**[Hadoop-2.0命令手册](http://www.cnblogs.com/xd502djj/p/3625799.html)**

1.  FS Shell

1.1     简介

调用文件系统(FS)Shell命令应使用 bin/hadoop fs <args>的形式。 所有的的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，不同之处会在下面介绍各命令使用详情时指出。出错信息会输出到stderr，其他信息输出到stdout。

1.2     常用命令使用

1)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-1。

2)chgrp

使用方法：hadoop fs -chgrp [-R]

改变文件所属的组。使用-R将使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。

3)chmod

使用方法：hadoop fs -chmod [-R] <MODE[,MODE]... | OCTALMODE> URI [URI …]

改变文件的权限。使用-R将使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。更多的信息请参见HDFS权限用户指南。

4)chown

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

改变文件的拥有者。使用-R将使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。更多的信息请参见HDFS权限用户指南。

5)copyFromLocal

使用方法：hadoop fs -copyFromLocal <localsrc> URI

除了限定源路径是一个本地文件外，和put命令相似。

6)copyToLocal

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

除了限定目标路径是一个本地文件外，和get命令类似。

7)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。这个命令允许有多个源路径，此时目标路径必须是一个目录。

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-1。

8)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-1。

9)dus

使用方法：hadoop fs -dus <args>

显示文件的大小。

10)expunge

使用方法：hadoop fs -expunge

清空回收站。请参考HDFS设计文档以获取更多关于回收站特性的信息。

11)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-1。

12)getmerge

使用方法：hadoop fs -getmerge <src> <localdst> [addnl]

接受一个源目录和一个目标文件作为输入，并且将源目录中所有的文件连接成本地目标文件。addnl是可选的，用于指定在每个文件结尾添加一个换行符。

13)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：

文件名 <副本数> 文件大小 修改日期 修改时间 权限 用户ID 组ID

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：

目录名 <dir> 修改日期 修改时间 权限 用户ID 组ID

示例：

hadoop fs -ls /user/hadoop/file1 /user/hadoop/file2 hdfs://host:port/user/hadoop/dir1 /nonexistentfile

返回值：

成功返回0，失败返回-1。

14)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。

15)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-1。

16)movefromLocal

使用方法：dfs -moveFromLocal <src> <dst>

输出一个”not implemented“信息。

17)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。这个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-1。

18)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 localfile2 /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。

返回值：

成功返回0，失败返回-1。

19)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-1。

20)rmr

使用方法：hadoop fs -rmr URI [URI …]

delete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-1。

21)setrep

使用方法：hadoop fs -setrep [-R] <path>

改变一个文件的副本系数。-R选项用于递归改变目录下所有文件的副本系数。

示例：

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

返回值：

成功返回0，失败返回-1。

22)  text

使用方法：hadoop fs -text <src>

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

2. jar

运行jar文件。用户可以把他们的Map Reduce代码捆绑到jar文件中，使用这个命令执行。

用法：hadoop jar <jar> [mainClass] args...

例如我们提交wordcount mapreduce的时候：

hadoop jar hadoop-examples.jar wordcount input output

3. archive

bin/hadoop archive

创建一个hadoop档案文件。参考 Hadoop Archives.

用法：hadoop archive -archiveName NAME <src>\* <dest>

-archiveName NAME   要创建的档案的名字。

src文件系统的路径名，和通常含正则表达的一样。

4.  distcp

bin/hadoop distcp

递归地拷贝文件或目录。参考DistCp指南以获取等多信息。

用法：hadoop distcp <srcurl> <desturl>

srcurl  源Url

desturl目标Url

daemonlog

获取或设置每个守护进程的日志级别。

用法：hadoop daemonlog -getlevel <host:port> <name>

用法：hadoop daemonlog -setlevel <host:port> <name> <level>

-getlevel <host:port> <name>    打印运行在<host:port>的守护进程的日志级别。这个命令内部会连接http://<host:port>/logLevel?log=<name>

-setlevel <host:port> <name> <level>    设置运行在<host:port>的守护进程的日志级别。这个命令内部会连接http://<host:port>/logLevel?log=<name>

5. bin/hdfs shell

dfs                  run a filesystem command on the file systems supported in Hadoop.

namenode -format     format the DFS filesystem

secondarynamenode    run the DFS secondary namenode

namenode             run the DFS namenode

zkfc                 run the ZK Failover Controller daemon

datanode             run a DFS datanode

dfsadmin             run a DFS admin client

haadmin              run a DFS HA admin client

fsck                 run a DFS filesystem checking utility

 balancer             run a cluster balancing utility

jmxget            get JMX exported values from NameNode or DataNode.

