

原

spark参数调优详解目录

置顶

2018年07月12日 10:30:30

Deegue

阅读数：10389

版权声明：本文为博主原创文章，未经博主允许不得转载。<https://blog.csdn.net/zyzzxycj/article/details/81011540>

spark参数调优需要对各个参数充分理解，没有一套可以借鉴的参数，因为每个集群规模都不一样，只有理解了参数的用途，调试出符合自己场景且能在扩大集群、业务的情况下，能够跟着修改参数才算是正确的参数调优。

本系列文章仅为个人理解与观点，如果任何地方有疑问与错误，欢迎评论指正！

（上次更新时间---2018-09-05 15：07）

1、背景

spark2.2.1版本(现升级到了2.3.1版本)，其实官方文档中，相关的参数已经写的很详细了，这边仔细阅读了官方文档，

并结合了实际情景，实践之后，整理了这篇博客。

官方文档：

<http://spark.apache.org/docs/2.2.1/configuration.html>

yarn:

<http://spark.apache.org/docs/2.2.1/running-on-yarn.html#configuration>

2、参数详解

① Application Properties 应用基本属性

<https://blog.csdn.net/zyzzxycj/article/details/82255979>

② Runtime Environment 运行环境

<https://blog.csdn.net/zyzzxycj/article/details/82256032>

③ Shuffle Behavior

<https://blog.csdn.net/zyzzxycj/article/details/82256081>

④ Spark UI

<https://blog.csdn.net/zyzzxycj/article/details/82256109>

⑤ Compression and Serialization 压缩和序列化

<https://blog.csdn.net/zyzzxycj/article/details/82256152>

⑥ Memory Management 内存管理

<https://blog.csdn.net/zyzzxycj/article/details/82256162>

⑦ Executor behavior 执行器

<https://blog.csdn.net/zyzzxycj/article/details/82256206>

⑧ Networking 网络

<https://blog.csdn.net/zyzzxycj/article/details/82256228>

⑨ Scheduling 调度

<https://blog.csdn.net/zyzzxycj/article/details/82256244>

⑩ Dynamic Allocation 动态分配

<https://blog.csdn.net/zyzzxycj/article/details/82256893>

11 Spark Streaming

<https://blog.csdn.net/zyzzxycj/article/details/82428031>

上海十年股民经验分享,只用1招锁定牛股 单月扭亏为盈!

中方信富 · 熯燚



想对作者说点什么

Spark调优-参数及配置

参数调优 1 资源类 11 num-executors 12 executor-memory 13 executor-cores 14 driver-memory 15 sparkdefaultpar...

4927

来自: 朱金华的专栏

Spark开发性能调优

Spark开发性能调优 标签 (空格分隔) : Spark –Write By Vin 1. 分配资源调优 Spark性能调优的王道就是分配资源,...

3684

来自: vinfly\_li的博客

Spark性能调优-性能监控、参数调整

以下是整理的Spark中的一些配置参数，官方文档请参考Spark Configuration。 Spark提供三个位置用来配置系统： ...

2643



发现了一个免费的云服务器,号称是永久的

百度广告

spark参数调优

最近在搞一个sparkjob，耗时真的是醉醉的，最终的reduce过程只占三分之一的时间，其他的占了三分之二的时间...

301

来自: a925907195的专栏

spark性能调优：开发调优

在大数据计算领域，Spark已经成为了越来越流行、越来越受欢迎的计算平台之一。Spark的功能涵盖了大数据领域...

2.1万

来自: LW\_ICE

spark调优(一)【spark参数介绍】

1 spark on yarn常用属性介绍 属性名 默认值 属性说明 spark.yarn.am.memory 512m 在客户端模式（client mode）...

681

来自: 极客神殿

Spark性能优化总结

优化的目标 保证大数据量下任务运行成功 降低资源消耗 提高计算性能 三个目标优先级依次递减，首要解决的是程...

