), output directory address(where to store the output files in HDFS), address of the directory where all mappers and reducers are located. If you do not need multiple mappers or reducers, you still need to take directory paths, and they will terminate with '/'. In the runner script you need to use the copyFromLocal command to copy the input. In single cluster mode (default and the one you might use on your system), you won't have to explicitly copy the input files or the code files to HDFS, but during the evals your scripts would run on a cluster of multiple nodes which won't work without copying from the local file-system to HDFS.

2.1 NATIONAL PARTY OR NOT

A regional party has its candidates in S number of states. It wants to expand its operations to become a national party. However, to become a national party, it has to contest in general elections and satisfy any one of the below-mentioned criteria. **Criterion:**

1. The party wins 3 seats from at least three different states.
2. At a general election, the party polls 6% of votes in any four or more states, and in addition, it wins four seats (in any state/s)
3. The party gets recognition as a state party in four states.

Each state is divided into multiple election zones. You are given the election results of a party from various election zones. Determine if it is now a national party or not.

Assume if it gets more than P% votes out of all votes at any state, then it wins one seat.

Input :

First 3 lines have values of S, P, Total no.of Zones (Z) respectively.
Next Z lines have 5 space separated values Zone, State, Votes_Polled, Total_Votes , Recognised_as_State_party for each zone respectively.

Output :
YES if it satisfies else NO.

Example

Sample Input :

```
4
15
5
A A1 100 2000 NO
B A1 150 1500 NO
C F1 200 1800 YES
D D1 1500 2000 YES
E E1 1200 2400 NO
```

Sample Output :

```
YES
```

Explanation :

- It is recognized as a state party in only 2 states. Hence condition 3 is not satisfied.
- Percentage of votes in state A1 = 7.14%, in F1 = 11.11%, in G1 = 75% - won 5 seats and in K1 = 50% - won 3 seats.
- It won 8 seats, but only from 2 states hence condition 1 is not satisfied.
- It won more than 6% votes in all states and won 8 seats. Hence condition 2 is satisfied.

Clarification : Across zones of a same state the number of votes to be recognised as a national party will be the same. If number of seats obtained by a party in a state is not a whole number then take floor of that value.

2.2 MATRIX MULTIPLICATION

Given two matrices A of size $m * n$ and B of size $n * p$. Output the matrix multiplication of A and B.

Input: The first line of the input contains m and n followed by m lines of elements belonging to matrix A. Then it is followed by n and p, followed by n lines of elements belonging to matrix B.

Output: The output should contain lines of elements belonging to output $A * B$.

Constraints:

```
1 <= n, m, p <= 100
```

The values of the matrix are between $-1e5$ and $1e5$

Example:
Sample Input:

```
2 2
1 3
2 4
2 2
4 2
3 1
```

Sample Output:

```
13 5
20 8
```

2.3 EULER'S CONSTANT

We will now estimate the value of Euler's constant (e) using a Monte Carlo method. Let X_1, X_2, ..., X_n be an infinite sequence of independent random variables drawn from the uniform distribution on $[0, 1]$. Let V be the least number n such that the sum of the first n samples exceeds 1:

$$V = \min\{n | X_1 + X_2 + ... + X_n > 1\}$$

The expected value of V is e: $E(V) = e$.

Each Map task will generate random points using a uniform distribution on $[0, 1]$ in order to find a fixed number of values of n. It will output the number of time each value of n has been produced. The Reduce task will sum the results, and using them calculate the expected value of V and print the result.

3 Submission Instructions

Your submission is expected to be a <RollNumber>.zip file containing a directory with the same name as your roll number that holds the following files:

- A directory for each of the mentioned problems with the name: <RollNumber><ProblemNumber>
- A brief report describing and analysing your solution as: README.md

You can have multiple mappers and reducers. They should be named mapper.py, mapper2.py, and so on. Similarly for reducers too. (extension depends upon your language).

Example structure

```
2020121011
├── 2020121011_1
│   ├── mapper
│   ├── reducer
│   └── runner_script
├── 2020121011_2
│   ├── mapper
│   ├── reducer
│   └── runner_script
└── README.md
```

NOTE: Strict actions would be taken against anyone found involved in any kind of plagiarism either from the internet or from other students.