Experiment No 1

Aim: Study and Installation of Hadoop Ecosystem

Objective:

The objective of this lab experiment is to familiarize students with the Hadoop ecosystem by guiding them through the installation and setup of core components. Students will gain hands-on experience in configuring a basic Hadoop cluster, understanding its architecture, and verifying its functionality.

Tools and Technologies:

- Hadoop: A framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models.
- Hadoop Ecosystem Components: HDFS (Hadoop Distributed File System), YARN (Yet Another Resource Negotiator), and MapReduce.

Pre-requisites:

- Basic understanding of Linux/Unix commands.
- Familiarity with Java programming (helpful but not mandatory).

Equipment Required:

- Virtual or physical machines capable of running a Linux distribution (e.g., Ubuntu, CentOS).
- Sufficient memory and disk space to accommodate Hadoop's requirements (minimum of 4GB RAM recommended per node).

Experiment Steps:

1. **Setting Up the Environment:**
    - Prepare the environment by setting up virtual machines (VMs) or physical machines with a Linux distribution (e.g., Ubuntu Server).
    - Ensure that each machine has a static IP address and can communicate with each other over the network.
2. **Installing Java Development Kit (JDK):**
    - Hadoop requires Java, so install JDK on all machines that will be part of the Hadoop cluster.
    - Example command to install OpenJDK:

      bash
      Copy code
      sudo apt-get update
      sudo apt-get install openjdk-8-jdk
3. **Downloading and Extracting Hadoop:**
    - Download the desired version of Hadoop from the Apache Hadoop website (https://hadoop.apache.org/releases.html).
    - Extract the downloaded Hadoop tarball to a suitable directory on each machine in your cluster.

      bash
      Copy code
      tar -xzvf hadoop-3.x.x.tar.gz -C /opt

4. **Configuring Hadoop Environment Variables:**
    - Set up Hadoop environment variables in the .bashrc or .bash_profile file for each user:

      bash
      Copy code
      export HADOOP_HOME=/opt/hadoop-3.x.x
      export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
5. **Configuring Hadoop Cluster:**
    - **HDFS Configuration:**
        - Edit core-site.xml to configure Hadoop core settings, including HDFS filesystem URI and default filesystem.
        - Edit hdfs-site.xml to define HDFS block size, replication factor, and namenode/datanode directories.
    - **YARN Configuration:**
        - Edit yarn-site.xml to configure YARN ResourceManager and NodeManager settings.
        - Optionally, configure mapred-site.xml for MapReduce framework settings if not managed by YARN.
    - **Setup SSH Authentication:**
        - Enable SSH access between nodes without requiring a password for seamless communication.
        - Generate SSH keys (ssh-keygen) and distribute the public key (ssh-copy-id) to each node.
6. **Starting Hadoop Cluster:**
    - Format the HDFS filesystem on the namenode:

      bash
      Copy code
      hdfs namenode -format

    - Start Hadoop daemons using the provided scripts:

      bash
      Copy code
      start-dfs.sh
      start-yarn.sh
7. **Verifying Hadoop Installation:**
    - Access the Hadoop web interfaces:
        - HDFS Namenode: http://namenode_host:9870/
        - YARN ResourceManager: http://resourcemanager_host:8088/
    - Run basic Hadoop commands to ensure functionality:

      bash
      Copy code
      hdfs dfs -ls / # List contents of root directory in HDFS
      yarn node -list # List nodes in the YARN cluster

8. **Performing a Simple MapReduce Job (Optional):**

    - Write a basic MapReduce program (e.g., WordCount) or use a pre-existing example.
    - Compile and package the program into a JAR file.
    - Submit the job to the YARN ResourceManager and monitor its progress using the web interface.

9. **Observations and Conclusion:**

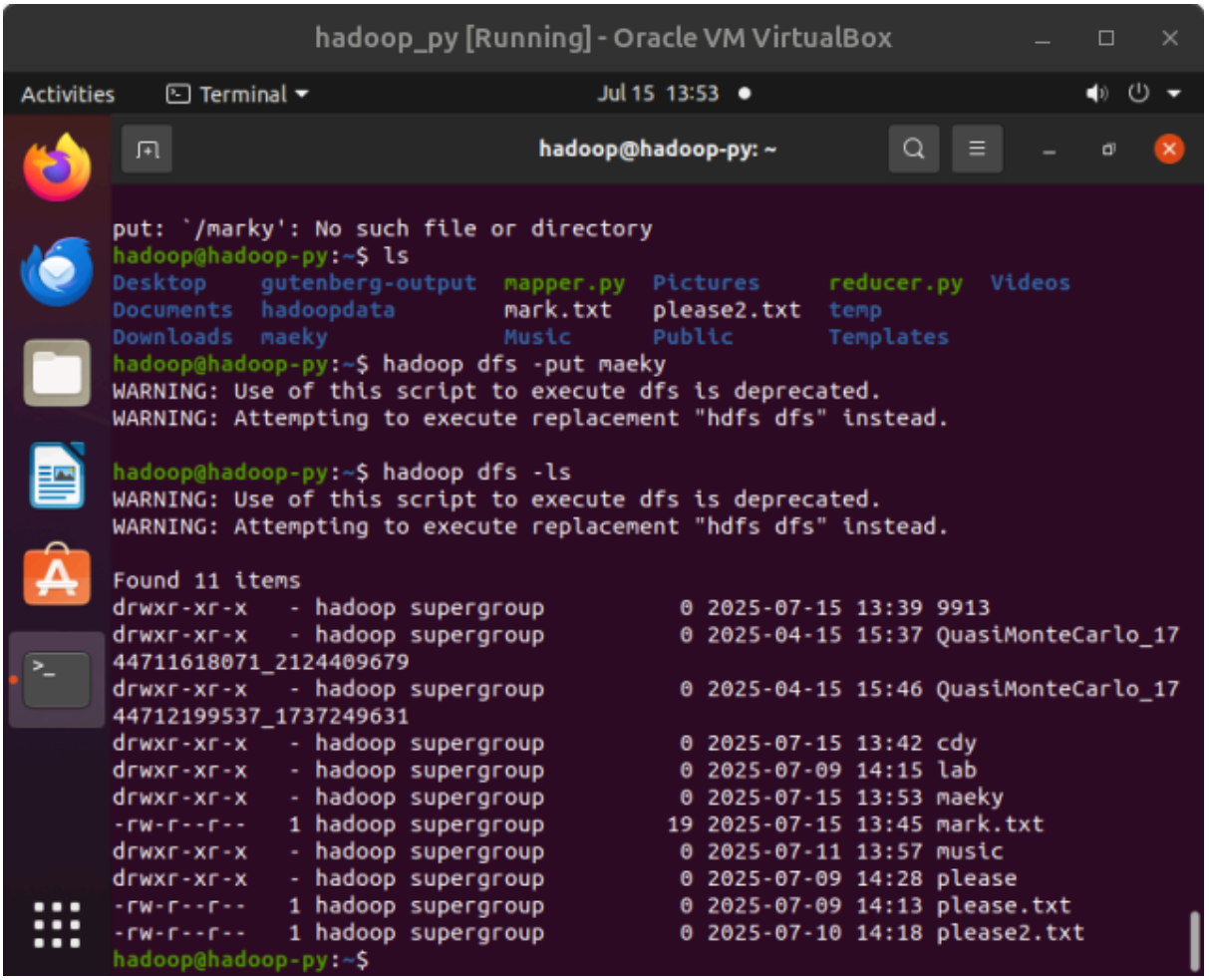
    - Document any issues encountered during setup and how they were resolved.
    - Discuss the scalability and fault-tolerance features provided by Hadoop.
    - Reflect on the importance of Hadoop in big data processing and its role in modern data architectures.

