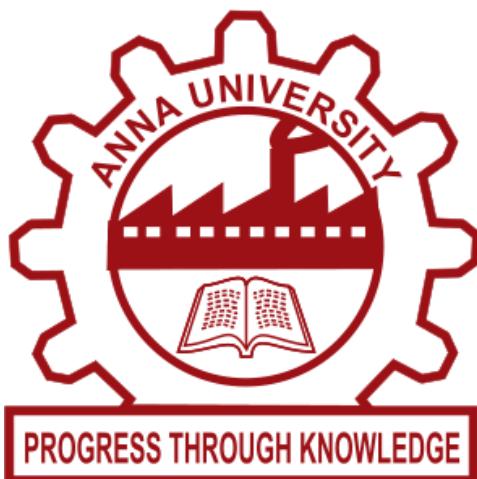


UNIVERSITY COLLEGE OF ENGINEERING NAGERCOIL

(ANNA UNIVERSITY CONSTITUENT COLLEGE)

KONAM, NAGERCOIL – 629 004



RECORD NOTE BOOK

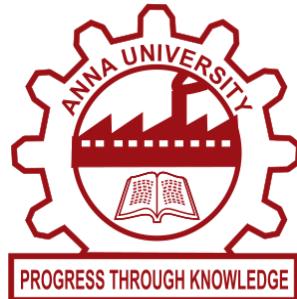
CCS334-BIG DATA ANALYTICS

REGISTER NO:

UNIVERSITY COLLEGE OF ENGINEERING NAGERCOIL

(ANNA UNIVERSITY CONSTITUENT COLLEGE)

KONAM, NAGERCOIL – 629 004



Register No:

*Certified that, this is the Bonafide record of work done by
Mr./Ms. of VI Semester
in Computer Science and Engineering of this college, in the
CCS334-BIG DATA ANALYTICS during academic year 2023-2024 in
partial fulfillment of the requirements of the B.E Degree course of
the Anna University Chennai.*

Staff-in-charge

Head of the Department

This record is submitted for the University Practical Examination
held on

Internal Examiner

External Examiner

INDEX

Exp No	Date	Title	Page	Sign
1.	01/02/2024	Downloading and installing Hadoop; Understanding different Hadoop modes. Startup scripts, Configuration files.	1	
2.	08/02/2024	Hadoop Implementation of file management tasks, such as Adding files and directories, retrieving files and Deleting files.	6	
3.	04/03/2024	Implement of Matrix Multiplication with Hadoop Map Reduce.	10	
4.	07/03/2024	Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.	17	
5.	14/03/2024	Installation of Hive along with practice examples.	19	
6.	26/03/2024	Installation of HBase, Installing thrift along with Practice examples.	22	
7.	28/03/2024	Practice importing and exporting data from various databases.	29	

EX NO:1

01/02/2024

DOWNLOADING AND INSTALLING HADOOP UNDERSTANDING DIFFERENT HADOOP MODES, STARTUP SCRIPTS, CONFIGURATION FILES

AIM:

To Downloading and installing Hadoop; Understanding different Hadoop modes. Startup scripts, Configuration files.

PREREQUISITES TO INSTALL HADOOP ON WINDOWS:

VIRTUAL BOX (For Linux): it is used for installing the operating system on it.

OPERATING SYSTEM: You can install Hadoop on Windows or Linux based operating systems. Ubuntu and CentOS are very commonly used.

JAVA: You need to install the Java 8 package on your system.

HADOOP: You require Hadoop latest version

1. Install Java

Java JDK Link to download here

Extract and install Java in C:\Java

Open cmd and type -> javac –version

2. Download Hadoop

Hadoop Link to download: <https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz>

Extract to C:\Hadoop

3. Set the path JAVA_HOME Environment variable

4. Set the path HADOOP_HOME Environment variable

5. Configure the required files in the C:\hadoop\etc as follows:

For core-site.xml

```
<configuration>
```

```
  <property>
```

```
    <name>fs.defaultFS</name>
```

```
    <value>hdfs://localhost:9000</value>
```

```
  </property>
```

```
</configuration>
```

```

# For hdfs-site.xml or https-site.xml

<configuration>

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

# For mapred-site.xml

<configuration>

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

# For yarn-site.xml

<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>

```

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

6.Create the following folders:

Create folder “data” under “C:\Hadoop”

Create folder “datanode” under “C:\Hadoop-3.3.0\data”

Create folder “namenode” under “C:\Hadoop-3.3.0\data”

7.Edit file hadoop-env.cmd:

C:/Hadoop-3.3.0/etc/hadoop/hadoop-env.cmd by closing the command line “JAVA HOME=%JAVA HOME%” instead of set ;

```
@rem The java implementation to use. Required.  
set JAVA_HOME=C:\progra~1\Java\jdk-1.8
```

8. Hadoop Configurations:

Download <https://github.com/s911415/apache-hadoop-3.1.0-winutil>

Copy folder bin and replace existing bin folder in C:\Hadoop-3.3.0\bin.

9.Format the NameNode:

Open cmd and type command “hdfs namenode –format”

```
C:\Windows\System32\cmd.exe  
Microsoft Windows [Version 10.0.19041.572]  
(c) 2020 Microsoft Corporation. All rights reserved.  
  
C:\hadoop-3.3.0\bin>hdfs namenode -format
```

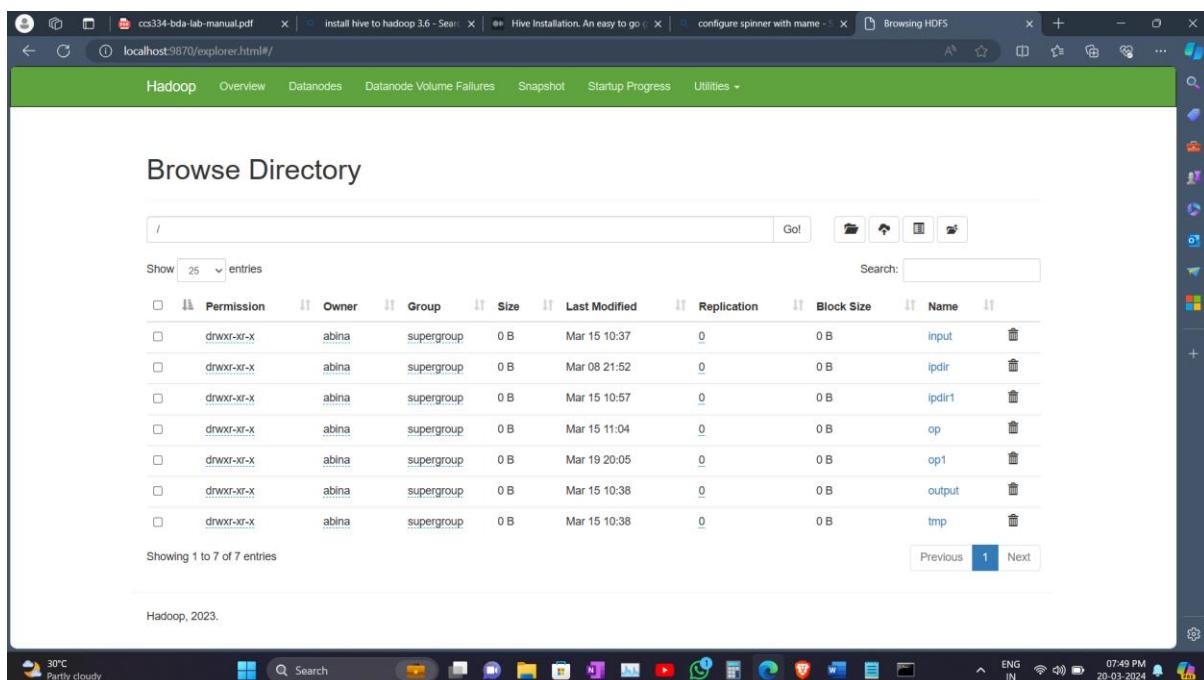
10. Start Hadoop:

Open cmd and type command “start-all.cmd”

Then verify the following windows appear without any warnings and errors.

11. Open localhost:

localhost:9870



localhost:8088

The screenshot shows the Hadoop Web UI running on port 8088. The title bar reads "All Applications". The left sidebar has a "Cluster" section with "About", "Nodes", "Node Labels", and "Applications" subsections. The "Applications" subsection lists stages: NEW, NEW_SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, and KILLED. Below this is a "Scheduler" section and a "Tools" section. The main content area displays "Cluster Metrics" with values: Apps Submitted: 0, Apps Pending: 0, Apps Running: 0, Apps Completed: 0, Containers Running: 0, and Used Resources: <memory:0 B, vCores:0>. It also shows "Cluster Nodes Metrics" with Active Nodes: 1, Decommissioning Nodes: 0, and Decommissioned Nodes: 0. Under "Scheduler Metrics", it shows Scheduler Type: Capacity Scheduler, Scheduling Resource Type: [memory-mb (unit=M), vcores], and Minimum Allocation: <memory:1024, vCores:1>. A table titled "Show 20 entries" is present but empty, stating "No data available in table". The bottom status bar shows the date and time as 20-03-2024, 07:50 PM.

Result:

To Downloading and installing Hadoop; Understanding different Hadoop modes. Startup scripts, Configuration files

EX NO:2**08/02/2024**

HADOOP IMPLEMENTATION OF FILE MANAGEMENT TASKS SUCH AS ADDING FILES AND DIRECTORIES, RETRIEVING FILES AND DELETING FILES

AIM:

To implement the following file management tasks in Hadoop:

1. Adding files and directories
2. Retrieving files
3. Deleting Files

PROCEDURE:

1.Create a directory in HDFS at given path:

Usage:

hadoop fs -mkdir <paths> Example:

hadoop fs -mkdir /user/saurzcode/dir1 /user/saurzcode/dir2

```
C:\hadoop\sbin>hadoop fs -mkdir /dir1  
C:\hadoop\sbin>
```



Browse Directory

The screenshot shows a Windows taskbar with several application icons visible, including a search bar, a file explorer icon, and icons for Google Chrome, Microsoft Edge, and others. On the right side of the taskbar, there are system status icons for battery, signal strength, and network, along with the date and time (29-03-2024, 07:49 PM).

2.List the contents of a directory:

Usage:

```
hadoop fs -ls <args>
```

Example:

```
hadoop fs -ls /user/saurzcode
```

```
C:\Windows\System32\cmd.exe
C:\hadoop\sbin>hadoop fs -touchz /dir1/1.txt

C:\hadoop\sbin>hadoop fs -ls /dir1
Found 1 items
-rw-r--r-- 1 user supergroup          0 2024-03-29 19:51 /dir1/1.txt

C:\hadoop\sbin>
```

3.Upload and download a file in HDFS:

Upload: hadoop fs -put:

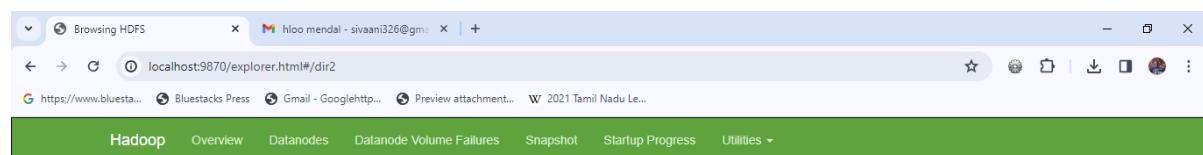
Copy single src file, or multiple src files from local file system to the Hadoop data file system Usage:

hadoop fs -put <localsrc>...<HDFS_dest_Path> Example:

```
hadoop fs -put /home/saurzcode/Samplefile.txt /user/saurzcode/dir3/
```

```
C:\hadoop\sbin>hadoop fs -put C:\Users\user\b.py /dir2

C:\hadoop\sbin>
```



Browse Directory

Browse Directory								
/dir2								
Go! 								
Search: <input type="text"/>								
Show	25	entries	Permission	Owner	Group	Size	Last Modified	Replication
<input type="checkbox"/>	-rw-r--r--	user	supergroup	0 B	Mar 29 19:58	1	128 MB	1.txt
<input type="checkbox"/>	-rw-r--r--	user	supergroup	2.61 KB	Mar 29 22:13	1	128 MB	b.py

Showing 1 to 2 of 2 entries

[Previous](#) [1](#) [Next](#)

Hadoop, 2023.



