SONA COLLEGE OF TECHNOLOGY
(AUTONOMOUS)
SONA NAGAR, SALEM – 636005.

[An Autonomous Institution]

Name : …….

Branch : ……..

Semester : ……..

Register No : ……..

Name of the Laboratory : ……..

Certified that this is the bona-fide Record of practical work done by the above student in _____ *during the year* _____-_____.

Lab-In-Charge Head of the Department

Submitted for University Practical Examination held on …….

Internal Examiner External Examiner

CONTENTS

S.No	Date	Name of the Experiment	Page No	Marks Awarded	Signature of the Staff

CONTENTS

S.No	Date	Name of the Experiment	Page No	Marks Awarded	Signature of the Staff

EX NO: 1 DATE:	Run the VirtualBox/VMware Workstation with different flavors of Linux or Windows OS

AIM:

To find the procedure to Install VirtualBox/VMware Workstation with different flavors of Linux or Windows OS.

PROCEDURE:

Step 1: Download and install Oracle VM VirtualBox/ VMware in Windows platform (Host OS) https://www.virtualbox.org/

Step 2: Visit the page http://www.ubuntu.com/download/ubuntu/download

Step 3: Choose the Latest version of Ubuntu and 32-bit and click “Start Download”

Step 4: Run VirtualBox by double-clicking the icon and Click “New” button on the top left corner

Step 5: Click “Continue” on the pop-up window and Type VM name, select Linux for the OS and choose “Ubuntu” for the version

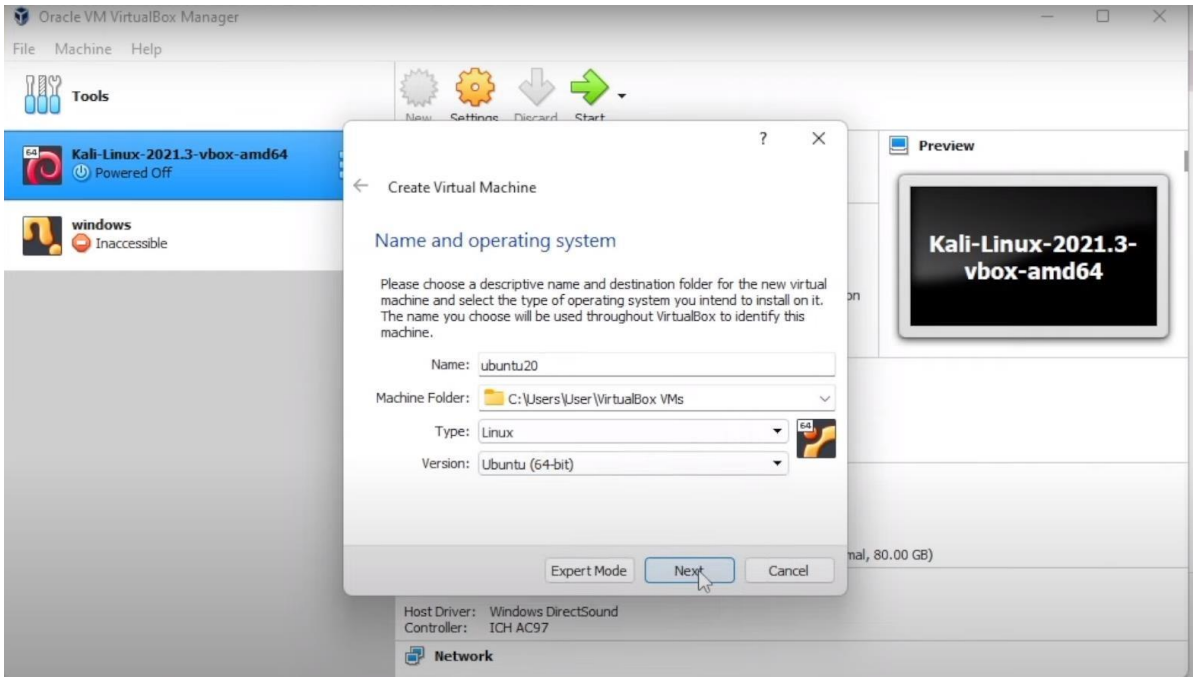
Step 6: Choose the amount of memory to allocate (choosing between 512 MB to 1024 MB) and Click Continue or Next

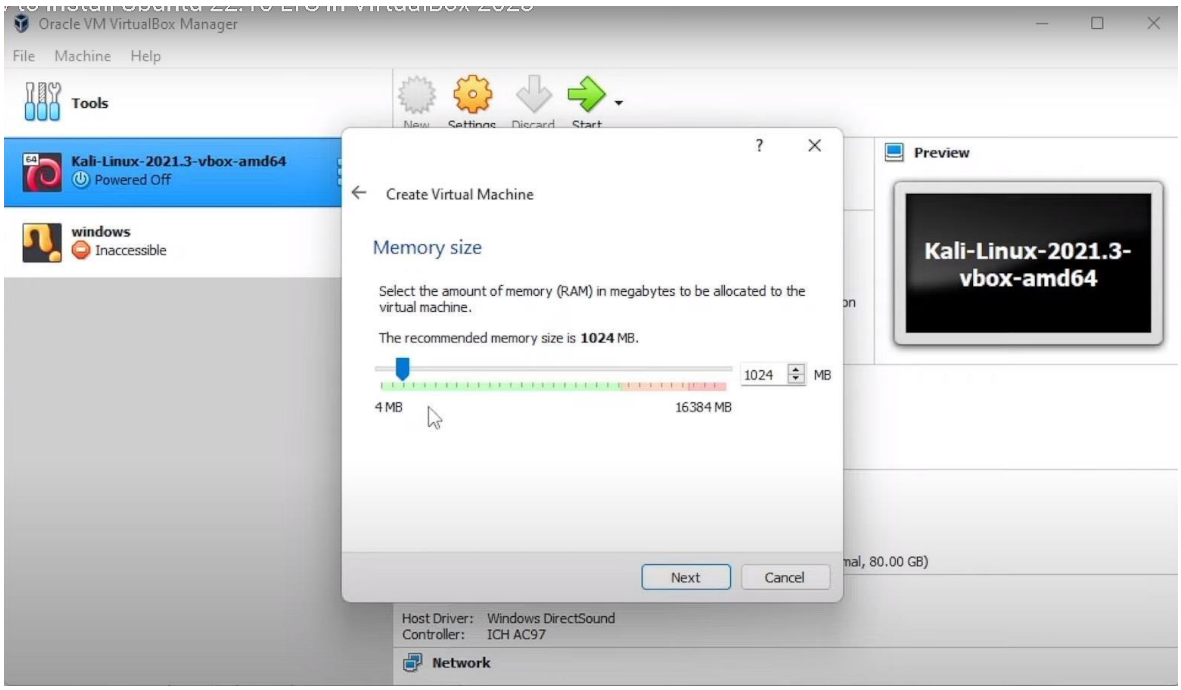
Step 7: Choose create a new virtual hard disk and Click Continue or Next Step 8: Choose VDI (VirtualBox Disk Image) and Click Continue or Create

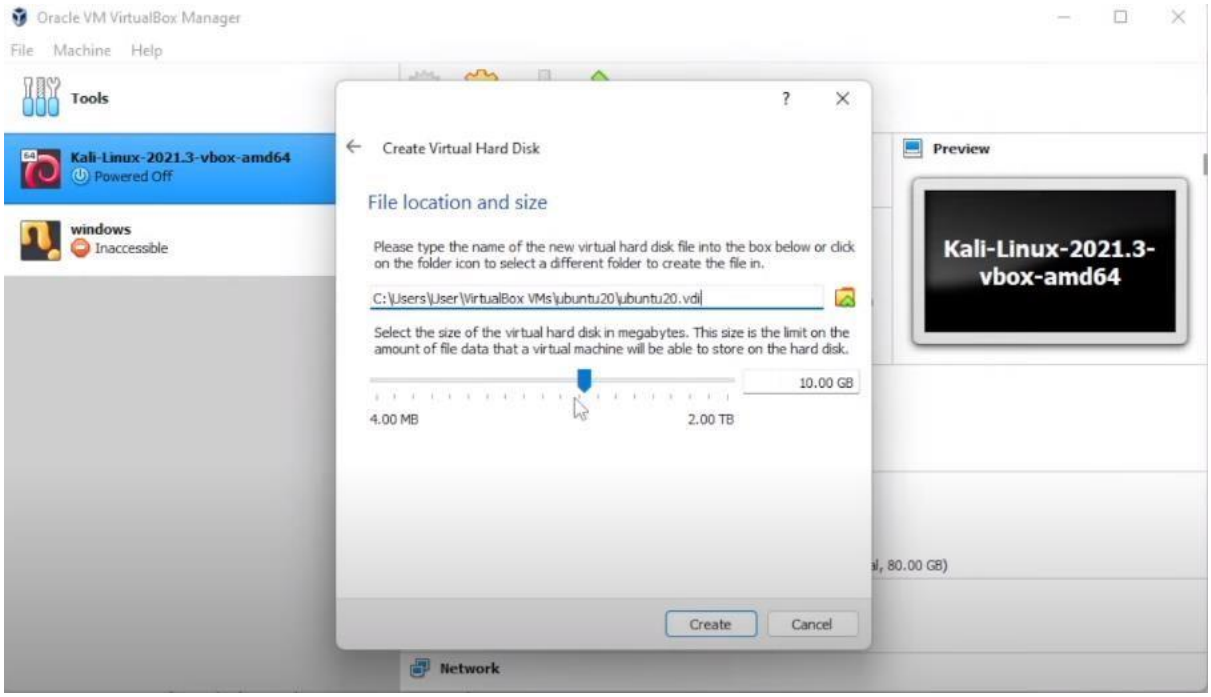
Step 9: Install Ubuntu OS in VM and when finished, click Restart and press Enter Step 5: Open terminal and update all packages

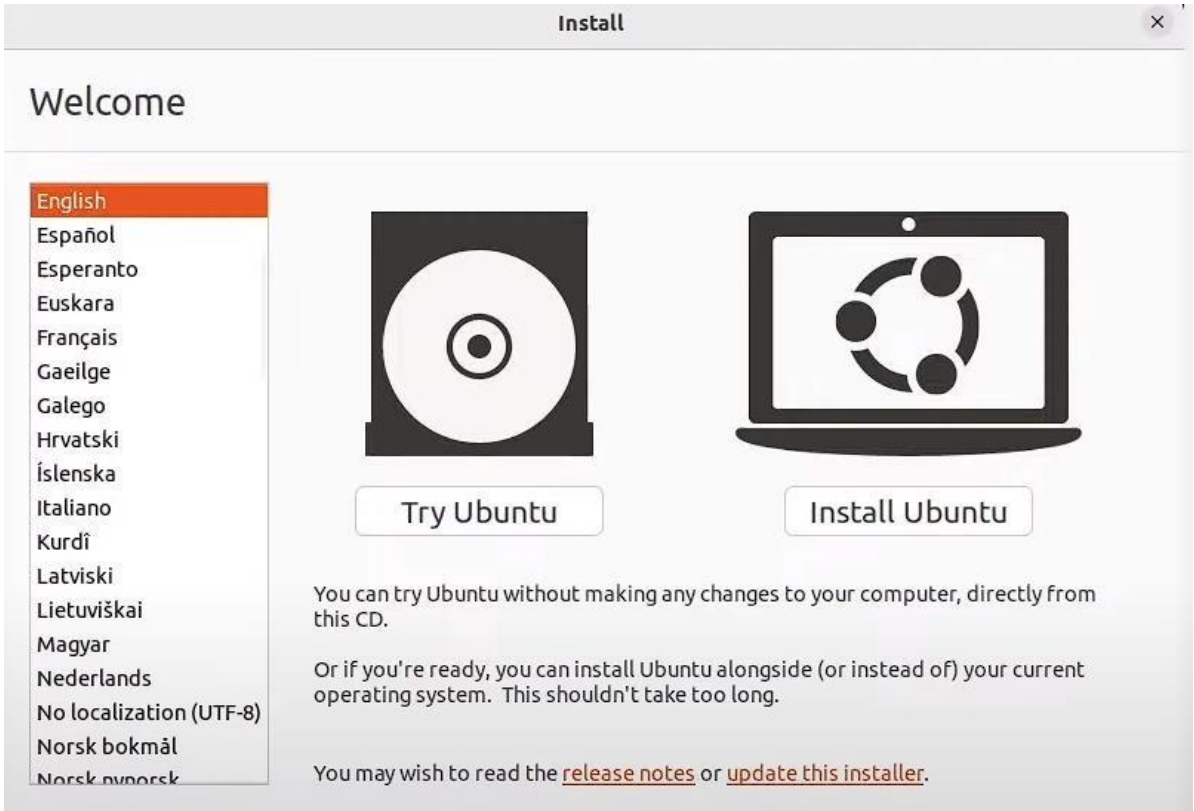
```
>sudo apt-get install update
```

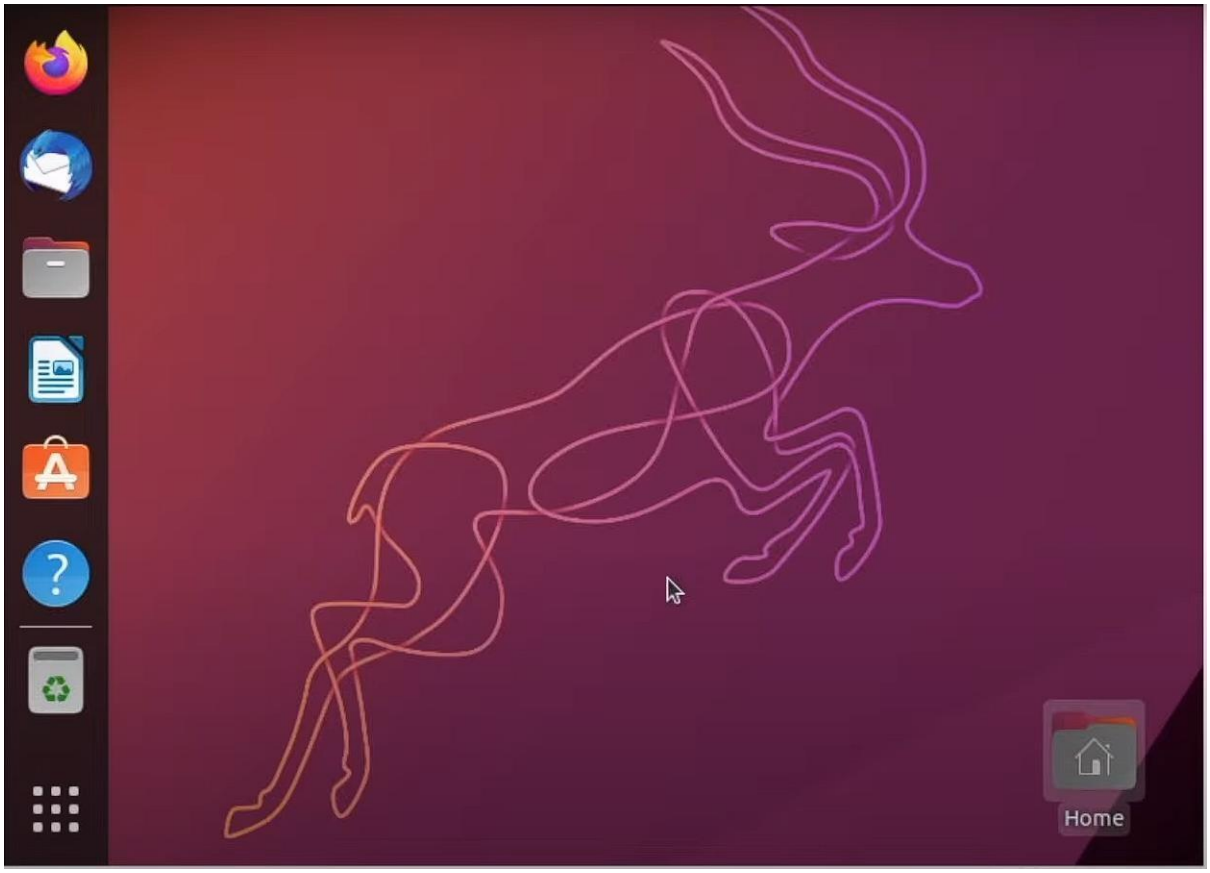
OUTPUT:


Oracle VM VirtualBox Manager
File Machine Help
Tools
Kali-Linux-2021.3-vbox-amd64
Powered Off
windows
Inaccessible
Preview
Kali-Linux-2021.3-vbox-amd64
Create Virtual Machine
Name and operating system
Please choose a descriptive name and destination folder 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: ubuntu20
Machine Folder: C:\Users\User\VirtualBox VMs
Type: Linux
Version: Ubuntu (64-bit)
Expert Mode
Next
Cancel
Host Driver: Windows DirectSound
Controller: ICH AC97
Network


Oracle VM VirtualBox Manager
File Machine Help
Tools
Kali-Linux-2021.3-vbox-amd64
Powered Off
windows
Inaccessible
Preview
Kali-Linux-2021.3-vbox-amd64
Create Virtual Machine
Memory size
Select the amount of memory (RAM) in megabytes to be allocated to the virtual machine.
The recommended memory size is 1024 MB.
1024 MB
4 MB
16384 MB
Next
Cancel
Host Driver: Windows DirectSound
Controller: ICH AC97
Network


Oracle VM VirtualBox Manager
File Machine Help
Tools
Kali-Linux-2021.3-vbox-amd64
Powered Off
windows
Inaccessible
Create Virtual Hard Disk
File location and size
Please type the name of the new virtual hard disk file into the box below or click on the folder icon to select a different folder to create the file in.
C:\Users\User\VirtualBox VMs\ubuntu20\ubuntu20.vdi
Select the size of the virtual hard disk in megabytes. This size is the limit on the amount of file data that a virtual machine will be able to store on the hard disk.
10.00 GB
4.00 MB
2.00 TB
Create
Cancel
Preview
Kali-Linux-2021.3-vbox-amd64
Network


Install
Welcome
English
Español
Esperanto
Euskara
Français
Gaeilge
Galego
Hrvatski
Íslenska
Italiano
Kurdî
Latviski
Lietuviškai
Magyar
Nederlands
No localization (UTF-8)
Norsk bokmål
Try Ubuntu
Install Ubuntu
You can try Ubuntu without making any changes to your computer, directly from this CD.
Or if you're ready, you can install Ubuntu alongside (or instead of) your current operating system. This shouldn't take too long.
You may wish to read the release notes or update this installer.



RESULT:

Hence the procedure to Install VirtualBox/VMware Workstation with different flavors of Linux or Windows OS is Successfully executed.

EX NO: 2 DATE:	**Install a C compiler in the Virtual Machine created using Virtual Box/ VMware and Executes Simple Programs**

AIM:

To install a C compiler in the Virtual Machine created using Virtual box/ VMware and execute a Simple Program.

PROCEDURE:

Step 1: Open Oracle VM Virtualbox/ VMware and Run Ubuntu OS

Step 2: Open Terminal and install gcc compiler

>sudo apt-get install gcc

Step 3: Check the gcc version you have installed

>gcc –version

Step 4: Open the editor

>sudo nano hello.c

Step 5: Type sample C Program

```
#include<stdio.h>

main()

{

printf("Hello World\n");

}
```

Step 5: Save the file by clicking Ctrl+X then type Y

Step 6: Type ls command on Terminal to see all files under current folder

Step 7: To Compile the program type

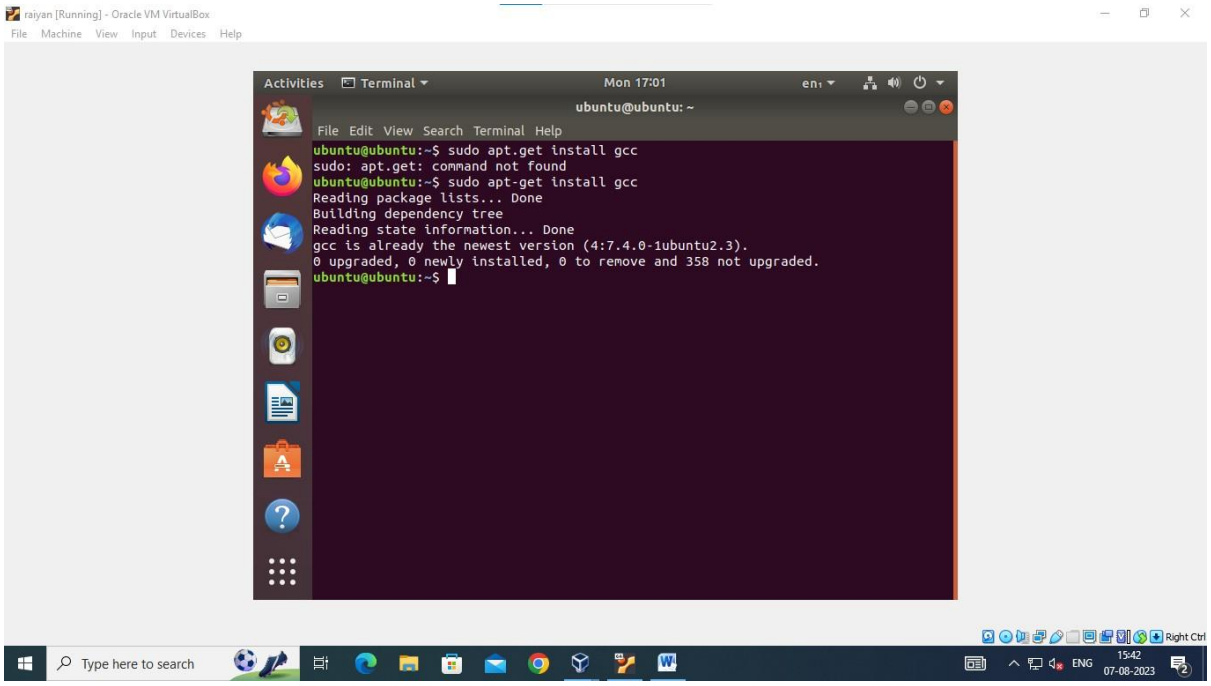
>sudo gcc hello.c

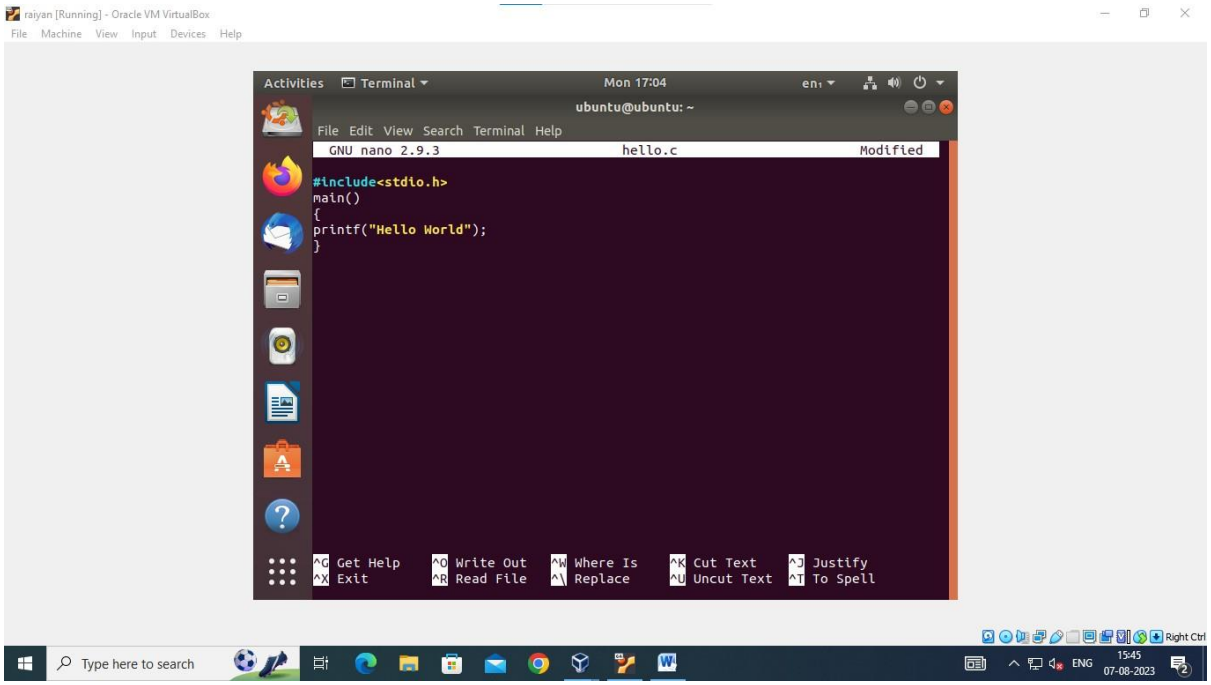
Step 8: To run the program type

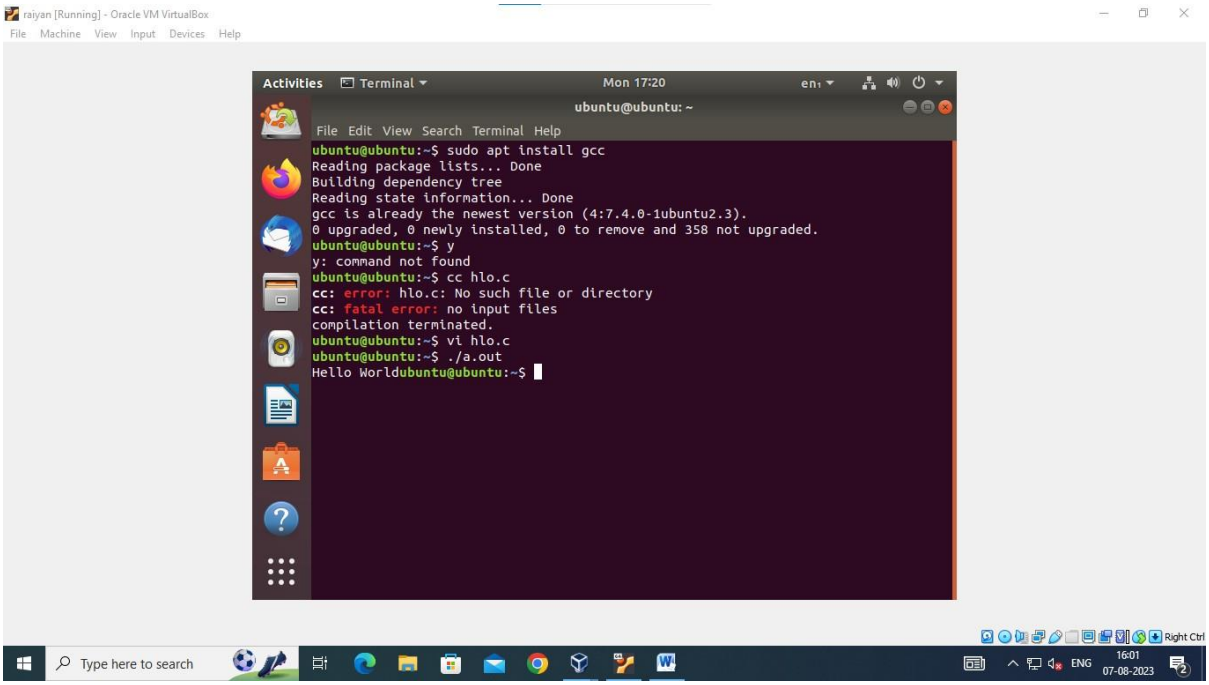
>./a.out

Step 9: If you see “Hello World” on the next line, successfully executed your First program.

OUTPUT:


raiyan [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Mon 17:01
ubuntu@ubuntu: ~
File Edit View Search Terminal Help
ubuntu@ubuntu:~$ sudo apt.get install gcc
sudo: apt.get: command not found
ubuntu@ubuntu:~$ sudo apt-get install gcc
Reading package lists... Done
Building dependency tree
Reading state information... Done
gcc is already the newest version (4:7.4.0-1ubuntu2.3).
0 upgraded, 0 newly installed, 0 to remove and 358 not upgraded.
ubuntu@ubuntu:~$
Type here to search
ENG
15:42
07-08-2023


raiyan [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Mon 17:04
ubuntu@ubuntu: ~
File Edit View Search Terminal Help
GNU nano 2.9.3 hello.c Modified
#include<stdio.h>
main()
{
printf("Hello World");
}
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Spell
Type here to search
ENG
15:45
07-08-2023



OUTPUT:

Hence to install a C compiler in the Virtual Machine created using Virtual box/ VMware and execute a Simple Program is successfully executed.

EX NO:3 DATE:	**Use VMware workstation to show the virtual machine migration based on the certain condition from one node to the other**

AIM:

To use VMWare workstation to show the virtual machine migration based on the certain condition from one node to the other.

PROCEDURE:

Step 1: Download & install Virtual box https://www.virtualbox.org/

Step 2: Download Ubuntu 18.04 ISO https://ubuntu.com/download/desktop

Step 3: Create Ubuntu VM in virtual box 4GB RAM, 2 vCPUs, Hard disk capacity of minimum 10 GB

Step 4: Create a new folder in the host OS (shared-VM)

Step 5: Open VM Virtualbox manager, Select your VM, go to settings Shared Folders option and the files to be shared from the host OS

Step 6: Login to guest OS (Ubuntu OS), goto devices menu install guest additions.

