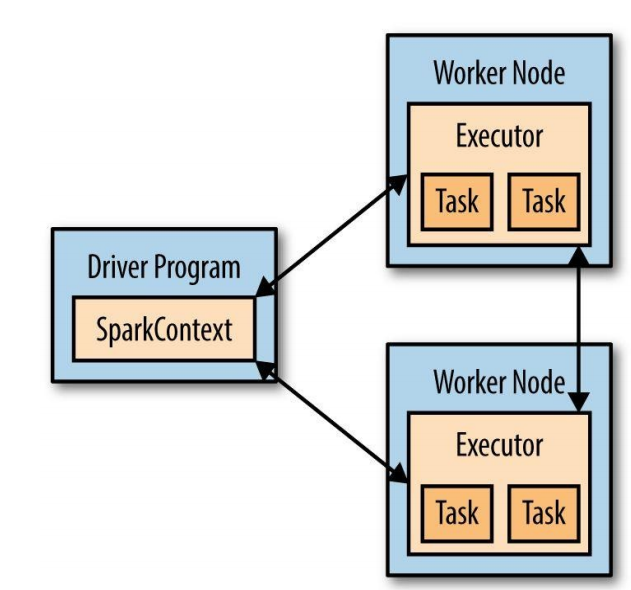
**Basic Definitions –**

Driver Program – Launches various parallel operations on the cluster. It defines distributed datasets and the actions on them.

Spark Context – Driver program connects to cluster using spark context object.



* We can package spark program using Maven or SBT to pass across various nodes.

Word Count – Scala

// Create a Scala Spark Context.

val conf = new SparkConf().setAppName(“wordCount”)

val sc = new SparkContext(conf)

// Load our input data.

val input = sc.textFile(inputFile)

// Split it up into words.

val words = input.flatMap(line => line.split(” “))

// Transform into pairs and count.

val counts = words.map(word => (word, 1)).reduceByKey{case (x, y) => x + y}

// Save the word count back out to a text file, causing evaluation.

counts.saveAsTextFile(outputFile)