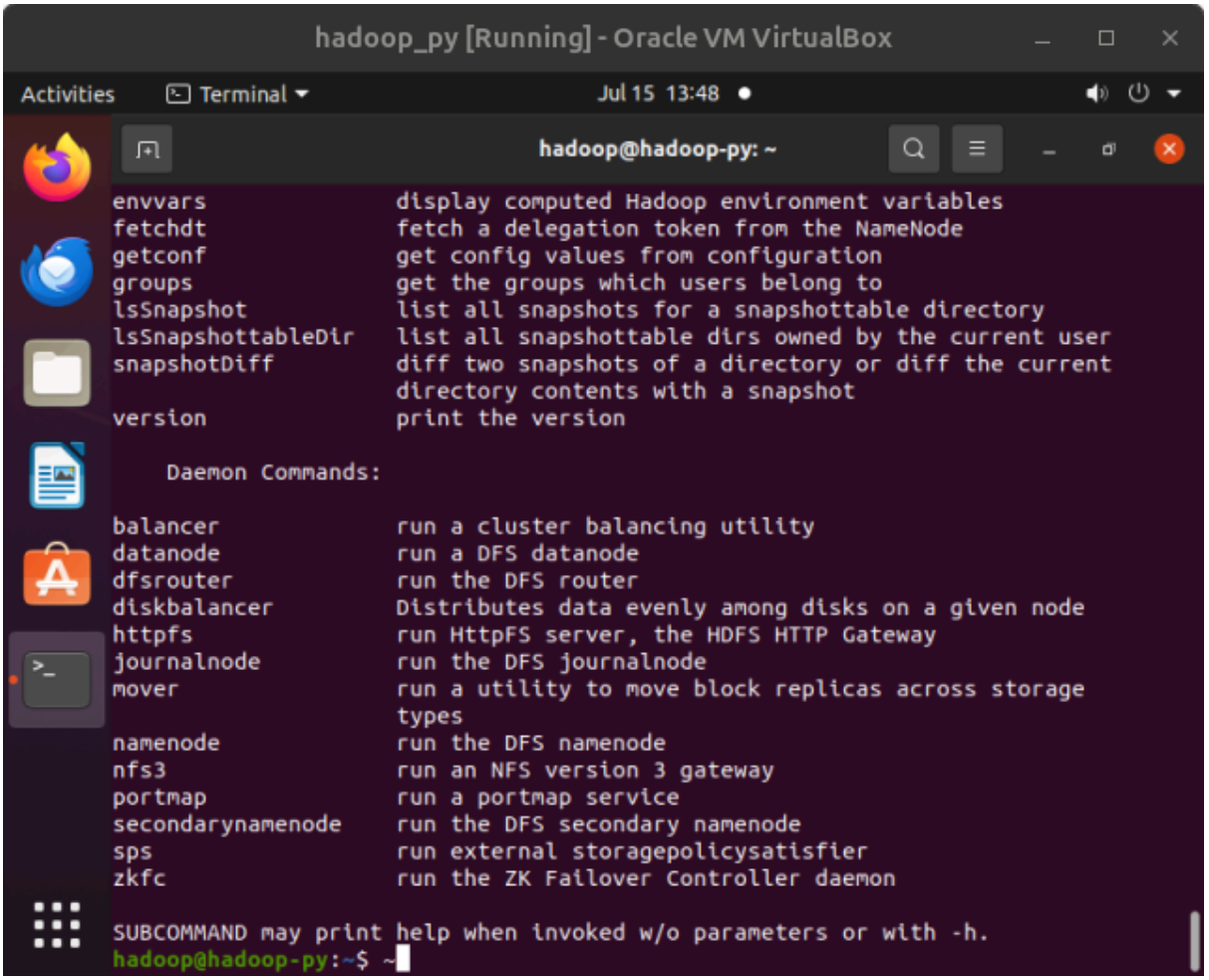
Expected Outcome:

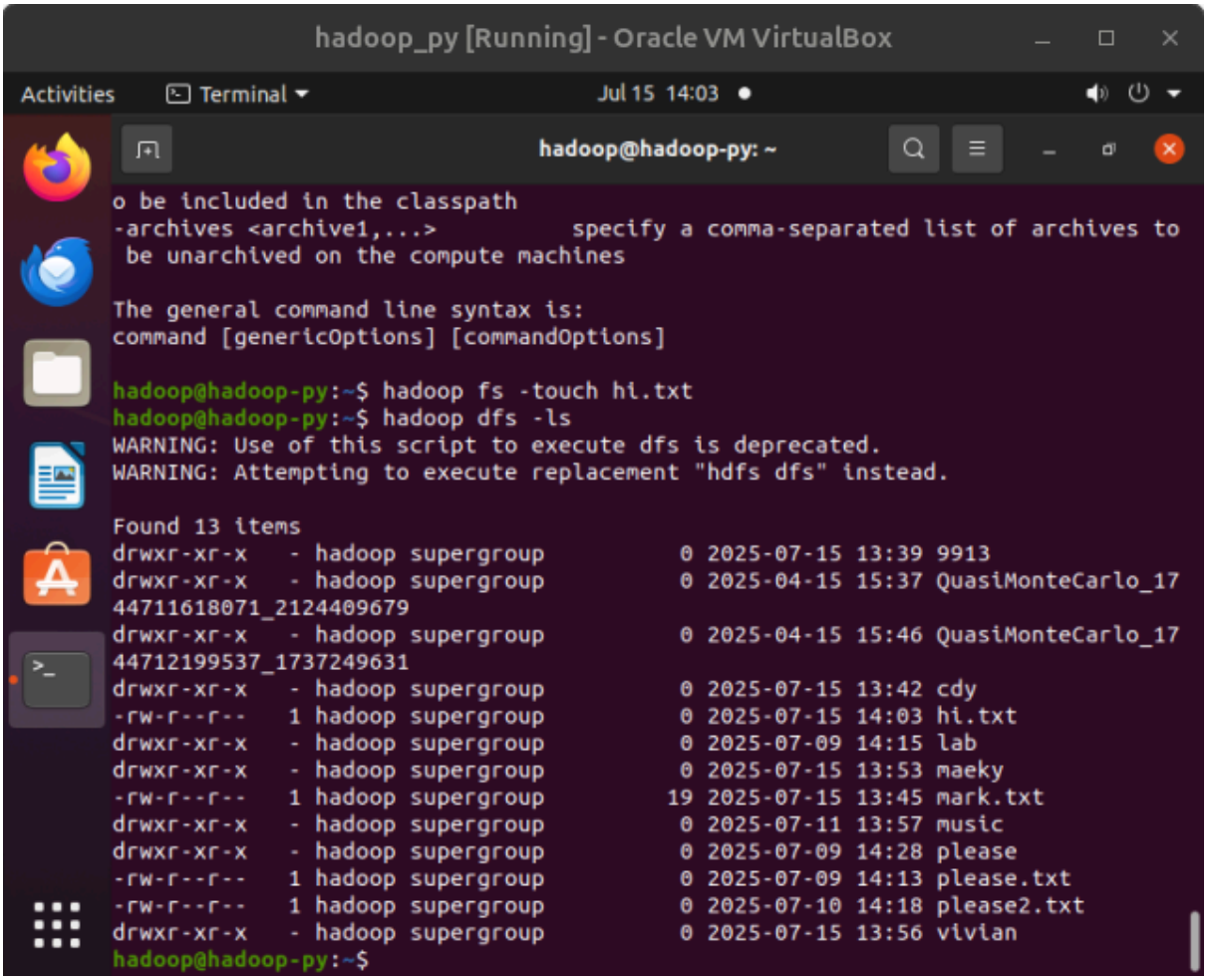
By the end of this experiment, students should have successfully set up a basic Hadoop cluster comprising HDFS and YARN components. They should be able to navigate Hadoop's web interfaces, execute basic Hadoop commands, and understand the distributed nature of Hadoop processing.

