

实 验 报 告

学 号	1902001 1038	姓 名	岳宇轩	专业班级	2019 级计算机科学与技术专业 慧与卓越工程师班
课程名称	大数据导论			学期	2022 年秋季学期
任课教师	刘洁 刘艳艳	完成日期	2022.9.20	上机课时间	周一 56 节（双周）
实 验 名 称	实验一. 搭建 Hadoop 集群				

一、实验要求（10%）

实验内容和要求：

搭建 Hadoop 集群，要求至少是 3 个节点的真分布式集群系统，包含 1 个 NameNode，两个 DataNode。如果机器性能不允许必须做伪分布式集群，必须跟老师提前说明。

1. 软件工具，参考版本，可选择更高版：

- 1) VMware16,
- 2) CentOS 7 镜像 或者 Ubuntu14.04 以上
- 3) JDK 1.8.0
- 4) Hadoop 2.7.7
- 5) Hadoop

2. 集群的安装配置大致为如下流程：

- 1) 准备 3 台客户机（关闭防火墙、静态 IP、主机名称）
- 2) 安装 JDK
- 3) 配置环境变量
- 4) 安装 Hadoop
- 5) 配置环境变量
- 6) 配置集群
- 7) 单点启动
- 8) 配置 ssh
- 9) 集群启动并测试集群

二、实验内容及步骤（80%）

VMware 的安转

按照实验一说明中的内容以及老师提供的下载地址即可

安装 Linux

按照教程中的步骤安装即可，我的虚拟机配置如下：

▼ 设备

内存	1 GB
处理器	1
硬盘 (SCSI)	20 GB

修改 ip 和主机名

关于配置 VM ware 的 IP 地址，按照配置说明中的内容进行即可

我这个好像是最小安装，因此需要先使用配置说明中的指令安装 vim

修改虚拟机 ip

```
TYPE="Ethernet"
PROXY_METHOD="none"
BROWSER_ONLY="no"
BOOTPROTO="static"
DEFROUTE="yes"
IPV4_FAILURE_FATAL="no"
IPV6INIT="yes"
IPV6_AUTOCONF="yes"
IPV6_DEFROUTE="yes"
IPV6_FAILURE_FATAL="no"
IPV6_ADDR_GEN_MODE="stable-privacy"
NAME="ens33"
UUID="0f2b4834-d6d9-4a61-b440-0c19282b1ef5"
DEVICE="ens33"
ONBOOT="yes"

IPADDR=192.168.10.100
GATEWAY=192.168.10.2
DNS1=192.168.10.2
```

这里要使用在操作系统课上学到的一些 vim 知识，比如：

按 esc 退出/进入编辑模式，按 i 进入编辑模式，

： w 保存

： q 退出

： wq 保存并退出

加！是强制命令。

修改主机名称，配置主机名 IP 地址映射也按照配置说明中操作即可。

验证如下：

ifconfig 命令进行验证:

```
[root@namenode sakana]# ifconfig
ens33: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.10.100 netmask 255.255.255.0 broadcast 192.168.10.255
    inet6 fe80::3d9e:6abb:90d7:87a prefixlen 64 scopeid 0x20<link>
    ether 00:0c:29:c0:2a:8f txqueuelen 1000 (Ethernet)
    RX packets 334 bytes 24868 (24.2 KiB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 146 bytes 12012 (11.7 KiB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10<host>
    loop txqueuelen 1000 (Local Loopback)
    RX packets 0 bytes 0 (0.0 B)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 0 bytes 0 (0.0 B)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

[root@namenode sakana]#
```

使用 hostname 命令进行验证:

```
[root@namenode sakana]# hostname
namenode
```

接下来的步骤都按照配置说明中的来操作即可, 在实际操作的过程中也没有遇到什么问题

创建完毕 module 和 software 之后, 验证如下

```
[root@namenode sakana]# cd /opt/
[root@namenode opt]# ll
total 0
drwxr-xr-x. 4 hadoop hadoop 46 Sep 16 14:06 module
drwxr-xr-x. 2 hadoop hadoop 67 Sep 16 14:02 software
[root@namenode opt]#
```

由于我是最小安装, 因此没有要卸载的 Jdk

我之前就用过 xshell 和 xftp, 所以就直接用来导入了。导入过程遇到的问题见心得总结 2

接下来的操作按照配置说明中执行即可, 最后测试 Jdk 是否安装成功:

```

namenode login: hadoop
Password:
Last login: Sat Sep 17 11:21:04 on tty1
[hadoop@namenode ~]$ java -version
java version "1.8.0_212"
Java(TM) SE Runtime Environment (build 1.8.0_212-b10)
Java HotSpot(TM) 64-Bit Server VM (build 25.212-b10, mixed mode)
[hadoop@namenode ~]$ _

```

成功安装 jdk。

安装 Hadoop 的过程也没有遇到什么大问题，照着配置说明操作即可。

下面进行虚拟机的克隆，照着配置说明来做即可。

ssh 配置也是，照着一步一步来操作。

下面来到配置集群，这四个集群配置文件里的内容我不能直接复制到虚拟机中，于是我现在 win 上修改好文件之后用 xftp 上传并替换掉了虚拟机中的文件。其它按照配置说明操作即可。

最后是启动集群，最终结果如下：

Overview 'namenode:8020' (active)

