

HOW TO SETUP HADOOP 3.2 ON WINDOWS 10

1. Download the Hadoop binary from the following url:
<https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.2/hadoop-3.2.2.tar.gz>
2. Extract the **hadoop-3.2.2.tar.gz** to **C:\opt**
3. Download the JDK 8 from the following link
<https://www.oracle.com/java/technologies/javase/javase-jdk8-downloads.html>
4. Install **jdk-8u281-windows-x64.exe** to **c:\opt**
5. Edit the **setenv.bat** with the following content:

```
set JAVA_HOME=c:\opt\jdk1.8.0_281
set PATH=%JAVA_HOME%\bin;%PATH%
set HADOOP_HOME=c:\opt\hadoop-3.2.2
set HADOOP_CONF_DIR=c:\opt\hadoop-3.2.2\etc\hadoop
set HADOOP_MAPRED_HOME=c:\opt\hadoop-3.2.2
set HADOOP_COMMON_HOME=c:\opt\hadoop-3.2.2
set HADOOP_HDFS_HOME=c:\opt\hadoop-3.2.2
set HADOOP_YARN_HOME=c:\opt\hadoop-3.2.2
set PATH=%HADOOP_HOME%\bin;%HADOOP_HOME%\sbin;%PATH%
```
6. Edit the following Hadoop configuration files with the following given content:

map-red.xml

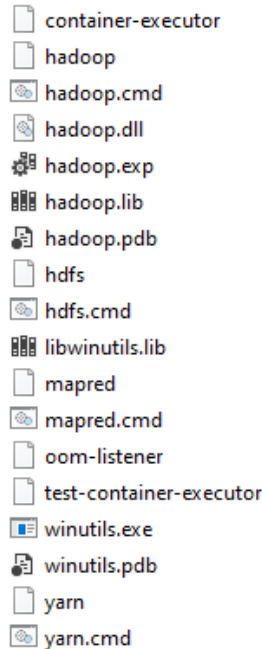
```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xml"?>
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

hdfs-site.xml

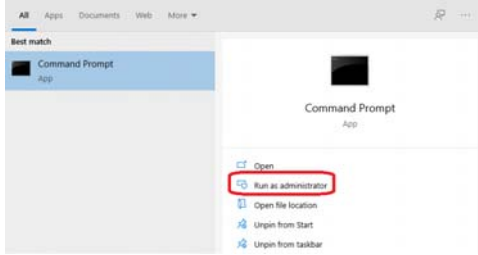
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xml"?>
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.permission</name>
    <value>>false</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///e:/var/big.data/tmp/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///e:/var/big.data/tmp/datanode</value>
```

<pre></property> </configuration></pre>
core-site.xml
<pre><?xml version="1.0" encoding="UTF-8"?> <?xml-stylesheet type="text/xsl" href="configuration.xsl"?> <configuration> <property> <name>fs.default.name</name> <value>hdfs://localhost:9000</value> </property> </configuration></pre>
yarn-site.xml
<pre><?xml version="1.0"?> <configuration> <property> <name>yarn.nodemanager.aux-services</name> <value>mapreduce_shuffle</value> </property> <property> <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name> <value>org.apache.hadoop.mapred.ShuffleHandler</value> </property> </configuration></pre>

- Download the **WinUtils for Hadoop 3.2** from the following url:
<https://courseware.deepcloudlabs.com/softtech/dcl700/lab-files/winutils-for-hadoop-3.2.zip>
- Unzip the file **winutils-for-hadoop-3.2.zip** to **c:\opt\hadoop-3.2.2\bin**



container-executor
hadoop
hadoop.cmd
hadoop.dll
hadoop.exp
hadoop.lib
hadoop.pdb
hdfs
hdfs.cmd
libwinutils.lib
mapred
mapred.cmd
oom-listener
test-container-executor
winutils.exe
winutils.pdb
yarn
yarn.cmd

