## **Practical 1:**

## Aim: Installation of Hadoop

Step 1: Download Binary File for Windows <a href="https://hadoop.apache.org/releases.html">https://hadoop.apache.org/releases.html</a>



We suggest the following location for your download:

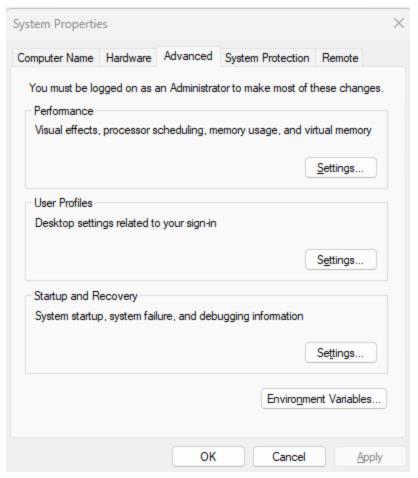
https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

Alternate download locations are suggested below.

It is essential that you verify the integrity of the downloaded file using the PGP signature ( .asc file) or a hash ( .mds or .sha\* file).

Step 2: Extract the files in C drive .

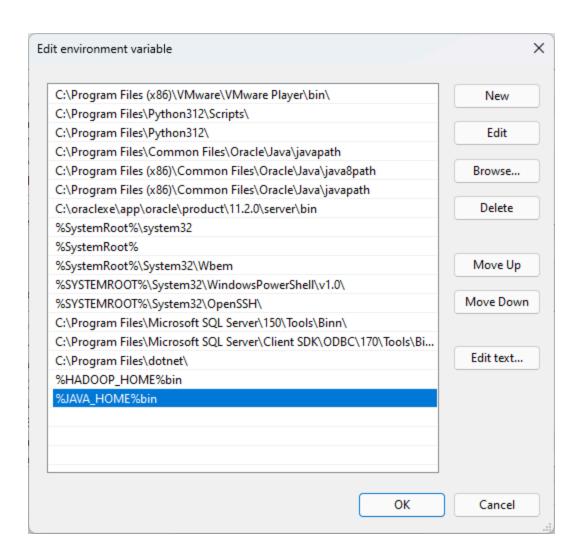
Step 3: Edit Environment Variables.



Step 4: Under System Variables click "New" and set "Variable name" as JAVA\_HOME and "Variable value" as the path of your JAVA JDK.

Step 5: Similarly add "HADOOP\_HOME" variable and download the bin folder from the below link

https://drive.google.com/drive/folders/1iURNbow2lglhAhSy3sfY5xxVfAg33NBW



Step 6: Extract the bin archive and replace the bin folder in Hadoop folder with the bin folder in this archive.

Name	Date modified	Туре	Size
hadoop	07-07-2020 00:16	File	9 KB
₲ hadoop	07-07-2020 00:16	Windows Comma	12 KB
hadoop.dll	01-08-2020 17:58	Application exten	85 KB
hadoop.exp	01-08-2020 17:58	Exports Library File	20 KB
hadoop.lib	01-08-2020 17:58	Object File Library	33 KB
‡ hadoop	01-08-2020 17:58	PDB File	684 KB
hdfs	14-08-2023 21:39	File	12 KB
	14-08-2023 21:39	Windows Comma	8 KB
libwinutils.lib	01-08-2020 17:58	Object File Library	1,283 KB
mapred	14-08-2023 21:39	File	7 KB
	14-08-2023 21:39	Windows Comma	7 KB
oom-listener	14-08-2023 21:39	File	29 KB
test-container-executor	07-07-2020 01:03	File	819 KB
winutils	14-08-2023 21:39	Application	110 KB
# winutils	01-08-2020 17:58	PDB File	1,156 KB
yarn	14-08-2023 21:39	File	13 KB
yarn	14-08-2023 21:39	Windows Comma	13 KB

Step 7: Check if "winutils" is working. If you get any dll error then download that dll and paste in the Windows -> System32 folder.

Step 8: Create a data folder in the hadoop home directory and add the folders datanode and namenode to it.

Name	Date modified	Туре	Size
namenode	31-07-2024 19:29	File folder	
datanode	31-07-2024 19:29	File folder	

9: Add the following path to "Path" under "System Variables" in "Edit Environment Variables" C:\hadoop-3.4.0\sbin

10: Make the changes to the following files as given, in "etc/hadoop" folder of hadoop home.

