

# TensorflowOnSpark : 1)Standalone集群初体验

2

2017年06月05日 19:28:46 标签 : tensorflow / spark

3160

## 实验环境

和Hadoop的集群搭建网上教程比较多，这里以最简洁的方法配置集群，针对tensorflow添加的额外配置，  
手进行强调(其实地上本没有坑，跌的人多了，也便成了Keng)

### 环境环境变量

```
export JAVA_HOME=/hadoop/jdk1.8.0_65
export HADOOP_HOME=/hadoop/hadoop-2.7.0
export SPARK_HOME=/hadoop/spark-1.6.0
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$SPARK_HOME/bin
CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar
```

在/etc/profile或者~/.bashrc中配置都行，CLASSPATH不能少(Keng1)

### 2>hadoop集群

需修改的配置文件都在\$HADOOP\_HOME/hadoop-2.7.0/etc/hadoop目录下

(1)修改hadoop-env.sh 文件

```
1 export JAVA_HOME=/hadoop/jdk1.8.0_65
```

(2)修改core-site.xml

```
1 <configuration>
2   <property>
3     <name>fs.defaultFS</name>
4     <value>hdfs://master:9000</value>
5   </property>
6   <property>
7     <name>hadoop.tmp.dir</name>
8     <value>/hadoop/hadoop-2.7.0/tmp</value>
9   </property>
10 </configuration>
```

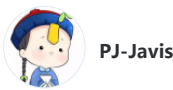
注意这里我把hdfs的namenode也设置在master节点上，hadoop.tmp.dir为hadoop的绝对路径

(3)修改文件hdfs-site.xml

```
1 <configuration>
2   <property>
3     <name>dfs.replication</name>
4     <value>2</value>
5   </property>
6   <property>
7     <name>dfs.namenode.name.dir</name>
8     <value>/hadoop/hadoop-2.7.0/dfs/name</value>
9   </property>
10  <property>
11    <name>dfs.datanode.data.dir</name>
12    <value>/hadoop/hadoop-2.7.0/dfs/data</value>
13  </property>
14 </configuration>
```

(4)修改slaves文件,配置DataNode节点地址

这里的hosts我已经配置好，所以输入你对应的hostname就行了



|     |    |    |
|-----|----|----|
| 原创  | 粉丝 | 喜欢 |
| 158 | 72 | 45 |



等级 : 博客 5 访问量 : 2  
积分 : 3868 排名 : 1万



### 人脸识别算法



### 他的最新文章

- pytorch笔记 : 04)resnet网络
- pytorch笔记 : 03)softmax和log x, 以及CrossEntropyLoss
- pytorch笔记02)模型的保存和加载
- BWA0.7+Samtools1.5+GATK4 数据集上的试验
- BWA0.7+Samtools1.5+GATK4 数据集上的试验

### 文章分类

- 数据结构和算法
- Scala
- Java
- Hadoop
- 数据库
- Spark

展开

### 文章存档

- 2018年3月
- 2018年2月
- 2017年8月
- 2017年7月
- 2017年6月

(5)格式化namenode并启动hdfs

```
1  hdfs namenode -format
2  $HADOOP_HOME/sbin/start-dfs.sh
```

3 点赞 1 收藏 1 评论 1 打赏

### Spark集群

Spark集群Standalone的配置非常简单，修改2个文件即可，在此之前记得重命名去掉template

(1)配置spark-env.sh

```
1  export JAVA_HOME=/hadoop/jdk1.8.0_65
2  export HADOOP_CONF_DIR=/hadoop/hadoop-2.7.0/etc/hadoop
3  export HADOOP_HDFS_HOME=/hadoop/hadoop-2.7.0
4  SPARK_MASTER_IP=master
5  SPARK_WORKER_CORES=4
6  SPARK_WORKER_MEMORY=12G
7  SPARK_EXECUTOR_MEMORY=8G
```

核数和内存根据自己的机器进行设置，环境变量HADOOP\_CONF\_DIR和HADOOP\_HDFS\_HOME不能少(Keng2)

(2)配置slaves

```
1  slave01
2  slave02
3  slave03
```

(3)启动spark集群

```
1  $SPARK_HOME/sbin/start-all.sh
```

| Worker Id | Cores      | Memory               |
|-----------|------------|----------------------|
| worker1   | 4 (0 Used) | 12.0 GB (0.0 B Used) |
| worker2   | 4 (0 Used) | 12.0 GB (0.0 B Used) |
| worker3   | 4 (0 Used) | 12.0 GB (0.0 B Used) |

集群总共3个worker-instance，每个worker4核12G，总12核，所有的环境配置均和master节点一致(Keng3)

## 2.Tensorflow安装

雅虎目前开源的框架是基于python2.7和Tensorflow0.12.1的，目前Tensorflow版本为1.2，但是考虑到兼容性，我们还是使用推荐的版本进行测试。

安装Tensorflow0.12.1

```
1  pip install https://storage.googleapis.com/tensorflow/linux/cpu/tensorflow-0.12.1-cp27-none-linux_x86_64.whl
```

测试tf，无异常说明安装成功

```
1  import tensorflow as tf
```

## 3.下载TensorflowOnSpark源码

```
1  git clone https://github.com/yahoo/TensorFlowOnSpark.git
2  cd TensorFlowOnSpark
3  export TFoS_HOME=$(pwd)
```

这里使用git进行下载，最后我会所有用到资源的百度云链接。顺便提一句，TensorflowOnSpark最近好像更新了，但是**指导文档**貌似有点问题，下面会进行说明。

### 他的热门文章

- Scala之旅-class和Object详解  
29297
- Spark核心RDD：combineByKey解  
11438
- Spark核心RDD：Sort排序详解  
10875
- Spark基础随笔：分区小结  
8863
- keras：4)LSTM函数详解  
7351
- keras：3)Embedding层详解  
7122
- SparkStream：4)foreachRDD詳  
5851
- Spark基础随笔：Spark1.6 Idea  
试的2种方法  
5326
- Spark机器学习：TF-IDF实例讲  
4947
- Spark机器学习：密集和稀疏向量  
4892



达内的



### 联系我们

请扫描二维码联系  
webmaster@csdn.net  
400-660-0101  
QQ客服 400-660-0101

关于 招聘 广告服务 注册  
©1999-2018 CSDN版权所有  
京ICP证09002463号

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

下载成功后，你会得到类似上面的文件夹，tfspark.zip是我们生成的python库文件，之后提交Spark的时候用到，其就是把tensorflowonspark所有文件进行了打包，在TensorFlowOnSpark目录运行如下的命令进行打包(Ke ng4)

```
1 zip -r tfspark.zip tensorflowonspark/*
```

## Spark集群测试

### 1 MNIST数据文件

```
 ${SPARK_HOME}/bin/spark-submit \  
 --master=local[*] \  
 ${TFoS_HOME}/examples/mnist/mnist_data_setup.py \  
--output examples/mnist/csv \  
5 --format csv
```

该命令的功能是把之前下载的.gz文件转换为对应的scv文件，网上有人说要修改源码才能正常运行，能偷懒的地方绝不放过。可以先看下mnist\_data\_setup.py的源码

```
1 def writeMNIST(sc, input_images, input_labels, output, format, num_partitions):  
2     """Writes MNIST image/label vectors into parallelized files on HDFS"""  
3     # load MNIST gzip into memory  
4     with open(input_images, 'rb') as f:  
5         images = numpy.array(mnist.extract_images(f))  
6         imageRDD = sc.parallelize(images.reshape(shape[0], shape[1] * shape[2]), num_partition  
7 s)  
8  
9 if not args.read:  
10     # Note: these files are inside the mnist.zip file  
11     writeMNIST(sc, "mnist/train-images-idx3-ubyte.gz", "mnist/train-labels-idx1-ubyte.gz",  
    args.output + "/train", args.format, args.num_partitions)  
    writeMNIST(sc, "mnist/t10k-images-idx3-ubyte.gz", "mnist/t10k-labels-idx1-ubyte.gz", ar  
gs.output + "/test", args.format, args.num_partitions)
```

使用python的IO流读取gz文件数据，显然gz文件肯定本地而非hdfs上，因此为了兼容源码，可以把mnist放在\$PARK\_HOME/bin下，然后使用本地模式进行数据转换即可

