

Apache Hadoop

Aim: Installing and running a MapReduce example on a Hadoop cluster.

Note:


- There are three ways of using Apache Hadoop; a standalone (local) installation, a pseudo cluster installation and a cluster installation.
- We will be performing a pseudo cluster installation here.
- Uninstall [YARN Package Manager \(Node.js\)](#) if it is installed as it conflicts with the YARN provided by Hadoop.

Installation Instructions:

➔ **Prerequisites:**

- Java version 1.8 (Install JDK if you want to write (and compile) JAR files else install a JRE).
- An archiver software (It is optional for Windows 10 as it can extract both tar.gz and zip files)

➔ Download binaries from the [releases](#) page. It is around ~ 600 MB. The version I've chosen is 3.3.2.


Apache Hadoop
Download
Documentation
Community
Development
Help
Apache Software Foundation

Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

Version	Release date	Source download	Binary download	Release notes
3.3.2	2022 Mar 3	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.2.2	2021 Jan 9	source (checksum signature)	binary (checksum signature)	Announcement
2.10.1	2020 Sep 21	source (checksum signature)	binary (checksum signature)	Announcement

To verify Hadoop releases using GPG:

1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the signature file `hadoop-X.Y.Z-src.tar.gz.asc` from [Apache](#).
3. Download the [Hadoop KEYS](#) file.
4. `gpg --import KEYS`
5. `gpg --verify hadoop-X.Y.Z-src.tar.gz.asc`

To perform a quick check using SHA-512:

1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the checksum `hadoop-X.Y.Z-src.tar.gz.sha512` or `hadoop-X.Y.Z-src.tar.gz.mds` from [Apache](#).
3. `shasum -a 512 hadoop-X.Y.Z-src.tar.gz`

➔ Extract these files using an archiver software of your choice. You should see the following directory structure after completion (It may take a few moments).

Name	Date modified	Type	Size
bin	06-03-2022 10:25	File folder	
data	06-03-2022 10:48	File folder	
etc	22-02-2022 00:14	File folder	
include	22-02-2022 02:12	File folder	
lib	22-02-2022 02:12	File folder	
libexec	22-02-2022 02:12	File folder	
licenses-binary	22-02-2022 02:12	File folder	
logs	06-03-2022 12:30	File folder	
sbin	22-02-2022 00:14	File folder	
share	22-02-2022 03:22	File folder	
LICENSE	15-01-2022 22:54	Text Document	15 KB
LICENSE-binary	15-01-2022 22:54	File	23 KB
NOTICE	02-12-2021 06:41	Text Document	2 KB
NOTICE-binary	31-01-2022 23:47	File	29 KB
README	02-12-2021 06:41	Text Document	1 KB

➔ Now, the native IO libraries that Hadoop provides do **not** work on Windows. So we need to add the libraries from [here](#). Download the repository (or clone it, your choice) and copy the binary files to the `bin/` subdirectory in the extracted Hadoop archive.

master
winutils / hadoop-3.2.2 / bin /
Go to file
Add file
...

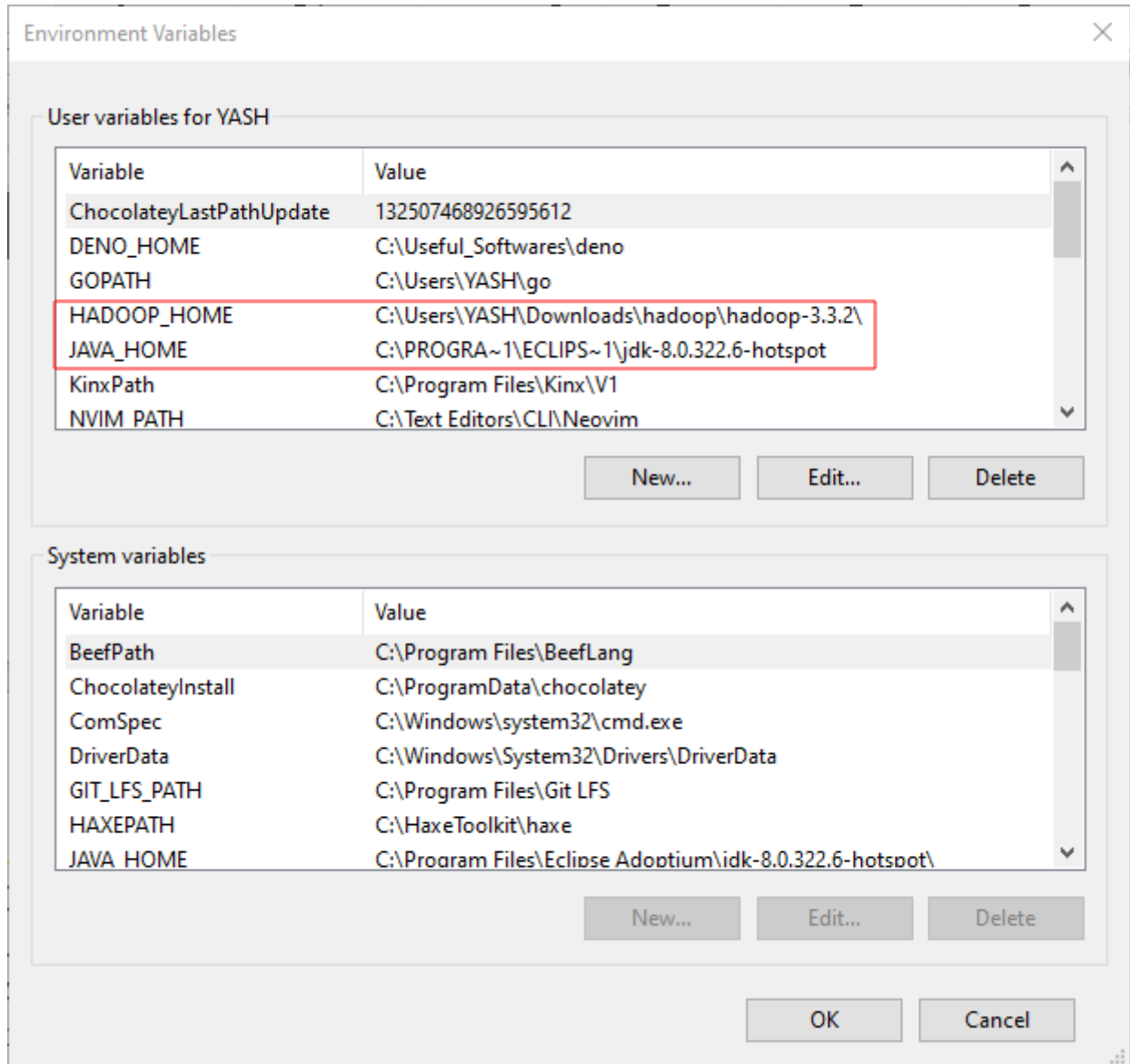
HanZhao-ii compile hadoop-3.2.2
2def4b7 on 13 Apr 2021
History

..

