

INFANT JESUS COLLEGE OF ENGINEERING

KAMARAJAR NAGAR

KEELAVALLANADU-628 851



CS8711 – Cloud Computing Laboratory

NAME : -----

REGISTER NUMBER : -----



INFANT JESUS COLLEGE OF ENGINEERING

KAMARAJAR NAGAR

KEELAVALLANADU-628 851



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NAME : -----

REGISTER NUMBER : -----ROLL NO: -----

BRANCH : -----

YEAR : -----

Bonafide Record of work done in the **CS 8711 Cloud Computing Laboratory** at Infant Jesus College of Engineering, Keelavallanadu during the year **2020-2021**

STAFF INCHARGE

H.O.D

Submitted for the Practical Examination held on -----

at Infant Jesus College of Engineering, Keelavallanadu.

INTERNAL EXAMINER

EXTERNAL EXAMINER

SI.No	Date	Experiment	Marks	Signature
1		Installation of Virtualbox		
2		Install a C compiler in the virtual machine		
3		C Program to implement Bubble Sort		
4		Install Google App Engine		
5		Use GAE launcher to launch the web applications		
6		Launching virtual machine using trystack		
7		Transfer a files from one virtual machine to another virtual machine		
8		Installing Hadoop single node cluster		
9		Word Count Program using Hadoop		

Ex.No : 1

Installation of Virtualbox

Date :

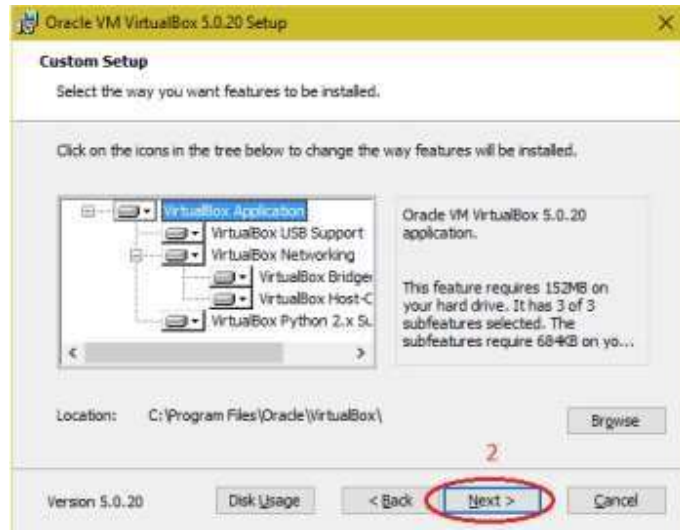
Aim:

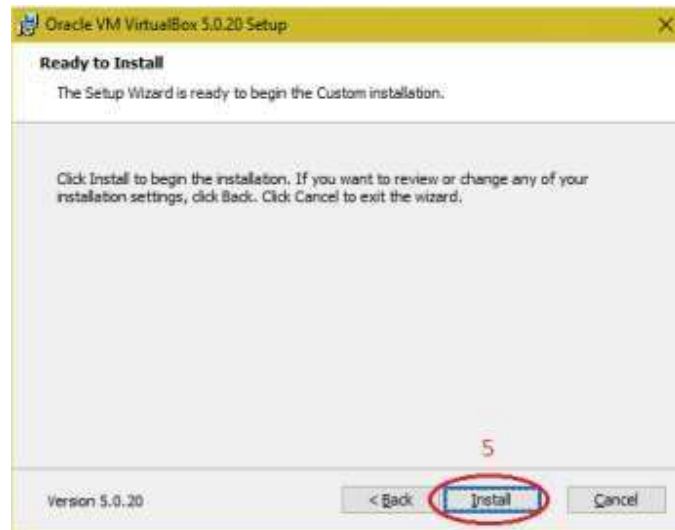
To Install VirtualBox with different flavors of Linux or windows OS on top of windows7 or 8

Procedure:

1. Download VirtualBox setup for installation from the below link
<https://www.virtualbox.org/wiki/Downloads>
2. Run the Virtual Box setup.

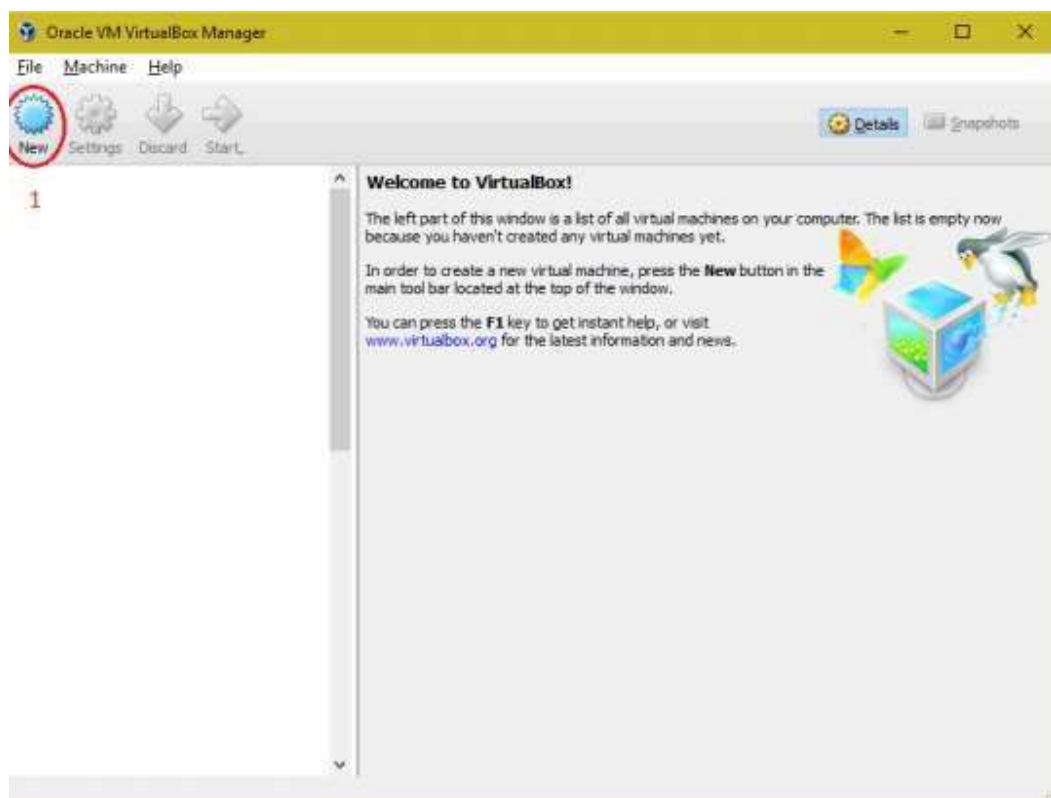






Installing "Ubuntu" as virtual machine in "Oracle VM VirtualBox"

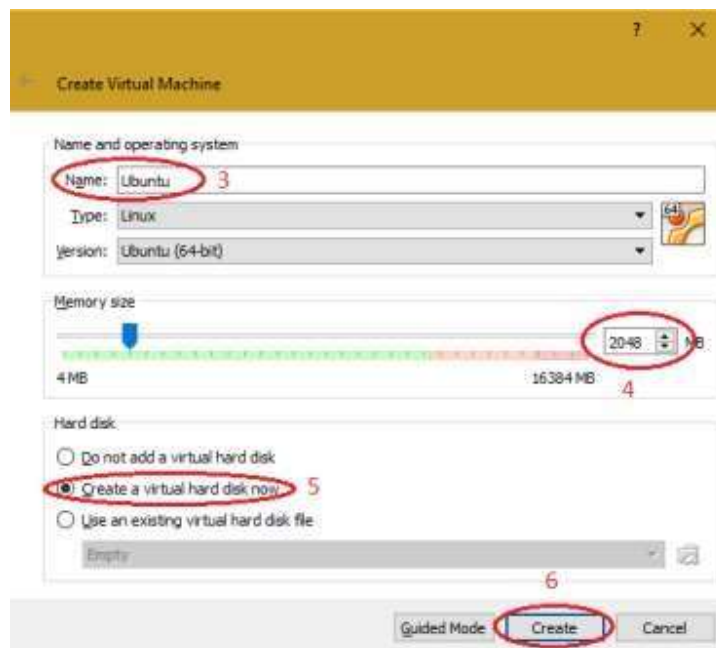
1. Open "Oracle VM VirtualBox Manager".



