

大数据管理技术 第一次上机

林汇平 1800013104

- 实习要求：完成基本的 HDFS Shell 命令操作。至少完成这些命令：
- cat chmod chown cp du get ls mkdir mv put rm setrep stat tail test touchz
- 实习环境：

虚拟机：Ubuntu 15.1.0 build-13591040

主机操作系统：Windows 10, 64-bit (Build 17134) 10.0.17134

内存：4GB

硬盘：20GB

CPU：Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz(1992 MHz)

1. cat命令：

- 选项名称：-cat
- 使用格式：-cat <hdfs路径>
- 作用：查看文件内容
- 使用示例：本例中，-cat input的作用为：查看input文件内容。

```
phoenix@Master:~$ hdfs dfs -put /usr/local/hadoop/hadoop-2.7.7/etc/hadoop/*.xml input
phoenix@Master:~$ hdfs dfs -cat input/*
<!--
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.
```

2. chmod命令：

- 选项名称：-chmod
- 使用格式：-chmod [-R] <权限模式> [路径]
- 作用：修改权限
- 使用示例：本例中，先将test文件夹的权限先设为755、后改为700，可以通过ls命令看到其权限有所改变。使用-R可以改变文件夹中全部文件的权限。

```

phoenix@Master:~/桌面/example$ hdfs dfs -chmod 755 test
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - phoenix supergroup 0 2020-07-12 10:52 input
drwxr-xr-x - phoenix supergroup 0 2020-07-12 11:34 test
phoenix@Master:~/桌面/example$ hdfs dfs -chmod 700 test
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - phoenix supergroup 0 2020-07-12 10:52 input
drwx----- - phoenix supergroup 0 2020-07-12 11:34 test
phoenix@Master:~/桌面/example$ hdfs dfs -chmod -R 700 test
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - phoenix supergroup 0 2020-07-12 10:52 input
drwx----- - phoenix supergroup 0 2020-07-12 11:34 test

```

3. chown命令:

- 选项名称: -chown
- 使用格式: -chown [-R] [属主][:[属组]] 路径
- 作用: 修改属主
- 使用示例: 本例中, 先创建了用户phoenix2, 然后将他加入组hadoop, 然后把test文件的属主从用户phoenix换成了用户phoenix2。

```

phoenix@Master:~/桌面/example$ sudo useradd -g hadoop phoenix2
phoenix@Master:~/桌面/example$ sudo usermod -a -G hadoop phoenix2
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - phoenix supergroup 0 2020-07-12 10:52 input
drwx----- - phoenix supergroup 0 2020-07-12 11:34 test
phoenix@Master:~/桌面/example$ hdfs dfs -chown phoenix2:phoenix2 test
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - phoenix supergroup 0 2020-07-12 10:52 input
drwx----- - phoenix2 phoenix2 0 2020-07-12 11:34 test
phoenix@Master:~/桌面/example$

```

4. cp命令:

- 选项名称: -cp
- 使用格式: -cp <源路径> <目的路径>
- 作用: 复制
- 使用示例: 本例中, 将input文件复制到了test文件夹中, 使用ls命令可以看到test文件夹中新出现了input文件。

```

phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - phoenix supergroup 0 2020-07-12 10:52 input
drwx----- - phoenix2 phoenix2 0 2020-07-12 11:34 test
phoenix@Master:~/桌面/example$ hdfs dfs -cp input test
phoenix@Master:~/桌面/example$ hdfs dfs -ls test
Found 1 items
drwxr-xr-x - phoenix phoenix2 0 2020-07-12 13:12 test/input

```

5. du命令:

- 选项名称: -du
- 使用格式: -du <路径>
- 作用: 统计目录下文件大小 (单位是字节)
- 使用示例: 本例中, 统计了test文件夹下的文件大小, 是28373字节。

```

phoenix@Master:~/桌面/example$ hdfs dfs -du test
28373 test/input

```

6. get命令:

- 选项名称: -get
- 使用格式: -get <源路径> <目的路径>
- 作用: 把hdfs中的文件下载到本地
- 使用示例: 本例中, 把hdfs中的input文件下载到了本地桌面/example/example中。

```
phoenix@Master:~/桌面/example$ ls
file1 file2
phoenix@Master:~/桌面/example$ hdfs dfs -get input example
phoenix@Master:~/桌面/example$ ls example/
capacity-scheduler.xml  hadoop-policy.xml  httpfs-site.xml  kms-site.xml  yarn-site.xml
core-site.xml           hdfs-site.xml      kms-acls.xml      mapred-site.xml
```

7. ls命令:

