

Deccan Education Society's
Kirti M. Doongursee College of Arts, Science and Commerce
[NAAC Accredited: "A Grade"]



M.Sc. [Computer Science]

Practical Journal

PAPER: PSCSP3042

Roll Number [_____]

Department of Computer Science and Information Technology

Department of Computer Science and Information Technology
Deccan Education Society's
Kirti M. Doongursee College of Arts, Science and Commerce
[NAAC Accredited: "A Grade"]

C E R T I F I C A T E

This is to certify that Mr./Mrs. _____
of M.Sc. (Computer Science) with Roll No. _____ has completed **5**
Practicals of Paper **PSCSP3042** under my supervision in this College during the
year 2022-2023.

Lecturer-In-Charge

H.O.D.

Dept of CS & IT

Date:

Date:

Examined by:

Remarks:

Date:

Index

Sr.No	Date	Title	Sign
1		Installing and setting environment variables for Working with Apache Hadoop.	
2		Implementing Map-Reduce Program for Word Count problem.	
3		Install Hive and use Hive Create and store structured databases.	
4		Install Pig and Load and Dump the content.	
5		Install SQOOP and list all databases.	

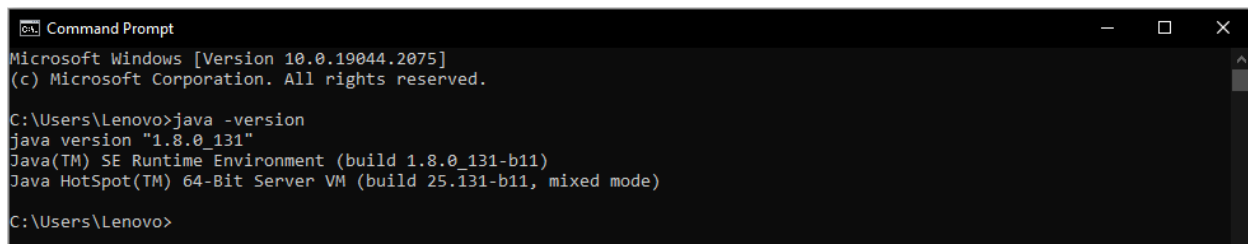
Practical No. 1

Aim: Installing and setting environment variables for Working with Apache Hadoop.

Step 1: Install Java JDK 1.8 in system.

- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/linux-i586/jdk-8u321-linux-aarch64.tar.gz
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/linux-i586/jdk-8u321-linux-arm32-vfp-hflt.tar.gz
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/linux-i586/jdk-8u321-linux-i586.rpm
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/linux-i586/jdk-8u321-linux-i586.tar.gz
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/linux-i586/jdk-8u321-linux-x64.rpm
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/linux-i586/jdk-8u321-linux-x64.tar.gz
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/unix-i586/jdk-8u321-macosx-x64.dmg
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/solaris-i586/jdk-8u321-solaris-sparcv9.tar.Z
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/solaris-i586/jdk-8u321-solaris-sparcv9.tar.gz
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/solaris-i586/jdk-8u321-solaris-x64.tar.Z
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/solaris-i586/jdk-8u321-solaris-x64.tar.gz
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/windows-i586/jdk-8u321-windows-i586.exe
- https://javadl.oracle.com/webapps/download/GetFile/1.8.0_321-b07/df5ad55fdd604472a86a45a217032c7d/windows-i586/jdk-8u321-windows-x64.exe

After downloading java check your java version through this command on cmd.



```
Microsoft Windows [Version 10.0.19044.2075]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>java -version
java version "1.8.0_131"
Java(TM) SE Runtime Environment (build 1.8.0_131-b11)
Java HotSpot(TM) 64-Bit Server VM (build 25.131-b11, mixed mode)

C:\Users\Lenovo>
```

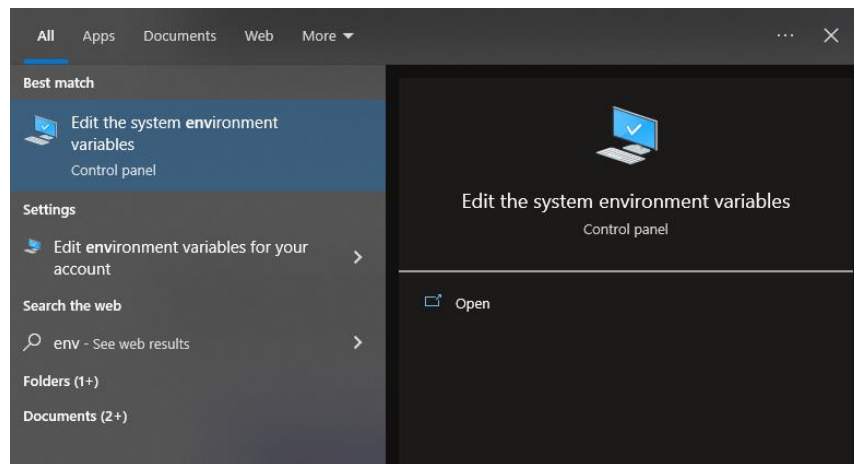
Step 2: Download Hadoop version 3.1 and extract it to C:\ drive.

Hadoop version 3.1 link: <https://archive.apache.org/dist/hadoop/common/hadoop-3.1.0/hadoop-3.1.0.tar.gz>

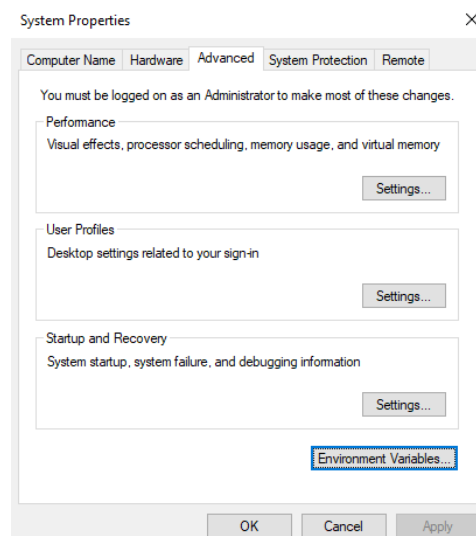
PC > Local Disk (C:) > hadoop > hadoop-3.1.0

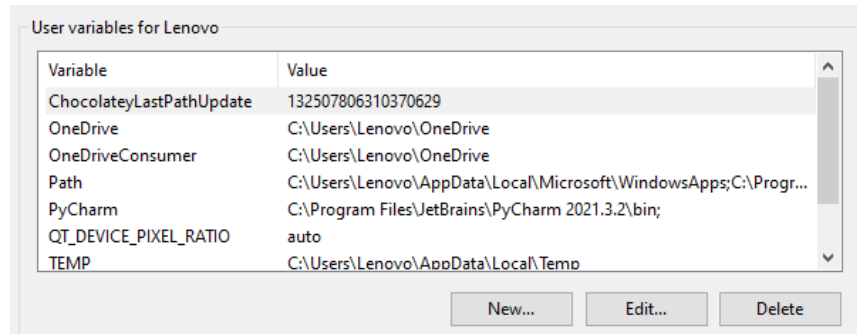
Name	Date modified	Type	Size
bin	4/12/2018 11:33 AM	File folder	
bin_old	11/21/2022 11:49 AM	File folder	
data	11/21/2022 12:34 PM	File folder	
etc	11/21/2022 11:49 AM	File folder	
include	11/21/2022 11:49 AM	File folder	
lib	11/21/2022 11:49 AM	File folder	
libexec	11/21/2022 11:49 AM	File folder	
logs	11/21/2022 4:48 PM	File folder	
sbin	11/21/2022 11:49 AM	File folder	
share	11/21/2022 11:47 AM	File folder	
LICENSE	3/21/2018 11:27 PM	Text Document	144 KB
NOTICE	3/21/2018 11:27 PM	Text Document	22 KB
README	3/21/2018 11:27 PM	Text Document	2 KB

Step 3: Setup System Environment Variables



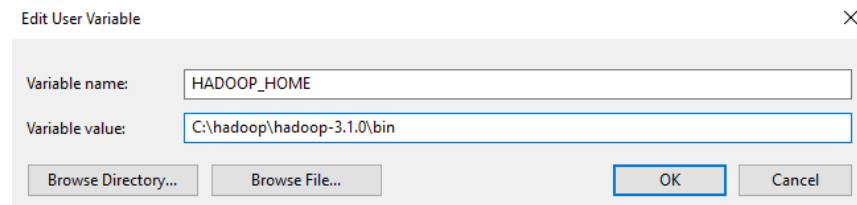
To edit the system environment variable, go to environment variable in system properties.



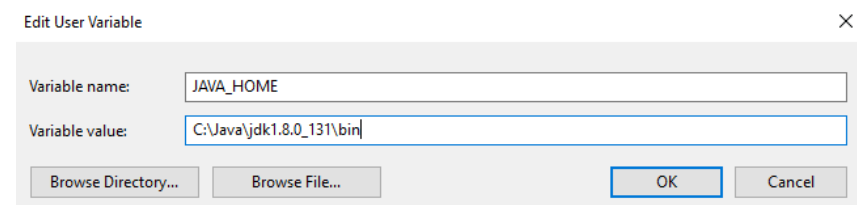


Create 2 new user variables:

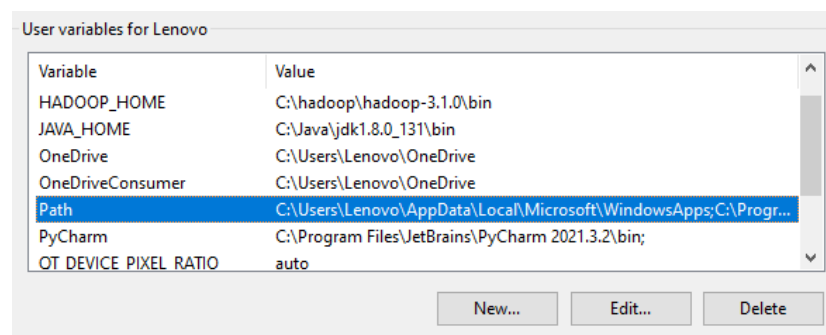
1. Variable Name: HADOOP_HOME
Variable Value: The path of bin folder where extracted Hadoop.



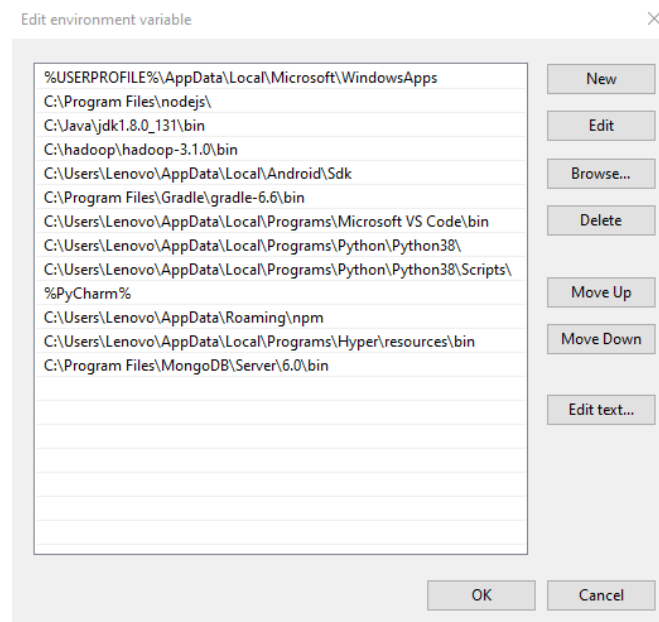
2. Variable Name: JAVA_HOME
Variable Value: The path of the bin folder in the Java directory.



To set Hadoop bin directory and Java bin directory path in system variable path, edit Path in the system variable.



Click on New and add the bin directory path of Hadoop and Java in it.



Step 4: Configuration

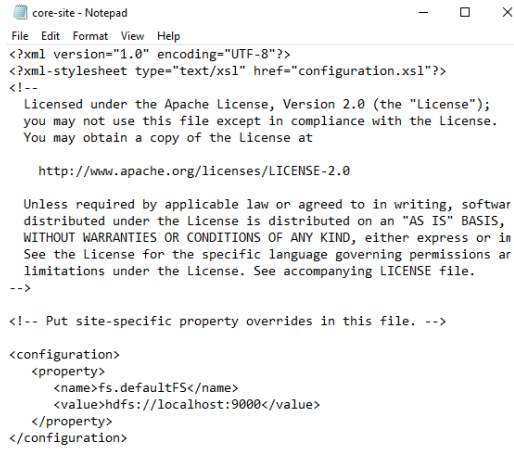
Edit some files located in the Hadoop directory of the etc folder.

PC > Local Disk (C:) > hadoop > hadoop-3.1.0 > etc > hadoop

Name	Date modified	Type	Size
shellprofile.d	11/21/2022 11:49 AM	File folder	
capacity-scheduler	3/30/2018 5:43 AM	XML File	8 KB
configuration	3/30/2018 5:44 AM	XSL Stylesheet	2 KB
container-executor.cfg	3/30/2018 5:43 AM	CFG File	2 KB
core-site	3/30/2018 5:31 AM	XML File	1 KB
hadoop-env	3/30/2018 5:31 AM	Windows Comma...	4 KB
hadoop-env	3/30/2018 5:52 AM	Shell Script	16 KB
hadoop-metrics2.properties	3/30/2018 5:31 AM	PROPERTIES File	4 KB
hadoop-policy	3/30/2018 5:31 AM	XML File	11 KB
hadoop-user-functions.sh.example	3/30/2018 5:31 AM	EXAMPLE File	4 KB
hdfs-site	3/30/2018 5:33 AM	XML File	1 KB
httpfs-env	3/30/2018 5:33 AM	Shell Script	2 KB
httpfs-log4j.properties	3/30/2018 5:33 AM	PROPERTIES File	2 KB
httpfs-signature.secret	3/30/2018 5:33 AM	SECRET File	1 KB
httpfs-site	3/30/2018 5:33 AM	XML File	1 KB
kms-acls	3/30/2018 5:31 AM	XML File	4 KB
kms-env	3/30/2018 5:31 AM	Shell Script	2 KB
kms-log4j.properties	3/30/2018 5:31 AM	PROPERTIES File	2 KB
kms-site	3/30/2018 5:31 AM	XML File	1 KB
log4j.properties	3/30/2018 5:31 AM	PROPERTIES File	14 KB
mapred-env	3/30/2018 5:44 AM	Windows Comma...	1 KB
mapred-env	3/30/2018 5:44 AM	Shell Script	2 KB
mapred-queues.xml.template	3/30/2018 5:44 AM	TEMPLATE File	5 KB
mapred-site	3/30/2018 5:44 AM	XML File	1 KB

- a. Edit the core-site.xml file in the Hadoop directory. Copy this xml property in the configuration in the file and save it.

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



```
core-site - Notepad
File Edit Format View Help
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

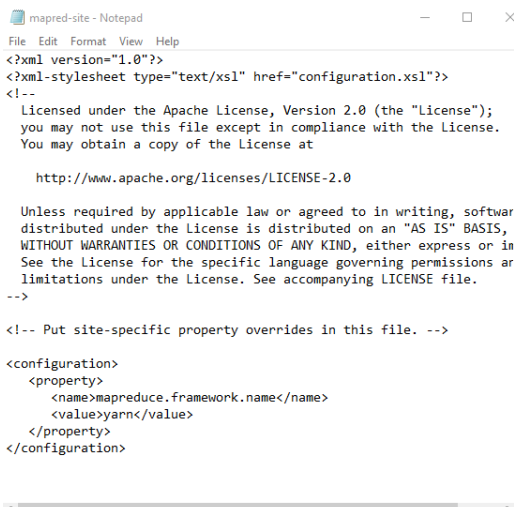
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

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

- b. Edit mapred-site.xml and copy this property in the configuration and save it.

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



```
mapred-site - Notepad
File Edit Format View Help
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

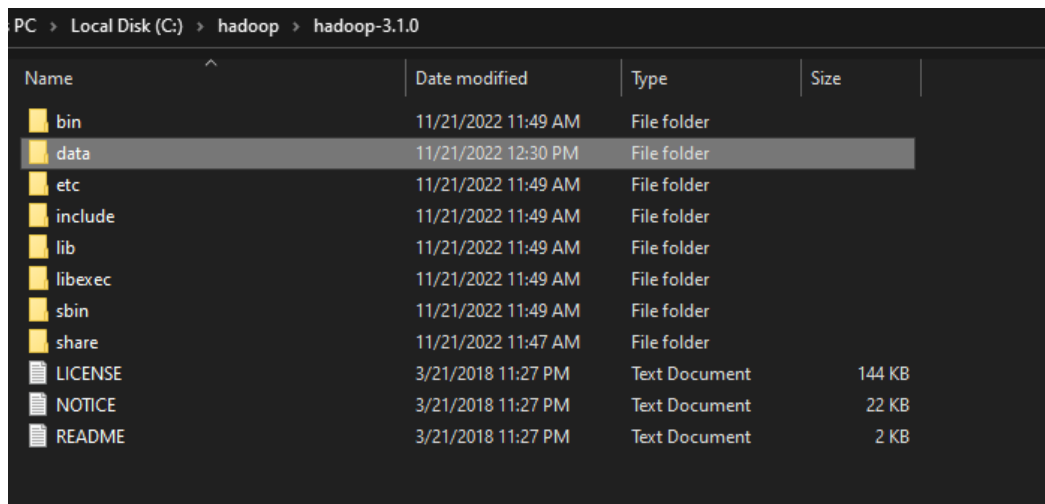
http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

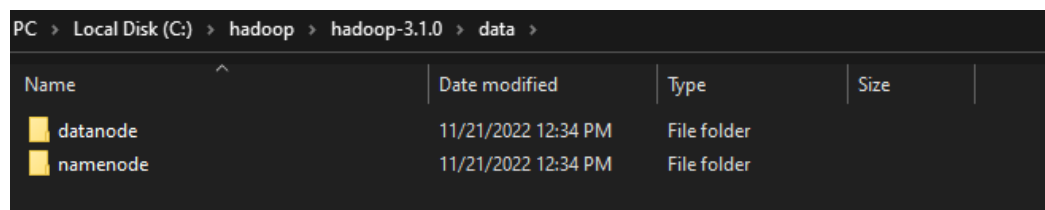
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```


Create a folder data in the Hadoop directory.



Name	Date modified	Type	Size
bin	11/21/2022 11:49 AM	File folder	
data	11/21/2022 12:30 PM	File folder	
etc	11/21/2022 11:49 AM	File folder	
include	11/21/2022 11:49 AM	File folder	
lib	11/21/2022 11:49 AM	File folder	
libexec	11/21/2022 11:49 AM	File folder	
sbin	11/21/2022 11:49 AM	File folder	
share	11/21/2022 11:47 AM	File folder	
LICENSE	3/21/2018 11:27 PM	Text Document	144 KB
NOTICE	3/21/2018 11:27 PM	Text Document	22 KB
README	3/21/2018 11:27 PM	Text Document	2 KB

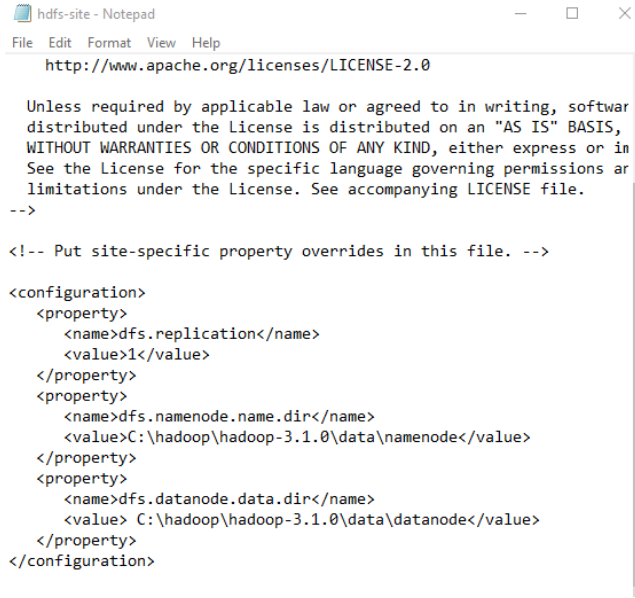
Inside data folder create two new empty folders with the names datanode and namenode



Name	Date modified	Type	Size
datanode	11/21/2022 12:34 PM	File folder	
namenode	11/21/2022 12:34 PM	File folder	

c. Edit the file hdfs-site.xml and add below property in the configuration and save it.

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value> C:\hadoop\hadoop-3.1.0\data\namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value> C:\hadoop\hadoop-3.1.0\data\namenode</value>
  </property>
</configuration>
```



```
hdfs-site - Notepad
File Edit Format View Help
http://www.apache.org/licenses/LICENSE-2.0

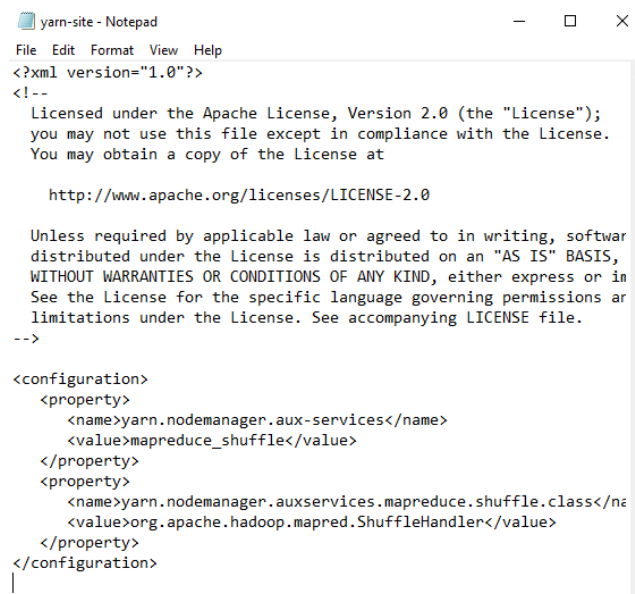
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or in
See the License for the specific language governing permissions ar
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>C:\hadoop\hadoop-3.1.0\data\namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value> C:\hadoop\hadoop-3.1.0\data\datanode</value>
  </property>
</configuration>
```

d. Edit the file yarn-site.xml and add below property in the configuration and save it.

```
<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>
```



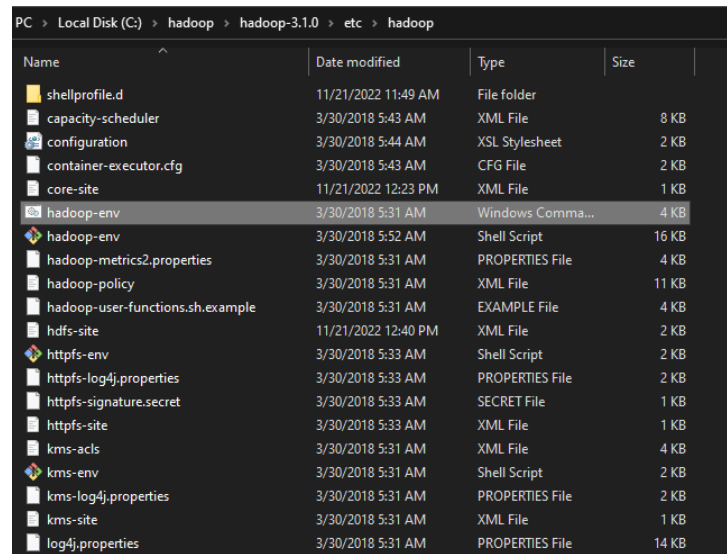
```
yarn-site - Notepad
File Edit Format View Help
<?xml version="1.0"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or in
See the License for the specific language governing permissions ar
limitations under the License. See accompanying LICENSE file.
-->

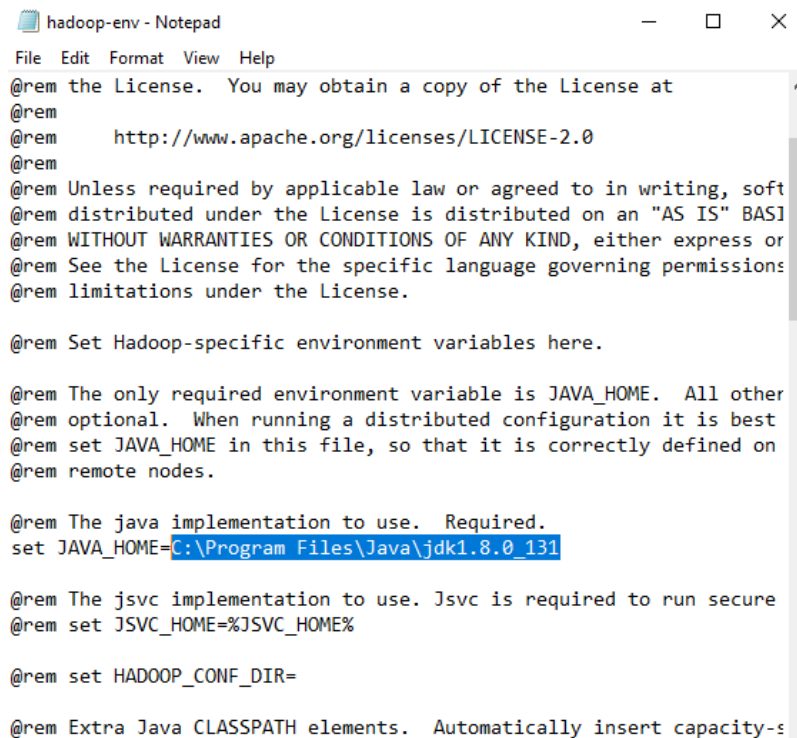
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</na
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
|
```

e. Edit hadoop-env.cmd



Name	Date modified	Type	Size
shellprofile.d	11/21/2022 11:49 AM	File folder	
capacity-scheduler	3/30/2018 5:43 AM	XML File	8 KB
configuration	3/30/2018 5:44 AM	XSL Stylesheet	2 KB
container-executor.cfg	3/30/2018 5:43 AM	CFG File	2 KB
core-site	11/21/2022 12:23 PM	XML File	1 KB
hadoop-env	3/30/2018 5:31 AM	Windows Comma...	4 KB
hadoop-env	3/30/2018 5:52 AM	Shell Script	16 KB
hadoop-metrics2.properties	3/30/2018 5:31 AM	PROPERTIES File	4 KB
hadoop-policy	3/30/2018 5:31 AM	XML File	11 KB
hadoop-user-functions.sh.example	3/30/2018 5:31 AM	EXAMPLE File	4 KB
hdfs-site	11/21/2022 12:40 PM	XML File	2 KB
httpfs-env	3/30/2018 5:33 AM	Shell Script	2 KB
httpfs-log4j.properties	3/30/2018 5:33 AM	PROPERTIES File	2 KB
httpfs-signature.secret	3/30/2018 5:33 AM	SECRET File	1 KB
httpfs-site	3/30/2018 5:33 AM	XML File	1 KB
kms-acls	3/30/2018 5:31 AM	XML File	4 KB
kms-env	3/30/2018 5:31 AM	Shell Script	2 KB
kms-log4j.properties	3/30/2018 5:31 AM	PROPERTIES File	2 KB
kms-site	3/30/2018 5:31 AM	XML File	1 KB
log4j.properties	3/30/2018 5:31 AM	PROPERTIES File	14 KB

Replace %JAVA_HOME% with the path of the java folder where the jdk1.8 is installed and save it.



```
hadoop-env - Notepad
File Edit Format View Help
@rem the License. You may obtain a copy of the License at
@rem
@rem http://www.apache.org/licenses/LICENSE-2.0
@rem
@rem Unless required by applicable law or agreed to in writing, soft
@rem distributed under the License is distributed on an "AS IS" BASI
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or
@rem See the License for the specific language governing permissions
@rem limitations under the License.

@rem Set Hadoop-specific environment variables here.

@rem The only required environment variable is JAVA_HOME. All other
@rem optional. When running a distributed configuration it is best
@rem set JAVA_HOME in this file, so that it is correctly defined on
@rem remote nodes.

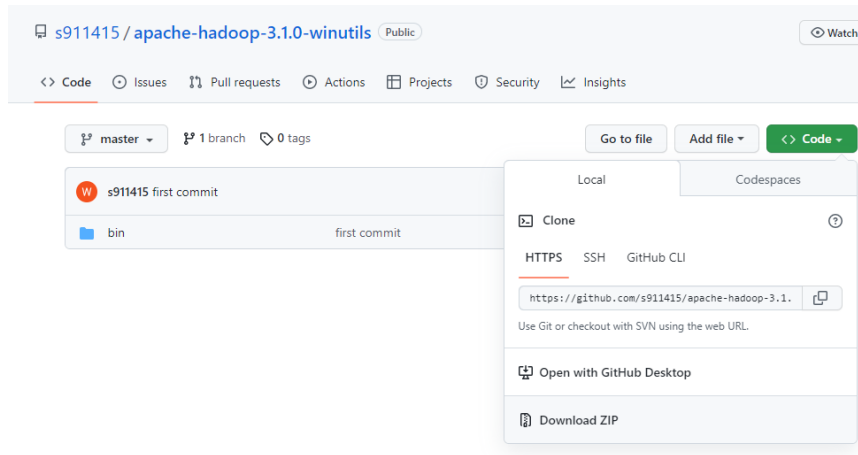
@rem The java implementation to use. Required.
set JAVA_HOME=C:\Program Files\Java\jdk1.8.0_131

@rem The jsvc implementation to use. Jsvc is required to run secure
@rem set JSVC_HOME=%JSVC_HOME%

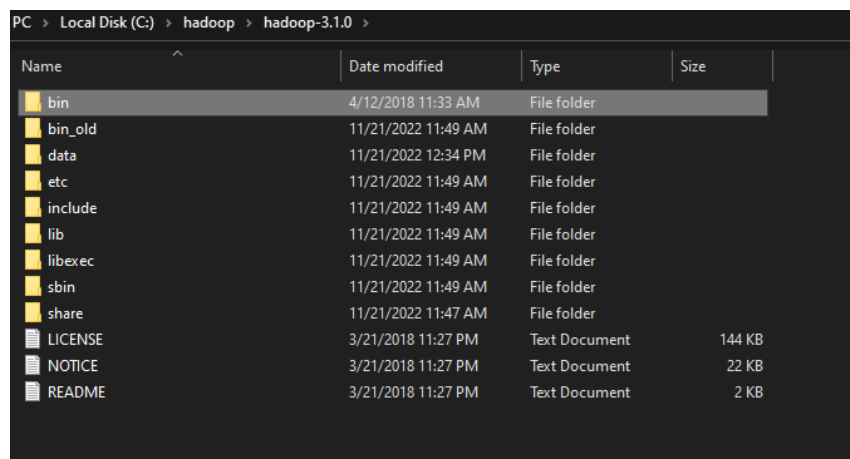
@rem set HADOOP_CONF_DIR=

@rem Extra Java CLASSPATH elements. Automatically insert capacity-s
```

Hadoop needs Windows OS-specific files which do not come with default download of Hadoop. To include those files, replace the bin folder in hadoop directory with the bin folder provided in this [GitHub](https://github.com/s911415/apache-hadoop-3.1.0-winutils) link. (<https://github.com/s911415/apache-hadoop-3.1.0-winutils>)



Download zip file. Extract it and copy the bin folder in it. Rename old bin folder like bin_old. Paste new downloaded bin folder in that directory.



Check whether hadoop is successfully installed by running this command hadoop version on cmd.

```
Microsoft Windows [Version 10.0.19044.2075]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>hadoop version
Hadoop 3.1.0
Source code repository https://github.com/apache/hadoop -r 16b70619a24cdc5d3b0fcf4b58ca77238ccbe6d
Compiled by centos on 2018-03-30T00:00Z
Compiled with protoc 2.5.0
From source with checksum 14182d20c972b3e2105580a1ad6990
This command was run using /C:/hadoop/hadoop-3.1.0/share/hadoop/common/hadoop-common-3.1.0.jar

C:\Users\Lenovo>
```

Hadoop is successfully installed in the system.

Step 5: Format the NameNode

Once the hadoop is installed, the NameNode is formatted. This is done to avoid deletion of all the data inside HDFS.

```
hdfs namenode -format
```

```

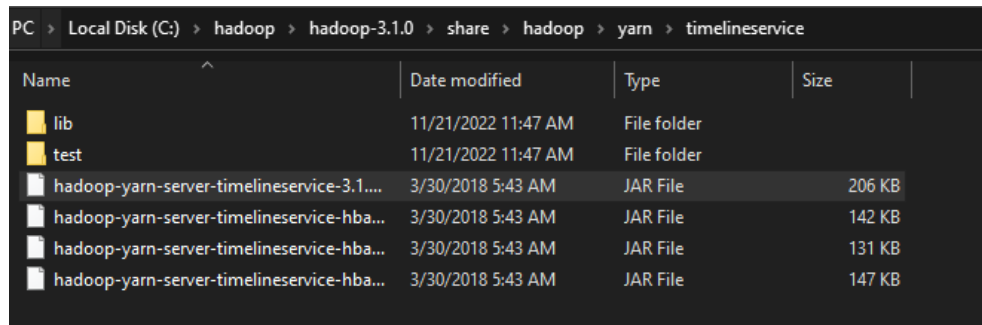
C:\Users\Lenovo>hdfs namenode -format
2022-11-21 14:31:16,978 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = LENOVO\192.168.43.163
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.1.0
STARTUP_MSG: classpath = C:\hadoop\hadoop-3.1.0\etc\hadoop;C:\hadoop\hadoop-3.1.0\share\hadoop\common;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\accessors-smart-1.2.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\asm-5.0.4.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\avro-1.7.7.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-beanutils-1.9.3.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-cli-1.2.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-codec-1.11.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-collections-3.2.2.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-compress-1.4.1.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-configuration-2.4.1.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-io-2.5.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-lang-2.6.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-lang3-3.4.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-logging-1.1.3.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-math3-3.1.1.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\commons-net-3.6.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\curator-client-2.12.0.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\curator-framework-2.12.0.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\curator-recipes-2.12.0.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\gson-2.2.4.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\guava-11.0.2.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\hadoop-annotations-3.1.0.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\hadoop-auth-3.1.0.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\hadoop-core-1.3.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\htrace-core4-4.1.0-incubating.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\httpclient-4.5.2.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\httpcore-4.4.4.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\jackson-annotations-2.7.8.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\jackson-core-2.7.8.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\jackson-core-asl-1.9.13.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\jackson-databind-2.7.8.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\jackson-jaxrs-1.9.13.jar;C:\hadoop\hadoop-3.1.0\share\hadoop\common\lib\jackson-mapper-asl-1.9.13.jar;C:\hadoop\hadoop-3.1.0
*****/

```

Copy `hadoop-yarn-server-timelineservice-3.1.0` from `timelineservice` folder in `yarn` directory which is in `hadoop` located in the `hadoop` directory of the `share` folder where we installed `hadoop` to `yarn` directory which is in `hadoop` located in the `hadoop` directory of the `share` folder where we installed `hadoop`.

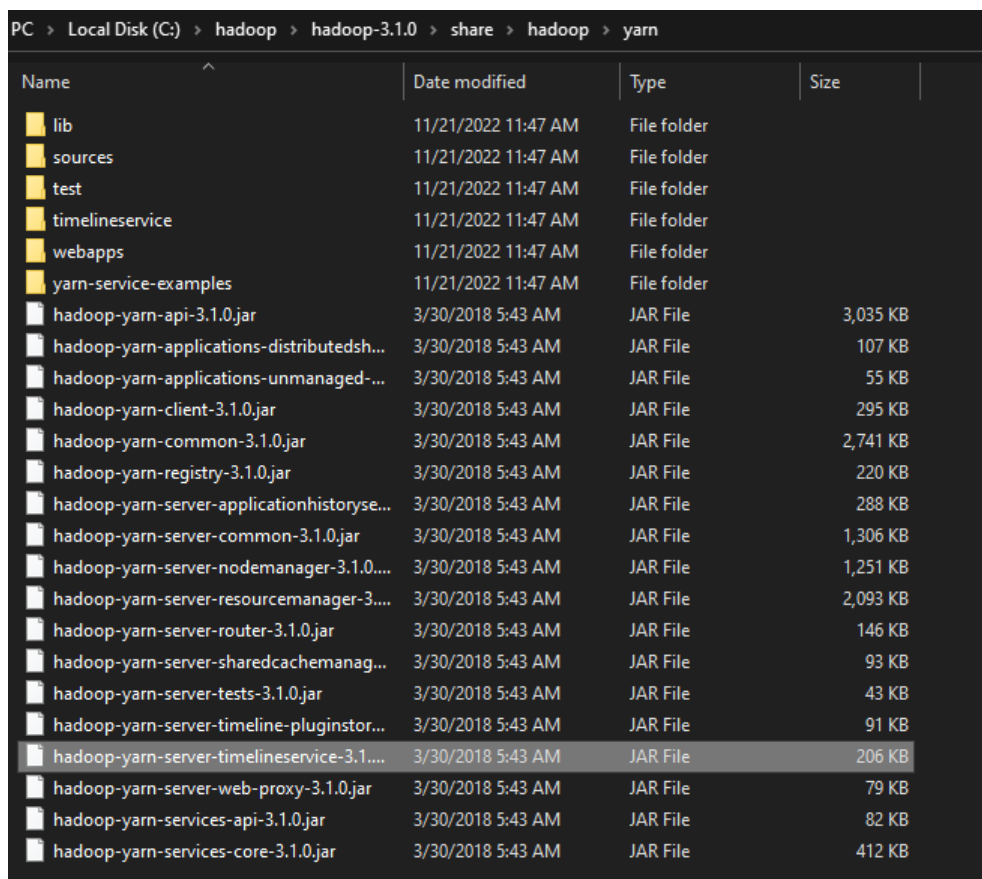
Name	Date modified	Type	Size
lib	11/21/2022 11:47 AM	File folder	
sources	11/21/2022 11:47 AM	File folder	
test	11/21/2022 11:47 AM	File folder	
timelineservice	11/21/2022 11:47 AM	File folder	
webapps	11/21/2022 11:47 AM	File folder	
yarn-service-examples	11/21/2022 11:47 AM	File folder	
hadoop-yarn-api-3.1.0.jar	3/30/2018 5:43 AM	JAR File	3,035 KB
hadoop-yarn-applications-distributedsh...	3/30/2018 5:43 AM	JAR File	107 KB
hadoop-yarn-applications-unmanaged-...	3/30/2018 5:43 AM	JAR File	55 KB
hadoop-yarn-client-3.1.0.jar	3/30/2018 5:43 AM	JAR File	295 KB
hadoop-yarn-common-3.1.0.jar	3/30/2018 5:43 AM	JAR File	2,741 KB
hadoop-yarn-registry-3.1.0.jar	3/30/2018 5:43 AM	JAR File	220 KB
hadoop-yarn-server-applicationhistoryse...	3/30/2018 5:43 AM	JAR File	288 KB
hadoop-yarn-server-common-3.1.0.jar	3/30/2018 5:43 AM	JAR File	1,306 KB
hadoop-yarn-server-nodemanager-3.1.0....	3/30/2018 5:43 AM	JAR File	1,251 KB
hadoop-yarn-server-resourcemanager-3....	3/30/2018 5:43 AM	JAR File	2,093 KB
hadoop-yarn-server-router-3.1.0.jar	3/30/2018 5:43 AM	JAR File	146 KB
hadoop-yarn-server-sharedcachemanag...	3/30/2018 5:43 AM	JAR File	93 KB
hadoop-yarn-server-tests-3.1.0.jar	3/30/2018 5:43 AM	JAR File	43 KB
hadoop-yarn-server-timeline-pluginstor...	3/30/2018 5:43 AM	JAR File	91 KB
hadoop-yarn-server-web-proxy-3.1.0.jar	3/30/2018 5:43 AM	JAR File	79 KB
hadoop-yarn-services-api-3.1.0.jar	3/30/2018 5:43 AM	JAR File	82 KB
hadoop-yarn-services-core-3.1.0.jar	3/30/2018 5:43 AM	JAR File	412 KB

Copy `hadoop-yarn-server-timelineservice-3.1.0`



Name	Date modified	Type	Size
lib	11/21/2022 11:47 AM	File folder	
test	11/21/2022 11:47 AM	File folder	
hadoop-yarn-server-timelineservice-3.1....	3/30/2018 5:43 AM	JAR File	206 KB
hadoop-yarn-server-timelineservice-hba...	3/30/2018 5:43 AM	JAR File	142 KB
hadoop-yarn-server-timelineservice-hba...	3/30/2018 5:43 AM	JAR File	131 KB
hadoop-yarn-server-timelineservice-hba...	3/30/2018 5:43 AM	JAR File	147 KB

Paste `hadoop-yarn-server-timelineservice-3.1.0` in `yarn` folder.



Name	Date modified	Type	Size
lib	11/21/2022 11:47 AM	File folder	
sources	11/21/2022 11:47 AM	File folder	
test	11/21/2022 11:47 AM	File folder	
timelineservice	11/21/2022 11:47 AM	File folder	
webapps	11/21/2022 11:47 AM	File folder	
yarn-service-examples	11/21/2022 11:47 AM	File folder	
hadoop-yarn-api-3.1.0.jar	3/30/2018 5:43 AM	JAR File	3,035 KB
hadoop-yarn-applications-distributedsh...	3/30/2018 5:43 AM	JAR File	107 KB
hadoop-yarn-applications-unmanaged-...	3/30/2018 5:43 AM	JAR File	55 KB
hadoop-yarn-client-3.1.0.jar	3/30/2018 5:43 AM	JAR File	295 KB
hadoop-yarn-common-3.1.0.jar	3/30/2018 5:43 AM	JAR File	2,741 KB
hadoop-yarn-registry-3.1.0.jar	3/30/2018 5:43 AM	JAR File	220 KB
hadoop-yarn-server-applicationhistoryse...	3/30/2018 5:43 AM	JAR File	288 KB
hadoop-yarn-server-common-3.1.0.jar	3/30/2018 5:43 AM	JAR File	1,306 KB
hadoop-yarn-server-nodemanager-3.1.0....	3/30/2018 5:43 AM	JAR File	1,251 KB
hadoop-yarn-server-resourcemanager-3....	3/30/2018 5:43 AM	JAR File	2,093 KB
hadoop-yarn-server-router-3.1.0.jar	3/30/2018 5:43 AM	JAR File	146 KB
hadoop-yarn-server-sharedcachemanag...	3/30/2018 5:43 AM	JAR File	93 KB
hadoop-yarn-server-tests-3.1.0.jar	3/30/2018 5:43 AM	JAR File	43 KB
hadoop-yarn-server-timeline-pluginstor...	3/30/2018 5:43 AM	JAR File	91 KB
hadoop-yarn-server-timelineservice-3.1....	3/30/2018 5:43 AM	JAR File	206 KB
hadoop-yarn-server-web-proxy-3.1.0.jar	3/30/2018 5:43 AM	JAR File	79 KB
hadoop-yarn-services-api-3.1.0.jar	3/30/2018 5:43 AM	JAR File	82 KB
hadoop-yarn-services-core-3.1.0.jar	3/30/2018 5:43 AM	JAR File	412 KB

Step 6: To start run all the Apache Hadoop Distribution

change the directory in cmd to `sbin` folder of `hadoop` directory.

Command: `cd C:\hadoop\hadoop-3.1.0\sbin`

