Prepear

Download

```
wget --no-cookies --no-check-certificate --header "Cookie: gpw_e24=http%3A%2F%2Fwww.oracle.com%2F; oraclelicense=accept-securebackup-cookie" "http://download.oracle.com/otn-pub/java/jdk/7u79-b15/jdk-7u79-linux-x64.tar.gz"
wget --no-cookies --no-check-certificate --header "Cookie: gpw_e24=http%3A%2F%2Fwww.oracle.com%2F; oraclelicense=accept-securebackup-cookie" "http://download.oracle.com/otn-pub/java/jdk/8u66-b17/jdk-8u66-linux-x64.tar.gz"
wget https://dl.bintray.com/sbt/native-packages/sbt/0.13.9/sbt-0.13.9.tgz
wget http://ftp.jaist.ac.jp/pub/apache/maven/maven-3/3.3.9/binaries/apache-maven-3.3.9-bin.tar.gz
wget http://downloads.typesafe.com/scala/2.11.7/scala-2.11.7.tgz
wget http://d3kbcqa49mib13.cloudfront.net/spark-1.5.2-bin-hadoop2.6.tgz
wget http://ftp.riken.jp/net/apache/hadoop/common/hadoop-2.6.2/hadoop-2.6.2.tar.gz
wget http://archive.cloudera.com/cdh5/cdh/5/hadoop-2.6.0-cdh5.5.1.tar.gz
wget https://archive.apache.org/dist/spark/spark-2.1.0/spark-2.1.0-bin-hadoop2.7.tgz
```

unzip

```
tar -zxvf jdk-7u79-linux-x64.tar.gz
tar -zxvf jdk-8u66-linux-x64.tar.gz
tar -zxvf sbt-0.13.9.tgz
tar -zxvf apache-maven-3.3.9-bin.tar.gz
tar -zxvf scala-2.11.7.tgz
tar -zxvf spark-1.5.2-bin-hadoop2.6.tgz
tar -zxvf hadoop-2.6.2.tar.gz
tar -zxvf hadoop-2.6.0-cdh5.5.1.tar.gz
```

Spark Standalone

modify profile

```
vim ~/.bash_profile

export SOFT_BASE_PATH=/app/soft
# Spark Standalone
export SPARK_BASE_PATH=/app/spark-standalone
export JAVA_HOME=$SOFT_BASE_PATH/jdk1.8.0_66
export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar
export SCALA_HOME=$SOFT_BASE_PATH/scala-2.11.7
export SPARK_HOME=$SPARK_BASE_PATH/spark-1.6.0-bin-hadoop2.6
export HADOOP_HOME=$SOFT_BASE_PATH/hadoop-2.7.1
export HADOOP_CONF_DIR=$SOFT_BASE_PATH/hadoop-2.7.1/etc/hadoop
export PATH=$PATH:$JAVA_HOME/bin:$SCALA_HOME/bin:$SPARK_HOME/bin:$HADOOP_HOM
E/bin:$HADOOP_HOME/sbin
```

load profile

```
source ~/.bash_profile
```

确认java，scala环境

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

配置文件spark-env.sh

```
cp $SPARK_HOME/conf/spark-env.sh.template $SPARK_HOME/conf/spark-env.sh
vim $SPARK_HOME/conf/spark-env.sh
```

添加

```
export SCALA_HOME=/app/soft/scala-2.11.7
export SPARK_MASTER_IP=spark-centos-01
export SPARK_WORKER_MEMORY=2G
export JAVA_HOME=/app/soft/jdk1.8.0_66
# export HIVE_HOME=/app/soft/apache-hive-1.2.1-bin
# export SPARK_CLASSPATH=$HIVE_HOME/lib/mysql-connector-java-5.1.15-bin.jar:
$SPARK_CLASSPATH
```

配置文件slaves

```
cp $SPARK_HOME/conf/slaves.template $SPARK_HOME/conf/slaves
vim $SPARK_HOME/conf/slaves
```

在slaves最后添加下面

```
spark-centos-01
spark-centos-02
spark-centos-03
```

使用scp命令，将配置修改后的spark代码发送到其他节点（spark-centos-02、spark-centos-03）

```
scp -r /app/spark-standalone root@spark-centos-02:/app/
scp -r /app/spark-standalone root@spark-centos-03:/app/
```