1164

来自: gongpulin的博客

Spark性能优化：资源调优篇

在开发完Spark作业之后，就该为作业配置合适的资源了。Spark的资源参数，基本都可以在spark-submit命令中作...

5.1万

来自: 铭霏的记事本

led霓虹灯字母定制制作玻璃管霓虹灯酒吧门头广告牌

北京东都华易 · 熯燚

Spark调优（数据序列化和内存调优）

翻译自Spark官网文档：https://spark.apache.org/docs/2.1.0/tuning.html 前言 由于大多数Spark计算的内存使用...

3725

来自: junerli的博客

文章热词

Spark Spark培训 Spark课程 Spark视频教程 Spark学习

相关热词

androidstudio 项目目录详解 c++ mlib spark 调用 android调试和调优技能 bootstrap-table详解 c++ vector详解 spark教程+python spark教程python

spark性能优化：数据倾斜调优

调优概述 有的时候，我们可能会遇到大数据计算中一个最棘手的问题——数据倾斜，此时Spark作业的性能会比期...

2.6万

来自: LW\_ICE



段智华

847篇文章

排名:1105

关注



Gamer\_gyt

263篇文章

排名:1378

关注



kwu\_ganymede

252篇文章

排名:3303

关注



xwc35047

43篇文章

排名:39503

关注

Spark Dynamic Allocation 分析

动态资源调配这是spark1.5以后新加入的功能为standalone模式和mesos的粗粒度模式提供了executor的动态管理， ...

5077

来自: 偷闲小苑

<div>关于<b>spark</b>程序动态资源分配的一些理解</div> <div>环境： cdh5.7.1 cdh5.7.1中的spark版本为spark1.6 关于如何配置动态资源分配， 参见： http://spark.apache.org/do...</div> <div>一步一步完成如何在现有的<b>CDH</b>集群中部署一个与<b>CDH</b>版本不同的<b>spark</b></div> <div>首先当然是下载一个spark源码， 在http://archive.cloudera.com/cdh5/cdh/5/中找到属于自己的源码， 自己编译打包,...</div>	<div>来自： xiao_jun_0820的专栏</div> <div>来自： xiao_jun_0820的专栏</div>
<div>上海股王8年追涨停定律“1272”曝光，震惊众人！</div> <div>陕西bce · 熾燚</div>	
<div>Apache Zeppelin on CDH搭建</div> <div>本文基于centos6.4、CDH版本5.7.6、spark版本为1.6.0 1.环境准备 git1.7.1、maven3.3.9、JDK1.8 2.下载最新版z...</div>	<div>来自： 公子洲的博客</div>
<div>Spark Sql性能测试及调优</div> <div>1 问题描述 内存不足时group by操作失败。 正常应该速度变慢，而不是失败，因为还有磁盘可用 错误日志： T...</div>	<div>来自： DF_XIAO的专栏</div>
<div>Spark 性能相关参数配置详解—任务调度篇</div> <div>随着Spark的逐渐成熟完善，越来越多的可配置参数被添加到Spark中来，本文试图通过阐述这其中部分参数的工作原...</div>	<div>来自： stark_summer的专栏</div>
<div>大数据套件之HBase性能测试</div> <div>转载地址:http://komi.leanote.com/post/tpcds%E5%92%8Cycsb%E8%B0%83%E4%BC%98%E6%80%BB%E7%B...</div>	<div>来自： 知识的力量</div>
<div>neo4j写数据遭遇OOM</div> <div>aaaa</div>	<div>来自： 白乔专栏</div>
<div>重大通知！12月起，上海上班族可自考985/211本科学历</div> <div>尚德机构 · 顶新</div>	
<div>Spark性能优化：JVM参数调优</div> <div>Spark性能优化：JVM参数调优 年轻代：主要是用来存放新生的对象。老年代：主要存放应用程序中生命周期长的...</div>	<div>来自： Ganymede的Hadoop世界</div>
<div>Spark性能调优2-总结分享</div> <div>1、Spark调优背景 目前Zeppelin已经上线一段时间，Spark作为底层SQL执行引擎，需要进行整体性能调优，来提...</div>	<div>来自： 简单就好</div>
<div>spark之jvm调优</div> <div>转自:https://blog.csdn.net/lxhandlbb/article/details/52987928一、性能调优分类：1.常规性能调优： 分配资源，并...</div>	<div>来自： weixin_41804049的博客</div>
<div>[Spark2.0]ML 调优：模型选择和超参数调优</div> <div>本节讲述如何使用MLlib提供的工具来对ML算法和Pipeline进行调优。内建的交叉验证和其他工具允许用户在算法和Pi...</div>	<div>来自： yhao2014的专栏</div>
<div>Spark2 Linear Regression线性回归案例(参数调优)</div> <div>回归正则化方法（Lasso，Ridge和ElasticNet）在高维和数据集变量之间多重共线性情况下运行良好。 数学上，EI...</div>	<div>来自： tiansheng1225的博客</div>
<div>一站式服务,先服务后收费</div> <div>上海隆助 · 顶新</div>	
<div>spark submit参数调优</div> <div>摘要 1.num-executors 2.executor-memory 3.executor-cores 4.driver-memory 5.spark.default...</div>	<div>来自： chenjieit619的博客</div>
<div>spark 参数调优详解（持续更新中）</div> <div>spark参数调优需要对各个参数充分理解，没有一套可以借鉴的参数，因为每个集群规模都不一样，只有理解了参数...</div>	<div>来自： kwame211的博客</div>
<div>Spark性能优化：开发调优篇</div> <div>前言 在大数据计算领域，Spark已经成为了越来越流行、越来越受欢迎的计算平台之一。Spark的功能涵盖了大数据...</div>	<div>来自： 铭霏的记事本</div>
<div>Spark调优的策略</div> <div>1. RDD的持久化 cahce() persist() checkpoint() 2. 避免创建重复的RDD 3.尽可能复用同一个RDD 类似于多个RDD...</div>	<div>来自： fanyao4144的博客</div>
<div>Spark性能调优-总结分享</div> <div>1、Spark调优背景 目前Zeppelin已经上线一段时间，Spark作为底层SQL执行引擎，需要进行整体性能调优，来提...</div>	<div>来自： shieh的专栏</div>