		-	
core-site.xml	31-07-2024 08:44	xmlfile	1 KB
🖫 hadoop-env	04-03-2024 12:06	Windows Comma	4 KB
■ hadoop-env	04-03-2024 13:35	SH Source File	17 KB
hadoop-metrics2	04-03-2024 12:06	Properties Source	4 KB
hadoop-policy.xml	04-03-2024 12:06	xmlfile	14 KB
hadoop-user-functions.sh.example	04-03-2024 12:06	EXAMPLE File	4 KB
hdfs-rbf-site.xml	04-03-2024 12:37	xmlfile	1 KB
hdfs-site.xml	04-03-2024 12:13	xmlfile	1 KB
httpfs-env	04-03-2024 12:22	SH Source File	2 KB
httpfs-log4j	04-03-2024 12:22	Properties Source	2 KB
httpfs-site.xml	04-03-2024 12:22	xmlfile	1 KB
kms-acls.xml	04-03-2024 12:08	xmlfile	4 KB
kms-env	04-03-2024 12:08	SH Source File	2 KB
kms-log4j	04-03-2024 12:08	Properties Source	2 KB
kms-site.xml	04-03-2024 12:08	xmlfile	1 KB
🖹 log4j	04-03-2024 12:06	Properties Source	15 KB
3 mapred-env	04-03-2024 13:00	Windows Comma	1 KB
mapred-env	04-03-2024 13:00	SH Source File	2 KB
mapred-queues.xml.template	04-03-2024 13:00	TEMPLATE File	5 KB
mapred-site.xml	04-03-2024 13:00	xmlfile	1 KB

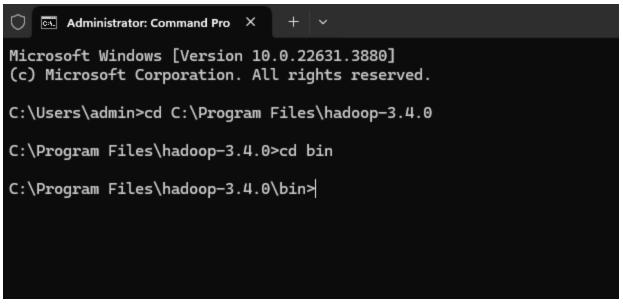
core-site.xml

mapred-site.xml

```
</property>
      </configuration>
      hdfs-site.xml
      <configuration>
             property>
                   <name>dfs.replication</name>
                   <value>1</value>
             property>
                   <name>dfs.namenode.name.dir</name>
                   <value>C:\hadoop-3.4.0\data\namenode</value>
             property>
                   <name>dfs.datanode.data.dir</name>
                   <value>C:\hadoop-3.4.0\data\datanode</value>
             </property>
      </configuration>
      yarn-site.xml
      <configuration>
             cproperty>
                   <name>yarn.nodemanager.aux-services</name>
                   <value>mapreduce_shuffle</value>
             property>
                   <name>yarn.nodemanager.auxservice.mapreduce.shuffle.class</name>
                   <value>org.apache.hadoop.mapred.shuffleHandler</value>
             </configuration>
Step 11: Go to hadoop-env.cmd file in /etc/hadoop folder and replace the set
JAVA_HOME=%JAVA_HOME% line with the following:
set JAVA_HOME=C:\Progra~1\Java\jdk-21
```

Step 13: Go to Admin Command prompt and type "hadoop" to see if the server is recognized.

Step 12: Restart your PC for the changes to take effect.



```
○ Administrator: Command Pro × + ∨
C:\Users\admin>hadoop
Usage: hadoop [--config confdir] [--loglevel loglevel] COMMAND where COMMAND is one of:
                          run a generic filesystem user client print the version
 fs
 version
                          run a jar file
note: please use "yarn jar" to launch
  jar <jar>
 YARN applications, not this command. checknative [-a|-h] check native hadoop and compression libraries availability
                          validate configuration XML files
  conftest
 distch path:owner:group:permisson
                          distributed metadata changer
  distcp <srcurl> <desturl> copy file or directories recursively
  archive -archiveName NAME -p parent path> <src>* <dest> create a hadoop archive
classpath prints the class path needed to get the
                          Hadoop jar and the required libraries
  credential
                          interact with credential providers
  jnipath
                          prints the java.library.path
 kerbname
                          show auth_to_local principal conversion
  kdiag
                          diagnose kerberos problems
                          manage keys via the KeyProvider
 key
                          view and modify Hadoop tracing settings
get/set the log level for each daemon
 trace
 daemonlog
                          run the class named CLASSNAME
Most commands print help when invoked w/o parameters.
C:\Users\admin>
```

Step 14: Type "hdfs namenode -format" to format the namenode.

