CLOUD COMPUTING AND BIG DATA LABORATORY

Exercises:

1. Install Virtualbox/VMware Workstation with different flavours of linux or windows OS on topofwindows7or8.
2. Install a C compiler in the virtual machine created using virtual box and execute Simple Programs
3. Install Google Cloud SDK CLI. Create hello world app and execute Simple Web applications using python/java.
4. Use Google App Engine Launcher to launch the web applications
5. Installation of Single Node Hadoop Cluster on Ubuntu 22.04 LTS.
6. Hadoop Programming: Word Count MapReduce Program Using Eclipse
7. File Management tasks in Hadoop using HDFS commands
   a. Adding files to HDFS
   b. Retrieving files from HDFS
   c. Deleting files from HDFS
8. Creation of DataFrames using CreateDataFrame working with an example related to FIFA dataset.

EX.No:1 Install Virtual box/VMware Workstation

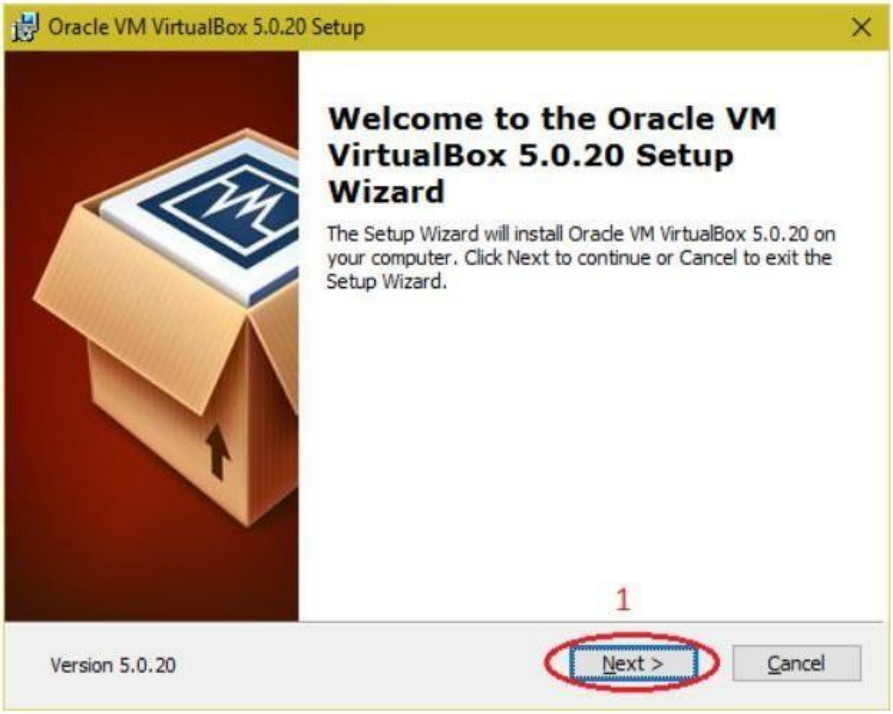
Aim: Find procedure to Install Virtualbox/VMware Workstation with different flavours of linux or windows OS on top of windows7or8.

This experiment is to be performed through portal.

PROCEDURE TO INSTALL

1. Download and Install VirtualBox. Using the link-
2. Download and Install Ubuntu. Using the link-

Run the virtual box setup and click on “Next” Button.



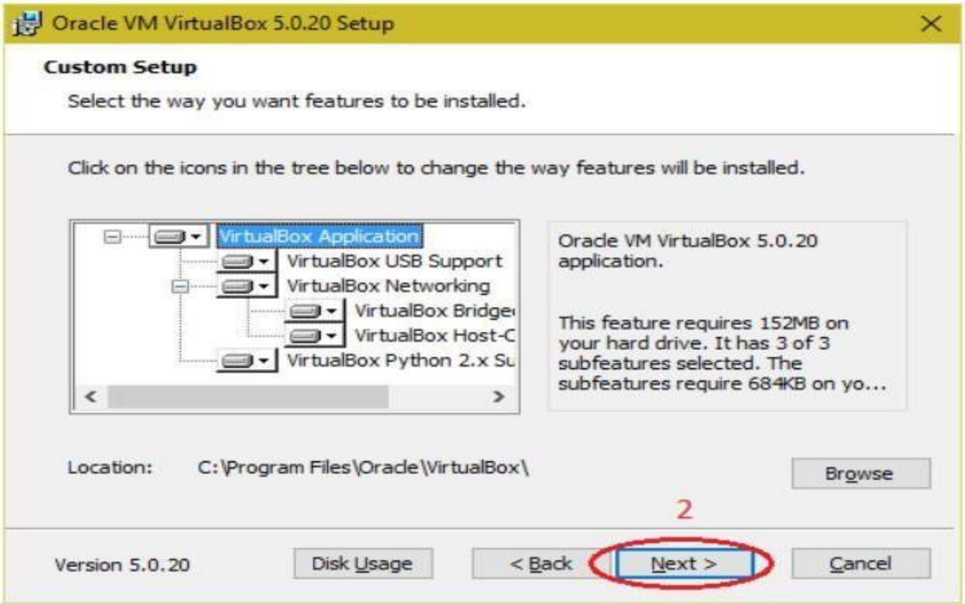
Oracle VM VirtualBox 5.0.20 Setup

Welcome to the Oracle VM VirtualBox 5.0.20 Setup Wizard

The Setup Wizard will install Oracle VM VirtualBox 5.0.20 on your computer. Click Next to continue or Cancel to exit the Setup Wizard.

Version 5.0.20

Click on “Next” button.



Oracle VM VirtualBox 5.0.20 Setup

Custom Setup

Select the way you want features to be installed.

Click on the icons in the tree below to change the way features will be installed.

- VirtualBox Application
  - VirtualBox USB Support
  - VirtualBox Networking
    - VirtualBox Bridge
    - VirtualBox Host-C
  - VirtualBox Python 2.x Su

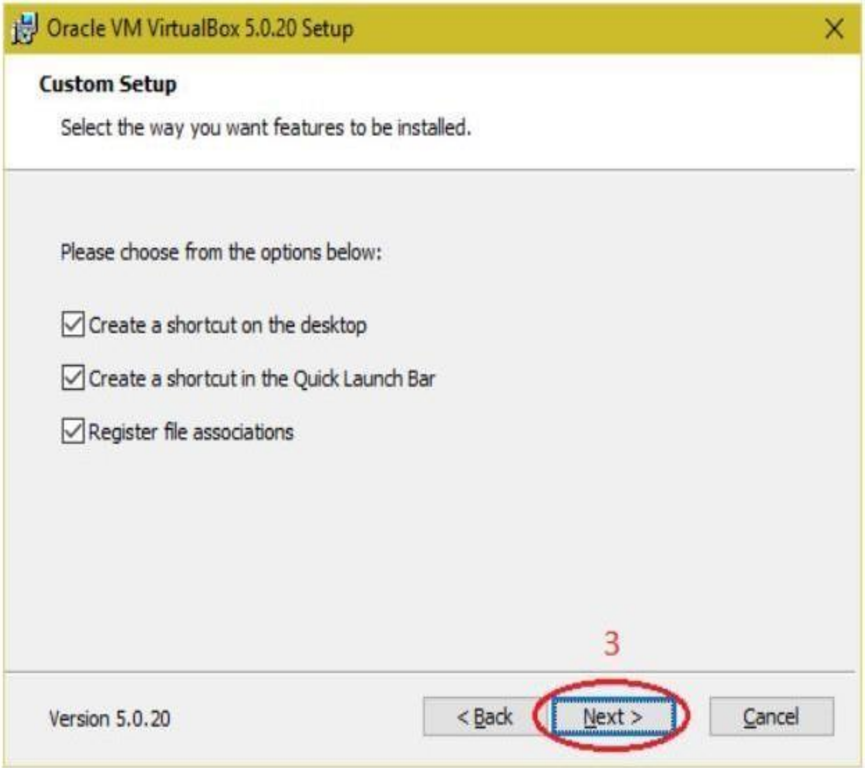
Oracle VM VirtualBox 5.0.20 application.

This feature requires 152MB on your hard drive. It has 3 of 3 subfeatures selected. The subfeatures require 684KB on yo...

Location: C:\Program Files\Oracle\VirtualBox\

Version 5.0.20

Click on “Next” button.



Oracle VM VirtualBox 5.0.20 Setup

Custom Setup

Select the way you want features to be installed.

Please choose from the options below:

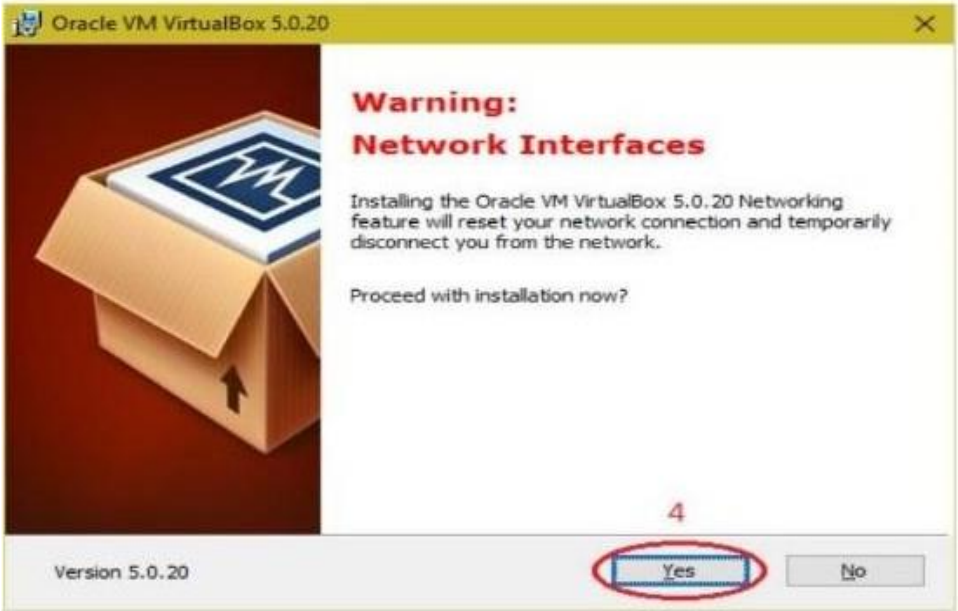
☑ Create a shortcut on the desktop

☑ Create a shortcut in the Quick Launch Bar

☑ Register file associations

Version 5.0.20

3



Oracle VM VirtualBox 5.0.20

Warning: Network Interfaces

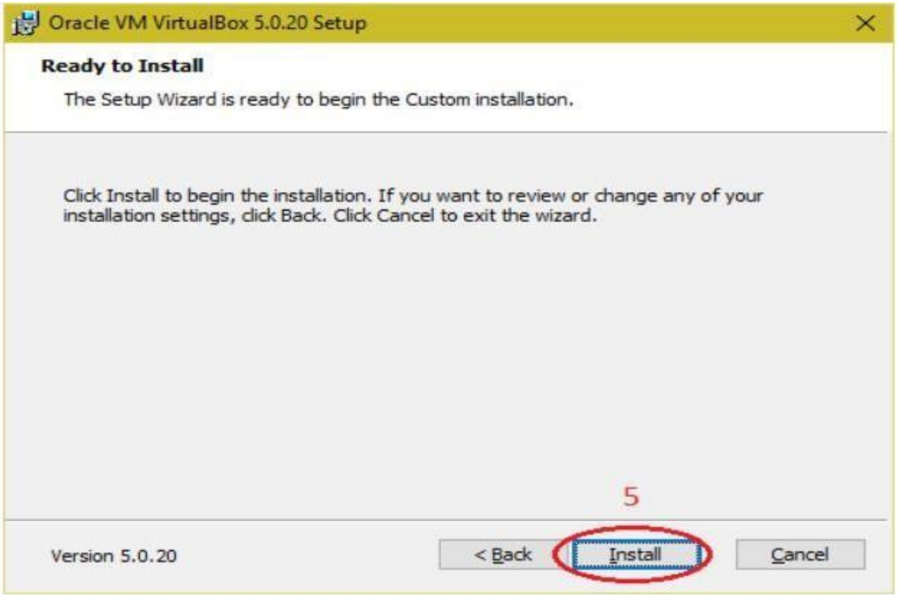
Installing the Oracle VM VirtualBox 5.0.20 Networking feature will reset your network connection and temporarily disconnect you from the network.

Proceed with installation now?

Version 5.0.20

4

Click on “Yes” button



Oracle VM VirtualBox 5.0.20 Setup

Ready to Install

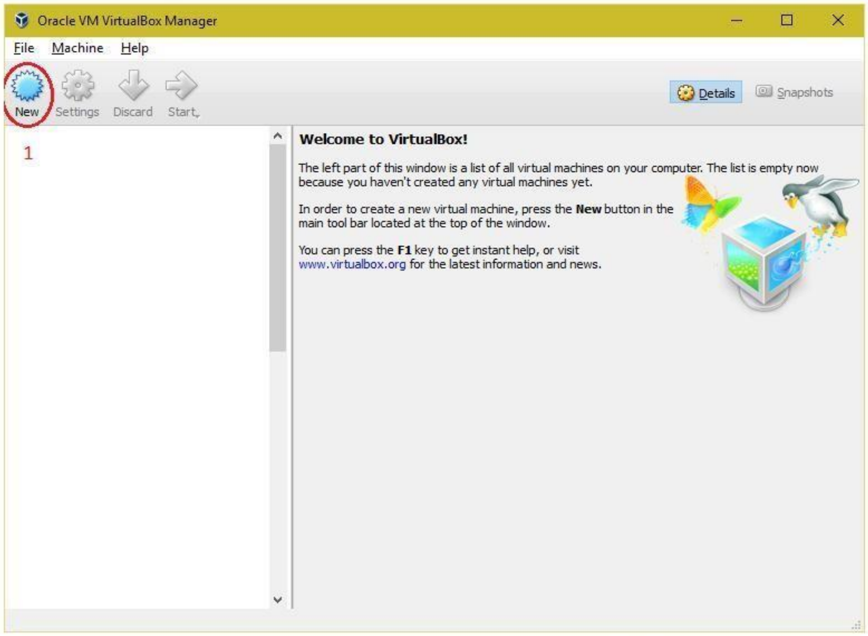
The Setup Wizard is ready to begin the Custom installation.

Click Install to begin the installation. If you want to review or change any of your installation settings, click Back. Click Cancel to exit the wizard.

Version 5.0.20

Installing “Ubuntu”as Virtual machine in “Oracle VM VirtualBox”.

1. Open “Oracle VM VirtualBoxManager”.



Oracle VM VirtualBox Manager

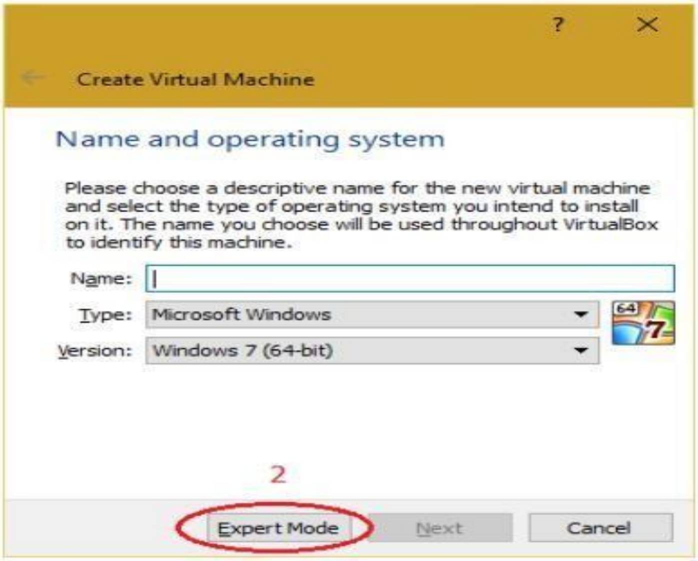
Welcome to VirtualBox!

The left part of this window is a list of all virtual machines on your computer. The list is empty now because you haven't created any virtual machines yet.

In order to create a new virtual machine, press the **New** button in the main tool bar located at the top of the window.

You can press the **F1** key to get instant help, or visit www.virtualbox.org for the latest information and news.

