链滴

spark wordcount—IDEA

作者：hadoop

原文链接：https://ld246.com/article/1464161366786

来源网站：链滴

许可协议：署名-相同方式共享 4.0 国际 (CC BY-SA 4.0)

```
<pre>package com.admin;

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;
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);
        }

        SparkConf sparkConf = new SparkConf().setAppName("JavaWordCount");
        JavaSparkContext ctx = new JavaSparkContext(sparkConf);
        JavaRDD<String> lines = ctx.textFile(args[0], 1);

        JavaRDD<String> words = lines.flatMap(new FlatMapFunction<String, String>() {
            @Override
            public Iterable<String> call(String s) {
                return Arrays.asList(SPACE.split(s));
            }
        });

        JavaPairRDD<String, Integer> ones = words.mapToPair(new PairFunction<String, String,
nteger>() {
            @Override
            public Tuple2<String, Integer> call(String s) {
                return new Tuple2<String, Integer>(s, 1);
            }
        });

        JavaPairRDD<String, Integer> counts = ones.reduceByKey(new Function2<Integer, Integ
r, Integer>() {
            @Override
            public Integer call(Integer i1, Integer i2) {
                return i1 + i2;
            }
        });

        List<Tuple2<String, Integer>> output = counts.collect();
        for (Tuple2<?, ?> tuple : output) {
```

```
            System.out.println(tuple._1() + ": " + tuple._2());
        }
        ctx.stop();
    }
}</pre>
```