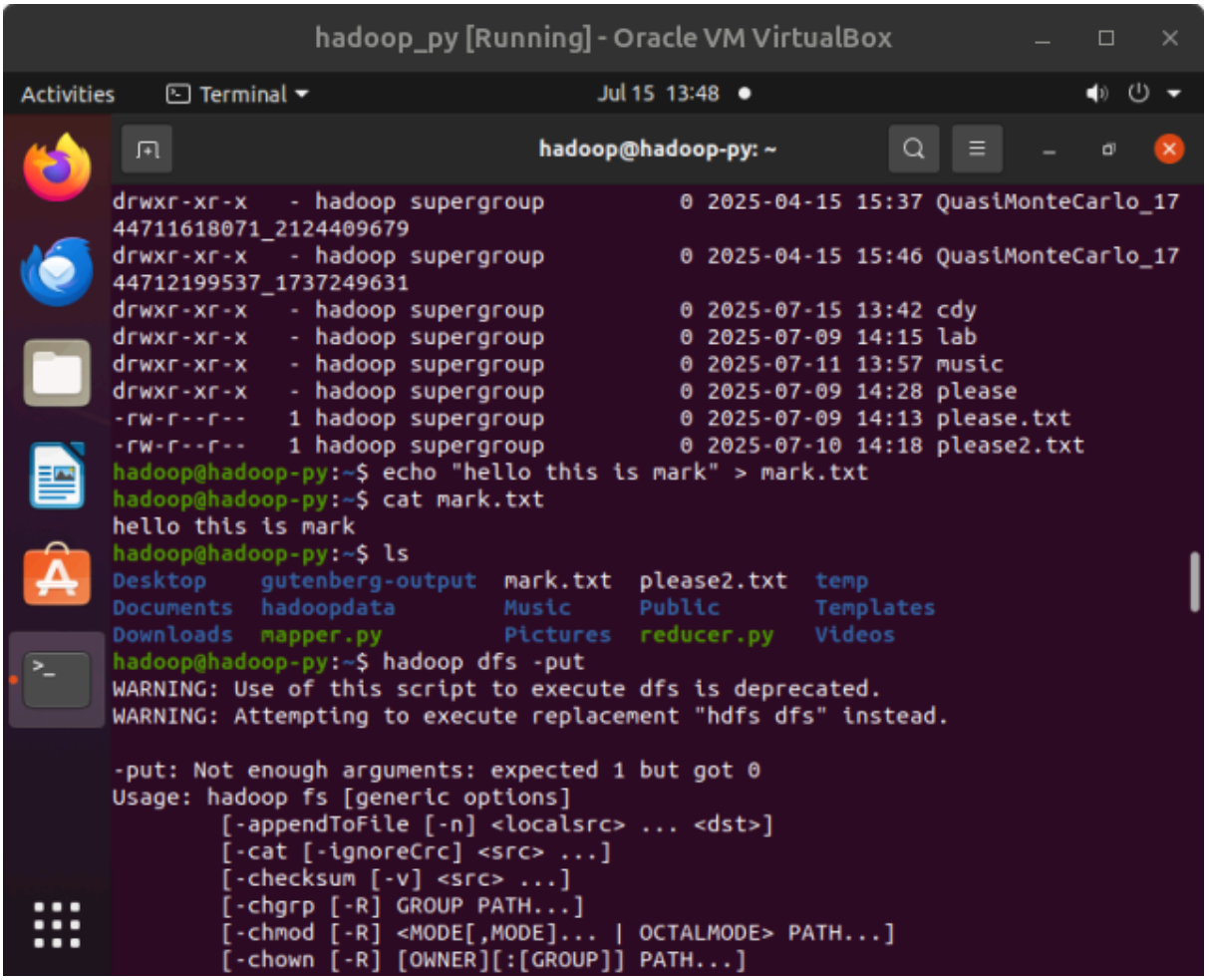
Conclusion:

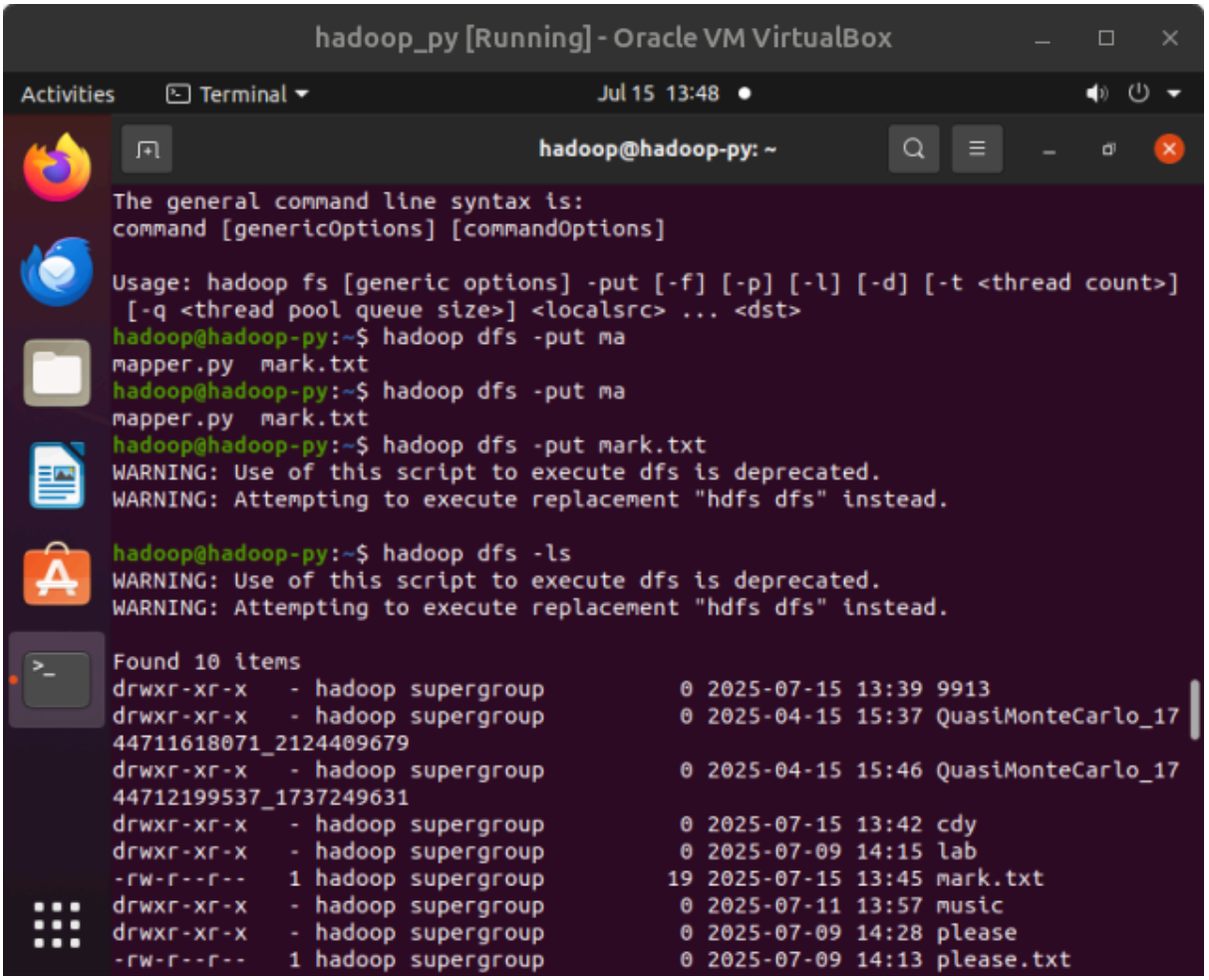
In this experiment, we successfully installed and configured a basic Hadoop ecosystem, including HDFS and YARN. Through hands-on setup of environment variables, SSH authentication, and cluster configuration, we gained practical understanding of Hadoop’s distributed architecture.

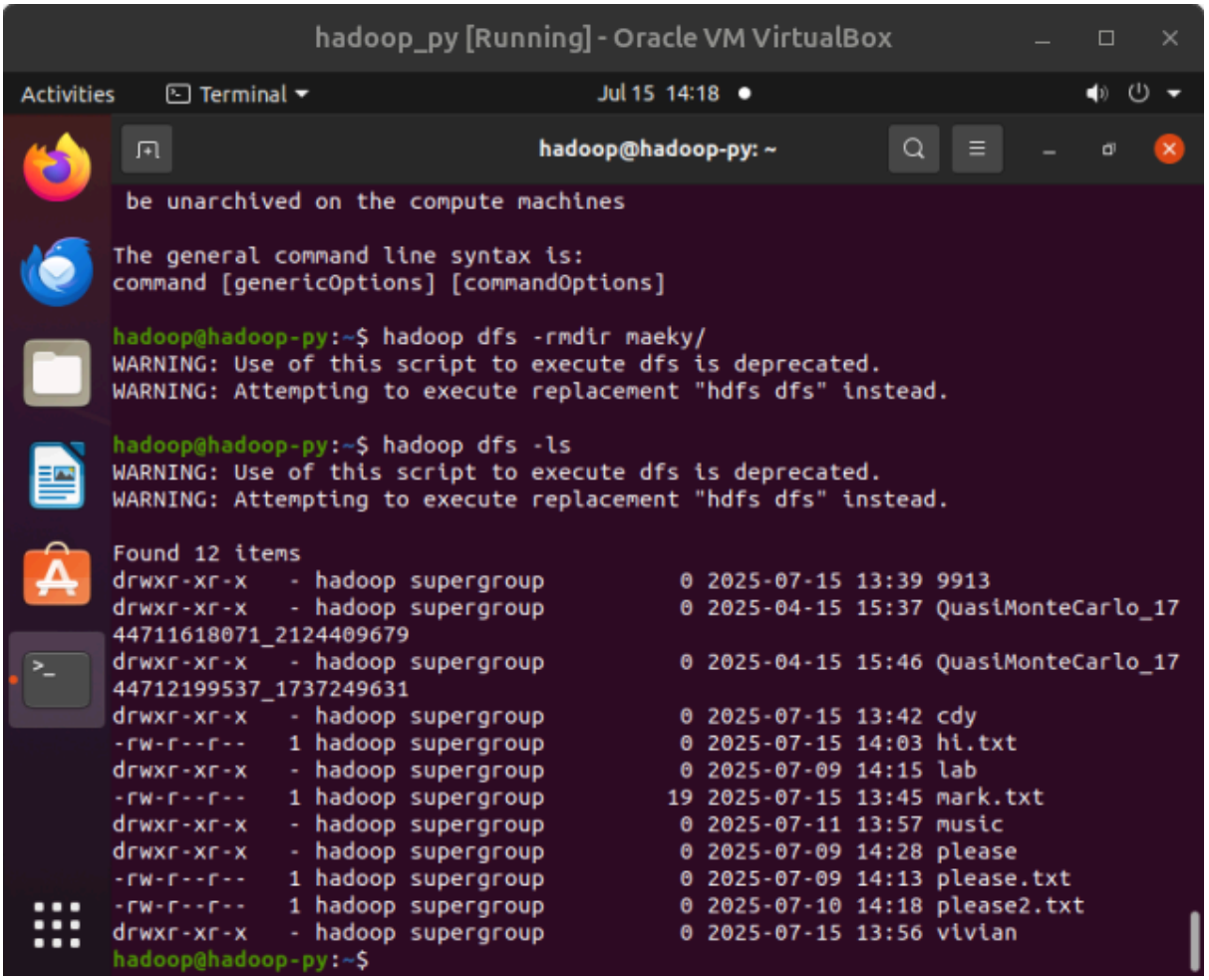

hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 13:53
hadoop@hadoop-py: ~
put: `/marky': No such file or directory
hadoop@hadoop-py:~$ ls
Desktop gutenberg-output mapper.py Pictures reducer.py Videos
Documents hadoopdata mark.txt please2.txt temp
Downloads maeky Music Public Templates
hadoop@hadoop-py:~$ hadoop dfs -put maeky
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
hadoop@hadoop-py:~$ hadoop dfs -ls
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
Found 11 items
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:39 9913
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:37 QuasiMonteCarlo_17
44711618071_2124409679
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:46 QuasiMonteCarlo_17
44712199537_1737249631
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:42 cdy
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:15 lab
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:53 maeky
-rw-r--r-- 1 hadoop supergroup 19 2025-07-15 13:45 mark.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-11 13:57 music
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:28 please
-rw-r--r-- 1 hadoop supergroup 0 2025-07-09 14:13 please.txt
-rw-r--r-- 1 hadoop supergroup 0 2025-07-10 14:18 please2.txt
hadoop@hadoop-py:~$

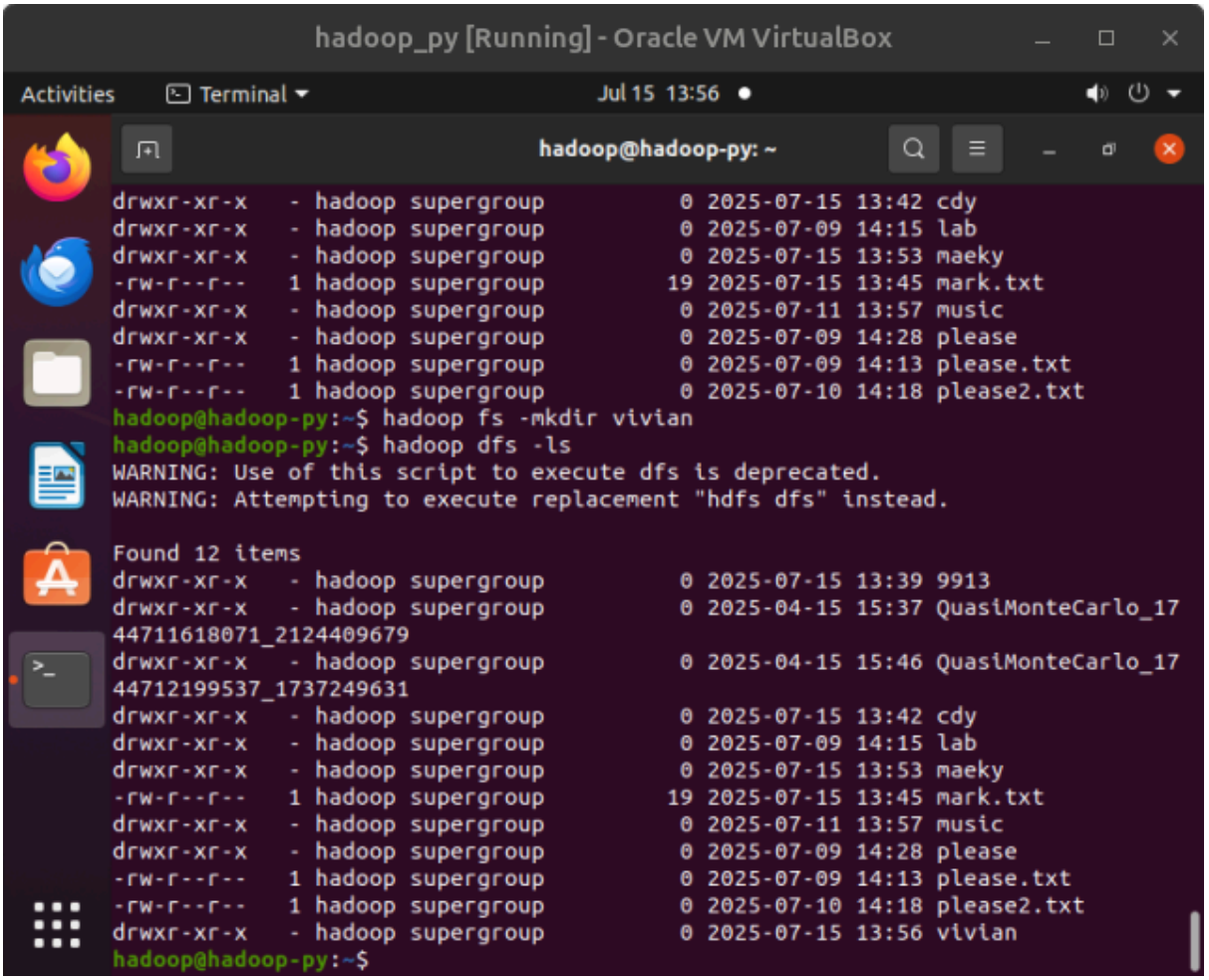

hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 13:48
hadoop@hadoop-py: ~
envvars display computed Hadoop environment variables
fetchdt fetch a delegation token from the NameNode
getconf get config values from configuration
groups get the groups which users belong to
lsSnapshot list all snapshots for a snapshottable directory
lsSnapshottableDir list all snapshottable dirs owned by the current user
snapshotDiff diff two snapshots of a directory or diff the current
directory contents with a snapshot
version print the version
Daemon Commands:
balancer run a cluster balancing utility
datanode run a DFS datanode
dfsrouter run the DFS router
diskbalancer Distributes data evenly among disks on a given node
httpfs run HttpFS server, the HDFS HTTP Gateway
journalnode run the DFS journalnode
mover run a utility to move block replicas across storage
types
namenode run the DFS namenode
nfs3 run an NFS version 3 gateway
portmap run a portmap service
secondarynamenode run the DFS secondary namenode
sps run external storagepolicysatisfier
zkfc run the ZK Failover Controller daemon
SUBCOMMAND may print help when invoked w/o parameters or with -h.
hadoop@hadoop-py:~$ ~

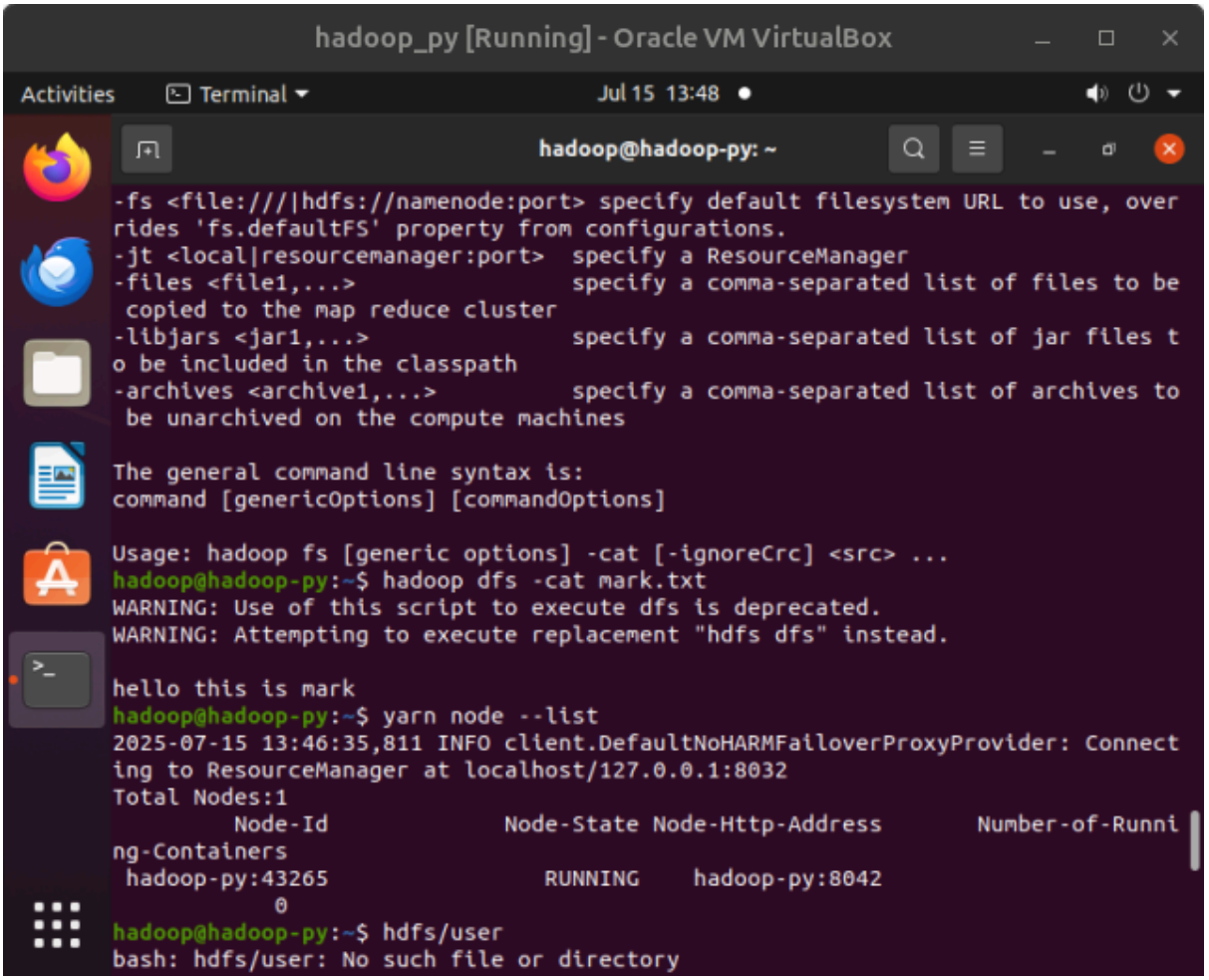

hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 14:03
hadoop@hadoop-py: ~
o be included in the classpath
-archives <archive1,...>   specify a comma-separated list of archives to
 be unarchived on the compute machines
The general command line syntax is:
command [genericOptions] [commandOptions]
hadoop@hadoop-py:~$ hadoop fs -touch hi.txt
hadoop@hadoop-py:~$ hadoop dfs -ls
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
Found 13 items
drwxr-xr-x   - hadoop supergroup   0 2025-07-15 13:39 9913
drwxr-xr-x   - hadoop supergroup   0 2025-04-15 15:37 QuasiMonteCarlo_17
44711618071_2124409679
drwxr-xr-x   - hadoop supergroup   0 2025-04-15 15:46 QuasiMonteCarlo_17
44712199537_1737249631
drwxr-xr-x   - hadoop supergroup   0 2025-07-15 13:42 cdy
-rw-r--r--   1 hadoop supergroup   0 2025-07-15 14:03 hi.txt
drwxr-xr-x   - hadoop supergroup   0 2025-07-09 14:15 lab
drwxr-xr-x   - hadoop supergroup   0 2025-07-15 13:53 maeky
-rw-r--r--   1 hadoop supergroup   19 2025-07-15 13:45 mark.txt
drwxr-xr-x   - hadoop supergroup   0 2025-07-11 13:57 music
drwxr-xr-x   - hadoop supergroup   0 2025-07-09 14:28 please
-rw-r--r--   1 hadoop supergroup   0 2025-07-09 14:13 please.txt
-rw-r--r--   1 hadoop supergroup   0 2025-07-10 14:18 please2.txt
drwxr-xr-x   - hadoop supergroup   0 2025-07-15 13:56 vivian
hadoop@hadoop-py:~$

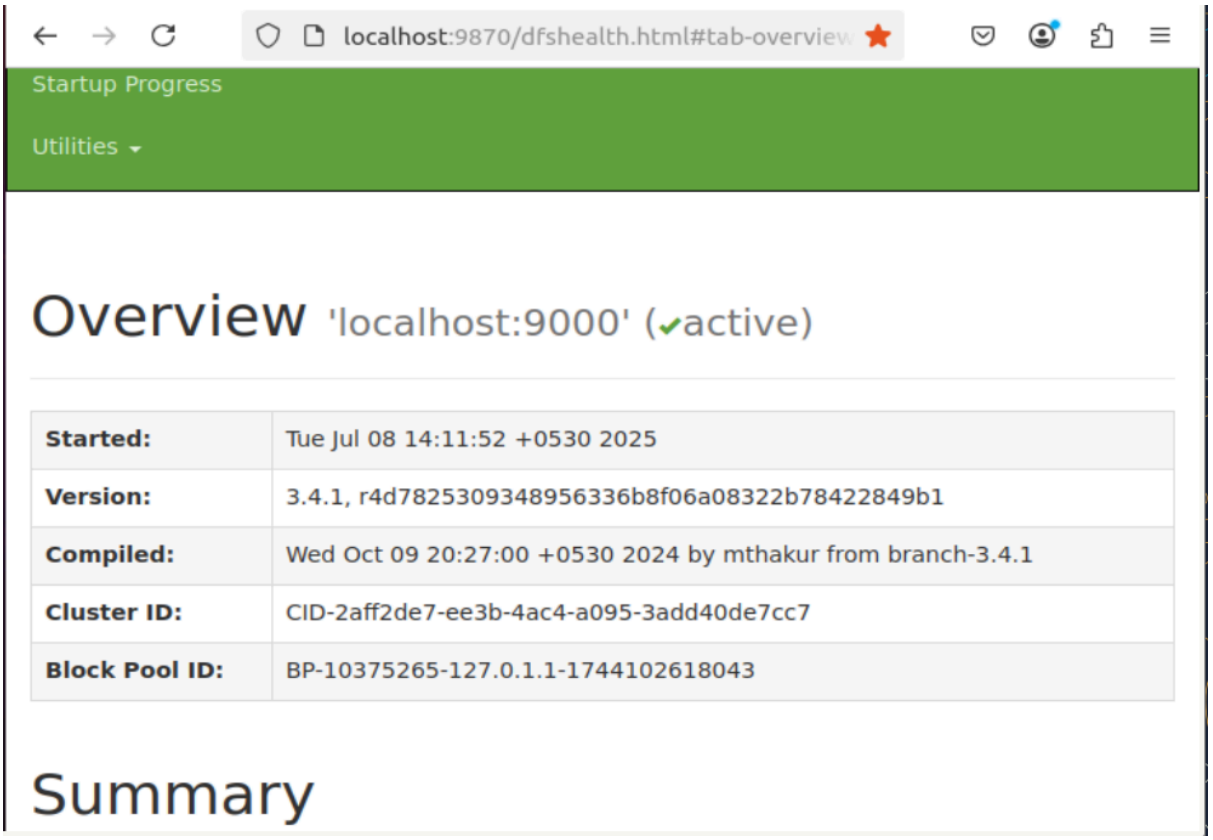

hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 13:48
hadoop@hadoop-py: ~
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:37 QuasiMonteCarlo_1744711618071_2124409679
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:46 QuasiMonteCarlo_1744712199537_1737249631
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:42 cdy
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:15 lab
drwxr-xr-x - hadoop supergroup 0 2025-07-11 13:57 music
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:28 please
-rw-r--r-- 1 hadoop supergroup 0 2025-07-09 14:13 please.txt
-rw-r--r-- 1 hadoop supergroup 0 2025-07-10 14:18 please2.txt
hadoop@hadoop-py:~$ echo "hello this is mark" > mark.txt
hadoop@hadoop-py:~$ cat mark.txt
hello this is mark
hadoop@hadoop-py:~$ ls
Desktop gutenberg-output mark.txt please2.txt temp
Documents hadoopdata Music Public Templates
Downloads mapper.py Pictures reducer.py Videos
hadoop@hadoop-py:~$ hadoop dfs -put
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
-put: Not enough arguments: expected 1 but got 0
Usage: hadoop fs [generic options]
[-appendToFile [-n] <localsrc> ... <dst>]
[-cat [-ignoreCrc] <src> ...]
[-checksum [-v] <src> ...]
[-chgrp [-R] GROUP PATH...]
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
[-chown [-R] [OWNER][:[GROUP]] PATH...]


hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 13:48
hadoop@hadoop-py: ~
The general command line syntax is:
command [genericOptions] [commandOptions]
Usage: hadoop fs [generic options] -put [-f] [-p] [-l] [-d] [-t <thread count>]
[-q <thread pool queue size>] <localsrc> ... <dst>
hadoop@hadoop-py:~$ hadoop dfs -put ma
mapper.py mark.txt
hadoop@hadoop-py:~$ hadoop dfs -put ma
mapper.py mark.txt
hadoop@hadoop-py:~$ hadoop dfs -put mark.txt
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
hadoop@hadoop-py:~$ hadoop dfs -ls
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
Found 10 items
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:39 9913
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:37 QuasiMonteCarlo_17
44711618071_2124409679
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:46 QuasiMonteCarlo_17
44712199537_1737249631
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:42 cdy
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:15 lab
-rw-r--r-- 1 hadoop supergroup 19 2025-07-15 13:45 mark.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-11 13:57 music
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:28 please
-rw-r--r-- 1 hadoop supergroup 0 2025-07-09 14:13 please.txt

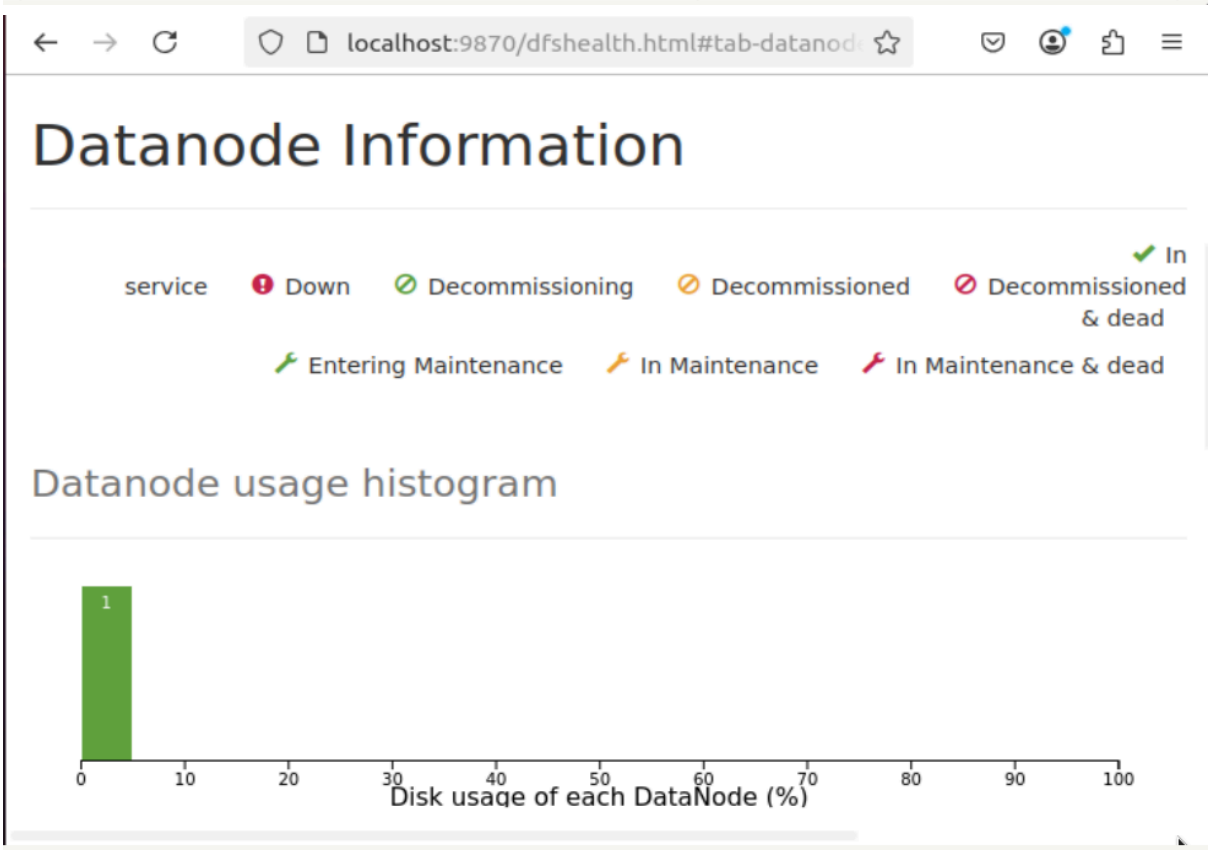

hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 14:18
hadoop@hadoop-py: ~
be unarchived on the compute machines
The general command line syntax is:
command [genericOptions] [commandOptions]
hadoop@hadoop-py:~$ hadoop dfs -rmdir maeky/
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
hadoop@hadoop-py:~$ hadoop dfs -ls
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
Found 12 items
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:39 9913
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:37 QuasiMonteCarlo_17
44711618071_2124409679
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:46 QuasiMonteCarlo_17
44712199537_1737249631
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:42 cdy
-rw-r--r-- 1 hadoop supergroup 0 2025-07-15 14:03 hi.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:15 lab
-rw-r--r-- 1 hadoop supergroup 19 2025-07-15 13:45 mark.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-11 13:57 music
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:28 please
-rw-r--r-- 1 hadoop supergroup 0 2025-07-09 14:13 please.txt
-rw-r--r-- 1 hadoop supergroup 0 2025-07-10 14:18 please2.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:56 vivian
hadoop@hadoop-py:~$

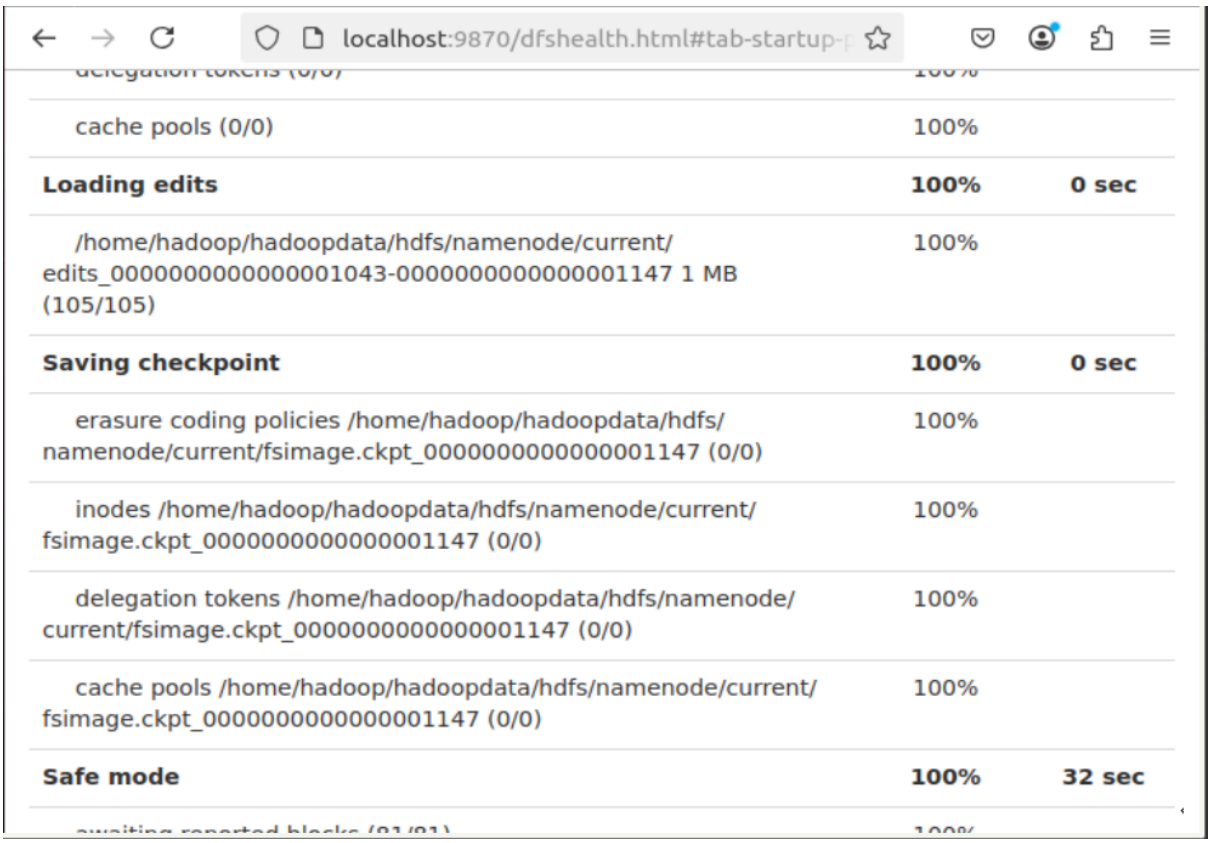

hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 13:56
hadoop@hadoop-py: ~
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:42 cdy
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:15 lab
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:53 maeky
-rw-r--r-- 1 hadoop supergroup 19 2025-07-15 13:45 mark.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-11 13:57 music
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:28 please
-rw-r--r-- 1 hadoop supergroup 0 2025-07-09 14:13 please.txt
-rw-r--r-- 1 hadoop supergroup 0 2025-07-10 14:18 please2.txt
hadoop@hadoop-py:~$ hadoop fs -mkdir vivian
hadoop@hadoop-py:~$ hadoop dfs -ls
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
Found 12 items
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:39 9913
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:37 QuasiMonteCarlo_17
44711618071_2124409679
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:46 QuasiMonteCarlo_17
44712199537_1737249631
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:42 cdy
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:15 lab
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:53 maeky
-rw-r--r-- 1 hadoop supergroup 19 2025-07-15 13:45 mark.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-11 13:57 music
drwxr-xr-x - hadoop supergroup 0 2025-07-09 14:28 please
-rw-r--r-- 1 hadoop supergroup 0 2025-07-09 14:13 please.txt
-rw-r--r-- 1 hadoop supergroup 0 2025-07-10 14:18 please2.txt
drwxr-xr-x - hadoop supergroup 0 2025-07-15 13:56 vivian
hadoop@hadoop-py:~$