1. Click on ”New “button and select “Expert Mode” .



Create Virtual Machine

Name and operating system

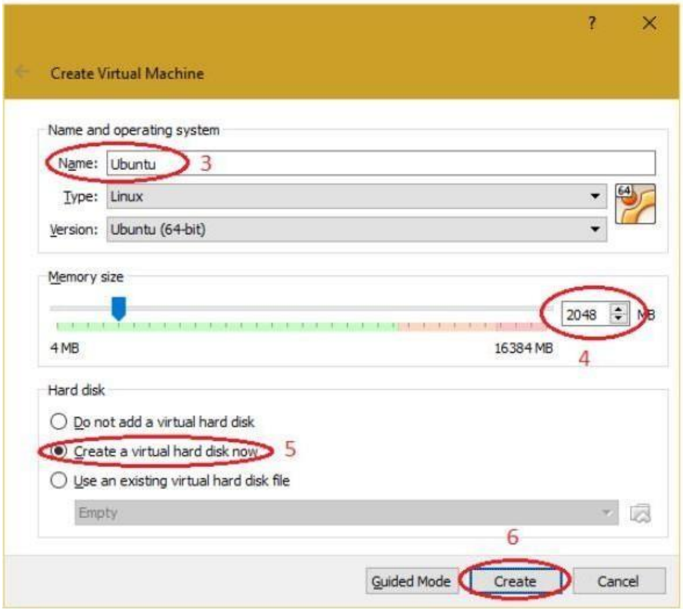
Please choose a descriptive name for the new virtual machine and select the type of operating system you intend to install on it. The name you choose will be used throughout VirtualBox to identify this machine.

Name:

Type: Microsoft Windows

Version: Windows 7 (64-bit)

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



Create Virtual Machine

Name and operating system

Name: Ubuntu

Type: Linux

Version: Ubuntu (64-bit)

Memory size: 2048 MB

Hard disk

○ Do not add a virtual hard disk

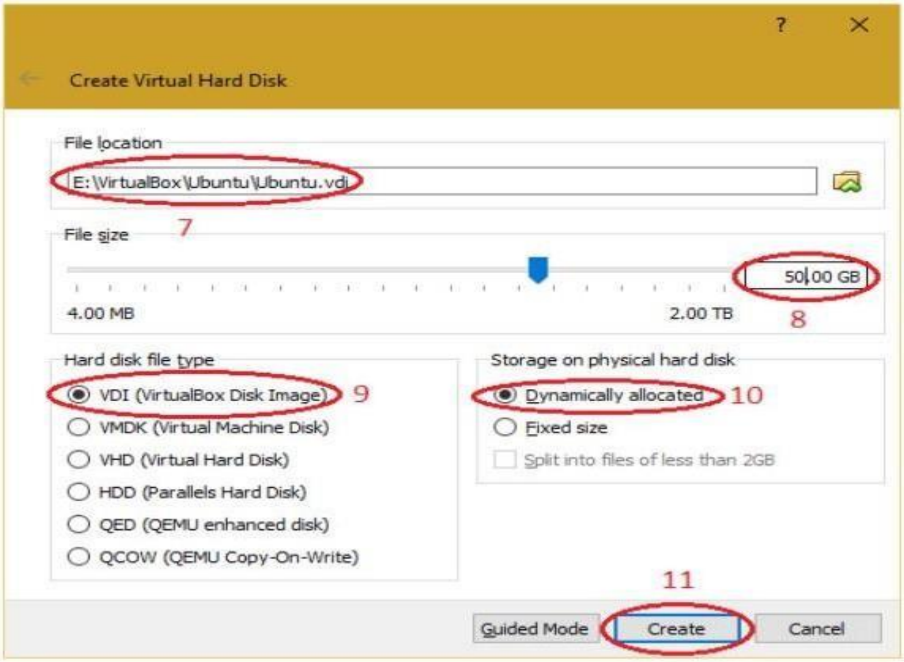
● Create a virtual hard disk now

○ Use an existing virtual hard disk file

Empty

Note: Before installing 64-bit operating system, Intel VT-x/AMD-V must be enabled in "BIOS" on the system. To enable Intel VT-x/AMD-V, open BIOS and search for "Intel Virtualization Technology" or "AMD-V", save the BIOS and boot the PC/Laptop.

3.Select the path for the virtual hard disk and Click on “create”button.



Create Virtual Hard Disk

File location

E:\VirtualBox\Ubuntu\Ubuntu.vdi

File size

50.00 GB

4.00 MB

2.00 TB

Hard disk file type

● VDI (VirtualBox Disk Image)

○ VMDK (Virtual Machine Disk)

○ VHD (Virtual Hard Disk)

○ HDD (Parallels Hard Disk)

○ QED (QEMU enhanced disk)

○ QCOW (QEMU Copy-On-Write)

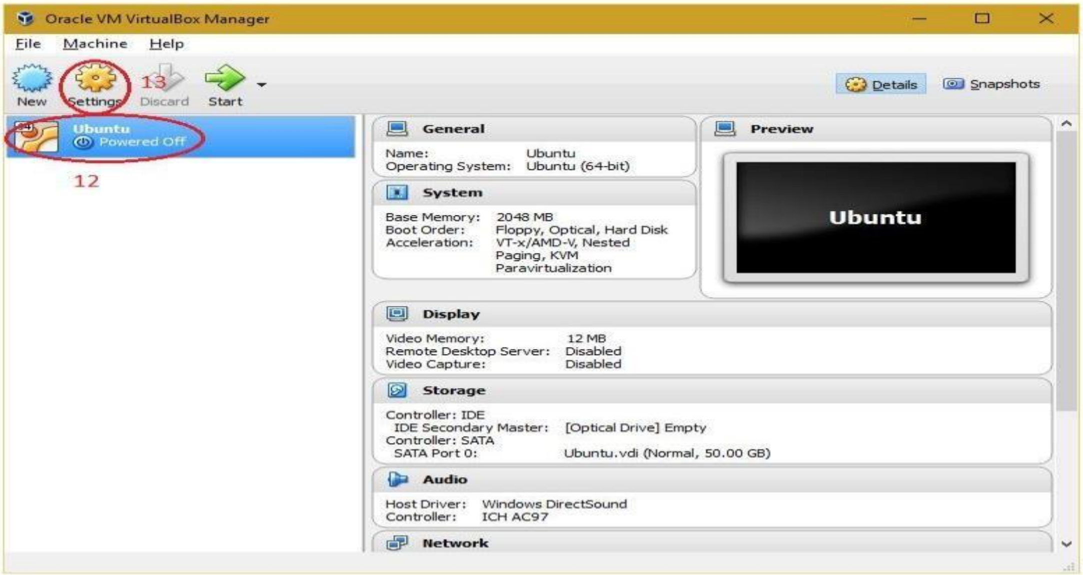
Storage on physical hard disk

● Dynamically allocated

○ Fixed size

□ Split into files of less than 2GB

4. Select the virtual machine from the Virtual box manager and click on “settings “ button.



Oracle VM VirtualBox Manager

Ubuntu
Powered Off

General

Name: Ubuntu
Operating System: Ubuntu (64-bit)

System

Base Memory: 2048 MB
Boot Order: Floppy, Optical, Hard Disk
Acceleration: VT-x/AMD-V, Nested Paging, KVM Paravirtualization

Preview

Ubuntu

Display

Video Memory: 12 MB
Remote Desktop Server: Disabled
Video Capture: Disabled

Storage

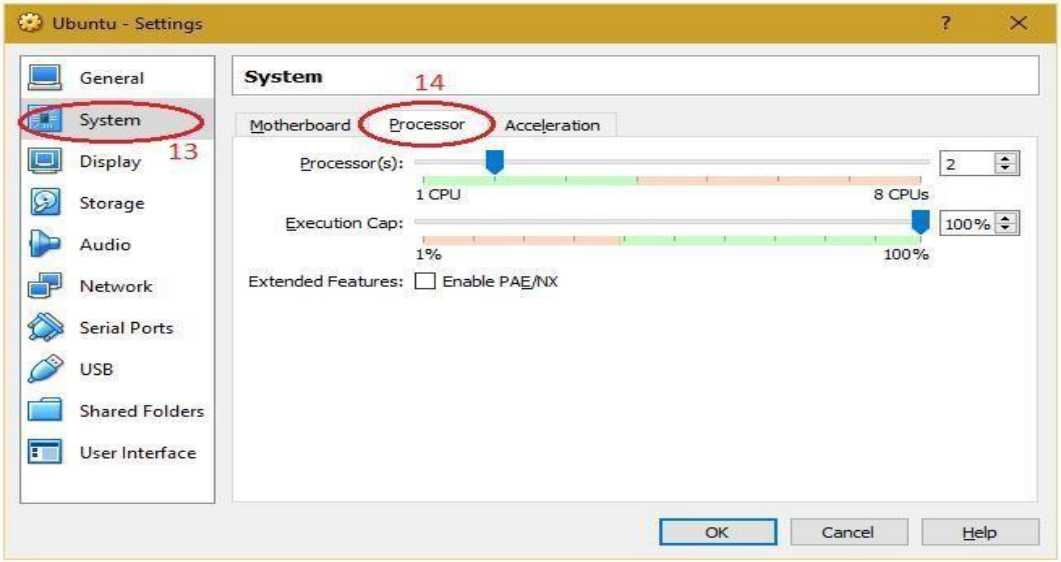
Controller: IDE
IDE Secondary Master: [Optical Drive] Empty
Controller: SATA
SATA Port 0: Ubuntu.vdi (Normal, 50.00 GB)

Audio

Host Driver: Windows DirectSound
Controller: ICH AC97

Network

5. Select “System” and navigate to “processor” tab to adjust number of processor virtual machine for better performance.



Ubuntu - Settings

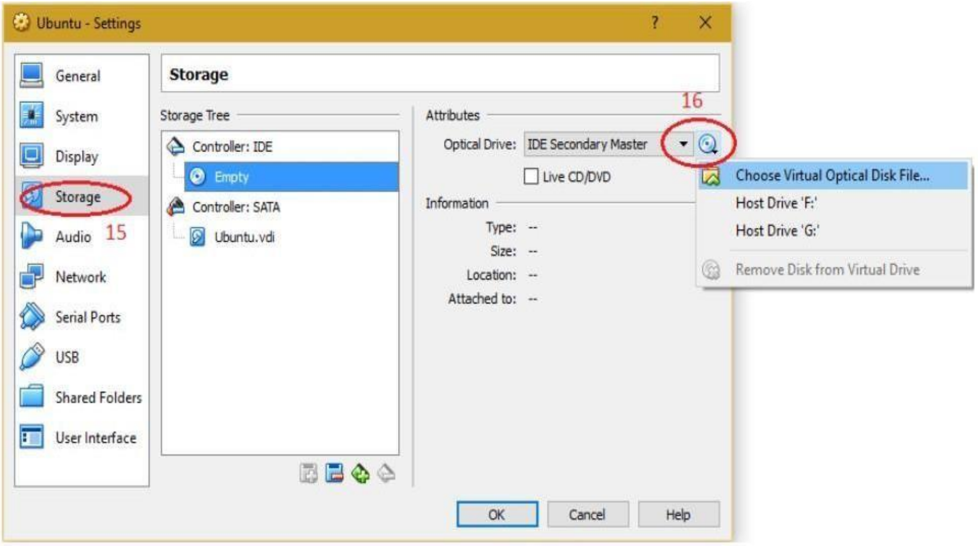
System

Processor(s): 2

Execution Cap: 100%

Extended Features: □ Enable PAE/NX

6. Select “storage” and choose the installation media of Operation System (ISO/CD/DVD). Preferred Linux “.iso” can be download from CC ftp site. Also many different flavours of Linux are available on the internet- Fedora , CentOS , Ubuntu, Debian, Mageia, openSUSE, Arch Linux, Slackware Linux, etc.



Ubuntu - Settings

Storage

Storage Tree

Controller: IDE
- Empty

Controller: SATA
- Ubuntu.vdi

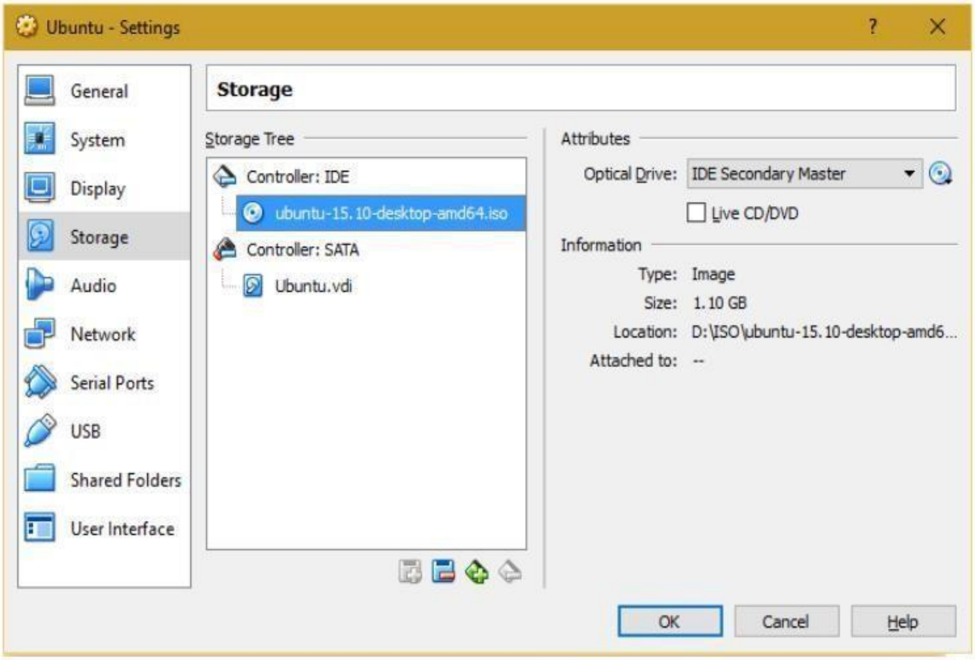
Attributes

Optical Drive: IDE Secondary Master

□ Live CD/DVD

Information

Type: --
Size: --
Location: --
Attached to: --



Storage

Optical Drive: IDE Secondary Master

□ Live CD/DVD

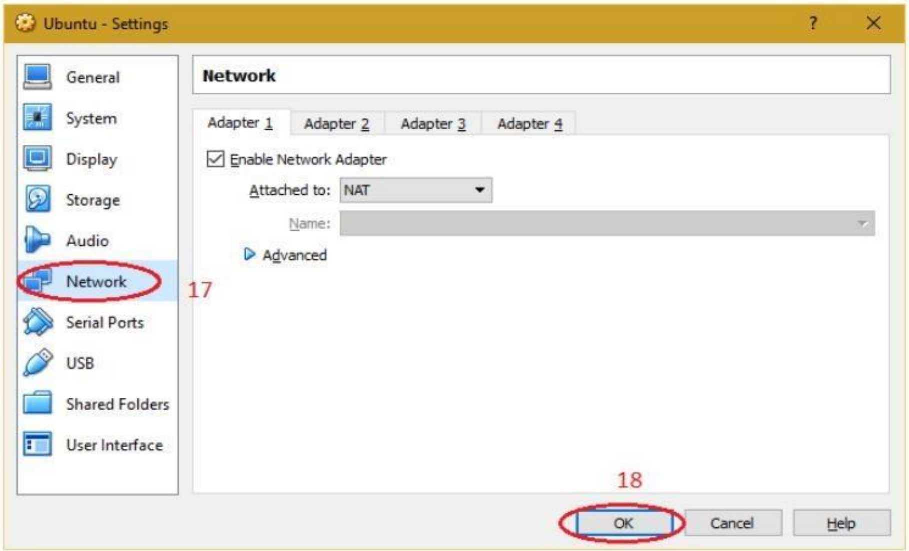
Type: Image

Size: 1.10 GB

Location: D:\ISO\ubuntu-15.10-desktop-amd6...

