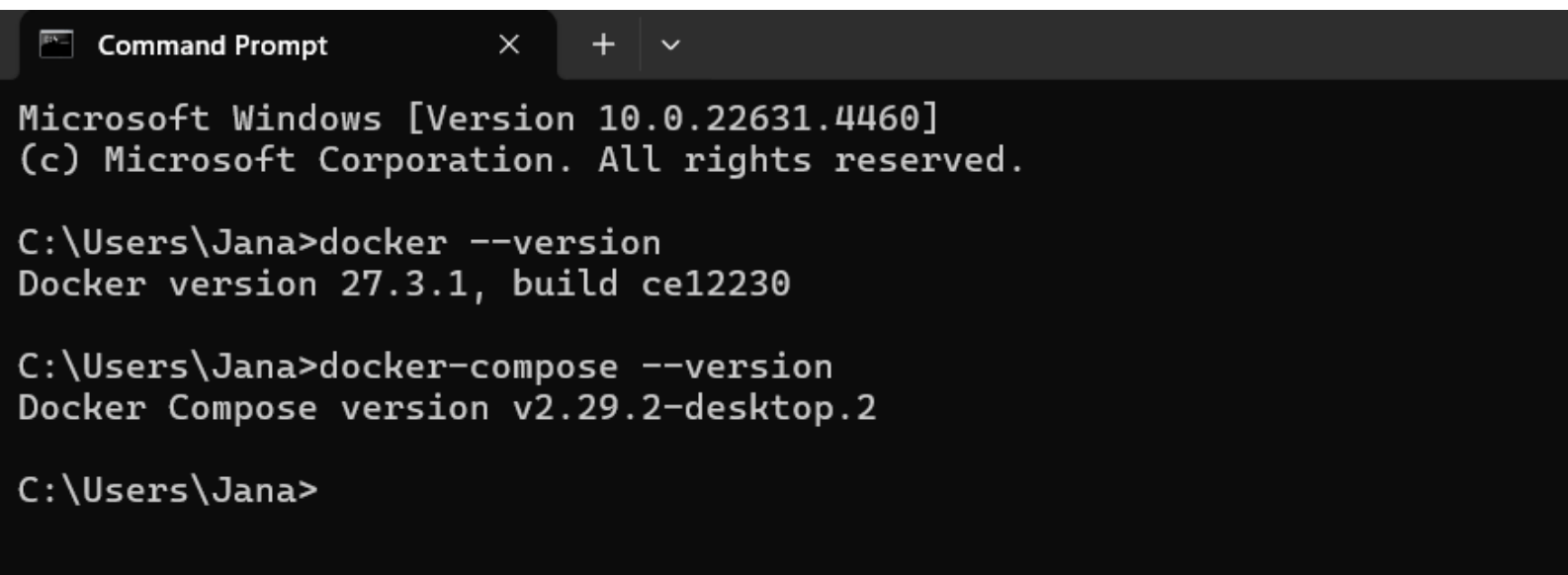


01. Version Check



```
Microsoft Windows [Version 10.0.22631.4460]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Jana>docker --version
Docker version 27.3.1, build ce12230

C:\Users\Jana>docker-compose --version
Docker Compose version v2.29.2-desktop.2

C:\Users\Jana>
```

02. Step 2: Deploy the Cluster

`docker -compose up -d`

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> code .
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker-compose up -d
time="2024-11-25T08:31:03+05:30" level=warning msg="C:\\Users\\Jana\\Desktop\\Lecture Notes\\3 Year 2 Semester\\SE6103 Parallel and Distributed Systems\\Practical\\Day 4\\docker-compose.yml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 8/8
  ✔ datanode Pulled                                6.6s
  ✔ 5e185246c615 Download complete                 0.8s
  ✔ b91d0b0b68c8 Download complete                 1.4s
  ✔ 4bf0ae3d5cc8 Download complete                 1.2s
  ✔ historyserver Pulled                           8.0s
  ✔ f3f6b02c1935 Download complete                 1.1s
  ✔ 84560426d8fd Download complete                 0.8s
  ✔ 78d381637ee0 Download complete                 1.4s
[+] Running 6/6
  ✔ Network day4_default Created                   0.1s
  ✔ Volume "day4_namenode-data" Created            0.0s
  ✔ Volume "day4_datanode-data" Created            0.0s
  ✔ Container namenode Started                    0.0s
  ✔ Container datanode Started                    1.0s
  ✔ Container historyserver Started               1.3s
```

03. Uploading Data to HDFS

1) Download Sample Data