Adding files to a directory:

hadoop fs -touchz:

Adds files to a directory.

Usage:

hadoop fs -touchz /directory_name/file

Example:

hadoop fs -touchz /dir1/1.txt

```
C:\Windows\System32\cmd.exe
C:\hadoop\sbin>hadoop fs -touchz /dir1/1.txt
C:\hadoop\sbin>
```

Browse Directory

/dir1							Go!	File Operations
Show 25 entries							Search:	
	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
	-rw-r--r--	user	supergroup	0 B	Mar 29 19:51	1	128 MB	1.txt
Showing 1 to 1 of 1 entries								
Hadoop, 2023.								



4. Remove a file or directory in HDFS.

Remove files specified as argument. Deletes directory only when it is empty Usage:

hadoop fs -rm <arg> Example:

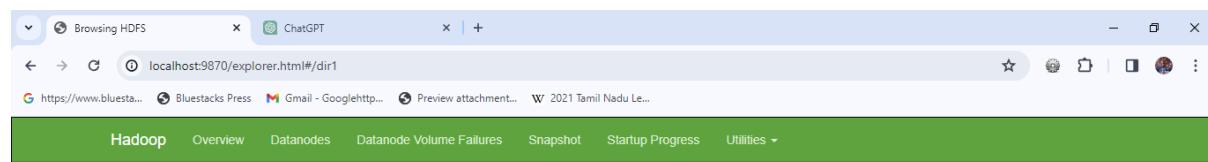
hadoop fs -rm /user/saurzcode/dir1/abc.txt

Recursive version of delete.

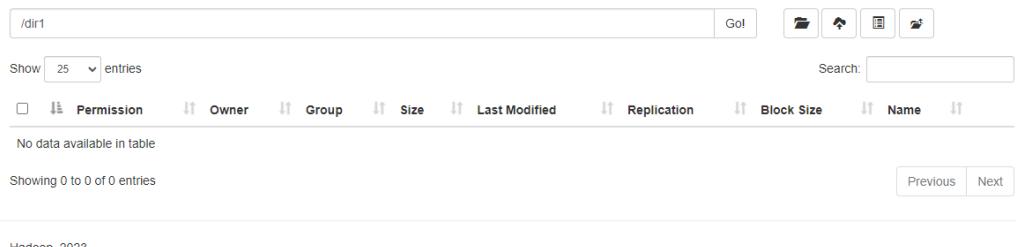
Usage:

hadoop fs -rm <arg> Example: hadoop fs -rmdir /user/saurzcode/

```
C:\Windows\System32\cmd.exe
C:\hadoop\sbin>hadoop fs -rm /dir1/1.txt
Deleted /dir1/1.txt
C:\hadoop\sbin>
```



Browse Directory



Hadoop, 2023.



RESULT:

Thus, the Hadoop Implementation of file management tasks, such as Adding files and directories, retrieving files and Deleting files is executed successfully.

IMPLEMENTATION OF MATRIX MULTIPLICATION USING MAP REDUCE

AIM:

To implement matrix multiplication using mapreduce in Hadoop.

PROGRAM:**Map.java**

```
package matrix;

import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
//import org.apache.hadoop.mapreduce.Mapper;
import java.io.IOException;

public class Map extends org.apache.hadoop.mapreduce.Mapper<LongWritable, Text, Text, Text> {

    @Override
    public void map(LongWritable key, Text value, Context context)
            throws IOException, InterruptedException {
        Configuration conf = context.getConfiguration();
        int m = Integer.parseInt(conf.get("m"));
        int p = Integer.parseInt(conf.get("p"));

        String line = value.toString();
        // (M, i, j, Mij);
        String[] indicesAndValue = line.split(",");
        Text outputKey = new Text();
        Text outputValue = new Text();
        if (indicesAndValue[0].equals("M")) {
            for (int k = 0; k < p; k++) {
                outputKey.set(indicesAndValue[1] + "," + k);
```

```

// outputKey.set(i,k);
outputValue.set(indicesAndValue[0] + "," + indicesAndValue[2]
+ "," + indicesAndValue[3]);
// outputValue.set(M,j,Mij);
context.write(outputKey, outputValue);
}
} else {
    // (N, j, k, Njk);
    for (int i = 0; i < m; i++) {
        outputKey.set(i + "," + indicesAndValue[2]); outputValue.set("N," +
indicesAndValue[1] + ","
+ indicesAndValue[3]); context.write(outputKey, outputValue);
    }
}
}
}

```

MatrixMultiply.java

```

package matrix;

import org.apache.hadoop.conf.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

public class MatrixMultiply {
    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.err.println("Usage: MatrixMultiply <in_dir> <out_dir>");
            System.exit(2);
        }
    }
}

```

```

Configuration conf = new Configuration();
conf.set("m", "1000");
conf.set("n", "100");
conf.set("p", "1000");
@SuppressWarnings("deprecation")
Job job = new Job(conf, "MatrixMultiply");
job.setJarByClass(MatrixMultiply.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(Text.class);
job.setMapperClass(Map.class);
job.setReducerClass(Reduce.class);
job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
job.waitForCompletion(true);
}
}

```

Reduce.java

```

package matrix;
import org.apache.hadoop.io.Text;
// import org.apache.hadoop.mapreduce.Reducer;
import java.io.IOException;
import java.util.HashMap;
public class Reduce
extends org.apache.hadoop.mapreduce.Reducer<Text, Text, Text, Text> { @Override
public void reduce(Text key, Iterable<Text> values, Context context)
throws IOException, InterruptedException {
String[] value;

```

```

//key=(i,k),
//Values = [(M/N,j,V/W),..]

HashMap<Integer, Float> hashA = new HashMap<Integer, Float>(); HashMap<Integer,
Float> hashB = new HashMap<Integer, Float>(); for (Text val : values) {
    value = val.toString().split(",");
    if (value[0].equals("M")) {
        hashA.putInt(Integer.parseInt(value[1]), Float.parseFloat(value[2])); } else {
        hashB.putInt(Integer.parseInt(value[1]), Float.parseFloat(value[2]));
    }
}
int n = Integer.parseInt(context.getConfiguration().get("n"));
float result = 0.0f;
float m_ij;
float n_jk;
for (int j = 0; j < n; j++) {
    m_ij = hashA.containsKey(j) ? hashA.get(j) : 0.0f; n_jk = hashB.containsKey(j) ?
    hashB.get(j) : 0.0f; result += m_ij * n_jk;
}
if (result != 0.0f) {
    context.write(null,
    new Text(key.toString() + "," + Float.toString(result)));
}
}
}
}

```

PROCEDURE:

Step 1:create a directory in hadoop

Input the two text files with the matrix files int into the direcrory in Hadoop.

Using

Hadoop fs -put <file location>/<input directory>.

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	⋮
□	-rw-r--r--	abina	supergroup	34 B	Apr 05 21:16	1	128 MB	1.txt	✖
□	-rw-r--r--	abina	supergroup	34 B	Apr 05 21:16	1	128 MB	2.txt	✖

```

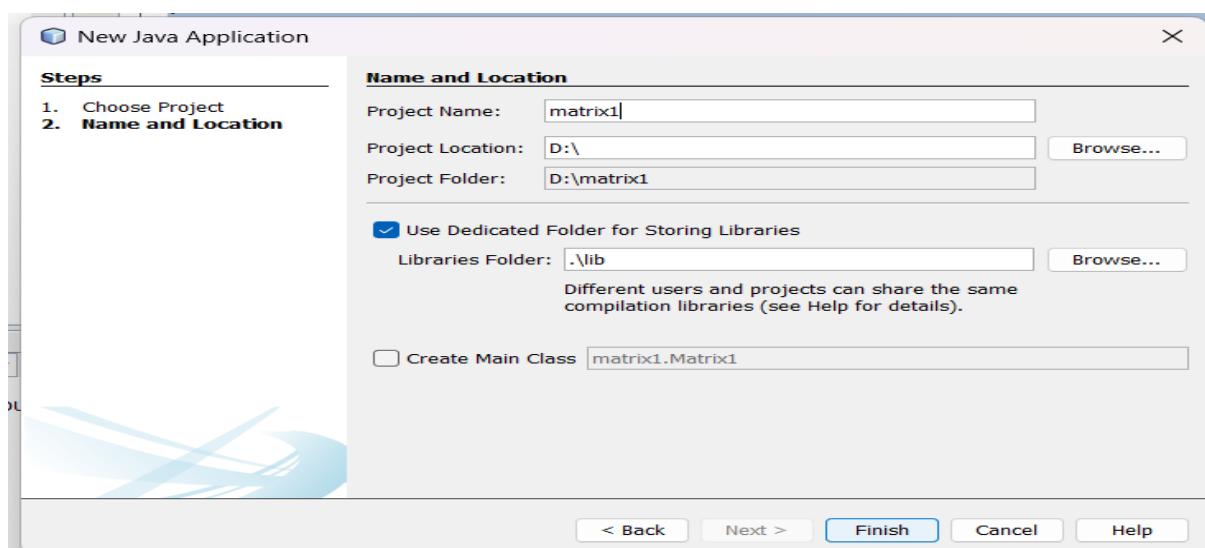
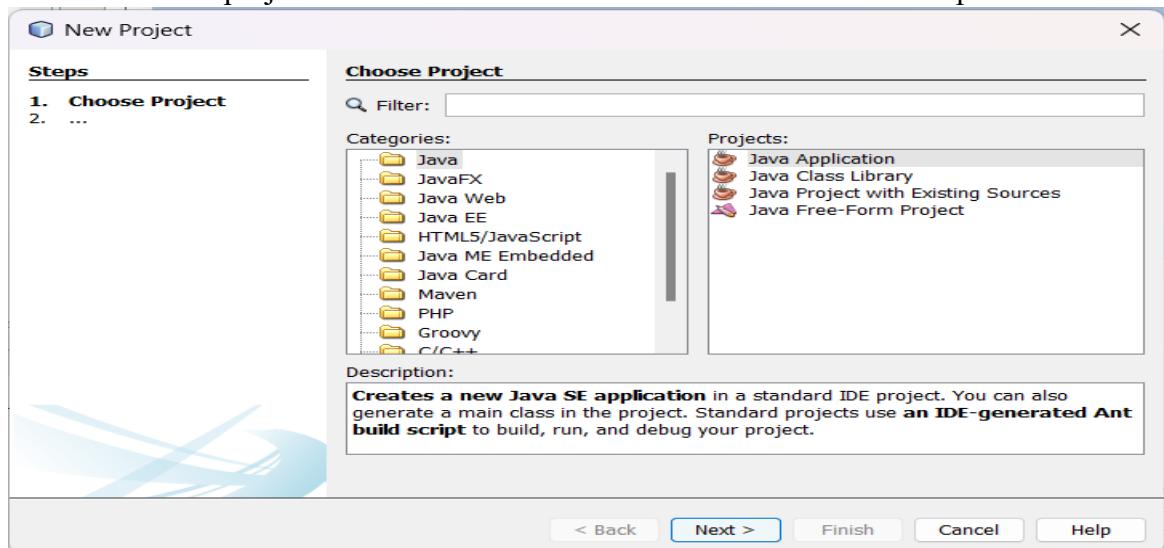
C:\hadoop\sbin>hadoop dfs -cat /ipdir/1.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
M,0,0,1
M,0,0,2
M,0,0,3
M,0,0,4
C:\hadoop\sbin>hadoop dfs -cat /ipdir/2.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
N,0,1,2
N,0,1,3
N,0,1,4
N,0,1,5
C:\hadoop\sbin>

```

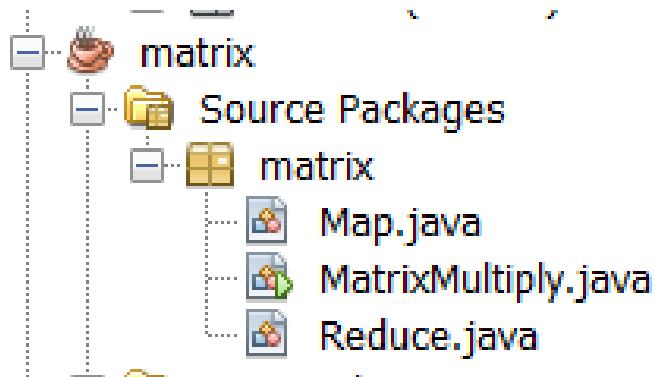
Step 2: Open the netbeans application and open a new project.