Attached to: --

7. Select “Network” to make changes required for network setting of virtual machine and click on “OK”.



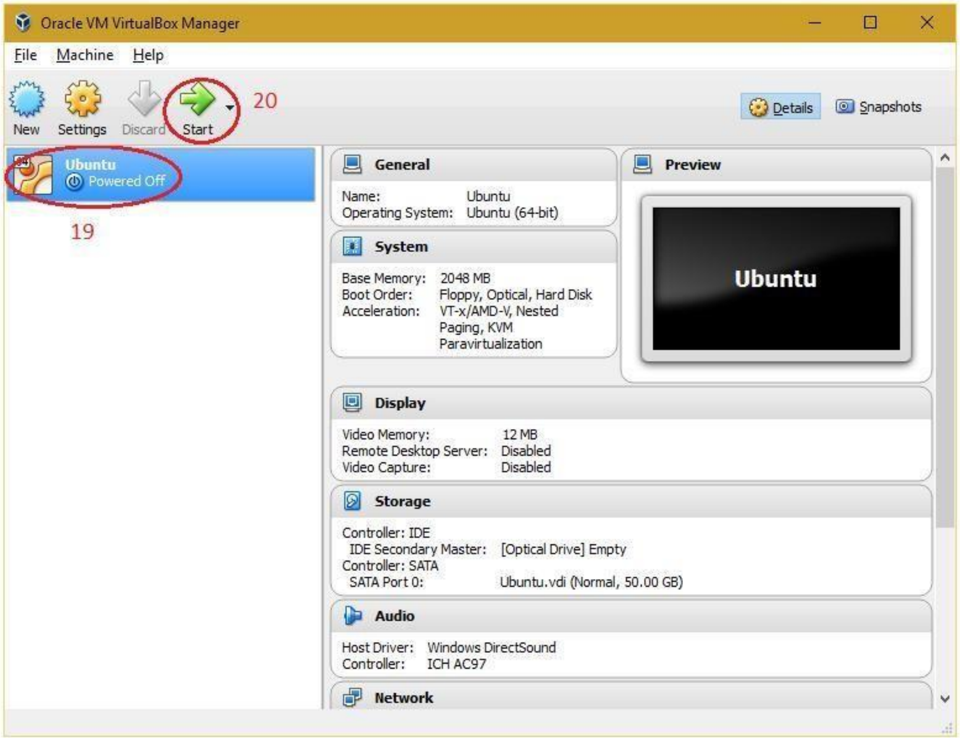
Network

☑ Enable Network Adapter

Attached to: NAT

Name:

8. Select the created Virtual machine and click on “Start” button.



General
Name: Ubuntu
Operating System: Ubuntu (64-bit)

System
Base Memory: 2048 MB
Boot Order: Floppy, Optical, Hard Disk
Acceleration: VT-x/AMD-V, Nested Paging, KVM Paravirtualization

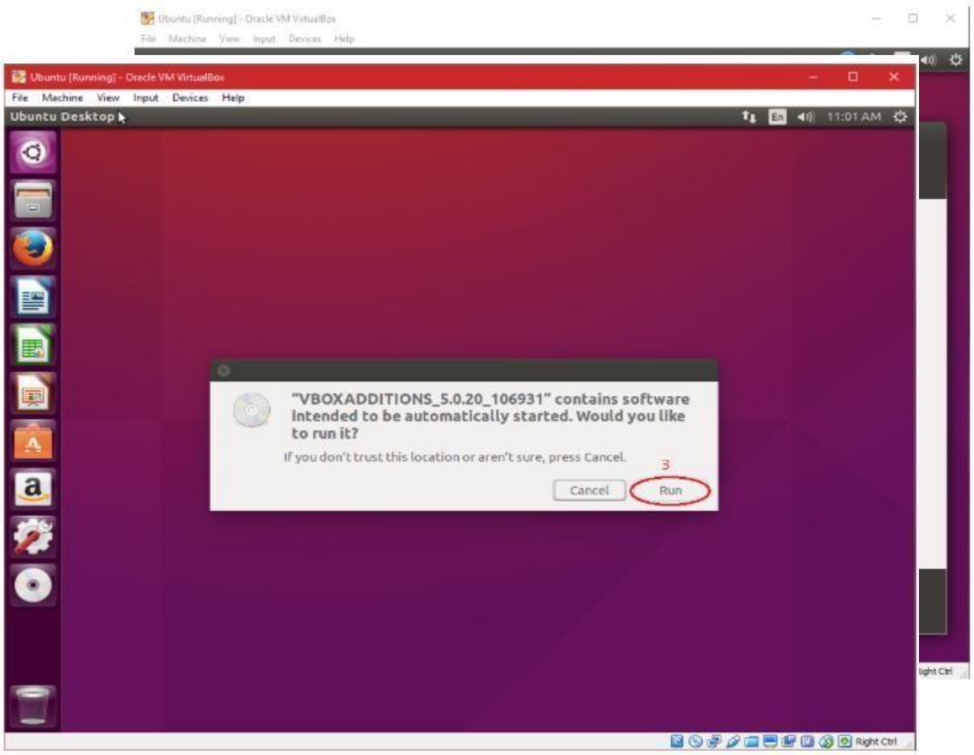
Display
Video Memory: 12 MB
Remote Desktop Server: Disabled
Video Capture: Disabled

Storage
Controller: IDE
IDE Secondary Master: [Optical Drive] Empty
Controller: SATA
SATA Port 0: Ubuntu.vdi (Normal, 50.00 GB)

Audio
Host Driver: Windows DirectSound
Controller: ICH AC97

Network

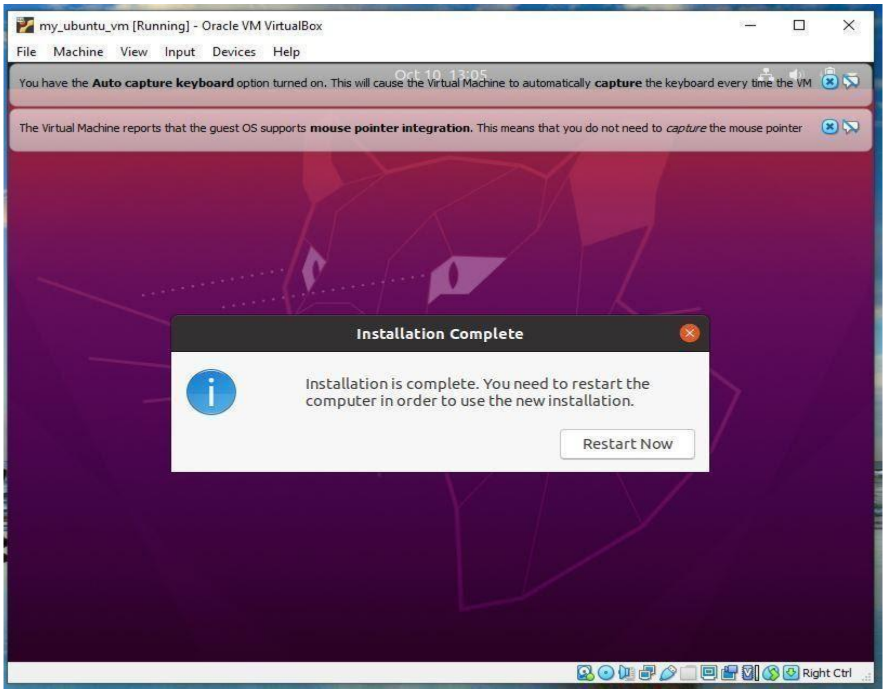
9. Proceed with the installation of operating system in virtual machine.


Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Ubuntu Desktop
11:01 AM
3
Cancel
Run
Right Ctrl

“VBOXADDITIONS_5.0.20_106931” contains software intended to be automatically started. Would you like to run it?

If you don't trust this location or aren't sure, press Cancel.

“Run “ to install on Ubuntu virtual machine. Once the installation Finishes installation, restart the virtual system.

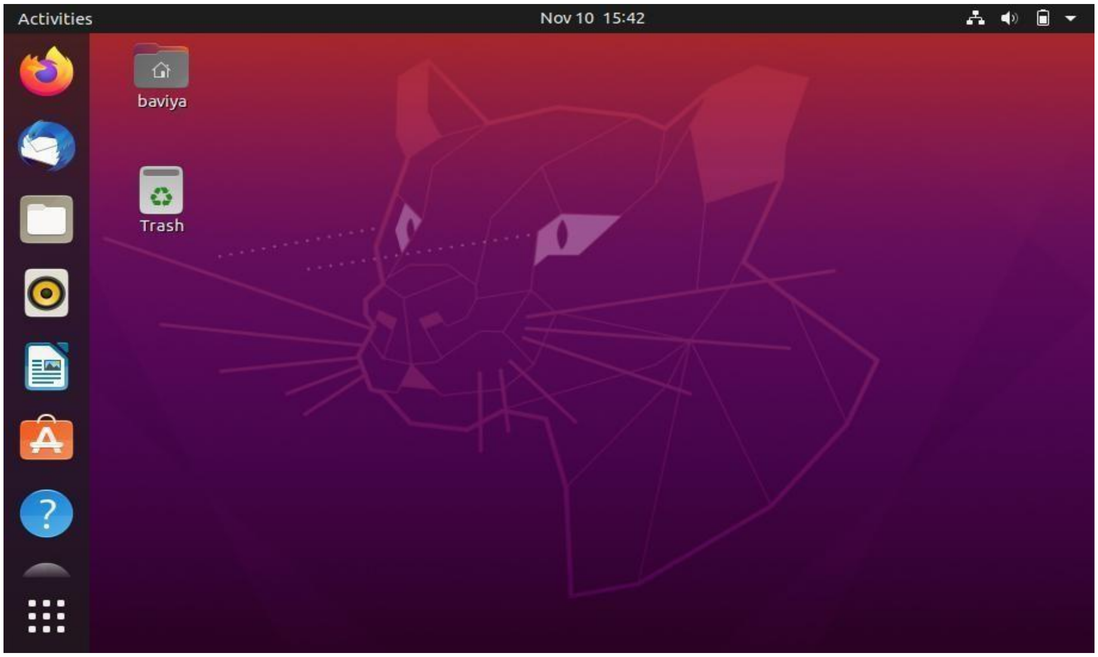

my_ubuntu_vm [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Oct 10 13:05
Installation Complete
Restart Now
Right Ctrl

You have the **Auto capture keyboard** option turned on. This will cause the Virtual Machine to automatically **capture** the keyboard every time the VM

The Virtual Machine reports that the guest OS supports **mouse pointer integration**. This means that you do not need to *capture* the mouse pointer

Installation is complete. You need to restart the computer in order to use the new installation.

Complete the installation process.



RESULT:

Successfully.Thus, the VirtualBox was installed with Ubuntu OS over Windows

Ex No: 2

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

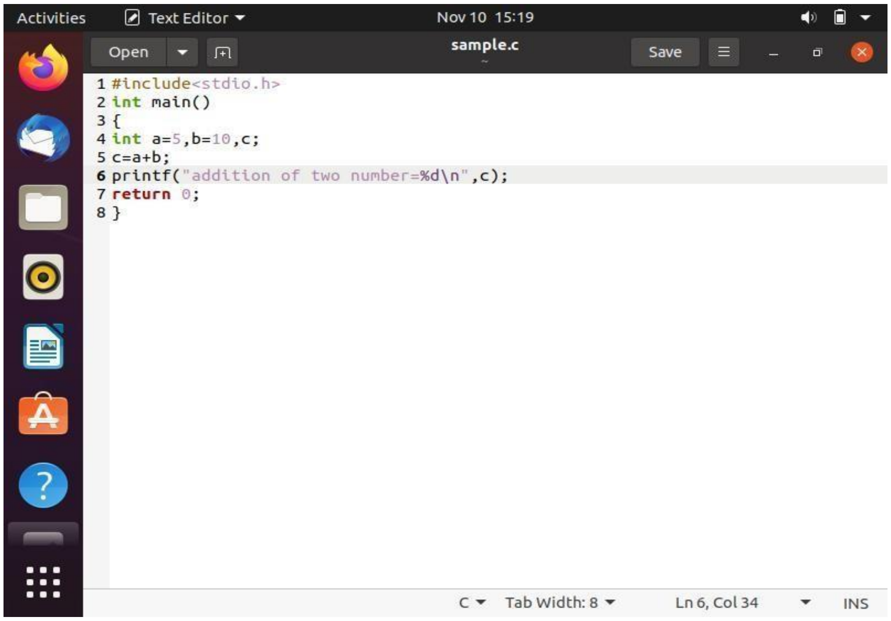
AIM: To install a C compiler in the virtual machine created and execute simple programs.

1.Open the terminal in your Virtual Machine and install sudo and gcc using thecommands-

- For sudo installation -$ apt-get install sudo
- For gcc installation -$ sudo apt update

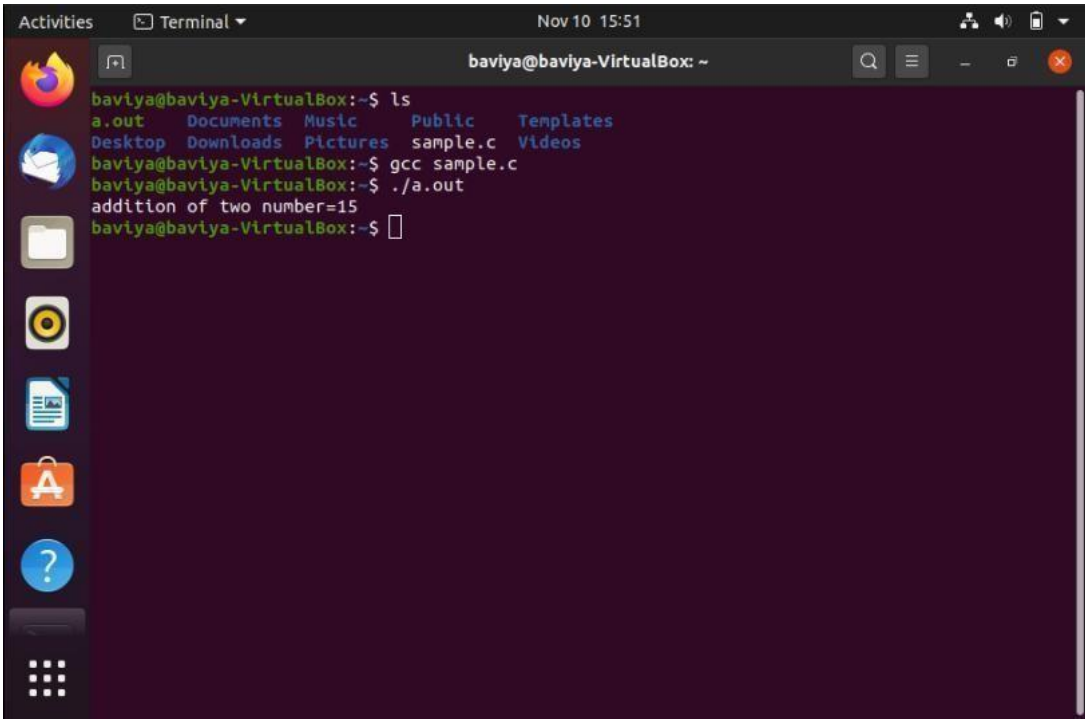
3.After installation, run and type a simple C program.

