**第5章  Hadoop配置操作**

**1.实验步骤**

1.1  **安装配置JDK 1.8**

**在模板中，我们已经将相应的JDK安装包jdk-8u60-linux-x64.tar.gz放到了/usr/software/目录下，同学们不需要再次下载，可以直接使用。**

我们可以在Oracle JDK的官网下载相版本的JDK，官网地址为：<http://www.oracle.com/technetwork/java/javase/downloads/index.html>（如果有需要，同学们需要在新的标签页打开）

步骤1.   **创建工作路径**

首先需要在终端中输入下列命令，在/usr目录下建立cx工作路径：

mkdir /usr/cx

步骤2.   **解压安装包**

我们可以在/usr/software/目录下找到jdk-8u60-linux-x64.tar.gz安装包，通过下列命令将其解压到/usr/cx/目录下，执行如下命令：

tar -zxvf /usr/software/jdk-8u60-linux-x64.tar.gz -C /usr/cx

步骤3.   **配置环境变量**

1.   通过下列命令使用vi编辑器打

开 ~/.bashrc文件：

vi ~/.bashrc

打开的~/.bashrc文件内容如下所示：

# Source global definitions

if [ -f /etc/bashrc ]; then

        . /etc/bashrc

fi

(----------------注：需要在此处增加内容-------------------)

2.   在文件中写入下列内容（默认进入vi编辑器后是无法进行操作的，我们需要输入小写字母i进入编辑模式，然后通过键盘上下左右键，定位到需要更改的位置更改内容即可。）：

export JAVA\_HOME=/usr/cx/jdk1.8.0\_60

export PATH=$JAVA\_HOME/bin:$PATH

export CLASSPATH=.:$JAVA\_HOME/jre/lib/rt.jar:$JAVA\_HOME/jre/lib/tools.jar

编辑完成后保存文件并退出vi编辑器（先按Esc键，然后再按冒号：，最后输入小写的wq!,按回车键即可进行文件内容保存并退出）。

步骤4.   **更新环境变量**

执行如下命令，更新环境变量：

source  ~/.bashrc

步骤5.   **验证JDK是否配置成功**

通过下列命令验证JDK是否安装并配置成功：

java -version

1.2  **主机名配置**

步骤1.   **编辑主机名**

1.   通过下列命令使用vi编辑器编辑主机名配置文件：

vi /etc/sysconfig/network

打开后的文件内容如下所示：

NETWORKING=yes

HOSTNAME=CentOS6.5                                   (注：需要将此行内容修改为实际的主机名master)

2.   在文件中进行内容更改，将HOSTNAME字段内容配置成master。：

HOSTNAME=master

编辑完成后保存文件并退出vi编辑器。

3.     更改后的内容会在下次系统重启的时候生效，通过下列命令重新启动系统：

reboot

步骤2.   **IP地址与主机名映射文件配置**

1.   通过下列命令使用vi编辑器编辑hosts文件：

vi /etc/hosts

2.   增加主机名与IP地址的映射关系：

127.0.0.1 master

更改后的文件内容如下所示：

127.0.0.1   master localhost localhost.localdomain localhost4 localhost4.localdomain4

::1         localhost localhost.localdomain localhost6 localhost6.localdomain6

编辑完成后保存文件并退出vi编辑器。

3.   通过下列命令检测主机名与IP映射是否配置成功：

ping master -c 4

1.3  **SElinux安全配置**

步骤1.   **关闭SElinux**

CentOS默认启用了SElinux，在网络服务方面权限要求比较严格，因此我们需要将SElinux关闭。

1.   通过下列命令使用vi编辑器打开SElinux配置文件：

vi /etc/selinux/config

打开后的文件内容如下所示：

# This file controls the state of SELinux on the system.

# SELINUX= can take one of these three values:

#     enforcing - SELinux security policy is enforced.

#     permissive - SELinux prints warnings instead of enforcing.

#     disabled - No SELinux policy is loaded.

SELINUX=permissive                                                      (注：需要更改此行内容)

# SELINUXTYPE= can take one of these two values:

#     targeted - Targeted processes are protected,

#     mls - Multi Level Security protection.

SELINUXTYPE=targeted

2.   在文件中进行内容更改，将SELINUX=permissive改写成SELINUX=disabled：

更改后的文件内容如下所示：

# This file controls the state of SELinux on the system.

# SELINUX= can take one of these three values:

#     enforcing - SELinux security policy is enforced.

#     permissive - SELinux prints warnings instead of enforcing.

#     disabled - No SELinux policy is loaded.

SELINUX=disabled

# SELINUXTYPE= can take one of these two values:

#     targeted - Targeted processes are protected,

#     mls - Multi Level Security protection.

SELINUXTYPE=targeted

编辑完成后保存文件并退出vi编辑器。

步骤2.   **SElinux配置强制生效**

更改后的内容会在下次系统重启的时候生效，为了操作便捷性，我们可以使用下列命令使更改即时生效，而不需要重新启动系统：

setenforce 0

1.4  **配置SSH免密码登录**

步骤1.   **生成秘钥**

1.   输入下面的命令，生成本机密钥文件：

ssh-keygen -t dsa

当出现提示的时候，我们直接按回车键即可，默认会将秘钥文件生成到~/.ssh/目录下（由于我们实验所使用的登录用户为root，因此~/.ssh/等同于/root/.ssh/）：

[root@master ~]# ssh-keygen -t dsa                          （注：通过dsa加密方式对秘钥进行加密）

Generating public/private dsa key pair.

Enter file in which to save the key (/root/.ssh/id\_dsa): （注：按回车即可）

Enter passphrase (empty for no passphrase):               （注：按回车即可）

Enter same passphrase again:                                （注：按回车即可）

Your identification has been saved in /root/.ssh/id\_dsa. （注：生成的私钥文件位置）

Your public key has been saved in /root/.ssh/id\_dsa.pub. （注：生成的公钥文件位置）

The key fingerprint is:

21:68:a9:03:1e:18:40:79:ed:d2:68:d2:63:22:fe:25 root@master

The key's randomart image is:

+--[ DSA 1024]----+

|=.. .            |

|.o . +           |

|o.o B . .        |

|o+.X o . .       |

|o.B o   S        |

| . E .           |

|  . o            |

|   .             |

|                 |

+-----------------+

[root@master ~]#

2.   通过下列命令查看~/.ssh目录下的文件：

ls ~/.ssh

步骤2.   **秘钥分发**

把当前节点的公钥文件id\_dsa.pub内容输出追加到任意节点的~/.ssh/authorized\_keys文件的末尾，则在被添加的节点上便可以免密码登录到当前的节点（由于我们是单节点部署，因此直接追加到当前节点的~/.ssh/authorized\_keys文件中即可），输入命令如下：

Cat ~/.ssh/id\_dsa.pub >> ~/.ssh/authorized\_keys

步骤3.   **验证免密码登录是否配置成功**

使用下列命令通过设置的主机名进行连接，可以验证免密码登录是否配置成功：

ssh master

第一次登录的时候，会询问是否继续连接，输入yes即可进入，当第二次进行连接的时候则不会再出现此提示（***如果无法正常连接，请检查是否完成秘钥分发步骤操作***）：

连接成功后，我们需要通过下列命令退出连接：

exit

1.5  **安装配置Hadoop**

**在模板中，我们已经将相应的Hadoop安装包hadoop-2.7.1.tar.gz放到/usr/software/目录下，同学们不需要再次下载，可以直接使用。**

步骤1.   **解压安装文件**

通过下列命令解压Hadoop安装文件，将文件解压到/usr/cx目录下：

tar -zxvf /usr/software/hadoop-2.7.1.tar.gz -C /usr/cx

步骤2.   **配置Hadoop环境变量**

1.   通过下列命令使用vi编辑器编辑~/.bashrc文件：

vi  ~/.bashrc

打开后的文件内容如下所示：

# .bashrc

# User specific aliases and functions

alias rm='rm -i'

alias cp='cp -i'

alias mv='mv -i'

# Source global definitions

if [ -f /etc/bashrc ]; then

        . /etc/bashrc

fi

export JAVA\_HOME=/usr/cx/jdk1.8.0\_60

export PATH=$JAVA\_HOME/bin:$PATH

export CLASSPATH=.:$JAVA\_HOME/jre/lib/rt.jar:$JAVA\_HOME/jre/lib/tools.jar

(----------------在此处增加内容-------------------)

2.   在~/.bashrc文件中增加以下内容：

export HADOOP\_HOME=/usr/cx/hadoop-2.7.1

export PATH=$PATH:$HADOOP\_HOME/bin:$PATH

export PATH=$PATH:$HADOOP\_HOME/sbin:$PATH

编辑完成后保存文件并退出vi编辑器。

3.   执行如下命令，更新环境变量：

source ~/.bashrc

4.   通过下列命令验证Hadoop环境变量是否配置成功：

hadoop

步骤3.   **编辑Hadoop配置文件：**

1.   使用vi命令打开hadoop-env.sh配置文件进行编辑：

vi /usr/cx/hadoop-2.7.1/etc/hadoop/hadoop-env.sh

打开后的文件内容如下所示：

(-------------------省略------------------------)

# Set Hadoop-specific environment variables here.

# The only required environment variable is JAVA\_HOME.  All others are

# optional.  When running a distributed configuration it is best to

# set JAVA\_HOME in this file, so that it is correctly defined on

# remote nodes.

# The java implementation to use.

export JAVA\_HOME=${JAVA\_HOME}             (注：需要对此行内容进行更改，为Hadoop绑定Java运行环境)

# 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

# data transfer protocol using non-privileged ports.

#export JSVC\_HOME=${JSVC\_HOME}

export HADOOP\_CONF\_DIR=${HADOOP\_CONF\_DIR:-"/etc/hadoop"}

(-------------------省略------------------------)

在文件中进行下列内容更改，将JAVA\_HOME对应的值改成实际的JDK安装目录：

export JAVA\_HOME=/usr/cx/jdk1.8.0\_60

编辑完成后保存文件并退出vi编辑器。

2.使用vi命令打开core-site.xml配置文件进行编辑：

vi /usr/cx/hadoop-2.7.1/etc/hadoop/core-site.xml

打开后的文件内容如下所示：

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!--

  Licensed under the Apache License, Version 2.0 (the "License");

  you may not use this file except in compliance with the License.

  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software

  distributed under the License is distributed on an "AS IS" BASIS,

  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

  See the License for the specific language governing permissions and

  limitations under the License. See accompanying LICENSE file.

-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

(注：需要在此处进行相关内容配置)

</configuration>

在文件中<configuration>和</configuration>之间增加下列内容：

/\*设置默认的HDFS访问路径\*/

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

</property>

/\*缓冲区大小：io.file.buffer.size默认是4KB\*/

<property>

<name>io.file.buffer.size</name>

<value>131072</value>

</property>

/\*临时文件夹路径设置\*/

<property>

<name>hadoop.tmp.dir</name>

<value>file:/usr/tmp</value>

</property>

/\*设置使用hduser用户可以代理所有主机用户进行任务提交\*/

<property>

<name>hadoop.proxyuser.hduser.hosts</name>

<value>\*</value>

</property>

/\*设置使用hduser用户可以代理所有组用户进行任务提交\*/

<property>

<name>hadoop.proxyuser.hduser.groups</name>

<value>\*</value>

</property>

编辑完成后保存文件并退出vi编辑器。

3.   使用vi命令打开yarn-site.xml文件进行配置：

vi /usr/cx/hadoop-2.7.1/etc/hadoop/yarn-site.xml

打开后的文件内容如下所示：

<?xml version="1.0"?>

<!--

  Licensed under the Apache License, Version 2.0 (the "License");

  you may not use this file except in compliance with the License.

  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software

  distributed under the License is distributed on an "AS IS" BASIS,

  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

  See the License for the specific language governing permissions and

  limitations under the License. See accompanying LICENSE file.

-->

<configuration>

(注：需要在此处进行相关内容配置)

<!-- Site specific YARN configuration properties -->

</configuration>

在文件中<configuration>和</configuration>之间增加下列内容：

/\*设置NodeManager上运行的附属服务，需配置成mapreduce\_shuffle才可运行MapReduce程序\*/

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

/\*设置客户端与ResourceManager的通信地址\*/

<property>

<name>yarn.resourcemanager.address</name>

<value>master:8032</value>

</property>

/\*设置ApplicationMaster调度器与ResourceManager的通信地址\*/

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>master:8030</value>

</property>

/\*设置NodeManager与ResourceManager的通信地址\*/

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>master:8031</value>

</property>

/\*设置管理员与ResourceManager的通信地址\*/

<property>

<name>yarn.resourcemanager.admin.address</name>

<value>master:8033</value>

</property>

/\* ResourceManager的Web地址，监控资源调度\*/

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>master:8088</value>

</property>

编辑完成后保存文件并退出vi编辑器。

4.   使用下列命令复制mapred-site.xml.template文件并重命名为mapred-site.xml：

cp /usr/cx/hadoop-2.7.1/etc/hadoop/mapred-site.xml.template /usr/cx/hadoop-2.7.1/etc/hadoop/mapred-site.xml

5.   使用vi命令打开mapred-site.xml文件进行配置：

vi /usr/cx/hadoop-2.7.1/etc/hadoop/mapred-site.xml

打开后的文件内容如下所示：

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!--

  Licensed under the Apache License, Version 2.0 (the "License");

  you may not use this file except in compliance with the License.

  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software

  distributed under the License is distributed on an "AS IS" BASIS,

  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

  See the License for the specific language governing permissions and

  limitations under the License. See accompanying LICENSE file.

-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

(注：需要在此处进行相关内容配置)

</configuration>

在文件中<configuration>和</configuration>之间增加下列内容：

/\*Hadoop对MapReduce运行框架一共提供了3种实现，在mapred-site.xml中通过"mapreduce.framework.name"这个属性来设置为"classic"、"yarn"或者"local"\*/

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

/\*MapReduce JobHistory Server地址\*/

<property>

<name>mapreduce.jobhistory.address</name>

<value>master:10020</value>

</property>

/\*MapReduce JobHistory Server Web UI访问地址\*/

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>master:19888</value>

</property>

编辑完成后保存文件并退出vi编辑器。

6.   执行以下命令创建Hadoop的数据存储目录namenode和datanode：

mkdir -p /hdfs/namenode

mkdir -p /hdfs/datanode

7.   使用vi命令打开hdfs-site.xml文件进行配置：

vi /usr/cx/hadoop-2.7.1/etc/hadoop/hdfs-site.xml

打开后的文件内容如下所示：

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!--

  Licensed under the Apache License, Version 2.0 (the "License");

  you may not use this file except in compliance with the License.

  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software

  distributed under the License is distributed on an "AS IS" BASIS,

  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

  See the License for the specific language governing permissions and

  limitations under the License. See accompanying LICENSE file.

-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

(注：需要在此处进行相关内容配置)

</configuration>

在文件中<configuration>和</configuration>之间增加下列内容：

/\*配置SecondaryNameNode地址\*/

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>master:9001</value>

</property>

/\*配置NameNode的数据存储目录，需要与上文创建的目录相对应\*/

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/hdfs/namenode</value>

</property>

/\*配置DataNode的数据存储目录，需要与上文创建的目录相对应\*/

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/hdfs/datanode</value>

</property>

/\*配置数据块副本数\*/

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

/\*将dfs.webhdfs.enabled属性设置为true，否则就不能使用webhdfs的LISTSTATUS、LIST FILESTATUS等需要列出文件、文件夹状态的命令，因为这些信息都是由namenode保存的\*/

<property>

<name>dfs.webhdfs.enabled</name>

<value>true</value>

</property>

编辑完成后保存文件并退出vi编辑器。

8.   使用vi命令打开slaves文件进行配置（要与我们前文设置的主机名相互一致，否则将会引起Hadoop相关进程无法正确启动）：

vi /usr/cx/hadoop-2.7.1/etc/hadoop/slaves

打开后的文件内容如下所示：

localhost（注：需要对此内容进行更改，配置为Slave节点的实际主机名）

将文件中的内容更改为下列内容：

master

编辑完成后保存文件并退出vi编辑器。

步骤4.   **格式化HDFS**

通过下列命令格式化HDFS文件系统（***如果格式化失败，会有相关的错误日志输出，根据输出内容进行更改即可***）：

hadoop namenode -format

命令运行后的部分显示内容如下所示：

…successfully formatted.   （注：当出现successfully formatted字样时，则说明格式化成功）

1.6  **Hadoop运行及测试**

**步骤1.**  **启动Hadoop**

1.   通过下列命令启动Hadoop：

start-all.sh

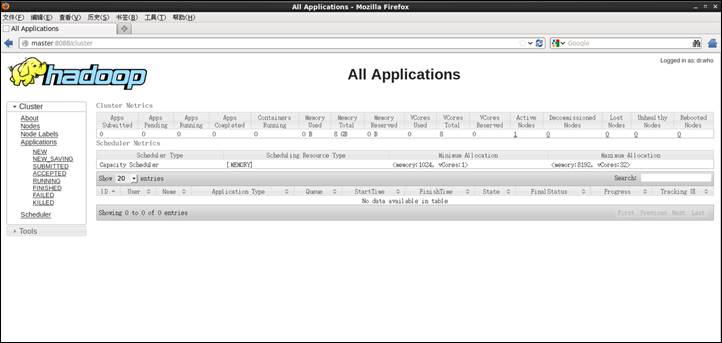
2.   通过下列命令，查看相应的JVM进程确定Hadoop是否配置及启动成功：

jps

当有5个进程启动时，则说明Hadoop已经成功启动.

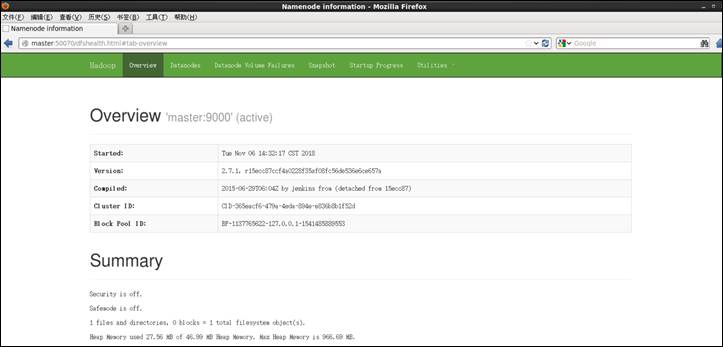
**步骤2.**  **Web页面测试**

1.   当Hadoop成功启动后，我们打开浏览器，输入网址[***http://master:8088***](http://master:8088/)便可以访问任务调度的Web管理页面：



**图 1.6 访问Hadoop任务调度页面**

2.   当Hadoop成功启动后，我们打开浏览器，输入网址***http://master:50070***便可以访问HDFS的Web管理页面：



**图 1.7 访问HDFS Web管理页面**