Step 7: Open terminal and go

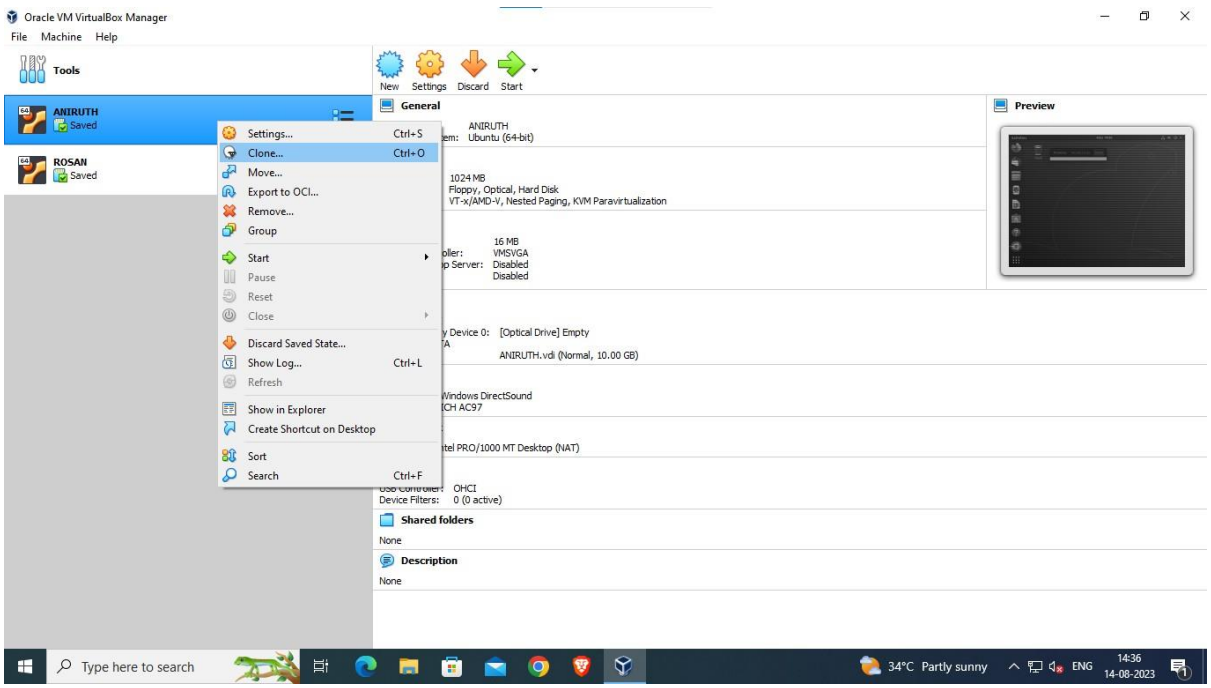
> sudo adduser vboxuser vboxsf

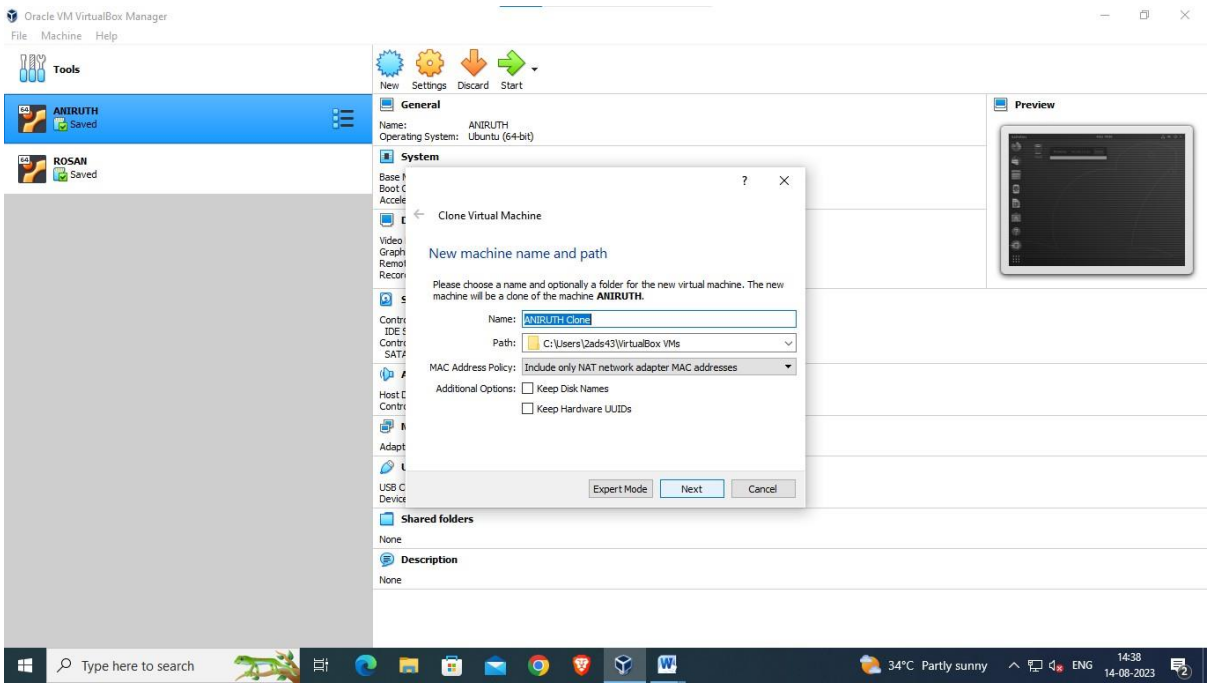
> sudo reboot

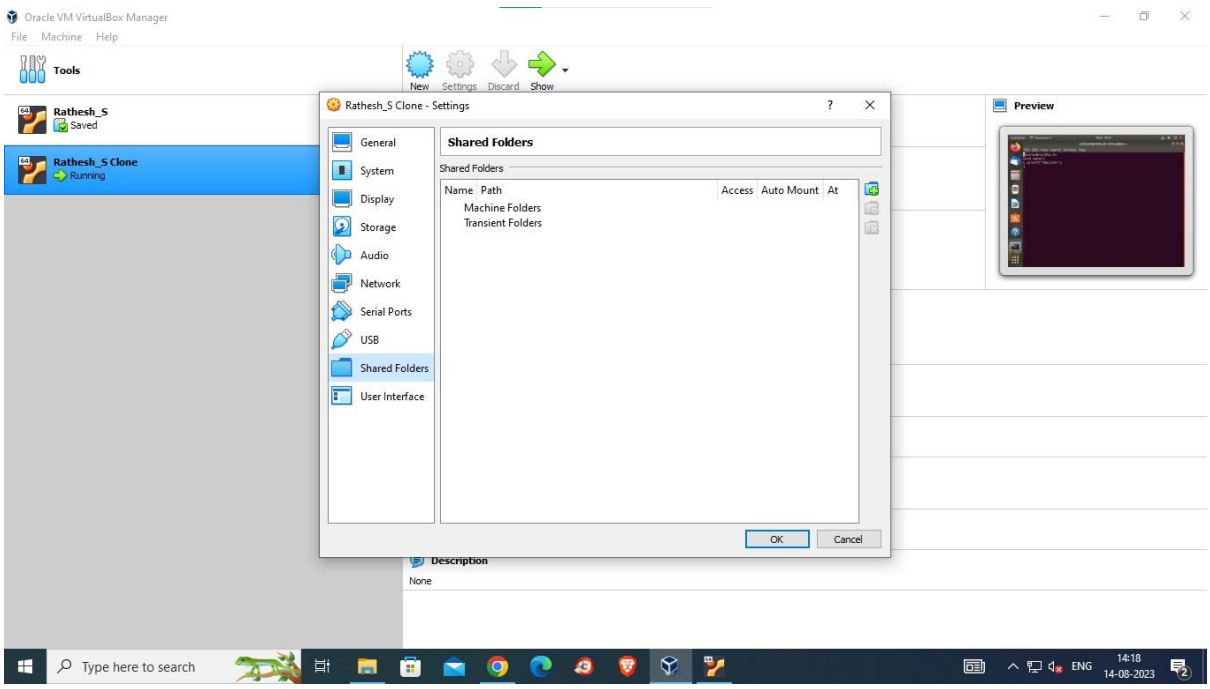
Step 8: Login to guest OS (Ubuntu OS), goto home folder to find the shared files

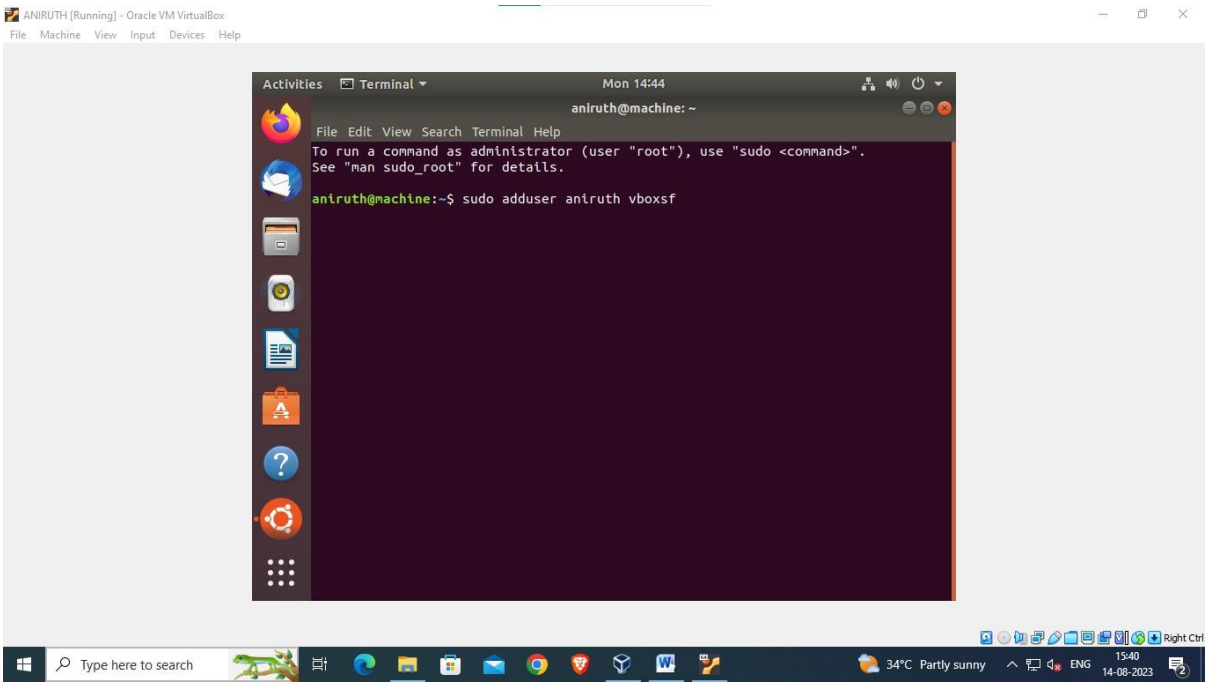
Step 9: Create a new file to share from guest OS to host OS.

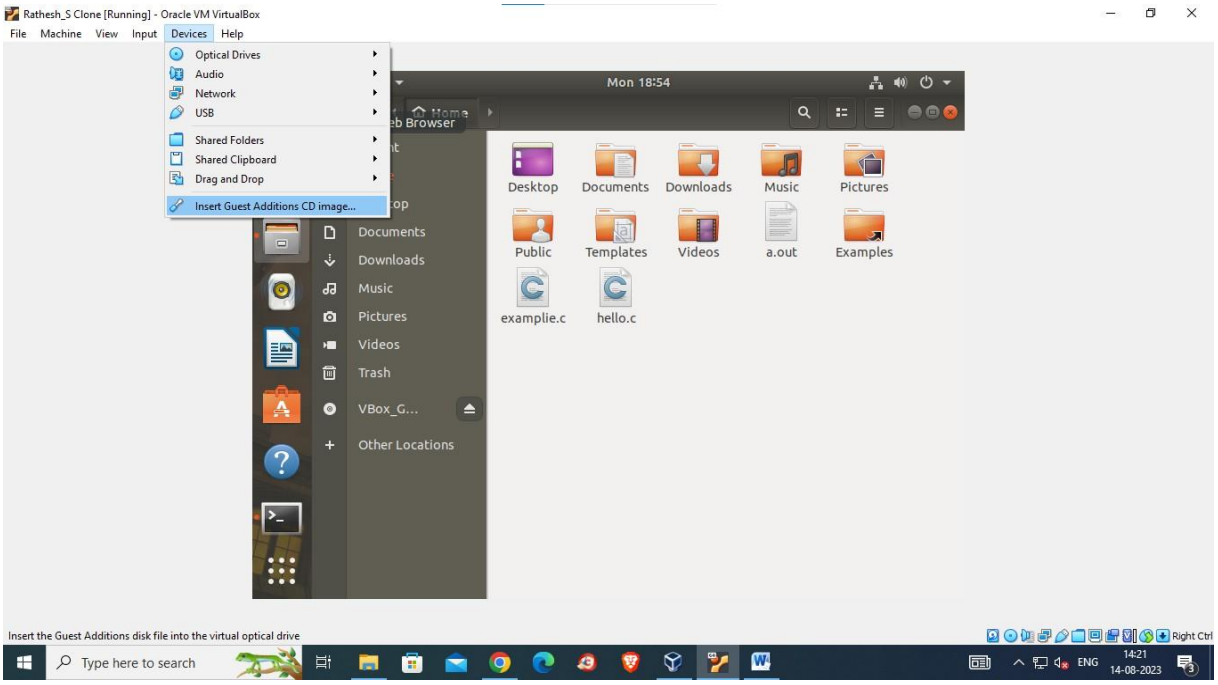
OUTPUT:

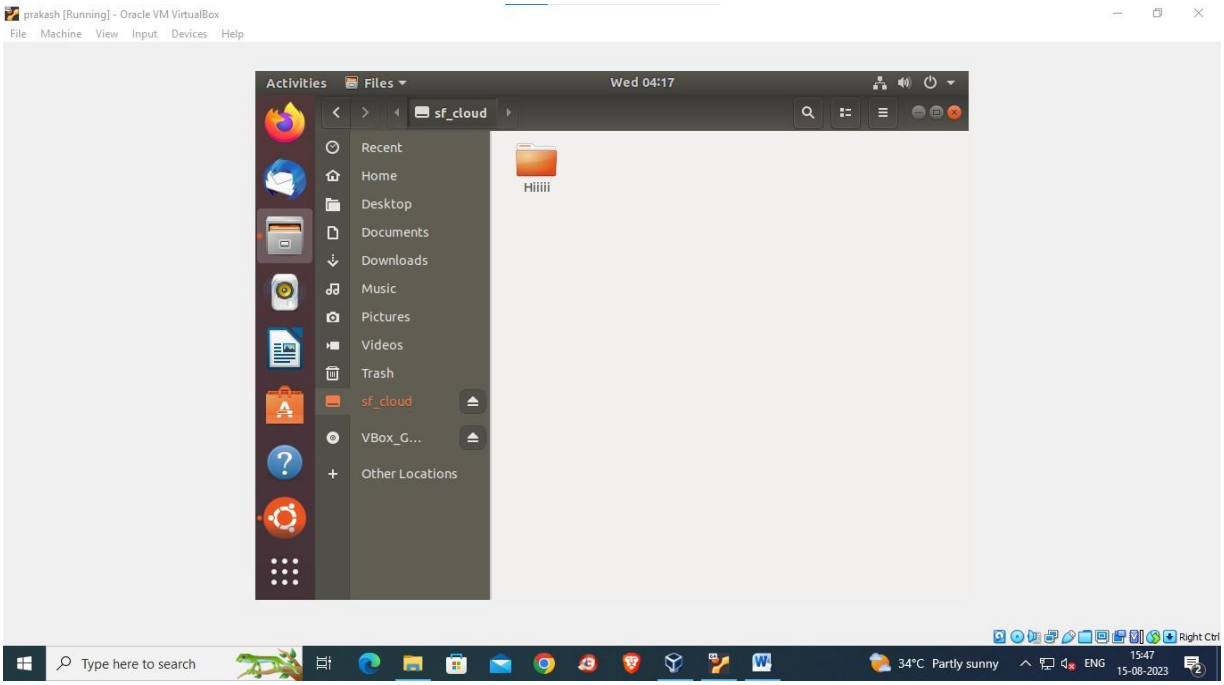

Oracle VM VirtualBox Manager
File Machine Help
Tools
ANIRUTH
Saved
ROSAN
Saved
Settings... Ctrl+S
Clone... Ctrl+O
Move...
Export to OCI...
Remove...
Group
Start
Pause
Reset
Close
Discard Saved State...
Show Log... Ctrl+L
Refresh
Show in Explorer
Create Shortcut on Desktop
Sort
Search Ctrl+F
New Settings Discard Start
General
Preview
Shared folders
None
Description
None
Type here to search
34°C Partly sunny
ENG
14:36
14-08-2023


Oracle VM VirtualBox Manager
File Machine Help
Tools
ANIRUTH
Saved
ROSAN
Saved
New Settings Discard Start
General
Name: ANIRUTH
Operating System: Ubuntu (64-bit)
System
Preview
Clone Virtual Machine
New machine name and path
Please choose a name and optionally a folder for the new virtual machine. The new machine will be a clone of the machine ANIRUTH.
Name: ANIRUTH Clone
Path: C:\Users\2ads43\VirtualBox VMs
MAC Address Policy: Include only NAT network adapter MAC addresses
Additional Options: Keep Disk Names
Keep Hardware UUIDs
Expert Mode
Next
Cancel
Shared folders
None
Description
None
Type here to search
34°C Partly sunny
ENG
14:38
14-08-2023


Oracle VM VirtualBox Manager
File Machine Help
Tools
Rathesh_S
Saved
Rathesh_S Clone
Running
New Settings Discard Show
Rathesh_S Clone - Settings
General
System
Display
Storage
Audio
Network
Serial Ports
USB
Shared Folders
User Interface
Shared Folders
Shared Folders
Name Path Access Auto Mount At
Machine Folders
Transient Folders
OK
Cancel
Preview
Description
None
Type here to search
ENG
14:18
14-08-2023


ANIRUTH [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities
Terminal
Mon 14:44
aniruth@machine: ~
File Edit View Search Terminal Help
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.
aniruth@machine:~$ sudo adduser aniruth vboxsf
Right Ctrl
Type here to search
34°C Partly sunny
ENG
15:40
14-08-2023





Result:

Hence to show the virtual machine migration based on the certain condition from one node to theother is successfully executed.

EX NO:4 DATE:	**Install Google App Engine. Create hello world app and other simple web applications using python/java. Use GAE launcher to launch the web applications**

AIM:

To Install Google App Engine, create hello world app and other simple web applications using python/java and use GAE launcher to launch the web applications.

PROCEDURE:

Step 1: In the Google Cloud console, on the project selector page, select or create a Google Cloud project

Step 2: Goto cloudshell terminal and check your project information

> gcloud projects list

Step 3: Navigate to APIs & Services menu to enable app engine, select App Engine Admin API enable it

Step 4: Goto cloudshell terminal to set your project

> gcloud config set project <project name>

Step 5: In the navigation menu select App Engine to create new applicationStep 6: Goto cloudshell terminal for deploying the application

> git clone https://github.com/GoogleCloudPlatform/python-docs-samples

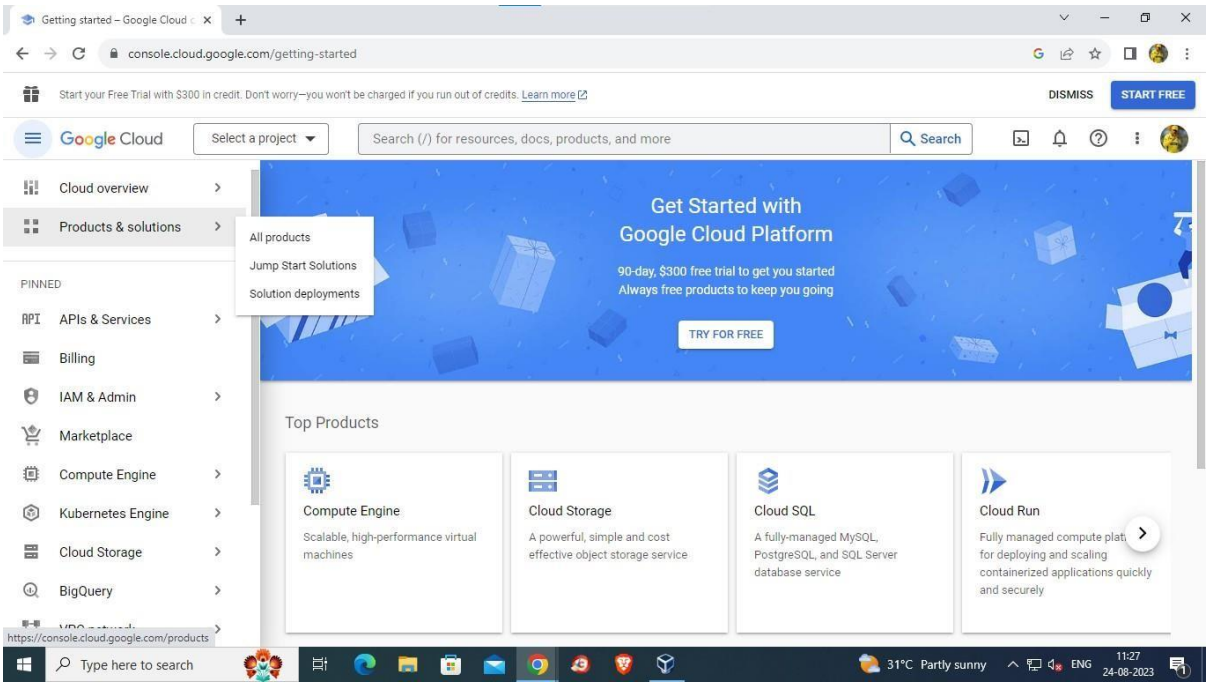
> cd python-docs-samples/appengine/ standard_python3/ hello_world/

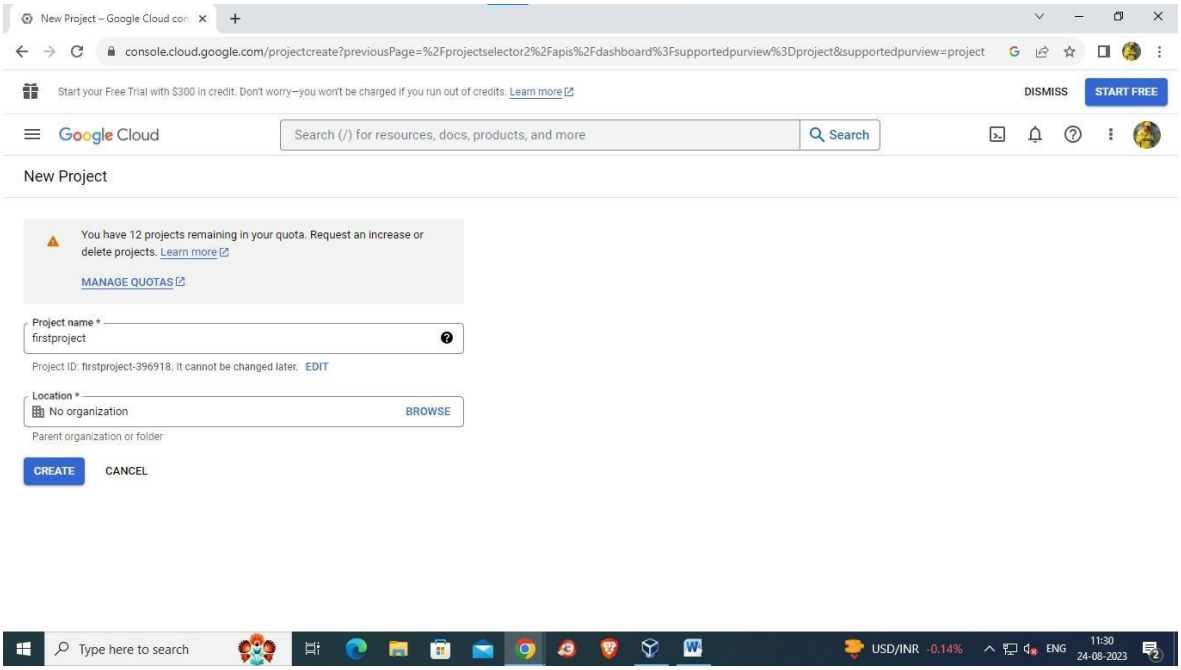
> cat main.py

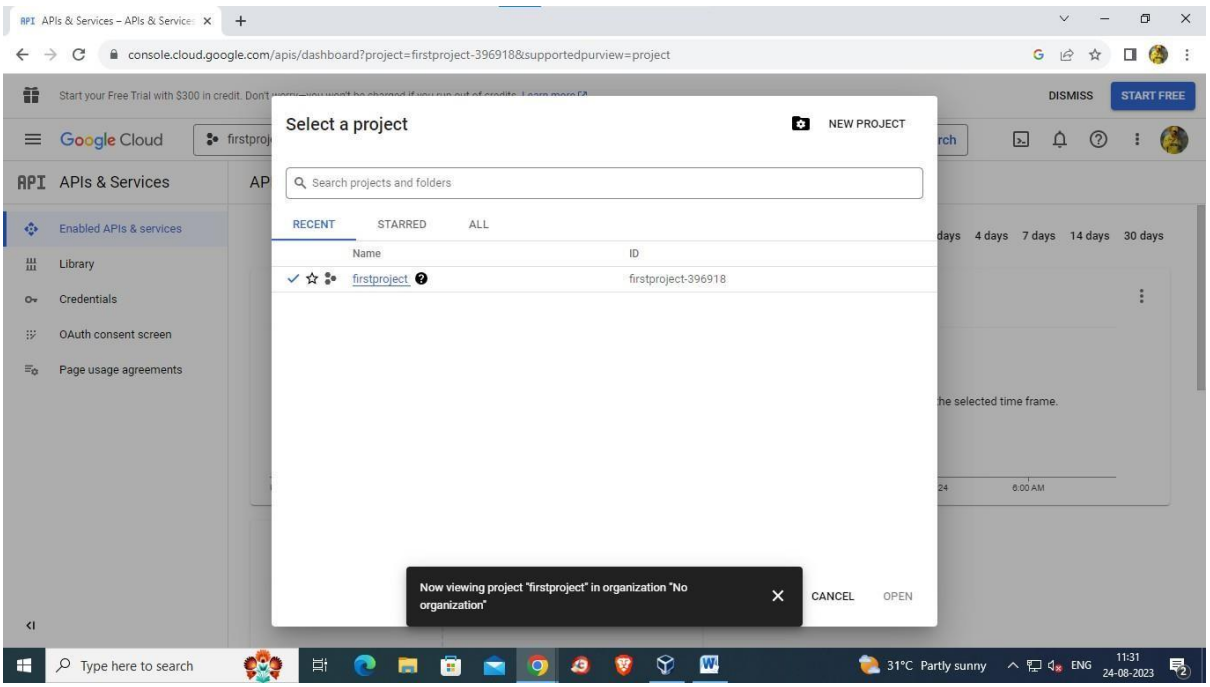
> python3 main.py

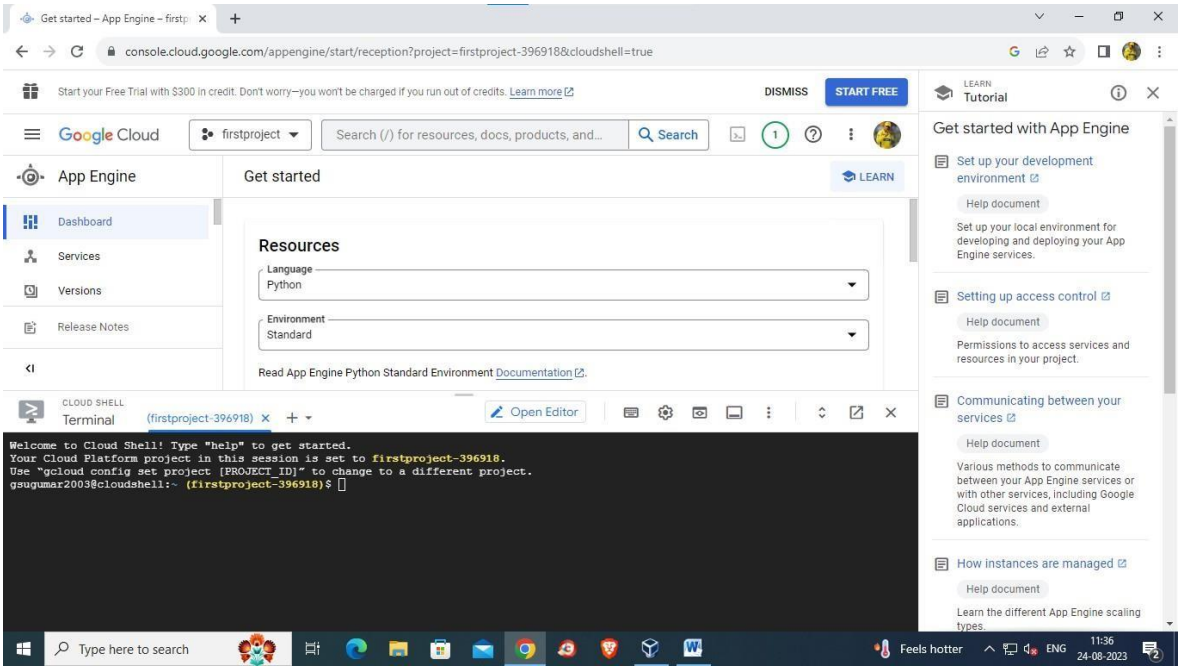
Running on http://127.0.0.1:8080

Output: 4-(a) – To install GAE launcher and develop a web application to display a message "hello world".

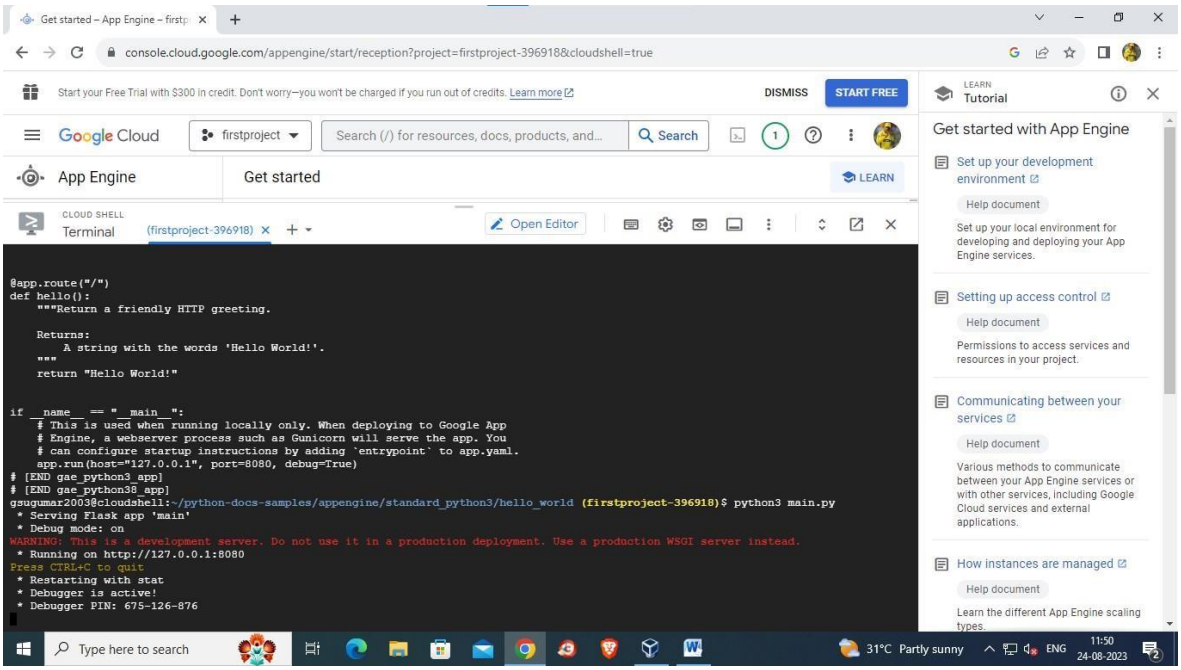


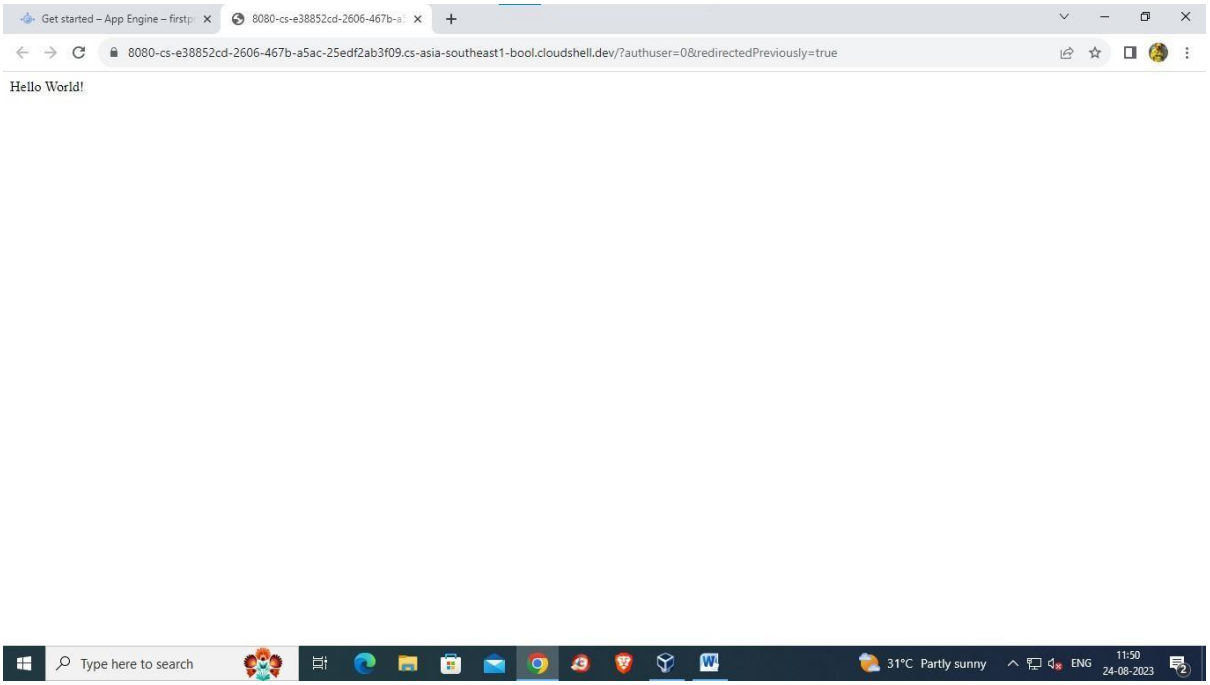



APIs & Services – APIs & Services
console.cloud.google.com/apis/dashboard?project=firstproject-396918&supportedpurview=project
Start your Free Trial with $300 in credit.
DISMISS
START FREE
Google Cloud
Select a project
NEW PROJECT
Search projects and folders
RECENT
STARRED
ALL
Name
ID
firstproject
firstproject-396918
APIs & Services
Enabled APIs & services
Library
Credentials
OAuth consent screen
Page usage agreements
Now viewing project "firstproject" in organization "No organization"
CANCEL
OPEN
Type here to search
31°C Partly sunny
ENG
11:31
24-08-2023

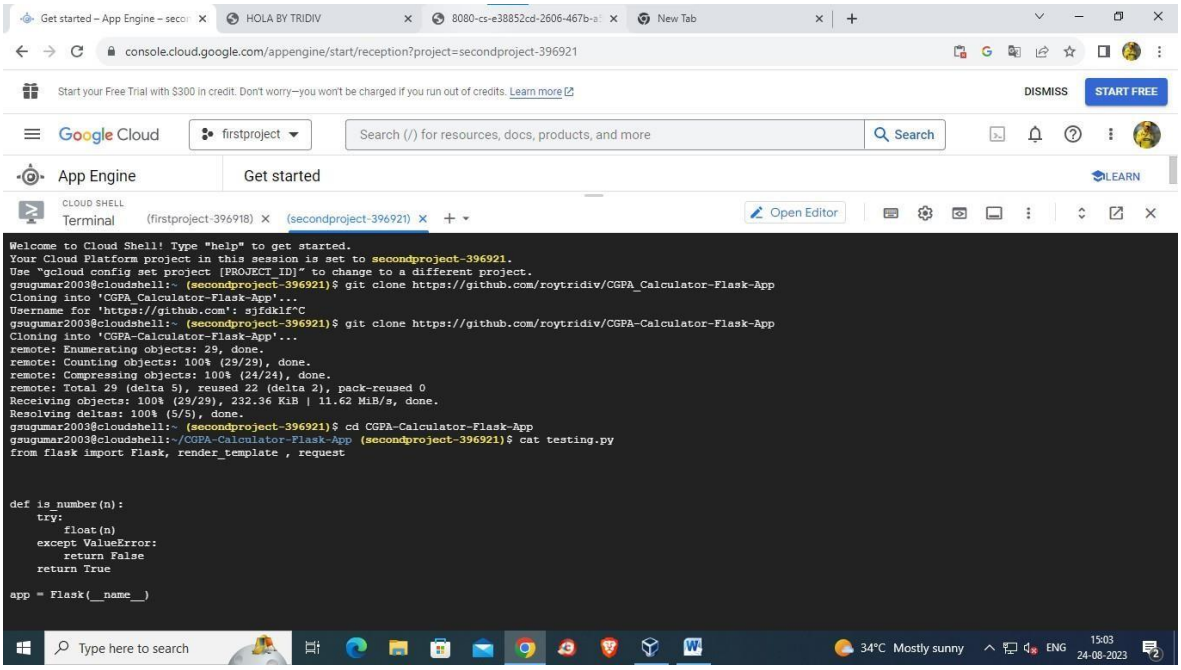

Get started – App Engine – firstp
console.cloud.google.com/appengine/start/reception?project=firstproject-396918&cloudshell=true
Start your Free Trial with $300 in credit. Don't worry—you won't be charged if you run out of credits. Learn more
DISMISS
START FREE
Google Cloud
firstproject
Search (/) for resources, docs, products, and...
Search
App Engine
Get started
LEARN
Dashboard
Services
Versions
Release Notes
Resources
Language
Python
Environment
Standard
Read App Engine Python Standard Environment Documentation.
CLOUD SHELL
Terminal
(firstproject-396918)
Open Editor
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to firstproject-396918.
Use "gcloud config set project [PROJECT_ID]" to change to a different project.
gsugumar2003@cloudshell:~ (firstproject-396918)$
LEARN
Tutorial
Get started with App Engine
Set up your development environment
Help document
Set up your local environment for developing and deploying your App Engine services.
Setting up access control
Help document
Permissions to access services and resources in your project.
Communicating between your services
Help document
Various methods to communicate between your App Engine services or with other services, including Google Cloud services and external applications.
How instances are managed
Help document
Learn the different App Engine scaling types.
Type here to search
Feels hotter
ENG
11:36
24-08-2023

```
gsugumar2003@cloudshell:~ (firstproject-396918)$ git clone https://github.com/GoogleCloudPlatform/python-docs-samples
Cloning into 'python-docs-samples'...
remote: Enumerating objects: 105268, done.
remote: Counting objects: 100% (105268/105268), done.
remote: Compressing objects: 100% (32739/32739), done.
remote: Total 105268 (delta 63561), reused 104239 (delta 63038), pack-reused 0
Receiving objects: 100% (105268/105268), 213.70 MiB | 17.81 MiB/s, done.
Resolving deltas: 100% (63561/63561), done.
Updating files: 100% (4545/4545), done.
gsugumar2003@cloudshell:~ (firstproject-396918)$ cd python-docs-samples/appengine/standard_python3/hello_world/
gsugumar2003@cloudshell:~/python-docs-samples/appengine/standard_python3/hello_world (firstproject-396918)$ cat main.py
# Copyright 2018 Google LLC
#
# 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.

# [START gae_python38_app]
```





Output: 4-(b) – Develop an app to calculate CGPA using java / python and deploy it on theGoogle App Engine using GAE launcher on windows



console.cloud.google.com/appengine/start/reception?project=secondproject-396921

Start your Free Trial with $300 in credit. Don't worry—you won't be charged if you run out of credits. Learn more

DISMISS START FREE

Google Cloud firstproject Search (/) for resources, docs, products, and more Search

CLOUD SHELL Terminal (firstproject-396918) (secondproject-396921) Open Editor

You are viewing Cloud Shell in full view. Click here to restore Cloud Shell.

Got it!

```
if __name__ == '__main__':
    app.run(debug=True)
gsugumar2003@cloudshell:~/CGPA-Calculator-Flask-App (secondproject-396921)$ python3 testing.py
 * Serving Flask app 'testing'
 * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: 675-126-876
127.0.0.1 - - [23/Aug/2023 09:32:35] "GET /?authuser=0&redirectedPreviously=true HTTP/1.1" 200 -
127.0.0.1 - - [23/Aug/2023 09:32:35] "GET /static/styles/main.css HTTP/1.1" 200 -
127.0.0.1 - - [23/Aug/2023 09:32:35] "GET /static/logo.png HTTP/1.1" 200 -
127.0.0.1 - - [23/Aug/2023 09:32:35] "GET /static/pic.png HTTP/1.1" 200 -
127.0.0.1 - - [23/Aug/2023 09:32:37] "GET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [23/Aug/2023 09:33:57] "POST /result HTTP/1.1" 200 -
127.0.0.1 - - [23/Aug/2023 09:33:57] "GET /static/logo.png HTTP/1.1" 304 -
127.0.0.1 - - [23/Aug/2023 09:33:57] "GET /static/styles/main.css HTTP/1.1" 304 -
127.0.0.1 - - [23/Aug/2023 09:33:57] "GET /static/styles/{{%20url_for('static',filename='styles/digital-7.ttf')%20}} HTTP/1.1" 404 -
127.0.0.1 - - [23/Aug/2023 09:34:06] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [23/Aug/2023 09:34:06] "GET /static/logo.png HTTP/1.1" 304 -
127.0.0.1 - - [23/Aug/2023 09:34:06] "GET /static/styles/main.css HTTP/1.1" 304 -
127.0.0.1 - - [23/Aug/2023 09:34:06] "GET /static/pic.png HTTP/1.1" 304 -
```

Type here to search Hot weather ENG 15:04 24-08-2023

5000-cs-e38852cd-2606-467b-a5ac-25edf2ab3f09.cs-asia-southeast1-bool.cloudshell.dev/?

CGPA CALCULATOR

This is a cumulative GPA calculator. There are options for six courses input for calculaton. You can leave any course input box empty if you have taken fewer courses. But you must fill the **Previous CGPA** and **Current credit** (they are required fields)

Current CGPA

Enter current CGPA

Current Credits

Enter current credit

Course 1 Credit

Enter Credit of course 1

Course 1 GPA

Enter Credit of gpa 1

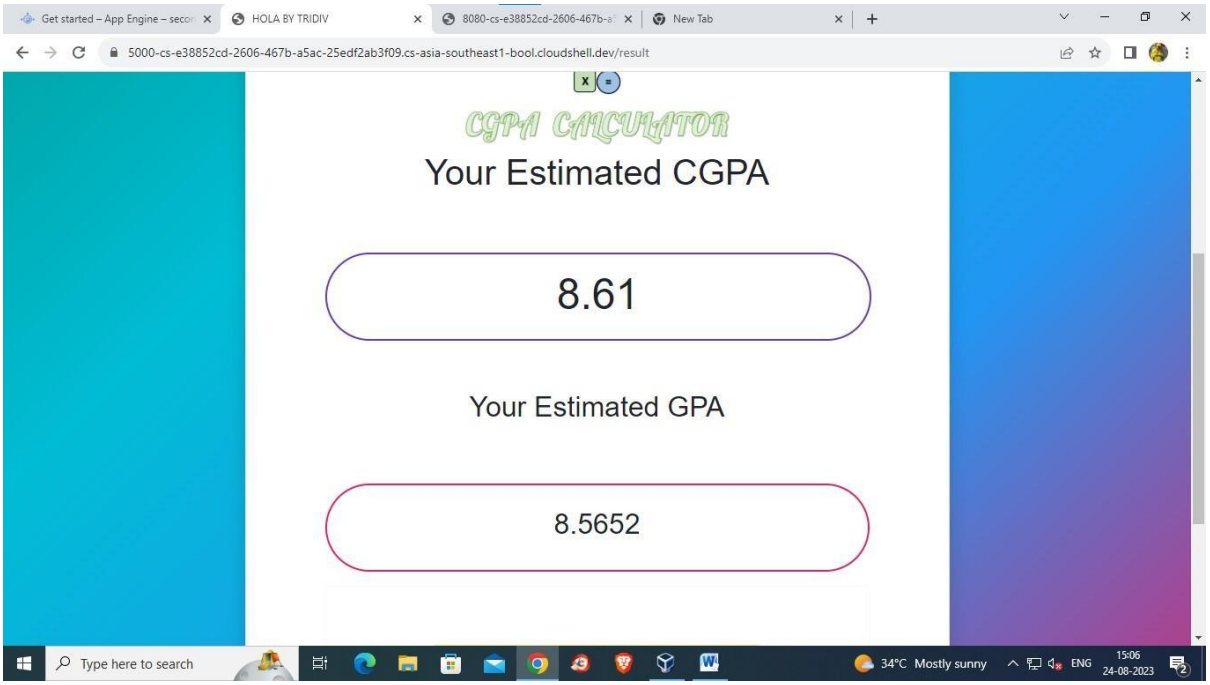
Course 2 Credit

Enter Credit of course 2

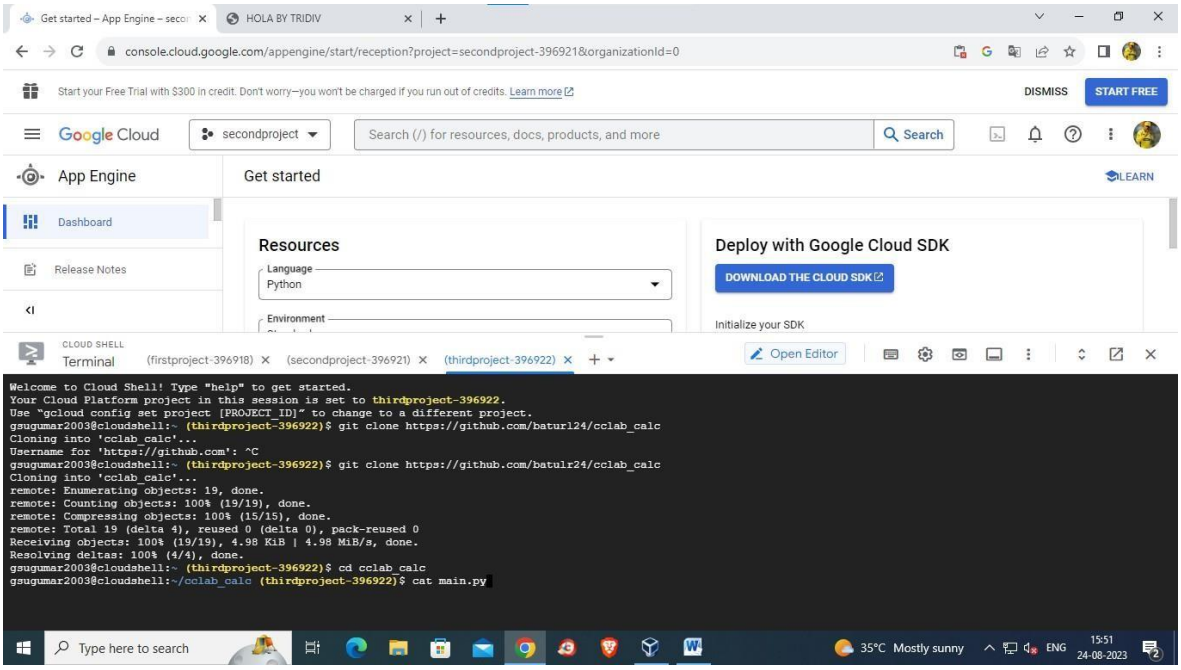
Course 2 GPA

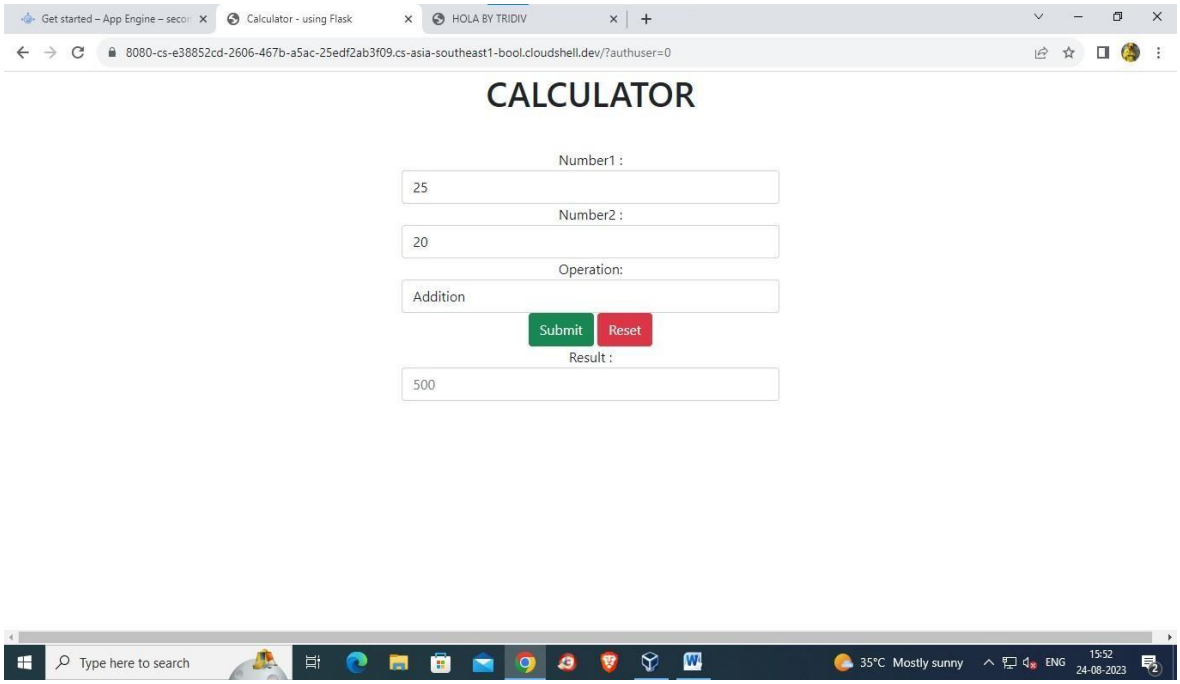
Enter Credit of gpa 2

Type here to search Hot weather ENG 15:04 24-08-2023



4-(c) To install a Google App Engine and create the web application for the calculatorusing python:





Result:

Hence to Install Google App Engine, create hello world app and other simple web applications using python/java and use GAE launcher to launch the web applications is successfully executed.

EX NO: 5 DATE:	Install and Configure Hadoop

AIM:

To install and configure Hadoop using single node cluster.

PROCEDURE:

Step 1: Download & install Hadoop 3.3.6 & Java JDK 1.8.0Setup Java and Hadoop

Step 2: Check either Java 1.8.0 is already installed on your system or not, use "Javac -version"to check.

Step 3: If Java is not installed on your system, then first install java under "C:\JAVA"

Step 4: Extract file Hadoop 3.3.6.tar.gz or Hadoop-3.3.6.zip and place under "C:\Hadoop 3.3.6". Step 5: Set the path HADOOP_HOME Environment variable on windows 7/10

Configuration

Step 6: Edit file C:/Hadoop-3.3.6/etc/hadoop/core-site.xml, paste below xml paragraph and savethis file.

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

Step 7: Rename "mapred-site.xml.template" to "mapred-site.xml" and edit this file C:/Hadoop 3.3.6/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>
```

Step 8: Create folder "data" under "C:\Hadoop-3.3.6"

- Create folder "datanode" under "C:\Hadoop-3.3.6\data"
- Create folder "namenode" under "C:\Hadoop-3.3.6\data"

Step 9: Edit file C:\Hadoop-3.3.6/etc/hadoop/hdfs-site.xml, paste below xml paragraph and savethis file.

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>/hadoop-3.3.6/data/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/hadoop-3.3.6/data/datanode</value>
</property>
</configuration>
```

Step 9: Edit file C:/Hadoop-3.3.6/etc/hadoop/yarn-site.xml, paste below xml paragraph and savethis 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>
```

Step 10: Edit file C:/Hadoop-3.3.6/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)

Step 11: Hadoop Configuration

- Dowload file Hadoop Configuration.zip
- Delete file bin on C:\Hadoop-3.3.6\bin, replaced by file bin on file just download

(from Hadoop Configuration.zip).

- Open cmd and typing command “hdfs namenode –format”.Step 12: Testing

Open cmd and change directory to “C:\Hadoop-3.3.6\sbin” and type “startall.cmd” to start apache.

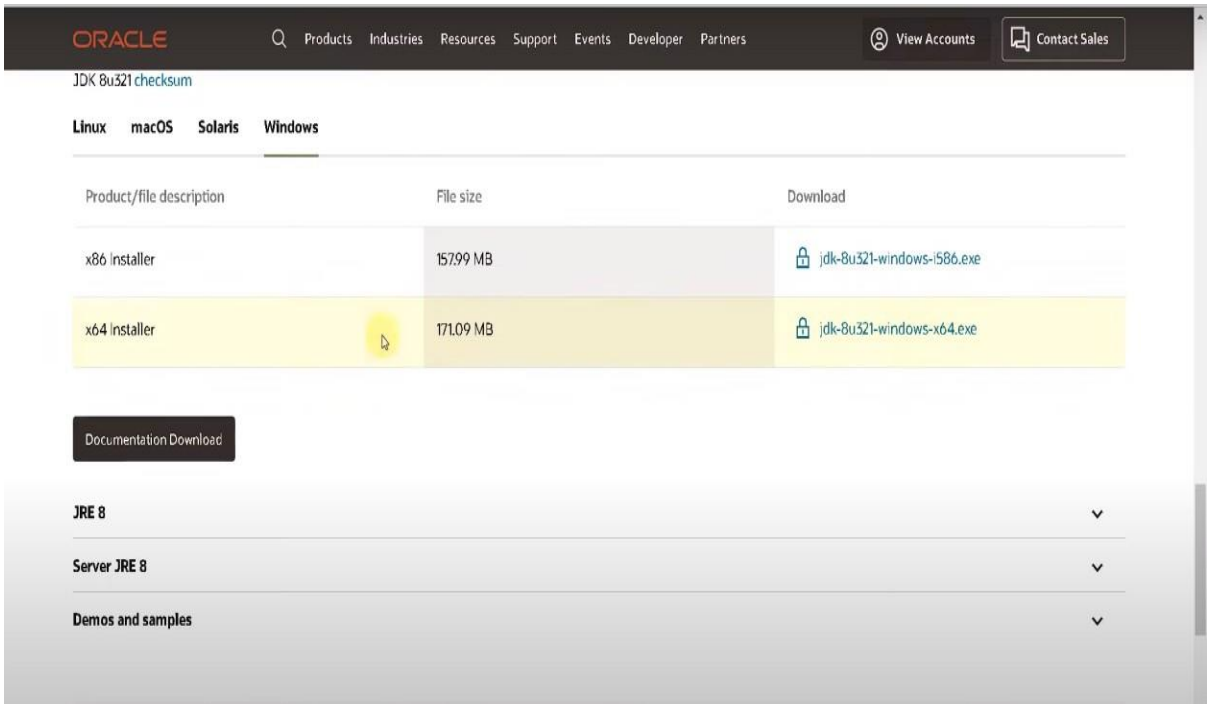
Make sure these apps are running

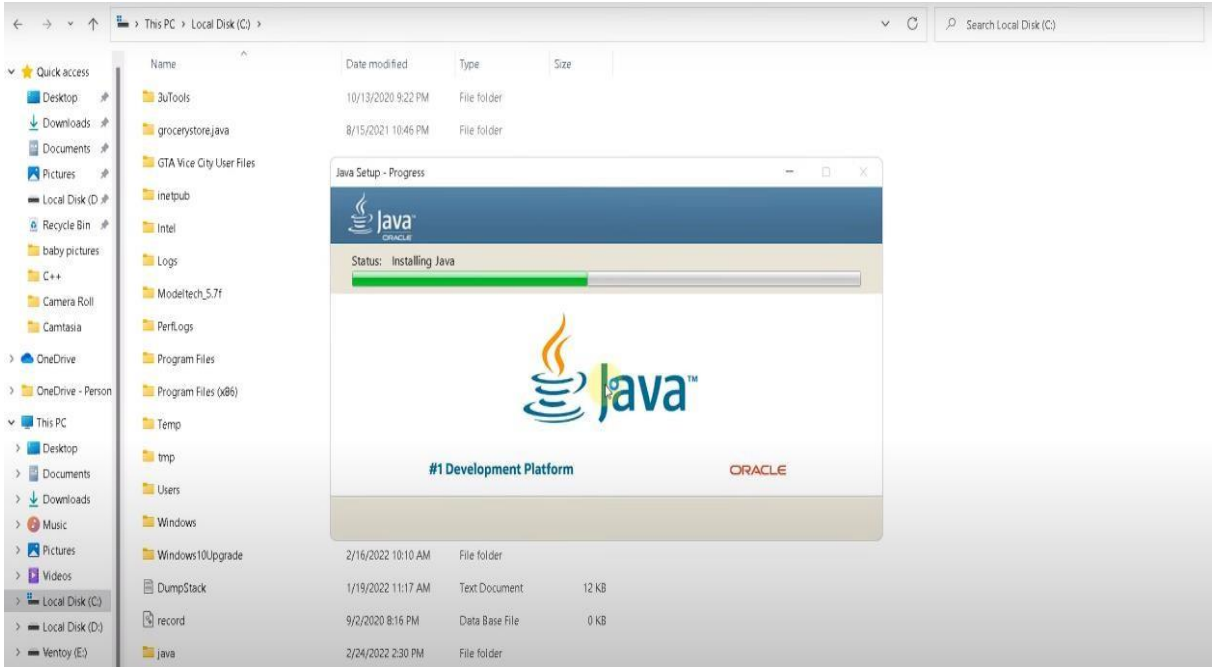
- Hadoop Namenode
- Hadoop datanode
- YARN Resourc Manager
- YARN Node Manager

Step 13: Open: http://localhost:8088 Step 14: Open: http://localhost:50070

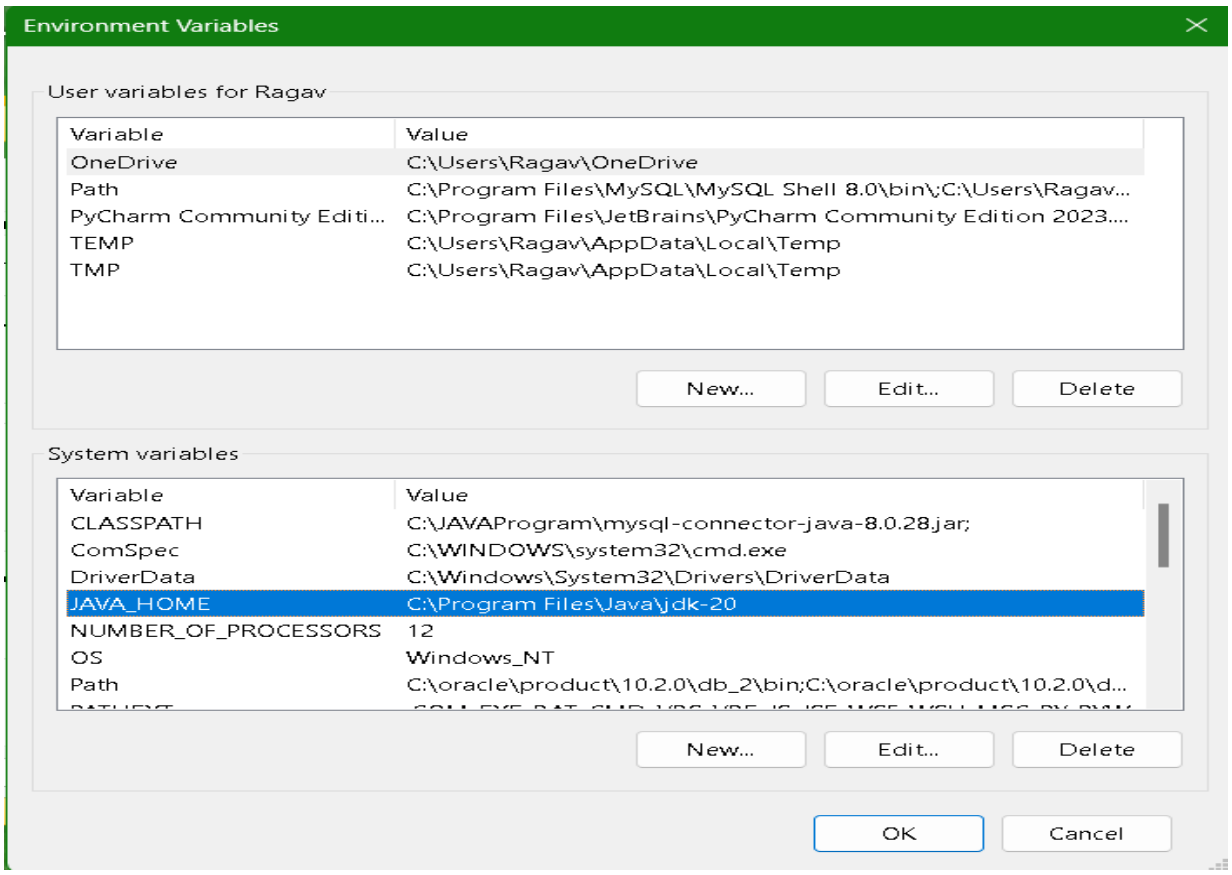
OUTPUT:

STEP 1: Installing JAVA jdk

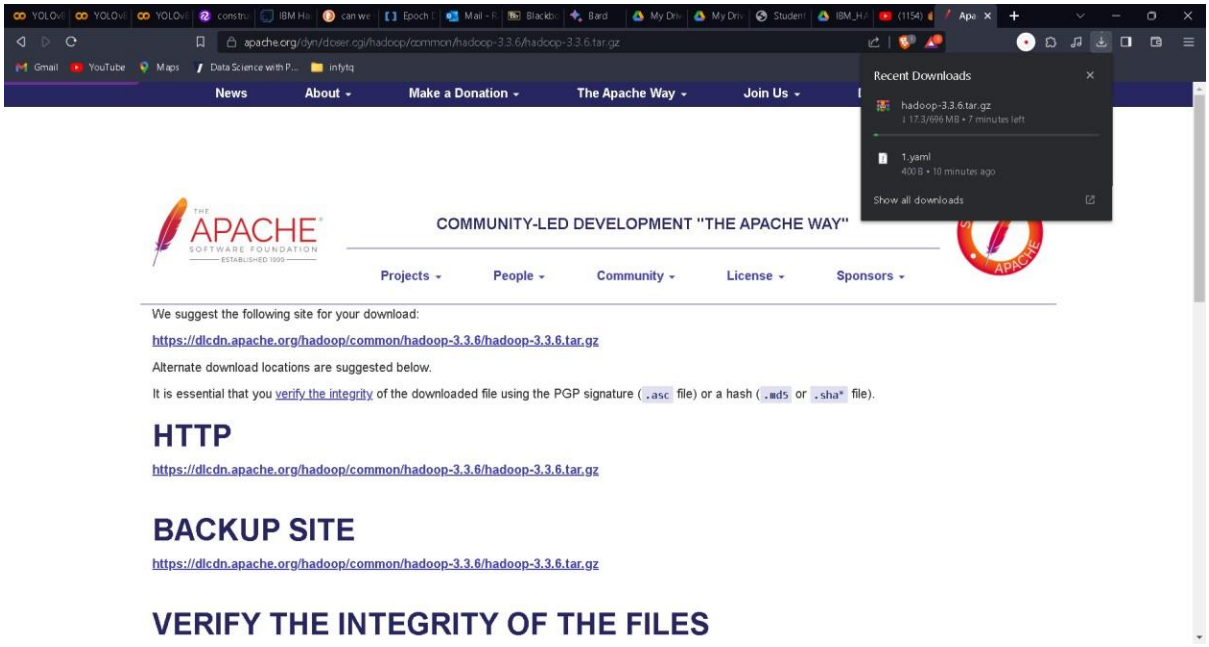


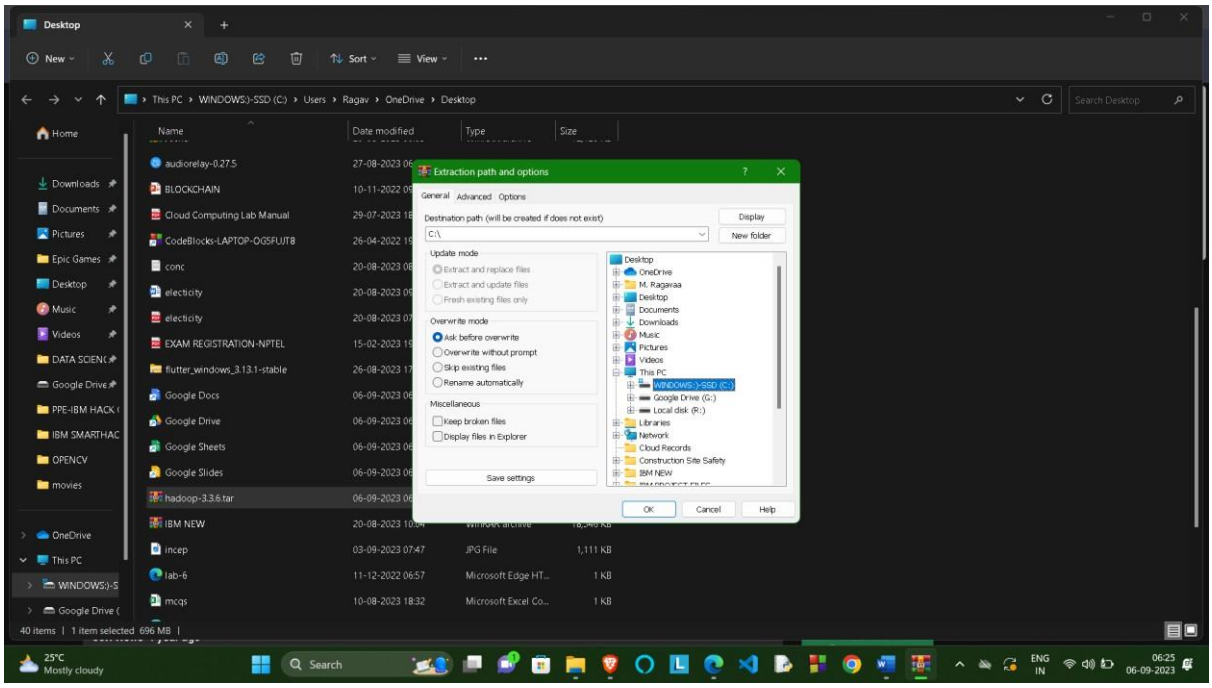


Setting Environmental variables:

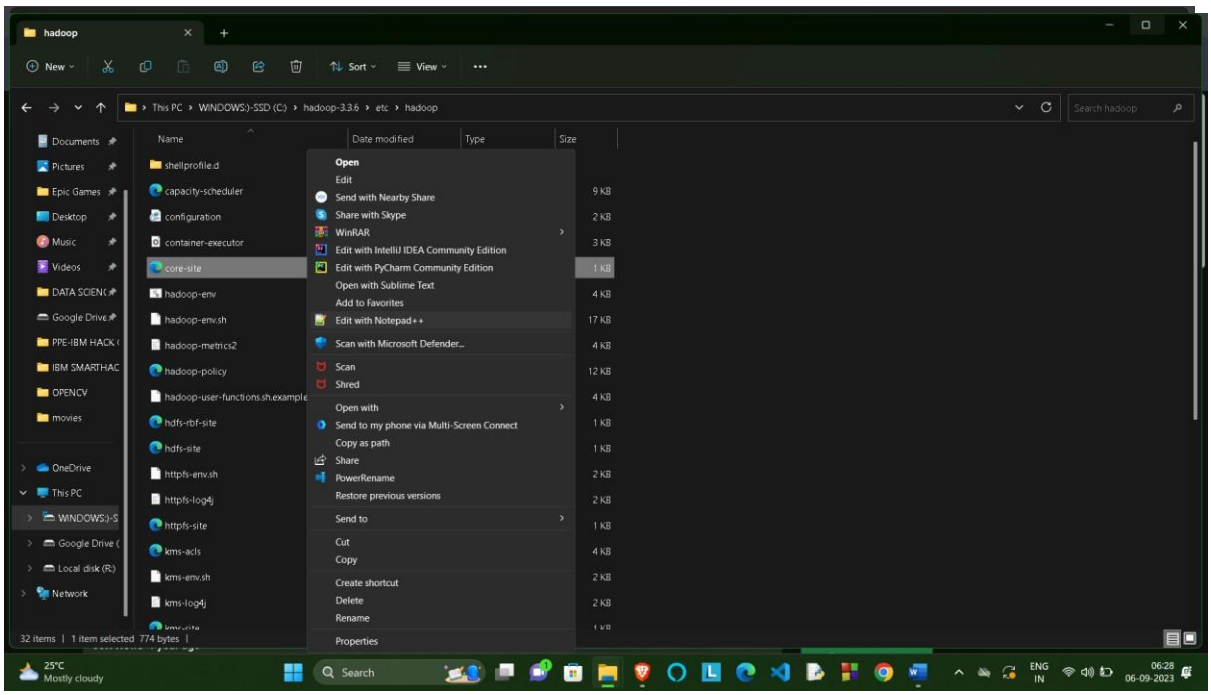


STEP 2: Installing and Configuring hadoop.





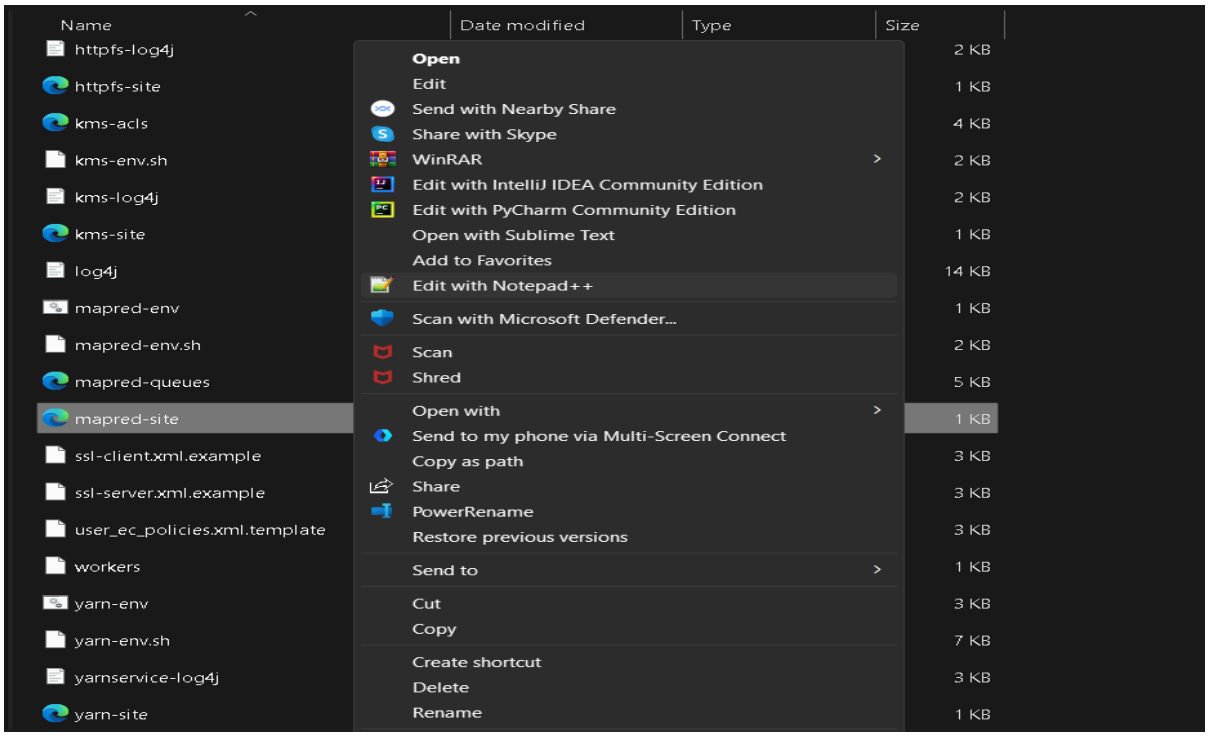
EDITTING ‘core-site.xml’:



core-site.xml

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

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

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

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

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

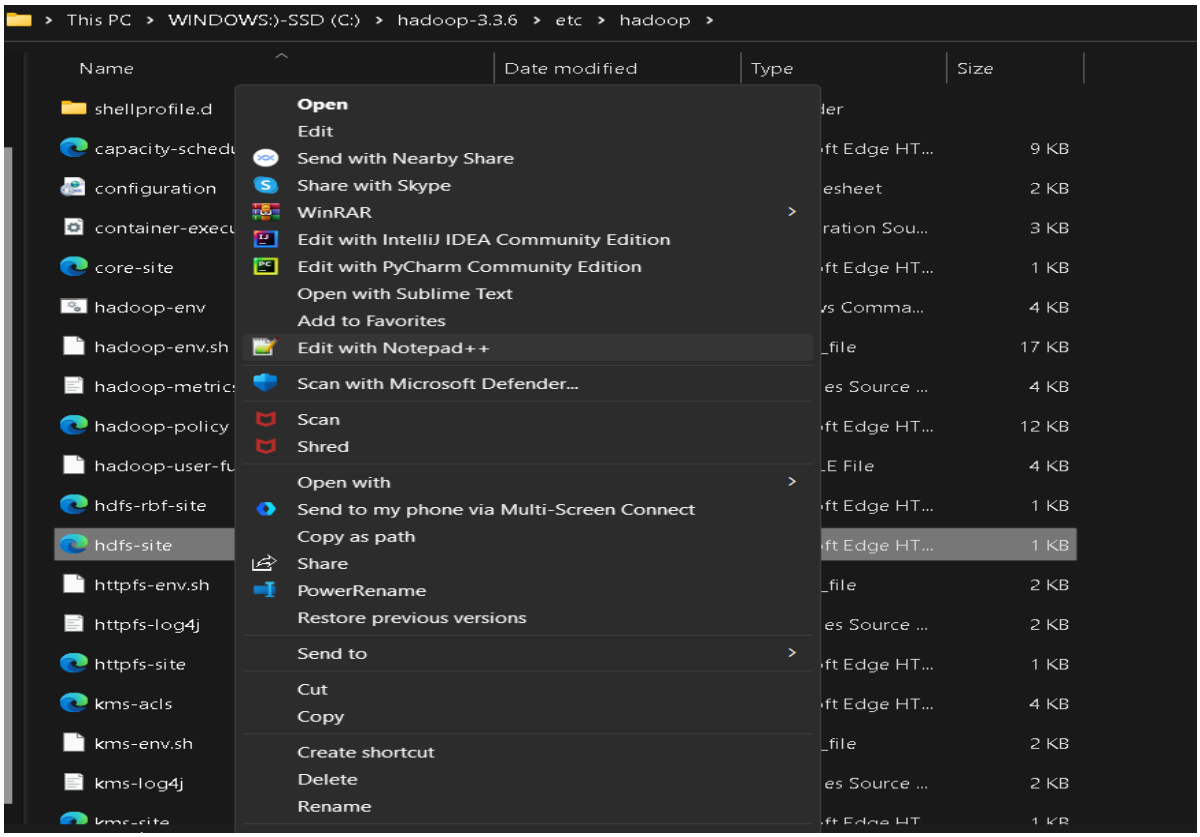
EDITTING ‘mapred-site.xml’:

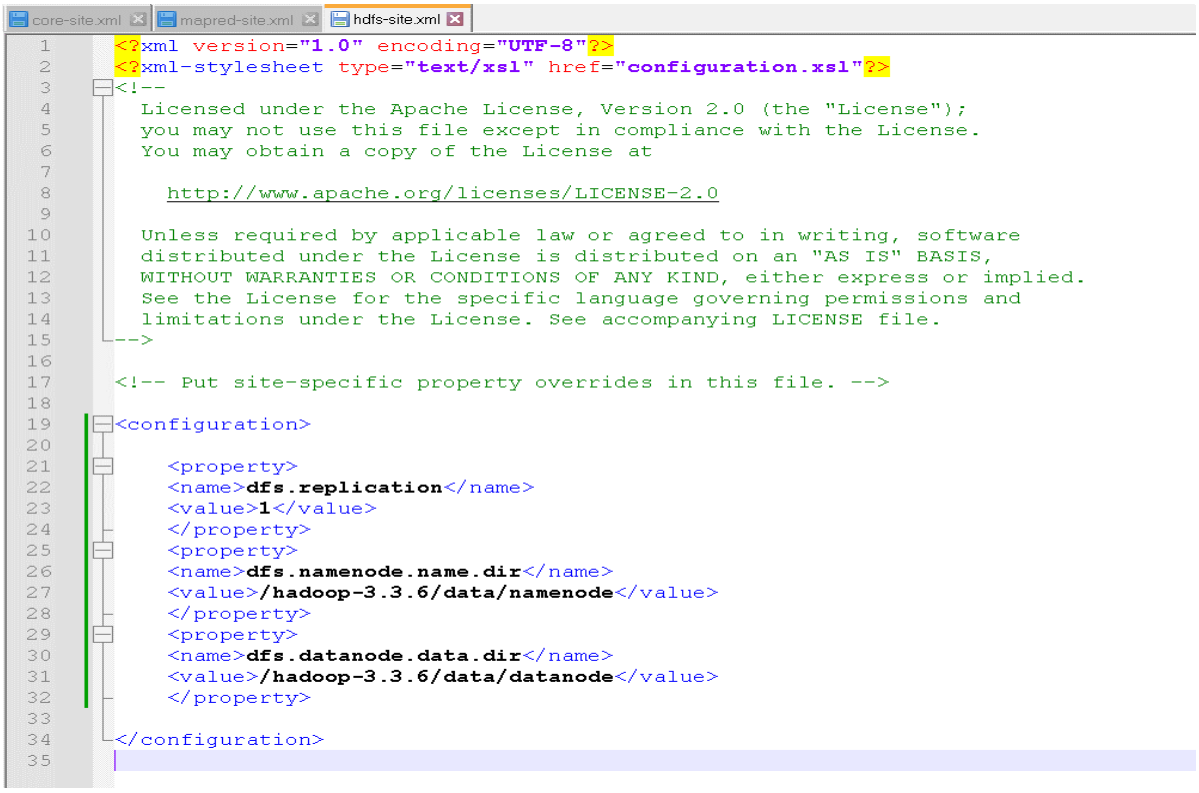


core-site.xml | mapred-site.xml

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

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

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

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

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

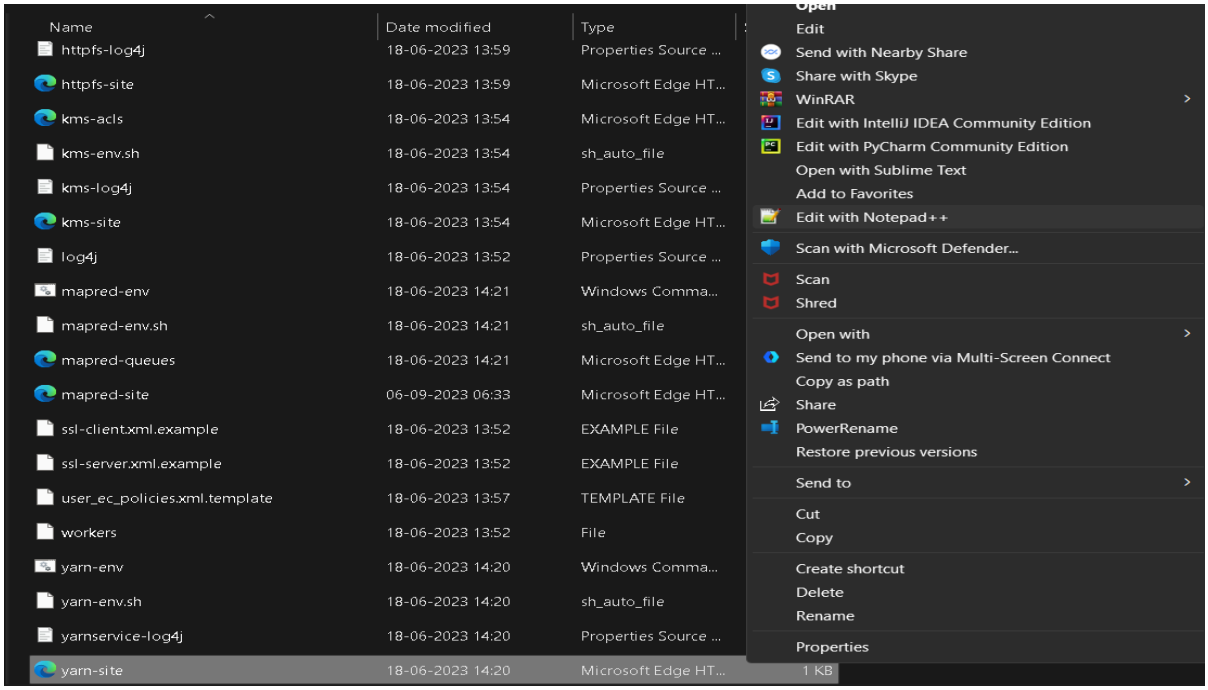
EDITTING ‘hdfs-site.xml’:






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

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

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

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

<configuration>

    <property>
    <name>dfs.replication</name>
    <value>1</value>
    </property>
    <property>
    <name>dfs.namenode.name.dir</name>
    <value>/hadoop-3.3.6/data/namenode</value>
    </property>
    <property>
    <name>dfs.datanode.data.dir</name>
    <value>/hadoop-3.3.6/data/datanode</value>
    </property>

</configuration>
```

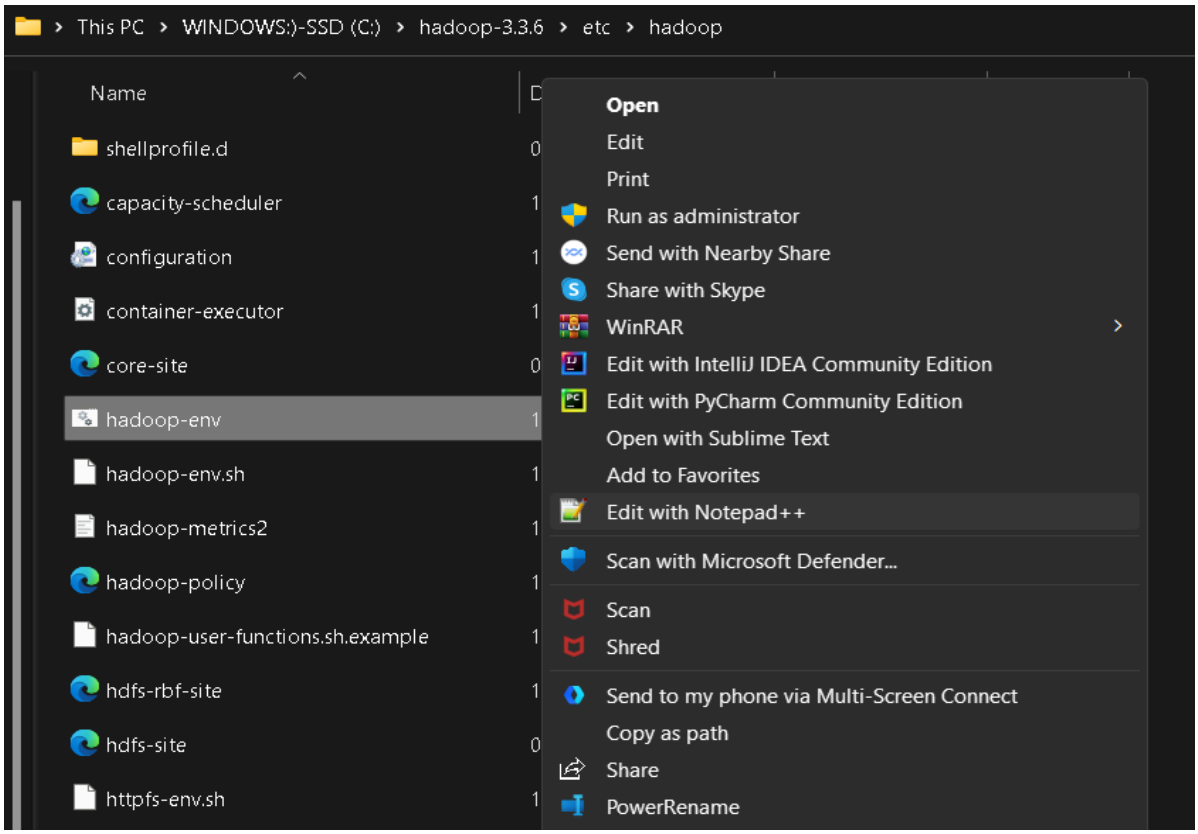
EDITTING ‘yarn-site.xml’

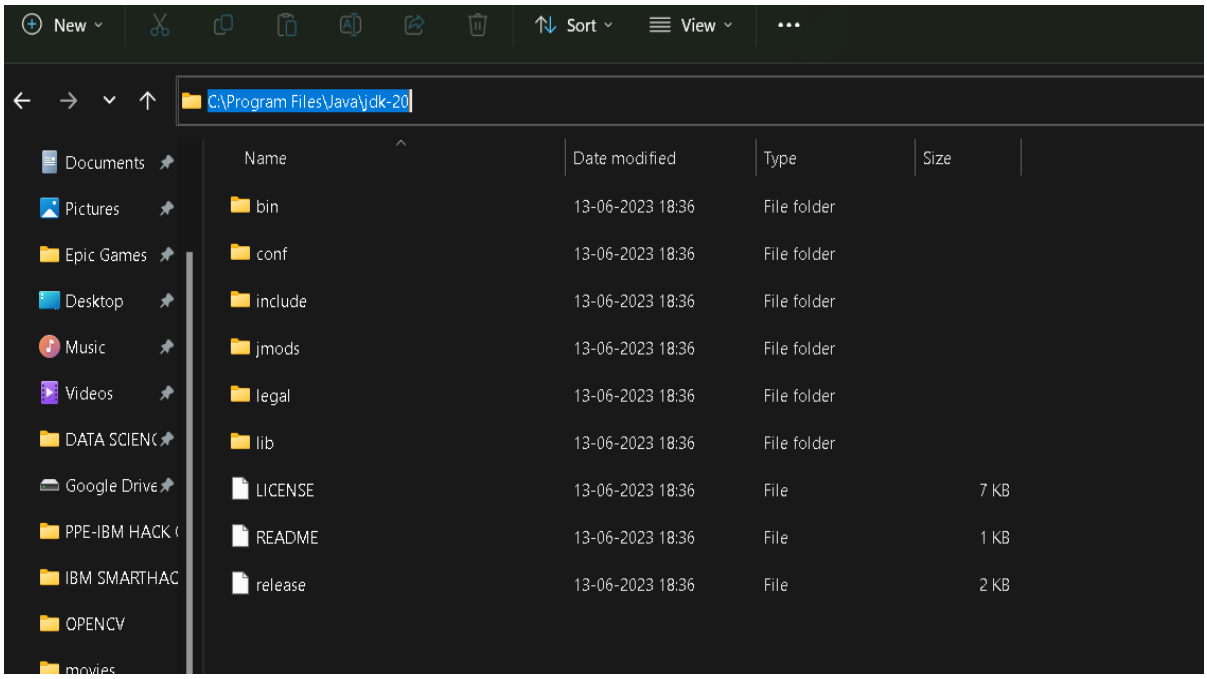


core-site.xml | mapred-site.xml | hdfs-site.xml | yarn-site.xml

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

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

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

<property>

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

</configuration>
```

RUNNIG ‘Hadoop Environment (Hadoop-env)’:



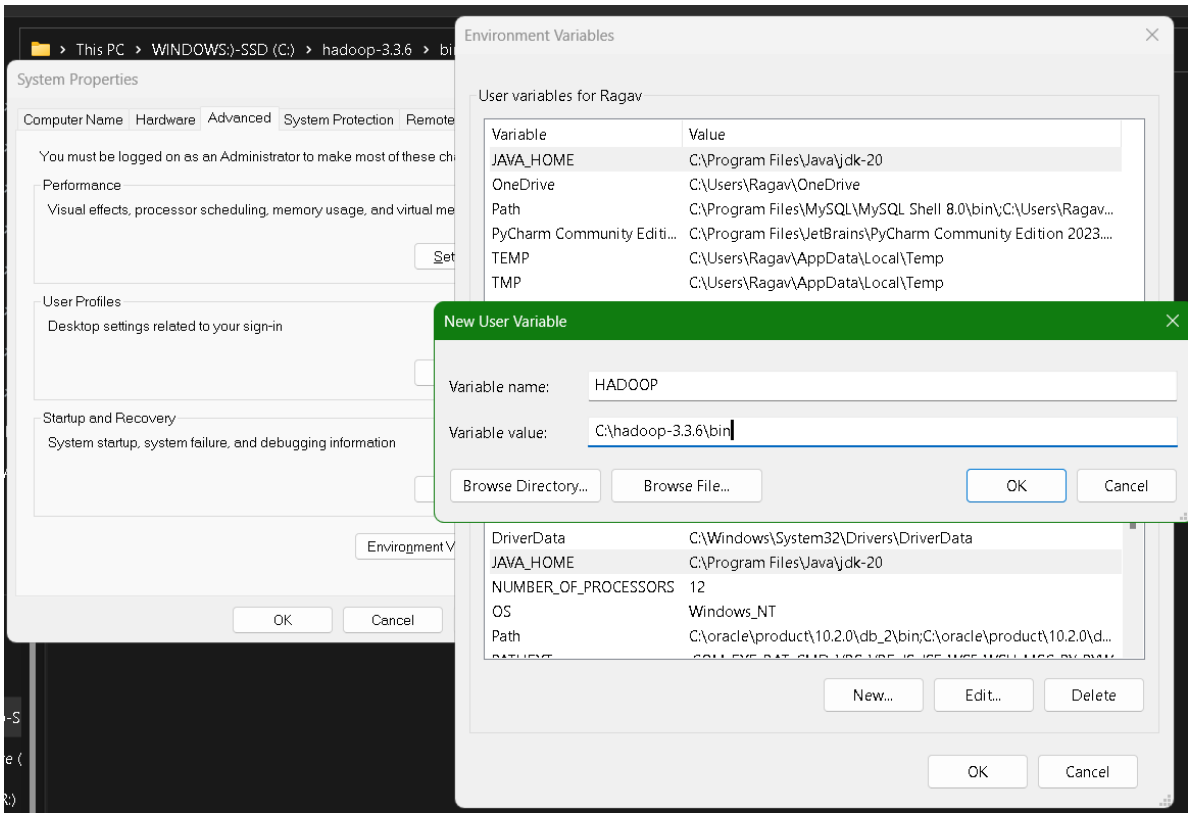


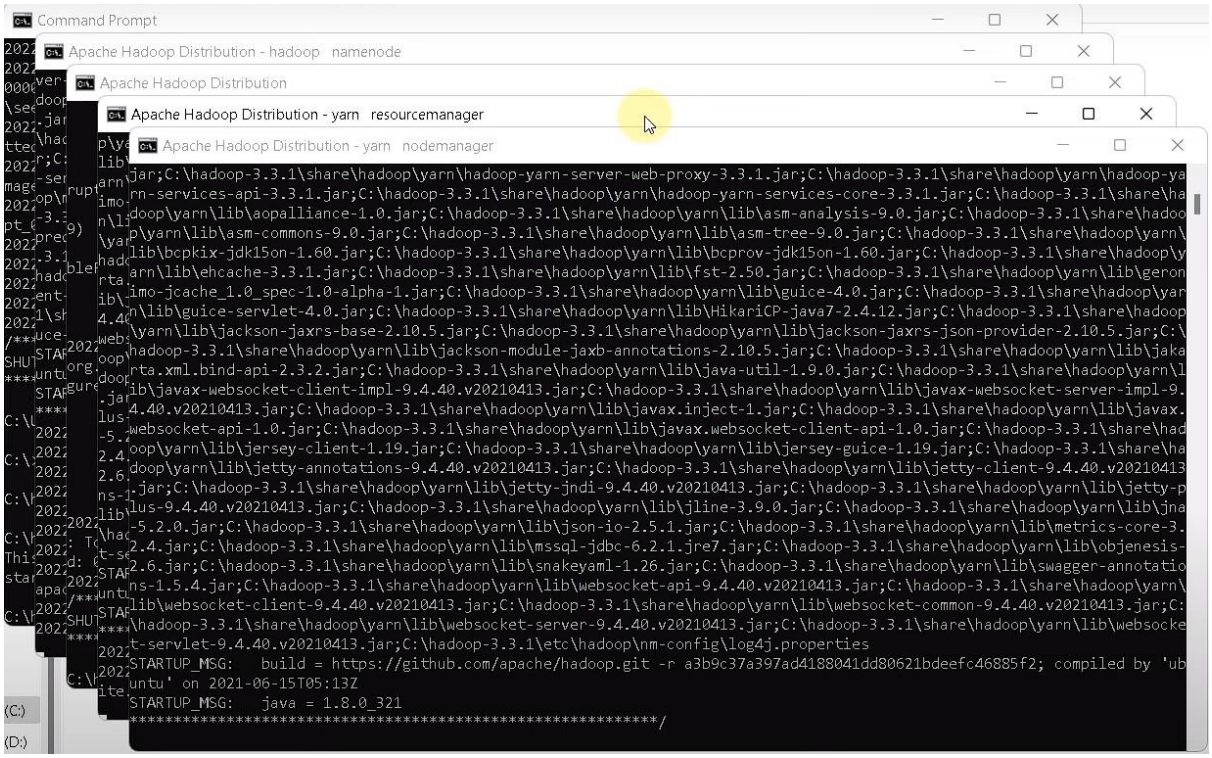
Setting Path for Java-jdk:

core-site.xml | mapred-site.xml | hdfs-site.xml | yarn-site.xml | hadoop-env.cmd

```
@echo off
@rem Licensed to the Apache Software Foundation (ASF) under one or more
@rem contributor license agreements.  See the NOTICE file distributed with
@rem this work for additional information regarding copyright ownership.
@rem The ASF licenses this file to You under the Apache License, Version 2.0
@rem (the "License"); you may not use this file except in compliance with
@rem the License.  You may obtain a copy of the License at
@rem
@rem     http://www.apache.org/licenses/LICENSE-2.0
@rem
@rem Unless required by applicable law or agreed to in writing, software
@rem distributed under the License is distributed on an "AS IS" BASIS,
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
@rem See the License for the specific language governing permissions and
@rem limitations under the License.

@rem Set Hadoop-specific environment variables here.

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

@rem The java implementation to use.  Required.
set JAVA_HOME= C:\Program Files\Java\jdk-20
```

Setting Environmental Variable for Hadoop:





RESULT:

Hence Hadoop is installed and configured successfully.

EX NO : 6 DATE:	Use Apache Hadoop to write a program to use the API's of Hadoop to interact with it

AIM:

To use Apache Hadoop to write a program to use the API's of Hadoop to interact with it.

PROCEDURE:

Step 1: Once you have downloaded a test dataset, we can write an application to read a file from the local file system and write the contents to Hadoop Distributed File System.

```
package com.hadoop.hdfs.writer;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.util.Tool;
import java.io.BufferedInputStream;
import java.io.FileInputStream;
import java.io.InputStream;
import java.io.OutputStream;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IOUtils;
import org.apache.hadoop.util.ToolRunner;
public class HdfsWriter extends Configured implements Tool {
public static final String FS_PARAM_NAME = "fs.defaultFS";
public int run(String[] args) throws Exception {
if (args.length < 2) {
System.err.println("HdfsWriter [local input path] [hdfs output path]");
return 1;
}
String localInputPath = args[0];
Path outputPath = new Path(args[1]);
 Configuration conf = getConf();
System.out.println("configured filesystem = " + conf.get(FS_PARAM_NAME)); FileSystem
fs = FileSystem.get(conf);
if (fs.exists(outputPath)) {
System.err.println("output path exists");
return 1;
}
```

```
OutputStream os = fs.create(outputPath);
InputStream is = new BufferedInputStream(new FileInputStream(localInputPath));
IOUtils.copyBytes(is, os, conf);
return 0;
}
public static void main( String[] args ) throws Exception {
int returnCode = ToolRunner.run(new HdfsWriter(), args);
System.exit(returnCode);
}
}
```

Step 2: Export the Jar file and run the code from terminal to write a sample file to HDFS.

```
[training@localhost ~]$ hadoop jar HdfsWriter.jar com.hadoop.hdfs.writer.HdfsWriter
sample.txt
/user/training/HdfsWriter_sample.txt
```

Step 3: Verify whether the file is written into HDFS and check the contents of the file.

```
[training@localhost ~]$ hadoop fs -cat /user/training/HdfsWriter_sample.txt
```

Step 4: Next, we write an application to read the file we just created in Hadoop Distributed File System and write its contents back to the local file system.

```
package com.hadoop.hdfs.reader;
import java.io.BufferedOutputStream;
import java.io.FileOutputStream;
import java.io.InputStream;
import java.io.OutputStream;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.conf.Configured;

import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IOUtils;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
public class HdfsReader extends Configured implements Tool {
public static final String FS_PARAM_NAME = "fs.defaultFS";
 public int run(String[] args) throws Exception {
if (args.length < 2) {
        System.err.println("HdfsReader [hdfs input path] [local output path]");
         return 1;
}
Path inputPath = new Path(args[0]); String localOutputPath = args[1];
Configuration conf = getConf();
```

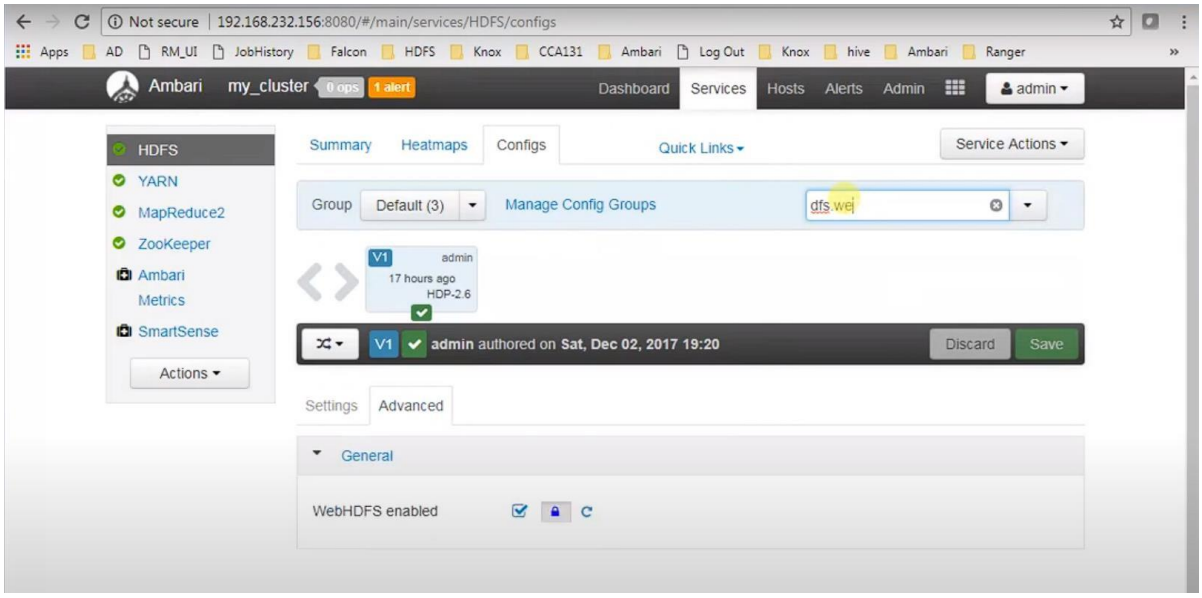
```
System.out.println("configured filesystem = " + conf.get(FS_PARAM_NAME)); FileSystem
fs = FileSystem.get(conf);
InputStream is = fs.open(inputPath);
OutputStream os = new BufferedOutputStream(new FileOutputStream(localOutputPath));
IOUtils.copyBytes(is, os, conf);
return 0;
}
public static void main( String[] args ) throws Exception {
int returnCode = ToolRunner.run(new HdfsReader(), args);
System.exit(returnCode);
}
}
```

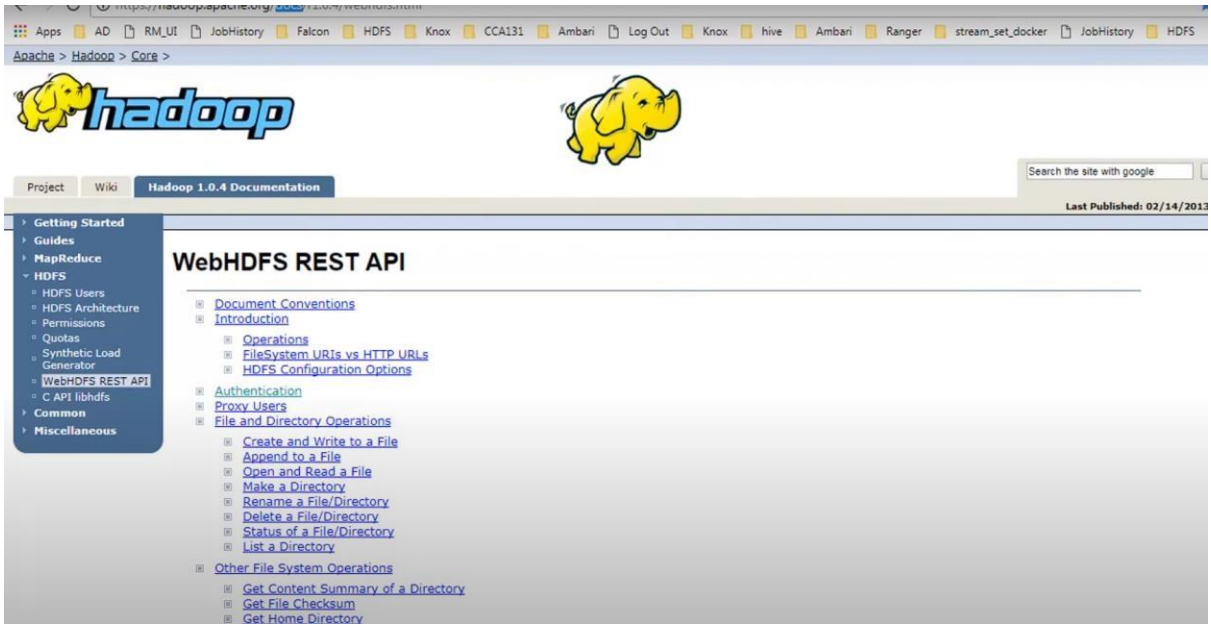
Step 5: Export the Jar file and run the code from terminal to write a sample file to HDFS.
[training@localhost ~]$ hadoop
jarHdfsReader.jarcom.hadoop.hdfs.reader.HdfsReader/user/training/HdfsWriter_sample.txt
/home/training/HdfsReader_sample.txt

Step 6: Verify whether the file is written back into local file system. [training@localhost ~]$ hadoop fs -cat /user/training/HdfsWriter_sample.txt
FileSystem is an abstract class that represents a generic file system. Most Hadoop file system implementations can be accessed and updated through the FileSystem object. To create an instance of the HDFS, you call the method FileSystem.get(). The FileSystem.get() method will look at the
URI assigned to the fs.defaultFS parameter of the Hadoop configuration files on your classpath and choose the correct implementation of the FileSystem class to instantiate. Thefs.defaultFS parameter of HDFS has the value hdfs://.
Once an instance of the FileSystem class has been created, the HdfsWriter class calls
the create() method to create a file in HDFS. The create() method return an OutputStream object, which can be manipulated using normal Java I/O methods. Similarly HdfsReader calls the method open() to open a file in HDFS, which returns an InputStream object that can be used to read the contents of the file.
The FileSystem API is extensive. To demonstrate some of the other methods available in the API,
we can add some error checking to the HdfsWriter and HdfsReader classes we created

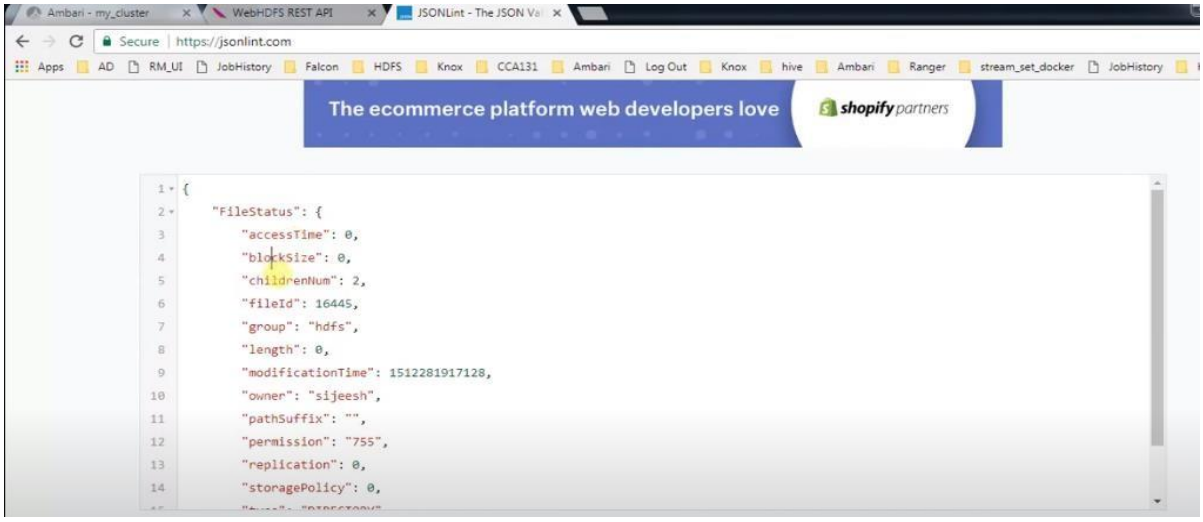
To check whether the file exists before we call create(), use: boolean exists =
fs.exists(inputPath);
To check whether the path is a file, use:
boolean isFile = fs.isFile(inputPath); To rename a file that already exits, use:
boolean renamed = fs.rename(inputPath, new Path("old_file.txt"));

OUTPUT:






```
[sijeesh@master ~]$ hdfs dfs -ls /user
Found 2 items
drwxrwx---   - ambari-qa hdfs          0 2017-12-02 06:02 /user/ambari-qa
drwxr-xr-x   - sijeesh   hdfs          0 2017-12-02 22:18 /user/sijeesh
[sijeesh@master ~]$ curl -i "http://master.hadoop.com:50070/webhdfs/v1/user/sijeesh?user.name=sijeesh&op=GETFILESTATUS"
HTTP/1.1 200 OK
Cache-Control: no-cache
Expires: Sun, 03 Dec 2017 07:11:50 GMT
Date: Sun, 03 Dec 2017 07:11:50 GMT
Pragma: no-cache
Expires: Sun, 03 Dec 2017 07:11:50 GMT
Date: Sun, 03 Dec 2017 07:11:50 GMT
Pragma: no-cache
Content-Type: application/json
X-FRAME-OPTIONS: SAMEORIGIN
Set-Cookie: hadoop.auth="u=sijeesh&p=sijeesh&t=simple&e=1512321110450&s=KrvzIkxW9ut742vjF6ku3s1ouUw="; Path=/; HttpOnly
Transfer-Encoding: chunked
Server: Jetty(6.1.26.hwx)

{"FileStatus":{"accessTime":0,"blockSize":0,"childrenNum":2,"fileId":16445,"group":"hdfs","length":0,"modificationTime":1512281917128,
"owner":"sijeesh","pathSuffix":"","permission":"755","replication":0,"storagePolicy":0,"type":"DIRECTORY"}}[sijeesh@master ~]$
```




```
[sijeesh@master ~]$ ls
ambari.properties  Desktop  Documents  Downloads  Music  Pictures  Public  Templates  test.txt  Videos
[sijeesh@master ~]$ hdfs dfs -ls /user/sijeesh
Found 4 items
drwxr-xr-x   - hdfs    hdfs             0 2017-12-02 20:46 /user/sijeesh/abcd
-rwxrwxrwx   3 sijeesh sijeesh       7156 2017-12-02 22:18 /user/sijeesh/ambari.properties
drwxr-xr-x   - sijeesh hdfs             0 2017-12-02 23:33 /user/sijeesh/new_dir
drwxr-xr-x   - hdfs    hdfs             0 2017-12-02 23:34 /user/sijeesh/new_dir2
[sijeesh@master ~]$ curl -X PUT "http://master.hadoop.com:50070/webhdfs/v1/user/sijeesh/test.txt?user.name=sijeesh&op=CREATE"
[sijeesh@master ~]$ curl -i -X PUT "http://master.hadoop.com:50070/webhdfs/v1/user/sijeesh/test.txt?user.name=sijeesh&op=CREATE"
HTTP/1.1 307 TEMPORARY_REDIRECT
Cache-Control: no-cache
Expires: Sun, 03 Dec 2017 07:39:06 GMT
Date: Sun, 03 Dec 2017 07:39:06 GMT
Pragma: no-cache
Expires: Sun, 03 Dec 2017 07:39:06 GMT
Date: Sun, 03 Dec 2017 07:39:06 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Set-Cookie: hadoop.auth="u=sijeesh&p=sijeesh&t=simple&e=1512322746112&s=DR4RM3aNBh5abChzKQcMia7LWY4="; Path=/; HttpOnly
Location: http://master.hadoop.com:50075/webhdfs/v1/user/sijeesh/test.txt?op=CREATE&user.name=sijeesh&namenoderpcaddress=master.hadoop
.com:8020&createflag=&createparent=true&overwrite=false
Content-Type: application/octet-stream
Content-Length: 0
Server: Jetty(6.1.26.hwx)

[sijeesh@master ~]$
```

```
[sijeesh@master ~]$ hdfs dfs -cat /user/sijeesh/test.txt
hello this is test file
[sijeesh@master ~]$
[sijeesh@master ~]$ curl -i -L "http://master.hadoop.com:50070/webhdfs/v1/user/sijeesh/test.txt?user.name=sijeesh&op=OPEN"
HTTP/1.1 307 TEMPORARY_REDIRECT
Cache-Control: no-cache
Expires: Sun, 03 Dec 2017 07:42:30 GMT
Date: Sun, 03 Dec 2017 07:42:30 GMT
Pragma: no-cache
Expires: Sun, 03 Dec 2017 07:42:30 GMT
Date: Sun, 03 Dec 2017 07:42:30 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Set-Cookie: hadoop.auth="u=sijeesh&p=sijeesh&t=simple&e=1512322950080&s=C1Q7/2DYac+cUeicVJbKxUqogBE="; Path=/; HttpOnly
Location: http://master.hadoop.com:50075/webhdfs/v1/user/sijeesh/test.txt?op=OPEN&user.name=sijeesh&namenoderpcaddress=master.hadoop.c
om:8020&offset=0
Content-Type: application/octet-stream
Content-Length: 0
Server: Jetty(6.1.26.hwx)

HTTP/1.1 200 OK
Access-Control-Allow-Methods: GET
Access-Control-Allow-Origin: *
Content-Type: application/octet-stream
Connection: close
Content-Length: 24

hello this is test file
[sijeesh@master ~]$
```

RESULT:

Hence the program to use Apache Hadoop to write a program to use the API's of Hadoop to interact with it is successfully executed.

EX NO: 7 DATE:	**Use Apache Hadoop write a word count program to demonstrate the use of Map and Reduce tasks**

AIM:

To use Apache hadoop write a word count program to demonstrate the use of Map and Reduce tasks.

PROCEDURE:

Step 1: Download MapReduceClient.jar

https://github.com/MuhammadBilalYar/HADOOP-INSTALLATION-ON-WINDOW10/blob/master/MapReduceClient.jar

Step 2: Download Input_file.txt

https://github.com/MuhammadBilalYar/HADOOP-INSTALLATION-ON-WINDOW10/blob/master/input_file.txt

Place both files in "C:/"

Step 3: Open cmd in Administrative mode and move to "C:/Hadoop-2.8.0/sbin" and start cluster

Start-all.cmd

Step 4: Create an input directory in HDFShadoop fs -mkdir /input_dir

Step 5: Copy the input text file named input_file.txt in the input directory (input_dir) of HDFS

hadoop fs -put C:/input_file.txt /input_dir

Step 6: Verify input_file.txt available in HDFS input directory (input_dir)hadoop fs -ls /input_dir/

Step 7: Verify content of the copied filehadoop dfs -cat /input_dir/input_file.txt

Step 8: Run MapReduceClient.jar and also provide input and out directorieshadoop jar C:/MapReduceClient.jar wordcount /input_dir /output_dir

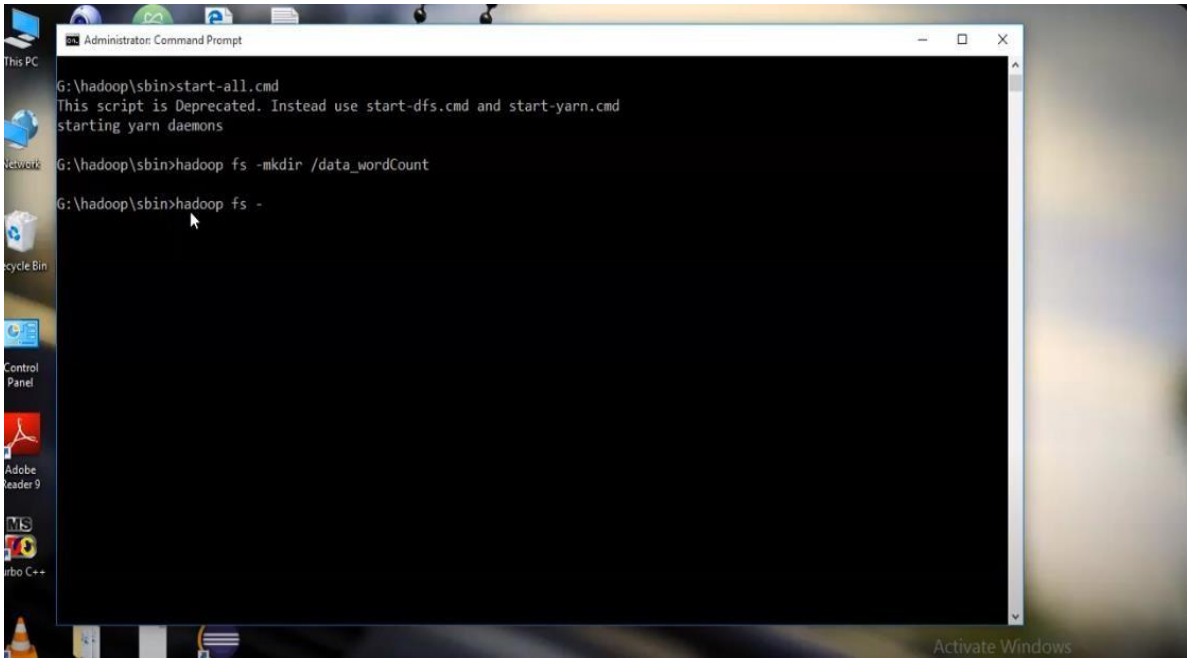
Step 9: Verify content for generated output fileSome Other useful commands in hadoop:

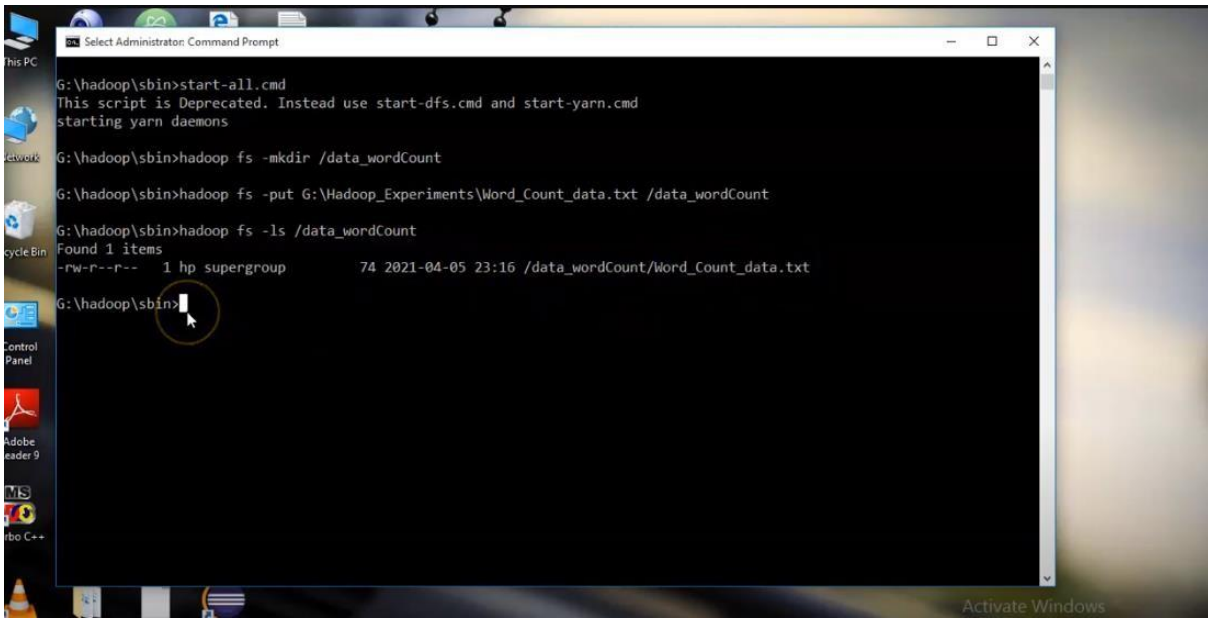
To leave Safe mode

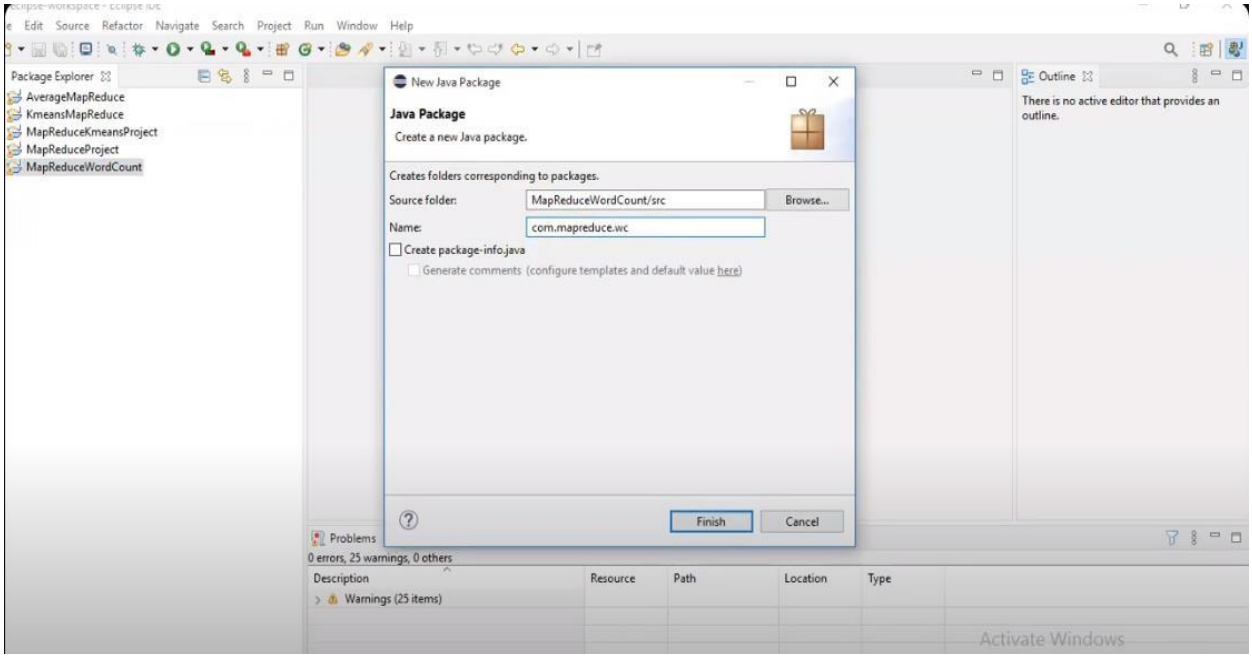
hadoop dfsadmin –safemode leave To delete file from HDFS directory

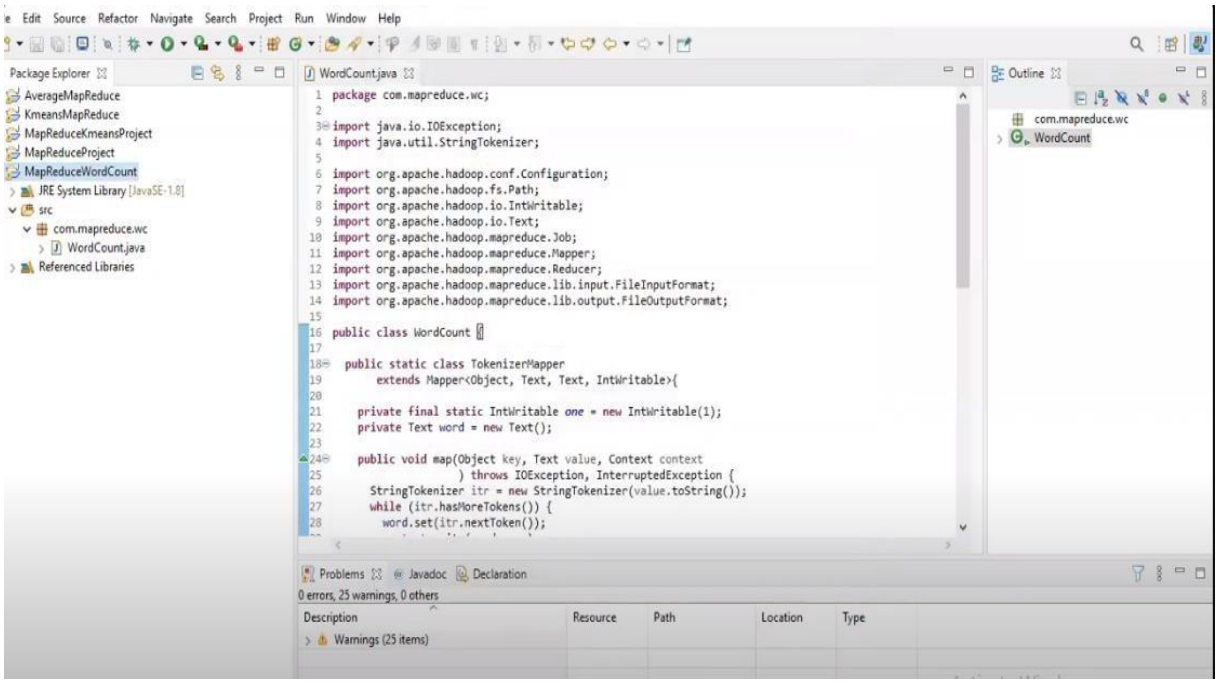
hadoop fs -rm -r /iutput_dir/input_file.txtTo delete directory from HDFS directoryhadoop fs -rm -r /iutput_dir

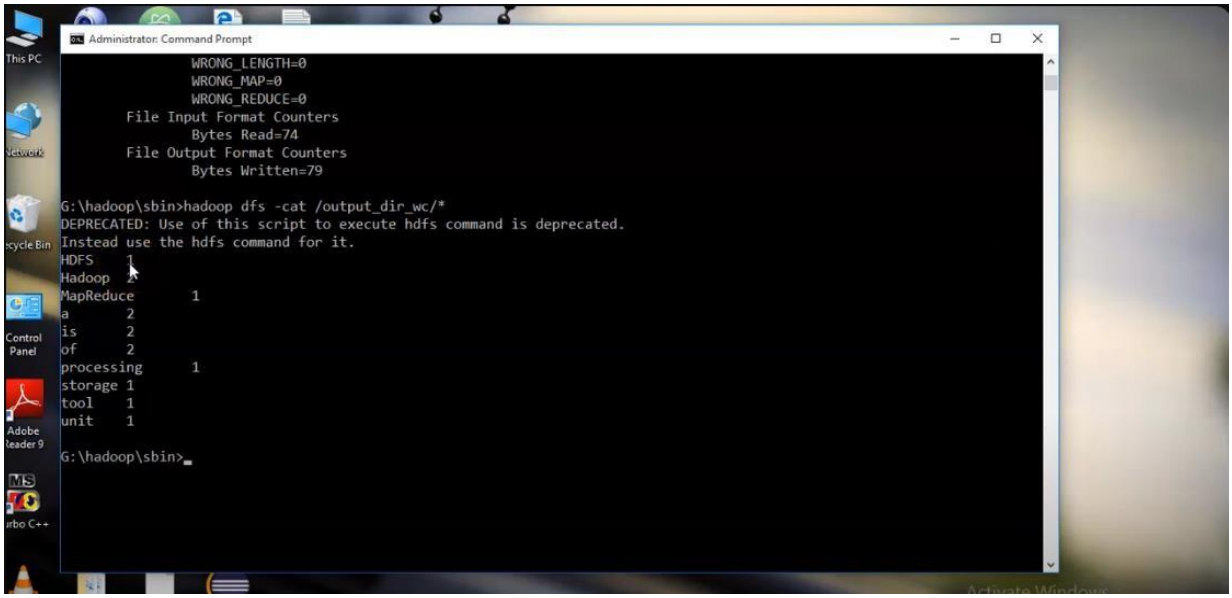
OUTPUT:

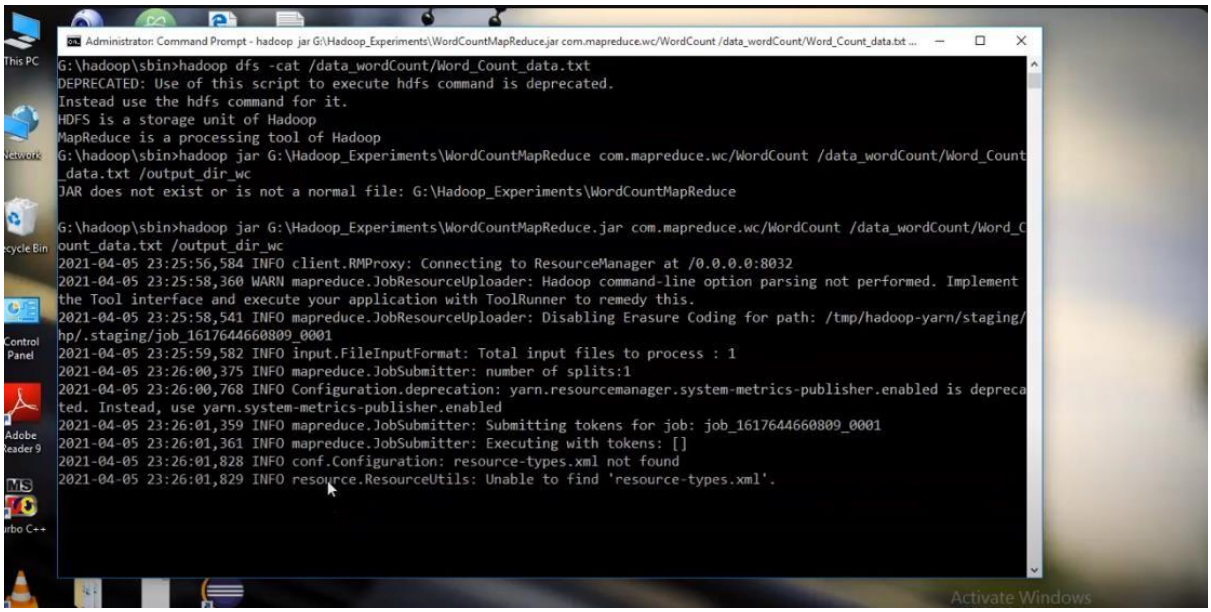





Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
AverageMapReduce
KmeansMapReduce
MapReduceKmeansProject
MapReduceProject
MapReduceWordCount
New Java Package
Java Package
Create a new Java package.
Creates folders corresponding to packages.
Source folder:
MapReduceWordCount/src
Browse...
Name:
com.mapreduce.wc
Create package-info.java
Generate comments (configure templates and default value here)
Finish
Cancel
Outline
There is no active editor that provides an outline.
Problems
0 errors, 25 warnings, 0 others
Description
Resource
Path
Location
Type
Warnings (25 items)
Activate Windows


Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
AverageMapReduce
KmeansMapReduce
MapReduceKmeansProject
MapReduceProject
MapReduceWordCount
JRE System Library [JavaSE-1.8]
src
com.mapreduce.wc
WordCount.java
Referenced Libraries
WordCount.java
package com.mapreduce.wc;
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class WordCount {
public static class TokenizerMapper
extends Mapper<Object, Text, Text, IntWritable>{
private final static IntWritable one = new IntWritable(1);
private Text word = new Text();
public void map(Object key, Text value, Context context
) throws IOException, InterruptedException {
StringTokenizer itr = new StringTokenizer(value.toString());
while (itr.hasMoreTokens()) {
word.set(itr.nextToken());
Outline
com.mapreduce.wc
WordCount
Problems
Javadoc
Declaration
0 errors, 25 warnings, 0 others
Description
Resource
Path
Location
Type
Warnings (25 items)





RESULT:

Hence the experiment to use apache hadoop write a word count program to demonstrate the useof Map and Reduce tasks.

EX NO: 8 DATE:	**Installation Of Aneka**

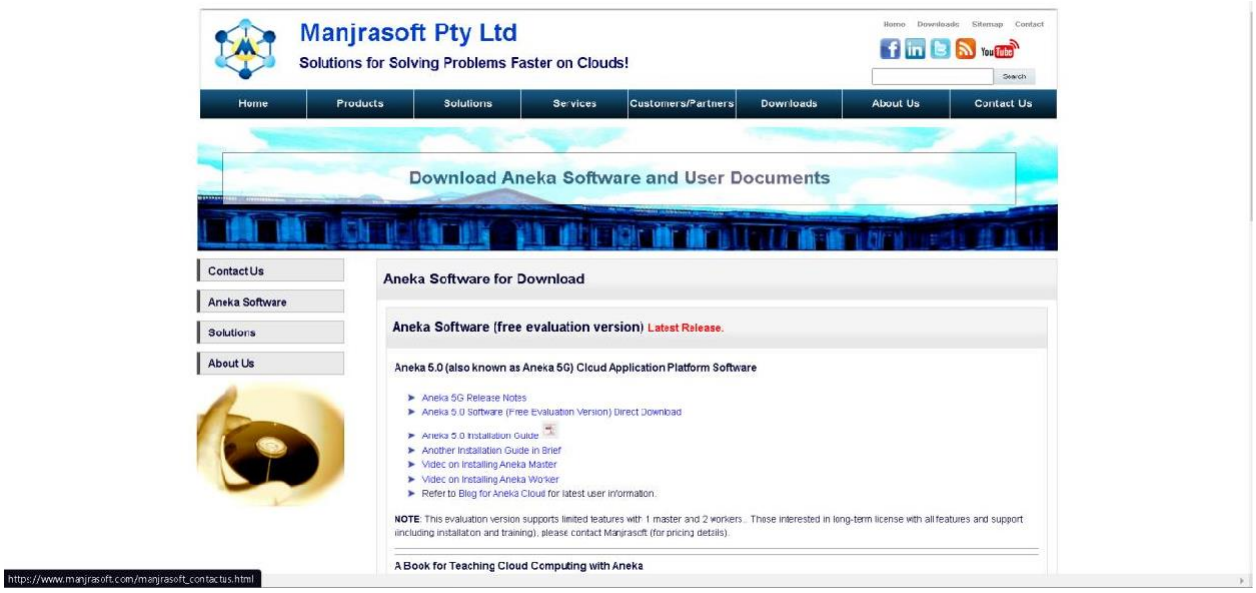
AIM:

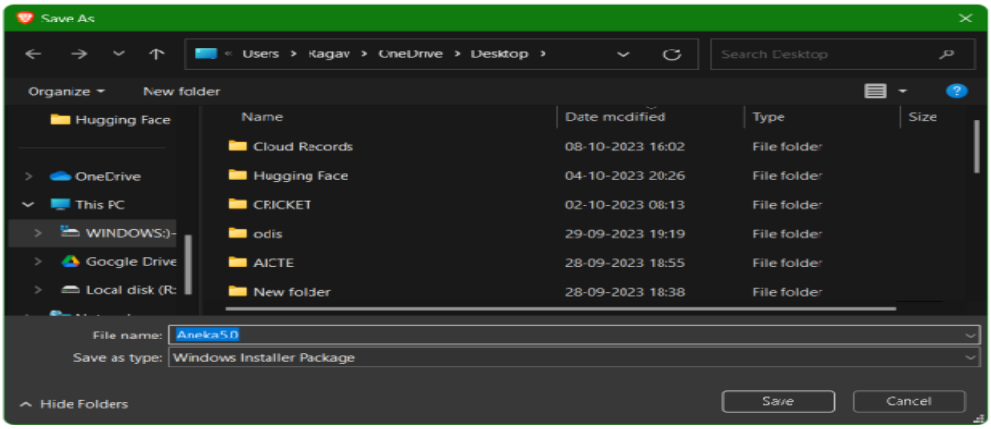
To install Aneka on your PC.

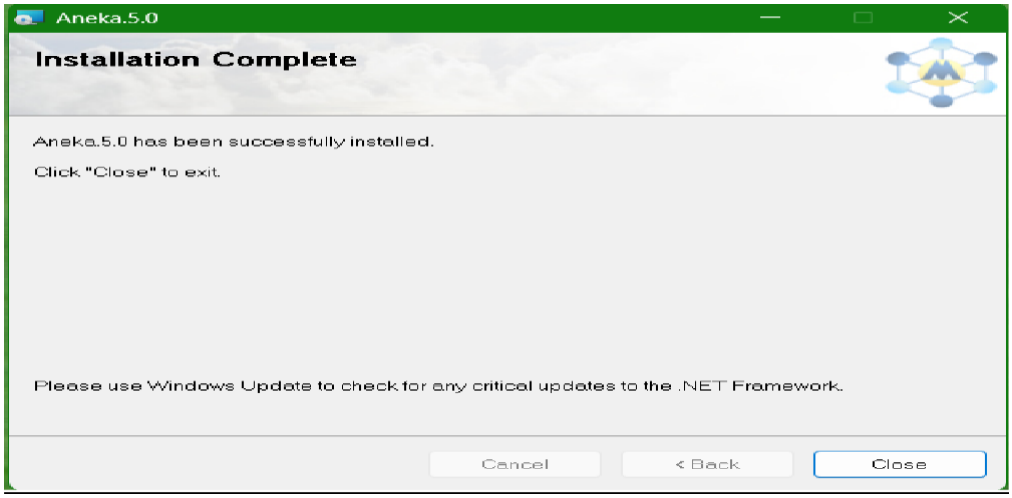
PROCEDURE:

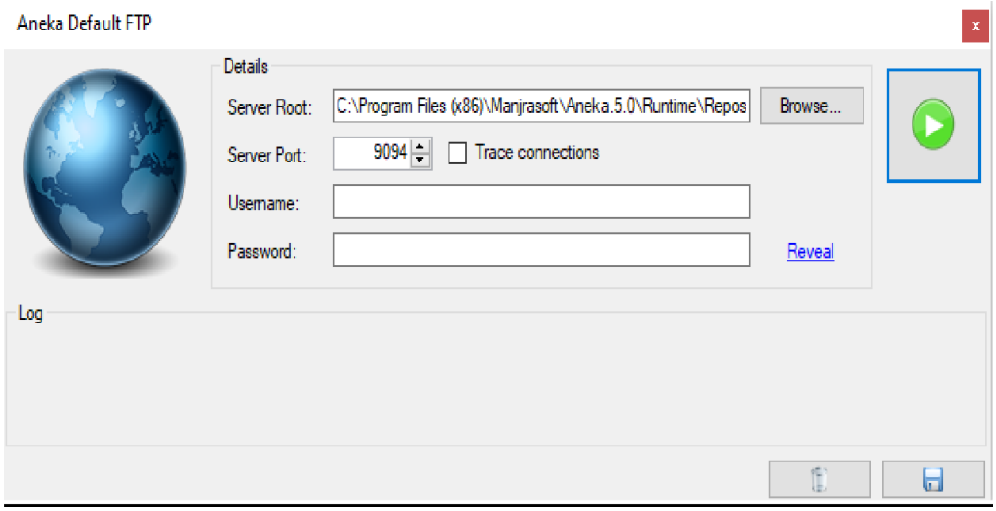
i. Install the aneka cloud management studio from the websites and open
ii. Run the installer packages to start the setup wizard
iii. Click on next
iv. Specify the installation folder.
v. Confirm and start the installation by clicking next.
vi. After the installation gets completed close the wizard
vii. Launch aneka management studio from the start menu.
viii. Click "yes" We should add a repository to do remote installation so, specify the wizardby name method.
ix. access URL access user access password and click "ok"
x. Now, the repository will be present in the left navigation panel.
xi. Click the file menu and choose add machine and add your required machine
xii. Click on the machine credential and create the credentials.

OUTPUT:









RESULT:

Hence Aneka is successfully installed.

EX NO: 9 DATE:	**Case study on Amazon Web Services (AWS)**

Amazon.com Case Study:

Amazon.com is the world's largest online retailer. In 2011, Amazon.com switched from tape backup to using Amazon Simple Storage Service (Amazon S3) for backing up the majority of itsOracle databases. This strategy reduces complexity and capital expenditures, provides faster backup and restore performance, eliminates tape capacity planning for backup and archive, and frees up administrative staff for higher value operations. The company was able to replace their backup tape infrastructure with cloud-based Amazon S3 storage, eliminate backup software, andexperienced a 12X performance improvement, reducing restore time from around 15 hours to 2.5hours in select scenarios.

The Challenge:

As Amazon.com grows larger, the sizes of their Oracle databases continue to grow, and so doesthe sheer number of databases they maintain. This has caused growing pains related to backing up legacy Oracle databases to tape and led to the consideration of alternate strategies including the use of Cloud services of Amazon Web Services (AWS), a subsidiary of Amazon.com. Someof the business challenges Amazon.com faced included:

Utilization and capacity planning is complex. and time and capital expense budget are at a premium. Significant capital expenditures were required over the years for tape hardware, data center space for this hardware, and enterprise licensing fees for tape software. During that time, managing tape infrastructure required highly skilled staff to spend time with setup. certification and engineering archive planning instead of on higher value projects. And at the end of every fiscal year, projecting future capacity requirements required time consuming audits, forecasting,and budgeting.

The cost of backup software required to support multiple tape devices sneaks up on you Tape robots provide basic read/write capability, but in order to fully utilize them, you must invest in proprietary tape backup software. For Amazon.com, the cost of the software had been high, andadded significantly to overall backup costs. The cost of this software was an ongoing budgetingpain point, but one that was difficult to address as long as backups needed to be written to tape devices.

Maintaining reliable backups and being fast and efficient when retrieving data requires a lot of time and effort with tape. When data needs to be durably stored on tape, multiple copies are required. When everything is working correctly, and there is minimal contention for tape resources, the tape robots and backup software can easily find the required data. However, if there is a hardware failure, human intervention is necessary to restore from tape. Contention for tape drives resulting from multiple users' tape requests slows down restore processes even more.This adds to the recovery time objective (RTO) and makes achieving it more challenging compared to backing up to Cloud storage.

Why Amazon Web Services?

Amazon.com initiated the evaluation of Amazon S3 for economic and performance improvements related to data backup. As part of that evaluation, they considered security. availability, and performance aspects of Amazon S3 backups. Amazon.com also executed a cost-benefit analysis to ensure that a migration to Amazon S3 would be financially worthwhile. That cost benefit analysis included the following elements:

- Performance advantage and cost competitiveness. It was important that the overall costsof the backups did not increase. At the same time, Amazon.com required faster backup and recovery performance. The time and effort required for backup and for recovery operations proved to be a significant improvement over tape, with restoring from Amazon S3 running from two to twelve times faster than a similar restore from tape. Amazon.com required any new backup medium to provide improved performance whilemaintaining or reducing overall costs. Backing up to on-premises disk based storage would have improved performance, but missed on cost competitiveness. Amazon S3 Cloud based storage met both criteria.
- Greater durability and availability. Amazon S3 is designed to provide 99.999999999% durability and 99.99% availability of objects over a given year. Amazon.com comparedthese figures with those observed from their tape infrastructure, and determined that Amazon S3 offered significant improvement.
- Less operational friction. Amazon.com DBAs had to evaluate whether Amazon S3 backups would be viable for their database backups. They determined that using AmazonS3 for backups was easy to implement because it worked seamlessly with Oracle RMAN.
- Strong data securityAmazon.com found that AWS met all of their requirements for physical security, security accreditations, and security processes, protecting data in flight.data at rest, and utilizing suitable encryption standards.

The Benefits:

With the migration to Amazon S3 well along the way to completion. Amazon.com has realized several benefits, including,

- Elimination of complex and time-consuming tape capacity planning. Amazon.com is growing larger and more dynamic each year, both organically and as a result of acquisitions. AWS has enabled Amazon.com to keep pace with this rapid expansion, andto do so seamlessly. Historically, Amazon.com business groups have had to write annual backup plans, quantifying the amount of tape storage that they plan to use for the year and the frequency with which they will use the tape resources. These plans are then used to charge each organization for their tape usage, spreading the cost among many teams. With Amazon S3, teams simply pay for what they use, and are billed for their usage as they go. There are virtually no upper limits as to how much data can be stored in Amazon S3, and so there are no worries about running out of resources. For teams adopting Amazon S3 backups, the need for formal planning has been all but eliminated.
- Reduced capital expenditures. Amazon.com no longer needs to acquire tape robots, tape drives, tape inventory, data center space, networking gear, enterprise backup software, or

predict future tape consumption. This eliminates the burden of budgeting for capital equipment well in advance as well as the capital expense.

- Immediate availability of data for restoring - no need to locate or retrieve physical tapes. Whenever a DBA needs to restore data from tape, they face delays. The tape backup software needs to read the tape catalog to find the correct files to restore, locate the correct tape, mount the tape, and read the data from it. In almost all cases the data is spread across multiple tapes, resulting in further delays. This, combined with contention for tape drives resulting from multiple users' tape requests, slows the process down even more. This is especially severe during critical events such as a data center outage, when many databases must be restored simultaneously and as soon as possible. None of these problems occur with Amazon S3. Data restores can begin immediately, with no waiting or tape queuing- and that means the database can be recovered much faster.

- Backing up a database to Amazon S3 can be two to twelve times faster than with tape drives. As one example, in a benchmark test a DBA was able to restore 3.8 terabytes in

2.5 hours over gigabit Ethernet. This amounts to 25 gigabytes per minute, or 422MB per second. In addition, since Amazon.com uses RMAN data compression, the effective restore rate was 3.37 gigabytes per second. This 2.5 hours compares to, conservatively. 10-15 hours that would be required to restore from tape.

- Easy implementation of Oracle RMAN backups to Amazon S3. The DBAs found it easy to start backing up their databases to Amazon S3. Directing Oracle RMAN backups to Amazon S3 requires only a configuration of the Oracle Secure Backup Cloud (SBC) module. The effort required to configure the Oracle SBC module amounted to an hour or less per database. After this one-time setup, the database backups were transparently redirected to Amazon S3.

- . Durable data storage provided by Amazon S3, which is designed for 11 nines durability. On occasion, Amazon.com has experienced hardware failures with tape infrastructure - tapes that break, tape drives that fail, and robotic components that fail. Sometimes this happens when a DBA is trying to restore a database, and dramatically increases the mean time to recover (MTTR). With the durability and availability of Amazon S3, these issues are no longer a concern.

- Freeing up valuable human resources. With tape infrastructure. Amazon.com had to seek out engineers who were experienced with very large tape backup installations - a specialized, vendor-specific skill set that is difficult to find. They also needed to hire data center technicians and dedicate them to problem-solving and troubleshooting hardware issues-replacing drives, shuffling tapes around shipping and tracking tapes, and so on. Amazon S3 allowed them to free up these specialists from day-to-day operations so that they can work on more valuable, business-critical engineering tasks.

- Elimination of physical tape transport to off-site location. Any company that has been storing Oracle backup data offsite should take a hard look at the costs involved in transporting, securing and storing their tapes offsite these costs can be reduced or possibly eliminated by storing the data in Amazon S3.

As the world's largest online retailer, Amazon.com continuously innovates in order to provide improved customer experience and offer products at the lowest possible prices. One such innovation has been to replace tape with Amazon S3 storage for database backups. This innovation is one that can be easily replicated by other organizations that back up their Oracle databases to tape.

About Amazon.com

Amazon.com is the world's largest online retailer.

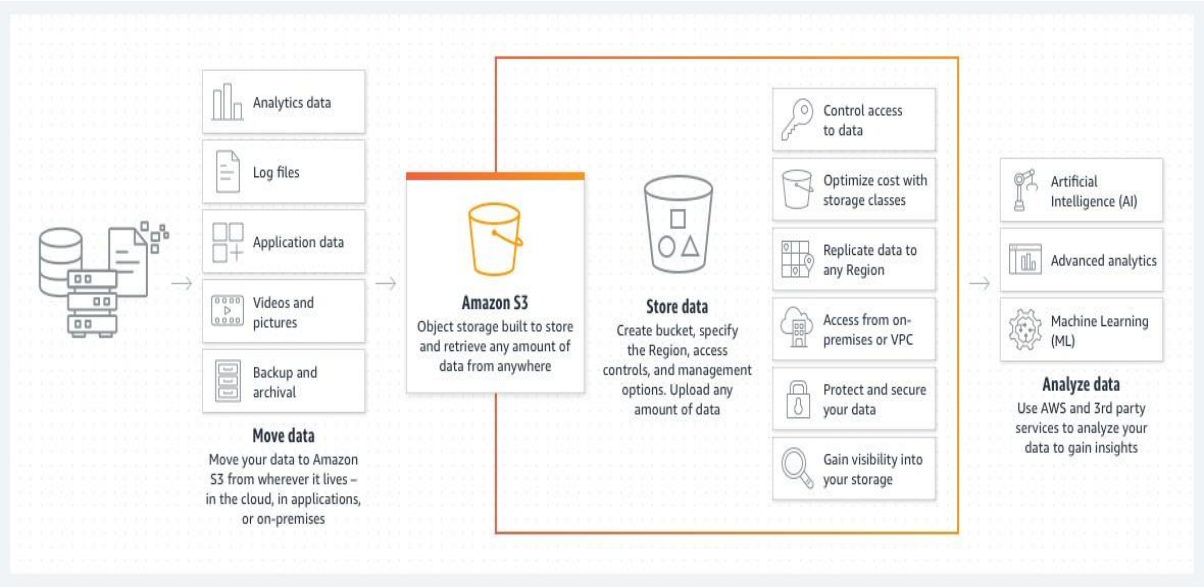
Benefits of AWS

- Eliminated complex. time-consuming tape-capacity planning
- Reduced capital expenditures
- Gained immediate access to data for restoring
- Reduced database-backup times by 12x

AWS Services Used: Amazon S3

How it works

Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance. Customers of all sizes and industries can store and protect any amount of data for virtually any use case, such as data lakes, cloud-native applications, and mobile apps. With cost-effective storage classes and easy-to-use management features, you can optimize costs, organize data, and configure fine-tuned access controls to meet specific business, organizational, and compliance requirements.



Use Case:

1. Build a data lake

   Run big data analytics, artificial intelligence (AI), machine learning (ML), and high performance computing (HPC) applications to unlock Learn more about building a data lake on Amazon S3 »>

2. Back up and restore critical data

   Meet Recovery Time Objectives (RTO). Recovery Point Objectives (RPO), and compliance requirements with S3's robust replication features.

   Learn more about backup and restore >>

3. Archive data at the lowest cost

   Move data archives to the Amazon S3 Glacier storage classes to lower costs, eliminate operational complexities, and gain new insights.
   Learn more about archiving with the s3 Glacier storage classes

4. Run cloud-native applications

   Build fast, powerful mobile and web-based cloud-native apps that scale automatically in a highly available configuration.

   Learn more about cloud-native applications >>

Customers:

Georgia-Pacific

NASDAQ

RYANAIR

zalando

How to get started:

Find out how Amazon S3 works

Amazon S3 Features:

Amazon S3 has various features you can use to organize and manage your data in ways that support specific use cases, enable cost efficiencies, enforce security, and meet compliance requirements. Data is stored as objects within resources called -buckets!, and a single object can be up to 5 terabytes in size. S3 features include capabilities to append metadata tags to objects, move and store data across the S3 Storage Classes, configure and enforce data access controls, secure data against unauthorized users, run big data analytics, monitor data at the object and bucket levels, and view storage usage and activity trends across your organization. Objects can be accessed through S3 Access Points or directly through the bucket hostname.

Storage management and monitoring:

Amazon S3's flat, non-hierarchical structure and various management features are helping customers of all sizes and industries organize their data in ways that are valuable to their businesses and teams. All objects are stored in S3 buckets and can be organized with shared names called prefixes. You can also append up to 10 key-value pairs called S3 object tags to each object, which can be created, updated, and deleted throughout an object's lifecycle. To keep track of objects and their respective tags, buckets, and prefixes, you can use an S3 Inventory report that lists your stored objects within an S3 bucket or with a specific prefix. and their respective metadata and encryption status. S3 Inventory can be configured to generate reports on a daily or a weekly basis.

Storage management

With S3 bucket names, prefixes, object tags, and S3 Inventory, you have a range of ways to categorize and report on your data, and subsequently can configure other S3 features to take action. Whether you store thousands of objects or a billion. S3 Batch Operations makes it simple to manage your data in Amazon S3 at any scale. With S3 Batch Operations, you can copy objects between buckets, replace object tag sets, modify access controls, and restore archived objects from S3 Glacier Flexible Retrieval and S3 Glacier Deep Archive storage classes, with a single S3 API request or a few clicks in the S3 console. You can also use S3 Batch Operations to run AWS Lambda functions across your objects to execute custom business logic, such as processing data or transcoding image files. To get started, specify a list of target objects by using an S3 Inventory report or by providing a custom list, and then select the desired operation from a pre-populated menu. When an S3 Batch Operation requestis done, you'll receive a notification and a completion report of all changes made. Learn moreabout S3 Batch Operations by watching the video tutorials.

Amazon S3 also supports features that help maintain data version control, prevent accidental deletions, and replicate data to the same or different AWS Region. With S3 Versioning, you can easily preserve, retrieve, and restore every version of an object stored in Amazon S3, which allows you to recover from unintended user actions and application failures.

If you try to delete an object stored in an MFA Delete-enabled bucket, it will require twoforms of authentication: your AWS account credentials and the concatenation of a valid serial number, a space, and the six-digit code displayed on an approved authentication device, like a hardware key fob or a Universal 2nd Factor (U2F) security key.

With S3 Replication, you can replicate objects (and their respective metadata and object tags) to one or more destination buckets into the same or different AWS Regions for reduced latency, compliance, security, disaster recovery, and other use cases. You can configure S3 Cross-Region Replication (CRR) to replicate objects from a source S3 bucket to one or more destination buckets in different AWS Regions. S3 Same-Region Replication (SRR) replicates objects between buckets in the same AWS Region. While live replication like CRR and SRR automatically replicates newly uploaded objects as they are written to your bucket. S3 Batch Replication allows you to replicate existing objects. You can use S3 Batch Replication to backfill a newly created bucket with existing objects, retry objects that were previously unable to replicate, migrate data across accounts, or add new buckets to your data lake.
Amazon S3 Replication Time Control (S3 RTC) helps you meet compliance requirements for data replication by providing an SLA and visibility into replication times.

Amazon S3 Multi-Region Access Points accelerate performance by up to 60% when accessing data sets that are replicated across multiple AWS Regions. Based on AWS Global Accelerator. S3 Multi-Region Access Points consider factors like network congestion and thelocation of the requesting application to dynamically route your requests over the AWS network to the lowest latency copy of your data. S3 Multi-Region Access Points provide a single global endpoint that you can use to access a replicated data set, spanning multiple buckets in S3. This allows you to build multi-region applications with the same simple architecture that you would use in a single region, and then to run those applications anywhere in the world.

You can also enforce write-once-read-many (WORM) policies with S3 Object Lock. This S3 management feature blocks object version deletion during a customer-defined retention period so that you can enforce retention policies as an added layer of data protection or to meet compliance obligations. You can migrate workloads from existing WORM systems into Amazon S3, and configure S3 Object Lock at the object- and bucket-levels to prevent object version deletions prior to a pre-defined Retain Until Date or Legal Hold Date, Objects with S3 Object Lock retain WORM protection, even if they are moved to different storage classes

with an S3 Lifecycle policy. To track what objects have S3 Objeet Lock, you can refer to an S3 Inventory report that includes the WORM status of objects. S3 Object Lock can be configured in one of two modes. When deployed in Governance mode, AWS accounts with specific IAM permissions are able to remove S3 Object Lock from objects. If you require stronger immutability in order to comply with regulations, you can use Compliance Mode. In Compliance Mode, the protection cannot be removed by any user, including the root account.

Storage monitoring:

In addition to these management capabilities, use Amazon S3 features and other AWS services to monitor and control your S3 resources. Apply tags to S3 buckets to allocate costsacross multiple business dimensions (such as cost centers, application names, or owners), then use AWS Cost Allocation Reports to view the usage and costs aggregated by the buckettags. You can also use Amazon Cloud Watch to track the operational health of your AWS resources and configure billing alerts for estimated charges that reach a user-defined threshold.

Use AWS CloudTrail to track and report on bucket- and object-level activities, and configure S3 Event Notifications to trigger workflows and alerts or invoke AWS Lambda when a specific change is made to your S3 resources. S3 Event Notifications automatically transcodes media files as they're uploaded to S3. processes data files as they become available, and synchronizes objects with other data stores. Additionally, you can verify integrity of data transferred to and from Amazon S3, and can access the checksum information at any time using the GetObjectAttributes S3 API or an S3 Inventory report. Youcan choose from four supported checksum algorithms (SHA-1, SHA-256, CRC32, or CRC32C) for data integrity checking on your upload and download requests depending on your application needs. Learn more about S3 storage management and monitoring

Storage analytics and insights

S3 Storage Lens

S3 Storage Lens delivers organization-wide visibility into object storage usage, activity trends, and makes actionable recommendations to improve cost-efficiency and apply data protection best practices. S3 Storage Lens is the first cloud storage analytics solution to provide a single view of object storage usage and activity across hundreds, or even thousands. of accounts in an organization. with drill-downs to generate insights at the account, bucket, or even prefix level. Drawing from more than 14 years of experience helping customers optimize their storage, S3 Storage Lens analyzes organization-wide metrics to deliver contextual recommendations to find ways to reduce storage costs and apply best practices on data protection. To learn more, visit the storage analytics and insights page.

S3 Storage classes Analysis:

Amazon S3 Storage Class Analysis analyzes storage access patterns to help you decide when to transition the right data to the right storage class. This Amazon S3 feature observes data access patterns to help you determine when to transition less frequently accessed storage to a lower-cost storage class. You can use the results to help improve your S3 Lifecycle policies. You can configure storage class analysis to analyze all the objects in a bucket. Or, you can configure filters to group objects together for analysis by common prefix, by object tags, or by both prefix and tags. To learn more, visit the storage analytics and insights page.

Storage classes:

With Amazon S3, you can store data across a range of different S3 storage classes purpose-built for specific use cases and access patterns: S3 Intelligent-Tiering, S3 Standard, S3 Standard-Infrequent Access (S3 Standard-IA), S3 One Zone-Infrequent Access (S3 One Zone-IA), S3 Glacier Instant Retrieval, S3 Glacier Flexible Retrieval, S3 Glacier Deep Archive, and S3 Outposts.

Every S3 storage class supports a specific data access level at corresponding costs orgeographic location.

For data with changing, unknown, or unpredictable access patterns, such as data lakes, analytics, or new applications, use S3 Intelligent-Tiering, which automatically optimizes your storage costs. S3 Intelligent-Tiering automatically moves your data between three lowlatency access tiers optimized for frequent, infrequent, and rare access. When subsets of objects become archived over time, you can activate the archive access tier designed for asynchronous access.

For more predictable access patterns, you can store mission-critical production data in S3 Standard for frequent access, save costs by storing infrequently accessed data in S3 Standard-IA or S3 One Zone-IA, and archive data at the lowest costs in the archival storage classes- S3 Glacier Instant Retrieval. S3 Glacier Flexible Retrieval, and S3 Glacier Deep Archive. You can use S3 Storage Class Analysis to monitor access patterns across objects to discover data that should be moved to lower-cost storage classes. Then you can use this information to configure an S3 Lifecycle policy that makes the data transfer. S3 Lifecycle policies can also be used to expire objects at the end of their lifecycles.

If you have data residency requirements that can't be met by an existing AWS Region, you can use the S3 Outposts storage class to store your S3 data on premises using S3 on Outposts.Learn more by visiting S3 Storage Classes. S3 Storage Class Analysis, and S3 Lifecycle management

Access Management and Security

Access Management:

To protect your data in Amazon S3, by default, users only have access to the S3 resources they create. You can grant access to other users by using one or a combination of the following access management features: AWS Identity and Access Management (IAM) to create users and manage their respective access; Access Control Lists (ACLS) to make individual objects accessible to authorized users; bucket policies to configure permissions forall objects within a single S3 bucket: S3 Access Points to simplify managing data access to shared data sets by creating access points with names and permissions specific to each application or sets of applications; and Query String Authentication to grant time-limited access to others with temporary URLs. Amazon S3 also supports Audit Logs that list the requests made against your S3 resources for complete visibility into who is accessing what data

Amazon S3 offers flexible security features to block unauthorized users from accessing your data Use VPC endpoints to connect to S3 resources from your Amazon Virtual Private Cloud (Amazon VPC) and from on-premises. Amazon S3 supports both server-side encryption (with three key management options) and client-side encryption for data uploads. Use S3 Inventory to check the encryption status of your S3 objects (see storage management for more information on S3 Inventory).

S3 Block Public Access is a set of security controls that ensures S3 buckets and objects do not have public access. With a few clicks in the Amazon S3 Management Console, you canapply the S3 Block Public Access settings to all buckets within your AWS account or to specific S3 buckets.

Once the settings are applied to an AWS account, any existing or new buckets and objects associated with that account inherit the settings that prevent public access. S3 Block Public Access settings override other S3 access permissions, making it easy for the account administrator to enforce a -no public access! policy regardless of how an object is added, how a bucket is created, or if there are existing access permissions. S3 Block Public Access controls are auditable, provide a further layer of control, and use AWS Trusted Advisor bucket permission checks, AWS CloudTrail logs, and Amazon CloudWatch alarms. You should enable Block Public Access for all accounts and buckets that you do not want publicly accessible.

S3 Object Ownership is a feature that disables Access Control Lists (ACLS), changing ownership for all objects to the bucket owner and simplifying access management for data stored in S3. When you configure the S3 Object Ownership Bucket owner enforced setting. ACLS will no longer affect permissions for your bucket and the objects in it. All access control will be defined using resource-based policies, user policies, or some combination ofthese.

Using S3 Access Points that are restricted to a Virtual Private Cloud (VPC), you can easily firewall your S3 data within your private network. Additionally, you can use AWS Service Control Policies to require that any new S3 Access Point in your organization is restricted to VPC-only access.

IAM Access Analyzer for S3 is a feature that helps you simplify permissions management as you set, verify, and refine policies for your S3 buckets and access points. Access Analyzer for S3 monitors your existing bucket access policies to verify that they provide only the required access to your S3 resources. Access Analyzer for S3 evaluates your bucket access policies so that you can swiftly remediate any buckets with access that isn't required. When reviewing results that show potentially shared access to a bucket, you can Block Public Access to the bucket with a single click in the S3 console. For auditing purposes, you can download Access Analyzer for S3 findings as a CSV report. Additionally, the S3 console reports security warnings, errors, and suggestions from IAM Access Analyzer as you author your S3 policies. The console automatically runs more than 100 policy checks to validate your policies. These checks save you time, guide you to resolve errors, and help you apply security best practices.

IAM makes it easier for you to analyze access and reduce permissions to achieve least privilege

by providing the timestamp when a user or role last used S3 and the associated actions. Use this last accessed! information to analyze S3 access, identify unused permissions, and remove them confidently. To learn more see Refining Permissions UsingLast Accessed Data.

You can use Amazon Macie to discover and protect sensitive data stored in Amazon S3. Macie automatically gathers a complete S3 inventory and continually evaluates every bucket to alert on any publicly accessible buckets, unencrypted buckets, or buckets shared or replicated with AWS accounts outside of your organization. Then, Macie applies machine learning and pattern matching techniques to the buckets you select to identify and alert you tosensitive data, such as personally identifiable information (PII). As security findings are generated, they are pushed out to the Amazon CloudWatch Events, making it easy to integrate with existing workflow systems and to trigger automated remediation with services like AWS Step Functions to take action like closing a public bucket or adding resource tags.

AWS Privatel ink for S3 provides private connectivity between Amazon S3 and on-premises. You can provision interface VPC endpoints for S3 in your VPC to connect your on-premises applications directly with S3 over AWS Direct Connect or AWS VPN. Requests to interface VPC endpoints for S3 are automatically routed to S3 over the Amazon network. You can set security groups and configure VPC endpoint policies for your interface VPC endpoints for additional access controls. Learn more by visiting S3 access management and security and protecting data in Amazon S3

Data processing

S3 Object Lambda

With S3 Object Lambda you can add your own code to S3 GET, HEAD, and LIST requests to modify and process data as it is returned to an application. You can use custom code to modify the data returned by standard S3 GET requests to filter rows, dynamically resize images, redact confidential data, and much more. You can also use S3 Object Lambda to modify the output of S3 LIST requests to create a custom view of objects in a bucket and S3HEAD requests to modify object metadata like object name and size. Powered by AWS Lambda functions, your code runs on infrastructure that is fully managed by AWS. eliminating the need to create and store derivative copies of your data or to run expensive proxies, all with no changes required to applications.

S3 Object Lambda uses AWS Lambda functions to automatically process the output of a standard S3 GET, HEAD, or LIST request. AWS Lambda is a serverless compute service that runs customer-defined code without requiring management of underlying compute resources. With just a few clicks in the AWS Management Console, you can configure a Lambda function and attach it to a S3 Object Lambda Access Point. From that point forward. S3 will automatically call your Lambda function to process any data retrieved through the S3Object Lambda Access Point, returning a transformed result back to the application. You canauthor and execute your own custom Lambda functions, tailoring S3 Object Lambda's data transformation to your specific use case.

Learn more by visiting the S3 Object Lambda feature page.

Query in place

Amazon S3 has a built-in feature and complementary services that query data without needing to copy and load it into a separate analytics platform or data warehouse. This meansyou can run big data analytics directly on your data stored in Amazon S3. S3 Select is an S3feature designed to increase query performance by up to 400%, and reduce querying costs asmuch as 80%. It works by retrieving a subset of an object's data (using simple SQL expressions) instead of the entire object, which can be up to 5 terabytes in size.

Amazon S3 is also compatible with AWS analytics services Amazon Athena and Amazon Redshift Spectrum. Amazon Athena queries your data in Amazon S3 without needing to extract and load it into a separate service or platform. It uses standard SQL expressions toanalyze your data, delivers results within seconds, and is commonly used for ad hoc data discovery. Amazon Redshift Spectrum also runs SQL queries directly against data at rest in Amazon S3, and is more appropriate for complex queries and large data sets (up to exabytes).

Because Amazon Athena and Amazon Redshift share a common data catalog and data formats, you can use them both against the same data sets in Amazon S3. Learn more byvisiting building big data storage solutions and S3 Select >>

Data transfer

AWS provides a portfolio of data transfer services to provide the right solution for any data migration project. The level of connectivity is a major factor in data migration, and AWS has offerings that can address your hybrid cloud storage, online data transfer, and offline data transfer needs.

Hybrid cloud storage: AWS Storage Gateway is a hybrid cloud storage service that lets you seamlessly connect and extend your on-premises applications to AWS Storage. Customers use Storage Gateway to seamlessly replace tape libraries with cloud storage, provide cloud storage-backed file shares, or create a low-latency cache to access data in AWS for on- premises applications.

Online data transfer: AWS DataSync makes it easy and efficient to transfer hundreds of terabytes and millions of files into Amazon S3, up to 10x faster than open-source tools.
DataSync automatically handles or eliminates many manual tasks, including scripting copy jobs, scheduling and monitoring transfers, validating data, and optimizing network utilization. Additionally, you can use AWS DataSync to copy objects between a bucket on S3on Outposts and a bucket stored in an AWS Region. The AWS Transfer Family provides fully managed, simple, and seamless file transfer to Amazon S3 using SFTP, FTPS, and FTP.Amazon S3 Transfer Acceleration enables fast transfers of files over long distances between your client and your Amazon S3 bucket.

Offline data transfer: The AWS Snow Family is purpose-built for use in edge locations where network capacity is constrained or nonexistent and provides storage and computing.capabilities in harsh environments. The AWS Snowball service uses ruggedized, portable storage and edge computing devices for data collection, processing, and migration.

Customers can ship the physical Snowball device for offline data migration to AWS. AWS Snowmobile is an exabyte-scale data transfer service used to move massive volumes of datato the cloud, including video libraries, image repositories, or even a complete data center migration.

Customers can also work with third-party providers from the AWS Partner Network (APN)to deploy hybrid storage architectures, integrate Amazon S3 into existing applications and workflows, and transfer data to and from the AWS Cloud. Learn more by visiting AWS

cloud data migration services». AWS Storage Gateway ». AWS DataSync », AWS Transfer Family » Amazon S3 Transfer Acceleration »>, AWS Snow Family≫

Performance

Amazon S3 provides industry leading performance for cloud object storage. Amazon S3 supports parallel requests, which means you can scale your S3 performance by the factor of your compute cluster, without making any customizations to your application. Performance scales per prefix. so you can use as many prefixes as you need in parallel to achieve the required throughput.

There are no limits to the number of prefixes. Amazon S3 performance supports at least 3.500 requests per second to add data and 5.500 requests per second to retrieve data. Each S3 prefix can support these request rates, making it simple to increase performance significantly.

To achieve this S3 request rate performance you do not need to randomize object prefixes to achieve faster performance. That means you can use logical or sequential naming patterns in S3 object naming without any performance implications. Refer to the Performance Guidelines for Amazon S3 and Performance Design Patterns for Amazon S3 for the most current information about performance optimization for Amazon S3.

Consistency

Amazon S3 delivers strong read-after-write consistency automatically for all applications. without changes to performance or availability, without sacrificing regional isolation for applications, and at no additional cost. With strong consistency, S3 simplifies the migration of on-premises analytics workloads by removing the need to make changes to applications. and reduces costs by removing the need for extra infrastructure to provide strong consistency.

Any request for S3 storage is strongly consistent. After a successful write of a new object oran overwrite of an existing object, any subsequent read request immediately receives the latest version of the object. S3 also provides strong consistency for list operations, so after awrite, you can immediately perform a listing of the objects in a bucket with any changes reflected.

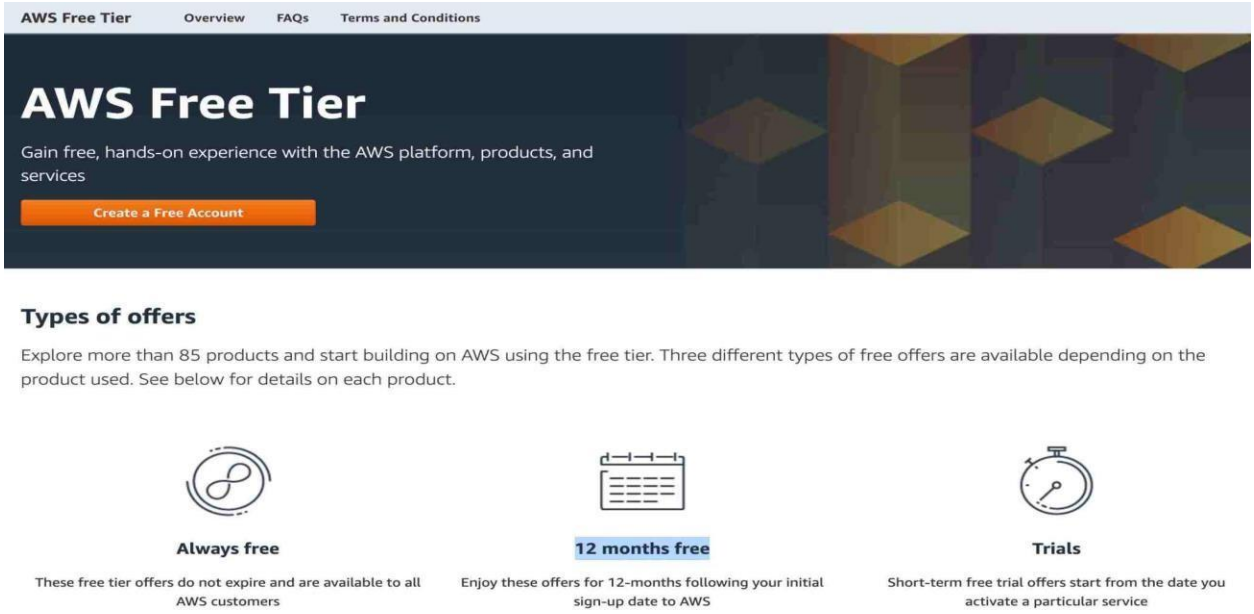
Sign up for a free account

Pay nothing or try for free while learning the fundamentals and building on AWS.

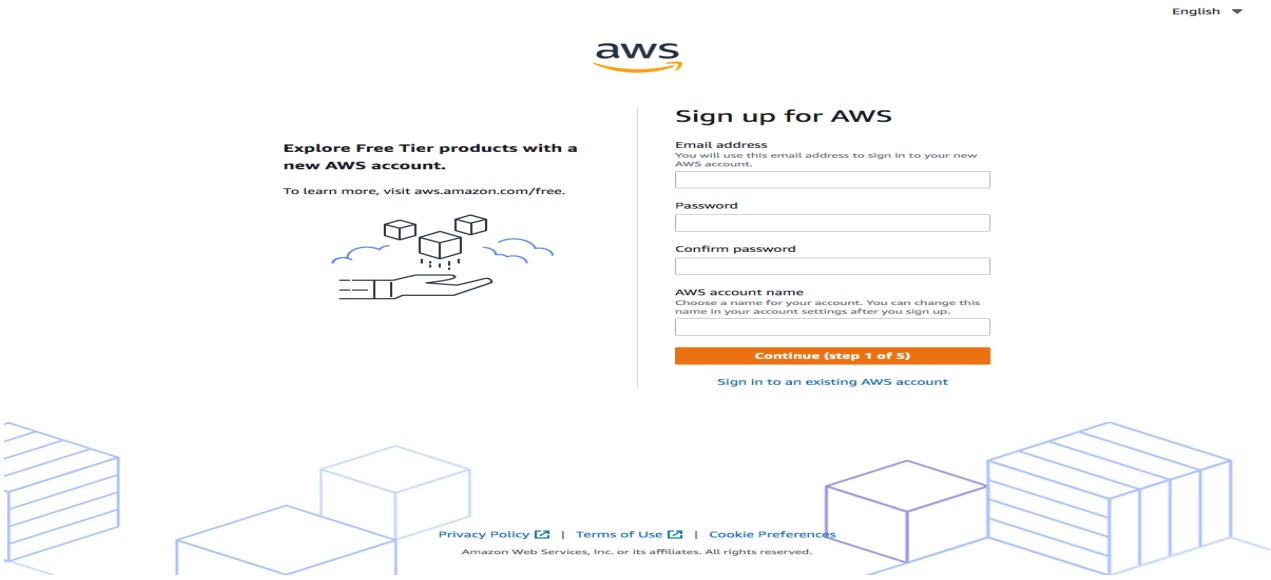
Types Of Offers

Explore more than 100 products and start building on AWS using the Free Tier. Threedifferent types of free offers are availabledepending on the product used. Click iconbelow to explore our offers.

AWS Free Tier:



SIGN UP FOR AWS



EX NO: 10 DATE:	Case study on Google Cloud Platform(GCP)

Google Cloud Platform (GCP), offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, Google Drive, and YouTube. Alongside a set of management tools, it provides a series of modular cloud services including computing, data storage, data analytics andmachine learning. Registration requires a credit card or bank account details.

Google Cloud Platform provides infrastructure as a service, platform as and serverless computing environments.

In April 2008, Google announced App Engine, a platform for developing and hosting web applications in Google-managed data centers, which was the first cloud computing service from the company. The service became generally available in November 2011. Since the announcement of App Engine, Google added multiple cloud services to the platform.

Google Cloud Platform is a part of Google Cloud, which includes the Google Cloud Platform public cloud infrastructure, as well as Google Workspace (G Suite), enterprise versions of Android and Chrome OS, and application programming interfaces (APIs) for machine learning and enterprise mapping services.

Products

Conference presentation on Google Container Engine Kubernetes Google lists over 100 productsunder the Google Cloud brand. Some of the key services are listed below.

Compute

- App Engine Platform as a Service to deploy Java, PHP, Node.js, Python, C#, Net, Rubyand Go applications.
- Compute Engine Infrastructure as a Windows and Linux virtual machines Service to run Microsoft
- Google Kubernetes Engine (GKE) or GKE on-prem offered as part of Anthosplatform - Containers as a Service based on Kubernetes.
- Cloud Functions - Functions as a Service to run event-driven code written inNode.js,Java, Python, or Go.
- Cloud Run - Compute execution environment based on Knative. Offered as Cloud Run(fully managed)[9] or as Cloud Run for Anthos. Currently supports GCP, AWS and VMware management.

Storage & Databases

- Cloud Storage - Object storage with integrated edge caching to store unstructured data.Cloud SQL Database as a Service based on MySQL, PostgreSQL and Microsoft SQL Server.
- Cloud Bigtable - Managed NoSQL database service.
- Cloud Spanner Horizontally scalable, strongly consistent, relational database service [12]
- Cloud Datastore - NoSQL database for web and mobile applications. Persistent Disk -Block

storage for Compute Engine virtual machines.

- Local SSD: High-performance, transient, local block storage.
- Filestore: High-performance file storage for Google Cloud users.
- AlloyDB: Fully managed PostgreSQL database service.

Networking

- VPC - Virtual Private Cloud for managing the software defined network of cloudresources.
- Cloud Load Balancing Software-defined, managed service for load balancing the traffic.
- Cloud Armor - Web application firewall to protect workloads from DDoS attacks.
- Cloud CDN - Content Delivery Network based on Google's globally distributed edgepoints of presence. Cloud Interconnect - Service to connect a data center with GoogleCloud Platform
- Cloud DNS - Managed, authoritative DNS hosting service running on the same infrastructure as Google. Network Service Tiers Option to choose Premium vs Standardnetwork tier for - higher-performing network.

Big Data

- BigQuery - Scalable, managed enterprise data warehouse for analytics.
- Cloud Dataflow - Managed service based on Apache Beam for stream and batch data processing.
- Dataproc - Big data platform for running Apache Hadoop and Apache Spark jobs.
- Cloud Composer - Managed workflow orchestration service built on Apache Airflow Cloud Datalab Tool for data exploration, analysis, visualization and machine learning.This is a fully managed Jupyter Notebook service.
- Cloud Dataprep - Data service based on Trifacta to visually explore, clean, and preparedata for analysis.
- Cloud Pub/Sub - Scalable event ingestion service based on message queues. Cloud Data Studio Business intelligence tool to visualize data through dashboards and reports.

Cloud AI

- Cloud AutoML - Service to train and deploy custom machine learning models. As ofSeptember 2018, the service is in Beta.
- Cloud TPU- Accelerators used by Google to train machine learning models.
- Cloud Machine Learning Engine - Managed service for traming and building machinelearning models based on mainstream frameworks.
- Cloud Talent Solution (formerly Cloud Job Discovery) Service hased on Google's searchand machine learning capabilities for the recruiting ecosystem.
- Dialogflow Enterprise Development environment based on Google's machine learning for building conversational interfaces . Cloud Natural Language Text analysis service based on Google Deep Learning models.
- Cloud Speech-to-Text-Speech to text conversion service based on machine learning.Cloud Text-to-Speech-Text to speech conversion service based on machine learning.

- Cloud Translation API - Service to dynamically translate between thousands of available language pairs.
- Cloud Vision API- Image analysis service based on machine learning. Cloud VideoIntelligence - Video analysis service based on machine learning Management Tools
- Operations suite (formerly Stackdriver) - Monitoring, logging, and diagnostics forapplications on Google Cloud Platform and AWS.
- Cloud Deployment Manager - Tool to deploy Google Cloud Platform resources definedin templates created in YAML, Python or Jinja2
- Cloud Console - Web interface to manage Google Cloud Platform resources.
- Cloud Shell Browser-based shell command-line access to manage Google Cloud Platform resources.
- Cloud Console Mobile App Android and iOS application to manage Google CloudPlatform resources.
- Cloud APIS - APIs to programmatically access Google Cloud Platform resources

Identity & Security:

- Cloud Identity-Single sign-on (SSO) service based on SAML 2.0 and OpenID
- Cloud IAM Identity & Access Management (IAM) service for defining policies based onrole-based access control. Cloud Identity-Aware Proxy Service to control access to cloud applications running on Google Cloud Platform without using a VPN.
- Cloud Data Loss Prevention API - Service to automatically discover, classify, and redact sensitive data
- Security Key Enforcement -Two-step verification service based on a security key. CloudKey Management Service - Cloud-hosted key management service integrated with IAM and audit logging.
- Cloud Resource Manager - Service to manage resources by project, folder, andorganization based on the hierarchy.
- Cloud Security Command Center-Security and data risk platform for data and servicesrunning in Google Cloud Platform
- Cloud Security Scanner - Automated vulnerability scanning service for applicationsdeployed in App Engine.
- Access Transparency - Near real-time audit logs providing visibility to Google CloudPlatform administrators.
- VPC Service Controls - Service to manage security perimeters for sensitive data inGoogle Cloud Platform services.

IOT:

- Cloud loT Core - Secure device connection and management service for Internet of
- Things Edge TPU - Purpose-built ASIC designed to run inference at the edge. As ofSeptember 2018, this product is in private beta. Cloud loT Edge - Brings Al to the edge computing layer.

API Platform

- Maps Platform - APIs for maps, routes, and places based on Google Maps.
- Apigee API Platform Lifecycle management platform to design, secure, deploy,monitor, and scale APIs.
- API Monetization Tool for API providers to create revenue models, reports, payment gateways, and developer portal integrations.
- Apigee Sense - Enables API security by identifying and alerting administrators to suspicious API behaviours. Cloud Endpoints - An NGINX-based proxy to deploy andmanage APIs
- Service Infrastructure A set of foundational services for building Google Cloudproducts

GCP Customers

https://cloud google.com/customers/#/

About PayPal

PayPal has remained at the forefront of the digital payment revolution for more than 20 years. By leveraging technology to make financial services and commerce more convenient, inclusive, and secure, the PayPal platform is empowering more than 377 million consumers and merchants in more than 200 markets to join and thrive in the global economy. Industries: Financial Services & Insurance

Location: United States

PayPal leverages Google Cloud to flawlessly manage surges in financial transactions.

PayPal recently migrated the horizontally scalable applications of its payment platform to Google Cloud a significant shift in strategy to cloud-based transactions to help ensure easyand reliable services for its customers.

Google Cloud results

- Shifted the horizontally scalable applications of its payment platform and transactionworkloads to the cloud
- Completed a month's worth of on-premises work in minutes in the cloud
- Experienced significant cost savings over what it would cost to capacitize for peakvolumes on-premises
- Seamlessly handled the pandemic and holiday season transaction surges in 2020
- PayPal was one of the first businesses to move money over the internet, securely andswiftly. Assisted by cloud computing, it wants to bring a better financial life to everypossible point-and person on the planet.
- With revenues reaching S2 trillion in 2020, the global payments industry is attractinga wide range of companies looking to win customers and expand markets by providing simpler and safer ways to make online transactions. Competing in this fast-changing landscape requires a global reach and high degree of efficiency. PayPal keeps its edge through an unwavering focus on agility and innovation.
- Today, PayPal has over 377 million active accounts in 200 global markets. Its customers transacted more than S936 billion in total payment volume (TPV) in 2020.The company has accomplished this through a strong technology-driven culture, running largely on its own infrastructure and third-party networks. In previous years.total payment volume (TPV) was growing at 25% annually. However, due to a significant uptick in digital commerce and user traffic triggered by the 2020 pandemic, PayPal accelerated its infrastructure strategy to support their customers evolving needs.
- PayPal worked with Deloitte to define and execute a multiyear change in its infrastructure strategy, a bold move for a large-scale company intending to grow its global footprint in the middle of a pandemic. As part of its mission to move to zero data center ownership, PayPal collaborated with Deloitte to drive the exit of its non-strategic data centers. Moving its mission-critical workloads from primarily on- premises infrastructure to the cloud is a key enabler of this mission. This means by the end of 2021, PayPal expects to completely shut down one of its Salt Lake City data centers. While the process of adding capacity to an on-premises infrastructure typically takes months, with Google Cloud. PayPal was able to do three month's worth of work in just hours.
- "With this shift, there was initially some concern among the PayPal team of engineers managing the workloads," says Wes Hummel, Vice President, Site Reliability and Cloud Engineering at PayPal. "But once they saw how easy it was toadd capacity through the highest peak holiday season we have ever had and reduce capacity when it was no longer needed, they were excited about the value of movingto the cloud."
- • Deloitte provided assistance in migrating PayPal's workloads onto Google Cloud byhelping to define PayPal's overall migration strategy and providing additional engineering support. Manual effort was minimized through the design, development, and implementation of automation scripts, reducing PayPal's migration timeline.

- "Our multiyear journey with PayPal and Google Cloud originated from a shared interest in democratizing the future of financial services through the use of cloud technologies. Over the past several years, the Deloitte team has been proud to providestrategy and engineering talent to complement and augment the Google Cloud and PayPal teams. In 2020, our ambitious strategy paid off, as Google Cloud provided thenecessary scale for PayPal products when consumers and merchants needed them the most," said Dan Grayson, Cloud Engineering Principal at Deloitte.
- Handles transaction surge quickly through a shift to the cloud Previously, the PayPal team had ported its internally developed containerization platform to the cloud, which added performance to its microservices, without the needfor new learning. Now the company needed to act quickly to manage the impending bursts in holiday transactions

  The Salt Lake City region is Google Cloud's sixth region in the United States and one of the 23 regions Google Cloud operated in 2020. It was chosen for proximity tomany knowledge-intensive industries like healthcare, information technology, and financial services. Enterprises like PayPal that choose to run their platforms throughGoogle Cloud regions can gain access to lower latency, enhanced scalability, and better end-user customer experiences.

- Customers can also take advantage of hybrid cloud solutions, or if they wish, splitworkloads across Western U.S. facilities in Los Angeles and Oregon, ensuring reliability at high speeds.
- Speed matters for both operational efficiency and customer delight. Even a few hundred unnecessary milliseconds in processing a transaction can result in customerdissatisfaction. When the majority of operational activity is happening on Google Cloud's secure, high-performing proprietary network, latency is minimized.
- After the move to Google Cloud, the PayPal team saw immediate improvements. During peak traffic times, like Black Friday and Cyber Monday, PayPal processed 1,000 payments per second. Google Cloud helped manage these high-traffic events bystaggering workloads while also scaling up or down to manage off-peak times effectively.

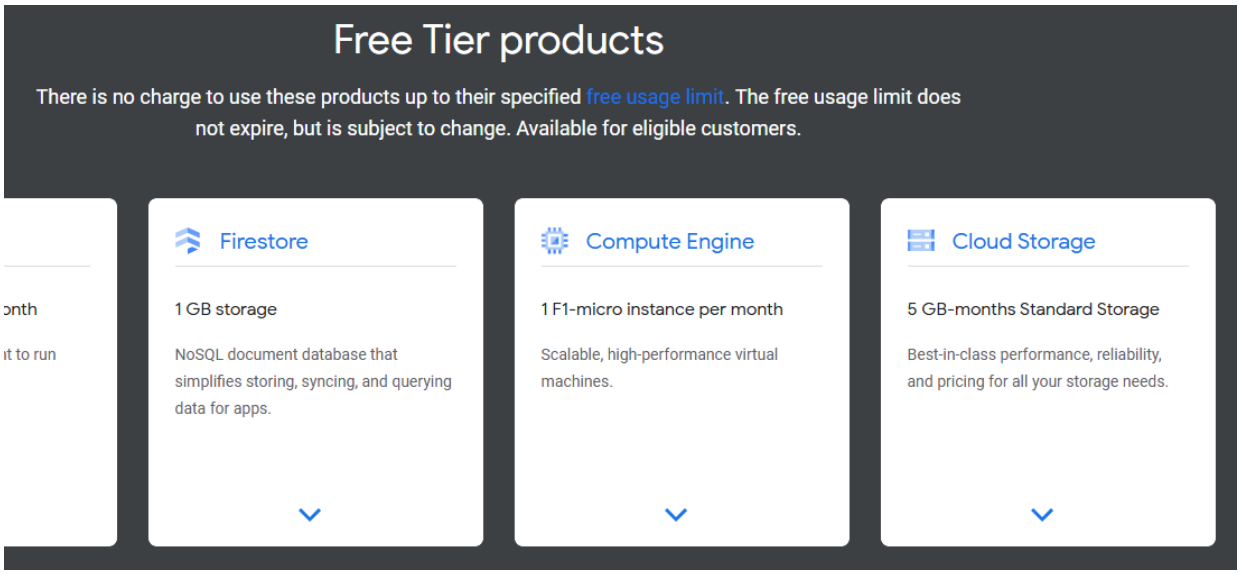
Flexibly adjusts to infrastructure requirement ebbs and flows

- When PayPal was handling transactions entirely on-premises and in its own data centers, it had to build onto its infrastructure to prepare for peak transaction load and maintain it at all times. With its payment platform now on Google Cloud, PayPal canquickly adjust to handle changes in traffic without the cost of continually maintainingthat infrastructure. In fact, the company experienced significant savings during the 2020 holiday season by utilizing the Google Cloud.
- "We can only develop fast, build fast, and deploy fast if we have infrastructure that's as nimble as we are," says PayPal's CTO, Sri Shivananda. "Google Cloud allows us toinnovate and serve our customers at the speed we require. We can bring as much data as we want to the platform and manage it in a secure, cost-effective way."
- With these changes, PayPal continues to democratize ecommerce as it scales, affording big-retail services to smaller, entrepreneurial retailers. By serving thismarket, PayPal can give consumers and retailers everywhere more choice, moreopportunity, and more tools with which to improve their lives and businesses.

Envisions a Google Cloud future

Approximately 20% of PayPal's core infrastructure is now on Google Cloud, withmore to come. PayPal, in collaboration with Google Cloud and Deloitte, is continuing to implement a multiyear cloud transformation that will shift mission- critical workloads to the cloud, tap into cutting-edge advanced analyticscapabilities, and increase its cost savings and resources.

"Democratizing financial services is our mission and working with the Google Cloud team is enabling us to focus on providing the best products, capabilities andservices to our customers instead of worrying about managing our infrastructure,"says Sri Shivananda, EVP, Chief Technology Officer at PayPal. "As a part of our strategic partnership, we're working side by side with Google Cloud to scale and secure our infrastructure for the future."



EX NO: 11 DATE:	**Procedure To Launch Virtual Machine Using Trystack**

AIM:

To find a procedure to launch virtual machine using trystack (Online Openstack Demo Version).

PROCEDURE:

Step 1: Download & install Virtual boxhttps://www.virtualbox.org/

Step 2: Download Ubuntu 18.04 ISO https://ubuntu.com/download/desktop

Step 3: Create Ubuntu VM in virtual box

4GB RAM, 2 vCPUs, Hard disk capacity of minimum 10 GB

Step 4: Update and Upgrade the SystemOpen terminal in Ubuntu

> sudo su

> sudo apt-get update –y && apt-get upgrade –y

> sudo reboot

> sudo useradd –s /bin/bash –d /opt/stack –m stack

Step 5: Create Stack user and assign sudo privilege

>echo “stack ALL=(ALL) NOPASSWD: ALL” | sudo tee /etc/sudoers.d/stack

Step 6: Install git and download DevStack

>sudo su – stack

If git not installed, install >sudo apt-get install git –y

sudo git clone https://github.com/openstack-dev/devstack.git -b stable/pike devstack/

Step 7: Create devstack configuration file

>cd devstack

Step 8: Then create a local.conf configuration file

>sudo nano local.conf

>cat . local.conf << EOF

[[local|localrc]] ADMIN_PASSWORD=secret

DATABASE_PASSWORD=\$ADMIN_PASSWORD

RABBIT_PASSWORD=\$ADMIN_PASSWORD

SERVICE_PASSWORD=\$ADMIN_PASSWORD HOST_IP=10.0.2.15

RECLONE=yesEOF

>sudo apt-get install net-tools

Step 9: Install OpenStack with Devstack

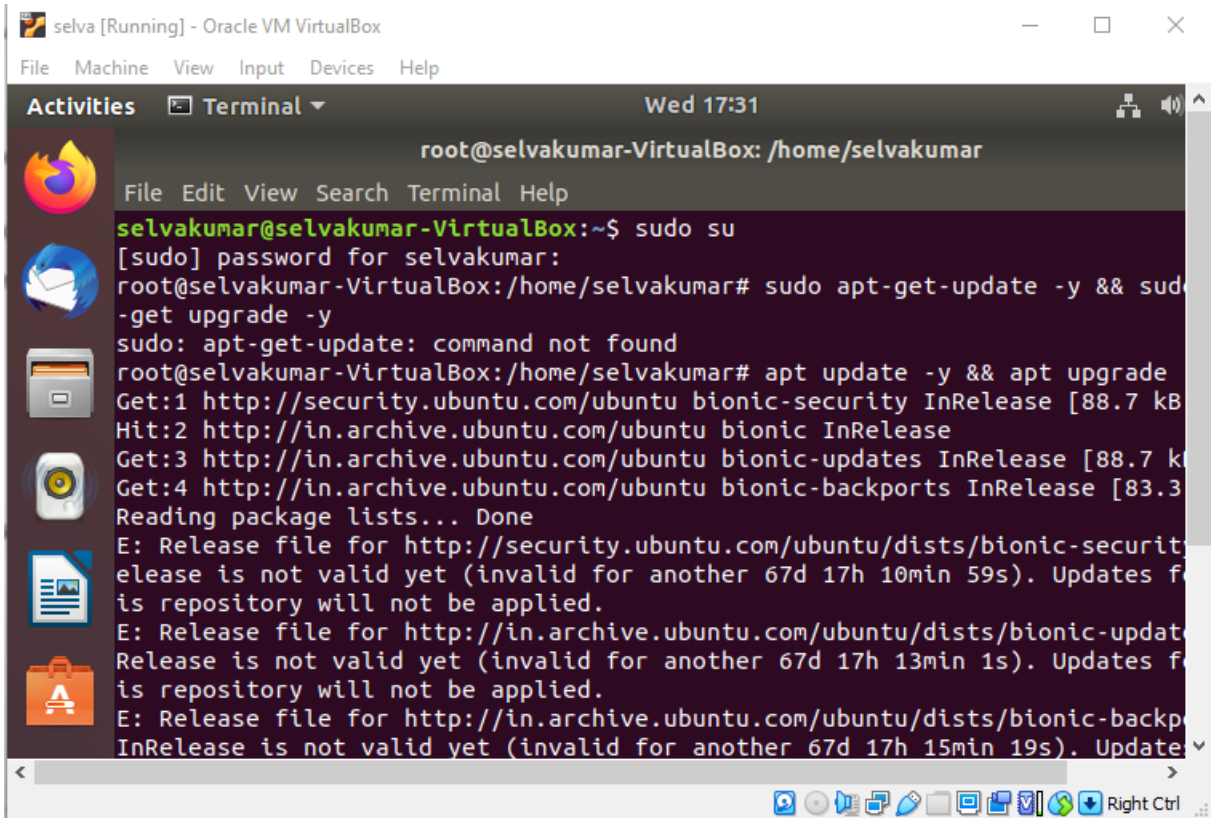
>./stack.sh

> FORCE=yes ./stack.sh

Step 10: Accessing OpenStack on a web browser

To access OpenStack via a web browser browse your Ubuntu's IP address as shown https://server-ip/dashboard

OUTPUT:



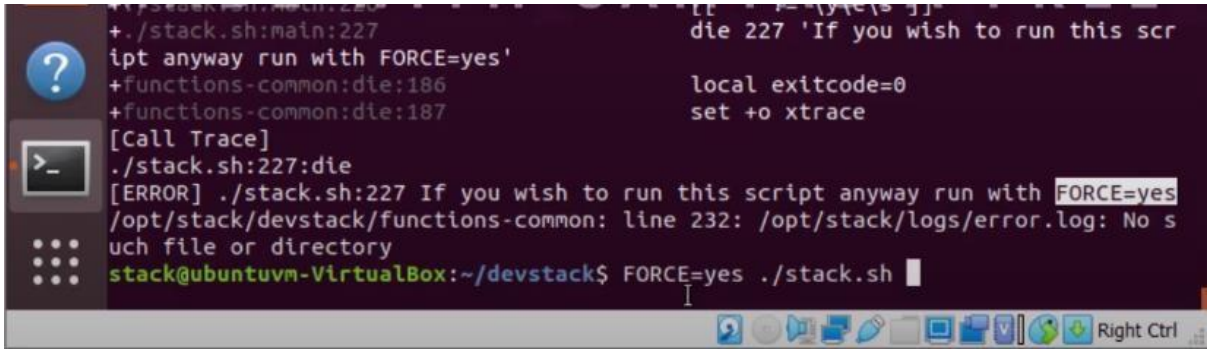
```
selvakumar@selvakumar-VirtualBox:~$ ^C
selvakumar@selvakumar-VirtualBox:~$ sudo adduser -s /bin/bash -d /opt/stack -m
stack
[sudo] password for selvakumar:
Option s is ambiguous (shell, system)
Option d is ambiguous (debug, disabled-login, disabled-password)
Unknown option: m
adduser [--home DIR] [--shell SHELL] [--no-create-home] [--uid ID]
[--firstuid ID] [--lastuid ID] [--gecos GECOS] [--ingroup GROUP | --gid ID]
[--disabled-password] [--disabled-login] [--add_extra_groups]
[--encrypt-home] USER
   Add a normal user

adduser --system [--home DIR] [--shell SHELL] [--no-create-home] [--uid ID]
[--gecos GECOS] [--group | --ingroup GROUP | --gid ID] [--disabled-password]
[--disabled-login] [--add_extra_groups] USER
   Add a system user

adduser --group [--gid ID] GROUP
addgroup [--gid ID] GROUP
   Add a user group

addgroup --system [--gid ID] GROUP
   Add a system group

adduser USER GROUP
   Add an existing user to an existing group

general options:
```

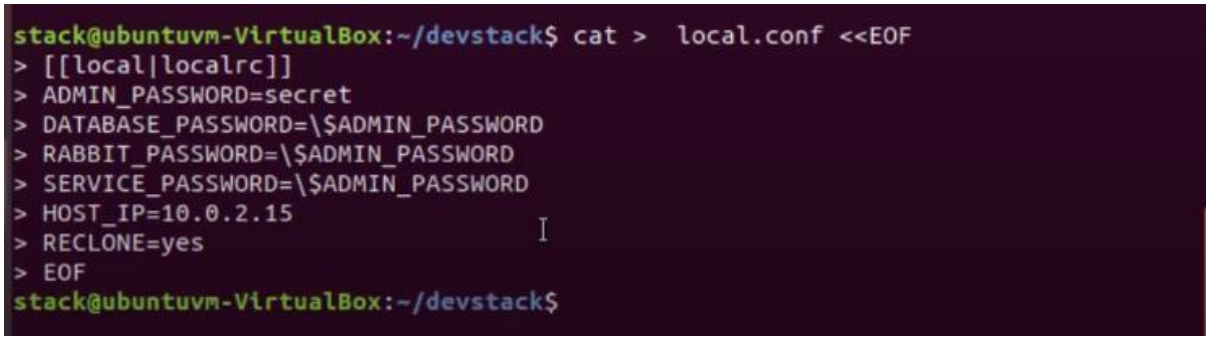
```
                    NAME_REGEX[_SYSTEM] configuration variable
  --extrausers      uses extra users as the database
  --help | -h       usage message
  --version | -v    version number and copyright
  --conf | -c FILE  use FILE as configuration file

selvakumar@selvakumar-VirtualBox:~$ echo "stack ALL=(ALL) NOPASSWD = ALL" | sud
o tee/etc/sudoers.d/stack
sudo: tee/etc/sudoers.d/stack: command not found
selvakumar@selvakumar-VirtualBox:~$ echo "stack ALL=(ALL) NOPASSWD = ALL" | sud
o tee/etc/sudoers.d/stack
sudo: tee/etc/sudoers.d/stack: command not found
selvakumar@selvakumar-VirtualBox:~$ sudo apt get install git-y
E: Invalid operation get
selvakumar@selvakumar-VirtualBox:~$ sudo apt get-install git -y
E: Invalid operation get-install
selvakumar@selvakumar-VirtualBox:~$ sudo -su stack
sudo: unknown user: stack
sudo: unable to initialize policy plugin
selvakumar@selvakumar-VirtualBox:~$ sudo apt -get install git -y
E: Command line option 'g' [from -get] is not understood in combination with th
e other options.
selvakumar@selvakumar-VirtualBox:~$ sudo apt-get install git-y
Reading package lists... Done
Building dependency tree
Reading state information... Done
E: Unable to locate package git-y
selvakumar@selvakumar-VirtualBox:~$ S
```

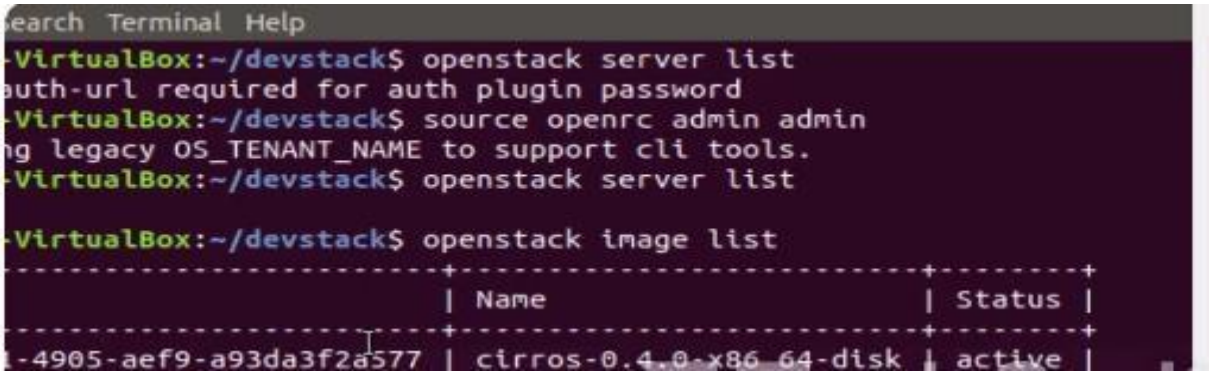
```
selvakumar@selvakumar-VirtualBox:~$ sudo apt install git
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  git-man liberror-perl
Suggested packages:
  git-daemon-run | git-daemon-sysvinit git-doc git-el git-email git-gui gitk
  gitweb git-cvs git-mediawiki git-svn
The following NEW packages will be installed:
  git git-man liberror-perl
0 upgraded, 3 newly installed, 0 to remove and 52 not upgraded.
Need to get 4,817 kB of archives.
After this operation, 34.3 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://in.archive.ubuntu.com/ubuntu bionic/main amd64 liberror-perl all 0
.17025-1 [22.8 kB]
Get:2 http://in.archive.ubuntu.com/ubuntu bionic-updates/main amd64 git-man all
 1:2.17.1-1ubuntu0.18 [804 kB]
Get:3 http://in.archive.ubuntu.com/ubuntu bionic-updates/main amd64 git amd64 1
:2.17.1-1ubuntu0.18 [3,990 kB]
Fetched 4,817 kB in 4s (1,231 kB/s)
Selecting previously unselected package liberror-perl.
(Reading database ... 167743 files and directories currently installed.)
```

```
stack@ubuntuvm-VirtualBox:~$ git clone https://github.com/openstack-dev/devstac
k.git -b stable/pike devstack/
Cloning into 'devstack'...
remote: Enumerating objects: 264, done.
remote: Counting objects: 100% (264/264), done.
remote: Compressing objects: 100% (193/193), done.
remote: Total 45869 (delta 148), reused 150 (delta 69), pack-reused 45605
Receiving objects: 100% (45869/45869), 14.72 MiB | 90.00 KiB/s, done.
Resolving deltas: 100% (31915/31915), done.
stack@ubuntuvm-VirtualBox:~$ cd devstack/
stack@ubuntuvm-VirtualBox:~/devstack$
```

```
selvakumar@selvakumar-VirtualBox:~$ sudo apt install net-tools
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  net-tools
0 upgraded, 1 newly installed, 0 to remove and 52 not upgraded.
Need to get 194 kB of archives.
After this operation, 803 kB of additional disk space will be used.
Get:1 http://in.archive.ubuntu.com/ubuntu bionic/main amd64 net-tools amd64 1.6
0+git20161116.90da8a0-1ubuntu1 [194 kB]
Fetched 194 kB in 2s (95.4 kB/s)
Selecting previously unselected package net-tools.
(Reading database ... 168650 files and directories currently installed.)
Preparing to unpack .../net-tools_1.60+git20161116.90da8a0-1ubuntu1_amd64.deb .
..
Unpacking net-tools (1.60+git20161116.90da8a0-1ubuntu1) ...
Setting up net-tools (1.60+git20161116.90da8a0-1ubuntu1) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
selvakumar@selvakumar-VirtualBox:~$
```







RESULT:

Hence the procedure to launch Virtual Machine Using Trystack is executed successfully.