

# Hadoop-v2FsshellIntroduction

创建： 畅晋，最新修改： 系统管理员 于 2014-10-26

- 简介
- 命令详细介绍
  - mkdir
  - ls
  - lsr
  - put
  - copyFromLocal
  - moveFromLocal
  - get
  - copyToLocal
  - rm
  - rmr
  - test
  - du
  - dus
  - touchz
  - mv
  - cp
  - cat
  - tail
  - setrep
  - stat
  - getmerge
  - text
  - chmod
  - chown
  - chgrp
  - moveToLocal
  - count
  - expunge
  - help
- 限速

## 简介

HDFS（Hadoop Distributed Filesystem）提供一个类似于Unix Shell的命令接口FSShell，用于client和HDFS进行数据交互。

调用FSShell命令的方式：

**\$HADOOP\_HOME/bin/hadoop fs <args>**

其中，\$HADOOP\_HOME指hadoop所在的目录。所有的FS shell命令使用URI路径作为参数。URI格式是scheme://path。对HDFS文件系统，scheme是hdfs，对本地文件系统，scheme是file。其中scheme是可选的，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）。命令的出错信息会输出到stderr，其他信息输出到stdout。

## 命令详细介绍

下面是对常用命令的详细介绍（本文档只适用于Hadoop-v2版本，其他版本的命令使用及返回值存在差异）：

## mkdir

---

使用方法: `hadoop fs -mkdir <paths>`

接受路径指定的uri作为参数, 创建这些目录。其行为类似于Unix的`mkdir -p`, 它会创建路径中的各级父目录。

示例:

- `hadoop fs -mkdir /user/hadoop/dir1 /user/hadoop/dir2`
- `hadoop fs -mkdir hdfs://host1:port1/user/hadoop/dir hdfs://host2:port2/user/hadoop/dir`

返回值:

- 成功返回0, 失败返回255 (目录已存在; 失去与Namenode的连接或Namenode处于安全模式返回)。

## ls

---

使用方法: `hadoop fs -ls <args>`

如果是文件, 则按照如下格式返回文件信息:

权限 副本数 用户ID 组ID 文件大小 修改日期 修改时间 文件名

如果是目录, 则返回它直接子文件的一个列表, 就像在Unix中一样。目录返回列表的信息如下:

权限 副本数 用户ID 组ID 0 修改日期 修改时间 目录名

示例:

- `hadoop fs -ls /user/hadoop/file1 /user/hadoop/file2 ...`

返回值:

- 成功返回0, 失败返回255 (文件或目录不存在; 失去与Namenode的连接)。

## lsr

---

使用方法: `hadoop fs -lsr <args>`

ls命令的递归版本。类似于Unix中的`ls -R`。

## put

---

使用方法: `hadoop fs -put <localsrc> ... <dst>`

从本地文件系统中复制单个或多个源路径到目标文件系统; 也支持从标准输入中读取输入写入目标文件系统。

- `hadoop fs -put localfile /user/hadoop/hadoopfile`
- `hadoop fs -put localfile1 localdir1 /user/hadoop/hadoopdir`
- `hadoop fs -put localfile hdfs://host:port/hadoop/hadoopfile`
- `hadoop fs -put - hdfs://host:port/hadoop/hadoopfile` (从标准输入中读取输入, 以ctrl+D结束输入)

返回值:

- 成功返回0, 失败返回255 (源文件不存在或目标文件已存在; 失去与Namenode的连接或Namenode处于安全模式返回)。

## copyFromLocal

---

使用方法: `hadoop fs -copyFromLocal <localsrc> ... <dst>`

和put命令相似。

## moveFromLocal

---

使用方法: `dfs -moveFromLocal <localsrc> ... <dst>`

除了拷贝完成源路径将被删除外, 和`put`和`copyFromLocal`命令相似。

## get

---

使用方法: `hadoop fs -get [-ignorecrc] [-crc] <src> <localdst>`

复制文件到本地文件系统。可用`-ignorecrc`选项复制CRC校验失败的文件。使用`-crc`选项复制文件以及CRC信息。

示例:

- `hadoop fs -get /user/hadoop/file localfile`
- `hadoop fs -get hdfs://host:port/user/hadoop/file localfile`

返回值:

- 成功返回0, 失败返回255 (源文件不存在或目标文件已存在; 失去与Namenode的连接)。

## copyToLocal

---

使用方法: `hadoop fs -copyToLocal [-ignorecrc] [-crc] <src> <localdst>`

和`get`命令类似。

## rm

---

使用方法: `hadoop fs -rm URI [URI ...]`

删除指定的文件。只删除非空目录和文件。请参考`rmr`命令了解递归删除。

示例:

- `hadoop fs -rm hdfs://host:port/file /user/hadoop/emptydir`

返回值:

- 成功返回0, 失败返回255 (文件不存在; 删除目录; 失去与Namenode的连接)。

## rmr

---

使用方法: `hadoop fs -rmr URI [URI ...]`

删除的递归版本。

示例:

- `hadoop fs -rmr /user/hadoop/dir`
- `hadoop fs -rmr hdfs://host:port/user/hadoop/dir`

返回值:

- 成功返回0, 失败返回255 (文件或目录不存在; 失去与Namenode的连接)。

## test

---

使用方法: `hadoop fs -test [-ezd] URI`

选项:

- `-e` 检查文件是否存在。如果存在则返回0; 否则返回1。
- `-z` 检查文件是否是0字节。如果是则返回0; 否则返回1。
- `-d` 检查路径是否是目录。如果是则返回0; 否则返回1。

失去与Namenode的连接返回255。

示例:

- `hadoop fs -test -e filename`

## du

---

使用方法: `hadoop fs -du URI [URI ...]`

显示目录中所有文件的大小, 或者当只指定一个文件时, 显示此文件的大小。

示例:

- `hadoop fs -du /user/hadoop/dir1 /user/hadoop/file1 hdfs://host:port/user/hadoop/dir1`

返回值:

- 成功返回0, 失败返回255 (文件或目录不存在; 失去与Namenode的连接)。

## dus

---

使用方法: `hadoop fs -dus URI [URI ...]`

显示目录中所有文件的总大小, 而不会分别显示出目录中各个文件的大小; 当只指定一个文件时, 显示此文件的大小。

## touchz

---

使用方法: `hadoop fs -touchz URI [URI ...]`

创建一个0字节的空文件; 如果文件已存在则修改相应的mtime。

示例:

- `hadoop -touchz pathname`

返回值:

- 成功返回0, 失败返回255 (失去与Namenode的连接或Namenode处于安全模式)。

## mv

---

使用方法: `hadoop fs -mv URI [URI ...] <dest>`

将文件从源路径移动到目标路径。该命令允许有多个源路径, 此时目标路径必须是一个目录。不允许在不同的文件系统间移动文件。

示例:

- `hadoop fs -mv /user/hadoop/file1 /user/hadoop/file2`
- `hadoop fs -mv hdfs://host:port/file1 hdfs://host:port/file2 hdfs://host:port/file3 hdfs://host:port/dir1`

返回值:

- 成功返回0, 失败返回255 (源路径不存在或目标路径已存在; 失去与Namenode的连接或Namenode处于安全模式)。

## cp

---

使用方法: `hadoop fs -cp URI [URI ...] <dest>`

将文件从源路径复制到目标路径。该命令允许有多个源路径, 此时目标路径必须是一个目录。需要注意的是, 当目标文件已存在时, 仍然会拷贝并覆盖。

示例:

- `hadoop fs -cp /user/hadoop/file1 /user/hadoop/file2`

- `hadoop fs -cp /user/hadoop/file1 /user/hadoop/file2 /user/hadoop/dir`

返回值:

- 成功返回0，失败返回255（源路径不存在；失去与Namenode的连接或Namenode处于安全模式）。

---

## cat

使用方法: `hadoop fs -cat URI [URI ...]`

将路径指定文件的内容输出到stdout。

示例:

- `hadoop fs -cat hdfs://host1:port1/file1 hdfs://host2:port2/file2`
- `hadoop fs -cat file:///file3 /user/hadoop/file4`

返回值:

- 成功返回0，失败返回255（路径不是文件或不存在；失去与Namenode的连接）。

---

## tail

使用方法: `hadoop fs -tail [-f] URI`

将文件尾部1K字节的内容输出到stdout。支持-f选项，行为和Unix中一致。

示例:

- `hadoop fs -tail pathname`

返回值:

- 成功返回0，失败返回255（路径不是文件或不存在；失去与Namenode的连接）。

---

## setrep

使用方法: `hadoop fs -setrep [-R] [-w] [-d] <rep> <path>`

改变一个文件或目录的副本系数。

选项:

- **-R** 递归改变目录下所有文件的副本系数。
- **-w** 等待副本系数调整完毕后返回。
- **-d** 只修改目录的副本系数，而不修改其第一级文件的副本系数。

示例:

- `hadoop fs -setrep -R 3 /user/hadoop/dir1`

返回值:

- 成功返回0，失败返回255（路径不存在；失去与Namenode的连接或Namenode处于安全模式）。

---

## stat

使用方法: `hadoop fs -stat [format] URI`

返回指定路径的统计信息。

format选项:

- **%b** 统计指定文件（或目录）占用的数据块数目。
- **%n** 显示指定文件（或目录）的名称。
- **%o** 显示指定文件（或目录）一个数据块的大小。
- **%r** 显示指定文件（或目录）的副本系数。

- %y, %Y 显示指定文件（或目录）的修改时间。
- 示例：
- `hadoop fs -stat path`
- 返回值：
- 成功返回0，失败返回255（路径不存在；失去与Namenode的连接）。

## getmerge

使用方法：`hadoop fs -getmerge <src> <localdst>`

接受一个源目录和一个本地目标文件作为输入，并且将源目录中所有的文件连接成本地目标文件。

示例：

- `hadoop fs -getmerge mergedir localfile`

返回值：

- 成功返回0，失败返回255。

2.12.3版本后的getmerge进行了优化，通过多线程方式加快了getmerge执行速度。有以下参数控制getmerge的并发表现：

参数名	含义	默认值
getmerge.thread.num	getmerge并发线程数	10
getmerge.thread.buf.size	getmerge单线程缓存，单位Byte	1024*1024， 1M
getmerge.speed.limit	getmerge进程总体限速，单位Byte	30*1024*1024， 30M

经测试，getmerge.thread.buf.size的值比src文件中大部分稍大时，getmerge性能较优。

## text

使用方法：`hadoop fs -text <path>`

将文件输出为文本格式。允许的格式是zip和TextRecordInputStream。

## chmod

使用方法：`hadoop fs -chmod [-R] MODE URI [URI ...]`

改变文件的权限。使用-R让改变在目录结构下递归进行。命令的使用者必须是文件的所有者或者超级用户。

## chown

使用方法：`hadoop fs -chown [-R] [OWNER][:[GROUP]] URI [URI ...]`

改变文件的拥有者。使用-R让改变在目录结构下递归进行。命令的使用者必须是超级用户。

## chgrp

使用方法：`hadoop fs -chgrp [-R] GROUP URI [URI ...]`

改变文件所属的组。使用-R让改变在目录结构下递归进行。命令的使用者必须是文件的所有者或者超级用户。

## moveToLocal

使用方法: `dfs -moveToLocal [-crc] <src> <localdst>`

输出一个“not implemented”信息。

## count

---

使用方法: `dfs -count [-q] <path>`

如果没有指定-q, 输出<path>的如下信息:

目录数 文件数 长度 文件名

如果指定了-q, 则在上述信息之前, 输出quota信息, 格式如下:

quota数 quota剩余 spacequota数 spacequota剩余

正常返回0, 异常返回255。

## expunge

---

使用方法: `dfs -expunge`

触发对当前用户trash的一次checkpoint。

正常返回0, 异常返回255。

## help

---

更多命令信息, 可以通过-help命令查看。

## 限速

---

**FsShell**提供限速功能, 在命令行用-Dspeed.limit.kb=<限速值>、或在hadoop-site.xml中配置（[点这里看如何进行配置](#)）可以生效。

以下命令支持限速功能:

- put/copyFromLocal
- get/copyToLocal
- cp
- text

getmerge的限速功能由getmerge.speed.limit进行控制, 详情参考getmerge部分。

赞 成为第一个赞同者

无标签

## 评论

---



系统管理员 发表:

本页面由老系统（<http://wiki.babel.baidu.com/>）迁移而来, 原始页面请访问: [Hadoop-v2FsshellIntroduction](#), 迁移时间: 2014-10-25 13:05:34