

云计算实验报告

班级： 物联网 1603 学号： 20161670320 姓名： 郭治洪 指导老师： 魏蔚

一、实验题目

Linux 下 Hadoop 的安装与配置。

二、实验内容

1. 下载 OpenJDK，解压并且配置环境变量
2. 下载 Hadoop，解压并且配置环境变量
3. 查看 Hadoop 官方文档，了解如何运行示例程序。

三、实验步骤

1. 新建脚本 `vim hadoop.sh`

```
#!/bin/bash

#Download JDK and Hadoop, then insatll then and run hadoop's
example.

### Please set the jdk and hadoop download url
JDK_DOWNLOAD_URL="https://download.java.net/java/GA/jdk12/GPL/
openjdk-12_linux-x64_bin.tar.gz"
Hadoop_DOWNLOAD_URL="http://mirrors.tuna.tsinghua.edu.cn/apach
e/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz"

### Do not edit the following sections.

[ $(id -u) != "0" ] && { echo "${CFAILURE}Error: You must run this
as root or sudo.${CEND}"; exit 1; }

# If you need to use again, please del all old folder and user.
# rm -rf /home/hadoop
# userdel -rf hadoop

#Create user hadoop to use.
#Thanks for sunrise's suggestion

echo "Create hadoop user,the password is hadoop."
useradd -m hadoop -s /bin/bash
echo hadoop:hadoop|chpasswd
```

```
# Edit /etc/sudoers to open wheel group then add user hadoop into wheel group
```

```
sed -i 's/#%wheel ALL=(ALL)  ALL/%wheel ALL=(ALL)  ALL/g' /etc/sudoers
usermod -a -G wheel hadoop
```

```
echo "Set the authorized_keys for user hadoop."
sudo -u hadoop ssh-keygen -t rsa -P '' -f /home/hadoop/.ssh/id_rsa
sudo -u hadoop touch /home/hadoop/.ssh/authorized_keys
sudo -u hadoop cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
sudo -u hadoop chmod 0600 /home/hadoop/.ssh/authorized_keys
```

```
workpath=/home/hadoop/software
echo "Set the software path: ${workpath}"
mkdir -p ${workpath}
cd ${workpath}
```

```
echo "Install the base tools"
yum -y install wget tar ssh sshd pdsh rsync
```

```
echo "Try to download JDK."
wget -t 3 -T 30 --no-check-certificate -O "${workpath}/jdk.tar.gz" "${JDK_DOWNLOAD_URL}"
echo "Try to download Hadoop."
wget -t 3 -T 30 --no-check-certificate -O "${workpath}/hadoop.tar.gz" "${HADOOP_DOWNLOAD_URL}"
```

```
echo "Unzip JDK and Hadoop."
tar -xvf ${workpath}/jdk.tar.gz
tar -xvf ${workpath}/hadoop.tar.gz
```

```
echo "Rename their folder."
mv ${workpath}/jdk-* ${workpath}/jdk
mv ${workpath}/hadoop-* ${workpath}/hadoop
```

```
echo "Set environment variables"
export JAVA_HOME=${workpath}/jdk
export HADOOP_HOME=${workpath}/hadoop
```

```

export
CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar
export
HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$CLASSPATH:~$HADOOP_HOME/bin/hadoop classpath`
export
PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$JAVA_HOME/bin:${PATH}
export PDSH_RCMD_TYPE=ssh
echo "export JAVA_HOME=${workpath}/jdk" >> /home/hadoop/.bashrc
echo "export HADOOP_HOME=${workpath}/hadoop" >>
/home/hadoop/.bashrc
echo "export
CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar" >>
/home/hadoop/.bashrc
echo "export
HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$CLASSPATH:~$HADOOP_HOME/bin/hadoop classpath`" >> /home/hadoop/.bashrc
echo
"PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$JAVA_HOME/bin:${PATH}
}" >> /home/hadoop/.bashrc
echo "PDSH_RCMD_TYPE=ssh" >> /home/hadoop/.bashrc
source /home/hadoop/.bashrc