Started:	Sun Sep 25 16:53:24 +0800 2022
Version:	3.1.3, rba631c436b067289ec2f54ab1e289526c90579
Compiled:	Thu Sep 12 10:47:00 +0800 2019 by zhang from branch-3.1.3
Cluster ID:	CID-558e0bc6-309f-4fbb-9482-b527219d524
Block Pool ID:	BP-193889274-192.168.10.100-1663385342745

Summary

Security is off.
 Safemode is off.
 1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
 Heap Memory used 23.11 MB of 28.63 MB Heap Memory. Max Heap Memory is 235.68 MB.
 Non Heap Memory used 45.57 MB of 46.66 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	41.89 GB
Configured Remote Capacity:	0 B
DFS Used:	28 KB (0%)
Non DFS Used:	8.83 GB
DFS Remaining:	33.06 GB (79.92%)
Block Pool Used:	28 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%

[Live Nodes](#) 3 (Decommissioned: 0, In Maintenance: 0)

修改主机名称

执行以下命令，进入之后修改主机名称为 namenode

```
[root@**** ~]#vim /etc/hostname
```

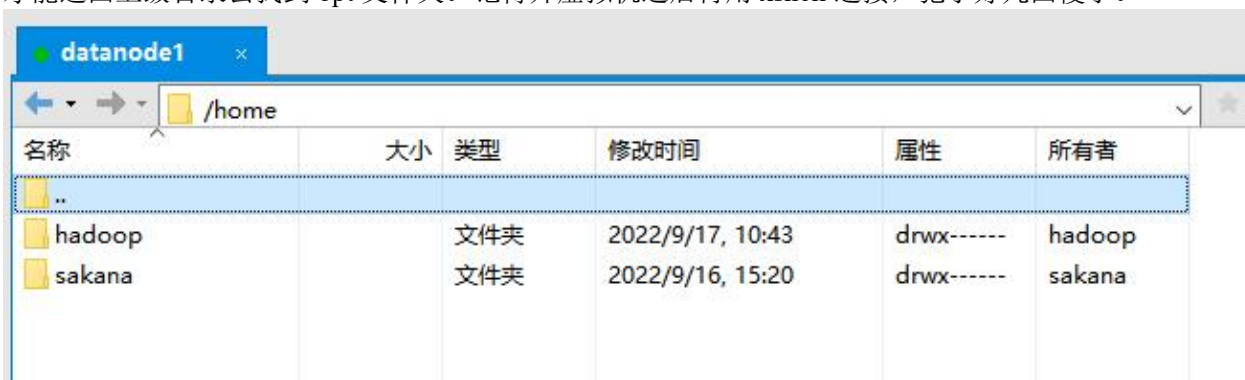
在用 vim 打开这个文件后，我发现的是 localhost.localsakana
因为我用户名就是 sakana，所以我认为前面的 localhost.local 是不能修改的，所以我就改成了 localhost.localnamenode

但后来发现只要把一整行全改成 namenode 就行了



```
namenode
```

我遇到的第二个问题是，在导入 Jdk 的过程中，我没有办法连接到我的虚拟机，经过一番尝试之后发现，我在 xshell 的连接登录时应该输入的是 hadoop 的用户账号和密码，否则是没有权限的。然后我访问虚拟机之后一开始是进到了一个由 hadoop 文件夹和 sakana 文件夹的地方，需要点最上面的..才能返回上级目录去找到 opt 文件夹。记得开虚拟机之后再用 xshell 连接，犯了好几回傻了。



名称	大小	类型	修改时间	属性	所有者
..					
hadoop		文件夹	2022/9/17, 10:43	drwx-----	hadoop
sakana		文件夹	2022/9/16, 15:20	drwx-----	sakana

我遇到的第三个问题是在集群配置时使用 scp 指令，

```
[hadoop@namenode hadoop]$ scp -r /opt/module/hadoop-3.1.3/etc/hadoop/
hadoop@datanode1:/opt/module/hadoop-3.1.3/etc/hadoop/

[hadoop@namenode hadoop]$ scp -r
/opt/module/hadoop-3.1.3/etc/hadoop/
hadoop@datanode2:/opt/module/hadoop-3.1.3/etc/
```

上图是配置说明中给出的 scp 指令，但我照着输入时遇到了问题，没有办法在正确的移动，所以我新建了一个文件夹用所实验，尝试修改上述 scp 指令为

```
scp -r hadoop@namenode:/opt/module/hadoop-3.1.3/etc/hadoop/
hadoop@datanode1:/opt/module/hadoop-3.1.3/etc/
```

在上述命令中，与配置说明中有出路的是命令的第 3、4 个参数，其中我在 3、4 参数中都标注了用户名

为 `hadoop` 以及各自的主机名，最让人迷惑不清的是后面的两个地址。

首先我们要清楚，我们的目的是把已经配置好的 `namenode` 主机上的 `hadoop` 文件夹替换给 `datanode1` 主机，所以我们的第 3 个参数最后一定要落到 `hadoop` 文件夹上（不用最后加/），而第四个参数最后的 `etc` 是 `hadoop` 的上级文件夹，再加一个/就表示打开这个 `etc` 文件夹。如此就是正确的做法了。

我遇到的第四个问题是，我在启动集群时，`yarn` 可以正常启动，但是 `hdfs` 不行，会报错提示 `permission denied`，我思考了一下，认为是权限的问题，按照步骤排查，再加上上网搜索资料，重新配置了一遍 `ssh` 之后就好了。

在本次实验中，我更加熟悉了 `linux` 系统的操作，`vim` 的使用，`xshell` 和 `xftp` 的使用。对于如何搭建 `hadoop` 集群也有了一定的体会。