We can initialize ... Their main data object is the **ndarray**, an n-dimensional array type that describes a collection of 'items' of the same type.

An array of two dimensions/index/subscript is termed as a 2D (two-dimensional) array.

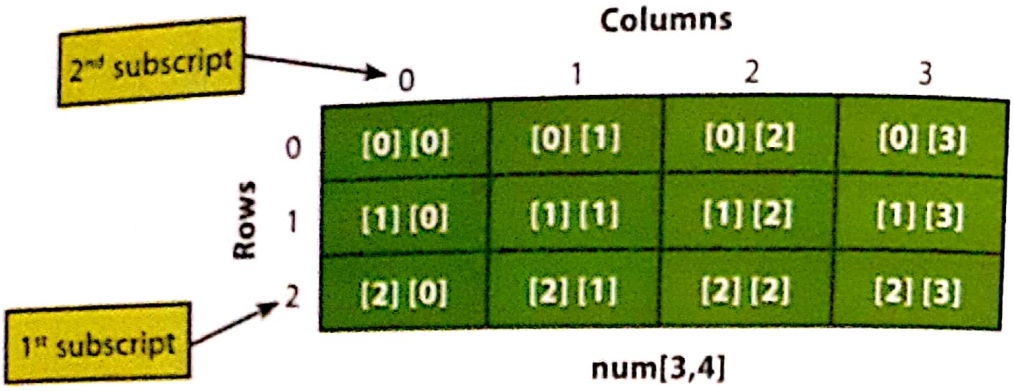


Fig. 1.4: 2D Representation in Memory

In Fig. 1.4, num is an array of two dimensions with 3 rows and 4 columns. Subscript of rows is from 0 to 2 and columns are from 0 to 3.

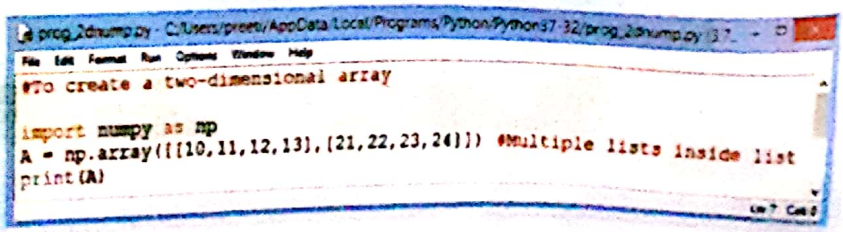
Table 1.1: Difference between NumPy Array and List

NumPy Array	List
NumPy array works on homogeneous (same) types.	Python lists are made up of heterogeneous (different) types.
NumPy array does not support addition and removal of elements.	Python list supports adding and removal of elements.
Can't contain elements of different types	Can contain elements of different types
Less memory consumption	More memory consumption
Faster runtime execution	Runtime execution is comparatively slower than arrays

1.4 HOW TO CREATE A NumPy ARRAY

There are several ways to create a NumPy array. Let us see a few of them in brief.

S.No.	Method	Description	Examples

1.	array()	To create a one-dimensional/two-dimensional NumPy array, we can simply pass a Python list to the **array** method.	`#To create a one-dimensional array`
			`import numpy as np`
			`l1 = [2, 3, 4, 5, 6]`
			`arr1 = np.array(l1)`
			`print(arr1)`
			 

2. **fromstring()** This function creates one-dimensional array from String.

```
import numpy as np
data =np.fromstring('1 2 10 12', dtype=int, sep=' ')
print(data)
```

3. **empty()** **empty()** function can be used to create empty array or an uninitialized array of specified shape and dtype. Syntax is:

```
numpy.empty
(Shape, [dtype=
<datatype>,]
[order = 'C' or
'F'])
```

where, dtype: is a data type of Python or NumPy to set initial values. Shape: is dimension meaning row-wise arrangement of data. C means C-like. Order :'C' means column-wise arrangement of data. 'F' means Fortran-like. Also, the parameters dtype and order are optional. If you do not specify dtype, then, by default, dtype is taken as float. Similarly, default order is taken as 'C'.

```
import numpy as np
arr = np.empty([3,2], dtype=int)
print(arr)
```

```
>>>
 RESTART: C:/Users/preeti/AppData/Local
arr.py
[[6357069 6815843]
 [7209065 4653157]
 [6881397      100]]
>>>
```

4. **zeros()** The **zeros()** function works in the same manner as empty(), but the only difference is that it creates and returns a new array with all zeroes as its elements as per the specified size and type. Its syntax is:

```
numpy.
zeros(shape,
dtype=float,
order='C')
```

Here, shape refers to shape of an empty array in int or sequence of integer values. dtype is the desired output data type and is optional.

```
# array of five zeros. Default dtype is float

import numpy as np
arr1 = np.zeros(5)
print(arr1)
```

```
>>>
 RESTART: C:/Users/preeti/AppData/Local/
oes.py
[0. 0. 0. 0. 0.]
>>>
```

5. ones()

The **ones()** function works in the same manner as empty(), creates and returns a new array with all ones as its elements, as per the specified size and type. The syntax is:

```
numpy.
ones(shape,
dtype=None,
order='C')
```

```
# array of five ones. Default dtype is float

import numpy as np
arr1 = np.ones(5)
print(arr1)
```

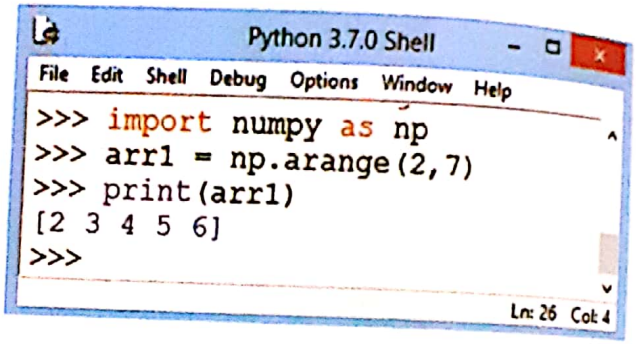
```
>>>
 RESTART: C:/Users/preeti/AppData/Local
s.py
[1. 1. 1. 1. 1.]
```

6. arange()

The **arange()** function is used to create array from a range. Its syntax is:

```
<arrayname> =
numpy.arange([
start],stop,[s
tep],[dtype])
```

Another commonly-used method for creating a NumPy array is the arange method. This method takes the start index of the array, the end index and the step size (which is optional).

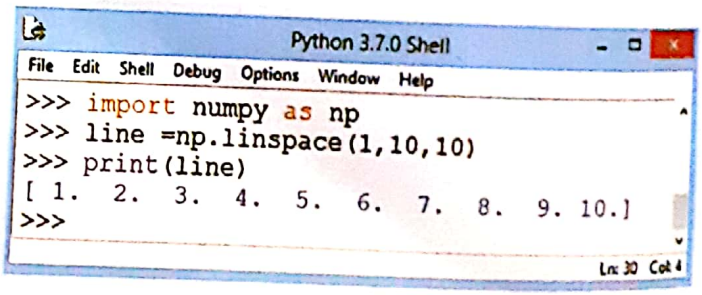


7. linspace()

Another very useful method to create NumPy array is the **linspace** method. This function can be used to prepare array of range. Its syntax is:

```
<arrayname> =
numpy.linspace
([start],stop,
[dtype])
```

This method takes three arguments: a start index, end index and the number of linearly-spaced numbers that we want between the specified range. For instance, if the first index is 1, the last index is 10, and we need 10 equally-spaced elements within this range, we can use the linspace method.



8. copy()

The **copy()** function is used to create a copy of an existing array. If we just assign a portion of an array to another array, the new array we just created refers to the parent array in memory. Its syntax is:

```
<ndarray>.
copy(order='C')
```

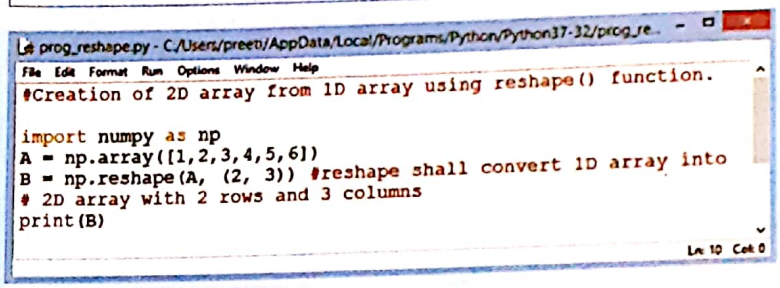
```
#Copy function is used to create the copy of the existing array.

import numpy as np
x = np.array([1, 2, 3])
y = x
z = np.copy(x)
x[0] = 10
print(x)
print(y)
print(z)
```

```
>>>
 RESTART: C:/Users/preeti/AppData/Local
.py
[10  2  3]
[10  2  3]
[1 2 3]
>>>
```

9. reshape()

We can create 2D array from 1D array using **reshape()** function. Reshaping means changing the arrangement of items so that the shape of the array changes while maintaining the same number of dimensions. Its syntax is:

```
<ndarray>.
reshape(<shape
tuple>)
```




```
>>>
 RESTART: C:/Users/preeti/AppData/Local
.py
[[1 2 3]
 [4 5 6]]
>>>
```

10. eye()/ identity()

The **eye** method can be used to create an identity matrix, which can be very useful to perform a variety of operations using arrays. An identity matrix is a matrix with zeroes across rows and columns except the diagonal. It contains all 1s as its diagonal elements. Its syntax is:

```
numpy.identity
(n,dtype=None)
```

or

```
numpy.eye
(n,dtype=
<class'float'>
```

```
#create a 4x4 identity matrix using the eye method
import numpy as np

# 4x4 matrix with 1's as all diagonal elements
mat1 = np.identity(4)  Alternatively, np.eye(4)
print("\nMatrix 1 : \n", mat1)

#Alternatively
mat2 = np.identity(4, dtype = float)
print("Matrix 2 : \n", mat2)
```

```
 RESTART: C:/Users/preeti
y1.py

Matrix 1 :
 [[1. 0. 0. 0.]
 [0. 1. 0. 0.]
 [0. 0. 1. 0.]
 [0. 0. 0. 1.]]
Matrix 2 :
 [[1. 0. 0. 0.]
 [0. 1. 0. 0.]
 [0. 0. 1. 0.]
 [0. 0. 0. 1.]]
```

MEMORY BYTES

- A NumPy array is simply a grid that contains values of the same/homogeneous type.
- NumPy arrays are also referred to as ndarrays.
- NumPy (Numerical Python) is a linear algebra library in Python.
- NumPy is very useful for performing mathematical and logical operations on arrays.
- Like the Python built-in range() function, we can create a NumPy array using arange().
- We can extract specific portions on an array using indexing starting with 0, something similar to how you would do with Python lists using Slicing.
- copy() method is used for creating another array from the parent array.
- Joining of two or more one-dimensional arrays is possible with the help of concatenate() function of NumPy object.
- NumPy provides var() function to calculate variance, cov() function to calculate covariance, and corrcoef() function to calculate correlation.
- Linear regression is a method used to find a relationship between a dependent variable and independent variable(s).

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.

(a) A NumPy ..... is simply a grid that contains values of the same/homogeneous type.

(b) The dimensions of a NumPy array are referred to as ..... .

(c) ..... is the first axis and runs along the rows.

(d) ..... is the second axis and runs along the columns.
(e) The number of axes in an ndarray is called its ..... .
(f) The ..... of an ndarray tells about the number of elements along each axis of it.
(g) The ..... method is useful for creating ndarrays from existing lists and other sequences.
(h) ..... function provides evenly-spaced elements between two given limits.
(i) ..... is used to create an uninitialized array of specified shape and data type.
(j) Array ..... refers to the process of extracting a subset of elements as another array from an existing array.
(k) A NumPy array always begins with index (from left to right) as ..... .
(l) NumPy arrays consume ..... memory as compared to lists.

Answers: (a) array (b) axes (c) axis=0 (d) axis=1 (e) rank (f) shape
(g) array() (h) linspace() (i) empty() (j) slicing (k) zero (l) less

2. State whether the following statements are True or False.
(a) Elements of arrays are stored in contiguous memory locations.
(b) Axes of an array are numbered starting with index 1 for ndarrays.
(c) array() function is used to create only 1D-arrays in NumPy.
(d) The rank of an ndarray tells about the number of elements in each axis.
(e) A list contains elements of heterogeneous type and an array contains homogeneous data.
(f) shape attribute returns the result in the form of tuple.
(g) High positive variance between 2 datasets means they are very dissimilar.
(h) By default, data type of ndarray is int 32.
(i) empty() function creates ndarray with random values.
(j) Using ones function creates an array filled with ones.
(k) Mod function of NumPy returns result in zeros and ones form.
(l) itemsize attribute of ndarray returns the length of each element in bytes.

