

华东师范大学数据科学与工程学院实验报告

课程名称： 分布式模型与编程	年级： 2016 级	上机实践成绩：
指导教师： 徐辰	姓名： 张宏伟	
上机实践名称： Hadoop 的部署与常用命令	学号： 10165101180	上机实践日期：
上机实践编号： 03	组号：	上机实践时间：

一、实验目的

熟悉 Hadoop 的部署与常用命令的使用

二、实验任务

尝试搭建 Hadoop 的三种模式：本地，伪分布式，分布式

三、使用环境

本地，伪分布式：hadoop3.1.1+Ubuntu18.04

分布式：hadoop2.7.7+Docker+ubuntu

四、实验过程

1.本地

一开始配 JAVA 环境，用的 zsh，source 有问题，始终无法使 JAVA_HOME 等环境变量生效，后面把 shell 切换到 bash 后正常。

NameNode 用 SSH 启动集群中机器的 Hadoop 守护进程，故需要配置 SSH 的免密钥登陆，在配置中最后一步验证 ssh local 时报错，发现只有一个 ssh 的进程，说明 openssh-server 没安装，安装后正常。

```
hadoop@PC-honwee:~$ ssh localhost
ssh: connect to host localhost port 22: Connection refused
hadoop@PC-honwee:~$ ps -e|grep ssh
1083 ?        00:00:00 ssh-agent
hadoop@PC-honwee:~$ sudo apt-get install openssh-server
```

下载 hadoop 并安装成功。

```
hadoop@PC-honwee:/usr/local$ cd hadoop/
hadoop@PC-honwee:/usr/local/hadoop$ ./bin/hadoop version
Hadoop 3.1.1
Source code repository https://github.com/apache/hadoop -r 2b9a8c1d3a2caf1e733d57f346af3ff0d5ba529c
Compiled by leftnoteasy on 2018-08-02T04:26Z
Compiled with protoc 2.5.0
From source with checksum f76ac55e5b5ff0382a9f7df36a3ca5a0
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3.1.1.jar
```

Hadoop 默认为本地模式，故安装后用样例库中的 grep 测试是否可以运行。

将/bin/hadoop 下的所有 xml 文件复制到 input，查找这些文件内以 dfs 开头的单词并统计出现次数。

```
hadoop@PC-honwee:/usr/local/hadoop$ mkdir input
hadoop@PC-honwee:/usr/local/hadoop$ cp ./etc/hadoop/*.xml ./input
hadoop@PC-honwee:/usr/local/hadoop$ ./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.1.jar grep ./input ./output 'dfs[a-z.]+'
```

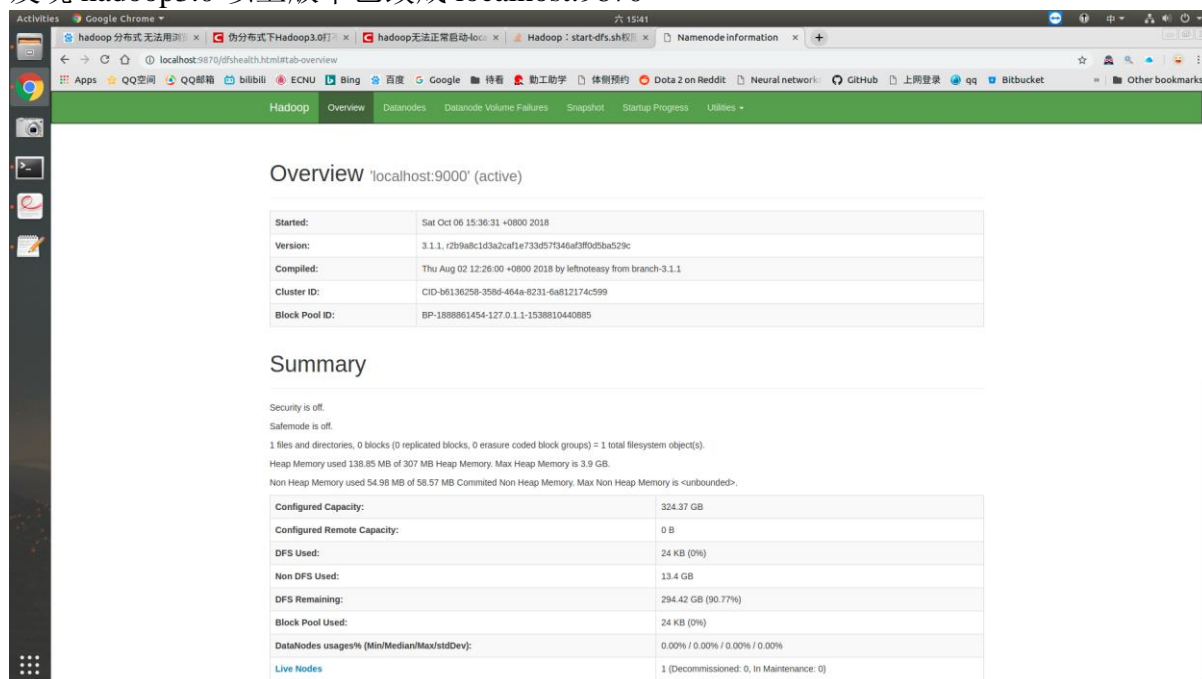
结果

```
hadoop@PC-honwee:/usr/local/hadoop$ cat ./output/*
1      dfsadmin
```

