

Hadoop集群用户的常用命令

hadoop archive l 创建一个名为NAME的hadoop档案文件.

命令选项	描述
-archiveName NAME	要创建档案的名字
-src	文件系统的路径名
-dest	保存档案文件的目标目录

hadoop distcp <srcurl> <desturl> | 递归地拷贝文件或目录，srcurl为源Url，desturl为目标Url.

hadoop fsck | 运行HDFS文件系统检查工具.

命令选项	描述
<path>	检查的起始目录
-move	移动受损文件
-delete	删除受损文件
-openforwrite	打印写入的文件
-files	打印正被检查的文件
-blocks	打印模块信息报告
-locations	打印每个模块的位置信息
-racks	打印出data-node的网络结构

hadoop pipes | 运行pipes作业.

命令选项	描述
-conf <path>	作业的配置
-input <path>	输入目录
-output <path>	输出目录
-jar <jar file>	Jar文件名
-inputformat <class>	InputFormat类
-map <class>	Java Map类
-partitioner <class>	Java Partitioner
-reduce <class>	Java Reduce类
-writer <class>	Java RecordWriter
-program <executable>	可执行程序的URI
-reduces <num>	reduce个数

hadoop fs | 运行一个常规的文件系统客户端.

hadoop jar <jar> [mainClass] args... | 运行jar文件.

hadoop version | 打印版本信息.

hadoop CLASSNAME | 运行名字为CLASSNAME的类.

hadoop job | 用于和Map Reduce作业交互和命令.

命令选项	描述
-submit <job-file>	提交任务
-status <job-id>	打印map和reduce完成度和计数器
-counter <job-id>	打印计数器的值
-kill <job-id>	终止指定任务
-history <jobOutputDir>	打印任务的细节、失败及被杀死原因
-list [all]	显示所有任务
-kill-task <task-id>	终止任务
-fail-task <task-id>	使任务失败

Hadoop集群管理员常用命令

hadoop balancer -threshold <threshold> | 运行集群平衡工具（管理员可以通过Ctrl加C来停止平衡过程）.

hadoop daemonlog -getlevel <host:port> <name> | 获取每个守护进程的日志级别.

hadoop daemonlog -setlevel <host:port> <name> | 设置每个守护进程的日志级别.

hadoop secondarynamenode | 运行HDFS的secondary namenode.

命令选项	描述
-checkpoint	当EditLog>= fs.checkpoint.size，启动检查点过程
-geteditsize	打印EditLog大小

hadoop namenode | 运行namenode.

命令选项	描述
-format	格式化namenode(启动-格式化-关闭)
-upgrade	namenode以upgrade选项启动
-rollback	将namenode回滚到前一版本
-finalize	删除文件系统的前一状态
-importCheckpoint	从检查点目录装载镜像并保存到当前检查点目录

hadoop tasktracker | 运行MapReduce的task Tracker节点.

hadoop jobtracker | 运行MapReduce job Tracker节点.

hadoop datanode -rollback | 运行一个HDFS的datanode, 将datanode回滚到前一个版本.

hadoop dfsadmin | 运行一个HDFS的dfsadmin客户端.

命令选项	描述
-report	报告文件系统的基本和统计信息
-safemode enter	进入管理安全模式
-safemode leave	离开安全模式
-refreshNodes	重新读取hosts和exclude文件
-finalizeUpgrade	完结HDFS的升级操作
-setQuota <dirname>	设置每个目录<dirname>配额设定
-clrQuota <dirname>	清除每个目录<dirname>配额设定
-metasave filename	保存Namenode到hadoop.log.dir目录下
-help [cmd]	显示给定命令的帮助信息

Hadoop Shell常用命令

URI表示文件、目录的路径，可以是相对路径或绝对路径.

hadoop fs -cat URI [URI ...] | 查看fs文件内容.

例如：**hadoop fs -cat /home/hadoop/input/content.txt**

hdfs dfs -cp URI [URI ...] | 将文件从源路径复制到目标路径.

hadoop fs -du URI [URI ...] | 显示目录中所有文件的大小.

hadoop fs -expunge | 清空回收站.

hadoop fs -ls <path> | 返回文件path的统计信息.

hadoop fs -df <path> | 显示Hadoop所使用的文件系统的大小.

hadoop fs -mv <src> <dst> | 从src移动到dst，允许多个源移动到同一个dst，dst必须是目录.

hadoop fs -rm [-skipTrash] <path> | 删除文件，不能删除目录，-skipTrash：直接删除文件.

hadoop fs -mkdir <path> | 创建path文件夹，如果path的父目录不存在，会迭代创建.

hadoop fs -test [-ezd] <path> | 测试文件的目录属性，-e：测试文件是否存在，-z：文件大小是否为0，-d：测试是否是目录.

hadoop fs -getmerge <src> <localdst> [addnl] | 合并src和dst文件，addnl：会在每行结尾添加newline字符.

