

Experiment No: 3

Name: Asad Shaikh

Branch: MTech CE

Registration ID: 242050023

Aim- Setup single node Hadoop cluster and apply HDFS commands on single node Hadoop Cluster.

Steps Involved-

We installed Hadoop 3.2.1 single node and learned the basic HDFS commands

Background

Hadoop

Apache Hadoop is a collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model

HDFS

The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets. HDFS relaxes a few POSIX requirements to enable streaming access to file system data. HDFS was originally built as infrastructure for the Apache Nutch web search engine project. HDFS is now an Apache Hadoop subproject.

Hadoop Installation

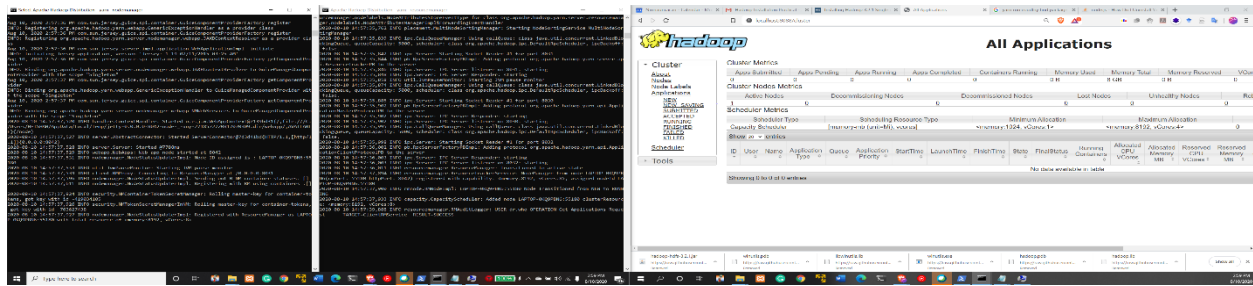
1) **Hadoop -version**, shows the Hadoop version installed in the system

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop -version
java version "1.8.0_261"
Java(TM) SE Runtime Environment (build 1.8.0_261-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.261-b12, mixed mode)
PS E:\Desktop\hadoop-3.2.1\sbin>
```

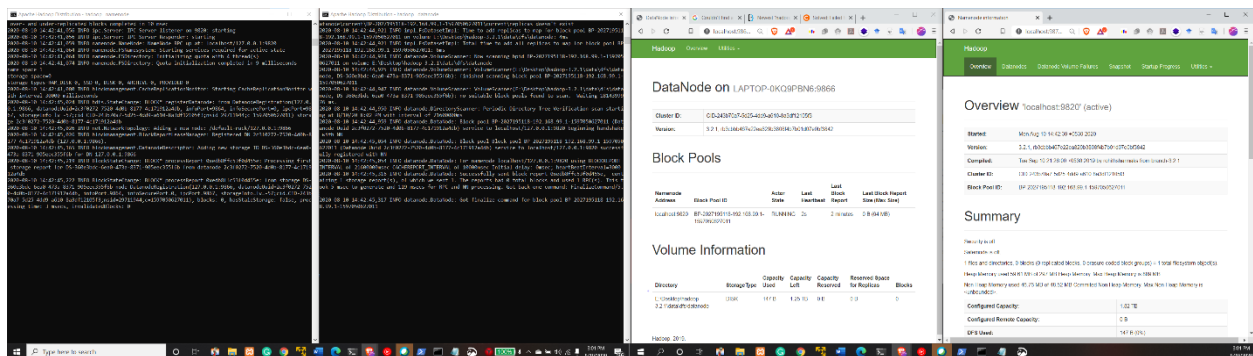
2) **Jps**- lists the instrumented Java HotSpot VMs on the target system

```
PS E:\Desktop\hadoop-3.2.1\sbin> jps
10100 NameNode
32132 ResourceManager
29800 NodeManager
33896 Jps
18524 DataNode
PS E:\Desktop\hadoop-3.2.1\sbin>
```

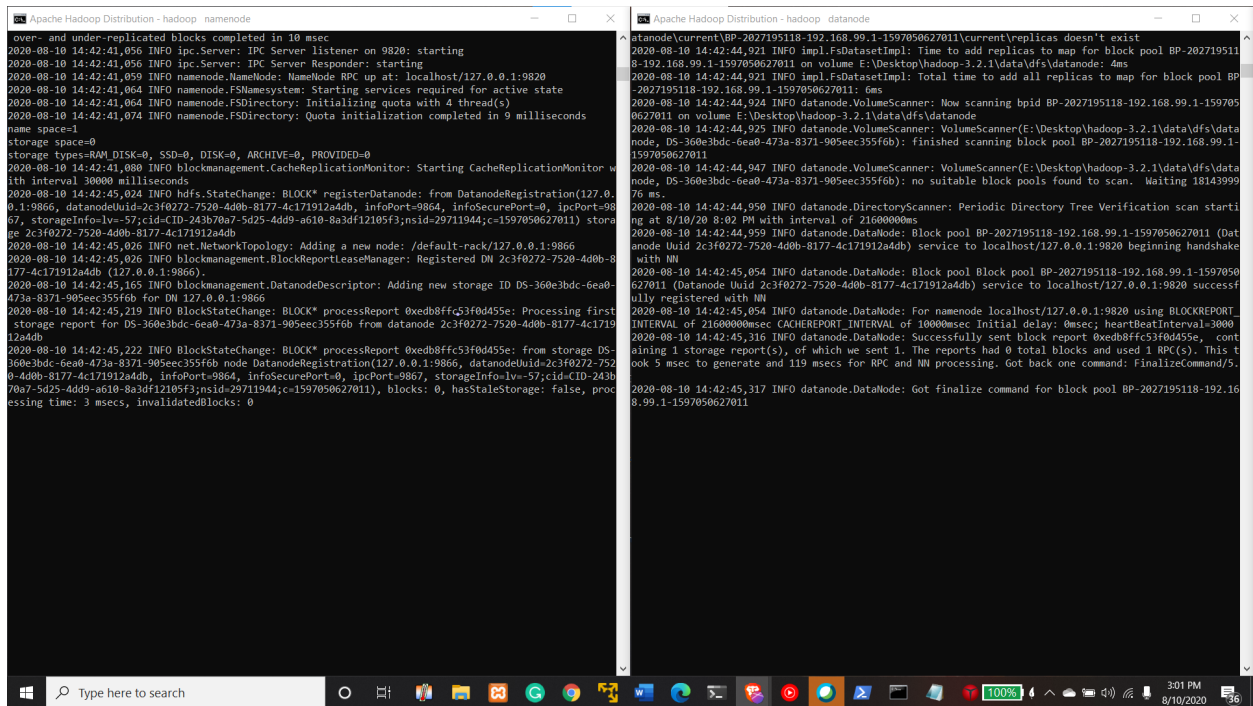
3) **Localhost:8080/cluster**



4) **Localhost 9820 and Localhost 9860** showing the data done and the system overview



5) Hadoop NameNode and the DataNode running side by side



The image shows two terminal windows side-by-side, both titled "Apache Hadoop Distribution - hadoop". The left window is titled "namenode" and the right window is titled "datanode". Both windows display log output from 2020-08-10 14:42:41.000 to 14:42:45.317.

Left Window (namenode) Log:

```
over, and under-replicated blocks completed in 10 msec
2020-08-10 14:42:41,000 INFO ipc.Server: IPC Server listener on 9820: starting
2020-08-10 14:42:41,056 INFO ipc.Server: IPC Server Responder: starting
2020-08-10 14:42:41,059 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9820
2020-08-10 14:42:41,064 INFO namenode.FSNamesystem: Starting services required for active state
2020-08-10 14:42:41,064 INFO namenode.FSDirectory: Initializing quota with 4 thread(s)
2020-08-10 14:42:41,074 INFO namenode.FSDirectory: Quota initialization completed in 9 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2020-08-10 14:42:41,080 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds
2020-08-10 14:42:45,024 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(127.0.0.1:9866, datanodeUuid=2c3f0272-7520-4d0b-8177-4c171912a4db, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-243b70a7-5d25-4dd9-a610-8a3df12105f3;nsid=29711944;c=1597050627011) storage 2c3f0272-7520-4d0b-8177-4c171912a4db
2020-08-10 14:42:45,026 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2020-08-10 14:42:45,026 INFO blockmanagement.BlockReportLeaseManager: Registered DN 2c3f0272-7520-4d0b-8177-4c171912a4db (127.0.0.1:9866).
2020-08-10 14:42:45,165 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-360e3bdc-6ea0-473a-8371-905eec355f6b for DN 127.0.0.1:9866
2020-08-10 14:42:45,219 INFO blockStateChange: BLOCK* processReport 0xedb8ffcc53f0d455e: Processing first storage report for DS-360e3bdc-6ea0-473a-8371-905eec355f6b from datanode 2c3f0272-7520-4d0b-8177-4c171912a4db
2020-08-10 14:42:45,222 INFO blockStateChange: BLOCK* processReport 0xedb8ffcc53f0d455e: from storage DS-360e3bdc-6ea0-473a-8371-905eec355f6b node DatanodeRegistration(127.0.0.1:9866, datanodeUuid=2c3f0272-7520-4d0b-8177-4c171912a4db, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-243b70a7-5d25-4dd9-a610-8a3df12105f3;nsid=29711944;c=1597050627011), blocks: 0, hasStateStorage: false, processing time: 3 msec, invalidatedBlocks: 0
```

Right Window (datanode) Log:

```
atanodeCurrent(BP-2027195118-192.168.99.1-1597050627011)current\replicas doesn't exist
2020-08-10 14:42:44,921 INFO impl.FsDatasetImpl: Time to add replicas to map for block pool BP-2027195118-192.168.99.1-1597050627011 on volume E:\Desktop\hadoop-3.2.1\data\dfs\datanode: 4ms
2020-08-10 14:42:44,921 INFO impl.FsDatasetImpl: Total time to add all replicas to map for block pool BP-2027195118-192.168.99.1-1597050627011: 6ms
2020-08-10 14:42:44,924 INFO datanode.VolumeScanner: Now scanning bpid BP-2027195118-192.168.99.1-1597050627011 on volume E:\Desktop\hadoop-3.2.1\data\dfs\datanode
2020-08-10 14:42:44,925 INFO datanode.VolumeScanner: VolumeScanner(E:\Desktop\hadoop-3.2.1\data\dfs\datanode, DS-360e3bdc-6ea0-473a-8371-905eec355f6b): finished scanning block pool BP-2027195118-192.168.99.1-1597050627011
2020-08-10 14:42:44,947 INFO datanode.VolumeScanner: VolumeScanner(E:\Desktop\hadoop-3.2.1\data\dfs\datanode, DS-360e3bdc-6ea0-473a-8371-905eec355f6b): no suitable block pools found to scan. Waiting 1814399976 ms.
2020-08-10 14:42:44,950 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting at 8/10/20 8:02 PM with interval of 21600000ms
2020-08-10 14:42:44,959 INFO datanode.DataNode: Block pool BP-2027195118-192.168.99.1-1597050627011 (Datanode Uuid 2c3f0272-7520-4d0b-8177-4c171912a4db) service to localhost/127.0.0.1:9820 beginning handshake with NN
2020-08-10 14:42:45,054 INFO datanode.DataNode: Block pool BP-2027195118-192.168.99.1-1597050627011 (Datanode Uuid 2c3f0272-7520-4d0b-8177-4c171912a4db) service to localhost/127.0.0.1:9820 successfully registered with NN
2020-08-10 14:42:45,054 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9820 using BLOCKREPORT_INTERVAL of 21600000msec CACHEREPORT_INTERVAL of 10000msec Initial delay: 0msec; heartbeatInterval=3000
2020-08-10 14:42:45,316 INFO datanode.DataNode: Successfully sent block report 0xedb8ffcc53f0d455e, containing 1 storage report(s), of which we sent 1. The reports had 0 total blocks and used 1 RPC(s). This took 5 msec to generate and 119 msec for RPC and NN processing. Got back one command: FinalizeCommand/5.
2020-08-10 14:42:45,317 INFO datanode.DataNode: Got finalize command for block pool BP-2027195118-192.168.99.1-1597050627011
```

6) Hadoop Yarn Manager is open

```
Select Apache Hadoop Distribution - yarn nodemanager
Aug 10, 2020 2:57:36 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class
Aug 10, 2020 2:57:36 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Aug 10, 2020 2:57:36 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19 02/11/2015 03:25 AM'
Aug 10, 2020 2:57:37 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"
Aug 10, 2020 2:57:37 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
Aug 10, 2020 2:57:37 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"
2020-08-10 14:57:37,520 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@7139bd31[/file:///C:/Users/HETSHAH/AppData/Local/Temp/jetty-0.0.0.0-8042-node-_-any-2718572728817070309.dir/webapp/,AVAILABLE][node]
2020-08-10 14:57:37,527 INFO server.AbstractConnector: Started ServerConnector@7d3d41bd(HTTP/1.1,[http/1.1]){0.0.0.0:8042}
2020-08-10 14:57:37,528 INFO server.Server: Started @7788ms
2020-08-10 14:57:37,529 INFO webapp.WebApp: Web app node started at 8042
2020-08-10 14:57:37,531 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : LAPTOP-0KQ9PBNG:55180
2020-08-10 14:57:37,533 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2020-08-10 14:57:37,539 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8031
2020-08-10 14:57:37,596 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM container statuses: []
2020-08-10 14:57:37,611 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers :[]
2020-08-10 14:57:37,924 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id -419034105
2020-08-10 14:57:37,926 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id -763627430
2020-08-10 14:57:37,927 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as LAPTOP-0KQ9PBNG:55180 with total resource of <memory:8192, vCores:8>

Apache Hadoop Distribution - yarn resourcemanager
ResourceManager, nodeLabels, NodeAttributesStoreEventType for class org.apache.hadoop.yarn.server.resourcemanager.nodeLabels.NodeAttributesManagerImpl$ForwardingEventHandler
2020-08-10 14:57:35,762 INFO placement.MultiNodeSortingManager: Starting NodeSortingService=MultiNodeSortingManager
2020-08-10 14:57:35,839 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2020-08-10 14:57:35,842 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2020-08-10 14:57:35,844 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server
2020-08-10 14:57:35,846 INFO ipc.Server: IPC Server listener on 8031: starting
2020-08-10 14:57:35,848 INFO ipc.Server: IPC Server Responder: starting
2020-08-10 14:57:35,858 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2020-08-10 14:57:35,874 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2020-08-10 14:57:35,885 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2020-08-10 14:57:35,902 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server
2020-08-10 14:57:35,902 INFO ipc.Server: IPC Server Responder: starting
2020-08-10 14:57:35,902 INFO ipc.Server: IPC Server listener on 8030: starting
2020-08-10 14:57:35,995 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2020-08-10 14:57:35,998 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2020-08-10 14:57:36,002 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
2020-08-10 14:57:36,003 INFO ipc.Server: IPC Server Responder: starting
2020-08-10 14:57:36,003 INFO ipc.Server: IPC Server listener on 8032: starting
2020-08-10 14:57:36,011 INFO resourcemanager.ResourceManager: Transitioned to active state
2020-08-10 14:57:37,894 INFO resourcemanager.ResourceTrackerService: NodeManager from node LAPTOP-0KQ9PBNG:55180(httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeid LAPTOP-0KQ9PBNG:55180
2020-08-10 14:57:37,900 INFO rmnode.RMNodeImpl: LAPTOP-0KQ9PBNG:55180 Node Transitioned from NEM to RUNNING
2020-08-10 14:57:37,933 INFO capacity.CapacityScheduler: Added node LAPTOP-0KQ9PBNG:55180 clusterResource: <memory:8192, vCores:8>
2020-08-10 14:57:39,088 INFO resourcemanager.RMAuditLogger: USER=dr.who OPERATION=Get Applications Request TARGET=ClientRMSERVICE RESULT=SUCCESS
```

HDFS commands

Hadoop fsck -/

```

drwxr-xr-x 2 HETSHAH supergroup 4096 2020-08-17 14:38 /user/
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fsck - /
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Connecting to namenode via http://localhost:9870/fsck?ugi=HETSHAH&path=%2F
FSCK started by HETSHAH (auth:SIMPLE) from /127.0.0.1 for path / at Mon Aug 24 09:22:12 IST 2020

Status: HEALTHY
  Number of data-nodes: 1
  Number of racks:      1
  Total dirs:           6
  Total symlinks:       0

Replicated Blocks:
  Total size: 0 B
  Total files: 0
  Total blocks (validated): 0
  Minimally replicated blocks: 0
  Over-replicated blocks: 0
  Under-replicated blocks: 0
  Mis-replicated blocks: 0
  Default replication factor: 1
  Average block replication: 0.0
  Missing blocks: 0
  Corrupt blocks: 0
  Missing replicas: 0

Erasure Coded Block Groups:
  Total size: 0 B
  Total files: 0
  Total block groups (validated): 0
  Minimally erasure-coded block groups: 0
  Over-erasure-coded block groups: 0
  Under-erasure-coded block groups: 0
  Unsatisfactory placement block groups: 0
  Average block group size: 0.0
  Missing block groups: 0
  Corrupt block groups: 0
  Missing internal blocks: 0
FSCK ended at Mon Aug 24 09:22:12 IST 2020 in 10 milliseconds

```

Hadoop Balancer

```

PS E:\Desktop\hadoop-3.2.1\sbin> hadoop balancer
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
2020-08-24 09:22:25,309 INFO balancer.Balancer: namenodes = [hdfs://localhost:9820]
2020-08-24 09:22:25,312 INFO balancer.Balancer: parameters = Balancer.BalancerParameters [BalancingPolicy.Node, threshold = 10.0, max idle iteration = 1, included nodes = 0, #included nodes = 0, #source nodes = 0, #blockpools = 0, run during upgrade = false]
2020-08-24 09:22:25,312 INFO balancer.Balancer: included nodes = []
2020-08-24 09:22:25,319 INFO balancer.Balancer: excluded nodes = []
2020-08-24 09:22:25,319 INFO balancer.Balancer: source nodes = []
Time Stamp      Iteration#  Bytes Already Moved  Bytes Left To Move  Bytes Being Moved
2020-08-24 09:22:27,754 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2020-08-24 09:22:27,955 INFO balancer.Balancer: dfs.balancer.movedwinwidth = 5400000 (default=5400000)
2020-08-24 09:22:27,955 INFO balancer.Balancer: dfs.balancer.moverThreads = 1000 (default=1000)
2020-08-24 09:22:27,957 INFO balancer.Balancer: dfs.balancer.dispatcherThreads = 200 (default=200)
2020-08-24 09:22:27,966 INFO balancer.Balancer: dfs.datanode.balance.max.concurrent.moves = 50 (default=50)
2020-08-24 09:22:27,966 INFO balancer.Balancer: dfs.balancer.getBlockSize.size = 2147483648 (default=2147483648)
2020-08-24 09:22:27,967 INFO balancer.Balancer: dfs.balancer.getBlockSize.min-block-size = 10485760 (default=10485760)
2020-08-24 09:22:27,974 INFO balancer.Balancer: dfs.balancer.max-size-to-move = 10737418240 (default=10737418240)
2020-08-24 09:22:27,974 INFO balancer.Balancer: dfs.blocksize = 134217728 (default=134217728)
2020-08-24 09:22:28,025 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2020-08-24 09:22:28,027 INFO balancer.Balancer: 0 over-utilized: []
2020-08-24 09:22:28,028 INFO balancer.Balancer: 0 underutilized: []
The cluster is balanced. Exiting...
Aug 24, 2020 9:22:28 AM          0      0 B          0 B          0 B
Aug 24, 2020 9:22:28 AM Balancing took 3.436 seconds

```

Hadoop fs -mkdir /user/hadoop

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -mkdir /user/hadoop
```

Hadoop fs -put /E:/Desktop/hi.txt /user/Hadoop

Hadoop fs -ls /user/Hadoop

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -put /E:/Desktop/hi.txt /user/hadoop
2020-08-24 09:33:09,580 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /user/hadoop
Found 1 items
-rw-r--r-- 1 HETSHAH supergroup      11 2020-08-24 09:33 /user/hadoop/hi.txt
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -put /E:/Desktop/temp /user/hadoop
```

Hadoop fs -put /E:/Desktop/temp /user/Hadoop

Hadoop fs -ls

Hadoop fs -ls /user/

Hadoop fs -ls /user/hadoop

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -put /E:/Desktop/temp1 /user/hadoop
2020-08-24 09:36:22,387 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /user/
Found 2 items
drwxr-xr-x - HETSHAH supergroup      0 2020-08-17 14:38 /user/HETSHAH
drwxr-xr-x - HETSHAH supergroup      0 2020-08-24 09:36 /user/hadoop
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /
Found 4 items
drwxr-xr-x - HETSHAH supergroup      0 2020-08-24 09:28 /desktop
drwxr-xr-x - HETSHAH supergroup      0 2020-08-17 14:39 /het
drwxr-xr-x - HETSHAH supergroup      0 2020-08-24 09:22 /system
drwxr-xr-x - HETSHAH supergroup      0 2020-08-24 09:29 /user
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /user/hadoop
Found 3 items
-rw-r--r-- 1 HETSHAH supergroup      11 2020-08-24 09:33 /user/hadoop/hi.txt
drwxr-xr-x - HETSHAH supergroup      0 2020-08-24 09:34 /user/hadoop/temp
drwxr-xr-x - HETSHAH supergroup      0 2020-08-24 09:36 /user/hadoop/temp1
PS E:\Desktop\hadoop-3.2.1\sbin> █
```