```
docker exec -it namenode hdfs dfs -mkdir -p /input
```

```
docker cp "C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4\sample.txt" namenode:/sample.txt
```

```
docker exec -it namenode ls /
```

```
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker cp "C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4\sample.txt" namenode:/sample.txt
Successfully copied 2.05kB to namenode:/sample.txt
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec -it namenode ls /
KEYS boot entrypoint.sh hadoop home lib64 mnt proc run sample.txt srv tmp var
bin dev etc hadoop-data lib media opt root run.sh sbin sys usr

What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
```

2) Upload Data to HDFS

```
docker exec -it namenode hdfs dfs -mkdir -p /input
```

```
docker exec -it namenode hdfs dfs -put /sample.txt /input
```

```
docker exec -it namenode hdfs dfs -ls /input
```

```
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec -it namenode hdfs dfs -mkdir -p /input
```

What's next:

Try Docker Debug for seamless, persistent debugging tools in any container or image → [docker debug namenode](#)

Learn more at <https://docs.docker.com/go/debug-cli/>

```
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec -it namenode hdfs dfs -put /sample.txt /input
2024-11-25 03:22:30,980 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
```

What's next:

Try Docker Debug for seamless, persistent debugging tools in any container or image → [docker debug namenode](#)

Learn more at <https://docs.docker.com/go/debug-cli/>

```
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec -it namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r-- 3 root supergroup          57 2024-11-25 03:22 /input/sample.txt
```

What's next:

Try Docker Debug for seamless, persistent debugging tools in any container or image → [docker debug namenode](#)

Learn more at <https://docs.docker.com/go/debug-cli/>

04. Running a MapReduce Job

1) Run the Word Count MapReduce Job

`docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output`

```
Windows PowerShell
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output
2024-11-25 09:52:43,445 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-25 09:52:43,653 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-25 09:52:43,653 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-25 09:52:44,050 INFO input.FileInputFormat: Total input files to process : 1
2024-11-25 09:52:44,145 INFO mapreduce.JobSubmitter: number of splits:1
2024-11-25 09:52:44,442 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1368330908_0001
2024-11-25 09:52:44,442 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-25 09:52:44,572 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-25 09:52:44,573 INFO mapreduce.Job: Running job: job_local1368330908_0001
2024-11-25 09:52:44,574 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-25 09:52:44,586 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 09:52:44,586 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-25 09:52:44,587 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-25 09:52:44,635 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-25 09:52:44,637 INFO mapred.LocalJobRunner: Starting task: attempt_local1368330908_0001_m_000000_0
2024-11-25 09:52:44,676 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 09:52:44,677 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-25 09:52:44,730 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-11-25 09:52:44,743 INFO mapred.MapTask: Processing split: hdfs://namenode:8020/input/sample.txt:0+57
2024-11-25 09:52:45,420 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-11-25 09:52:45,420 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-25 09:52:45,421 INFO mapred.MapTask: soft limit at 83886080
2024-11-25 09:52:45,421 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-11-25 09:52:45,421 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-11-25 09:52:45,442 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-25 09:52:45,579 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-11-25 09:52:45,583 INFO mapreduce.Job: Job job_local1368330908_0001 running in uber mode : false
2024-11-25 09:52:45,586 INFO mapreduce.Job: map 0% reduce 0%
2024-11-25 09:52:45,750 INFO mapred.LocalJobRunner:
2024-11-25 09:52:45,762 INFO mapred.MapTask: Starting flush of map output
2024-11-25 09:52:45,762 INFO mapred.MapTask: Spilling map output
2024-11-25 09:52:45,762 INFO mapred.MapTask: bufstart = 0; bufend = 86; bufvoid = 104857600
2024-11-25 09:52:45,763 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214368(104857472); length = 29/6553600
2024-11-25 09:52:45,799 INFO mapred.MapTask: Finished spill 0
2024-11-25 09:52:45,815 INFO mapred.Task: Task:attempt_local1368330908_0001_m_000000_0 is done. And is in the process of committing
2024-11-25 09:52:45,823 INFO mapred.LocalJobRunner: map
2024-11-25 09:52:45,823 INFO mapred.Task: Task 'attempt_local1368330908_0001_m_000000_0' done.
```