1. Choose categories as java and projects as java application.

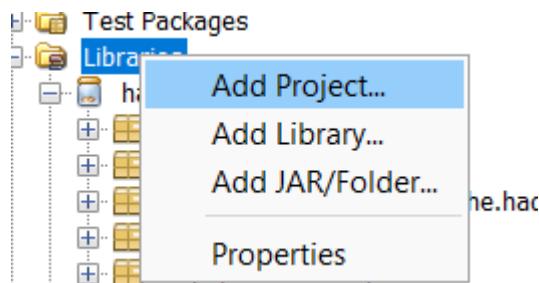
2. Set the project name as matrix and untick the create main class option.



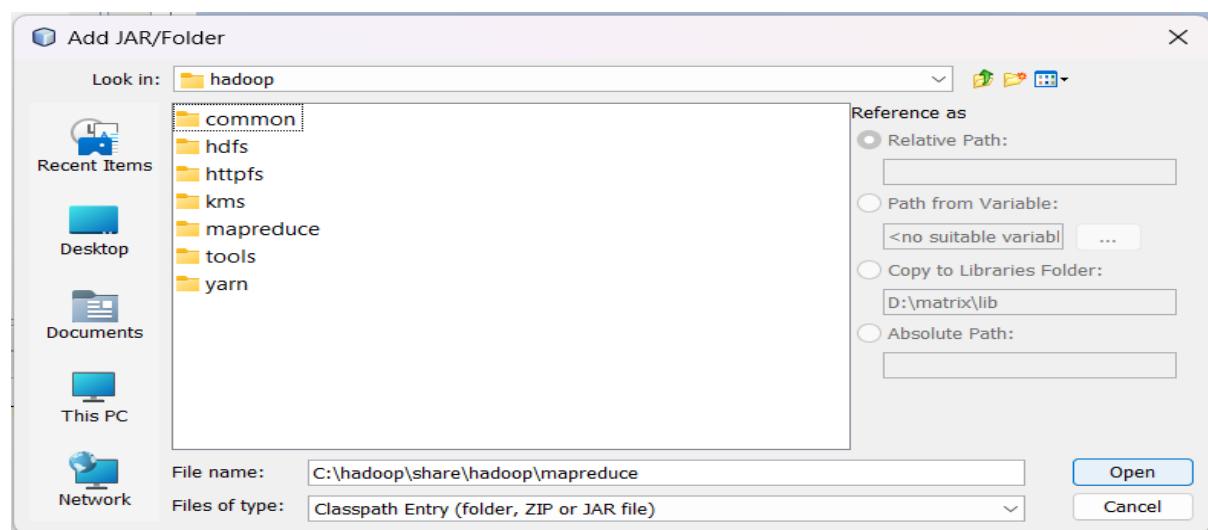
3.Add a new package a matrix and add the above three programs as three different class files



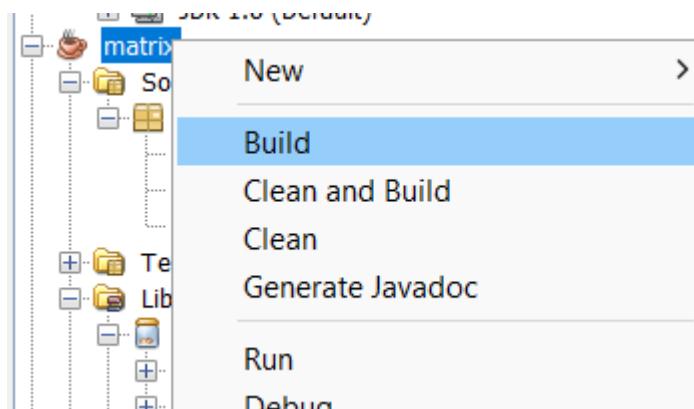
4.Add libraries to the package by right clicking in the libraries tab an select add JAR/FOLDER.



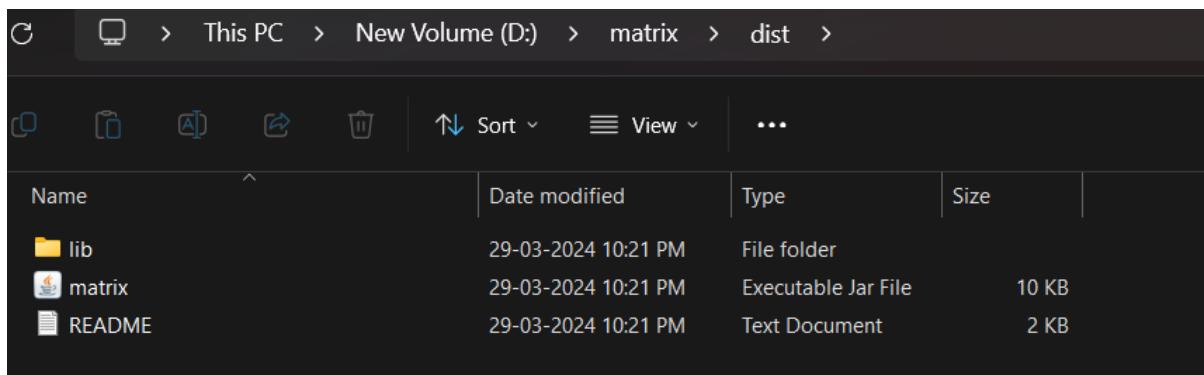
5.Add the folders like mapreduce and common.



6.Right click on the matrix project and clean and build the project.



7.The jar file will be added to the following location.



Step 3 : Input the following jar file into the same directory where we inserted the matrix files.

Using command hadoop jar <jar file name.jar jar name.package name.main class name> / input directory /output directory

```
D:\matrix\dist>hadoop jar matrix.jar matrix.MatrixMultiply /ipdir /op
```

Step 4:Get the output matrix

Using

```
hadoop dfs -cat /<output directory>/*
```

```
C:\hadoop\sbin>hadoop dfs -cat /op/*
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
0,1,8.0
```

RESULT:

Thus the implementation of matrix multiplication using map reduce was executed successfully.

EX NO:4

07/03/2024

RUN A BASIC WORD COUNT MAP REDUCE PROGRAM TO UNDERSTAND MAP REDUCE PARADIGM

AIM:

To run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.

PREREQUISITES:

VIRTUAL BOX (For Linux): it is used for installing the operating system on it.

OPERATING SYSTEM: You can install Hadoop on Windows or Linux based operating systems. Ubuntu and CentOS are very commonly used.

JAVA: You need to install the Java 8 package on your system.

HADOOP: You require Hadoop latest version

PROCEDURE:

Step 1: Install Hadoop with java and required configuration files.

Step 2: Create a directory in Hadoop.

```
C:\hadoop\sbin>hadoop fs -mkdir /ipdir1
```

Step 3: Upload a word file into the directory.

```
C:\hadoop\sbin>hadoop fs -mkdir /ipdir1
C:\hadoop\sbin>hadoop fs -put D:\word.txt /ipdir1
```

Step 4: View the file using cat option.

```
C:\hadoop\sbin>hadoop dfs -cat /ipdir1/word.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
apple is a red color fruit. it is sweet in taste.
```

Step 5: Go to file location C:\hadoop\share\hadoop\mapreduce and open cmd. And type the following cmd.

```
C:\hadoop\share\hadoop\mapreduce>hadoop jar hadoop-mapreduce-examples-3.3.6.jar
An example program must be given as the first argument.
Valid program names are:
aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
dbcount: An example job that count the pageview counts from a database.
distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
grep: A map/reduce program that counts the matches of a regex in the input.
join: A job that effects a join over sorted, equally partitioned datasets
multifilewc: A job that counts words from several files.
pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
randomwriter: A map/reduce program that writes 10GB of random data per node.
secondarysort: An example defining a secondary sort to the reduce.
sort: A map/reduce program that sorts the data written by the random writer.
sudoku: A sudoku solver.
teragen: Generate data for the terasort
terasort: Run the terasort
teravalidate: Checking results of terasort
wordcount: A map/reduce program that counts the words in the input files.
wordmean: A map/reduce program that counts the average length of the words in the input files.
wordmedian: A map/reduce program that counts the median length of the words in the input files.
wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.

C:\hadoop\share\hadoop\mapreduce>hadoop jar hadoop-mapreduce-examples-3.3.6.jar wordcount /ipdir1
Usage: wordcount <in> [<in>...] <out>

C:\hadoop\share\hadoop\mapreduce>hadoop jar hadoop-mapreduce-examples-3.3.6.jar wordcount /ipdir1 /op1
24/04/07 22:45:20 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
24/04/07 22:45:20 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
```

Step 6: View the output file using cat option.

```
C:\hadoop\share\hadoop\mapreduce>hadoop dfs -cat /op1/*
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

a      1
apple  1
color  1
fruit. 1
in     1
is     2
it     1
red    1
sweet  1
taste. 1
```

RESULT:

Thus running a basic Word Count Map Reduce program to understand Map Reduce Paradigm was executed successfully.

INSTALLATION OF HIVE ALONG WITH PRACTICE EXAMPLES

AIM:

To install HIVE along with practice examples.

PREREQUISITES:

1. Java Development Kit (JDK8) installed and the JAVA_HOME environment variable set.
2. Hadoop (version 2.9.2) installed and configured on your Windows system.
3. Apache Derby installed and DERBY_HOME environment variable set.

PROCEDURE:**1.Download HIVE:**

Visit the Apache Hive website and download the stable version (version 3.1.2) of Hive.
 Apache Hive website: <https://hive.apache.org/>

2. Extract the Downloaded Hive Archive to a Directory on Your Windows Machine:
 C:\hive.**3.Configure Hive:**

1. Download a Hive configuration file (hive-site.xml) in the conf folder of the extracted hive.

Download link: <https://drive.google.com/file/d/1tsBbHdvM1fFktmn9O0-u0pbG1vWWFoyE/view>

2. Copy the library files in the derby lib folder and paste them into the hive lib folder.
3. Replace the bin folder in Hive with the fixed bin folder from this website:
[GitHub - HadiFadl/Hive-cmd: All cmd files needed to run Hive on windows \(taken from https://svn.apache.org/repos/asf/hive/trunk/bin/\)](https://svn.apache.org/repos/asf/hive/trunk/bin/)

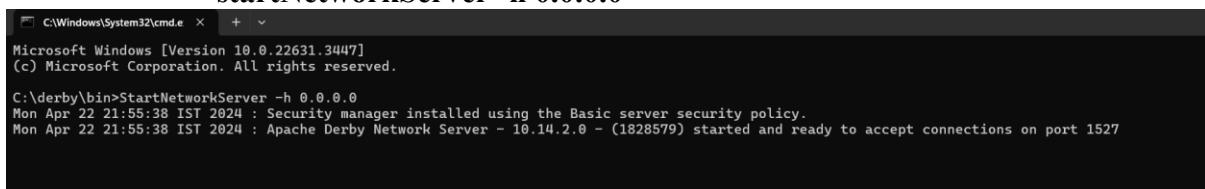
4.Environment Variable Setup:

1. Add the HIVE binary directory (C:\hive\bin) to your PATH environment variable.
2. Set the HIVE_HOME, HIVE_LIB and HIVE_BIN environment variables.
3. Add the Path variable HADOOP_USER_CLASSPATH_FIRST = true in both User and System variables.

5.Start Hive:

1. Run Hadoop in the command prompt.
2. Open a new command prompt and navigate to the Derby installation directory i.e., C:\derby\bin and execute the command:

>startNetworkServer -h 0.0.0.0



```

C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.22631.3447]
(c) Microsoft Corporation. All rights reserved.

C:\derby\bin>StartNetworkServer -h 0.0.0.0
Mon Apr 22 21:55:38 IST 2024 : Security manager installed using the Basic server security policy.
Mon Apr 22 21:55:38 IST 2024 : Apache Derby Network Server - 10.14.2.0 - (1828579) started and ready to accept connections on port 1527
  
```

3. Open a new command prompt and navigate to the Hive installation directory i.e., C:\hive\bin and execute the command:

```
>hive
```

```
C:\hive\bin>hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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.log4j.Log4jLoggerFactory]
2024-04-22T21:56:07,328 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Found configuration file null
2024-04-22T21:56:09,539 WARN [main] org.apache.hadoop.hive.common.LogUtils - hive-site.xml not found on CLASSPATH
Hive Session ID = 0d511a3a-583e-4613-b7b8-9ca4798701dd

Logging initialized using configuration in jar:file:/C:/hive/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
2024-04-22T21:56:11,835 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/abina/0d511a3a-583e-4613-b7b8-9ca4798701dd
2024-04-22T21:56:11,841 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created local directory: C:/Users/abina/AppData/Local/Temp/abina/0d511a3a-583e-4613-b7b8-9ca4798701dd
2024-04-22T21:56:11,855 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/abina/0d511a3a-583e-4613-b7b8-9ca4798701dd/_tmp_space.db
2024-04-22T21:56:11,889 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 0d511a3a-583e-4613-b7b8-9ca4798701dd main
Hive Session ID = ebd83d1-d80b-4c01-9180-c8c286b6d2af
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
2024-04-22T21:56:22,683 INFO [0d511a3a-583e-4613-b7b8-9ca4798701dd main] CliDriver - Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
2024-04-22T21:56:22,705 INFO [pool-9-thread-1] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/abina/ebdf83d1-d80b-4c01-9180-c8c286b6d2af
2024-04-22T21:56:22,709 INFO [pool-9-thread-1] org.apache.hadoop.hive.ql.session.SessionState - Created local directory: C:/Users/abina/AppData/Local/Temp/abina/ebdf83d1-d80b-4c01-9180-c8c286b6d2af
2024-04-22T21:56:22,715 INFO [pool-9-thread-1] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/abina/ebdf83d1-d80b-4c01-9180-c8c286b6d2af/_tmp_space.db
hive> |
```

EXAMPLES:

Create a Database:

To create a new database HIVE, use the following syntax:

```
CREATE DATABASE database_name;
```

Example:

```
CREATE DATABASE ipdir;
```

```
hive> CREATE DATABASE ipdir;
2024-04-22T22:11:36,830 INFO [0d511a3a-583e-4613-b7b8-9ca4798701dd main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id : 0d511a3a-583e-4613-b7b8-9ca4798701dd
2024-04-22T22:11:37,849 WARN [0d511a3a-583e-4613-b7b8-9ca4798701dd main] org.apache.hadoop.hive.ql.session.SessionState - METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFactory.
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. Database ipdir already exists
2024-04-22T22:11:37,910 [0d511a3a-583e-4613-b7b8-9ca4798701dd main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id : 0d511a3a-583e-4613-b7b8-9ca4798701dd
2024-04-22T22:11:37,910 INFO [0d511a3a-583e-4613-b7b8-9ca4798701dd main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
hive> |
```

Show Database:

To display a list of available databases in HIVE, use the following syntax:

```
SHOW DATABASES;
```

```
hive> SHOW DATABASES;
2024-04-21T20:59:39,034 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id : 8
2024-04-21T20:59:39,034 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 420d081b-OK
2024-04-21T20:59:39,470 INFO [420d081b-b756-4997-b01e-5fcc85825228 main] org.apache.hadoop.conf.Configuration.deprecatedProperties: mapreduce.input.fileinputformat.inputdir
abc
default
demo
hive
ipdir
ipdir1
temp
Time taken: 0.415 seconds, Fetched: 7 row(s)
```

Use a Database:

To use a specific database in HIVE, use the following syntax:

```
USE database_name;
```

Example:

```
USE ipdir;
```

```
hive> USE ipdir;
2024-04-21T21:00:19,625 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 420d081b-b756-4997-b01e-5fcc85825228
8
2024-04-21T21:00:19,625 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 420d081b-b756-4997-b01e-5fcc85825228 main
OK
Time taken: 0.049 seconds
```

Create a Table:

To create a table in HIVE, use the following syntax:

```
CREATE TABLE table_name(column1 datatype, column 2 datatype, ...);
```

Example:

```
CREATE TABLE student(id int, name string, age int);
```

```
hive> CREATE TABLE student(id int, name string, age int);
2024-04-21T21:01:15,742 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 420d081b-b756-4997-b01e-5fcc85825228
8
2024-04-21T21:01:15,742 INFO [main] org.apache.hadoop.hive.session.SessionState - Updating thread name to 420d081b-b756-4997-b01e-5fcc85825228 main
OK
Time taken: 1.16 seconds
```

Insert data into a Table:

To insert data into a table in HIVE, use the following syntax:

```
INSERT INTO table_name (column1, column2, ..) VALUES(value1, value2, ..);
```

Example:

```
INSERT INTO student(id, name, age) VALUES(1, 'John Doe', 25);
```

```
hive> INSERT INTO student(id, name, age) VALUES(001, 'manu ', 25);
2024-04-21T21:02:59,938 INFO [main] org.apache.hadoop.conf.HiveConf - Using the default value passed in for log id: 420d081b-b756-4997-b01e-5fcc85825228
8
2024-04-21T21:03:03,085 INFO [420d081b-b756-4997-b01e-5fcc85825228 main] org.apache.hadoop.common.FileUtils - Creating directory if it doesn't exist: h
dfs://localhost:9000/user/hive/warehouse/ipdir.db/student/_hive-staging_hive_2024-04-21-02-59_960_79167738667490999885-1
2024-04-21T21:03:03,314 INFO [420d081b-b756-4997-b01e-5fcc85825228 main] org.apache.hadoop.common.FileUtils - Creating directory if it doesn't exist: f
ile:/C:/Users/abina/AppData/Local/Temp/abina/420d081b-b756-4997-b01e-5fcc85825228/hive_2024-04-21_21-03-03_159_2527877755128861028-1/-mr-10000/.hive-staging
_hive_2024-04-21_21-03-03_159_2527877755128861028-1
Query ID = abina_20240421210259_0ba4ef63-20b1-4d34-b6e4-e3ba5f60acc6
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
```

Select Data from a Table:

To select data from a table in HIVE, use the following syntax:

```
SELECT * FROM table_name;
```

Example:

```
SELECT * FROM student;
```

```
hive> SELECT * FROM student;
2024-04-21T21:04:01,400 INFO [main] org.apache.hadoop.conf.HiveConf - Using the default value passed in for log id: 420d081b-b756-4997-b01e-5fcc85825228
8
2024-04-21T21:04:01,400 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 420d081b-b756-4997-b01e-5fcc85825228 main
2024-04-21T21:04:01,629 INFO [420d081b-b756-4997-b01e-5fcc85825228 main] org.apache.hadoop.common.FileUtils - Creating directory if it doesn't exist: f
ile:/C:/Users/abina/AppData/Local/Temp/abina/420d081b-b756-4997-b01e-5fcc85825228/hive_2024-04-21_04-01_425_7343547800977414264-1/-mr-10001/.hive-staging
_hive_2024-04-21_04-01_425_7343547800977414264-1
OK
1      manu    25
Time taken: 0.263 seconds, Fetched: 1 row(s)
2024-04-21T21:04:01,711 INFO [420d081b-b756-4997-b01e-5fcc85825228 main] org.apache.hadoop.conf.HiveConf - Using the default value passed in for log id
: 420d081b-b756-4997-b01e-5fcc85825228
2024-04-21T21:04:01,711 INFO [420d081b-b756-4997-b01e-5fcc85825228 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id
```

RESULT:

Thus, the installation of HIVE along with practice examples was executed successfully.

EX.NO:6a

26/03/2024

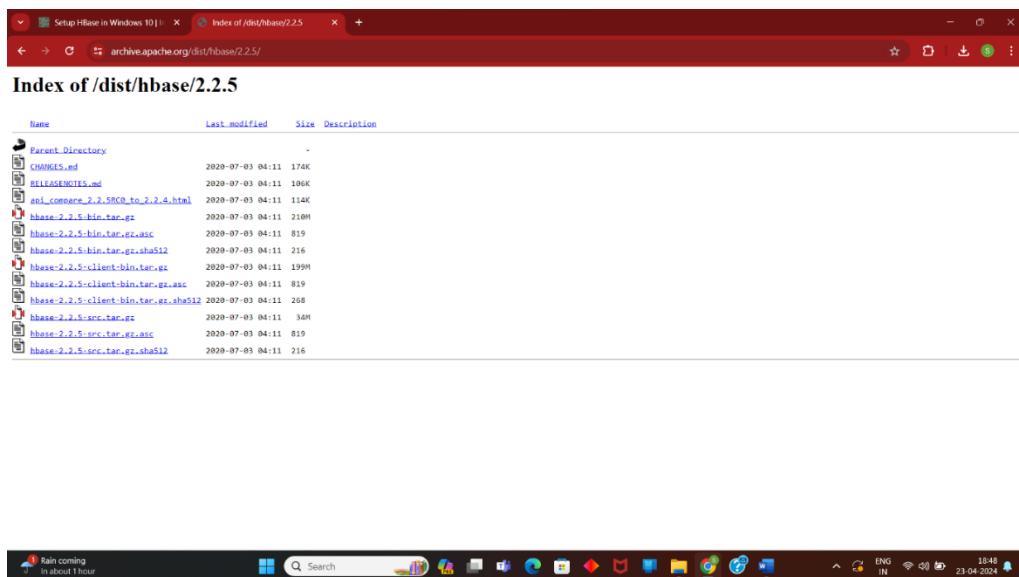
INSTALLATION OF HBASE ALONG WITH PRACTICE EXAMPLES

AIM:

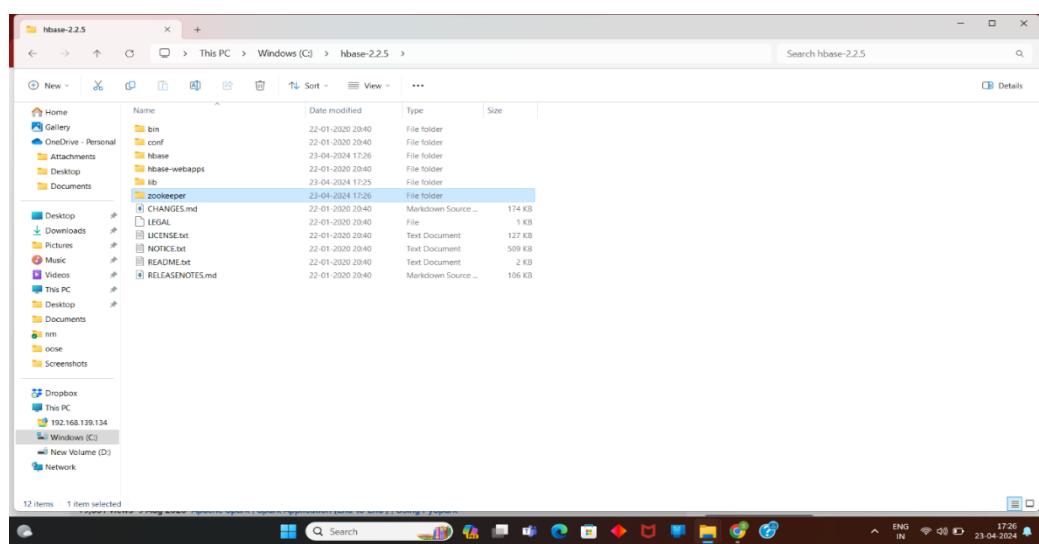
To install hbase along with practice examples.

