

```
Command Prompt
-v, --version          Print version information and quit

Run 'docker COMMAND --help' for more information on a command.

For more help on how to use Docker, head to https://docs.docker.com/go/guides/

C:\Users\user>docker images
REPOSITORY              TAG          IMAGE ID       CREATED        SIZE
thisfristone            latest       de0fd248b0fe  2 weeks ago   226MB
jenkins/jenkins         latest       6e28580fa377  3 weeks ago   788MB
ubuntu                  latest       99c35190e22d  5 weeks ago   117MB
nginx                   latest       28402db69fec  6 weeks ago   279MB
docker/welcome-to-docker latest       eedaff45e3c7  12 months ago 29.5MB
hello-world             latest       d211f485f2dd  18 months ago 24.4kB
bde2020/hadoop-namenode latest       fdf741108051  4 years ago   2.05GB

C:\Users\user>
```

docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
-v, --version          Print version information and quit

Run 'docker COMMAND --help' for more information on a command.

For more help on how to use Docker, head to https://docs.docker.com/go/guides/

C:\Users\user>docker images
REPOSITORY              TAG          IMAGE ID       CREATED        SIZE
thisfristone            latest       de0fd248b0fe  2 weeks ago   226MB
jenkins/jenkins         latest       6e28580fa377  3 weeks ago   788MB
ubuntu                  latest       99c35190e22d  5 weeks ago   117MB
nginx                   latest       28402db69fec  6 weeks ago   279MB
docker/welcome-to-docker latest       eedaff45e3c7  12 months ago 29.5MB
hello-world             latest       d211f485f2dd  18 months ago 24.4kB
bde2020/hadoop-namenode latest       fdf741108051  4 years ago   2.05GB

C:\Users\user>docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
- Setting fs.defaultFS=hdfs://9f3066d6281e:8020
Configuring hdfs
- Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@9f3066d6281e:/#
```

## hdfs namenode -format

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
2024-11-18 08:09:58,945 INFO namenode.FSDirectory: ACLs enabled? false
2024-11-18 08:09:58,945 INFO namenode.FSDirectory: POSIX ACL inheritance enabled? true
2024-11-18 08:09:58,945 INFO namenode.FSDirectory: XAttrs enabled? true
2024-11-18 08:09:58,945 INFO namenode.NameNode: Caching file names occurring more than 10 times
2024-11-18 08:09:58,950 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCaptureAccessTimeOnlyChange: false, s
napshotDiffAllowSnapRootDescendant: true, maxSnapshotLimit: 65536
2024-11-18 08:09:58,952 INFO snapshot.SnapshotManager: SkipList is disabled
2024-11-18 08:09:58,956 INFO util.GSet: Computing capacity for map cachedBlocks
2024-11-18 08:09:58,956 INFO util.GSet: VM type = 64-bit
2024-11-18 08:09:58,956 INFO util.GSet: 0.25% max memory 2.6 GB = 6.6 MB
2024-11-18 08:09:58,956 INFO util.GSet: capacity = 2^20 = 1048576 entries
2024-11-18 08:09:58,963 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2024-11-18 08:09:58,963 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2024-11-18 08:09:58,963 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-11-18 08:09:58,967 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2024-11-18 08:09:58,967 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000
milliseconds
2024-11-18 08:09:58,970 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-11-18 08:09:58,970 INFO util.GSet: VM type = 64-bit
2024-11-18 08:09:58,970 INFO util.GSet: 0.029999999329447746% max memory 2.6 GB = 811.0 KB
2024-11-18 08:09:58,970 INFO util.GSet: capacity = 2^17 = 131072 entries
2024-11-18 08:09:58,997 INFO namenode.FSImage: Allocated new BlockPoolId: BP-712404385-172.17.0.2-1731917398991
2024-11-18 08:09:59,012 INFO common.Storage: Storage directory /hadoop/dfs/name has been successfully formatted.
2024-11-18 08:09:59,039 INFO namenode.FSImageFormatProtobuf: Saving image file /hadoop/dfs/name/current/fsimage.ckpt_000000000000000000
0 using no compression
2024-11-18 08:09:59,136 INFO namenode.FSImageFormatProtobuf: Image file /hadoop/dfs/name/current/fsimage.ckpt_00000000000000000000 of si
ze 399 bytes saved in 0 seconds .
2024-11-18 08:09:59,145 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-11-18 08:09:59,151 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-11-18 08:09:59,151 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at 9f3066d6281e/172.17.0.2
*****/
root@9f3066d6281e:/#
```