oiv                  apply the offline fsimage viewer to an fsimage

oev                  apply the offline edits viewer to an edits file

fetchdt              fetch a delegation token from the NameNode

getconf              get config values from configuration

groups               get the groups which users belong to

1)balancer

运行集群平衡工具。管理员可以简单的按Ctrl-C来停止平衡过程。参考Rebalancer了解更多。

用法：hadoop balancer [-threshold <threshold>]

-threshold <threshold>  磁盘容量的百分比。这会覆盖缺省的阀值。

2)datanode

运行一个HDFS的datanode。

用法：hadoop datanode [-rollback]

-rollback 将datanode回滚到前一个版本。这需要在停止datanode，分发老的hadoop版本之后使用。

3)dfsadmin

bin/hdfs dfsadmin

Usage: java DFSAdmin

Note: Administrative commands can only be run as the HDFS superuser.

           [-report]

           [-safemode enter | leave | get | wait]

           [-saveNamespace]

           [-restoreFailedStorage true|false|check]

           [-refreshNodes]

           [-finalizeUpgrade]

           [-upgradeProgress status | details | force]

           [-metasave filename]

           [-refreshServiceAcl]

           [-refreshUserToGroupsMappings]

           [-refreshSuperUserGroupsConfiguration]

           [-printTopology]

           [-refreshNamenodes datanodehost:port]

           [-deleteBlockPool datanode-host:port blockpoolId [force]]

           [-setQuota <quota> <dirname>...<dirname>]

           [-clrQuota <dirname>...<dirname>]

           [-setSpaceQuota <quota> <dirname>...<dirname>]

           [-clrSpaceQuota <dirname>...<dirname>]

           [-setBalancerBandwidth <bandwidth in bytes per second>]

           [-fetchImage <local directory>]

           [-help [cmd]]

参数说明：

-report报告文件系统的基本信息和统计信息。

-safemode enter | leave | get | wait    安全模式维护命令。安全模式是Namenode的一个状态，这种状态下，Namenode

1. 不接受对名字空间的更改(只读)

2. 不复制或删除块

Namenode会在启动时自动进入安全模式，当配置的块最小百分比数满足最小的副本数条件时，会自动离开安全模式。安全模式可以手动进入，但是这样的话也必须手动关闭安全模式。

-refreshNodes   重新读取hosts和exclude文件，更新允许连到Namenode的或那些需要退出或入编的Datanode的集合。

-finalizeUpgrade    终结HDFS的升级操作。Datanode删除前一个版本的工作目录，之后Namenode也这样做。这个操作完结整个升级过程。

-upgradeProgress status | details | force   请求当前系统的升级状态，状态的细节，或者强制升级操作进行。

-metasave filename  保存Namenode的主要数据结构到hadoop.log.dir属性指定的目录下的<filename>文件。对于下面的每一项，<filename>中都会一行内容与之对应

1. Namenode收到的Datanode的心跳信号

2. 等待被复制的块

3. 正在被复制的块

4. 等待被删除的块

-setQuota <quota> <dirname>...<dirname>为每个目录 <dirname>设定配额<quota>。目录配额是一个长整型整数，强制限定了目录树下的名字个数。

命令会在这个目录上工作良好，以下情况会报错：

1. N不是一个正整数，或者

2. 用户不是管理员，或者

3. 这个目录不存在或是文件，或者

4. 目录会马上超出新设定的配额。

-clrQuota <dirname>...<dirname>为每一个目录<dirname>清除配额设定。

命令会在这个目录上工作良好，以下情况会报错：

1. 这个目录不存在或是文件，或者

2. 用户不是管理员。

如果目录原来没有配额不会报错。

-help [cmd]显示给定命令的帮助信息，如果没有给定命令，则显示所有命令的帮助信息。

4)   namenode

运行namenode。有关升级，回滚，升级终结的更多信息请参考升级和回滚。

用法：namenode [-backup] | [-checkpoint] | [-format [-clusterid cid ] [-force] [-nonInteractive] ] | [-upgrade] | [-rollback] | [-finalize] | [-importCheckpoint] | [-initializeSharedEdits] | [-bootstrapStandby] | [-recover [ -force ] ]

-format格式化namenode。它启动namenode，格式化namenode，之后关闭namenode。

-upgrade    分发新版本的hadoop后，namenode应以upgrade选项启动。

-rollback   将namenode回滚到前一版本。这个选项要在停止集群，分发老的hadoop版本后使用。

-finalize   finalize会删除文件系统的前一状态。最近的升级会被持久化，rollback选项将再不可用，升级终结操作之后，它会停掉namenode。

-importCheckpoint   从检查点目录装载镜像并保存到当前检查点目录，检查点目录由fs.checkpoint.dir指定。