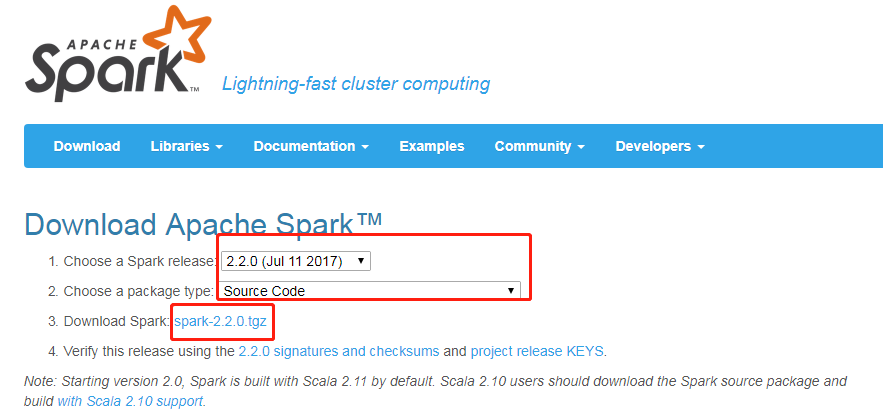
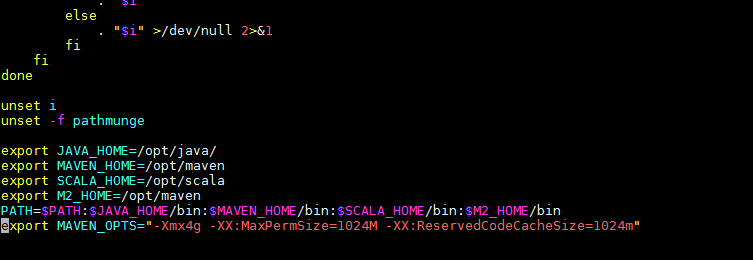
Spark2.2.0源码编译打包

1. 在Spark官网下载源码，选着Source Code下载即可（也可以git clone spark源码），下载之后的tar包传入机器对应位置解压即可。



1. 在环境变量里设置maven内存大小：

export MAVEN\_OPTS="-Xmx4g -XX:MaxPermSize=1024M -XX:ReservedCodeCacheSize=1024m"



1. 官方编译命令：

# Apache Hadoop 2.2.X

./build/mvn -Pyarn -Phadoop-2.2 -DskipTests clean package

# Apache Hadoop 2.3.X

./build/mvn -Pyarn -Phadoop-2.3 -Dhadoop.version=2.3.0 -DskipTests clean package

# Apache Hadoop 2.4.X or 2.5.X

./build/mvn -Pyarn -Phadoop-2.4 -Dhadoop.version=2.4.0 -DskipTests clean package

# Apache Hadoop 2.6.X

./build/mvn -Pyarn -Phadoop-2.6 -Dhadoop.version=2.6.0 -DskipTests clean package

# Apache Hadoop 2.7.X and later

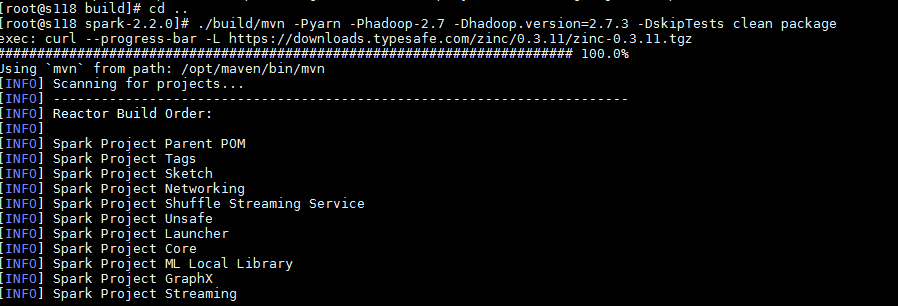
./build/mvn -Pyarn -Phadoop-2.7 -Dhadoop.version=2.7.0 -DskipTests clean package

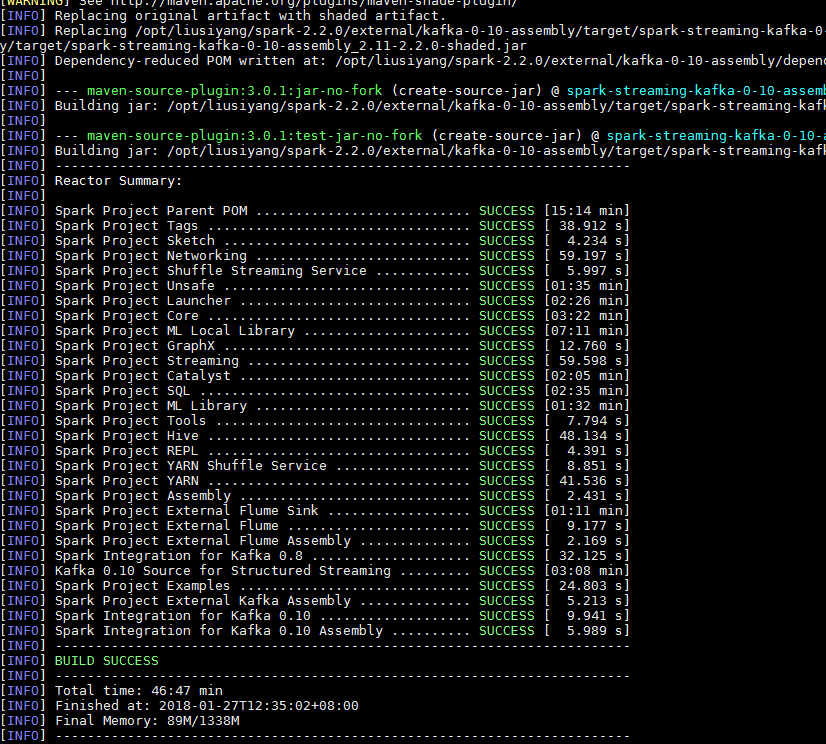
# Different versions of HDFS and YARN

./build/mvn -Pyarn -Phadoop-2.3 -Dhadoop.version=2.3.0 -Dyarn.version=2.2.0 -DskipTests clean package

进入Spark源码解压后目录执行（若报错，多试几次）：

/build/mvn -Pyarn -Phadoop-2.7 -Dhadoop.version=2.7.3 -DskipTests clean package





1. 编译安装好spark-2.2.0以后,可以进入不bin目录中,使用spark-shell启动,如下：

spark-shell命令及其常用的参数如下：

./bin/spark-shell --master <master-url>

Spark的运行模式取决于传递给SparkContext的Master URL的值。Master URL可以是以下任一种形式：

\* local 使用一个Worker线程本地化运行SPARK(完全不并行)

\* local[\*] 使用逻辑CPU个数数量的线程来本地化运行Spark

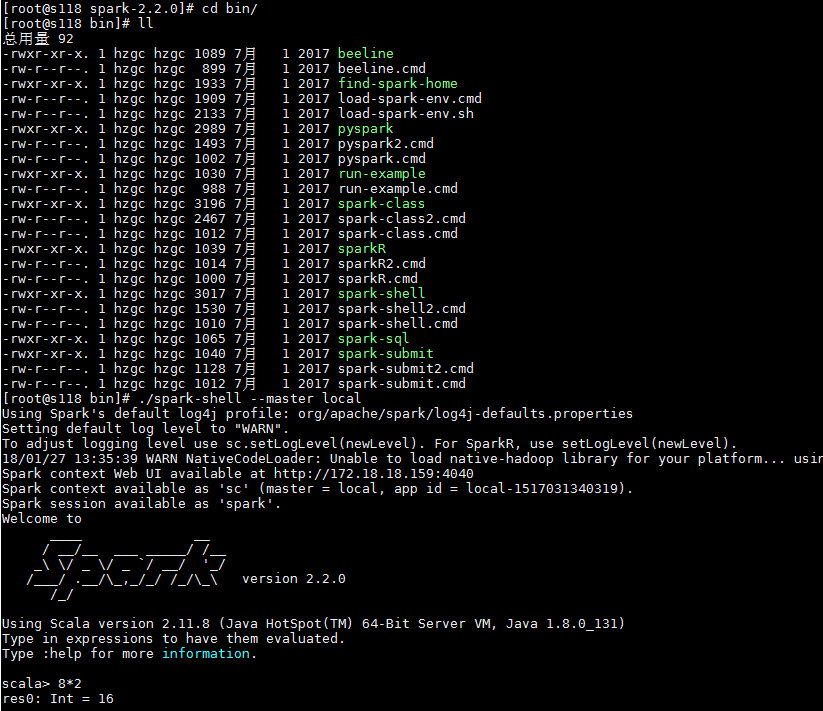
\* local[K] 使用K个Worker线程本地化运行Spark（理想情况下，K应该根据运行机器的CPU核数设定）

\* spark://HOST:PORT 连接到指定的Spark standalone master。默认端口是7077.

\* yarn-client 以客户端模式连接YARN集群。集群的位置可以在HADOOP\_CONF\_DIR 环境变量中找到。

\* yarn-cluster 以集群模式连接YARN集群。集群的位置可以在HADOOP\_CONF\_DIR 环境变量中找到。

\* mesos://HOST:PORT 连接到指定的Mesos集群。默认接口是5050



1. 可以使用dev/make-distribution.sh脚本进行打包

./make-distribution.sh --name custom-spark --tgz --mvn ${MAVEN\_HOME}/bin/mvn -Phadoop-2.4 -Phive -Phive-thriftserver -Pmesos -Pyarn

其中参数含义为：

--name custom-spark //指定压缩包名字

--tgz //生成部署包

--mvn ${MAVEN\_HOME}/bin/mvn //指定使用自己本地配置的mvn，不指定默认使用spark-source/build下面的mvn

我们这里使用的打包命令为：

./make-distribution.sh --tgz --name 2.2.0 -Pyarn -Phadoop-2.7 -Dhadoop.version=2.7.2 -Phive -Phive-thriftserve

1. 最后在Spark-source根目录生成安装包