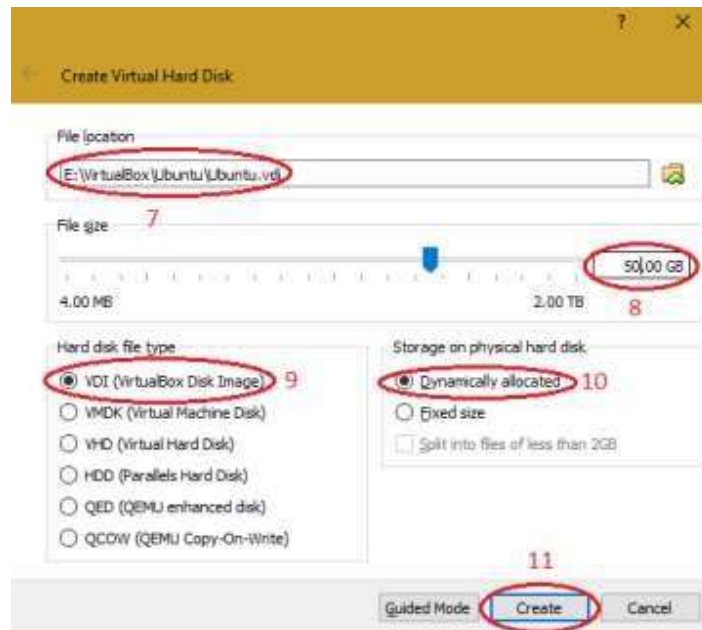
2. Click on "New" button



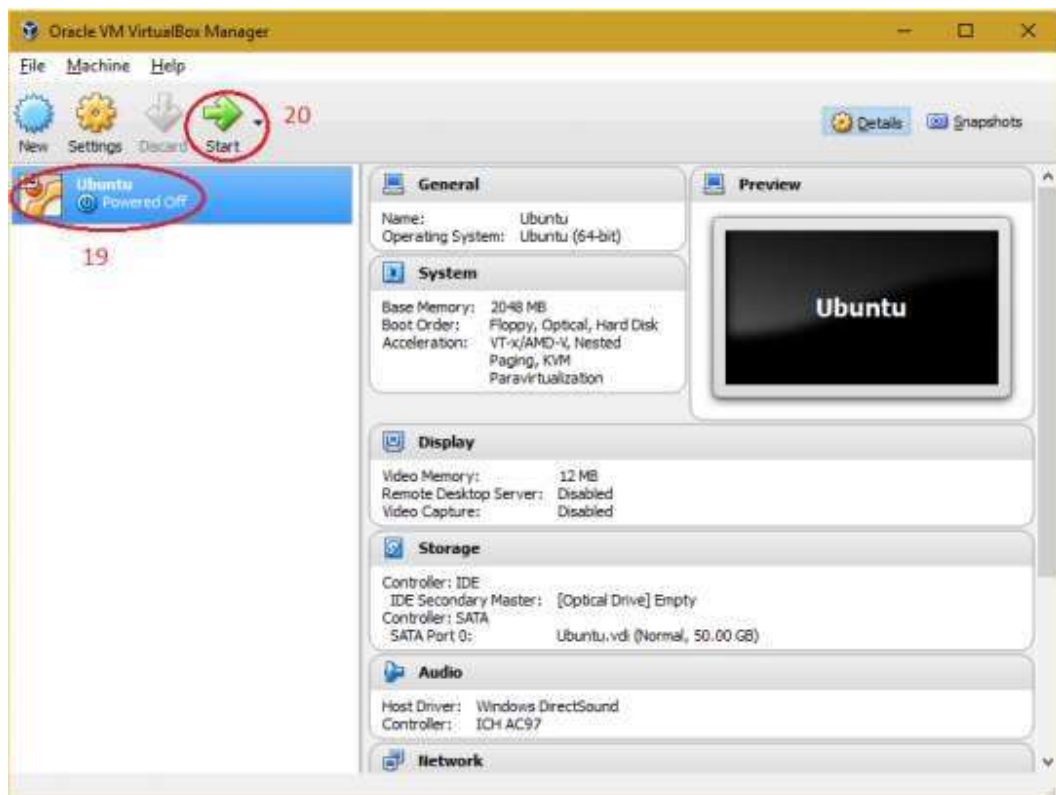
3. Provide the name and operating system information for virtual machine



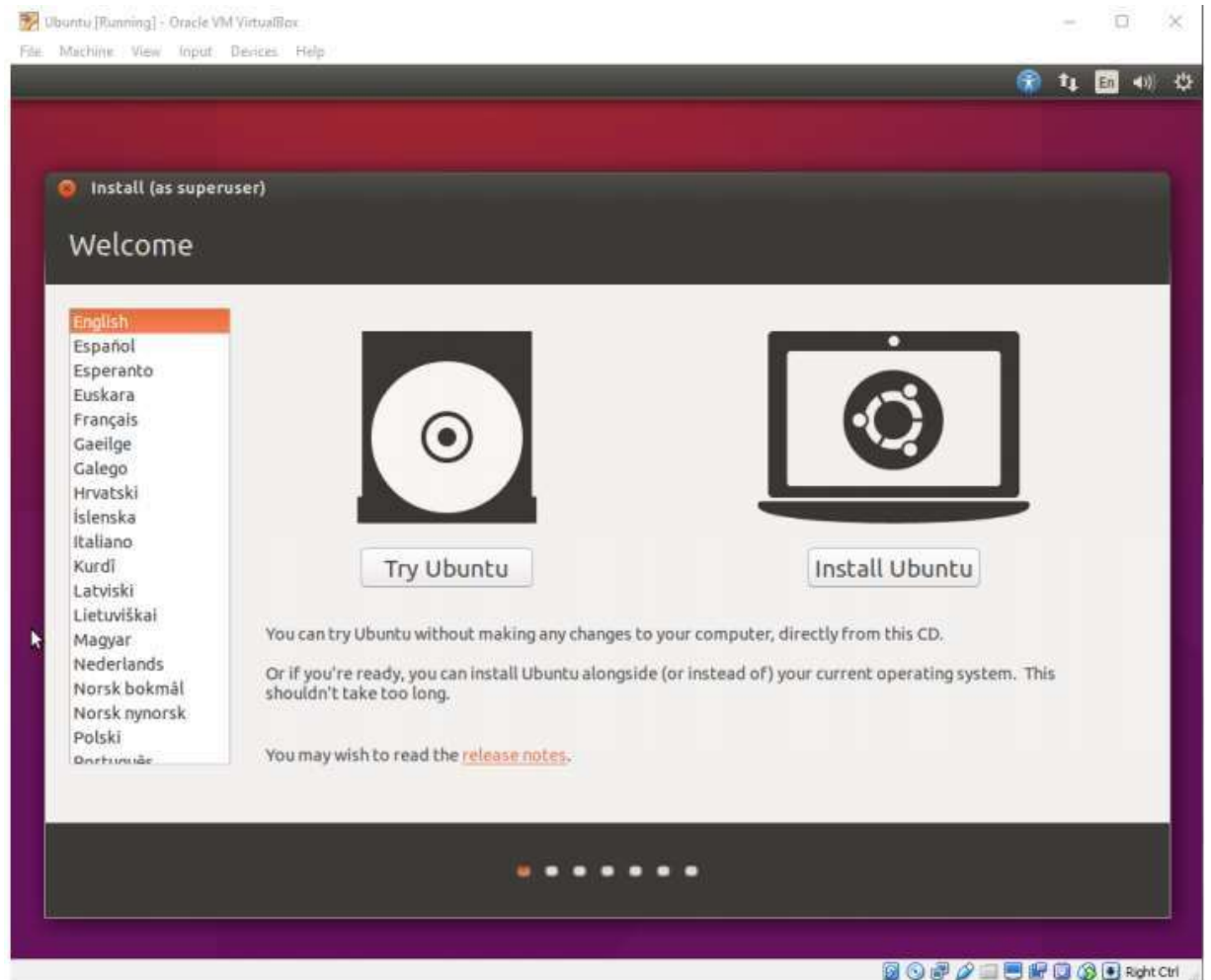
4. Select the path for the virtual hard disk and click on "Create" button.



5. Select the created virtual machine and click on "Start" button



6. Proceed with the installation of operating system in virtual machine



Result:

Thus, the Virtual Box is installed with ubuntu OS Successfully

Ex.No : 2 Install a C compiler in the virtual machine

Date :

Aim:

To Install a C compiler in the virtual machine created using virtual box and execute Simple Programs

Procedure:

Installing gcc in Ubuntu

1. Start by updating the packages list `sudo apt update`
2. Install the build-essential package by typing `sudo apt install build-essential`

The command installs a bunch of new packages including gcc

3. Install the manual pages about using GNU/Linux for development `sudo apt-get install manpages-dev`
4. To validate that the GCC compiler is successfully installed `gcc -version` gcc (Ubuntu 7.4.0-1ubuntu1~18.04) 7.4.0

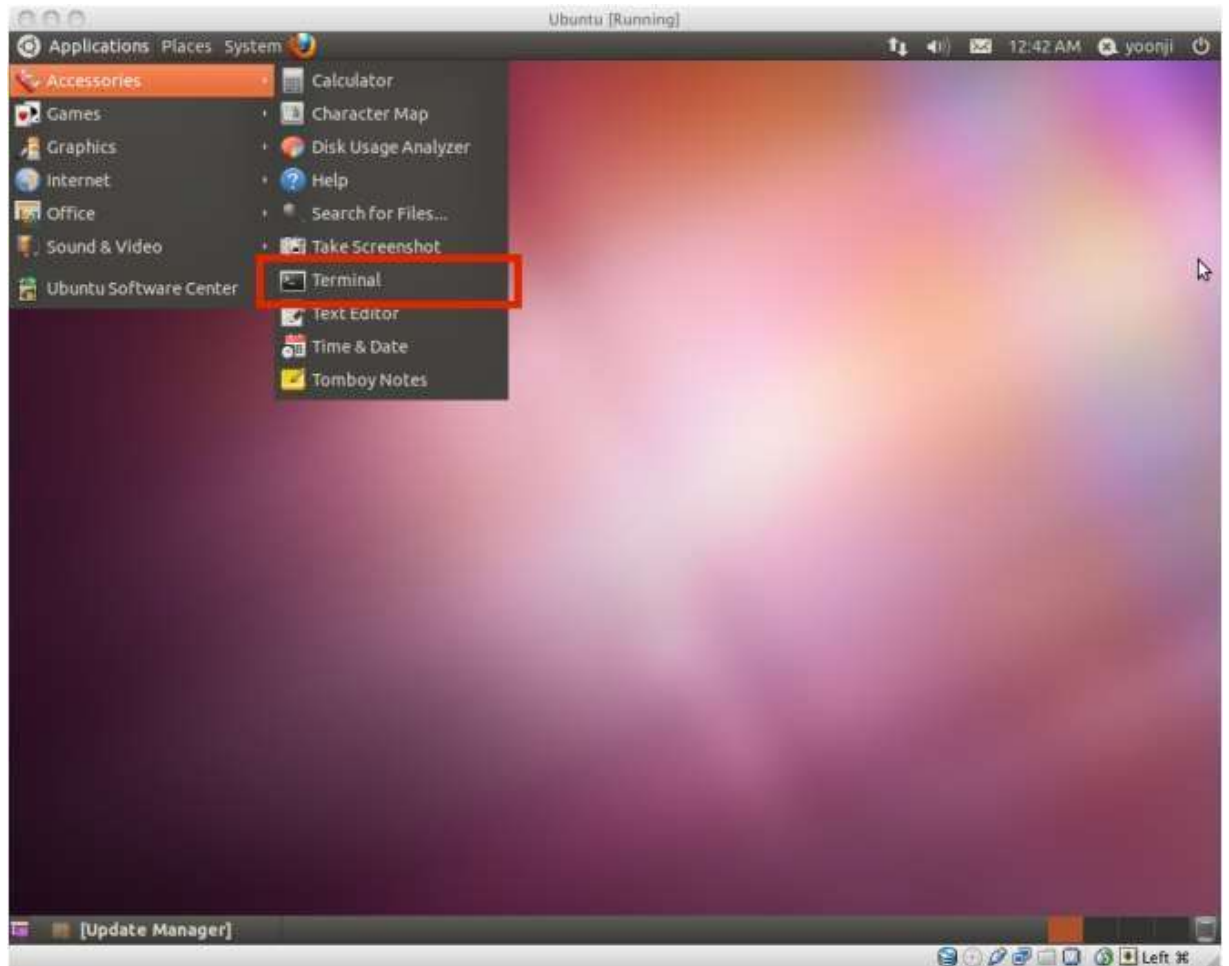
Copyright (C) 2017 Free Software Foundation, Inc.

This is free software; see the source for copying conditions. There is NO

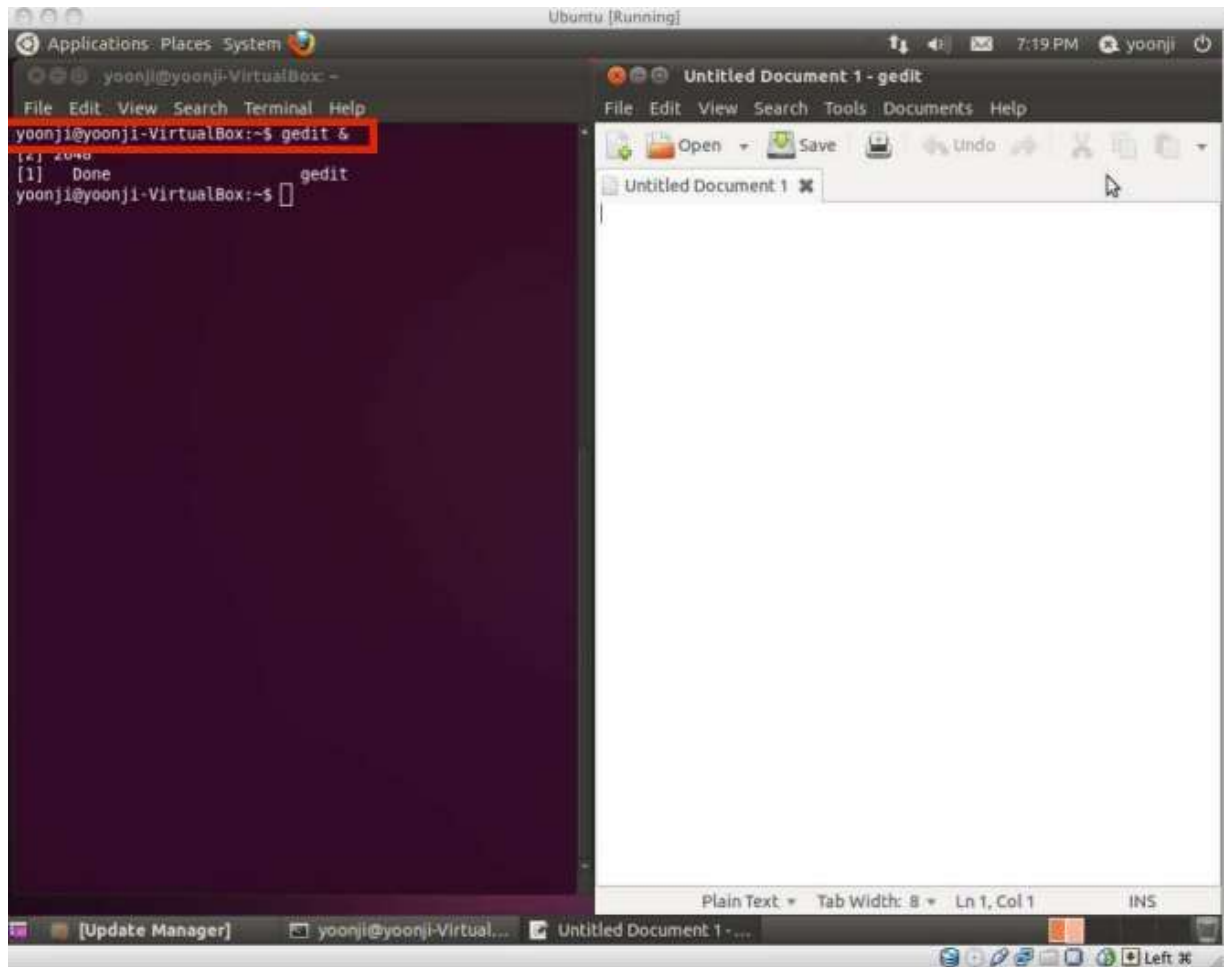
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

Running C Program

1. Open Terminal (Applications-Accessories-Terminal)



2. Open gedit by typing "gedit &" on terminal



3. Type the C Program
4. Compilation gcc filename.c
5. Execution ./a.out

Result:

Thus the C Compiler installed successfully in the Virtual Machine

Ex.No : 3 C Program to implement Bubble Sort

Date :

Aim:

To Write a C Program to implement Bubble Sort Algorithm:

Program:

```
#include <stdio.h>
int main() {
    int array[100], n, c, d, swap;
    printf("Enter number of elements\n");
    scanf("%d", &n);
    printf("Enter %d integers\n", n);
    for (c = 0; c < n; c++) scanf("%d", &array[c]);
    for (c = 0; c < n - 1; c++)
    {
        for (d = 0; d < n - c - 1; d++)
        {
            if (array[d] > array[d+1]) /* For decreasing order use < */
            {
                swap = array[d];
                array[d] = array[d+1];
                array[d+1] = swap;
            }
        }
    }
    printf("Sorted list in ascending order:\n");
    for (c = 0; c < n; c++) printf("%d\n", array[c]);
    return 0;
}
```

Output:

```
$ gcc bub.c
```

```
$ ./a.out
```

```
Enter number of elements
```

```
5
```

```
Enter 5 integers
```

```
23
```

```
34
```

```
12
```

```
10
```

```
56
```

```
Sorted list in ascending order:
```

```
10
```

```
12
```

```
23
```

```
34
```

```
56
```

Result:

Thus the C Program to implement Bubble sort was successfully executed

Ex.No : 4 Install Google App Engine

Date :

Aim:

To Install Google App Engine and create a simple hello world application using python Procedure:

1. Install Python
2. Download Google App Engine SDK from the Following link

<https://cloud.google.com/appengine/downloads>

Install an SDK for App Engine

Set up your computer for developing, deploying, and managing your apps in App Engine.