启动停止命令

```
# 启动全部节点
$SPARK_HOME/sbin/start-all.sh
# 启动master
$SPARK_HOME/sbin/start-master.sh
# 启动worker
$SPARK_HOME/sbin/start-slaves.sh
# 停止全部节点
$SPARK_HOME/sbin/stop-all.sh
```

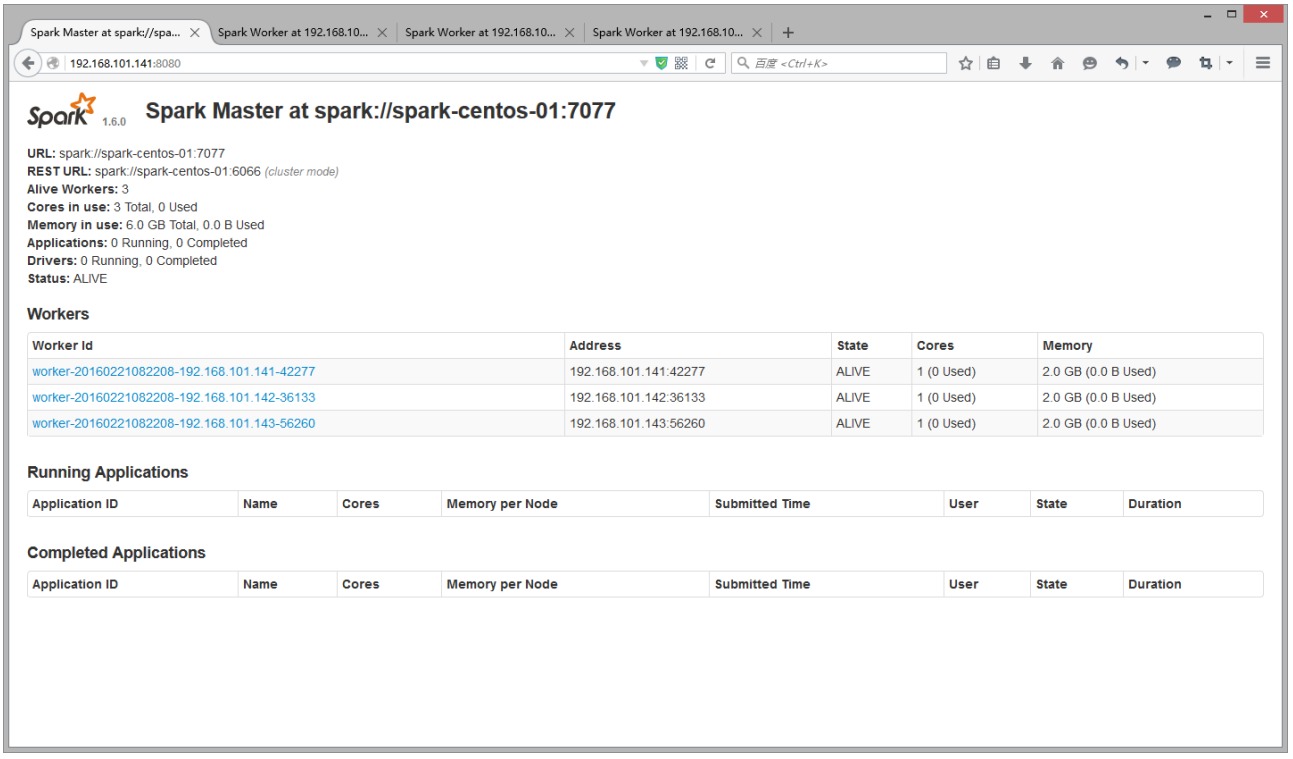
启动后的截图

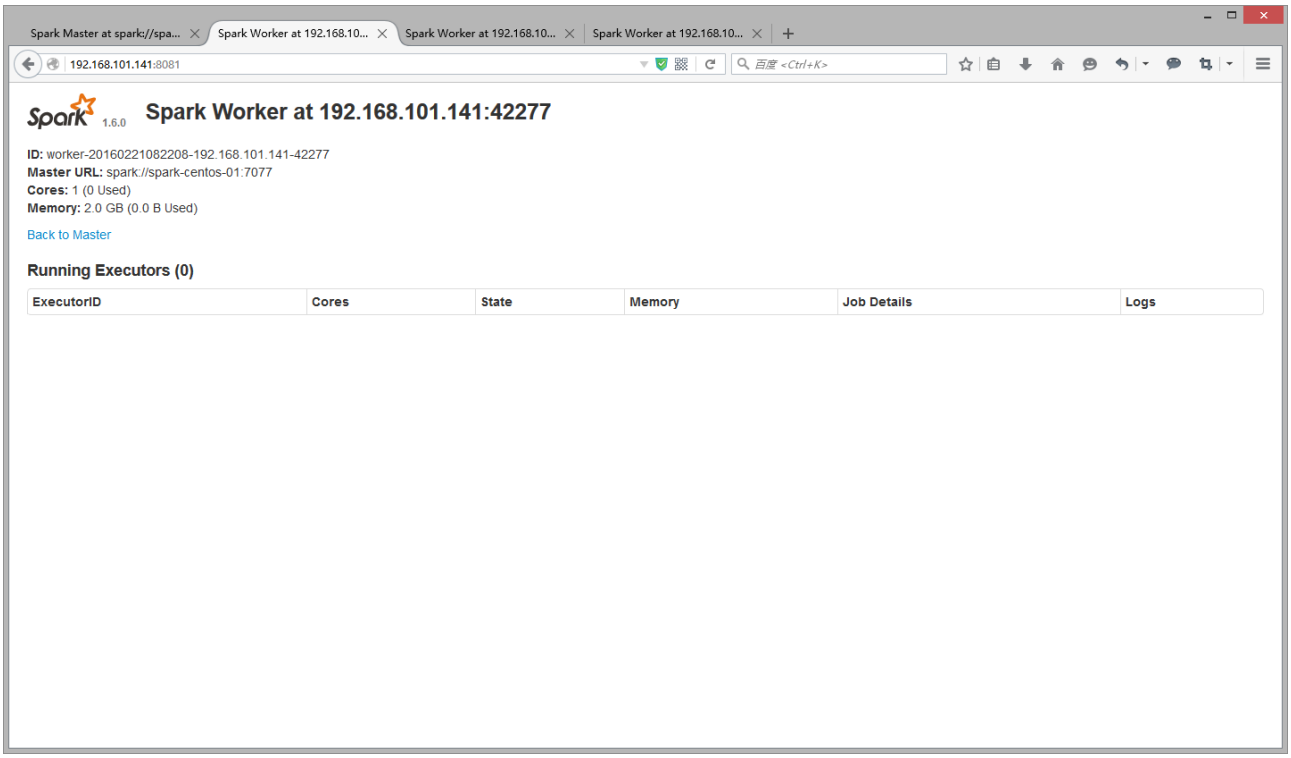
master节点：http://192.168.101.141:8080/

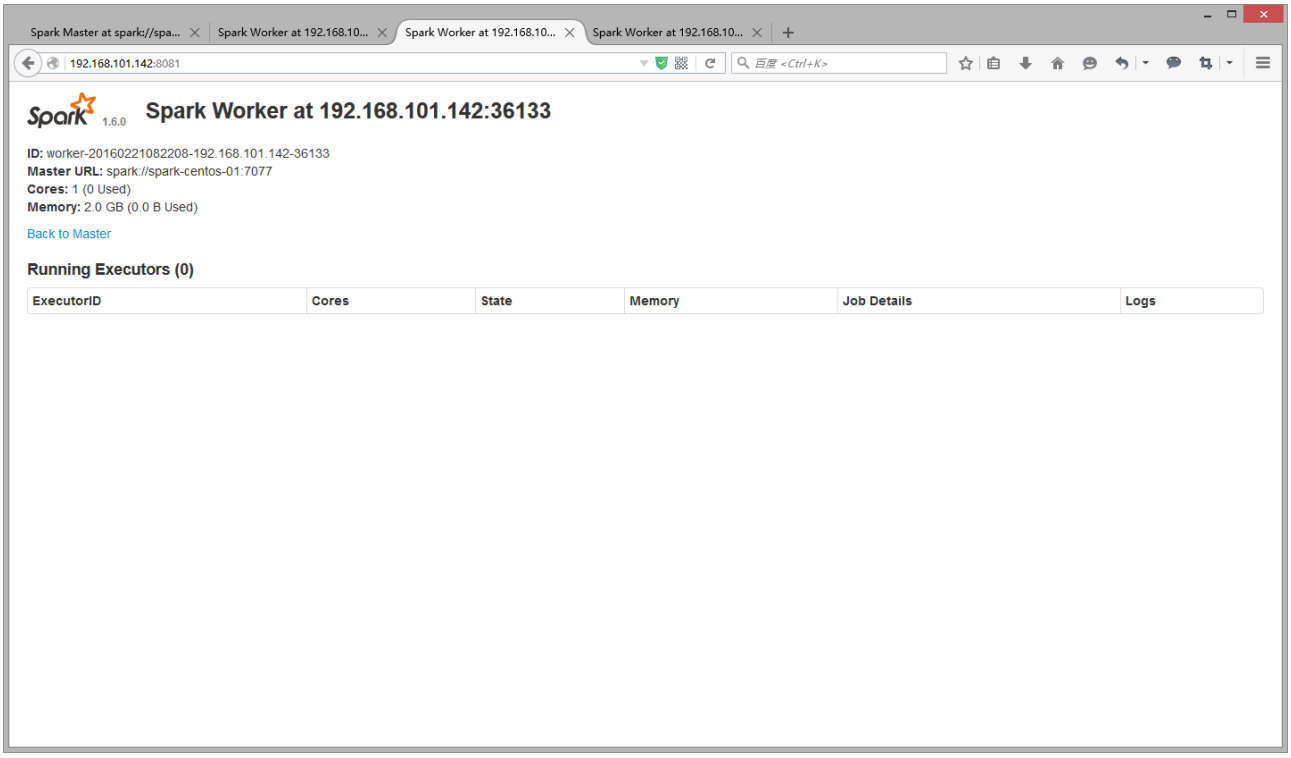
slave-01节点：http://192.168.101.141:8081/

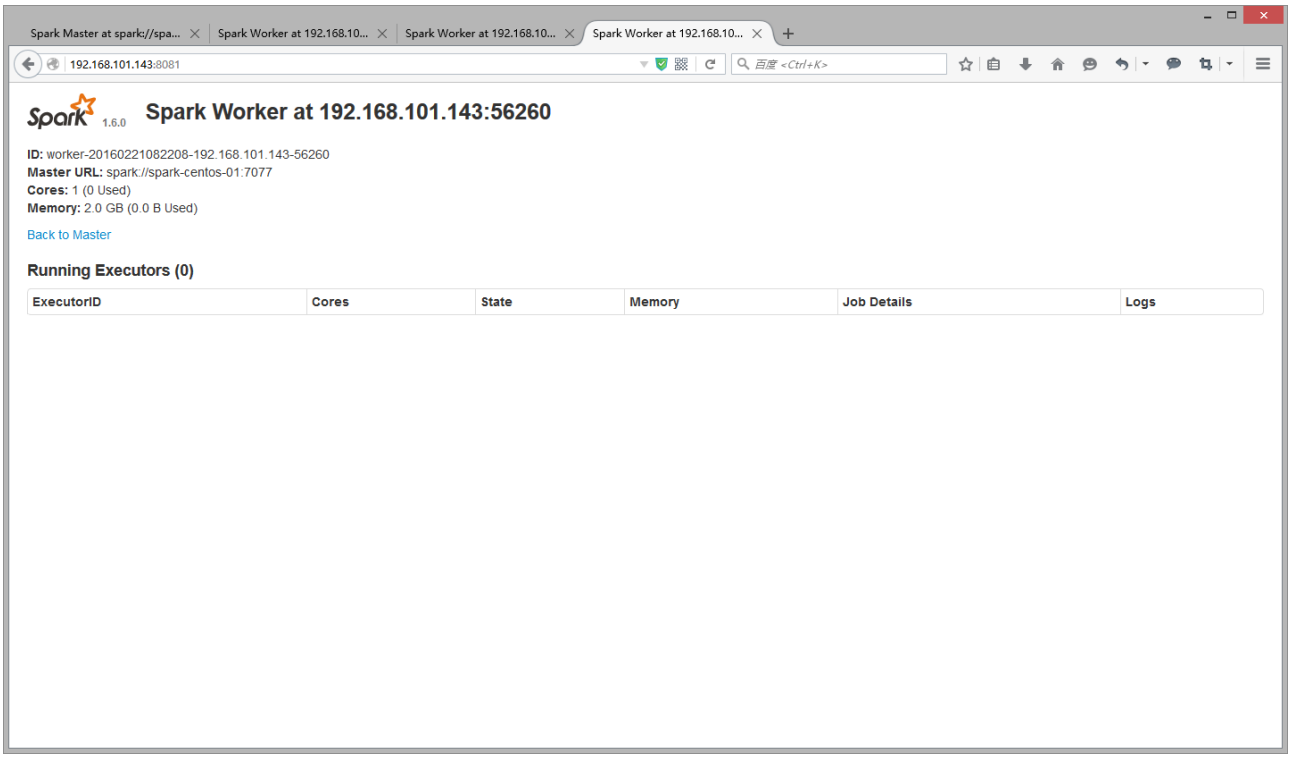
slave-02节点：http://192.168.101.142:8081/

slave-03节点：http://192.168.101.143:8081/