## hdfs namenode &

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
2024-11-18 08:10:23,806 INFO namenode.FSImage: Loaded image for txid 0 from /hadoop/dfs/name/current/fsimage_00000000000000000000
2024-11-18 08:10:23,809 INFO namenode.FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=false, isRollingUpgrade=f
alse)
2024-11-18 08:10:23,810 INFO namenode.FSEditLog: Starting log segment at 1
2024-11-18 08:10:23,899 INFO namenode.NameCache: initialized with 0 entries 0 lookups
2024-11-18 08:10:23,899 INFO namenode.FSNamesystem: Finished loading FSImage in 293 msec
2024-11-18 08:10:24,108 INFO namenode.NameNode: RPC server is binding to 0.0.0.0:8020
2024-11-18 08:10:24,131 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 1000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:10:24,139 INFO ipc.Server: Starting Socket Reader #1 for port 8020
2024-11-18 08:10:24,272 INFO namenode.FSNamesystem: Registered FSNamesystemState, ReplicatedBlocksState and ECBlockGroupsState MBeans.
2024-11-18 08:10:24,279 INFO namenode.LeaseManager: Number of blocks under construction: 0
2024-11-18 08:10:24,286 INFO blockmanagement.BlockManager: initializing replication queues
2024-11-18 08:10:24,286 INFO hdfs.StateChange: STATE* Leaving safe mode after 0 secs
2024-11-18 08:10:24,287 INFO hdfs.StateChange: STATE* Network topology has 0 racks and 0 datanodes
2024-11-18 08:10:24,287 INFO hdfs.StateChange: STATE* UnderReplicatedBlocks has 0 blocks
2024-11-18 08:10:24,297 INFO blockmanagement.BlockManager: Total number of blocks = 0
2024-11-18 08:10:24,297 INFO blockmanagement.BlockManager: Number of invalid blocks = 0
2024-11-18 08:10:24,297 INFO blockmanagement.BlockManager: Number of under-replicated blocks = 0
2024-11-18 08:10:24,297 INFO blockmanagement.BlockManager: Number of over-replicated blocks = 0
2024-11-18 08:10:24,298 INFO blockmanagement.BlockManager: Number of blocks being written = 0
2024-11-18 08:10:24,298 INFO hdfs.StateChange: STATE* Replication Queue initialization scan for invalid, over- and under-replicated blo
cks completed in 12 msec
2024-11-18 08:10:24,309 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:10:24,310 INFO ipc.Server: IPC Server listener on 8020: starting
2024-11-18 08:10:24,313 INFO namenode.NameNode: NameNode RPC up at: 9f3066d6281e/172.17.0.2:8020
2024-11-18 08:10:24,316 INFO namenode.FSNamesystem: Starting services required for active state
2024-11-18 08:10:24,316 INFO namenode.FSDirectory: Initializing quota with 4 thread(s)
2024-11-18 08:10:24,322 INFO namenode.FSDirectory: Quota initialization completed in 6 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2024-11-18 08:10:24,327 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds
```

## hdfs datanode &

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
8991: 4ms
2024-11-18 08:11:03,413 INFO datanode.VolumeScanner: Now scanning bpid BP-712404385-172.17.0.2-1731917398991 on volume /tmp/hadoop-root/dfs/data