9. Run the **cmd** as Administrator10. Run the **setenv.bat**11. Format the **namenode****hdfs namenode -format**12. Start the **hdfs****start-dfs**

```

Apache Hadoop Distribution - hadoop namenode
30 milliseconds
2021-03-22 15:40:56,555 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(127.0.0.1:9866, datanode
deUId=3ed615c4-5cd4-484f-9bc0-33f9c9360914, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-3
BF75554-84f7-4692-9206-a31a3fd1ff82;nsid=1036503250;c=1616414262343) storage 3ed615c4-5cd4-484f-9bc0-33f9c9360914
2021-03-22 15:40:56,561 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2021-03-22 15:40:56,562 INFO blockmanagement.BlockReportLeaseManager: Registered DN 3ed615c4-5cd4-484f-9bc0-33f9c9360914
(127.0.0.1:9866).
2021-03-22 15:40:56,847 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-0132c221-73ca-4683-9cd7-cb6469
74d64f for DN 127.0.0.1:9866
2021-03-22 15:40:56,928 INFO BlockStateChange: BLOCK* processReport 0xd398b4045be2c795: Processing first storage report
for DS-0132c221-73ca-4683-9cd7-cb646974d64f from datanode 3ed615c4-5cd4-484f-9bc0-33f9c9360914
2021-03-22 15:40:56,931 INFO BlockStateChange: BLOCK* processReport 0xd398b4045be2c795: from storage DS-0132c221-73ca-46
83-9cd7-cb646974d64f node DatanodeRegistration(127.0.0.1:9866, datanodeUId=3ed615c4-5cd4-484f-9bc0-33f9c9360914, infoPo
rt=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-3BF75554-84f7-4692-9206-a31a3fd1ff82;nsid=1036503250
;c=1616414262343), blocks: 0, hasStaleStorage: false, processing time: 3 msecs, invalidatedBlocks: 0
2021-03-22 16:17:18,355 INFO namenode.FSEditLog: Number of transactions: 2 Total time for transactions(ms): 338 Number o
f transactions batched in Syncs: 0 Number of syncs: 2 SyncTimes(ms): 276
2021-03-22 16:20:43,953 INFO namenode.FSEditLog: Number of transactions: 3 Total time for transactions(ms): 372 Number o
f transactions batched in Syncs: 0 Number of syncs: 3 SyncTimes(ms): 315
2021-03-22 16:20:44,152 INFO hdfs.StateChange: BLOCK* allocate blk_1073741825_1001, replicas=127.0.0.1:9866 for /files/c
ities2.csv._COPYING_
2021-03-22 16:20:44,752 INFO namenode.FSNamesystem: BLOCK* blk_1073741825_1001 is COMMITTED but not COMPLETE(numNodes= 0
< minimum = 1) in file /files/cities2.csv._COPYING_
2021-03-22 16:20:45,218 INFO hdfs.StateChange: DIR* completeFile: /files/cities2.csv._COPYING_ is closed by DFSClient_NO
MAPREDUCE_824237270_1
2021-03-22 16:24:25,749 INFO namenode.FSEditLog: Number of transactions: 9 Total time for transactions(ms): 373 Number o
f transactions batched in Syncs: 1 Number of syncs: 3 SyncTimes(ms): 704
2021-03-22 16:26:49,814 INFO namenode.FSEditLog: Number of transactions: 12 Total time for transactions(ms): 373 Number
of transactions batched in Syncs: 2 Number of syncs: 10 SyncTimes(ms): 2029