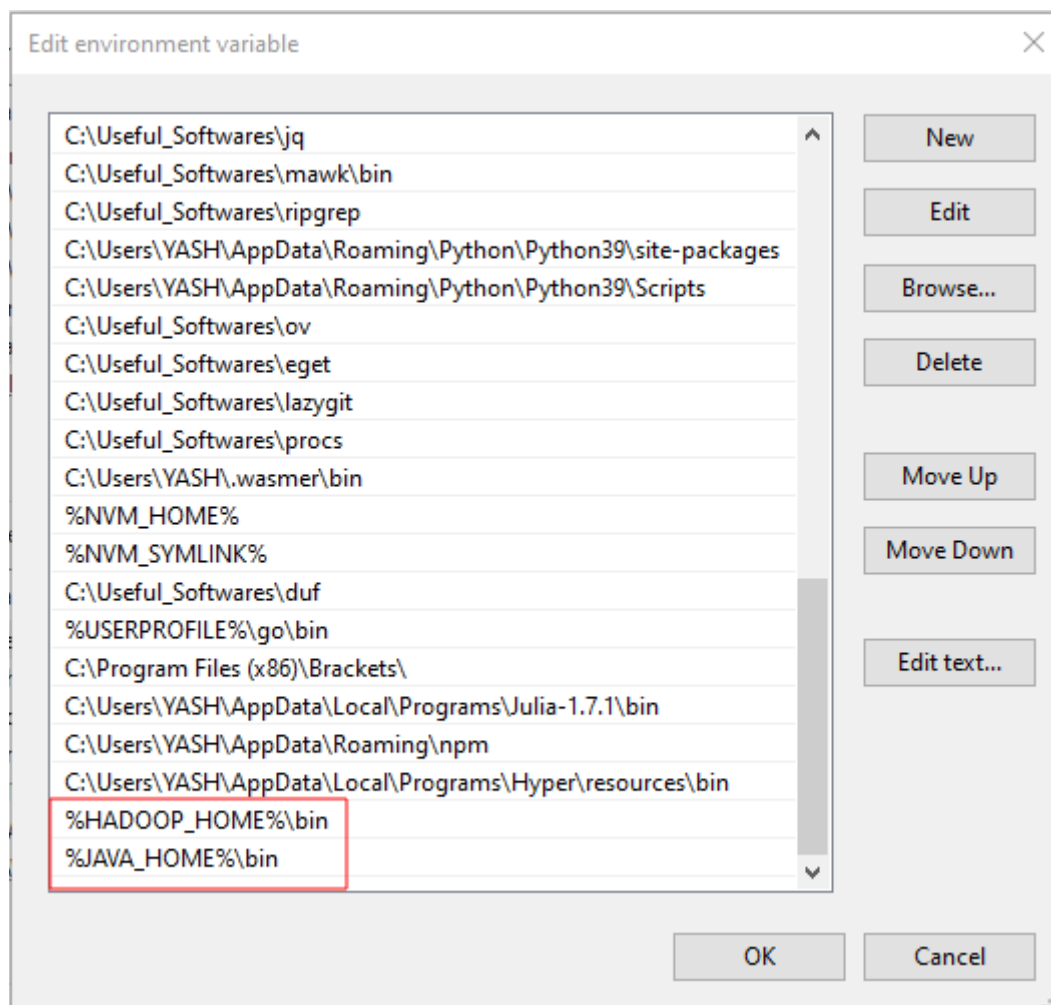
hadoop	compile hadoop-3.2.2	11 months ago
hadoop.cmd	compile hadoop-3.2.2	11 months ago
hadoop.dll	compile hadoop-3.2.2	11 months ago
hadoop.exp	compile hadoop-3.2.2	11 months ago
hadoop.lib	compile hadoop-3.2.2	11 months ago
hadoop.pdb	compile hadoop-3.2.2	11 months ago
hdfs	compile hadoop-3.2.2	11 months ago
hdfs.cmd	compile hadoop-3.2.2	11 months ago
libwinutils.lib	compile hadoop-3.2.2	11 months ago
mapred	compile hadoop-3.2.2	11 months ago
mapred.cmd	compile hadoop-3.2.2	11 months ago
winutils.exe	compile hadoop-3.2.2	11 months ago
winutils.pdb	compile hadoop-3.2.2	11 months ago
yarn	compile hadoop-3.2.2	11 months ago
yarn.cmd	compile hadoop-3.2.2	11 months ago

Name	Date modified	Type	Size
container-executor	22-02-2022 01:45	File	785 KB
hadoop	29-09-2021 13:42	File	9 KB
hadoop	29-09-2021 13:42	Windows Comma...	11 KB
hadoop.dll	29-09-2021 13:42	Application exten...	94 KB
hadoop.exp	29-09-2021 13:42	EXP File	25 KB
hadoop.lib	29-09-2021 13:42	LIB File	41 KB
hadoop.pdb	29-09-2021 13:42	PDB File	820 KB
hdfs	29-09-2021 13:42	File	12 KB
hdfs	29-09-2021 13:42	Windows Comma...	8 KB
libwinutils.lib	29-09-2021 13:42	LIB File	1,561 KB
mapred	29-09-2021 13:42	File	7 KB
mapred	29-09-2021 13:42	Windows Comma...	6 KB
oom-listener	22-02-2022 01:45	File	33 KB
test-container-executor	22-02-2022 01:45	File	818 KB
winutils	29-09-2021 13:42	Application	116 KB
winutils.pdb	29-09-2021 13:42	PDB File	1,324 KB
yarn	29-09-2021 13:42	File	12 KB
yarn	29-09-2021 13:42	Windows Comma...	13 KB

- ➔ Move the extracted directory to a location (lets call it HADOOP_HOME_DIR).
- ➔ Now, set the JAVA_HOME and HADOOP_HOME environment variable. Check [this website](#) for instructions regarding adding environent variables. Add JAVA_HOME\bin and HADOOP_HOME\bin to the PATH.

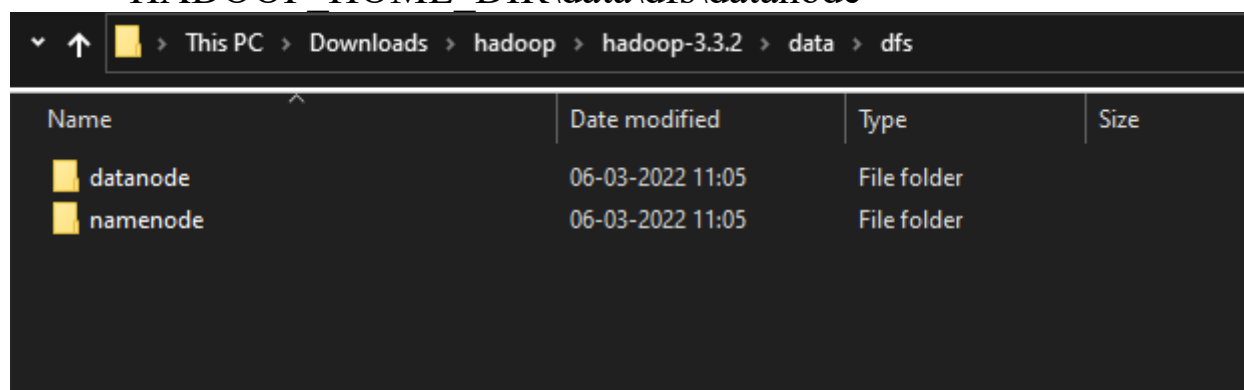


Use Windows 8.3 notation if the path contains spaces or other special characters.



