

## Step 2: Start the Hadoop Container

### Run the Container:

```
C:\Windows\System32>docker run -it --name new-hadoop-cluster -p 8:9870 -p 9:8088 -p 10:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
- Setting fs.defaultFS=hdfs://f29660bda025:8020
Configuring hdfs
- Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@f29660bda025:/#
```

### Start Hadoop Services:

```
Administration: Command Prompt: docker run -it --name new-hadoop-cluster -p 8:9870 -p 9:8088 -p 10:50070 bde2020/hadoop-namenode:latest /bin/bash
2024-11-18 08:25:11,415 INFO datanode.DataNode: dslserName = root
2024-11-18 08:25:11,417 INFO datanode.DataNode: supergroup = supergroup
2024-11-18 08:25:11,517 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 1000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-11-18 08:25:11,539 INFO ipc.Server: Starting Socket Reader #1 for port 9867
2024-11-18 08:25:11,622 INFO datanode.DataNode: Opened IPC server at /f29660bda025:9867
2024-11-18 08:25:11,854 INFO datanode.DataNode: Refresh request received for nameservices: null
2024-11-18 08:25:11,870 INFO datanode.DataNode: Starting BPOfferServices for nameservices: <default>
2024-11-18 08:25:11,888 INFO datanode.DataNode: Block pool <registering> (Datanode Uid unassigned) service to f29660bda025/172.17.0.3:8020 starting to offer service
2024-11-18 08:25:11,900 INFO ipc.Server: IPC Server Responder: starting
2024-11-18 08:25:11,901 INFO ipc.Server: IPC Server Listener on 9867: starting
2024-11-18 08:25:12,127 INFO datanode.DataNode: Acknowledging ACTIVE Namenode during handshakeBlock pool <registering> (Datanode Uid unassigned) service to f29660bda025/172.17.0.3:8020
2024-11-18 08:25:12,232 INFO common.Storage: Using 1 threads to upgrade data directories (dfs.datanode.parallel.volumes.load.threads.num=1, dataDir=/)
2024-11-18 08:25:12,344 INFO common.Storage: Lock on /tmp/hadoop-root/dfs/data/in.use.lock acquired by nodeName 3159f29660bda025
2024-11-18 08:25:12,346 INFO common.Storage: Storage directory with location [DISK]file:/tmp/hadoop-root/dfs/data is not formatted for namespace 1980485915. Formatting...
2024-11-18 08:25:12,349 INFO common.Storage: Generated new storageID DS-19d7646a-dc6b-4694-95f7-9a048a5ecf26 for directory /tmp/hadoop-root/dfs/data
2024-11-18 08:25:12,389 INFO common.Storage: Analyzing storage directories for bpid BP-131007328-172.17.0.3-1731918255688
2024-11-18 08:25:12,393 INFO common.Storage: Locking is disabled for /tmp/hadoop-root/dfs/data/current/BP-131007328-172.17.0.3-1731918255688
2024-11-18 08:25:12,396 INFO common.Storage: Block pool storage directory for location [DISK]file:/tmp/hadoop-root/dfs/data and block pool id BP-131007328-172.17.0.3-1731918255688 is not formatted. Formatting ...
2024-11-18 08:25:12,396 INFO common.Storage: Formatting block pool BP-131007328-172.17.0.3-1731918255688 directory /tmp/hadoop-root/dfs/data/current/BP-131007328-172.17.0.3-1731918255688/current
2024-11-18 08:25:12,406 INFO datanode.DataNode: Setting up storage: nsid=1980485915;bpid=BP-131007328-172.17.0.3-1731918255688;lv=-57;nsInfo=lv=-65;cid=CID-570b52a3-2e8f-4fc0-9164-c613a22b7d3;nsid=1980485915;c=1731918255688;bpid=BP-131007328-172.17.0.3-1731918255688;duuid=null
2024-11-18 08:25:12,410 INFO datanode.DataNode: Generated and persisted new Datanode UUID 39ac3410-2fa1-46bf-98fb-d4343313d03b
2024-11-18 08:25:12,566 INFO impl.FSDataSetImpl: Added new volume: DS-19d7646a-dc6b-4694-95f7-9a048a5ecf26
2024-11-18 08:25:12,567 INFO impl.FSDataSetImpl: Added volume - [DISK]file:/tmp/hadoop-root/dfs/data, StorageType: DISK
2024-11-18 08:25:12,572 INFO impl.FSDataSetImpl: Registered FSDataSetState MBean
2024-11-18 08:25:12,588 INFO checker.ThrottledAsyncChecker: Scheduling a check for /tmp/hadoop-root/dfs/data
2024-11-18 08:25:12,602 INFO checker.DataseVolumeChecker: Scheduled health check for volume /tmp/hadoop-root/dfs/data
2024-11-18 08:25:12,604 INFO impl.FSDataSetImpl: Adding block pool BP-131007328-172.17.0.3-1731918255688
2024-11-18 08:25:12,682 INFO impl.FSDataSetImpl: Scanning block pool BP-131007328-172.17.0.3-1731918255688 on volume /tmp/hadoop-root/dfs/data...
2024-11-18 08:25:12,728 INFO impl.FSDataSetImpl: Time taken to scan block pool BP-131007328-172.17.0.3-1731918255688 on /tmp/hadoop-root/dfs/data: 4ms
2024-11-18 08:25:12,729 INFO impl.FSDataSetImpl: Total time to scan all replicas for block pool BP-131007328-172.17.0.3-1731918255688: 32ms
2024-11-18 08:25:12,732 INFO impl.FSDataSetImpl: Adding replicas to map for block pool BP-131007328-172.17.0.3-1731918255688 on volume /tmp/hadoop-root/dfs/data...
2024-11-18 08:25:12,732 INFO impl.FSDataSetImpl: Total time to add all replicas to map for block pool BP-131007328-172.17.0.3-1731918255688: 3ms
2024-11-18 08:25:12,736 INFO impl.FSDataSetImpl: Time to add replicas to map for block pool BP-131007328-172.17.0.3-1731918255688 on volume /tmp/hadoop-root/dfs/data: 4ms
2024-11-18 08:25:12,736 INFO impl.FSDataSetImpl: Total time to add all replicas to map for block pool BP-131007328-172.17.0.3-1731918255688: 5ms
2024-11-18 08:25:12,739 INFO datanode.VolumeScanner: Now scanning bpid BP-131007328-172.17.0.3-1731918255688 on volume /tmp/hadoop-root/dfs/data
2024-11-18 08:25:12,742 INFO datanode.VolumeScanner: VolumeScanner(/tmp/hadoop-root/dfs/data, DS-19d7646a-dc6b-4694-95f7-9a048a5ecf26): finished scanning block pool BP-131007328-172.17.0.3-1731918255688
2024-11-18 08:25:12,762 INFO datanode.VolumeScanner: VolumeScanner(/tmp/hadoop-root/dfs/data, DS-19d7646a-dc6b-4694-95f7-9a048a5ecf26): no suitable block pools found to scan. Waiting 1814399977 ms.
2024-11-18 08:25:12,764 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting at 11/18/24 1:04 PM with interval of 2160000ms
2024-11-18 08:25:12,775 INFO datanode.DataNode: Block pool BP-131007328-172.17.0.3-1731918255688 (Datanode Uid 39ac3410-2fa1-46bf-98fb-d4343313d03b) service to f29660bda025/172.17.0.3:8020 beginning handshake with NN
2024-11-18 08:25:12,829 INFO hdfs.HdfsStateChange: from DatanodeRegistration(172.17.0.3:9866, datanodeUid=39ac3410-2fa1-46bf-98fb-d4343313d03b, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-570b52a3-2e8f-4fc0-9164-c613a22b7d3;nsid=1980485915;c=1731918255688) storage 39ac3410-2fa1-46bf-98fb-d4343313d03b
2024-11-18 08:25:12,831 INFO net.NetworkTopology: Adding a new node: /default-rack/172.17.0.3:9866
2024-11-18 08:25:12,831 INFO blockmanagement.BlockReportLeaseManager: Registered DN 39ac3410-2fa1-46bf-98fb-d4343313d03b (172.17.0.3:9866).
2024-11-18 08:25:12,848 INFO datanode.DataNode: Block pool Block pool BP-131007328-172.17.0.3-1731918255688 (Datanode Uid 39ac3410-2fa1-46bf-98fb-d4343313d03b) service to f29660bda025/172.17.0.3:8020 successfully registered with NN
2024-11-18 08:25:12,848 INFO datanode.DataNode: For namenode f29660bda025/172.17.0.3:8020 using BLOCKEXPORT_INTERVAL of 2160000ms CACHEEXPORT_INTERVAL of 10000ms Initial delay: 0ms; heartbeatInterval=3000
2024-11-18 08:25:12,945 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-19d7646a-dc6b-4694-95f7-9a048a5ecf26 for DN 172.17.0.3:9866
2024-11-18 08:25:13,023 INFO blockStateChange: BLOCK processReport 0xca45851519dbdbee: Processing first storage report for DS-19d7646a-dc6b-4694-95f7-9a048a5ecf26 from datanode 39ac3410-2fa1-46bf-98fb-d4343313d03b
2024-11-18 08:25:13,026 INFO blockStateChange: BLOCK processReport 0xca45851519dbdbee: from storage DS-19d7646a-dc6b-4694-95f7-9a048a5ecf26 node DatanodeRegistration(172.17.0.3:9866, datanodeUid=39ac3410-2fa1-46bf-98fb-d4343313d03b, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-570b52a3-2e8f-4fc0-9164-c613a22b7d3;nsid=1980485915;c=1731918255688), blocks: 0, hasStaleStorage: false, processing time: 3 msec, invalidatedBlocks: 0
2024-11-18 08:25:13,084 INFO datanode.DataNode: Successfully sent block report 0xca45851519dbdbee, containing 1 storage report(s), of which we sent 1. The reports had 0 total blocks and used 1 RPC(s). This took 6 msec to generate and 95 msec for RPC and NN processing. Got back one command: FinalizeCommand/5
2024-11-18 08:25:13,084 INFO datanode.DataNode: Got Finalize command for block pool BP-131007328-172.17.0.3-1731918255688
```