C:\Users\admin>hdfs namenode -format

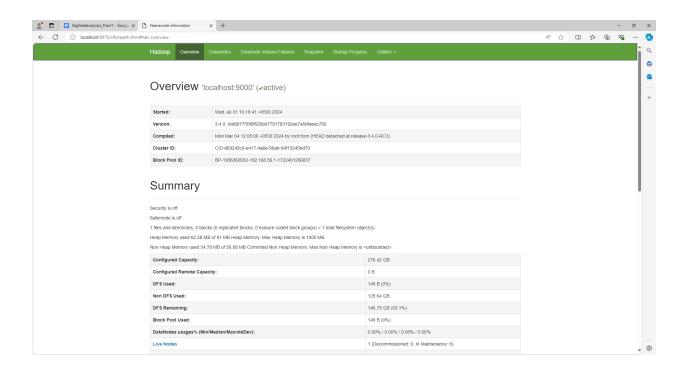
```
O Administrator: Command Pro X + V
 2024-07-31 10:17:41,894 INFO util.GSet: 0.25% max memory 1000 MB = 2.5 MB
2024-07-31 10:17:41,894 INFO util.Gset: 0.23% max memory 1000 MB = 2.5 MB
2024-07-31 10:17:41,894 INFO util.Gset: capacity = 2^18 = 262144 entries
2024-07-31 10:17:41,898 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2024-07-31 10:17:41,898 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2024-07-31 10:17:41,899 INFO metrics. TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-07-31 10:17:41,899 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-07-31 10:17:41,901 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2024-07-31 10:17:41,901 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry
time is 600000 millis
time is 6000000 mills
2024-07-31 10:17:41,902 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-07-31 10:17:41,902 INFO util.GSet: VM type = 64-bit
2024-07-31 10:17:41,902 INFO util.GSet: 0.02999999329447746% max memory 1000 MB = 307.2 KB
2024-07-31 10:17:41,902 INFO util.GSet: capacity = 2^15 = 32768 entries
2024-07-31 10:17:46,837 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1936360632-192.168.56.1-1722401266837
2024-07-31 10:17:46,865 INFO common.Storage: Storage directory C:\hadoop-3.4.0\data\namenode has been successfully forma
tted
2024-07-31 10:17:46,878 INFO namenode.FSImageFormatProtobuf: Saving image file C:\hadoop-3.4.0\data\namenode\current\fsi
 2024-07-31 10:17:46,920 INFO namenode.FSImageFormatProtobuf: Image file C:\hadoop-3.4.0\data\namenode\current\fsimage.ck
pt_00000000000000000000 of size 400 bytes saved in 0 seconds .
 .
2024-07-31 10:17:46,935 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-07-31 10:17:46,938 INFO blockmanagement.DatanodeManager: Slow peers collection thread shutdown 2024-07-31 10:17:46,947 INFO namenode.FSNamesystem: Stopping services started for active state 2024-07-31 10:17:46,948 INFO namenode.FSNamesystem: Stopping services started for standby state 2024-07-31 10:17:46,950 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown. 2024-07-31 10:17:46,950 INFO namenode.NameNode: SHUTDOWN_MSG:
  /***************
SHUTDOWN_MSG: Shutting down NameNode at 31D-LAB4-11/192.168.56.1
  *****************
C:\Users\admin>
```