```
C:\ Command Prompt

Microsoft Windows [Version 10.0.19044.2075]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>cd C:\hadoop\hadoop-3.1.0\sbin

C:\hadoop\hadoop-3.1.0\sbin>
```

Start namenode and datanode

Command: start-dfs.cmd

```
C:\ Command Prompt

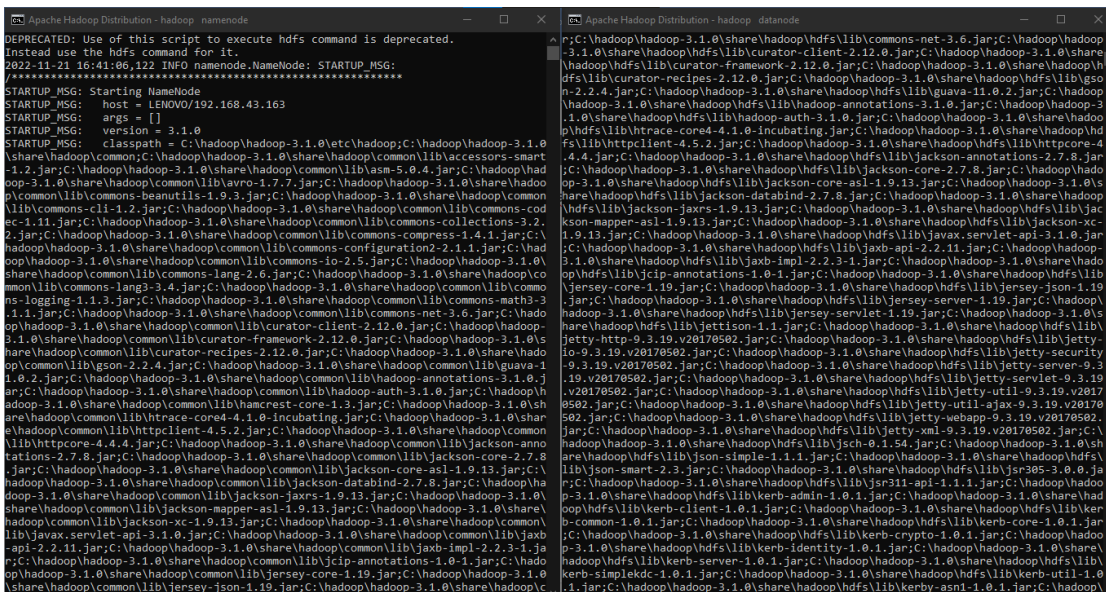
Microsoft Windows [Version 10.0.19044.2075]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>cd C:\hadoop\hadoop-3.1.0\sbin

C:\hadoop\hadoop-3.1.0\sbin>start-dfs.cmd

C:\hadoop\hadoop-3.1.0\sbin>
```

Two cmd windows will open for NameNode and DataNode



Start yarn

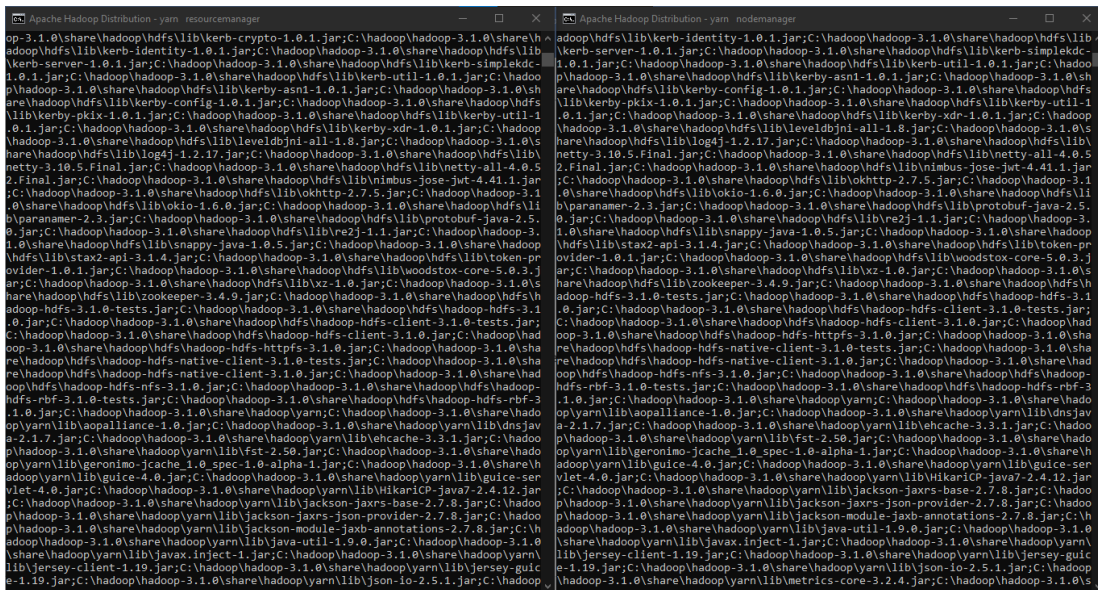
Command: start-yarn.cmd

```
C:\ Command Prompt

C:\hadoop\hadoop-3.1.0\sbin>start-yarn.cmd
starting yarn daemons

C:\hadoop\hadoop-3.1.0\sbin>
```

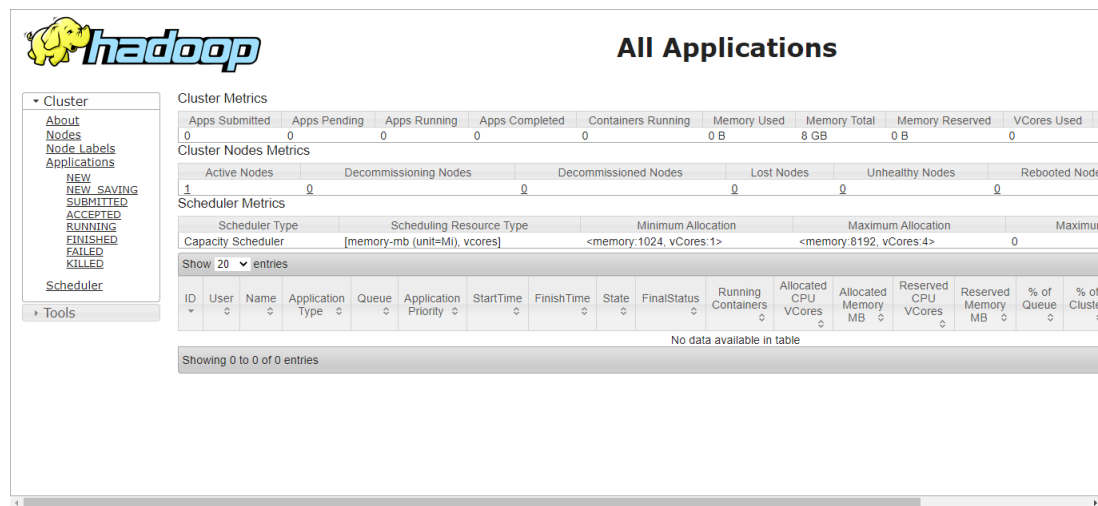

Two more windows will open, one for yarn resource manager and one for yarn node manager



The image shows two terminal windows side-by-side. The left window is titled 'Apache Hadoop Distribution - yarn_resourcemanager' and the right window is titled 'Apache Hadoop Distribution - yarn_nodemanager'. Both windows display a long list of file paths, likely representing the classpath for the Hadoop daemons. The paths include various Hadoop binaries and configuration files, such as 'hadoop-hdfs', 'hadoop-yarn', 'hadoop-common', and 'hadoop-mapreduce'.

Step 7: Verification

To access information about resource manager current jobs, successful and failed jobs, go to this link in browser- <http://localhost:8088/cluster>



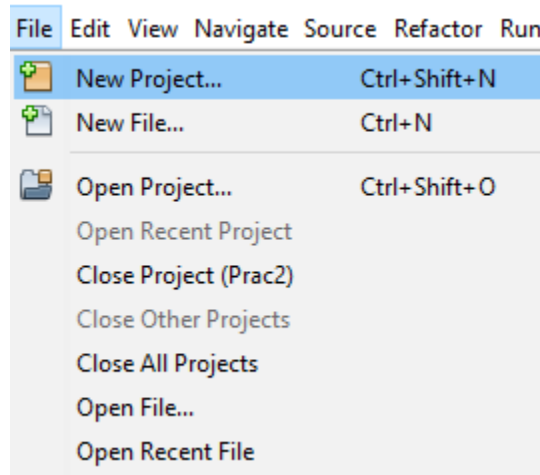
The screenshot shows the 'All Applications' page in the Hadoop web interface. The page has a sidebar on the left with navigation links: Cluster, About, Nodes, Node Labels, Applications, NEW, NEW SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED, and Scheduler. The main content area displays 'Cluster Metrics' and 'Cluster Nodes Metrics'. The 'Cluster Metrics' table shows various statistics: Apps Submitted (0), Apps Pending (0), Apps Running (0), Apps Completed (0), Containers Running (0 B), Memory Used (8 GB), Memory Total (0 B), Memory Reserved (0 B), and VCoers Used (0). The 'Cluster Nodes Metrics' table shows Active Nodes (1), Decommissioning Nodes (0), Decommissioned Nodes (0), Lost Nodes (0), Unhealthy Nodes (0), and Rebooted Node (0). Below these tables, there is a 'Scheduler Metrics' section showing Capacity Scheduler, Scheduling Resource Type (memory-mb (unit=Mi), vcores), Minimum Allocation (<memory:1024, vCores:1>), Maximum Allocation (<memory:8192, vCores:4>), and Maximum (0). At the bottom, there is a table with columns for ID, User, Name, Application, Queue, Application Priority, StartTime, FinishTime, State, FinalStatus, Running Containers, Allocated CPU VCoers, Allocated Memory MB, Reserved CPU VCoers, Reserved Memory MB, % of Queue, and % of Cluste. The table currently shows 'No data available in table'.

To check the details about the hdfs (namenode and datanode), go to this link <http://localhost:50070/> in browser.

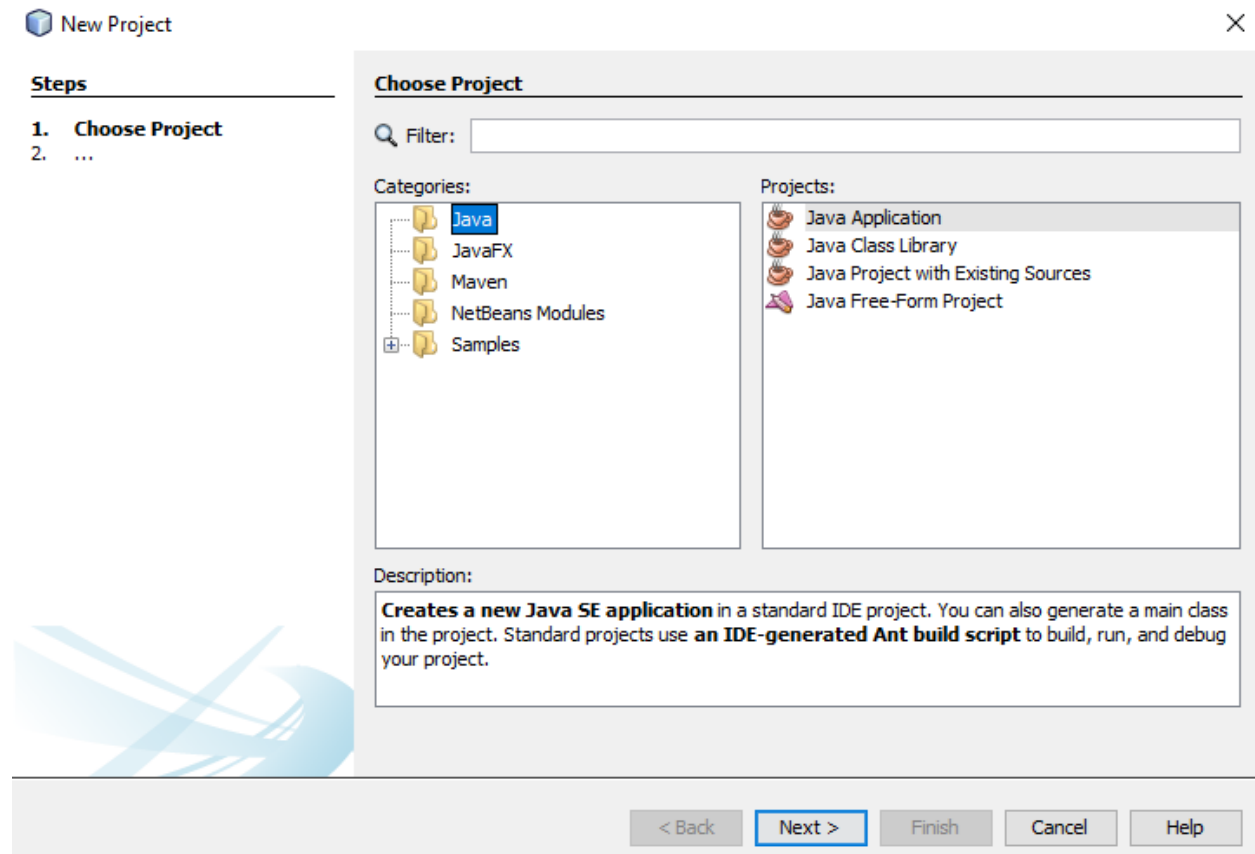
Practical No. 2

Aim: Implementing Map-Reduce Program for Word Count problems.

Step 1: Open NetBeans. Click on File and Select New Project.



Select Category Java and Project Java Application.



Give the Project Name and Click Finish.

New Java Application

Steps

1. Choose Project
2. **Name and Location**

Name and Location

Project Name: WordCount1

Project Location: C:\Users\DELL\Documents\NetBeansProjects Browse...

Project Folder: C:\Users\DELL\Documents\NetBeansProjects\WordCount1

☐ Use Dedicated Folder for Storing Libraries

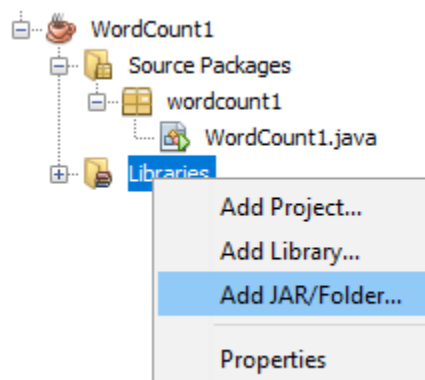
Libraries Folder: Browse...

Different users and projects can share the same compilation libraries (see Help for details).

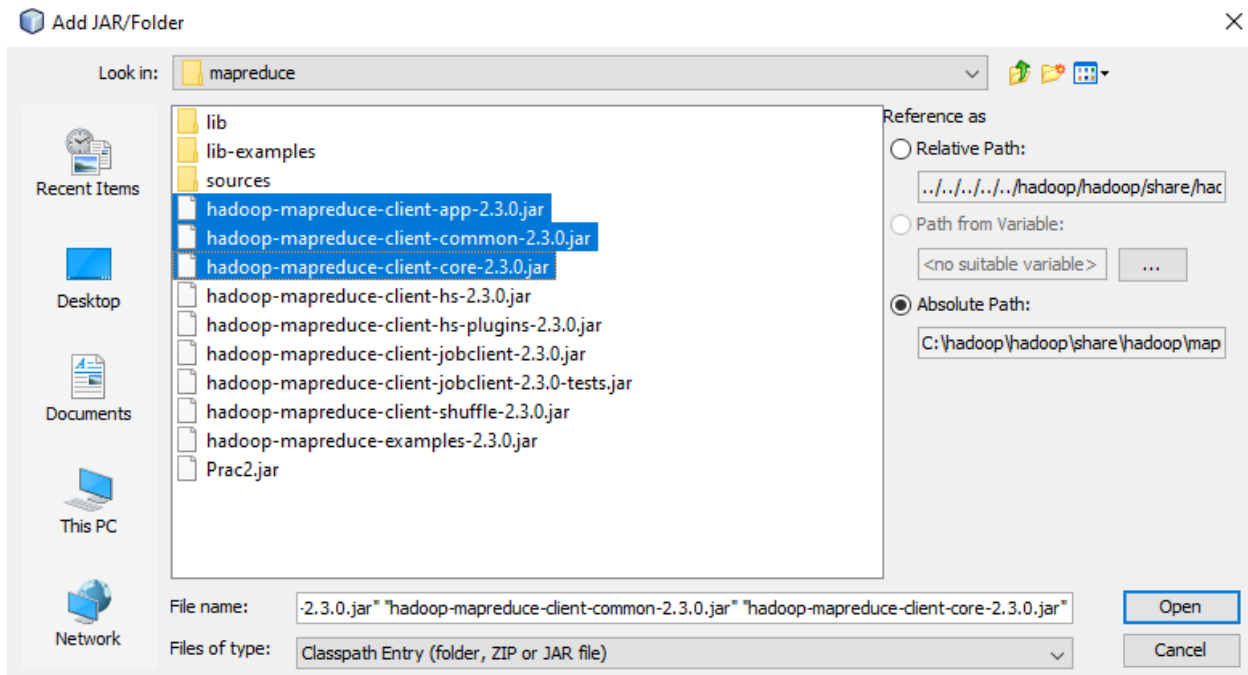
☒ Create Main Class wordcount1.WordCount1

< Back Next > **Finish** Cancel Help

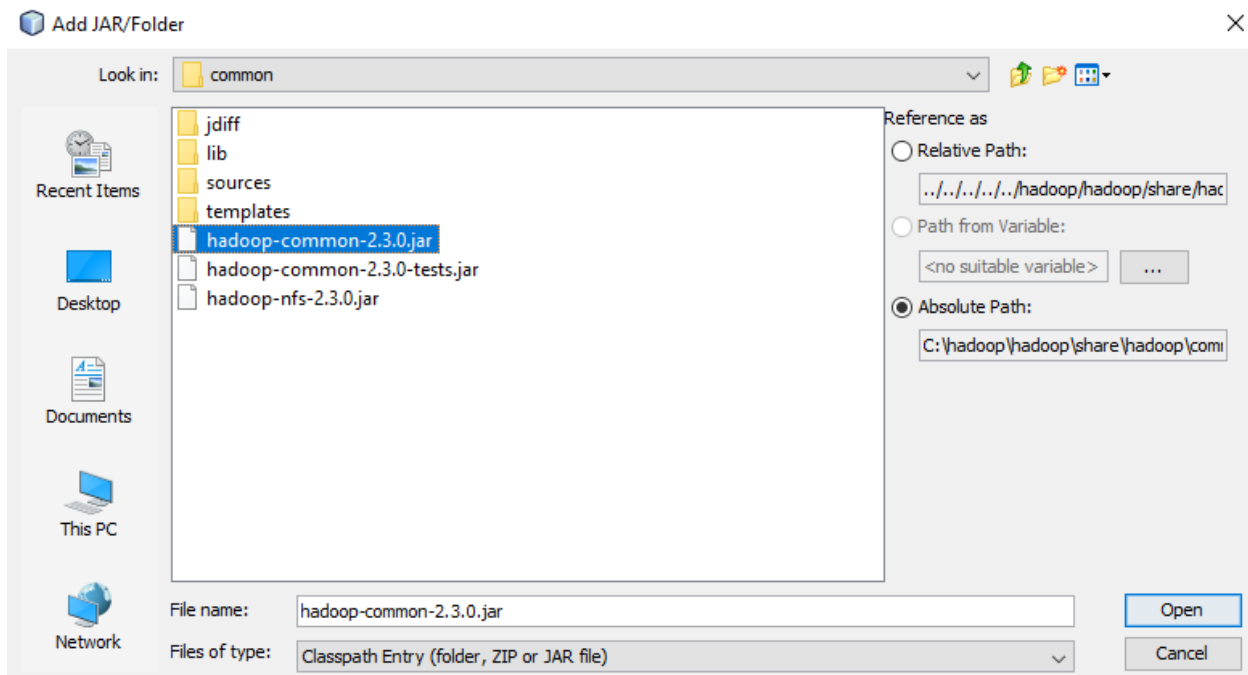
Right Click on Libraries and Select Add JAR/Folder option.



Then Go to C:\hadoop\hadoop\share\hadoop\mapreduce this path and select First three JAR files and Click on Open.



Again, Go to Libraries and Select Add JAR/Folder option. Then go to C:\hadoop\hadoop\share\hadoop\common this path and select JAR file.



Step 2: Write the following code

Code:

```
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
```

```
public class WordCount1 {

    public static class TokenizerMapper
        extends Mapper<Object, Text, Text, IntWritable>{

        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(Object key, Text value, Mapper.Context context
            ) throws IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }
}
```

```

    }
}

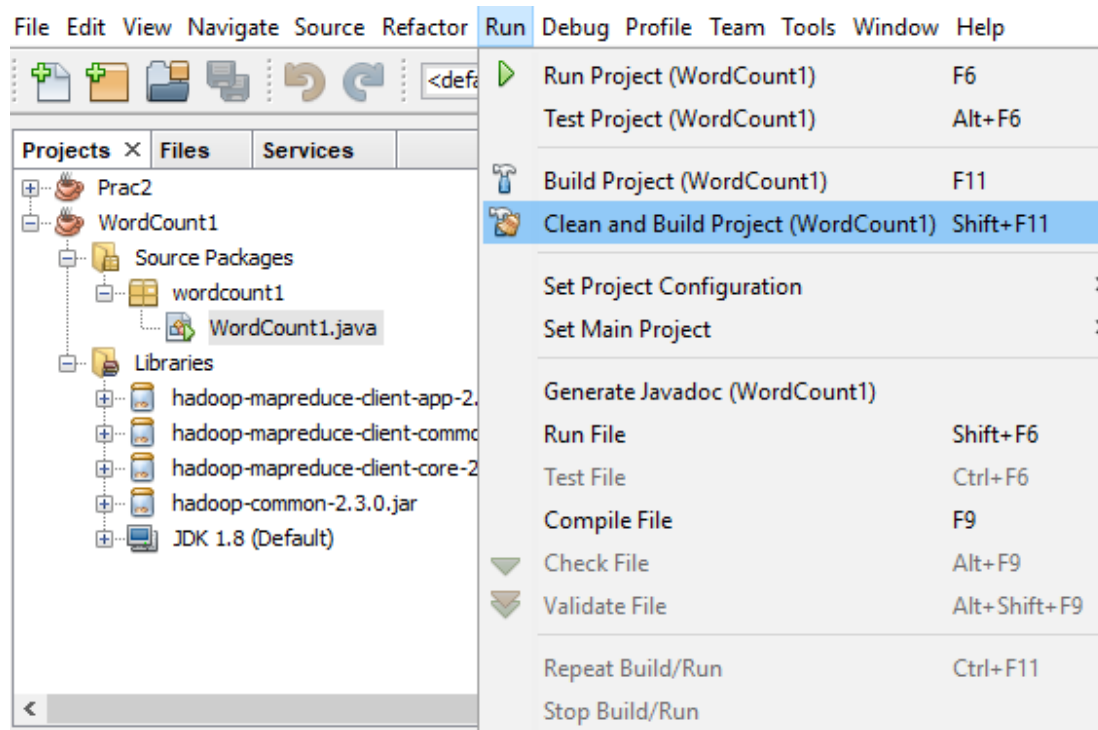
public static class IntSumReducer
    extends Reducer<Text, IntWritable, Text, IntWritable> {
    private IntWritable result = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values, Reducer.Context context)
        throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
            context.write(key, result);
        }
    }
}

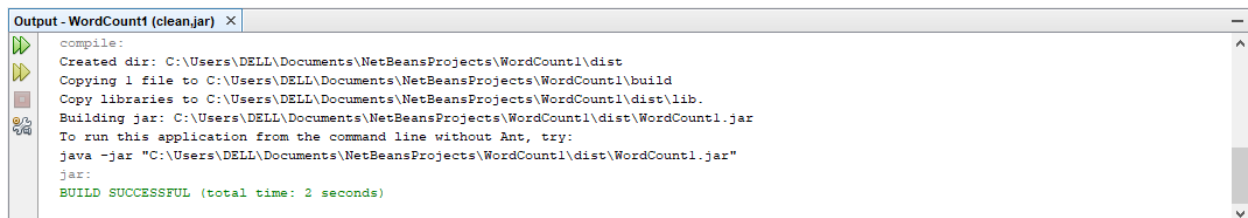
public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "wordcounter");
    job.setJarByClass(WordCount1.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

Save file and then Go to run and select Clean and Build Project.

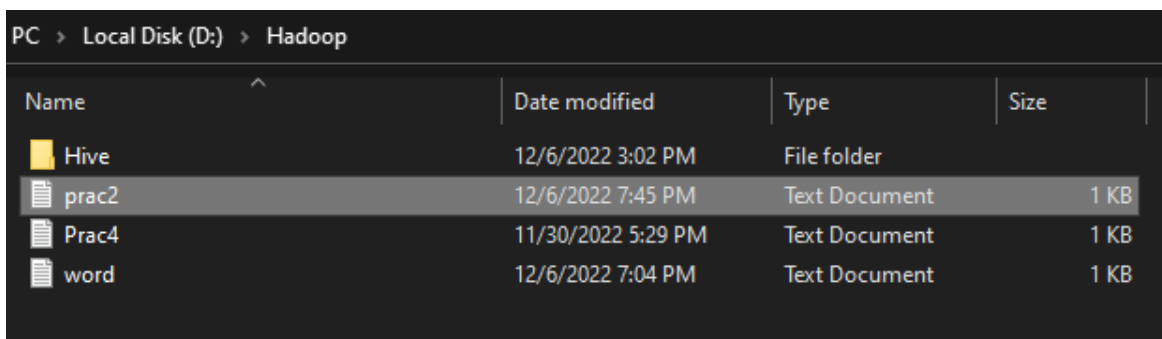


The Message appear Build Successful. It successfully builds WordCount1.jar

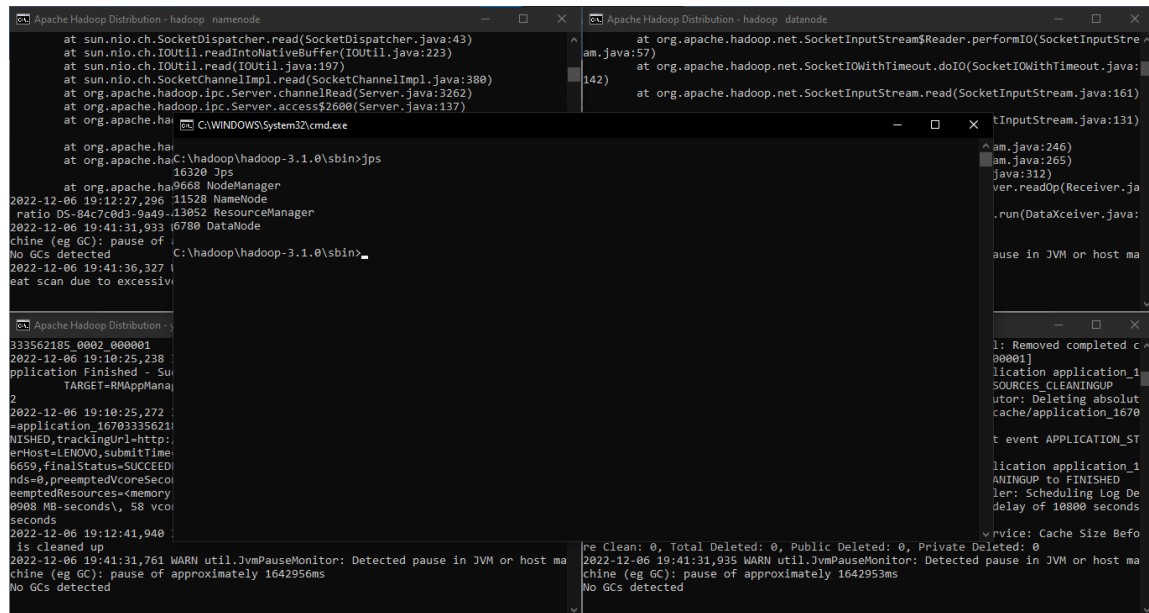


Step 3: Use this JAR for word count.

Create New File and Write content inside the file.



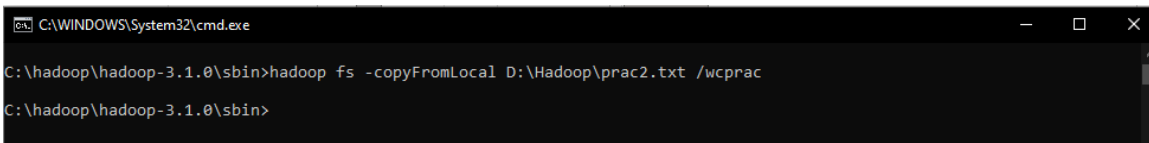
Step 4: Start all Hadoop Services.

The screenshot shows three overlapping terminal windows from the 'Apache Hadoop Distribution' application. The leftmost window is the 'namenode' log, showing startup messages for the NameNode and DataNode, including the command 'C:\hadoop\hadoop-3.1.0\sbin>jps'. The middle window is the 'datanode' log, showing startup messages for the DataNode. The rightmost window is a Java application log, showing messages from 'am.java:246' through 'am.java:312', including 'ver.readOp(Receiver.java:312)' and 'run(DataXceiver.java:312)'. The logs indicate that the Hadoop services are running successfully.

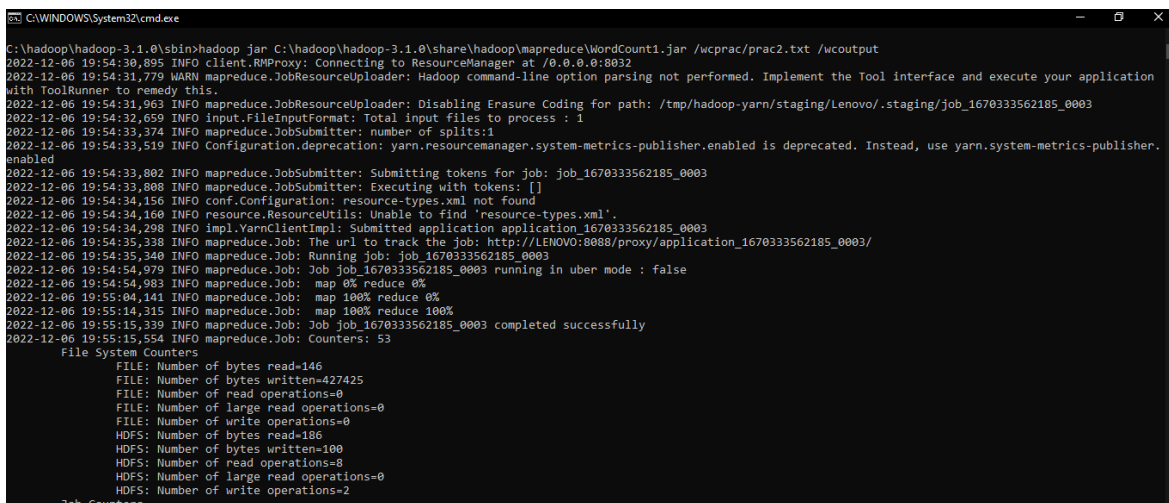
Create new directory in HDFS.

The screenshot shows a Windows command prompt window titled 'C:\WINDOWS\System32\cmd.exe'. The user has entered the command 'C:\hadoop\hadoop-3.1.0\sbin>hadoop fs -mkdir /wcprac' and the prompt has moved to the next line, indicating the command was executed successfully.

Copy prac2.txt file in /wcprac

The screenshot shows a Windows command prompt window titled 'C:\WINDOWS\System32\cmd.exe'. The user has entered the command 'C:\hadoop\hadoop-3.1.0\sbin>hadoop fs -copyFromLocal D:\Hadoop\prac2.txt /wcprac' and the prompt has moved to the next line, indicating the command was executed successfully.

Count the words from prac2.txt file using above WordCount1.jar

The screenshot shows a Windows command prompt window titled 'C:\WINDOWS\System32\cmd.exe'. The user has entered the command 'C:\hadoop\hadoop-3.1.0\sbin>hadoop jar C:\hadoop\hadoop-3.1.0\share\hadoop\mapreduce\WordCount1.jar /wcprac/prac2.txt /wcoutput'. The output shows the progress of the WordCount application, including messages from 'client.RMPProxy', 'mapreduce.JobResourceUploader', 'mapreduce.JobSubmitter', and 'mapreduce.Job'. The final output shows the word counts for the file 'prac2.txt' in the directory '/wcprac'.

See the word count using -cat command.

```
C:\WINDOWS\System32\cmd.exe
C:\hadoop\hadoop-3.1.0\sbin>hadoop fs -ls /wcoutput
Found 2 items
-rw-r--r-- 1 Lenovo supergroup 0 2022-12-06 19:55 /wcoutput/_SUCCESS
-rw-r--r-- 1 Lenovo supergroup 100 2022-12-06 19:55 /wcoutput/part-r-00000

C:\hadoop\hadoop-3.1.0\sbin>hadoop fs -cat /wcoutput/part-r-00000
After 1
Downloading 1
Hive, 1
Unzip 1
apache-hive-3.1.2-bin.tar.gz 1
file. 1
need 1
the 2
to 1
we 1

C:\hadoop\hadoop-3.1.0\sbin>
```









Practical No. 3

Aim: Install Hive and use Hive Create and store structured databases.

