第 1 篇：安装 Hadoop

1. 确保 Linux 系统和 Windows 系统能够共享文件夹

需要重新安装 vmware-tools

具体参考文章：https://www.cnblogs.com/ygh1229/p/6379817.html

2.安装 SSH

Ubunt 已经安装了 SSH 客户端，现在只需要安装 SSH 服务器端

执行下列命令

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

然后使用如下命令登录本机

```
ssh localhost
```

最后去掉 ssh 登录的密码

先执行

```
exit  #退出 ssh 登录
```

然后执行以下 3 条命令

```
cd ~/.ssh/   #切换到 ssh 目录，加"."表示隐藏目录
ssh-keygen -t rsa   #生成 2 个密钥 id_rsa 和 id_rsa.pub
cat ./id_rsa.pub >> ./authorized_keys   #显示前面文件内容并追加到后面文件内容的末尾
```

其中：“./” 是当前目录

再次执行：

ssh localhost

无需密码即可登录

3.安装 JDK7

因为 Ubuntu16 版自带的是 JDK8，要想安装 JDK7，需要执行：

```
sudo add-apt-repository ppa:openjdk-r/ppa     #repository 是仓库的意思
sudo apt-get update   #更新 apt
sudo apt-get install openjdk-7-jre openjdk-7-jdk
```

4.下载安装文件

(1)目录的含义

“./” 是当前目录，“/”代表根目录、“..”代表上一级目录、“~”代表 HOME 目录、“-”代表前一目录。

(2)解压 Hapdoop2.7

```
sudo tar -zxf ~/下载/hadoop-2.7.1.tar.gz -C /usr/local  #解压
cd /usr/local/                  #切换目录
sudo mv ./hadoop-2.7.1/ ./hadoop   #重命名
sudo chown -R hadoop ./hadoop   #修改文件权限
```

(3)检查 Hadoop 是否可用

```
cd /usr/local/hadoop
./bin/hadoop version
```

5.伪分布式模式配置

（1）修改配置文件

Hadoop 的配置文件位于/usr/local/hadoop/etc/hadoop/中，需要修改 2 个配置文件，即 core-site.xml 和 hdfs-site.xml4

1）用 vim 编辑器打开 core-site.xml。改变内容如下：

```
<configuration>
```

```
<property>
   <name>hadoop.tmp.dir</name>
   <value>file:/usr/local/hadoop/tmp</value>
   <description>Abase for other temporary directories.</description>
</property>
<property>
   <name>fs.defaultFS</name>
   <value>hdfs://localhost:9000</value>
</property>
</configuration>
```

2）用 vim 编辑器打开 hdfs-site.xml。改变内容如下：

```
<configuration>
<property>
   <name>dfs.replication</name>
   <value>1</value>
</property>
<property>
   <name>dfs.namenode.name.dir</name>
   <value>file:/usr/local/hadoop/tmp/dfs/name</value>
</property>
<property>
   <name>dfs.datanode.data.dir</name>
   <value>file:/usr/local/hadoop/tmp/dfs/data</value>
</property>
</configuration>
```

(2)执行名称节点格式化

输入如下命令：

```
cd /usr/local/hadoop
./bin/hdfs namenode -format
```

(3)启动 Hadoop

执行如下命令：

```
cd /usr/local/hadoop
./sbin/start-dfs.sh
```

(4)判断 Hadoop 是否启动成功

执行命令

```
jps
```

列出

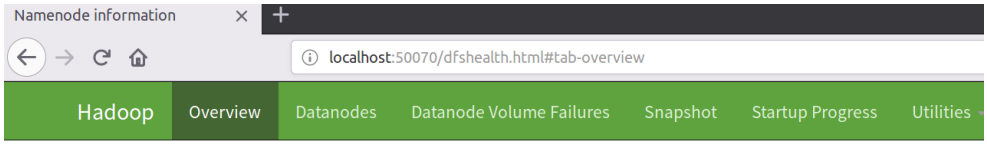
NameNode、DataNode 和 SecondaryNameNode 进程

```
hadoop@ubuntu:/usr/local/hadoop$ jps
4043 NameNode
4189 DataNode
4366 SecondaryNameNode
4485 Jps
```

