LAB 1

PERFORM FOLLOWING OPERATION IN HIVE

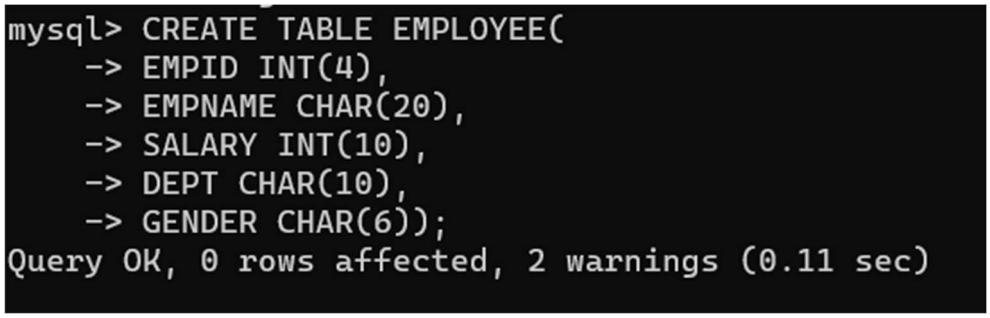
a) **Table creation and population**
b) **Creating and switching databases**
c) **Querying tables**
d) **Modifying table information: Altering and dropping a table**

Solutions:

a) Table creation and population

Query:-

```
>>CREATE TABLE EMPLOYEE (
        EMPID  INT(4),
        EMPNAME  CHAR(20),
        SALARY  INT(10),
        DEPT  CHAR(10),
        GENDER  CHAR(6));
```




```
>>INSERT INTO EMPLOYEE
   VALUES(1,'ANUJ',20000,'MCA','M');

>>INSERT INTO EMPLOYEE
   VALUES(2,'AJAY',30000,'MBA','M');

>>INSERT INTO EMPLOYEE
```

VALUES(3,’AMRITA’,35000,’MCA’,’F’);

```
mysql> INSERT INTO EMPLOYEE
    -> VALUES(1,'ANUJ',20000,'MCA','M');
Query OK, 1 row affected (0.02 sec)

mysql> INSERT INTO EMPLOYEE
    -> VALUES(2,'AJAY',30000,'MBA','M');
Query OK, 1 row affected (0.01 sec)

mysql> INSERT INTO EMPLOYEE
    -> VALUES(3,'AMRITA',35000,'MCA','F');
Query OK, 1 row affected (0.01 sec)

mysql>
```

>>SELECT * FROM EMPLOYEE;

```
mysql> SELECT * FROM EMPLOYEE;
+-------+---------+--------+------+--------+
| EMPID | EMPNAME | SALARY | DEPT | GENDER |
+-------+---------+--------+------+--------+
|     1 | ANUJ    |  20000 | MCA  | M      |
|     2 | AJAY    |  30000 | MBA  | M      |
|     3 | AMRITA  |  35000 | MCA  | F      |
+-------+---------+--------+------+--------+
3 rows in set (0.02 sec)

mysql>
```

b) Creating and switching databases

Query: -

>>CREATE DATABASE COMPANY;

>>USE COMPANY;

```
mysql> CREATE DATABASE COMPANY;
Query OK, 1 row affected (0.02 sec)

mysql> USE COMPANY;
Database changed
mysql>
```

c) Querying tables

```
mysql> SELECT EMPNAME,SALARY FROM EMPLOYEE
    -> WHERE DEPT='MCA';
+---------+--------+
| EMPNAME | SALARY |
+---------+--------+
| ANUJ    |  20000 |
| AMRITA  |  35000 |
+---------+--------+
2 rows in set (0.02 sec)

mysql> SELECT EMPID,EMPNAME,GENDER FROM EMPLOYEE
    -> ;
+-------+---------+--------+
| EMPID | EMPNAME | GENDER |
+-------+---------+--------+
|     1 | ANUJ    | M      |
|     2 | AJAY    | M      |
|     3 | AMRITA  | F      |
+-------+---------+--------+
3 rows in set (0.00 sec)

mysql>
```

d) Modifying table information: Altering and dropping a table

Altering: -

```
mysql> ALTER TABLE EMPLOYEE
    -> ADD EMAIL VARCHAR(50);
Query OK, 0 rows affected (0.06 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> DESC EMPLOYEE;
+---------+-------------+------+-----+---------+-------+
| Field   | Type        | Null | Key | Default | Extra |
+---------+-------------+------+-----+---------+-------+
| EMPID   | int         | YES  |     | NULL    |       |
| EMPNAME | char(20)    | YES  |     | NULL    |       |
| SALARY  | int         | YES  |     | NULL    |       |
| DEPT    | char(10)    | YES  |     | NULL    |       |
| GENDER  | char(6)     | YES  |     | NULL    |       |
| EMAIL   | varchar(50) | YES  |     | NULL    |       |
+---------+-------------+------+-----+---------+-------+
6 rows in set (0.05 sec)
```

DROP: -

```
mysql> DROP TABLE EMPLOYEE;
Query OK, 0 rows affected (0.04 sec)
```

LAB 2

PERFORM FOLLOWING OPERATION IN HIVE

a) **Mathematical & Aggregate function**
b) **Operators- arithmetic , relational ,logical**
c) **Use of order by**

Solutions: -

a) **Mathematical Operations**

```
mysql> SELECT CEILING(1.25);
+---------------+
| CEILING(1.25) |
+---------------+
|             2 |
+---------------+
1 row in set (0.01 sec)

mysql> SELECT ROUND(4.782);
+--------------+
| ROUND(4.782) |
+--------------+
|            5 |
+--------------+
1 row in set (0.01 sec)

mysql> SELECT FLOOR(3.778);
+--------------+
| FLOOR(3.778) |
+--------------+
|            3 |
+--------------+
1 row in set (0.01 sec)

mysql> SELECT POW(2,4);
+----------+
| POW(2,4) |
+----------+
|       16 |
+----------+
1 row in set (0.01 sec)
```

Aggregate: -

```
mysql> SELECT COUNT(PNAME) FROM MARKET;
+--------------+
| COUNT(PNAME) |
+--------------+
|            4 |
+--------------+
1 row in set (0.00 sec)

mysql> SELECT SUM(PRICE) FROM MARKET;
+------------+
| SUM(PRICE) |
+------------+
|      16500 |
+------------+
1 row in set (0.00 sec)

mysql> SELECT MAX(DISCOUNT) FROM MARKET;
+---------------+
| MAX(DISCOUNT) |
+---------------+
|            40 |
+---------------+
1 row in set (0.00 sec)

mysql> SELECT MIN(QUANT) FROM MARKET;
+------------+
| MIN(QUANT) |
+------------+
|          4 |
+------------+
1 row in set (0.00 sec)

mysql> SELECT AVG(PRICE) FROM MARKET;
+------------+
| AVG(PRICE) |
+------------+
|  4125.0000 |
+------------+
1 row in set (0.00 sec)
```

b) Operators- arithmetic , relational ,logical

Arithmetic Operators: -

Like: - ‘+’ , ‘-’ , ‘x’ , ‘/’ , ‘%’ etc.

```
mysql> SELECT PNAME,PRICE*2 FROM MARKET
    -> WHERE PNAME='MOUSE';
+-------+---------+
| PNAME | PRICE*2 |
+-------+---------+
| MOUSE |    3000 |
+-------+---------+
1 row in set (0.01 sec)

mysql> SELECT PNAME,QUANT+5 AS 'NEW QUANT' FROM MARKET
    -> WHERE PID=4;
+-----------+-----------+
| PNAME     | NEW QUANT |
+-----------+-----------+
| MOUSE PAD |        25 |
+-----------+-----------+
1 row in set (0.00 sec)

mysql> SELECT PNAME,DISCOUNT-10 AS 'NEW DISCOUNT' FROM MARKET
    -> WHERE PID=3;
+---------+--------------+
| PNAME   | NEW DISCOUNT |
+---------+--------------+
| MONITOR |           30 |
+---------+--------------+
1 row in set (0.01 sec)
```

Relational Operators: -

Like: - ‘=’ , ’>’ , ’<’ , ‘>=’ , ‘<=’ , ‘< >’

```
mysql> SELECT PNAME,PRICE FROM MARKET
    -> WHERE PID=2;
+-------+-------+
| PNAME | PRICE |
+-------+-------+
| MOUSE |  1500 |
+-------+-------+
1 row in set (0.00 sec)

mysql> SELECT PNAME,DISCOUNT FROM MARKET
    -> WHERE PRICE>3000;
+---------+----------+
| PNAME   | DISCOUNT |
+---------+----------+
| MONITOR |       40 |
+---------+----------+
1 row in set (0.01 sec)

mysql> SELECT PNAME,QUANT FROM MARKET
    -> WHERE DISCOUNT>=20;
+---------+-------+
| PNAME   | QUANT |
+---------+-------+
| MOUSE   |    10 |
| MONITOR |     4 |
+---------+-------+
2 rows in set (0.00 sec)

mysql> SELECT PNAME,PRICE FROM MARKET
    -> WHERE QUANT<=15;
+----------+-------+
| PNAME    | PRICE |
+----------+-------+
| KEYBOARD |  2000 |
| MOUSE    |  1500 |
| MONITOR  | 12000 |
+----------+-------+
3 rows in set (0.00 sec)
```

Logical Operators: -

Like: - ‘OR’ , ‘AND’ , ‘IN’ , ‘OR’ , ‘NOT’ , ‘LIKE’ , ‘ANY’ etc.

```
mysql> SELECT * FROM MARKET
    -> WHERE PID=4 AND PNAME='CPU';
+------+-------+-------+----------+-------+
| PID  | PNAME | PRICE | DISCOUNT | QUANT |
+------+-------+-------+----------+-------+
|    4 | CPU   | 35000 |       25 |     6 |
+------+-------+-------+----------+-------+
1 row in set (0.00 sec)

mysql> SELECT * FROM MARKET
    -> WHERE PNAME LIKE 'M%';
+------+-----------+-------+----------+-------+
| PID  | PNAME     | PRICE | DISCOUNT | QUANT |
+------+-----------+-------+----------+-------+
|    2 | MOUSE     |  1500 |       20 |    10 |
|    3 | MONITOR   | 12000 |       40 |     4 |
|    4 | MOUSE PAD |  1000 |        5 |    20 |
+------+-----------+-------+----------+-------+
3 rows in set (0.01 sec)

mysql> SELECT * FROM MARKET
    -> WHERE PNAME NOT LIKE 'M%';
+------+----------+-------+----------+-------+
| PID  | PNAME    | PRICE | DISCOUNT | QUANT |
+------+----------+-------+----------+-------+
|    1 | KEYBOARD |  2000 |       10 |     5 |
|    4 | CPU      | 35000 |       25 |     6 |
|    5 | SSD      |  6000 |       12 |    16 |
+------+----------+-------+----------+-------+
3 rows in set (0.01 sec)

mysql> SELECT * FROM MARKET
    -> WHERE PNAME='SSD' OR PID=4;
+------+-----------+-------+----------+-------+
| PID  | PNAME     | PRICE | DISCOUNT | QUANT |
+------+-----------+-------+----------+-------+
|    4 | MOUSE PAD |  1000 |        5 |    20 |
|    4 | CPU       | 35000 |       25 |     6 |
|    5 | SSD       |  6000 |       12 |    16 |
+------+-----------+-------+----------+-------+
3 rows in set (0.00 sec)
```

c) Order By

```
mysql> SELECT * FROM MARKET
    -> ORDER BY PRICE;
+------+-----------+-------+----------+-------+
| PID  | PNAME     | PRICE | DISCOUNT | QUANT |
+------+-----------+-------+----------+-------+
|    4 | MOUSE PAD |  1000 |        5 |    20 |
|    2 | MOUSE     |  1500 |       20 |    10 |
|    1 | KEYBOARD  |  2000 |       10 |     5 |
|    5 | SSD       |  6000 |       12 |    16 |
|    3 | MONITOR   | 12000 |       40 |     4 |
|    4 | CPU       | 35000 |       25 |     6 |
+------+-----------+-------+----------+-------+
6 rows in set (0.00 sec)

mysql> SELECT * FROM MARKET
    -> ORDER BY DISCOUNT DESC;
+------+-----------+-------+----------+-------+
| PID  | PNAME     | PRICE | DISCOUNT | QUANT |
+------+-----------+-------+----------+-------+
|    3 | MONITOR   | 12000 |       40 |     4 |
|    4 | CPU       | 35000 |       25 |     6 |
|    2 | MOUSE     |  1500 |       20 |    10 |
|    5 | SSD       |  6000 |       12 |    16 |
|    1 | KEYBOARD  |  2000 |       10 |     5 |
|    4 | MOUSE PAD |  1000 |        5 |    20 |
+------+-----------+-------+----------+-------+
6 rows in set (0.00 sec)
```

LAB 3

USE OF GROUP BY & HAVING CLAUSE

a) **Find the sum of salaries Dept. wise from employee table.**
b) **Find the sum of salaries based on Department having sum>=75000.**

Solutions: -

```
mysql> SELECT * FROM COLLEGE;
+------+-------+------+--------+
| ID   | NAME  | DEPT | SALARY |
+------+-------+------+--------+
|    1 | ABHAY | MCA  |  30000 |
|    2 | AMIT  | MBA  |  20000 |
|    3 | AMIT  | ENGG |  25000 |
|    4 | RAHUL | MCA  |  50000 |
|    5 | ROHAN | MCA  |  40000 |
|    6 | ROHIT | MBA  |  30000 |
+------+-------+------+--------+
6 rows in set (0.00 sec)
```

1. Find the sum of salaries Dept. wise from employee table.

```
mysql> SELECT DEPT,SUM(SALARY) FROM COLLEGE
    -> GROUP BY DEPT;
+------+-------------+
| DEPT | SUM(SALARY) |
+------+-------------+
| MCA  |      120000 |
| MBA  |       50000 |
| ENGG |       25000 |
+------+-------------+
3 rows in set (0.00 sec)
```

2. Find the sum of salaries based on Department having sum>=75000..

```
mysql> SELECT DEPT,SUM(SALARY) FROM COLLEGE
    -> GROUP BY DEPT
    -> HAVING SUM(SALARY)>=75000;
+------+-------------+
| DEPT | SUM(SALARY) |
+------+-------------+
| MCA  |      120000 |
+------+-------------+
1 row in set (0.01 sec)
```

LAB 4

CREATE VIEW OF A TABLE

Solution: -

```
mysql> CREATE VIEW EMPLY AS
    -> SELECT NAME,DEPT
    -> FROM COLLEGE;
Query OK, 0 rows affected (0.02 sec)

mysql> SELECT * FROM EMPLY;
+-------+------+
| NAME  | DEPT |
+-------+------+
| ABHAY | MCA  |
| AMIT  | MBA  |
| AMIT  | ENGG |
| RAHUL | MCA  |
| ROHAN | MCA  |
| ROHIT | MBA  |
+-------+------+
6 rows in set (0.01 sec)
```

LAB 5

UNDERSTANDING THE CONCEPT OF INDEX ON A TABLE

Indexes are used to retrieve data from the database more quickly than otherwise. The users cannot see the indexes, they are just used to speed up searches/queries.

```
mysql> CREATE INDEX STU_INDEX
    -> ON STUDENT(NAME,DEPT);
Query OK, 0 rows affected (0.07 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> SELECT * FROM STUDENT;
+--------+------+
| NAME   | DEPT |
+--------+------+
| ANIKET | ENGG |
| ANUJ   | MBA  |
| ROHAN  | BCA  |
| ROHIT  | MCA  |
+--------+------+
4 rows in set (0.00 sec)
```

LAB 6

WRITE DOWN THE STEPS TO INSTALL A VIRTUAL MACHINE TO SET UP HADOOP ENVIRONMENT

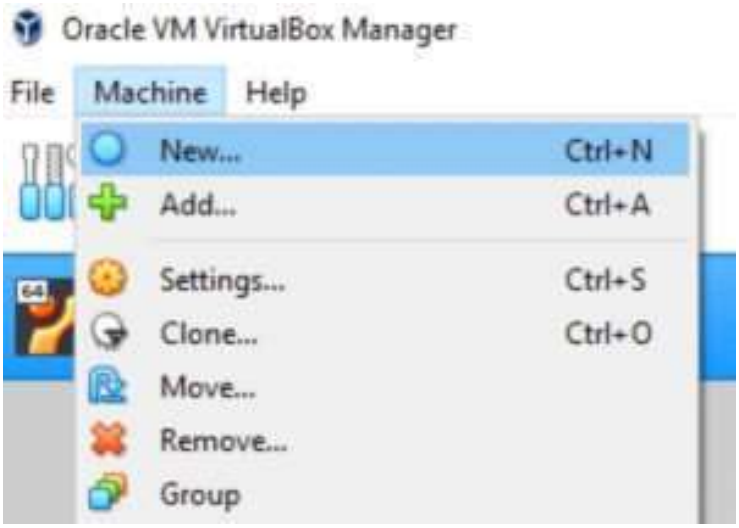
Prerequisites

Apache recommends that a test cluster have the following minimum specifications:

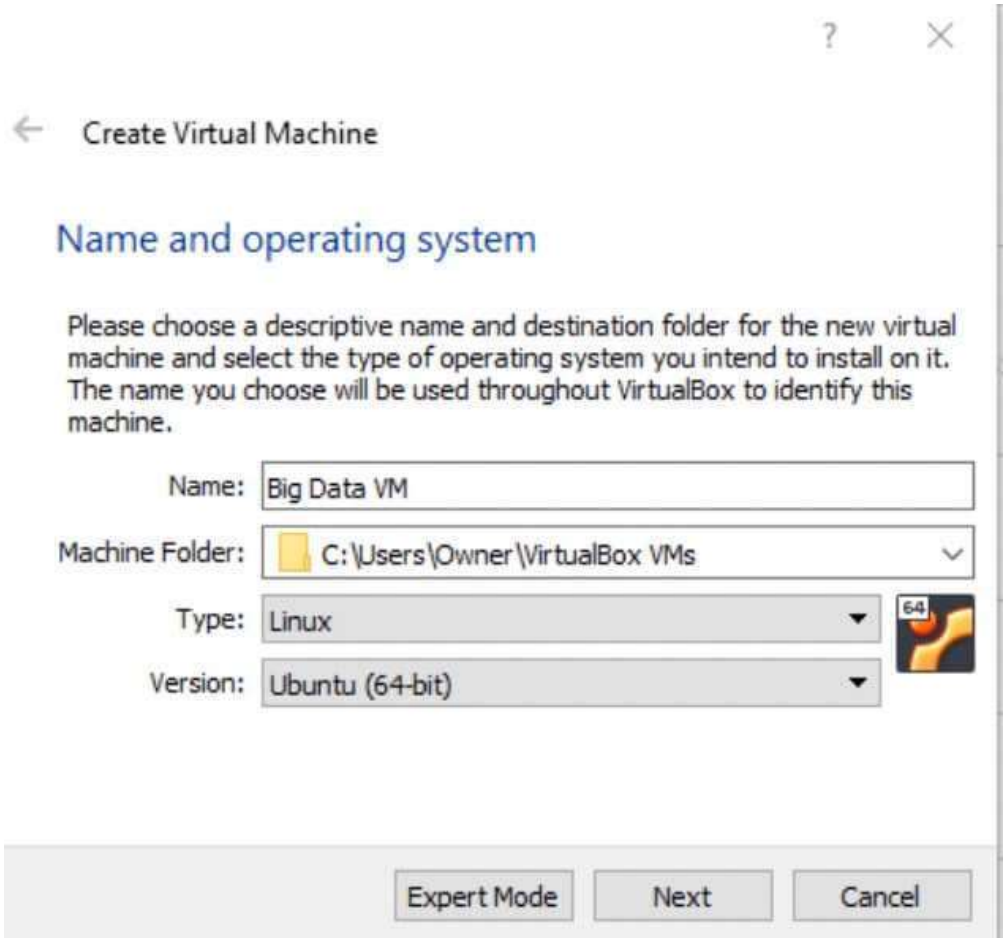
- 2 CPU Cores
- 8 GB RAM
- 30 GB Hard Drive Space

Setting up the Virtual Machine

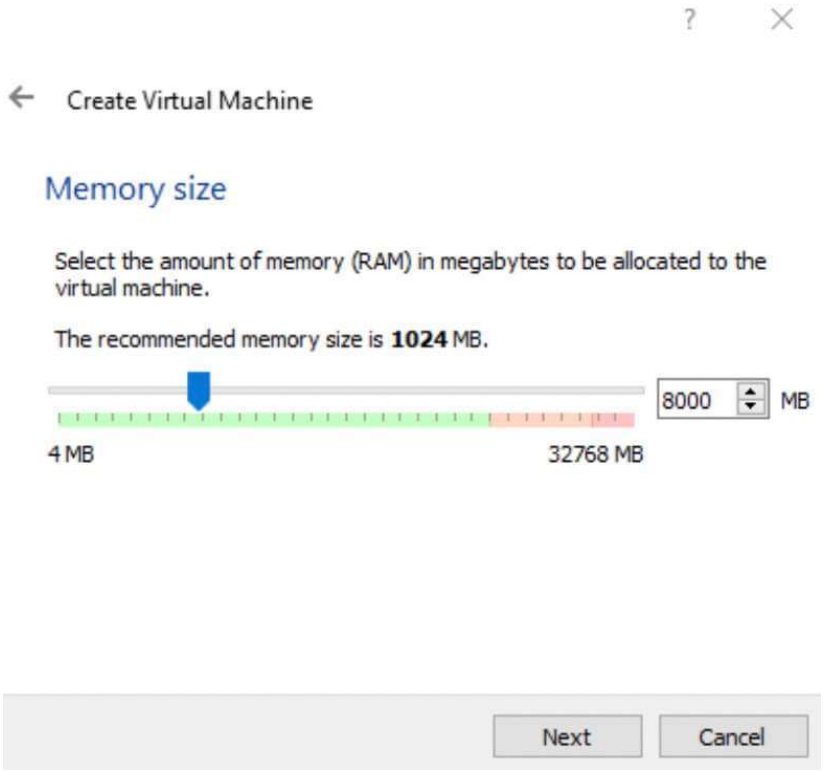
- Download and install the Oracle Virtual Box (Virtual Machine host)
- Download the Linux VM image. There is no need to install it yet, just have it downloaded. Take note of the download location of the iso file, as you will need it in a later step for installation.
- This tutorial will be using Ubuntu 18.04.2 LTS. You may choose to use another Linux platform such as RedHat, however, the commands and screenshots used in this tutorial will be relevant to the Ubuntu platform.
- The Ubuntu iso can be found here.
- Now, open up the Oracle VM VirtualBox Manager and select Machine [wp-svg-icons icon=”arrow-right-2″ wrap=”i”] New.



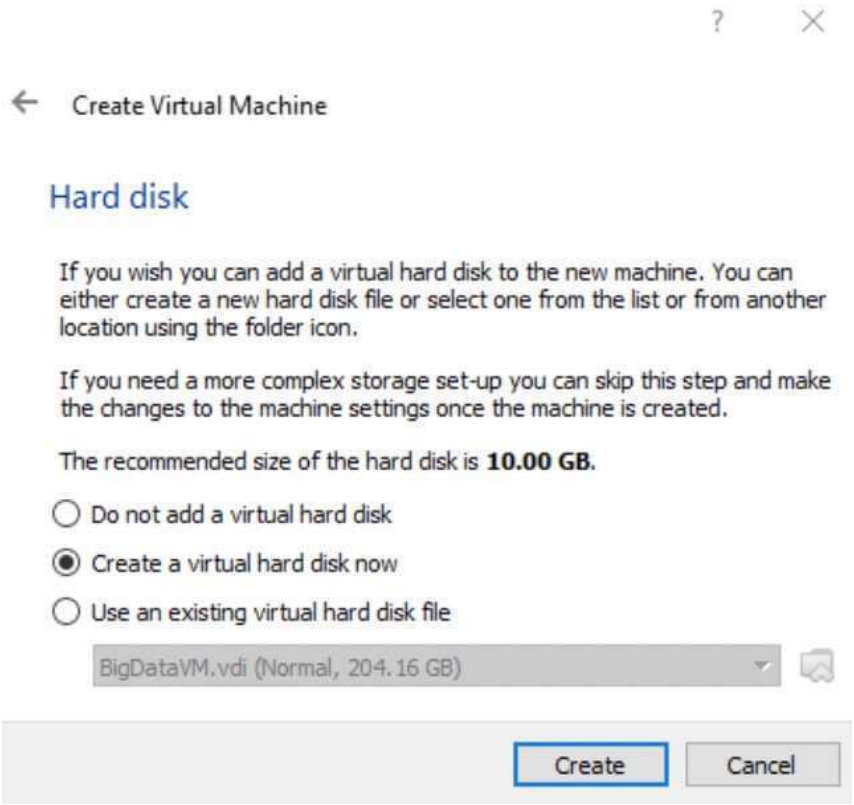
- Choose a Name and Location for your Virtual Machine. Then select the Type as ‘Linux’ and the version as Ubuntu (64-bit). Select ‘Next’ to go to the next dialogue.



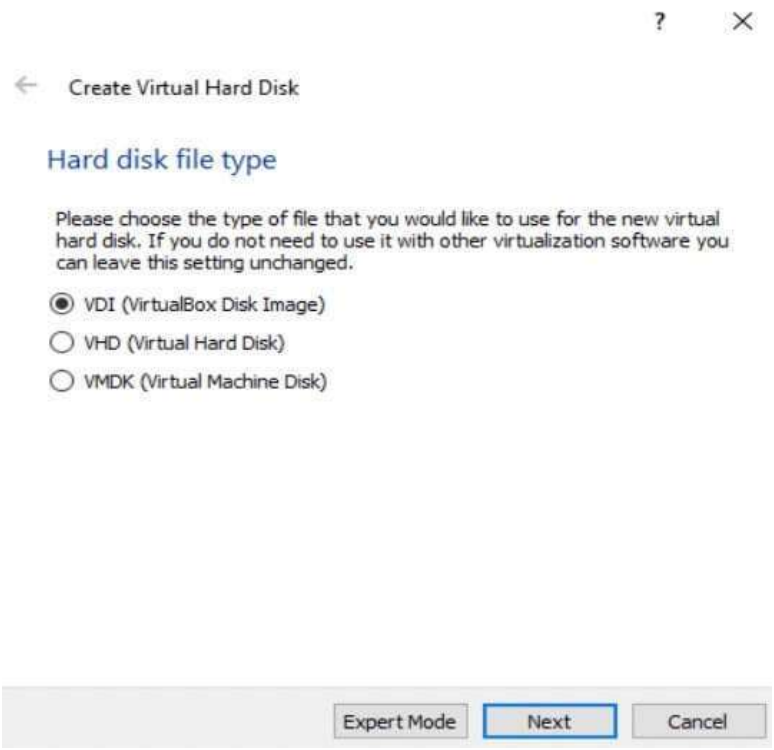
- Select the appropriate memory size for your Virtual Machine. Be mindful of the minimum specs outlined in the prerequisite section of this article. Click Next to go onto the next dialogue.



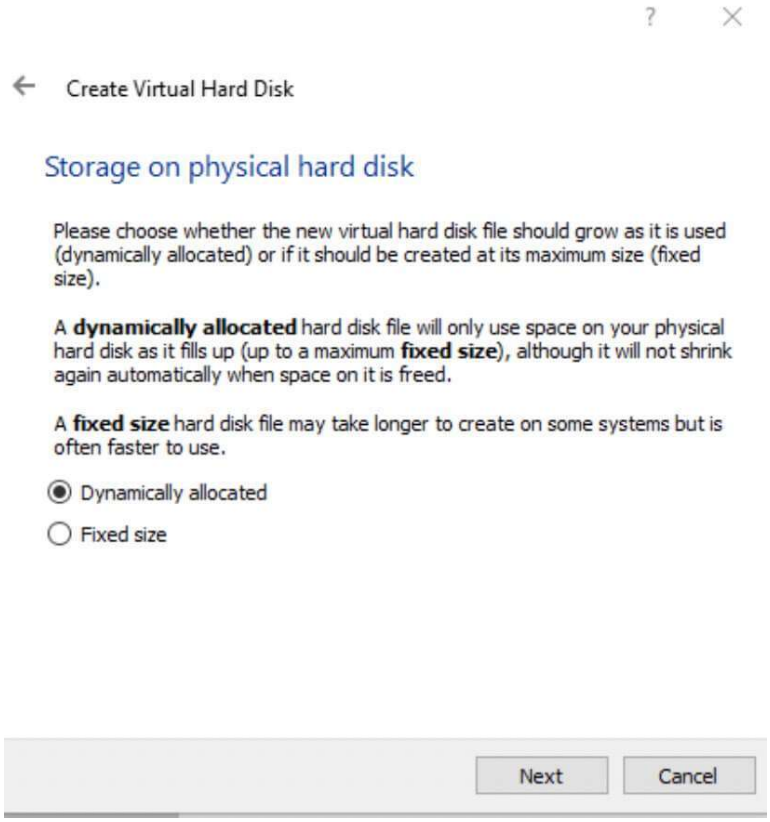
- Choose the default, which is ‘Create a virtual hard disk now ‘. Click the ‘Create’ button.



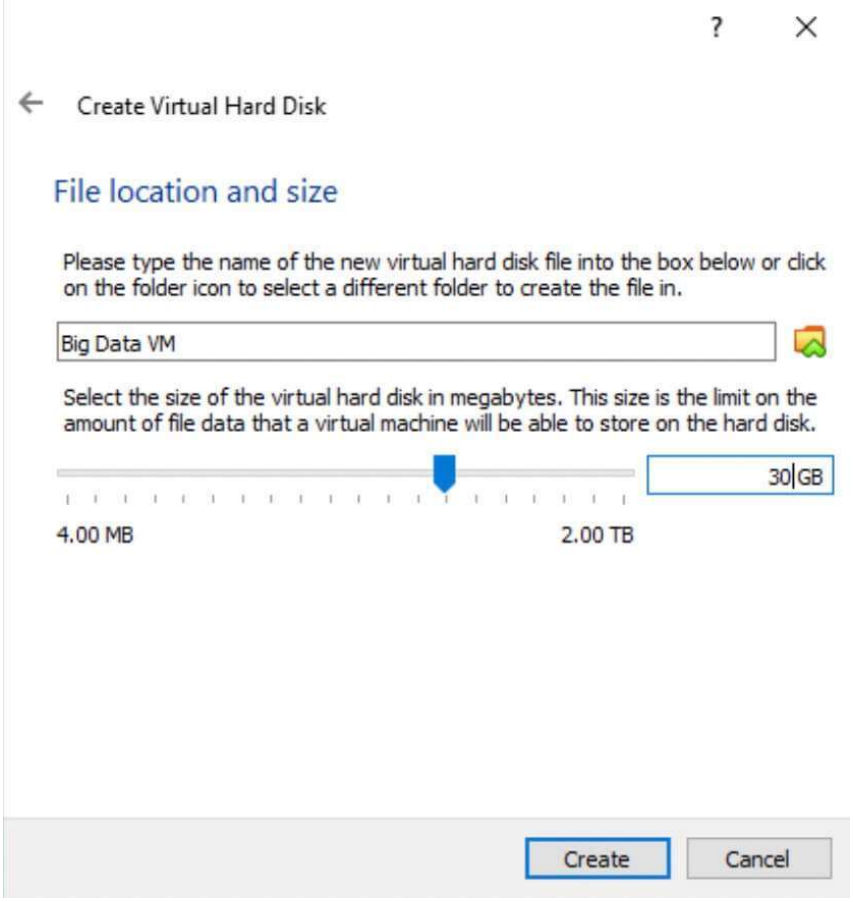
- Choose the VDI Hard Disk file type and Click ‘Next’.



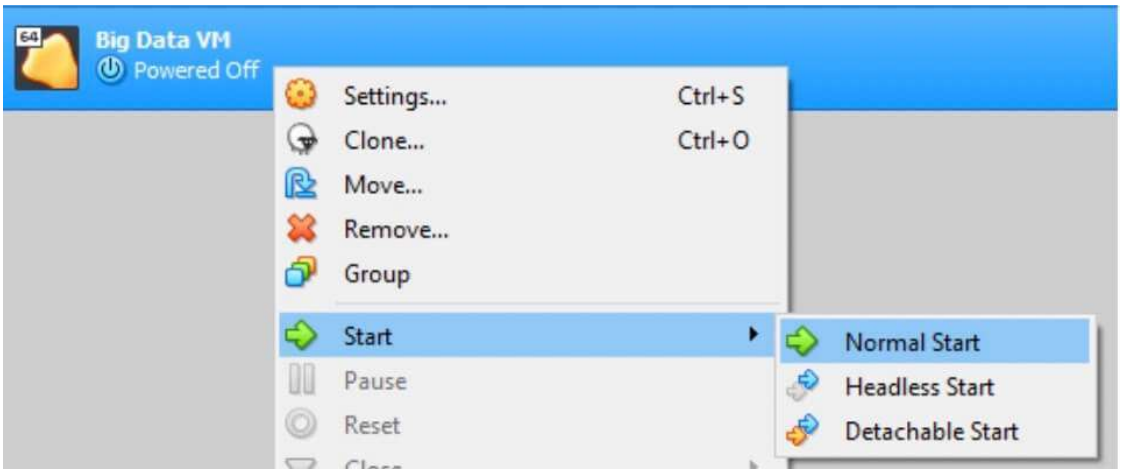
- Choose Dynamically allocated and Select ‘Next’.



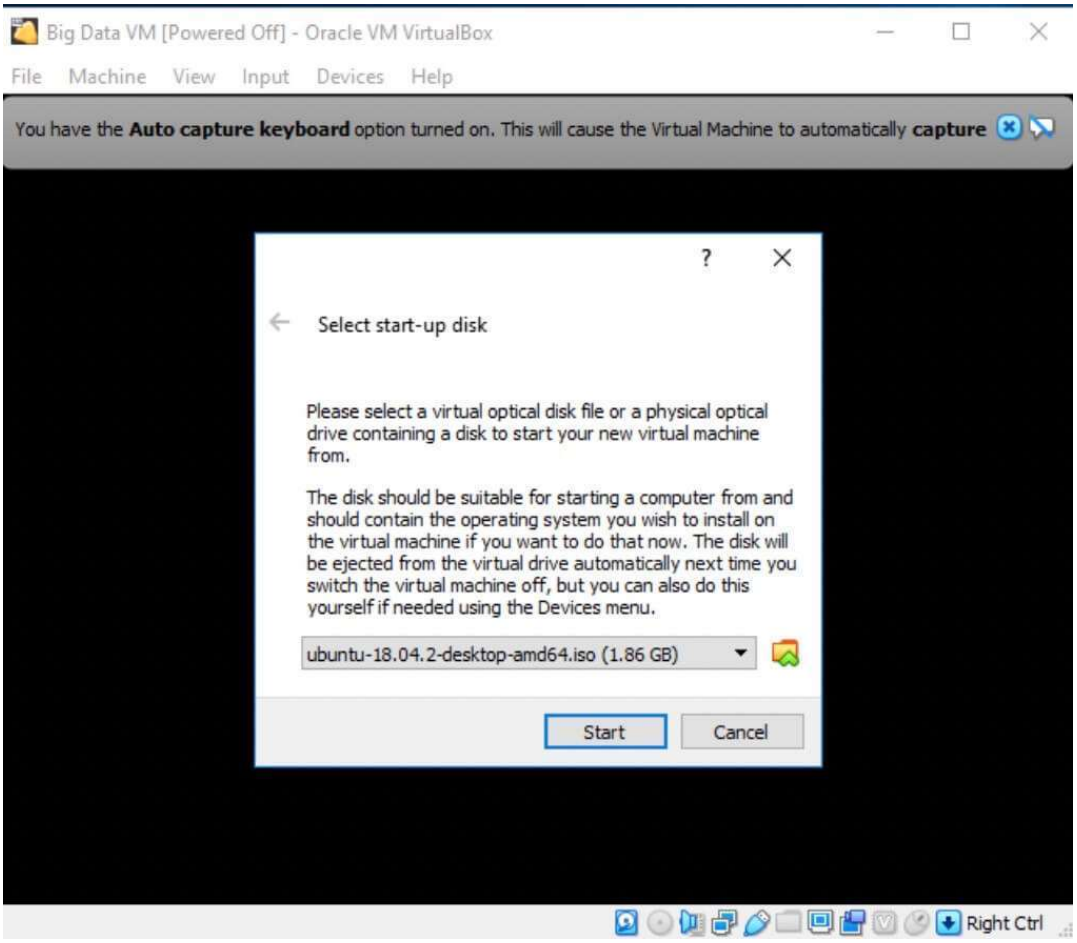
- Choose the Hard drive space reserved by the Virtual Machine and hit ‘Create’.



- At this point, your VM should be created! Now go back to the Oracle VM VirtualBox Manager and start the Virtual Machine. You can start your machine by right clicking your new instance choosing Start [wp-svg-icons icon=”arrow-right-2″ wrap=”i”] Normal Start.



- After selecting Start, you will be prompted to add a Start-up disk. You will need to navigate on your file system to where you saved your Ubuntu ISO file.



At this point, you will be taken to an Ubuntu installation screen. The process is straightforward and should be self-explanatory. The installation process will only take a few minutes. We're getting close to starting up our Hadoop instance!

LAB 7
INSTALL HADOOP IN STANDALONE MODE

Setup Hadoop

Prerequisite Installations

Next, it's necessary to first install some prerequisite software. Once logged into your Linux VM, simply run the following commands in Linux Terminal Window to install the software.

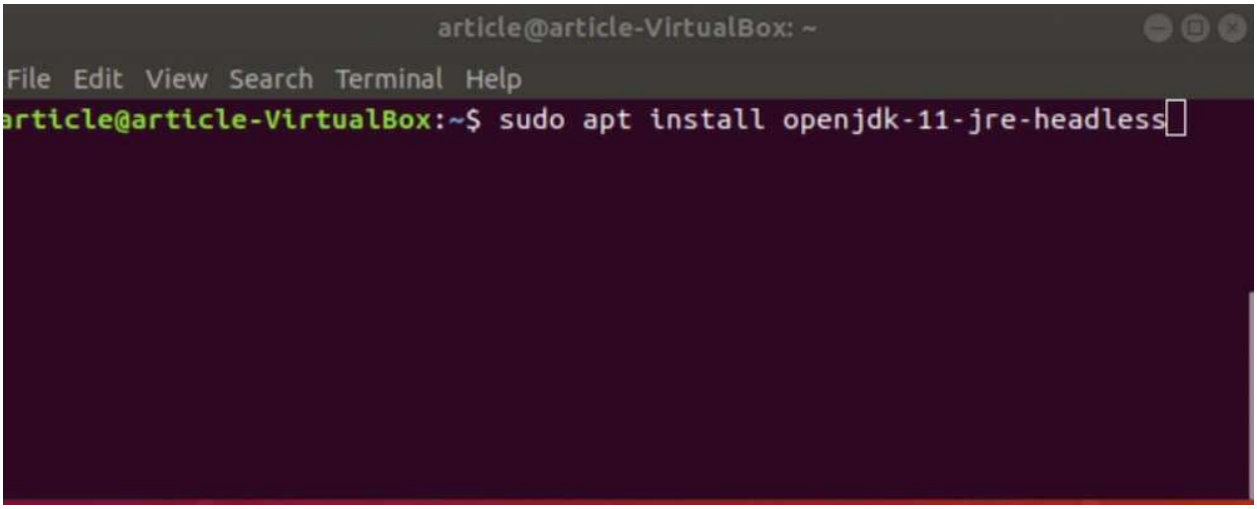
- JAVA: Terminal Command:

  $ sudo apt install openjdk-11-jre-headless
- ssh: Terminal Command:

  $ sudo apt-get install ssh

- pdsh: Terminal Command:

  $ sudo apt-get install pdsh



Download and Unpack Hadoop

Now let's download and unpack Hadoop.

- To **download Hadoop**, enter the following command in the terminal window:

  $ wget http://www.gtlib.gatech.edu/pub/apache/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz

- To **unpack Hadoop**, enter the following commands in the terminal window:

  $ tar -xvf hadoop-3.3.0.tar.gz

  $ mv hadoop-3.3.0 hadoop

  $ sudo mv hadoop/ /usr/share/

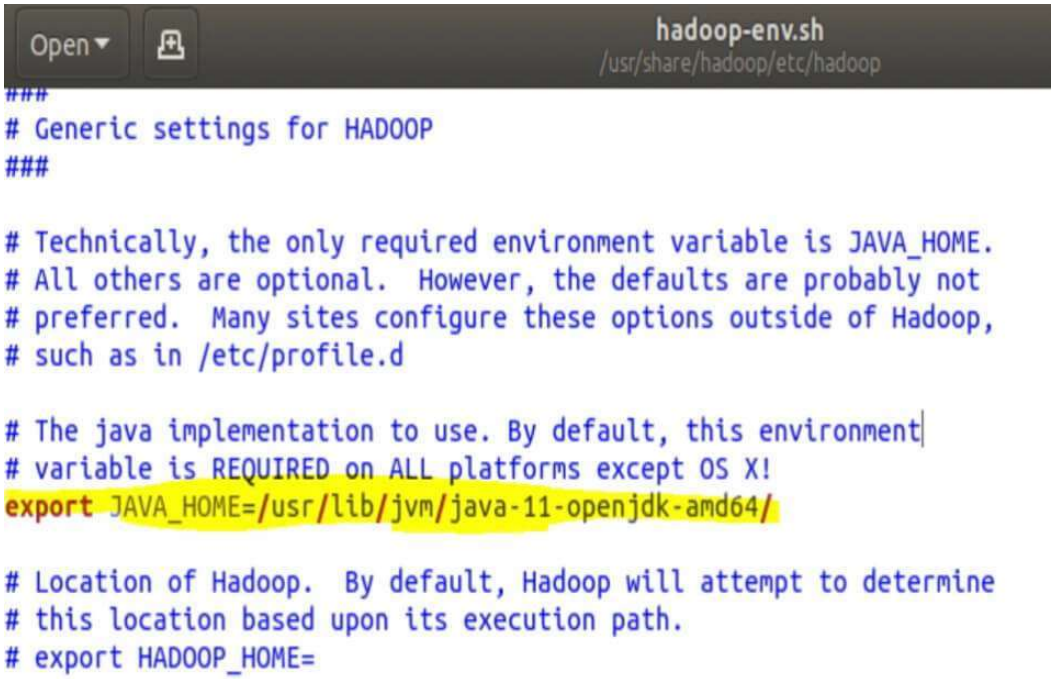
  $ export HADOOP_HOME=/usr/share/Hadoop

Setting the JAVA_HOME Environment Variable

Navigate to the 'etc/hadoop/hadoop-env.sh' file and open it up in a text editor. Find the 'export JAVA_HOME' statement and replace it with the following line:

export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64/

It should look like the picture below.



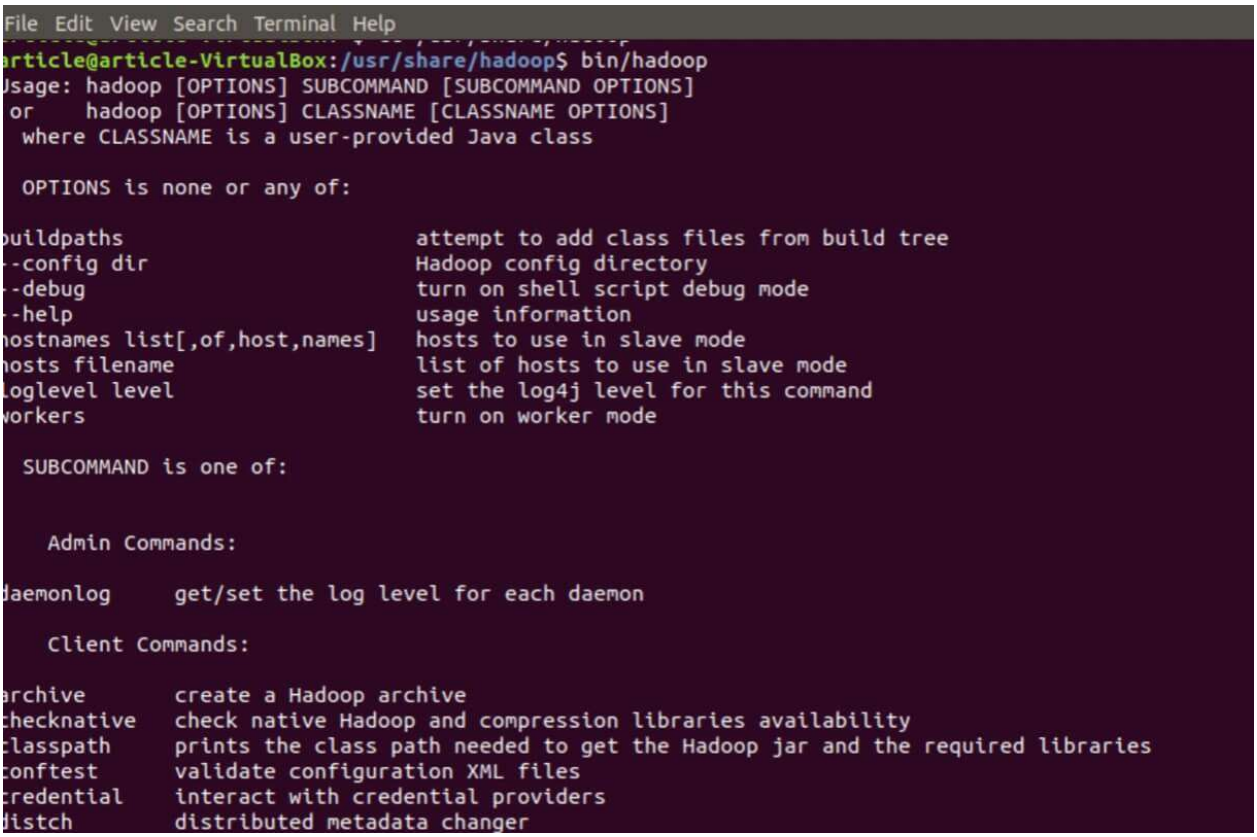
Standalone Operation

The first mode we will be looking at is Local (Standalone) Mode. This method allows you to run a single JAVA process in non-distributed mode on your local instance. It is not run by any Hadoop Daemons or services.

- Navigate to your Hadoop Directory by entering the following command in the terminal window:
  $cd /usr/share/Hadoop

- Next, run the following command:
  $ bin/Hadoop

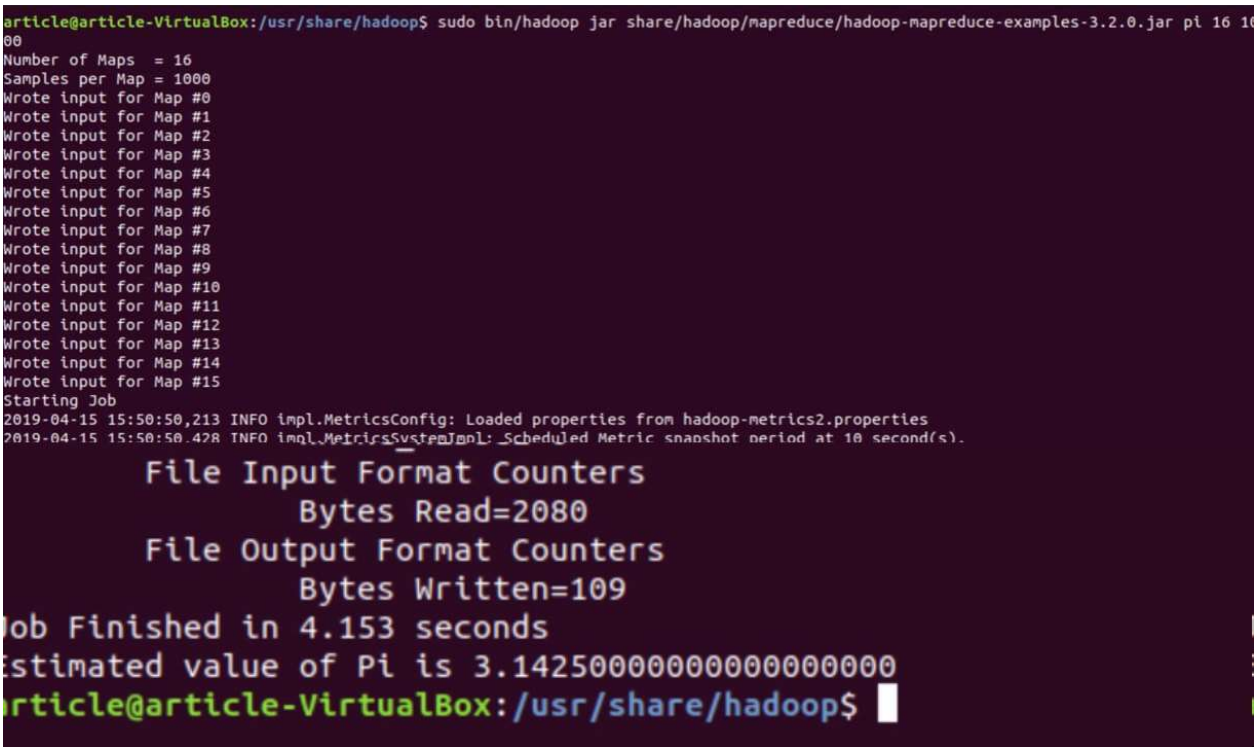
The output should look similar to the following:



- Next, we will try running a simple PI estimator program, which is included in the Hadoop Release. Try running the following command in the Terminal Window:

  sudo bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.0.jar pi 16 1000

The output should look similar to the following:



LAB 8

PERFORM THE BELOW HADOOP MANAGEMENT TASKS

(a) Add and delete directories

(b) Add and delete files

Solution (a):

- To add a directory in HDFS, you can use the 'hadoop fs –mkdir' command. Here is the syntax:

  $ hadoop fs -mkdir <directory_path>

  Example:

  $ hadoop fs -mkdir /mydir

  You can also create nested directories:

  $ hadoop fs -mkdir /mydir/subdir

- To delete a directory in HDFS, you can use the 'hadoop fs –rm' or 'hadoop fs –rmdir' command. The '-rm' command is used to delete files or directories recursively, while the '–rmdir' command is used to delete an empty directory. Here is the syntax:

  $ hadoop fs -rm -r <file_or_directory_path>

  $ hadoop fs -rmdir <directory_path>

  Example:

  $ hadoop fs -rm -r /mydir

  To delete an empty directory called "subdir," you would run:

  $ hadoop fs -rmdir /mydir/subdir

Solution (b):

- To add a file to HDFS, you can use the 'hadoop fs –copyFromLocal' command. Here is the syntax:

  $ hadoop fs -copyFromLocal <local_file_path> <hdfs_destination_path>

Example:

$ hadoop fs -copyFromLocal localfile.txt /mydir/

- To delete a file in HDFS, you can use the 'hadoop fs –rm' command. Here is the syntax:

  $ hadoop fs -rm <hdfs_file_path>

  Example:

  $ hadoop fs -rm /mydir/myfile.txt

LAB 9

INSTALLATION OF HIVE

Pre-requisite

- **Java Installation -** Check whether the Java is installed or not using the following command.

  $ java -version

- **Hadoop Installation -** Check whether the Hadoop is installed or not using the following command.

  $ hadoop version

Steps to Install Apache Hive

- Download the Apache Hive tar file.

  http://mirrors.estointernet.in/apache/hive/hive-1.2.2/

- DUnzip the downloaded tar file.

  tar -xvf apache-hive-1.2.2-bin.tar.gz

- DOpen the bashrc file.

  $ sudo nano ~/.bashrc

- DNow, provide the following HIVE_HOME path.

  export HIVE_HOME=/home/codegyani/apache-hive-1.2.2-bin
  export PATH=$PATH:/home/codegyani/apache-hive-1.2.2-bin/bin

- DUpdate the environment variable.

  $ source ~/.bashrc

- DLet's start the hive by providing the following command.

  $ hive

LAB 10

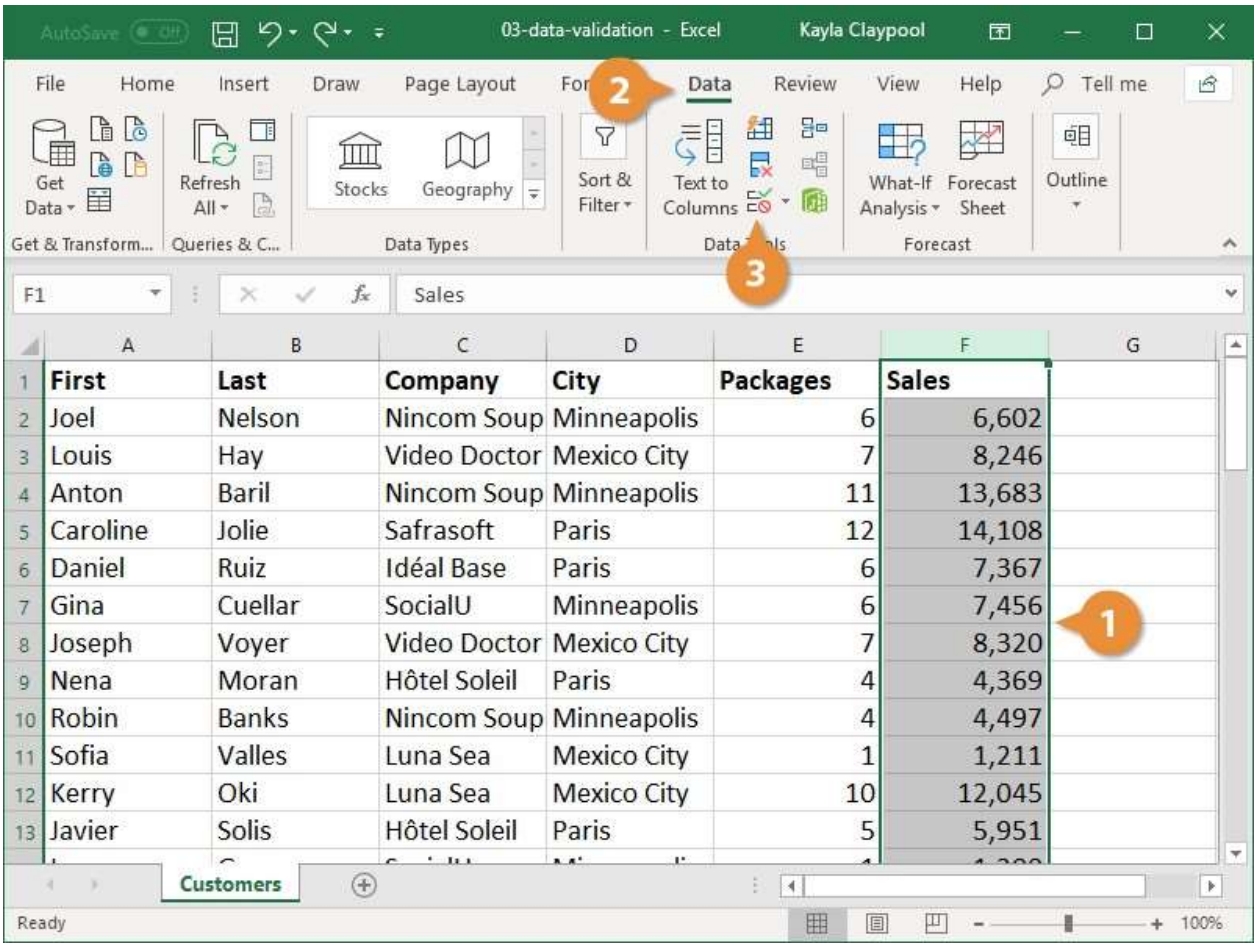
CREATE AND APPLY DATA VALIDATION IN EXCEL SHEET

Create a Validation Rule:

1. Select the cells you want to validate.

2. Click the **Data** tab.

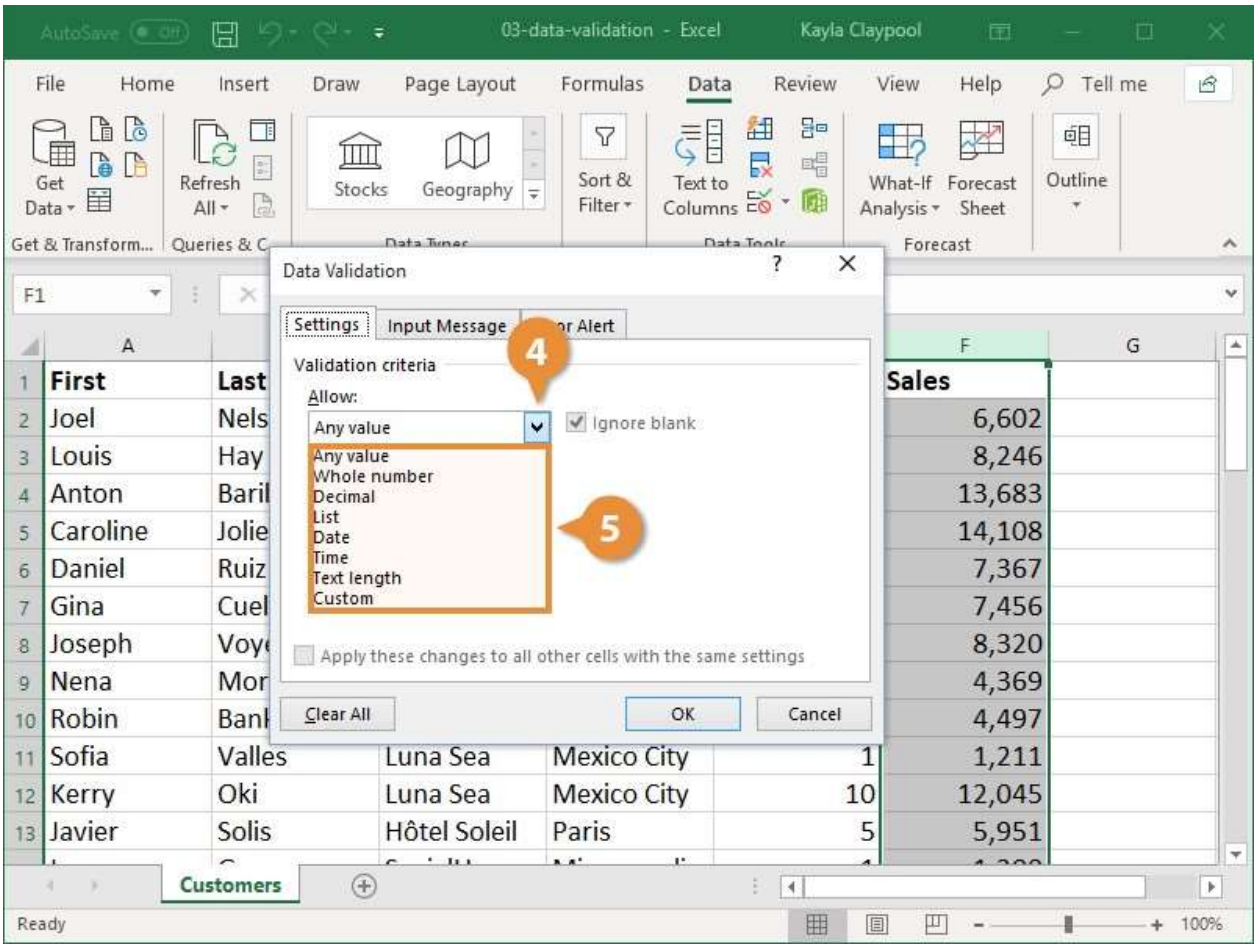
3. Click the **Data Validation** button.



	First	Last	Company	City	Packages	Sales
2	Joel	Nelson	Nincom Soup	Minneapolis	6	6,602
3	Louis	Hay	Video Doctor	Mexico City	7	8,246
4	Anton	Baril	Nincom Soup	Minneapolis	11	13,683
5	Caroline	Jolie	Safrasoft	Paris	12	14,108
6	Daniel	Ruiz	Idéal Base	Paris	6	7,367
7	Gina	Cuellar	SocialU	Minneapolis	6	7,456
8	Joseph	Voyer	Video Doctor	Mexico City	7	8,320
9	Nena	Moran	Hôtel Soleil	Paris	4	4,369
10	Robin	Banks	Nincom Soup	Minneapolis	4	4,497
11	Sofia	Valles	Luna Sea	Mexico City	1	1,211
12	Kerry	Oki	Luna Sea	Mexico City	10	12,045
13	Javier	Solis	Hôtel Soleil	Paris	5	5,951

4. Click the **Allow** list arrow.

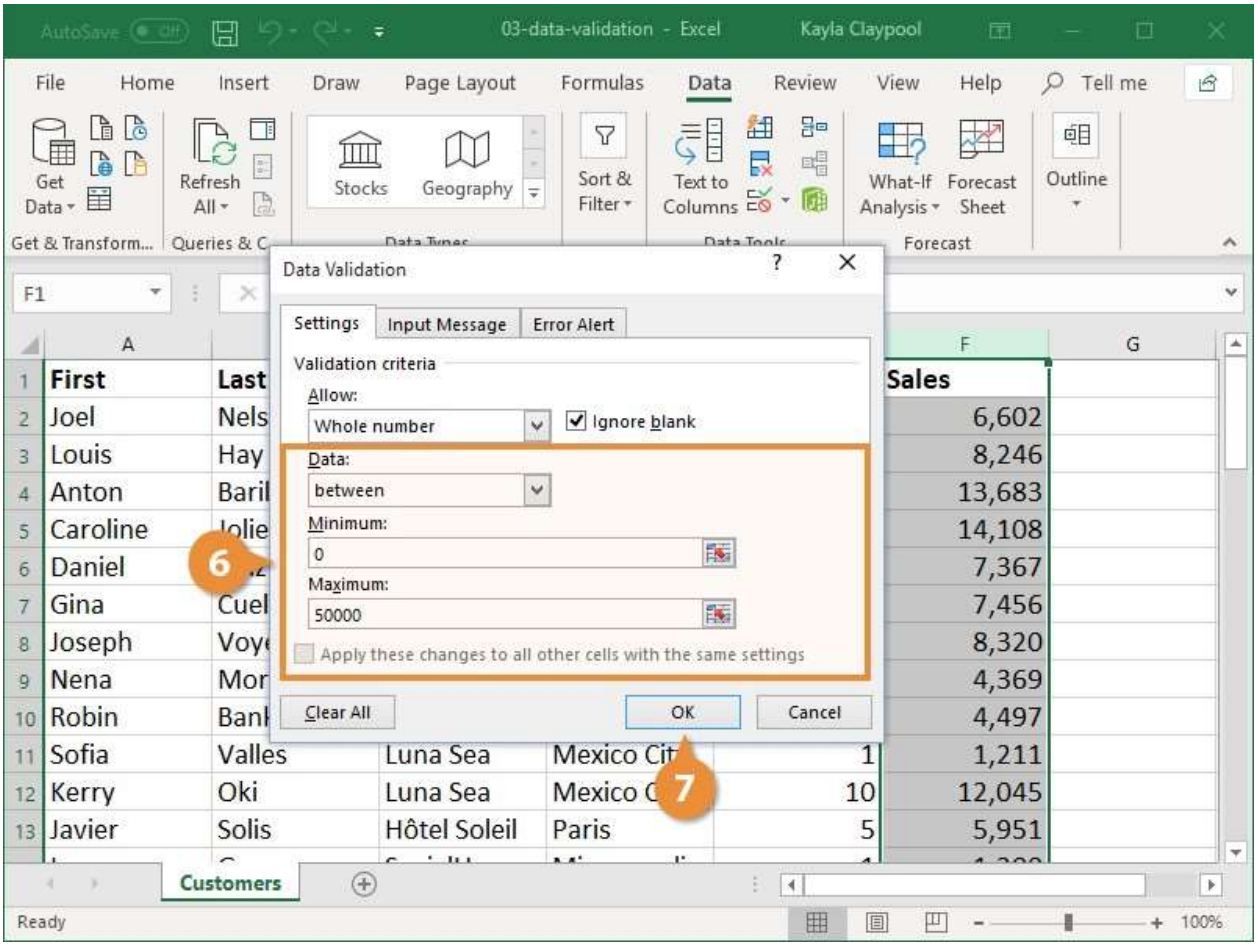
5. Select the type of data you want to allow.
    - **Any value:** No validation criteria applied.
    - **Whole number:** Allows a whole number between the minimum and maximum limits set.
    - **Decimal:** Allows a decimal or a percent entered as a decimal between the set limits.
    - **List:** Allows a value from a list of choices. A list arrow appears in the cell, and users can choose from the list.
    - **Date:** Allows a date within set limits.
    - **Time:** Allows a time within set limits.
    - **Text length:** Allows text containing a certain number of characters.
    - **Custom:** Allows a formula to be entered to calculate what is allowed in the cell.



6. Specify the data validation rules.

The validation options will vary depending on the option selected in the Allow field.

7. Click **OK**.



The data validation is set for the selected cell(s). When a user tries to enter data that is not valid, Excel will prevent the entry and display a message about the cell being restricted.

To find validated data in a worksheet, click the **Find & Select** button in the Editing group on the Home tab and select **Data Validation**. The validated cells are highlighted.

Add Input and Error Messages (Apply Data Validation):

Prevent data validation issues by setting up Excel to display a message whenever a cell or range of cells is selected. These messages are useful when other people will be entering data in your worksheet. An error message can be configured to appear when data is entered that does not match a data validation rule.

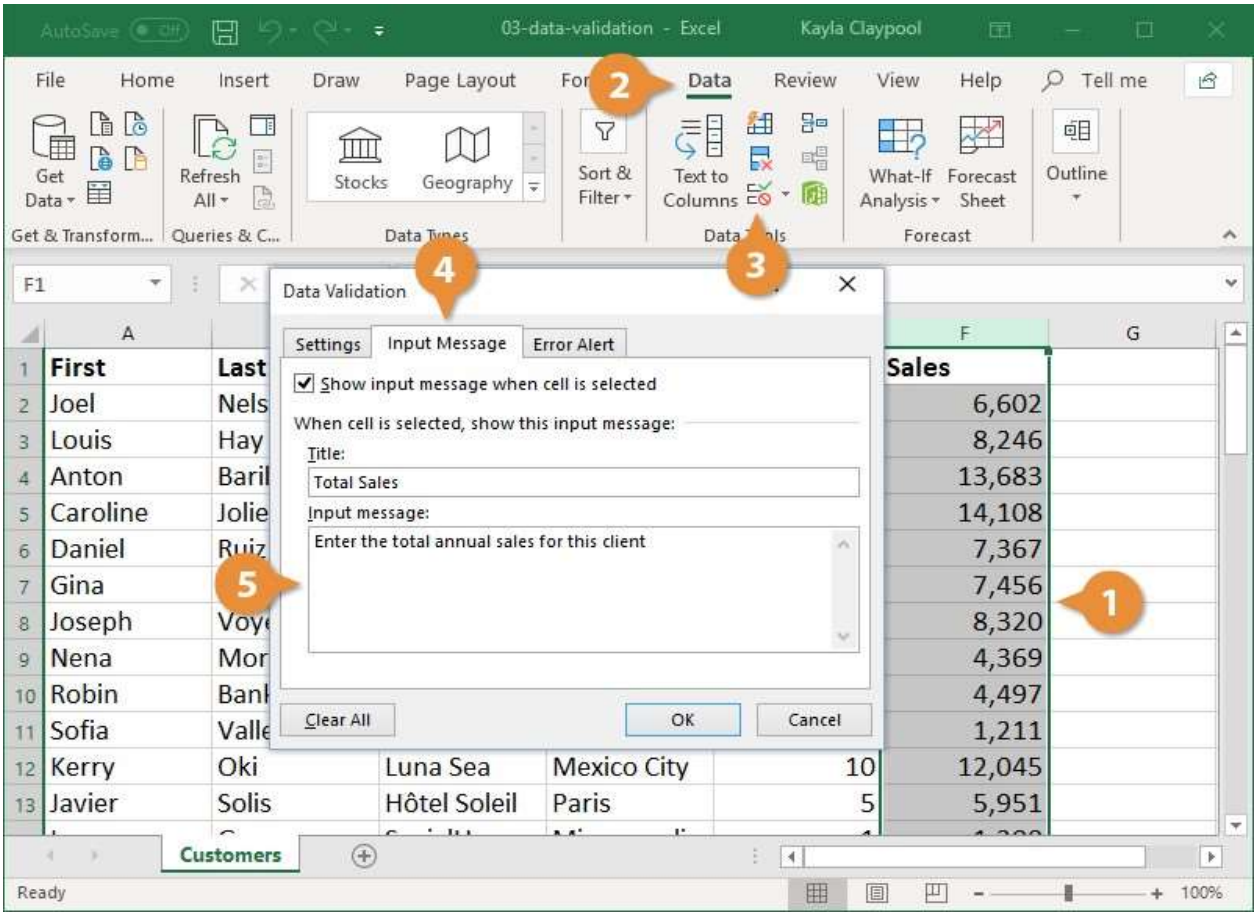
1. Select the cells where you want an input message to appear.