2024-11-18 08:11:03,415 INFO datanode.VolumeScanner: VolumeScanner(/tmp/hadoop-root/dfs/data, DS-ee42f06c-4fc3-4af1-99e8-617e2e720e9b): finished scanning block pool BP-712404385-172.17.0.2-1731917398991
2024-11-18 08:11:03,428 INFO datanode.VolumeScanner: VolumeScanner(/tmp/hadoop-root/dfs/data, DS-ee42f06c-4fc3-4af1-99e8-617e2e720e9b): no suitable block pools found to scan. Waiting 1814399985 ms.
2024-11-18 08:11:03,431 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting at 11/18/24 12:34 PM with interval of 21600000ms
2024-11-18 08:11:03,437 INFO datanode.DataNode: Block pool BP-712404385-172.17.0.2-1731917398991 (Datanode Uuid b5818208-c523-4fd6-9bd7-83b6c8a4c1bb) service to 9f3066d6281e/172.17.0.2:8020 beginning handshake with NN
2024-11-18 08:11:03,468 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(172.17.0.2:9866, datanodeUuid=b5818208-c523-4fd6-9bd7-83b6c8a4c1bb, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-f5cf278d-c772-49af-bca9-a1cddb6d301d;nsid=1474860137;c=1731917398991) storage b5818208-c523-4fd6-9bd7-83b6c8a4c1bb
2024-11-18 08:11:03,470 INFO datanode.Net.NetworkTopology: Adding a new node: /default-rack/172.17.0.2:9866
2024-11-18 08:11:03,470 INFO blockmanagement.BlockReportLeaseManager: Registered DN b5818208-c523-4fd6-9bd7-83b6c8a4c1bb (172.17.0.2:9866).
2024-11-18 08:11:03,481 INFO datanode.DataNode: Block pool Block pool BP-712404385-172.17.0.2-1731917398991 (Datanode Uuid b5818208-c523-4fd6-9bd7-83b6c8a4c1bb) service to 9f3066d6281e/172.17.0.2:8020 successfully registered with NN
2024-11-18 08:11:03,481 INFO datanode.DataNode: For namenode 9f3066d6281e/172.17.0.2:8020 using BLOCKREPORT_INTERVAL of 21600000msec CA ChereportInterval of 10000msec Initial delay: 0msec; heartbeatInterval=3000
2024-11-18 08:11:03,532 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-ee42f06c-4fc3-4af1-99e8-617e2e720e9b for DN 172.17.0.2:9866
2024-11-18 08:11:03,562 INFO BlockStateChange: BLOCK* processReport 0xfea286c194c9c2e0: Processing first storage report for DS-ee42f06c-4fc3-4af1-99e8-617e2e720e9b from datanode b5818208-c523-4fd6-9bd7-83b6c8a4c1bb
2024-11-18 08:11:03,564 INFO BlockStateChange: BLOCK* processReport 0xfea286c194c9c2e0: from storage DS-ee42f06c-4fc3-4af1-99e8-617e2e720e9b node DatanodeRegistration(172.17.0.2:9866, datanodeUuid=b5818208-c523-4fd6-9bd7-83b6c8a4c1bb, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-f5cf278d-c772-49af-bca9-a1cddb6d301d;nsid=1474860137;c=1731917398991), blocks: 0, hasStaleStorage: false, processing time: 2 msecs, invalidatedBlocks: 0
2024-11-18 08:11:03,590 INFO datanode.DataNode: Successfully sent lock report 0xfea286c194c9c2e0, containing 1 storage report(s), of which we sent 1. The reports had 0 total blocks and used 1 RPC(s). This took 2 msec to generate and 40 msecs for RPC and NN processing. Got back one command: FinalizeCommand/5.
2024-11-18 08:11:03,590 INFO datanode.DataNode: Got finalize command for block pool BP-712404385-172.17.0.2-1731917398991
```

