# Hadoop分布式集群部署实验报告

## 实验一：CenOS系统安装及配置

|  |
| --- |
| 【实验名称】 CenOS系统安装及配置 |
| 【实验目的】 |
| 掌握CenOS的安装   1. 安装CenOS 2. 配置静态ip |
| 【实验原理】 |
| 在虚拟机软件VMware中安装CentOS |
| 【实验环境】 |
| VMware® Workstation 14 Pro、CentOS-7-x86\_64-Minimal-1804  提示：本实验中Linux系统用户名为个人姓名全拼，密码为123456。 |
| 【实验步骤】 |
| 1. 创建虚拟机   （1）打开VMware  1.1  （2）在虚拟机软件中创建虚拟机。  1.2  选择典型，点击下一步。  1.3  （3）选择要安装镜像。  1.4  （4）命名虚拟机及选择虚拟机安装位置。安装位置选择一个空间有100G以上的分区（不能选择C盘），后面还有两个虚拟机需要存放在同一分区。  1.5  （5）选择虚拟机磁盘大小，默认20G。  1.6  1.7  二、安装CentOS系统  （1）上面步骤完成后虚拟机创建完成，虚拟机会自动重启进入安装系统阶段。媒介测试，默认一分钟后自动进入安装界面。  1.8  1.9   1. 选择语言，默认选择English。   1.10   1. 选择时区。选择Asia，Shanghai,时间为外部Windows的时间，完成后点Done。   1.12  1.11   1. 系统分区。选择自动分区，即不做修改直接点Done。   1.14   1. 开始安装系统。点Begin Installation。   1.13  （7）创建root密码。设置为123456方便记忆，完成后点Done。  1.15  1.16  （8）创建普通用户，用户名用本人姓名拼音，密码用123456。  1.17  1.18  （9）安装完毕，重启。  1.20  三、系统设置  （1）增加sudo权限。此时普通用户没有sudo使用权限。后面很多命令用到该权限。先切换到root。  1.21  之后修改/etc/sudoers文件  1.22  找到这一 行："root ALL=(ALL) ALL"，在下面添加"jiangyanping ALL=(ALL) ALL"(jiangyanping是用户名)，然后保存退出。  1.23  退出root用户。  1.24  （2）设置固定ip。  输入ip查询命令 ip addr  查看ip，发现 ens33 没有 inet 这个属性，那么就没法通过IP地址连接虚拟机。  1.25  给虚拟机配置静态ip，修改配置文件ifcfg-ens33    #修改BOOTPROTO=static #这里将dhcp换成static  #修改ONBOOT=yes #将no换成yes  #新增  IPADDR=192.168.1.100 #静态IP  GATEWAY=192.168.1.2 #默认网关  NETMASK=255.255.255.0 #子网掩码  保存退出。  1.27  重启网络服务:  # service network restart  1.28  查看ip,此时有inet属性。  1.29  （3）VMware网络设置，选择NAT模式，点击更改设置。  1.33  然后点NAT设置  1.34  1.35  以上，配置好保存退出。  如果在Windows可ping通192.168.1.100，则网络配置完成。  1.36  （4）通过SSH终端工具登录CentOS。打开SSH终端工具SecureCRTPortable.exe。  1.37  建立到虚拟机的连接。  1.38  1.39  点击连接后，选择接受并保存，然后输入密码，输完后点击确定。  1.40  1.41  登录成功，后面的操作大部分都在该终端完成。  1.42  （5）关闭防火墙，以免影响后面的一些应用的正常工作。  1.43  以上步骤完成了系统的安装及配置。 |

## 实验二：搭建Hadoop单机环境

|  |
| --- |
| 【实验名称】 搭建Hadoop环境 |
| 【实验目的】 |
| 掌握Hadoop环境搭建 |
| 【实验原理】 |
| 在一台虚拟机上安装java、hadoop以及其他工具，为实验三通过虚拟机克隆的方法克隆多台虚拟机，继而完成集群的搭建做准备。 |
| 【实验环境】 |
| CentOS，VMware |
| 【实验步骤】  在家目录创建一个software文件夹，用以存放后面用到的软件包。打开文件上传工具，并连接。  2.1  E:\360MoveData\Users\Administrator\Desktop\截图\2.2.PNG2.2  2.3  上传文件zookeeper-3.4.10.tar.gz、jdk-8u181-linux-x64.tar.gz、hbase-1.2.6.1-bin.tar.gz、hadoop-2.7.3.tar.gz到CentOS系统的software文件夹。  E:\360MoveData\Users\Administrator\Desktop\截图\2.4.PNG2.4  2.5   1. 安装java环境   （1）进入/home/jiangyanping/software/目录，解压jdk-8u181-linux-x64.tar.gz到/usr/local目录  E:\360MoveData\Users\Administrator\Desktop\截图\2.6.PNG2.6到/usr/local查看是否有jdk1.8.0\_181目录  E:\360MoveData\Users\Administrator\Desktop\截图\2.7.PNG2.7  （2）配置JDK环境变量，修改家目录下的配置文件.bashrc。  E:\360MoveData\Users\Administrator\Desktop\截图\2.8.PNG2.8  在末尾添加如下内容，保存退出。  E:\360MoveData\Users\Administrator\Desktop\截图\2.9.PNG2.9  export JAVA\_HOME=/usr/local/jdk1.8.0\_181  export JRE\_HOME=${JAVA\_HOME}/jre  export CLASSPATH=.:${JAVA\_HOME}/lib:${JRE\_HOME}/lib  export PATH=${JAVA\_HOME}/bin:$PATH  用命令Source /home/jiangyanping/.bashrc 刷新配置，使配置立即生效  （3）用命令java -version检验是否安装成功，如下有版本信息输出为正确。  E:\360MoveData\Users\Administrator\Desktop\截图\2.10.PNG2.10   1. 安装hadoop   （1）进入/home/panyongguang/software/目录，解压hadoop-2.7.3.tar.gz到/usr/local目录。  E:\360MoveData\Users\Administrator\Desktop\截图\2.11.PNG2.11  解压完成后，切换到目标目录查看解压好的内容。  E:\360MoveData\Users\Administrator\Desktop\截图\2.12.PNG2.12  （2）修改hadoop-2.7.3的操作权限，后面步骤需要在该目录建立目录与文件，需要写权限。切换到/usr/local目录，执行权限修改命令。  E:\360MoveData\Users\Administrator\Desktop\截图\2.13.PNG2.13  （3）配置hadoop的java工作环境，修改配置文件hadoop-env.sh。该配置文件在hadoop安装目录的etc/hadoop目录下。  找到export JAVA\_HOME，并指定java的安装路径：  E:\360MoveData\Users\Administrator\Desktop\截图\2.14.PNG2.14  2.15  （4）把hadoop加入到环境变量中，修改家目录下的配置文件.bashrc。在末尾添加如下内容，保存退出。  E:\360MoveData\Users\Administrator\Desktop\截图\2.16.PNG2.16  刷新配置，使配置立即生效，并检验hadoop是否安装成功。输出如下版本信息，说明安装成功。  E:\360MoveData\Users\Administrator\Desktop\截图\2.17.PNG2.17  （5）修改hadoop配置文件  切换到/usr/local/hadoop-2.7.3/etc/hadoop/目录，需要修改如下配置，指定集群入口地址以及数据保存目录。  1）etc/hadoop/core-site.xml:  <configuration>  <property>  <name>fs.defaultFS</name>  <value>hdfs://Master:9000</value>  </property>  <property>     <name>hadoop.tmp.dir</name>    <value>/usr/local/hadoop-2.7.3/</value>  </property>  </configuration>  查看文件  E:\360MoveData\Users\Administrator\Desktop\截图\2.18.PNG2.18  修改core-site.xml文件  2.19  2.20  保存退出。 |
| 2） etc/hadoop/mapred-site.xml:  把mapred-site.xml.template复制一份并改名为mapred-site.xml，然后添加内容。  E:\360MoveData\Users\Administrator\Desktop\截图\2.21.PNG2.21  E:\360MoveData\Users\Administrator\Desktop\截图\2.22.PNG2.22  3）etc/hadoop/yarn-site.xml:  E:\360MoveData\Users\Administrator\Desktop\截图\2.23.PNG2.23  2.24   1. etc/hadoop/hdfs-site.xml:   2.25  E:\360MoveData\Users\Administrator\Desktop\截图\2.26.PNG2.26   1. 配置主机名到IP的静态映射，修改配置文件/etc/hosts。   注释掉前面两行，添加192.168.1.100 Master  E:\360MoveData\Users\Administrator\Desktop\截图\2.27.PNG2.27  （7）修改主机名，修改文件/etc/hostname，把原来的主机名删除掉，改为Master  E:\360MoveData\Users\Administrator\Desktop\截图\2.28.PNG2.28  重启虚拟机，主机名则为上面新改的主机名。  E:\360MoveData\Users\Administrator\Desktop\截图\2.29.PNG2.29  （8）配置免密登录  $ ssh-keygen -t rsa -P "" -f ~/.ssh/id\_rsa  $ cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys  $ chmod 0600 ~/.ssh/authorized\_keys  E:\360MoveData\Users\Administrator\Desktop\截图\2.31.PNG2.31  配置完成，测试是否配置成功  E:\360MoveData\Users\Administrator\Desktop\截图\2.32.PNG2.32  登录成功后需要退出，回到原用户继续下面的实验。  三、启动hadoop，测试前面步骤是否完全配置成功  （1）格式化namenode。格式化的过程是创建初始目录和文件系统结构的过程。执行以下命令：  E:\360MoveData\Users\Administrator\Desktop\截图\2.33.PNG2.33  成功后有以下提示信息。  E:\360MoveData\Users\Administrator\Desktop\截图\2.34.PNG2.34  （2）启动HDFS，用 jps命令验证，正确启动会出现NameNode，DataNode，SecondaryNameNode三个进程：  E:\360MoveData\Users\Administrator\Desktop\截图\2.35.PNG2.35  （3）启动yarn，用 jps命令验证，正确启动将出现ResourceManager、NodeManager两个进程：  E:\360MoveData\Users\Administrator\Desktop\截图\2.36.PNG2.36  四、安装HBase单机环境。  （1）进入/home/jiangyanping/software/目录，解压hbase-1.2.6.1-bin.tar.gz到/usr/local目录。  E:\360MoveData\Users\Administrator\Desktop\截图\2.37.PNG2.37  切换到/usr/local目录可看到hbase目录。  E:\360MoveData\Users\Administrator\Desktop\截图\2.39.PNG2.39  （2）把Hbase加入到环境中，修改家目录下的.bashrc文件  E:\360MoveData\Users\Administrator\Desktop\截图\2.40.PNG2.40  （3）配置Hbase安装目录下配置文件conf/hbase-env.sh。修改JAVA\_HOME，找到对应位置把注释符号#去掉，填上java的安装路径，HBase是java编写的，需要java支持。  E:\360MoveData\Users\Administrator\Desktop\截图\2.41.PNG2.41  pids文件保存路径，把注释去掉。默认保存在/tmp，容易丢失。  E:\360MoveData\Users\Administrator\Desktop\截图\2.42.PNG2.42  （4）Hbase服务依赖zookeeper，现在还没有安装zookeeper，此处开启使用Hbase自带的zookeeper，方便后面测试。找到对应行，把前面的注释符号去掉。  E:\360MoveData\Users\Administrator\Desktop\截图\2.43.PNG2.43  （5）修改Hbase安装目录下配置文件hbase-site.xml，指定HBase的入口。  E:\360MoveData\Users\Administrator\Desktop\截图\2.44.PNG2.44  2.45  （6）修改hbase安装目录的操作权限，后面步骤需要在该目录建立目录与文件，当前用户需要写权限。切换到/usr/local目录，执行权限修改命令。  E:\360MoveData\Users\Administrator\Desktop\截图\2.46.PNG2.46  （7）启动HBase  E:\360MoveData\Users\Administrator\Desktop\截图\2.47.PNG2.47  以上，HBase 正确启动。不能正常启动时查看各步骤查找原因，正常后方可往下实验。  （8）停止所有进程  分别关闭HDFS和YARN  $ stop-dfs.sh  $ stop-yarn.sh  或者使用以下命令停止所有进程：  $ stop-all.sh  用命令jps查看：关闭了Hadoop所有进程  $ jps  Jps  2.48 |

## 实验三：搭建Hadoop的集群环境

|  |
| --- |
| 【实验名称】 搭建Hadoop集群环境 |
| 【实验目的】 |
| 掌握Hadoop集群搭建   1. 搭建Hadoop集群 |
| 【实验原理】 |
| 在单机模式安装好的情况下，克隆两个虚拟机。搭建一个三个节点的集群。 |
| 【实验环境】 |
| VMware® Workstation 14 Pro、CentOS7系统 |
| 【实验步骤】 |
| 一、克隆虚拟机前还需要做一些配置工作，免得克隆好后再一台台做配置，可节省一些时间。  （1）上面进行了一次单机版的格式化，后面集群需要重新格式化，所以前面格式化产生的文件需要删除掉。  E:\360MoveData\Users\Administrator\Desktop\截图\3.1.PNG3.1   1. 实验二hadoop是单机版，我们要搭建完全分布式系统。需要配置slaves文件，指定数据节点。slaves文件在hadoop安装目录的etc/hadoop下。   Cd 进入etc/hadoop下  3.2  用sudo vi 修改slaves文件  3.3  把原有内容删掉，加上Master，slave1，slave2三行  E:\360MoveData\Users\Administrator\Desktop\截图\3.4.PNG3.4  （3）配置主机名到IP的静态映射，修改配置文件/etc/hosts，注释前面两行，加上三条映射。  E:\360MoveData\Users\Administrator\Desktop\截图\3.5.PNG3.5  （4）上面配置的免密登录在集群中需要重新配置，所以先把配置好的相关文件删除掉。  E:\360MoveData\Users\Administrator\Desktop\截图\3.6.PNG3.6  （5）前面步骤配置的Hbase为单机版。分布式部署还需要做一些配置。  修改Hbase安装目录下配置文件conf/hbase-env.sh。  后面我们安装并使用独立的zookeeper，不使用HBase自带的zookeeper。需要关掉HBase自带的zookeeper自动启动的功能。找到对应行，把true设置为false。  Cd 进入conf目录，查看文件  E:\360MoveData\Users\Administrator\Desktop\截图\3.7.PNG3.7  修改hbase-env.sh  3.8   1. 修改Hbase安装目录下配置文件hbase-site.xml   3.9  E:\360MoveData\Users\Administrator\Desktop\截图\3.10.PNG3.10  参数说明：  1）hbase.rootdir，HDFS的入口地址，地址和端口要和你的hadoop配置一样（core-site.xml中的 <name>fs.default.name</name>），所有节点公用地址。  2）hbase.cluster.distributed，ture表示分布式  3）hbase.zookeeper.property.clientPort ， zookeeper端口  4）hbase.zookeeper.quorum ， zookeeper节点  5）hbase.zookeeper.property.dataDir ， zookeeper 保持信息的文件，默认为/tmp 重启会丢失 。  （7）修改配置conf/regionservers ，这个相当于hadoop的slaves配置，指定HBase服务器。  E:\360MoveData\Users\Administrator\Desktop\截图\3.11.PNG3.11  3.12   1. 安装zookeeper   （1）切换到/home/jiangyanping/software，解压zookeeper-3.4.10.tar.gz到/usr/local目录  E:\360MoveData\Users\Administrator\Desktop\截图\3.13.PNG3.13  解压完成切换到/usr/local目录可看到解压好的内容。  3.14  （2）把zookeeper添加到环境中  E:\360MoveData\Users\Administrator\Desktop\截图\3.15.PNG3.15   1. 将zookeeper安装目录中conf文件夹下的zoo\_sample.cfg重命名为zoo.cfg，重命名后zookeeper便可以识别到该文件。修改zoo.cfg文件。   3.16  E:\360MoveData\Users\Administrator\Desktop\截图\3.18.PNG3.18  注释掉原来的dataDir，加上dataDir和dataLogDir的路径    （4）上面的dataDir目录和dataLogDir目录默认是没有的，需要自己预先建立好。  E:\360MoveData\Users\Administrator\Desktop\截图\3.20.PNG3.20  修改上面建立好的文件夹的权限。给当前用户写权限。  E:\360MoveData\Users\Administrator\Desktop\截图\3.21.PNG3.21 （5）在上面配置的dataDir目录下新建一个名为 myid的文件，文件内容填写 1，zookeeper是根据该文件来决定zookeeper集群各个机器的身份分配。 E:\360MoveData\Users\Administrator\Desktop\截图\3.22.PNG3.22  以上步骤完成后开始克隆虚拟机。克隆虚拟机相当于把上面的系统及刚才所有的配置做完整的复制。  四、克隆虚拟机  （1）克隆。先关闭虚拟机，开机或挂起的情况下无法克隆。  E:\360MoveData\Users\Administrator\Desktop\截图\3.23.png3.23  E:\360MoveData\Users\Administrator\Desktop\截图\3.24.PNG3.24  虚拟机名称填slave1，新建一个slave1文件夹来保存虚拟机，与Master保存在同一级目录。  E:\360MoveData\Users\Administrator\Desktop\截图\3.25.PNG3.25  开始克隆Slave1    （2）同样方法克隆Slave2，虚拟机保存在Slave2文件夹。  E:\360MoveData\Users\Administrator\Desktop\截图\3.26.PNG3.26  五、配置并启动hadoop集群  （1）克隆完成后，三个虚拟机都开启。此时三个虚拟机的ip和主机名是一样的，不能用SSH终端连接。先修改Slave1的主机名，配置文件/etc/hostname。  E:\360MoveData\Users\Administrator\Desktop\截图\3.28.PNG3.28   1. 同样操作修改Slave2的主机名。修改完成都重启使配置生效。   3.29  （3）修改两个Slave的IP。先修改Slave1的。  修改Slave1的配置文件ifcfg-ens33  E:\360MoveData\Users\Administrator\Desktop\截图\3.30.PNG3.30  修改ip为：  IPADDR=192.168.1.101 #静态IP  保存退出。  E:\360MoveData\Users\Administrator\Desktop\截图\3.31.PNG3.31  重启网络服务:  # service network restart  E:\360MoveData\Users\Administrator\Desktop\截图\3.32.PNG3.32  查看ip，如果前面配置正确此时可看到正确ip。  E:\360MoveData\Users\Administrator\Desktop\截图\3.33.PNG3.33  （4）同样的步骤修改Slave2的ip为192.168.1.102。  3.34  3.35  3.36  （5）修改完毕可以通过SSH终端工具连接虚拟机了。  E:\360MoveData\Users\Administrator\Desktop\截图\3.37.PNG3.37  E:\360MoveData\Users\Administrator\Desktop\截图\3.38.PNG3.38  （6）配置免密登录。  在完全分布式模式下，集群内任一机器需要免密登录到集群内所有机器，即实现两两免密登录。免密登录的方法和伪分布的免密登录方法一样，分别在Master、Slave1、Slave2主机上生成公/私密钥对，然后将公钥发送给集群内的所有机器，以Master免密登录其他集群内所有机器为例。  在Master主机生成密钥对  命令：ssh-keygen -t rsa  其中，rsa表示加密算法，键入上面一条命令后连续敲击三次回车键，系统会自动在~/.ssh目录下生成公钥(id\_rsa.pub)和私钥(id\_rsa)，通过命令$ ls ~/.ssh查看  $ ls ~/.ssh  id\_rsa id\_rsa.pub  E:\360MoveData\Users\Administrator\Desktop\截图\3.39.PNG3.39  将Master公钥id\_rsa.pub复制到Master、Slave1和Slave2主机上。  E:\360MoveData\Users\Administrator\Desktop\截图\3.40.PNG3.40  3.41  3.42  验证免密登录：在Master主机输入以下命令验证，注意主机名称的变化。  $ ssh Master  $ ssh Slave1  $ ssh Slave2  3.43  3.44  3.45  成功登录后都需要退出返回到原来节点。  （7）在完成Master主机免密登录集群主机后，其他两台主机仿照Master步骤完成免密码登录设置。  （8） 格式化HDFS，在主节点即Master上格式化文件系统。  E:\360MoveData\Users\Administrator\Desktop\截图\3.46.PNG3.46  格式化成功后有如下信息  E:\360MoveData\Users\Administrator\Desktop\截图\3.47.PNG3.47  （9）启动HDFS及YARN，在主节点上输入启动命令即可启动系统。  启动成功后，主节点上有NameNode，DataNode，SecondaryNameNode  ResourceManager、NodeManager进程。如果没有，证明前面步骤出错，返回检查出错原因并修改。  E:\360MoveData\Users\Administrator\Desktop\截图\3.48.PNG3.48  两个从节点上有DataNode，NodeManager进程。如果没有，证明前面步骤出错，返回检查出错原因并修改。  E:\360MoveData\Users\Administrator\Desktop\截图\3.49.PNG3.49  E:\360MoveData\Users\Administrator\Desktop\截图\3.50.PNG3.50  （10）配置zookeeper，我们已经配置了Master的id，即在dataDir目录下新建一个名为 myid的文件，文件内容填写 1，zookeeper是根据该文件来决定zookeeper集群各个机器的身份分配。现在我们需要配置Slave1和Slave2的id，分别设置为2和3。配置方法与Master的配置一样。  E:\360MoveData\Users\Administrator\Desktop\截图\3.51.PNG3.51  E:\360MoveData\Users\Administrator\Desktop\截图\3.52.PNG3.52  （11）启动zookeeper，需要分别在每台机器上进行启动操作。命令为zkServer.sh start。启动成功后每台服务器上都有QuorumPeerMain进程。E:\360MoveData\Users\Administrator\Desktop\截图\3.53.PNG3.53  分别查看每台服务器zookeeper状态，三个台服务器只有一台是leader，其他的为follower。  E:\360MoveData\Users\Administrator\Desktop\截图\3.54.PNG3.54  E:\360MoveData\Users\Administrator\Desktop\截图\3.55.PNG3.55  E:\360MoveData\Users\Administrator\Desktop\截图\3.56.PNG3.56  （12）启动HBase。在主节点启动HBase，启动命令start-hbase.sh  启动成功后主节点上有HRegionServer和HMaster进程，Slave节点上有HRegionServer进程。  E:\360MoveData\Users\Administrator\Desktop\截图\3.57.PNG3.57  E:\360MoveData\Users\Administrator\Desktop\截图\3.58.PNG3.58  E:\360MoveData\Users\Administrator\Desktop\截图\3.59.PNG3.59  到此，整个hadoop和HBase搭建完毕。  六、在集群上做实验  1、HDFS操作  1)把第三章实验一，实验三在集群上操作一遍。实验三代码需打包上传到集群运行。  实验一：通过shell命令访问HDFS  (1)建立用户目录，类似于Linux下的家目录。  hdfs dfs -mkdir -p /user/jiangyanping  G:\2017级实践周\截图\4.8.PNG4.8  目录已存在，可直接查看根目录下的文件及文件夹。  hdfs dfs -ls /  4.9   1. 列出HDFS文件下名为user的文件夹中的文件     (3) 上传文件  在本地建立jiangyanping\_local文件，并写入个人姓名。将本地jiangyanping\_local文件上传到HDFS上并重命名为jiangyanping\_HDFS，并查看是否上传成功。  hdfs dfs -put jiangyanping\_local jiangyanping\_HDFS  4.10  4.11  4.12  (4)下载文件  将HDFS中的jiangyanping\_HDFS文件复制到本地系统并命名为4.13jiangyanping\_HDFS\_to\_local：  hdfs dfs -get jiangyanping\_HDFS jiangyanping\_HDFS\_to\_local  (5) 查看文件  查看HDFS下jiangyanping\_HDFS文件中的内容：  hdfs dfs -cat jiangyanping\_HDFS  G:\2017级实践周\截图\4.14.PNG4.14  (6) 删除文档  删除HDFS下名为jiangyanping\_HDFS的文件（参数r为递归删除），删除后会移动到回收站.trash。  hdfs dfs -rm jiangyanping\_HDFS  4.15  (7) 在HDFS的当前用户目录下建立新目录，本例为/user/jiangyanping。名为dir\_jiangyanping。  