**Big Data Frameworks CSE3120**

Lab – 7 Spark Word Count Program

**Name:** Naveen Nidadavolu

**Roll No:** 22MIA1049

**Aim**

To implement a Word Count program using Apache Spark to count the occurrences of words in a given text file.

**Algorithm**

1. **Start**
2. Initialize the Spark environment.
3. Load the input text file into an RDD (Resilient Distributed Dataset).
4. Split each line of the text file into individual words.
5. Map each word to a key-value pair in the form of (word, 1).
6. Use the **reduceByKey()** function to sum up the counts of each word.
7. Collect the final word count results.
8. Display the results.
9. **End**

**Procedure**

1. **Create a sample text file named "Ex1" and save it on the desktop.**

A screenshot of a computer

AI-generated content may be incorrect.

1. **. Give the command: spark-shell, to check the spark framework version**

A screenshot of a computer

AI-generated content may be incorrect.

Apache Spark is an open source, in-memory distributed computing engine created to address the problem of processing large datasets for data analytics and machine learning. Spark is written in Scala and it's native integration with Spark APIs

1. **Execute Word Count in Spark**

* Load the text file Ex1 into Spark
* Split the lines into words, map each word to a key-value pair, and count occurrences
* Display the word count results

A computer screen with white text

AI-generated content may be incorrect.

1. **Check the Spark Web UI (local host) for monitoring job execution and results.**A screenshot of a computer

   AI-generated content may be incorrect.A screenshot of a computer

   AI-generated content may be incorrect.A screenshot of a computer

   AI-generated content may be incorrect.

**Results:**

The Word Count program in Spark successfully counts the occurrences of words in the given input file.