



博客 (<http://blog.csdn.net/?ref=toolbar>)

学院 (<http://edu.csdn.net?ref=toolbar>)

下载 (<http://download.csdn.net?ref=toolbar>)

GitChat (<http://gitbook.cn/?ref=csdn>) ...

0

写博客

发Chat

登录

注册

<https://passport.csdn.net/account/login?ref=toolbar>

<https://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister>

spark 自定义排序详解

原创 2015年12月03日 21:54:07

- 标签 : [spark](http://so.csdn.net/so/search/s.do?q=spark&t=blog) /
- [自定义排序](http://so.csdn.net/so/search/s.do?q=自定义排序&t=blog) /
- [排序](http://so.csdn.net/so/search/s.do?q=排序&t=blog) /
- [按key排序](http://so.csdn.net/so/search/s.do?q=按key排序&t=blog) /
- [按value排序](http://so.csdn.net/so/search/s.do?q=按value排序&t=blog)

1335

目的:
排序输出数据是数据处理的常见操作，本篇博客示例解决元组、字符串按照key、value进行单值、多值 自定义排序问题。

代码

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

```
[plain]
1. package LoadTest
2.
3. import org.apache.log4j.{Logger, Level}
4. import org.apache.spark.{SparkConf, SparkContext}
5.
6. def main (args: Array[String]){
7.
8.     //屏蔽不必要的日志显示在终端上
9.     Logger.getLogger("org.apache.spark").setLevel(Level.WARN)
10.    Logger.getLogger("org.eclipse.jetty.server").setLevel(Level.OFF)
11.
12.    // 设置运行环境
13.    val conf = new SparkConf().setAppName("CustomSortBy").setMaster("local")
14.    val sc = new SparkContext(conf)
15.
16.    //装载数据集
17.    //val data = sc.textFile("F:/HDFSinputfile/customSortBy")
18.    val data = List((1,4),(4,8),(0,4),(12,8))
19.    val rdd = sc.parallelize(data)
20.
21.    implicit val st = new Ordering[Int]{
22.        override def compare(a:Int,b:Int): Int ={
23.            a.toString.compare(b.toString)
24.        }
25.    }
26.    println("stage1")
27.    rdd.sortBy(x=>x._1).collect().foreach(println)
28.
29.    //stage2
30.    //val data2 = List("a 1","a 3","a 11","b 2","c 5")//切分字符串方式 自定义排序
31.    val data2=List(("a",1),("a",3),("a",11),("b",2),("c",5))
32.    val rdd2 = sc.parallelize(data2)
33.
34.    // 字符串切分后自定义排序
35.    implicit val st2 = new Ordering[String]{
36.        override def compare(a:String,b:String): Int ={
37.            val a1=a.split(" ")[0]
38.            val a2=a.split(" ")[1]
```

huozi07 (<http://blog.csdn.net/huozi07>)

+ 关注

原创

粉丝

喜欢

未开通

57

18

1

(<https://github.com>)



- 他的最新文章
- 更多文章 (<http://blog.csdn.net/huozi07>)
- word2vec_gensim

(<http://blog.csdn.net/huozi07/article/details/62499782>)
- 使用Genism进行词向量训练-实践版

(<http://blog.csdn.net/huozi07/article/details/51393934>)
- 使用Genism进行词向量训练-教程版

(<http://blog.csdn.net/huozi07/article/details/51393506>)
- 【python】Mapping Foreign direct in vestment, net inflows in 2014

(<http://blog.csdn.net/huozi07/article/details/5055293>)
- 【python】获取三维图形的某个截面

(<http://blog.csdn.net/huozi07/article/details/50544575>)

- 文章分类
- data mining/machine learni...

4篇
- Python

(<http://blog.csdn.net>)

4篇
- data visualization—d3.js/R (...)

11篇
- statistics

(<http://blog.csdn.net>)

1篇
- hadoop/spark

(<http://blog.csdn.net>)

13篇

```
39. //      val a2=a.split(" ")(1).toInt
40. //      val b1=b.split(" ")(0)
41. //      val b2=b.split(" ")(1).toInt
42. //      if(a1==b1) a2.compare(b2)
43. //      else a1.compare(b1)
44. //      }
45. //    }
46. //元组自定义排序
47. implicit val st3 = new Ordering[(String,Int)]{
48.   override def compare(a:(String,Int),b:(String,Int)): Int ={
49.     if(a._1==b._1) a._2.compare(b._2)
50.     else a._1.compare(b._1)
51.   }
52. }
53.
54. println("stage2")
55. rdd2.sortBy(x=>x).collect().foreach(println)
56.
57. sc.stop()
58.
59. }
60.
61. }
```

结果：



展开

文章存档

2017年3月 (http://blog.csdn.net/huozi07/article/details/50538749)	1篇
2016年5月 (http://blog.csdn.net/huozi07/article/details/50133259)	2篇
2016年1月 (http://blog.csdn.net/huozi07/article/details/48177079)	5篇
2015年12月 (http://blog.csdn.net/huozi07/article/details/45504465)	7篇
2015年11月 (http://blog.csdn.net/huozi07/article/details/45504465)	7篇

展开

他的热门文章

- 【python】一维二维插值 (<http://blog.csdn.net/huozi07/article/details/50538749>)
11214
- 基于GARCH模型的股市研究与危机预警——R语言实现 (<http://blog.csdn.net/huozi07/article/details/48176587>)
7225
- Spark的Transform与Action操作(import ant) (<http://blog.csdn.net/huozi07/article/details/50133259>)
4229
- 3D动态词云—JavaScript (<http://blog.csdn.net/huozi07/article/details/48177079>)
4101
- 解常微分方程组 python (<http://blog.csdn.net/huozi07/article/details/45504465>)
3601



手机排行榜



spark 自定义排序详解

huozi07 2015年12月03日 21:54 1335

目的： 排序输出数据是数据处理的常见操作，本篇博客示例解决元组、字符串按照key、value进行单值、多值 自定义排序问题。代码 package LoadTest im...
(<http://blog.csdn.net/huozi07/article/details/50166823>)

lucene4.3 自定义排序

yinxing0906404009 2013年07月31日 14:26 1458

今日在学习lucene，参考书籍是lucene in action,但因为lucene 版本更新很快，此书中很多地方因api的变动而使得我在测试代码例子的时候时常出现错误，而网上对lucene4.3的...
(<http://blog.csdn.net/yinxing0906404009/article/details/9667317>)

联系我们

- 网站客服 (<http://wpa.qq.com/msgrd?v=3&uin=2431299880&site=qq&nr>)
- 微博客服 (<http://e.weibo.com/csdnsupport/p>)
- webmaster@csdn.net (mailto:webmaster@csdn.net)
- 400-660-0108



霸气！吴恩达说：女儿识字后就教她学Python！

Python的火爆最近挡不住了，李飞飞表示人人都要会Python，就连身边多年工作经验的朋友都开始学Python了！背后的原因竟然是....

(http://www.baidu.com/cb.php?c=IgF_pyfqHmknjnvPjc0IZ0qnfK9ujYzP1nYPH0k0Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1YsPAmzmHuhPhf1nhm3uHFB0AwY5HDdnHR1n103PH00IgF_5y9YIZ0IQzq-uZR8mLPbUB48ugfEIAqspynETZ-YpAq8nWqdiAdxTvqdThP-5yF_UvTkn0KzujY1n0KBuHYs0ZKz5H00ly-b5HDdP1f1PWD0Uv-b5HDzrH63nHf0mv-b5HTzPWb1n6KEIv3qn0KsXHYznjm0mLFW5HRkPHRd)

【转】jQuery.dataTables 自定义排序 u010067685 2016年08月18日 16:20 1722

原文http://wuchong.me/blog/2014/02/25/jquery-datatable-sort-plugin/?utm_source=tuicool&utm_medium=refe... (http://blog.csdn.net/u010067685/article/details/52242120)

大数据IMF传奇 第19课 spark 二次排序 使用JAVA自定义key 进行二次排序

scala> sc.textFile("/README.txt").flatMap(_split(" ")).map(_1).reduceByKey(_+_).map(x => (x._2,x....

duan_zhihua 2016年01月24日 20:56 1004 (http://blog.csdn.net/duan_zhihua/article/details/50575226)

详解spark sql用户自定义函数:UDF与UDAF xgjianstart 2017年02月09日 20:57 4368

场景 UDAF = USER DEFINED AGGREGATION FUNCTION11 上一篇文章已经介绍了spark sql的窗口函数，并知道Spark sql提供了丰富的内置函数供猿友... (http://blog.csdn.net/xgjianstart/article/details/54956413)

迈出成为抢手机器学习工程师第1步



【免费试听】迈出成为抢手机器学习工程师第1步

详解spark sql用户自定义函数:UDF与UDAF

参考 Spark官网 王家林DT大数据梦工厂 场景 UDAF = USER DEFINED AGGREGATION FUNCTION1 上一篇文章已经介绍了spark sql的窗口函数，... (http://blog.csdn.net/weixin_38655836/article/details/78485638)

[2.5]详解spark sql用户自定义函数:UDF与UDAF

参考Spark官网 王家林DT大数据梦工厂场景上一篇文章已经介绍spark sql的窗口函数，并且了解到spark sql提供了丰富的内置函数供猿友们使用，辣为何还要用户自定义函数呢？实际的业务场景...

PENGYUCHENG32109 2016年06月29日 21:58 6357 (http://blog.csdn.net/PENGYUCHENG32109/article/details/51786244)

spark自定义分区

2017年12月08日 17:18 384KB 下载





【集合详解】TreeSet源码解析及自定义排序 u010176014 2016年08月02日 17:54 472

上篇博客中，我们已经了解到TreeSet为有序集合，按照Comparator接口定义的排序顺序进行排序。一些基础类型，都已经实现了该接口，默认是按照自然排序进行排列。而我们自定义类型需要手动实现接口。...

京ICP证09002463号
(http://www.miibeian.gov.cn/)
关于
(http://www.csdn.net/company/about.h
招聘
(http://www.csdn.net/company/recruit.h
广告服务
(http://www.csdn.net/company/marketi
阿里云
Copyright © 1999-2018
CSDN.NET, All Rights Reserved

(<http://blog.csdn.net/u010176014/article/details/52096593>)



mapreduce(三):自定义二次排序流程实例详解

一、概述 MapReduce框架对处理结果的输出会根据key值进行默  u012592062 2016年07月01日 11:34  1371
认的排序，这个默认排序可以满足一部分需求，但是也是十分有限
的。在我们实际的需求当中，往往有要对reduce输出结果进行二次排序的需求...

(<http://blog.csdn.net/u012592062/article/details/51799979>)

Hadoop之MapReduce自定义二次排序流程实例详解



一、概述 MapReduce框架对处理结果的输出会根据key值进行默认的排序，这个默认排序可以满足一部分需求，但是也是十分有限的。在我们实际的需求当中，往往有要对reduce输出结果进行二次排序...

 fly_time2012 2016年02月14日 16:40  578

(http://blog.csdn.net/fly_time2012/article/details/50663435)

Hadoop之MapReduce自定义二次排序流程实例详解

一、概述 MapReduce框架对处理结果的输出会根据key值进行默认的排序，这个默认排序可以满足一部分需求，但是也是十分有限的。在我们实际的需求当中，往往有要对reduce输出结果进行二次排序的需求...

 w1014074794 2016年07月04日 17:37  1296

(<http://blog.csdn.net/w1014074794/article/details/51821727>)

java数组自定义排序

2017年10月31日 15:35 680B [下载](#)



自定义制作数据排序数据;

2013年07月10日 19:28 192KB [下载](#)



spark自定义分区及使用方法

 high2011 2017年03月30日 19:05  3561

spark自定义分区及使用方法

(<http://blog.csdn.net/high2011/article/details/68491115>)

js 自定义表格 支持调整列宽 排序 数据校验 添加/删除行 回车跳转 ...

2010年09月15日 20:23 93KB [下载](#)



自定义Gridview 拖动控件排序移动textview

2015年12月22日 16:52 3.46MB [下载](#)



Spark编写UDAF自定义函数

 mljava1111 2017年04月18日 18:03  574

package main.scalaimport org.apache.spark.SparkContext import org.apache.spark.SparkConf import org....

(<http://blog.csdn.net/mljava1111/article/details/70231057>)

Div仿制table送自定义Scrollbar支持排序

2014年12月25日 22:30 82KB [下载](#)



Flexigrid For Asp.Net 多选,客户端排序,自定义

2013年02月18日 16:42 1.32MB [下载](#)