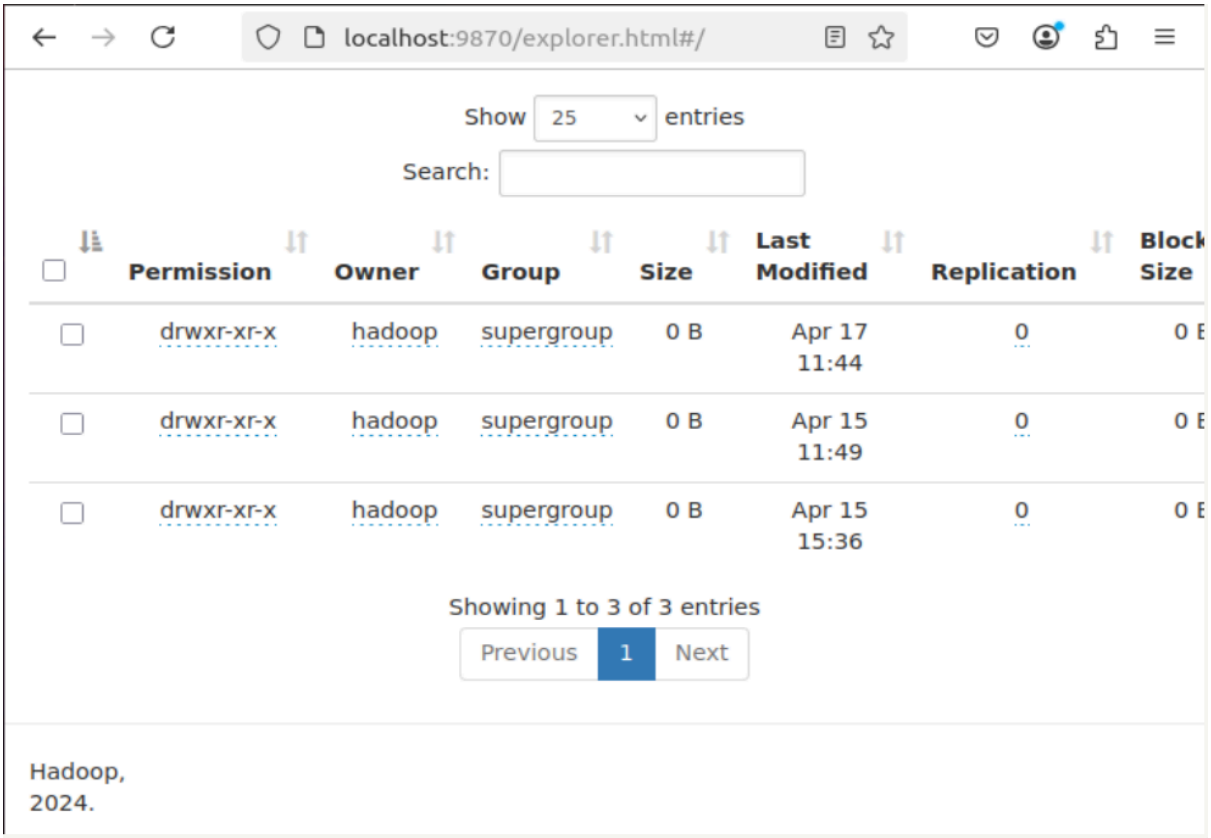

hadoop_py [Running] - Oracle VM VirtualBox
Activities
Terminal
Jul 15 13:48
hadoop@hadoop-py: ~
-fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
-jt <local|resourcemanager:port> specify a ResourceManager
-files <file1,...> specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...> specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...> specify a comma-separated list of archives to be unarchived on the compute machines
The general command line syntax is:
command [genericOptions] [commandOptions]
Usage: hadoop fs [generic options] -cat [-ignoreCrc] <src> ...
hadoop@hadoop-py:~$ hadoop dfs -cat mark.txt
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.
hello this is mark
hadoop@hadoop-py:~$ yarn node --list
2025-07-15 13:46:35,811 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at localhost/127.0.0.1:8032
Total Nodes:1
Node-Id Node-State Node-Http-Address Number-of-Running-Containers
hadoop-py:43265 RUNNING hadoop-py:8042 0
hadoop@hadoop-py:~$ hdfs/user
bash: hdfs/user: No such file or directory

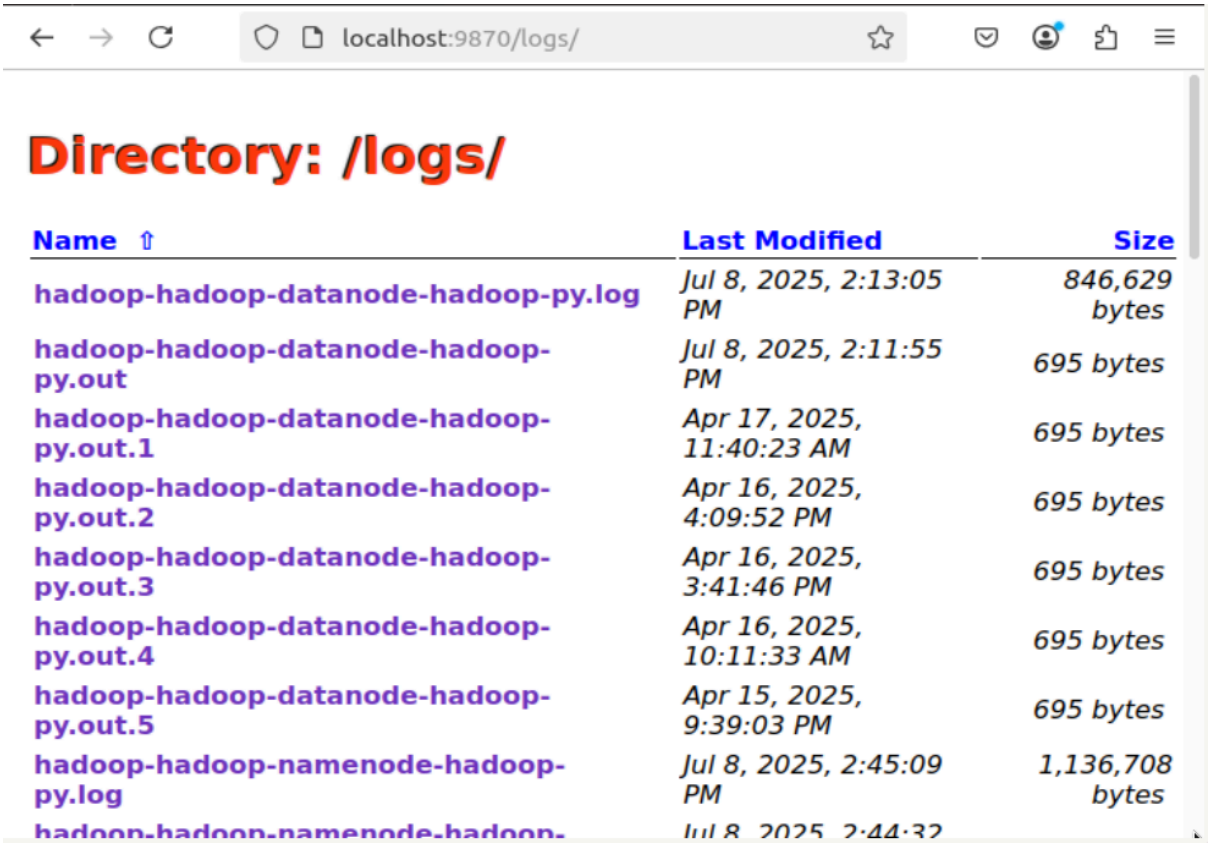

localhost:9870/dfshealth.html#tab-overview
Startup Progress
Utilities
Overview 'localhost:9000' (active)
Started: Tue Jul 08 14:11:52 +0530 2025
Version: 3.4.1, r4d7825309348956336b8f06a08322b78422849b1
Compiled: Wed Oct 09 20:27:00 +0530 2024 by mthakur from branch-3.4.1
Cluster ID: CID-2aff2de7-ee3b-4ac4-a095-3add40de7cc7
Block Pool ID: BP-10375265-127.0.1.1-1744102618043
Summary

