

## 实验环境搭建——Spark

软件准备：Spark 安装包

### 1、Java + Scala 安装

首先确定系统中已安装 Java 和 Scala, 可以通过 `java -version` 和 `scala -version` 来验证版本。  
若 Scala 未安装, 可通过命令进行一键安装。

```
$ sudo apt install scala
```

### 2、安装 Spark

直接解压 Spark 安装包

```
$ tar zxvf <spark 安装包>
```

然后把解压的整个目录移到指定位置。

### 3、配置 Spark

这里默认已经安装好 Hadoop。

Spark 需要配置 `spark-env.sh` 和 `slaves` 文件。Spark 默认没有上述两个文件, 可以自己建, 也可以直接复制 `spark-env.sh.template` 和 `slaves.template`。

`spark-env.sh` 的配置内容如下:

```
JAVA_HOME=/usr/local/jdk1.8.0_181
HADOOP_CONF_DIR=/home/hadoop/bigdata/hadoop-2.8.5/etc/hadoop
SPARK_WORKER_DIR=/home/hadoop/bigdata/spark-2.4.0-bin-hadoop2.7/runtime/worker
SPARK_LOG_DIR=/home/hadoop/bigdata/spark-2.4.0-bin-hadoop2.7/runtime/logs
SPARK_PID_DIR=/home/hadoop/bigdata/spark-2.4.0-bin-hadoop2.7/runtime/pid
```

`slaves` 的配置内容参考 hadoop 中 `slaves` 的配置内容, 即所有 worker 节点的 host 名字。

```
# A Spark Worker will be started on each of the machines listed below.
hadoop1
hadoop2
hadoop3
```

将 spark 配置分发到所有节点上。

```
$ scp -r <spark 目录> /hadoop2:/home/hadoop/bigdata/
```

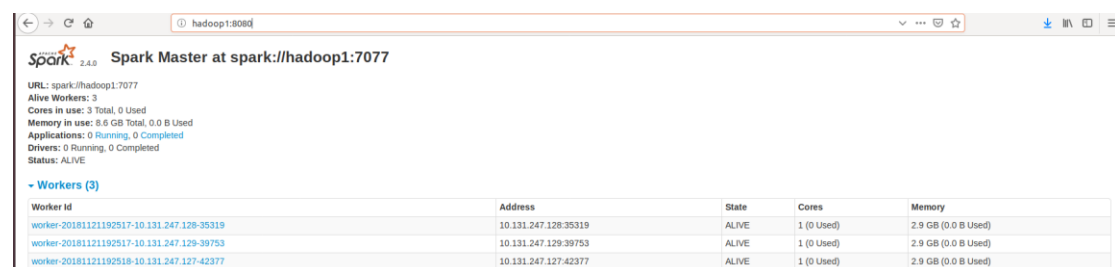
```
$ scp -r <spark 目录> /hadoop3:/home/hadoop/bigdata/
```

### 4、测试

首先确保 hadoop 已成功启动, 启动 spark 的命令为

```
$ sbin/start-all.sh
```

启动成功后, 输入 `<主机名>:8080` 可以看到 spark 的启动信息。



Workers (3)				
Worker Id	Address	State	Cores	Memory
worker-20181121192517-10.131.247.128-35319	10.131.247.128:35319	ALIVE	1 (0 Used)	2.9 GB (0.0 B Used)
worker-20181121192517-10.131.247.129-39753	10.131.247.129:39753	ALIVE	1 (0 Used)	2.9 GB (0.0 B Used)
worker-20181121192518-10.131.247.127-42377	10.131.247.127:42377	ALIVE	1 (0 Used)	2.9 GB (0.0 B Used)

可以看到 spark 已成功启动。