Step 1: Download Apache Hive and Apache Derby







Download Apache Hive link: <https://archive.apache.org/dist/hive/hive-2.1.0/apache-hive-2.1.0-bin.tar.gz>

Index of /dist/hive/hive-2.1.0

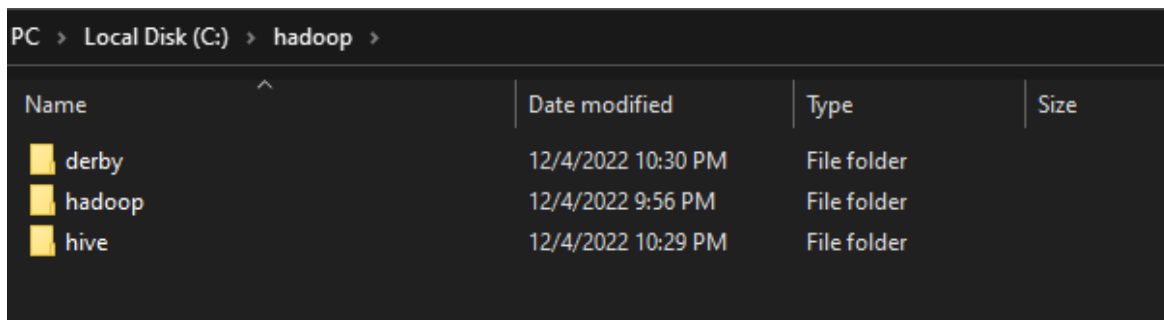
Name	Last modified	Size	Description
 Parent Directory		-	
 apache-hive-2.1.0-bin.tar.gz	2016-06-21 01:26	143M	
 apache-hive-2.1.0-bin.tar.gz.asc	2016-06-21 01:26	819	
 apache-hive-2.1.0-bin.tar.gz.md5	2016-06-21 01:26	70	
 apache-hive-2.1.0-src.tar.gz	2016-06-21 01:26	18M	
 apache-hive-2.1.0-src.tar.gz.asc	2016-06-21 01:26	819	
 apache-hive-2.1.0-src.tar.gz.md5	2016-06-21 01:26	70	

Download Apache Derby link: <https://archive.apache.org/dist/db/derby/db-derby-10.12.1.1/db-derby-10.12.1.1-bin.tar.gz>

Index of /dist/db/derby/db-derby-10.12.1.1

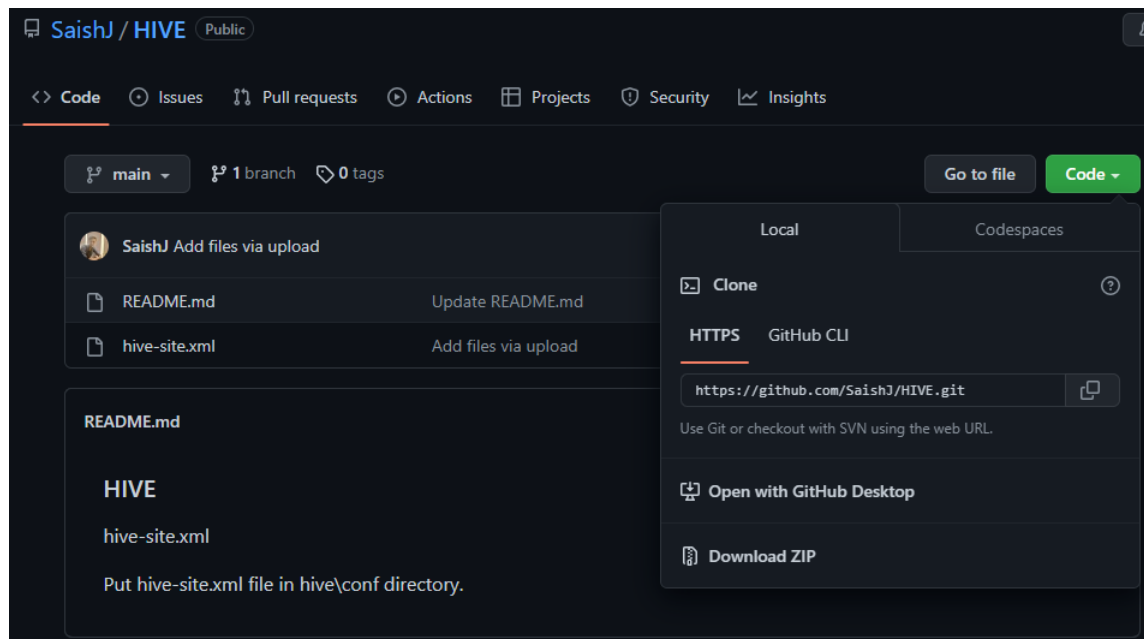
Name	Last modified	Size	Description
 Parent Directory		-	
 db-derby-10.12.1.1-bin.tar.gz	2015-10-10 14:38	18M	
 db-derby-10.12.1.1-bin.tar.gz.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-bin.tar.gz.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-bin.zip	2015-10-10 14:38	20M	
 db-derby-10.12.1.1-bin.zip.asc	2015-10-10 14:38	194	
 db-derby-10.12.1.1-bin.zip.md5	2015-10-10 14:38	33	
 db-derby-10.12.1.1-lib-debug.tar.gz	2015-10-10 14:38	14M	

After downloading Hive and Derby, Extract it in C:\hadoop\ directory.



Name	Date modified	Type	Size
derby	12/4/2022 10:30 PM	File folder	
hadoop	12/4/2022 9:56 PM	File folder	
hive	12/4/2022 10:29 PM	File folder	

Step 2: Download hive-site.xml file. link: <https://github.com/saishj/hive>



SaishJ / HIVE (Public)

<> Code Issues Pull requests Actions Projects Security Insights

main 1 branch 0 tags

Go to file Code

Local Codespaces

Clone

HTTPS GitHub CLI

<https://github.com/SaishJ/HIVE.git>

Use Git or checkout with SVN using the web URL.

Open with GitHub Desktop

Download ZIP

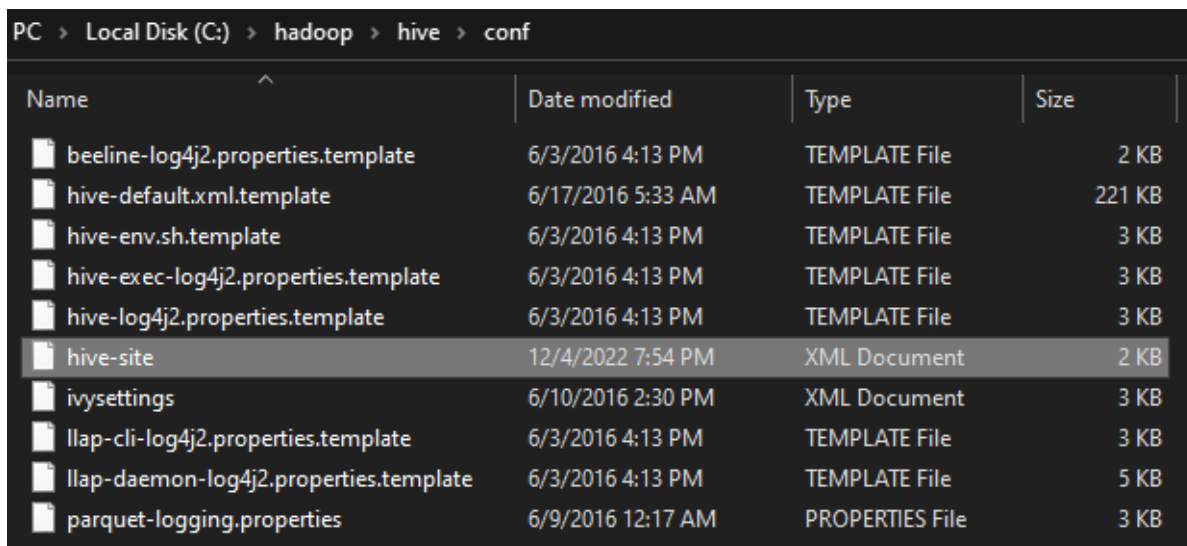
README.md

HIVE

hive-site.xml

Put hive-site.xml file in hive\conf directory.

And put it in the C:\hadoop\hive\conf directory.



Name	Date modified	Type	Size
beeline-log4j2.properties.template	6/3/2016 4:13 PM	TEMPLATE File	2 KB
hive-default.xml.template	6/17/2016 5:33 AM	TEMPLATE File	221 KB
hive-env.sh.template	6/3/2016 4:13 PM	TEMPLATE File	3 KB
hive-exec-log4j2.properties.template	6/3/2016 4:13 PM	TEMPLATE File	3 KB
hive-log4j2.properties.template	6/3/2016 4:13 PM	TEMPLATE File	3 KB
hive-site	12/4/2022 7:54 PM	XML Document	2 KB
ivysettings	6/10/2016 2:30 PM	XML Document	3 KB
llap-cli-log4j2.properties.template	6/3/2016 4:13 PM	TEMPLATE File	3 KB
llap-daemon-log4j2.properties.template	6/3/2016 4:13 PM	TEMPLATE File	5 KB
parquet-logging.properties	6/9/2016 12:17 AM	PROPERTIES File	3 KB

Step 3: Copy Derby libraries C:\hadoop\derby\lib

PC > Local Disk (C:) > hadoop > derby > lib

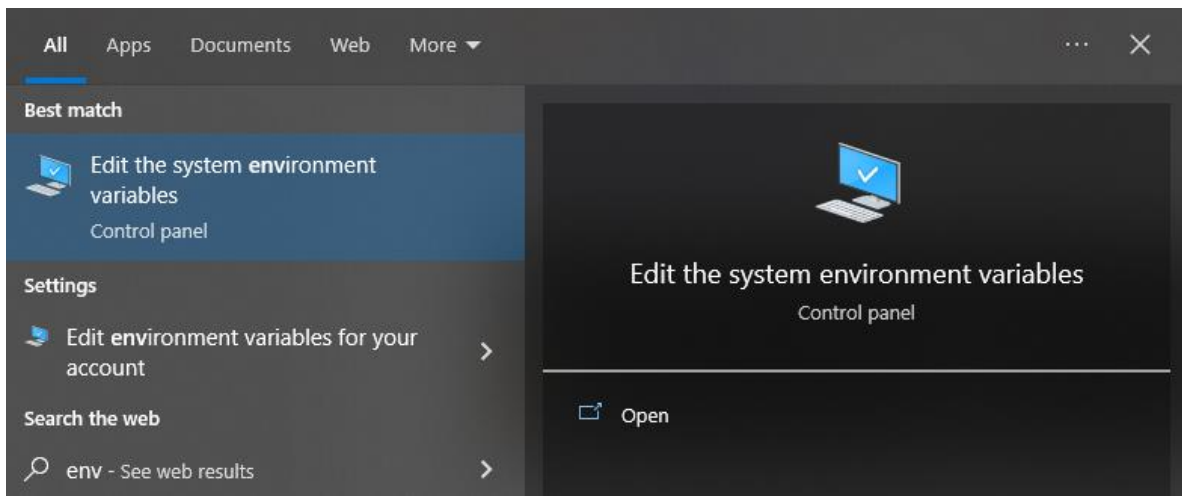
Name	Date modified	Type	Size
derby.jar	9/20/2015 7:25 PM	JAR File	3,150 KB
derby.war	9/20/2015 7:25 PM	WAR File	2 KB
derbyclient.jar	9/20/2015 7:25 PM	JAR File	577 KB
derbyLocale_cs.jar	9/20/2015 7:25 PM	JAR File	93 KB
derbyLocale_de_DE.jar	9/20/2015 7:25 PM	JAR File	110 KB
derbyLocale_es.jar	9/20/2015 7:25 PM	JAR File	104 KB
derbyLocale_fr.jar	9/20/2015 7:25 PM	JAR File	110 KB
derbyLocale_hu.jar	9/20/2015 7:25 PM	JAR File	94 KB
derbyLocale_it.jar	9/20/2015 7:25 PM	JAR File	104 KB
derbyLocale_ja_JP.jar	9/20/2015 7:25 PM	JAR File	121 KB
derbyLocale_ko_KR.jar	9/20/2015 7:25 PM	JAR File	115 KB
derbyLocale_pl.jar	9/20/2015 7:25 PM	JAR File	92 KB
derbyLocale_pt_BR.jar	9/20/2015 7:25 PM	JAR File	89 KB
derbyLocale_ru.jar	9/20/2015 7:25 PM	JAR File	119 KB
derbyLocale_zh_CN.jar	9/20/2015 7:25 PM	JAR File	107 KB
derbyLocale_zh_TW.jar	9/20/2015 7:25 PM	JAR File	109 KB
derbynet.jar	9/20/2015 7:25 PM	JAR File	267 KB
derbyoptionaltools.jar	9/20/2015 7:25 PM	JAR File	60 KB
derbyrun.jar	9/20/2015 7:25 PM	JAR File	10 KB
derbytools.jar	9/20/2015 7:25 PM	JAR File	225 KB

And paste in Hive libraries folder C:\hadoop\hive\lib

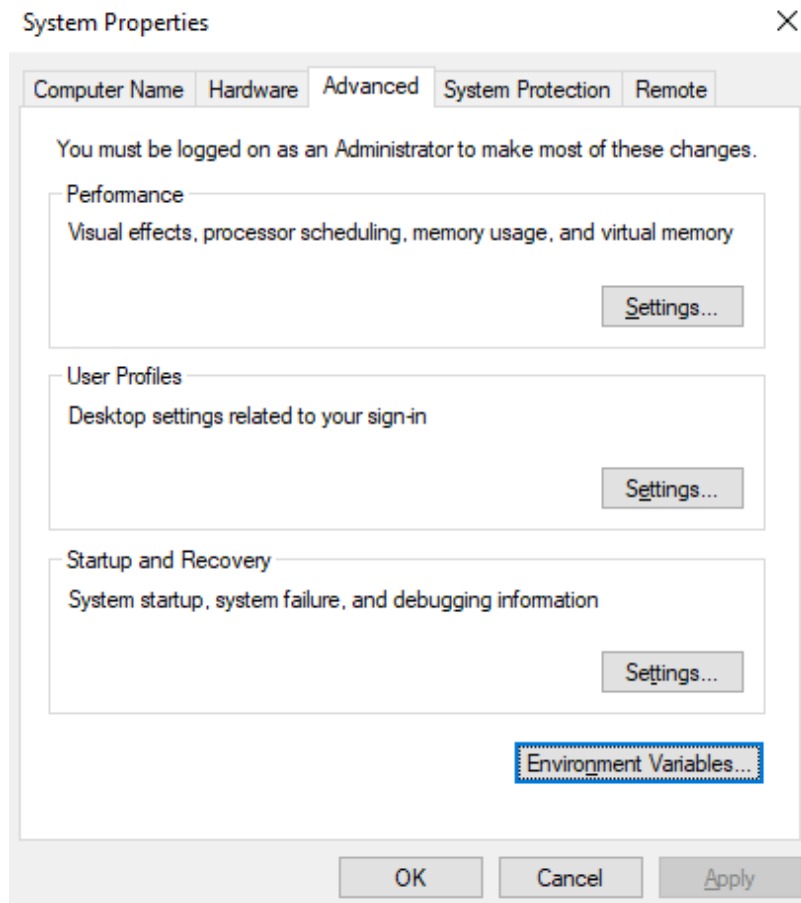
PC > Local Disk (C:) > hadoop > hive > lib

Name	Date modified	Type	Size
derby.jar	9/20/2015 7:25 PM	JAR File	3,150 KB
derby.war	9/20/2015 7:25 PM	WAR File	2 KB
derby-10.10.2.0.jar	2/25/2016 6:59 PM	JAR File	2,773 KB
derbyclient.jar	9/20/2015 7:25 PM	JAR File	577 KB
derbyLocale_cs.jar	9/20/2015 7:25 PM	JAR File	93 KB
derbyLocale_de_DE.jar	9/20/2015 7:25 PM	JAR File	110 KB
derbyLocale_es.jar	9/20/2015 7:25 PM	JAR File	104 KB
derbyLocale_fr.jar	9/20/2015 7:25 PM	JAR File	110 KB
derbyLocale_hu.jar	9/20/2015 7:25 PM	JAR File	94 KB
derbyLocale_it.jar	9/20/2015 7:25 PM	JAR File	104 KB
derbyLocale_ja_JP.jar	9/20/2015 7:25 PM	JAR File	121 KB
derbyLocale_ko_KR.jar	9/20/2015 7:25 PM	JAR File	115 KB
derbyLocale_pl.jar	9/20/2015 7:25 PM	JAR File	92 KB
derbyLocale_pt_BR.jar	9/20/2015 7:25 PM	JAR File	89 KB
derbyLocale_ru.jar	9/20/2015 7:25 PM	JAR File	119 KB
derbyLocale_zh_CN.jar	9/20/2015 7:25 PM	JAR File	107 KB
derbyLocale_zh_TW.jar	9/20/2015 7:25 PM	JAR File	109 KB
derbynet.jar	9/20/2015 7:25 PM	JAR File	267 KB
derbyoptionaltools.jar	9/20/2015 7:25 PM	JAR File	60 KB
derbyrun.jar	9/20/2015 7:25 PM	JAR File	10 KB
derbytools.jar	9/20/2015 7:25 PM	JAR File	225 KB
disruptor-3.3.0.jar	2/25/2016 7:08 PM	JAR File	78 KB

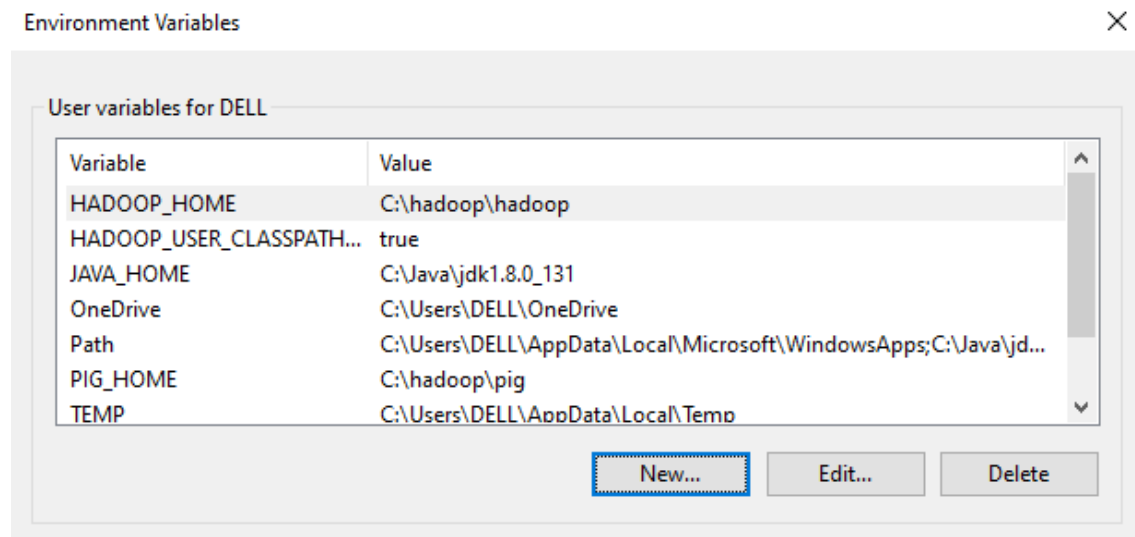
Step 4: Setting Environment Variables



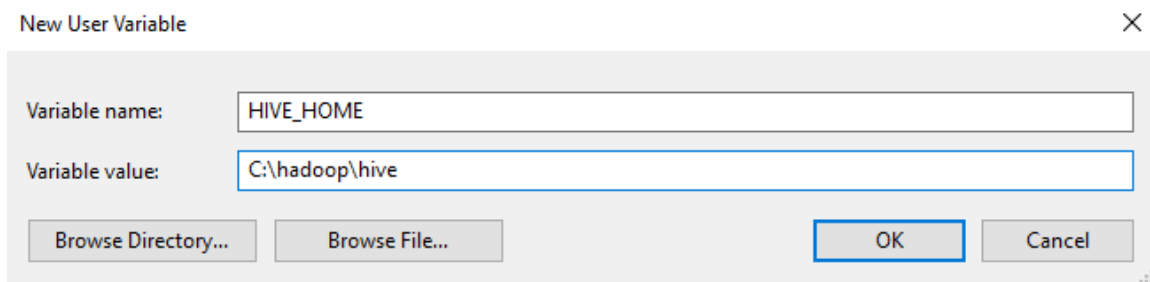
To edit the system environment variable, go to environment variable in system properties.



