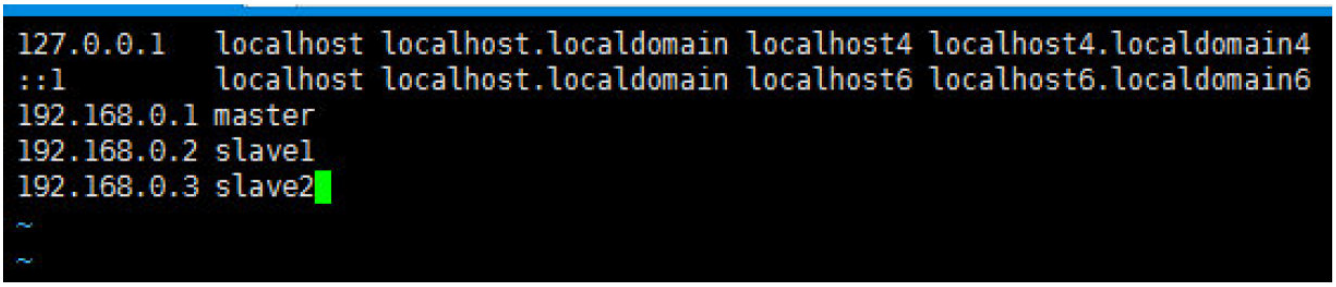
Spark StandAlone 部署

1. 配置host文件（所有节点）

```
vi /etc/hosts
192.168.0.1 master
192.168.0.2 slave1
192.168.0.3 slave2
```



2. 配置ssh

- 关闭防火墙（所有节点）

```
service iptables stop
chkconfig iptables off
```

- 生成分发ssh

```
ssh-keygen -t rsa
#一路回车
cat /root/.ssh/id_rsa.pub >> /root/.ssh/authorized_keys
scp /root/.ssh/authorized_keys root@slave1：/root/.ssh/
scp /root/.ssh/authorized_keys root@slave2：/root/.ssh/
```

- 设置ssh目录权限(所有节点)

```
chmod 700 /root/.ssh
chmod 600 /root/.ssh/authorized_keys
```

3. 安装java,scala（所有节点）

- 卸载open jdk

```
rpm -qa | grep java
```

显示如下信息（或其它）：

```
java-1.7.0-openjdk-1.7.0.45-2.4.3.3.el6.x86_64
tzdata-java-2013g-1.el6.noarch
java-1.6.0-openjdk-1.6.0.0-1.66.1.13.0.el6.x86_64
```

卸载：

```
rpm -e --nodeps java-1.7.0-openjdk-1.7.0.45-2.4.3.3.el6.x86_64
rpm -e --nodeps tzdata-java-2013g-1.el6.noarch
rpm -e --nodeps java-1.6.0-openjdk-1.6.0.0-1.66.1.13.0.el6.x86_64
```

- 解压

```
tar -zxvf jdk-8u101-linux-x64.tar.gz -C /home/cloud
tar -zxvf scala-2.10.6.tgz -C /home/cloud
```

- 配置环境变量

```
vi /etc/profile
export JAVA_HOME=/home/cloud/jdk1.8.0_101
export SCALA_HOME=/home/cloud/scala-2.10.6
export JRE_HOME=/home/cloud/jdk1.8.0_101/jre
export CLASSPATH=.:$JAVA_HOME/lib:$JRE_HOME/lib:$CLASSPATH
export PATH=$JAVA_HOME/bin:$JRE_HOME/bin:$SCALA_HOME/bin:$PATH
```

- 使环境变量生效

```
source /etc/profile
```

- 查看是否安装成功，版本号是否与安装一致

```
java -version
scala -version
```

4. 安装spark

- 解压安装包到指定目录

```
tar -zxvf spark-1.4.0-bin-hadoop2.6.tgz -C /home/cloud
```

- 修改配置文件

```
cd /home/cloud/spark-1.4.0-bin-hadoop2.6/conf
# 配置从节点
cp slaves.template slaves
vi slaves
slave

# 配置默认环境
cp spark-env.sh.template spark-env.sh
vi spark-env.sh
export JAVA_HOME=/home/cloud/jdk1.8.0_101
export SCALA_HOME=/home/cloud/scala-2.10.6
export HADOOP_HOME=/home/cloud/hadoop-2.6.0
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export SPARK_WORKER_OPTS="-Dspark.worker.cleanup.enabled=true
-Dspark.worker.cleanup.interval=864000 -Dspark.worker.cleanup.appDataTtl=864000"
export SPARK_LOCAL_HOSTNAME=`hostname`
```

- 分发spark

```
scp -r /home/cloud/spark-1.4.0-bin-hadoop2.6 root@slave1:/home/cloud/
scp -r /home/cloud/spark-1.4.0-bin-hadoop2.6 root@slave2:/home/cloud/
```

5. 启动spark

```
cd /home/cloud/spark-1.4.0-bin-hadoop2.6/sbin/
./start-all.sh
```