```

```

chown -R hadoop:hadoop /home/hadoop

```

```

# The set environment will lost when user used command sudo, so
i have to use other way
#echo "Start to test hadoop by example"
#sudo -u hadoop hadoop version
#sudo -u hadoop mkdir ${workpath}/test
#sudo -u hadoop cd ${workpath}/test
#sudo -u hadoop mkdir input
#sudo -u hadoop cp ${workpath}/hadoop/etc/hadoop/*.xml input
#sudo -u hadoop jar
${workpath}/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-exa
mples-*.jar grep input output 'dfs[a-z.]+' grep input output
'dfs[a-z.]+'
#sudo -u hadoop cat output/*

```

```

echo "Start to test hadoop by example"
su hadoop << TEST
hadoop version
mkdir ${workpath}/test

```

```

cd ${workpath}/test
mkdir input
cp ${workpath}/hadoop/etc/hadoop/*.xml input
hadoop jar
${workpath}/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep input output 'dfs[a-z.]+' grep input output 'dfs[a-z.]+'
cat output/*

TEST

```

2. 赋予脚本权限 `chmod a+x hadoop.sh`

3. 等待脚本执行完成

四、实验结果

```

2019-03-23 12:19:23,819 INFO mapred.LocalJobRunner: Finishing task: attempt_local24552267_0002_r_000000_0
2019-03-23 12:19:23,820 INFO mapred.LocalJobRunner: reduce task executor complete.
2019-03-23 12:19:24,626 INFO mapreduce.Job: Job job_local24552267_0002 running in uber mode : false
2019-03-23 12:19:24,626 INFO mapreduce.Job: map 100% reduce 100%
2019-03-23 12:19:24,626 INFO mapreduce.Job: Job job_local24552267_0002 completed successfully
2019-03-23 12:19:24,632 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=1337718
  FILE: Number of bytes written=3261566
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Map output bytes=17
  Map output materialized bytes=25
  Input split bytes=130
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=25
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=30
  Total committed heap usage (bytes)=370753536
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=123
File Output Format Counters
  Bytes Written=23
1 dfsadmin
[root@Aliyun-HK ~]#

```

```
CentOS7-VM - Xshell 6 (Free for Home/School)
文件(F) 编辑(E) 查看(V) 工具(T) 选项卡(B) 窗口(W) 帮助(H)
ssh://root:*****@192.168.216.131:22
要添加当前会话，点击左侧的箭头按钮。
1 CentOS7-VM x +
Total committed heap usage (bytes)=184324096
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Output Format Counters
  Bytes Written=23
2019-03-22 18:57:21,674 INFO mapred.LocalJobRunner: Finishing task: attempt_local2045077908_0002_r_000000_0
2019-03-22 18:57:21,675 INFO mapred.LocalJobRunner: reduce task executor complete.
2019-03-22 18:57:22,341 INFO mapreduce.Job: Job job_local2045077908_0002 running in uber mode : false
2019-03-22 18:57:22,342 INFO mapreduce.Job: map 100% reduce 100%
2019-03-22 18:57:22,342 INFO mapreduce.Job: Job job_local2045077908_0002 completed successfully
2019-03-22 18:57:22,345 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=1335226
  FILE: Number of bytes written=3265534
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Map output bytes=17
  Map output materialized bytes=25
  Input split bytes=109
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=25
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=36
  Total committed heap usage (bytes)=368648192
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=123
File Output Format Counters
  Bytes Written=23
[root@localhost ~]# cat output/*
1
dfsadmin
[root@localhost ~]#
```

五、实验的心得体会

我首先在自己的虚拟机配置了 hadoop 并且取得了成功，于是我就在想能不能更加智能写一个脚本来实现，和我们学长姚东阳帮助得知他也有这方面的研究和想法，于是他给我提了一些建议和参考，在他帮助下，我成功实现。

我会更深入的。