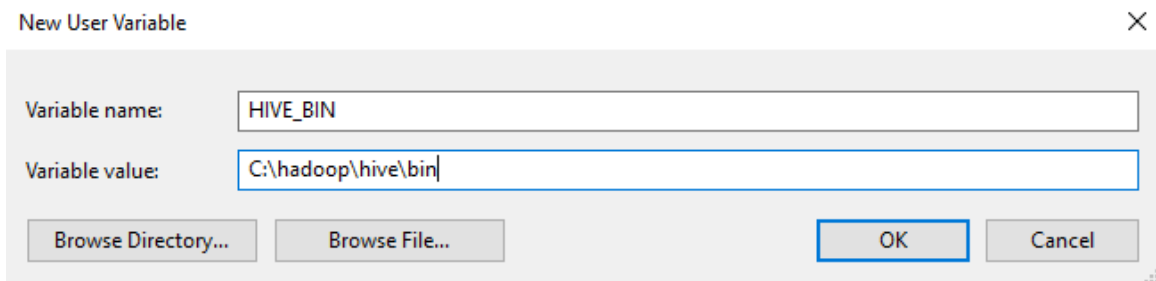
Add New User variable



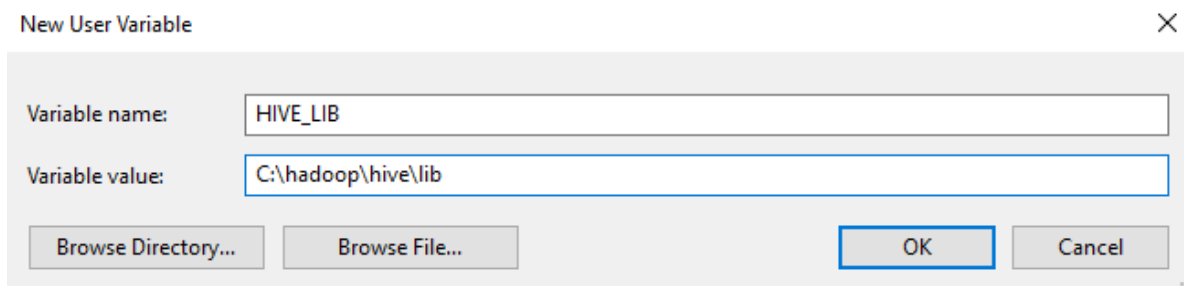
HIVE_HOME: C:\hadoop\hive



HIVE_BIN: C:\hadoop\hive\bin



HIVE_LIB: C:\hadoop\hive\lib



DERBY_HOME: C:\hadoop\derby

New User Variable ✕

Variable name:

Variable value:

HADOOP_USER_CLASSPATH_FIRST: true

Edit User Variable ✕

Variable name:

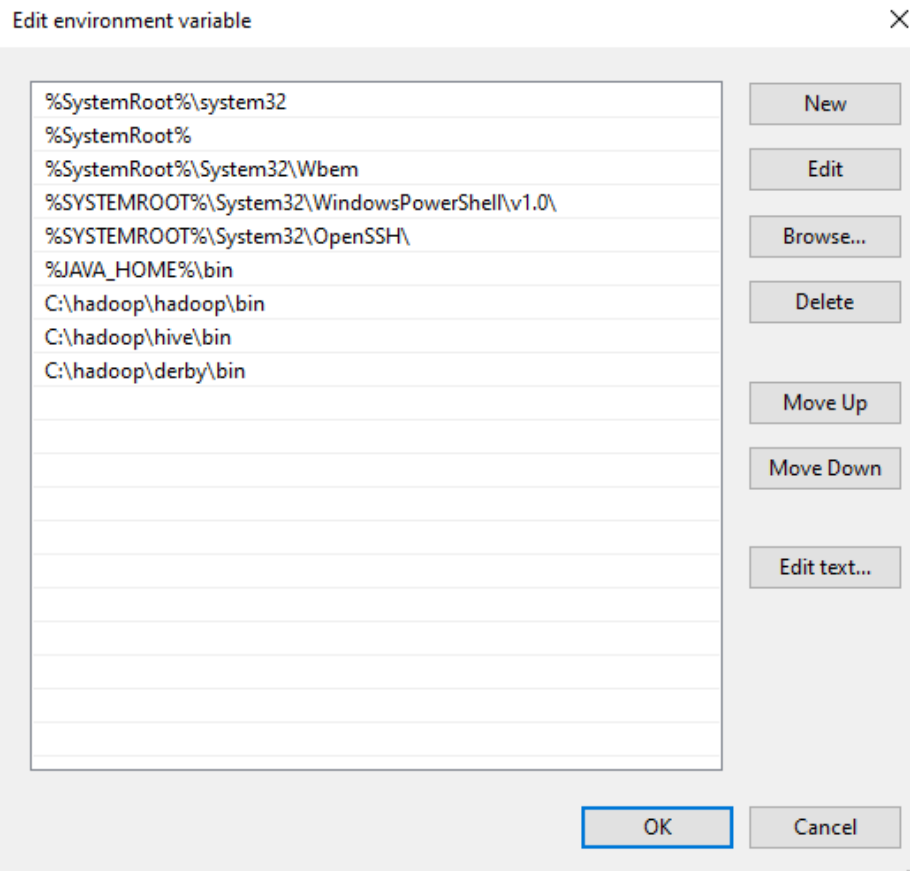
Variable value:

Edit Path in System Variable

System variables

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_131
NUMBER_OF_PROCESSORS	4
OS	Windows_NT
Path	C:\Windows\system32;C:\Windows;C:\Windows\System32\Wbem;...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64
PROCESSOR_IDENTIFIER	Intel64 Familly 6 Model 78 Stepping 3. GenuineIntel

Add New Path: C:\hadoop\hive\bin
C:\hadoop\derby\bin



Step 5: Start all Hadoop services

```
22/12/06 16:00:06 INFO datanode.DataNode: Namenode Block pool BP-1047024942-127.0.0.1-1670322569421 (Datanode Uuid 5ca1091b-5b5b-4046-bdd1-1a48c21f7e50) service to Dell11/127.0.0.1:9000 trying to claim ACTIVE state with txid=1
22/12/06 16:00:06 INFO datanode.DataNode: Acknowledging ACTIVE Namenode Block pool BP-1047024942-127.0.0.1-1670322569421 (Datanode Uuid 5ca1091b-5b5b-4046-bdd1-1a48c21f7e50) service to Dell11/127.0.0.1:9000
22/12/06 16:00:06 INFO Administrator: Command Prompt
22/12/06 16:00:06 INFO C:\hadoop\hadoop\sbin>start-dfs.cmd
22/12/06 16:00:06 INFO C:\hadoop\hadoop\sbin>start-yarn.cmd
22/12/06 16:00:06 INFO starting yarn daemons
22/12/06 16:00:06 INFO C:\hadoop\hadoop\sbin>jps
22/12/06 16:00:06 INFO 8060 ResourceManager
22/12/06 16:00:06 INFO 11652 Jps
22/12/06 16:00:06 INFO 5724 NameNode
22/12/06 16:00:06 INFO 1-1670322569421 to block-9532 DataNode
C:\hadoop\hadoop\sbin>
22/12/06 16:00:32 INFO /yarn/hadoop-yarn-common.jar
22/12/06 16:00:33 INFO p/Jetty 0.0.0.0:8042
22/12/06 16:00:33 INFO
22/12/06 16:00:33 INFO
22/12/06 16:00:33 INFO
22/12/06 16:00:33 INFO
22/12/06 16:00:33 INFO
22/12/06 16:00:33 INFO finished containers :|
22/12/06 16:00:33 INFO on container-tokens, g
22/12/06 16:00:33 INFO -tokens, got key with id : -684787314
22/12/06 16:00:33 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceM
anager as Dell:62801 with total resource of <memory:8192, vCores:8>
22/12/06 16:00:33 INFO nodemanager.NodeStatusUpdaterImpl: Notifying ContainerManage
r to unblock new container-requests
22/12/06 16:00:05 INFO blockmanagement.CacheReplicationMonitor: Rescanning because
of pending operations
22/12/06 16:00:05 INFO blockmanagement.CacheReplicationMonitor: Scanned 0 directive
(s) and 0 block(s) in 2 millisecond(s).
22/12/06 16:00:06 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegi
18c21f7e50, infoPort=580
4de3-92d7-ad98801957ed;
adding new storage ID DS
processReport: Received
or becoming active. Its
rt: from DatanodeRegistr
21f7e50, infoPort=58075,
s3-92d7-ad98801957ed;nsi
Dell:62801 Node Transi
tioned from NEW to
RUNNING
22/12/06 16:00:33 INFO capacity.CapacityScheduler: Added node Dell:62801 clusterRes
ource: <memory:8192, vCor
```

Step 6: Start the Derby network server on the localhost using the following command:

Command: StartNetworkServer -h 0.0.0.0

```
Administrator: Command Prompt - StartNetworkServer -h 0.0.0.0
C:\hadoop\derby\bin>StartNetworkServer -h 0.0.0.0
Tue Dec 06 16:09:28 IST 2022 : Security manager installed using the Basic server security policy.
Tue Dec 06 16:09:28 IST 2022 : Apache Derby Network Server - 10.12.1.1 - (1704137) started and ready to accept connections on port 1527
```

Step 7: Starting Apache Hive

Open Command Prompt and go to the HIVE binaries directory. (C:\hadoop\hive\bin) and run the following command:

Command: hive

```
Administrator: Command Prompt - hive
C:\hadoop\hive\bin>hive
ERROR StatusLogger No log4j2 configuration file found. Using default configuration: logging only errors to the console.
Connecting to jdbc:hive2://
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop/hive/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connected to: Apache Hive (version 2.1.0)
Driver: Hive JDBC (version 2.1.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 2.1.0 by Apache Hive
hive> _
```

Step 8: Create Database

```
hive> CREATE DATABASE BigData;
OK
No rows affected (0.062 seconds)
hive>
```

Step 9: Show Databases

```
hive> SHOW DATABASES;
OK
a
bda
bigdata
default
4 rows selected (0.391 seconds)
hive> _
```

Step 10: Create Table and Insert 5 Records.

```
hive> CREATE TABLE students (id INT, name STRING, city STRING)
. . > ROW FORMAT DELIMITED
. . > FIELDS TERMINATED BY '|'
. . > STORED AS TEXTFILE;
OK
No rows affected (0.818 seconds)
hive>
```


Create a text file and add 5 records in it.



Load the above text file in students table.

```
hive> LOAD DATA LOCAL INPATH 'd:/students.txt' INTO TABLE students;
Loading data to table default.students
OK
No rows affected (1.071 seconds)
hive>
```

See the table.









```
hive> SELECT * FROM students;
OK
1 Saish Mumbai
2 Pratik Pune
3 Prafulla Kolhapur
4 Siddhi Goa
5 Mayuresh Thane
5 rows selected (1.441 seconds)
hive> _
```

Practical No. 4

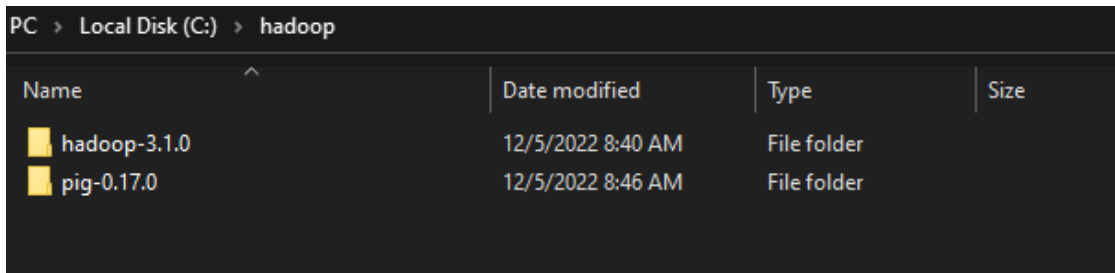
Aim: Install Pig and Load and Dump the content.

Step 1: Download the Apache Pig link: <https://downloads.apache.org/pig/latest/pig-0.17.0.tar.gz>

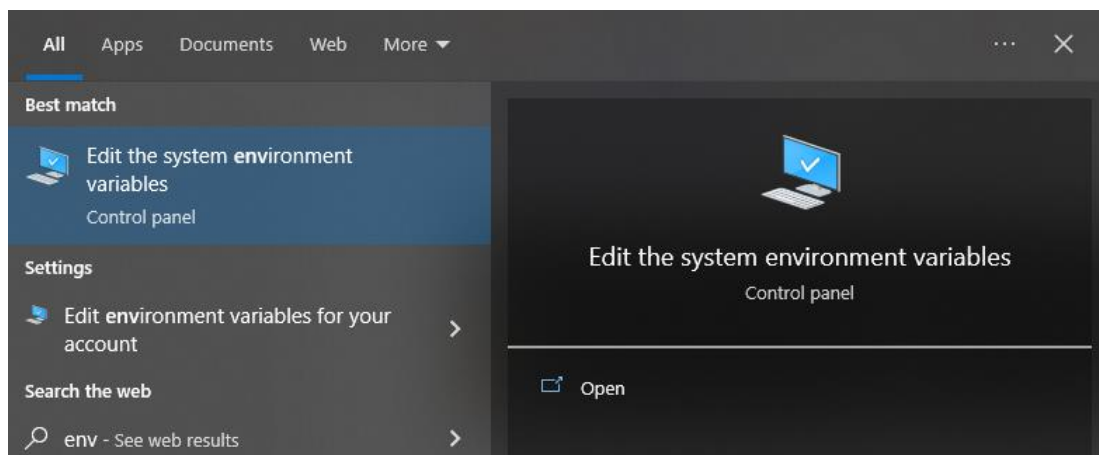
Index of /pig/latest

Name	Last modified	Size	Description
 Parent Directory		-	
 RELEASE_NOTES.txt	2017-06-16 18:10	1.9K	
 pig-0.17.0-src.tar.gz	2017-06-16 18:11	15M	
 pig-0.17.0-src.tar.gz.asc	2017-06-16 18:11	488	
 pig-0.17.0-src.tar.gz.md5	2017-06-16 18:11	56	
 pig-0.17.0.tar.gz	2017-06-16 18:10	220M	
 pig-0.17.0.tar.gz.asc	2017-06-16 18:11	488	
 pig-0.17.0.tar.gz.md5	2017-06-16 18:11	52	

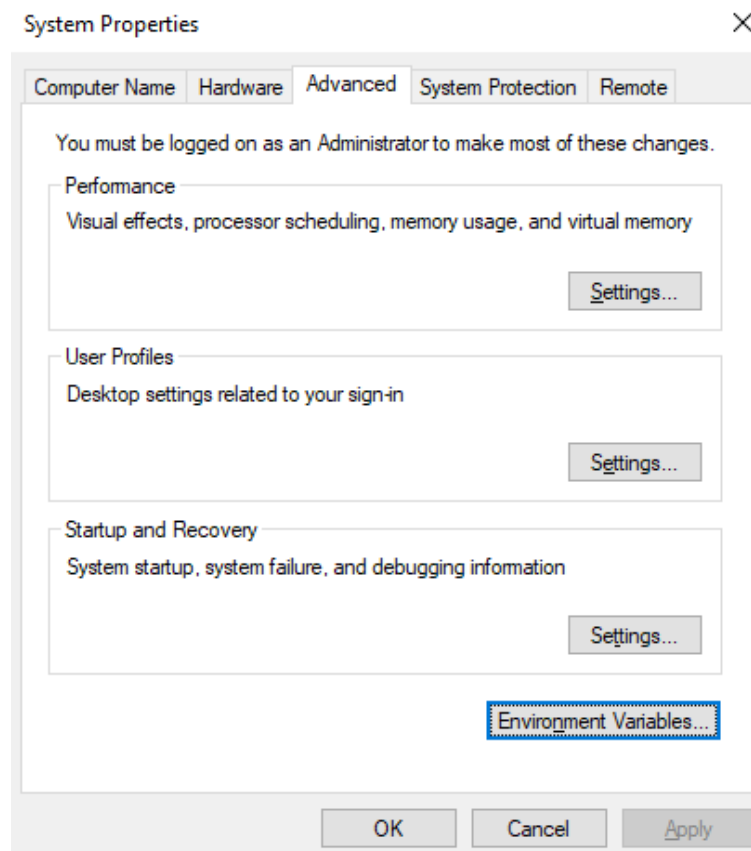
After the file is downloaded, Extract it in C:\hadoop\ directory.



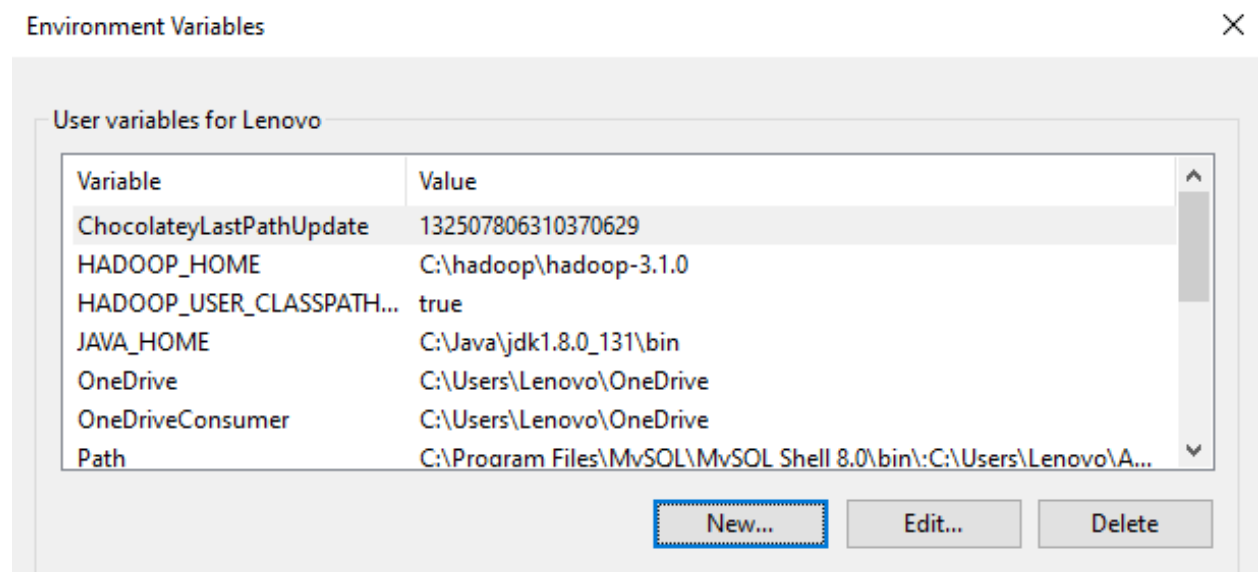
Step 2: Setting Environment Variables



To edit the system environment variable, go to environment variable in system properties.