为啥90后都买了房，是因为这种新套路！

光谷金融 · 熯燚

Spark性能优化-----资源调优

为作业配置合适的资源，Spark的资源参数，基本都可以在spark-submit命令中作为参数设置。Spark作业基本运行...

816

来自： 闭雨哲的专栏

SparkML模型选择（超参数调整）与调优

Spark ML模型选择与调优本文主要讲解如何使用Spark MLlib的工具去调优ML算法和Pipelines。内置的交叉验证和...

1646

来自： Spark高级玩法

sparksql调优之第一弹

1, jvm调优这个是扯不断，理还乱。建议能加内存就加内存，没事调啥JVM，你都不了解JVM和你的任务数据。sp...

2721

来自： Spark高级玩法

hive on spark 性能参数调优

select \* from stg\_bankcard\_auth\_apply where length(idcardno) &gt;= 1 and length(idcardno) &lt;= 18

120

来自： AntKengElephant的博客

Spark Streaming详解---概述、基本概念、性能调优

本文主要讲述SparkStreaming概念原理、基本概念、以及调优等一些知识点。1 概述1.1 SparkStreaming是什...

482

来自： 香菇的博客



30个数据可视化超级工具

百度广告

SparkSQL性能调优

最近在学习spark时，觉得Spark SQL性能调优比较重要，所以自己写下来便于更过的博友查看，同时也希望大家给...

6584

来自： YQlakers的博客

spark-shuffer参数调优

转自：https://tech.meituan.com/spark-tuning-pro.html 部分 shuffle调优 调优概述 大多数Spark作业的性能主要就是...

977

来自： 卡奥斯道的博客

Spark性能优化：shuffle调优

shuffle调优 调优概述 大多数Spark作业的性能主要就是消耗在了shuffle环节，因为该环节包含了大量的磁盘IO...

