

javastart的专栏

目录视图

摘要视图

RSS 订阅

个人资料



javastart

访问：93897次

积分：1500

等级：BLOG-4

排名：第16381名

原创：26篇

转载：185篇

译文：0篇

评论：9条

文章搜索

文章分类

java (11)

javascript (1)

jsp (1)

jtee (1)

mssql (1)

tomcat (6)

数据库开发 (2)

架构 (16)

系统架构 (11)

大数据 (122)

oracle (1)

mysql (23)

nosql (13)

linux (5)

hadoop (2)

spark (2)

hbase (0)

clouderamanger (1)

sqoop (0)

算法 (3)

文章存档

2016年01月 (9)

2015年12月 (12)

2015年11月 (23)

2015年10月 (12)

2016软考项目经理实战班 学院周年礼-顶尖课程钜惠呈现 有奖征文：在云上开发的无限可能 2015年度社区之星 放榜啦

Spark History

置使用(zhuan)

2015-02-11 16:20 2001人阅读 评论(0) 收藏 举报

分类： 大数据 (121)

Spark History Server配置使用

Spark history Server产生背景

以standalone运行模式为例，在运行Spark Application的时候，Spark会提供一个WEBUI列出应用程序的运行信息；但该WEBUI随着Application的完成(成功/失败)而关闭，也就是说，Spark Application运行完(成功/失败)后，将无法查看Application的历史记录；

Spark history Server就是为了应对这种情况而产生的，通过配置可以在Application执行的过程中记录下了日志事件信息，那么在Application执行结束后，WEBUI就能重新渲染生成UI界面展现出该Application在执行过程中的运行时信息；

Spark运行在yarn或者mesos之上，通过spark的history server仍然可以重构出一个已经完成的Application的运行参数信息（假如Application运行的事件日志信息已经记录下来）；

配置&使用Spark History Server

以默认配置的方式启动spark history server:

cd \$SPARK_HOME/sbin

start-history-server.sh

报错:



starting org.apache.spark.deploy.history.HistoryServer, logging to /home/spark/software/source/compile/deploy_spark/sbin/../logs/spark-spark-org.apache.spark.deploy.history.HistoryServer-1-hadoop000.out

failed to launch org.apache.spark.deploy.history.HistoryServer:

at org.apache.spark.deploy.history.FsHistoryProvider.<init>(FsHistoryProvider.scala:44)

... 6 more



需要在启动时指定目录:

start-history-server.sh hdfs://hadoop000:8020/directory

hdfs://hadoop000:8020/directory可以配置在配置文件中，那么在启动history-server时就不需要指定，后续介绍怎么配置；

注：该目录需要事先在hdfs上创建好，否则history-server启动报错。

启动完成之后可以通过WEBUI访问，默认端口是18080：http://hadoop000:18080

http://blog.csdn.net/javastart/article/details/43735343

1/5

2015年09月 (18)

展开

阅读排行

数据库—Cobar分布式数

(10534)

Windows下IntelliJ IDEA

(5792)

spark 案例集群测试整理

(4771)

构建高并发高可用的电商

(3918)

spark1.3.0__for_hadoop

(3895)

Java环境配置

(3501)

Spark与Mysql(JdbcRDC

(3345)

金蝶 eas webservice 接

(2582)

Java编程:配置tomcat5数

(2105)

Spark History Server配置

(1999)

评论排行

Windows下IntelliJ IDEA

(4)

[原]Spark MLlib Deep Le

(2)

构建高并发高可用的电商

(1)

Spark MLlib Deep Learn

(1)

Spark RDD写入RMDB(M

(1)

12个最重要的J2EE最佳

(0)

Tomcat配置技巧Top10

(0)

不负众望 Java API文档

(0)

基于Spark MLlib平台的

(0)

Java编程:配置tomcat5数

(0)

推荐文章

*算法与数据结构学习资源大搜罗——良心推荐

*架构设计：系统间通信（17）——服务治理与Dubbo 中篇（分析）

*数据库性能优化之SQL语句优化

*Android应用开发allowBackup敏感信息泄露的一点反思

*Linux多线程实践（四）线程的特定数据

* Android性能优化之加快应用启动速度

最新评论

Spark RDD写入RMDB(Mysql)方

kaka953443858: 您好,我们现在用的是sqlserver数据库,现在向数据库插入数据 insertintoJD...

Windows下IntelliJ IDEA中调试S

flyingsky2008: @yy2314703:你知道怎么用idea远程提交了么?

Spark MLlib Deep Learning Ne

hxwj110: 您好 我想咨询下/* ** sigm激活函数 * X = 1./(1+exp(-P));...

构建高并发高可用的电商平台架

guoxiaodiu_: 很赞!

