Assignment 2: Map Reduce

Distributed Systems

February 5, 2022
Due: 11:55 PM, February 21, 2022

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).

- Documentation: `https://hadoop.apache.org/docs/r1.2.1/streaming.html`
- References: `https://www.geeksforgeeks.org/what-is-hadoop-streaming/`

2 Problems

You will be given an input file which you can change in your runner script the way your mapper needs it.

2.1 Problem 1 (40 points)

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:
$1 <= n, m, p <= 100$

Example:

Sample Input:
2 2
1 3
2 4
2 2
4 2
3 1

Sample Output:
13 5
20 8

2.2 Problem 2 (60 points)

Find the connected components in the graph. You will be required to give the Nodes in each component in each new line. Component of size 1 is possible. Any reasonably fast algorithm is fine.

Input: The input file will contain edges only.

Output: The output should be a set of lines where each line contains the nodes in a connected component.

Constraints:
2 <= Number of Nodes <= 100,
1 <= Number of Edges <= 1000

Example:

Sample Input:
1 2
1 3
2 3
4 5

Sample Output:
1 2 3
4 5

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 analyzing your solution as: **README.md**

Example structure

```
2018101077
├── 2018101077_1
│   ├── mapper
│   ├── reducer
│   └── runner script
├── 2018101077_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.