**spark计算框架搭建**

版本:2.0

Victors团队

## 功能任务

spark定时将elasticsearch中的数据取出并进行计算，最后把计算结果存入elasticsearch供其他模块使用。其中spark-master负责计算任务的分发。

## 主节点spark安装

1. **适用节点：**
   1. 概括：spark框架主节点
   2. 数量：1台
   3. 主机名：主节点：spark-master；
   4. IP：主节点：192.168.1.100；
2. **Spark下载：**
3. wget http://d3kbcqa49mib13.cloudfront.net/spark-2.1.0-bin-hadoop2.7.tgz
4. tar –zxvf spark-\*
5. **环境依赖：**
   1. python依赖

树莓派内置python版本为2.7.11版本，故无需升级。

* 1. java环境配置

jdk下载及安装详见基础环境部分java配置文档。

[..\..\基础环境部分\java环境配置.docx](../../基础环境部分/java环境配置.docx)

## 主节点spark配置

1. **进入spark主目录**
2. mv spark-\* spark
3. cd spark
4. **配置spark-env.sh**
   1. cp conf/spark-env.sh.template conf/spark-env.sh
   2. 文档中添加如下配置：

export SPARK\_MASTER\_PORT=7077

export SPARK\_MASTER\_IP=192.168.1.100

export SPARK\_MASTER\_WEBUI\_PORT=8080

export JAVA\_HOME=/usr/local/java/jdk

export CLASSPATH=.:JAVA\_HOME/lib/dt.jar:JAVA\_HOME/lib/tools.jar

export SPARK\_CLASSPATH=/home/pi/spark/jars

export SPARK\_HOME=/home/pi/spark

1. **配置slaves**
2. cp conf/slaves.template conf/slaves
3. vim conf/slaves
4. 删除文件中localhost并增加如下配置

spark-node1

spark-node2

spark-node3

………

spark-node9

1. **配置spark-defaults.conf**
   1. cp spark-defaults.conf.template spark-defaults.conf
   2. 文档中添加如下配置

spark.master spark://192.168.1.50:7077

spark.driver.memory 500m

spark.executor.memory 500m

## 服务启动

1. **守护进程启动**

./sbin/start-all.sh

1. **计算任务启动（python版）**

./bin/spark-sublit test.py

test.py表示python程序，此处仅叙述启动方式，test.py无实际意义。

1. 启动注意：

start-all.sh会启动spark所有节点的后台守护进程。故启动时，若所有从节点都未配置完成，则会导致启动失败。（有一台从节点配置好即可启动成功，但建议所有节点配置完成后在开启服务）

## 注意事项：

* 1. 查看当前集群状态可登录spark发布的网页查看，网址为

<http://192.168.1.100:8080>（spark主节点）

* 1. spark作业查看可通过spark发布的网页查看，但作业结束后该页面就会关闭。

<http://192.168.1.100:8080（spark>主节点）

1. 附录：

spark计算程序：<spark计算程序.txt>

1. 此处的服务启动指该项功能的启动方式，系统具体的调用方式参见构建书综合功能部分相关源码。