

原

Spark的Kryo序列化注册

2017年01月13日 10:32:02 阅读数：2094

版权声明：分享的快乐。。 <https://blog.csdn.net/baolibin528/article/details/54406286>

Spark的Kryo序列化注册

Spark序列化可以将RDD序列化来减少内存占用。对于优化网络性能极为重要

1、java序列化：

```
spark.serializer=org.apache.spark.serializer.JavaSerialization
```

Spark默认使用Java自带的ObjectOutputStream 框架来序列化对象，这样任何实现了 java.io.Serializable 接口的对象，都能序列化。Java序列化很灵活但性能差速度很慢，同时序列化后占用的字节数也较多。

2、Kryo序列化：

```
spark.serializer=org.apache.spark.serializer.KryoSerialization
```

KryoSerialization速度快,产生的结果更为紧凑（通常能提高10倍）。但Kryo不支持所有实现了java.io.Serializable 接口的类型，它需要你程序中 register 需要序列化的类型，以得到最佳性能。

在 SparkConf 初始化的时候调用

```
conf.set("spark.serializer", "org.apache.spark.serializer.KryoSerializer")
```

使用 Kryo。这个设置不仅控制各个worker节点之间的混洗数据序列化格式，同时还控制RDD存到磁盘上的序列化格式。需要在使用时注册需要序列化的类型，建议在对网络敏感的应用场景下使用Kryo。

一个例子：

```
1 val sparkConf = new SparkConf()
2   /*.setMaster("Local")*/
3   .setAppName("Test")
4   .set("spark.serializer", "org.apache.spark.serializer.KryoSerializer")
5   /*.set("spark.kryo.registrationRequired", "true")*/
6   .registerKryoClasses(Array(
7     classOf[Array[String]],
8     classOf[util.HashMap[String, String]],
9     classOf[util.ArrayList[String]],
10    classOf[MyTest]
11  ))
```

在集群跑的时候把/* .set("spark.kryo.registrationRequired", "true")*/注释掉，否则会报如下错误：

```
Class is not registered: scala.reflect.ClassTag$$anon$1或者Class is not registered: scala.reflect.ManifestFactory
```


最后，如果你不注册需要序列化的自定义类型，Kryo也能工作，不过每一个对象实例的序列化结果都会包含一份完整的类名，这很浪费没必要

老股民酒后无意说漏：20年炒股 坚持只看1指标

柘晋 · 顶新


想对作者说点什么？


我来说一句


spark之kryo 序列化  245
几乎所有的资料都显示kryo 序列化方式优于java自带的序列化方式，而且在spark2.*版本中都是默认采用kryo 序...


Spark性能调优之——在实际项目中使用Kryo序列化  3221
set ("spark.serializer","org.apache.spark.serializer.KryoSerializer") Java的序列化机制，ObjectOutputStream/...

 **你不知道区块链开发都是这样练成的？**
区块链DApp开发学习路线图，月薪4万很轻松


Spark kyro Serialization  3811
序列化在分布式系统中扮演着重要的角色，优化Spark程序时，首当其冲的就是对序列化方式的优化。Spark为...

Kryo序列化  1981
在上一篇博文中提到，Spark任务中，使用序列化数据的方式对增加其执行效率有很大的作用。并且由于Kryo序...


Spark中使用kyro序列化  787
序列化在分布式系统中扮演着重要的角色，优化Spark程序时，首当其冲的就是对序列化方式的优化。Spark为...


Spark性能调优：使用Kryo序列化库、RDD的压缩  794
使用Kryo序列化的原因默认情况下，Spark内部是通过Java序列化（默认的序列化方式）形成一个二进制字节数...

女性得了静脉曲张变成蚯蚓腿怎么办？用这方法坚持3个月全恢复！
水英电器 · 顶新

spark中的性能调优5(使用kryo序列化)  56
问题1>为什么要使用kryo序列化呢？ 主要是解决数据在内存中的占用，和网络传输的性能问题2&...

spark从入门到放弃十九:Spark 性能优化(2)高性能序列化类库  374
文章地址：http://www.haha174.top/article/details/254551 在任何分布式系统里面序列化都是扮演一个很重要的...

[Spark优化]在Spark中使用Kryo序列化  9080
conf.set("spark.serializer", "org.apache.spark.serializer.KryoSerializer") conf.registerKryoClasses...

spark 使用kryo  2131
kryo vs nokryo背景我今天突发奇想，说测一下kryo如果没有注册会咋样，结果发现没有注册的Kryo性能低于直...

相关热词 spark和 gp和spark spark中文网 spark取时间 spark子句

个人资料

 **鲍礼彬**

关注

原创	粉丝	喜欢	评论
338	167	13	44

等级：  访问：61万+

积分：9745 排名：2520

勋章： 



阅读量：9566 9 篇

个人简介

说明：小青年在奋斗中
一个爱学习的小菜鸟。。。

专业：养猪
年级：小学未毕业
所在地：村头干活那小伙就是我~
姓名：鲍礼彬

QQ号：
257-8858-653

呵呵哒
么么哒

最新文章

[spark的aggregateByKey简单用法](#)
[剑指offer面试题](#)
[jieba分词](#)
[面试题42-题目2：字符串左旋转](#)
[面试题42：翻转单词顺序](#)

个人分类

Bigdatda-Hive	43篇
Bigdatda-HBase	24篇
Bigdatda-Zookeeper	10篇
Bigdatda-MapReduce	25篇
Rindatda-HDFS	11篇

展开

归档

2017年7月	1篇
2017年3月	1篇
2017年2月	46篇
2017年1月	9篇
2016年12月	2篇

展开

雨水收集系统



联系我们



请扫描二维码联系客服

 webmaster@csdn.net

 400-660-0108

 QQ客服

 客服论坛

关于 招聘 广告服务 网站地图

©2018 CSDN版权所有 京ICP证09002463号

 百度提供搜索支持

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

https://blog.csdn.net/baolibin528/article/details/54406286

4/4