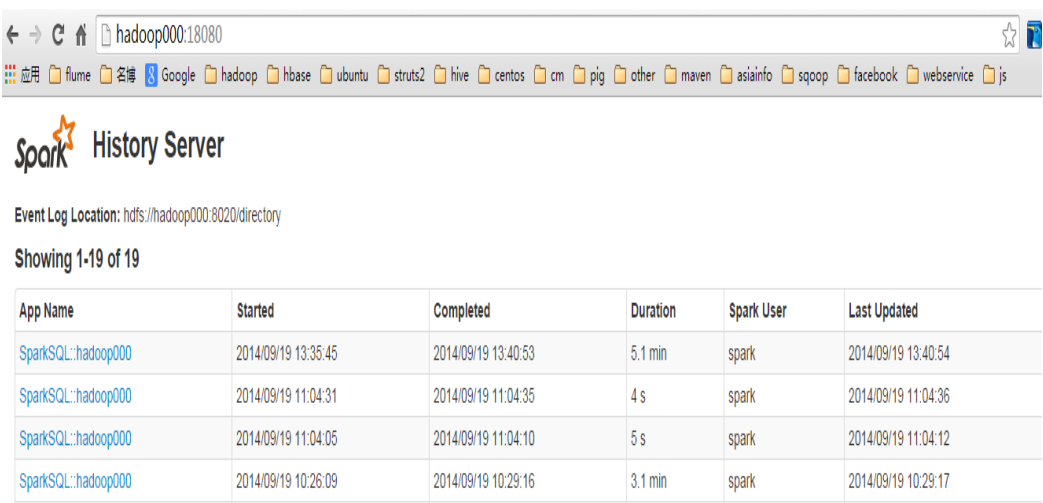
Windows下IntelliJ IDEA中调试S

yy2314703: 代码一样, 集群没问题. 第一个pi的例子我这一跑就提示集群没有资源 master设为local就没问...

[原]Spark MLlib Deep Learning

suihenglight: 1065075152@qq.com

默认界面列表信息是空的，下面截图是我跑了几次spark-sql测试后出现的。



history server相关的配置参数描述

- 1) spark.history.updateInterval
默认值：10
以秒为单位，更新日志相关信息的时间间隔
- 2) spark.history.retainedApplications
默认值：50
在内存中保存Application历史记录个数，如果超过这个值，旧的应用程序信息将被删除，当再次访问已被删除的应用信息时需要重新构建页面。
- 3) spark.history.ui.port
默认值：18080
HistoryServer的web端口
- 4) spark.history.kerberos.enabled
默认值：false
是否使用kerberos方式登录访问HistoryServer，对于持久层位于安全集群的HDFS上是有用的，如果设置为true，就要配置下面的两个属性
- 5) spark.history.kerberos.principal
默认值：用于HistoryServer的kerberos主体名称
- 6) spark.history.kerberos.keytab
用于HistoryServer的kerberos keytab文件位置
- 7) spark.history.ui.acls.enable
默认值：false
授权用户查看应用程序信息的时候是否检查acl。如果启用，只有应用程序所有者和spark.ui.view.acls指定的用户可以查看应用程序信息;否则，不做任何检查
- 8) spark.eventLog.enabled
默认值：false
是否记录Spark事件，用于应用程序在完成重构webUI
- 9) spark.eventLog.dir
默认值：file:///tmp/spark-events
保存日志相关信息的路径，可以是hdfs://开头的HDFS路径，也可以是file://开头的本地路径，都需要提前创建
- 10) spark.eventLog.compress
默认值：false
是否压缩记录Spark事件，前提spark.eventLog.enabled为true，默认使用的是snappy

[原]Spark MLlib Deep Learning
suifenglight: 大神，为什么有很多内容都是空白呢？可以发我QQ邮箱吗？麻烦了！！！！

Windows下IntelliJ IDEA中调试S
guotong1988: 学习

Windows下IntelliJ IDEA中调试S
cughuhao: 学习了，谢谢~

以spark.history开头的需要配置在spark-env.sh中的SPARK_HISTORY_OPTS，以spark.eventLog开头的配置在spark-defaults.conf

我在测试过程中的配置如下：

spark-defaults.conf

```
spark.eventLog.enabled true
spark.eventLog.dir hdfs://hadoop000:8020/directory
spark.eventLog.compress true
```

spark-env.sh

```
export SPARK_HISTORY_OPTS="-Dspark.history.ui.port=7777 -Dspark.history.retainedApplications=3 -Dspark.history.fs.logDirectory=hdfs://hadoop000:8020/directory"
```

