

HDFS Commands: A Hadoop Storage File System

The Hadoop Distributed File System (HDFS) is a distributed, scalable, and portable file system written in Java for the Hadoop framework. It is designed to store very large files across multiple machines in a reliable and fault-tolerant manner. HDFS is a crucial component of Hadoop, providing high-throughput access to application data and is suitable for applications with large data sets.

SA by Sarika Alladi

HADOOP HDFS COMMANDS CHEATSHEET

ories for the given hdfs destination path
as plain files. In this case, this command will list the details of
human-readable fashion (eg 64.0m instead of 67108864)
s in hadoop directory and all subdirectories in hadoop

ing the pattern. In this case, it will list all the files inside
ch starts with 'dat'

takes a source file and outputs the file in text format on the
formats are zip and TextRecordInputStream
isplay the content of the HDFS file test on your stdout
of a local file test1 to a hdfs file test2

ical file system to HDFS

ical file system to HDFS, and in case the local already exists in
path, using -f option with put command will overwrite it
ical file system to HDFS. Allow DataNode to lazily persist the
lication factor of 1

ical file system to HDFS. Passing -p preserves access and
wnership and the mode
DFS to local file system

DFS to local file system. Passing -p preserves access and
wnership and the mode
tching the pattern from local file system to HDFS

put command, except that the source is restricted to a local

put command, except that the destination is restricted to a

put command, except that the source is deleted after it's

e to destination on HDFS. In this case, copying file1 from
adoop1 directory

e to destination on HDFS. Passing -p preserves access and
wnership and the mode

hdfs dfs -cp -f /hadoop/file1 /hadoop1	Copies file from destination wh
hdfs dfs -mv /hadoop/file1 /hadoop1	Move files tha
	When moving
hdfs dfs -rm /hadoop/file1	Deletes the file
hdfs dfs -rm -r /hadoop	Deletes the di
hdfs dfs -rm -R /hadoop	
hdfs dfs -rmr /hadoop	
hdfs dfs -rm -skipTrash /hadoop	The -skipTrash immediately
hdfs dfs -rm -f /hadoop	If the file does status to refle
hdfs dfs -rmdir /hadoop1	Delete a direct
hdfs dfs -mkdir /hadoop2	Create a direct
hdfs dfs -mkdir -f /hadoop2	Create a direct the directory a
hdfs dfs -touchz /hadoop3	Creates a file o <path>

OWNERSHIP AND VALIDATION

hdfs dfs -checksum /hadoop/file1	Dump checksu
hdfs dfs -chmod 755 /hadoop/file1	Changes perm
hdfs dfs -chmod -R 755 /hadoop	Changes perm
hdfs dfs -chown ubuntu:ubuntu /hadoop	Changes owne group
hdfs dfs -chown -R ubuntu:ubuntu /hadoop	Changes owne
hdfs dfs -chgrp ubuntu /hadoop	Changes group
hdfs dfs -chgrp -R ubuntu /hadoop	Changes group

FILESYSTEM

hdfs dfs -df /hadoop	Shows the cap
hdfs dfs -df -h /hadoop	Shows the cap
	Formats the si
hdfs dfs -du /hadoop/file	Show the amo file pattern
hdfs dfs -du -s /hadoop/file	Rather than sh shows the tota
hdfs dfs -du -h /hadoop/file	Show the amo specified file p

ADMINISTRATION

hdfs balancer -threshold 30	Runs a cluster default thresh
hadoop version	To check the v
hdfs fsck /	It checks the h
hdfs dfsadmin -safemode leave	The command
hdfs dfsadmin -refreshNodes	Re-read the h
	allowed to cor
	decommission
hdfs namenode -format	Formats the N

Create a directory structure in HDFS

1

Step 1: Accessing HDFS

In order to create a directory structure in HDFS, you first need to access the Hadoop Distributed File System using the appropriate commands and permissions.

2

Step 2: Creating Directories

Once inside the HDFS, use the command `hadoop fs -mkdir -p /user/your_username` to create a directory structure, ensuring the `-p` flag is used for creating parent directories.

3

Step 3: Verifying the Structure

After creating the directory structure, it's essential to verify that the directories have been created as intended. This can be done using the appropriate commands to list and navigate the HDFS.