CentOS

已安装

相关配置

用户

添加普通用户并设置密码

```
[root@localhost dhw]# sudo passwd dhw
Changing password for user dhw.
New password:
BAD PASSWORD: The password is shorter than 8 characters
Retype new password:
Sorry, passwords do not match.
New password:
BAD PASSWORD: The password is shorter than 8 characters
Retype new password:
passwd: all authentication tokens updated successfully.
[root@localhost dhw]#
```

安装 JDK

下载安装包
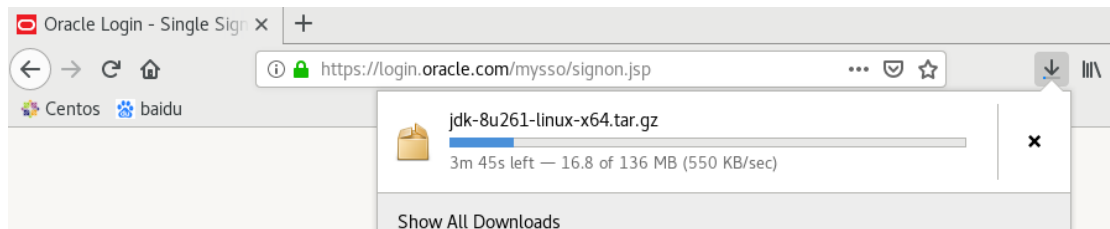

```
[dhw@localhost Downloads]$ ll
total 475676
-rw-rw-r--. 1 dhw dhw 218720521 Oct 13 19:53 hadoop-2.7.7.tar.gz
-rw-rw-r--. 1 dhw dhw 143111803 Oct 13 20:20 jdk-8u261-linux-x64.tar.gz
[dhw@localhost Downloads]$
```

删除原有的

```
[dhw@localhost Downloads]$ rpm -qa | grep java
java-1.7.0-openjdk-headless-1.7.0.251-2.6.21.1.el7.x86_64
java-1.8.0-openjdk-1.8.0.262.b10-0.el7_8.x86_64
javapackages-tools-3.4.1-11.el7.noarch
java-1.7.0-openjdk-1.7.0.251-2.6.21.1.el7.x86_64
python-javapackages-3.4.1-11.el7.noarch
tzdata-java-2020a-1.el7.noarch
java-1.8.0-openjdk-headless-1.8.0.262.b10-0.el7_8.x86_64
[dhw@localhost Downloads]$ rpm -e --nodeps java-1.7.0-openjdk
error: can't create transaction lock on /var/lib/rpm/.rpm.lock (Permission denie
d)
[dhw@localhost Downloads]$ sudo rpm -e --nodeps java-1.7.0-openjdk
[sudo] password for dhw:
```

安装解压

```
dhw@localhost:/usr/local/jdk1.8
File  Edit  View  Search  Terminal  Help
[dhw@localhost Downloads]$ sudo tar -zxf jdk-8u261-linux-x64.tar.gz -C /usr/loca
l/
[dhw@localhost Downloads]$ cd /usr/local/
[dhw@localhost local]$ mv jdk1.8.0_261/ jdk1.8/
mv: cannot move 'jdk1.8.0_261/' to 'jdk1.8/': Permission denied
[dhw@localhost local]$ sudo mv jdk1.8.0_261/ jdk1.8/
[dhw@localhost local]$ cd jdk1.8/
[dhw@localhost jdk1.8]$ pwd
/usr/local/jdk1.8
[dhw@localhost jdk1.8]$
```

配置环境变量

```
# java
export JAVA_HOME=/usr/local/jdk1.8
export PATH=$PATH:$JAVA_HOME/bin

:wq
```

若出现下述问题改为 root 用户权限即可

```
"/etc/profile"
"/etc/profile" E212: Can't open file for writing
Press ENTER or type command to continue
```

验证:


```
dhw@localhost:/usr/local/jdk1.8
File  Edit  View  Search  Terminal  Help
[dhw@localhost jdk1.8]$ vim /etc/profile
[dhw@localhost jdk1.8]$ sudo vim /etc/profile
[dhw@localhost jdk1.8]$ java -version
bash: java: command not found...
[dhw@localhost jdk1.8]$ java -version
bash: java: command not found...
[dhw@localhost jdk1.8]$ sudo vim /etc/profile
[dhw@localhost jdk1.8]$ source /etc/profile
[dhw@localhost jdk1.8]$ java -version
java version "1.8.0_261"
Java(TM) SE Runtime Environment (build 1.8.0_261-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.261-b12, mixed mode)
[dhw@localhost jdk1.8]$
```

安装 Hadoop

下载安装包
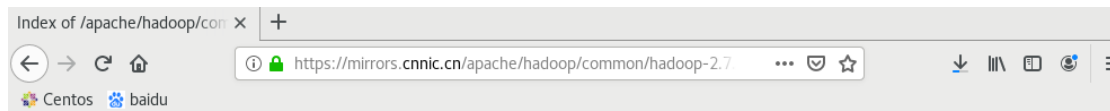
Index of /apache/hadoop/common/hadoop-2.7.7

Name	Last modified	Size	Description
Parent Directory		-	
hadoop-2.7.7-src.tar.gz	2018-07-20 09:11	18M	
hadoop-2.7.7.tar.gz	2018-07-20 09:11	209M	

---

dhw@localhost:~/Downloads   _ □ ✕

File Edit View Search Terminal Help

```
[dhw@localhost ~]$ pwd
/home/dhw
[dhw@localhost ~]$ ll
total 33536
-rw-------. 1 dhw dhw 34340864 Oct  2 14:12 core.2431
drwxr-xr-x. 2 dhw dhw        6 Oct  2 13:46 Desktop
drwxr-xr-x. 2 dhw dhw        6 Oct  2 13:46 Documents
drwxr-xr-x. 2 dhw dhw       65 Oct 13 19:49 Downloads
drwxr-xr-x. 2 dhw dhw        6 Oct  2 13:46 Music
drwxr-xr-x. 2 dhw dhw        6 Oct  2 13:46 Pictures
drwxr-xr-x. 2 dhw dhw        6 Oct  2 13:46 Public
drwxr-xr-x. 2 dhw dhw        6 Oct  2 13:46 Templates
drwxr-xr-x. 2 dhw dhw        6 Oct  2 13:46 Videos
[dhw@localhost ~]$ cd Downloads/
[dhw@localhost Downloads]$ ll
total 311232
-rw-rw-r--. 1 dhw dhw 218720521 Oct 13 19:53 hadoop-2.7.7.tar.gz
[dhw@localhost Downloads]$
```

解压并改名

```
[dhw@localhost Downloads]$ sudo tar -zxf hadoop-2.7.7.tar.gz -C /usr/local
[sudo] password for dhw:
[dhw@localhost Downloads]$ cd /usr/local
[dhw@localhost local]$ ll
total 0
drwxr-xr-x. 2 root root   6 Apr 11  2018 bin
drwxr-xr-x. 2 root root   6 Apr 11  2018 etc
drwxr-xr-x. 2 root root   6 Apr 11  2018 games
drwxr-xr-x. 9 dhw  ftp  149 Jul 19  2018 hadoop-2.7.7
drwxr-xr-x. 2 root root   6 Apr 11  2018 include
drwxr-xr-x. 2 root root   6 Apr 11  2018 lib
drwxr-xr-x. 2 root root   6 Apr 11  2018 lib64
drwxr-xr-x. 2 root root   6 Apr 11  2018 libexec
drwxr-xr-x. 2 root root   6 Apr 11  2018 sbin
drwxr-xr-x. 5 root root  49 Oct  2 11:36 share
drwxr-xr-x. 2 root root   6 Apr 11  2018 src
[dhw@localhost local]$ sudo mv hadoop-2.7.7/ hadoop
```

修改文件权限

```
drwxr-xr-x. 5 root root  49 Oct  2 11:36 share
drwxr-xr-x. 2 root root   6 Apr 11  2018 src
[dhw@localhost local]$ sudo chown -R hadoop ./hadoop/
chown: invalid user: 'hadoop'
[dhw@localhost local]$ sudo chown -R dhw ./hadoop/
[dhw@localhost local]$

[dhw@localhost local]$ cd hadoop/
[dhw@localhost hadoop]$ ./bin/hadoop version
Error: JAVA_HOME is not set and could not be found.
[dhw@localhost hadoop]$
```

安装 JDK（写到上面了），之后验证安装

```
                    dhw@localhost:/usr/local/hadoop              _   □   ×

 File  Edit  View  Search  Terminal  Help

[dhw@localhost hadoop]$ ./bin/hadoop version
Hadoop 2.7.7
Subversion Unknown -r c1aad84bd27cd79c3d1a7dd58202a8c3ee1ed3ac
Compiled by stevel on 2018-07-18T22:47Z
Compiled with protoc 2.5.0
From source with checksum 792e15d20b12c74bd6f19a1fb886490
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-2
.7.7.jar
[dhw@localhost hadoop]$ ▮
```

伪分布模式配置

```
                 dhw@localhost:/usr/local/hadoop/etc/hadoop          _   □   ×

 File  Edit  View  Search  Terminal  Help

[dhw@localhost hadoop]$ cd  /usr/local/hadoop/etc/hadoop/
[dhw@localhost hadoop]$ vim core-site.xml
[dhw@localhost hadoop]$ vim hdfs-site.xml
[dhw@localhost hadoop]$ ▮
```

