Hadoop集群安装

依赖

> JDK1.8

- 安装OpenJDK或JDK

```
sudo apt-get install openjdk-8-jdk
```

基础配置

- 配置hostname

```
sudo hostname master(每个节点配置自己单独的名字)
```

- 配置host

```
sudo vim /etc/hosts
```

将内网ip对应修改的hosts中的name写入文件

```
master的内网ip master
slave01的内网ip slave01
slave02的内网ip slave02
```

例如

```
192.168.1.1 master
192.168.1.2 slave01
192.168.1.3 slave02
```

- 关闭防火墙

```
sudo ufw disable
```

- 配置无密钥证书

```
ssh-keygen -t rsa
ssh-copy-id -i /home/ubuntu/.ssh/id_rsa.pub master
ssh-copy-id -i /home/ubuntu/.ssh/id_rsa.pub slave01
ssh-copy-id -i /home/ubuntu/.ssh/id_rsa.pub slave02
```

以上所有内容需要在三个节点都进行配置！！！

安装Hadoop集群

1. 获取hadoop压缩包

```
wget https://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/hadoop-2.7.7/hadoop-2.7.7.tar.gz
```

2. 解压

```
tar -zxvf hadoop-2.7.7.tar.gz
```

3. 创建dfs目录
   - 打开路径指令

```
cd filename
```

   - 进入hadoop文件夹

```
cd hadoop-2.7.7
```

   - 创建路径

```
mkdir -p dfs/name
mkdir -p dfs/data
mkdir -p dfs/namesecondary
```

4. 更改hadoop配置文件
   - 进入配置文件路径

```
cd etc/hadoop
```

   - 编辑文件指令

```
sudo vim filename
```

   - 修改core-site.xml，在<configuration></configuration>中增加如下内容

```
<property>
<name>fs.defaultFS</name>
<value>hdfs://master:9000</value>
<description>NameNode URI.</description>
</property>

<property>
<name>io.file.buffer.size</name>
<value>131072</value>
<description>Size of read/write buffer used inSequenceFiles.</description>
</property>
```

   - 修改hdfs-site.xml，在<configuration></configuration>中增加如下内容

```
<property>
<name>dfs.namenode.secondary.http-address</name>
```

```
<value>master:50090</value>
<description>The secondary namenode http server address andport.
</description>
</property>

<property>
<name>dfs.namenode.name.dir</name>
<value>file:///home/ubuntu/hadoop-2.7.7/dfs/name</value>
<description>Path on the local filesystem where the NameNodestores the
namespace and transactions logs persistently.</description>
</property>

<property>
<name>dfs.datanode.data.dir</name>
<value>file:///home/ubuntu/hadoop-2.7.7/dfs/data</value>
<description>Comma separated list of paths on the local filesystemof a
DataNode where it should store its blocks.</description>
</property>

<property>
<name>dfs.namenode.checkpoint.dir</name>
<value>file:///home/ubuntu/hadoop-2.7.7/dfs/namesecondary</value>
<description>Determines where on the local filesystem the DFSsecondary name
node should store the temporary images to merge. If this is acomma-delimited
list of directories then the image is replicated in all of thedirectories
for redundancy.</description>
</property>

<property>
<name>dfs.replication</name>
<value>2</value>
</property>
```

- 修改mapred-site.xml

1. 复制模板文件

```
cp mapred-site.xml.template mapred-site.xml
```

2. 在<configuration></configuration>中增加如下内容

```
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
<description>Theruntime framework for executing MapReduce jobs. Can be one
of local, classic oryarn.</description>
</property>

<property>
<name>mapreduce.jobhistory.address</name>
<value>master:10020</value>
<description>MapReduce JobHistoryServer IPC host:port</description>
</property>

<property>
<name>mapreduce.jobhistory.webapp.address</name>
<value>master:19888</value>
```

```
<description>MapReduce JobHistoryServer Web UI host:port</description>
</property>

<property>
<name>mapreduce.application.classpath</name>
<value>
/home/ubuntu/hadoop-2.7.7/etc/hadoop,
/home/ubuntu/hadoop-2.7.7/share/hadoop/common/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/common/lib/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/hdfs/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/hdfs/lib/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/mapreduce/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/mapreduce/lib/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/yarn/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/yarn/lib/*
</value>
</property>
```

- 修改yarn-site.xml，在<configuration></configuration>中增加如下内容

```
<property>
<name>yarn.resourcemanager.hostname</name>
<value>master</value>
<description>The hostname of theRM.</description>
</property>

<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
<description>Shuffle service that needs to be set for Map
Reduceapplications.</description>
</property>

<property>
<name>yarn.application.classpath</name>
<value>
/home/ubuntu/hadoop-2.7.7/etc/hadoop,
/home/ubuntu/hadoop-2.7.7/share/hadoop/common/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/common/lib/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/hdfs/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/hdfs/lib/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/mapreduce/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/mapreduce/lib/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/yarn/*,
/home/ubuntu/hadoop-2.7.7/share/hadoop/yarn/lib/*
</value>
</property>

<property>
    <name>yarn.nodemanager.resource.memory-mb</name>
    <value>1024</value>
</property>

<property>
    <name>yarn.nodemanager.resource.cpu-vcores</name>
    <value>1</value>
</property>
```

- 修改hadoop-env.sh中java路径

```
JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64
```

- 填写masters文件(部分版本不需要)和works文件

```
master
```

```
slave01
slave02
```

5. 复制文件到子节点

```
scp -r /home/ubuntu/hadoop-2.7.7 ubuntu@slave01:/home/ubuntu/
scp -r /home/ubuntu/hadoop-2.7.7 ubuntu@slave02:/home/ubuntu/
```

6. 环境变量

```
sudo vi /etc/profile
```

```
PATH=$PATH:$HOME/bin
export HADOOP_HOME=/home/ubuntu/hadoop-2.7.7
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

配置环境可能需要重启连接！！！

7. 启动和停止

```
start-all.sh
```

```
stop-all.sh
```

8. 验证

hadoop验证

进程验证

Master节点验证

```
[hadoop@master ~]$ jps
29633 NameNode
30071 Jps
29820 SecondaryNameNode
29965 ResourceManager
```

Slave01节点验证

```
[hadoop@slave01 ~]$ jps
28083 NodeManager
27978 DataNode
28158 Jps
```

Slave02节点验证

```
[hadoop@slave02 ~]$ jps
28176 Jps
28054 NodeManager
27947 DataNode
```

web界面验证

http://169.254.1.100:50070