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
share/hadoop/yarn/hadoop-yarn-common-3.2.1.jar!/webapps/static,AVAILABLE}
Nov 18, 2024 8:11:49 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices as a root resource class
Nov 18, 2024 8:11:49 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class
Nov 18, 2024 8:11:49 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Nov 18, 2024 8:11:49 AM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19 02/11/2015 03:25 AM'
Nov 18, 2024 8:11:50 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
Nov 18, 2024 8:11:50 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
Nov 18, 2024 8:11:50 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the scope "Singleton"
2024-11-18 08:11:50,468 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@ce9b9a9{/,file:///tmp/jetty-0.0.0.0-8042-node-__any-6210024053321463913.dir/webapp/,AVAILABLE}{/node}
2024-11-18 08:11:50,473 INFO server.AbstractConnector: Started ServerConnector@167279d1{HTTP/1.1,[http/1.1]}{0.0.0.0:8042}
2024-11-18 08:11:50,473 INFO server.Server: Started @3147ms
2024-11-18 08:11:50,473 INFO webapp.WebApps: Web app node started at 8042
2024-11-18 08:11:50,475 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : 9f3066d6281e:37635
2024-11-18 08:11:50,476 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-11-18 08:11:50,488 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8031
2024-11-18 08:11:50,534 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM container statuses: []
2024-11-18 08:11:50,547 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers: []
2024-11-18 08:11:51,591 INFO ipc.Client: Retrying connect to server: 0.0.0.0/0.0.0.0:8031. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-11-18 08:11:52,593 INFO ipc.Client: Retrying connect to server: 0.0.0.0/0.0.0.0:8031. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-11-18 08:11:53,595 INFO ipc.Client: Retrying connect to server: 0.0.0.0/0.0.0.0:8031. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
```



start-all.sh

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenodelatest /bin/bash
2024-11-18 08:12:21,498 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2024-11-18 08:12:21,500 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server
2024-11-18 08:12:21,500 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,500 INFO ipc.Server: IPC Server listener on 8031: starting
2024-11-18 08:12:21,529 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-11-18 08:12:21,540 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:12:21,545 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-11-18 08:12:21,550 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server
2024-11-18 08:12:21,551 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,551 INFO ipc.Server: IPC Server listener on 8030: starting
2024-11-18 08:12:21,635 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:12:21,637 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-11-18 08:12:21,641 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
2024-11-18 08:12:21,642 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,643 INFO ipc.Server: IPC Server listener on 8032: starting
2024-11-18 08:12:21,659 INFO resourcemanager.ResourceManager: Transitioned to active state
2024-11-18 08:12:31,100 INFO resourcemanager.ResourceTrackerService: NodeManager from node 9f3066d6281e(cmfPort: 37635 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId 9f3066d6281e:37635
2024-11-18 08:12:31,120 INFO rmnode.RMNodeImpl: 9f3066d6281e:37635 Node Transitioned from NEW to RUNNING
2024-11-18 08:12:31,174 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id 1962839724
2024-11-18 08:12:31,179 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id 1378724301
2024-11-18 08:12:31,183 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as 9f3066d6281e:37635 with total resource of <memory:8192, vCores:8>
2024-11-18 08:12:31,220 INFO capacity.CapacityScheduler: Added node 9f3066d6281e:37635 clusterResource: <memory:8192, vCores:8>

root@9f3066d6281e:/# start-all.sh
bash: start-all.sh: command not found
root@9f3066d6281e:/#

Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenodelatest /bin/bash
for class org.apache.hadoop.yarn.nodelabels.CommonNodeLabelsManager$ForwardingEventHandler
2024-11-18 08:12:21,459 INFO store.AbstractFSNodeStore: Created store directory :file:/tmp/hadoop-yarn-root/node-attribute
2024-11-18 08:12:21,472 INFO store.AbstractFSNodeStore: Finished write mirror at:file:/tmp/hadoop-yarn-root/node-attribute/nodeattribut
e.mirror
2024-11-18 08:12:21,472 INFO store.AbstractFSNodeStore: Finished create editlog file at:file:/tmp/hadoop-yarn-root/node-attribute/nodea
ttribute.editlog
2024-11-18 08:12:21,483 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.server.resourcemanager.nodelabels.NodeAttr
ibutesStoreEventType for class org.apache.hadoop.yarn.server.resourcemanager.nodelabels.NodeAttributesManagerImpl$ForwardingHandle
r
2024-11-18 08:12:21,484 INFO placement.MultiNodeSortingManager: Starting NodeSortingService=MultiNodeSortingManager
2024-11-18 08:12:21,497 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:12:21,498 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2024-11-18 08:12:21,500 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the serv
er
2024-11-18 08:12:21,500 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,500 INFO ipc.Server: IPC Server listener on 8031: starting
2024-11-18 08:12:21,529 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-11-18 08:12:21,540 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:12:21,545 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-11-18 08:12:21,550 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the s
erver
2024-11-18 08:12:21,551 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,551 INFO ipc.Server: IPC Server listener on 8030: starting
2024-11-18 08:12:21,635 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:12:21,637 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-11-18 08:12:21,641 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the s
erver
2024-11-18 08:12:21,642 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,643 INFO ipc.Server: IPC Server listener on 8032: starting
2024-11-18 08:12:21,659 INFO resourcemanager.ResourceManager: Transitioned to active state
```

