### windows系统手动编译spark-2.2.0-bin-2.6.0-cdh5.7.0.tgz

#### Spark 2.2.0源码的下载地址：

https://mirrors.tuna.tsinghua.edu.cn/apache/spark/spark-2.2.0/spark-2.2.0.tgz

编译Spark源码的文档(参考官方文档)：

http://spark.apache.org/docs/latest/building-spark.html

#### **编译Spark源码的前置要求：**

Maven 3.3.9 or newer

Java 8+

Scala

Git(后文会通过分析make-distribution.sh脚本，来阐述为何需要安装Git)

使用的软件的版本：

apache-maven-3.3.9-bin.tar.gz

hadoop-2.6.0-cdh5.7.0.tar.gz

jdk-8u91-linux-x64.tar.gz

scala-2.11.8.tgz

#### **2.2. 解压**

tar -zxvf spark-2.2.0.tgz

#### **3. 进入主目录，修改编译文件，进行编译**

修改spark-2.2.0/dev目录下的make-distribution.sh ，注释掉原来的指定版本，可以节省时间

VERSION=2.2.0

SCALA\_VERSION=2.11

SPARK\_HADOOP\_VERSION=2.6.0-cdh5.7.0

SPARK\_HIVE=1

温馨提示：

该make-distribution.sh文件中如图所示，czf前少个“-”，需要自己修改，添加上去

如果你用的hadoop版本是cdh的，那么需要修改spark根目录pom.xml文件，添加cdh的依赖

<repository>

<id>cloudera</id>

<name>cloudera Repository</name>

<url>https://repository.cloudera.com/artifactory/cloudera-repos/</url>

</repository>

还有一点很重要：

在你的maven目录修改 /conf/setting.xml，，否则很慢

修改/conf/setting.xml

<mirror>

<id>alimaven</id>

<name>aliyun maven</name>

<url>http://maven.aliyun.com/nexus/content/groups/public/</url>

<mirrorOf>central</mirrorOf>

</mirror>

也可以修改指定的版本仓库到指定的文件夹

最后用Git进入spark根目录，执行

./dev/make-distribution.sh \

--name 2.6.0-cdh5.7.0 \

--tgz \

-Dhadoop.version=2.6.0-cdh5.7.0 \

-Phadoop-2.6 \

-Phive -Phive-thriftserver \

-Pyarn \

-DskipTests clean package

等待吧，别断网，编译所有的东西需要有一段的时间

最后添加，自己编译好的地址，

链接：https://pan.baidu.com/s/1syfzjdC7llhLhI\_Ul26wOg 密码：bwnv