```
Windows PowerShell
2024-11-25 09:52:45,833 INFO mapred.Task: Final Counters for attempt_local1368330908_0001_m_000000_0: Counters: 24
File System Counters
  FILE: Number of bytes read=316693
  FILE: Number of bytes written=842045
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=57
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=5
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=1
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=3
  Map output records=8
  Map output bytes=86
  Map output materialized bytes=96
  Input split bytes=102
  Combine input records=8
  Combine output records=7
  Spilled Records=7
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=283639808
File Input Format Counters
  Bytes Read=57
2024-11-25 09:52:45,833 INFO mapred.LocalJobRunner: Finishing task: attempt_local1368330908_0001_m_000000_0
2024-11-25 09:52:45,834 INFO mapred.LocalJobRunner: map task executor complete.
2024-11-25 09:52:45,837 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2024-11-25 09:52:45,839 INFO mapred.LocalJobRunner: Starting task: attempt_local1368330908_0001_r_000000_0
2024-11-25 09:52:45,848 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 09:52:45,848 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup fa
ilures: false
2024-11-25 09:52:45,849 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-11-25 09:52:45,852 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@4f5c7eb
2024-11-25 09:52:45,854 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-11-25 09:52:45,875 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=1272394496, maxSingleShuffleLimit=318098624, mergeThreshold=839780416, ioSo
rtFactor=10, memToMemMergeOutputsThreshold=10
2024-11-25 09:52:45,879 INFO reduce.EventFetcher: attempt_local1368330908_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
```



```
Windows PowerShell
2024-11-25 09:52:45,905 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local1368330908_0001_m_000000_0 decomp: 92 len: 96 t
o MEMORY
2024-11-25 09:52:45,910 INFO reduce.InMemoryMapOutput: Read 92 bytes from map-output for attempt_local1368330908_0001_m_000000_0
2024-11-25 09:52:45,911 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 92, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMe
mory -> 92
2024-11-25 09:52:45,914 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2024-11-25 09:52:45,916 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-11-25 09:52:45,916 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2024-11-25 09:52:45,926 INFO mapred.Merger: Merging 1 sorted segments
2024-11-25 09:52:45,926 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 78 bytes
2024-11-25 09:52:45,928 INFO reduce.MergeManagerImpl: Merged 1 segments, 92 bytes to disk to satisfy reduce memory limit
2024-11-25 09:52:45,928 INFO reduce.MergeManagerImpl: Merging 1 files, 96 bytes from disk
2024-11-25 09:52:45,929 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2024-11-25 09:52:45,930 INFO mapred.Merger: Merging 1 sorted segments
2024-11-25 09:52:45,931 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 78 bytes
2024-11-25 09:52:45,932 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-11-25 09:52:45,965 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2024-11-25 09:52:46,043 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-11-25 09:52:46,548 INFO mapred.Task: Task:attempt_local1368330908_0001_r_000000_0 is done. And is in the process of committing
2024-11-25 09:52:46,563 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-11-25 09:52:46,564 INFO mapred.Task: Task attempt_local1368330908_0001_r_000000_0 is allowed to commit now
2024-11-25 09:52:46,590 INFO mapreduce.Job: map 100% reduce 0%
2024-11-25 09:52:46,614 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1368330908_0001_r_000000_0' to hdfs://namenode:8020/output
2024-11-25 09:52:46,618 INFO mapred.LocalJobRunner: reduce > reduce
2024-11-25 09:52:46,618 INFO mapred.Task: Task 'attempt_local1368330908_0001_r_000000_0' done.
2024-11-25 09:52:46,619 INFO mapred.Task: Final Counters for attempt_local1368330908_0001_r_000000_0: Counters: 30
File System Counters
  FILE: Number of bytes read=316917
  FILE: Number of bytes written=842141
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=57
  HDFS: Number of bytes written=62
  HDFS: Number of read operations=10
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=3
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Combine input records=0
  Combine output records=0
```

Windows PowerShell

