Module 3: Hadoop MapReduce Framework - I

Assignment

Version 2.0

Module 3: Hadoop MapReduce Framework - I

Assignment – Calculate the size of a word and the number of words of that size in a text file

Table of Contents

Introduction

Apply your MapReduce programming knowledge and write a MapReduce program to process two text files. You need to calculate the size of each word and count the number of words of that size in the text file.

The dataset for this problem is the text file **'alphabets'** available in your LMS.

Problem statement

Let's understand the problem through a sample text file content:

```
"Hello everyone this is a sample dataset. Calculate the word
size and count the number of words of that size in this text
file."
```

Your MapReduce program should process this text file and should provide output as follows:

Sample Output

Word Size	Word Count
1	**1** (As the word of size 1 is: a)
2	**4** (As the words of size 2 are: is, of, of, in)
3	**3** (As the words of size 3 are: the, and, the)
4	**6** (As the words of size 4 are: this, word, size, that, size)

Your task in this assignment is to process the 'alphabets' text file using MapReduce program. Your program should calculate the size of each word (Number of alphabets in the word) and count the number of words of that size in this text file.

You can review the solution in your LMS.