•



```c
#include<stdio.h>
int main()
{
int a=5,b=10,c;
c=a+b;
printf("addition of two number=%d\n",c);
return 0;
}
```

1. Run the program using the command (Objects can also be created)-
   - ./a.out



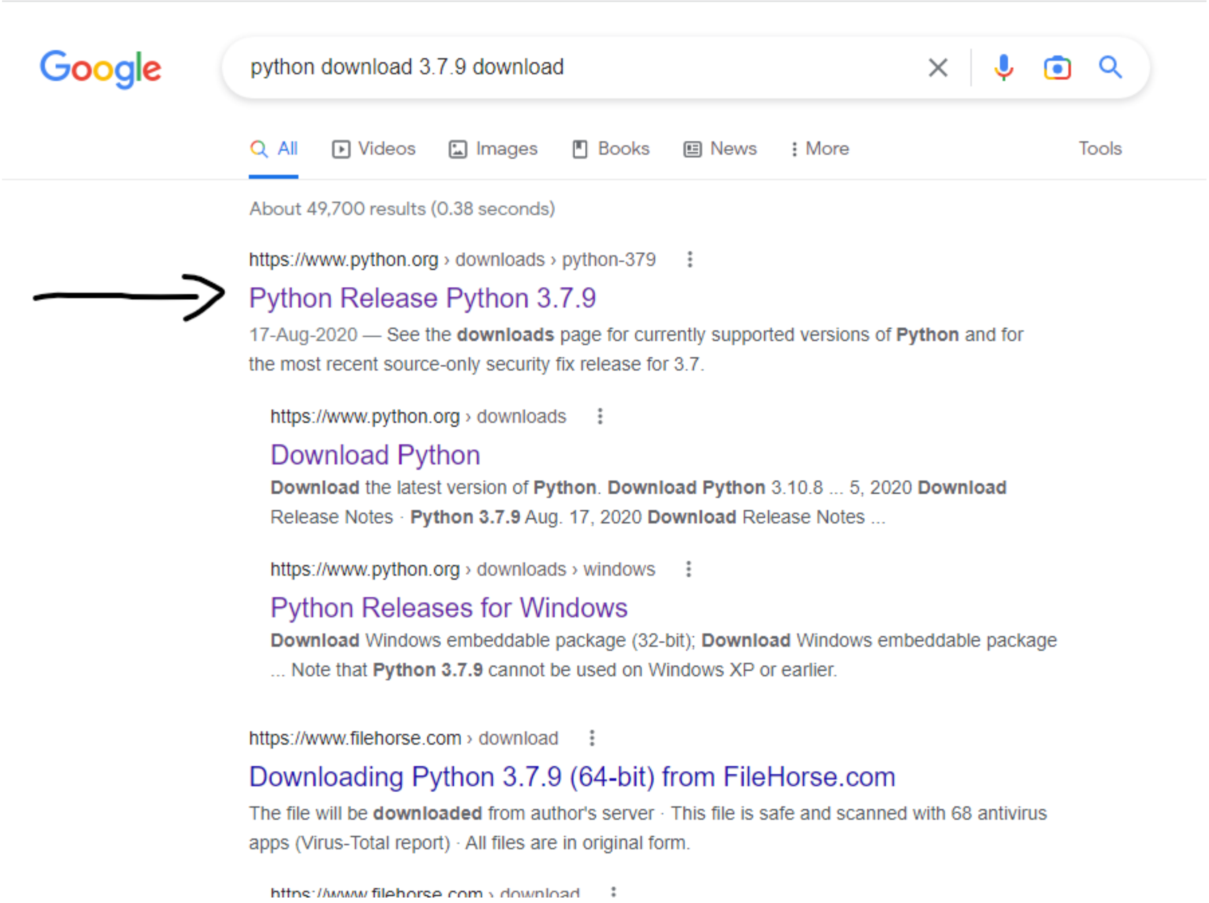
```
baviya@baviya-VirtualBox:~$ ls
a.out    Documents  Music     Public    Templates
Desktop  Downloads  Pictures  sample.c  Videos
baviya@baviya-VirtualBox:~$ gcc sample.c
baviya@baviya-VirtualBox:~$ ./a.out
addition of two number=15
baviya@baviya-VirtualBox:~$ 
```

RESULT:

Thus, a C compiler was installed in the virtual machine and simple programs were executed successfully.

EX 3: Install Google SDK CLI. Create hello world app and execute Simple Web applications using python/java.

Step1: go to Google chrome and Search python download 3.7.9 download click on first option has shown in the figure



About 49,700 results (0.38 seconds)

https://www.python.org › downloads › python-379

Python Release Python 3.7.9

17-Aug-2020 — See the **downloads** page for currently supported versions of **Python** and for the most recent source-only security fix release for 3.7.

https://www.python.org › downloads

Download Python

Download the latest version of **Python**. **Download Python** 3.10.8 ... 5, 2020 **Download** Release Notes · **Python 3.7.9** Aug. 17, 2020 **Download** Release Notes ...

https://www.python.org › downloads › windows

Python Releases for Windows

Download Windows embeddable package (32-bit); **Download** Windows embeddable package ... Note that **Python 3.7.9** cannot be used on Windows XP or earlier.

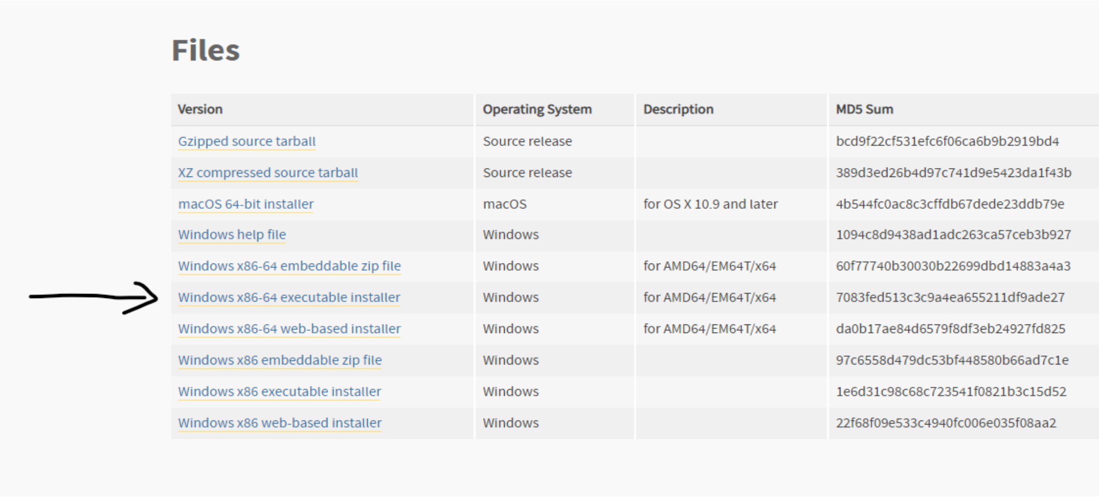
https://www.filehorse.com › download

Downloading Python 3.7.9 (64-bit) from FileHorse.com

The file will be **downloaded** from author's server · This file is safe and scanned with 68 antivirus apps (Virus-Total report) · All files are in original form.

https://www.filehorse.com › download

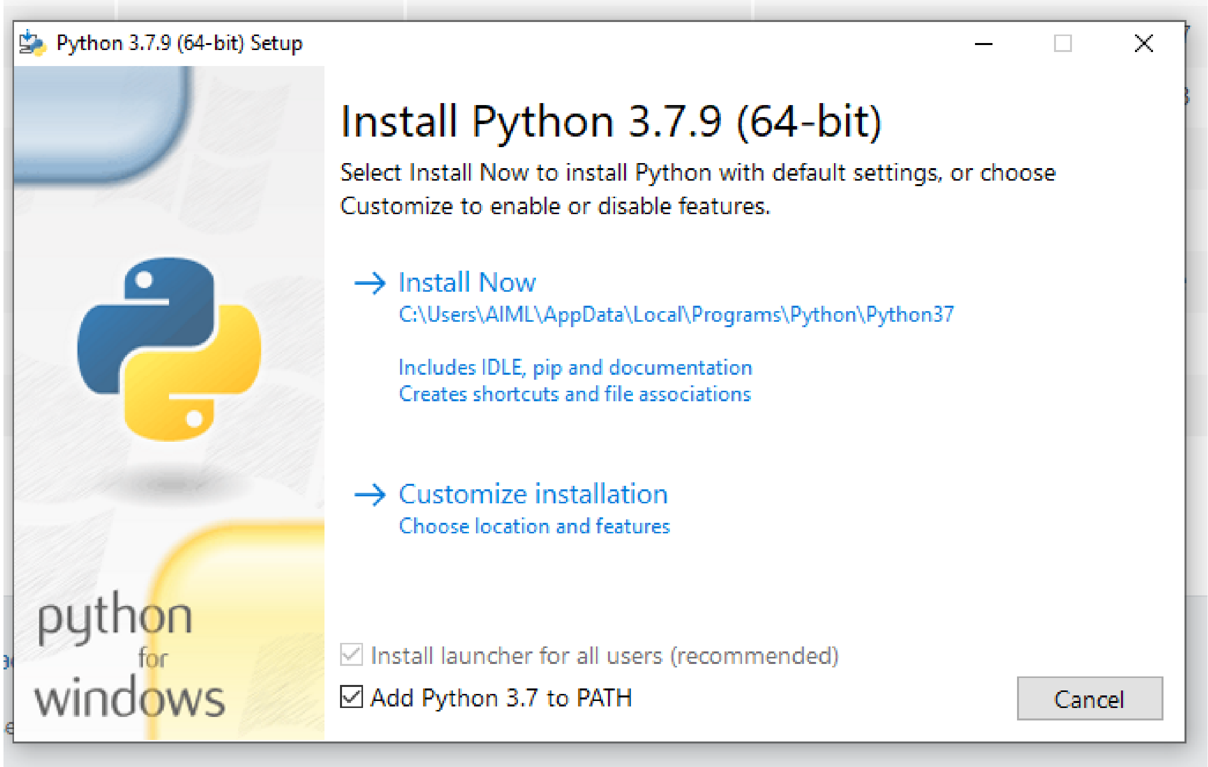
Step2: scroll down click on Windows x86-64 executable installer **download and install**



Files

Version	Operating System	Description	MD5 Sum
Gzipped source tarball	Source release		bcd9f22cf531efc6f06ca6b9b2919bd4
XZ compressed source tarball	Source release		389d3ed26b4d97c741d9e5423da1f43b
macOS 64-bit installer	macOS	for OS X 10.9 and later	4b544fc0ac8c3cffdb67dede23ddb79e
Windows help file	Windows		1094c8d9438ad1adc263ca57ceb3b927
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	60f77740b30030b22699dbd14883a4a3
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64	7083fed513c3c9a4ea655211df9ade27
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	da0b17ae84d6579f8df3eb24927fd825
Windows x86 embeddable zip file	Windows		97c6558d479dc53bf448580b66ad7c1e
Windows x86 executable installer	Windows		1e6d31c98c68c723541f0821b3c15d52
Windows x86 web-based installer	Windows		22f68f09e533c4940fc006e035f08aa2

Installing of python need click on add python 3.7 to path click then give install



Install Python 3.7.9 (64-bit)

Select Install Now to install Python with default settings, or choose Customize to enable or disable features.

→ Install Now
C:\Users\AIML\AppData\Local\Programs\Python\Python37

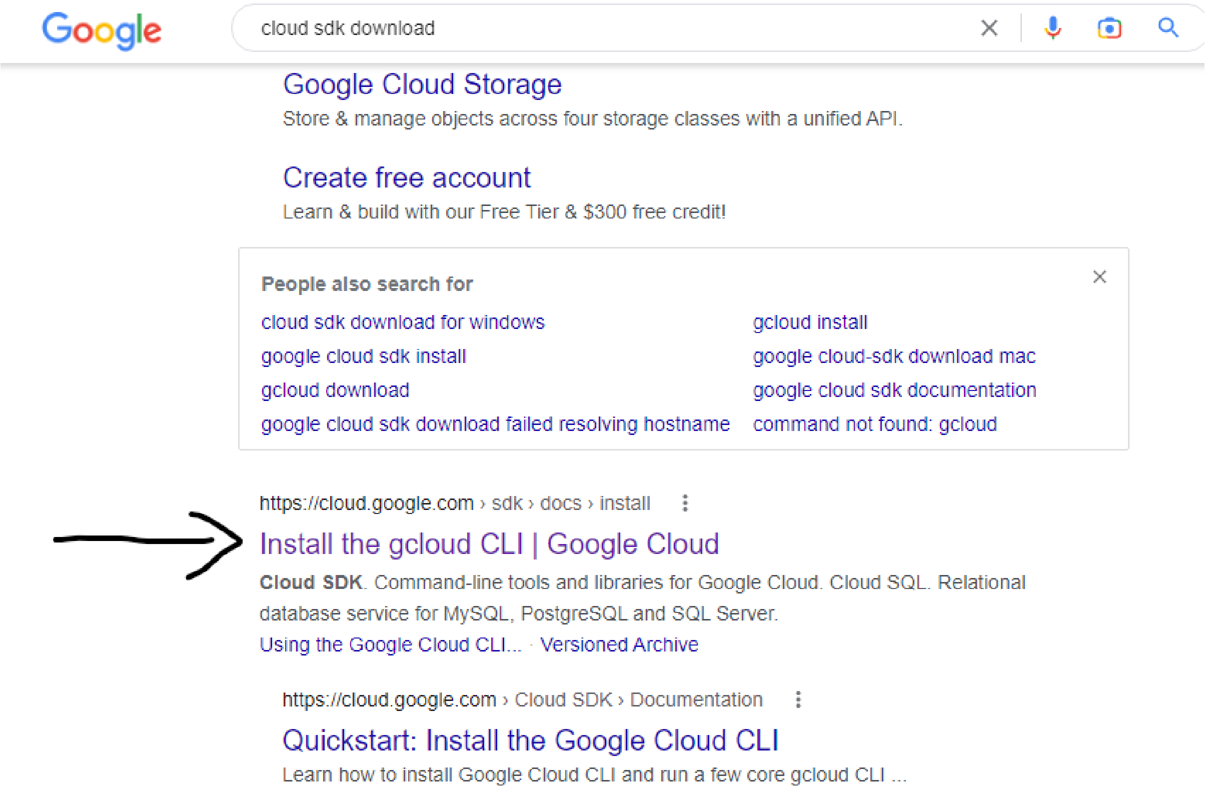
Includes IDLE, pip and documentation
Creates shortcuts and file associations

→ Customize installation
Choose location and features

☑ Install launcher for all users (recommended)

☑ Add Python 3.7 to PATH

Step3: go to Google chrome cloud SDk download and Search download click on first option has shown in the figure



Google Cloud Storage
Store & manage objects across four storage classes with a unified API.

Create free account
Learn & build with our Free Tier & $300 free credit!

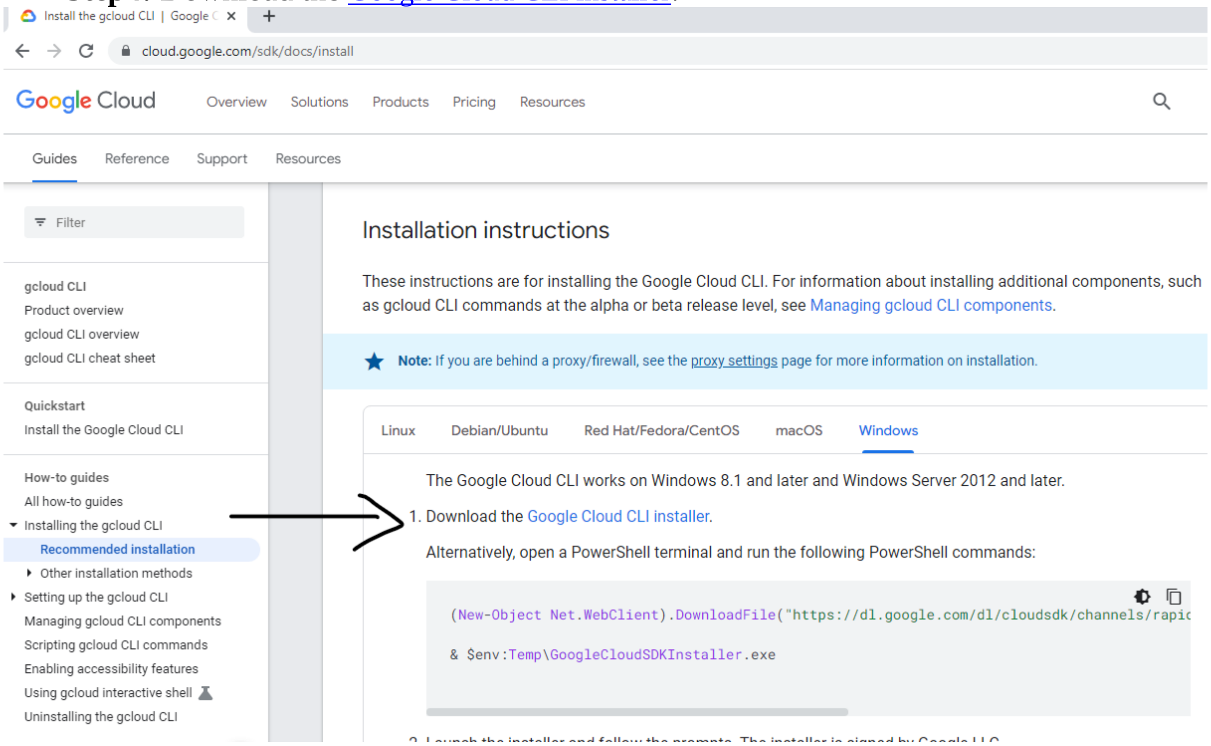
People also search for

cloud sdk download for windows
google cloud sdk install
gcloud download
google cloud sdk download failed resolving hostname
gcloud install
google cloud-sdk download mac
google cloud sdk documentation
command not found: gcloud

https://cloud.google.com › sdk › docs › install
Install the gcloud CLI | Google Cloud
Cloud SDK. Command-line tools and libraries for Google Cloud. Cloud SQL. Relational database service for MySQL, PostgreSQL and SQL Server.
Using the Google Cloud CLI... · Versioned Archive

https://cloud.google.com › Cloud SDK › Documentation
Quickstart: Install the Google Cloud CLI
Learn how to install Google Cloud CLI and run a few core gcloud CLI ...