➔ Now, create the following two directories:

- HADOOP_HOME_DIR\data\dfs\namenode
- HADOOP_HOME_DIR\data\dfs\datanode



- ➔ Now, open the 'hdfs-site.xml' file from %HADOOP_HOME%\etc\hadoop directory and add the following within the <configuration></configuration> tags.

```
<property>
    <name>dfs.replication</name>
    <value>1</value>
</property>
<property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///HADOOP_HOME_DIR/data/dfs/namenode</value>
</property>
<property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///HADOOP_HOME_DIR/data/dfs/datanode</value>
</property>
```

- ➔ Now, open the 'core-site.xml' file from %HADOOP_HOME%\etc\hadoop directory and add the following within the <configuration></configuration> tags.

```
<property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9820</value>
</property>
```

- ➔ Open the 'mapred-site.xml' file from %HADOOP_HOME%\etc\hadoop directory and add the following within the <configuration></configuration> tags.

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
  <description>MapReduce framework name</description>
</property>
```

- ➔ Now, open the 'yarn-site.xml' file from %HADOOP_HOME%\etc\hadoop directory and add the following within the <configuration></configuration> tags.

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
  <description>Yarn Node Manager Aux Service</description>
</property>
```

- ➔ Now, we format the namenode using the following command:
`hdfs namenode -format`

```
2020-04-17 22:02:58,422 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2020-04-17 22:02:58,423 INFO util.GSet: VM type = 64-bit
2020-04-17 22:02:58,424 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
2020-04-17 22:02:58,425 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory root= E:\hadoop-env\hadoop-3.2.1\data\dfs\namenode; location= null ? (Y or N)
y
```

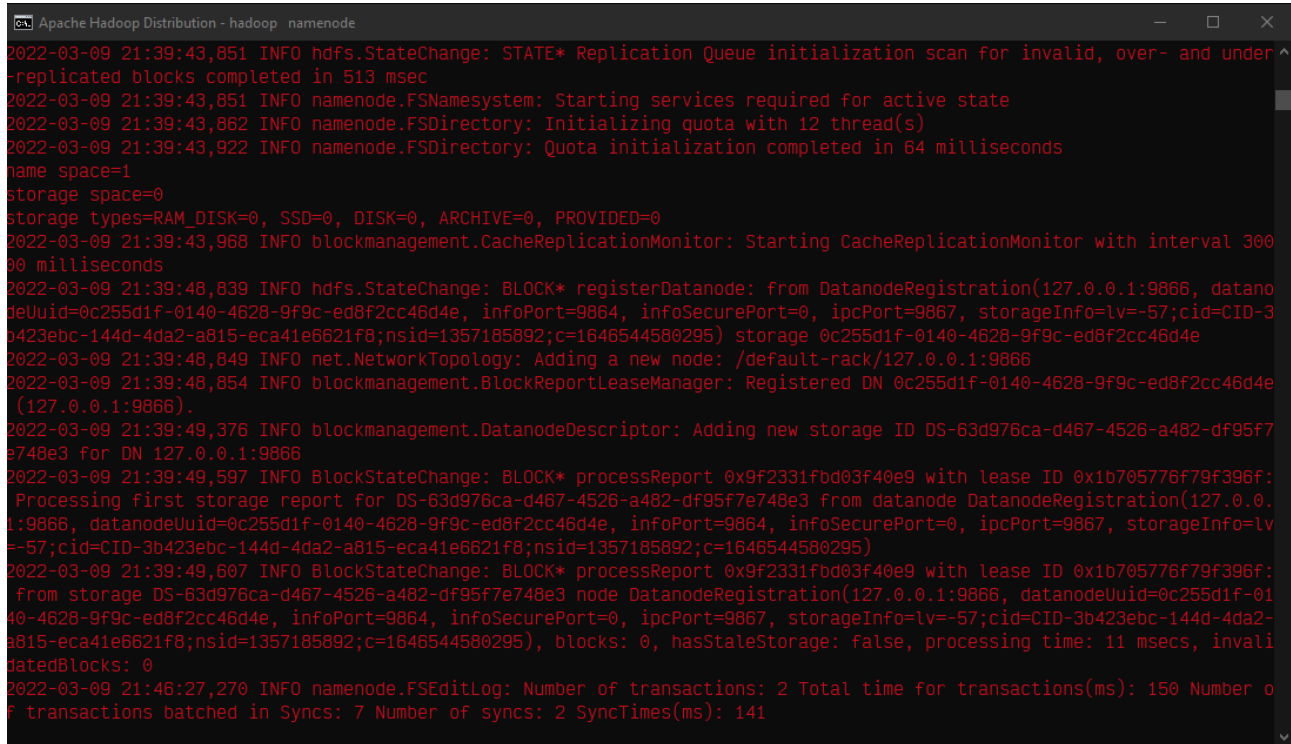
Key in 'Y' when it asks for formatting the filesystem. If it runs successfully, you should see this screen.

```
2020-04-17 22:14:17,206 INFO namenode.FSImage: Allocated new BlockPoolId: BP-2032026115-192.168.1.105-1587150857190
2020-04-17 22:14:17,207 INFO common.Storage: Will remove files: []
2020-04-17 22:14:17,275 INFO common.Storage: Storage directory E:\hadoop-env\hadoop-3.2.1\data\dfs\namenode has been successfully formatted.
2020-04-17 22:14:17,331 INFO namenode.FSImageFormatProtobuf: Saving image file E:\hadoop-env\hadoop-3.2.1\data\dfs\namenode\current\fsimage.ckpt_000000000000000000 using no compression
2020-04-17 22:14:17,531 INFO namenode.FSImageFormatProtobuf: Image file E:\hadoop-env\hadoop-3.2.1\data\dfs\namenode\current\fsimage.ckpt_000000000000000000 of size 400 bytes saved in 0 seconds .
2020-04-17 22:14:17,555 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2020-04-17 22:14:17,580 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2020-04-17 22:14:17,580 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at
*****/
PS C:\Windows\system32>
```

- ➔ To start HDFS, navigate to %HADOOP_HOME%\sbin and run the start-dfs.cmd file.

```
C:\Users\YASH>cd %HADOOP_HOME%\sbin  
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\sbin>start-dfs
```