Add New user variable



PIG_HOME: C:\hadoop\pig-0.17.0

Edit User Variable ✕

Variable name:

Variable value:

Edit the Path user variable to add the following paths

Edit environment variable ✕

%SystemRoot%\system32	<input type="button" value="New"/> <input type="button" value="Edit"/> <input type="button" value="Browse..."/> <input type="button" value="Delete"/> <input type="button" value="Move Up"/> <input type="button" value="Move Down"/> <input type="button" value="Edit text..."/>
%SystemRoot%	
%SystemRoot%\System32\Wbem	
%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\	
%SYSTEMROOT%\System32\OpenSSH\	
%JAVA_HOME%\bin	
C:\apache-ant-1.9.6\bin	
%ANDROID_HOME%\tools;%ANDROID_HOME%\platform-tools	
C:\Program Files\Git\cmd	
C:\Users\Lenovo\AppData\Local\Programs\Python\Python38\Scripts\	
C:\Users\Lenovo\AppData\Local\Programs\Python\Python38\	
C:\Program Files\cygwin64\bin	
%GRADLE_HOME%\bin	
C:\Program Files\nodejs\	
C:\Program Files\MongoDB\Server\6.0\bin	
C:\hadoop\pig-0.17.0\bin	




Start all Hadoop services

Open a Command Prompt as Administrator and execute the following command

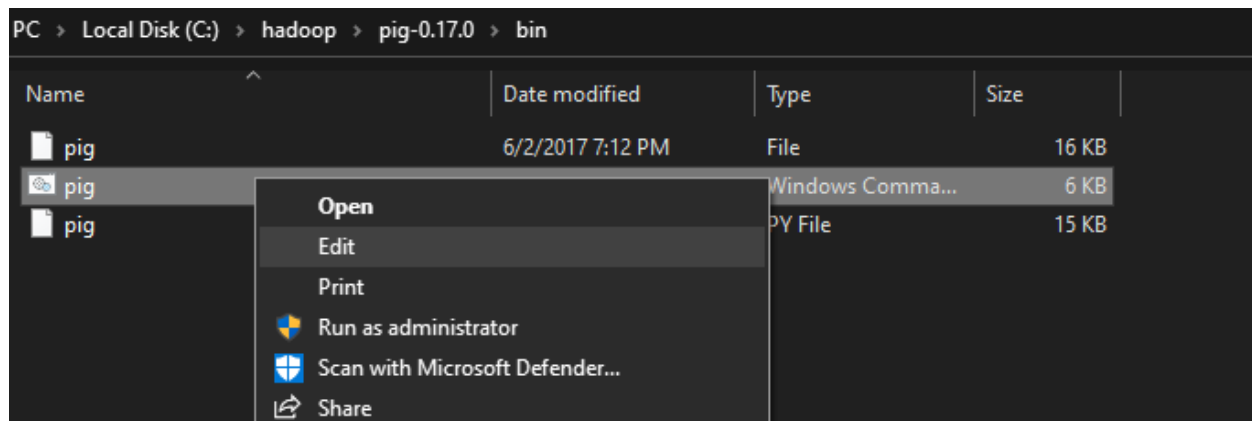
```
pig -version
```

Fix this error. Go to C:\hadoop\pig-0.17.0\bin and edit pig.cmd

PC > Local Disk (C:) > hadoop > pig-0.17.0 > bin

Name	Date modified	Type	Size
 pig	6/2/2017 7:12 PM	File	16 KB
 pig	12/6/2022 1:33 AM	Windows Comma...	6 KB
 pig	6/2/2017 7:12 PM	PY File	15 KB

Right click on pig.cmd and Click on Edit



Changing the HADOOP_BIN_PATH value from “%HADOOP_HOME%\bin”

```
pig - Notepad
File Edit Format View Help
::
:: PIG_OPTS Extra Java runtime options.
::
:: PIG_CONF_DIR Alternate conf dir. Default is ${PIG_HOME}/cc
::
:: HBASE_CONF_DIR - Optionally, the HBase configuration to run a
:: when using HBaseStorage
::

setlocal enabledelayedexpansion

set HADOOP_BIN_PATH=%HADOOP_HOME%\bin
```

To “%HADOOP_HOME%\libexec”

```
pig - Notepad
File Edit Format View Help
::
:: PIG_OPTS Extra Java runtime options.
::
:: PIG_CONF_DIR Alternate conf dir. Default is ${PIG_HOME}/cc
::
:: HBASE_CONF_DIR - Optionally, the HBase configuration to run a
:: when using HBaseStorage
::

setlocal enabledelayedexpansion

set HADOOP_BIN_PATH=%HADOOP_HOME%\libexec
```

Run pig -version

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.19044.2075]
(c) Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>pig -version
Apache Pig version 0.17.0 (r1797386)
compiled Jun 02 2017, 15:41:58

C:\WINDOWS\system32>_
```

The simplest way to write PigLatin statements is using Grunt shell which is an interactive tool where we write a statement and get the desired output. There are two modes to involve Grunt Shell:


1. Local: All scripts are executed on a single machine without requiring Hadoop. (command: pig -x local)
2. MapReduce: Scripts are executed on a Hadoop cluster (command: pig -x MapReduce)

```
Administrator: Command Prompt - pig -x local
C:\WINDOWS\system32>pig -x local
2022-12-06 01:48:47,596 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2022-12-06 01:48:47,597 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2022-12-06 01:48:48,050 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2022-12-06 01:48:48,050 [main] INFO org.apache.pig.Main - Logging error messages to: C:\hadoop\hadoop-3.1.0\logs\pig_1670271528048.log
2022-12-06 01:48:48,238 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file C:\Users\Lenovo/.pigbootup not found
2022-12-06 01:48:48,737 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-12-06 01:48:48,740 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2022-12-06 01:48:49,264 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-12-06 01:48:49,295 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-f77d0853-b648-46d5-be24-1fc7684ea96e
2022-12-06 01:48:49,295 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>
```

Load the file and Dump the content from D drive.

Create a new text file and write some content.

PC > Local Disk (D:) > Hadoop

Name	Date modified	Type	Size
 Prac4	11/30/2022 5:29 PM	Text Document	1 KB

Load the file using the following command:

text = LOAD '/d:/Hadoop/Prac4.txt/' AS (line:chararray);

```
Administrator: Command Prompt - pig -x local

grunt> text = LOAD '/d:/Hadoop/Prac4.txt/' AS (line:chararray);
2022-12-06 01:59:18,859 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
grunt>
```

See the content using the following command:

```
Administrator: Command Prompt - pig -x local

grunt> DUMP text;
2022-12-06 02:01:12,813 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-12-06 02:01:12,827 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-12-06 02:01:12,827 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-12-06 02:01:12,828 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2022-12-06 02:01:12,829 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-12-06 02:01:12,830 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-12-06 02:01:12,830 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-12-06 02:01:12,839 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-12-06 02:01:12,840 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-12-06 02:01:12,842 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-12-06 02:01:12,843 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-12-06 02:01:12,845 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-12-06 02:01:12,846 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-12-06 02:01:12,846 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-12-06 02:01:12,846 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Distributed cache not supported or needed in local mode. Setting key [pig.schematuple.local.dir] with code temp directory: C:\Users\Lenovo\AppData\Local\Temp\1670272272845-0
2022-12-06 02:01:12,860 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-12-06 02:01:12,864 [JobControl] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-12-06 02:01:13,142 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2022-12-06 02:01:13,152 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat
2022-12-06 02:01:13,157 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-12-06 02:01:13,161 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-12-06 02:01:13,165 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2022-12-06 02:01:13,278 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-12-06 02:01:13,351 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_local116663689_0002
2022-12-06 02:01:13,357 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Executing with tokens: []
2022-12-06 02:01:14,601 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8080/
2022-12-06 02:01:14,603 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_local116663689_0002
2022-12-06 02:01:14,603 [Thread-10] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - OutputCommitter set in config null
2022-12-06 02:01:14,605 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases text
2022-12-06 02:01:14,623 [Thread-10] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-12-06 02:01:14,624 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: text[2,7],text[-1,-1] C: R
2022-12-06 02:01:14,627 [Thread-10] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime
job_local116663689_0002 1 0 n/a n/a n/a n/a 0 0 0 0 text MAP_ONLY file:/tmp/temp1847705514/tmp2101070529,

Input(s):
Successfully read 4 records from: "/d:/Hadoop/Prac4.txt"

Output(s):
Successfully stored 4 records in: "file:/tmp/temp1847705514/tmp2101070529"

Counters:
Total records written : 4
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job_local116663689_0002

2022-12-06 02:01:15,492 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-12-06 02:01:15,495 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-12-06 02:01:15,498 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-12-06 02:01:15,505 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-12-06 02:01:15,506 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-12-06 02:01:15,508 [main] INFO org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-12-06 02:01:15,515 [main] WARN org.apache.hadoop.io.nativeio.NativeIO - NativeIO.getStat error (3): The system cannot find the path specified.
-- file path: tmp/temp1847705514/tmp2101070529/part-m-000000
2022-12-06 02:01:15,591 [main] WARN org.apache.hadoop.io.nativeio.NativeIO - NativeIO.getStat error (3): The system cannot find the path specified.
-- file path: tmp/temp1847705514/tmp2101070529/SUCCESS
2022-12-06 02:01:15,668 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-12-06 02:01:15,669 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Whenever private cloud resources are unable to meet users' quality-of-service requirements,)
(hybrid computing systems, partially composed of public cloud resources and privately owned infrastructures, are created to serve the organization's needs. These are often referred as hybrid clouds,)
(which are becoming a common way for many stakeholders to start exploring the possibilities)
(offered by cloud computing.)
grunt>
```


Practical No. 5

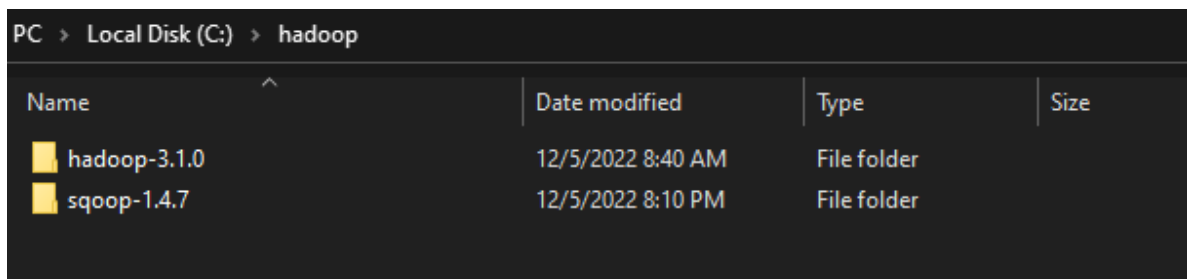
Aim: Install Sqoop and list all databases

Step 1: Download Sqoop link: https://archive.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz

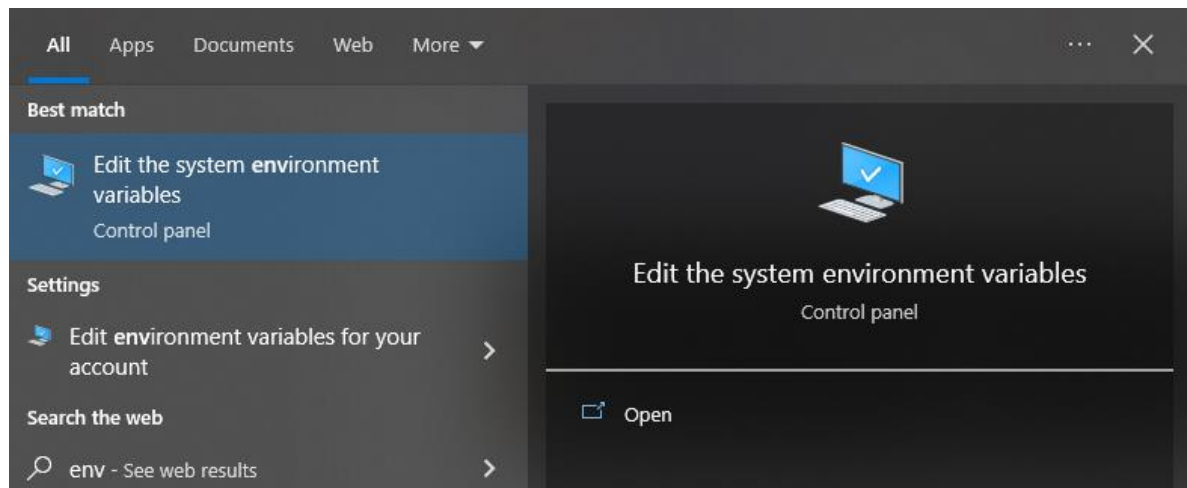
Index of /dist/sqoop/1.4.7

Name	Last modified	Size	Description
 Parent Directory		-	
 sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz	2020-07-06 15:19	17M	
 sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz.asc	2020-07-06 15:20	819	
 sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz.md5	2020-07-06 15:19	71	
 sqoop-1.4.7.tar.gz	2020-07-06 15:20	1.1M	
 sqoop-1.4.7.tar.gz.asc	2020-07-06 15:19	819	
 sqoop-1.4.7.tar.gz.md5	2020-07-06 15:20	53	

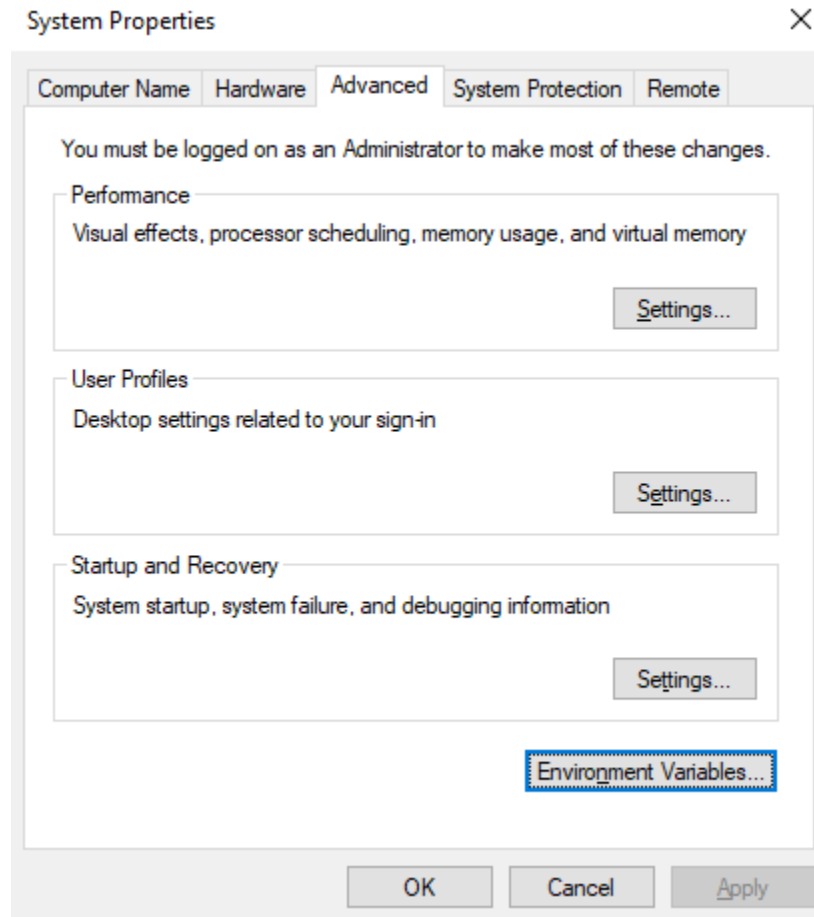
After the file is Downloaded. Extract it in C:\hadoop\ directory



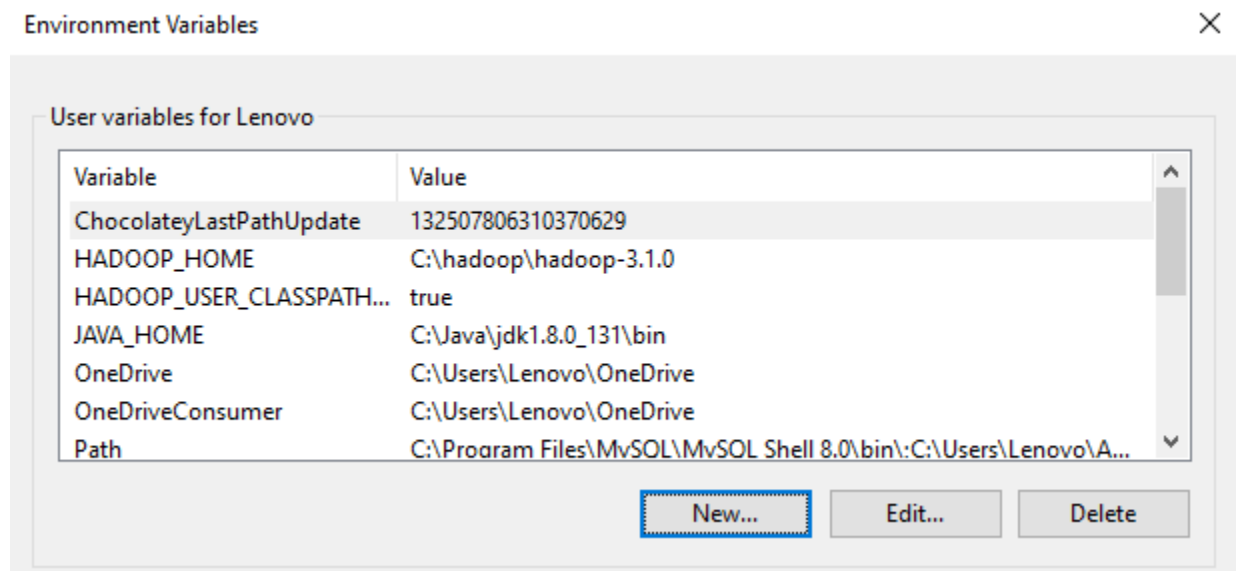
Step 2: Setting Up Environment Variables



To edit the System Environment variable, go to Environment Variable in system properties.



Add New User Variable



SQOOP_HOME: C:\hadoop\sqoop-1.4.7

Edit User Variable

Variable name:

Variable value:

Edit Path in System Variable

System variables

Variable	Value
JUNIT_HOME	C:\JUNIT
NUMBER_OF_PROCESSORS	4
OS	Windows_NT
Path	C:\WINDOWS\system32;C:\WINDOWS;C:\WINDOWS\System32\Wb...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64
PROCESSOR_IDENTIFIER	Intel64 Family 6 Model 142 Stepping 10. GenuineIntel

Add New Path: C:\hadoop\sqoop-1.4.7\bin

Edit environment variable

%SystemRoot%\system32
%SystemRoot%
%SystemRoot%\System32\Wbem
%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\
%SYSTEMROOT%\System32\OpenSSH\
%JAVA_HOME%\bin
C:\apache-ant-1.9.6\bin
%ANDROID_HOME%\tools;%ANDROID_HOME%\platform-tools
C:\Program Files\Git\cmd
C:\Users\Lenovo\AppData\Local\Programs\Python\Python38\Scripts\
C:\Users\Lenovo\AppData\Local\Programs\Python\Python38\
C:\Program Files\cygwin64\bin
%GRADLE_HOME%\bin
C:\Program Files\nodejs\
C:\Program Files\MongoDB\Server\6.0\bin
C:\hadoop\sqoop-1.4.7\bin

Step 3: Verify the Path

Open CMD and Run following Command

```
echo %SQOOP_HOME%
```

```
Command Prompt
Microsoft Windows [Version 10.0.19044.2075]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>echo %SQOOP_HOME%
C:\hadoop\sqoop-1.4.7