Step4: Download the Google Cloud CLI installer.



Install the gcloud CLI | Google C
cloud.google.com/sdk/docs/install
Google Cloud
Overview
Solutions
Products
Pricing
Resources
Guides
Reference
Support
Resources
Filter
gcloud CLI
Product overview
gcloud CLI overview
gcloud CLI cheat sheet
Quickstart
Install the Google Cloud CLI
How-to guides
All how-to guides
Installing the gcloud CLI
Recommended installation
Other installation methods
Setting up the gcloud CLI
Managing gcloud CLI components
Scripting gcloud CLI commands
Enabling accessibility features
Using gcloud interactive shell
Uninstalling the gcloud CLI
Linux
Debian/Ubuntu
Red Hat/Fedora/CentOS
macOS
Windows

Installation instructions

These instructions are for installing the Google Cloud CLI. For information about installing additional components, such as gcloud CLI commands at the alpha or beta release level, see Managing gcloud CLI components.

Note: If you are behind a proxy/firewall, see the proxy settings page for more information on installation.

The Google Cloud CLI works on Windows 8.1 and later and Windows Server 2012 and later.

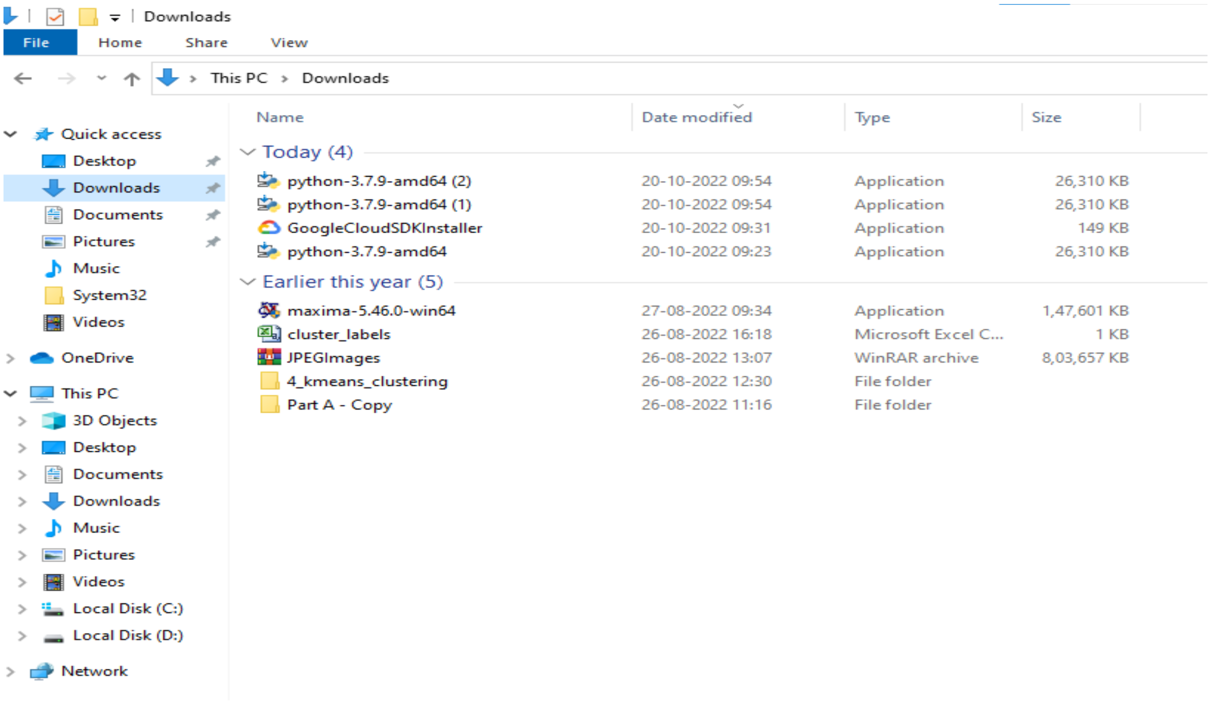
1. Download the Google Cloud CLI installer.

Alternatively, open a PowerShell terminal and run the following PowerShell commands:

```
(New-Object Net.WebClient).DownloadFile("https://dl.google.com/dl/cloudsdk/channels/rapi

& $env:Temp\GoogleCloudSDKInstaller.exe
```

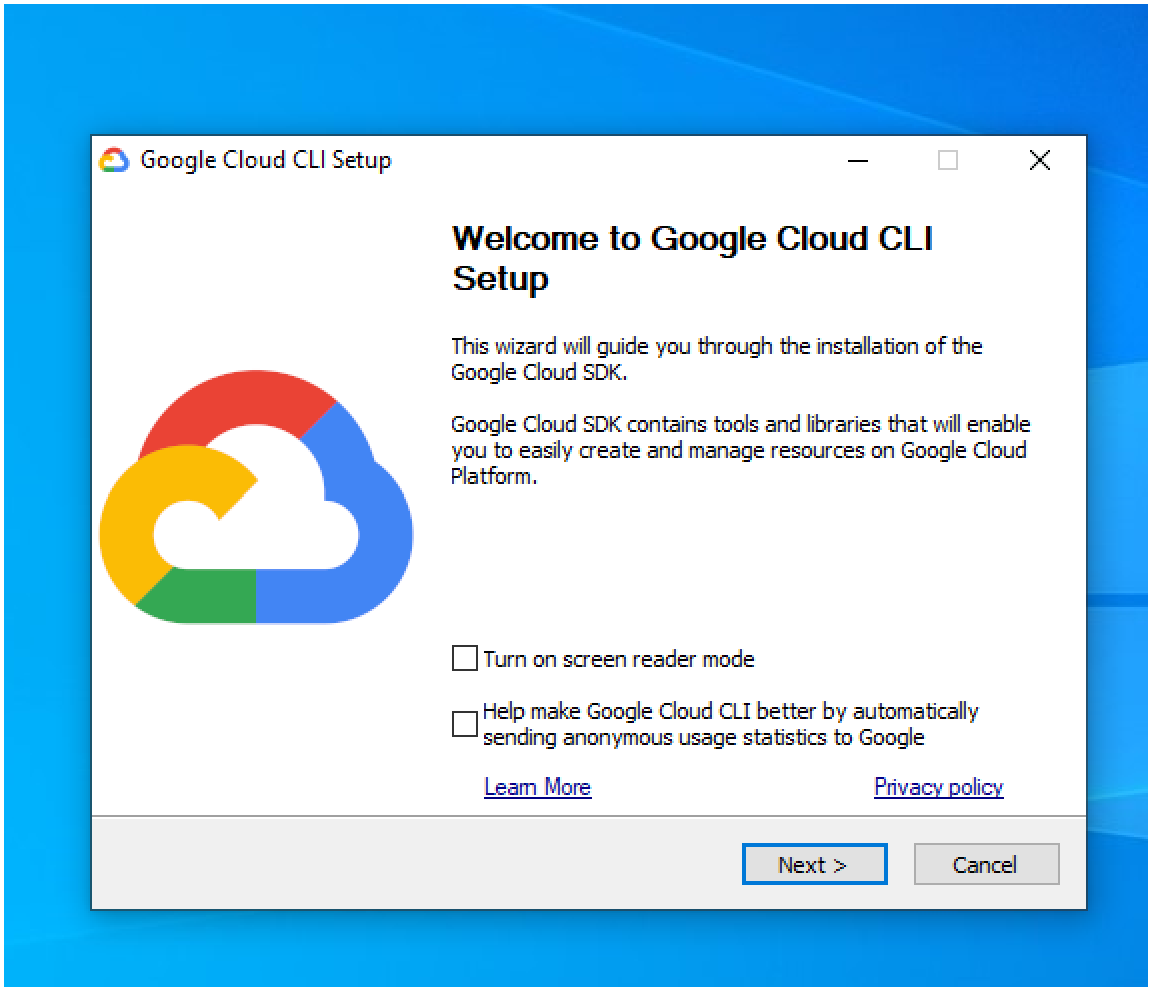
2. Launch the installer and follow the prompts. The installer is signed by Google LLC.

Step5: Double click on the Google App Engine installer.



Name	Date modified	Type	Size
Today (4)			
python-3.7.9-amd64 (2)	20-10-2022 09:54	Application	26,310 KB
python-3.7.9-amd64 (1)	20-10-2022 09:54	Application	26,310 KB
GoogleCloudSDKInstaller	20-10-2022 09:31	Application	149 KB
python-3.7.9-amd64	20-10-2022 09:23	Application	26,310 KB
Earlier this year (5)			
maxima-5.46.0-win64	27-08-2022 09:34	Application	1,47,601 KB
cluster_labels	26-08-2022 16:18	Microsoft Excel C...	1 KB
JPEGImages	26-08-2022 13:07	WinRAR archive	8,03,657 KB
4_kmeans_clustering	26-08-2022 12:30	File folder	
Part A - Copy	26-08-2022 11:16	File folder	

Step6: Click on the Next to Install



Welcome to Google Cloud CLI Setup

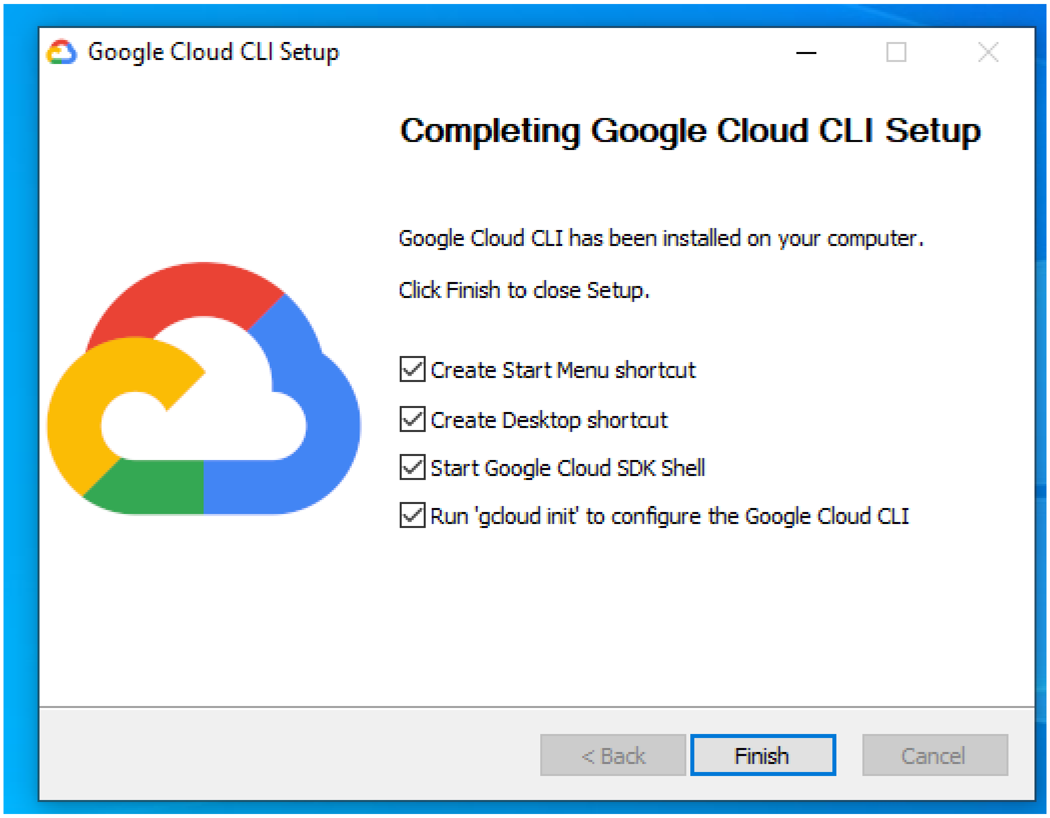
This wizard will guide you through the installation of the Google Cloud SDK.

Google Cloud SDK contains tools and libraries that will enable you to easily create and manage resources on Google Cloud Platform.

□ Turn on screen reader mode

□ Help make Google Cloud CLI better by automatically sending anonymous usage statistics to Google

Step7: Complete install of Google cloud CLI click on finish



Completing Google Cloud CLI Setup

Google Cloud CLI has been installed on your computer.

Click Finish to close Setup.

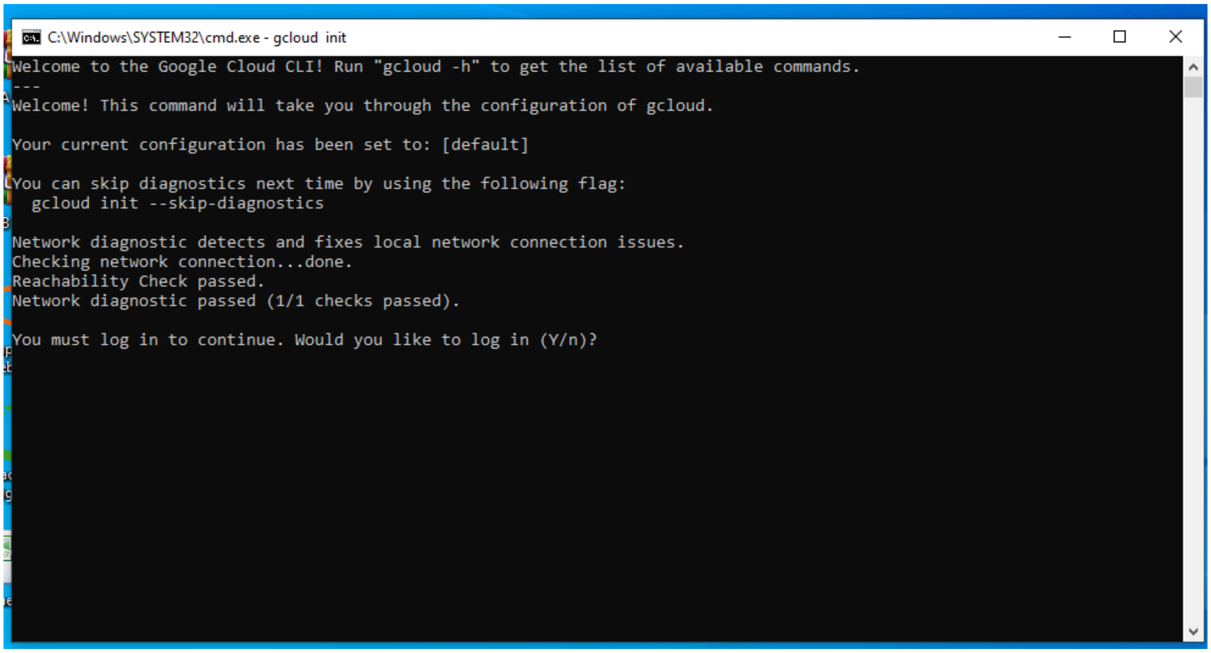
☑ Create Start Menu shortcut

☑ Create Desktop shortcut

☑ Start Google Cloud SDK Shell

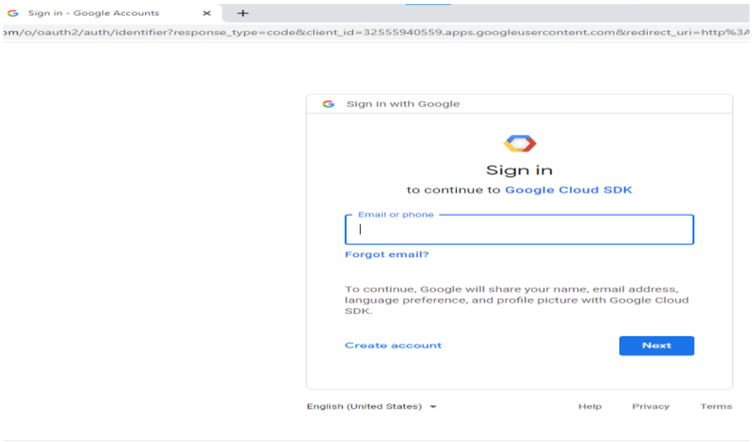
☑ Run 'gcloud init' to configure the Google Cloud CLI

Step 8: click on Google cloud SDK Shell in desktop.



```
Welcome to the Google Cloud CLI! Run "gcloud -h" to get the list of available commands.
---
Welcome! This command will take you through the configuration of gcloud.

Your current configuration has been set to: [default]

You can skip diagnostics next time by using the following flag:
  gcloud init --skip-diagnostics

Network diagnostic detects and fixes local network connection issues.
Checking network connection...done.
Reachability Check passed.
Network diagnostic passed (1/1 checks passed).

You must log in to continue. Would you like to log in (Y/n)?
```

Step9: Sign in Google cloud SDk with your Email-Id click on Next



Sign in - Google Accounts

om/o/oauth2/auth/identifier?response_type=code&client_id=32555940559.apps.googleusercontent.com&redirect_uri=http%3A

Sign in with Google

Sign in

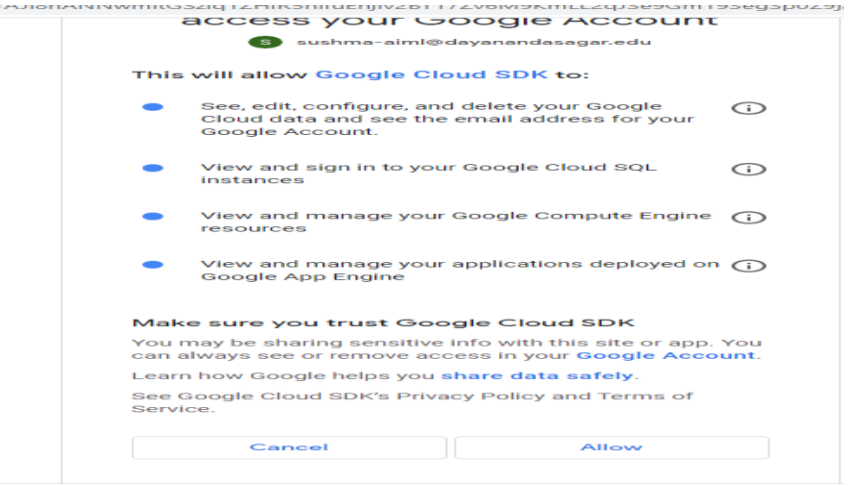
to continue to Google Cloud SDK

Email or phone

Forgot email?

To continue, Google will share your name, email address, language preference, and profile picture with Google Cloud SDK.

Step10: click on allow



access your Google Account

sushma-aiml@dayanandasagar.edu

This will allow Google Cloud SDK to:

- See, edit, configure, and delete your Google Cloud data and see the email address for your Google Account.
- View and sign in to your Google Cloud SQL instances
- View and manage your Google Compute Engine resources
- View and manage your applications deployed on Google App Engine

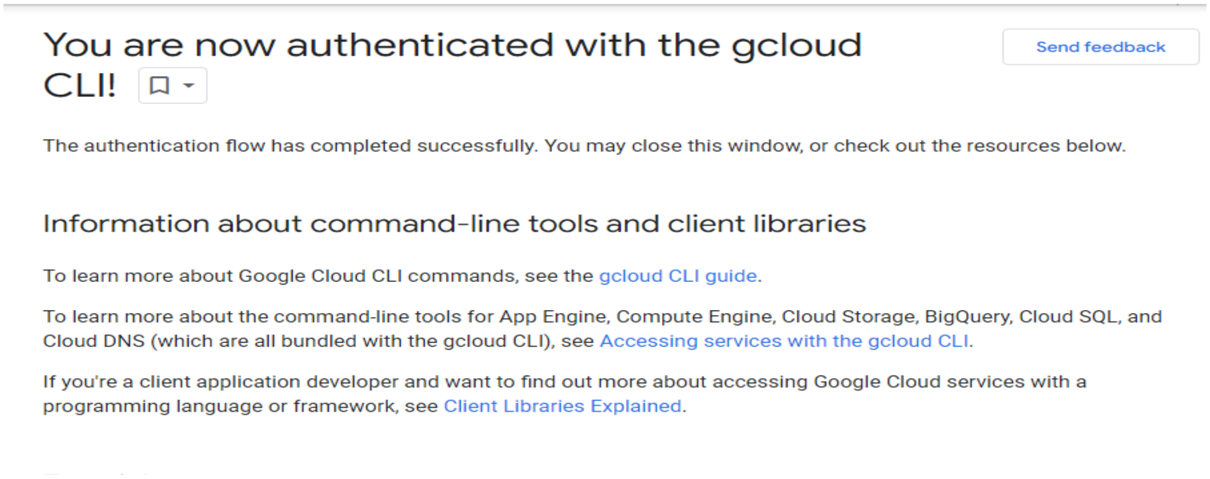
Make sure you trust Google Cloud SDK

You may be sharing sensitive info with this site or app. You can always see or remove access in your Google Account.

Learn how Google helps you share data safely.

See Google Cloud SDK's Privacy Policy and Terms of Service.

Step 11: clicking on allow you get this authenticated window .



You are now authenticated with the gcloud CLI!

The authentication flow has completed successfully. You may close this window, or check out the resources below.

Information about command-line tools and client libraries

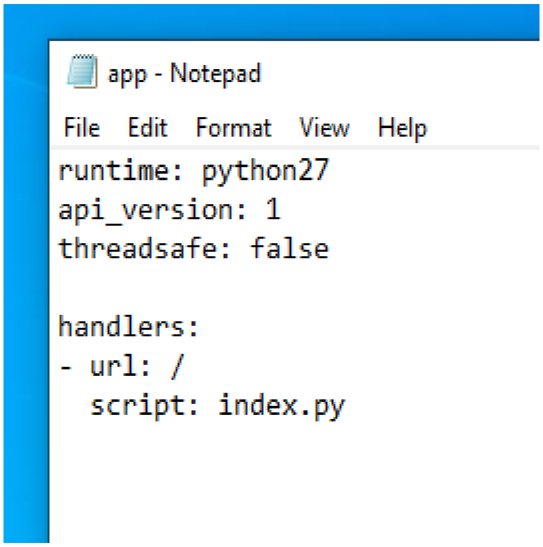
To learn more about Google Cloud CLI commands, see the gcloud CLI guide.

To learn more about the command-line tools for App Engine, Compute Engine, Cloud Storage, BigQuery, Cloud SQL, and Cloud DNS (which are all bundled with the gcloud CLI), see Accessing services with the gcloud CLI.

If you're a client application developer and want to find out more about accessing Google Cloud services with a programming language or framework, see Client Libraries Explained.

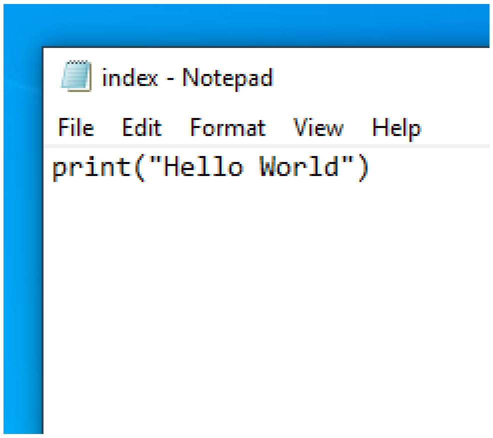
Step12: create an new folder in desktop

Step13: Need to type the program in notepad & save as app.yaml



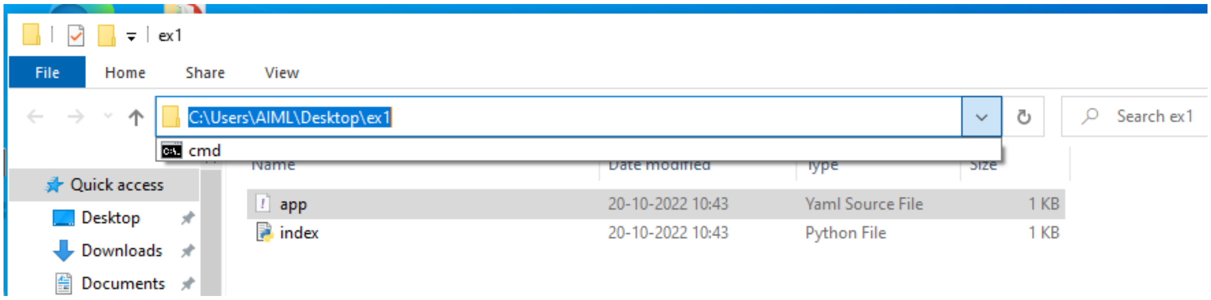
```
runtime: python27
api_version: 1
threadsafe: false

handlers:
- url: /
  script: index.py
```

Step13: save as index.py



```
print("Hello World")
```

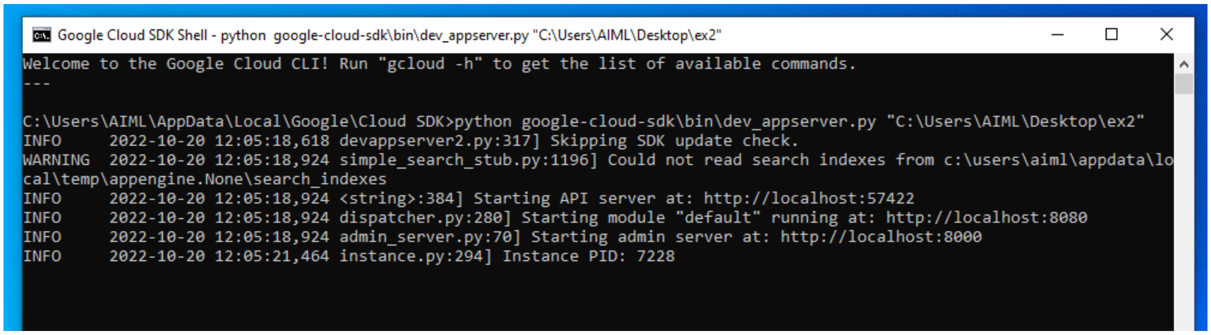
Step14: open the file manger click on desktop copy the file path.



C:\Users\AIML\Desktop\ex1

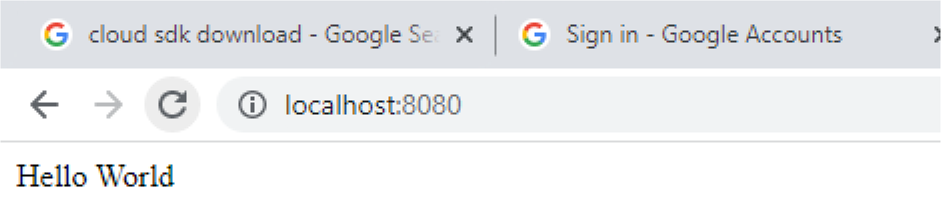
Name	Date modified	Type	Size
app	20-10-2022 10:43	Yaml Source File	1 KB
index	20-10-2022 10:43	Python File	1 KB

Step15: type the command python google-cloud-sdk\bin\dev_appserver.py "C:\Users\AIML\Desktop\ex2".py paste the file path copied for notepad.


Google Cloud SDK Shell - python google-cloud-sdk\bin\dev_appserver.py "C:\Users\AIML\Desktop\ex2"

```
Welcome to the Google Cloud CLI! Run "gcloud -h" to get the list of available commands.
---

C:\Users\AIML\AppData\Local\Google\Cloud SDK>python google-cloud-sdk\bin\dev_appserver.py "C:\Users\AIML\Desktop\ex2"
INFO     2022-10-20 12:05:18,618 devappserver2.py:317] Skipping SDK update check.
WARNING  2022-10-20 12:05:18,924 simple_search_stub.py:1196] Could not read search indexes from c:\users\aiml\appdata\lo
cal\temp\appengine.None\search_indexes
INFO     2022-10-20 12:05:18,924 <string>:384] Starting API server at: http://localhost:57422
INFO     2022-10-20 12:05:18,924 dispatcher.py:280] Starting module "default" running at: http://localhost:8080
INFO     2022-10-20 12:05:18,924 admin_server.py:70] Starting admin server at: http://localhost:8000
INFO     2022-10-20 12:05:21,464 instance.py:294] Instance PID: 7228
```

Step16: copy the url http://localhost:8080& paste in Google chrome


cloud sdk download - Google Se
Sign in - Google Accounts
localhost:8080

```
Hello World
```

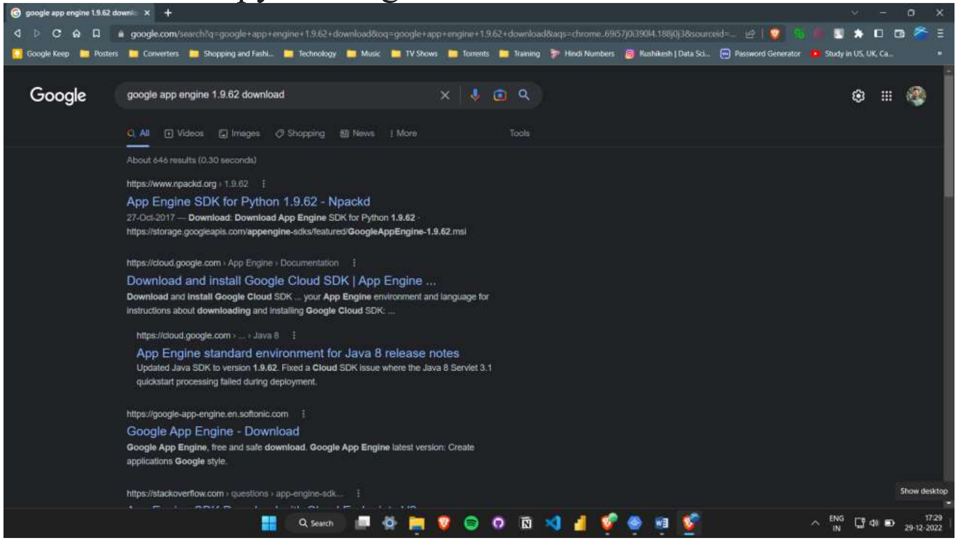
Ex. No. : 4 Use GAE Launcher to launch the web applications

Aim:

To use GAE launcher to launch the web applications

Procedure:

1. Install Google App Engine:
   a. Install Google App Engine 1.9.62
      i. Go To : "https://www.npackd.org/p/com.google.AppEnginePythonSDK/1.9.62" and install the 1.9.62 version
      Make sure Python 2.7 version is installed in your system, if not download it on python.org



google app engine 1.9.62 download

About 646 results (0.30 seconds)

https://www.npackd.org › 1.9.62
App Engine SDK for Python 1.9.62 - Npackd
27-Oct-2017 — Download: Download App Engine SDK for Python 1.9.62 · https://storage.googleapis.com/appengine-sdks/featured/GoogleAppEngine-1.9.62.msi

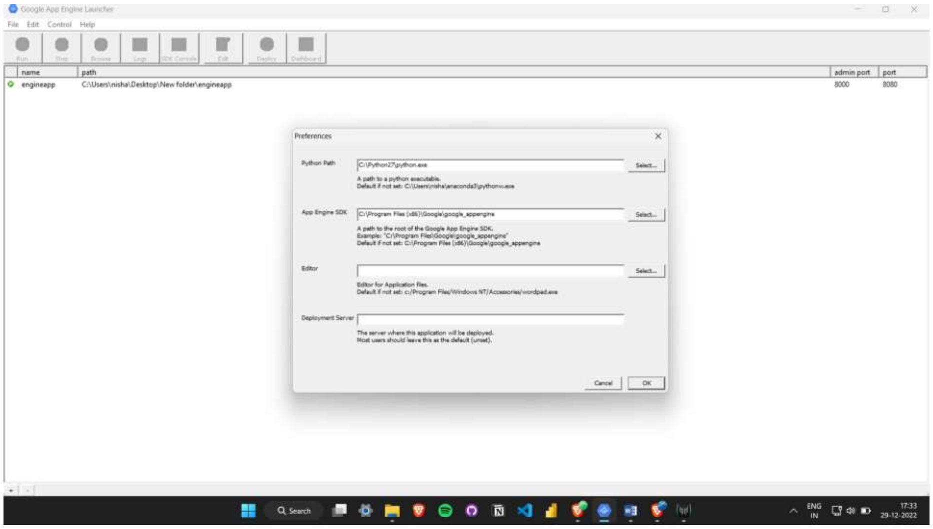
https://cloud.google.com › App Engine › Documentation
Download and install Google Cloud SDK | App Engine ...
Download and install Google Cloud SDK ... your App Engine environment and language for instructions about downloading and installing Google Cloud SDK: ...