Answers: (a) True (b) False (c) False (d) False (e) True (f) False
(g) False (h) False (i) True (j) True (k) False (l) False

3. Multiple Choice Questions (MCQs)
(a) Which of the following commands is used to install NumPy?
(i) pip install python-numpy (ii) python install pandas
(iii) python install python (iv) pip install numpy
(b) One-dimensional arrays are also known as:
(i) Vectors (ii) Matrices (iii) List (iv) Tuple
(c) Multidimensional arrays are also known as:
(i) Vectors (ii) Matrices (iii) List (iv) Tuple
(d) Which of the following statements is true?
(i) Size of NumPy array can't be changed but size of list is changeable.
(ii) NumPy array contains elements of homogeneous data type but list can contain heterogeneous type.
(iii) NumPy array is space-efficient than list.
(iv) All of these
(e) Which function is used to reshape an ndarray?
(i) shape() (ii) reshape() (iii) arange() (iv) range()
(f) Which function is used to create an array with random values?
(i) shape() (ii) reshape() (iii) empty() (iv) range()
(g) Which function is used to create an array with zero values?
(i) zeros() (ii) reshape() (iii) empty() (iv) range()
(h) In slicing ([<start>:<stop>:<step>]) operator, what is the default value of step?
(i) 2 (ii) 0 (iii) 1 (iv) –3

(i) If we omit value of axis for concatenate() function, then arrays will be joined:
(i) Row-wise (ii) Column-wise (iii) Will generate error (iv) None of these

(j) A high positive covariance between two datasets shows:
(i) Very strong similarity (ii) Very less similarity
(iii) It does not have any relation with similarity (iv) None of these

(k) Which function is used to find correlation?
(i) polyfit() (ii) corrcoef() (iii) cov() (iv) None of these

(l) Which function is used to find linear regression?
(i) polyfit() (ii) corrcoef() (iii) cov() (iv) None of these

Answers: (a) (iv) (b) (i) (c) (ii) (d) (iv) (e) (ii) (f) (iii)
(g) (i) (h) (iii) (i) (i) (j) (i) (k) (ii) (l) (i)

SOLVED QUESTIONS

1. Why are NumPy arrays used over lists?

Ans. (a) NumPy arrays have contiguous memory allocation. Thus, the same elements stored as list will require more space as compared to arrays.
(b) They are speedier to work with and hence, are more efficient than the lists.
(c) They are more convenient to deal with.

2. Write a NumPy program to get the NumPy version.

Ans.
```
import numpy as np
print(np.__version__)
```

3. Find the output of the following program: [CBSE Sample Paper 2020]
```
import numpy as np
d=np.array([10,20,30,40,50,60,70])
print(d[-4:])
```

Ans. `[40 50 60 70]`

4. Fill in the blank with appropriate NumPy method to calculate and print the variance of an array.
```
import numpy as np
data=np.array([1,2,3,4,5,6])
print(np.___(data,ddof=0)
```

Ans. `print(np.var(data,ddof=0))`

5. Write the output of the following code:
```
import numpy as np
array1=np.array([10,12,14,16,18,20,22])
array2=np.array([10,12,15,16,12,20,12])
a=(np.where(array1==array2))
print(array1[a])
```

Ans. `[10 12 16 20]`

6. Write a NumPy program to test whether none of the elements of a given array is zero.

Ans.
```
import numpy as np
x = np.array([1, 2, 3, 4])
print("Original array:")
print(x)
print("Test if none of the elements of the said array is zero:")
print(np.all(x))
x = np.array([0, 1, 2, 3])
print("Original array:")
print(x)
print("Test if none of the elements of the said array is zero:")
print(np.all(x))
```

Output:

```
Original array:
[1 2 3 4]
Test if none of the elements of the said array is zero: True
Original array:
[0 1 2 3]
Test if none of the elements of the said array is zero: False
```

7. Write a NumPy program to create a 3×3 identity matrix, *i.e.*, diagonal elements are 1, the rest ar
Replace all 0 to a random number from 10 to 20.

Ans.

```
import numpy as np
array1=np.identity(3)
print(array1)
x=np.where(array1==0)
for i in x:
  array1[x]=np.random.randint(low=10,high=20)
print(array1)
```

8. Write a NumPy program to create a 3x3 matrix with the elements between 0 and 9. Check for e
numbers in the matrix and replace them with a random number between 10 and 20.

Ans.

```
import numpy as np
Z = np.arange(9).reshape(3,3)
print(Z)
x=np.where((Z%2)==0)
for i in x:
    Z[x]=np.random.randint(low=10,high=20)
print(Z)
```

9. Write a NumPy program to create a 10x10 matrix in which the elements on the borders will be equal
1, and inside 0.

Ans.

```
import numpy as np
x = np.ones((10, 10))
x[1:-1, 1:-1] = 0
print(x)
```

Output:

```
[[1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 0. 0. 0. 0. 1.]
 [1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]]
```

10. Write a NumPy program to find the number of rows and columns of the given matrix.

Output:

```
Original matrix:
[[10 11 12 13]
 [14 15 16 17]
 [18 19 20 21]]
Number of rows and columns of the said matrix:
(3, 4)
```

Ans.

```
import numpy as np
m= np.arange(10,22).reshape((3, 4))
print("Original matrix:")
print(m)
print("Number of rows and columns of the said matrix:")
print(m.shape)
```

11. Write a NumPy program to compute sum of all elements, sum of each column and sum of each row of a given array.

Output:

```
Original array:
[[0 1]
 [2 3]]
Sum of all elements:
6
Sum of each column:
[2 4]
Sum of each row:
[1 5]
```

Ans.

```
import numpy as np
x = np.array([[0,1],[2,3]])
print("Original array:")
print(x)
print("Sum of all elements:")
print(np.sum(x))
print("Sum of each column:")
print(np.sum(x, axis=0))
print("Sum of each row:")
print(np.sum(x, axis=1))
```

12. Write a NumPy program to convert a given array into a list and then convert it into a list again.

Ans.

```
import numpy as np
a = [[1, 2], [3, 4]]
x = np.array(a)
a2 = x.tolist()
print(a == a2)
```

Output:

```
True
```

16. How can we create a Boolean array?

Ans.
```
import numpy as np
arr1 = np.full((3, 3), True, dtype=bool)
print(arr1)
```
Alternate method:
```
import numpy as np
arr1 = np.ones((3, 3), dtype=bool)
print(arr1)
```
Output:
```
[[ True, True, True]
 [ True, True, True]
 [ True, True, True]]
```

17. Write a NumPy program to extract all odd numbers from an array.

Input:
```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```
Desired output:
```
[1, 3, 5, 7, 9]
```

Ans.
```
import numpy as np
# Input
arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
# Solution
print(arr[arr % 2 == 1])
```
Output:
```
[1, 3, 5, 7, 9]
```

18. Write a NumPy program to replace all odd numbers in an array with –1 without changing it.

Input:
```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```
Desired output:
```
[ 0, -1, 2, -1, 4, -1, 6, -1, 8, -1]
```

Ans.
```
import numpy as np
arr = np.arange(10)
out = np.where(arr % 2 == 1, -1, arr)
print(out)
```

19. Write a NumPy program to convert a 1D array into a 2D array with 2 rows.

Input:
```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```
Output:
```
[[0 1 2 3 4]
 [5 6 7 8 9]]
```

Ans.
```
import numpy as np
arr = np.arange(10)
print(arr.reshape(2, -1))# Setting to -1 automatically decides the number of cols
```

Ans.
```
import numpy as np
x = np.array([0, 1, 3])
y = np.array([2, 4, 5])
print("\nOriginal array1:")
print(x)
print("\nOriginal array2:")
print(y)
print("\nCross-correlation of the said arrays:\n",np.cov(x, y))
```

23. Write a NumPy program to compute Pearson's product-moment correlation coefficients of the two given arrays.

```
Original array1:
[0 1 3]
Original array2:
[2 4 5]
Pearson product-moment correlation coefficients of the said arrays:
[[1. 0.92857143]
[0.92857143 1.]]
```

Ans.
```
import numpy as np
x = np.array([0, 1, 3])
y = np.array([2, 4, 5])
print("\nOriginal array1:")
print(x)
print("\nOriginal array2:")
print(y)
print("\nPearson product-moment correlation coefficients of the said
      arrays:\n",np.corrcoef(x, y))
```

24. Differentiate between covariance and correlation.

Ans.

Covariance	Correlation
Covariance is a measure of how much two random variables vary together.	Correlation is a statistical measure that indicates how strongly two variables are related.
Involves the relationship between two variables or datasets.	Involves the relationship between multiple variables as well.
Lies between –infinity and +infinity.	Lies between –1 and +1.
Measure of correlation provides direction of relationship.	Scaled version of covariance provides direction and strength of relationship.

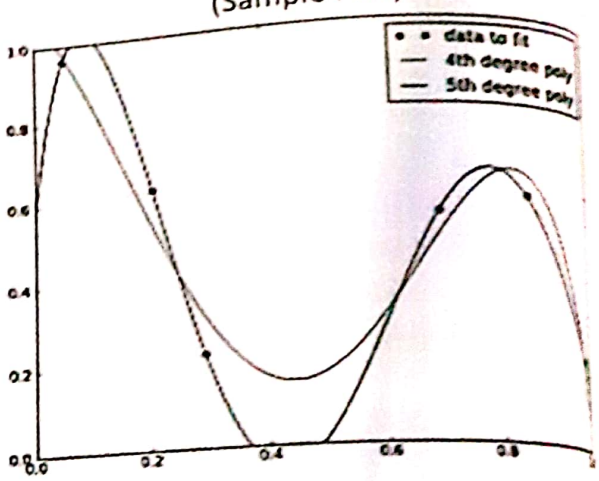
25. Compute the linear relationship between head size and brain weight using the concept of linear regression. The datasets are as under:

Head Size(cm^3)	Brain Weight(grams)
4512	1530

```
0.31  156.91
```

26. Plot linear regression using polyfit() method for 4th and 5th degree polynomial. Extract the plot points using rand() function.

(Sample Plot)



Ans.

```
from numpy import *
import matplotlib.pyplot as plt
# data to fit
x = random.rand(6)
y = random.rand(6)
# fit the data with a 4th degree polynomial
z4 = polyfit(x, y, 4)
p4 = poly1d(z4) # construct the polynomial
z5 = polyfit(x, y, 5)
p5 = poly1d(z5)
xx = linspace(0, 1, 100)
plt.plot(x, y, 'o', xx, p4(xx),'g', xx, p5(xx),'b')
plt.legend(['data to fit', '4th degree poly', '5th degree poly'])
plt.axis([0,1,0,1])
plt.show()
```

27. To test linear relationship of y (dependent) and x (independent) continuous variables, which of the following plots is best suited amongst scatter, histogram or bar chart?

Ans. To test the linear relationship between continuous variables, scatter plot is a good option. We can find out how one variable is changing with respect to another variable. A scatter plot displays the relationship between two quantitative variables.

28. Given two ndarrays:

```
A=np.array([1,2,3,4,5])
B=np.array([3,4,0,-1,-3])
```

Write code to calculate:

(i) Covariance (ii) Correlation (iii) Regression

Ans. *(i)*

```
cov=np.cov(A,B)
```

(ii)

```
cor=np.corrcoef(A,B)
```

(iii)

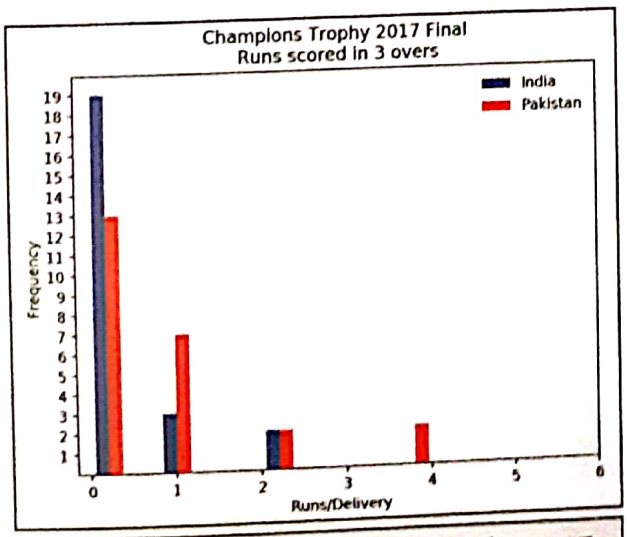
```
np.polyfit(A,B,1)
```

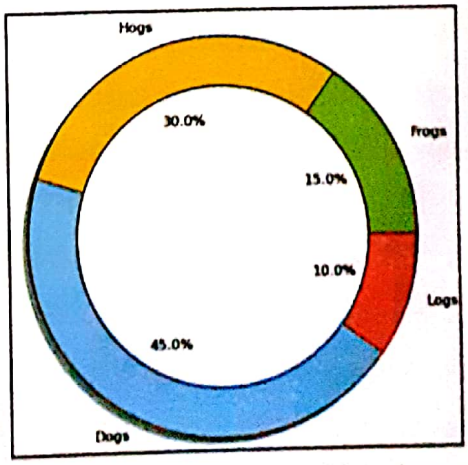
Data Visualization Using Pyplot

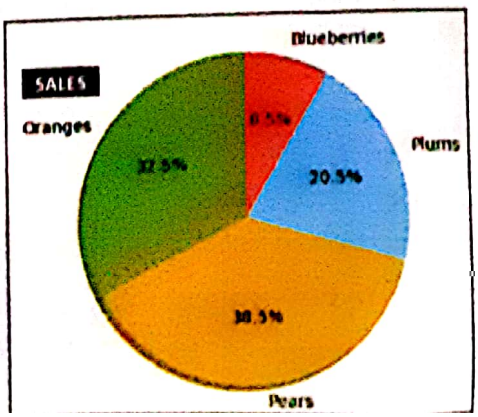
INTRODUCTION

"A picture is worth a thousand words."

all know that images or visuals are a powerful form of communication. We often use them to lerstand a situation better or to condense pieces of information into a graphical representation. ualization is the easiest way to analyze and absorb information. It is the first step for any kind lata analysis work. Visuals or better called *Data Visualization* help us to easily understand a nplex problem and see certain patterns. They also help in identifying patterns, relationships l outliers in data and in understanding business problems better and quickly. Insights gathered m the visuals help in building strategies for businesses. For this reason, data visualization hniques have gained popularity. ***Data visualization*** basically refers to the graphical or visual resentation of information and data using visual elements like charts, graphs, maps, etc.







CTM: A line chart or line graph is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments.

In order to draw a line plot, the steps to be followed are as under:

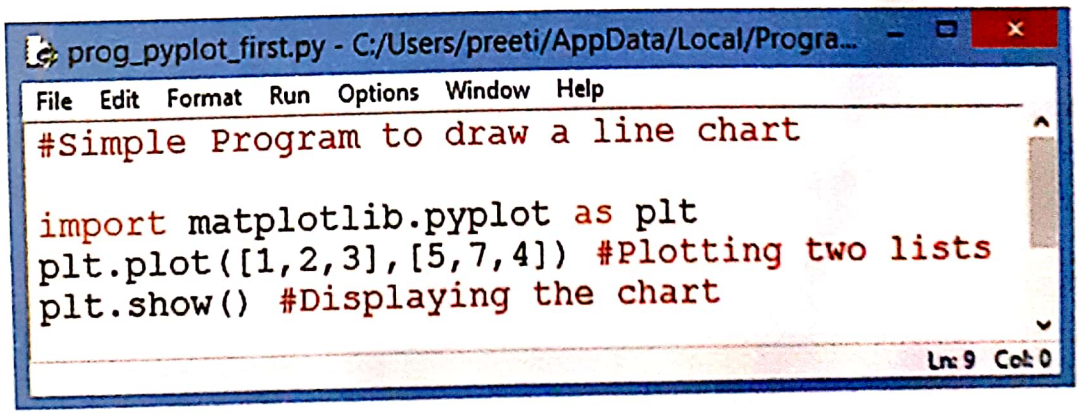
Steps:

1. Importing matplotlib.
2. plt.plot(x, y, color, others) Plot y versus x as lines and/or markers.
3. plt.xlabel("Your Text") Set the X-axis label of the current axes.
4. plt.ylabel("Your Text") Set the Y-axis label of the current axes.
5. plt.set_title("Your Title") Set a title of the current axes.
6. plt.show() Display a figure.

This can be better understood through the Practical Implementation that follows.

Practical Implementation–1

To plot a simple line chart using two lists.



2.8.1 Multiple Plots

If we want to plot multiple lines in one chart, we can simply call the plot() function multipl
times.

Practical Implementation–2

To add legends, titles and labels to a line plot with multiple lines.

prog_line1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_line1.py (3.7.0)

File Edit Format Run Options Window Help

```
#Program to draw two lines along with proper titles and legends
import matplotlib.pyplot as plt
x = [1,2,3]
y = [5,7,4]
plt.plot(x,y,label = 'First line') #Plotting two lists with appropriate label
x2 = [1,2,3]
y2 = [10,11,14]
plt.plot(x2,y2,label= 'Second line')
plt.xlabel('Plot number')
plt.ylabel('Important variables')
plt.title('New Graph')
plt.legend() #involves a legend

plt.show() #Displaying the chart
```

Ln: 17 C

CTM: Legends can be dynamically changed.

Here is a table of the abbreviations used to select colours:

Colour abbreviation	Colour name
b	blue
c	cyan
g	green
k	black
m	magenta
r	red
w	white
y	yellow

Apart from using colour abbreviations, we can also give complete names for colours like yellow, red, blue, etc., and hence, the output shown below is obtained.

Learning Tip: Even if you skip the colour information in **plot()**, Python will plot multiple lines in the same plot with different colours but these colours are decided internally by matplotlib (python).

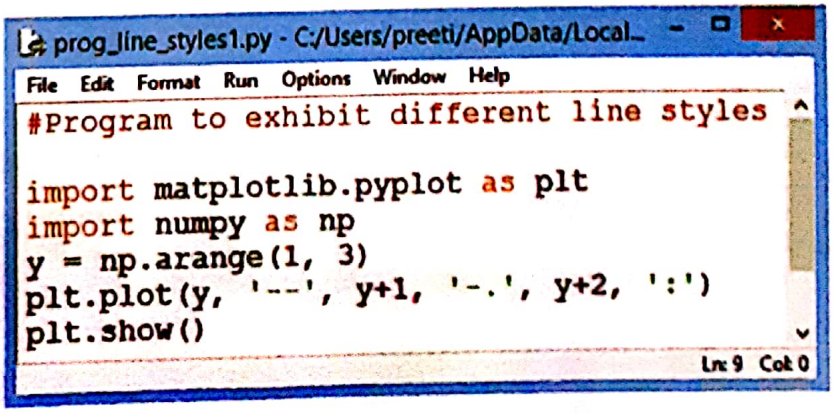
Also, the line styles can also be changed as per the needs of the user. Matplotlib allows us to use different line styles. All the available styles are listed in the following table:

Style abbreviation	Style
-	solid line
--	dashed line
-.	dash-dot line
:	dotted line

All the lines seen until now were proper ones without any dots or dashes. Matplotlib allows us to use different line styles which are implemented in Practical Implementation–9.

Practical Implementation–9

To plot lines with different styles using plot() function.

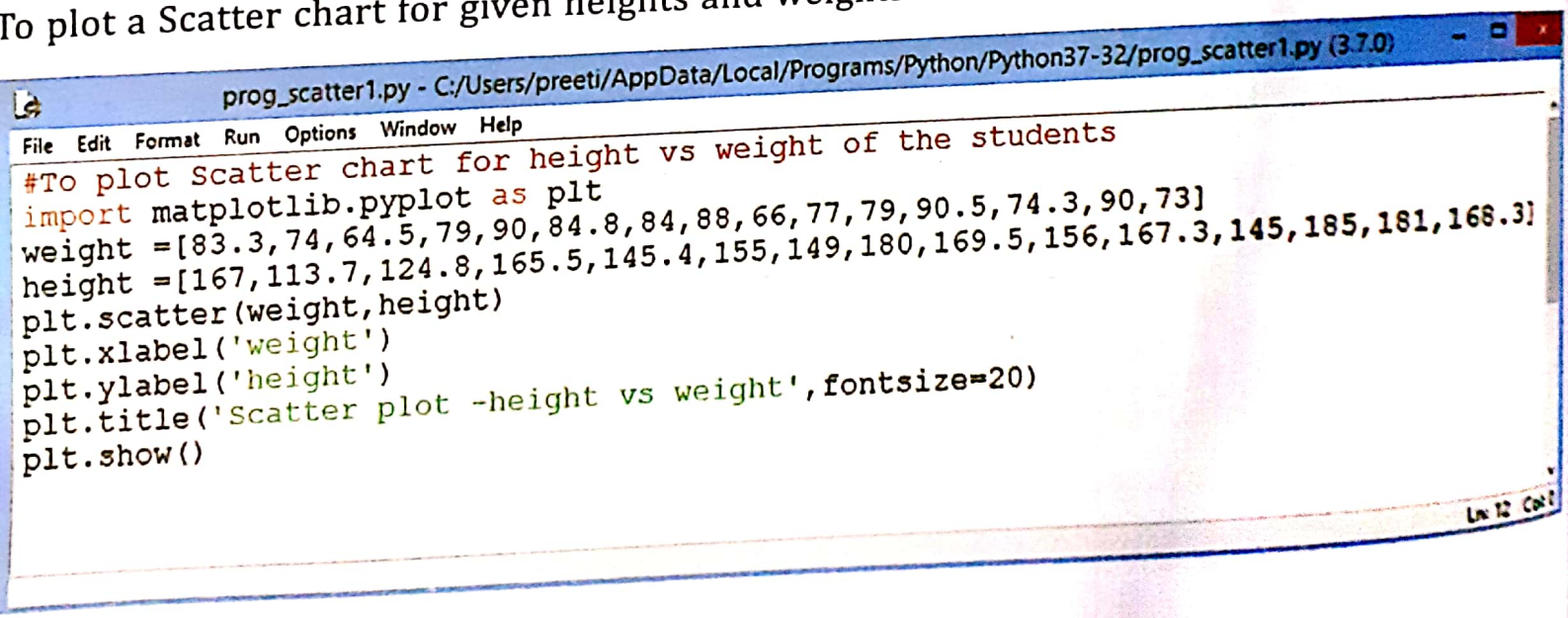


2.9 SCATTER CHART

A scatter plot is a two-dimensional data visualization that uses dots to represent the value
obtained for two different variables—one plotted along the X-axis and the other plotted alon
the Y-axis. The data visualization is done as a collection of points not connected by lines. Each o
them has its coordinates determined by the value of the variables (one variable determines th
X position, the other the Y position). A scatter plot is often used to identify potential associatio
between two variables.

Practical Implementation–10

To plot a Scatter chart for given heights and weights of 15 students.



2.10 BAR PLOT/CHART

A bar chart represents categorical data with rectangular bars. Each bar has a height which corresponds to the value it represents. It is useful when we want to compare a given numeric value on different categories. It can also be used with two data series. The bars can be plotted vertically or horizontally.

A bar chart/bar graph is a very commonly-used two-dimensional data visualization made up of rectangular bars, each for a specific category, with its length representing the value of that category.

Additionally, we can also configure other characteristics for the chart, like width of the bars, colour, etc., among others. The X-axis will be a range with the same quantity of items as the Y-axis. Let us take a simple example where we will store the configurations we want in variables and then will pass them to the bar() function:

To make a bar chart with matplotlib, we need to use the plt.bar() function.

Practical Implementation–14

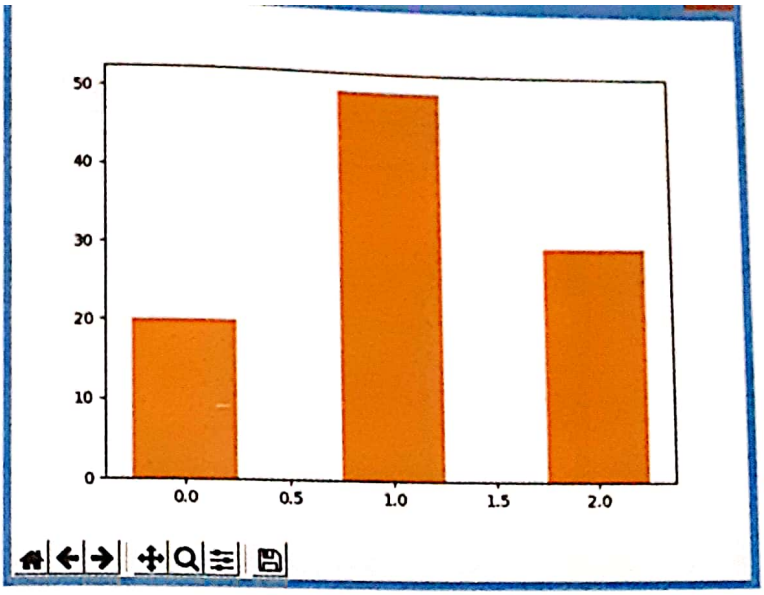
To plot a simple bar chart with orange in colour.

prog_bar2.py - C:/Users/preeti/AppData/Local/Programs/Python/Pyt...

File Edit Format Run Options Window Help

```
#To plot a simple bar chart
import matplotlib.pyplot as plt

# Variables for the bar chart
y_axis = [20,50,30]
x_axis = range(len(y_axis))
plt.bar(x_axis, y_axis, width=.5, color='orange')
plt.show()
```

Ln: 13 Col: 0



Matplotlib charts can be horizontal. To create a horizontal bar chart, type the following code as shown in Practical Implementation-15.

Practical Implementation–15

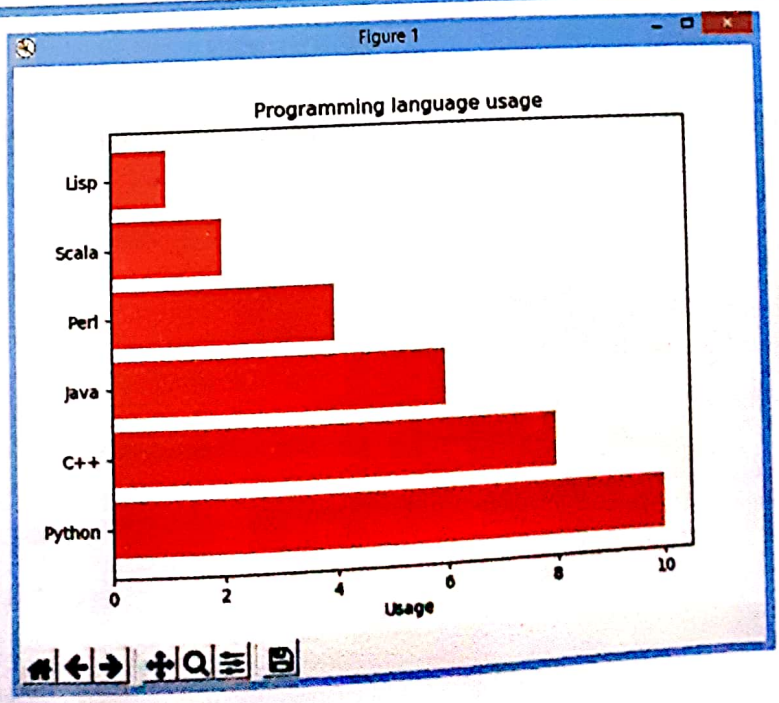
To plot a bar chart horizontally.

prog_bar_horizon.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-3_

File Edit Format Run Options Window Help