Spark Master at spark://spa...
Spark Worker at 192.168.10...
Spark Worker at 192.168.10...
Spark Worker at 192.168.10...
192.168.101.141:8081
Spark 1.6.0
Spark Worker at 192.168.101.141:42277
ID: worker-20160221082208-192.168.101.141-42277
Master URL: spark://spark-centos-01:7077
Cores: 1 (0 Used)
Memory: 2.0 GB (0.0 B Used)
Back to Master
Running Executors (0)
ExecutorID
Cores
State
Memory
Job Details
Logs


Spark Master at spark://spa...
Spark Worker at 192.168.10...
Spark Worker at 192.168.10...
Spark Worker at 192.168.10...
192.168.101.142:8081
Spark 1.6.0
Spark Worker at 192.168.101.142:36133
ID: worker-20160221082208-192.168.101.142-36133
Master URL: spark://spark-centos-01:7077
Cores: 1 (0 Used)
Memory: 2.0 GB (0.0 B Used)
Back to Master
Running Executors (0)
ExecutorID
Cores
State
Memory
Job Details
Logs



spark-shell

```
$SPARK_HOME/bin/spark-shell --master spark://spark-centos-01:7077
# -Dspark.master=spark://spark-centos-01:7077
# -Dspark.master=local
```

HelloWorld

```
# test localhost file
scala > val textFile = sc.textFile("file:///app/spark-standalone/spark-1.6.0-bin-hadoop2.6/README.md")
scala > textFile.count()
# test hdfs file
scala > val textFile = sc.textFile("hdfs://hadoop-centos-01:9000/input/README.txt")
scala > textFile.count()
```

spark-submit

```
$SPARK_HOME/bin/spark-submit --master spark://spark-centos-01:7077 --class org.apache.spark.examples.SparkPi --executor-memory 2g --total-executor-cores 2 lib/spark-examples-1.6.0-hadoop2.6.0.jar 1000
```