https://cloud.google.com › ... › Java 8
App Engine standard environment for Java 8 release notes
Updated Java SDK to version 1.9.62. Fixed a Cloud SDK issue where the Java 8 Servlet 3.1 quickstart processing failed during deployment.

https://google-app-engine.en.softonic.com
Google App Engine - Download
Google App Engine, free and safe download. Google App Engine latest version: Create applications Google style.

https://stackoverflow.com › questions › app-engine-sdk...

   b. Create a new folder on Desktop
   c. Go to Google App Engine and click on Edit-> Preferences-> Set the path of Google App Engine and Python present in Program Files


Google App Engine Launcher
File Edit Control Help
Run Stop Browse Logs SDK Console Edit Deploy Dashboard
17:33
29-12-2022

	name	path	admin port	port
●	engineapp	C:\Users\nisha\Desktop\New folder\engineapp	8000	8080

Preferences

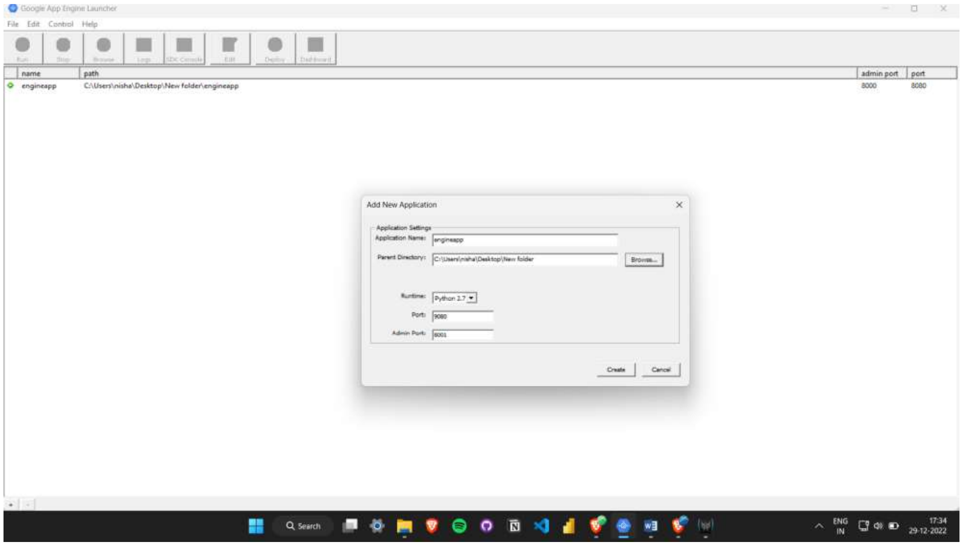
Python Path: C:\Python27\python.exe
A path to a python executable.
Default if not set: C:\Users\nisha\anaconda3\pythonw.exe

App Engine SDK: C:\Program Files (x86)\Google\google_appengine
A path to the root of the Google App Engine SDK.
Example: "C:\Program Files\Google\google_appengine"
Default if not set: C:\Program Files (x86)\Google\google_appengine

Editor:
Editor for Application files.
Default if not set: c:/Program Files/Windows NT/Accessories/wordpad.exe

Deployment Server:
The server where this application will be deployed.
Most users should leave this as the default (unset).

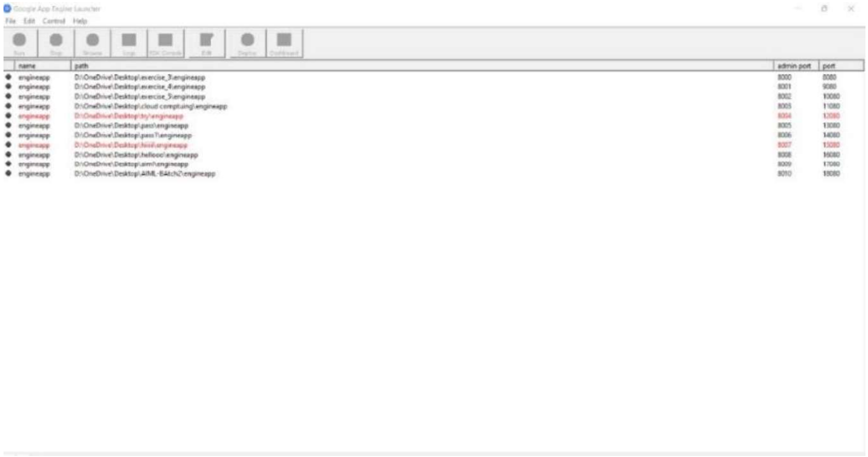
d. Click on File->Create new Application and select the path of the folder created earlier.


Google App Engine Launcher
File Edit Control Help
Run Stop Browse Logs SDK Console Edit Deploy Dashboard
17:34
29-12-2022

	name	path	admin port	port
●	engineapp	C:\Users\nisha\Desktop\New folder\engineapp	8000	8080

Add New Application

Application Settings
Application Name: engineapp
Parent Directory: C:\Users\nisha\Desktop\New folder
Runtime: Python 2.7
Port: 9080
Admin Port: 8001

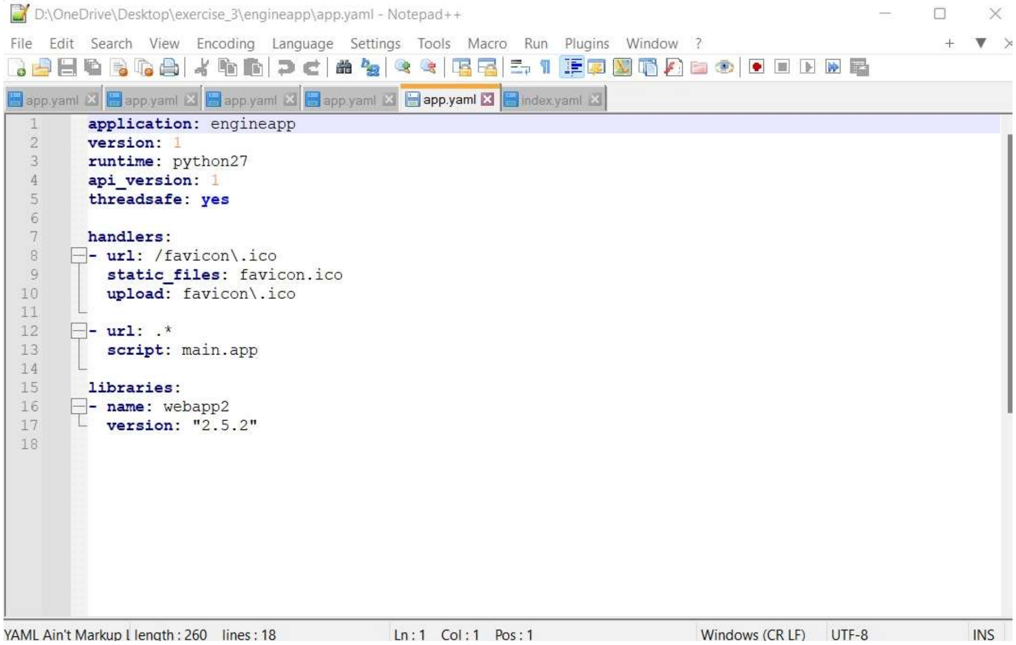

Google App Engine Launcher
File Edit Control Help
Run Stop Browse Logs SDK Console Edit Deploy Dashboard

	name	path	admin port	port
●	engineapp	D:\OneDrive\Desktop\exercise_3\engineapp	8000	8080
●	engineapp	D:\OneDrive\Desktop\exercise_4\engineapp	8001	9080
●	engineapp	D:\OneDrive\Desktop\exercise_5\engineapp	8002	10080
●	engineapp	D:\OneDrive\Desktop\cloud comptuing\engineapp	8003	11080
●	engineapp	D:\OneDrive\Desktop\try\engineapp	8004	12080
●	engineapp	D:\OneDrive\Desktop\pass\engineapp	8005	13080
●	engineapp	D:\OneDrive\Desktop\pass1\engineapp	8006	14080
●	engineapp	D:\OneDrive\Desktop\hiiii\engineapp	8007	15080
●	engineapp	D:\OneDrive\Desktop\hellooo\engineapp	8008	16080
●	engineapp	D:\OneDrive\Desktop\aiml\engineapp	8009	17080
●	engineapp	D:\OneDrive\Desktop\AIML-BAtch2\engineapp	8010	18080

d. Open the created folder :

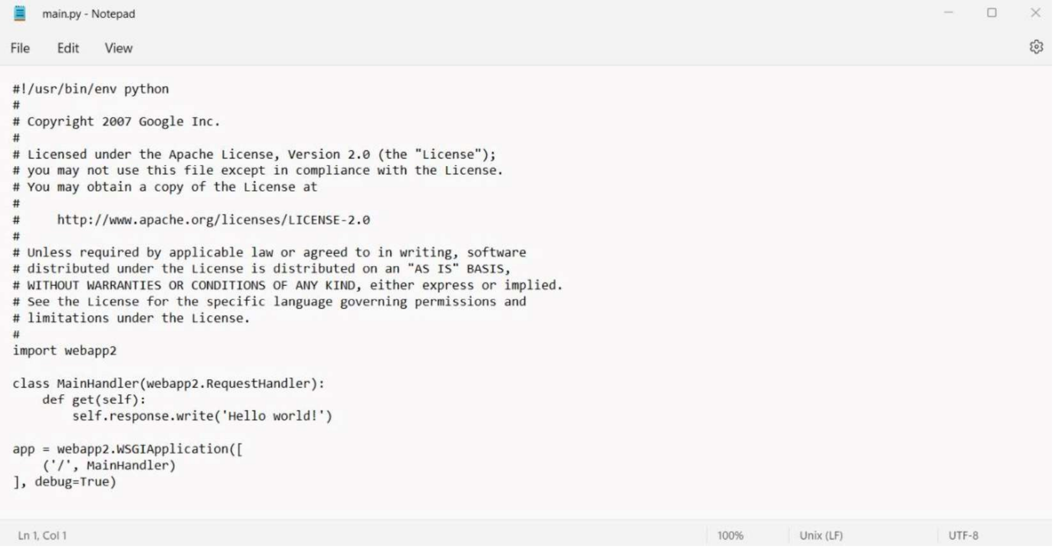
i. Engineapp created automatically

ii. Folder or Project directory must contain `app.yaml` file which contains instruction for GoogleApp Engine To provision the resources for your app.


D:\OneDrive\Desktop\exercise_3\engineapp\app.yaml - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
app.yaml
app.yaml
app.yaml
app.yaml
app.yaml
index.yaml
YAML Ain't Markup L length : 260 lines : 18
Ln : 1 Col : 1 Pos : 1
Windows (CR LF) UTF-8 INS

```
application: engineapp
version: 1
runtime: python27
api_version: 1
threadsafe: yes

handlers:
- url: /favicon\.ico
  static_files: favicon.ico
  upload: favicon\.ico

- url: .*
  script: main.app

libraries:
- name: webapp2
  version: "2.5.2"
```

iii. Contain `main.py` or any other file will handle all the logic.



```
#!/usr/bin/env python
#
# Copyright 2007 Google Inc.
#
# Licensed under the Apache License, Version 2.0 (the "License");
# you may not use this file except in compliance with the License.
# You may obtain a copy of the License at
#
#     http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#
import webapp2

class MainHandler(webapp2.RequestHandler):
    def get(self):
        self.response.write('Hello world!')

app = webapp2.WSGIApplication([
    ('/', MainHandler)
], debug=True)
```

Deploying web service:

- Run this app in Google App Engine Launcher
- Launch your browser and access your web service in localhost:port_no of your app

Result:

A web application has been successfully deployed to Google App Engine

PRACTICAL NO – 1

Aim: Installation of Single Node Hadoop Cluster on Ubuntu

```
Prerequisite Test
============================
sudo apt update
sudo apt install openjdk-8-jdk -y

java -version; javac -version
sudo apt install openssh-server openssh-client -y
sudo adduser hdoop
su - hdoop
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys
ssh localhost
```

Downloading Hadoop (Check the path where Hadoop has installed and set HADOOP_HOME to this path)

```
==============================
wget https://downloads.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
tar xzf hadoop-3.2.3.tar.gz
```

```
Editng 6 important files
================================
1st file
==========================
sudo nano .bashrc -  here you might face issue saying hdoop is not sudo user
if this issue comes then
su - aman
sudo adduser hdoop sudo

sudo nano .bashrc
#Add below lines in this file

#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.2.3
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS"-Djava.library.path=$HADOOP_HOME/lib/nativ"


source ~/.bashrc

2nd File
==============================
sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh

#Add below line in this file in the end

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

3rd File
==================================
sudo nano $HADOOP_HOME/etc/hadoop/core-site.xml

#Add below lines in this file(between "<configuration>" and "<"/configuration>")
  <property>
      <name>hadoop.tmp.dir</name>
      <value>/home/hdoop/tmpdata</value>
      <description>A base for other temporary directories.</description>
   </property>
   <property>
      <name>fs.default.name</name>
      <value>hdfs://localhost:9000</value>
      <description>The name of the default file system></description>
   </property>

4th File
========================================
sudo nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml

#Add below lines in this file(between "<configuration>" and "<"/configuration>")


<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/namenode</value>
```

```
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>



5th File
===============================================

sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml

#Add below lines in this file(between "<configuration>" and "<"/configuration>")

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

6th File
=================================================
sudo nano $HADOOP_HOME/etc/hadoop/yarn-site.xml

#Add below lines in this file(between "<configuration>" and "<"/configuration>")

<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
</property>
<property>
```

```
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
  <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
```

Launching Hadoop (Change to sbin path and run)
===================================

```
hdfs namenode -format

./start-dfs.sh
./start-yarn.sh
```

To check all the daemons of Hadoop s running give JPS and check

Dayananda Sagar College of Engineering

An Autonomous Institute Affiliated to VTU, Belagavi

Approved by AICTE & UGC, Accredited by NAAC with 'A' Grade and NBA

Department of Artificial Intelligence & Machine Learning

PRACTICAL NO – 2

AIM: Hadoop Programming: Word Count MapReduce Program Using Eclipse

The entire MapReduce program can be fundamentally divided into three parts:

- Driver Code
- Mapper Phase Code
- Reducer Phase Code

1. Driver Code

```
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;
    }
    public static void main(String args[]) throws Exception
    {
        int exitCode = ToolRunner.run(new WCDriver(), args);
        System.out.println(exitCode);
    }
}
```

2. Mapper Code

```
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>
{
    public void map(LongWritable key, Text value, OutputCollector<Text,
        IntWritable> output, Reporter rep) throws IOException
        {
            String line = value.toString();

     // Splitting the line on spaces
     for (String word : line.split(" "))
     {
        if (word.length() > 0)
        {
           output.collect(new Text(word), new IntWritable(1));
        }
     }
   }
    }
```

3. Reducer Code

```
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>
{
            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));
}
}
```

Run the MapReduce code:

The command for running a MapReduce code is:

```
hadoop      jar     hadoop-mapreduce-example.jar WordCount
/sample/input /sample/output
```

Dayananda Sagar College of Engineering

An Autonomous Institute Affiliated to VTU, Belagavi

Approved by AICTE & UGC, Accredited by NAAC with 'A' Grade and NBA

Department of Artificial Intelligence & Machine Learning

PRACTICAL NO – 3

AIM: File Management tasks in Hadoop using HDFS commands

a. Adding files to HDFS
b. Retrieving files from HDFS
c. Deleting files from HDFS

1. Create a directory in HDFS atgiven path(s).

Usage:

```
hadoop fs -mkdir <paths>
```

Example:

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

2. List the contents of adirectory.

Usage :

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

Example:

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

3. Upload and download a file inHDFS.

Upload: **hadoop fs**

–put:

Copy single src file, or multiple src files from local file systemto the Hadoop data file system

Usage:

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

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

Download:

hadoop fs -get:

Copies/Downloads files to the local file system

Usage:

hadoop fs -get <hdfs_src> <localdst>

Example: hadoop fs get

/user/saurzcode/dir3/Samplefile.txt /home/

4. See contents of a file

Same as unix cat command:

Usage:

hadoop fs -cat <path[filename]>Example:

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

5. Remove a file or directory inHDFS.

Remove files specified as argument. Deletes directory onlywhen it is empty

Usage :

hadoop fs -rm <arg>

Example: `hadoop fs -rm /user/saurzcode/dir1/abc.txt`

6. Display last few lines of a file.

Similar to tail command in Unix.

Usage :

```
hadoop fs -tail <path[filename]>Example:
```

```
hadoop fs –tail

/user/saurzcode/dir/

ab.txt
```

Dayananda Sagar College of Engineering

An Autonomous Institute Affiliated to VTU, Belagavi

Approved by AICTE & UGC, Accredited by NAAC with 'A' Grade and NBA

Department of Artificial Intelligence & Machine Learning

PRACTICAL NO – 4

AIM: Creation of DataFrames using CreateDataFrame working with an example related to FIFA dataset.

Creation of DataFrame in Spark

- Let us use the following code to create a new DataFrame.
- Here, we shall create a new DataFrame using the createDataFrame method.
- we shall design the schema for the data that we will read from fifa csv file.
- Finally, let us use the createDataFrame method to create our DataFrame
- Hence, we create DataFrame and display it by using the .show method.

Before we read the data from a CSV file, we need to import certain libraries which we need for processing the DataFrames in Spark.

```
import org.apache.spark._

import org.apache.spark.sql._

import org.apache.spark.sql.types._

import org.apache.spark.storage.StorageLevel

import scala.collection.mutable.HashMap

import java.io.File

import org.apache.spark.sql.Row

import org.apache.spark.util.IntParam

import scala.collection.mutable.ListBuffer

import org.apache.spark.sql.types.{StructType, StructField, StringType, IntegerType};
```

- We design the schema for our CSV file once we import libraries,

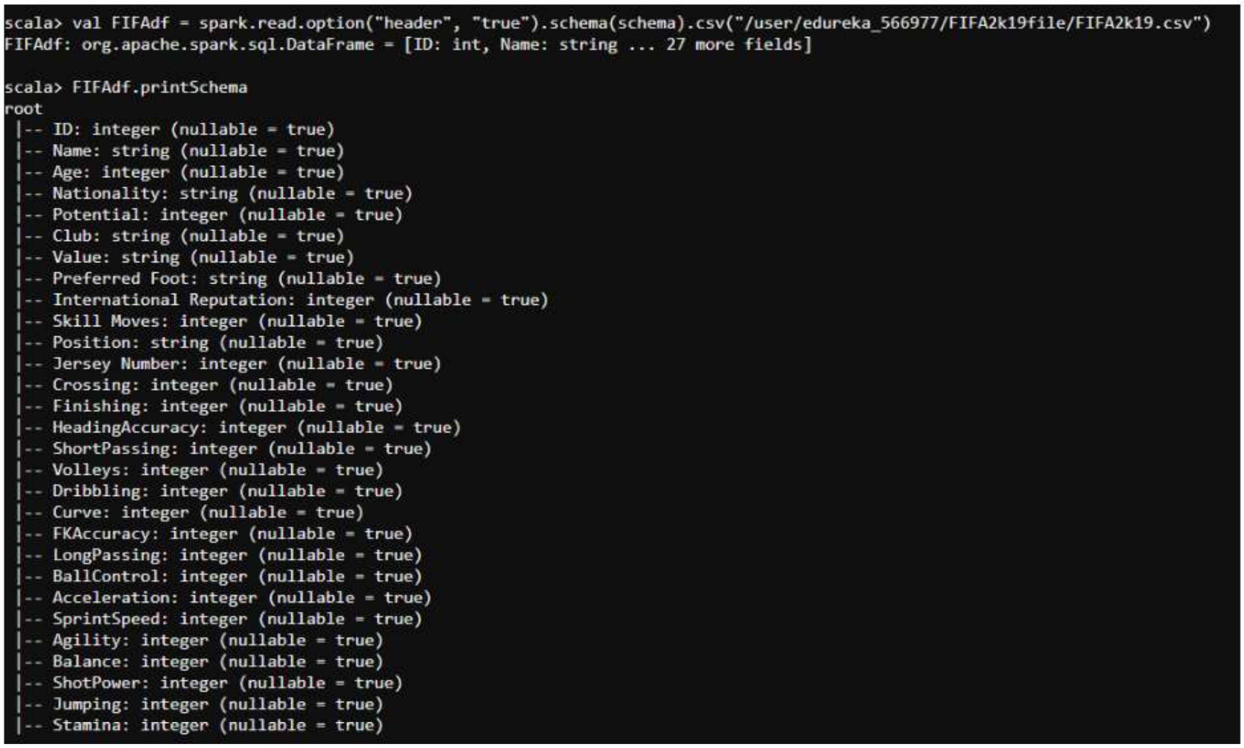
```
val schema =


StructType(Array(StructField("ID",IntegerType,true),StructField("Name",StringType,true),Struc
```

tField("Age",IntegerType,true),StructField("Nationality",StringType,true),StructField("Potential",IntegerType,true),StructField("Club",StringType,true),StructField("Value",StringType,true),StructField("PreferredFoot",StringType,true),StructField("InternationalReputation",IntegerType,true),StructField("SkillMoves",IntegerType,true),StructField("Position",StringType,true),StructField("JerseyNumber",IntegerType,true),StructField("Crossing",IntegerType,true),StructField("Finishing",IntegerType,true),StructField("HeadingAccuracy",IntegerType,true),StructField("ShortPassing",IntegerType,true),StructField("Volleys",IntegerType,true),StructField("Dribbling",IntegerType,true),StructField("Curve",IntegerType,true),StructField("FKAccuracy",IntegerType,true),StructField("LongPassing",IntegerType,true),StructField("BallControl",IntegerType,true),StructField("Acceleration",IntegerType,true),StructField("SprintSpeed",IntegerType,true),StructField("Agility",IntegerType,true),StructField("Balance",IntegerType,true),StructField("ShotPower",IntegerType,true),StructField("Jumping",IntegerType,true),StructField("Stamina",IntegerType,true)))

- Let us load the Fifa data from a CSV file from the HDFS as shown below. We are first going to use Spark.read.format("csv") method for reading our CSV file from our HDFS.

Val FifAdfspark.read.format("csv").option("header",true).load("/home/rash/Downloads/players_16.csv")

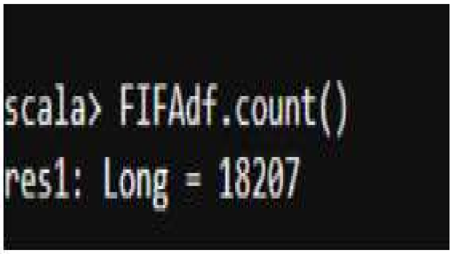
- Let us use .printSchema() method to see the schema of our CSV file.



```
scala> val FIFAdf = spark.read.option("header", "true").schema(schema).csv("/user/edureka_566977/FIFA2k19file/FIFA2k19.csv")
FIFAdf: org.apache.spark.sql.DataFrame = [ID: int, Name: string ... 27 more fields]

scala> FIFAdf.printSchema
root
 |-- ID: integer (nullable = true)
 |-- Name: string (nullable = true)
 |-- Age: integer (nullable = true)
 |-- Nationality: string (nullable = true)
 |-- Potential: integer (nullable = true)
 |-- Club: string (nullable = true)
 |-- Value: string (nullable = true)
 |-- Preferred Foot: string (nullable = true)
 |-- International Reputation: integer (nullable = true)
 |-- Skill Moves: integer (nullable = true)
 |-- Position: string (nullable = true)
 |-- Jersey Number: integer (nullable = true)
 |-- Crossing: integer (nullable = true)
 |-- Finishing: integer (nullable = true)
 |-- HeadingAccuracy: integer (nullable = true)
 |-- ShortPassing: integer (nullable = true)
 |-- Volleys: integer (nullable = true)
 |-- Dribbling: integer (nullable = true)
 |-- Curve: integer (nullable = true)
 |-- FKAccuracy: integer (nullable = true)
 |-- LongPassing: integer (nullable = true)
 |-- BallControl: integer (nullable = true)
 |-- Acceleration: integer (nullable = true)
 |-- SprintSpeed: integer (nullable = true)
 |-- Agility: integer (nullable = true)
 |-- Balance: integer (nullable = true)
 |-- ShotPower: integer (nullable = true)
 |-- Jumping: integer (nullable = true)
 |-- Stamina: integer (nullable = true)
```

- Let us find out the total number of rows we have using the following code.

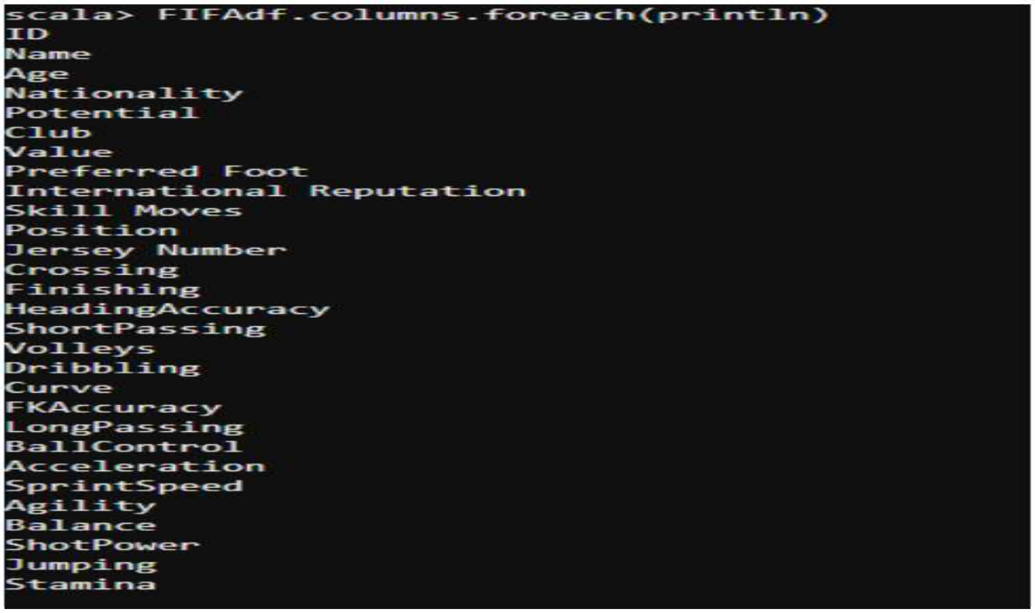
  FifaAdf.count()



```
scala> FIFAdf.count()
res1: Long = 18207
```

- Let us now find the columns we have in our CSV file. We shall use the following code.

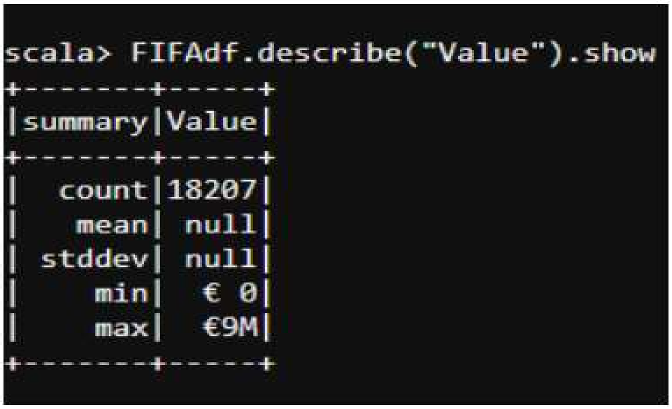
  FifaAdf.columns.foreach(println)



```
scala> FIFAdf.columns.foreach(println)
ID
Name
Age
Nationality
Potential
Club
Value
Preferred Foot
International Reputation
Skill Moves
Position
Jersey Number
Crossing
Finishing
HeadingAccuracy
ShortPassing
Volleys
Dribbling
Curve
FKAccuracy
LongPassing
BallControl
Acceleration
SprintSpeed
Agility
Balance
ShotPower
Jumping
Stamina
```

- If you wish to look at the summary of a particular column in a DataFrame, we can apply to describe command. This command will give us the **statistical summary** of a particular selected column if nothing is specified, and then it provides the statistical information of the DataFrame.
- Let us find out the description of the Value column to know the minimum and maximum values present in it.

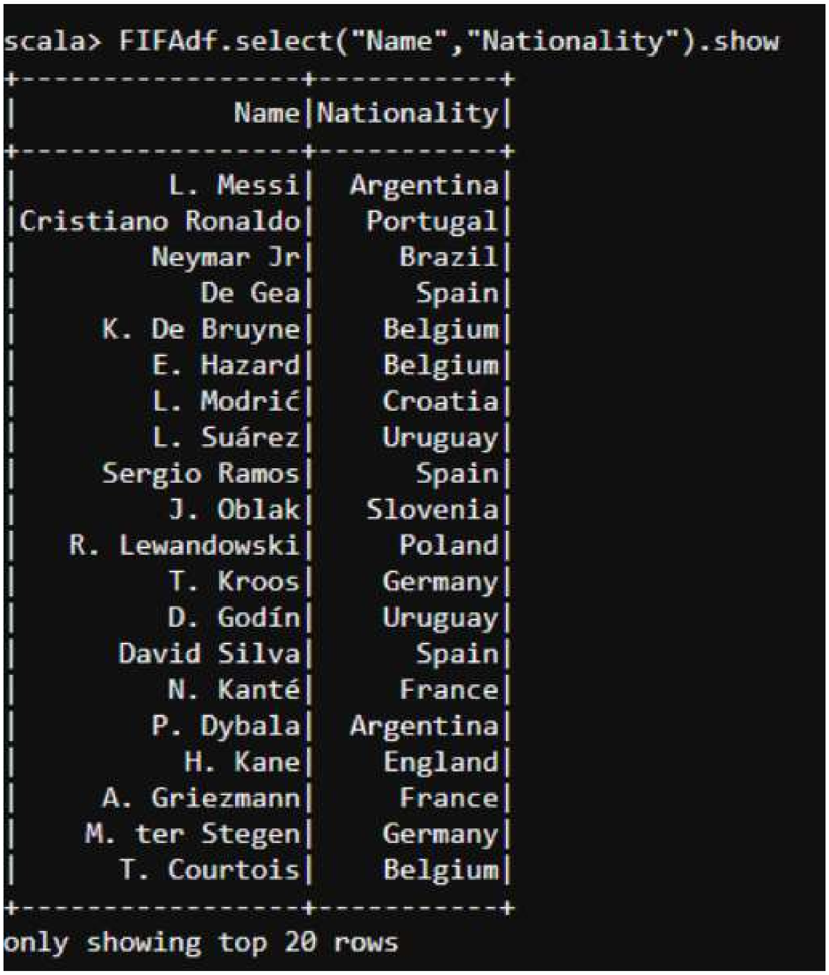
  FifaAdf.describe("Value").show



```
scala> FIFAdf.describe("Value").show
+-------+-----+
|summary|Value|
+-------+-----+
|  count|18207|
|   mean| null|
| stddev| null|
|    min|  € 0|
|    max|  €9M|
+-------+-----+
```

- We shall find out the Nationality of a particular player by using the select command.

  FifaAdf.select("name","Nationality").show



```
scala> FIFAdf.select("Name","Nationality").show
+-----------------+-----------+
|             Name|Nationality|
+-----------------+-----------+
|         L. Messi|  Argentina|
|Cristiano Ronaldo|   Portugal|
|        Neymar Jr|     Brazil|
|           De Gea|      Spain|
|     K. De Bruyne|    Belgium|
|        E. Hazard|    Belgium|
|        L. Modrić|    Croatia|
|        L. Suárez|    Uruguay|
|     Sergio Ramos|      Spain|
|         J. Oblak|   Slovenia|
|   R. Lewandowski|     Poland|
|         T. Kroos|    Germany|
|         D. Godín|    Uruguay|
|      David Silva|      Spain|
|         N. Kanté|     France|
|        P. Dybala|  Argentina|
|          H. Kane|    England|
|     A. Griezmann|     France|
|    M. ter Stegen|    Germany|
|      T. Courtois|    Belgium|
+-----------------+-----------+
only showing top 20 rows
```