hdfs dfs -mkdir -p /user/hadoop/input

```
Command Prompt - docker run -it --name hadoop-cluster -p 9070:9070 -p 8080:8080 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
2024-11-18 08:12:21,500 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,500 INFO ipc.Server: IPC Server listener on 8031: starting
2024-11-18 08:12:21,529 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-11-18 08:12:21,540 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:12:21,545 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-11-18 08:12:21,550 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the s
erver
2024-11-18 08:12:21,551 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,551 INFO ipc.Server: IPC Server listener on 8030: starting
2024-11-18 08:12:21,635 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000
, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:12:21,637 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-11-18 08:12:21,641 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the s
erver
2024-11-18 08:12:21,642 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:12:21,643 INFO ipc.Server: IPC Server listener on 8032: starting
2024-11-18 08:12:21,659 INFO resourcemanager.ResourceManager: Transitioned to active state
2024-11-18 08:12:31,100 INFO resourcemanager.ResourceTrackerService: NodeManager from node 9f3066d6281e(cmPort: 37635 httpPort: 8042) r
egistered with capability: <memory:8192, vCores:8>, assigned nodeId 9f3066d6281e:37635
2024-11-18 08:12:31,120 INFO rmnode.RMNodeImpl: 9f3066d6281e:37635 Node Transitioned from NEW to RUNNING
2024-11-18 08:12:31,174 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id 196283972
4
2024-11-18 08:12:31,179 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id 1378724301
2024-11-18 08:12:31,183 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as 9f3066d6281e:37635 with total resour
ce of <memory:8192, vCores:8>
2024-11-18 08:12:31,220 INFO capacity.CapacityScheduler: Added node 9f3066d6281e:37635 clusterResource: <memory:8192, vCores:8>

root@9f3066d6281e:/# start-all.sh
bash: start-all.sh: command not found
root@9f3066d6281e:/# hdfs dfs -mkdir -p /user/hadoop/input
2024-11-18 08:13:17,624 INFO namenode.FSEditLog: Number of transactions: 4 Total time for transactions(ms): 11 Number of transactions b
atched in Syncs: 0 Number of syncs: 2 SyncTimes(ms): 10
root@9f3066d6281e:/#
```