PROCEDURE:

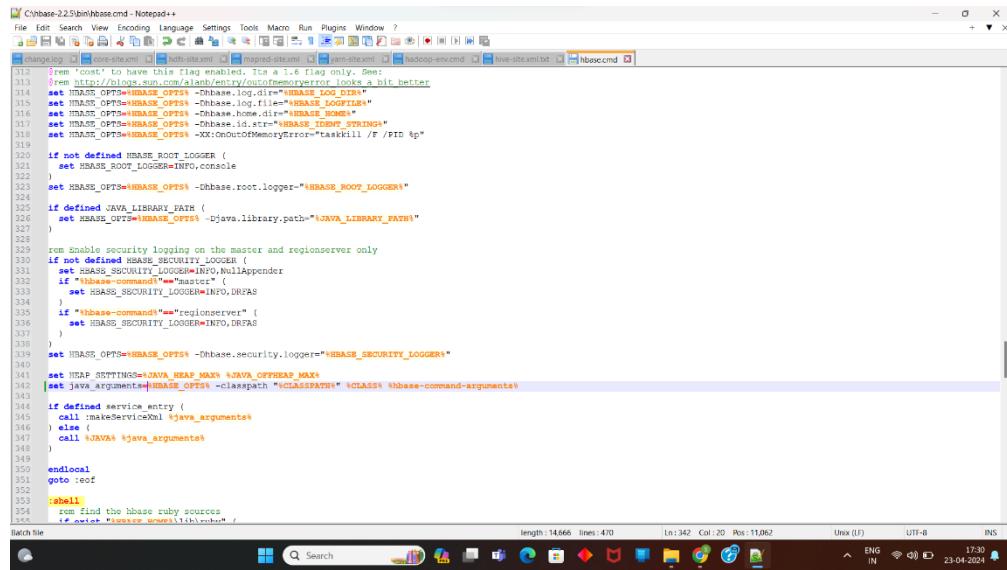
Step 1: Download the hbase 2.2.5 bin file from apache website and unzip the file and extract the file.



Step 2: Create a folder as shown below inside root folder for HBase data and zookeeper.



Step 3: Open C:/ hbase-2.2.5/bin/hbase.cmd in notepad++. Search for below given lines and remove **%HEAP_SETTINGS%** from that line



```

C:\hbase-2.2.5\bin\hbase.cmd - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
File Explorer Task View Start Taskbar Help
changeLog config-chp.cmd hbase-env.cmd hbase.cmd hadoop-env.cmd live-shumit.txt Hbasecmd

312 *coast* to have this flag enabled. It's a 1.6 flag only. See:
313 http://blogs.sun.com/alternentry/outofmemoryerror looks a bit better
314 set HBASE_OPTS=%HBASE_OPTS% -Dhbase.log.dir="%HBASE_HOME%\log"
315 set HBASE_OPTS=%HBASE_OPTS% -Dhbase.homes.dir="%HBASE_HOME%\homes"
316 set HBASE_OPTS=%HBASE_OPTS% -Dhbase.id.str="%HBASE_IDENT_STRINGS%"
317 set HBASE_OPTS=%HBASE_OPTS% -XX:OnOutOfMemoryError="taskkill /F /Td %p"
318
319 if not defined HBASE_ROOT_LOGGER (
320     set HBASE_ROOT_LOGGER=INFO,console
321 )
322 set HBASE_OPTS=%HBASE_OPTS% -Dhbase.root.logger=%HBASE_ROOT_LOGGER%
323
324 if defined JAVA_LIBRARY_PATH (
325     set HBASE_OPTS=%HBASE_OPTS% -Djava.library.path=%JAVA_LIBRARY_PATH%
326 )
327
328 enable security logging on the master and regionserver only
329 if not defined HBASE_SECURITY_LOGGER (
330     set HBASE_SECURITY_LOGGER=INFO,DRFASender
331     if "%hbbase_command%"=="master"
332         set HBASE_SECURITY_LOGGER=INFO,DRFAS
333
334     if "%hbbase_command%"=="regionserver" {
335         set HBASE_SECURITY_LOGGER=INFO,DRFAS
336     }
337 )
338
339 set HBASE_OPTS=%HBASE_OPTS% -Dhbase.security.logger=%HBASE_SECURITY_LOGGER%
340
341 set HEAP_SETTINGS=%JAVA_HEAP_MAX% %JAVA_OFFHEAP_MAX%
342 set JAVA_ARGUMENTS=%HBASE_OPTS% -classpath "%CLASSPATH%" %CLASS% %hbbase-command-arguments%
343
344 if defined service entry (
345     call :makeServiceXml %JAVA_arguments%
346 ) else (
347     call %JAVA% %JAVA_arguments%
348 )
349
350 endlocal
351 goto :eof
352
353 #shell
354 # Find the hbase ruby sources
355 if exist %hbase_home%\lib\client-facing-thirdparty\*

```

Step 4: Open C:/hbase-2.2.5/conf/hbase-env.cmd n notepad++. Add the below lines to the file after the comment session

- set JAVA_HOME=%JAVA_HOME%
- set HBASE_CLASSPATH=%HBASE_HOME%\lib\client-facing-thirdparty*
- set HBASE_HEAPSIZE=8000
- set HBASE_OPTS="-XX:+UseConcMarkSweepGC" "-Djava.net.preferIPv4Stack=true"
- set SERVER_GC_OPTS="-verbose:gc" "-XX:+PrintGCDetails" "-XX:+PrintGCDateStamps" %HBASE_GC_OPTS%
- set HBASE_USE_GC_LOGFILE=true
- set HBASE_JMX_BASE="-Dcom.sun.management.jmxremote.ssl=false" "-Dcom.sun.management.jmxremote.authenticate=false"
- set HBASE_MASTER_OPTS=%HBASE_JMX_BASE% "-Dcom.sun.management.jmxremote.port=10101"
- set HBASE_REGIONSERVER_OPTS=%HBASE_JMX_BASE% "-Dcom.sun.management.jmxremote.port=10102"
- set HBASE_THRIFT_OPTS=%HBASE_JMX_BASE% "-Dcom.sun.management.jmxremote.port=10103"
- set HBASE_ZOOKEEPER_OPTS=%HBASE_JMX_BASE% "-Dcom.sun.management.jmxremote.port=10104"
- set HBASE_REGIONSVERS=%HBASE_HOME%\conf\regionservers

- set HBASE_LOG_DIR=%HBASE_HOME%\logs
 - set HBASE_IDENT_STRING=%USERNAME%
 - set HBASE_MANAGES_ZK=true

```
C:\base-2.2.0\conf\hbase-env.cmd - Notepad++
```

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

File core-site.xml hdfs-site.xml mapred-site.xml xml-site.xml hadoop-env.cmd hive-site.xml hbase-env.cmd hbase-site.xml

```
7 %rem * "license.txt"; you may not use this file except in compliance
8 * with the License. You may obtain a copy of the License at
9 * http://www.apache.org/licenses/LICENSE-2.0
10 %rem *
11 %rem * Unless required by applicable law or agreed to in writing, software
12 %rem * distributed under the license is distributed on an "AS IS" BASIS,
13 %rem * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 %rem * See the License for the specific language governing permissions and
15 %rem * limitations under the license.
16 %rem */
17
18 %rem Set environment variables here.
19
20 %rem The Java implementation to use. Java 1.8+ required.
21 %rem set JAVA_HOME=C:\app\java
22 %rem set HBASE_CLASSPATH=%JAVA_HOME%\lib\client-facing-thirdparty*
23 %set HBASE_HEAPSIZE=8000
24 %set HBASE_GC_OPTS=-XX:+UseConcMarkSweepGC -XX:Java.net.preferIPv4Stack=true
25 %set SERVER_OPTS="-Xms1024m -Xmx1024m -XX:+PrintGCTimeStamps %HBASE_GC_OPTS"
26 %set HBASE_USE_GC_LOGFILE=true
27
28 %set HBASE_JMX_BASES="com.sun.management.jmxremote.ssl=false" -Dcom.sun.management.jmxremote.authenticate=false"
29
30 %set HBASE_MONITOR_OPTS=%HBASE_JMX_BASES% -Dcom.sun.management.jmxremote.port=10101
31 %set HBASE_REGIONSERVER_OPTS=%HBASE_JMX_BASES% -Dcom.sun.management.jmxremote.port=10102"
32 %set HBASE_MASTER_OPTS=%HBASE_JMX_BASES% -Dcom.sun.management.jmxremote.port=10103"
33 %set HBASE_ZOOKEEPER_OPTS=%HBASE_JMX_BASES% -Dcom.sun.management.jmxremote.port=10104"
34 %set HBASE_REGIONCOUNTERS=%HBASE_JMX_BASES%\conf\regionserver
35 %set HBASE_LOG_DIR=%HBASE_HOME%\logs
36 %set HBASE_PID_DIR=%HBASE_HOME%\pid
37 %set HBASE_MANAGES_ZK=true
38
39 %rem Extra Java CLASSPATH elements. Optional.
40 %rem see HBASE_CLASSPATH=
41
42 %rem The maximum amount of heap to use. Default is left to JVM default.
43 %rem set HBASE_HEAPSIZE=1000
44
45 %rem Uncomment below if you intend to use off heap cache. For example, to allocate 8G of
46 %rem offheap, set the value to "8G".
47 %rem set HBASE_OFFHEAPSIZE=1000
48
49 %rem For avast to allocate 8G of offheap to 8G.
```

Step 5: Open C:/hbase-2.2.5/conf/hbase-site.xml notepad++. Add the below lines inside <configuration> tag.

```
<property>
    <name>hbase.rootdir</name>
    <value>C:/hbase-2.2.5/hbase</value>
</property>
<property>
    <name>hbase.zookeeper.property.dataDir</name>
    <value>C:/hbase-2.2.5/zookeeper</value>
</property>
<property>
    <name> hbase.zookeeper.quorum</name>
    <value>localhost</value>
</property>
```

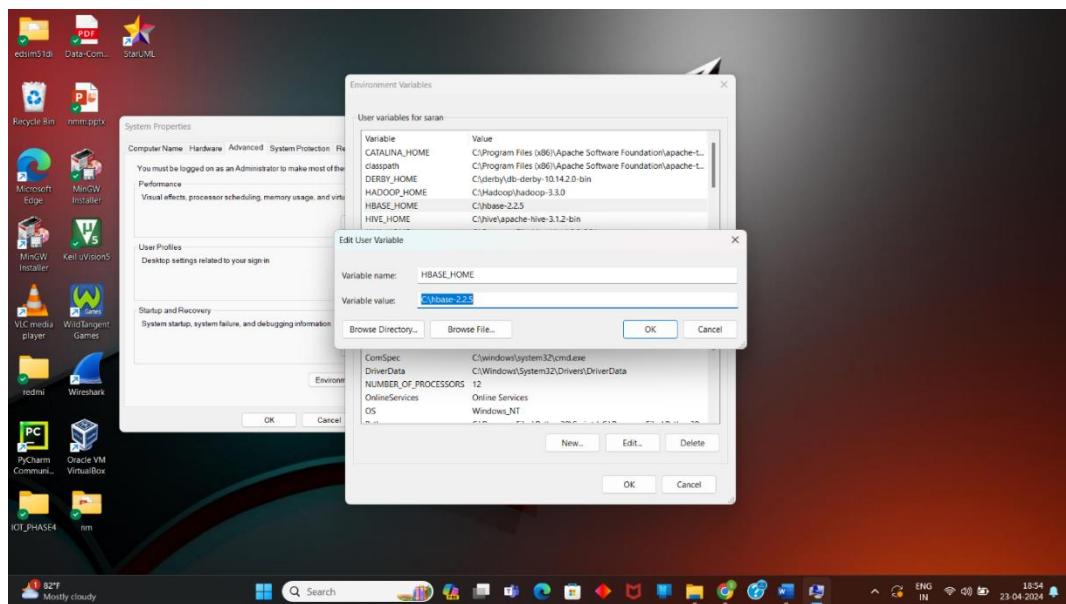
```

28 local filesystem, in a path under the value configured for 'hbasesite.tmp.dir'.
29 This value is overridden from its default value of '/tmp' because many
30 systems clean '/tmp' on a regular basis. Instead, it points to a path within
31 this HBase installation directory.
32
33 Running against the 'LocalFilesystem', as opposed to a distributed
34 filesystem, runs the risk of data integrity issues and data loss. Normally
35 this is mitigated by running in an environment where
36 'hbasesite.unsafe.stream.capability.enforce' to 'false' overrides this behavior,
37 permitting operation. This configuration is for the developer workstation
38 only and _should not be used in production!
39
40 See also https://hbase.apache.org/book.html#standalone\_dist
41 -->
42 <property>
43   <name>hbasesite.cluster.distributed</name>
44   <value>false</value>
45 </property>
46 <property>
47   <name>hbasesite.tmp.dir</name>
48   <value>/tmp</value>
49 </property>
50 <property>
51   <name>hbasesite.unsafe.stream.capability.enforce</name>
52   <value>false</value>
53 </property>
54 <property>
55   <name>hbasesite.rootdir</name>
56   <value>C:/hbase-2.2.5/hbase</value>
57 </property>
58 <property>
59   <name>hbasesite.zookeeper.property.dataDir</name>
60   <value>C:/hbase-2.2.5/zookeeper</value>
61 </property>
62 <property>
63   <name>hbasesite.zookeeper.quorum</name>
64   <value>localhost</value>
65 </property>
66 </configuration>
67

```

oSXe Markup Language file

Step 6: Setup the Environment variable for HBASE_HOME



Step 7: Open terminal and direct to the hbase bin path.