启动后的截图

master节点：http://192.168.101.141:8080/

slave-01节点：http://192.168.101.141:8081/

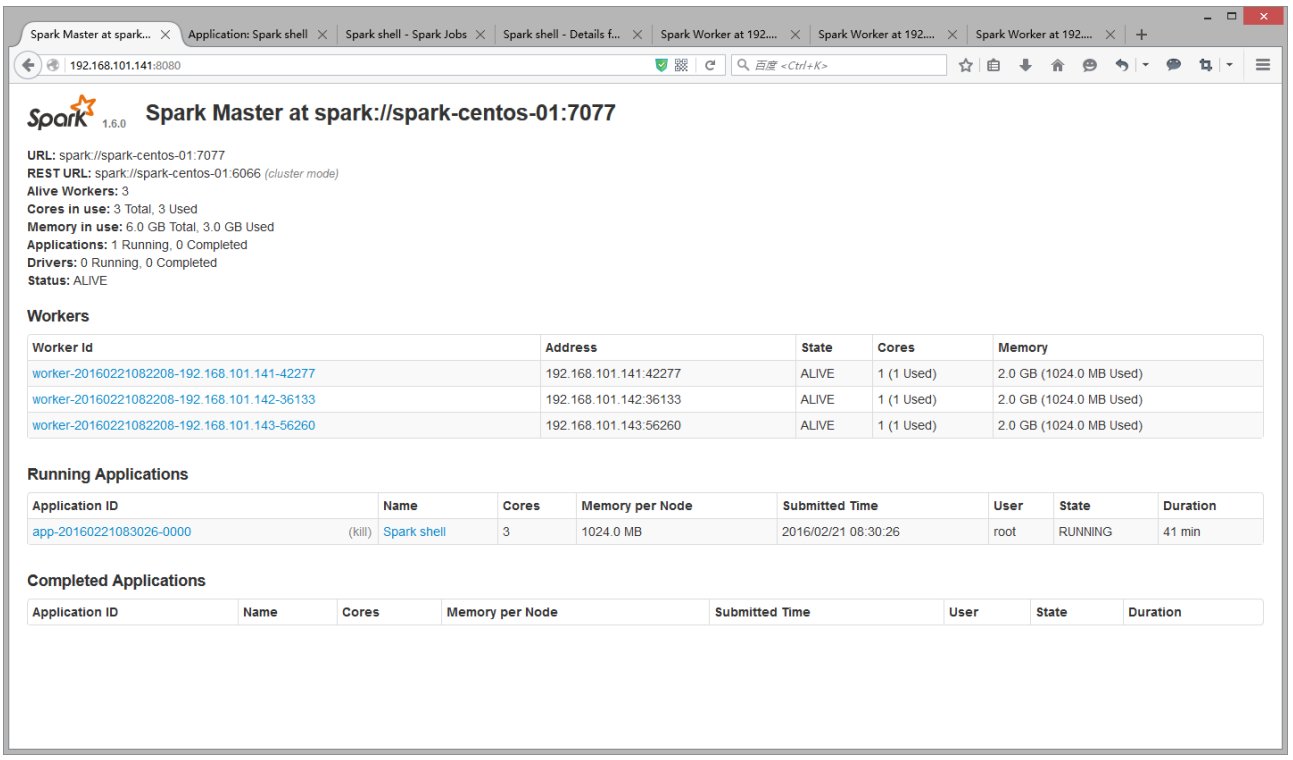
slave-02节点：http://192.168.101.142:8081/