```
Map output bytes=86
Map output materialized bytes=96
Input split bytes=102
Combine input records=8
Combine output records=7
Reduce input groups=7
Reduce shuffle bytes=96
Reduce input records=7
Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=16
Total committed heap usage (bytes)=567279616
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=57
File Output Format Counters
Bytes Written=62
```

What's next:

Try Docker Debug for seamless, persistent debugging tools in any container or image → [docker debug namenode](#)
Learn more at <https://docs.docker.com/go/debug-cli/>

2) View Job Output

```
docker exec -it namenode hdfs dfs -ls /output
```

3) Download the results

```
docker exec -it namenode hdfs dfs -cat /output/part-r-00000
```

```
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec -it namenode hdfs dfs -ls /out
put
Found 2 items
-rw-r--r-- 3 root supergroup 0 2024-11-25 09:52 /output/_SUCCESS
-rw-r--r-- 3 root supergroup 62 2024-11-25 09:52 /output/part-r-00000

What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec-it namenode hdfs dfs-cat /out
put/part-r-00000
docker: 'exec-it' is not a docker command.
See 'docker --help'
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker exec -it namenode hdfs dfs -cat /ou
put/part-r-00000
2024-11-25 09:59:02,406 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Distributed 1
Hadoop 1
Hello 2
Systems 1
World 1
are 1
cool 1

What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
```

05. Analyze and Clean Up

docker-compose down

```
PS C:\Users\Jana\Desktop\Lecture Notes\3 Year 2 Semester\SE6103 Parallel and Distributed Systems\Practical\Day 4> docker-compose down
time="2024-11-25T15:33:58+05:30" level=warning msg="C:\\Users\\Jana\\Desktop\\Lecture Notes\\3 Year 2 Semester\\SE6103 Parallel and Distributed Systems\\Practical\\Day 4\\docker-compose.yml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 4/4
✔ Container historyserver Removed 10.6s
✔ Container datanode Removed 10.6s
✔ Container namenode Removed 10.6s
✔ Network day4_default Removed 0.2s
```

Overview 'namenode:8020' (active)

Started:	Mon Nov 25 08:31:16 +0530 2024
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-84bd8e8a-f2f7-488e-ad59-8a0e5ace2d1a
Block Pool ID:	BP-984741370-172.19.0.2-1732503674544

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 152.6 MB of 244.5 MB Heap Memory. Max Heap Memory is 1.69 GB.

Non Heap Memory used 45.39 MB of 46.56 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	1006.85 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	4.44 GB
DFS Remaining:	951.19 GB (94.47%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 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
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Mon Nov 25 08:31:16 +0530 2024
Last Checkpoint Time	Mon Nov 25 08:31:14 +0530 2024
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 1

Journal Manager	State
FileJournalManager(root=/hadoop/dfs/name)	EditLogFileOutputStream(/hadoop/dfs/name/current/edits_inprogress_0000000000000000001)

NameNode Storage

Storage Directory	Type	State
/hadoop/dfs/name	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	1006.85 GB	24 KB (0%)	951.19 GB (94.47%)	24 KB	1

Browse Directory

Show 25 entries




Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	root	supergroup	57 B	Nov 25 08:52	3	128 MB	sample.txt

Showing 1 to 1 of 1 entries

Browse Directory

Go!




Show

25

 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 25 08:34	0	0 B	Input	

Showing 1 to 1 of 1 entries

Previous

1

Next

File information - sample.txt



[Download](#)

[Head the file \(first 32K\)](#)

[Tail the file \(last 32K\)](#)

Block information —

Block 0 ▾

Block ID: 1073741825

Block Pool ID: BP-984741370-172.19.0.2-1732503674544

Generation Stamp: 1001

Size: 57

Availability:

- 99359a6035d0

Close