- Type “start-hbase.cmd” to start the hbase
- Type “hbase version” to verify the hbase server is running

```
Microsoft Windows [Version 10.0.22631.3447]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd..

C:\Windows>cd C:\hbase-2.2.5\bin

C:\hbase-2.2.5\bin>start-hbase.cmd
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hbase-2.2.5/lib/client-facing-thirdparty/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/Hadoop/hadoop-3.3.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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.slf4j.impl.Log4jLoggerFactory]

C:\hbase-2.2.5\bin>
```

Step 8: Type “hbase shell” to open the hbase shell.

Create a table by using “create” command

```
at org.jruby.runtime.callsite.CachingCallSite.cacheAndCall(CachingCallSite.java:338)
at org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.java:163)
at org.jruby.ir.interpreter.InterpreterEngine.processCall(InterpreterEngine.java:314)
at org.jruby.ir.interpreter.StartupInterpreterEngine.interpret(StartupInterpreterEng
at org.jruby.ir.interpreter.InterpreterEngine.interpret(InterpreterEngine.java:77)
at org.jruby.internal.runtime.methods.MixedModeIRMethod.INTERPRET_METHOD(MixedModeIR
at org.jruby.internal.runtime.methods.MixedModeIRMethod.call(MixedModeIRMethod.java:
at org.jruby.internal.runtime.methods.DynamicMethod.call(DynamicMethod.java:192)
at org.jruby.runtime.callsite.CachingCallSite.cacheAndCall(CachingCallSite.java:318)
at org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.java:131)
at C_3a_.Users.as5272.Documents.hbase_minus_2_dot_2_dot_5.bin.hirb.invokeOther172:pr
2.2.5\bin\hirb.Rb:190)
at C_3a_.Users.as5272.Documents.hbase_minus_2_dot_2_dot_5.bin.hirb.RUBY$script(C:\Us
b:190)
at java.lang.invoke.MethodHandle.invokeWithArguments(MethodHandle.java:627)
at org.jruby.ir.Compiler$1.load(Compiler.java:95)
at org.jruby.Ruby.runScript(Ruby.java:828)
at org.jruby.Ruby.runNormally(Ruby.java:747)
at org.jruby.Ruby.runNormally(Ruby.java:765)
at org.jruby.Ruby.runFromMain(Ruby.java:578)
at org.jruby.Main.doRunFromMain(Main.java:417)
at org.jruby.Main.internalRun(Main.java:305)
at org.jruby.Main.run(Main.java:232)
at org.jruby.Main.main(Main.java:204)

Took 0.0130 seconds
'stty' is not recognized as an internal or external command,
operable program or batch file.
hbase(main):001:0> create 'demo1','check'
Created table demo1
Took 4.3780 seconds
=> Hbase::Table - demo1
hbase(main):002:0>
```

RESULT:

Thus the installation of hbase along with practice examples was executed successfully.

AIM:

To install hbase along with practice examples.

PROCEDURE:

Step 1: Install dependencies

- To use Java you will need to install Apache Ant
 - sudo apt-get install ant
- Installing required tools and libraries
 - sudo apt-get install libboost-dev libboost-test-dev libboost-program-options-dev libboost-filesystem-dev libboost-thread-dev libevent-dev automake libtool flex bison pkgconfig g++ libssl-dev
- You can check for specific requirements for each language you wish to use here:
<http://thrift.apache.org/docs/install/>

Step 2: Download Thrift

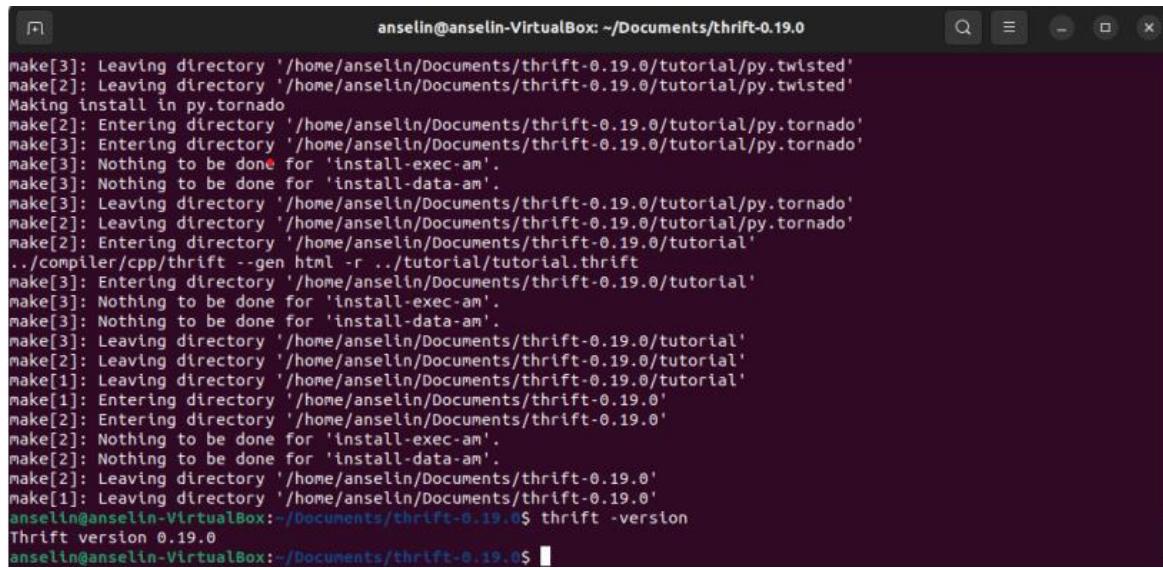
- Download Thrift: <http://thrift.apache.org/download>
- Copy the downloaded file into the desired directory and untar the file
“tar -xvf thrift-0.9.3.tar.gz”

Step 3: Configure Thrift

- For an Ubuntu linux distribution you just need to go to the thrift directory and type
 - ./bootstrap.sh
 - ./configure
- At the end of the output you should be able to see a list of all the libraries that are currently built in your system and ready to use with your desired programming languages. If a component is missing you should download the missing language and repeat the above step.
- Here <http://thrift.apache.org/docs/install/debian/> you can find all the packages you might need to support your desired language in case some of them are missing.

Step 4: Build and Run Thrift

- On the same directory run make to build Thrift
 - sudo make
- (Optional) Run the test suite if you want
 - sudo make check
- And finally you are ready to install Thrift by running
 - sudo make install
- To verify that you have successfully installed Thrift just type “thrift –version”



```
anselin@anselin-VirtualBox: ~/Documents/thrift-0.19.0
```

```
make[3]: Leaving directory '/home/anselin/Documents/thrift-0.19.0/tutorial/py.twisted'
make[2]: Leaving directory '/home/anselin/Documents/thrift-0.19.0/tutorial/py.twisted'
Making install in py.tornado
make[2]: Entering directory '/home/anselin/Documents/thrift-0.19.0/tutorial/py.tornado'
make[3]: Entering directory '/home/anselin/Documents/thrift-0.19.0/tutorial/py.tornado'
make[3]: Nothing to be done for 'install-exec-am'.
make[3]: Nothing to be done for 'install-data-am'.
make[3]: Leaving directory '/home/anselin/Documents/thrift-0.19.0/tutorial/py.tornado'
make[2]: Leaving directory '/home/anselin/Documents/thrift-0.19.0/tutorial/py.tornado'
make[2]: Entering directory '/home/anselin/Documents/thrift-0.19.0/tutorial'
.../compiler/cpp/thrift --gen html -r ./tutorial/tutorial.thrift
make[3]: Entering directory '/home/anselin/Documents/thrift-0.19.0/tutorial'
make[3]: Nothing to be done for 'install-exec-am'.
make[3]: Nothing to be done for 'install-data-am'.
make[3]: Leaving directory '/home/anselin/Documents/thrift-0.19.0/tutorial'
make[2]: Leaving directory '/home/anselin/Documents/thrift-0.19.0/tutorial'
make[1]: Entering directory '/home/anselin/Documents/thrift-0.19.0/tutorial'
make[1]: Entering directory '/home/anselin/Documents/thrift-0.19.0'
make[2]: Entering directory '/home/anselin/Documents/thrift-0.19.0'
make[2]: Nothing to be done for 'install-exec-am'.
make[2]: Nothing to be done for 'install-data-am'.
make[2]: Leaving directory '/home/anselin/Documents/thrift-0.19.0'
make[1]: Leaving directory '/home/anselin/Documents/thrift-0.19.0'
anselin@anselin-VirtualBox:~/Documents/thrift-0.19.0$ thrift -version
Thrift version 0.19.0
anselin@anselin-VirtualBox:~/Documents/thrift-0.19.0$
```

RESULT:

Thus the installation of hbase along with practice examples was executed successfully.

EX NO:7

28/03/2024

PRACTICE IMPORTING AND EXPORTING DATA FROM VARIOUS DATABASES

AIM:

Practice importing and exporting data from various databases.

PREREQUISITES:

JAVA: You need to install the Java 8 package on your system.

HADOOP: You require Hadoop latest version.

Mysql.

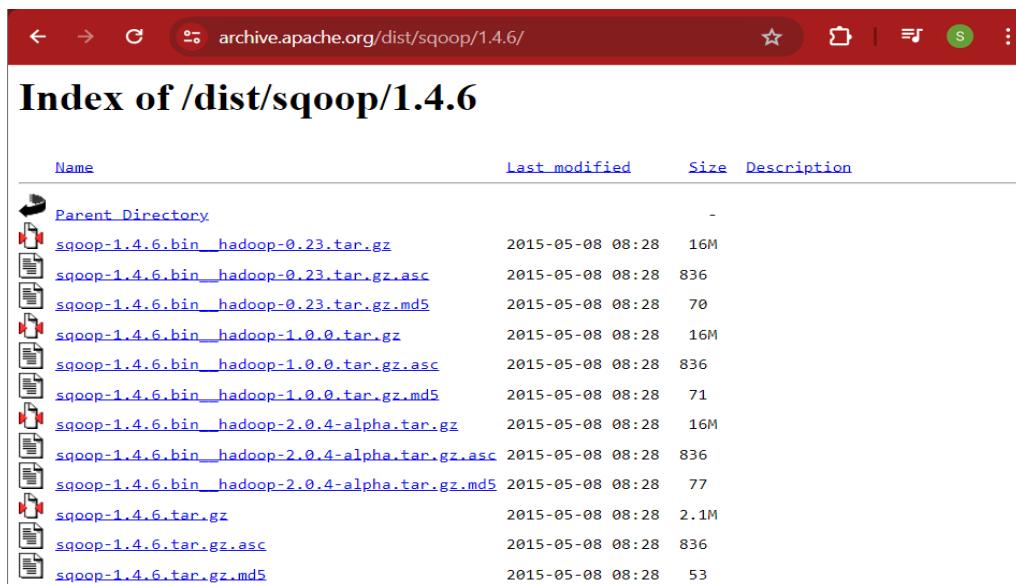
Hive.

