CDH 伪分布式环境搭建（适用于学习测试开发集群模式）

安装环境服务部署规划

服务器 IP	192.168.52.100	192.168.52.110	192.168.52.120
HDFS	NameNode		
	Secondary NameNode		
	DataNode	DataNode	DataNode
YARN	ResourceManager		
	NodeManager	NodeManager	NodeManager
MapReduce	JobHistoryServer		

cdh 所有软件下载地址：

http://archive.cloudera.com/cdh5/cdh/5/
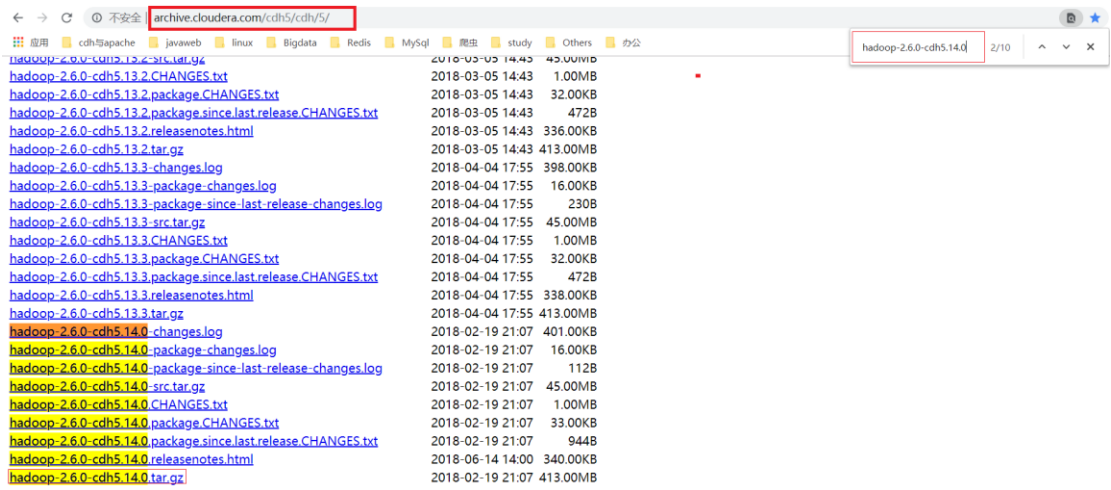
第一步：上传压缩包并解压

将我们重新编译之后支持 snappy 压缩的 hadoop 包上传到第一台服务器并解压

第一台机器执行以下命令

```
cd /export/softwares/
```
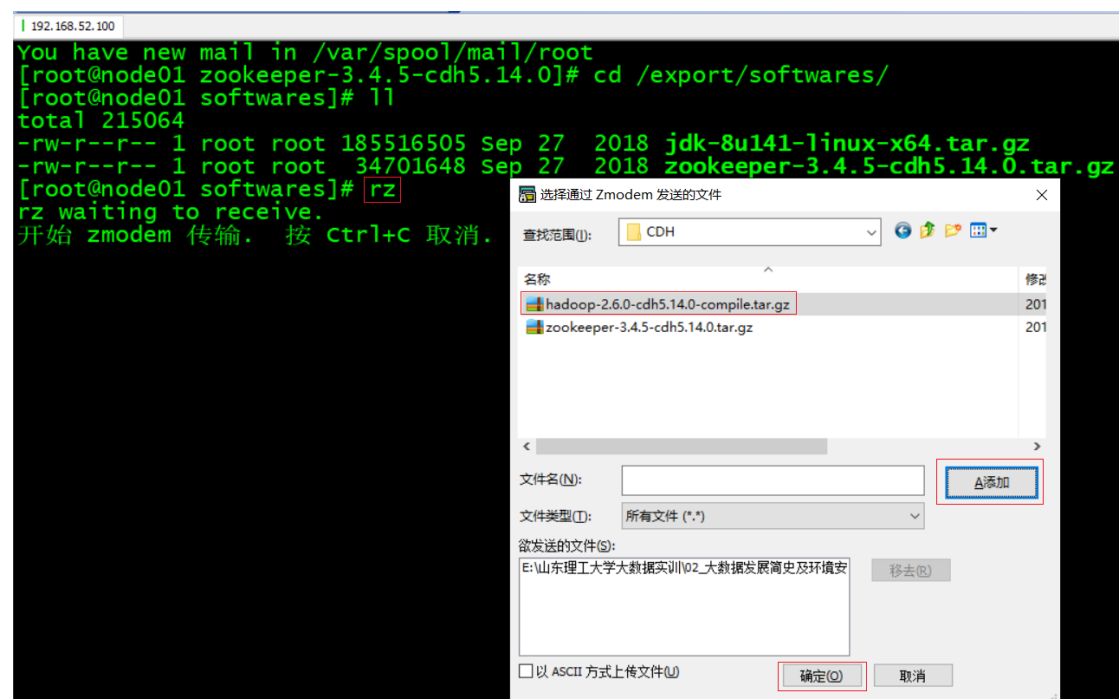


```
mv     hadoop-2.6.0-cdh5.14.0-compile.tar.gz      hadoop-2.6.0-
cdh5.14.0.tar.gz
```


第一步：上传压缩包并解压

将我们重新编译之后支持 snappy 压缩的 hadoop 包上传到第一台服务器并解压

第一台机器执行以下命令

```
tar -zxvf hadoop-2.6.0-cdh5.14.0.tar.gz -C ../servers/
```

```
[root@node01 softwares]# tar -zxvf hadoop-2.6.0-cdh5.14.0.tar.gz -C ../servers/
```

```
[root@node01 softwares]# cd /export/servers/
[root@node01 servers]# ll
total 12
drwxr-xr-x  9 root root 4096 May  8  2018 hadoop-2.6.0-cdh5.14.0
drwxr-xr-x  8 uucp  143 4096 Jul 12  2017 jdk1.8.0_141
drwxr-xr-x 18 root root 4096 Jul  7 07:19 zookeeper-3.4.5-cdh5.14.0
[root@node01 servers]#
```

第二步：查看 hadoop 支持的压缩方式以及本地库

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0

bin/hadoop checknative
```

```
| 192.168.52.100
[root@node01 hadoop-2.6.0-cdh5.14.0]# bin/hadoop checknative
19/07/07 07:38:32 INFO bzip2.Bzip2Factory: Successfully loaded & initialized native-bzip2 library system-native
19/07/07 07:38:32 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
Native library checking:
hadoop:  true /export/servers/hadoop-2.6.0-cdh5.14.0/lib/native/libhadoop.so.1.0.0
zlib:    true /lib64/libz.so.1
snappy:  true /usr/lib64/libsnappy.so.1
lz4:     true revision:10301
bzip2:   true /lib64/libbz2.so.1
openssl: false Cannot load libcrypto.so (libcrypto.so: cannot open shared object file: No such file or directory)!
[root@node01 hadoop-2.6.0-cdh5.14.0]#
```
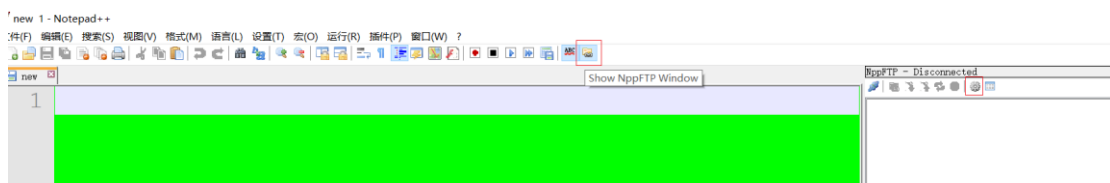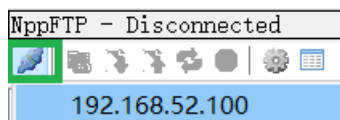
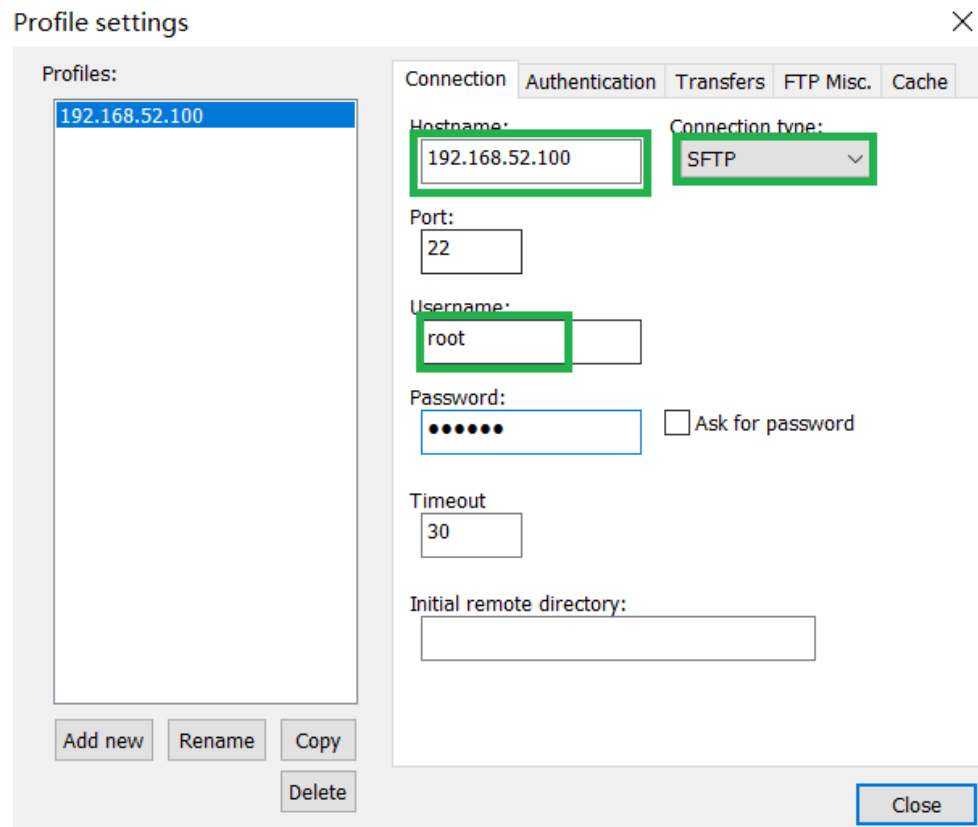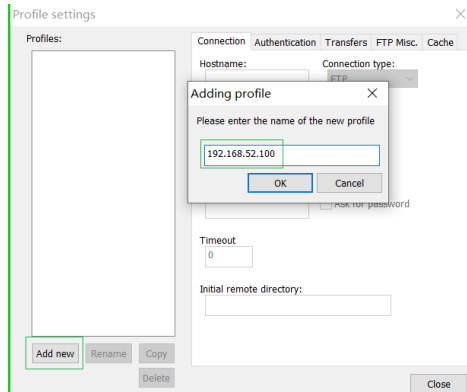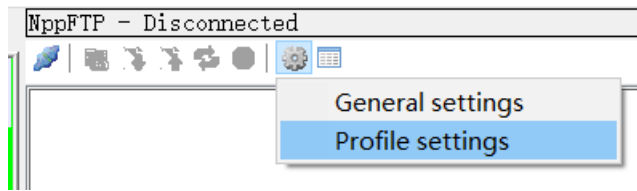如果出现 openssl（安全通信）为 false，那么所有机器在线安装 openssl 即可，执

行以下命令，虚拟机联网之后就可以在线进行安装了

```
yum -y install openssl-devel
```

```
[root@node01 hadoop-2.6.0-cdh5.14.0]# bin/hadoop checknative
19/07/07 07:43:31 INFO bzip2.Bzip2Factory: Successfully loaded & initialized native-bzip2 library system-native
19/07/07 07:43:31 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
Native library checking:
hadoop:  true /export/servers/hadoop-2.6.0-cdh5.14.0/lib/native/libhadoop.so.1.0.0
zlib:    true /lib64/libz.so.1
snappy:  true /usr/lib64/libsnappy.so.1
lz4:     true revision:10301
bzip2:   true /lib64/libbz2.so.1
openssl: true /usr/lib64/libcrypto.so
[root@node01 hadoop-2.6.0-cdh5.14.0]#
```

第三步：修改配置文件

Notepad++更改 linux 配置：

NppFTP - Disconnected

General settings
Profile settings

Profile settings

Profiles:

Connection | Authentication | Transfers | FTP Misc. | Cache

Hostname: Connection type:
                           FTP

Adding profile

Please enter the name of the new profile

192.168.52.100

OK   Cancel

Ask for password

Timeout
0

Initial remote directory:

Add new   Rename   Copy

Delete

Close

Profile settings

Profiles:

192.168.52.100

Connection | Authentication | Transfers | FTP Misc. | Cache

Hostname: Connection type:
192.168.52.100   SFTP

Port:
22

Username:
root

Password: Ask for password
••••••

Timeout
30

Initial remote directory:

Add new   Rename   Copy

Delete

Close

NppFTP - Disconnected

192.168.52.100

修改 core-site.xml

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0/etc/hadoop
vim core-site.xml
```

```
<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://node01:8020</value>
    </property>
    <property>
        <name>hadoop.tmp.dir</name>
```

```
        <value>/export/servers/hadoop-2.6.0-
cdh5.14.0/hadoopDatas/tempDatas</value>
    </property>


    <property>
        <name>io.file.buffer.size</name>
        <value>4096</value>
    </property>


    <property>
        <name>fs.trash.interval</name>
        <value>10080</value>
    </property>
</configuration>
```

修改 hdfs-site.xml

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0/etc/hadoop

vim hdfs-site.xml
```

```
<configuration>
    <property>
            <name>dfs.namenode.secondary.http-address</name>
            <value>node01:50090</value>
    </property>


    <property>
        <name>dfs.namenode.http-address</name>
        <value>node01:50070</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:///export/servers/hadoop-2.6.0-
cdh5.14.0/hadoopDatas/namenodeDatas</value>
    </property>
```

```xml
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:///export/servers/hadoop-2.6.0-
cdh5.14.0/hadoopDatas/datanodeDatas</value>
    </property>

    <property>
        <name>dfs.namenode.edits.dir</name>
        <value>file:///export/servers/hadoop-2.6.0-
cdh5.14.0/hadoopDatas/dfs/nn/edits</value>
    </property>
    <property>
        <name>dfs.namenode.checkpoint.dir</name>
        <value>file:///export/servers/hadoop-2.6.0-
cdh5.14.0/hadoopDatas/dfs/snn/name</value>
    </property>
    <property>
        <name>dfs.namenode.checkpoint.edits.dir</name>
        <value>file:///export/servers/hadoop-2.6.0-
cdh5.14.0/hadoopDatas/dfs/nn/snn/edits</value>
    </property>
    <property>
        <name>dfs.replication</name>
        <value>3</value>
    </property>
    <property>
        <name>dfs.permissions</name>
        <value>false</value>
    </property>
<property>
        <name>dfs.blocksize</name>
        <value>134217728</value>
    </property>
</configuration>
```

修改 hadoop-env.sh

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0/etc/hadoop
vim hadoop-env.sh
```

```
export JAVA_HOME=/export/servers/jdk1.8.0_141
```

修改 mapred-site.xml

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0/etc/hadoop
vim mapred-site.xml
```

```xml
<configuration>
    <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
    </property>

    <property>
        <name>mapreduce.job.ubertask.enable</name>
        <value>true</value>
    </property>

    <property>
        <name>mapreduce.jobhistory.address</name>
        <value>node01:10020</value>
    </property>

    <property>
        <name>mapreduce.jobhistory.webapp.address</name>
        <value>node01:19888</value>
    </property>
</configuration>
```

修改 yarn-site.xml

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0/etc/hadoop
vim yarn-site.xml
```

```xml
<configuration>
    <property>
        <name>yarn.resourcemanager.hostname</name>
        <value>node01</value>
    </property>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>

    <property>
        <name>yarn.log-aggregation-enable</name>
        <value>true</value>
    </property>
    <property>
        <name>yarn.log-aggregation.retain-seconds</name>
        <value>604800</value>
    </property>
</configuration>
```

修改 slaves 文件

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0/etc/hadoop
vim slaves
```

```
node01
node02
node03
```

第四步：创建文件存放目录

第一台机器执行以下命令

node01 机器上面创建以下目录

```
mkdir -p /export/servers/hadoop-2.6.0-cdh5.14.0/hadoopDatas/tempDatas

mkdir -p /export/servers/hadoop-2.6.0-cdh5.14.0/hadoopDatas/namenodeDatas

mkdir -p /export/servers/hadoop-2.6.0-cdh5.14.0/hadoopDatas/datanodeDatas

mkdir -p /export/servers/hadoop-2.6.0-cdh5.14.0/hadoopDatas/dfs/nn/edits

mkdir -p /export/servers/hadoop-2.6.0-cdh5.14.0/hadoopDatas/dfs/snn/name

mkdir -p /export/servers/hadoop-2.6.0-cdh5.14.0/hadoopDatas/dfs/nn/snn/edits
```

第五步：安装包的分发

第一台机器执行以下命令

```
cd /export/servers/
scp -r hadoop-2.6.0-cdh5.14.0/ node02:$PWD
scp -r hadoop-2.6.0-cdh5.14.0/ node03:$PWD
```

第六步：配置 hadoop 的环境变量

三台机器都要进行配置 hadoop 的环境变量

三台机器执行以下命令

```
vim  /etc/profile
```

```
export HADOOP_HOME=/export/servers/hadoop-2.6.0-cdh5.14.0
export PATH=:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
```

配置完成之后生效

```
source /etc/profile
```

第七步：集群启动

要启动 Hadoop 集群，需要启动 HDFS 和 YARN 两个集群。

注意：首次启动 HDFS 时，必须对其进行格式化操作。本质上是一些清理和准备工作，因为此时的 HDFS 在物理上还是不存在的。

bin/hdfs namenode   -format

脚本一键启动

如果配置了 etc/hadoop/slaves 和 ssh 免密登录，则可以使用程序脚本启动所有 Hadoop 两个集群的相关进程，在主节点所设定的机器上执行。

启动集群

node01 节点上执行以下命令

第一台机器执行以下命令

```
cd /export/servers/hadoop-2.6.0-cdh5.14.0/

sbin/start-dfs.sh

sbin/start-yarn.sh

sbin/mr-jobhistory-daemon.sh start historyserver
```

停止集群：没事儿不要去停止集群

```
sbin/stop-dfs.sh

sbin/stop-yarn.sh

sbin/mr-jobhistory-daemon.sh stop historyserver
```

第八步：浏览器查看启动页面

hdfs 集群访问地址

192.168.52.100:50070/dfshealth.html#tab-overview

yarn 集群访问地址

192.168.52.100:8088/cluster

jobhistory 访问地址：

192.168.52.100:19888/jobhistory