1. HDFS的概念和特性

首先，它是一个文件系统，用于存储文件，通过统一的命名空间——目录树来定位文件

其次，它是分布式的，由很多服务器联合起来实现其功能，集群中的服务器有各自的角色；

重要特性如下：

（1）HDFS中的文件在物理上是分块存储（block），块的大小可以通过配置参数( dfs.blocksize)来规定，默认大小在hadoop2.x版本中是128M，老版本中是64M

（2）HDFS文件系统会给客户端提供一个统一的抽象目录树，客户端通过路径来访问文件，形如：hdfs://namenode:port/dir-a/dir-b/dir-c/file.data

（3）目录结构及文件分块信息(元数据)的管理由namenode节点承担

——namenode是HDFS集群主节点，负责维护整个hdfs文件系统的目录树，以及每一个路径（文件）所对应的block块信息（block的id，及所在的datanode服务器）

（4）文件的各个block的存储管理由datanode节点承担

---- datanode是HDFS集群从节点，每一个block都可以在多个datanode上存储多个副本（副本数量也可以通过参数设置dfs.replication）

（5）联邦HDFS,解决单个namenode的内存限制整个集群的横向扩展

HDFS是设计成适应一次写入，多次读出的场景，且不支持文件的修改

1. Shell操作

1.参数：

[-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>]

[-getfacl [-R] <path>]

[-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>]

[-renameSnapshot <snapshotDir> <oldName> <newName>]

[-rm [-f] [-r|-R] [-skipTrash] <src> ...]

[-rmdir [--ignore-fail-on-non-empty] <dir> ...]

[-setfacl [-R] [{-b|-k} {-m|-x <acl\_spec>} <path>]|[--set <acl\_spec> <path>]]

[-setrep [-R] [-w] <rep> <path> ...]

[-stat [format] <path> ...]

[-tail [-f] <file>]

[-test -[defsz] <path>]

[-text [-ignoreCrc] <src> ...]

[-touchz <path> ...]

[-usage [cmd ...]]

1. 参数详解

|  |
| --- |
| -help  功能：输出这个命令参数手册 |
| **-ls**  **功能：显示目录信息**  *示例： hadoop fs -ls hdfs://hadoop-server01:9000/*  *备注：这些参数中，所有的hdfs路径都可以简写*  *-->hadoop fs -ls / 等同于上一条命令的效果* |
| **-mkdir**  **功能：在hdfs上创建目录**  *示例：hadoop fs -mkdir -p /aaa/bbb/cc/dd* |
| **-moveFromLocal**  **功能：从本地剪切粘贴到hdfs**  *示例：hadoop fs - moveFromLocal /home/hadoop/a.txt /aaa/bbb/cc/dd*  **-moveToLocal**  **功能：从hdfs剪切粘贴到本地**  *示例：hadoop fs - moveToLocal /aaa/bbb/cc/dd /home/hadoop/a.txt* |
| **--appendToFile**  **功能：追加一个文件到已经存在的文件末尾**  *示例：hadoop fs -appendToFile ./hello.txt hdfs://hadoop-server01:9000/hello.txt*  *可以简写为：*  *Hadoop fs -appendToFile ./hello.txt /hello.txt* |
| **-cat**  **功能：显示文件内容**  *示例：hadoop fs -cat /hello.txt*  **-tail**  **功能：显示一个文件的末尾**  *示例：hadoop fs -tail /weblog/access\_log.1*  **-text**  **功能：以字符形式打印一个文件的内容**  *示例：hadoop fs -text /weblog/access\_log.1* |
| **-chgrp**  **-chmod**  **-chown**  **功能：linux文件系统中的用法一样，对文件所属权限**  *示例：*  *hadoop fs -chmod 666 /hello.txt*  *hadoop fs -chown someuser:somegrp /hello.txt* |
| **-copyFromLocal**  **功能：从本地文件系统中拷贝文件到hdfs路径去**  *示例：hadoop fs -copyFromLocal ./jdk.tar.gz /aaa/*  **-copyToLocal**  **功能：从hdfs拷贝到本地**  *示例：hadoop fs -copyToLocal /aaa/jdk.tar.gz* |
| **-cp**  **功能：从hdfs的一个路径拷贝hdfs的另一个路径**  *示例： hadoop fs -cp /aaa/jdk.tar.gz /bbb/jdk.tar.gz.2*  **-mv**  **功能：在hdfs目录中移动文件**  *示例： hadoop fs -mv /aaa/jdk.tar.gz /* |
| **-get**  **功能：等同于copyToLocal，就是从hdfs下载文件到本地**  示例：hadoop fs -get /aaa/jdk.tar.gz  **-getmerge**  **功能：合并下载多个文件**  *示例：比如hdfs的目录 /aaa/下有多个文件:log.1, log.2,log.3,...*  hadoop fs -getmerge /aaa/log.\* ./log.sum |
| **-put**  **功能：等同于copyFromLocal**  *示例：hadoop fs -put /aaa/jdk.tar.gz /bbb/jdk.tar.gz.2* |
| **-rm**  **功能：删除文件或文件夹**  *示例：hadoop fs -rm -r /aaa/bbb/*  **-rmdir**  **功能：删除空目录**  *示例：hadoop fs -rmdir /aaa/bbb/ccc* |
| **-df**  **功能：统计文件系统的可用空间信息**  *示例：hadoop fs -df -h /*  **-du**  **功能：统计文件夹的大小信息**  *示例：*  *hadoop fs -du -s -h /aaa/\** |
| **-count**  **功能：统计一个指定目录下的文件节点数量**  *示例：hadoop fs -count /aaa/* |
| **-setrep**  **功能：设置hdfs中文件的副本数量**  *示例：hadoop fs -setrep 3 /aaa/jdk.tar.gz* |