PROCEDURE:

Step 1: Start hadoop

Step 2: Download sqoop

Download sqoop-1.4.6.bin.tar.gz file from apache website. And extract the files.

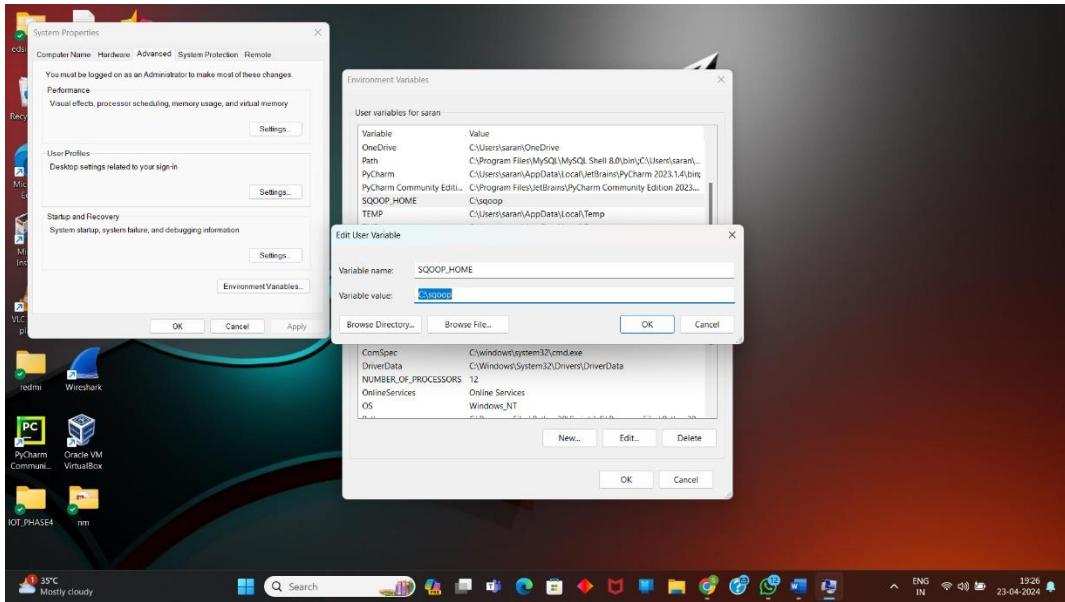


The screenshot shows a web browser window with the URL archive.apache.org/dist/sqoop/1.4.6/. The page title is "Index of /dist/sqoop/1.4.6". Below the title is a table listing various files and their details. The columns are "Name", "Last modified", "Size", and "Description".

Name	Last modified	Size	Description
 Parent Directory		-	
 sqoop-1.4.6.bin_hadoop-0.23.tar.gz	2015-05-08 08:28	16M	
 sqoop-1.4.6.bin_hadoop-0.23.tar.gz.asc	2015-05-08 08:28	836	
 sqoop-1.4.6.bin_hadoop-0.23.tar.gz.md5	2015-05-08 08:28	70	
 sqoop-1.4.6.bin_hadoop-1.0.0.tar.gz	2015-05-08 08:28	16M	
 sqoop-1.4.6.bin_hadoop-1.0.0.tar.gz.asc	2015-05-08 08:28	836	
 sqoop-1.4.6.bin_hadoop-1.0.0.tar.gz.md5	2015-05-08 08:28	71	
 sqoop-1.4.6.bin_hadoop-2.0.4-alpha.tar.gz	2015-05-08 08:28	16M	
 sqoop-1.4.6.bin_hadoop-2.0.4-alpha.tar.gz.asc	2015-05-08 08:28	836	
 sqoop-1.4.6.bin_hadoop-2.0.4-alpha.tar.gz.md5	2015-05-08 08:28	77	
 sqoop-1.4.6.tar.gz	2015-05-08 08:28	2.1M	
 sqoop-1.4.6.tar.gz.asc	2015-05-08 08:28	836	
 sqoop-1.4.6.tar.gz.md5	2015-05-08 08:28	53	

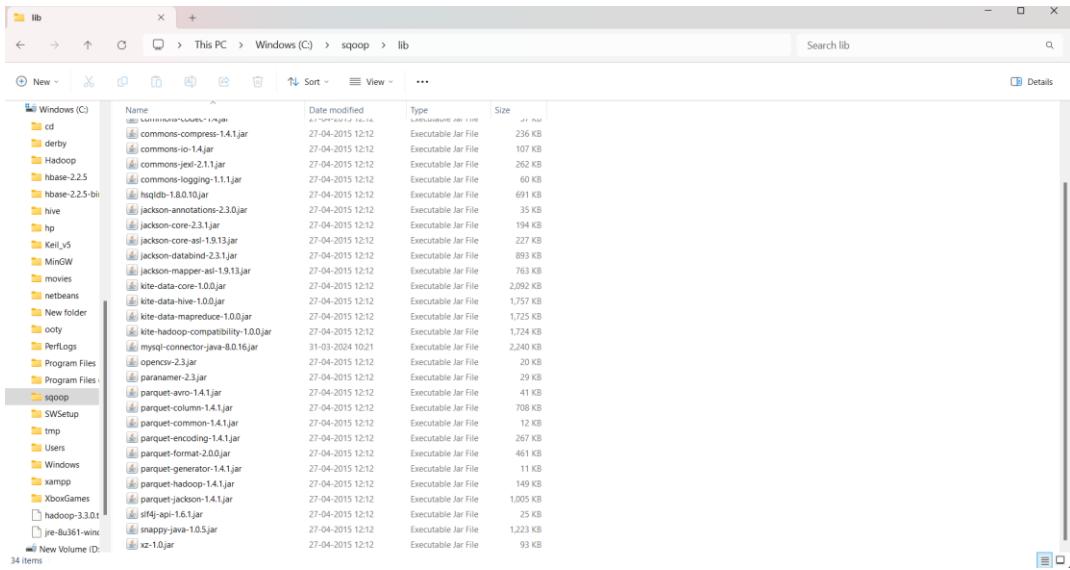
Step 3: Configure user variables

Variable name: SQQOUP_HOME



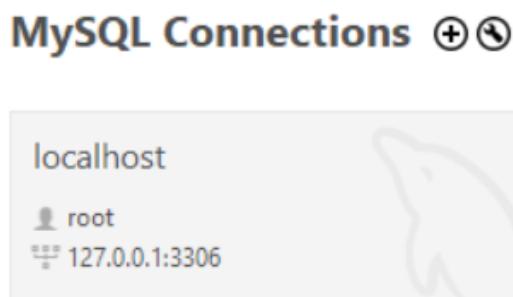
Step 4: download mysql-connector

- download mysql-connector-java-5.1.42.zip
- extract the file and copy the file and paste it in the path “C:\scoop\lib”

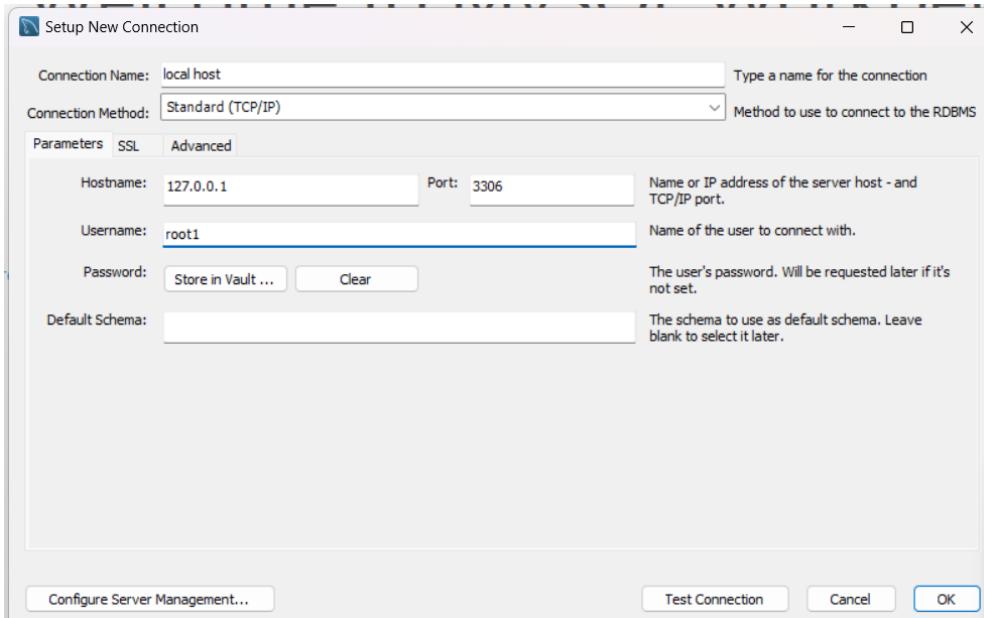


Step 5: Download mysql work bench.

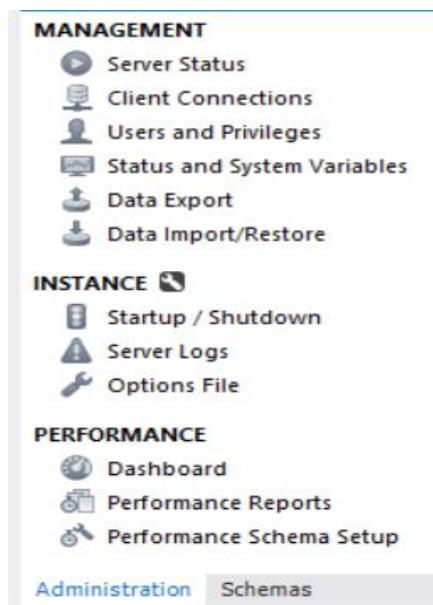
- To create a new connection click on to the plus icon.



- Configure as the following to set up the connection.



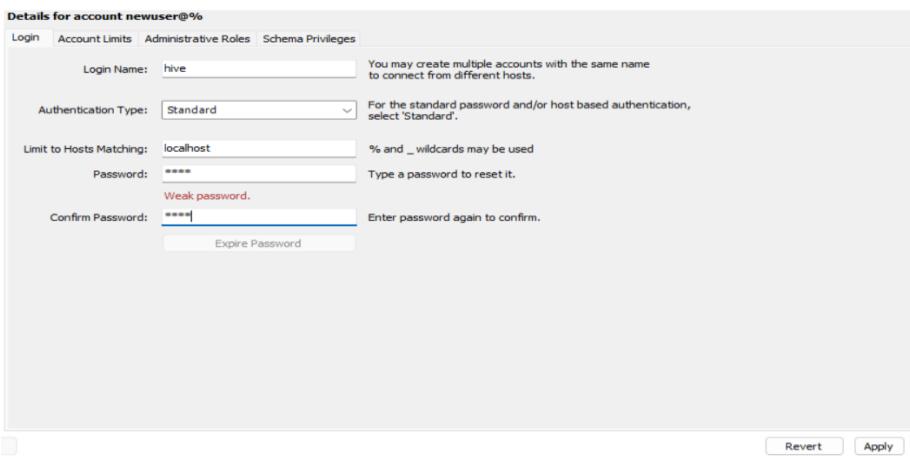
- Click the created localhost and in the left side go to administration setting and go to users and privileges



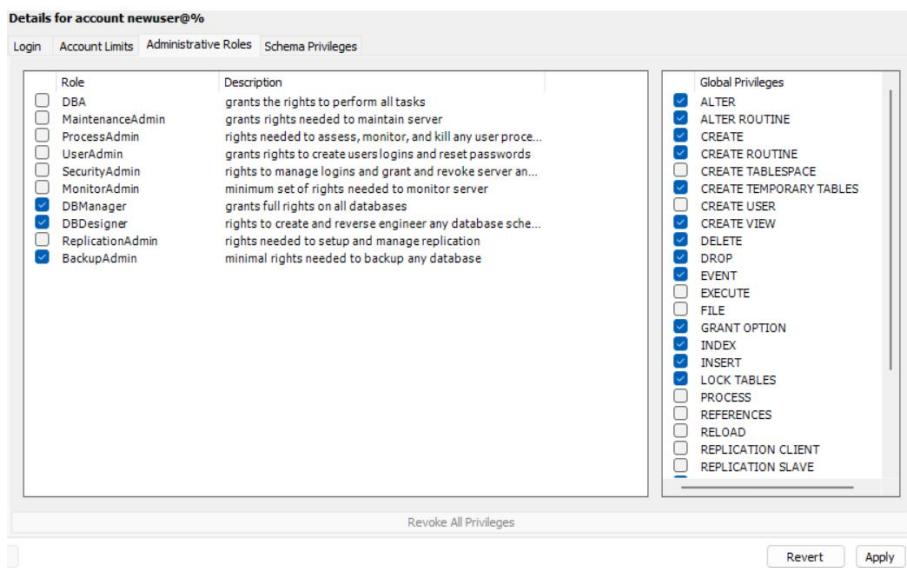
- Go to add account.