## Access Hadoop Web Interfaces

### HDFS Web Interface:

The screenshot displays the Hadoop Overview page in a web browser. The page title is "Overview 'f29660bda025:8020' (active)". It provides a summary of the cluster's health and configuration.

**Overview 'f29660bda025:8020' (active)**

Started:	Mon Nov 18 13:54:31 +0530 2024
Version:	3.2.1-dc322b7d3-2e8f-4fc0-9164-c613a22b7d3
Compiled:	Tue Sep 10 21:28:00 +0530 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-570b52a3-2e8f-4fc0-9164-c613a22b7d3
Block Pool ID:	BP-131007328-172.17.0.3-1731918255688

**Summary**

Security is off  
Salveto is off

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total Hadoop system objects

Heap Memory used 153.54 MB of 233 MB Heap Memory. Max Heap Memory is 1.0 GB

Non Heap Memory used 47.3 MB of 48.5 MB Committed Non Heap Memory. Max Non Heap Memory is 'unbounded'

Configured Capacity:	1008.05 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	5.44 GB

```
root@f2966bda025:/opt/hadoop-3.2.1/etc/hadoop# yarn nodemanager &
[3] 1588
root@f2966bda025:/opt/hadoop-3.2.1/etc/hadoop# 2024-11-18 09:02:05,036 INFO nodemanager.NodeManager: STARTUP_MSG:
*****
STARTUP_MSG: Starting NodeManager
STARTUP_MSG: host = f2966bda025/172.17.0.3
STARTUP_MSG: args = []
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-json-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/netty-3.10.5.Final.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-net-3.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-webapp-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-math-3.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-webapp-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-server-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/asm-5.0.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jettison-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/error_prone_annotations-2.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jaxb-impl-2.2.5-1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/stax2-api-1.1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-server-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-http-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/accessors-smart-1.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-util-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-asn1-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jul-to-slf4j-1.7.25.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/snappy-java-1.0.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/htrace-core4-1.0-incubating.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-common-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-admin-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/avro-1.7.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-collections-3.2.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/listenablefuture-9999.0-empty-to-avoid-conflict-with-guava.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/javax-servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-core-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/annotations-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/httpclient-4.5.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/paranamer-2.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr305-3.0.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jobjen-annotations-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/metrics-core-3.2.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-io-2.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-util-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-configuration2-2.11.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-text-1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/failureaccess-1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-lang3-3.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-compress-1.18.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-mapper-asl-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jaxb-api-2.2.11.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/metrics-jersey-server-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jcip-annotations-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-client-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-servlet-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-collections-3.2.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-simplekdc-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-xc-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-security-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jch-0.1.54.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-framework-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-config-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-pkix-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-recipes-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-client-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-crypto-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/opt/hadoop-3.2.1/share/hadoop/common/hadoop-hdfs-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/picom/hadoop-common-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/hadoop-common-3.2.1-tests.jar:/opt/hadoop-3.2.1/share/hadoop/hadoop-kws-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/hdfs-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/hdfs-3.2.1-tests.jar:/opt/hadoop-3.2.1/share/hadoop/hdfs-3.2.1-tests.jar
```

```
root@f2966bda025:/opt/hadoop-3.2.1/etc/hadoop# yarn resourcemanager &
[4] 1668
root@f2966bda025:/opt/hadoop-3.2.1/etc/hadoop# 2024-11-18 09:03:10,666 INFO resourcemanager.ResourceManager: STARTUP_MSG:
*****
STARTUP_MSG: Starting ResourceManager
STARTUP_MSG: host = f2966bda025/172.17.0.3
STARTUP_MSG: args = []
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-json-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/netty-3.10.5.Final.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-net-3.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-webapp-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-math-3.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-webapp-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-server-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/asm-5.0.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jettison-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/error_prone_annotations-2.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jaxb-impl-2.2.5-1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/stax2-api-1.1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-server-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-http-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/accessors-smart-1.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-util-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-asn1-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jul-to-slf4j-1.7.25.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/snappy-java-1.0.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/htrace-core4-1.0-incubating.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-common-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-admin-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/avro-1.7.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-collections-3.2.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/listenablefuture-9999.0-empty-to-avoid-conflict-with-guava.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/javax-servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-core-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/annotations-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/httpclient-4.5.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/paranamer-2.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr305-3.0.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jobjen-annotations-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/metrics-core-3.2.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-io-2.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-util-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-configuration2-2.11.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-text-1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/failureaccess-1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-lang3-3.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-compress-1.18.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-mapper-asl-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jaxb-api-2.2.11.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/metrics-jersey-server-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jcip-annotations-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-client-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-servlet-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-collections-3.2.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-simplekdc-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-xc-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-security-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jch-0.1.54.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-framework-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-config-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-pkix-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-recipes-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-client-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-crypto-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/opt/hadoop-3.2.1/share/hadoop/common/hadoop-hdfs-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/picom/hadoop-common-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/hadoop-common-3.2.1-tests.jar:/opt/hadoop-3.2.1/share/hadoop/hadoop-kws-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/hdfs-3.2.1.jar:/opt/hadoop-3.2.1/share/hadoop/hdfs-3.2.1-tests.jar:/opt/hadoop-3.2.1/share/hadoop/hdfs-3.2.1-tests.jar
```