hdfs dfs -put \$HADOOP\_HOME/etc/hadoop/\*.xml /user/hadoop/input

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
991:blk_1073741831_1007, duration(ns): 4714042
2024-11-18 08:13:47,119 INFO datanode.DataNode: PacketResponder: BP-712404385-172.17.0.2-1731917398991:blk_1073741831_1007, type=LAST_I
N_PIPELINE terminating
2024-11-18 08:13:47,122 INFO namenode.FSNamesystem: BLOCK* blk_1073741831_1007 is COMMITTED but not COMPLETE(numNodes= 0 < minimum = 1
) in file /user/hadoop/input/kms-site.xml._COPYING_
2024-11-18 08:13:47,528 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/kms-site.xml._COPYING_ is closed by DFSCClient_NONM
APREDUCE_1769506109_1
2024-11-18 08:13:47,582 INFO hdfs.StateChange: BLOCK* allocate blk_1073741832_1008, replicas=172.17.0.2:9866 for /user/hadoop/input/map
red-site.xml._COPYING_
2024-11-18 08:13:47,594 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = fa
lse
2024-11-18 08:13:47,598 INFO datanode.DataNode: Receiving BP-712404385-172.17.0.2-1731917398991:blk_1073741832_1008 src: /172.17.0.2:44
136 dest: /172.17.0.2:9866
2024-11-18 08:13:47,619 INFO DataNode.clienttrace: src: /172.17.0.2:44136, dest: /172.17.0.2:9866, bytes: 841, op: HDFS_WRITE, cliID: D
FSCClient_NONMAPREDUCE_1769506109_1, offset: 0, srvID: b5818208-c523-4fd6-9bd7-83b6c8a4c1bb, blockid: BP-712404385-172.17.0.2-1731917398
991:blk_1073741832_1008, duration(ns): 13822489
2024-11-18 08:13:47,622 INFO datanode.DataNode: PacketResponder: BP-712404385-172.17.0.2-1731917398991:blk_1073741832_1008, type=LAST_I
N_PIPELINE terminating
2024-11-18 08:13:47,628 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/mapred-site.xml._COPYING_ is closed by DFSCClient_N
ONMAPREDUCE_1769506109_1
2024-11-18 08:13:47,680 INFO hdfs.StateChange: BLOCK* allocate blk_1073741833_1009, replicas=172.17.0.2:9866 for /user/hadoop/input/yar
n-site.xml._COPYING_
2024-11-18 08:13:47,689 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = fa
lse
2024-11-18 08:13:47,693 INFO datanode.DataNode: Receiving BP-712404385-172.17.0.2-1731917398991:blk_1073741833_1009 src: /172.17.0.2:44
144 dest: /172.17.0.2:9866
2024-11-18 08:13:47,708 INFO DataNode.clienttrace: src: /172.17.0.2:44144, dest: /172.17.0.2:9866, bytes: 1031, op: HDFS_WRITE, cliID:
DFSCClient_NONMAPREDUCE_1769506109_1, offset: 0, srvID: b5818208-c523-4fd6-9bd7-83b6c8a4c1bb, blockid: BP-712404385-172.17.0.2-173191739
8991:blk_1073741833_1009, duration(ns): 8853318
2024-11-18 08:13:47,709 INFO datanode.DataNode: PacketResponder: BP-712404385-172.17.0.2-1731917398991:blk_1073741833_1009, type=LAST_I
N_PIPELINE terminating
2024-11-18 08:13:47,716 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/yarn-site.xml._COPYING_ is closed by DFSCClient_NON
MAPREDUCE_1769506109_1
root@9f3066d6281e:/#
```

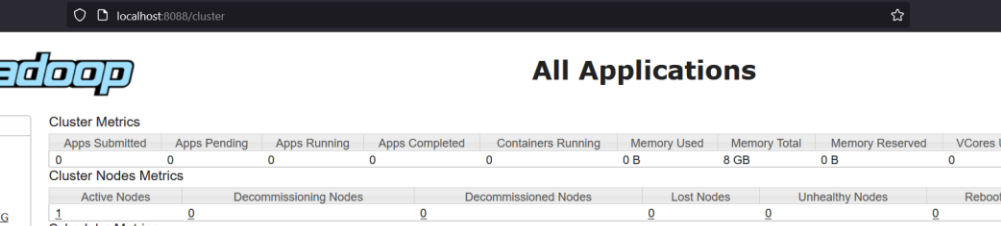
hadoop jar \$HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-\*.jar wordcount  
/user/hadoop/input /user/hadoop/output

