

电 子 科 技 大 学

实 验 报 告

学生姓名：Lolipop 学号：2018091202000 指导教师：xx

实验地点：信软学院楼西 400 实验时间：2020.11.18

一、实验名称：构建 Hadoop 单机环境模式

二、实验学时：4 学时

三、实验目的：

1. 掌握 Hadoop 的环境搭建；
2. 掌握 Hadoop、Linux 的基本命令。

四、实验原理：

Apache Hadoop 是一款支持数据密集型分布式应用程序并以 Apache 2.0 许可协议发布的开源软件框架。它支持在商用硬件构建的大型集群上运行的应用程序。Hadoop 是根据谷歌公司发表的 MapReduce 和 Google 文件系统的论文自行实现而成。所有的 Hadoop 模块都有一个基本假设，即硬件故障是常见情况，应该由框架自动处理。

Hadoop 框架透明地为应用提供可靠性和数据移动。它实现了名为 MapReduce 的编程范型：应用程序被分割成许多小部分，而每个部分都能在集群中的任意节点上运行或重新运行。此外，Hadoop 还提供了分布式文件系统，用以存储所有计算节点的数据，这为整个集群带来了非常高的带宽。MapReduce 和分布式文件系统的设计，使得整个框架能够自动处理节点故障。它使应用程序与成千上万的独立计算的电脑和 PB 级的数据连接起来。现在普遍认为整个 Apache Hadoop“平

台”包括 Hadoop 内核、MapReduce、Hadoop 分布式文件系统（HDFS）以及一些相关项目，有 Apache Hive 和 Apache HBase 等等。

五、实验内容：

在 Ubuntu 系统上搭建 Hadoop 环境，并在此基础上学习使用基本的 Hadoop 以及 Linux 命令。

六、实验器材（设备、元器件）：

1. Ubuntu 20.04
2. jdk 1.8
3. hadoop 2.10.1

七、实验步骤：

1. 创建用户；
2. 安装 ssh；
3. 安装 java；
4. 安装 hadoop；
5. 测试 wordcount。

八、实验结果与分析（含重要数据结果分析或核心代码流程分析）

1. 创建用户 hadoop。结果如图 1-1 所示。
 - a) 打开终端窗口，创建新用户。
 - b) 修改新创建用户的密码（密码为学号）。
 - c) 接着为新用户增加管理员的权限，以免之后出现权限问题。

```

lolipop@lolipop-VirtualBox:~$ sudo useradd -m hadoop -s /bin/bash
[sudo] password for lolipop:
lolipop@lolipop-VirtualBox:~$ sudo passwd hadoop
New password:
Retype new password:
passwd: password updated successfully
lolipop@lolipop-VirtualBox:~$ sudo adduser hadoop sudo
Adding user `hadoop' to group `sudo' ...
Adding user hadoop to group sudo
Done.

```

图 1-1 创建用户 hadoop

2. 安装 ssh。

a) 切换用户为 hadoop。

b) 安装 ssh。

```
$ sudo apt-get install openssh-server
```

c) 打开这个 ssh 服务。

```
$ sudo /etc/init.d/ssh start
```

d) 查看 ssh 服务是否正确启动，结果如图 2-1 所示。

```
$ ps -e | grep ssh
```

```

hadoop@lolipop-VirtualBox:~$ sudo /etc/init.d/ssh start
Starting ssh (via systemctl): ssh.service.
hadoop@lolipop-VirtualBox:~$ ps -e | grep ssh
  944 ?          00:00:00 ssh-agent
 2328 ?          00:00:00 sshd

```

图 2-1 查看 ssh 服务

e) 设置免密码登录，生成私钥和公钥。生成两个文件 id_rsa、id_rsa.pub，前者是私钥，后者是公钥。

```
$ ssh-keygen -t rsa -P ""
```

f) 接着将公钥追加到 authorized_keys 里面，它保存所有允许以当前用户身

份登录到 ssh 客户端用户的公钥内容。结果如图 2-2 所示。

```
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
hadoop@lolipop-VirtualBox:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa):
Created directory '/home/hadoop/.ssh'.
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:sqVPZ8aL/Nkuql1bfHehXQnJZAP82KRFQ0YsxqVq/xc hadoop@lolipop-VirtualBox
The key's randomart image is:
+---[RSA 3072]---+
|                 |
|  o*+=          |
|  . *.+oo       |
|  = oB+         |
|  o .o o. .     |
|  .oS          o.|
|  .=. . . Eo o  |
|  o ..* o.o.o   |
|  = *o* o . .   |
|  ..*o=++       |
+-----[SHA256]-----+
hadoop@lolipop-VirtualBox:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
```

图 2-2 生成密钥并配置免密登录

g) 登陆 ssh。结果如图 2-3 所示。

```
$ssh hadoop@localhost
```

```
hadoop@lolipop-VirtualBox:~$ ssh hadoop@localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:x8z3aaUgwAPV63B47tN2G4CXQeYLPJJJoqGdtRNAZbF0.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
hadoop@localhost's password:
Welcome to Ubuntu 20.04.1 LTS (GNU/Linux 5.4.0-56-generic x86_64)
```

图 2-3 登录 ssh

3. 安装 java。

a) 使用以下命令在 Ubuntu 20.04 系统中安装 Java 8, 并配置环境变量。

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

b) 检验 Java 是否安装成功, 如图 3-1 所示。

```
hadoop@lolipop-VirtualBox:/usr/local/hadoop$ java -version
openjdk version "1.8.0_275"
OpenJDK Runtime Environment (build 1.8.0_275-8u275-b01-0ubuntu1~20.04-b01)
OpenJDK 64-Bit Server VM (build 25.275-b01, mixed mode)
```

图 3-1 检验 Java 安装成功

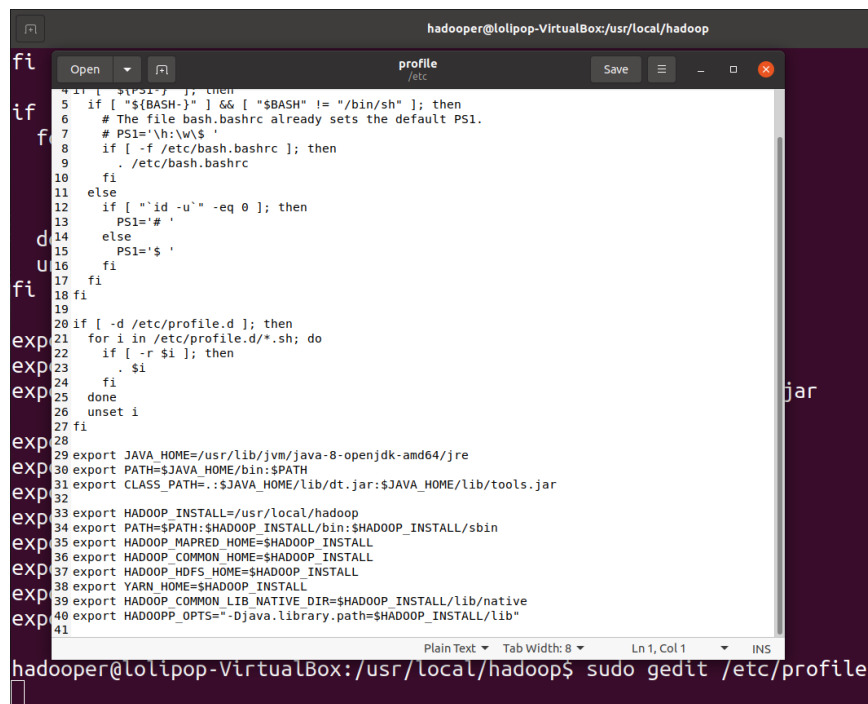
4. 安装 hadoop。

- a) 解压 hadoop-2.10.1 版本的文件到/usr/local 目录下，为方便配置，修改文件夹名为 hadoop。如图 4-1 所示。

```
hadoop@lolipop-VirtualBox:~$ cd /usr/local
hadoop@lolipop-VirtualBox:/usr/local$ ls
bin  etc  games  hadoop-2.10.1  include  lib  man  sbin  share  src
hadoop@lolipop-VirtualBox:/usr/local$ sudo mv hadoop-2.10.1 hadoop
hadoop@lolipop-VirtualBox:/usr/local$ ls
bin  etc  games  hadoop  include  lib  man  sbin  share  src
hadoop@lolipop-VirtualBox:/usr/local$ sudo chown hadoop:sudo -R -f /usr/local/hadoop
```

图 4-1 安装 hadoop 2.10.1

- b) 配置各种文件，最后编辑配置文件/etc/profile 添加命令语句。如图 4-2 所示。



```
hadoop@lolipop-VirtualBox:/usr/local/hadoop$ sudo gedit /etc/profile
```

```
41 if [ "${PS1}" = " " ]; then
5   if [ "${BASH}" = " " ] && [ "${BASH}" != "/bin/sh" ]; then
6     # The file bash.bashrc already sets the default PS1.
7     # PS1='\h:\w$ '
8     if [ -f /etc/bash.bashrc ]; then
9       . /etc/bash.bashrc
10    fi
11  else
12    if [ "`id -u`" = "0" ]; then
13      PS1='# '
14    else
15      PS1='$ '
16    fi
17  fi
18 fi
19
20 if [ -d /etc/profile.d ]; then
21   for i in /etc/profile.d/*.sh; do
22     if [ -r $i ]; then
23       . $i
24     fi
25   done
26   unset i
27 fi
28
29 export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/jre
30 export PATH=$JAVA_HOME/bin:$PATH
31 export CLASS_PATH=.:$JAVA_HOME/lib/rt.jar:$JAVA_HOME/lib/tools.jar
32
33 export HADOOP_INSTALL=/usr/local/hadoop
34 export PATH=$PATH:$HADOOP_INSTALL/bin:$HADOOP_INSTALL/sbin
35 export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
36 export HADOOP_COMMON_HOME=$HADOOP_INSTALL
37 export HADOOP_HDFS_HOME=$HADOOP_INSTALL
38 export YARN_HOME=$HADOOP_INSTALL
39 export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
40 export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
41
```

图 4-2 编辑配置文件

5. 测试 wordcount。

a) 应用配置。

```
source /etc/profile
```

b) 进入 hadoop 目录。

```
cd /usr/local/hadoop
```

c) 创建目录和测试用例。

```
$ sudo mkdir input  
$ sudo cp README.txt input
```

d) 运行测试样例：

```
sudo bin/hadoop jar  
share/hadoop/mapreduce/sources/hadoop-mapreduce-examples-2.7.3-sources  
.jar org.apache.hadoop.examples.WordCount input output
```

e) 执行 `cat output/*`，查看字符统计结果。结果如图 4-3 所示。

```
hadooper@lolipop-VirtualBox:/usr/local/hadoop$ cat output/*  
(BIS), 1  
(ECCN) 1  
(TSU) 1  
(see 1  
5D002.C.1, 1  
740.13) 1  
<http://www.wassenaar.org/> 1  
Administration 1  
Apache 1  
BEFORE 1  
BIS 1  
Bureau 1  
Commerce, 1  
Commodity 1  
Control 1  
Core 1  
Department 1  
ENC 1  
Exception 1  
Export 2  
For 1  
Foundation 1  
Government 1  
Hadoop 1
```

图 4-3 测试 WordCount

九、总结及心得体会：

// removed

十、对本实验过程及方法、手段的改进建议：

// removed

报告评分：

指导教师签字：