It should spawn two Command Prompt windows (one for namenode and the other for datanode).



```
Apache Hadoop Distribution - hadoop namenode  
2022-03-09 21:39:43,851 INFO hdfs.StateChange: STATE* Replication Queue initialization scan for invalid, over- and under-  
-replicated blocks completed in 513 msec  
2022-03-09 21:39:43,851 INFO namenode.FSNamesystem: Starting services required for active state  
2022-03-09 21:39:43,862 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)  
2022-03-09 21:39:43,922 INFO namenode.FSDirectory: Quota initialization completed in 64 milliseconds  
name space=1  
storage space=0  
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0  
2022-03-09 21:39:43,968 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 300  
00 milliseconds  
2022-03-09 21:39:48,839 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(127.0.0.1:9866, datano  
deUuid=0c255d1f-0140-4628-9f9c-ed8f2cc46d4e, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-3  
b423ebc-144d-4da2-a815-eca41e6621f8;nsid=1357185892;c=1646544580295) storage 0c255d1f-0140-4628-9f9c-ed8f2cc46d4e  
2022-03-09 21:39:48,849 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866  
2022-03-09 21:39:48,854 INFO blockmanagement.BlockReportLeaseManager: Registered DN 0c255d1f-0140-4628-9f9c-ed8f2cc46d4e  
(127.0.0.1:9866).  
2022-03-09 21:39:49,376 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-63d976ca-d467-4526-a482-df95f7  
e748e3 for DN 127.0.0.1:9866  
2022-03-09 21:39:49,597 INFO BlockStateChange: BLOCK* processReport 0x9f2331fbd03f40e9 with lease ID 0x1b705776f79f396f:  
Processing first storage report for DS-63d976ca-d467-4526-a482-df95f7e748e3 from datanode DatanodeRegistration(127.0.0.  
1:9866, datanodeUuid=0c255d1f-0140-4628-9f9c-ed8f2cc46d4e, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv  
=-57;cid=CID-3b423ebc-144d-4da2-a815-eca41e6621f8;nsid=1357185892;c=1646544580295)  
2022-03-09 21:39:49,607 INFO BlockStateChange: BLOCK* processReport 0x9f2331fbd03f40e9 with lease ID 0x1b705776f79f396f:  
from storage DS-63d976ca-d467-4526-a482-df95f7e748e3 node DatanodeRegistration(127.0.0.1:9866, datanodeUuid=0c255d1f-01  
40-4628-9f9c-ed8f2cc46d4e, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-3b423ebc-144d-4da2-  
a815-eca41e6621f8;nsid=1357185892;c=1646544580295), blocks: 0, hasStaleStorage: false, processing time: 11 msecs, invali  
datedBlocks: 0  
2022-03-09 21:46:27,270 INFO namenode.FSEditLog: Number of transactions: 2 Total time for transactions(ms): 150 Number o  
f transactions batched in Syncs: 7 Number of syncs: 2 SyncTimes(ms): 141
```



```
Apache Hadoop Distribution - hadoop datanode
644580295 on volume C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode...
2022-03-09 21:39:48,164 INFO impl.BlockPoolSlice: Replica Cache file: C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode\current\BP-1035977273-192.168.43.145-1646544580295\current\replicas doesn't exist
2022-03-09 21:39:48,173 INFO impl.FsDatasetImpl: Time to add replicas to map for block pool BP-1035977273-192.168.43.145-1646544580295 on volume C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode: 10ms
2022-03-09 21:39:48,175 INFO impl.FsDatasetImpl: Total time to add all replicas to map for block pool BP-1035977273-192.168.43.145-1646544580295: 15ms
2022-03-09 21:39:48,178 INFO checker.ThrottledAsyncChecker: Scheduling a check for C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode
2022-03-09 21:39:48,225 INFO checker.DatasetVolumeChecker: Scheduled health check for volume C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode
2022-03-09 21:39:48,322 INFO datanode.VolumeScanner: VolumeScanner(C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode, DS-63d976ca-d467-4526-a482-df95f7e748e3): no suitable block pools found to scan. Waiting 1517133421 ms.
2022-03-09 21:39:48,339 WARN datanode.DirectoryScanner: dfs.datanode.directoryscan.throttle.limit.ms.per.sec set to value above 1000 ms/sec. Assuming default value of -1
2022-03-09 21:39:48,341 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting in 13352888ms with interval of 21600000ms and throttle limit of -1ms/s
2022-03-09 21:39:48,370 INFO datanode.DataNode: Block pool BP-1035977273-192.168.43.145-1646544580295 (Datanode Uuid 0c255d1f-0140-4628-9f9c-ed8f2cc46d4e) service to localhost/127.0.0.1:9820 beginning handshake with NN
2022-03-09 21:39:48,923 INFO datanode.DataNode: Block pool BP-1035977273-192.168.43.145-1646544580295 (Datanode Uuid 0c255d1f-0140-4628-9f9c-ed8f2cc46d4e) service to localhost/127.0.0.1:9820 successfully registered with NN
2022-03-09 21:39:48,925 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9820 using BLOCKREPORT_INTERVAL of 21600000msecs CACHEREPORT_INTERVAL of 10000msecs Initial delay: 0msecs; heartBeatInterval=3000
2022-03-09 21:39:49,733 INFO datanode.DataNode: Successfully sent block report 0x9f2331fbd03f40e9 with lease ID 0x1b705776f79f396f to namenode: localhost/127.0.0.1:9820, containing 1 storage report(s), of which we sent 1. The reports had 0 total blocks and used 1 RPC(s). This took 21 msecs to generate and 233 msecs for RPC and NN processing. Got back one command: FinalizeCommand/5.
2022-03-09 21:39:49,744 INFO datanode.DataNode: Got finalize command for block pool BP-1035977273-192.168.43.145-1646544580295
```

➔ Now, start the YARN services by running the start-yarn.cmd file.

```
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\sbin>start-yarn
starting yarn daemons
```

This should spawn two more Command Prompt windows (one for resourcemanager and the other for nodemanager)

```
Apache Hadoop Distribution - C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\bin\yarn nodemanager
Mar 09, 2022 9:41:36 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19 02/11/2015 03:25 AM'
Mar 09, 2022 9:41:37 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with
the scope "Singleton"
Mar 09, 2022 9:41:38 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Sin
gleton"
Mar 09, 2022 9:41:41 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the s
cope "Singleton"
2022-03-09 21:41:41,909 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@7979b8b7{node/,/,file:///C:/Users/YASH
/AppData/Local/Temp/jetty-0_0_0-8042-hadoop-yarn-common-3_3_2_jar-_any-1188022385021473506/webapp/,AVAILABLE}{jar:fil
e:/C:/Users/YASH/Downloads/hadoop/hadoop-3.3.2/share/hadoop/yarn/hadoop-yarn-common-3.3.2.jar!/webapps/node}
2022-03-09 21:41:41,961 INFO server.AbstractConnector: Started ServerConnector@7a7471ce{HTTP/1.1, (http/1.1)}{0.0.0.0:80
42}
2022-03-09 21:41:41,964 INFO server.Server: Started @71184ms
2022-03-09 21:41:41,965 INFO webapp.WebApps: Web app node started at 8042
2022-03-09 21:41:41,970 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : LAPTOP-M0821FQC:19650
2022-03-09 21:41:42,021 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2022-03-09 21:41:42,083 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8031
2022-03-09 21:41:45,056 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key wi
th id 6588897223
2022-03-09 21:41:45,085 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id
258164966
2022-03-09 21:41:45,101 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as LAPTOP-M0821FQC:19650
with total resource of <memory:8192, vCores:8>
2022-03-09 21:51:18,074 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public
```

```
Apache Hadoop Distribution - C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\bin\yarn resourcemanager
2022-03-09 21:41:24,108 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTracke
rPB to the server
2022-03-09 21:41:24,136 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2022-03-09 21:41:25,256 INFO ipc.Server: IPC Server Responder: starting
2022-03-09 21:41:25,284 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2022-03-09 21:41:25,259 INFO ipc.Server: IPC Server listener on 8031: starting
2022-03-09 21:41:25,314 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2022-03-09 21:41:25,391 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProt
ocolPB to the server
2022-03-09 21:41:25,398 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2022-03-09 21:41:25,476 INFO ipc.Server: IPC Server listener on 8030: starting
2022-03-09 21:41:25,476 INFO ipc.Server: IPC Server Responder: starting
2022-03-09 21:41:26,575 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2022-03-09 21:41:26,649 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProt
ocolPB to the server
2022-03-09 21:41:26,682 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2022-03-09 21:41:26,927 INFO ipc.Server: IPC Server listener on 8032: starting
2022-03-09 21:41:26,927 INFO ipc.Server: IPC Server Responder: starting
2022-03-09 21:41:36,668 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-9F092a8a-0b8e-462f-91c1-d65105657a1b
2022-03-09 21:41:37,398 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2022-03-09 21:41:37,466 INFO resourcemanager.ResourceManager: Transitioned to active state
2022-03-09 21:41:44,879 INFO resourcemanager.ResourceTrackerService: NodeManager from node LAPTOP-M0821FQC(cmPort: 19650
, httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId LAPTOP-M0821FQC:19650
2022-03-09 21:41:44,913 INFO rmnode.RMNodeImpl: LAPTOP-M0821FQC:19650 Node Transitioned from NEW to RUNNING
2022-03-09 21:41:45,170 INFO capacity.CapacityScheduler: Added node LAPTOP-M0821FQC:19650 clusterResource: <memory:8192,
vCores:8>
2022-03-09 21:51:23,843 INFO scheduler.AbstractYarnScheduler: Release request cache is cleaned up
```

➔ Now, open any web browser and navigate to <http://localhost:9870/dfshealth.html>

Overview 'localhost:9820' (✓active)

Started:	Wed Mar 09 22:22:41 +0530 2022
Version:	3.3.2, r0bcb014209e219273cb6fd4152df7df713cbac61
Compiled:	Tue Feb 22 00:09:00 +0530 2022 by chao from branch-3.3.2
Cluster ID:	CID-3b423ebc-144d-4da2-a815-eca41e6621f8
Block Pool ID:	BP-1035977273-192.168.43.145-1646544580295

Summary

Security is off.

Safemode is off.

25 files and directories, 10 blocks (10 replicated blocks, 0 erasure coded block groups) = 35 total filesystem object(s).

Heap Memory used 79.04 MB of 122 MB Heap Memory. Max Heap Memory is 889 MB.

➔ Also navigate to this link: <http://localhost:9864/datanode.html>

DataNode on LAPTOP-MO821FQC:9866


Cluster ID:	CID-3b423ebc-144d-4da2-a815-eca41e6621f8
Started:	Wed Mar 09 22:23:30 +0530 2022
Version:	3.3.2, r0bcb014209e219273cb6fd4152df7df713cbac61

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9820	BP-1035977273-192.168.43.145-1646544580295	RUNNING	2s	31 minutes	60 B (128 MB)

Volume Information

➔ And finally, navigate to this link: <http://localhost:8088/cluster>



All Applications

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

Cluster Metrics									
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCor	
0	0	0	0	0	0 B	8 GB	0 B	0	

Cluster Nodes Metrics					
Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rel
1	0	0	0	0	

Scheduler Metrics			
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Reserved CPU VCoers	Reserved Memory MB
No data available in table															

Showing 0 to 0 of 0 entries

Source Code:

```
import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {

    public static class TokenizerMapper extends Mapper<Object,
Text, Text, IntWritable>{

        private final static IntWritable one = new
IntWritable(1);
        private Text word = new Text();

        public void map(Object key, Text value, Context context)
throws IOException, InterruptedException {
            StringTokenizer itr = new
StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }
}
```

```

    }

    public static class IntSumReducer extends
Reducer<Text,IntWritable,Text,IntWritable> {
        private IntWritable result = new IntWritable();

        public void reduce(Text key, Iterable<IntWritable>
values, Context context) throws IOException, InterruptedException
{
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }

    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "word count");
        job.setJarByClass(WordCount.class);
        job.setMapperClass(TokenizerMapper.class);
        job.setCombinerClass(IntSumReducer.class);
        job.setReducerClass(IntSumReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}

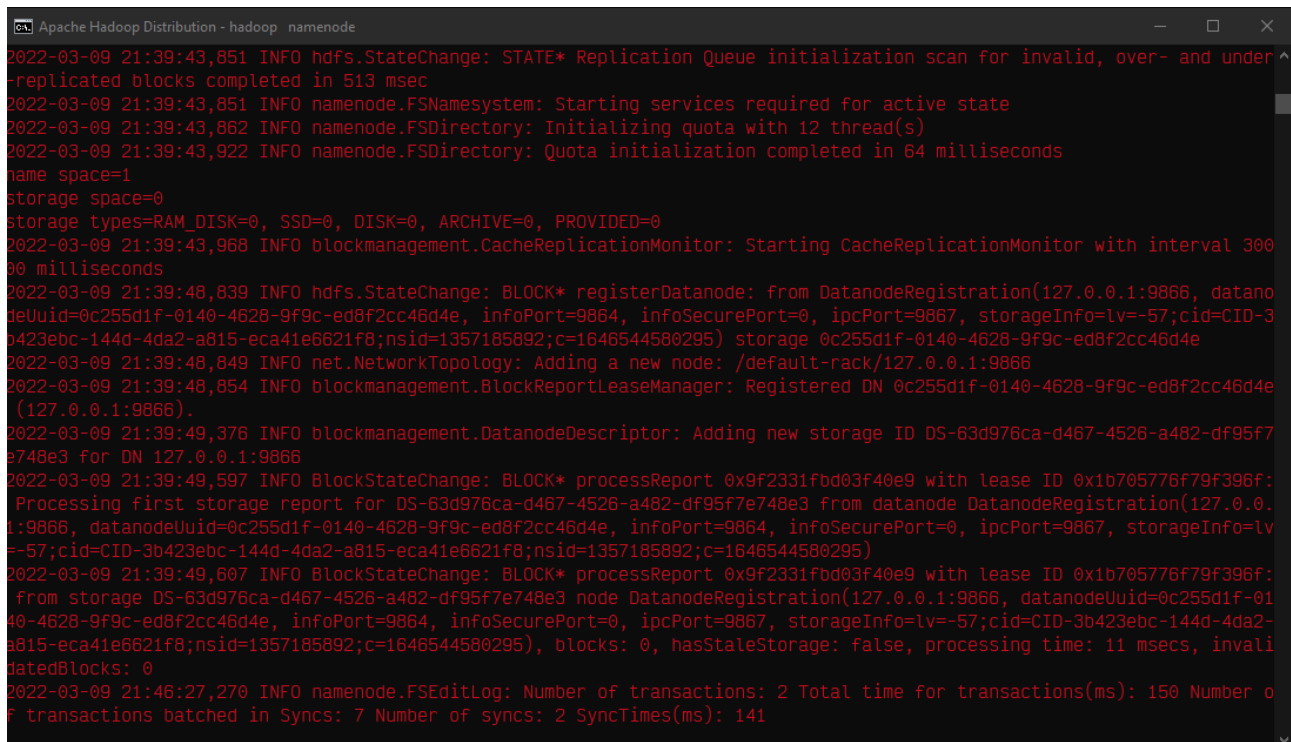
```

