# 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-2.1.0-bin-hadoop2.7.tgz
```

# Spark Standalone

modify profile

```
vim ~/.bashrc

export SOFT_BASE_PATH=/opt
# Spark Standalone
export SPARK_BASE_PATH=/opt
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=$SPARK_BASE_PATH/spark-2.1.0-bin-hadoop2.7
export HADOOP_HOME=$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 ~/.bashrc
```

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=/opt/scala-2.11.7
export SPARK_MASTER_IP=Spark-Master
export SPARK_WORKER_MEMORY=2G
export JAVA_HOME=/usr/lib/jvm/java-8-oracle/
```

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-Worker-1
Spark-Worker-2
```

使用scp命令，将配置修改后的spark代码发送到其他节点（Spark-Worker-1、Spark-Worker-2）

```
scp -r /app/spark-standalone root@Spark-Worker-1:/app/
scp -r /app/spark-standalone root@Spark-Worker-2:/app/
```

启动停止命令

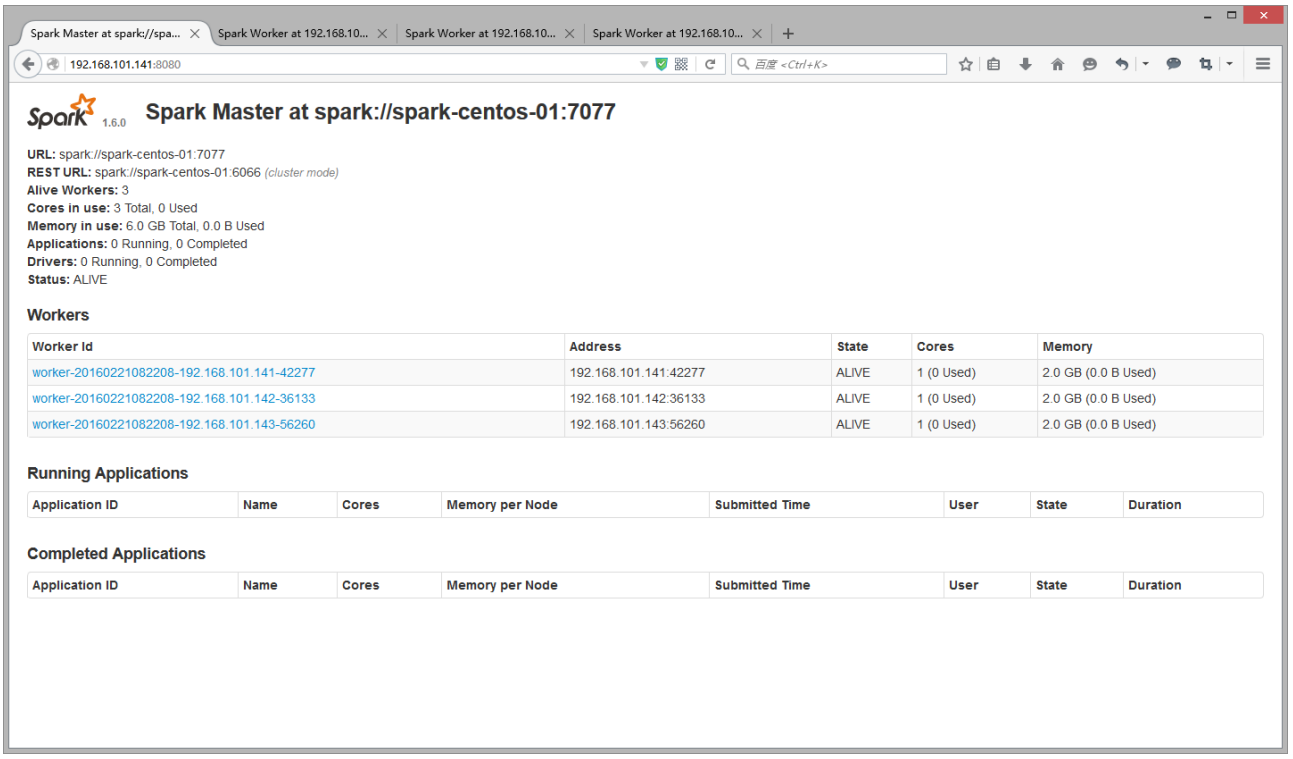
```
# 启动全部节点
$SPARK_HOME/sbin/start-all.sh
# 启动master
$SPARK_HOME/sbin/start-master.sh
# 启动worker
$SPARK_HOME/sbin/start-slaves.sh
# 停止全部节点
$SPARK_HOME/sbin/stop-all.sh
```

启动后的截图

Spark-Master 节点：http://192.168.101.231:8080/

Spark-Worker-1 节点：http://192.168.101.232:8081/

Spark-Worker-2 节点：http://192.168.101.233:8081/




spark-shell

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

HelloWorld

```
# test localhost file
scala > val textFile = sc.textFile("file:///opt/spark-2.1.0-bin-hadoop2.7/README.md")
scala > textFile.count()

# test hdfs file
scala > val textFile = sc.textFile("hdfs://Hadoop-NameNode:9000/input/README.txt")
scala > textFile.count()
```

spark-submit

```
$SPARK_HOME/bin/spark-submit --master spark://Spark-Master:7077 --class org.apache.spark.examples.SparkPi --executor-memory 2g --total-executor-cores 2 $SPARK_HOME/examples/jars/spark-examples_2.11-2.1.0.jar 1000
```

```
17/03/26 20:14:29 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@389adf1d{/stages/json,null,UNAVAILABLE}
17/03/26 20:14:29 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@7bf9b098{/stages,null,UNAVAILABLE}
17/03/26 20:14:29 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@72e34f77{/jobs/job/json,null,UNAVAILABLE}
17/03/26 20:14:29 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@6e9319f{/jobs/job,null,UNAVAILABLE}
17/03/26 20:14:29 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@6fa590ba{/jobs/json,null,UNAVAILABLE}
17/03/26 20:14:29 INFO handler.ContextHandler: Stopped o.s.j.s.ServletContextHandler@2416a51{/jobs,null,UNAVAILABLE}
17/03/26 20:14:29 INFO ui.SparkUI: Stopped Spark web UI at http://192.168.101.231:4040
17/03/26 20:14:29 INFO cluster.StandaloneSchedulerBackend: Shutting down all executors
17/03/26 20:14:29 INFO cluster.CoarseGrainedSchedulerBackend$DriverEndpoint: Asking each executor to shut down
17/03/26 20:14:29 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
17/03/26 20:14:29 INFO memory.MemoryStore: MemoryStore cleared
17/03/26 20:14:29 INFO storage.BlockManager: BlockManager stopped
17/03/26 20:14:29 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
17/03/26 20:14:29 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
17/03/26 20:14:29 INFO spark.SparkContext: Successfully stopped SparkContext
17/03/26 20:14:29 INFO util.ShutdownHookManager: Shutdown hook called
17/03/26 20:14:29 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-2b85e71f-d457-4994-8da9-51660c5c3a69
root@Spark-Master:/opt#
```

Spark-Master 节点：http://192.168.101.231:8080/
Spark-Worker-1 节点：http://192.168.101.232:8081/
Spark-Worker-2 节点：http://192.168.101.233:8081/

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://192.168.101.232:4040/jobs/