hdfs dfs -mkdir /user/jiangyanping/dir\_jiangyanping  4.16  (8) 本地文件复制到HDFS（注意与put的区别）  hdfs dfs -copyFromLocal jiangyanping\_local jiangyanping\_HDFS  4.17   1. 本地文件移动到HDFS。   hdfs dfs -moveFromLocal jiangyanping\_local jiangyanping\_local\_mv\_HDFS  G:\2017级实践周\截图\4.18.PNG4.18   1. 文件或文件夹复制   hdfs dfs -cp jiangyanping\_HDFS jiangyanping\_HDFS\_copy  4.19   1. 文件或文件夹的移动   hdfs dfs -mv jiangyanping\_HDFS\_copy jiangyanping\_HDFS\_mv  4.20   1. copyToLocal 复制文件到本地   hdfs dfs -copyToLocal jiangyanping\_HDFS jiangyanping\_HDFS\_to\_local  4.21   1. touchz 创建一个空文件test.txt   hdfs dfs -touchz test.txt  4.22   1. 把本地文件追加到HDFS文件末尾，并查看内容  hdfs dfs -appendToFile jiangyanping\_HDFS\_to\_local test.txt   G:\2017级实践周\截图\4.23.PNG4.23   1. 命令的综合运用   （1）在本地创建一个文件，以个人姓名命名，如file\_jiangyanping，查看file\_jiangyanping是否创建成功，往file\_jiangyanping内写一些内容。  4.24  4.25  （2）在HDFS上创建一个文件夹folder\_jiangyanping，把file\_jiangyanping上传到folder\_jiangyanping中。  4.26（3）查看是否上传成功。  4.27  （4）把file\_jiangyanping下载到本地，查看本地是否成功。  G:\2017级实践周\截图\4.28.PNG4.28  （5）把folder\_jiangyanping删除，并查看是否删除成功。  4.29  （6）最后把本地的file\_jiangyanping删除。  4.30  **实验三**  打包实验三的代码：  （1）选择HDFS,右键选择Export导出。  I:\2017级实践周\截图\7.1.PNG7.1  （2）选择JAR file,然后点next。  I:\2017级实践周\截图\7.2.PNG7.2  （3）选择要导出到的位置并命名（命名格式（文件名.jar）），然后点Finish。  I:\2017级实践周\截图\7.3.PNG7.3  导出完成，查看。  7.4  （4）将导出文件复制到Windows桌面上。  I:\2017级实践周\截图\7.5.PNG7.5  （5）用SecureFXportable.exe将java-files文件夹上传到集群。  6.7  I:\2017级实践周\截图\7.6.PNG7.6  （6）查看是否上传成功。  I:\2017级实践周\截图\7.7.PNG7.7  （7）运行创建文件CreateFile,运行完后hdfs dfs -ls 查看是否创建成功。  I:\2017级实践周\截图\7.8.PNG7.8  查看文件内容：  6.13  （8）运行文件的下载FileDownload，在本地查看是否下载成功。  I:\2017级实践周\截图\7.9.PNG7.9（9）运行文件的上传FileUpload，在HDFS查看是否上传成功。  I:\2017级实践周\截图\7.10.PNG7.10  （10）运行文件的删除DeleteFile，查看发现file\_jiangyanping.txt被删除了。  I:\2017级实践周\截图\7.11.PNG7.11  （11）运行文件属性查看GetFileAttribute。  I:\2017级实践周\截图\7.12.PNG7.12  （12）运行文件读取ReadFile。  I:\2017级实践周\截图\7.13.PNG7.13  2、HBase操作  HBase shell模式下可能不能退格，这个跟secureCRT有关，在secureCR工具栏上点击选项然后选择会话选项，在终端-仿真选Linux，然后每次需要删除时Ctrl+Backspace同时用即可实现Backspace功能。  G:\2017级实践周\截图\4.7.PNG4.7  1）启动hbase shell，  5.1  2）查看HBase状态  5.2  3）查看版本  5.3  4）使用help命令  5.4  5） 创建一个表，以个人姓名为表名，查看表结构。插入一个列族info，在info下插入以下列，分别为学号（student number），电话（phone number），QQ，身高(height)，体重(weight)，生日（年月）(birthday)。  5.5    6）全表扫描,删除生日信息。 |