1.8万

来自： 铭睿的记事本

spark参数配置调优

SPARK配置参数的两个地方： 1. \$SPARK\_HOME/conf/spark-env.sh 脚本上配置。配置格式如下： export SPAR...

2.8万

来自： 提笔写忧伤

spark性能调优--jvm调优

一.问题切入 调用spark 程序的时候，在获取数据库连接的时候总是报 内存溢出 错误 （在ideal上运行的时候设置jv...

453

来自： keyuquan的博客

紧急通知：吸这两种烟的人，快去医院检查！

华佰科技 · 熯燚

第1章 对运行在YARN上的Spark进行性能调优

第1章 对运行在YARN上的Spark进行性能调优1.1 运行环境Jar包管理及数据本地性原理调优实践1.1.1运行环境J...

1152

来自： 段智华的博客

说说我对spark-sql的job的参数调优

首先要说的是机器配置 namenode/datanode: cpu : 8core memory: 56G 我现在是基于yarn+spark的架构来说。现... 来自： Kelvin Liang Study Footp...

1693

Spark on yarn模式的参数设置即调优

1 启动方式 执行命令./spark-shell --master yarn默认运行的是client模式。执行./spark-shell --master yarn-client或者...

1261

Spark2.0机器学习系列之4：随机森林介绍、关键参数分析

随机森林算法介绍 Spark2.0中的参数设置 代码分析

4026

来自： qq\_34531825的博客

spark2.x-jvm调优实战（以tomcat访问日志分析为例）

集群环境 节点 xyz01.aiso.com 硬件环境（VMware 10 虚拟机）（三台） 内存 6G CPU 4个1核处理器 硬盘SCSI 5...

170

来自： program哲学



上海大数据公司

百度广告

Spark（十二）--性能调优篇

一段程序只能成功能是没有用的，只能够稳定、高效率地运行才是生成环境所需要的。本篇记录了Spark各个...

3616

来自： 小黑

<b>Spark Shuffle 过程参数优化深入剖析-Spark商业调优实战</b> 本套系列博客从真实商业环境抽取案例进行总结和分享，并给出Spark商业应用实战指导，请持续关注本套博客。版...	来自: 秦凯新的技术社区
<b>Spark应用程序开发参数调优深入剖析-Spark商业调优实战</b> 本套系列博客从真实商业环境抽取案例进行总结和分享，并给出Spark商业应用实战指导，请持续关注本套博客。版...	来自: 秦凯新的技术社区
<b>Spark ML Pipeline模型选择及超参数评估调优深入剖析 -Spark商业ML实战</b> 本套技术专栏是作者（秦凯新）平时工作的总结和升华，通过从真实商业环境抽取案例进行总结和分享，并给出商...	来自: 秦凯新的技术社区
<b>spark sql 性能优化</b> 一 设置shuffle的并行度 我们可以通过属性spark.sql.shuffle.partitions设置shuffle并行度 二 Hive数据仓库建设的时...	来自: happy19870612's blog
<b>spark SQL 参数调优</b> 1 spark.sql.codegen 默认值为false，当它设置为true时，Spark SQL会把每条查询的语句在运行时编译为java的二...	来自: 不管大小写的博客
<b>Spark性能优化：数据倾斜调优</b> 本文转载自：http://tech.meituan.com/spark-tuning-basic.html (美团技术点评团队) 前言 继基础篇讲解了每个Spark...	来自: 铭霏的记事本
<b>职场寒冬 请不要让码农只活在青春里～</b> 互联网兴起今天，让信息高速的传播，让很多人可能都已经忘了这万千焦点中的微不足道的一个小小插曲。分享这...	来自: 首席填坑官·苏南的专栏
<b>集中式架构，垂直拆分，分布式，微服务概念，大白话解释</b> 3分钟读懂系统架构演变 了解下最火的微服务概念 本人将从大到小给你讲授系统架构的演变（此处的大小不是对...	来自: 熊局长的博客
<b>整理了10个干净、好用的BT、磁力链搜索网站给大家</b> 现在越来越流行在线看视频了，但是对于我得收藏癖爱好者，还是希望可以有比较好的资源网站的，尤其是种子、...	来自: YXAPP的技术分享
<b>[亲测]mock.js使用</b> 一、Mock.js入门 1. 什么是mock.js? Mock.js （官网http://mockjs.com/）是一款模拟数据生成器，旨在帮助前端攻...	来自: Armymans的博客
<b>【亲测】SpringCloud之Eureka注册中心</b> 一.Eureka注册中心 1.1.认识Eureka Eureka做什么？ Eureka就好比是滴滴，负责管理、记录服务提供者的信息。服...	来自: Armymans的博客
<b>docker入门（五）</b> 如在文档中遇到什么问题请联系作者 QQ：1172796094 本人正在找深圳Java实习工作，求大佬带飞 —————...	
<b>Java抽象类中的静态方法</b> 抽象类中是可以有静态方法的，在被子类继承后可以通过子类的实例化对象调用；也可以直接通过抽象类和子类的...	来自: qq_33699659的博客
<b>批处理添加字段和删除字段</b> 在平时的工作中你或许会遇到这样的问题，给某个文件夹下面的所有的图片添加同一个字符串，如果一个一个去F2...	来自: pyf_914406232的博客
<b>最新迅雷“应版权方要求，文件无法下载”的解决办法</b> 迅雷下载有的电影电视剧的时候会出现：应版权方要求，文件无法下载，或者显示迅雷任务包含违规内容 无法继续...	来自: 徐奕的专栏
<b>容器将成为下一个“Linux”</b> ...	来自: Docker的专栏
<b>很黄很暴力的十个网站</b> 13岁的北京学生张某，在去年12月27日19时新闻联播一则关于净化网络视听的新闻里，接受采访时说的话激起了轩...	来自: Kinb_huangwei的专栏
<b>pyCharm最新2018激活码</b> 本教程对jetbrains全系列可用例：IDEA、WebStorm、phpstorm、clion等 因公司的需求，需要做一个爬取最近上映...	来自: 昌昌
<b>2018最新Web前端经典面试试题及答案</b> 本篇收录了一些面试中经常会遇到的经典面试题以及自己面试过程中遇到的一些问题，并且都给出了我在网上收集...	来自: wdlhao的博客
<b>webstorm 2018 激活破解方法大全</b> webstorm 作为最近最火的前端开发工具,也确实对得起那个价格,但是秉着勤俭节约的传统美德,我们肯定是能省则省...	来自: 唐大帅的编程之路