- 选项名称: -ls
- 使用格式: -get <路径>
- 作用: 查看指定路径的当前目录结构
- 使用示例: 本例中, 查看了input文件夹中的内容。

```
phoenix@Master:~/桌面/example$ hdfs dfs -ls input
Found 9 items
-rw-r--r--  2 phoenix supergroup      4436 2020-07-12 10:52 input/capacity-scheduler.xml
-rw-r--r--  2 phoenix supergroup      1122 2020-07-12 10:52 input/core-site.xml
-rw-r--r--  2 phoenix supergroup      9683 2020-07-12 10:52 input/hadoop-policy.xml
-rw-r--r--  2 phoenix supergroup      1361 2020-07-12 10:52 input/hdfs-site.xml
-rw-r--r--  2 phoenix supergroup       620 2020-07-12 10:52 input/httpfs-site.xml
-rw-r--r--  2 phoenix supergroup      3518 2020-07-12 10:52 input/kms-acls.xml
-rw-r--r--  2 phoenix supergroup      5540 2020-07-12 10:52 input/kms-site.xml
-rw-r--r--  2 phoenix supergroup      1175 2020-07-12 10:52 input/mapred-site.xml
-rw-r--r--  2 phoenix supergroup       918 2020-07-12 10:52 input/yarn-site.xml
```

8. mkdir命令:

- 选项名称: -mkdir
- 使用格式: -mkdir <hdfs路径>
- 作用: 创建空白文件夹
- 使用示例: 本例中, 创建了new文件夹, 通过ls命令可以看到。

```
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 2 items
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 10:52 input
drwx----- - phoenix2 phoenix2      0 2020-07-12 13:12 test
phoenix@Master:~/桌面/example$ hdfs dfs -mkdir new
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 3 items
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 10:52 input
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 14:20 new
drwx----- - phoenix2 phoenix2      0 2020-07-12 13:12 test
```

9. mv命令:

- 选项名称: -mv
- 使用格式: -mv <源路径> <目的路径>
- 作用: 移动
- 使用示例: 本例中, 用mv命令将test/input/core-site.xml文件移动到了new文件夹中。

```
phoenix@Master:~/桌面/example$ hdfs dfs -ls test/input
Found 9 items
-rw-r--r--  2 phoenix phoenix2      4436 2020-07-12 13:12 test/input/capacity-scheduler.xml
-rw-r--r--  2 phoenix phoenix2      1122 2020-07-12 13:12 test/input/core-site.xml
-rw-r--r--  2 phoenix phoenix2      9683 2020-07-12 13:12 test/input/hadoop-policy.xml
-rw-r--r--  2 phoenix phoenix2      1361 2020-07-12 13:12 test/input/hdfs-site.xml
-rw-r--r--  2 phoenix phoenix2       620 2020-07-12 13:12 test/input/httpfs-site.xml
-rw-r--r--  2 phoenix phoenix2      3518 2020-07-12 13:12 test/input/kms-acls.xml
-rw-r--r--  2 phoenix phoenix2      5540 2020-07-12 13:12 test/input/kms-site.xml
-rw-r--r--  2 phoenix phoenix2      1175 2020-07-12 13:12 test/input/mapred-site.xml
-rw-r--r--  2 phoenix phoenix2       918 2020-07-12 13:12 test/input/yarn-site.xml
phoenix@Master:~/桌面/example$ hdfs dfs -mv test/input/core-site.xml new
phoenix@Master:~/桌面/example$ hdfs dfs -ls new
Found 1 items
-rw-r--r--  2 phoenix phoenix2      1122 2020-07-12 13:12 new/core-site.xml
```

10. put命令:

- 选项名称: -put

- 使用格式: `-put <多个linux上的文件> <hdfs路径>`
- 作用: 上传文件
- 使用示例: 本例中, 用put命令将linux文件系统中的桌面/example/file1文件, 上传到了hdfs文件系统中。

```
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 4 items
-rw-r--r--  2 phoenix supergroup      16 2020-07-12 14:45 file2
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 10:52 input
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 14:42 new
drwx----- - phoenix2 phoenix2      0 2020-07-12 13:12 test
phoenix@Master:~/桌面/example$ hdfs dfs -put file1
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 5 items
-rw-r--r--  2 phoenix supergroup      16 2020-07-12 14:45 file1
-rw-r--r--  2 phoenix supergroup      16 2020-07-12 14:45 file2
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 10:52 input
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 14:42 new
drwx----- - phoenix2 phoenix2      0 2020-07-12 13:12 test
```

11. rm命令:

- 选项名称: `-rm`
- 使用格式: `-rm [-skipTrash] <路径>`
- 作用: 删除文件/空白文件夹
- 使用示例: 本例中, 用rm命令将file1文件删除。但是注意此处只能删除文件或者空白文件夹, 不能删除非空文件夹。

```
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 5 items
-rw-r--r--  2 phoenix supergroup      16 2020-07-12 14:45 file1
-rw-r--r--  2 phoenix supergroup      16 2020-07-12 14:45 file2
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 10:52 input
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 14:42 new
drwx----- - phoenix2 phoenix2      0 2020-07-12 13:12 test
phoenix@Master:~/桌面/example$ hdfs dfs -rm file1
20/07/12 14:47:56 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted file1
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 4 items
-rw-r--r--  2 phoenix supergroup      16 2020-07-12 14:45 file2
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 10:52 input
drwxr-xr-x  - phoenix supergroup      0 2020-07-12 14:42 new
drwx----- - phoenix2 phoenix2      0 2020-07-12 13:12 test
```

12. setrep命令:

- 选项名称: `-setrep`
- 使用格式: `-setrep [-R] [-w] <副本数> <路径>`
- 作用: 修改副本数量
- 使用示例: 本例中, 我们先put了file1, 显示其副本数量是2。通过setrep将副本数量修改为3, 再查看时副本数量已经变成3。

如果最后的路径表示文件夹, 那么需要跟选项-R, 表示对文件夹中的所有文件都修改副本。选项-w表示等待副本操作结束才退出命令。

```
phoenix@Master:~/桌面/example$ hdfs dfs -put file1
phoenix@Master:~/桌面/example$ hdfs dfs -ls file1
-rw-r--r--  2 phoenix supergroup      16 2020-07-12 14:52 file1
phoenix@Master:~/桌面/example$ hdfs dfs -setrep 3 file1
Replication 3 set: file1
phoenix@Master:~/桌面/example$ hdfs dfs -ls file1
-rw-r--r--  3 phoenix supergroup      16 2020-07-12 14:52 file1
```

13. stat命令:

- 选项名称: `-stat`
- 使用格式: `-stat [format] <路径>`
- 作用: 显示文件统计信息

- 使用示例：本例中，-stat命令选项显示文件file1的一些统计信息，其中[format]可以填入不同的格式。此处的格式'%b %n %o %r %Y'依次表示文件大小、文件名称、块大小、副本数、访问时间。

```
phoenix@Master:~/桌面/example$ hdfs dfs -stat '%b %n %o %r %Y' file1
16 file1 134217728 3 1594536742375
```

14. tail命令：

- 选项名称：-tail
- 使用格式：-tail [-f] <文件>
- 作用：查看文件尾部信息
- 使用示例：该命令选项显示文件最后1K字节的内容。一般用于查看日志。如果带有选项-f，那么当文件内容变化时，也会自动显示。

```
phoenix@Master:~/桌面/example$ hdfs dfs -ls
Found 4 items
-rw-r--r-- 3 phoenix supergroup 16 2020-07-12 14:52 file1
drwxr-xr-x - phoenix supergroup 0 2020-07-12 10:52 input
drwxr-xr-x - phoenix supergroup 0 2020-07-12 15:15 new
drwx----- - phoenix2 phoenix2 0 2020-07-12 13:12 test
phoenix@Master:~/桌面/example$ hdfs dfs -tail file1
This is file 1.
```

15. test命令：

- 选项名称：-test
- 使用格式：-test [-defsz] <路径>
- 作用：某个文件或者目录是否存在
- 具体用法：-d return 0 if <路径> is a directory.
 -e return 0 if <路径> exists.
 -f return 0 if <路径> is a file.
 -s return 0 if file <路径> is greater than zero bytes in size.
 -z return 0 if file <路径> is zero bytes in size.
 else, return 1.
- 使用示例：本例中，我们先用-f询问，file1是否是文件？结果返回0，表示它是。再用-d询问，file1是否是目录？结果返回1，表示它不是。其他用法类似。

```
phoenix@Master:~/桌面/example$ hdfs dfs -test -f file1
phoenix@Master:~/桌面/example$ echo $?
0
phoenix@Master:~/桌面/example$ hdfs dfs -test -d file1
phoenix@Master:~/桌面/example$ echo $?
1
```

16. touchz命令：

- 选项名称：-test
- 使用格式：-touchz <文件路径>
- 作用：创建空白文件
- 使用示例：本例中，new文件夹原本为空。我们用touchz语句为它创建了emptyfile空文件，使用ls语句可以看到。

```
phoenix@Master:~/桌面/example$ hdfs dfs -ls new
phoenix@Master:~/桌面/example$ hdfs dfs -touchz new/emptyfile
phoenix@Master:~/桌面/example$ hdfs dfs -ls new
Found 1 items
-rw-r--r-- 2 phoenix supergroup 0 2020-07-12 15:15 new/emptyfile
```