

Hadoop 分布式文件系统操作

上海交通大学 OmniLab

<http://omnilab.sjtu.edu.cn>

2013 年 11 月 21 日更新

目录

1	基本命令	2
2	管理员命令	4
3	使用 alias 简化命令	4
4	参考资料	4

1 基本命令

操作 Hadoop 分布式文件系统 HDFS 的基本命令是 `hdfs`，命令行参数如下：

```
$ 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

                        Use -help to see options
```

`hdfs` 命令的子功能很多，如 `dfs` 处理一般文件系统命令，`fsck` 检查文件系统完整性，`dfsadmin` 用于文件系统管理。

本教程着重向用户介绍 HDFS 的一般用法，以 `dfs` 子命令为主要内容。HDFS 的管理在其他文档介绍。

```
$ hdfs dfs
Usage: hadoop fs [generic options]
      [-cat [-ignoreCrc] <src> ...]
      [-chgrp [-R] GROUP PATH...]
      [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
      [-chown [-R] [OWNER][:[GROUP]] PATH...]
      [-copyFromLocal <localsrc> ... <dst>]
      [-copyToLocal [-ignoreCrc] [-crc] <src> ... <localdst>]
      [-count [-q] <path> ...]
```

```

[-cp <src> ... <dst>]
[-df [-h] [<path> ...]]
[-du [-s] [-h] <path> ...]
[-expunge]
[-get [-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 <localsrc> ... <dst>]
[-rm [-f] [-r|-R] [-skipTrash] <src> ...]
[-rmdir [--ignore-fail-on-non-empty] <dir> ...]
[-setrep [-R] [-w] <rep> <path/file> ...]
[-stat [format] <path> ...]
[-tail [-f] <file>]
[-test [-ezd] <path>]
[-text [-ignoreCrc] <src> ...]
[-touchz <path> ...]
[-usage [cmd ...]]

```

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
-jt <local jobtracker:port>	specify a job tracker
-files <comma separated list of files>	specify comma separated files to be copied to the map re
-libjars <comma separated list of jars>	specify comma separated jar files to include in the cla
-archives <comma separated list of archives>	specify comma separated archives to be unarchived

The general command line syntax is

```
bin/hadoop command [genericOptions] [commandOptions]
```

2 管理员命令

3 使用 `alias` 简化命令

4 参考资料

- “Hadoop MapReduce Tutorial v1.2.1” http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html