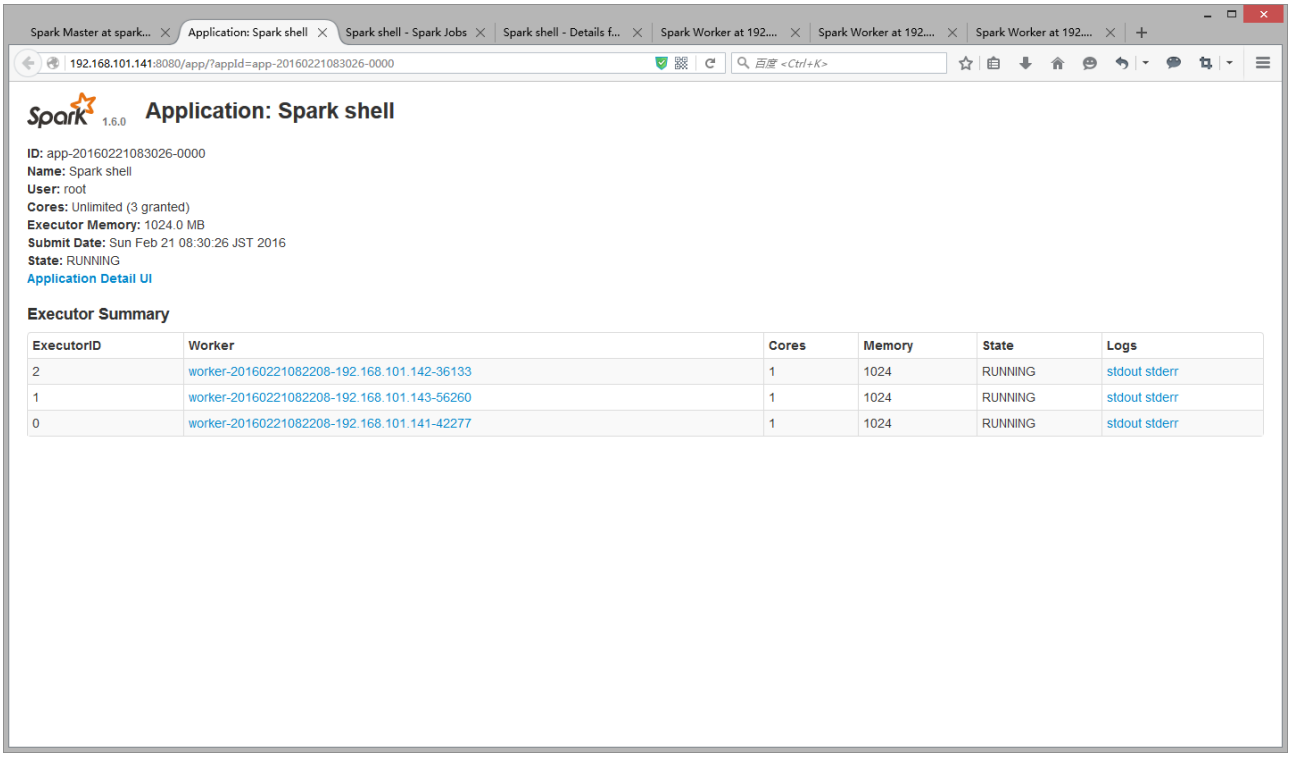
slave-03节点：http://192.168.101.143:8081/

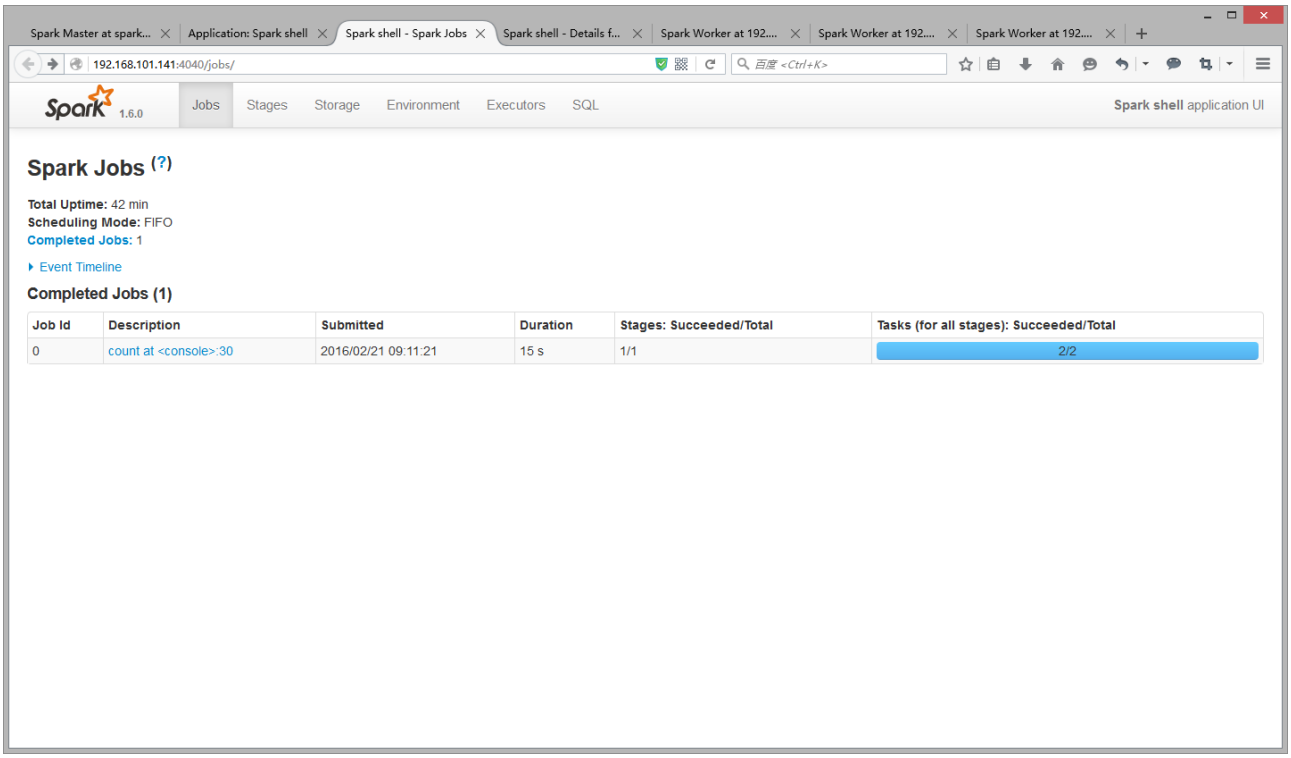
启动多个spark shell后，监控界面端口4040,4041自动依次递增（第二个spark shell启动的时

