

## Practical 1:

### Aim: Installation of Hadoop

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

---



---

We suggest the following location for your download:

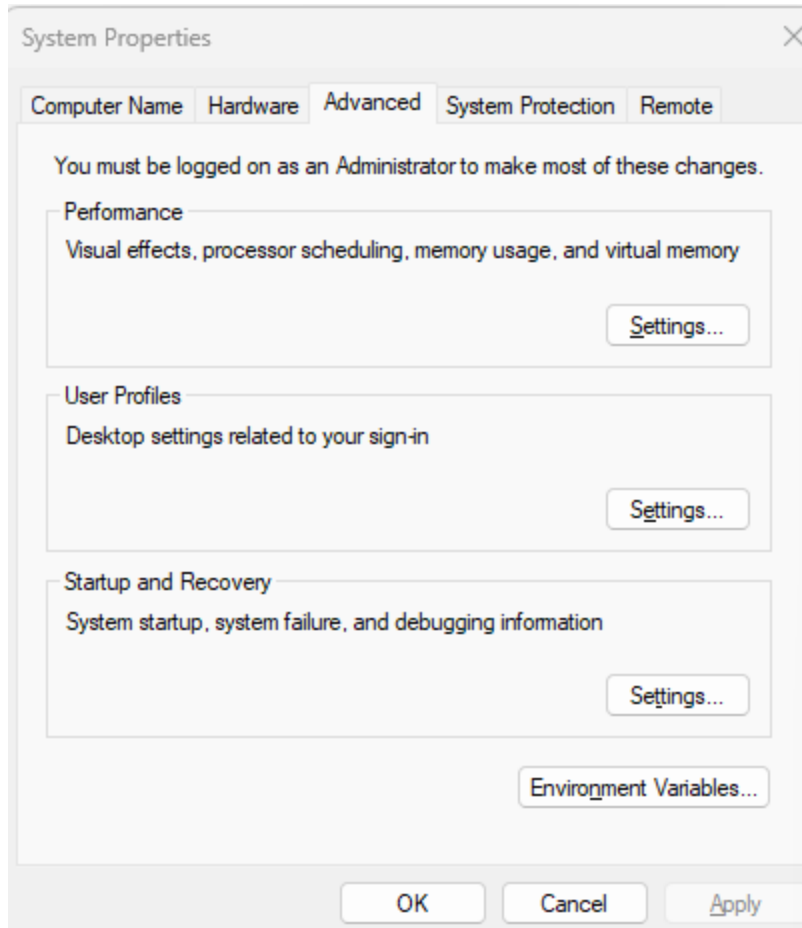
<https://d1cdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>

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 ( `.md5` or `.sha*` file).

Step 2: Extract the files in C drive .