## Running a Sample MapReduce Job

## Upload Sample Data to HDFS:

```
root@f2966bda025:/# hdfs dfs -mkdir -p /user/hadoop/input
2024-11-18 08:33:48,462 INFO namenode.FSEditLog: Number of transactions: 4 Total time for transactions(ms): 26 Number of transactions batched in Syncs: 0 Number of syncs: 2 SyncTimes(ms): 40
root@f2966bda025:/#
```

```
root@f2966bda025:/# hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input

2024-11-18 08:34:28,992 INFO hdfs.StateChange: BLOCK* allocate blk 1073741825_1001, replicas=172.17.0.3:9866 for /user/hadoop/input/capacity-scheduler.xml_COPYING_
2024-11-18 08:34:29,049 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 08:34:29,320 INFO datanode.DataNode: Receiving BP-131007328-172.17.0.3-1731918255688:blk 1073741825_1001 src: /172.17.0.3:44992 dest: /172.17.0.3:9866
2024-11-18 08:34:29,417 INFO datanode.DataNode: ClientTrace: src: /172.17.0.3:44992, dest: /172.17.0.3:9866, bytes: 8260, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_978383230_1, offset: 0, srvID: 39ac3410-2fa1-46bf-98fb-d4-p-131007328-172.17.0.3-1731918255688:blk 1073741825_1001, duration(ns): 26701649
2024-11-18 08:34:29,418 INFO datanode.DataNode: PacketResponder: BP-131007328-172.17.0.3-1731918255688:blk 1073741825_1001, type=LAST_IN_PIPELINE terminating
2024-11-18 08:34:29,429 INFO namenode.FSNamesystem: BLOCK* blk 1073741825_1001 is COMMITTED but not COMPLETE(numNodes= 0 < minimum = 1) in file /user/hadoop/input/capacity-scheduler.xml_COPYING_
2024-11-18 08:34:29,842 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/capacity-scheduler.xml_COPYING_ is closed by DFSClient_NONMAPREDUCE_978383230_1
2024-11-18 08:34:29,938 INFO hdfs.StateChange: BLOCK* allocate blk 1073741826_1002, replicas=172.17.0.3:9866 for /user/hadoop/input/core-site.xml_COPYING_
2024-11-18 08:34:29,944 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 08:34:29,946 INFO datanode.DataNode: Receiving BP-131007328-172.17.0.3-1731918255688:blk 1073741826_1002 src: /172.17.0.3:44992 dest: /172.17.0.3:9866
2024-11-18 08:34:29,953 INFO datanode.DataNode: ClientTrace: src: /172.17.0.3:44992, dest: /172.17.0.3:9866, bytes: 860, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_978383230_1, offset: 0, srvID: 39ac3410-2fa1-46bf-98fb-d4-p-131007328-172.17.0.3-1731918255688:blk 1073741826_1002, duration(ns): 3262308
2024-11-18 08:34:29,953 INFO datanode.DataNode: PacketResponder: BP-131007328-172.17.0.3-1731918255688:blk 1073741826_1002, type=LAST_IN_PIPELINE terminating
2024-11-18 08:34:29,956 INFO namenode.FSNamesystem: BLOCK* blk 1073741826_1002 is COMMITTED but not COMPLETE(numNodes= 0 < minimum = 1) in file /user/hadoop/input/core-site.xml_COPYING_
2024-11-18 08:34:30,360 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/core-site.xml_COPYING_ is closed by DFSClient_NONMAPREDUCE_978383230_1
2024-11-18 08:34:30,427 INFO hdfs.StateChange: BLOCK* allocate blk 1073741827_1003, replicas=172.17.0.3:9866 for /user/hadoop/input/hadoop-policy.xml_COPYING_
2024-11-18 08:34:30,434 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 08:34:30,436 INFO datanode.DataNode: Receiving BP-131007328-172.17.0.3-1731918255688:blk 1073741827_1003 src: /172.17.0.3:44992 dest: /172.17.0.3:9866
2024-11-18 08:34:30,446 INFO datanode.DataNode: ClientTrace: src: /172.17.0.3:44992, dest: /172.17.0.3:9866, bytes: 11392, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_978383230_1, offset: 0, srvID: 39ac3410-2fa1-46bf-98fb-d4-p-131007328-172.17.0.3-1731918255688:blk 1073741827_1003, duration(ns): 26701649
2024-11-18 08:34:30,446 INFO datanode.DataNode: PacketResponder: BP-131007328-172.17.0.3-1731918255688:blk 1073741827_1003, type=LAST_IN_PIPELINE terminating
2024-11-18 08:34:30,449 INFO namenode.FSNamesystem: BLOCK* blk 1073741827_1003 is COMMITTED but not COMPLETE(numNodes= 0 < minimum = 1) in file /user/hadoop/input/hadoop-policy.xml_COPYING_
2024-11-18 08:34:30,852 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/hadoop-policy.xml_COPYING_ is closed by DFSClient_NONMAPREDUCE_978383230_1
2024-11-18 08:34:30,888 INFO hdfs.StateChange: BLOCK* allocate blk 1073741828_1004, replicas=172.17.0.3:9866 for /user/hadoop/input/hdfs-site.xml_COPYING_
2024-11-18 08:34:30,894 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 08:34:30,896 INFO datanode.DataNode: Receiving BP-131007328-172.17.0.3-1731918255688:blk 1073741828_1004 src: /172.17.0.3:44992 dest: /172.17.0.3:9866
2024-11-18 08:34:30,902 INFO datanode.DataNode: ClientTrace: src: /172.17.0.3:44992, dest: /172.17.0.3:9866, bytes: 1385, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_978383230_1, offset: 0, srvID: 39ac3410-2fa1-46bf-98fb-d4-p-131007328-172.17.0.3-1731918255688:blk 1073741828_1004, duration(ns): 3287509
2024-11-18 08:34:30,902 INFO datanode.DataNode: PacketResponder: BP-131007328-172.17.0.3-1731918255688:blk 1073741828_1004, type=LAST_IN_PIPELINE terminating
2024-11-18 08:34:30,905 INFO namenode.FSNamesystem: BLOCK* blk 1073741828_1004 is COMMITTED but not COMPLETE(numNodes= 0 < minimum = 1) in file /user/hadoop/input/hdfs-site.xml_COPYING_
2024-11-18 08:34:31,360 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/hdfs-site.xml_COPYING_ is closed by DFSClient_NONMAPREDUCE_978383230_1
2024-11-18 08:34:31,364 INFO hdfs.StateChange: BLOCK* allocate blk 1073741829_1005, replicas=172.17.0.3:9866 for /user/hadoop/input/ntfs-site.xml_COPYING_
2024-11-18 08:34:31,369 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 08:34:31,371 INFO datanode.DataNode: Receiving BP-131007328-172.17.0.3-1731918255688:blk 1073741829_1005 src: /172.17.0.3:45000 dest: /172.17.0.3:9866
2024-11-18 08:34:31,377 INFO datanode.DataNode: ClientTrace: src: /172.17.0.3:45000, dest: /172.17.0.3:9866, bytes: 620, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_978383230_1, offset: 0, srvID: 39ac3410-2fa1-46bf-98fb-d4-p-131007328-172.17.0.3-1731918255688:blk 1073741829_1005, duration(ns): 26701649
2024-11-18 08:34:31,377 INFO datanode.DataNode: PacketResponder: BP-131007328-172.17.0.3-1731918255688:blk 1073741829_1005, type=LAST_IN_PIPELINE terminating
2024-11-18 08:34:31,381 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/ntfs-site.xml_COPYING_ is closed by DFSClient_NONMAPREDUCE_978383230_1
```