查看spark web ui界面： localhost:8080

Spark 1.4.0 **Spark Master at spark://DMMaster:7077**

URL: spark://DMMaster:7077
REST URL: spark://DMMaster:6066 *(cluster mode)*
Workers: 3
Cores: 24 Total, 0 Used
Memory: 43.7 GB Total, 0.0 B Used
Applications: 0 Running, 12 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address
worker-20170904143106-DMSlave1-34813	DMSlave1:34813
worker-20170904143106-DMSlave2-42074	DMSlave2:42074
worker-20170904143106-DMSlave3-52979	DMSlave3:52979

6. 测试spark

```
cd /home/cloud/spark-1.4.0-bin-hadoop2.6/bin
 ./spark-submit  --class org.apache.spark.examples.SparkPi
 --master spark://master:7077 ../lib/spark-examples-1.4.0-hadoop2.6.0.jar  4
```

```
17/09/19 08:57:35 INFO TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, 192.168.0.180, PROCESS_LOCAL, 1447 bytes)
17/09/19 08:57:35 INFO TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, 192.168.0.180, PROCESS_LOCAL, 1447 bytes)
17/09/19 08:57:35 INFO TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, 192.168.0.180, PROCESS_LOCAL, 1447 bytes)
17/09/19 08:57:35 INFO SparkDeploySchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@192.168.0.179:46012/user/Executor#-1700703654]) with ID 1
17/09/19 08:57:35 INFO SparkDeploySchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@192.168.0.184:38668/user/Executor#1272379333]) with ID 0
17/09/19 08:57:35 INFO SparkDeploySchedulerBackend: Registered executor: AkkaRpcEndpointRef(Actor[akka.tcp://sparkExecutor@192.168.0.185:48356/user/Executor#-518312216]) with ID 2
17/09/19 08:57:35 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.0.180:49884 with 265.1 MB RAM, BlockManagerId(3, 192.168.0.180, 49884)
17/09/19 08:57:35 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.0.179:52751 with 265.1 MB RAM, BlockManagerId(1, 192.168.0.179, 52751)
17/09/19 08:57:35 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.0.184:50865 with 265.1 MB RAM, BlockManagerId(0, 192.168.0.184, 50865)
17/09/19 08:57:35 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.0.185:58234 with 265.1 MB RAM, BlockManagerId(2, 192.168.0.185, 58234)
17/09/19 08:57:36 INFO BlockManagerInfo: Added broadcast_0_piece0 in memory on 192.168.0.180:49884 (size: 1202.0 B, free: 265.1 MB)
17/09/19 08:57:37 INFO TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 2259 ms on 192.168.0.180 (1/4)
17/09/19 08:57:37 INFO TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 2284 ms on 192.168.0.180 (2/4)
17/09/19 08:57:37 INFO TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 2315 ms on 192.168.0.180 (3/4)
17/09/19 08:57:37 INFO TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 2287 ms on 192.168.0.180 (4/4)
17/09/19 08:57:37 INFO DAGScheduler: ResultStage 0 (reduce at SparkPi.scala:35) finished in 3.434 s
17/09/19 08:57:37 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
17/09/19 08:57:37 INFO DAGScheduler: Job 0 finished: reduce at SparkPi.scala:35, took 3.760591 s
Pi is roughly 3.14022
17/09/19 08:57:37 INFO SparkUI: Stopped Spark web UI at http://192.168.0.178:4040
17/09/19 08:57:37 INFO DAGScheduler: Stopping DAGScheduler
17/09/19 08:57:37 INFO SparkDeploySchedulerBackend: Shutting down all executors
17/09/19 08:57:37 INFO SparkDeploySchedulerBackend: Asking each executor to shut down
17/09/19 08:57:37 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
17/09/19 08:57:37 INFO Utils: path = /home/develop/tmp/sparktmp/localdir/spark-f7623b67-5533-44bf-8588-f5c6a825d6c5/blockmgr-4264b0b1-07d3-4430-a900-54bb71873afd, already present as root for del
17/09/19 08:57:37 INFO MemoryStore: MemoryStore cleared
17/09/19 08:57:37 INFO BlockManager: BlockManager stopped
17/09/19 08:57:37 INFO BlockManagerMaster: BlockManagerMaster stopped
```

7. 关闭spark

```
cd /home/cloud/spark-1.4.0-bin-hadoop2.6/sbin/
./stop-all.sh
```

Spark程序环境

1. 运行环境

网络环境：能连接到集群

系统配置：hosts文件中配置hbase集群，spark集群的ip到主机名映射
语言环境: jdk 1.8, scala 2.10.6(如果用scala写程序)

2. 工程导入依赖jar包

> hbase jar包：hbase安装目录下的lib文件夹中
> guava-12.0.1.jar
> hbase-client-1.0.3.jar
> hbase-common-1.0.3.jar
> hbase-prefix-tree-1.0.3.jar
> hbase-protocol-1.0.3.jar
> hbase-server-1.0.3.jar
> htrace-core-3.1.0-incubating.jar
> httpclient-4.2.5.jar
> httpcore-4.1.3.jar
> zookeeper-3.4.6.jar

> spark jar包：spark安装目录下的lib文件夹中
> spark-assembly-1.4.0-hadoop2.6.0.jar

> scala sdk (如果用scala写程序，导入时选择scala安装目录)