

Spark 0.9.1 Standalone模式分布式部署

在本博客的[《Spark 0.9.1源码编译》](#)和[《Spark源码编译遇到的问题解决》](#)两篇文章中，分别讲解了如何编译Spark源码以及在编译源码过程中遇到的一些问题及其解决方法。今天来说说如何部署分布式的Spark集群，在本篇文章中，我主要是介绍如何部署Standalone模式。

一、修改配置文件

1、将\$SPARK_HOME/conf/spark-env.sh.template文件复制一份到spark-env.sh，并作以下修改（可选）：

```
export SCALA_HOME=/export1/spark/scala-2.10.3
export HADOOP_HOME=/home/q/hadoop-2.2.0
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
```

```
SPARK_WORKER_INSTANCES=3
SPARK_MASTER_PORT=8081
SPARK_MASTER_WEBUI_PORT=8090（这个默认端口是8080）
SPARK_WORKER_PORT=8091
SPARK_MASTER_IP=192.168.24.72
SPARK_WORKER_DIR=/export1/spark/worker
```

2、在\$SPARK_HOME/conf目录下新建slaves文件，这个是worker节点，每行一个节点：

```
worker1
worker2
worker3
```

二、打包\$SPARK_HOME目录下的文件，并分发到各个子节点

```
[wyp@master spark-0.9.1]$ tar -zcf spark-0.9.1.tar.gz spark-0.9.1
[wyp@master spark-0.9.1]$ scp spark-0.9.1.tar.gz worker1:/export1/spark
[wyp@master spark-0.9.1]$ scp spark-0.9.1.tar.gz worker2:/export1/spark
```

```
[wyp@master spark-0.9.1]$ scp spark-0.9.1.tar.gz worker3:/export1/spark
```

然后在worker1、worker2和worker3上的export1/spark目录下解压刚刚发过来的spark-0.9.1.tar.gz。

spark-0.9.1需要用到scala-2.10.3，所以你需要在Spark所有节点上安装好：

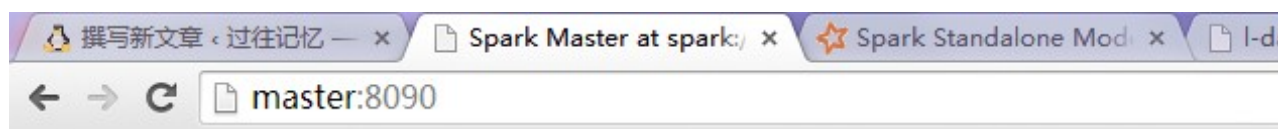
```
[wyp@master spark]$ wget http://www.scala-lang.org/files/archive/scala-2.10.3.tgz
[wyp@master spark]$ tar -zxf scala-2.10.3.tgz
[wyp@master spark]$ export SCALA_HOME=/export1/spark/scala-2.10.3
```

三、启动master和worker节点

如果你的master机器可以ssh登录到worker1、worker2和worker3，则可以用下面的脚本启动Spark master和worker节点：

```
[wyp@master spark-0.9.1]$ ./sbin/start-all.sh
```

然后可以在master:8090下看到以下页面：



Spark Master at spark://192.168.24.72:8081

URL: spark://192.168.24.72:8081

Workers: 9

Cores: 216 Total, 0 Used

Memory: 273.8 GB Total, 0.0 B Used

Applications: 0 Running, 1 Completed

Drivers: 0 Running, 0 Completed

Workers

Id
worker-20140421151419-worker1-55397
worker-20140421151530-worker1-46238
worker-20140421151537-worker1-33159
worker-20140421151628-worker2-27470
worker-20140421151636-worker2-50714
worker-20140421151646-worker2-35078
worker-20140421151704-worker3-27241
worker-20140421151710-worker3-28671
worker-20140421151717-worker3-23908

Spark web UI

如果你用./sbin/start-all.sh无法正常启动相关的进程，可以在\$SPARK_HOME/logs目录下查看相关的错误信息。其实，你还可以像Hadoop一样单独启动相关的进程，在master节点上运行下面的命令：

```
[wyp@master spark-0.9.1]$ ./sbin/start-master.sh
```

然后在worker1、worker2和worker3上运行下面的命令：

```
[wyp@master spark-0.9.1]$ ./sbin/start-slave.sh 3 W  
spark://192.168.24.72:8081 W  
--webui-port 8093
```

注意上面的端口和\$SPARK_HOME/conf/spark-env.sh配置文件要一致。

你还可以使用下面命令启动Master节点：

```
[wyp@master spark-0.9.1]$ bin/spark-class org.apache.spark.deploy.master.Master
```

使用下面命令启动Worker：

```
bin/spark-class org.apache.spark.deploy.worker.Worker spark://iteblog.com:7077
```

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接：【】（）