参数描述：

spark.history.ui.port=7777 调整WEBUI访问的端口号为7777

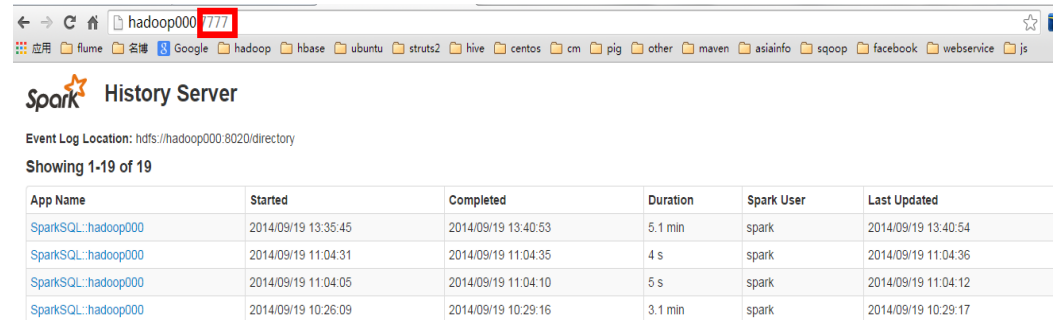
spark.history.fs.logDirectory=hdfs://hadoop000:8020/directory 配置了该属性后，在start-history-server.sh时就无需再显示的指定路径

spark.history.retainedApplications=3 指定保存Application历史记录个数，如果超过这个值，旧的应用程序信息将被删除

调整参数后启动start-history-server.sh

start-history-server.sh

访问WEBUI： http://hadoop000:7777



在使用spark history server的过程中产生的几个疑问：

疑问1： spark.history.fs.logDirectory和spark.eventLog.dir指定目录有啥区别？

经测试后发现：

spark.eventLog.dir： Application在运行过程中所有的信息均记录在该属性指定的路径下；

spark.history.fs.logDirectory： Spark History Server页面只展示该指定路径下的信息；

比如： spark.eventLog.dir刚开始时指定的是hdfs://hadoop000:8020/directory，而后修改成hdfs://hadoop000:8020/directory2

那么spark.history.fs.logDirectory如果指定的是hdfs://hadoop000:8020/directory，就只能显示出该目录下的所有Application运行的日志信息；反之亦然。

疑问**2**： **spark.history.retainedApplications=3** 貌似没生效？ ？ ？ ？ ？

The History Server will list all applications. It will just retain a max number of them in memory. That option does not control how many applications are show, it controls how much memory the HS will need.

注意：该参数并不是也页面中显示的**application**的记录数，而是存放在内存中的个数，内存中的信息在访问页面时直接读取渲染既可；

比如说该参数配置了**10**个，那么内存中就最多只能存放**10**个**applicaiton**的日志信息，当第**11**个加入时，第一个就会被踢除，当再次访问第**1**个**application**的页面信息时就需要重新读取指定路径上的日志信息来渲染展示页面。

详见官方文档：<http://spark.apache.org/docs/latest/monitoring.html>

分类: [Spark](#)

标签: [spark](#)

顶 0 踩 0

上一篇 [spark 案例集群测试整理](#)

下一篇 [初学HBase的几个问题](#)

我的同类文章

大数据（121）

- clouderamagner 中hue 工作流调用sqoop1 问题处理
- java连接hdfs ha和调用mapreduce jar示例
- Hive 使用解析Json 文件格式
- Hadoop NameNode 高可用 (High Availability) 实...
- Host Monitor和服务 Monitor服务无法启动，可...

- 带有HA功能的Hadoop Client端RPC实现原理与代...
- client如何访问HA HDFS
- 如何在Hive中使用Json格式数据
- hadoop 参数配置
- 使用 Spring、Elasticsearch 及 Logstash 构建企业...

更多

主题推荐 spark url ui server 应用 重构 界面

猜你在找

- 移动APP测试基础到进阶

Spark零基础入门（1）：Scala基本数据类型及程序控制

CSDN“攒课”第二期：Spark基础与编程模型

Spark应用实战

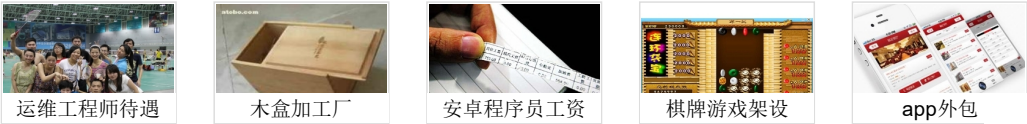
Linux快速学习以及监控分析实战【小强测试出品】
- C# WinForm 应用程序的 SQL Server 连接配置界面 动

spark SQL Running the Thrift JDBCDBC server

spark_job_server

spark解决beeline连接thrift-server加载数据权限问题

Apache Spark源码走读之18 —— 使用IntelliJ idea调试



查看评论

暂无评论

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

核心技术类目

全部主题

Hadoop

AWS

移动游戏

Java

Android

iOS

Swift

智能硬件

Docker

OpenStack

公司简介 | 招贤纳士 | 广告服务 | 银行汇款帐号 | 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

400-600-2320 | 北京创新乐知信息技术有限公司 版权所有 | 江苏乐知网络技术有限公司 提供商务支持

京 ICP 证 09002463 号 | Copyright © 1999-2014, CSDN.NET, All Rights Reserved



12