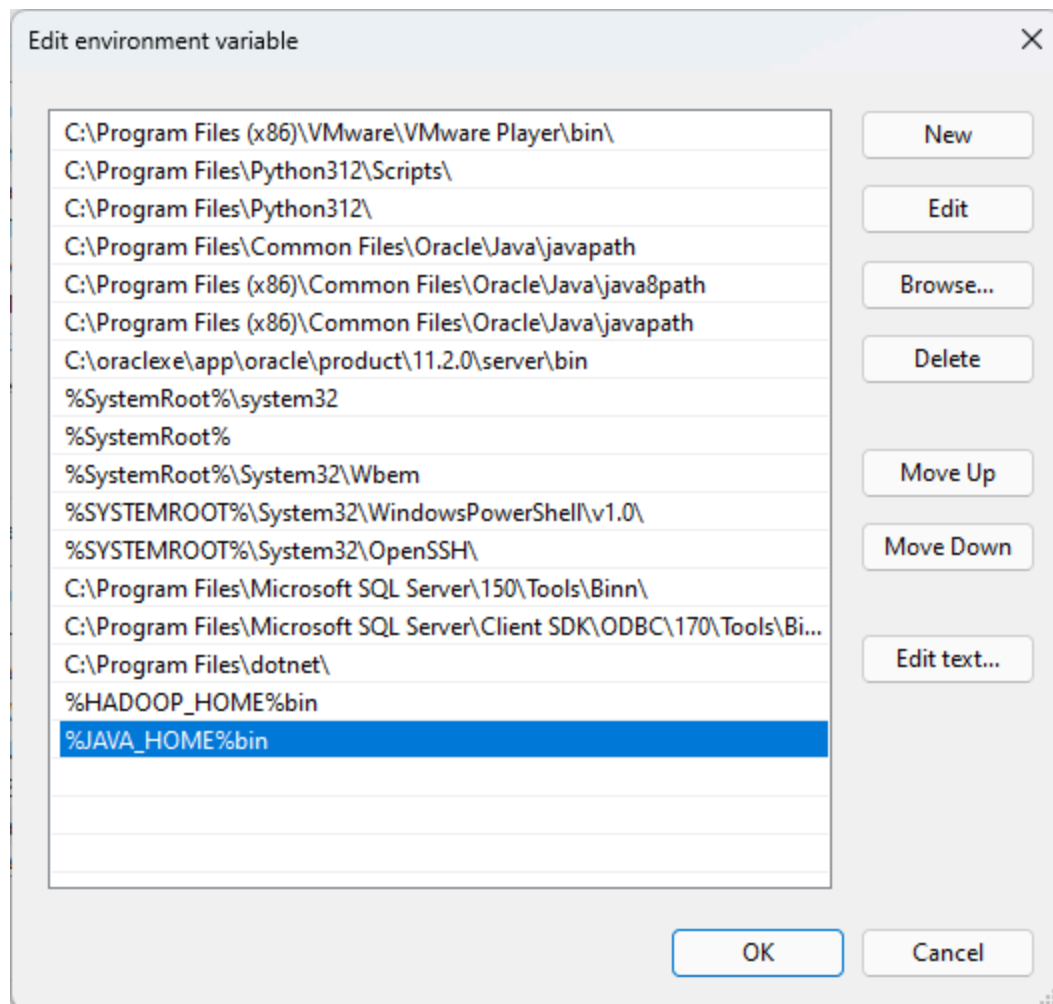
Step 3: Edit Environment Variables.




















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/1iURNbow2IglhAhSy3sfY5xxVfAg33NBW>





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

Name	Date modified	Type	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
 hdfs	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
 mapred	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	Type	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
 https-env	04-03-2024 12:22	SH Source File	2 KB
 https-log4j	04-03-2024 12:22	Properties Source ...	2 KB
 https-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
 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

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

mapred-site.xml

```
<configuration>
  <property>
    <name>mapred.framework.name</name>
    <value>yarn</value>
```

```
    </property>
</configuration>
```

hdfs-site.xml

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>

  <property>
    <name>dfs.namenode.name.dir</name>
    <value>C:\hadoop-3.4.0\data\namenode</value>
  </property>

  <property>
    <name>dfs.datanode.data.dir</name>
    <value>C:\hadoop-3.4.0\data\datanode</value>
  </property>
</configuration>
```

yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.auxservice.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.shuffleHandler</value>
  </property>
</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 12: Restart your PC for the changes to take effect.

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

```
Administrator: Command Pro X + v
Microsoft Windows [Version 10.0.22631.3880]
(c) Microsoft Corporation. All rights reserved.

C:\Users\admin>cd C:\Program Files\hadoop-3.4.0

C:\Program Files\hadoop-3.4.0>cd bin

C:\Program Files\hadoop-3.4.0\bin>
```

```
Administrator: Command Pro X + v
C:\Users\admin>hadoop
Usage: hadoop [--config confdir] [--loglevel loglevel] COMMAND
where COMMAND is one of:
  fs                run a generic filesystem user client
  version           print the version
  jar <jar>         run a jar file
                   note: please use "yarn jar" to launch
                   YARN applications, not this command.
  checknative [-a|-h] check native hadoop and compression libraries availability
  conftest         validate configuration XML files
  distch path:owner:group:permission 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
  key             manage keys via the KeyProvider
  trace           view and modify Hadoop tracing settings
  daemonlog       get/set the log level for each daemon
  or
  CLASSNAME       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
```





```
Apache Hadoop Distribution - yarn  resourcemanager
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, queueCapacity: 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.ApplicationMasterProtocolPB 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 11:07:31,243 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 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.ApplicationClientProtocolPB 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: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
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, queueCapacity: 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.ApplicationMasterProtocolPB 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 11:07:31,243 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 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.ApplicationClientProtocolPB 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: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
```

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: