

Hadoop HDFS Cheatsheet

Connection

hadoop fs -ls <path>
List files and directories in HDFS

hdfs dfs -ls <path>
Alternative syntax to list files and directories

hdfs getconf -confKey <key>
Get configuration information

File Operations

hadoop fs -put <local-path> <hdfs-path>
Upload files from local to HDFS

hadoop fs -get <hdfs-path> <local-path>
Download files from HDFS to local

hadoop fs -cp <source-path> <dest-path>
Copy files within HDFS

hadoop fs -mv <source-path> <dest-path>
Move files within HDFS

hadoop fs -rm <file-path>
Delete a file from HDFS

hadoop fs -rm -r <directory-path>
Delete a directory recursively

hadoop fs -mkdir <directory-path>
Create a directory in HDFS

hadoop fs -cat <file-path>
View file content

hadoop fs -tail <file-path>
View the last part of a file

hadoop fs -touchz <file-path>
Create an empty file

File System Status

hadoop fs -df -h
Show disk usage with human-readable format

hadoop fs -du -h <path>
Show disk usage for specified files/directories

hadoop fs -count <path>
Count files in directory

hadoop fsck <path>
Check HDFS file system health

hdfs dfsadmin -report
Get overall HDFS status

hdfs dfsadmin -safemode enter|leave|get|wait
Manage safe mode

Permissions & Ownership

hadoop fs -chmod <permissions> <path>
Change file permissions

hadoop fs -chown <owner>:<group> <path>
Change file ownership

hadoop fs -chgrp <group> <path>
Change file group

Quotas & Snapshots

hdfs dfsadmin -setQuota <quota> <path>
Set file count quota

hdfs dfsadmin -setSpaceQuota <quota> <path>
Set space quota

hdfs dfsadmin -clrQuota <path>
Clear file count quota

hdfs dfsadmin -clrSpaceQuota <path>
Clear space quota

hdfs dfs -createSnapshot <path>
Create snapshot

hdfs dfs -deleteSnapshot <path> <name>
Delete snapshot

Cluster Management

hdfs haadmin -getServiceState <serviceId>
Get HA state of namenode

hdfs haadmin -failover <serviceId1> <serviceId2>
Initiate failover

hadoop job -list
List running jobs

hadoop job -kill <job-id>
Kill a running job