Hadoop: MapReduce
Assignment - 4

Problem Statement:

You work as a Big Data Engineer at GrapeVine Pvt. Ltd. Your company is currently working on onomatology analytics using Hadoop. You have been assigned certain tasks for the fulfillment of this activity with the help of MapReduce.

Dataset: :

https://drive.google.com/file/d/1xjGPmdB3UsV2Faspffd2mNON87PXJKFI/view?usp=sharing

Dataset Description:

Use the 'Baby Name Data' dataset to get child naming statistics in the United States throughout the years (1800–present). The following are the field descriptions:

- Id – Serial number
- Name
- Year – The year of name usage
- Gender – M/F
- Count – The frequency of name usage in that particular year

Tasks To Be Performed:

1. Create an output with all the names starting with the letter, 'G'
2. Create an output with all the names having only two vowels
3. Create an output with names without any vowels, if any
4. Create an output with names containing the syllable, 'red'
5. Create an output with the total count of each name, throughout all years
6. Find the most popular female name in 1918
7. Find the most popular male name in 1945