## Run the WordCount Job:

```
root@f29660bda025:/# hadoop jar
Usage: hadoop jar <jar> [mainClass] args...
```

```
root@f29660bda025:/# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount /user/hadoop/input /user/hadoop/output
2024-11-18 08:36:25,835 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 08:36:25,939 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 08:36:25,939 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 08:36:26,493 INFO input.FileInputFormat: Total input files to process : 9
2024-11-18 08:36:26,561 INFO mapreduce.JobSubmitter: number of splits:9
2024-11-18 08:36:26,726 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local781974939_0001
2024-11-18 08:36:26,726 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-18 08:36:26,871 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-18 08:36:26,871 INFO mapreduce.Job: Running job: job_local781974939_0001
2024-11-18 08:36:26,873 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-18 08:36:26,885 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 08:36:26,885 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-18 08:36:26,886 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-18 08:36:26,899 INFO namenode.FSEditLog: Number of transactions: 61 Total time for transactions(ms): 36 Number of transactions batched in Syncs: 12 Number of syncs: 47 SyncTimes(ms): 263
2024-11-18 08:36:26,950 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-18 08:36:26,951 INFO mapred.LocalJobRunner: Starting task: attempt_local781974939_0001_m_000000_0
2024-11-18 08:36:26,983 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 08:36:26,983 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-18 08:36:27,000 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-11-18 08:36:27,017 INFO mapred.MapTask: Processing split: hdfs://f29660bda025:8020/user/hadoop/input/hadoop-policy.xml:0+11392
2024-11-18 08:36:27,093 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-11-18 08:36:27,094 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-18 08:36:27,094 INFO mapred.MapTask: soft limit at 83886080
2024-11-18 08:36:27,094 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-11-18 08:36:27,094 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-11-18 08:36:27,100 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-18 08:36:27,145 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-11-18 08:36:27,323 INFO mapred.LocalJobRunner:
2024-11-18 08:36:27,328 INFO mapred.MapTask: Starting flush of map output
2024-11-18 08:36:27,328 INFO mapred.MapTask: Spilling map output
2024-11-18 08:36:27,328 INFO mapred.MapTask: bufstart = 0; bufend = 15783; bufvoid = 104857600
2024-11-18 08:36:27,328 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26209160(104836640); length = 5237/6553600
2024-11-18 08:36:27,365 INFO mapred.MapTask: Finished spill 0
2024-11-18 08:36:27,379 INFO mapred.Task: Task:attempt_local781974939_0001_m_000000_0 is done. And is in the process of committing
2024-11-18 08:36:27,385 INFO mapred.LocalJobRunner: map
2024-11-18 08:36:27,386 INFO mapred.Task: Task 'attempt_local781974939_0001_m_000000_0' done.
2024-11-18 08:36:27,396 INFO mapred.Task: Final Counters for attempt_local781974939_0001_m_000000_0: Counters: 24
  File System Counters
    FILE: Number of bytes read=317864
    FILE: Number of bytes written=845822
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=11392
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=5
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=1
    HDFS: Number of bytes read erasure-coded=0
  Map-Reduce Framework
    Map input records=265
    Map output records=1310
```

