Hadoop: Spark Assignment - 1

Problem Statement:

You work as a Big Data Engineer at GrapeVine Pvt. Ltd. Your company is currently working on English language analytics using Hadoop. Having performed this task already, the company requires you to increase the computational efficiency using Apache Spark. You have been assigned certain tasks for the fulfillment of this activity.

Dataset: :

https://intellipaat-course-attachments.s3.ap-south1.amazonaws.com/Hadoop/hadoop_dataset.rar

Tasks To Be Performed:

1. Find out the count of each word in the ‘Shakespeare.txt’ dataset.
2. Display the most commonly used words (words with the count over 100 are considered common).
3. Display the words that are rarely used (words with the count below 30 are considered rare).
4. Display the most commonly used word.
5. Display the least used word.