2. Click the **Data** tab.

3. Click the **Data Validation** button.

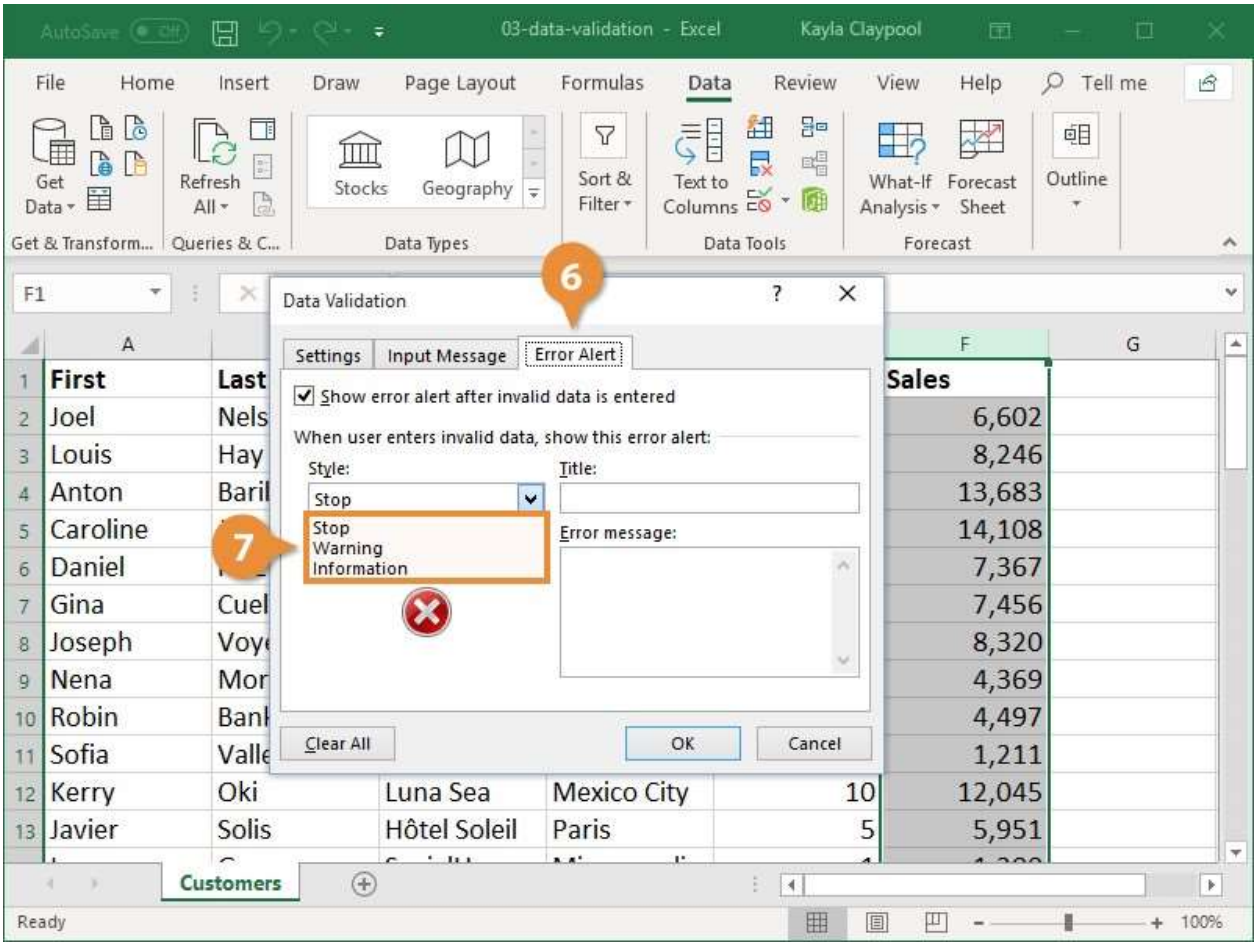
4. Click the **Input Message** tab.

5. Enter an input message.



6. Click the **Error Alert** tab.

7. Select an error alert style.

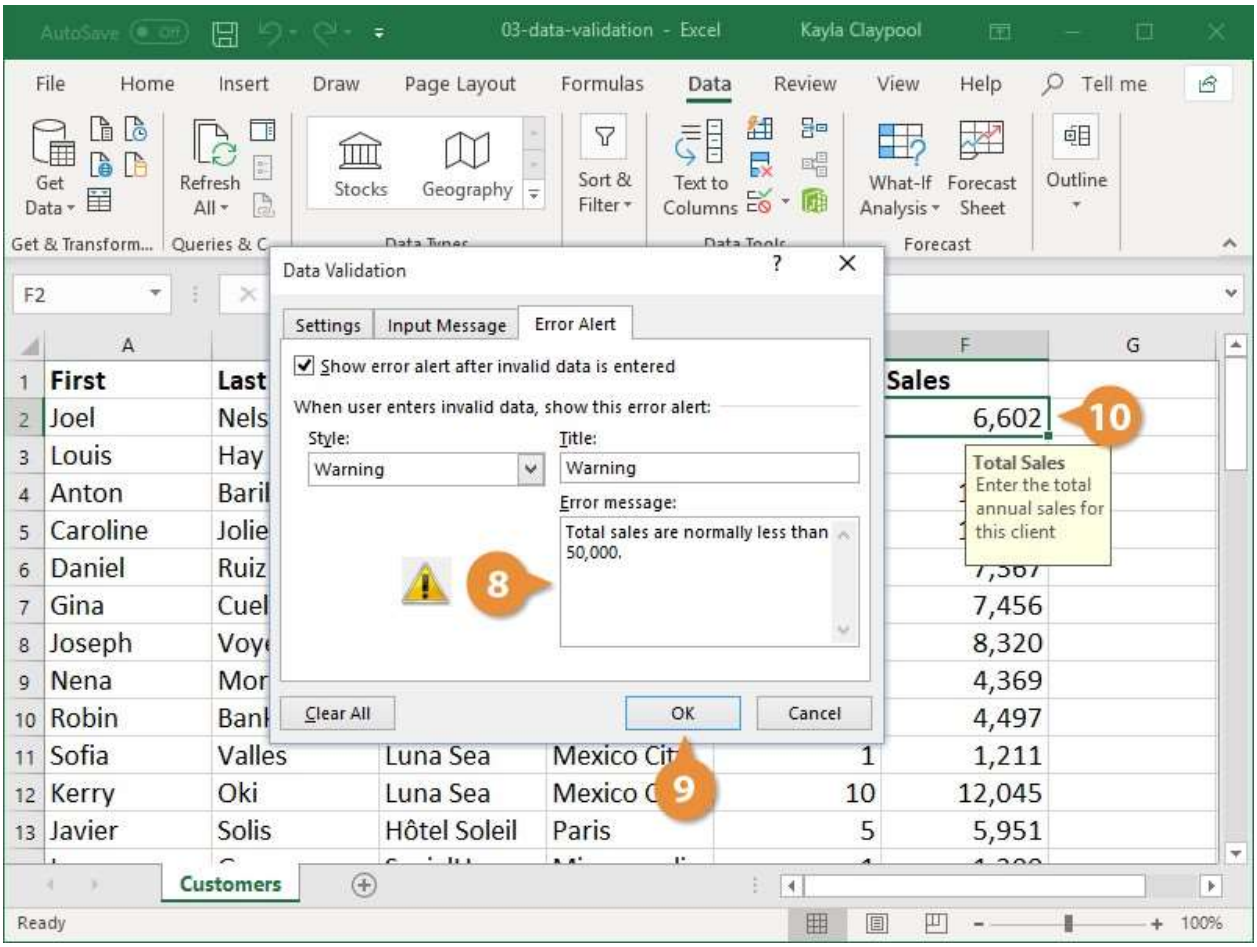


- **Stop:** Prevents users from adding invalid data in a cell.
- **Warning:** Warns that the data entered is invalid, but users can click **Yes** to accept the invalid entry, **No** to edit it, or **Cancel** to remove it.
- **Information:** Informs users that the data entered is invalid, but users can click **OK** to accept the invalid entry or **Cancel** to remove it.

8. Enter an error alert message.

9. Click **OK**.

10. Select a cell with an input message.



Now when a cell in the range is selected, the title and message display. If you enter an invalid value, a custom error message appears.

LAB 11

WRITE A PROGRAM IN PYTHON FOR EXTRACTING DATA FROM SOCIAL MEDIA PLATFORM LIKE FACEBOOK

Getting Started

To use the Facebook API, you will need to create a Facebook Developer account and create an app. Follow the steps below to get started:

1. Create a Facebook Developer account: Go to the Facebook Developer website and sign up for an account. You will need to provide some basic information, including your name and email address.

2. Create a new app: Once you have created your account, click on the Create App button and choose the platform you want to develop for. For this tutorial, we will be developing for the web.

3. Set up your API key: After creating your app, you will be given an API key. This key will be used to authenticate your requests to the Facebook API.

Now that you have set up your Facebook Developer account and created an app, you are ready to start using the Facebook API.

Facebook Graph API

Facebook provide a powerful and comprehensive RESTful API known as the Facebook Graph API. The Graph API allows developers to interact with Facebook data, such as users, pages, posts, and more, using HTTP requests.

To use the Facebook Graph API with Python, you can make HTTP requests to the API endpoints using libraries like requests or httplib2, and then parse the JSON responses to extract the data you need. Additionally, there are several third-party libraries available that simplify working with the Facebook Graph API in Python, such as facebook-sdk and python-graph-api.

We will use requests library to interact with Facebook Graph API.