Hadoop fs -du -s h /user/hadoop

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -du -s -h /user/hadoop
10.0 M 10.0 M /user/hadoop
PS E:\Desktop\hadoop-3.2.1\sbin>
```

Hadoop fs -rm /user/Hadoop/temp1/hi.txt

Hadoop fs -ls /user/Hadoop/temp1

Hadoop -ls user/Hadoop

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -rm /user/hadoop/temp1/hi.txt
Deleted /user/hadoop/temp1/hi.txt
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /user/hadoop/temp1
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /user/hadoop
Found 3 items
```

Hadoop fs -rm /Hadoop/temp

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -rm /user/hadoop/temp/*
rm: /user/hadoop/temp/Local-File-Search-Engine-master': Is a directory
Deleted /user/hadoop/temp/Local-File-Search-Engine-master.zip
rm: /user/hadoop/temp/Text-Search-Engine-master': Is a directory
rm: /user/hadoop/temp/search.py': Is a directory
Deleted /user/hadoop/temp/search.py.zip
rm: /user/hadoop/temp/text-search-engine-master (1)': Is a directory
PS E:\Desktop\hadoop-3.2.1\sbin> _
```

Hadoop fs -expunge

```
rm: /user/hadoop/temp/text-search-engine-master (1)': Is a directory
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -expunge
PS E:\Desktop\hadoop-3.2.1\sbin> _
```

Hadoop fs -rmr /user/Hadoop/temp

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -rmr /user/hadoop/temp
rmr: DEPRECATED: Please use '-rm -r' instead.
Deleted /user/hadoop/temp
PS E:\Desktop\hadoop-3.2.1\sbin> _
```

Hadoop fs -ls /user/hadoop

```
Deleted /user/hadoop/temp
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /user/hadoop
Found 2 items
-rw-r--r-- 1 HETSHAH supergroup 11 2020-08-24 09:33 /user/hadoop/hi.txt
drwxr-xr-x - HETSHAH supergroup 0 2020-08-24 09:43 /user/hadoop/temp1
PS E:\Desktop\hadoop-3.2.1\sbin> _
```

Hadoop fs -copyfromLocal E:\Desktop\hi.txt /user/Hadoop/temp1

```
drwxr-xr-x - HETSHAH supergroup 0 2020-08-24 09:43 /user/hadoop/temp1
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -copyFromLocal E:\Desktop\hi.txt /user/hadoop/temp1
2020-08-24 09:49:19,585 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
PS E:\Desktop\hadoop-3.2.1\sbin> _
```

Hadoop fs cat /user/Hadoop/temp1/hi.txt

```
2020-08-24 09:49:19,585 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -cat /user/hadoop/temp1/hi.txt
2020-08-24 09:49:56,584 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
hello
PS E:\Desktop\hadoop-3.2.1\sbin> _
```

Hadoop fs -copytoLocal /user/Hadoop/hi.txt E:/Desktop/hi.txt

```
command [generalOptions] [commandOptions]
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -copyToLocal /user/hadoop/hi.txt E:/Desktop/
copyToLocal: /Desktop/hi.txt': File exists
PS E:\Desktop\hadoop-3.2.1\sbin> _
```

Hadoop fs -cp /user/Hadoop/hi.txt /user/Hadoop/temp1

```
copyToLocal: /Desktop/hi.txt': File exists
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -cp /user/hadoop/hi.txt /user/hadoop/temp1
cp: /user/hadoop/temp1/hi.txt': File exists
PS E:\Desktop\hadoop-3.2.1\sbin> _
```


Hadoop fs -get /user/Hadoop/hi.txt /user/Hadoop/temp

```
get: /desktop/temp/hi.txt: File exists
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -get /user/hadoop/hi.txt E:/Desktop/temp
2020-08-24 09:53:09,711 INFO sas1.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
PS E:\Desktop\hadoop-3.2.1\sbin>
```

Hadoop fs -tail /user/Hadoop/hi.txt

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -tail /user/hadoop/hi.txt
2020-08-24 09:53:41,443 INFO sas1.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
hello
he
PS E:\Desktop\hadoop-3.2.1\sbin>
```

Hadoop fs -chmod 600 /user/Hadoop/hi.txt

Hadoop fs -chown root:root /user/Hadoop/hi.txt

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -chmod 600 /user/hadoop/hi.txt
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -chown root:root /user/hadoop/hi.txt
PS E:\Desktop\hadoop-3.2.1\sbin>
```

Hadoop fs -chgrp training /user/Hadoop/hi.txt

Hadoop fs -ls/

Hadoop fs -ls /user/hadoop

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -chmod 600 /user/hadoop/hi.txt
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -chown root:root /user/hadoop/hi.txt
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -chgrp training /user/hadoop/hi.txt
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /
Found 4 items
drwxr-xr-x - HETSHAH supergroup 0 2020-08-24 09:28 /desktop
drwxr-xr-x - HETSHAH supergroup 0 2020-08-17 14:39 /het
drwxr-xr-x - HETSHAH supergroup 0 2020-08-24 09:22 /system
drwxr-xr-x - HETSHAH supergroup 0 2020-08-24 09:29 /user
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -ls /user/hadoop
Found 2 items
-rw----- 1 root training 11 2020-08-24 09:33 /user/hadoop/hi.txt
drwxr-xr-x - HETSHAH supergroup 0 2020-08-24 09:49 /user/hadoop/temp1
PS E:\Desktop\hadoop-3.2.1\sbin>
```

Hadoop fs -setrep -w 2 /user/Hadoop/hi.txt

```
Usage: hadoop fs [generic options] -setrep [-R] [-w] <rep> <path> ...
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -setrep -w 2 /user/hadoop/hi.txt
Replication 2 set: /user/hadoop/hi.txt
Waiting for /user/hadoop/hi.txt .....
```


Hadoop dfsadmin -safemode leave

```
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs -expunge
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop dfsadmin -safemode leave
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Safe mode is OFF
PS E:\Desktop\hadoop-3.2.1\sbin> █
```

Hadoop fs

```
Safe mode is OFF
PS E:\Desktop\hadoop-3.2.1\sbin> hadoop fs
Usage: hadoop fs [generic options]
    [-appendToFile <localsrc> ... <dst>]
    [-cat [-ignoreCrc] <src> ...]
    [-checksum <src> ...]
    [-chgrp [-R] GROUP PATH...]
    [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
    [-chown [-R] [OWNER][:[GROUP]] PATH...]
    [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] <localsrc> ... <dst>]
    [-copyToLocal [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-count [-q] [-h] [-v] [-t <storage type>]] [-u] [-x] [-e] <path> ...]
    [-cp [-f] [-p | -p[topax]] [-d] <src> ... <dst>]
    [-createSnapshot <snapshotDir> [<snapshotName>]]
    [-deleteSnapshot <snapshotDir> <snapshotName>]
    [-df [-h] [<path> ...]]
    [-du [-s] [-h] [-v] [-X] <path> ...]
    [-expunge [-immediate]]
    [-find <path> ... <expression> ...]
    [-get [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-getfacl [-R] <path>]
    [-getfattr [-R] {-n name | -d} [-e en] <path>]
    [-getmerge [-nl] [-skip-empty-file] <src> <localdst>]
    [-head <file>]
    [-help [cmd ...]]
    [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]
    [-mkdir [-p] <path> ...]
    [-moveFromLocal <localsrc> ... <dst>]
    [-moveToLocal <src> <localdst>]
    [-mv <src> ... <dst>]
    [-put [-f] [-p] [-l] [-d] <localsrc> ... <dst>]
    [-renameSnapshot <snapshotDir> <oldName> <newName>]
    [-rm [-f] [-r] [-R] [-skipTrash] [-safely] <src> ...]
    [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
    [-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>]| [--set <acl_spec> <path>]]
    [-setfattr {-n name [-v value] | -x name} <path>]
    [-setrep [-R] [-w] <rep> <path> ...]
    [-stat [format] <path> ...]
```

Conclusion

In this practical we learned how to install Hadoop in our systems and we also used the HDFS commands for file manipulation to and fro from the system to the HDFS and vice versa.