This code is precompiled as %HADOOP_HOME%\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.3.2.jar.

Execution:

- ◆ Navigate to %HADOOP_HOME%\sbin on the command-line (either Command Prompt or PowerShell) and start HDFS and YARN services.

```
C:\Users\YASH>cd %HADOOP_HOME%\sbin  
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\sbin>start-dfs
```



The screenshot shows a terminal window titled "Apache Hadoop Distribution - hadoop namenode". It displays a series of log messages from the namenode's startup process. The messages include information about the replication queue initialization, the starting of services for the active state, the initialization of the file system directory with a quota of 12 threads, and the registration of a new datanode at 127.0.0.1:9866. It also shows the processing of a storage report for the new datanode and the initialization of the edit log.

```
2022-03-09 21:39:43,851 INFO hdfs.StateChange: STATE* Replication Queue initialization scan for invalid, over- and under-replicated blocks completed in 513 msec  
2022-03-09 21:39:43,851 INFO namenode.FSNamesystem: Starting services required for active state  
2022-03-09 21:39:43,862 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)  
2022-03-09 21:39:43,922 INFO namenode.FSDirectory: Quota initialization completed in 64 milliseconds  
name space=1  
storage space=0  
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0  
2022-03-09 21:39:43,968 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds  
2022-03-09 21:39:48,839 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(127.0.0.1:9866, datanodeUuid=0c255d1f-0140-4628-9f9c-ed8f2cc46d4e, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-3b423ebc-144d-4da2-a815-eca41e6621f8;nsid=1357185892;c=1646544580295) storage 0c255d1f-0140-4628-9f9c-ed8f2cc46d4e  
2022-03-09 21:39:48,849 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866  
2022-03-09 21:39:48,854 INFO blockmanagement.BlockReportLeaseManager: Registered DN 0c255d1f-0140-4628-9f9c-ed8f2cc46d4e (127.0.0.1:9866).  
2022-03-09 21:39:49,376 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-63d976ca-d467-4526-a482-df95f7e748e3 for DN 127.0.0.1:9866  
2022-03-09 21:39:49,597 INFO BlockStateChange: BLOCK* processReport 0x9f2331fbd03f40e9 with lease ID 0x1b705776f79f396f: Processing first storage report for DS-63d976ca-d467-4526-a482-df95f7e748e3 from datanode DatanodeRegistration(127.0.0.1:9866, datanodeUuid=0c255d1f-0140-4628-9f9c-ed8f2cc46d4e, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-3b423ebc-144d-4da2-a815-eca41e6621f8;nsid=1357185892;c=1646544580295)  
2022-03-09 21:39:49,607 INFO BlockStateChange: BLOCK* processReport 0x9f2331fbd03f40e9 with lease ID 0x1b705776f79f396f: from storage DS-63d976ca-d467-4526-a482-df95f7e748e3 node DatanodeRegistration(127.0.0.1:9866, datanodeUuid=0c255d1f-0140-4628-9f9c-ed8f2cc46d4e, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-3b423ebc-144d-4da2-a815-eca41e6621f8;nsid=1357185892;c=1646544580295), blocks: 0, hasStaleStorage: false, processing time: 11 msecs, invalidatedBlocks: 0  
2022-03-09 21:46:27,270 INFO namenode.FSEditLog: Number of transactions: 2 Total time for transactions(ms): 150 Number of transactions batched in Syncs: 7 Number of syncs: 2 SyncTimes(ms): 141
```

```
Apache Hadoop Distribution - hadoop datanode
644580295 on volume C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode...
2022-03-09 21:39:48,164 INFO impl.BlockPoolSlice: Replica Cache file: C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode\current\BP-1035977273-192.168.43.145-1646544580295\current\replicas doesn't exist
2022-03-09 21:39:48,173 INFO impl.FsDatasetImpl: Time to add replicas to map for block pool BP-1035977273-192.168.43.145-1646544580295 on volume C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode: 10ms
2022-03-09 21:39:48,175 INFO impl.FsDatasetImpl: Total time to add all replicas to map for block pool BP-1035977273-192.168.43.145-1646544580295: 15ms
2022-03-09 21:39:48,178 INFO checker.ThrottledAsyncChecker: Scheduling a check for C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode
2022-03-09 21:39:48,225 INFO checker.DatasetVolumeChecker: Scheduled health check for volume C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode
2022-03-09 21:39:48,322 INFO datanode.VolumeScanner: VolumeScanner(C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\data\dfs\datanode, DS-63d976ca-d467-4526-a482-df95f7e748e3): no suitable block pools found to scan. Waiting 1517133421 ms.
2022-03-09 21:39:48,339 WARN datanode.DirectoryScanner: dfs.datanode.directoryscan.throttle.limit.ms.per.sec set to value above 1000 ms/sec. Assuming default value of -1
2022-03-09 21:39:48,341 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting in 13352688ms with interval of 21600000ms and throttle limit of -1ms/s
2022-03-09 21:39:48,370 INFO datanode.DataNode: Block pool BP-1035977273-192.168.43.145-1646544580295 (Datanode Uuid 0c255d1f-0140-462b-9f9c-ed8f2cc46d4e) service to localhost/127.0.0.1:9820 beginning handshake with NN
2022-03-09 21:39:48,923 INFO datanode.DataNode: Block pool BP-1035977273-192.168.43.145-1646544580295 (Datanode Uuid 0c255d1f-0140-462b-9f9c-ed8f2cc46d4e) service to localhost/127.0.0.1:9820 successfully registered with NN
2022-03-09 21:39:48,925 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9820 using BLOCKREPORT_INTERVAL of 21600000msecs CACHEREPORT_INTERVAL of 10000msecs Initial delay: 0msecs; heartbeatInterval=3000
2022-03-09 21:39:49,733 INFO datanode.DataNode: Successfully sent block report 0x9f2331fbd03f40e9 with lease ID 0x1b705776f79f396f to namenode: localhost/127.0.0.1:9820, containing 1 storage report(s), of which we sent 1. The reports had 0 total blocks and used 1 RPC(s). This took 21 msecs to generate and 233 msecs for RPC and NN processing. Got back one command: FinalizeCommand/5.
2022-03-09 21:39:49,744 INFO datanode.DataNode: Got finalize command for block pool BP-1035977273-192.168.43.145-1646544580295
```