<b>Postman 使用方法详解</b>	👁 180264
一、Postman背景介绍 用户在开发或者调试网络程序或者是网页B/S模式的程序的时候是需要一些方法来跟踪网页...	来自: <a href="#">fxb123的博客</a>
<b>史上最简单的 SpringCloud 教程   终章</b>	👁 1229368
转载请标明出处： http://blog.csdn.net/forezp/article/details/70148833 本文出自方志朋的博客 错过了这一篇，你可...	来自: <a href="#">方志朋的专栏</a>
<b>门罗币 xmr 超级详细的CPU xmr挖矿教程</b>	👁 56723
门罗币 xmr 最详细的CPU 挖矿教程 基础 CUP 挖矿教程 如何挖矿？ Step1:获得一个钱包地址 钱包分为两个部分讲，...	来自: <a href="#">qq_39863517的博客</a>
<b>Python 爬取爱奇艺腾讯视频二十五万条数据分析为什么李诞不值得了？</b>	👁 3982
Python 分析 在《Python 爬取爱奇艺 52432 条数据分析谁才是《奇葩说》的焦点人物？》这篇文章中，我们从爱奇...	来自: <a href="#">大罗说事</a>
<b>史上最全Java面试题（带全部答案）</b>	👁 112353
今天要谈的主题是关于求职，求职是在每个技术人员的生涯中都要经历多次。对于我们大部分人而言，在进入自己...	来自: <a href="#">林老师带你学编程</a>
<b>军事理论课答案（西安交大版）</b>	👁 1027270
1.1 1 【单选题】我国陆地领土面积排名世界第几？（C） A、1 B、2 C、3 D、4 2 【单选题】以下哪个国家不属于...	来自: <a href="#">ling_wang的博客</a>
<b>关系型数据库与非关系型数据库详细比较</b>	👁 1355
关系型数据库与非关系型数据库详细比较 1.关系型数据库通过外键关联来建立表与表之间的关系， 2.非关系型数据...	来自: <a href="#">Armymans的博客</a>
<b>Java子类实例化总是调用父类的无参构造方法</b>	👁 554
子类实例化时，默认调用父类的无参构造方法（不管子类的构造器有没有参数，因为子类继承的是父类的属性和方...	来自: <a href="#">qq_33699659的博客</a>
<b>2018最好用百度云破解版，百度网盘不限速下载，教你如何解决百度网盘限速的方法。亲测完美使用</b>	👁 30185
百度网盘不限速 点击下载 提取码：jsk0 百度网盘不限速 点击下载 提取码：jsk0 对于大多数人来说，每次在百度网...	来自: <a href="#">qq_41925894的博客</a>
<b>java23种设计者模式 + 例子</b>	👁 2300
java的设计模式大体上分为三大类： *创建型模式（4种）： 工厂方模式，单例模式，建造者模式，原型模式 *结构...	来自: <a href="#">Armymans的博客</a>
<b>docker入门(一)</b>	👁 2219
docker入门（一） 如在文档中遇到什么问题请联系作者 QQ：1172796094 本人正在找深圳实习工作，求大佬带飞 ...	
<b>【《Unity Shader入门精要》提炼总结】(十二)第十二章·实现透明的两种方法&amp;深度缓冲概念介绍&amp;渲染...</b>	👁 2136
本文由@唐三十胖子出品，转载请注明出处。 文章链接：https://blog.csdn.net/iceSony/article/details/8466719...	来自: <a href="#">唐三十胖子的博客</a>
<b>SpringCloud搭建Hystrix熔断机制（亲测）</b>	👁 584
1. Hystrix 在高并发领域，在分布式的系统中，可能因为一个小小的功能抗不主压力，宕机了，导致其他服务也随着...	来自: <a href="#">weixin_43866856的博客</a>
<b>java缓冲区</b>	👁 743
1 缓冲区的分类 ByteBuffer CharBuffer ShortBuffer IntBuffer LongBuffer FloatBuffer DoubleBuffer 2 ByteBuffe...	来自: <a href="#">weixin_43694144的博客</a>
<b>2019届华为秋招数字芯片面试经验</b>	👁 2147
2019届秋招，2018年10月16日，坐标杭州香格里拉饭店西楼小礼堂。距离拿到offer已半月有余，闲来无事，回忆...	来自: <a href="#">weixin_43343190的博客</a>
<b>java练习总结</b>	👁 579
记录平时使用java时发现的一些细节问题。	来自: <a href="#">qq_33699659的博客</a>
<b>docker入门（七）</b>	👁 1754
如在文档中遇到什么问题请联系作者 QQ：1172796094 本人正在找深圳Java实习工作，求大佬带飞 —————...	
<b>docker入门（八）</b>	👁 1769
如在文档中遇到什么问题请联系作者 QQ：1172796094 本人正在找深圳Java实习工作，求大佬带飞 —————...	
<b>【Linux从入门到放弃】Linux开发第一天&amp;文件与文件夹的创建&amp;文本的读写</b>	👁 2144
Linux第一天 1.创建一个目录/data mkdir /data或者cd /;mkdir data 2.在data目录下创建一个test.txt touch /data/test.t...	来自: <a href="#">唐三十胖子的博客</a>
<b>手把手教你如何安装Pycharm——靠谱的Pycharm安装详细教程</b>	👁 266100
今天小编给大家分享如何在本机上下载和安装Pycharm，具体的教程如下：1、首先去Pycharm官网，或者直接输入...	来自: <a href="#">pdfighting的博客</a>



各大磁力种子搜索引擎对比

424594

现在磁力种子搜索引擎质量参差不齐，现在就重点整理几个常用的种子搜索站，做个对比分析 1.屌丝搜-最懂屌丝的...