（5）使用 Web 界面查看 HDFS 信息

Hadoop 成功启动后，打开 Linux 下火狐浏览器，在地址栏输入：

http://localhost:50070 可以看看如下页面：



Overview 'localhost:9000' (active)

Started:	Fri Jul 27 14:52:33 GMT+08:00 2018
Version:	2.7.1, r15ecc87ccf4a0228f35af08fc56de536e6ce657a
Compiled:	2015-06-29T06:04Z by jenkins from (detached from 15ecc87)
Cluster ID:	CID-aef2a570-a786-4aa3-be3b-30f0449eeaed
Block Pool ID:	BP-387006502-127.0.1.1-1532673936122

（6）运行 Hadoop 伪分布式实例

1）在 HDFS 中创建用户目录

HDFS 的操作命令如下：

```
cd /usr/local/hadoop
./bin/hdfs dfs -mkdir -p /user/hadoop
```

2）把本地文件系统的/usr/local/Hadoop/etc/hadoop 目录中的所有 xml 文件作为输入文件，复制到分布式文件系统 HDFS 中的/user/Hadoop/input 目录中，命令如下：

```
cd /usr/local/hadoop
./bin/hdfs dfs -mkdir input
./bin/hdfs dfs -put ./etc/hadoop/*.xml input
```

3）复制完成后，查看 HDFS 中的文件列表

```
./bin/hdfs dfs -ls input
```

```
hadoop@ubuntu:/usr/local/hadoop$ ./bin/hdfs dfs -mkdir -p /user/hadoop
hadoop@ubuntu:/usr/local/hadoop$ ./bin/hdfs dfs -mkdir input
hadoop@ubuntu:/usr/local/hadoop$ ./bin/hdfs dfs -put ./etc/hadoop/*.xml input
hadoop@ubuntu:/usr/local/hadoop$ ./bin/hdfs dfs -ls input
Found 8 items
-rw-r--r--   1 hadoop supergroup       4436 2018-07-27 15:48 input/capacity-sche
duler.xml
-rw-r--r--   1 hadoop supergroup       1034 2018-07-27 15:48 input/core-site.xml
-rw-r--r--   1 hadoop supergroup       9683 2018-07-27 15:48 input/hadoop-policy
.xml
-rw-r--r--   1 hadoop supergroup       1079 2018-07-27 15:48 input/hdfs-site.xml
-rw-r--r--   1 hadoop supergroup        620 2018-07-27 15:48 input/httpfs-site.x
ml
-rw-r--r--   1 hadoop supergroup       3518 2018-07-27 15:48 input/kms-acls.xml
-rw-r--r--   1 hadoop supergroup       5511 2018-07-27 15:48 input/kms-site.xml
-rw-r--r--   1 hadoop supergroup        690 2018-07-27 15:48 input/yarn-site.xml
```

4）运行 Hadoop 自带的 grep 程序

■ 查看 Hadoop 所有例子

```
./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.1.jar
```

■ 创建 input 目录，复制配置文件到 input 目录下

```
mkdir input
cp ./etc/hadoop/*.xml ./input
```

■ 运行 Hadoop 自带的 grep 程序

```
./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep ./input ./output 'dfs[a-z.]+'
```

5）运行结束后，可以通过如下命令查看 HDFS 中的 output 文件夹中的内容：

```
./bin/hdfs dfs -cat output/*
```

6）再次执行 grep 程序前，删除 HDFS 中的 output 文件夹

```
./bin/hdfs dfs -rm -r output    #删除 output 文件夹
```

（6）配置 PATH 变量

打开~/.bashrc 这个文件，然后在文件最前面位置加入如下单独一行：

```
exprot PATH=$PATH:/usr/local/hadoop/sbin:/usr/local/hadoop/bin
```

运行下面命令使配置生效

```
source ~/.bashrc
```

(7) 停止 Hadoop 命令

```
cd /usr/local/hadoop
./sbin/stop-dfs.sh
```