```
Command Prompt - docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
HDFS: Number of read operations=143
HDFS: Number of large read operations=0
HDFS: Number of write operations=12
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=751
  Map output records=3063
  Map output bytes=38519
  Map output materialized bytes=21327
  Input split bytes=1104
  Combine input records=3063
  Combine output records=1223
  Reduce input groups=588
  Reduce shuffle bytes=21327
  Reduce input records=1223
  Reduce output records=588
  Spilled Records=2446
  Shuffled Maps =9
  Failed Shuffles=0
  Merged Map outputs=9
  GC time elapsed (ms)=5
  Total committed heap usage (bytes)=7185367040
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=28589
File Output Format Counters
  Bytes Written=10327
root@9f3066d6281e:/#
```

```

type="text/xsl" 4
u:user:%user 1
under 28
unique 1
updating 1
use 11
used 24
user 48
user. 2
user? 1
users 24
users,wheel". 21
uses 2
using 1
v2 1
valid. 1
value 30
value, 2
values 2
version="1.0" 5
version="1.0"?> 3
via 1
well 1
when 3
which 10
while 1
who 3
will 12
with 28
work 1
writing, 9
you 10
zero 2
root@9f3066d6281e:/#

```



The screenshot shows the Hadoop web interface at localhost:8088/cluster. The main heading is "All Applications". On the left, there's a sidebar menu with options like Cluster, About Nodes, Node Labels, Applications, NEW, NEW\_SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED, Scheduler, and Tools. The main content area displays several metrics tables:

- Cluster Metrics**: A single row showing various cluster-wide statistics.
- Cluster Nodes Metrics**: A table with columns for Active Nodes, Decommissioning Nodes, Decommissioned Nodes, Lost Nodes, Unhealthy Nodes, and Rebooted Nodes.
- Scheduler Metrics**: A table showing Capacity Scheduler details, including Scheduling Resource Type, Minimum Allocation, Maximum Allocation, and Maximum Cluster Application.
- Applications Table**: A large table listing individual applications with columns for ID, User, Name, Application Type, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, FinalStatus, Running Containers, Allocated CPU VCores, Allocated Memory MB, Reserved CPU VCores, Reserved Memory MB, % of Queue, % of Cluster, and Progress.

In the screenshot, the Applications table is currently empty, displaying "Showing 0 to 0 of 0 entries".

http://localhost:9870

Problem loading pageAll ApplicationsNameNode information

localhost:9870/dfshealth.html#tab-overview

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

### Overview '9f3066d6281e:8020' (active)

Started:	Mon Nov 18 13:40:23 +0530 2024
Version:	3.2.1, rb3cbbb467e22ea829b38084b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-f5cf278d-c772-49af-bca9-a1cddb6d301d
Block Pool ID:	BP-712404385-172.17.0.2-1731917398991

### Summary

Security is off.  
Safemode is off.  
16 files and directories, 10 blocks (10 replicated blocks, 0 erasure coded block groups) = 26 total filesystem object(s).  
Heap Memory used 148.03 MB of 407.5 MB Heap Memory. Max Heap Memory is 2.58 GB.  
Non Heap Memory used 49.93 MB of 51.19 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	1006.85 GB
Configured Remote Capacity:	0 B
DFS Used:	140 KB (0%)

Type here to search

1:58 PM11/18/2024