1] Create a new project having a package called com.YOURNAME and inside that would be this class.

**package** com.evenkat;

**import** org.apache.spark.SparkConf;

**import** org.apache.spark.api.java.JavaPairRDD;

**import** org.apache.spark.api.java.JavaRDD;

**import** org.apache.spark.api.java.JavaSparkContext;

**import** org.apache.spark.api.java.function.FlatMapFunction;

**import** org.apache.spark.api.java.function.Function2;

**import** org.apache.spark.api.java.function.PairFunction;

**import** scala.Tuple2;

**import** java.util.Arrays;

**public** **class** WordCount {

**private** **static** **final** FlatMapFunction<String, String> *WORDS\_EXTRACTOR* =

**new** FlatMapFunction<String, String>() {

@Override

**public** Iterable<String> call(String s) **throws** Exception {

**return** Arrays.*asList*(s.split(" "));

}

};

**private** **static** **final** PairFunction<String, String, Integer> *WORDS\_MAPPER* =

**new** PairFunction<String, String, Integer>() {

@Override

**public** Tuple2<String, Integer> call(String s) **throws** Exception {

**return** **new** Tuple2<String, Integer>(s, 1);

}

};

**private** **static** **final** Function2<Integer, Integer, Integer> *WORDS\_REDUCER* =

**new** Function2<Integer, Integer, Integer>() {

@Override

**public** Integer call(Integer a, Integer b) **throws** Exception {

**return** a + b;

}

};

**public** **static** **void** main(String[] args) {

**if** (args.length < 1) {

System.*err*.println("Please provide the input file full path as argument");

System.*exit*(0);

}

SparkConf conf = **new** SparkConf().setAppName("com.evenkat.WordCount").setMaster("local");

JavaSparkContext context = **new** JavaSparkContext(conf);

JavaRDD<String> file = context.textFile(args[0]);

JavaRDD<String> words = file.flatMap(*WORDS\_EXTRACTOR*);

JavaPairRDD<String, Integer> pairs = words.mapToPair(*WORDS\_MAPPER*);

JavaPairRDD<String, Integer> counter = pairs.reduceByKey(*WORDS\_REDUCER*);

counter.saveAsTextFile(args[1]);

}

}

2] Add the spark-assembly-1.6.0-hadoop2.6.0.jar as External Jar file

3] Create a jar of your program and move it to Ubuntu

4] Create a file called words and add some content to it.

5] From the SparkInstall\bin location directory run the following:

spark-submit --class com.evenkat.WordCount --master local[2] /home/notroot/data/SparkHelloWorld.jar /home/notroot/data/words /home/notroot/data/wcount