```
import numpy as np
import matplotlib.pyplot as plt

objects = ('Python', 'C++', 'Java', 'Perl', 'Scala', 'Lisp')
y_pos = np.arange(len(objects))
performance = [10,8,6,4,2,1]

plt.barh(y_pos, performance, align='center', color='r')
plt.yticks(y_pos, objects)
plt.xlabel('Usage')
plt.title('Programming language usage')

plt.show()
```

Ln: 17 Col: 0



2.11 HISTOGRAMS

A histogram is a powerful technique in data visualization. It is an accurate graphical representation of the distribution of numerical data. It was first introduced by Karl Pearson. It is an estimate of the distribution of a continuous variable (quantitative variable). It is similar to a bar graph.

In other words, we can say that histogram charts are a graphical display of frequencies, represented as bars. They show what portion of the dataset falls into each category, usually specified as non-overlapping intervals called ***bins***.

To construct a histogram, the first step is to "bin" the range of values, *i.e.*, divide the entire range of values into a series of intervals and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often (but are not required to be) of equal size.

- **Difference between a bar chart/graph and a histogram**

  A bar chart majorly represents categorical data (data that has some labels associated with it); they are usually represented using rectangular bars with lengths proportional to the values that they represent. Histograms, on the other hand, are used to describe distributions. Given a set of data, what are their distributions.

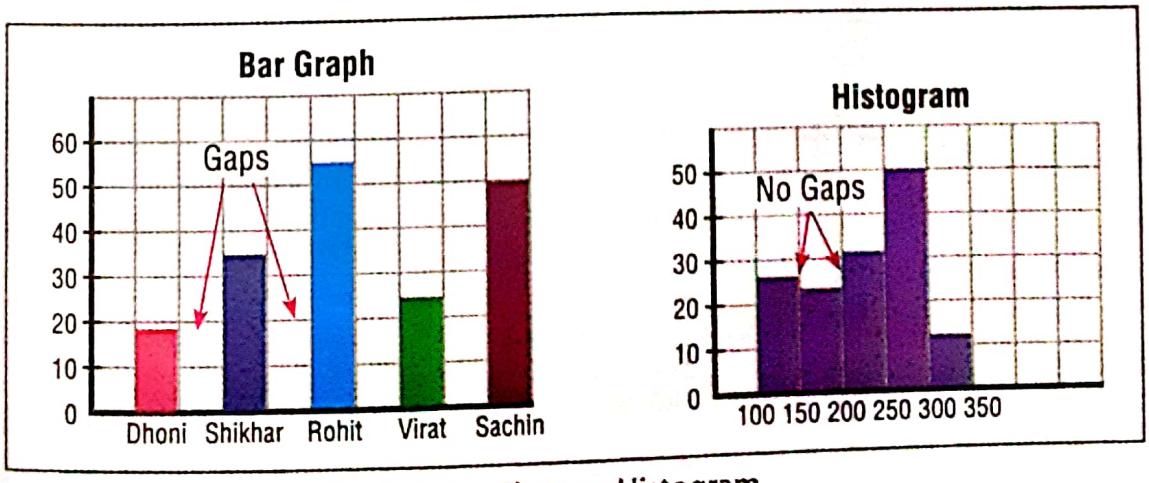


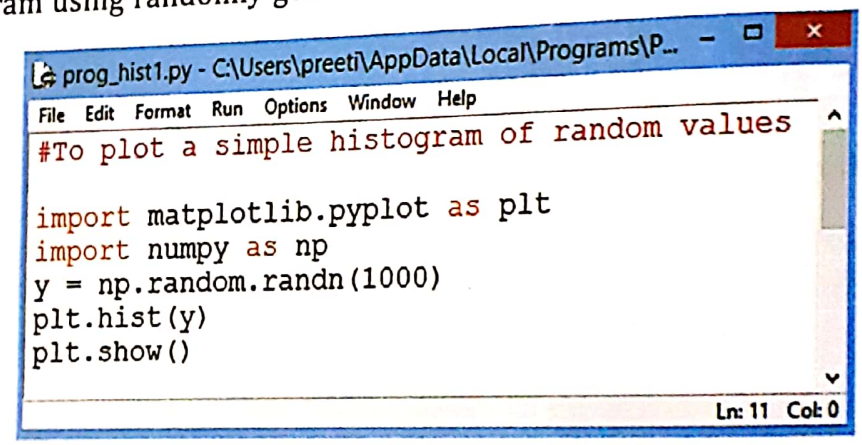
Fig. 2.3: Bar Chart vs Histogram

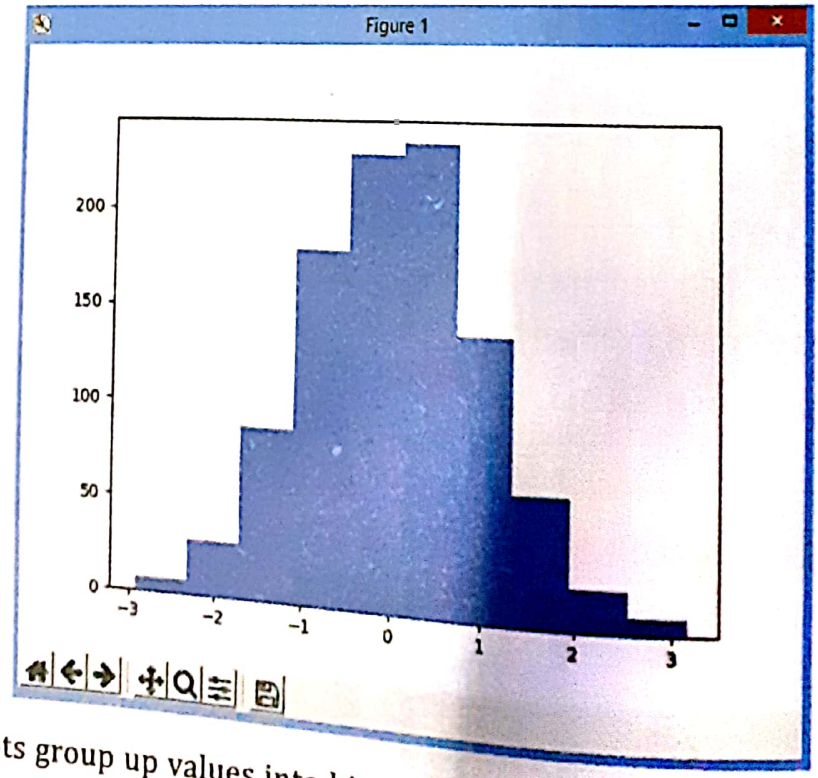
Drawing a histogram in Python is quite simple. ***hist()*** function is called for creating a histogram. Before we start drawing a histogram, a few concepts should be clear in your mind. The components of a histogram plot constitute:

- **Title:** To display heading of the histogram.
- **Colour:** To show the colour of the bar.
- **Axis:** Y-axis and X-axis.
- **Data:** The data can be represented as an array.
- **Height and width of bars:** This is determined based on the analysis. The width of the bar is called bin or intervals.
- **Border colour:** To display border colour of the bar.

Practical Implementation–18

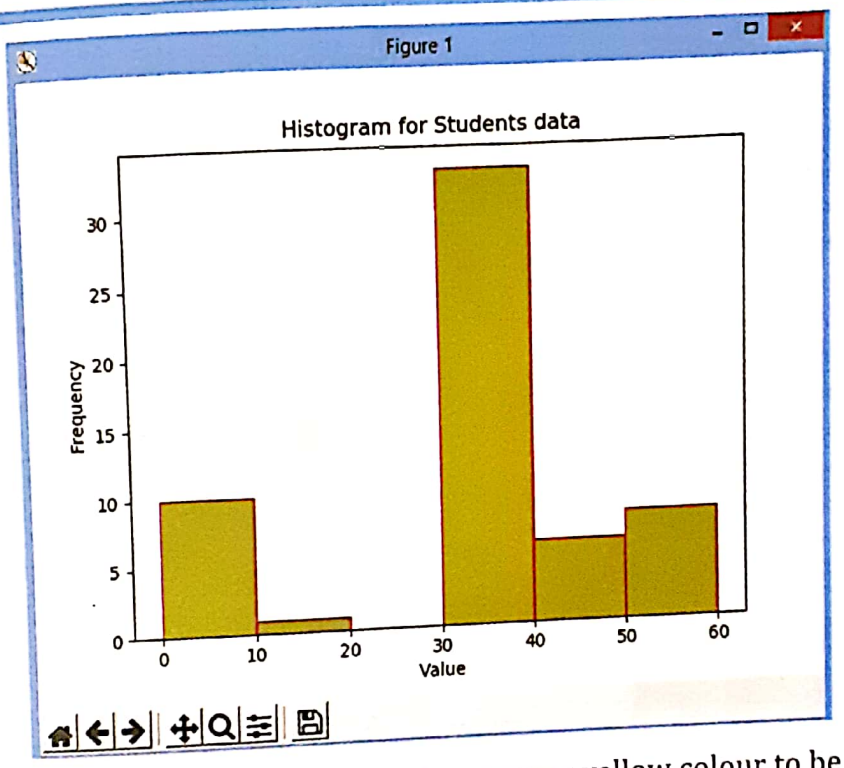
To plot a histogram using randomly generated datasets.





Histogram plots group up values into bins of values. By default, hist() uses a bin value of 10 (so only ten categories, or bars, are computed), but we can customize it, either by passing an additional parameter, *for example*, in hist(y, <bins>), or using the bin keyword argument as hist(y, bin=<bins>).

```
prog_hist_datastud.py - C:\Users\preeti\AppData\Local\Programs\Python\Python37-32\prog_hist_datastud.py (3...
File Edit Format Run Options Window Help
#To plot a histogram with different formatting and suitable title and labels

import numpy as np
import matplotlib.pyplot as plt

data_students=[1,11,21,31,41,51]
plt.hist(data_students,bins=[0,10,20,30,40,50,60],weights=[10,1,0,33,6,8],
         facecolor='y',edgecolor="red")
plt.title("Histogram for Students data")
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.savefig("student.png")
plt.show()
                                                                    Ln: 17  Col: 0
```



In the above code, we are passing 'y' as face colour means yellow colour to be displayed in bars with edges in red colour.

To give a name to the histogram, given statement is to be added before calling show()

```
plt.title("HistogramHeading")
```

For x and y labels, type the given code:

```
plt.xlabel('Value')
plt.ylabel('Frequency')
```

This will generate a histogram accordingly as shown above.

2.12 SAVING PLOTS TO FILE

The active figure can be saved to file using plt.savefig() method.

The histogram created in the above implementation can be saved by clicking on the Save button on the GUI panel. It can also be done using savefig() and typing the following code:

```
plt.savefig("student.png")
```

:m selected 14.4 KB

2.13 FREQUENCY POLYGONS

Frequency polygons are a graphical device for understanding the shapes of distributions. They serve the same purpose as histograms, but are especially helpful for comparing sets of data. Frequency polygons are also a good choice for displaying *cumulative frequency distributions*.

In a frequency polygon, the number of observations is marked with a single point at the midpoint of an interval. A straight line then connects each set of points. Frequency polygons make it easy to compare two or more distributions on the same set of axes.

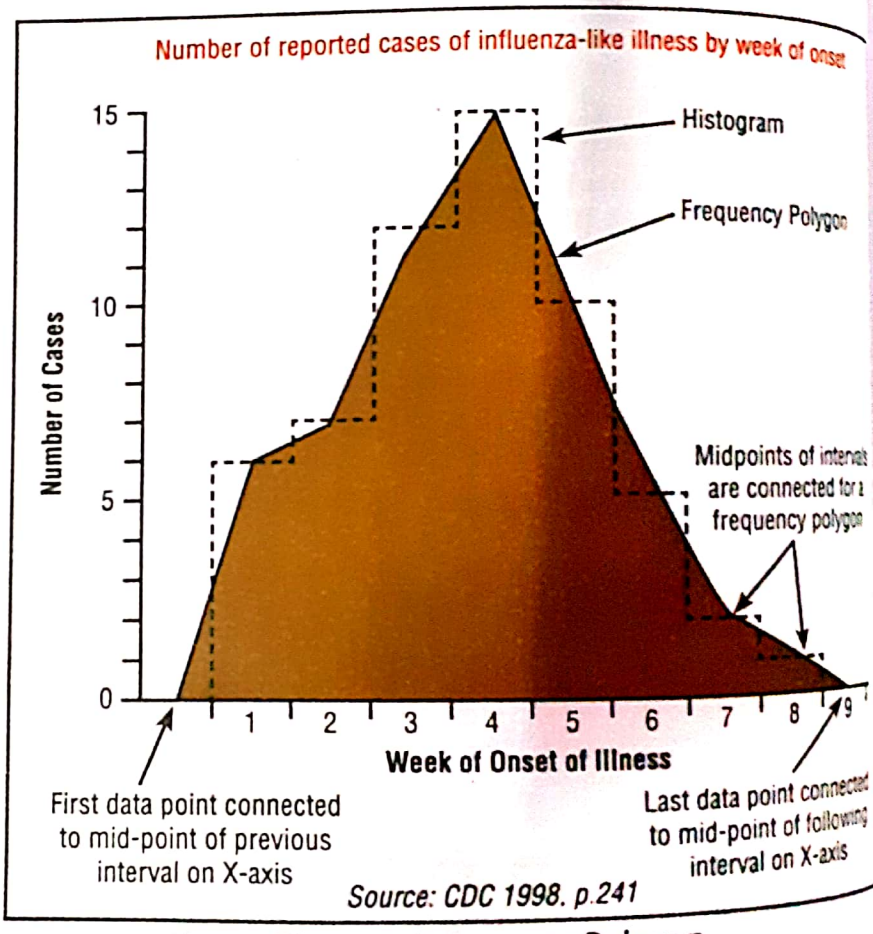


Fig. 2.4: Frequency Polygon

Practical Implementation–23

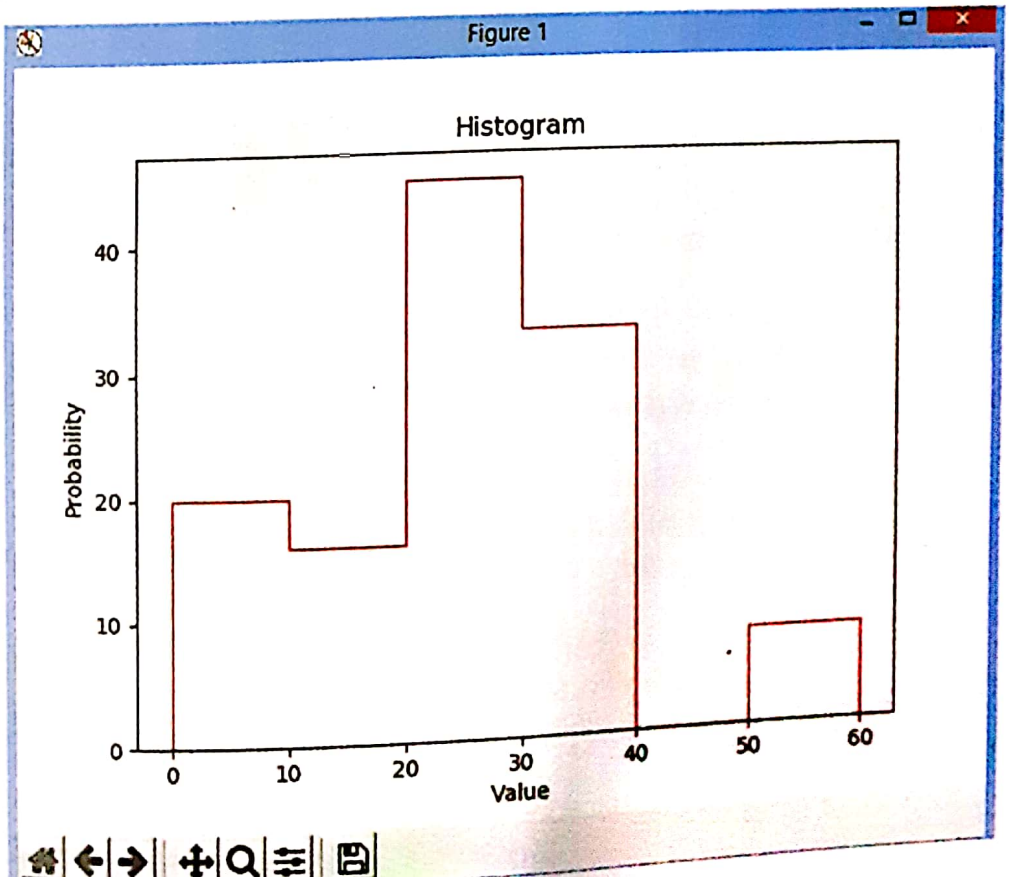
To generate a frequency polygon with reference to the histogram created in Practical Implementation–22.

prog_freq_poly.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_freq_poly.py (3.7.0)

File Edit Format Run Options Window Help

```
#To construct a Frequency polygon on the basis
#of student_data v/s weight histogram

import numpy as np
import matplotlib.pyplot as plt
data = [5,15,25,35,15, 55]
plt.hist(data bins=[0,10,20,30,40,50, 60], weights=[20,10,45,33,6,8],
        edgecolor="red",histtype='step')
plt.xlabel('Value')
plt.ylabel('Probability')
plt.title('Histogram')
plt.show()
```

Ln: 15 Col: 0



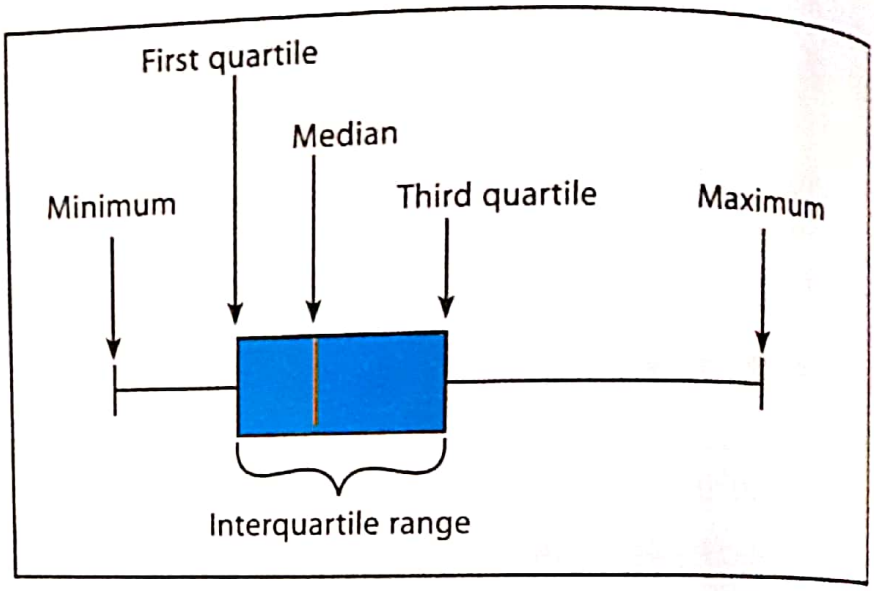
2.14 BOX PLOT

Box plots are descriptive diagrams that help compare the distribution of different series of data. They are *descriptive* because they show measures (*e.g.*, the *median*) which do not assume an underlying probability distribution.

A Box plot is the visual representation of the statistical five-number summary of a given dataset.

A Five-Number Summary includes:

- Minimum
- First Quartile (Q1)
- Median (Second Quartile-Q2)
- Third Quartile (Q3)
- Maximum



➢ What is a Box plot

A box plot, also called a box and whisker plot, is a way to show the **spread** and **centres** of a dataset. **Measures of spread** include the **interquartile range** and the **mean** of the dataset. Measures of centre include the mean or **average** and **median** (the middle of a dataset).

The box and whisker chart shows you how your data is spread out.

Five pieces of information (the "five-number summary") are generally included in the chart:

- The minimum (the smallest number in the dataset). The minimum is shown at the far left of the chart, at the end of the left "whisker."
- First quartile, Q1, is the far left of the box (or the far right of the left whisker).
- The median is shown as a line in the centre of the box.
- Third quartile, Q3, shown at the far right of the box (at the far left of the right whisker).
- The maximum (the largest number in the dataset), shown at the far right of the box.

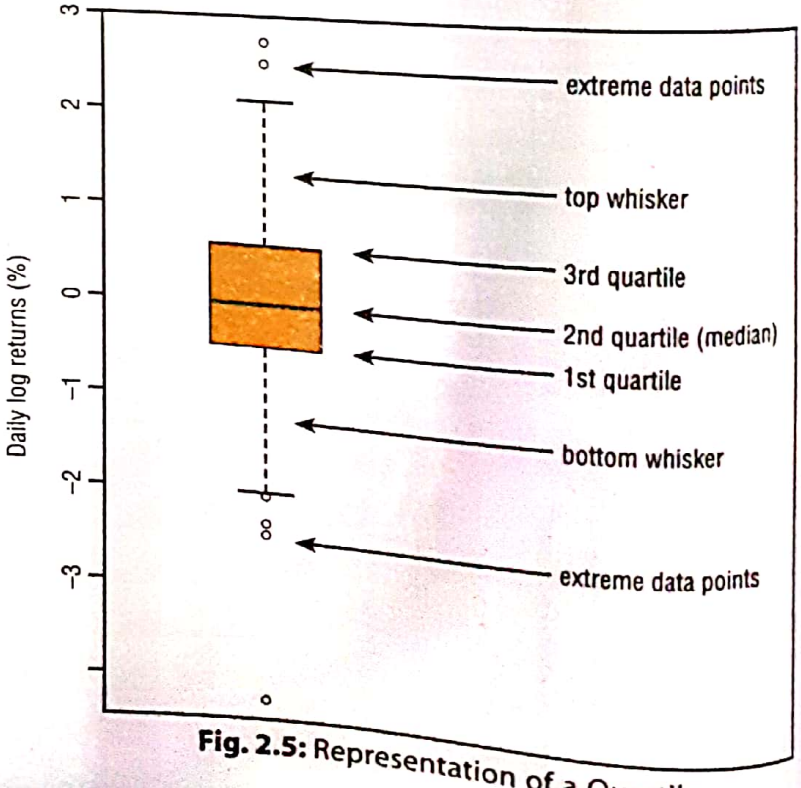
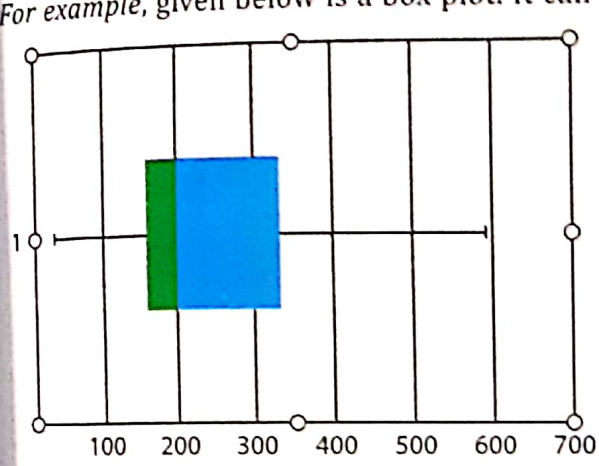


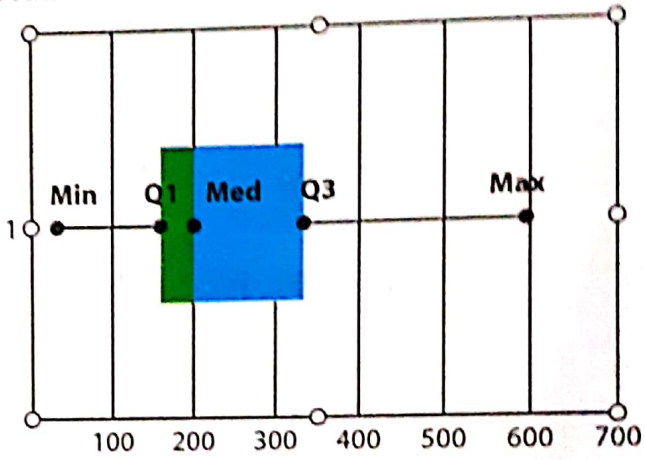
Fig. 2.5: Representation of a Quartile

How to Read a Box Plot

A box plot is a way to show a five-number summary in a chart. The main part of the chart (the "box") shows where the middle portion of the data is: the interquartile range. **The Interquartile Range or IQR is the distance between the Upper and Lower Quartile.** At the ends of the box, you find the first quartile (the 25% mark) and the third quartile (the 75% mark). The far left of the chart (at the end of the left "whisker") is the minimum (the smallest number in the set) and the far right is the maximum (the largest number in the set). Finally, the median is represented by a vertical bar in the centre of the box. All these components to be read can be described as an "**artist**".

For example, given below is a box plot. It can be read and summarized as:





The above image shows a box and whiskers chart with the following information:

- Minimum: 20
- Q_1: 160
- Median: 200
- Q_3: 330
- Maximum: 590

Box plots aren't used that much in real life. However, they can be a useful tool for getting a quick summary of data.

The box plot is generated by using **boxplot()** method, with the data set values, to be passed as an argument to it.

Practical Implementation–24

To plot a simple box plot by passing list of lists.

list1 = [43,76,34,63,56,82,87,55,64,87,95,23,14,65,67,25,23,85]

list2 = [34,45,34,23,43,76,26,18,24,74,23,56,23,23,34,56,32,23]

prog_boxplot.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_boxplo...

File Edit Format Run Options Window Help