Step 15: Type start-all.cmd to start all hadoop processes (make sure you have added set JAVA\_HOME=C:\Progra~1\Java\jdk-22)

```
Anache Hadoon Distribution . X +
be Anacha Hadoon Distribution . X +
                                                                                                                                                                                                                           □ ×
20 20 Administrator Anache Hadon X +
                                                                                                                                                                                                                             - п x
                Administrator: Apache Hadoo X
20 de
20 20
20 e
20 20
                                at com.google.inject.internal.cglib.core.internal.$LoadingCache$2.call(LoadingCache.java:54)
                               at java.base/java.util.concurrent.FutureTask.run(FutureTask.)ava:317)
at com.google.inject.internal.cglib.core.internal.$LoadingCache.createEntry(LoadingCache.java:61)
at com.google.inject.internal.cglib.core.internal.$LoadingCache.get(LoadingCache.java:34)
at com.google.inject.internal.cglib.core.$AbstractClassGenerator$ClassLoaderData.get(AbstractClassGenerator.java
st 20
st e
20 20
00 wi
20 20
de 04
69 00
20 20
20 ba
                 :119)
                                at com.google.inject.internal.cglib.core.$AbstractClassGenerator.create(AbstractClassGenerator.java:294)
                                at com.google.inject.internal.cglib.reflect.$FastClass$Generator.create(FastClass.java:65) at com.google.inject.internal.BytecodeGen.newFastClassForMember(BytecodeGen.java:258)
                                at com.google.inject.internal.BytecodeGen.nem/astClass/oxhem/ex(bytecodeGen.java:255)
at com.google.inject.internal.BytecodeGen.nem/astClass/oxhem/ex(BytecodeGen.java:207)
at com.google.inject.internal.ProviderMethod.create(ProviderMethod.java:69)
at com.google.inject.internal.ProviderMethodsModule.createProviderMethod(ProviderMethodsModule.java:327)
                                at com.google.inject.internal.ProviderMethodsModule.getProviderMethods(ProviderMethodsModule.java:135)
                               at com.google.inject.internal.ProviderMethodsModule.getProviderMethods(ProviderMethodsModule. at com.google.inject.internal.ProviderMethodsModule.configure(ProviderMethodsModule.java:105) at com.google.inject.spi.Elements$RecordingBinder.install(Elements.java:347) at com.google.inject.spi.Elements$RecordingBinder.install(Elements.java:356) at com.google.inject.AbstractModule.install(AbstractModule.java:103) at com.google.inject.servlet.ServletModule.configure(ServletModule.java:49) at com.google.inject.AbstractModule.configure(AbstractModule.java:61)
( 20
20 04
 ac 1:
20 20
     00
     20
                                at com.google.inject.spi.Elements$RecordingBinder.install(Elements.java:347) at com.google.inject.spi.Elements.getElements(Elements.java:104) at com.google.inject.internal.InjectorShell$Builder.build(InjectorShell.java:137)
      :9
 f a7
      t 20
                                 at com.google.inject.internal.InternalInjectorCreator.build(InternalInjectorCreator.java:105)
96 an /*
at 20 SH
                                      12 more
                 2024-07-31 10:19:04,581 INFO nodemanager.NodeManager: SHUTDOWN_MSG:
                 SHUTDOWN_MSG: Shutting down NodeManager at 31D-LAB4-11/192.168.56.1
                 *******************
                 C:\Users\admin>
                    C:\Users\ad
```

```
Apache Hadoop Distribution - yarn resourcemanager
                                                                                                                                                                                                П
                                                                                                                                                                                                          ×
2024-08-08 11:07:30,879 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-08-08 11:07:30,926 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, que
 capacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false, ipcFailOver: false.
2024-08-08 11:07:30,948 INFO ipc.Server: Listener at 0.0.0.0:8030
2024-08-08 11:07:30,950 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-08-08 11:07:30,960 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProt
ocolPB to the server
2024-08-08 11:07:31,023 INFO ipc.Server: IPC Server Responder: starting
2024-08-08 11:07:31,023 INFO ipc.Server: IPC Server listener on 8030: starting
2024-08-08 11:07:31,023 INFO ipc.Server: IPC Server listener on 8030: starting
2024-08-08 11:07:31,243 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false, ipcFailOver: false.
2024-08-08 11:07:31,244 INFO ipc.Server: Listener at 0.0.0.0:8032
2024-08-08 11:07:31,254 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-08-08 11:07:31,256 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProt
 colPB to the server
2024-08-08 11:07:31,258 INFO ipc.Server: IPC Server Responder: starting
2024-08-08 11:07:31,259 INFO ipc.Server: IPC Server listener on 8032: starting
2024-08-08 11:07:32,117 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-9bffb8a3-6efe-4c00-b2cd-e0d105ef686c
2024-08-08 11:07:32,323 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2024-08-08 11:07:32,356 INFO resourcemanager.ResourceManager: Transitioned to active state
2024-08-08 11:07:33,889 INFO resourcemanager.ResourceTrackerService: NodeManager from node 31D-LAB5-26.SVV.local(cmPort
62661 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId 31D-LAB5-26.SVV.local:62661 2024-08-08 11:07:33,905 INFO rmnode.RMNodeImpl: 31D-LAB5-26.SVV.local:62661 Node Transitioned from NEW to RUNNING
 024-08-08 11:07:33,934 INFO capacity.AbstractLeafQueue: LeafQueue: root.default update max app related, maxApplication:
-10000, maxApplicationsPerUser=10000, Abs Cap:1.0, Cap: 1.0, MaxCap : 1.0
2024-08-08 11:07:33,937 INFO capacity.CapacityScheduler: Added node 31D-LAB5-26.SVV.local:62661 clusterResource: <memor
:8192, vCores:8>
2024-08-08 11:07:33,938 INFO capacity.AbstractLeafQueue: LeafQueue: root.default update max app related, maxApplications
 10000, maxApplicationsPerUser=10000, Abs Cap:1.0, Cap: 1.0, MaxCap : 1.0
 Apache Hadoop Distribution - yarn resourcemanager
                                                                                                                                                                                                П
                                                                                                                                                                                                          ×
2024-08-08 11:07:30,879 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-08-08 11:07:30,926 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, que
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false, ipcFailOver: false.
2024-08-08 11:07:30,948 INFO ipc.Server: Listener at 0.0.0.0:8030
2024-08-08 11:07:30,950 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-08-08 11:07:30,960 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProt
2024-08-08 11:07:31,023 INFO ipc.Server: IPC Server Responder: starting
2024-08-08 11:07:31,023 INFO ipc.Server: IPC Server listener on 8030: starting
2024-08-08 11:07:31,243 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
2024-08-08 11:07:31,243 INFO 1pc.CailQueueManager: Using cailQueue: Class Java.util.concurrent.linkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false, ipcFailOver: false.
2024-08-08 11:07:31,244 INFO ipc.Server: Listener at 0.0.0:8032
2024-08-08 11:07:31,254 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-08-08 11:07:31,256 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProt
ocolPB to the server
2024-08-08 11:07:31,258 INFO ipc.Server: IPC Server Responder: starting
2024-08-08 11:07:31,259 INFO ipc.Server: IPC Server listener on 8032: starting
2024-08-08 11:07:32,117 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-9bffb8a3-6efe-4c00-b2cd-e0d105ef686c
2024-08-08 11:07:32,323 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2024-08-08 11:07:32,356 INFO resourcemanager.ResourceManager: Transitioned to active state
2024-08-08 11:07:33,889 INFO resourcemanager.ResourceTrackerService: NodeManager from node 31D-LAB5-26.SVV.local(cmPort:
62661 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId 31D-LAB5-26.SVV.local:62661
2024-08-08 11:07:33,905 INFO rmnode.RMNodeImpl: 31D-LAB5-26.SVV.local:62661 Node Transitioned from NEW to RUNNING
2024-08-08 11:07:33,934 INFO capacity.AbstractLeafQueue: LeafQueue: root.default update max app related, maxApplications
=10000, maxApplicationsPerUser=10000, Abs Cap:1.0, Cap: 1.0, MaxCap : 1.0
2024-08-08 11:07:33,937 INFO capacity.CapacityScheduler: Added node 31D-LAB5-26.SVV.local:62661 clusterResource: <memory
2024-08-08 11:07:33,938 INFO capacity.AbstractLeafQueue: LeafQueue: root.default update max app related, maxApplications
 10000, maxApplicationsPerUser=10000, Abs Cap:1.0, Cap: 1.0, MaxCap : 1.0
```

Step 16: Go to your browser and type localhost:9870 to view Hadoop Page.



Step 17: Now, go to cmd and type start-yarn.cmd:

```
C:\hadoop-3.4.0\bin>start-dfs.cmd
C:\hadoop-3.4.0\bin>start-yarn.cmd
starting yarn daemons
C:\hadoop-3.4.0\bin>_
```

Step 18: Now, go to localhost:8088 and observe accordingly: