

Apache Spark Word Count in Scala

WordCount.scala

```
import org.apache.spark.sql.SparkSession

object WordCount {
  def main(args: Array[String]): Unit = {
    val spark = SparkSession.builder
      .appName("Simple Word Count")
      .master("local[*]") // Use local mode with all cores
      .getOrCreate()

    val sc = spark.sparkContext

    // Read the input file
    val lines = sc.textFile("sample.txt")

    // Split lines into words, map each word to (word, 1), then reduce by key
    val wordCounts = lines
      .flatMap(line => line.split("\s+"))
      .map(word => (word.toLowerCase, 1))
      .reduceByKey(_ + _)

    // Collect and print the result
    wordCounts.collect().foreach { case (word, count) =>
      println(s"$word: $count")
    }

    spark.stop()
  }
}
```

Input File (sample.txt)

```
hello spark
hello scala
spark is fast
```

Expected Output

```
hello: 2
spark: 2
scala: 1
is: 1
fast: 1
```