```
#To plot a simple boxplot

import matplotlib.pyplot as plt

list1 = [43,76,34,63,56,82,87,55,64,87,95,23,14,65,67,25,23,85]
list2 = [34,45,34,23,43,76,26,18,24,74,23,56,23,23,34,56,32,23]
data = [list1,list2]    #Passing nested lists (lists inside a list)

plt.boxplot(data)
plt.show()
```

Ln: 13 Col: 0

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.

(a) ..... refers to the graphical or visual representation of information and data using visua elements like charts, graphs and maps, etc.

(b) ..... is a collection of methods with matplotlib library which allows the user to construct 2D plots easily and interactively.

(c) The ..... chart is a graph of plotted points on two axes that shows the relationship betwee two sets of data.

(d) ..... is the text that appears on the top of the plot and defines what the chart is about.

(e) The axes of a plot can be labelled using ..... and ..... functions.

(f) A ..... is a summarization tool for discrete or continuous data.

(g) Pyplot module's ..... lets you create histograms.

(h) In a ....., the number of observations is marked with a single point at the midpoint of an interval.

(i) The ..... is used to show the range and the middle half of the ranked data.

(j) The ..... module of Pyplot lets you draw box plots.

(k) Two functions of Pyplot library used to create scatter charts are ..... and ..... .

(l) Barh() function is used to create ..... bar chart.

(m) The area on which actual plot will appear is defined by ..... .

(n) ..... describe the number of data points that fall within a specified range of values.

(o) To change the orientation of the histogram, we can use ..... argument with hist().

Answers: (a) Data Visualization (b) Pyplot (c) scatter
(d) Title (e) xlabel(), ylabel() (f) histogram
(g) hist() (h) frequency polygon (i) box plot
(j) boxplot() (k) plot(), scatter() (l) horizontal
(m) axes (n) Bins (o) orientation

2. State whether the following statements are True or False.
   (a) The matplotlib is a Python interface.
   (b) To save the plot, we have to use save graph() function.
   (c) Frequency polygons are drawn with respect to the histogram created.
   (d) The box plot is also described as five-number summary plot.
   (e) In box plot, the highest and lowest scores are not joined to the box by straight lines.
   (f) Plot can be saved in a pdf format.
   (g) We can specify different colours for different bars of a bar chart.
   (h) To use Pyplot for data visualization, we have to import it by giving import command for matplotlib.pyplot.
   (i) Pyplot is a Python library.
   (j) Frequency polygon is a type of frequency distribution graph.
   (k) To specify a common width for all bars in a bar graph, we have to use thick argument.
   (l) To add a Title to the plot, we have to call function header().
   (m) Markers are data points in the graphs.
   (n) Line style argument of plot() function is not required in scatter chart.
   (o) When we don't specify X or Y limits for a plot, then Pyplot does not automatically decide limits as per values being plotted.

Answers: (a) False (b) False (c) True (d) True (e) False (f) True
(g) True (h) True (i) False (j) True (k) False (l) False
(m) True (n) True (o) False

3. Multiple Choice Questions (MCQs)
   (a) Which Python package is used for 2D graphics?
   (i) matplotlib.pyplot (ii) matplotlib.pip
   (iii) matplotlib.numpy (iv) matplotlib.plt
   (b) The most popular data visualization library in Python is:
   (i) pip (ii) matinfolib (iii) matplotlib (iv) matpiplib
   (c) Matplotlib allows you to create:
   (i) table (ii) charts (iii) maps (iv) infographics
   (d) Which of the following is not a visualization under matplotlib?
   (i) Scatter plot (ii) Histogram (iii) Box plot (iv) Table plot
   (e) Which plot displays the distribution of data based on the five-number summary?
   (i) Scatter plot (ii) Line plot (iii) Box plot (iv) Chart plot
   (f) Which of the following commands is used to install matplotlib for coding?
   (i) import plt.matplotlib as plot (ii) import plot.matplotlib as pt
   (iii) import matplotlib.plt as plot (iv) import matplotlib.pyplot as plt
   (g) Which of the following methods should be employed in the code to display a plot()?
   (i) show() (ii) display() (iii) execute() (iv) plot()
   (h) Which of the following statements is used to create a histogram of 'step' type with 20 bins?
   (i) plt.hist(x, bins=20,histype="barstacked") (ii) plt.hist(x, bins=20)
   (iii) plt.hist(x, bins=20,histype="step") (iv) plt.hist(x, bins=20,histype=hist()
   (i) Which of the following plots makes it easy to compare two or more distributions on the same set of axes?
   (i) Box plot (ii) Histogram (iii) Frequency Polygon (iv) Bar chart

(j) The part of chart which identifies different sets of data plotted on plot by using different colours is called:
(i) legends (ii) title (iii) axes (iv) figure

(k) Which of the following is an incorrect example of savefig() function?
(i) plt.savefig("bar1.pdf") (ii) plt.savefig("bar1.png")
(iii) plt.savefig("bar1.eps") (iv) plt.savefig("bar1.ppt")

(l) Which of the following plots makes it easy to compare two or more distributions on the same set of axes?
(i) Box plot (ii) Histogram (iii) Frequency polygon (iv) Bar chart

(m) Consider the snippet given below:

```
import matplotlib.pyplot as plt
#arr1,arr2 defined here
Colors=['r','b','k','g','m']
Sizes=[50,120,220,150,80]
plt.scatter(arr1,arr2, c=colors, s=sizes, marker="s")
```

With reference to the above code, what will be the shape of marker?
(i) square (ii) circle (iii) star (iv) diamond

(n) With reference to the code in (m), what will be the colour of the last point?
(i) Red (ii) Blue (iii) Black (iv) Magenta

(o) The scatter() function:
(i) is a powerful method of creating scatter plots than plot() function
(ii) can create line graph
(iii) can create bar graph
(iv) None of the above

Answers: (a) (i) (b) (iii) (c) (ii) (d) (iv) (e) (iii) (f) (iv)
(g) (i) (h) (iii) (i) (iii) (j) (i) (k) (iv) (l) (iii)
(m) (i) (n) (iv) (o) (i)

SOLVED QUESTIONS

1. What is Python matplotlib?

Ans. **matplotlib.pyplot** is a plotting library used for 2D graphics in Python programming language. It can be used in Python scripts, shell, web application servers and other graphical user interface toolkits.

2. How can we install matplotlib?

Ans. It is easy to install Python matplotlib library with pip statement – pip install matplotlib.

3. What are the various types of plots offered by matplotlib?

Ans. Matplotlib offers several types of plots:

- Line Graph
- Bar Graph
- Histogram
- Scatter Plot
- Area Plot
- Pie Chart

4. What is data visualization? What is its significance?

Ans. Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. In simple words, Data visualization is the process of displaying data/information in graphical charts, figures and bars.

5. Name the functions you will use to create a (i) line chart, (ii) bar chart, (iii) scatter chart.

Ans. (i) matplotlib.pyplot.plot()
(ii) matplotlibIpyplot.bar()
(iii) matplotlib.pyplot.plot() and matplotlib.pyplot.scatter()

6. Mr. Sanjay wants to plot a bar graph for the given set of values of subject on x-axis and number of students who opted for that subject on y-axis. [CBSE Sample Paper 2020]

Complete the code to perform the following:

(i) To plot the bar graph in statement 1

(ii) To display the graph in statement 2

```
import matplotlib.pyplot as plt
x=['Hindi', 'English', 'Science', 'SST']
y=[10,20,30,40]
_____ Statement 1
_____ Statement 2
```

Ans. (i) `plt.bar(x, y)`

(ii) `plt.show()`

7. Mr. Harry wants to draw a line chart using a list of elements named LIST. Complete the code to perform the following operations:

(i) To plot a line chart using the given LIST,

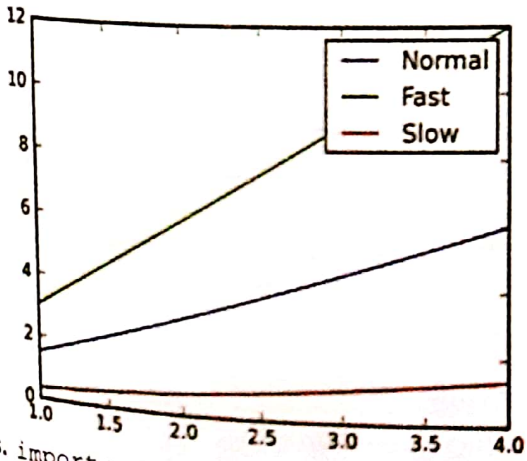
(ii) To give a y-axis label to the line chart named "Sample Numbers".

```
import matplotlib.pyplot as PLINE
LIST=[10,20,30,40,50,60]
_____ Statement 1
_____ Statement 2
PLINE.show()
```

Ans. (i) `PLINE.plot(LIST)`

(ii) `PLINE.ylabel("Sample Numbers")`

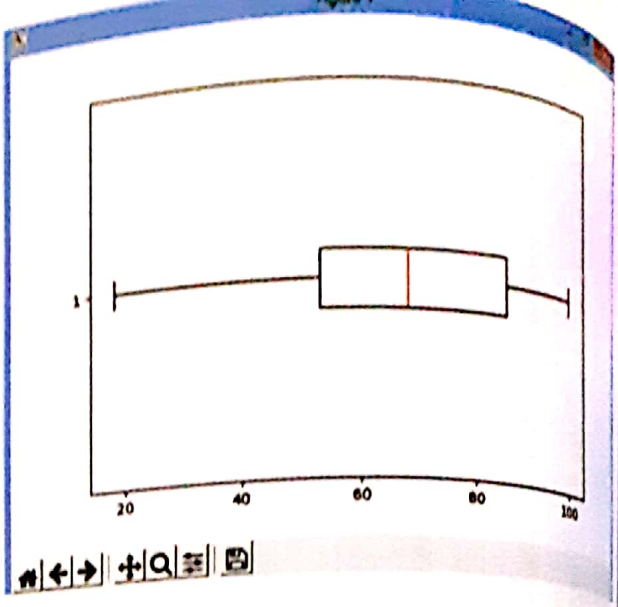
8. Write a code to plot the speed of a passenger train as shown in the figure given below:



Ans.
```
import matplotlib.pyplot as plt
import numpy as np
x = np.arange(1, 5)
plt.plot(x, x*1.5, label='Normal')
plt.plot(x, x*3.0, label='Fast')
plt.plot(x, x/3.0, label='Slow')
plt.legend()
plt.show()
```

17. Create the same box plot as previous question, but change the orientation to horizontal.

Ans.
```
import matplotlib.pyplot as plt
a = [34, 18, 100, 27, 54, 52, 93, 59,
     61, 87, 68, 85, 78, 82, 91]
plt.boxplot(a, vert=False)
plt.show()
```



18. How are bar charts represented using matplotlib?

Ans. Bar charts display rectangular bars (either vertical or horizontal) with their length proportional to the values they represent. They are commonly used to visually compare two or more values. The bar function is used to generate bar charts in Matplotlib. The function expects two lists of values: the coordinates that are the positions of the bar's left margin and the height of the bars.

19. What is the purpose of plot function?

Ans. plot() is a versatile command and takes an arbitrary number of arguments. *For example*, plot (x, y) to plot versus y.

20. Which function is required to plot a bar graph?

Ans. Bar in place of plot(), *i.e.*, bar() function is used to plot a bar graph.

21. List the methods used with pyplot.

Ans. Various methods used with pyplot object are:

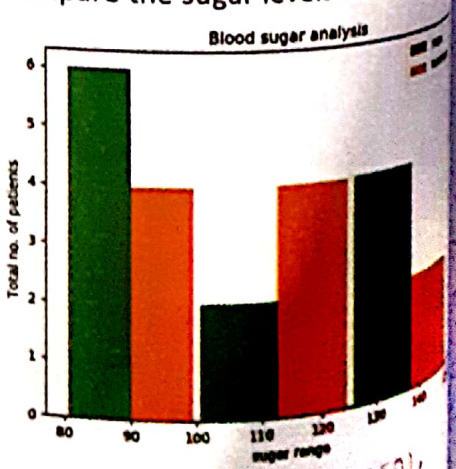
- plot()
- show()
- title()
- xlabel()
- ylabel()
- explode()
- bar()
- hist()
- scatter()
- box plot()

22. Given below are the sugar levels for men and women in a city. Compare the sugar levels amongst the through a histogram.

Men: [113,85,90,150,149,88,93,115,135,80,77,82,129]

Women: [67,98,89,120,133,150,84,69,89,79,120,112,100]

Ans.
```
import matplotlib.pyplot as plt
blood_sugar_men = [113,85,90,150,149,88,93,
                   115,135,80,77,82,129]
blood_sugar_women = [67,98,89,120,133,150,84,
                     69,89,79,120,112,100]
plt.xlabel('sugar range')
plt.ylabel('Total no. of patients')
plt.title('Blood sugar analysis')
plt.hist([blood_sugar_men, blood_sugar_women], bins=[80,100,125,150],
         rwidth=0.95, color=['green','orange'], label=['men','women'])
plt.legend()
plt.show()
```



3.2 PANDAS

Pandas is a Python module that makes data science or data analysis easy and effective. It is the most famous Python package for data science that offers powerful and flexible data structures which make data analysis and manipulation easy. Pandas makes data importing and data analyzing easier.

It is a high-level data manipulation tool developed by **Wes McKinney** and built on packages like NumPy and matplotlib to give us a single and convenient place for data analysis and visualization work.

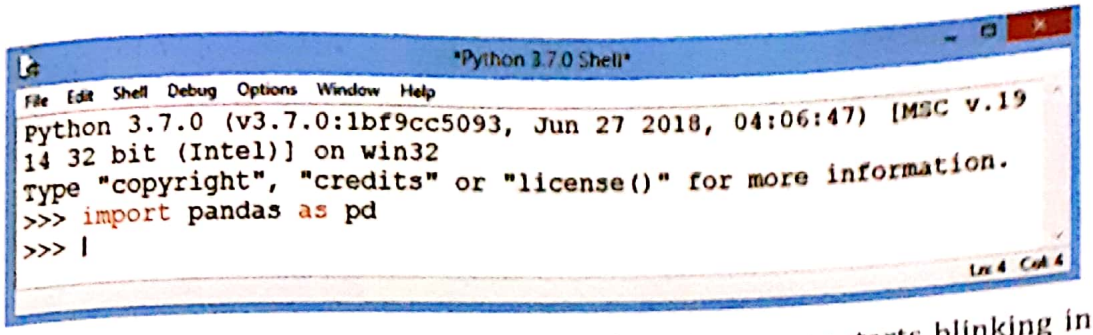
The term Pandas is derived from "Panel data system", which is an ecometric term for multidimensional, structured dataset. Pandas is an open-source Python library which provides high-performance, easy-to-use data structures and data analysis tools. It is used in a wide range of fields including academic and commercial domains which include finance, economics, statistics, analytics, etc.

CTM: Pandas is a software library written for the Python programming language for data manipulation and analysis.

3.2.1 Features of Pandas

Pandas is the most popular library in scientific Python ecosystem for doing data analysis. Pandas can handle several tasks related to data processing and offers the following salient features:

1. It can read or write in many different data formats (integer, float, double, etc.).
2. Columns from a Pandas data structure can be deleted or inserted.
3. It supports group by operation for data aggregation and transformations and allows high performance merging and joining of data.
4. It offers good IO (Input-Output) capabilities as it easily pulls data from a MySQL database directly into a dataframe.
5. It can easily select subsets of data from bulky datasets and can even combine multiple data sets together.
6. It has the functionality to find and fill missing data.
7. It allows us to apply operations to independent groups within the data.
8. It supports reshaping of data into different forms.
9. It supports advanced time-series functionality (which is the use of a model to predict future values based on previously observed values).
10. It supports visualization by integrating libraries such as matplotlib and seaborn, etc. Pandas is best at handling huge tabular datasets comprising different data formats.



As shown in the above figure, after pressing Enter key, the cursor starts blinking in front of Python shell prompt which indicates that Pandas has been installed.

3.4 DATA STRUCTURES IN PANDAS

A data structure is a way of storing and organizing data in a computer so that it can be accessed and worked with in an appropriate way. Pandas provides and deals with the following three data structures:

- **Series:** It is a one-dimensional structure storing homogeneous (same data type) mutable (which can be modified) data.
- **Dataframes:** It is a two-dimensional structure storing heterogeneous (multiple data type) mutable data.
- **Panel:** It is a three-dimensional way of storing items.

We shall discuss Series and dataframes data structures only as Panel is beyond the scope of this book.

3.4.1 Series

Series is like a one-dimensional structure with homogeneous (same type) data. It contains a sequence of values and an associated position of data labels called its index.

For example, the following series is a collection of integers.

49	55	10	79	67

Thus, series in Pandas is like a one-dimensional structure capable of holding data of any type (integer, string, float, Python objects, etc.). A series can also be described as an ordered dictionary with mapping of index values to data values.

Index	Data
0	22
1	–14
2	52
3	100

Index	Data
Jan	31
Feb	28
Mar	31
April	30

Index	Data
'Sunday'	1
'Monday'	2
'Tuesday'	3
'Wednesday'	4

Fig. 3.3: Examples of Series-type Objects

One important point to remember is that the **Series data** is **mutable** (*i.e.*, can be changed). But the **size of Series data** is **immutable**. Once declared, we cannot change the size of a series.

A **Series** in Pandas can be created using '**Series()**' method. Also, any list or dictionary data can be converted into series using this method.

CTM: Series is a one-dimensional labelled structure capable of holding any data type (integers, strings, floating point numbers, Python objects, etc.).

3.10 DATAFRAMES

In the previous sections, we have discussed series in detail. One limitation of series is that it is not able to handle the data in the form of 2D or multidimensional related to real time.

For such tasks, Python Pandas provides another data structure called dataframes.

Dataframe objects of Pandas can store 2D heterogeneous data. **It is a two-dimensional data structure, just like any table (with rows & columns)**. Dataframes are similar to spreadsheets or SQL tables. While working with Pandas, dataframes are the most commonly-used data structures.

The basic features of dataframes are:

(i) Columns can be of different types, *i.e.*, it is possible to have any kind of data in columns.

(ii) Size of dataframe is mutable, *i.e.*, the number of rows and columns can be increased or decreased any time.

Columns

Rows	A	B	C	D	E
a	0	0	0	0	0
b	0	0	0	0	0
c	0	0	0	0	0
d	0	0	0	0	0
e	0	0	0	0	0

(iii) Its data/values are also mutable and can be changed any time.

(iv) Labelled axes (rows/columns).

(v) Arithmetic operations on rows and columns.

(vi) Indexes may constitute numbers, strings or letters.

Dataframes can be created with the following constructs:

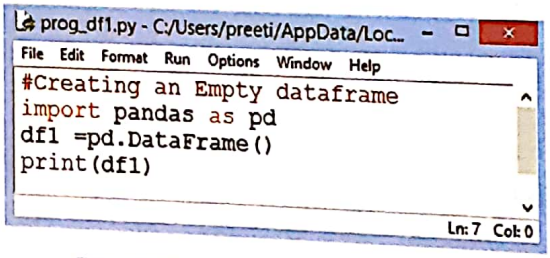
- **Lists**
- **Series**
- **Dictionary**
- **NumPy ndarrays**

Name	Marks	Index
Vijaya	80	B1
Rahul	92	A2
Meghna	67	C
Radhika	95	A1
Shaurya	97	A1

The table represents the data of students with their marks along with the respective grades. The data is represented in rows and columns. Each column represents an attribute (like name, marks, grade, etc.) and each row represents a student.

- **Creation of DataFrame and Display**

  Creating a dataframe begins with the creation of an empty dataframe. To create an empty dataframe, DataFrame() method is used. To display a dataframe, the command `print(<dataframe_name>)` is used.