2.伪分布式

配置 hadoop 环境。

按照教材和教程，尝试用 localhost:50070 在浏览器中打开 hadoop 失败，发现 hadoop3.0 以上版本已改成 localhost:9870



在伪分布式下运行 grep 样例

创建 input 文件夹时没有加-p 会报错。

```
hadoop@PC-honwee:/usr/local/hadoop$ ./bin/hdfs dfs -mkdir input
mkdir: `input': No such file or directory
hadoop@PC-honwee:/usr/local/hadoop$ ./bin/hdfs dfs -mkdir -p input
```

运行结果

```
hadoop@PC-honwee:/usr/local/hadoop$ ./bin/hdfs dfs -cat output/*
1      dfsadmin
1      dfs.replication
1      dfs.namenode.name.dir
1      dfs.datanode.data.dir
```

3.分布式

为了和教材演示的 hadoop 版本保持一致，分布式的 hadoop 版本换成 2.7.7，在 Docker 中制作镜像，并用做好的镜像创造多个容器进行 hadoop 分布式集群配置模拟。

制作镜像：更新软件列表，安装 vim, ssh, java-11-jdk, hadoop2.7.7，配置 ssh 免密钥登录，配置 hadoop 中的 core-site.xml, hdfs-site.xml, mapred-site.xml, 及 yarn-site.xml 文件，最后保存镜像。

用上述镜像创建三个容器，并命名为 master, slave01, slave02, hostname 的命名也是如此。分别配置三个容器的 hosts。

启动后，分别在各个容器中输入 jps 的结果：

```
root@master:/usr/local/hadoop# jps
498 SecondaryNameNode
280 NameNode
956 Jps
685 ResourceManager
root@master:/usr/local/hadoop#
root@slave01:/# jps
246 NodeManager
366 Jps
126 DataNode
root@slave01:/#
root@slave02:/# jps
246 NodeManager
366 Jps
126 DataNode
root@slave02:/#
```

分布式下运行样例：

```
root@master:/usr/local/hadoop# ./bin/hdfs dfs -cat output/*
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security
/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.7.jar) to
)
WARNING: Please consider reporting this to the maintainers of or
.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further
WARNING: All illegal access operations will be denied in a futur
1      dfsadmin
1      dfs.replication
1      dfs.namenode.name.dir
1      dfs.datanode.data.dir
root@master:/usr/local/hadoop#
```

五、总结

1. 第一次碰到因为 shell 不是 bash 出现问题，下次碰到同样的情况会有所警觉。
2. 安装版本最好和教材一致，避免因版本更新造成命令的变更。
3. 分布式部署实验中，Docker 比 VMware 虚拟机来得方便和高效。