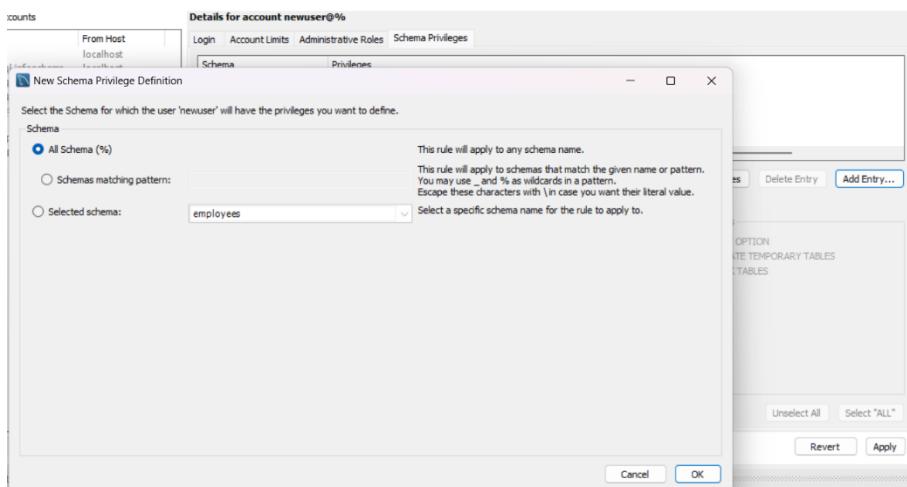
- Configure the following.



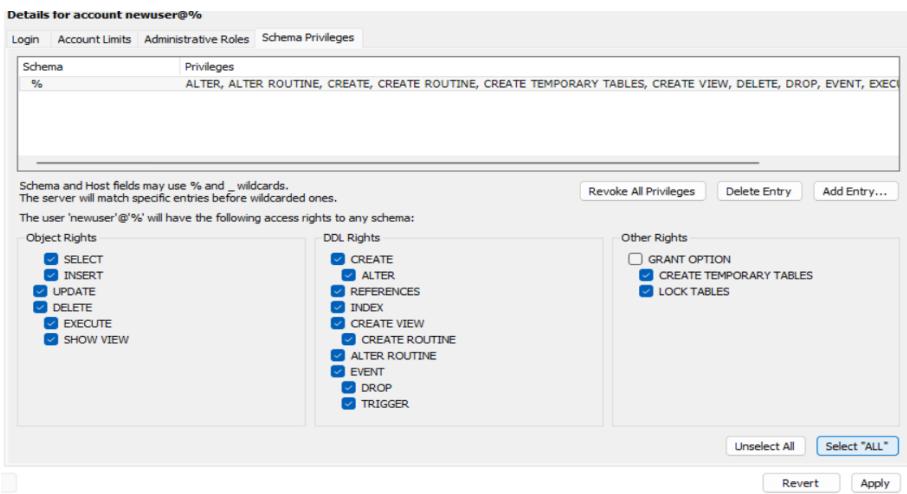
- In administrative roles select the following DB manager ,DB designer and Backup Admin.



- In the schema privileges select the employee data set an apply changes.



- In addition, select all and apply changes

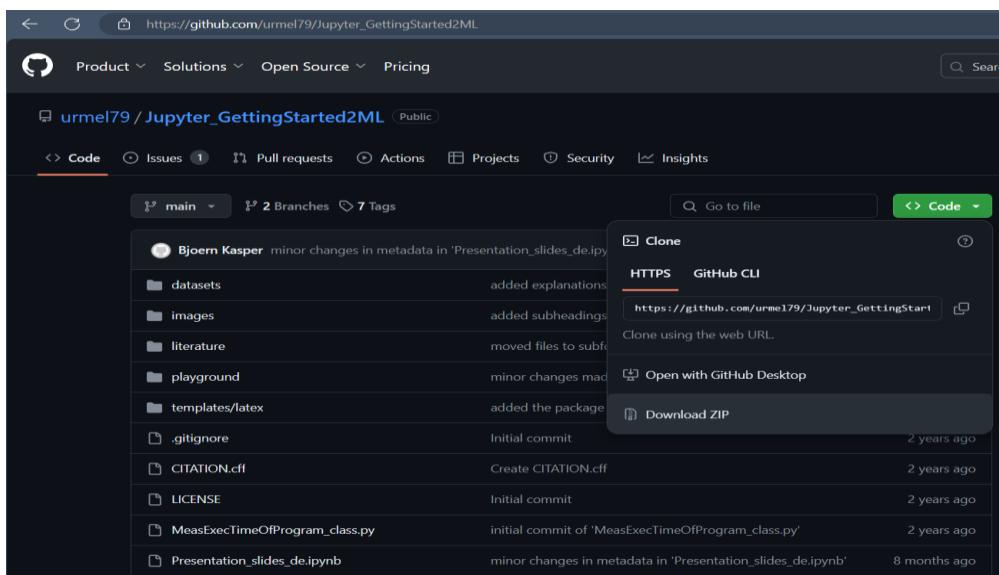


Step 6: Repeat step 5 for installation of Sqoop.

Step 7: To add the data set into the mysql database.

Do the following:

- Go to the following github link : [GitHub - urmel79/Jupyter_GettingStarted2ML: This Getting Started Tutorial systematically demonstrates the typical ML work process step-by-step using the powerful and performant Support Vector Classifier \(SVC\) and the beginner-friendly Iris Dataset. Furthermore, the selection of the correct SVC kernel and its parameters are described and their effects on the classification result are shown.](https://github.com/urmel79/Jupyter_GettingStarted2ML)
- Download the data set.



- Extract the zip file.
- Copy all the db master contents.

Name	Date modified	Type	Size
images	27-08-2023 12:16 AM	File folder	
sakila	27-08-2023 12:16 AM	File folder	
Changelog	27-08-2023 12:16 AM	File	1 KB
employees	27-08-2023 12:16 AM	SQL Text File	5 KB
employees_partitioned	27-08-2023 12:16 AM	SQL Text File	7 KB
employees_partitioned_5.1	27-08-2023 12:16 AM	SQL Text File	8 KB
load_departments.dump	27-08-2023 12:16 AM	DUMP File	1 KB
load_dept_emp.dump	27-08-2023 12:16 AM	DUMP File	13,829 KB
load_dept_manager.dump	27-08-2023 12:16 AM	DUMP File	2 KB
load_employees.dump	27-08-2023 12:16 AM	DUMP File	17,308 KB
load_salaries1.dump	27-08-2023 12:16 AM	DUMP File	38,874 KB
load_salaries2.dump	27-08-2023 12:16 AM	DUMP File	38,874 KB
load_salaries3.dump	27-08-2023 12:16 AM	DUMP File	38,165 KB
load_titles.dump	27-08-2023 12:16 AM	DUMP File	21,200 KB
objects	27-08-2023 12:16 AM	SQL Text File	5 KB
README	27-08-2023 12:16 AM	Markdown Source ...	5 KB
show_elapsed	27-08-2023 12:16 AM	SQL Text File	1 KB
sql_test	27-08-2023 12:16 AM	SH Source File	2 KB

- Go to the following folder and past the contents C:\Program Files\MySQL\MySQL Server 8.0\bin.
- After the above task open cmd in the same folder and type the below cmd.

```
C:\Program Files\MySQL\MySQL Server 8.0\bin>mysql -u root -t -p
Enter password: ****
```

```
mysql> source employees.sql
Query OK, 8 rows affected (0.13 sec)
```

- After this wait for the configuration message.

```
+-----+
| data_load_time_diff |
+-----+
| 00:01:07           |
+-----+
1 row in set (0.01 sec)
```

- Then refresh the mysql connector page



- It will import the data set into the mysql.

Step 8: Open **mysql command line** and grant all privileges to user hive and user sqoop.

Use the following cmd's.

```

MySQL 8.0 Command Line Cli  +  ▾
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 15
Server version: 8.0.36 MySQL Community Server - GPL

Copyright (c) 2000, 2024, 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> grant all privileges on employees.* to 'hive'@'localhost';
Query OK, 0 rows affected (0.03 sec)

mysql> grant all privileges on employees.* to 'sqoop'@'localhost';
Query OK, 0 rows affected (0.01 sec)

mysql> |

```

Step 9: run sqoop

- Open sqoop”c:\sqoop\bin\” in terminal
- Check the database “sqoop list-databases – connect jdbc:mysql://localhost/ --username sqoop –password sqooppassword”.

```

C:\sqoop\bin>sqoop list-databases -- connect jdbc:mysql://localhost/ -- username sqoop -password sqooppassword
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.
24/04/25 23:34:16 INFO sqoop: Running Sqoop version: 1.4.5
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Error parsing arguments for list-databases:
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: -
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: connect
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: jdbc:mysql://localhost/
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: --
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: username
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: sqoop
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: &password
24/04/25 23:34:16 ERROR tool.BaseSqoopTool: Unrecognized argument: sqooppassword

Try --help for usage instructions.

C:\sqoop\bin>

```

Step 10: **import data**

Import data from mysql server into warehouse by using “sqoop import –connect jdbc:mysql://localhost/employees –table departments –username hive –password hivepassword –hive-import –fields-terminated-by “,”

```
C:\sqoop\bin>sqoop import --connect jdbc:mysql://localhost/employees -table departments -username hive -password hivelpassword --hive-import --fields-terminated-by "
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: ACCUMULO_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.
24/04/25 23:35:47 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Error parsing arguments for import:
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: --connect
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: jdbc:mysql://localhost/employees
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: --table
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: departments
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: --username
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: hive
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: --password
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: hivelpassword
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: --hive-import
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: --fields-terminated-by "
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: "
24/04/25 23:35:47 ERROR tool.BaseSqoopTool: Unrecognized argument: "

Try --help for usage instructions.
```

Step 11: Run hive.

- Open terminal and direct to the path “c:/hive/bin” and run hive by using “hive” command.

Step 12: export data

- Check data in hive by using “**show table;**” command
 - Check data in department table by using “**select * from departments;**”

```
option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Connected to: Apache Hive (version 2.0.0)
Driver: Hive JDBC (version 2.0.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 2.0.0 by Apache Hive
hive> show tables;
OK
departments
1 row selected (1.466 seconds)
hive> select * from departments;
OK
d001 Marketing
d002 Finance
d003 Human Resources
d004 Production
d005 Development
d006 Quality Management
d007 Sales
d008 Research
d009 Customer Service
9 rows selected (1.791 seconds)
hive>
```

RESULT:

Thus practice importing and exporting data from various databases was executed successfully.