```
>>>
 RESTART: C:/Users/preeti/
Empty DataFrame
Columns: []
Index: []
>>>
```

Creation of DataFrame Options

The syntax for creating a Pandas dataframe is:

`pandas.DataFrame(data, index, columns, dtype, copy)`

- **data:** data can be represented as series, list, dict, constants or other dataframes.
- **index:** For the row labels, the index to be used for the resulting frame is optional. By default the index value is displayed from 0 to n–1 if no index is passed.
- **columns:** For column labels, the optional default syntax is: np.arrange(n). This is only true if no index is passed.
- **dtype:** dtype is for data type of each column. If no data type is defined, None is applied.
- **copy:** This command is used for copying data if the default is false.

3.17.6 quantile()

The quantile() function returns the value at the given quantiles over requested axis (axis 0 or 1).

We have already discussed quantiles in Chapter 2 under Boxplot.

The word "quantile" finds its origin from the word quantity which means, a quantile is where a sample is divided into equal-sized sub-groups (that's why it is sometimes called a "fractile"). That is also why it means dividing a probability distribution into areas of equal probability.

However, in this section, we shall be discussing quantiles in respect of Pandas dataframes.

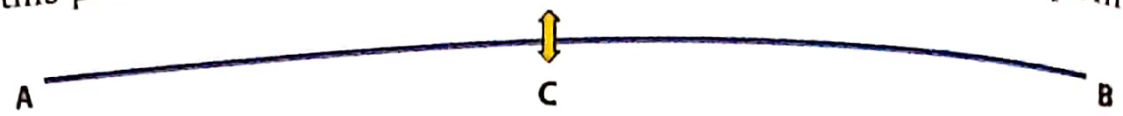
Quantile statistics is a part of a dataset. It is used to describe data in a clear and understandable way. The 0.30 quartile is basically saying that 30% of the observations in our dataset are below a given line. On the other hand, it is also stating that there are 70% remaining above the line we set. Let us understand this in detail.

What are Quartiles

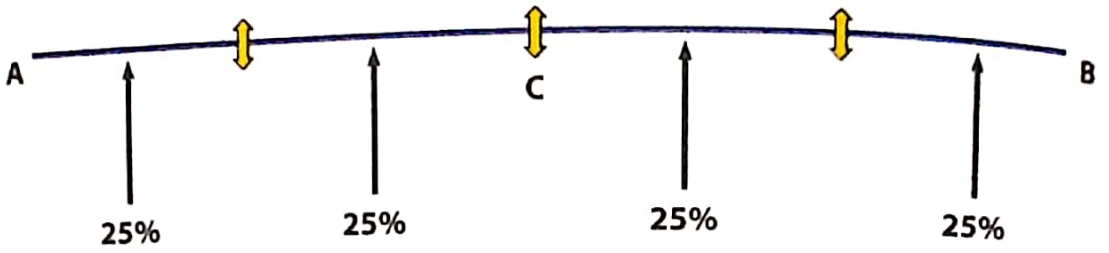
Quartiles in statistics are values that divide data into quarters. However, quartiles aren't shaped like pizza slices; instead they divide your data into four segments according to which the numbers fall on the number line.

Suppose we have to traverse a path from source point A to destination point B. This entire path is to be traced by four persons, one after the other. Therefore, each one has to cover 25% path of the entire stretch so that it is divided equally among the four.

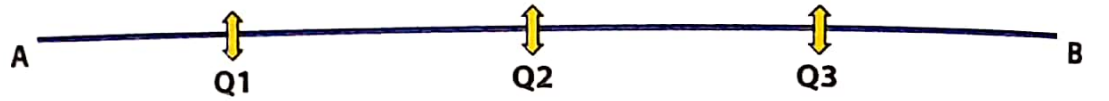
Now, we divide this path into two equal halves, *i.e.*, from the middle, say point C.



- Again, we divide it further between A and C and then between C and B such that each p covers the same distance.



Thus, defining quartile for the divided regions while tracing the path from A to B.



Thus, it has been clearly shown that the series is divided into 4 equal parts.

Q1 is the 1st Quartile (25th Percentile)	Q2 is the 2nd Quartile (50th Percentile) also called median	Q3 is the 3rd Quartile (75th Percentile)

Common Quantiles

Certain types of quantiles are used commonly enough to have specific names. Below is a list of these:

- The 2 quantiles are called the **median**.
- The 3 quantiles are called **terciles**.
- The 4 quantiles are called **quartiles**.
- The 5 quantiles are called **quintiles**.
- The 6 quantiles are called **sextiles**.
- The 7 quantiles are called **septiles**.
- The 8 quantiles are called **octiles**.
- The 10 quantiles are called **deciles**.
- The 12 quantiles are called **duodeciles**.
- The 20 quantiles are called **vigintiles**.
- The 100 quantiles are called **percentiles**.
- The 1000 quantiles are called **permilles**.

Let us now implement quartile using Pandas.

Practical Implementation-32

To implement quantile using Pandas dataframe.

prog_quantile1.py - C:\Users\preeti\AppData\Local\Programs\Python\Python37-32\prog_quantile1.py (3.7.0)

File Edit Format Run Options Window Help

```
#Program on Quantile
import pandas as pd
import numpy as np
df= pd.DataFrame(np.array([[1, 1], [2, 10], [3, 100], [4, 1000]]),columns=['a','b'])
print(df)
print(df.quantile(0.5))  #50% quantile or median
print(df.quantile([.1,.25,.5,.75],axis=0)) #25%,50% and 75% quantile
```

Ln: 11 Col: 0

```
>>>
 RESTART: C:\Users\preeti\AppData\Local
el.py
   a     b
0  1     1
1  2    10
2  3   100
3  4  1000
a     2.5
b    55.0
Name: 0.5, dtype: float64
          a       b
0.10   1.30    3.70
0.25   1.75    7.75
0.50   2.50   55.00
0.75   3.25  325.00
>>>
```

Calculate the Variance of the Specific Column:

```
>>> df.loc[:,"Age"].var() #Variance for specific column
7.7
>>>
```

Ln: 205 Col: 4

The above code calculates the variance of the "Age" column.

3.18 ADVANCED OPERATIONS ON DATAFRAMES

In the previous section, we have discussed aggregate and statistical functions in Pandas and their implementation. In this section, we will discuss the three important advanced operations on dataframes, *i.e.*, **pivoting**, **aggregation** and **sorting**.

Let us start with pivoting first.

ITEM	COMPANY	RUPEES	USD
LED	SAMSUNG	22000	600
LED	LG	25000	700
AC	SONY	15000	650
AC	SONY	18000	800

DATAFRAME

COMPANY	SAMSUNG	LG	SONY
ITEM			
LED	22000	25000	15000
AC	NaN	NaN	18000

PIVOT

Fig. 3.4: Pivoting in Pandas

There are two functions available in Python Pandas for pivoting dataframe:

1. pivot()
2. pivot_table()

1. pivot()

pivot() method creates a new dataframe after reshaping the data based on column values. This method takes 3 arguments—index, columns and values. Out of these three arguments, a minimum of two arguments have to be given.