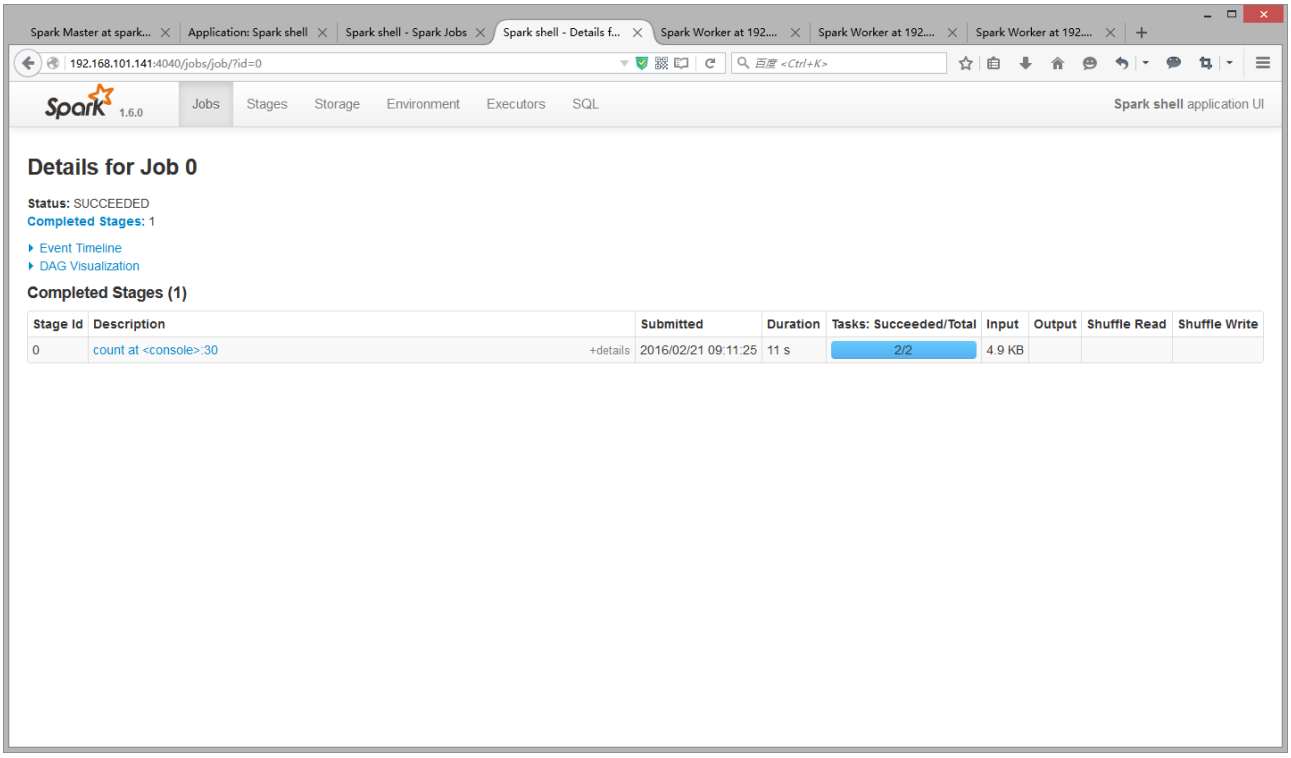
候会出现端口绑定错误）

spark shell job：http://192.168.101.141:4040/jobs/






Spark Master at spark...
Application: Spark shell
Spark shell - Spark Jobs
Spark shell - Details f...
Spark Worker at 192....
Spark Worker at 192....
Spark Worker at 192....
192.168.101.141:4040/jobs/
Spark 1.6.0
Jobs
Stages
Storage
Environment
Executors
SQL
Spark shell application UI
Spark Jobs (?)
Total Uptime: 42 min
Scheduling Mode: FIFO
Completed Jobs: 1
Event Timeline
Completed Jobs (1)
Job Id
Description
Submitted
Duration
Stages: Succeeded/Total
Tasks (for all stages): Succeeded/Total
0
count at <console>:30
2016/02/21 09:11:21
15 s
1/1
2/2


Spark Master at spark...
Application: Spark shell
Spark shell - Spark Jobs
Spark shell - Details f...
Spark Worker at 192....
Spark Worker at 192....
Spark Worker at 192....
192.168.101.141:4040/jobs/job/?id=0
Spark 1.6.0
Jobs
Stages
Storage
Environment
Executors
SQL
Spark shell application UI
Details for Job 0
Status: SUCCEEDED
Completed Stages: 1
Event Timeline
DAG Visualization
Completed Stages (1)
Stage Id
Description
Submitted
Duration
Tasks: Succeeded/Total
Input
Output
Shuffle Read
Shuffle Write
0
count at <console>:30
+details
2016/02/21 09:11:25
11 s
2/2
4.9 KB