来自: [lizhengnanhua的专栏](#)

Java 枚举(enum) 详解7种常见的用法

473024

JDK1.5引入了新的类型——枚举。在 Java 中它虽然算个“小”功能，却给我的开发带来了“大”方便。大师兄我又加...

来自: [请叫我大师兄](#)

智慧树走进故宫章节测试答案智慧树走进故宫刷课插件2018年智慧树走进故宫期末考试答案

94077

欢迎通过微信客户端搜索jiutu2019或者扫描下方二维码关注廖图公众号获取智慧树刷题插件。智慧树刷题插件获取...

来自: [柠檬很酸的博客](#)

Python 凭什么碾压Java、C++等老牌语言？

33061

随着计算机语言的发展，Python也跻身于语言排行的常青树。要是说Python是最目前最火爆的语言，应该没有人反...

来自: [CSDN学院](#)

MVX框架模式了解

94

MVX框架模式了解 MVX框架模式：MVC+MVP+MVVM 1.MVC：Model(模型)+View(视图)+controller(控制器)，主要...

来自: [猪精的博客](#)

tcp的java代码

722

服务器端 package com.car.client; import java.io.IOException; import java.io.OutputStream; import java.ne...

来自: [weixin\\_43694144的博客](#)



Deegue

关注

原创51

粉丝15

喜欢2

评论8

等级: 博客5

访问: 39万+

积分: 3720

排名: 1万+

勋章: 恒



最新文章

[Spark执行流程与原理](#)

[broadcast hash join和sort merge join](#)

[Apache Spark SQL自适应执行实践](#)

[使用jmap和MAT分析JVM堆内存](#)

[HBase基本架构及原理](#)

个人分类

个人感想

2篇

hadoop

29篇

git

2篇

linux

2篇

ACM

1篇

展开

归档

2018年11月	1篇
2018年10月	4篇
2018年9月	9篇
2018年8月	11篇
2018年7月	2篇
展开	

热门文章

spark 参数调优5-Compression and Serialization压缩与序列化

阅读量：17035

spark 参数调优7-Executor behavior

阅读量：17009

zeppelin on yarn 搭建遇到的问题总结

阅读量：13088

位图索引:原理（BitMap index）

阅读量：13068

Hadoop 过滤,映射,谓词下推基本概念

阅读量：12877

最新评论

Mac配置git管理神器-Sour...

zyzzxycj: [reply]u014355937[/reply] 所有csdn上vpn的文章都被封了，如果需要可...

Mac配置git管理神器-Sour...

u014355937: 最关键的地方却恰恰没写

Spark执行计划分析与研究

zyzzxycj: [reply]zlb872551601[/reply] 执行计划中，每个RDD操作前都会有编号；第...

Spark执行计划分析与研究

zlb872551601: 为什么第一个图中，从HadoopRDD到Union会经过两次的MapPartitions，感觉只需...

Spark执行计划分析与研究

zlb872551601: 楼主，为什么我在看执行计划的时候，只看到 == Physical Plan ==和HiveTabl...



IT 外包公司



联系我们



微信客服



QQ客服

 QQ客服

 kefu@csdn.net

 客服论坛

 400-660-0108

工作时间 8:00-22:00

关于我们

招聘

广告服务

网站地图

 百度提供站内搜索

京ICP证09002463号

©1999-2018 江苏乐知网络技术有限公司



江苏知之为计算机有限公司 北京创新乐知  
信息技术有限公司版权所有

网络110报警服务 经营性网站备案信息  
北京互联网违法和不良信息举报中心  
中国互联网举报中心