Spark 源码编译手册

## 源码获取

1. Spark官网下载

Spark官网下载地址：

http://spark.apache.org/downloads.html，根据自己需要下载对应版本即可。

1. GitHub下载

GitHub的下载地址：https://github.com/apache/spark

1. **Cloudera下载**

Cloudera下载地址：

http://archive-primary.cloudera.com/cdh5/cdh/5/

说明：本次我们下载1.6.1版本

## **编译**

## Spark官方提出的编译注意事项：Spark1.6版本要求Maven不能低于3.3.3，Java不能低于1.7。

## **Maven编译**

首先我们要自己搭建一个Maven环境。然后执行编译命令：

|  |
| --- |
| mvn -Pyarn -Phadoop-2.6 -Dhadoop.version=2.6.0-cdh5.10.0 -DskipTests clean package |

如果抛出内存溢出相关错误，执行如下名：

|  |
| --- |
| export MAVEN\_OPTS=”-Xmx2g -XX:MaxPermSize=512m -XX:PeservedCodeCacheSize=512m” |

## **make-distribution.sh编译**

Make-distribution.sh存放在Spark的根目录里，其底层也是调用Maven进行编译的，但其有好处是，编译成功后在Spark源码的根目录下会生成对应的tar.gz包，后边我们部署可以直接使用该包进行安装。

编译命令：

|  |
| --- |
| make-distribution.sh --name 2.6.0-cdh5.10.0 --tgz -Phadoop-2.6 -Dhadoop.version=2.6.0-chd5.10.0 -Pyarn |

参数说明：

1. --name NAME 和 --tgz结合可以生成spark-$VERSION-bin-$NAME.tgz的部署包，默认NAME为Hadoop的版本号。
2. --tgz：在根目录下生成spark-$VERSION-bin.tgz,默认不生成。