C:\Users\Lenovo>
```

Step 4: Configure SQOOP

A. Installing MySQL Database. <https://dev.mysql.com/downloads/installer/>

MySQL Community Downloads

MySQL Installer

General Availability (GA) Releases Archives ⓘ

MySQL Installer 8.0.31

Select Operating System:
Microsoft Windows ▼

[Looking for previous GA versions?](#)

Windows (x86, 32-bit), MSI Installer (mysql-installer-web-community-8.0.31.0.msi)	8.0.31	5.5M	Download
MD5: 7a83203e24f873b49fa2df2f1a58eca6 Signature			
Windows (x86, 32-bit), MSI Installer (mysql-installer-community-8.0.31.0.msi)	8.0.31	431.7M	Download
MD5: ef57176fcb01f01f4e87dbba9b87ac6f Signature			

We suggest that you use the MD5 checksums and GnuPG signatures to verify the integrity of the packages you download.

B. Download mysql-connector-java.jar and put it in the lib folder of SQOOP.

Home » mysql » mysql-connector-java » 8.0.16



MySQL Connector Java » 8.0.16

MySQL Connector/J is a JDBC Type 4 driver, which means that it is pure Java implementation of the MySQL protocol and does not rely on the MySQL client libraries. This driver supports auto-registration with the Driver Manager, standardized validity checks, categorized SQLExceptions, support for large update counts, support for local and offset date-time variants from the java.time package, support for JDBC-4.x XML processing, support for per connection client information and support for the NCHAR, NVARCHAR ...

License	GPL 2.0
Categories	JDBC Drivers
Tags	database sql jdbc driver connector mysql
Organization	Oracle Corporation
HomePage	http://dev.mysql.com/doc/connector-j/en/
Date	Apr 25, 2019
Files	pom (1 KB) jar (2.2 MB) View All
Repositories	Central
Ranking	#68 in MvnRepository (See Top Artifacts) #1 in JDBC Drivers

PC > Local Disk (C:) > hadoop > sqoop-1.4.7 > lib

Name	Date modified	Type	Size
commons-io-1.4.jar	12/19/2017 4:30 AM	JAR File	107 KB
commons-jexl-2.1.1.jar	12/19/2017 4:30 AM	JAR File	262 KB
commons-lang3-3.4.jar	12/19/2017 4:30 AM	JAR File	425 KB
commons-logging-1.1.1.jar	12/19/2017 4:30 AM	JAR File	60 KB
hsqldb-1.8.0.10.jar	12/19/2017 4:30 AM	JAR File	691 KB
jackson-annotations-2.3.1.jar	12/19/2017 4:30 AM	JAR File	36 KB
jackson-core-2.3.1.jar	12/19/2017 4:30 AM	JAR File	194 KB
jackson-core-asl-1.9.13.jar	12/19/2017 4:30 AM	JAR File	227 KB
jackson-databind-2.3.1.jar	12/19/2017 4:30 AM	JAR File	893 KB
jackson-mapper-asl-1.9.13.jar	12/19/2017 4:30 AM	JAR File	763 KB
kite-data-core-1.1.0.jar	12/19/2017 4:30 AM	JAR File	2,128 KB
kite-data-hive-1.1.0.jar	12/19/2017 4:30 AM	JAR File	1,760 KB
kite-data-mapreduce-1.1.0.jar	12/19/2017 4:30 AM	JAR File	1,727 KB
kite-hadoop-compatibility-1.1.0.jar	12/19/2017 4:30 AM	JAR File	1,725 KB
mysql-connector-java-8.0.16.jar	12/5/2022 8:25 PM	JAR File	2,240 KB
opencsv-2.3.jar	12/19/2017 4:30 AM	JAR File	20 KB
paranamer-2.7.jar	12/19/2017 4:30 AM	JAR File	34 KB

C. Creating Users in MySQL

To configuring SGOOP is to create users for MySQL. These Users are used for connecting SGOOP to MySQL Database for reading and writing data from it.

1. Open the MySQL Workbench and open the workspace.

Welcome to MySQL Workbench

MySQL Workbench is the official graphical user interface (GUI) tool for MySQL. It allows you to design, create and browse your database schemas, work with database objects and insert data as well as design and run SQL queries to work with stored data. You can also migrate schemas and data from other database vendors to your MySQL database.

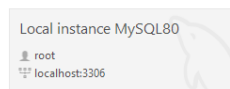
[Browse Documentation >](#)

[Read the Blog >](#)

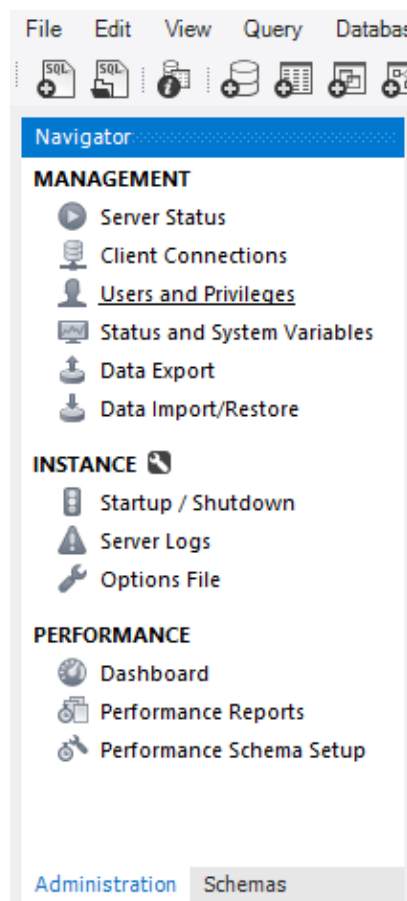
[Discuss on the Forums >](#)

MySQL Connections

9 |



2. Now Open the Administration option in the Workspace and select Users and privileges option under Management.



3. Creating SQOOP User in MySQL

Now select Add Account option and Create an new user with Login Name as sqoop and Limit to Host Mapping as the localhost and Password of your choice.

Local instance MySQL80
Users and Privileges

User Accounts

User	From Host
hive	localhost
mysql.infoschema	localhost
mysql.session	localhost
mysql.sys	localhost
root	localhost
sqoop	localhost

Details for account sqoop@localhost

Login Account Limits Administrative Roles Schema Privileges

Login Name: sqoop You may create multiple accounts with the same name to connect from different hosts.

Authentication Type: Standard For the standard password and/or host based authentication, select 'Standard'.

Limit to Hosts Matching: localhost % and _ wildcards may be used

Password: ***** Type a password to reset it.

Consider using a password with 8 or more characters with mixed case letters, numbers and punctuation marks.

Confirm Password: ***** Enter password again to confirm.

Expire Password

Add Account Delete Refresh Revert Apply

4. Define the roles for this user under Administrative Roles and select DBManager, DBDesigner and BackupAdmin Roles.

Local instance MySQL80
Users and Privileges

User Accounts

User	From Host
hive	localhost
mysql.infoschema	localhost
mysql.session	localhost
mysql.sys	localhost
root	localhost
sqoop	localhost

Details for account sqoop@localhost

Login Account Limits Administrative Roles Schema Privileges

Role	Description
<input type="checkbox"/> DBA	grants the rights to perform all tasks
<input type="checkbox"/> MaintenanceAdmin	grants rights needed to maintain server
<input type="checkbox"/> ProcessAdmin	rights needed to assess, monitor, and kill a process
<input type="checkbox"/> UserAdmin	grants rights to create users logins and reset passwords
<input type="checkbox"/> SecurityAdmin	rights to manage logins and grant and revoke privileges
<input type="checkbox"/> MonitorAdmin	minimum set of rights needed to monitor server
<input checked="" type="checkbox"/> DBManager	grants full rights on all databases
<input checked="" type="checkbox"/> DBDesigner	rights to create and reverse engineer any database
<input type="checkbox"/> ReplicationAdmin	rights needed to setup and manage replication
<input checked="" type="checkbox"/> BackupAdmin	minimal rights needed to backup any database

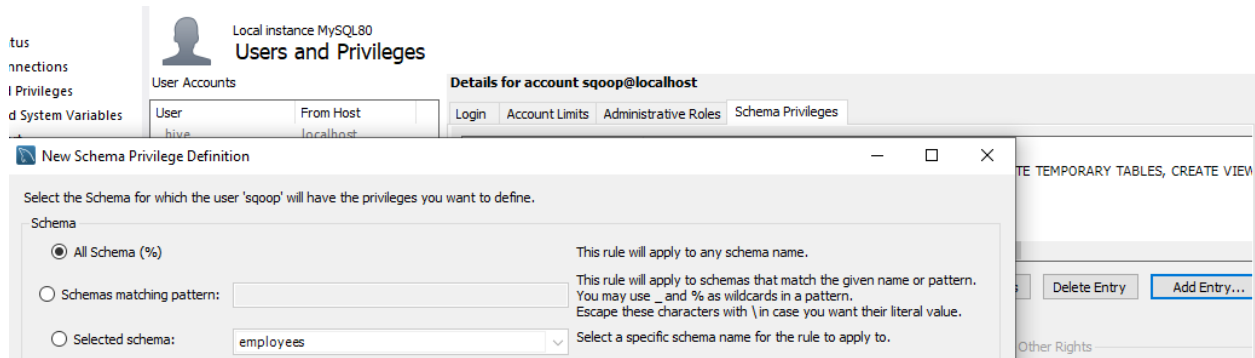
Global Privileges

- ☒ ALTER
- ☒ ALTER ROUTINE
- ☒ CREATE
- ☒ CREATE ROUTINE
- ☐ CREATE TABLESPACE
- ☐ CREATE TEMPORARY TABLES
- ☐ CREATE USER
- ☒ CREATE VIEW
- ☒ DELETE
- ☒ DROP
- ☒ EVENT
- ☐ EXECUTE
- ☐ FILE
- ☒ GRANT OPTION
- ☒ INDEX
- ☒ INSERT
- ☒ LOCK TABLES
- ☐ PROCESS
- ☐ REFERENCES
- ☐ RELOAD
- ☐ REPLICATION CLIENT

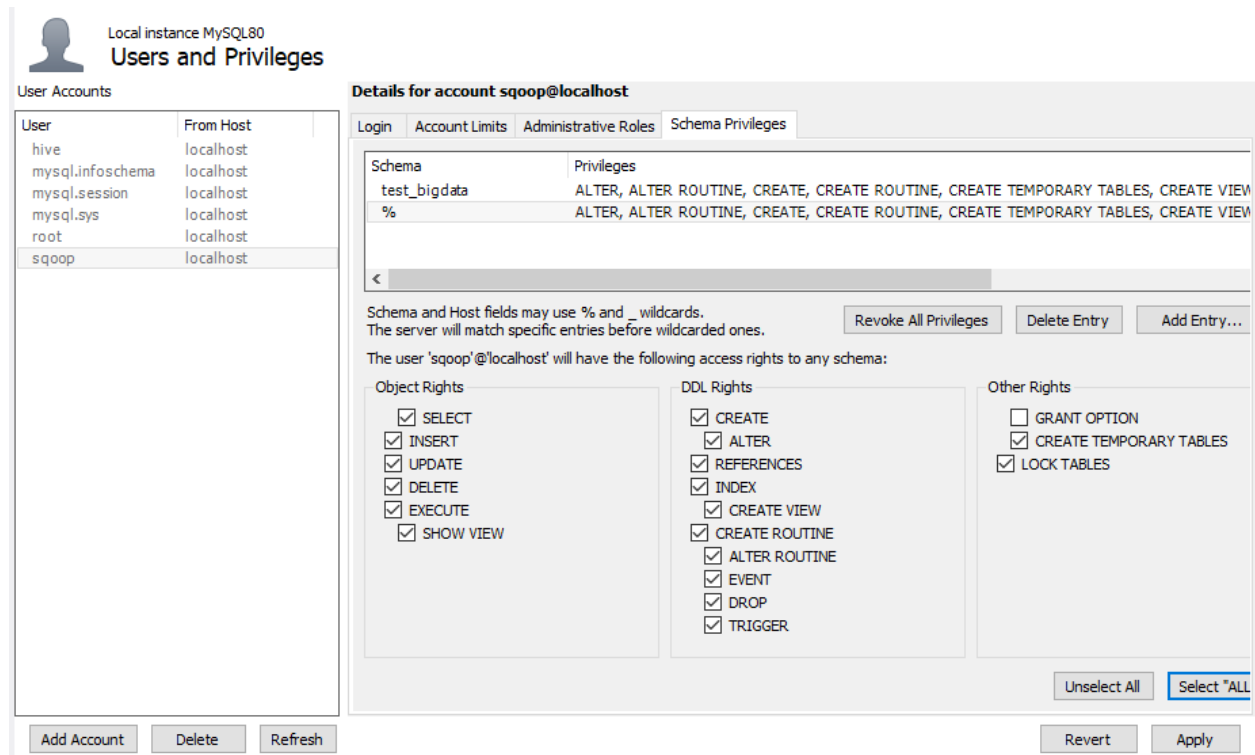
Revoke All Privileges

Add Account Delete Refresh Revert Apply

- Now grant schema privileges for the user by using Add Entry option and selecting the schemas.



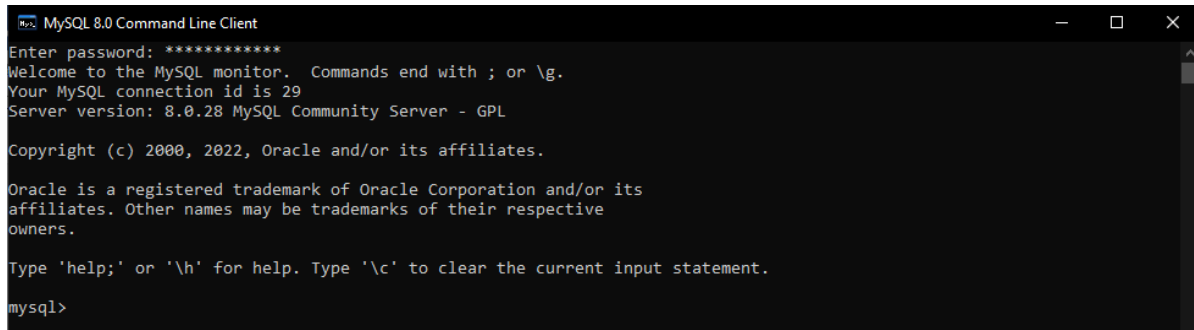
- After clicking OK to select All the privileges for this schema and Click Apply.



D. Granting permission to Users.

Once the user is created. Grant all privileges to the user for all the Tables.

1. Open the MySQL CMD window.



```
MySQL 8.0 Command Line Client
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 29
Server version: 8.0.28 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

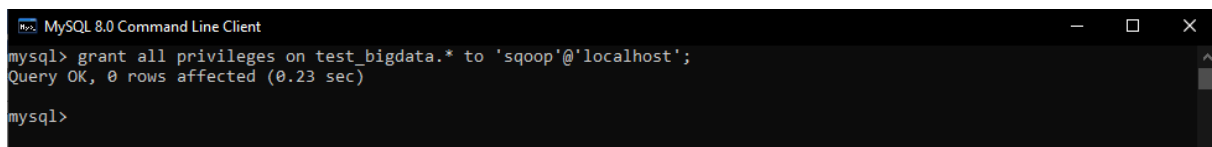
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
```

2. Run the following command

grant all privileges on test_bigdata.* to 'sqoop'@'localhost';



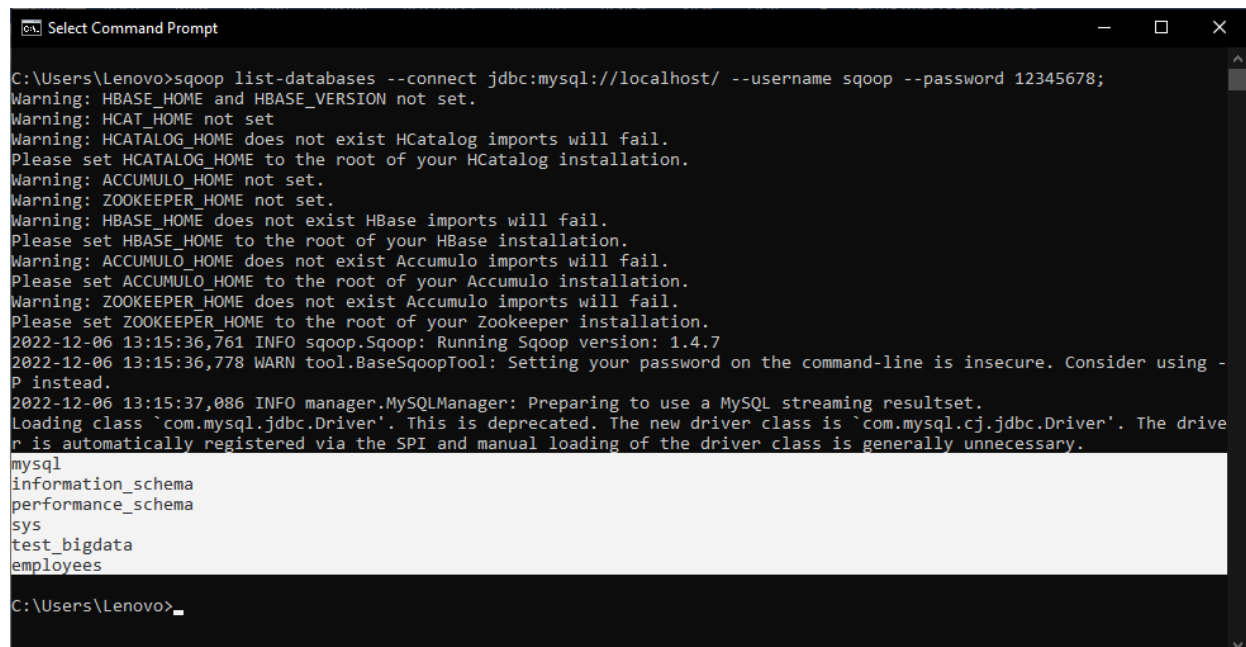
```
MySQL 8.0 Command Line Client
mysql> grant all privileges on test_bigdata.* to 'sqoop'@'localhost';
Query OK, 0 rows affected (0.23 sec)

mysql>
```

Step 5: Testing SQOOP

Open the Command Prompt and run the following command.

Command: sqoop list-databases --connect jdbc:mysql://localhost/ --username sqoop --password 12345678 (Running this command it will show the list of databases)



```
Select Command Prompt
C:\Users\Lenovo>sqoop list-databases --connect jdbc:mysql://localhost/ --username sqoop --password 12345678;
Warning: HBASE_HOME and HBASE_VERSION not set.
Warning: HCAT_HOME not set
Warning: HCATALOG_HOME does not exist HCatalog imports will fail.
Please set HCATALOG_HOME to the root of your HCatalog installation.
Warning: ACCUMULO_HOME not set.
Warning: ZOOKEEPER_HOME not set.
Warning: HBASE_HOME does not exist HBase imports will fail.
Please set HBASE_HOME to the root of your HBase installation.
Warning: ACCUMULO_HOME does not exist Accumulo imports will fail.
Please set ACCUMULO_HOME to the root of your Accumulo installation.
Warning: ZOOKEEPER_HOME does not exist Accumulo imports will fail.
Please set ZOOKEEPER_HOME to the root of your Zookeeper installation.
2022-12-06 13:15:36,761 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2022-12-06 13:15:36,778 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2022-12-06 13:15:37,086 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
mysql
information_schema
performance_schema
sys
test_bigdata
employees
C:\Users\Lenovo>
```