**Lab Assignment-2**

**Gatiganti Sai Venkatesh (Class Id :09)**

**Word Combinations**

Objective of this use case is to get most used combination of words in a given text file.

**Map Reduce** paradigm of the use case:

Shuffling

Final Result

Reducing

Sliding, mapping

Input File

Splitting

(text file,1)

(file for,1)

(for this,1)

(this use,1)

(use case,1)

(case only,1)

(This is,1)

(is a,1)

(a sample,1)

(sample text,1)

(text file,1)

text file for this use case only

This is a sample text file.

(text file,2)

(file for,1)

(for this,1)

(use case,1)

(case only,1)

(This is,1)

(this use,1)

(is a,1)

(a sample,1)

(sample text,1)

(text file,2)

(This is,1)

(this use,1)

(is a,1)

(a sample,1)

(sample text,1)

(file for,1)

(for this,1)

(use case,1)

(case only,1

(file for,1)

(for this,1)

(use case,1)

(case only,1)

(text file,1)

(text file,1)

(is a,1)

(a sample,1)

(sample text,1)

This is a sample text file. text file for this use case only

(This is,1)

(this use,1)

[Input](https://github.com/venkatesh-sg/Big-Data-Analytics-Lab-Assignments/blob/master/LabAssignment2/Source/WordCombination/About_UMKC): Input is a text file about UMKC.

[Output](https://github.com/venkatesh-sg/Big-Data-Analytics-Lab-Assignments/tree/master/LabAssignment2/Source/WordCombination/output): Combination of every two words used in a input text file in descending order.

[Spark Program for Word Combination](https://github.com/venkatesh-sg/Big-Data-Analytics-Lab-Assignments/blob/master/LabAssignment2/Source/WordCombination/src/main/scala/WordCombinations.scala): Both actions and transformations are used in the program.

**References:**

1)Tutorial lab2

2)[Sliding function](https://spark.apache.org/docs/1.3.1/api/java/org/apache/spark/mllib/rdd/RDDFunctions.html#sliding(int))