**Spark安装教程**

# 上传解压压缩文件

tar -zxvf spark-1.5.2.tgz

修改make-distribution.sh

注释掉原文件

VERSION=$("$MVN" help:evaluate -Dexpression=project.version $@ 2>/dev/null | grep -v "INFO" | tail -n 1)

SCALA\_VERSION=$("$MVN" help:evaluate -Dexpression=scala.binary.version $@ 2>/dev/null\

| grep -v "INFO"\

| tail -n 1)

SPARK\_HADOOP\_VERSION=$("$MVN" help:evaluate -Dexpression=hadoop.version $@ 2>/dev/null\

| grep -v "INFO"\

| tail -n 1)

SPARK\_HIVE=$("$MVN" help:evaluate -Dexpression=project.activeProfiles -pl sql/hive $@ 2>/dev/null\

| grep -v "INFO"\

| fgrep --count "<id>hive</id>";\

# Reset exit status to 0, otherwise the script stops here if the last grep finds nothing\

# because we use "set -o pipefail"

echo -n)

加入

VERSION=1.5.2 （具体版本）

SCALA\_VERSION=具体版本

SPARK\_HADOOP\_VERSION=具体版本

SPARK\_HIVE=1 （1为需要将hive打包进去，非1为不需要打包进去）

# 配置配置文件

## 配置域名解析服务器

vi /etc/resolv.conf

内容：

nameserver 8.8.8.8

nameserver 8.8.4.4

# 打包编译

### MAVEN 编译

$ /opt/modules/spark-1.3.0-src/build/mvn clean package -DskipTests -Phadoop-2.4 -Dhadoop.version=2.5.0-cdh5.3.6 -Pyarn -Phive-0.13.1 -Phive-thriftserver

OR

### 打包编译，使用CDH 5.3.6版本的HADOOP

./make-distribution.sh --tgz -Phadoop-2.4 -Dhadoop.version=2.5.0-cdh5.3.6 -Pyarn -Phive-0.13.1 -Phive-thriftserver

### 打包编译，使用Apache版本的HADOOP

./make-distribution.sh --tgz -Phadoop-2.4 -Dhadoop.version=2.5.0 -Pyarn -Phive-0.13.1 -Phive-thriftserver

# 解压编译好的压缩包

## 编辑spark-env.sh文件

### 编辑spark-env.sh文件

重命名spark-env.sh.template为spark-env.sh，添加：

JAVA\_HOME=/home/zkj/Java/jdk1.8.0\_144

SCALA\_HOME=/home/zkj/Scala/scala-2.12.3

HADOOP\_CONF\_DIR=/home/zkj/hadoop-2.8.1/etc/hadoop

SPARK\_MASTER\_HOST=MyDream

SPARK\_MASTER\_PORT=7077

SPARK\_MASTER\_WEBUI\_PORT=8080

SPARK\_WORKER\_CORES=1

SPARK\_WORKER\_MEMORY=2g

SPARK\_WORKER\_PORT=7078

SPARK\_WORKER\_WEBUI\_PORT=8081

### 编辑slave文件

重命名slaves.template为slaves

添加主机名

# 配置历史日志收集

## 服务端配置

配置spark-env.sh文件

SPARK\_HISTORY\_OPTS=-Dspark.history.fs.logDirectory=hdfs://MyDream:9000/spark/eventLogs

HDFS新建该路径

## 客户端配置

配置spark-defaults.conf文件

spark.eventLog.enabled true

spark.eventLog.dir hdfs://MyDream:9000/spark/eventLogs

spark.eventLog.compress true

## 启动历史日志收集

start-history-server.sh

# 编辑hadoop配置

若要运行在yarn上编辑yarn-site.xml文件，增加

<property>

<name>yarn.nodemanager.pmem-check-enabled</name>

<value>false</value>

</property>

<property>

<name>yarn.nodemanager.vmem-check-enabled</name>

<value>false</value>

</property>

否则分配给节点分配的内存少，yarn kill了[Spark](http://lib.csdn.net/base/spark) application。