

**数学与信息学院学生实验报告**

**实验课程名称：**大数据概论 **教师：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **实验项目名称** | **熟悉常用的Linux操作和Hadoop操作** | | | **实验成绩** |  |
| **学生姓名** |  | **学 号** |  | **年级专业班级** |  |
| **小组成员** | **无** | | | **实验日期** | **2019年9月25日** |

# 1. 实验目的和要求

## 1.1 实验目的

## Hadoop运行在Linux系统上，因此，需要学习实践一些常用的Linux命令。本实验旨在熟悉常用的Linux操作和Hadoop操作，为顺利开展后续其他实验奠定基础。

## 1.2 实验软硬件环境

① 操作系统：Linux Ubuntu 16.04

② Java：open-jdk-1.8

③ Apache Hadoop：3.2.0

## 1.3 实验要求

①在Linux系统上安装Hadoop(过程)，记录软件安装、测试的过程及结果。

②熟悉常用的Linux操作，实现并记录30项操作的过程及结果。

③熟悉常用的Hadoop操作，实现并记录4项操作的过程及结果。

# 2. 实验记录

2.1 Linux系统上安装Hadoop(过程)，记录软件安装、测试的过程及结果。

2.1.1 安装过程（简明说）：

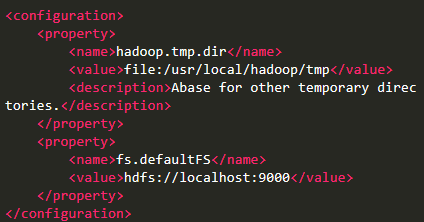
Java安装，SSH安装及配置SSH无密码登录 ==> Hadoop伪分布式安装

看起来过程极为清晰明了，步骤少之又少，但安装的过程是无比心酸的，Java和SSH安装网上教程很多就不多讲了。

伪分布式安装过程：

1. 伪分布式需要修改2个配置文件 core-site.xml 和 hdfs-site.xml

修改配置文件 core-site.xml：



修改配置文件 hdfs-site.xml：



1. 配置完成后，执行 NameNode 的格式化。

./bin/hdfs namenode -format

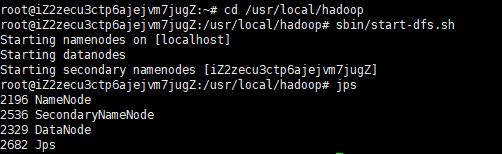
1. 接着开启 NameNode 和 DataNode 守护进程。

./sbin/start-dfs.sh

2.1.2 测试过程

1. 启动完成后，可以通过命令 jps 来判断是否成功启动，若成功启动则会列出如下进程: “NameNode”、”DataNode” 和 “SecondaryNameNode”。

2. 可以访问 Web 界面 [http://localhost:50070](http://localhost:50070/) 查看 NameNode 和 Datanode 信息，还可以在线查看 HDFS 中的文件。

2.1.3 测试结果： 

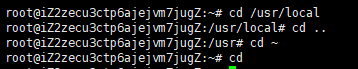
因为某种无法描述的原因，网页版的打不开。

# 2.2熟悉常用的Linux操作

**能一个截图表示的尽量一个截图表示，方便老师查看**

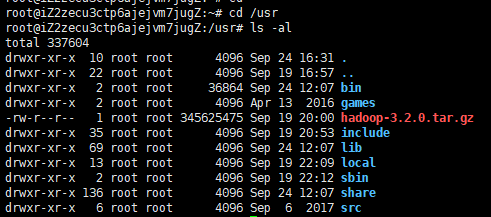
* cd命令：切换目录

1. 切换到目录“/usr/local”
2. 切换到当前目录的上一级目录
3. 切换到当前登录Linux系统的用户的自己的主文件夹



* ls命令：查看文件与目录

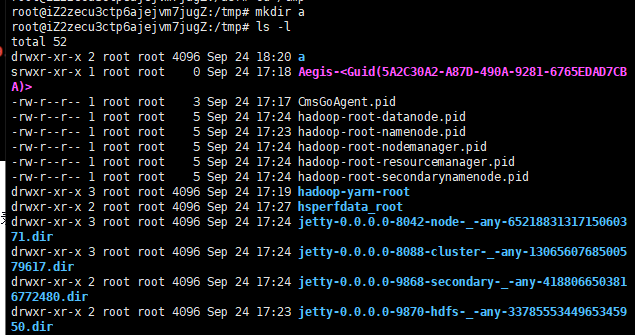
1. 查看目录“/usr”下的所有文件和目录



* mkdir命令：新建目录

（5）进入“/tmp”目录，创建一个名为“a”的目录，并查看“/tmp”目录下已经存在哪些目录

（6）进入“/tmp”目录，创建目录“a1/a2/a3/a4”





mkdir : -p或--parents 若所要建立目录的上层目录目前尚未建立，则会一并建立上层目录.

* rmdir命令：删除空的目录

（7）将上面创建的目录a（在“/tmp”目录下面）删除

（8）删除上面创建的目录“a1/a2/a3/a4” （在“/tmp”目录下面），然后查看“/tmp”目录下面存在哪些目录



rmdir: -p或--parents：删除指定目录后，若该目录的上层目录已变成空目录，则将其一并删除.

* cp命令：复制文件或目录

（9）将当前用户的主文件夹下的文件.bashrc复制到目录“/usr”下，并重命名为bashrc1

（10）在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下



cp : -R/r：递归处理，将指定目录下的所有文件与子目录一并处理

* mv命令：移动文件与目录，或更名

（11）将“/usr”目录下的文件bashrc1移动到“/usr/test”目录下

（12）将“/usr”目录下的test目录重命名为test2



mv: -b：当文件存在时，覆盖前，为其创建一个备份；

-f：若目标文件或目录与现有的文件或目录重复，则直接覆盖现有的文件或目录

-i：交互式操作，覆盖前先行询问用户，如果源文件与目标文件或目标目录中的文件同名，则询问用户是否覆盖目标文件。

* rm命令：移除文件或目录

（13）将“/usr/test2”目录下的bashrc1文件删除

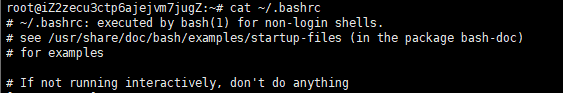
（14）将“/usr”目录下的test2目录删除



rm : -r或-R：递归处理，将指定目录下的所有文件与子目录一并处理；

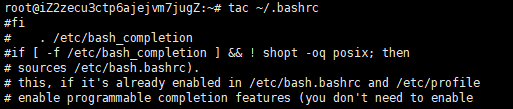
* cat命令：查看文件内容

1. 查看当前用户主文件夹下的.bashrc文件内容



* tac命令：反向查看文件内容

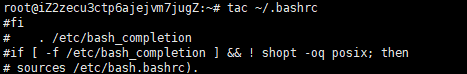
1. 反向查看当前用户主文件夹下的.bashrc文件的内容



一段一段倒着输出，不是倒着每句输出

* more命令：一页一页翻动查看

1. 翻页查看当前用户主文件夹下的.bashrc文件的内容





more： 从26%开始，使用鼠标的滑轮无法加载，只能按键。

按Space键：显示文本的下一屏内容。

按Enier键：只显示文本的下一行内容。

按斜线符|：接着输入一个模式，可以在文本中寻找下一个相匹配的模式。

按H键：显示帮助屏，该屏上有相关的帮助信息。

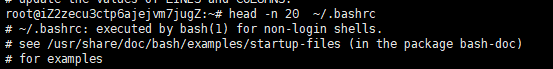
按B键：显示上一屏内容。

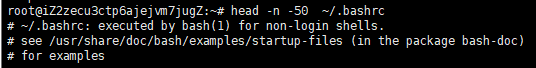
按Q键：退出rnore命令。

* head命令：取出前面几行

（18）查看当前用户主文件夹下.bashrc文件内容前20行

（19）查看当前用户主文件夹下.bashrc文件内容，后面50行不显示，只显示前面几行



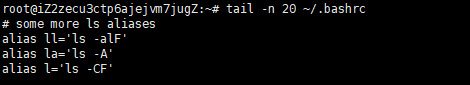


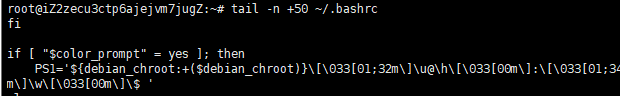
head : -n<数字>：指定显示头部内容的行数.

* tail命令：取出后面几行

（20）查看当前用户主文件夹下.bashrc文件内容最后20行

（21）查看当前用户主文件夹下.bashrc文件内容，并且只列出50行以后的数据





* touch命令：修改文件时间或创建新文件

（22）在“/tmp”目录下创建一个空文件hello，并查看文件时间

（23）修改hello文件，将文件时间整为5天前



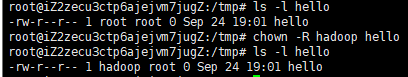




touch： -t：<日期时间> 使用指定的日期时间，而非现在的时间

* chown命令：修改文件所有者权限

1. 将hello文件所有者改为root帐号，并查看属性



chown : -R或——recursive：递归处理，将指定目录下的所有文件及子目录一并处理

* find命令：文件查找

1. 找出主文件夹下文件名为.bashrc的文件



* tar命令：压缩命令

（26）在根目录“/”下新建文件夹test，然后在根目录“/”下打包成test.tar.gz

（27）把上面的test.tar.gz压缩包，解压缩到“/tmp”目录





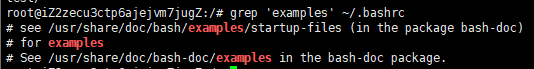
压　缩：tar -jcv -f filename.tar.bz2 要被压缩的文件或目录名称

查　询：tar -jtv -f filename.tar.bz2

解压缩：tar -jxv -f filename.tar.bz2 -C 欲解压缩的目录

* grep命令：查找字符串

1. 从“～/.bashrc”文件中查找字符串'examples'



‘Examples ‘加不加引号都没影响。

* 配置环境变量

（29）请在“～/.bashrc”中设置，配置Java环境变量

（30）查看JAVA\_HOME变量的值



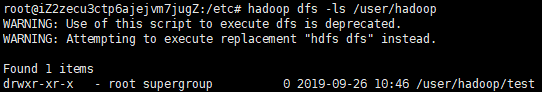




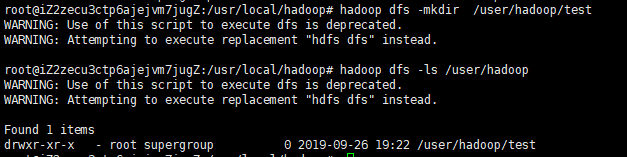


**2.3**熟悉常用的Hadoop操作，实现并记录4项操作的过程及结果。

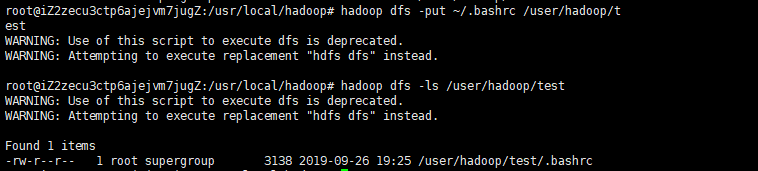
1. 使用hadoop用户登录Linux系统，启动Hadoop（Hadoop的安装目录为“/usr/local/hadoop”），为hadoop用户在HDFS中创建用户目录“/user/hadoop”

1. 接着在HDFS的目录“/user/hadoop”下，创建test文件夹，并查看文件列表

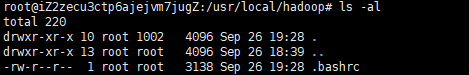


1. 将Linux系统本地的“～/.bashrc”文件上传到HDFS的test文件夹中，并查看test



（34）将HDFS文件夹test复制到Linux系统本地文件系统的“/usr/local/hadoop”目录下





# 总结

我先表扬一下自己，在林老师和朋友们的帮助下，历尽千辛万苦最终完成hadoop的下载安装，并顺利完成了实验Linux和HDFS的操作。当实验报告终于只剩下总结时，真心感慨良多，觉得自己很厉害诶，也增长了不少学习的兴趣和自信。

从大一开始就在不断尝试，学习安装各种软件，除了devC++之外，任何所需软件，每次的下载都让我痛苦万分，总是一个人，根据错误提示，查找网上各种信息进行修正，安装，但这一次比之前有了一点改变，虽说开始是一个人用一个下午的事件，看着黑屏中无穷无尽的各种错误提示中挣扎最后放手，第二天又不得不继续的状况下持续三天，终于在之后迎来了老师，朋友的帮助，在经历一个下午后终于完成了hadoop的安装。后续的Linux和hadoop很快地完成了，小 case。

要说我在这次安装中学到了什么，我总结成以下几点：

1. 对网上的安装教程不能按部就班.
2. 过程是痛苦的，结局是美好的.
3. 万事开头难，安装过程中，对Linux文件管理理解通透了，完成实验很轻松.
4. 将错误提示粘贴到网上去查时，没必要从第一个网页开始一个个打开看，应选出最符合自身条件的.
5. 安装Linux先要明确自己将hadoop解压到哪个子文件中，像etc,bin,sbin之内的文件有好几个，一弄错，就要耗费很长时间.
6. 实在解决不了时，立即问同学，老师；或休息几分钟再重新开始.
7. 不要退缩，要抓住本质.

说实话，在安装的每一步过程中都有错误，特别是格式化的时候，错误解决一个又一个，不断地修改各种配置文件，但修改的之后又完全不知道这种解决方法适不适用，会不会更糟，自己却只能不断地弄，感觉hadoop 文件包已经被我改得面目全非了，最后因为格式化太多次，安装好后，在做上传文件时，系统显示 因为被格式化太多次，磁盘损坏，让我删掉hadoop生成的所有文件，重新格式化，但是haddop的文件在哪我都不知道，都不知道删哪，最终我选择在安装一次（试不试特别厉害，其实我内心在哭泣），最后重新解压缩，重新配置文件，最后竟然顺利通过（不得不说挺幸运的）。这次实验我尽自己最大的努力，结果也没让我失望，是一次想哭想笑的经历，希望后续实验也能保持这样求知的心态。