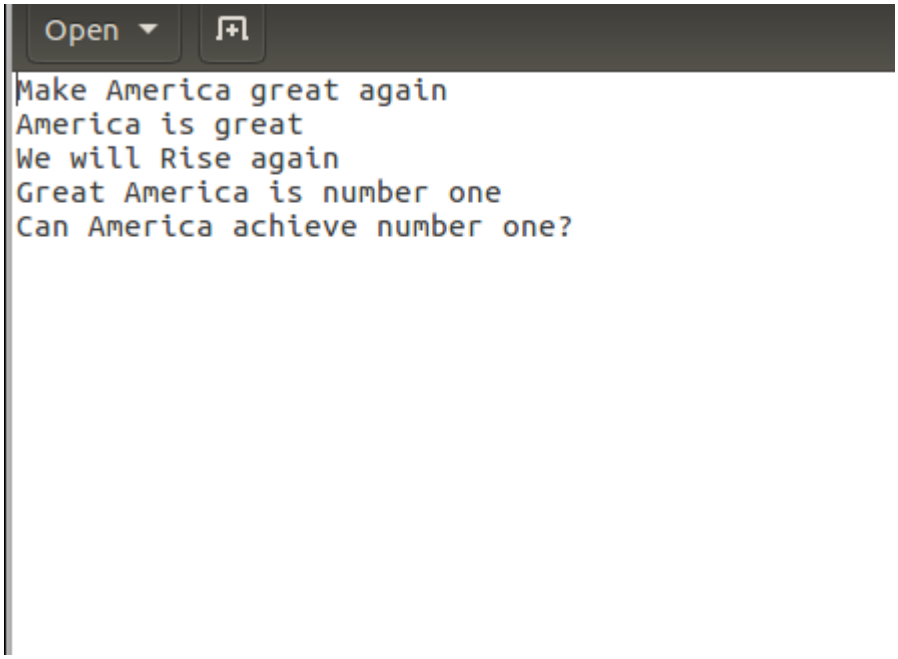
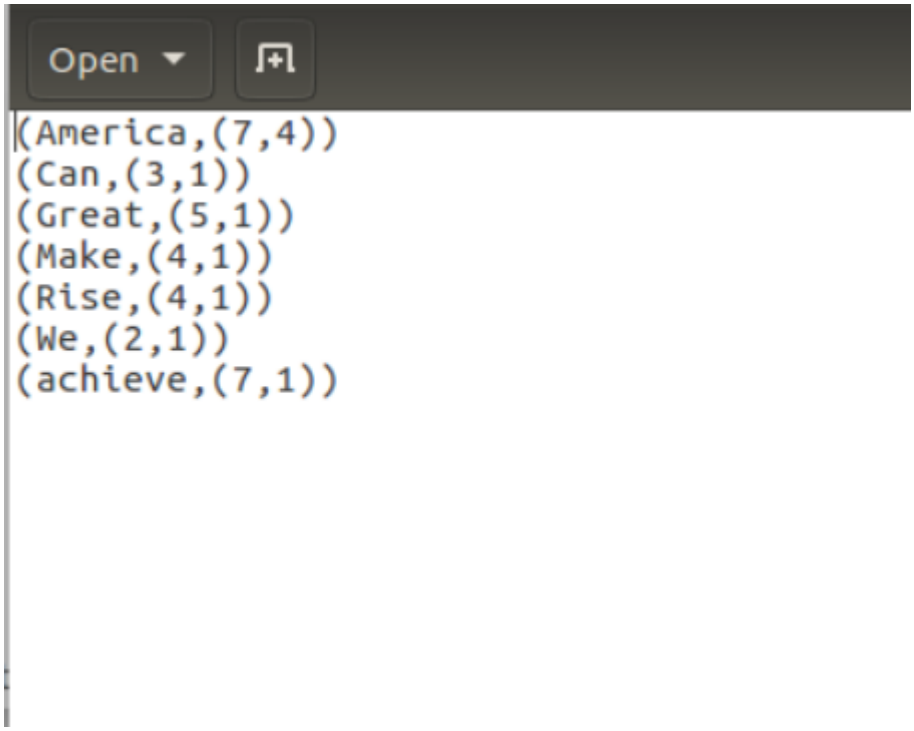
Lab 2 Report

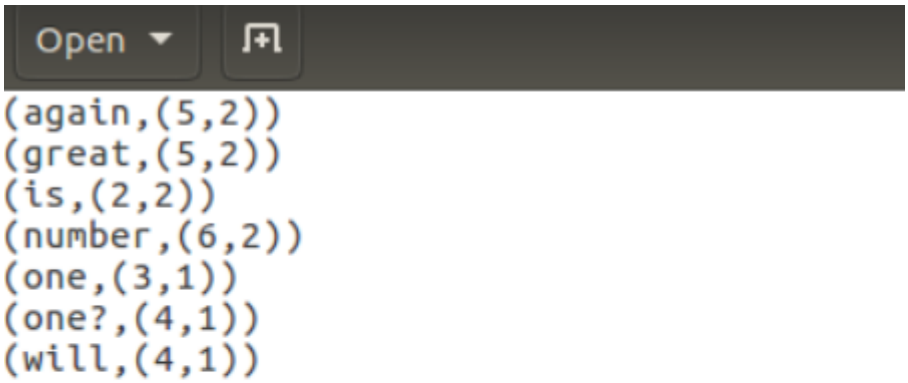
Input: Text Data as Input.



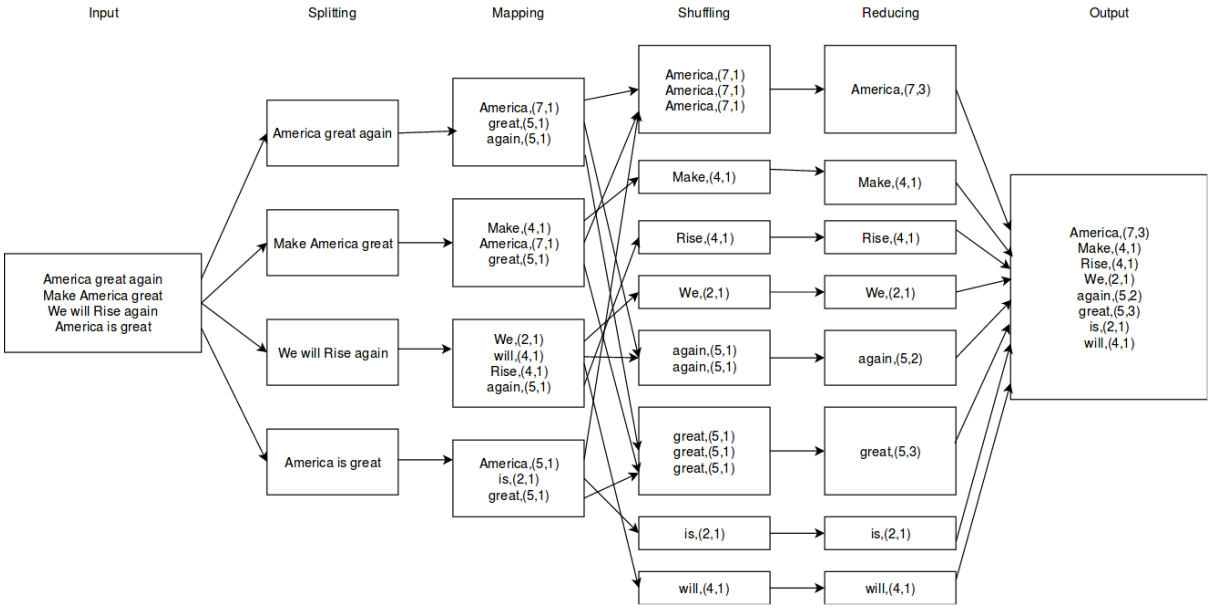
Word Count Output 1:



Output 2:



MapReduce Process:



Spark Program Code:

```
import org.apache.spark.{SparkConf, SparkContext}

object SparkSortWord {

  def main(args: Array[String]) {

     val sparkConf = new
SparkConf().setAppName("SparkSortWord").setMaster("local[*]")

     val sc = new SparkContext(sparkConf)

     val input = sc.textFile("inputtext")

     val wordsort=input.flatMap(line=>{line.split("
")}).map(word=>(word,word.size)).cache()      //Spark
Transformation 1 and 2

     val ignoreDuplicates=wordsort.reduceByKey((a,b)=>a)
//Spark Action 1

     val wordcount=input.flatMap(line=>{line.split("
")}).map(word=>(word,1)).cache()

     val output=wordcount.reduceByKey(_+_)

     val join=ignoreDuplicates.join(output).sortByKey()
//Spark Transformation 3 and 4

     join.saveAsTextFile("outputtext")
//Spark Action 2

   }

}
```