## Browse Directory

| Permission | Omer | Group      | Size    | Last Modified      | Replication | Block Size | Name       |
|------------|------|------------|---------|--------------------|-------------|------------|------------|
| -rw-r--r-- | root | supergroup | 0 B     | 2017/6/5 上午8:50:35 | 2           | 128 MB     | _SUCCESS   |
| -rw-r--r-- | root | supergroup | 1.73 MB | 2017/6/5 上午8:50:32 | 2           | 128 MB     | part-00000 |
| -rw-r--r-- | root | supergroup | 1.72 MB | 2017/6/5 上午8:50:32 | 2           | 128 MB     | part-00001 |
| -rw-r--r-- | root | supergroup | 1.73 MB | 2017/6/5 上午8:50:32 | 2           | 128 MB     | part-00002 |
| -rw-r--r-- | root | supergroup | 1.73 MB | 2017/6/5 上午8:50:32 | 2           | 128 MB     | part-00003 |
| -rw-r--r-- | root | supergroup | 1.73 MB | 2017/6/5 上午8:50:32 | 2           | 128 MB     | part-00004 |
| -rw-r--r-- | root | supergroup | 1.75 MB | 2017/6/5 上午8:50:32 | 2           | 128 MB     | part-00005 |
| -rw-r--r-- | root | supergroup | 1.76 MB | 2017/6/5 上午8:50:34 | 2           | 128 MB     | part-00006 |
| -rw-r--r-- | root | supergroup | 1.76 MB | 2017/6/5 上午8:50:34 | 2           | 128 MB     | part-00007 |
| -rw-r--r-- | root | supergroup | 1.77 MB | 2017/6/5 上午8:50:34 | 2           | 128 MB     | part-00008 |
| -rw-r--r-- | root | supergroup | 1.75 MB | 2017/6/5 上午8:50:34 | 2           | 128 MB     | part-00009 |

[illegible]

```

1 ${SPARK_HOME}/bin/spark-submit \
2 --master=spark://master:7077 \
3 --conf spark.executorEnv.LD_LIBRARY_PATH="${JAVA_HOME}/jre/lib/amd64/server" \
4 --conf spark.executorEnv.CLASSPATH="$($HADOOP_HOME/bin/hadoop classpath --glob):${CLASSPATH}" \
5 H}" \
6 --py-files ${TFoS_HOME}/examples/mnist/spark/mnist_dist.py,${TFoS_HOME}/tfspark.zip \
7 --conf spark.cores.max=12 \
8 --conf spark.task.cpus=4 \
9 ${TFoS_HOME}/examples/mnist/spark/mnist_spark.py \
10 --cluster_size 3 \
11 --images examples/mnist/csv/train/images \
12 --labels examples/mnist/csv/train/labels \
13 --format csv \
14 --mode train \
    --model mnist_model

```

-No.2\_因为本人的Spark集群和作者的不一样，这里建议设置`spark.cores.max`(集群总核数)和`spark.task.cpus`(worker节点分配核数)满足 $spark.cores.max / spark.task.cpus = workernumber$ ，否则会出现无尽等待的情况：

1 需要改一下mnist\_dist.py的第109行, 把logdir=logdir 改成 logdir=None

虽然解决了卡主的情况，但是训练完成后，不知道mnist\_model去哪了？本地和hdfs都找不到，继续执行测试集，发现准确度几乎是0%(○\_○)...之后查看worker的error日志，发现该信息一直都存在

```
1 INFO:tensorflow:Waiting for model to be ready. Ready_for_local_init_op: None, ready: Variables not initialized: hid_w, hid_b, sm_w, sm_b, Variable, hid_w/Adagrad, hid_b/Adagrad, sm_w/Adagrad, sm_b/Adagrad
2 2017-06-05 05:00:55,324 INFO (MainThread-31600) Waiting for model to be ready. Ready_for_local_init_op: None, ready: Variables not initialized: hid_w, hid_b, sm_w, sm_b, Variable, hid_w/Adagrad, hid_b/Adagrad, sm_w/Adagrad, sm_b/Adagrad
```

最终在github上找到了解决方法，原来python在写hdfs文件的时候，找不到对应的jar包，在提交的时候添加如下的配置信息

```
1 --conf spark.executorEnv.LD_LIBRARY_PATH="${JAVA_HOME}/jre/lib/amd64/server" \
2 --conf spark.executorEnv.CLASSPATH="($HADOOP_HOME/bin/hadoop classpath --glob):${CLASSPATH}" \
```

最终可以解决卡住的情况，终于在hdfs上面和model相遇了O(∩\_∩)O~

👍

2

⋮

🔖

💬

| /user/root/mnist_model |       |            |           |                    |             |            |  | Go! |
|------------------------|-------|------------|-----------|--------------------|-------------|------------|--|-----|
| Permission             | Owner | Group      | Size      | Last Modified      | Replication | Block Size | Name                                   |     |
| -rw-r--r--             | root  | supergroup | 218 B     | 2017/6/5 下午1:40:54 | 2           | 128 MB     | checkpoint                             |     |
| -rw-r--r--             | root  | supergroup | 84.91 KB  | 2017/6/5 下午1:41:01 | 2           | 128 MB     | events.out.tfevents.1496639022.slave00 |     |
| -rw-r--r--             | root  | supergroup | 85 KB     | 2017/6/5 下午1:47:56 | 2           | 128 MB     | events.out.tfevents.1496639449.slave00 |     |
| -rw-r--r--             | root  | supergroup | 116.14 KB | 2017/6/5 下午1:47:21 | 2           | 128 MB     | graph.pbtxt                            |     |
| -rw-r--r--             | root  | supergroup | 795.08 KB | 2017/6/5 下午1:40:24 | 2           | 128 MB     | model.ckpt-0.data-00000-of-00001       |     |
| -rw-r--r--             | root  | supergroup | 372 B     | 2017/6/5 下午1:40:24 | 2           | 128 MB     | model.ckpt-0.index                     |     |
| -rw-r--r--             | root  | supergroup | 44.78 KB  | 2017/6/5 下午1:40:25 | 2           | 128 MB     | model.ckpt-0.meta                      |     |
| -rw-r--r--             | root  | supergroup | 795.08 KB | 2017/6/5 下午1:40:24 | 2           | 128 MB     | model.ckpt-147.data-00000-of-00001     |     |
| -rw-r--r--             | root  | supergroup | 372 B     | 2017/6/5 下午1:40:34 | 2           | 128 MB     | model.ckpt-147.index                   |     |
| -rw-r--r--             | root  | supergroup | 44.78 KB  | 2017/6/5 下午1:40:35 | 2           | 128 MB     | model.ckpt-147.meta                    |     |
| -rw-r--r--             | root  | supergroup | 795.08 KB | 2017/6/5 下午1:40:44 | 2           | 128 MB     | model.ckpt-289.data-00000-of-00001     |     |
| -rw-r--r--             | root  | supergroup | 372 B     | 2017/6/5 下午1:40:44 | 2           | 128 MB     | model.ckpt-289.index                   |     |
| -rw-r--r--             | root  | supergroup | 44.78 KB  | 2017/6/5 下午1:40:45 | 2           | 128 MB     | model.ckpt-289.meta                    |     |
| -rw-r--r--             | root  | supergroup | 795.08 KB | 2017/6/5 下午1:40:54 | 2           | 128 MB     | model.ckpt-434.data-00000-of-00001     |     |
| -rw-r--r--             | root  | supergroup | 372 B     | 2017/6/5 下午1:40:54 | 2           | 128 MB     | model.ckpt-434.index                   |     |
| -rw-r--r--             | root  | supergroup | 44.78 KB  | 2017/6/5 下午1:40:55 | 2           | 128 MB     | model.ckpt-434.meta                    |     |

http://blog.csdn.net/jiangpeng59

3>模型测试

如此多的Keng 做铺垫，测试的时候终于一气呵成了！

```
1 ${SPARK_HOME}/bin/spark-submit \  
2 --master spark://master:7077 \  
3 --conf spark.executorEnv.LD_LIBRARY_PATH="${JAVA_HOME}/jre/lib/amd64/server" \  
4 --conf spark.executorEnv.CLASSPATH="(${HADOOP_HOME}/bin/hadoop classpath --glob):${CLASSPATH}" \  
5 H}" \  
6 --py-files ${TFoS_HOME}/tfspark.zip,${TFoS_HOME}/examples/mnist/spark/mnist_dist.py \  
7 --conf spark.cores.max=12 \  
8 --conf spark.task.cpus=4 \  
9 --conf spark.executorEnv.JAVA_HOME="${JAVA_HOME}" \  
10 ${TFoS_HOME}/examples/mnist/spark/mnist_spark.py \  
11 --cluster_size 3 \  
12 --images examples/mnist/csv/test/images \  
13 --labels examples/mnist/csv/test/labels \  
14 --mode inference \  
15 --format csv \  
16 --model mnist_model \  
--output predictions
```


部分结果如下：

```
1 [root@slave01 ~]# hadoop fs -cat /user/root/predictions/part-00000  
2 2017-06-05T05:48:00.385513 Label: 7, Prediction: 7  
3 2017-06-05T05:48:00.385574 Label: 2, Prediction: 2  
4 2017-06-05T05:48:00.385591 Label: 1, Prediction: 1  
5 2017-06-05T05:48:00.385625 Label: 0, Prediction: 0  
6 2017-06-05T05:48:00.385639 Label: 4, Prediction: 4  
7 2017-06-05T05:48:00.385653 Label: 1, Prediction: 1  
8 2017-06-05T05:48:00.385667 Label: 4, Prediction: 4  
9 2017-06-05T05:48:00.385680 Label: 9, Prediction: 9  
10 2017-06-05T05:48:00.385697 Label: 5, Prediction: 6  
11 2017-06-05T05:48:00.385711 Label: 9, Prediction: 9  
12 2017-06-05T05:48:00.385724 Label: 0, Prediction: 0  
13 2017-06-05T05:48:00.385736 Label: 6, Prediction: 6  
14 2017-06-05T05:48:00.385749 Label: 9, Prediction: 9  
15 2017-06-05T05:48:00.385762 Label: 0, Prediction: 0  
16 2017-06-05T05:48:00.385775 Label: 1, Prediction: 1  
17 2017-06-05T05:48:00.385788 Label: 5, Prediction: 5
```


解铃还须系铃人，问题来于Git解决于Git  
tf百度资源：<http://pan.baidu.com/s/1bpEhPHP>


参考：


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/jiangpeng59/article/details/72867368>





2

 目前您尚未登录，请 [登录](#) 或 [注册](#) 后进行评论



 [veixin\\_33563949](#) 2017-11-27 11:29 #1楼 [回复](#)

 设置了4/1=4, 还是无尽等待什么原因？



```
-conf spark.cores.max=4 \  
-conf spark.task.cpus=1 \  
--conf spark.executorEnv.JAVA_HOME=&quot;$JAVA_HOME&quot; \  
${TFoS_HOME}/examples/mnist/spark/mnist_spark.py \  
--cluster_size 4 \  

```

### TensorflowOnSpark 介绍与搭建


1.介绍TensorFlowOnSpark 为 Apache Hadoop 和 Apache Spark 集群带来可扩展的深度学习。通过结合深度学习框架 Tensor Flow 和大数据框架 Apac...

 [sinat\\_34233802](#) 2017年04月01日 11:41 

📖 4835

### 如何安装Spark & TensorflowOnSpark

JAVA+HADOOP+SCALA+SPARK+TensorFlowOnSpark的安装教程

 [FishSeeker](#) 2017年03月13日 15:37 

📖 9432

### 程序员不会英语怎么行？

老司机教你一个数学公式秒懂天下英语

广告



### TensorflowOnSpark 安装

Tensorflow on Spark 安装

 [hjh00](#) 2017年03月21日 13:14 

📖 3726

### TensorFlowOnSpark 初体验

点击下方链接阅读原文：TensorFlowOnSpark 初体验

 [dxmkkk](#) 2017年03月13日 09:18 

📖 1568

[传播公益品牌，支持慈善事业](#)  
为公益机构提供免费推广服务  
[open.baidu.com](http://open.baidu.com)

### Tensorflow on Spark爬坑指南

北京 上海巡回站 | NVIDIA DLI深度学习培训2018年1月26/1月12日NVIDIA 深度学习学院 带你快速进入火热的DL领域阅读全文


...

 [leadai](#) 2018年01月03日 00:00 

📖 169

### TensorflowOnSpark-集群版


参考：[https://github.com/yahoo/TensorFlowOnSpark/wiki/GetStarted\\_YARN](https://github.com/yahoo/TensorFlowOnSpark/wiki/GetStarted_YARN) <http://www.cnblogs.com/heimians...>

 [wc781708249](#) 2017年10月17日 08:43 

📖 435

### 关于建立tensorflowOnspark时候，需要做的环境搭建，看如下链接的文章

这是参考的搭建配置环境

 [u012869364](#) 2017年10月09日 10:45 

📖 202

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！



## 分布式TensorFlow：在Spark上将谷歌的深度学习库进行尺度变换

本文为数盟原创译文，转载请注明出处为“数盟社区”。介绍 Arimo的日益增长的数据科学团队包括研究和开发机器学习和深入学习新的方法和应用。 我们正在调查的一个主题是分布式的深度学习...

 u013886628

2016年07月04日 11:26

 8027

## 如何安装Spark & TensorflowOnSpark

对的 你没看错，这是我的一条龙服务，我在入坑填坑无数之后终于成功搭建起了Spark和TensorflowOnSpark的运行环境，并成功：三 示例程序（大概就是手写识别的训练和识别吧）。安装JAV...

 u010159842

2017年06月06日 14:52

 1165

## 雅虎开源TensorflowOnSpark，Ubuntu配置安装TensorflowOnSpark

1、 Yahoo在2016年2月左右开源了基于Spark / Hadoop的分布式深度学习框架—CaffeOnSpark，CaffeOnSpark被设计成为一个Spark深度学习包，Spark ML...

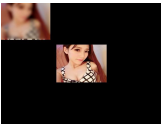
 u013041398

2017年02月15日 12:35

 2105

## 技术外文文献看不懂？教你一个公式秒懂英语

不背单词和语法，一个公式学好英语



## TensorFlowOnSpark 源码解析

 qq\_39161737

2017年10月23日 12:07

 264

这两天琢磨了下spark-deep-learning和spark-sklearn两个项目，但是感觉都不尽人如意。在training时，都需要把数据broadcast到各个节点进行并行训练，基本就失去实...

## Centos6安装TensorFlow及TensorFlowOnSpark

1. 需求描述在Centos6系统上安装Hadoop、Spark集群，并使用TensorFlowOnSpark的 YARN运行模式下执行TensorFlow的代码。（最好可以在不联网的集群中进行配置并...

 fansy1990

2017年10月27日 23:51

 463

## Spark(一): 基本架构及原理

 swing2008

2017年03月08日 11:26

 23215

Apache Spark是一个围绕速度、易用性和复杂分析构建的大数据处理框架，最初在2009年由加州大学伯克利分校的AMPLab开发，并于2010年成为Apache的开源项目之一，与Hadoop和St...

## spark使用文档

 qq\_30408111

2018年01月26日 14:19

 883

最近在看了一些spark的使用文档，以及官网简介，自己总结了一点使用文档，记录一下 快速启动spark 关于这一部分 Apache Spark 是专为大规模数据处理而设计的快速通用的计算引擎。S...

## Spark编程模型经典解析（一）

 WYpersist

2018年03月23日 19:16

 93

从Hadoop MR到Spark回顾hadoop -MR计算过程1. 文件 通过 split，split的原因：MapReduce是并行计算的，如果一个文件很大，500G的大小，做不了分片，那如何并...

## 社保代缴只需20元,正规可靠!

正规社保代理机构,10W+ 客户选择,20元起搞定社保代缴/补缴/开户/挂靠

百度广告



## Spark是什么？

 caiandyong

2015年10月15日 15:36

 3647

Spark是个通用的集群计算框架，通过将大量数据集计算任务分配到多台计算机上，提供高效内存计算。如果你熟悉Hadoop，那么你知道分布式计算框架要解决两个问题：如何分发数据和如何分发计算。Hadoo...

spark sql在项目中使用越来越多，spark sql都支持哪些功能？官网没有明确说明，只能在class SqlParserd代码中看到一些Keyword，所以准备测试下spark对常用sql的支...

Spark【案例】实现黑名单实时过滤

qq\_21989939 2018年03月16日 10:33 123

黑名单实时过滤一、实验介绍1.1 实验内容本节课主要讲解 Spark 的 RDD 操作，让您对 Spark 算子的特性快速了解。通过演示案例：黑名单过滤，让您亲身体会到 RDD 的强大功能，然后学以...



第...：SparkSQL JDBC实战详解学习笔记

slq1023 2016年04月08日 23:52 4381

第...：SparkSQL JDBC实战详解学习笔记 本期内容：1 SparkSQL操作关系型数据库的意义 2 SparkSQL操作关系型数据库实战 使用Spark通过JDBC操作...