The instructions for setting up your computer differ by App Engine environment, either the standard environment or the flexible environment. [For help with choosing, see a summary of differences between the two environments.](#)

Standard environment



3. Install Google App Engine SDK
4. Create a Folder "Sampleapplication"
5. Create a python program to print hello world and save the file as index.py

```
print ('Content-Type: text/plain')  
print (' ')  
print ('Hello World')
```
6. Create app.yaml file and save the following code inside the folder

```
runtime: python38  
api_version: 1  
threadsafe: false
```

```
handlers: - url:
```

```
/  script:
```

```
index.py
```

7. Open the Google App Engine SDK Terminal and execute the following command

```
> goolge-cloud-sdk\bin\dev_appserver.py "...\\Sampleapplication"
```

8. Open the localhost:8080 to see the output

Result:

Thus the Google App Engine is installed and hello world program is successfully executed

Ex.No : 5 Use GAE launcher to launch the web applications

Date :

Aim:

To Launch simple web applications using Google App Engine

Program:

temperature.py

```
import webapp2
def convert_temp(cel_temp):
    if cel_temp == "":
        return ""
    try:
        far_temp = float(cel_temp) * 9 / 5 + 32
        far_temp = round(far_temp, 3)
        cel_temp = self.request.get("cel_temp")
        far_temp = convert_temp(cel_temp)
        self.response.headers["Content-Type"] = "text/html"
        self.response.write("""
<html>
    <head>
        <title>Temperature Converter</title>
    </head>
    <body>
        <form action="/" method="get">
            Celsius temperature: <input type="text" name="cel_temp" value={}>
            <input type="submit" value="Convert"><br>
            Fahrenheit temperature: {}
        </form>
    </body>
</html>""".format(cel_temp, far_temp))
routes = [('/', MainPage)]
my_app = webapp2.WSGIApplication(routes, debug=True)
```

app.yaml

```
application: temperature-converter
version: 1 runtime: python27
api_version: 1 threadsafe: false
handlers: - url: /*    script:
temperature.my_app
```

Result:

Thus the web application to convert from Celsius to Fahrenheit is successfully launched using GAE

EX.NO :6 Launching virtual machine using trystack

Date :

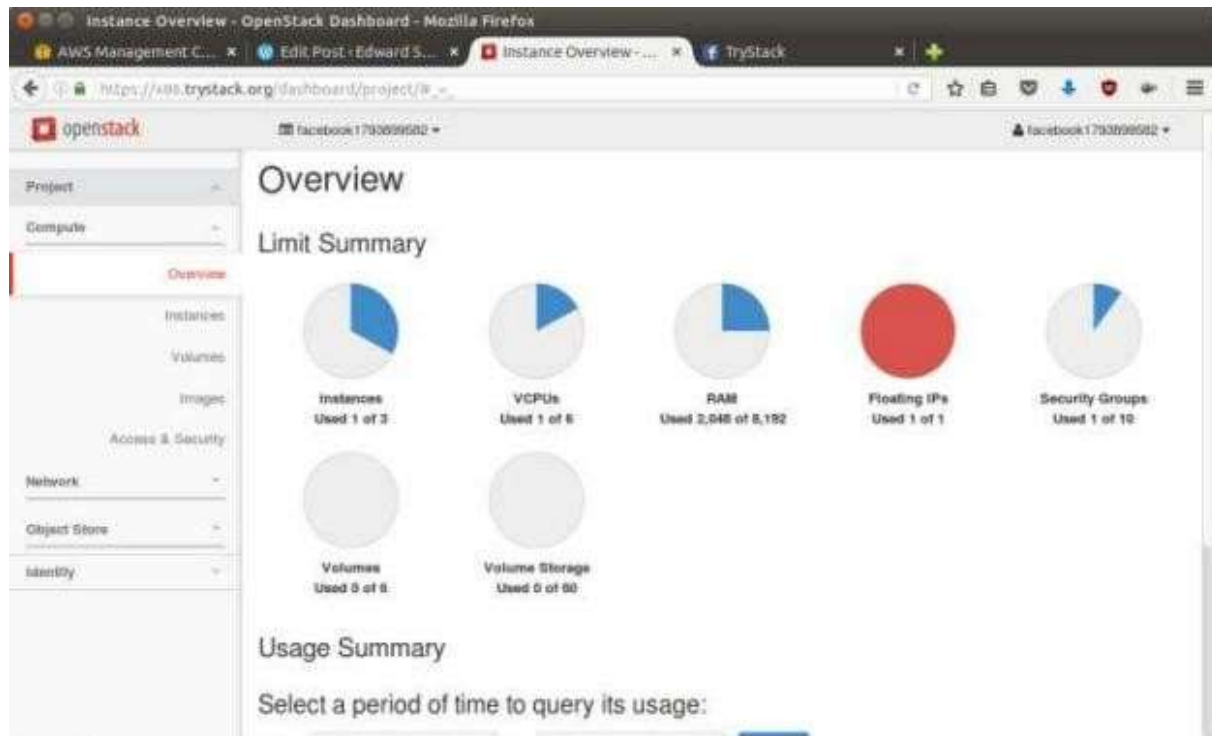
Aim:

To Find a procedure to launch virtual machine using trystack
(Online Openstack Demo Version) Procedure:

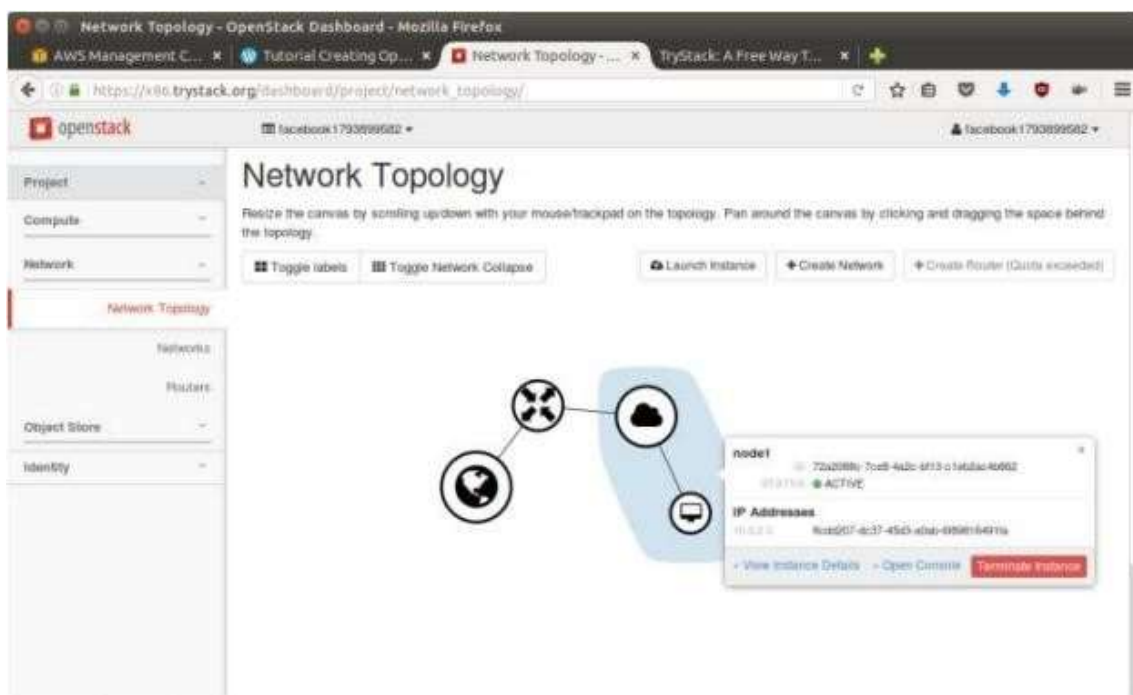
OpenStack is an open-source software cloud computing platform. OpenStack is primarily used for deploying an infrastructure as a service (IaaS) solution like Amazon Web Service (AWS).

In order to try OpenStack in TryStack, you must register yourself by joining TryStack Facebook Group. The acceptance of group needs a couple days because it's approved manually. After you have been accepted in the TryStack Group, you can log in TryStack.





OpenStack Compute Dashboard



Network topology

As you see from the image above, the instance will be connected to a local network and the local network will be connected to internet.

Creating Network

1. Go to Network > Networks and then click Create Network.
2. In Network tab, fill Network Name for example internal and then click Next.
3. In Subnet tab,
 1. Fill Network Address with appropriate CIDR, for example 192.168.1.0/24. Use privatenetwork CIDR block as the best practice.
 2. Select IP Version with appropriate IP version, in this case IPv4.
 3. Click Next.
4. In Subnet Details tab, fill DNS Name Servers with 8.8.8.8 (Google DNS) and then click Create.