## Check the Output:

```

root@f29660bda025:/# hdfs dfs -cat /user/hadoop/output/part-r-00000
2024-11-18 08:38:47,676 INFO sas1.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
"*)" 21
"AS" 9
"License"); 9
"alice,bob" 21
"clumping" 1
(ASF) 1
(root 1
(the 9
--> 18
-1 1
-1, 1
0.0 1
1-MAX_INT. 1
1. 1
1.0. 1
2.0 9
40 2
40+20=60 1
: 2
<!-- 18
</configuration> 9
</description> 33
</name> 2
</property> 57
<?xml 8
<?xml-stylesheet 4
<configuration> 9
<description> 31
<description>ACL 25
<description>Default 1
<name>default.key.acl.DECRYPT_EEK</name> 1
<name>default.key.acl.GENERATE_EEK</name> 1
<name>default.key.acl.MANAGEMENT</name> 1
<name>default.key.acl.READ</name> 1
<name>hadoop.kms.acl.CREATE</name> 1
<name>hadoop.kms.acl.DECRYPT_EEK</name> 1
<name>hadoop.kms.acl.DELETE</name> 1
<name>hadoop.kms.acl.GENERATE_EEK</name> 1
<name>hadoop.kms.acl.GET</name> 1
<name>hadoop.kms.acl.GET_KEYS</name> 1
<name>hadoop.kms.acl.GET_METADATA</name> 1
<name>hadoop.kms.acl.ROLLOVER</name> 1
<name>hadoop.kms.acl.SET_KEY_MATERIAL</name> 1
<name>security.admin.operations.protocol.acl</name> 1
<name>security.applicationclient.protocol.acl</name> 1
<name>security.applicationhistory.protocol.acl</name> 1
<name>security.applicationmaster-nodemanager.applicationmaster.protocol.acl</name> 1
<name>security.applicationmaster.protocol.acl</name> 1
<name>security.client.datanode.protocol.acl</name> 1
<name>security.client.protocol.acl</name> 1
<name>security.collector-nodemanager.protocol.acl</name> 1
<name>security.containermanagement.protocol.acl</name> 1

```