Making Requests

- Install requests library

  pip install requests

Program:

```
import requests

access_token =
'EAACknJspLqUBOx36ZAZCEuNGdDKfZC7uZAtumKuPvn2hQyR0K3oiTmEIwT8hUVNzW
1aWA7gz2D3IaSu2euCQTUB57hQnG5ZCCxGlDP8spXAC2C44BZAHDmqmR39fGozIHslt6
BnpSeqlBSqJYOaX4ZAlZA8vkJoZBRIGVi9YDpZABewcLk8E1PRTRRy6hwy9Du6AgsI8mJ
mO4bZBi7F2g1hbP9T8Rr6j50q'

user_id = 252902701073046

url = f'https://graph.facebook.com/{user_id}?access_token={access_token}'

response = requests.get(url)

print(response.json())
```

Output:

```
{'name': 'Anubhav Yadav', 'id': '252902701073046'}
```

Handling Responses

Program:

```
url = f'https://graph.facebook.com/{user_id}?access_token={access_token}'

response = requests.get(url)

data = response.json()

print(data['name'])
```

Output:

```
Anubhav Yadav
```

Explore Facebook Graph API

- **Getting Profile Details**

Program:

```
url = f'https://graph.facebook.com/{user_id}?fields=name,id,picture&access_token={access_token}'

response = requests.get(url)

data = response.json()

print(data)
```

Output:

```
{'name': 'Anubhav Yadav', 'id': '252902701073046', 'picture': {'data': {'height': 50, 'is_silhouett
e': False, 'url': 'https://platform-lookaside.fbsbx.com/platform/profilepic/?asid=252902701073046&hei
ght=50&width=50&ext=1698762965&hash=AeS-EgBLOIdauB5Oyw8', 'width': 50}}}
```

The user_id is a unique identifier for a Facebook user, which is assigned to each user when they create a Facebook account.
To get the user_id of a Facebook user, you can follow these steps:

1. Go to the Facebook profile of the user you want to get the user_id for.

2. Look at the URL in your browser's address bar.

3. The user_id is the string of numbers that comes after the "facebook.com/" part of the URL.

- **Getting Posts**

Program:

```
access_token =
'EAACknJspLqUBOx36ZAZCEuNGdDKfZC7uZAtumKuPvn2hQyR0K3oiTmEIwT8hUVNzW
1aWA7gz2D3IaSu2euCQTUB57hQnG5ZCCxGlDP8spXAC2C44BZAHDmqmR39fGozIHslt6
BnpSeqlBSqJYOaX4ZAlZA8vkJoZBRIGVi9YDpZABewcLk8E1PRTRRy6hwy9Du6AgsI8mJ
mO4bZBi7F2g1hbP9T8Rr6j50q'

page_id = 142033095648902

url=f'https://graph.facebook.com/{page_id}/posts?access_token={access_token}'

response = requests.get(url)

data = response.json()

for post in data['data']:
    print(post)
```

Output:

```
{'created_time': '2023-09-21T14:49:30+0000', 'message': 'Here is the complete roadmap of "Data Analys
t". Learn and boost your career...\n#python #learning #dataanalyst #dataanalysis', 'id': '14203309564
8902_122104808210046909'}
{'created_time': '2023-09-16T06:42:54+0000', 'message': 'Hello World! This is my first facebook pag
e.', 'id': '142033095648902_122098495712046909'}
{'created_time': '2023-09-16T06:39:19+0000', 'story': 'Tech Talks updated their profile picture.', 'i
d': '142033095648902_122098494062046909'}
```

LAB 12

WRITE A PROGRAM IN PYTHON FOR EXTRACTING DATA FROM DATA SOURCE AND PERFORM MANIPULATION ON DATASET

```
import pandas as pd

df = pd.read_csv('employees.csv')

df.head()
```

Output:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services

```
df.info()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   First Name        933 non-null    object
 1   Gender            855 non-null    object
 2   Start Date        1000 non-null   object
 3   Last Login Time   1000 non-null   object
 4   Salary            1000 non-null   int64
 5   Bonus %           1000 non-null   float64
 6   Senior Management 933 non-null    object
 7   Team              957 non-null    object
dtypes: float64(1), int64(1), object(6)
memory usage: 62.6+ KB
```

df.describe(include='object')

Output:

	First Name	Gender	Start Date	Last Login Time	Senior Management	Team
count	933	855	1000	1000	933	957
unique	200	2	972	720	2	10
top	Marilyn	Female	10/30/1994	1:35 PM	True	Client Services
freq	11	431	2	5	468	106

df.shape

Output:

```
df.shape
```

```
(1000, 8)
```

df.tail()

Output:

```
df.tail()
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

Removing 'Last Login Time' from dataframe

data = df.drop('Last Login Time',axis=1)

data.head()

Output:

```
# Removing 'Last Login Time' from dataframe
data = df.drop('Last Login Time',axis=1)
data.head()
```

	First Name	Gender	Start Date	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	101004	1.389	True	Client Services

Using groupby to get 'Team' column in groups

group_data = df.groupby(by='Team')

group_data.first()

Output:

Team	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management
Business Development	Frances	Female	8/8/2002	6:51 AM	139852	7.524	True
Client Services	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True
Distribution	Michael	Male	10/10/2008	11:25 AM	99283	2.665	True
Engineering	Angela	Female	11/22/2005	6:29 AM	95570	18.523	True
Finance	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False
Human Resources	Brandon	Male	12/1/1980	1:08 AM	112807	17.492	True
Legal	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False
Marketing	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True
Product	Ruby	Female	8/17/1987	4:20 PM	65476	10.012	True
Sales	Gary	Male	1/27/2008	11:40 PM	109831	5.831	False

LAB 13

WRITE A PROGRAM OF PRE-PROCESSING (HANDLING OF MISSING VALUES) IN PYTHON PLATFORM

```
# Missing or Null values in dataframe
df.isna().sum()
```

Output:

```
First Name            67
Gender               145
Start Date             0
Last Login Time        0
Salary                 0
Bonus %                0
Senior Management     67
Team                  43
dtype: int64
```

```
# Filling null values in 'Gender' column
df['Gender'].fillna('No Gender',axis=0,inplace=True)
df.isna().sum()
```

Output:

```
First Name            67
Gender                 0
Start Date             0
Last Login Time        0
Salary                 0
Bonus %                0
Senior Management     67
Team                  43
dtype: int64
```

Remove all null values rows from dataframe

df.dropna(axis=0,inplace=True)

df.isna().sum()

Output:

```
First Name           0
Gender               0
Start Date           0
Last Login Time      0
Salary               0
Bonus %              0
Senior Management    0
Team                 0
dtype: int64
```

df.shape

Output:

```
df.shape
```

```
(899, 8)
```

LAB 14

WRITE A PROGRAM TO PERFORM READING AND PROCESSING ANDROID SENSOR DATA USING PYTHON WITH CSV READ

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')

df = pd.read_csv('Orientation.csv')
df.head()
```

Output:

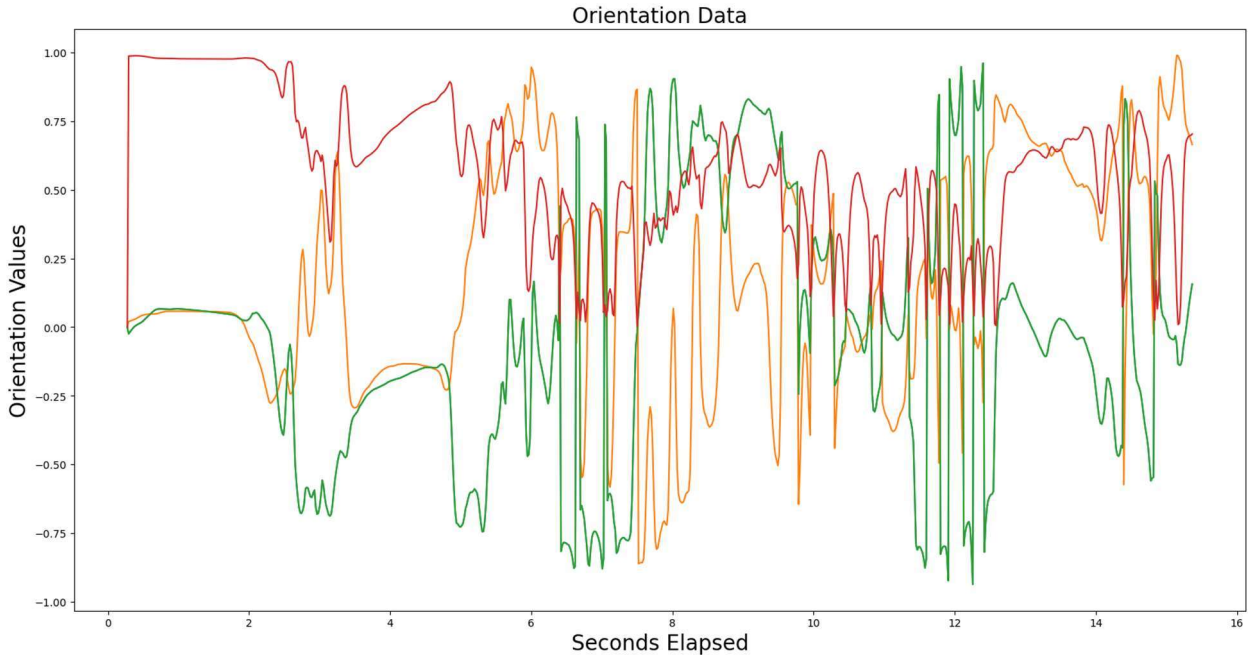
	time	seconds_elapsed	qz	qy	qx	qw	roll	pitch	yaw
0	1695389724915486700	0.276487	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	1695389724933778000	0.294778	-0.024466	0.019904	0.162414	0.986219	0.049837	-0.325072	0.057781
2	1695389724952067600	0.313068	-0.017869	0.021995	0.162239	0.986345	0.051926	-0.324949	0.044742
3	1695389724970358300	0.331358	-0.011539	0.023924	0.161571	0.986504	0.053750	-0.323860	0.032175
4	1695389724988644400	0.349644	-0.005513	0.025479	0.159345	0.986879	0.054851	-0.319644	0.020017

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 826 entries, 0 to 825
Data columns (total 9 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   time              826 non-null    int64
 1   seconds_elapsed   826 non-null    float64
 2   qz                826 non-null    float64
 3   qy                826 non-null    float64
 4   qx                826 non-null    float64
 5   qw                826 non-null    float64
 6   roll              826 non-null    float64
 7   pitch             826 non-null    float64
 8   yaw               826 non-null    float64
dtypes: float64(8), int64(1)
memory usage: 58.2 KB
```

```
plt.figure(figsize=(20, 10))

plt.plot(df['seconds_elapsed'], df['qz'], df['seconds_elapsed'], df['qy'], df['seconds_elapsed'], df['qz'],df['seconds_elapsed'],df['qw'])

plt.title('Orientation Data', fontsize=20)

plt.xlabel('Seconds Elapsed', fontsize=20)

plt.ylabel('Orientation Values', fontsize=20)
```

Output:



Alternative Method of Plotting the Graph

```
fig = plt.figure(figsize=(20,10))

ax = fig.add_axes([0.1,0.1,0.8,0.8])

ax1 = ax.plot(df['seconds_elapsed'],df['qx'],'r')

ax2 = ax.plot(df['seconds_elapsed'],df['qy'],'b')

ax3 = ax.plot(df['seconds_elapsed'],df['qz'],'g')

ax4 = ax.plot(df['seconds_elapsed'],df['qw'],'y')

ax.set_title('Orientation Data', fontsize=20)

ax.set_xlabel('Seconds Elapsed', fontsize=20)
```

```
ax.set_ylabel('Orientation Values', fontsize=20)
ax.legend(['qx','qy','qz','qw'],loc='lower left')
```

Output: