

第二课--Hadoop Shell操作

任务目标

相关知识

系统环境

任务内容

- 1、任务步骤
- 2、Hadoop主要命令
- 3、Hadoop Jobs

第二课--Hadoop Shell操作

任务目标

- 1、熟练掌握常用的Hadoop Shell命令

相关知识

调用文件系统(FS)Shell命令应使用 `hadoop fs` 的形式。所有的FS shell命令使用URI路径作为参数。URI格式是 `scheme://authority/path`。对HDFS文件系统，`scheme`是`hdfs`，对本地文件系统，`scheme`是`file`。其中`scheme`和`authority`参数都是可选的，如果未加指定，就会使用配置中指定的默认`scheme`。一个HDFS文件或目录比如`/parent/child`可以表示成 `hdfs://namenode:namenodeport/parent/child`，或者更简单的`/parent/child`（假设你配置文件中的默认值是`namenode:namenodeport`）。大多数FS Shell命令的行为和对应的Unix Shell命令类似，出错信息会输出到`stderr`，其他信息输出到`stdout`。

系统环境

Linux Ubuntu 16.04

Windows10 1903

hadoop-2.7.1

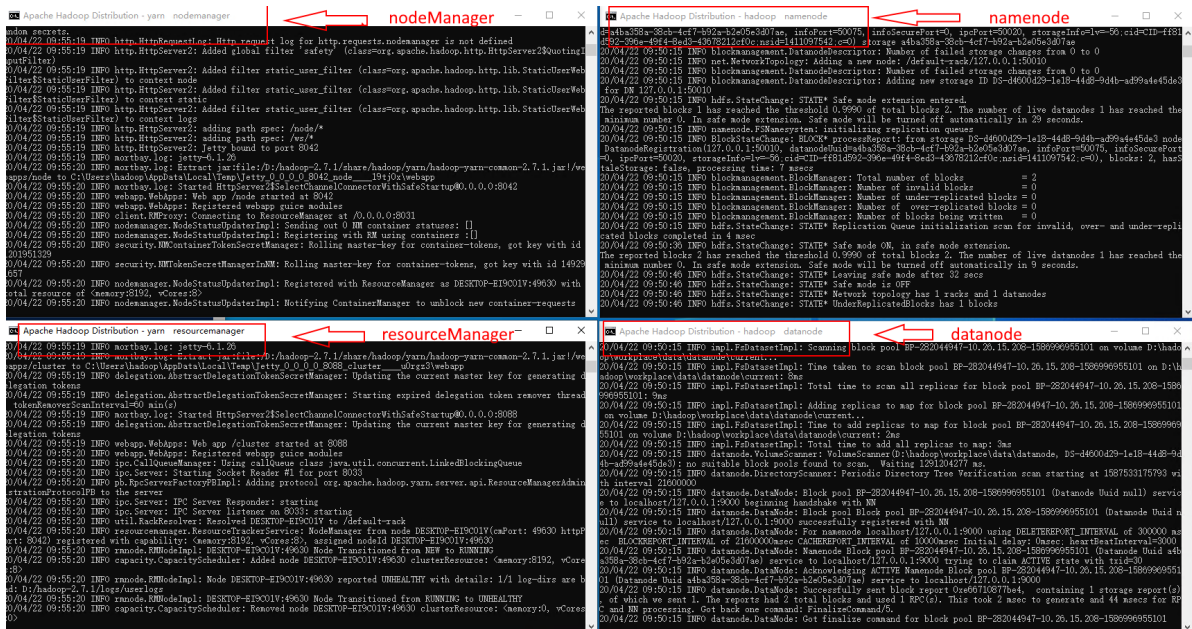
任务内容

- 1、学习在开启、关闭Hadoop
- 2、学习在Hadoop中创建、修改、查看、删除文件夹及文件
- 3、学习改变文件的权限及文件的拥有者
- 4、学习使用shell命令提交job任务
- 5、Hadoop安全模式的进入与退出