Creating Instance

1. Go to Compute > Instances and then click Launch Instance.
2. In Details tab,
 1. Fill Instance Name, for example Ubuntu 1.
 2. Select Flavor, for example m1.medium.
 3. Fill Instance Count with 1.
 4. Select Instance Boot Source with Boot from Image.
 5. Select Image Name with Ubuntu 14.04 amd64 (243.7 MB) if you want install Ubuntu14.04 in your virtual machine.
3. In Access & Security tab,
 1. Click [+] button of Key Pair to import key pair. This key pair is a public and private key that we will use to connect to the instance from our machine.
2. In Import Key Pair dialog,

1. Fill Key Pair Name with your machine name
2. Fill Public Key with your SSH public key (usually is in ~/.ssh/id_rsa.pub). See description in Import Key Pair dialog box for more information.
3. Click Import key pair.
3. In Security Groups, mark/check default.
4. In Networking tab,
 1. In Selected Networks, select network that have been created in Step 1, for example internal.
5. Click Launch.
6. If you want to create multiple instances, you can repeat step 1-5.

Result:

Thus, virtual machine using trystack was Successfully Launched

Ex.No: 7 Transfer a files from one virtual machine to another virtual machine

Date :

Aim:

To Find a procedure to transfer the files from one virtual machine to another virtual machine

Procedure:

Step 1: Start All the Virutal Machines

Step 2: Set Administrator Password for all the OS installed in VMs

Step 3: Note all the IP Address of all the VMs

Step 4: Create file in all the operating system and share that file by selecting properties of the file

Step 5: Check the connection by typing ping ipaddress

Step 6: access the shared file by typing [\\ipaddress](#) in the search option and by typing the administrator password

Result:

Thus the files from one virtual machine to another is successfully transferred

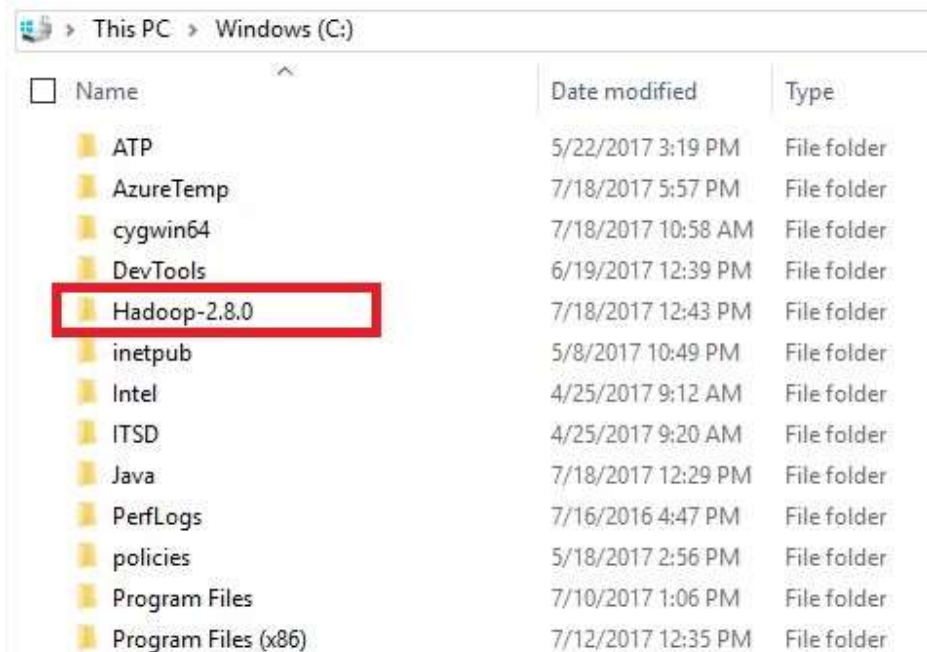
Ex.No. : 8 Installing Hadoop single node cluster

Date :

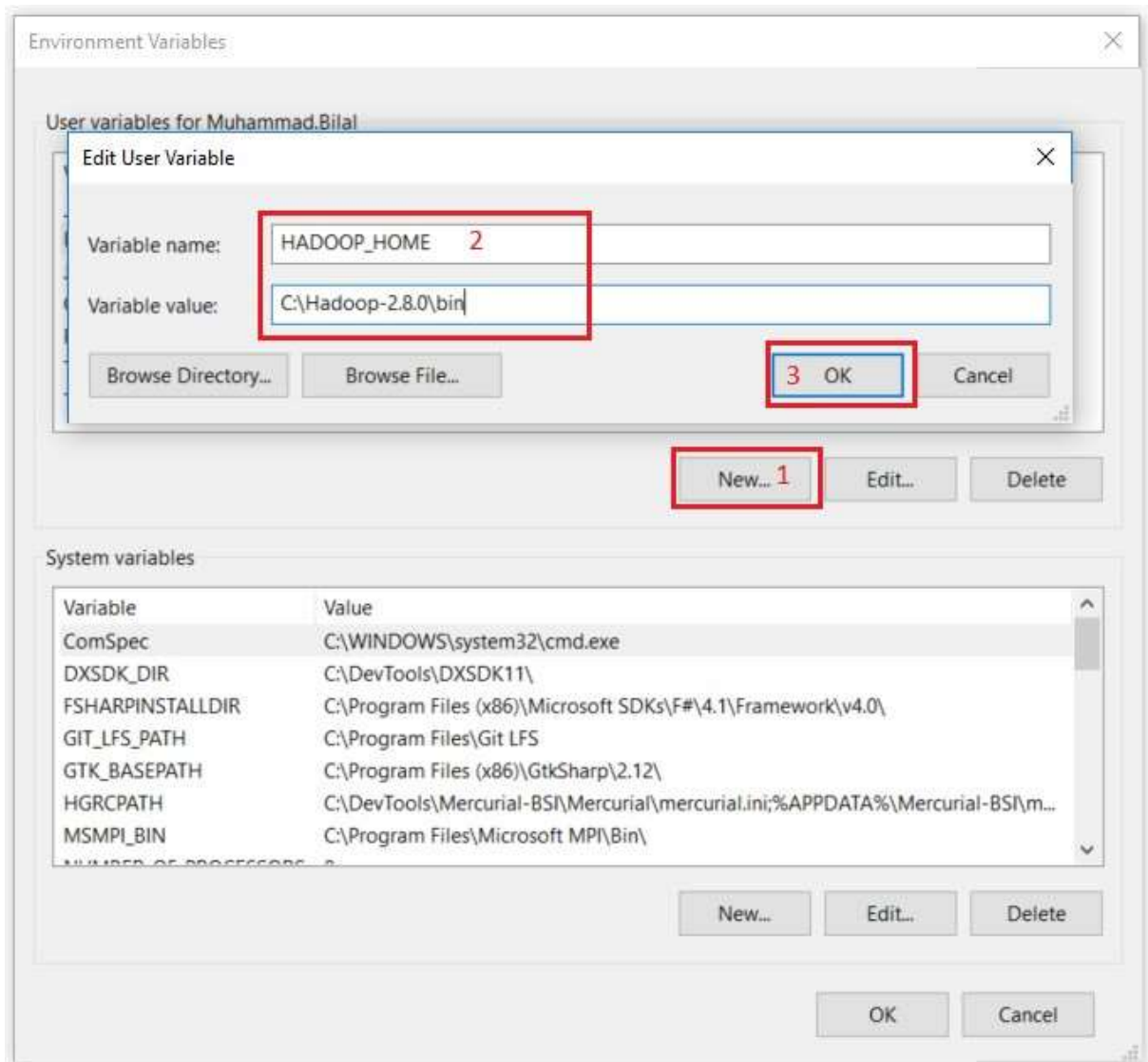
Aim:

To Install Hadoop single node cluster Procedure:

1. Download the Hadoop 2.8.0 by using the link
<https://hadoop.apache.org/docs/r2.8.0/index.html>
2. Check either Java 1.8.0 is already installed on your system or not, use “Javac -version” to check.
3. If Java is not installed on your system then first install java under “C:\JAVA”
4. Extract file Hadoop 2.8.0.tar.gz or Hadoop-2.8.0.zip and place under “C:\Hadoop-2.8.0”.