```

```

Apache Hadoop Distribution - hadoop datanode
2021-03-22 15:40:56,669 INFO datanode.DataNode: Block pool Block pool BP-873157547-172.27.128.1-1616414262343 (Datanode
UId 3ed615c4-5cd4-484f-9bc0-33f9c9360914) service to localhost/127.0.0.1:9000 successfully registered with NN
2021-03-22 15:40:56,669 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9000 using BLOCKREPORT_INTERVAL of 2160
0000msec CACHEREPORT_INTERVAL of 10000msec Initial delay: 0msec; heartBeatInterval=3000
2021-03-22 15:40:57,012 INFO datanode.DataNode: Successfully sent block report 0xd398b4045be2c795, containing 1 storage
report(s), of which we sent 1. The reports had 0 total blocks and used 1 RPC(s). This took 7 msec to generate and 120 m
secs for RPC and NN processing. Got back one command: FinalizeCommand/5.
2021-03-22 15:40:57,013 INFO datanode.DataNode: Got finalize command for block pool BP-873157547-172.27.128.1-1616414262
343
2021-03-22 16:20:44,540 INFO datanode.DataNode: Receiving BP-873157547-172.27.128.1-1616414262343:blk_1073741825_1001 sr
c: /127.0.0.1:50353 dest: /127.0.0.1:9866
2021-03-22 16:24:30,284 INFO impl.FsDatasetAsyncDiskService: Deleted BP-873157547-172.27.128.1-1616414262343 blk_1073741
825_1001 URI file:/var/big.data/tmp/datanode/current/BP-873157547-172.27.128.1-1616414262343/current/finaliz
ed/subdir0/subdir0/blk_1073741825_1001 for deletion
2021-03-22 16:24:30,288 INFO impl.FsDatasetAsyncDiskService: Scheduling blk_1073741825_1001 replica FinalizedReplica, bl
k_1073741825_1001, FINALIZED
getNumBytes() = 142021
getBytesOnDisk() = 142021
getVisibleLength() = 142021
getVolume() = e:\var\big.data\tmp\datanode
getBlockURI() = file:/var/big.data/tmp/datanode/current/BP-873157547-172.27.128.1-1616414262343/current/finaliz
ed/subdir0/subdir0/blk_1073741825_1001
2021-03-22 18:36:53,363 INFO datanode.DirectoryScanner: BlockPool BP-873157547-172.27.128.1-1616414262343 Total blocks:
0, missing metadata files:0, missing block files:0, missing blocks in memory:0, mismatched blocks:0

```



13. Start the yarn start-yarn

```

Apache Hadoop Distribution - yarn resourcemanager
2021-03-22 15:59:10,709 INFO placement.MultiNodeSortingManager: Starting NodeSortingService=MultiNodeSortingManager
2021-03-22 15:59:10,748 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2021-03-22 15:59:10,753 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2021-03-22 15:59:10,759 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server
2021-03-22 15:59:10,759 INFO ipc.Server: IPC Server Responder: starting
2021-03-22 15:59:10,759 INFO ipc.Server: IPC Server listener on 8031: starting
2021-03-22 15:59:10,778 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2021-03-22 15:59:10,817 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2021-03-22 15:59:10,832 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2021-03-22 15:59:10,882 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server
2021-03-22 15:59:10,887 INFO ipc.Server: IPC Server Responder: starting
2021-03-22 15:59:10,892 INFO ipc.Server: IPC Server listener on 8030: starting
2021-03-22 15:59:11,183 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2021-03-22 15:59:11,188 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2021-03-22 15:59:11,194 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
2021-03-22 15:59:11,195 INFO ipc.Server: IPC Server Responder: starting
2021-03-22 15:59:11,196 INFO ipc.Server: IPC Server listener on 8032: starting
2021-03-22 15:59:11,207 INFO resourcemanager.ResourceManager: Transitioned to active state
2021-03-22 15:59:15,018 INFO resourcemanager.ResourceTrackerService: NodeManager from node DESKTOP-HSPQDSO(cmPort: 49786, httpPort: 8042) registered with capability: <Memory:8192, vCores:8>, assigned nodeId DESKTOP-HSPQDSO:49786
2021-03-22 15:59:15,038 INFO rmnode.RMNodeImpl: DESKTOP-HSPQDSO:49786 Node Transitioned from NEW to UNHEALTHY
2021-03-22 15:59:15,043 ERROR capacity.CapacityScheduler: Attempting to remove non-existent node DESKTOP-HSPQDSO:49786
2021-03-22 16:09:10,719 INFO scheduler.AbstractYarnScheduler: Release request cache is cleaned up

```

```

Apache Hadoop Distribution - yarn nodemanager
Deleted: 0, Private Deleted: 0
2021-03-22 18:09:10,443 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 18:19:10,445 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 18:29:10,445 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 18:39:10,455 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 18:49:10,469 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 18:59:10,475 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 19:09:10,488 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 19:19:10,502 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 19:29:10,512 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 19:39:10,526 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 19:49:10,535 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 19:59:10,536 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 20:09:10,543 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0
2021-03-22 20:19:10,544 INFO localizer.ResourceLocalizationService: Cache Size Before Clean: 0, Total Deleted: 0, Public Deleted: 0, Private Deleted: 0

```