1、任务步骤

- 1、打开cmd，切换到`/hadoop/sbin`目录下，启动Hadoop。

```
cd d:/hadoop-2.7.1/sbin
start-dfs.cmd
start-yarn.cmd
```

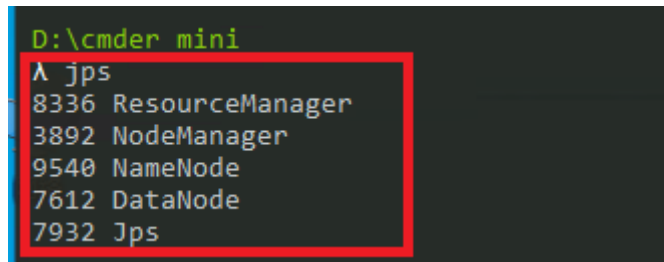


除了分步启动start-dfs.cmd和start-yarn.cmd，还可以直接执行start-all.cmd，还可以。

2、执行jps，检查一下Hadoop相关进程是否启动

打开Perferences，出现了Hadoop Map/Reduce 菜单，设置Hadoop installation directory，设置Hadoop的安装目录d:\hadoop-2.7.1。

jps



3、在Linux本地/data目录下，创建一个data.txt文件，并向其中写入hello hadoop！。

```
cd /data
touch data.txt
echo hello hadoop! >> data.txt
```

2、Hadoop主要命令

1、显示文件夹命令-ls

```
hdfs dfs -ls [-R] <dir>
```

-R 递归显示根目录下所有的文件和目录

```
D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-22 15:32 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/guest/file.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:55 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput
```

2、上传本地文件到hdfs文件系统-put

```
hdfs dfs -put <local file> <dst file>
```

通过vim新建test.txt文件，并将test.txt文件上传到HDFS的user/目录下

```
D:\cmdr_mini
λ ls
bin/          config/  LICENSE    vendor/
Cmder.exe*   icons/   test.txt   Version 1.3.14.982

D:\cmdr_mini
λ hdfs dfs -put test.txt hdfs://localhost:9000/user/
```

3、将hdfs文件系统的文件复制到本地目录-get

```
hdfs dfs -get <hdfs loc> <local loc>
```

将hdfs的/user/README.txt文件复制本地文件目录

```
D:\cmdr_mini
λ hdfs dfs -get hdfs://localhost:9000/user/README.txt ./
20/04/23 16:05:10 WARN hdfs.DFSClient: DFSInputStream has been closed already

D:\cmdr_mini
λ ls
bin/          config/  LICENSE    test.txt   Version 1.3.14.982
Cmder.exe*   icons/   README.txt  vendor/
```

4、cat/text

```
hdfs dfs -cat <file>
hdfs dfs -text <file>
```

显示HDFS文件的内容，并将文件的内容输出到标准输出中

```
D:\cmdr_mini
λ hdfs dfs -cat hdfs://localhost:9000/user/test.txt
hello python
hello hadoop
hello vim
```

Text命令可以读取压缩文件，并输出解压的数据。Text方法可以将源文件输出为文本格式。允许的格式是zip和TextRecordInputStream。

```
D:\cmdr_mini
λ hdfs dfs -text hdfs://localhost:9000/user/test.txt
hello python
hello hadoop
hello vim
```

Text命令与head命令结合使用

```
hdfs dfs -text <file> | head
```

```
D:\cmdr_mini
λ hdfs dfs -text hdfs://localhost:9000/user/test.txt | head
hello python
hello hadoop
hello vim
```

5、cp/mv 复制和移动

```
hdfs dfs -cp <source> <dst>
hdfs dfs -mv <source> <dst>
```

从HDFS文件系统的/guest/file.txt文件复制到/user/目录下。

```
D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-22 15:32 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/guest/file.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:01 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

D:\cmdr_mini
λ hdfs dfs -cp /guest/file.txt /user/
20/04/23 16:27:08 WARN hdfs.DFSClient: DFSInputStream has been closed already

D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-22 15:32 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/guest/file.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput
```

将/guest/file.txt文件剪切到/output/目录下

```

D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-22 15:32 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/guest/file.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

D:\cmdr_mini
λ hdfs dfs -mv /guest/file.txt /output/

D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

```

6、chmod

```
hdfs dfs -chmod [-R] <mode> <file/dir>
```

改变file/directory的权限，使用标准的Unix文件权限。


```

D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/output/file.txt
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
-rw-r--r-- 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

```

```

D:\cmdr_mini
λ hdfs dfs -chmod 777 hdfs://localhost:9000/user/file.txt

D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/output/file.txt
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
-rwxrwxrwx 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

```

7、getmerge

```
hdfs dfs -getmerge <source dir> <local file>
```

将/user/目录下所有的文件合并为user.zip文件

```

D:\cmdr_mini
λ hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/output/file.txt
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:55 hdfs://localhost:9000/user/README1.txt
-rwxrwxrwx 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

D:\cmdr_mini
λ hdfs dfs -getmerge /user user.zip

D:\cmdr_mini
λ ls
bin/          config/  LICENSE  test.txt  vendor/
Cmder.exe*   icons/  README.txt user.zip  Version 1.3.14.982

```

8、rm

```
hdfs dfs -rm [-r] [-skipTrash] <file/dir>
```

从HDFS文件系统中删除文件。

```

D:\cmdr_mini
hdfs dfs -rm /user/README1.txt
20/04/23 16:42:02 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /user/README1.txt

D:\cmdr_mini
hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/guest
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 15:32 hdfs://localhost:9000/guest/README1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 09:06 hdfs://localhost:9000/guest/input
-rw-r--r-- 1 hadoop supergroup 90 2020-04-16 09:06 hdfs://localhost:9000/guest/input/data.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/output/file.txt
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:42 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rwxrwxrwx 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

```

从HDFS文件系统递归删除guest目录。

```

D:\cmdr_mini
hdfs dfs -rm -r /guest/
20/04/23 16:44:07 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /guest

D:\cmdr_mini
hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/output/file.txt
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 14:05 hdfs://localhost:9000/test2/test1.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:42 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rwxrwxrwx 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

```

skipTrash删除文件

```

D:\cmdr_mini
hdfs dfs -rm -skipTrash /test2/test1.txt
Deleted /test2/test1.txt

D:\cmdr_mini
hdfs dfs -ls -R hdfs://localhost:9000/
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:31 hdfs://localhost:9000/output
-rw-r--r-- 3 hadoop supergroup 0 2020-04-16 16:45 hdfs://localhost:9000/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 0 2020-04-22 10:12 hdfs://localhost:9000/output/file.txt
-rw-r--r-- 3 hadoop supergroup 94 2020-04-16 16:45 hdfs://localhost:9000/output/part-r-00000
drwxr-xr-x - hadoop supergroup 0 2020-04-22 10:10 hdfs://localhost:9000/test1
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:46 hdfs://localhost:9000/test2
drwxr-xr-x - hadoop supergroup 0 2020-04-23 16:42 hdfs://localhost:9000/user
-rw-r--r-- 1 hadoop supergroup 1366 2020-04-22 14:27 hdfs://localhost:9000/user/README.txt
-rwxrwxrwx 1 hadoop supergroup 0 2020-04-23 16:27 hdfs://localhost:9000/user/file.txt
-rw-r--r-- 1 hadoop supergroup 39 2020-04-23 16:01 hdfs://localhost:9000/user/test.txt
drwxr-xr-x - hadoop supergroup 0 2020-04-16 14:52 hdfs://localhost:9000/user/wcinput

```

9、du

```
hdfs dfs -du [-s] [-h] <file/dir>
```

显示文件的尺寸，-h 和-s 参数option summarises all the files, instead of giving you individual file sizes.

One thing to note is that the size reported is un-replicated. If your replication factor is 3, the actual disk usage will be 3 times this size.

```

D:\cmdr_mini
λ hdfs dfs -du /user/test.txt
39 /user/test.txt

D:\cmdr_mini
λ hdfs dfs -du /user/README.txt
1366 /user/README.txt

D:\cmdr_mini
λ hdfs dfs -du -s -h /user/README.txt
1.3 K /user/README.txt

D:\cmdr_mini
λ hdfs dfs -du -s -h /user/
1.4 K /user

D:\cmdr_mini
λ hdfs dfs -du -s /user/
1405 /user

D:\cmdr_mini
λ hdfs dfs -du /user/
1366 /user/README.txt
0 /user/file.txt
39 /user/test.txt
0 /user/wcinput

```

10、count

```
hdfs dfs -count [-q] <dir>
```

对文件夹计数

```

D:\cmdr_mini
λ hdfs dfs -count /user/
      2      3      1405 /user

D:\cmdr_mini
λ hdfs dfs -count -q /user/
none      inf      none      inf      2      3      1405 /user

```

11、Admin Report

```
hadoop dfsadmin -report
```

显示集群的使用信息。


```

λ hadoop dfsadmin -report
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Configured Capacity: 255413186560 (237.87 GB)
Present Capacity: 249168480057 (232.06 GB)
DFS Remaining: 249168478208 (232.06 GB)
DFS Used: 1849 (1.81 KB)
DFS Used%: 0.00%
Under replicated blocks: 1
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0

-----
Live datanodes (1):

Name: 127.0.0.1:50010 (127.0.0.1)
Hostname: DESKTOP-EI9C01V
Decommission Status : Normal
Configured Capacity: 255413186560 (237.87 GB)
DFS Used: 1849 (1.81 KB)
Non DFS Used: 6244706503 (5.82 GB)
DFS Remaining: 249168478208 (232.06 GB)
DFS Used%: 0.00%
DFS Remaining%: 97.56%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Thu Apr 23 16:56:12 CST 2020

```

12、tail方法

```
hadoop fs -tail /test1/data.txt
```

tail方法是将文件尾部1K字节的内容输出。支持-f选项，行为和Unix中一致。

3、Hadoop Jobs

1、进入Hadoop安全模式

```
hdfs dfsadmin -safemode enter
```

2、退出Hadoop安全模式

```
hdfs dfsadmin -safemode leave
```

```

D:\cmdr_mini
λ hdfs dfsadmin -safemode enter
Safe mode is ON

D:\cmdr_mini
λ hdfs dfsadmin -safemode leave
Safe mode is OFF

```

3、Hadoop Jobs

直接通过jar包中mainclass启动Hadoop

```
hadoop jar <jar> [mainClass] args...
```

```
D:\vender_mini
λ hadoop jar hadoop-mapreduce-examples-2.7.1.jar wordcount /user/README.txt /out
20/04/23 17:22:13 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8020
20/04/23 17:22:14 INFO input.FileInputFormat: Total input paths to process : 1
20/04/23 17:22:14 INFO mapreduce.JobSubmitter: number of splits:1
20/04/23 17:22:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1587628388073_0001
20/04/23 17:22:15 INFO impl.YarnClientImpl: Submitted application application_1587628388073_0001
20/04/23 17:22:15 INFO mapreduce.Job: The url to track the job: http://DESKTOP-EI9C01V:8088/proxy/application_1587628388073_0001/
20/04/23 17:22:15 INFO mapreduce.Job: Running job: job_1587628388073_0001
```

Scala启动任务

```
hadoop jar <jar> com.twitter.scalding.Tool [mainClass] args
```

关闭mapred任务

```
mapred job -kill <jobid>
```

4、切换到/apps/hadoop/sbin目录下，关闭Hadoop

```
stop-yarn.cmd
stop-dfs.cmd
```