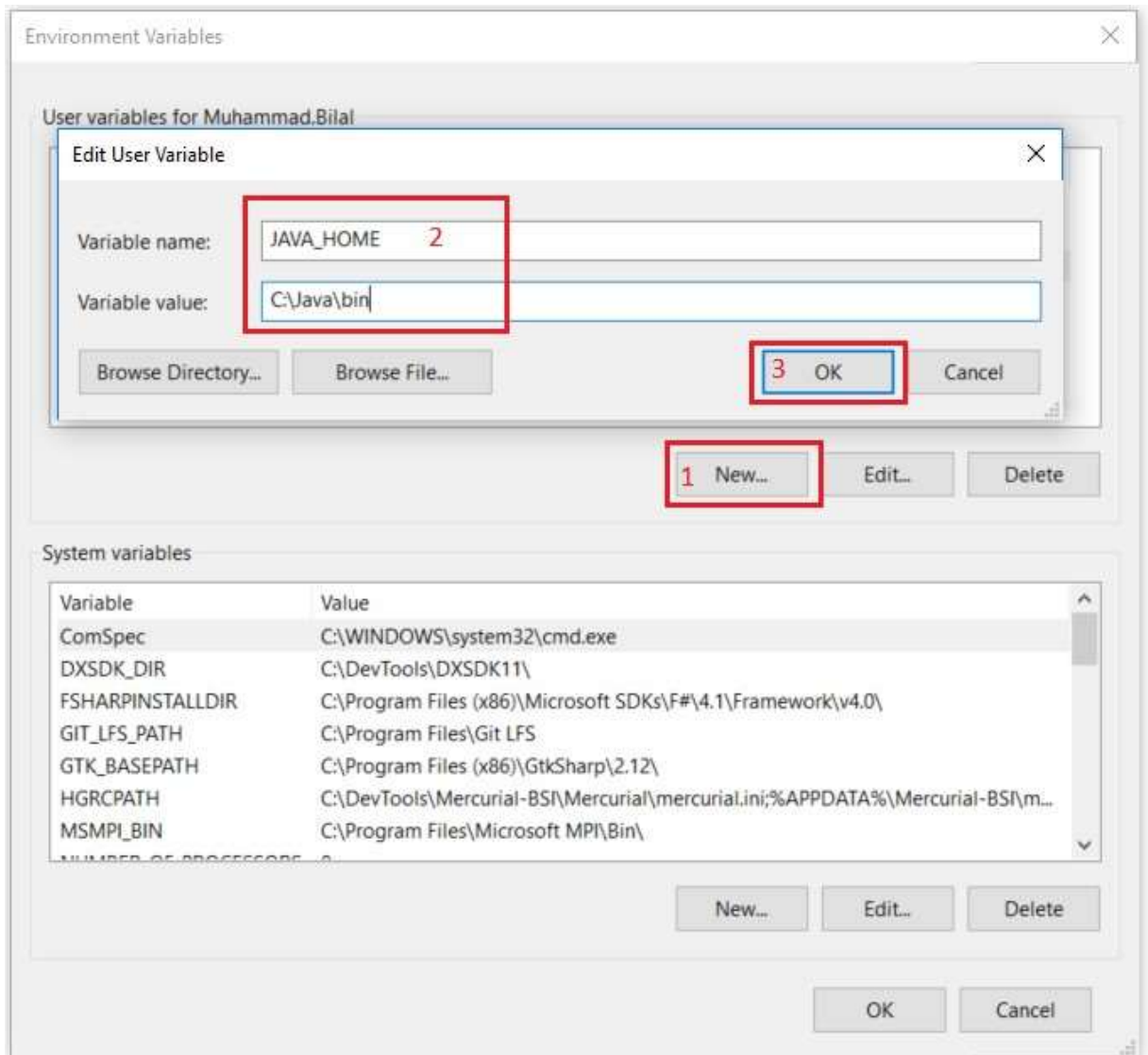


5. Set the path HADOOP_HOME Environment variable on windows 10



6.

Set the path JAVA_HOME Environment variable on windows 10



7. Next we set the Hadoop bin directory path and JAVA bin directory path.


```

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

```

2. Rename “mapred-site.xml.template” to “mapred-site.xml” and edit this file C:/Hadoop-2.8.0/etc/hadoop/mapred-site.xml, paste below xml paragraph and save this file.

```

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

```

3. Create folder “data” under “C:\Hadoop-2.8.0” Create folder “datanode” under “C:\Hadoop-2.8.0\data”
Create folder “namenode” under “C:\Hadoop-2.8.0\data”

<input type="checkbox"/> Name	Date modified	Type	Size
bin	7/20/2017 2:14 PM	File folder	
<input checked="" type="checkbox"/> data	7/20/2017 2:47 PM	File folder	
etc	7/20/2017 2:14 PM	File folder	
include	7/20/2017 2:14 PM	File folder	
lib	7/20/2017 2:14 PM	File folder	
libexec	7/20/2017 2:14 PM	File folder	
sbin	7/20/2017 2:14 PM	File folder	
share	7/20/2017 2:20 PM	File folder	
LICENSE.txt	3/17/2017 10:31 AM	TXT File	97 KB
NOTICE.txt	3/17/2017 10:31 AM	TXT File	16 KB
README.txt	3/17/2017 10:31 AM	TXT File	2 KB

4. Edit file C:\Hadoop-2.8.0/etc/hadoop/hdfs-site.xml, paste below xml paragraph and save this file.

```

<configuration>

<property>

```

```

<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name> <value>/hadoop-
2.8.0/data/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/hadoop-2.8.0/data/datanode</value>
</property>
</configuration>

```

5. Edit file C:/Hadoop-2.8.0/etc/hadoop/yarn-site.xml, paste below xml paragraph and save this file

```

<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>

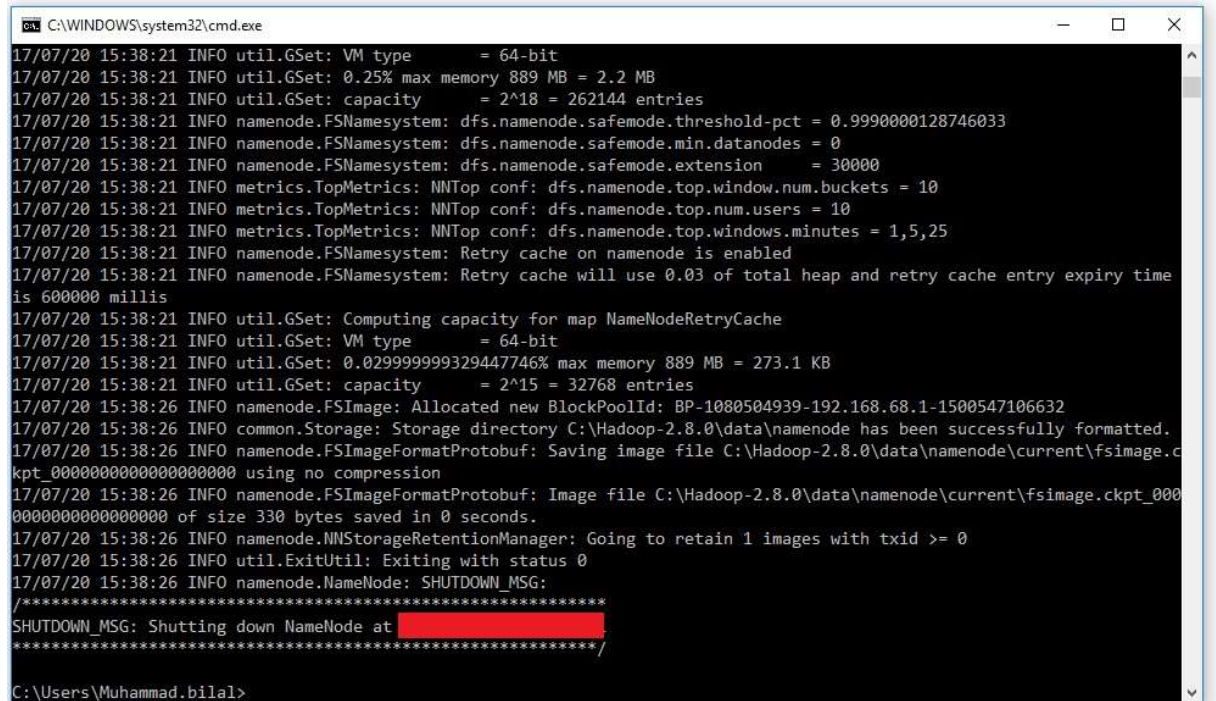
```

6. Edit file C:/Hadoop-2.8.0/etc/hadoop/hadoop-env.cmd by closing the command line "JAVA_HOME=%JAVA_HOME%" instead of set "JAVA_HOME=C:\Java" (On C:\java this is path to file jdk.18.0)

Hadoop Configuration

- Download file Hadoop Configuration.zip

- Delete file bin on C:\Hadoop-2.8.0\bin, replaced by file bin on file just download (from Hadoop Configuration.zip).
- Open cmd and typing command “hdfs namenode -format” .