配置文件 core.site.xml：

```
<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/usr/local/hadoop/temp</value>
    <description>Abase for other temporary directories.</description>
  </property>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
▮
-- INSERT --                                        30.1        Bot
```

其中第一个属性用于保存临时文件；第二个属性用于指定 HDFS 的访问地址，后面是端口号。

配置文件 hdfs-site.xml：

```
<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>
:wq
```

其中第一个属性用于指定副本的数量，但是这里是伪分布式模型，只有一个节点，因此只有

一个副本。第二个属性用于设定名称节点的元数据的保存目录。最后一个属性用于设定数据节点的数据保存目录。

（若要换回默认的单机模式，只需要删除 core-site.xml 中的配置项即可）

名称节点格式化：（最后提示信息为 0 表示成功）

```
                          dhw@localhost:/usr/local/hadoop              _  □  ✕

 File  Edit  View  Search  Terminal  Help

[dhw@localhost hadoop]$ pwd
/usr/local/hadoop/etc/hadoop
[dhw@localhost hadoop]$ cd /usr/local/hadoop/
[dhw@localhost hadoop]$ ./bin/hdfs namenode -format

20/10/13 20:54:10 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
20/10/13 20:54:10 INFO util.ExitUtil: Exiting with status 0
20/10/13 20:54:10 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at localhost/127.0.0.1
************************************************************/
[dhw@localhost hadoop]$
```

启动 Hadoop：

```
                          dhw@localhost:/usr/local/hadoop              _  □  ✕

 File  Edit  View  Search  Terminal  Help

[dhw@localhost hadoop]$ pwd
/usr/local/hadoop
[dhw@localhost hadoop]$ ./sbin/start-dfs.sh
```

若出现如下错误：（更改 hadoop 环境配置文件，如下图）

```
dhw@localhost's password:
localhost: Error: JAVA_HOME is not set and could not be found.
dhw@localhost's password:
localhost: Error: JAVA_HOME is not set and could not be found.
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is SHA256:+6kban/NZF1k0TGhejELmA43S5x/3Us1NWyjcnSU2l0.
ECDSA key fingerprint is MD5:b1:52:25:1f:60:43:24:24:13:d5:ce:55:66:fa:e1:2d.
Are you sure you want to continue connecting (yes/no)? yes
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts
.
dhw@0.0.0.0's password:
0.0.0.0: Error: JAVA_HOME is not set and could not be found.
[dhw@localhost hadoop]$
```

```
                       dhw@localhost:/usr/local/hadoop/etc/hadoop        _  □  ✕

 File  Edit  View  Search  Terminal  Help

[dhw@localhost hadoop]$ pwd
/usr/local/hadoop/etc/hadoop
[dhw@localhost hadoop]$ vim hadoop-env.sh


# The java implementation to use.

# export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=/usr/local/jdk1.8

# The jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol.  Jsvc is not required if SASL is configured for authentication of
-- INSERT --                                            27,35          10%
```
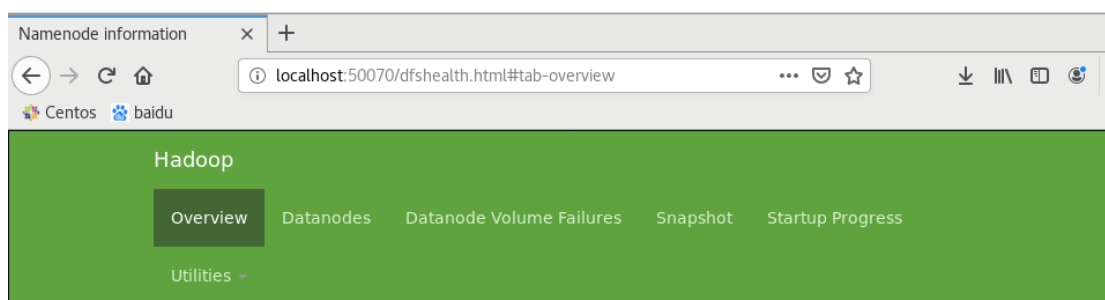
重新启动 Hadoop:（看到红框则表示成功）

启动成功后，在 Linux 系统打开浏览器可看到如下页面:（http://localhost:50070）

Overview 'localhost:9000' (active)

Started:	Tue Oct 13 21:06:32 CST 2020
Version:	2.7.7, rc1aad84bd27cd79c3d1a7dd58202a8c3ee1ed3ac
Compiled:	2018-07-18T22:47Z by stevel from branch-2.7.7
Cluster ID:	CID-e8d32b42-5227-4944-b1d1-54d56a1fb9f6
Block Pool ID:	BP-1774376895-127.0.0.1-1602593650117

Summary

在 HDFS 中创建用户目录:

将本地文件系统的/usr/local/Hadoop/etc/hadoop 目录中的所有 xml 文件复制到 HDFS 中的/user/Hadoop/input 目录下:（注意创建目录时加-p）

```
                    dhw@localhost:/usr/local/hadoop              _  □  ✕

 File  Edit  View  Search  Terminal  Help

[dhw@localhost hadoop]$ cd /usr/local/hadoop/
[dhw@localhost hadoop]$ ./bin/hdfs dfs -mkdir input
mkdir: `input': No such file or directory
[dhw@localhost hadoop]$ cd etc/hadoop/
[dhw@localhost hadoop]$ cd /usr/local/hadoop/
[dhw@localhost hadoop]$ ./bin/hdfs dfs -mkdir -p input
[dhw@localhost hadoop]$ ./bin/hdfs dfs -put ./etc/hadoop/*.xml input
[dhw@localhost hadoop]$ ./bin/hdfs dfs -ls input
Found 8 items
-rw-r--r--   1 dhw supergroup        4436 2020-10-13 21:20 input/capacity-schedul
er.xml
-rw-r--r--   1 dhw supergroup        1049 2020-10-13 21:20 input/core-site.xml
-rw-r--r--   1 dhw supergroup        9683 2020-10-13 21:20 input/hadoop-policy.xm
l
-rw-r--r--   1 dhw supergroup        1097 2020-10-13 21:20 input/hdfs-site.xml
-rw-r--r--   1 dhw supergroup         620 2020-10-13 21:20 input/httpfs-site.xml
-rw-r--r--   1 dhw supergroup        3518 2020-10-13 21:20 input/kms-acls.xml
-rw-r--r--   1 dhw supergroup        5540 2020-10-13 21:20 input/kms-site.xml
-rw-r--r--   1 dhw supergroup         690 2020-10-13 21:20 input/yarn-site.xml
[dhw@localhost hadoop]$ █
```

复制完成可查看文件列表（如上图），运行 Hadoop 自带的 grep 程序（并查看输出文件夹的内容）：

```
                    dhw@localhost:/usr/local/hadoop              _  □  ✕

 File  Edit  View  Search  Terminal  Help

[dhw@localhost hadoop]$ ./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapredu
ce-examples-2.7.7.jar grep input output 'dfs[a-z.]+'█


                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=219
        File Output Format Counters
                Bytes Written=77
[dhw@localhost hadoop]$ ./bin/hdfs dfs -cat output/*
1       dfsadmin
1       dfs.replication
1       dfs.namenode.name.dir
1       dfs.datanode.data.dir
[dhw@localhost hadoop]$ █
```

若要再次运行则需要删除 output 目录：

```
        dfs.datanode.data.dir
[dhw@localhost hadoop]$ ./bin/hdfs dfs -rm -r output
20/10/13 21:26:18 INFO fs.TrashPolicyDefault: Namenode trash configuration: Dele
tion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted output
[dhw@localhost hadoop]$
```

关闭 Hadoop：

```
                dhw@localhost:/usr/local/hadoop              _  □  ×

File  Edit  View  Search  Terminal  Help
[dhw@localhost hadoop]$ cd /usr/local/hadoop/
[dhw@localhost hadoop]$ ./sbin/st
start-all.cmd          start-yarn.cmd          stop-dfs.sh
start-all.sh           start-yarn.sh           stop-secure-dns.sh
start-balancer.sh      stop-all.cmd            stop-yarn.cmd
start-dfs.cmd          stop-all.sh             stop-yarn.sh
start-dfs.sh           stop-balancer.sh
start-secure-dns.sh    stop-dfs.cmd
[dhw@localhost hadoop]$ ./sbin/stop-dfs.sh
Stopping namenodes on [localhost]
dhw@localhost's password:
localhost: stopping namenode
dhw@localhost's password:
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
dhw@0.0.0.0's password:
0.0.0.0: stopping secondarynamenode
[dhw@localhost hadoop]$
```

下次启动时，无须进行名称节点的初始化。只需要直接运行 start-dfs.sh 即可

```
[dhw@localhost hadoop]$ start-dfs.sh
bash: start-dfs.sh: command not found...
[dhw@localhost hadoop]$ ./sbin/start-dfs.sh
Starting namenodes on [localhost]
dhw@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-dhw-namen
ode-localhost.localdomain.out
dhw@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-dhw-datan
ode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
dhw@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-dh
w-secondarynamenode-localhost.localdomain.out
[dhw@localhost hadoop]$ jps
20152 DataNode
20488 Jps
19993 NameNode
20350 SecondaryNameNode
[dhw@localhost hadoop]$ pwd
/usr/local/hadoop
[dhw@localhost hadoop]$
```

（注意：hadoop 安装目录里的 etc/hadoop 是定义的 hadoop 用户目录）