HDFS 2.0 Shell命令使用

讲师：董西成

博客：dongxicheng.org

微信号：hadoop-123

（二维码见右）

目录

HDFS Shell命令---概述

```
dongxicheng@dongxicheng-laptop:~/hadoop/hadoop-2.2.0$ bin/hdfs
Usage: hdfs [--config confdir] COMMAND
       where COMMAND is one of:
  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
  journalnode          run the DFS journalnode
  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
  snapshotDiff         diff two snapshots of a directory or diff the
                       current directory contents with a snapshot
  lsSnapshottableDir   list all snapshottable dirs owned by the current user
                                          Use -help to see options
  portmap              run a portmap service
  nfs3                 run an NFS version 3 gateway
```

HDFS Shell命令---文件操作命令

```
dongxicheng@dongxicheng-laptop:~/hadoop/hadoop-2.2.0$ bin/hdfs dfs
Usage: hadoop fs [generic options]
        [-appendToFile <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-copyFromLocal [-f] [-p] <localsrc> ... <dst>]
        [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-count [-q] <path> ...]
        [-cp [-f] [-p] <src> ... <dst>]
        [-createSnapshot <snapshotDir> [<snapshotName>]]
        [-deleteSnapshot <snapshotDir> <snapshotName>]
        [-df [-h] [<path> ...]]
        [-du [-s] [-h] <path> ...]
        [-expunge]
        [-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-getmerge [-nl] <src> <localdst>]
        [-help [cmd ...]]
        [-ls [-d] [-h] [-R] [<path> ...]]
        [-mkdir [-p] <path> ...]
        [-moveFromLocal <localsrc> ... <dst>]
        [-moveToLocal <src> <localdst>]
        [-mv <src> ... <dst>]
        [-put [-f] [-p] <localsrc> ... <dst>]
```

HDFS Shell命令---管理命令dfsadmin

```
dongxicheng@dongxicheng-laptop:~/hadoop/hadoop-2.2.0$ bin/hdfs dfsadmin
Usage: java DFSAdmin
Note: Administrative commands can only be run as the HDFS superuser.
        [-report]
        [-safemode enter | leave | get | wait]
        [-allowSnapshot <snapshotDir>]
        [-disallowSnapshot <snapshotDir>]
        [-saveNamespace]
        [-rollEdits]
        [-restoreFailedStorage true|false|check]
        [-refreshNodes]
        [-finalizeUpgrade]
        [-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>]
```

HDFS Shell命令---文件管理工具fsck

- ➤ 检查**hdfs**中文件的健康状况
- ➤ 查找缺失的块以及过少或过多复本的块
- ➤ 查看一个文件的所有数据块位置
- ➤ 删除损坏的数据块

```
dongxicheng@dongxicheng-laptop:~/hadoop/hadoop-2.2.0$ bin/hdfs fsck
Usage: DFSck <path> [-list-corruptfileblocks | [-move | -delete | -openforwrite] [-files [-blocks [-locations | -racks]]]]
        <path>  start checking from this path
        -move   move corrupted files to /lost+found
        -delete delete corrupted files
        -files  print out files being checked
        -openforwrite   print out files opened for write
        -list-corruptfileblocks print out list of missing blocks and files they belong to
        -blocks print out block report
        -locations      print out locations for every block
        -racks  print out network topology for data-node locations
                By default fsck ignores files opened for write, use -openforwrite to report such files. They are usually  tagg
lock allocation status

Generic options supported are
-conf <configuration file>      specify an application configuration file
-D <property=value>             use value for given property
-fs <local|namenode:port>       specify a namenode
```

6

HDFS Shell命令---文件管理工具fsck

```
dongxicheng@dongxicheng-laptop:~/hadoop/hadoop-2.2.0$ bin/hdfs fsck /home/dongxicheng/README.txt -files -blocks -locations
14/01/18 12:14:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
Connecting to namenode via http://YARN001:50070
FSCK started by dongxicheng (auth:SIMPLE) from /127.0.1.1 for path /home/dongxicheng/README.txt at Sat Jan 18 12:14:07 CST 2014
/home/dongxicheng/README.txt 1366 bytes, 1 block(s):  OK
0. BP-1658977241-127.0.1.1-1389428745643:blk_1073741825_1001 len=1366 repl=1 [127.0.1.1:50010]

Status: HEALTHY
 Total size:    1366 B
 Total dirs:    0
 Total files:   1
 Total symlinks:              0
 Total blocks (validated):    1 (avg. block size 1366 B)
 Minimally replicated blocks: 1 (100.0 %)
 Over-replicated blocks:      0 (0.0 %)
 Under-replicated blocks:     0 (0.0 %)
 Mis-replicated blocks:       0 (0.0 %)
 Default replication factor:  1
 Average block replication:   1.0
```

➢ 数据块重分布

- ✓ **sbin/start-balancer.sh -threshold <percentage of disk capacity>**

➢ **percentage of disk capacity**

- ✓ **HDFS**达到平衡状态的磁盘使用率偏差值
- ✓ 值越低各节点越平衡，但消耗时间也更长

加入新datanode/移除旧的datanode

- ➢ **加入新的datanode**
  - ✓ 步骤1：将已存在datanode上的安装包（包括配置文件等）拷贝到新datanode上；
  - ✓ 步骤2：启动新datanode：

  sbin/hadoop-deamon.sh start datanode

- ➢ **移除旧datanode**
  - ✓ 步骤1：将datanode加入黑名单，并更新黑名单，在NameNode上，将datanode的host或者ip加入配置选项dfs.hosts.exclude指定的文件中
  - ✓ 步骤2：移除datanode

    bin/hadoop dfsadmin -refreshNodes

总结

- **dfs**

- **dfsadmin**

- **fsck**

- **balancer**