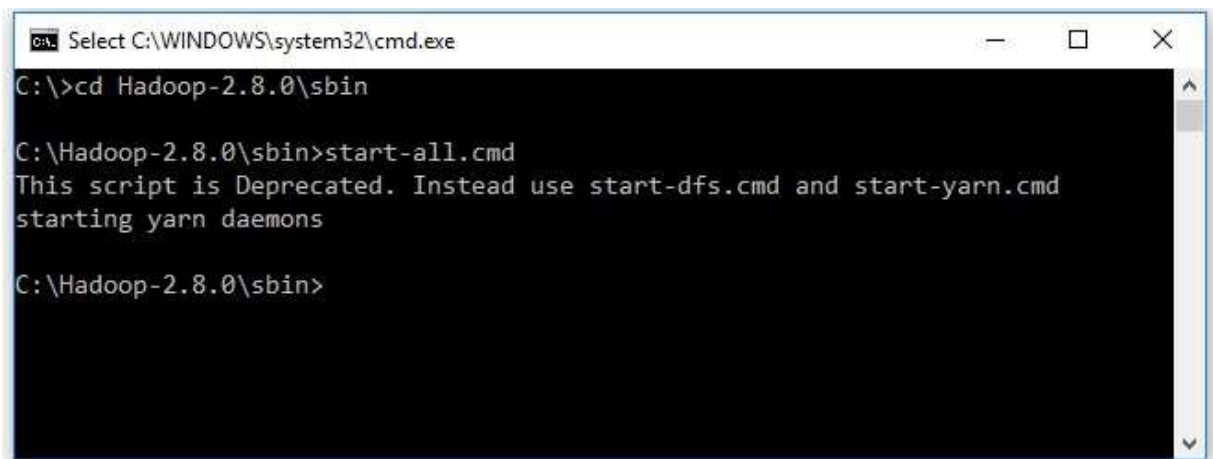
```

C:\WINDOWS\system32\cmd.exe
17/07/20 15:38:21 INFO util.GSet: VM type = 64-bit
17/07/20 15:38:21 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
17/07/20 15:38:21 INFO util.GSet: capacity = 2^18 = 262144 entries
17/07/20 15:38:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
17/07/20 15:38:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
17/07/20 15:38:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
17/07/20 15:38:21 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
17/07/20 15:38:21 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
17/07/20 15:38:21 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
17/07/20 15:38:21 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
17/07/20 15:38:21 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
17/07/20 15:38:21 INFO util.GSet: Computing capacity for map NameNodeRetryCache
17/07/20 15:38:21 INFO util.GSet: VM type = 64-bit
17/07/20 15:38:21 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
17/07/20 15:38:21 INFO util.GSet: capacity = 2^15 = 32768 entries
17/07/20 15:38:26 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1080504939-192.168.68.1-1500547106632
17/07/20 15:38:26 INFO common.Storage: Storage directory C:\Hadoop-2.8.0\data\namenode has been successfully formatted.
17/07/20 15:38:26 INFO namenode.FSImageFormatProtobuf: Saving image file C:\Hadoop-2.8.0\data\namenode\current\fsimage.ckpt_000000000000000000 using no compression
17/07/20 15:38:26 INFO namenode.FSImageFormatProtobuf: Image file C:\Hadoop-2.8.0\data\namenode\current\fsimage.ckpt_000000000000000000 of size 330 bytes saved in 0 seconds.
17/07/20 15:38:26 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
17/07/20 15:38:26 INFO util.ExitUtil: Exiting with status 0
17/07/20 15:38:26 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at [REDACTED]
*****/
C:\Users\Muhammad.bilal>

```

Testing

Open cmd and change directory to “C:\Hadoop-2.8.0\sbin” and type “start-all.cmd” to start apache.



```

C:\WINDOWS\system32\cmd.exe
C:\>cd Hadoop-2.8.0\sbin

C:\Hadoop-2.8.0\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\Hadoop-2.8.0\sbin>

```



```
17/07/2017 15:50:09 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:12 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:15 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:18 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:21 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:24 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:27 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:30 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:33 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:36 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:39 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:42 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:45 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:49 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:52 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:55 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:50:58 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:01 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:04 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:07 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:10 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:13 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:16 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:19 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:22 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:25 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:29 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:32 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/2017 15:51:35 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
```


Open: <http://localhost:8088>

← → ↻

localhost:8088/cluster

☆

:



Logged in as: dr.who

All Applications

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries

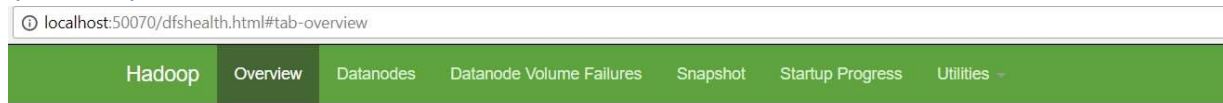
Search:

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																	

Showing 0 to 0 of 0 entries

First Previous Next Last

Open: <http://localhost:50070>



Overview 'localhost:9000' (active)

Started:	Thu Jul 20 15:44:11 +0500 2017
Version:	2.8.0, r91f2b7a13d1e97b[REDACTED]7cc29ac0009
Compiled:	Fri Mar 17 09:12:00 +0500 2017 by jdu from branch-2.8.0
Cluster ID:	CID-098b09fc-fc[REDACTED]df7b674
Block Pool ID:	BP-1080504[REDACTED]47106632

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 36.53 MB of 311 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 40.68 MB of 41.53 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	475.24 GB
DFS Used:	321 B (0%)
Non DFS Used:	261.08 GB

Result:

Thus the Hadoop single node cluster was successfully installed

Ex.No : 9 Word Count Program using Hadoop

Date :

Aim:

To implement a word count program using Hadoop Program:

WCMapper.java

```
// Importing Libraries
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

public class WCMapper extends MapReduceBase implements
Mapper < LongWritable, Text, Text, IntWritable >
{
    // Map function
    public void map(LongWritable key, Text value, OutputCollector < Text,
        IntWritable > output, Reporter rep) throws IOException {
        String line = value.toString();
        {
            if (word.length() > 0) {
                output.collect(new Text(word), new IntWritable(1));
            }
        }
    }
}
```

WCReducer.java

```
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;
```

```

public class WCReducer extends MapReduceBase implements Reducer < Text,
    IntWritable, Text, IntWritable >
{
    // Reduce function
    public void reduce(Text key, Iterator < IntWritable > value,
        OutputCollector < Text, IntWritable > output,
        Reporter rep) throws IOException {
        int count = 0;
        // Counting the frequency of each words while (value.hasNext())
        {
            IntWritable i = value.next();
            count += i.get();
        }
        output.collect(key, new IntWritable(count));
    }
}

```

WCDriver.java

```

import java.io.IOException;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
public class WCDriver extends Configured implements Tool {
    public int run(String args[]) throws IOException {
        if (args.length < 2) {
            System.out.println("Please give valid inputs");
            return -1;
        }
        JobConf conf = new JobConf(WCDriver.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));
        conf.setMapperClass(WCMapper.class);
        conf.setReducerClass(WCReducer.class);
        conf.setMapOutputKeyClass(Text.class);
        conf.setMapOutputValueClass(IntWritable.class);
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
    }
}

```

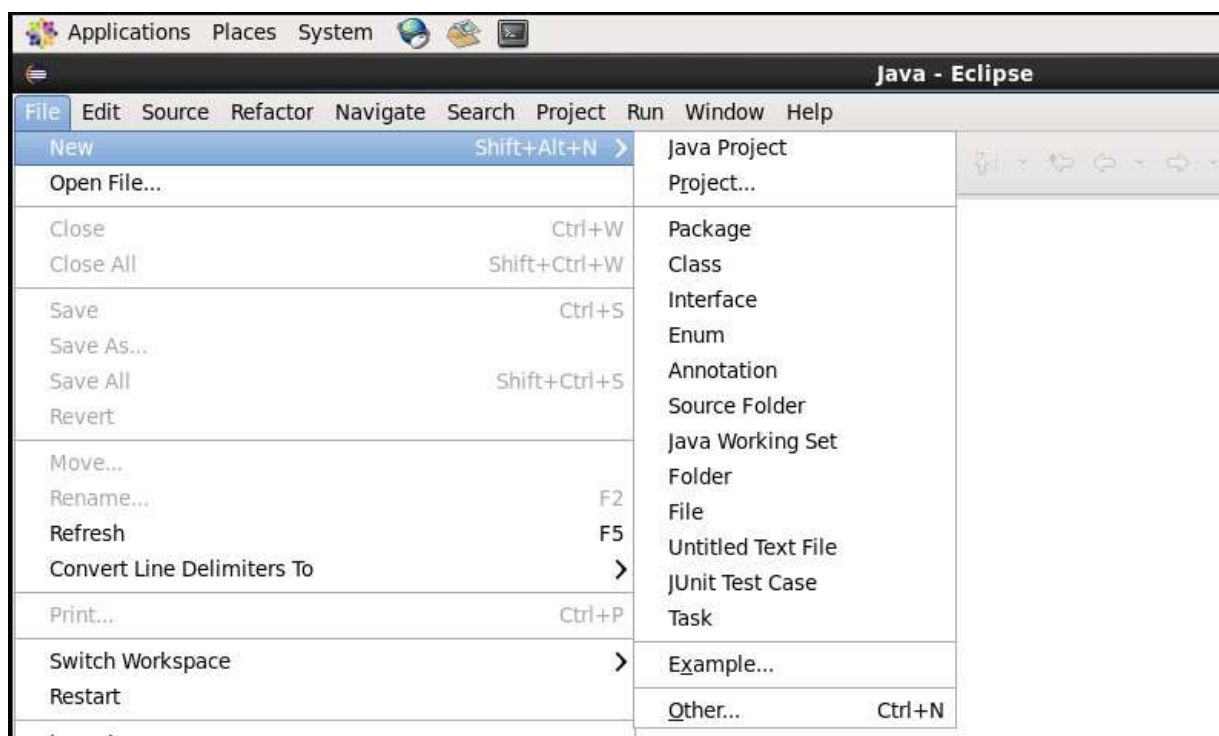
```

        JobClient.runJob(conf);
        return 0;
    }
    // Main Method
    public static void main(String args[]) throws Exception {
        int exitCode = ToolRunner.run(new WCDriver(), args);
        System.out.println(exitCode);
    }
}

```

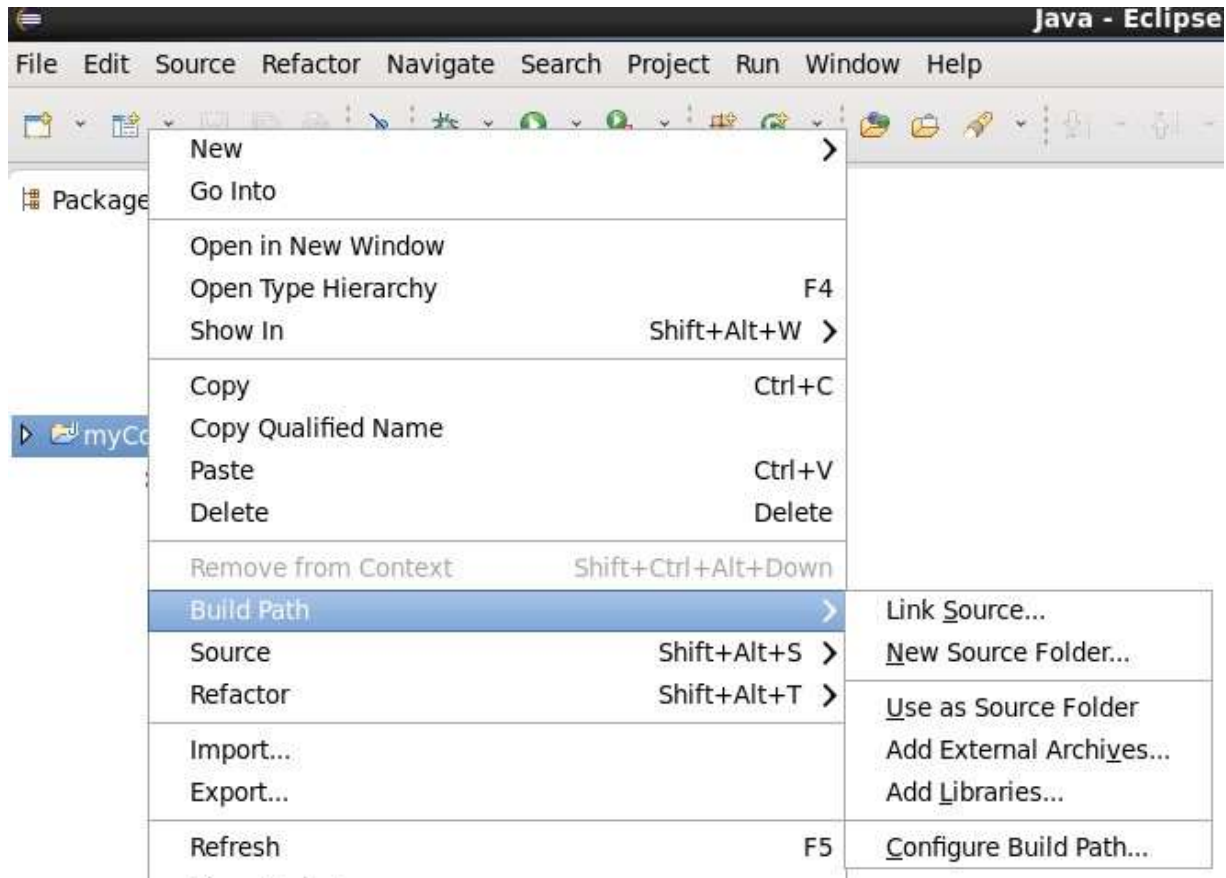
Steps:

First Open Eclipse -> then select File -> New -> Java Project ->Name it WordCount -> then Finish.



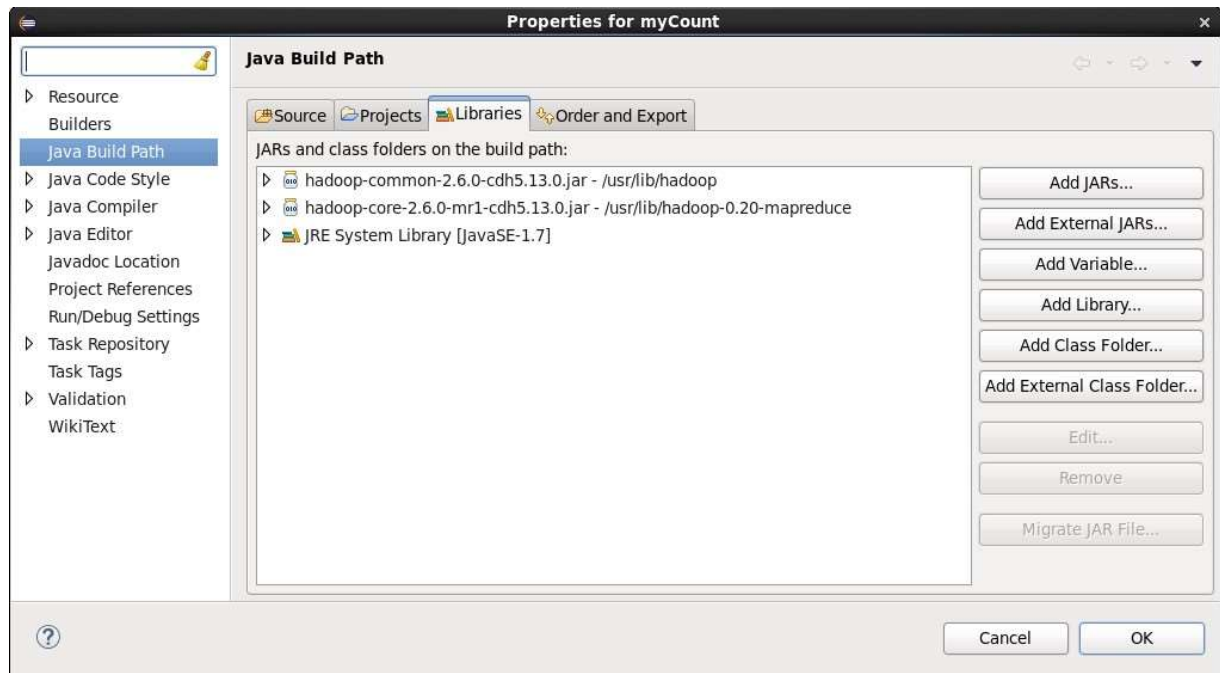
Create Three Java Classes into the project. Name them **WCDriver**(having the main function), **WCMapper**, **WCReducer**.

You have to include two Reference Libraries for that: Right Click on **Project** -> then select **Build Path**-> Click on **Configure Build Path**



In the above figure, you can see the Add External JARs option on the Right Hand Side. Click on it and add the below mention files. You can find these files in /usr/lib/

1. /usr/lib/hadoop-0.20-mapreduce/hadoop-core-2.6.0-mr1-cdh5.13.0.jar
2. /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.13.0.jar



Now you have to make a jar file. Right Click on Project-> Click on Export-> Select export destination as Jar File-> Name the jar File(WordCount.jar) -> Click on next -> at last Click on Finish. Now copy this file into the Workspace directory of Cloudera

Open the terminal on CDH and change the directory to the workspace.

You can do this by using “cd workspace/” command. Now, Create a text file(WCFile.txt) and move it to HDFS. For that open terminal and write this code(remember you should be in the same directory as jar file you have created just now).

```

cloudera@quickstart:~/workspace
File Edit View Search Terminal Help
[cloudera@quickstart workspace]$ cat WCFile.txt
Hello I am GeeksforGeeks
Hello I am an Intern
[cloudera@quickstart workspace]$

```

Now, run this command to copy the file input file into the HDFS.

`hadoop fs -put WCFile.txt WCFile.txt`

```

cloudera@quickstart:~/workspace
File Edit View Search Terminal Help
[cloudera@quickstart workspace]$ hadoop fs -put WCFile.txt WCFile.txt
[cloudera@quickstart workspace]$

```

```
cloudera@quickstart:~/workspace
File Edit View Search Terminal Help
[cloudera@quickstart workspace]$ hadoop jar wordCount.jar WCDriver WCFile.txt WCOuput
19/05/06 22:43:22 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/05/06 22:43:22 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
```

After Executing the code, you can see the result in WCOuput file or by writing following command on terminal. `hadoop fs -cat WCOuput/part-00000`

```
cloudera@quickstart:~/workspace
File Edit View Search Terminal Help
[cloudera@quickstart workspace]$ hadoop fs -cat WCOuput/part-00000
GeeksforGeeks 1
Hello 2
I 2
Intern 1
am 2
an 1
```

Result:

Thus the word count program using hadoop was executed successfully

