import scala.Tuple2;

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

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

import org.apache.spark.sql.SparkSession;

import java.util.Arrays;

import java.util.List;

import java.util.regex.Pattern;

public final class JavaWordCount {

private static final Pattern SPACE = Pattern.compile(" ");

public static void main(String[] args) throws Exception {

if (args.length < 1) {

System.err.println("Usage: JavaWordCount <file>");

System.exit(1);

}

SparkSession spark = SparkSession

.builder()

.appName("JavaWordCount")

.getOrCreate();

JavaRDD<String> lines = spark.read().textFile(args[0]).javaRDD(); JavaRDD<String> words = lines.flatMap(s -> Arrays.asList(SPACE.split(s)).iterator());

JavaPairRDD<String, Integer> ones = words.mapToPair(s -> new Tuple2<>(s, 1));

JavaPairRDD<String, Integer> counts = ones.reduceByKey((i1, i2) -> i1 + i2);

List<Tuple2<String, Integer>> output = counts.collect();

for (Tuple2<?,?> tuple : output) {

System.out.println(tuple.\_1() + ": " + tuple.\_2());

}

spark.stop();

}

}