C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\sbin>start-yarn
starting yarn daemons

```
Apache Hadoop Distribution - C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\bin\yarn nodemanager
Mar 09, 2022 9:41:36 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19 02/11/2015 03:25 AM'
Mar 09, 2022 9:41:37 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
Mar 09, 2022 9:41:38 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
Mar 09, 2022 9:41:41 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the scope "Singleton"
2022-03-09 21:41:41,909 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@7979b8b7{/node/,file:///C:/Users/YASH/AppData/Local/Temp/jetty-0_0_0-8042-hadoop-yarn-common-3_3_2_jar_-any-1188022385021473506/webapp/,AVAILABLE}{jar:file:/C:/Users/YASH/Downloads/hadoop/hadoop-3.3.2/share/hadoop/yarn/hadoop-yarn-common-3.3.2.jar!/webapps/node}
2022-03-09 21:41:41,961 INFO server.AbstractConnector: Started ServerConnector@7a7471ce{HTTP/1.1, (http/1.1)}{0.0.0.0:8042}
2022-03-09 21:41:41,964 INFO server.Server: Started @71184ms
2022-03-09 21:41:41,965 INFO webapp.WebApps: Web app node started at 8042
2022-03-09 21:41:41,970 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : LAPTOP-M0821FQC:19650
2022-03-09 21:41:42,021 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2022-03-09 21:41:42,083 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8031
2022-03-09 21:41:45,056 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id 658897223
2022-03-09 21:41:45,085 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id 258164966
2022-03-09 21:41:45,101 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as LAPTOP-M0821FQC:19650 with total resource of <memory:8192, vCores:8>
2022-03-09 21:51:18,074 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public
```



```
Apache Hadoop Distribution - C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\bin\yarn resourcemanager
2022-03-09 21:41:24,108 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server
2022-03-09 21:41:24,136 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2022-03-09 21:41:25,256 INFO ipc.Server: IPC Server Responder: starting
2022-03-09 21:41:25,284 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2022-03-09 21:41:25,259 INFO ipc.Server: IPC Server listener on 8031: starting
2022-03-09 21:41:25,314 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2022-03-09 21:41:25,391 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server
2022-03-09 21:41:25,398 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2022-03-09 21:41:25,476 INFO ipc.Server: IPC Server listener on 8030: starting
2022-03-09 21:41:25,476 INFO ipc.Server: IPC Server Responder: starting
2022-03-09 21:41:26,575 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2022-03-09 21:41:26,649 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
2022-03-09 21:41:26,682 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2022-03-09 21:41:26,927 INFO ipc.Server: IPC Server listener on 8032: starting
2022-03-09 21:41:26,927 INFO ipc.Server: IPC Server Responder: starting
2022-03-09 21:41:36,668 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-9f092a8a-0b8e-462f-91c1-d65105657a1b
2022-03-09 21:41:37,398 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2022-03-09 21:41:37,466 INFO resourcemanager.ResourceManager: Transitioned to active state
2022-03-09 21:41:44,879 INFO resourcemanager.ResourceTrackerService: NodeManager from node LAPTOP-M0821FQC(cmPort: 19650, httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId LAPTOP-M0821FQC:19650
2022-03-09 21:41:44,913 INFO rmnode.RMNodeImpl: LAPTOP-M0821FQC:19650 Node Transitioned from NEW to RUNNING
2022-03-09 21:41:45,170 INFO capacity.CapacityScheduler: Added node LAPTOP-M0821FQC:19650 clusterResource: <memory:8192, vCores:8>
2022-03-09 21:51:23,843 INFO scheduler.AbstractYarnScheduler: Release request cache is cleaned up
```

- ◆ Now navigate to the %HADOOP_HOME% directory.

```
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2\sbin>cd %HADOOP_HOME%
```

- ◆ Now, create a file that stores your input data. Lets call the file input_text_mapreduce.txt. I have chosen the complete paraphrase of the ‘Lorem Ipsum’ text as my input. Create a new directory called input_files in HDFS and move this file into this directory.

```
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2>%HADOOP_HOME%\bin\hdfs dfs -mkdir -p input_files
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2>%HADOOP_HOME%\bin\hdfs dfs -copyFromLocal C:/Temp/input_text_mapreduce.txt
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2>%HADOOP_HOME%\bin\hdfs dfs -ls
Found 2 items
drwxr-xr-x   - YASH supergroup          0 2022-03-09 21:46 input_files
-rw-r--r--   1 YASH supergroup        3046 2022-03-09 21:56 input_text_mapreduce.txt

C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2>%HADOOP_HOME%\bin\hdfs dfs -cp input_text_mapreduce.txt input_files/input.txt
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2>%HADOOP_HOME%\bin\hdfs dfs -ls
Found 2 items
drwxr-xr-x   - YASH supergroup          0 2022-03-09 22:10 input_files
-rw-r--r--   1 YASH supergroup        3046 2022-03-09 21:56 input_text_mapreduce.txt
```


- ◆ Now, run the example JAR. The output will be stored in the output directory.

```
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2>bin\yarn jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.2.jar wordcount input_files output
2022-03-09 22:26:23,837 INFO client.DefaultNoHARMFaoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2022-03-09 22:26:28,829 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/YASH/.staging/job_1646844810820_0001
2022-03-09 22:26:34,451 INFO input.FileInputFormat: Total input files to process : 1
2022-03-09 22:26:36,811 INFO mapreduce.JobSubmitter: number of splits:1
2022-03-09 22:26:39,670 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1646844810820_0001
2022-03-09 22:26:39,679 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-03-09 22:26:41,441 INFO conf.Configuration: resource-types.xml not found
2022-03-09 22:26:41,446 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-03-09 22:26:44,111 INFO impl.YarnClientImpl: Submitted application application_1646844810820_0001
2022-03-09 22:26:46,028 INFO mapreduce.Job: The url to track the job: http://LAPTOP-M0821FQC:8088/proxy/application_1646844810820_0001/
2022-03-09 22:26:46,091 INFO mapreduce.Job: Running job: job_1646844810820_0001
2022-03-09 22:28:20,164 INFO mapreduce.Job: Job job_1646844810820_0001 running in uber mode : false
2022-03-09 22:28:20,182 INFO mapreduce.Job: map 0% reduce 0%
2022-03-09 22:29:08,115 INFO mapreduce.Job: map 100% reduce 0%
2022-03-09 22:29:56,515 INFO mapreduce.Job: map 100% reduce 100%
2022-03-09 22:30:00,623 INFO mapreduce.Job: Job job_1646844810820_0001 completed successfully
2022-03-09 22:30:01,285 INFO mapreduce.Job: Counters: 54
    File System Counters
      FILE: Number of bytes read=2827
      FILE: Number of bytes written=560653
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=3164
```

```
      HDFS: Number of bytes written=1966
      HDFS: Number of read operations=8
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=2
      HDFS: Number of bytes read erasure-coded=0
    Job Counters
      Launched map tasks=1
      Launched reduce tasks=1
      Data-local map tasks=1
      Total time spent by all maps in occupied slots (ms)=43296
      Total time spent by all reduces in occupied slots (ms)=44011
      Total time spent by all map tasks (ms)=43296
      Total time spent by all reduce tasks (ms)=44011
      Total vcore-milliseconds taken by all map tasks=43296
      Total vcore-milliseconds taken by all reduce tasks=44011
      Total megabyte-milliseconds taken by all map tasks=44335104
      Total megabyte-milliseconds taken by all reduce tasks=45067264
    Map-Reduce Framework
      Map input records=9
      Map output records=455
      Map output bytes=4855
      Map output materialized bytes=2827
      Input split bytes=118
      Combine input records=455
      Combine output records=214
      Reduce input groups=214
      Reduce shuffle bytes=2827
      Reduce input records=214
```

```


Spilled Records=428
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=581
CPU time spent (ms)=4558
Physical memory (bytes) snapshot=434835456
Virtual memory (bytes) snapshot=563851264
Total committed heap usage (bytes)=273154048
Peak Map Physical memory (bytes)=267620352
Peak Map Virtual memory (bytes)=334045184
Peak Reduce Physical memory (bytes)=167215104
Peak Reduce Virtual memory (bytes)=229806080

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=3046
File Output Format Counters
Bytes Written=1966

```

- ◆ Now, check status of the job by navigating to <http://localhost:8088/cluster>.



All Applications

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources
1	0	0	1	0	<memory:0 B, vCores:0>	<memory:8 GB, vCores:8>

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy
1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit*Mt), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB
#application_1646844810820_0001	YASH	word count	MAPREDUCE		default	0	Wed Mar 9 22:26:42 +0550 2022	Wed Mar 9 22:26:51 +0550 2022	Wed Mar 9 22:29:59 +0550 2022	FINISHED	SUCCEEDED	N/A	N/A	N/A

Showing 1 to 1 of 1 entries

- ◆ Now, verify the output by printing all the files in the output directory.

```
C:\Users\YASH\Downloads\hadoop\hadoop-3.3.2>hdfs dfs -cat output/*
A      2
Aliquam 1
Aliquet 2
Augue  1
Commodo 1
Condimentum 1
Consectetur 1
Dictumst 1
Dignissim 1
Dolor 1
Donec 1
Eget 1
Enim 1
Erat 1
Est 1
Et 3
Fames 1
In 3
Justo 1
Lacus 1
Lectus 3
Lorem 1
Luctus 1
Maecenas 1
Malesuada 2
Mollis 2
Morbi 1
Neque 1
```

Nibh	1	
Nullam	1	
Nunc	1	
Orci	1	
Ornare	1	
Quis	1	
Sit	1	
Suspendisse		1
Tincidunt		2
Tortor	1	
Ultrices		1
Urna	1	
Vel	1	
Vitae	2	
Vivamus	1	
a	4	
a.	2	
ac	6	
ac.	1	
accumsan		1
accumsan.		1
adipiscing		3
adipiscing.		1
aenean.	1	
aliqua.	1	
aliquam	4	
aliquet	2	
amet	6	

amet,	1	
amet.	3	
arcu	5	
at	3	
augue	1	
bibendum		1
bibendum.		1
commodo	2	
commodo.		1
condimentum		2
congue	3	
consectetur		3
consequat		3
convallis		2
cras.	1	
curabitur		1
cursus	3	
diam	4	
dictum	2	
dignissim		1
do	1	
dolor	4	
dolor.	2	
dolore	1	
donec	3	
dui	2	
duis	2	
egetas	3	

egeſtas.	1
eget	10
eiusmod	1
elit	2
elit,	1
elit.	1
enim	2
erat	1
eſt	3
et	6
et.	1
etiam	1
eu	4
faciliſi	1
faciliſis	2
fames	3
faucibus	3
faucibus.	1
felis	4
fermentum	1
feugiat	2
fringilla	1
fuſce	1
fuſce.	1
gravidā	4
habitant	1
hendrerit	2
hendrerit.	1

iaculis	4	
id	7	
id.	2	
imperdiet		1
in	2	
incididunt		1
integer	1	
interdum		1
ipsum	4	
justo	1	
labore	1	
lacus	3	
lacus.	1	
laoreet	1	
lectus	5	
leo	5	
libero	1	
lobortis		1
lorem	2	
lorem.	2	
maecenas		2
magna	5	
magna.	1	
malesuada		3
malesuada.		1
massa	3	
massa.	2	
mattis	3	

mauris	2	
morbi	3	
morbi.	2	
nec	3	
netus	2	
nibh	5	
nisi	2	
nisl	4	
nisl.	1	
non	3	
non.	2	
nulla	2	
nulla.	1	
nullam	2	
nunc	5	
odio	1	
odio.	1	
orci	2	
ornare	1	
pellentesque	5	
pharetra	3	
placerat	1	
porttitor	5	
posuere	2	
posuere.	1	
praesent	3	
pretium	4	
proin	3	

pulvinar	1
quam 3	
quis 1	
quisque.	1
rhoncus 2	
rhoncus.	1
risus 3	
rutrum 1	
sagittis	3
sapien 2	
scelerisque.	1
sed 8	
semper 2	
senectus	1
sit 9	
sit. 3	
sodales 1	
sollicitudin	1
sollicitudin.	1
suspendisse	1
tempor 2	
tempor. 1	
tempus 1	
tempus. 1	
tincidunt	2
tristique	3
tristique.	1
turpis 5	

turpis.	1	
ullamcorper		5
ultrices		1
urna	2	
urna.	2	
ut	9	
varius	1	
varius.	1	
vel	2	
velit	1	
velit.	1	
venenatis		4
vestibulum		3
vitae	4	
viverra	2	
viverra.		2
volutpat		2
vulputate		2