

# 大数据架构Hadoop UNIT03-04

- 虚拟机安装
- Hadoop伪分布式安装

## Hadoop伪分布式安装

### 问题

如何在一台虚拟机上模拟Hadoop所有进程的使用（HDFS、YARN）

### 方案

使用Hadoop的伪分布式安装可以在一台虚拟机上模拟全部的进程

### 环境

CentOS 7环境

### 资源

jdk-8u281-linux-x64.tar.gz、hadoop-2.8.5.tar.gz、Xshell-6.0.0197p.exe

### 步骤

- 步骤一：关闭防火墙

首先通过下面的指令查看防火墙状态：

```
[root@hadoop ~]# firewall-cmd --state
```

此时可能会出现如下两种状态：

running	#防火墙处于运行状态
not running	#防火墙处于关闭状态

如果防火墙显示为running状态时，需要我们手动关闭防火墙：

```
[root@hadoop ~]# systemctl stop firewalld.service      #关闭防火墙  
[root@hadoop ~]# systemctl disable firewalld.service   #禁止开机启动防火墙
```

此时在查看防火墙状态，将显示not running 表示防火墙关闭完成。

- 步骤二：配置主机名

通过hostnamectl指令也可以设置主机名：

```
[root@xxxxx .ssh]# hostnamectl set-hostname hadoop
```

然后通过hostname查看主机名：

```
[root@xxxxx ~]# hostname  
hadoop
```

但是此时显示任然为旧的主机名，可以通过exit指令，先退出当前的登录，然后在重新登录之后就会显示设置完成的主机名。

- **步骤三：配置/etc/hosts 文件**

/etc/hosts文件主要用于配置主机名和IP地址的映射关系，作用相当于DNS服务器。便于在之后的大数据环境中，使用简洁的虚拟机主机名来代表冗长的IP地址。

使用vim、vi打开/etc/hosts文件

```
[root@hadoop ~]# vim/vi /etc/hosts
```

然后在文件的末尾添加如下所示的内容：

```
192.168.56.100    hadoop
```

保存退出

- **步骤四：配置免密登录**

# 生成公钥和私钥：

```
[root@hadoop ~]# ssh-keygen -t rsa
```

```
[root@hadoop ~]# ssh-keygen -t rsa  
Generating public/private rsa key pair. 回车  
Enter file in which to save the key (/root/.ssh/id_rsa): 回车  
Created directory '/root/.ssh'.  
Enter passphrase (empty for no passphrase): 回车  
Enter same passphrase again: 回车  
Your identification has been saved in /root/.ssh/id_rsa.  
Your public key has been saved in /root/.ssh/id_rsa.pub.  
The key fingerprint is:  
SHA256:AQVBH0BSQuwtdkjScXhEhZ7E7irngZ8thJwXKJYKn/M root@hadoop  
The key's randomart image is:  
+---[RSA 2048]---+  
|  +=X%B+       |  
|  . +=+o .     |  
|  . = B .o     |  
| oo. * * .     |  
| ++ = = S      |  
| . B.o .       |  
|  . = .        |  
|  ..E+         |  
|  ++..         |  
+---[SHA256]-----+  
出现该界面，表示私  
钥、公钥成功生成
```

在/root/.ssh/目录中查看生成的公钥、私钥

```
[root@hadoop ~]# cd /root/.ssh    #进入指定目录  
[root@hadoop .ssh]# ls           #查看内容
```

显示内容有：id\_rsa（私钥）、id\_rsa.pub（公钥）

将公钥拷贝到/root/.ssh/authorized\_keys文件中

```
[root@hadoop .ssh]# cp id_rsa.pub authorized_keys
```

测试免密登录，通过ssh hadoop登录时，发现没有要求输入用户名和密码，表示免密登录配置成功。  
(第一次登录时，可能会询问是否连接，输入yes即可)

```
[root@hadoop .ssh]# ssh hadoop
Last login: Wed Oct 2 14:24:40 2019 from 192.168.56.100
[root@hadoop ~]#
```

如上图所示：在登录之前在.ssh目录，登录之后进入~目录

使用 exit 指令退出当前登录状态。

- **步骤五：Hadoop分布式安装**

上传hadoop、jdk安装包到Linux的 /opt/ 目录下

在XSELL中新建选项卡(打开一个新的窗口)

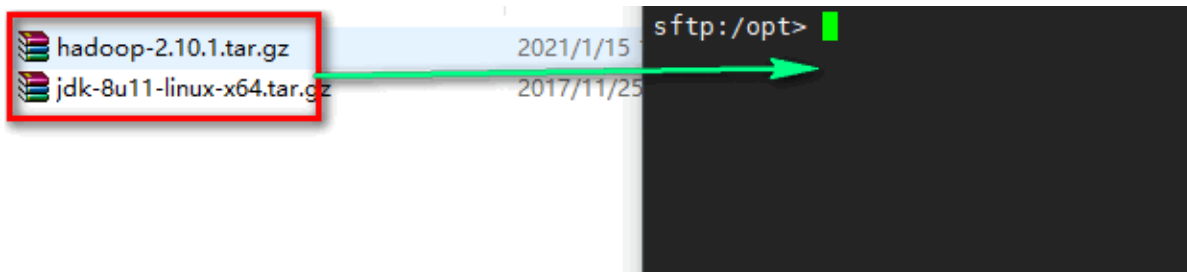
通过指令远程连接Linux系统:

```
[C:\~]$ sftp root@192.168.56.100
```

切换到Linux系统的 /opt 目录中:

```
sftp:/root> cd /opt/
sftp:/opt>
```

拖动JDK、Hadoop安装到，实现文件的上传



**解压安装包：**

```
# 进入 /opt/ 目录
[root@hadoop ~]# cd /opt

[root@hadoop ~]# tar -zxvf jdk-8u281-linux-x64.tar.gz
[root@hadoop ~]# tar -zxvf hadoop-2.8.5.tar.gz
```

**配置环境变量：**

```
vim /etc/profile
```

在文件末尾添加如下所示内容：

```
export JAVA_HOME=/opt/jdk1.8.0_281
export HADOOP_HOME=/opt/hadoop-2.8.5
export PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$JAVA_HOME/bin:$PATH:$HOME/bin
```

保存退出。然后执行下面的指令，是配置文件立即生效

```
[root@hadoop ~]# source /etc/profile
```

检验配置文件是否配置成功：

执行下面的指令，看是否能够输出hadoop和jdk的版本

```
[root@hadoop ~]# java -version
[root@hadoop ~]# hadoop version
```

- **步骤六：伪分布式配置**

**配置hadoop-env.sh文件：**

```
# 进入hadoop配置文件所在目录
[root@hadoop ~]# cd /opt/hadoop-2.8.5/etc/hadoop/

[root@hadoop hadoop]# vim hadoop-env.sh
```

在文件第25行处，编辑内容如下：

```
export JAVA_HOME=/opt/jdk1.8.0_281      #第25行处
```

保存处。

**配置core-site.xml文件：**

```
[root@hadoop hadoop]# vim core-site.xml
```

将<Configuration> 和 </configuration> 标签中的内容修改如下所示：

```
<!-- 指定Hadoop使用的文件系统为HDFS，并指定NameNode的地址 -->
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://hadoop:9000</value>
</property>

<!-- 指定Hadoop运行时产生的文件存储路径 -->
<property>
  <name>hadoop.tmp.dir</name>
  <value>/opt/hadoop-2.8.5/tmp/</value>
</property>
```

保存退出。

**配置hdfs-site.xml文件：**

```
[root@hadoop hadoop]# vim hdfs-site.xml
```

将 `<Configuration>` 和 `</configuration>` 标签中的内容修改如下所示:

```
<!-- 指定文件副本数量 -->
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>

<!-- 关闭权限检查 -->
<property>
  <name>dfs.permissions.enabled</name>
  <value>>false</value>
</property>
```

保存退出。

#### 配置mapred-site.xml文件:

mapred-site.xml文件在配置目录中没有, 只有mapred-site.xml.template文件, 所以需要拷贝一份:

拷贝mapred-site.xml.template为mapred-site.xml

```
[root@hadoop hadoop]# cp mapred-site.xml.template mapred-site.xml
[root@hadoop hadoop]# vim mapred-site.xml
```

将 `<Configuration>` 和 `</configuration>` 标签中的内容修改如下所示:

```
<!-- 指定MapReduce程序使用yarn组件协调 -->
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

保存退出。

#### 配置yarn-site.xml文件:

```
[root@hadoop hadoop]# vim yarn-site.xml
```

将 `<Configuration>` 和 `</configuration>` 标签中的内容修改如下所示:

```
<!-- 指定ResourceManager的主机名 -->
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>hadoop</value>
</property>

<!-- 指定map和reduce阶段数据传输使用shuffle -->
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
```

保存退出。

- 步骤七: 格式化namenode

刚安装好Hadoop之后（或者安装完HDFS），想要使用，在第一次启动Hadoop之前必须要进行格式化操作，用于生成HDFS文件系统的系统文件。

```
[root@hadoop ~]# hdfs namenode -format
```

首先检查在格式化过程中，有没有报错信息。如果没有报错信息，然后检查如下三个地方：

```
19/10/02 16:19:48 INFO util.GSet: Computing capacity for map NameNodeRetryCache
19/10/02 16:19:48 INFO util.GSet: VM type = 64-bit
19/10/02 16:19:48 INFO util.GSet: 0.0299999993294477463, max memory 966.7 MB - 207.0 KB
19/10/02 16:19:48 INFO util.GSet: capacity
19/10/02 16:19:48 INFO namenode.FSImage: Allocated new block pool ID: b1-950192049-192.168.56.100-1570004388829
19/10/02 16:19:49 INFO common.Storage: Storage directory /opt/hadoop-2.7.3/tmp/dfs/name has been successfully formatted.
19/10/02 16:19:49 INFO namenode.FSImageFormatProtobuf: Saving image file /opt/hadoop-2.7.3/tmp/dfs/name/current/fsimage.ckpt_00000000000000000000 using no compression
19/10/02 16:19:49 INFO namenode.FSImageFormatProtobuf: Image file /opt/hadoop-2.7.3/tmp/dfs/name/current/fsimage.ckpt_00000000000000000000 of size 351 bytes saved in 0 seconds.
19/10/02 16:19:49 INFO namenode.NNStorageRetentionManager: Going to idle txxid >= 0
19/10/02 16:19:49 INFO util.ExitUtil: Exiting with status 0
19/10/02 16:19:49 INFO namenode.NameNode: SHUTDOWN_MSG:
/*
*****
SHUTDOWN_MSG: Shutting down NameNode at hadoop/192.168.56.100
*****
*/
```

- 启动hadoop

```
[root@hadoop ~]# start-all.sh
```

使用 `jps` 指令，查看Hadoop的所有进程如下所示：

```
Jps
NameNode
DataNode
SecondaryNameNode
ResourceManager
NodeManager
```

此时表示Hadoop启动成功。

- 暂停Hadoop

```
[root@hadoop ~]# stop-all.sh
```