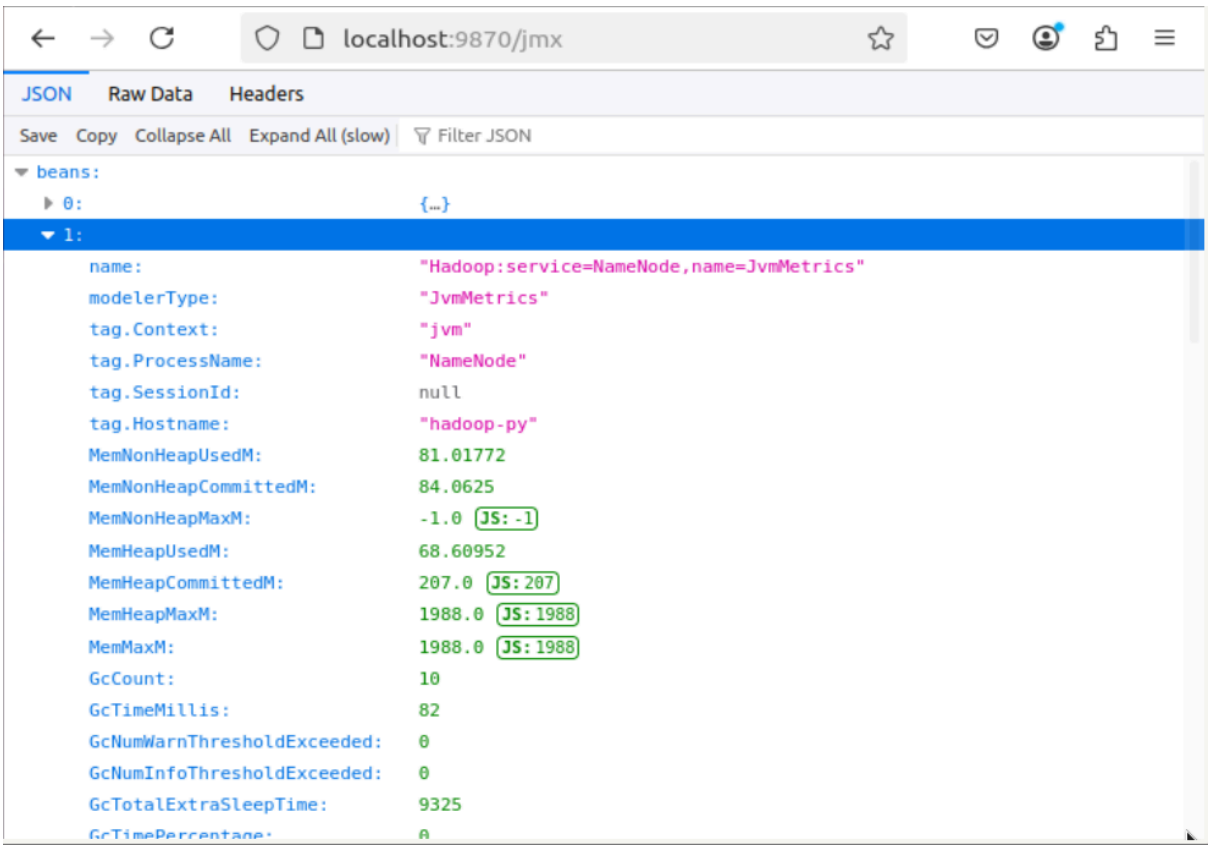
Configured Capacity:	19.02 GB
Configured Remote Capacity:	0 B
DFS Used:	10.64 MB (0.05%)
Non DFS Used:	13.06 GB
DFS Remaining:	4.96 GB (26.09%)
Block Pool Used:	10.64 MB (0.05%)
DataNodes usages% (Min/Median/Max/stdDev):	0.05% / 0.05% / 0.05% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0




localhost:9870/dfshealth.html#tab-startup-p
cache pools (0/0) 100%
Loading edits 100% 0 sec
/home/hadoop/hadoopdata/hdfs/namenode/current/edits_0000000000000001043-0000000000000001147 1 MB (105/105) 100%
Saving checkpoint 100% 0 sec
erasure coding policies /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_0000000000000001147 (0/0) 100%
inodes /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_0000000000000001147 (0/0) 100%
delegation tokens /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_0000000000000001147 (0/0) 100%
cache pools /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_0000000000000001147 (0/0) 100%
Safe mode 100% 32 sec


localhost:9870/explorer.html#/
Show 25 entries
Search:
Permission Owner Group Size Last Modified Replication Block Size
drwxr-xr-x hadoop supergroup 0 B Apr 17 11:44 0 0 B
drwxr-xr-x hadoop supergroup 0 B Apr 15 11:49 0 0 B
drwxr-xr-x hadoop supergroup 0 B Apr 15 15:36 0 0 B
Showing 1 to 3 of 3 entries
Previous 1 Next
Hadoop, 2024.


localhost:9870/logs/
Directory: /logs/
Name ⇧
Last Modified
Size
hadoop-hadoop-datanode-hadoop-py.log
Jul 8, 2025, 2:13:05 PM
846,629 bytes
hadoop-hadoop-datanode-hadoop-py.out
Jul 8, 2025, 2:11:55 PM
695 bytes
hadoop-hadoop-datanode-hadoop-py.out.1
Apr 17, 2025, 11:40:23 AM
695 bytes
hadoop-hadoop-datanode-hadoop-py.out.2
Apr 16, 2025, 4:09:52 PM
695 bytes
hadoop-hadoop-datanode-hadoop-py.out.3
Apr 16, 2025, 3:41:46 PM
695 bytes
hadoop-hadoop-datanode-hadoop-py.out.4
Apr 16, 2025, 10:11:33 AM
695 bytes
hadoop-hadoop-datanode-hadoop-py.out.5
Apr 15, 2025, 9:39:03 PM
695 bytes
hadoop-hadoop-namenode-hadoop-py.log
Jul 8, 2025, 2:45:09 PM
1,136,708 bytes


localhost:9870/jmx
JSON
Raw Data
Headers
Save
Copy
Collapse All
Expand All (slow)
Filter JSON
beans:
0: {…}
1:
name: "Hadoop:service=NameNode,name=JvmMetrics"
modelerType: "JvmMetrics"
tag.Context: "jvm"
tag.ProcessName: "NameNode"
tag.SessionId: null
tag.Hostname: "hadoop-py"
MemNonHeapUsedM: 81.01772
MemNonHeapCommittedM: 84.0625
MemNonHeapMaxM: -1.0 JS: -1
MemHeapUsedM: 68.60952
MemHeapCommittedM: 207.0 JS: 207
MemHeapMaxM: 1988.0 JS: 1988
MemMaxM: 1988.0 JS: 1988
GcCount: 10
GcTimeMillis: 82
GcNumWarnThresholdExceeded: 0
GcNumInfoThresholdExceeded: 0
GcTotalExtraSleepTime: 9325

Postlab:-

1. What are the main components of a Hadoop application?

HDFS (Hadoop Distributed File System):
Stores large files across multiple machines with fault tolerance using replication.

YARN (Yet Another Resource Negotiator):
Manages cluster resources and job scheduling.

MapReduce:
A programming model used for distributed data processing (map = split, reduce = aggregate).

Hadoop Common:
Provides essential Java libraries and utilities used by other modules.

2. Difference between NameNode, Backup Node, and Checkpoint Node:

Component	Function	Real-Time Sync	Failure Recovery Role
NameNode	Manages file system metadata like file names, directories, and block locations.	Yes	Acts as the master; essential for HDFS operation.
Backup Node	Maintains an in-memory, up-to-date copy of metadata from the NameNode.	Yes	Can immediately take over if NameNode fails.
Checkpoint Node	Periodically downloads and merges fsimage and edits, then sends a new fsimage to NameNode.	No	Reduces NameNode startup time, not used for failover.

3. Explain the use of `cat`, `du`, `du -s`:

- `cat` (concatenate):
  Used to view the contents of files in the terminal. Example: `cat file.txt`

- `du` (disk usage):
  Shows the space used by files and directories. Example: `du myfolder/`

- `du -s` (summary):
  Displays the total size of a folder, instead of listing all subdirectories. Example: `du -s myfolder/`