

HADOOP IN CONTAINER WITH PERSISTENT STORAGE AND CONFIGURATION OF HDFS FILESYSTEM

docker --version

```
root@kk:~# docker --version
Docker version 26.1.3, build 26.1.3-0ubuntu1~24.04.1
root@kk:~#
```

Create a volume in docker to use as persistent storage

sudo docker volume create --driver local --opt type=none --opt device=dockerVol --opt o=bind dockerVol

```
kk@kk:~$ sudo docker volume inspect dockerVol
[
  {
    "CreatedAt": "2025-02-13T11:15:10Z",
    "Driver": "local",
    "Labels": null,
    "Mountpoint": "/var/lib/docker/volumes/dockerVol/_data",
    "Name": "dockerVol",
    "Options": {
      "device": "/dockerVol",
      "o": "bind",
      "type": "none"
    },
    "Scope": "local"
  }
]
kk@kk:~$
```

Pull Hadoop image

docker pull westernscience/hadoop

```
root@kk:~#
root@kk:~# docker pull westernscience/hadoop
Using default tag: latest
latest: Pulling from westernscience/hadoop
e8091032bc12: Downloading [==>] 3.22MB/70.98MB
297e64efd81b: Downloading [===>] 1.786MB/29.13MB
9b6f815e70af: Download complete
6d454f4f220e: Waiting
a6f5bfd4f656: Waiting
8cf8a72f957c: Waiting
747f4410e9cf: Waiting
```

Run container with persistent storage and proper port mapping

```
# docker run -d --name=kk_hadoop -v /dockerVol:/usr/local/hdfs/datastore -p 9870:9870 westernscience/Hadoop
```

Interactive shell

```
# docker exec -it kk_hadoop /bin/bash
```

```
root@kk:~# docker exec -it kk_hadoop /bin/bash
[root@f1fbc8e02893 /]# ls
bin boot dev etc home lib lib64 lost+found media mnt opt proc root run sbin srv sys tmp usr var
[root@f1fbc8e02893 /]#
```

Add Hadoop binary path as PATH environment variables

```
# vi ~/.bashrc
```

```
export PATH=$PATH:$HADOOP_HOME/bin
```

```
# source ~/.bashrc
```

```
[root@f1fbc8e02893 /]# export PATH=$PATH:$HADOOP_HOME/bin
[root@f1fbc8e02893 /]# hadoop version
Hadoop 3.3.0
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r aa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled by brahma on 2020-07-06T18:44Z
Compiled with protoc 3.7.1
From source with checksum 5dc29b802d6ccd77b262ef9d04d19c4
This command was run using /usr/local/hadoop-3.3.0/share/hadoop/common/hadoop-common-3.3.0.jar
[root@f1fbc8e02893 /]#
```

Persistence volume inside container

```
[root@256blad4e301 /]# hadoop version
Hadoop 3.3.0
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r aa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled by brahma on 2020-07-06T18:44Z
Compiled with protoc 3.7.1
From source with checksum 5dc29b802d6ccd77b262ef9d04d19c4
This command was run using /usr/local/hadoop-3.3.0/share/hadoop/common/hadoop-common-3.3.0.jar
[root@256blad4e301 /]#
[root@256blad4e301 /]# ls /usr/local/hadoop/hdfs
datastore
[root@256blad4e301 /]#
```

HDFS Filesystem configuration

hdfs dfs -ls /

```
[root@a2745e82559e /]# hdfs dfs -ls /
Found 2 items
drwxrwx--- - root supergroup 0 2025-02-13 11:35 /tmp
drwxr-xr-x - root supergroup 0 2025-02-13 11:35 /user
[root@a2745e82559e /]#
```

hdfs dfs -mkdir /MDS

```
[root@256blad4e301 /]# hdfs dfs -mkdir /MDS
[root@256blad4e301 /]# hdfs dfs -ls /
Found 3 items
drwxr-xr-x - root supergroup 0 2025-02-14 06:58 /MDS
drwxrwx--- - root supergroup 0 2025-02-14 06:40 /tmp
drwxr-xr-x - root supergroup 0 2025-02-14 06:40 /user
[root@256blad4e301 /]#
```

hdfs dfs -touchz /MDS/Bigdata.txt

```
[root@256blad4e301 /]# hdfs dfs -touchz /MDS/Bigdata.txt
[root@256blad4e301 /]# hdfs dfs -ls /MDS
Found 1 items
-rw-r--r-- 1 root supergroup 0 2025-02-14 07:00 /MDS/Bigdata.txt
[root@256blad4e301 /]#
```

Access DataNode Dashboard

https://<IP_address>:9870

Browse Directory

Show entries

Search:

| <input type="checkbox"/> | Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name | |
|--------------------------|----------------------------|----------------------|----------------------------|------|---------------|-------------------|------------|-----------------------------|--|
| <input type="checkbox"/> | -rw-r--r-- | root | supergroup | 0 B | Feb 14 12:45 | 1 | 128 MB | Bigdata.txt | |

Showing 1 to 1 of 1 entries

hdfs dfs -copyFromLocal /kk.txt /MDS/

```
[root@256blad4e301 /]# hdfs dfs -copyFromLocal /kk.txt /MDS/
[root@256blad4e301 /]# hdfs dfs -ls /MDS
Found 2 items
-rw-r--r--  1 root supergroup          0 2025-02-14 07:00 /MDS/Bigdata.txt
-rw-r--r--  1 root supergroup        43 2025-02-14 07:30 /MDS/kk.txt
[root@256blad4e301 /]#
```

hdfs dfs -cat /MDS/kk.txt

```
[root@256blad4e301 /]# hdfs dfs -cat /MDS/kk.txt
Hello, My name is kaushal khatiwada
MDS
13
[root@256blad4e301 /]#
```

hdfs dfs -copyToLocal /MDS/Bigdata.txt /

```
[root@256blad4e301 /]# hdfs dfs -copyToLocal /MDS/Bigdata.txt /
[root@256blad4e301 /]# ls /
Bigdata.txt  bin  boot  dev  etc  home  kk.txt  lib  lib64  lost+found
[root@256blad4e301 /]#
```

hdfs dfs -mv /MDS/kk.txt /MDS/13.txt

```
[root@256blad4e301 /]# hdfs dfs -mv /MDS/kk.txt /MDS/13.txt
[root@256blad4e301 /]# hdfs dfs -ls /MDS
Found 2 items
-rw-r--r--  1 root supergroup        43 2025-02-14 07:30 /MDS/13.txt
-rw-r--r--  1 root supergroup          0 2025-02-14 07:00 /MDS/Bigdata.txt
[root@256blad4e301 /]#
```

hdfs dfs -du /MDS

```
[root@256blad4e301 /]# hdfs dfs -du /MDS
43 43 /MDS/13.txt
0  0 /MDS/Bigdata.txt
[root@256blad4e301 /]#
```

hdfs dfs -rm /MDS/13.txt

```
[root@256blad4e301 /]# hdfs dfs -rm /MDS/13.txt
Deleted /MDS/13.txt
[root@256blad4e301 /]# hdfs dfs -ls /MDS
Found 1 items
-rw-r--r--  1 root supergroup          0 2025-02-14 07:00 /MDS/Bigdata.txt
[root@256blad4e301 /]#
```