HDFS的Shell操作

基本格式

1. hadoop fs 具体命令 (我主要用这个)
2. hdfs dfs 具体命令

常用命令大全

- 启动Hadoop集群

```
strat-dfs.sh
start-yarn.sh
```

- -help : 输出这个命令参数

```
hadoop fs -help ls
```

```
[hadoop@master root]$ hadoop fs -help ls
-ls [-d] [-h] [-R] [<path> ...] :
  List the contents that match the specified file pattern. If path is not
  specified, the contents of /user/<currentUser> will be listed. Directory entries
  are of the form:
        permissions - userId groupId sizeOfDirectory(in bytes)
  modificationDate(yyyy-MM-dd HH:mm) directoryName

  and file entries are of the form:
        permissions numberOfReplicas userId groupId sizeOfFile(in bytes)
  modificationDate(yyyy-MM-dd HH:mm) fileName

  -d  Directories are listed as plain files.
  -h  Formats the sizes of files in a human-readable fashion rather than a number
      of bytes.
  -R  Recursively list the contents of directories.
```

- -ls:显示目录信息

```
hadoop fs -ls /
hadoop fs -ls -R /    递归查看
```

```
[hadoop@master root]$ hadoop fs -ls -R /
-rw-r--r--   2 hadoop  supergroup    4990963 2020-09-15 19:42 /a.txt
drwxr-xr-x   - hadoop  supergroup          0 2020-09-09 19:40 /home
drwxr-xr-x   - hadoop  supergroup          0 2020-09-14 23:33 /home/hadoop
-rw-r--r--   3 hadoop  supergroup          6 2020-09-14 23:33 /home/hadoop/hahaha.txt
drwxr-xr-x   - hadoop  supergroup          0 2020-09-15 19:29 /home/hadoop/input
-rw-r--r--   3 hadoop  supergroup    4538523 2020-09-09 19:49 /home/hadoop/input/data.txt
-rw-r--r--   3 root    supergroup          6 2020-09-15 19:30 /home/hadoop/input/hahaha.txt
-rw-r--r--   3 hadoop: supergroup          0 2020-09-09 20:18 /input
drwxrwx---   - root    supergroup          0 2020-09-18 20:16 /mr-history
drwxrwx---   - root    supergroup          0 2020-09-18 20:16 /mr-history/done
drwxr-xr-x   - root    supergroup          0 2020-09-18 20:16 /mr-history/mapred
drwxrwxrwt   - root    supergroup          0 2020-09-18 20:16 /mr-history/mapred/tmp
drwxr-xr-x   - root    supergroup          0 2020-09-09 21:16 /output
drwxr-xr-x   - root    supergroup          0 2020-09-09 21:17 /output/temp
drwxr-xr-x   - root    supergroup          0 2020-09-17 19:45 /root
```

- mkdir: 在HDFS上创建目录

```
hadoop fs -mkdir -p /input/word_data
```

- -moveFromLocal: 本地剪切粘贴到HDFS

```
touch new_data.txt 创建文件
hadoop fs -moveFromLocal ./new_data.txt /input/
```

- -appendoFile: 追加一个文件到已经存在的文件末尾

```
echo "helloword" >> helloword.txt
hadoop fs -appendToFile ./helloword.txt /input/new_data.txt
```

- -cat: 显示文件内容

```
hadoop fs -cat /input/new_data.txt
```

- -chgrp、-chmod、-chown: Linux文件系统中的用法一样，修改文件所属权限

```
hadoop fs -chomd 777 /input
hadoop fs -chown hadoop:hadoop /input
```

- -copyFromLocal: 本地文件系统中拷贝文件到HDFS路径去

```
hadoop fs -copyFromLocal ./helloword.txt /input
```

- -copyToLocal: 从HDFS拷贝到本地文件

```
hadoop fs -copyToLocal /input/helloeord.txt ./
```

- -cp: 从HDFS的一个路径拷贝到HDFS的另一个路径

```
hadoop fs -cp /input/helloword.txt /output/helloword.txt
```

- -mv: 在HDFS目录中移动文件

```
hadoop fs -mv /input/new_data.txt /output/new_data.txt
```

- -get: 等同于copyToLocal，就是从HDFS下载文件到本地

```
hadoop fs -get /input/hellowword.txt ./
```

- getmerge : 合并下载多个文件，比如HDFS的目录 /input/下有多个文件：input.1,input.2...

```
hadoop fs -getmerge /input/* ./together.txt
```

- -put: 等同于copyFromLocal

```
hadoop fs -put ./together.txt /input
```

- -tail: 显示一个文件的末尾

```
hadoop fs -tail /input/helloword.txt
```

- -rm: 删除文件或文件夹

```
hadoop fs -rm /output/new_data.txt
```

- -rmdir: 删除空目录

```
hadoop fs -mkdir /test
```

- -du 统计文件夹的大小信息

```
hadoop fs -du -h /input/helloword.txt
```

- -setrep: 设置HDFS中文件的副本数量

```
hadoop fs -setrep 10 /input/helloword.txt
```