安装 jdk

安装 hadoop

安装 scala

1. 下载 scala

   比如：

   wget http://downloads.typesafe.com/scala/2.11.4/scala-2.11.4.tgz?_ga=1.161621180.1629813778.1429158992

   重命名

   mv scala-2.11.4.tgz\?_ga\=1.161621180.1629813778.1429158992.1 scala-2.11.4.tgz

2. 解压

   tar -zxvf scala-2.11.4.tgz

   移动

   mv scala-2.11.4 /usr/local/

3. 配置 linux 环境

   vi ~/.bashrc

   ```
   [root@master local]# vi ~/.bashrc
   ```

   添加

   export SCALA_HOME=/usr/local/scala-2.11.4

   ```
   export SCALA_HOME=/usr/local/scala-2.11.4
   ```

   export PATH=$PATH:$SCALA_HOME/bin

   ```
   export PATH=$PATH:$SCALA_HOME/bin
   ```

   source 一下

   source ~/.bashrc

   ```
   [root@master local]# source ~/.bashrc
   ```

4. 测试

   scala –version

   ```
   [root@master local]# scala -version
   Scala code runner version 2.11.4 -- Copyright 2002-2013, LAMP/EPFL
   ```

安装 spark

1. 下载 spark

   eg.

   网址

   http://spark.apache.org/downloads.html

http://apache.fayea.com/spark/spark-1.3.0/spark-1.3.0-bin-hadoop2.4.tgz

wget http://apache.fayea.com/spark/spark-1.3.0/spark-1.3.0-bin-hadoop2.4.tgz

2. 解压

tar zxvf spark-1.3.0-bin-hadoop2.4.tgz

重命名

mv spark-1.3.0-bin-hadoop2.4 spark-1.3.0

移动

mv spark-1.3.0 /usr/local/

3. 配置 HOME、PATH

vi ~/.bashrc

添加

export SPARK_HOME=/usr/local/spark-1.3.0

export PATH=$PATH:$SPARK_HOME/bin

source ~/.bashrc

4. slaves

cd /usr/local/spark-1.3.0/conf

mv slaves.template slaves

vi slaves

删除原来的 localhost 加入节点比如

datanode1,datanode2

5. spark conf/spark-env.sh

mv spark-env.sh.template spark-env.sh

6. 复制到其他机器

scp -r spark-1.3.0/ root@datanode1:/usr/local/

scp -r spark-1.3.0/ root@datanode2:/usr/local/

scp -r scala-2.11.4/ root@datanode1:/usr/local/

scp -r scala-2.11.4/ root@datanode2:/usr/local/

scp ~/.bashrc root@datanode1:~/.bashrc

scp ~/.bashrc root@datanode2:~/.bashrc

ssh root@datanode1 "source ~/.bashrc"

```
ssh root@datanode2 "source ~/.bashrc"
```

7. 关闭防火墙

```
service iptables stop
ssh root@datanode1 "service iptables stop"
ssh root@datanode2 "service iptables stop"
```

8. 启动

启动 hdfs

```
start-dfs.sh
```

启动 spark

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

9. 运行 example

```
./bin/run-example org.apache.spark.examples.JavaWordCount h
dfs://master:9000/folder/file-1.txt
```

```
./bin/run-example org.apache.spark.examples.JavaWordCount h
dfs://master:9000/folder/file-1.txt
```

hdfs上的文件，没有的自行创建

java代码，在spark安装目录下的example中的java文件夹中