


Laurence的技术博客

目录视图

摘要视图

RSS 订阅

个人资料



bluishglc

关注

发私信

访问：3983036次

积分：30699

等级：BLOG

排名：第184名

原创：286篇

转载：8篇

译文：3篇

评论：684条

关于作者

架构师，CSDN博客专家，目前正从事大数据领域的研究和开发工作，对企业级应用架构、分布式存储、SaaS和领域驱动设计有丰富的实践经验，热衷函数式编程，喜欢摄影和旅行。

文章搜索

博客专栏



Scala语言

文章：27篇

阅读：161669



Linux与Shell编程

文章：16篇

阅读：433935



大数据专题

文章：36篇

阅读：370733

数据库分库分表(sharding)

Spark闭包与序列化

标签：spark 闭包 序列化 serializab not

2016-03-21 11:2711780人阅读评论(4)收藏举报

版权声明：本文为博主原创文章，未经博主允许不得转载。

目录(?)

[+]

本文原文出处: <http://blog.csdn.net/bluishglc/article/details/50945032> 严禁任何形式的转载，否则将委托CSDN官方维护权益！

Spark的官方文档再三强调那些将要作用到RDD上的操作，不管它们是一个函数还是一段代码片段，它们都是“闭包”，Spark会把这个闭包分发到各个worker节点上去执行，这里涉及到了一个容易被忽视的问题：闭包的“序列化”。

显然，闭包是有状态的，这主要是指它牵涉到的那些自由变量以及自由变量依赖到的其他变量，所以，在将一个简单的函数或者一段代码片段（就是闭包）传递给类似RDD.map这样的操作前，Spark需要检索闭包内所有的涉及到的变量（包括传递依赖的变量），正确地把这些变量序列化之后才能传递到worker节点并反序列化去执行。如果在涉及到的所有的变量中有任何不支持序列化或没有指明如何序列化自己时，你就会遇到这样的错误：

org.apache.spark.SparkException: Task not serializable

在下面的例子中，我们从kafka中持续地接收json消息，并在spark-streaming中将字符串解析成对应的实体：

```
1 object App {
2     private val config = ConfigFactory.load("my-streaming.conf")
3     case class Person (firstName: String,lastName: String)
4     def main(args: Array[String]) {
5         val zkQuorum = config.getString("kafka.zkQuorum")
6         val myTopic = config.getString("kafka.myTopic")
7         val myGroup = config.getString("kafka.myGroup")
8         val conf = new SparkConf().setAppName("my-streaming")
9         val ssc = new StreamingContext(conf, Seconds(1))
10        val lines = KafkaUtils.createStream(ssc, zkQuorum, myGroup, Map(myTopic -> 1))
11        //this val is a part of closure, and it's not serializable!
12        implicit val formats = DefaultFormats
13        def parser(json: String) = parse(json).extract[Person].firstName
14        lines.map(_._2).map(parser).print
15        ....
16        ssc.start()
17        ssc.awaitTerminationOrTimeout(10000)
18        ssc.stop()
19    }
20
21 }
```

<http://blog.csdn.net/bluishglc/article/details/50945032>

1/3



文章：7篇
阅读：561198

文章存档

- 2018年01月 (1)
- 2017年12月 (1)
- 2017年11月 (2)
- 2017年09月 (1)
- 2017年07月 (1)

展开

阅读排行

- 数据库分库分表(sharding)系... (151816)
- 数据库Sharding的基本思想和... (129334)
- RHEL7/CentOS7 Network S... (102349)
- 关于分布式事务、两阶段提交... (85050)
- 从MVC框架看MVC架构的设计 (79506)
- 数据库分库分表(sharding)系... (75813)
- Spring如何加载XSD文件(org.... (70457)
- 数据库分库分表(sharding)系... (69722)
- java命令行下如何设置classpa... (62893)
- navicat连接oracle数据库报O... (56789)

评论排行

- 从MVC框架看MVC架构的设计 (85)
- 数据库分库分表(sharding)系... (53)
- 数据库分库分表(sharding)系... (44)
- 数据库Sharding的基本思想和... (43)
- Spring如何加载XSD文件(org.... (34)
- HBase高性能复杂条件查询引擎 (28)
- 数据库分库分表(sharding)系... (25)
- 关于分布式事务、两阶段提交... (18)
- 分布式与集群的区别 (15)
- vsftpd：500 OOPS: vsftpd: r... (15)

最新评论

从函数字面量发现函数式编程
StruggleYang：- 把函数的“类型”和“值”联合起来，看看一个函数是如何像一个普通的变量那样被定义出来的val p l...

Spark闭包与序列化
Honxs_xu：图片那个例子，基本数据类型已经实现序列化，不应该会报错吧，即使param不是基本数据类型，按你说拷贝...

vsftpd 配置:chroot_local_user与chroot...
shuaikang666：博主讲的简单明了

vsftpd：500 OOPS: vsftpd: refusing to ...
LeeAtomic：非常感谢，帮了大忙

数据库Sharding的基本思想和切分策略
zyq070：good

Spring如何加载XSD文件(org.xml.sax.SA...
laizhouxiaoboge：很透彻，终于解决连外网的问题了

vsftpd：500 OOPS: vsftpd: refusing to ...
易默可：问题解决，讲的很好，感谢

我们为什么放弃SBT回归Maven
我邪故我在：gradle用的就是gradle和sbt是等价的，只是一个groovy语法，一个是scala语法体...

阅读全文

顶 5
踩 0

- 上一篇 Scala之Case Class
- 下一篇 Scala之偏函数Partial Function

相关文章推荐

- 理解spark闭包
- MySQL在微信支付下的高可用运营--莫晓东
- Spark学习——缓存、闭包及共享变量
- 容器技术在58同城的实践--姚远
- 理解Spark包中闭包
- SDCC 2017之容器技术实战线上峰会
- Spark编程指引(二)-----RDD介绍，闭包...
- SDCC 2017之数据库技术实战线上峰会
- Spark函数传递:闭包和单例模式
- 腾讯云容器服务架构实现介绍--董晓杰
- Spark 中的序列化
- 微博热点事件背后的数据库运维心得--张冬洪
- Spark中的序列化机制
- Spark性能优化第四季-序列化
- spark-java-task未序列化
- Spark 序列化问题全解

查看评论



Honxs_xu

4楼 2017-12-26 19:31发表

图片那个例子，基本数据类型已经实现序列化，不应该会报错吧，即使param不是基本数据类型，按你说拷贝一份到闭包中，这个拷贝动作不也是发生在闭包内吗？和你直接调到param有什么区别？



stoneFangCN

3楼 2017-08-15 08:38发表

为什么调用了成员变量，需要将整个类序列化



sinat_25347387

2楼 2017-02-13 20:06发表

非常好的文章，解决了我的问题，mark一下回去详细研究，多谢！



行巷带盐人

1楼 2016-12-28 15:24发表

写的很好 继续加油，写的这么好为什么没人评论。

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

关于“闭包”的杂文
z01367215352 : @z01367215352:就是例子看不到了

关于“闭包”的杂文
z01367215352 : 简洁到位

公司简介 | 招贤纳士 | 广告服务 | 联系方式 | 版权声明 | 法律顾问 | 问题报告 | 合作伙伴 | 论坛反馈

网站客服 杂志客服 微博客服 webmaster@csdn.net 400-660-0108 | 北京创新乐知信息技术有限公司 版权所有 | 江苏知之为计算机有限公司 |

江苏乐知网络技术有限公司

京 ICP 证 09002463 号 | Copyright © 1999-2017, CSDN.NET, All Rights Reserved 