# Hadoop的搭建流程

## 搭建服务器集群

### 1.1 在本地创建2~3台CentOS虚拟机（以3台为例）

在没有多台云服务器的情况下，在本地创建多台虚拟机模拟集群。由于虚拟机的配置基本相同，所以可以采用克隆技术先配置一台服务器，然后克隆出剩下2台虚拟机并修改相关配置。

### 1.2 修改虚拟机网络

#### 1.2.1 修改网卡配置文件

修改网络分为两种方式：动态IP、静态IP。在工作中，一般配置静态IP。在虚拟机中，编辑网卡：vi /etc/sysconfig/network-scripts/ifcfg-ens33，修改内容如下：

// 修改

BOOTPROTO=static

ONBOOT=yes

// 添加

IPADDR=192.168.70.110 // 虚拟机静态IP，注意前24位要和网关的前24位一致

NETMASK=255.255.255.0 // 子网掩码

GATEWAY=192.168.70.2 // 网关IP

DNS1=114.114.114.114 // 固定写法

DNS2=8.8.8.8 // 固定写法

#### 1.2.2 重启网络服务

编辑完网卡，保存退出后，执行：service network restart

#### 1.2.3 测试网络

通过ssh远程连接工具连接虚拟机，如果连接成功，则表示配置成功。

1.3 修改主机名  
虚拟机1号：hostnamectl set-hostname 'hadoop01'

### 1.4 克隆出另外2台虚拟机

克隆后，需要修改IP地址和主机名，然后重启即可。

### 1.5 服务器其他系统设置

#### 1.5.1 服务器互传资料、host映射配置

互传资料可以通过scp命令实现：scp -r /usr/local/jdk1.8.0\_211 192.168.70.120:/usr/local/。如果不想每次都输入IP地址，可以配置host映射。配置文件为/etc/hosts。

#### 1.5.2 免密登录

（1）生成密码

每台虚拟机都执行ssh-keygen

1. 将自己的密码告诉其他虚拟机（也要告诉自己一份）

ssh-copy-id hadoop01 // 如果没有配置host映射，要写IP地址

ssh-copy-id hadoop02

ssh-copy-id hadoop03

#### 1.5.3 关闭防火墙

在配置集群的时候，如果不关防火墙，会出现很多问题。

查看防火墙：systemctl status firewalld

临时关闭：systemctl stop firewalld

永久关闭防火墙 (记得先临时关闭，再永久关闭)：systemctl disable firewalld

## 安装Hadoop的依赖环境

### 安装jdk

由于Hadoop底层是Java，所以需要配置jdk环境。

#### 2.1.1 下载一个Linux的jdk的tar包

#### 2.1.2 解压tar包

为了方便管理，通常把压缩文件解压到/usr/local目录下。命令为：tar -zxvf XXXX.tar.gz -C /usr/local。

#### 2.1.3 配置环境变量

Linux环境变量在/etc/profile文件中设置。在文件的末尾添加：export JAVA\_HOME=/usr/local/jdk1.8.0\_211、export PATH=$PATH:$JAVA\_HOME/bin即可。配置完成后，还需要执行：source /etc/profile使其生效。

### 搭建MySQL

#### 去官网下载MySQL的在线安装包

下载下来的是一个.rpm文件。

#### 解压MySQL安装包

.rpm文件的解压命令是：rpm -ivh xxx.rpm。不需要添加归档目录参数，默认解压到 /etc/yum.repos.d。

#### 修改MySQL的在线安装包

编辑/etc/yum.repos.d/mysql-community.repo，将5.7的0改成1，把8.0的1改成0。

#### 在线安装MySQL服务

执行yum -y install mysql-community-server命令

#### 启动

执行service mysqld start命令

#### 登录MySQL并修改密码

在登录时，需要先查看系统提供的临时密码。执行命令：grep 'temporary password' /var/log/mysqld.log。获取到密码后进行登录，对密码进行修改。

在修改密码之前，要先设置密码长度、再降低MySQL安全等级，最后修改密码。依次执行如下命令：

set global validate\_password\_length=6;

set global validate\_password\_policy=0;

ALTER USER 'root'@'localhost' IDENTIFIED BY '123456';

#### MySQL的授权处理

登录MySQL后，依次执行：

use mysql;

update user set host='%' where user ='root';

FLUSH PRIVILEGES;

GRANT ALL PRIVILEGES ON \*.\* TO 'root'@'%'WITH GRANT OPTION;

## Hadoop框架的搭建

### 下载Hadoop的tar包并解压到/usr/local下

### 修改Hadoop的hadoop-env.sh配置文件

修改内容：export JAVA\_HOME=/usr/local/jdk1.8.0\_211

### 修改Hadoop的核心配置文件core-site.xml

<!-- 主节点和端口 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs://hadoop01:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/usr/local/hadoop-2.7.3/tmp</value>

</property>

### 修改Hadoop的hdfs-site.xml配置文件（修改保存元数据与真实数据的存储路径）

<!-- 元数据存储路径 -->

<property>

<name>dfs.namenode.name.dir</name>

<value>/usr/local/hadoop-2.7.3/data/name</value>

</property>

<!-- 真实数据存储路径 -->

<property>

<name>dfs.datanode.data.dir</name>

<value>/usr/local/hadoop-2.7.3/data/data</value>

</property>

<!-- 副本数，一般有几个datanode就设置为几 -->

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<property>

<name>dfs.secondary.http.address</name>

<value>hadoop01:50090</value>

</property>

### 修改mapred-site.xml

注意：该文件是一个临时文件，我们需要把临时文件变成非临时文件，执行命令：mv mapred-site.xml.template mapred-site.xml。

编辑该文件，修改内容如下：

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

### 修改yarn-site.xml

<!-- 配置resourcemanager -->

<property>

<name>yarn.resourcemanager.hostname</name>

<value>hadoop01</value>

</property>

<!-- 配置nodemanager -->

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

### 修改slaves文件，指定从节点是谁

修改内容如下：(把localhost删掉)

hadoop02

hadoop03

### 配置环境变量

export JAVA\_HOME=/usr/local/jdk1.8.0\_211

export HADOOP\_HOME=/usr/local/hadoop-2.7.3

export PATH=$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

### 初始化

hadoop namenode -format （该命令只执行一次：（是成功一次））

### 配置文件分发

将hadoop01上配置好的hadoop，通过scp命令分发到集群的其他服务器上，即hadoop02、hadoop03上

### 启动集群

执行start-all.sh命令。检查是否启动成功有两种办法：jps查看java进程，如果出现了namenode和datanode进程，证明启动成功；查看web界面（IP:50070），如果界面正常显示，证明配置成功。

注意：如果配置失败，要先执行stop-all.sh命令停止集群，再修改错误的地方。