```
pandas.pivot_table(DataFrame, values=None, index=None,
                   columns=None, aggfunc='mean', fill_value=None,
                   margins=False, dropna=True, margins_name='All')
```

Here, all the arguments are not always required to be given as some arguments have default values also.

- **Dataframe:** This is a Pandas dataframe.
- **values:** This is optional and holds the value of a column to be aggregated.
- **index:** This can be a column, grouper, array or list name.
- **columns:** This is a column, grouper, array or list.
- **aggfunc:** This is an aggregation function.
- **fill_value:** We can set default values using this, if the values are not given.
- **margins:** This is a Boolean attribute having default value as false. In case it is defined as true, then the sum of row and column is taken in the resulting dataframe.
- **dropna:** If this attribute is set as true, then it drops row having missing data.
- **margins_name='All':** If margins is set to true, then it keeps the name of the rows and column of all the dataset.

3.19 SORTING

Sorting means arranging the contents in ascending or descending order. Data of dataframe can be sorted according to values of row and column. By default, sorting is done on row labels in ascending order. There are two kinds of sorting available in Pandas (dataframe):

- By value (column)
- By index

Pandas dataframes provide two useful sort functions:

1. **sort_values():** It sorts the data of given column to the function in ascending or descending order.
2. **sort_index():** This function sorts rows (axis=0) or columns (axis=1).

1. sort_values()—Sorting by Value (Column)

   Sorting over dataframe column's elements is supported by sort_values() method. Its syntax is:

   ```
   DataFrame.sort_values(by = None, axis=0, ascending = True, inplace = False)
   ```

   Here—

   - **by:** Defines column to be sorted.
   - **axis:** Here, passing argument as 0 means sorting will be done row-wise and 1 will perform sorting column-wise.
   - **ascending:** By default, ascending is true.
   - **inplace:** Default is false. If you don't wish to create a new dataframe, then set it as true.

However, we will be using 'by' argument only as per CBSE curriculum.

Here, we will cover three aspects of sorting values of dataframe:

- Sort a Pandas dataframe in either ascending or descending order.
- Sort a Pandas dataframe on the basis of a single column.
- Sort a Pandas dataframe on the basis of multiple columns.

histogram.

3.21 FUNCTION APPLICATION

If we want to apply user-defined function or any library function to manipulate dataframes, Python Pandas provides mainly three important functions, namely pipe(), apply() and applymap().

Use of these functions depends on entire dataframe, row-column elements or respective or individual element.

- Table-wise Function Application: **pipe()**
- Row or Column-wise Function Application: **apply()**
- Element-wise Function Application: **applymap()**

In the following section, we will discuss the use and working of all three functions one by one.

3.21.1 pipe()

This function performs the custom operation for the entire dataframe. In other words, **pipe()** function performs the operation for the entire dataframe with the help of user-defined or library function. In the example given below, we are using pipe() function to add the value '5' to the entire dataframe.

Practical Implementation–47

To perform summation operation on 'Marks' dataframe by adding 5 marks to Science and English marks for all the students (elements of the dataframe).

Data for dataframe:

Science_marks	English_marks
22	89
55	87
63	67
85	55
47	47

In this program, we are implementing table-wise function application using pipe(). This involves creating a user-defined function adder for adding the value 5 to all the numeric elements of the dataframe.

```
prog_pipe1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_pipe1.py (3.7.0)
File Edit Format Run Options Window Help
#Illustrating the use of pipe()
import pandas as pd
import numpy as np
import math
#user defined function
def adder(adder1,adder2):
    return adder1+adder2

#Create a Dictionary of series
d={'science_marks':pd.Series([22,55,63,85,47]),
'english_marks':pd.Series([89,87,67,55,47])}
df=pd.DataFrame(d)
df1=df.pipe(adder,5)
print(df1)
                                                    Ln: 17  Col: 0
```

```
>>>
 RESTART: C:/Users/preeti/AppData/
y
   science_marks  english_marks
0             27             94
1             60             92
2             68             72
3             90             60
4             52             52
>>>
```

As can be observed from the output displayed, 5 marks are added to each element of the dataframe using pipe(). Thus, marks in Science and English are incremented by 5 for each student.

The basic idea behind pipe() function is that if we wish to apply a function to a dataframe or series, which is further applied to another function, which is again applied to some other function, and so on; in such a scenario, it will result in a sandwich-like structure:

```
df=fun3(fun2(fun1(df,arg1=1),arg2=2),arg3=3)
```

This is where pipe() implementation becomes mandatory and here it is implemented as pipelining, like chaining.

Practical Implementation–48

To illustrate pipelining or chaining of dataframe (Modification of Practical Implementation-47)

```
prog_chain1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_chain1.py (3.7.0)
File Edit Format Run Options Window Help
#Implementing pipelining/chaining by performing addition and division
#operation one after the other using pipe()

import pandas as pd
import numpy as np
import math
#user defined functions

def adder(adder1,adder2):
    return adder1+adder2

def divide(adder1,adder2):
    return adder1/adder2

#Create a Dictionary of series
d={'science_marks':pd.Series([22,55]),
'english_marks':pd.Series([89,87])}
df=pd.DataFrame(d)
print(df)
df1=df.pipe(adder,5).pipe(divide,2) #Pipelining/Chaining
print(df1)
```

```
>>>
 RESTART: C:/Users/preeti/AppData
py
   science_marks  english_marks
0             22             89
1             55             87
   science_marks  english_marks
0           13.5           47.0
1           30.0           46.0
>>>
```

In the above program, the two operations of addition and division are performed one after the other as in a pipeline. First, 5 is added to each and every element of marks scored in Science and English, followed by dividing each element of the resultant dataframe (after addition operation) by 2; hence, the output is obtained as shown.

3.21.2 apply()

This function performs the custom operation on the elements row-wise or column-wise. apply() function manipulates the data based on rows or columns of the dataframe.

Practical Implementation–49

To calculate the mean of marks obtained by all the students in both the subjects (Science and English) on the basis of rows.

prog_apply1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_apply_

File Edit Format Run Options Window Help

```
#Row wise Function in python pandas : apply()

import pandas as pd
import numpy as np
import math
d = {'science_marks':pd.Series([22,55]),
'english_marks':pd.Series([89,87])}
df= pd.DataFrame(d)
print(df)
r=df.apply(np.mean,axis=1) #axis=1 implies row-wise manipulation
print(r)
```

Ln: 14 Col: 0

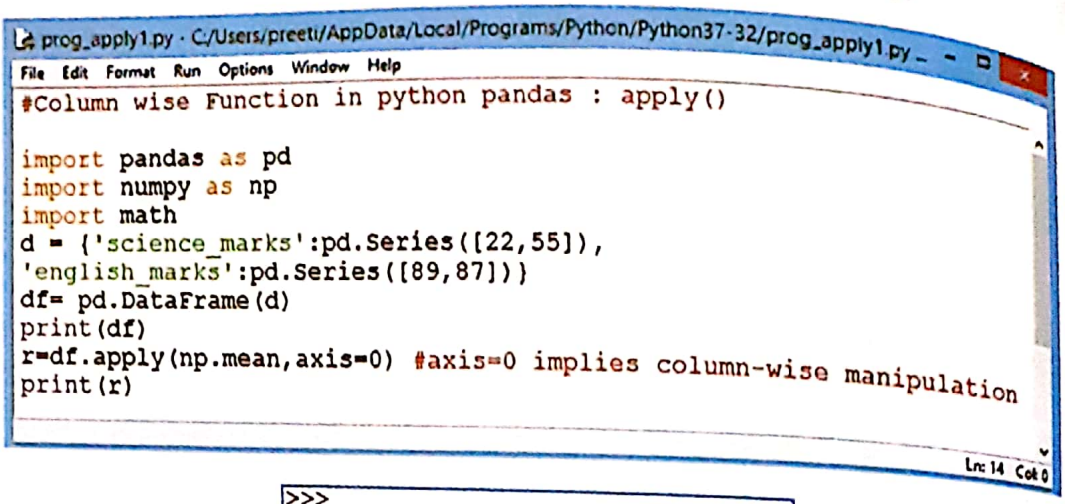
```
>>>
 RESTART: C:/Users/preeti/AppData
py
   science_marks  english_marks
0             22             89
1             55             87
0    55.5
1    71.0
dtype: float64
>>>
```

In the above program, mean of marks obtained in both the subjects is computed by passing argument axis=1 to function apply(), as axis=1 implies row-wise implementation.

Now we will implement apply() column-wise.

Practical Implementation–50

To calculate the mean column-wise for Science as well as English marks.



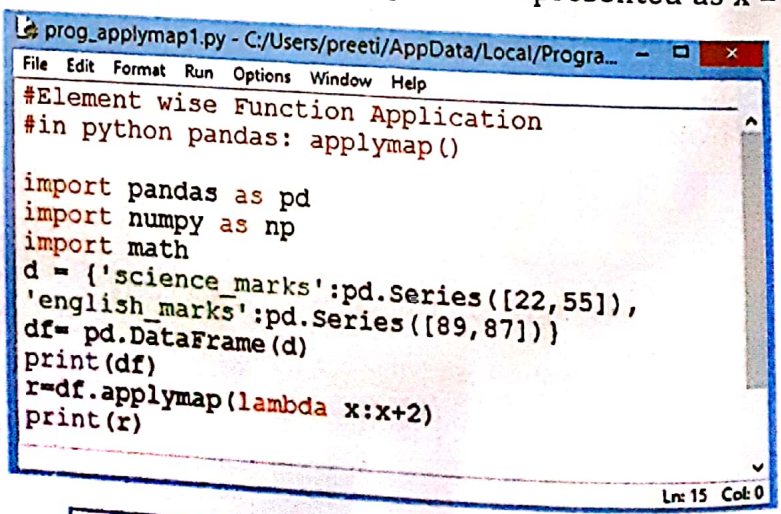

```
>>>
 RESTART: C:/Users/preeti/AppData
py
   science_marks  english_marks
0             22             89
1             55             87
science_marks    38.5
english_marks    88.0
dtype: float64
>>>
```

3.21.3 applymap()

applymap() function is used to perform element-wise Function Application in Python Pandas. applymap() function performs the specified operation for all the elements of the dataframe.

Practical Implementation–51

To illustrate applymap() function for a linear equation represented as x = x + 2.




```
>>>
 RESTART: C:/Users/preeti/AppData/Local
pl.py
   science_marks  english_marks
0             22             89
1             55             87
   science_marks  english_marks
0             24             91
1             57             89
>>>
```

In the above program, each element of the dataframe is manipulated on the basis of the linear expression, x = x+2, thus incrementing each element by a value of 2. Also, we have used a new function—lambda here. Let us discuss in brief what lambda is in Pandas.

Lambda

Python lets us create a function on the go but without really assigning a name to it. These "anonymous" functions are called "lambda functions". One typically writes a lambda function on the fly, when one wants to write a function for one-time use. Lambda functions come in handy in a variety of situations and are of great use in Python's built-in functional commands like map, reduce and filter.

One can see that the lambda function is like the regular function except that it does not have a name, "def ", to define a function and no return statements. The lambda function is written in a single line. One can assign a lambda function to a variable and use it later.

Thus, in the above program, lambda function is referred to for solving the expression on the basis of dataframe for subject marks in Science and English respectively.

3.21.4 groupby() in Pandas

For data analysis, we do segmentations many times using groupby() function in Pandas. For instance, if we wish to know the maximum value from a given dataframe, then grouping is to be done for each column (group).

Table: Data of Dataframe

	name	age	weight	height	runsscored
0	vishal	15	51	5.1	55
1	anil	16	48	5.2	25
2	mayur	15	49	5.1	71
3	viraj	17	51	5.3	53
4	mahesh	16	48	5.1	51

Now, when we perform grouping on the above dataframe, it will be something like:

DATA OF DATAFRAME

	name	age	weight	height	runsscored
0	vishal	15	51	5.1	55
1	anil	16	48	5.2	25
2	mayur	15	49	5.1	71
3	viraj	17	51	5.3	53
4	mahesh	16	48	5.2	51

age	name	weight	height	runsscored
15	vishal	51	5.1	71
16	mahesh	48	5.2	51
17	viraj	51	5.3	53

Maximum weight, height and runsscored in each age group

Here, all the records are grouped together on the basis of age along with application of aggregate function for finding the maximum value among weight, height and runsscored by all the players.

By "groupby" we are referring to a process involving one or more of the following steps:

- Splitting the data into groups based on some criteria.
- Applying a function to each group independently.

Let us implement it in Pandas.

Practical Implementation-52

To implement groupby() by grouping all the records based on age and displaying the maxi... age amongst each age group.

prog_groupby1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_groupby1.py (3.7.0)

File Edit Format Run Options Window Help

```
#Grouping in Pandas- group by

from collections import OrderedDict
from pandas import DataFrame
import pandas as pd
import numpy as np

table=OrderedDict((
("name",['vishal','anil','mayur','viraj','mahesh']),
('age',[15,16,15,17,16]),
('weight',[51,48,49,51,48]),
('height',[5.1,5.2,5.1,5.3,5.1]),
('runsscored',[55,25,71,53,51])
))
d=DataFrame(table)
print("DATA OF DATAFRAME")
print(d)
print(d.groupby('age').max()) #Grouping followed by finding the maximum value
# for each age group
```

Ln: 22 Col: 0

```
>>>
 RESTART: C:/Users/preeti/AppData/Local/Programs
1.py
DATA OF DATAFRAME
     name  age  weight  height  runsscored
0  vishal   15      51     5.1          55
1    anil   16      48     5.2          25
2   mayur   15      49     5.1          71
3   viraj   17      51     5.3          53
4  mahesh   16      48     5.1          51
        name  weight  height  runsscored
age
15    vishal      51     5.1          71
16    mahesh      48     5.2          51
17     viraj      51     5.3          53
>>>
```

It is to be kept in mind that for individual column values, we can use statements like:

```
print(d.groupby('age').max().name)
```

```
>>>
 RESTART: C:/Users/preeti/AppData/Local/Programs
1.py
DATA OF DATAFRAME
     name  age  weight  height  runsscored
0  vishal   15      51     5.1          55
1    anil   16      48     5.2          25
2   mayur   15      49     5.1          71
3   viraj   17      51     5.3          53
4  mahesh   16      48     5.1          51
age
15    vishal
16    mahesh
17     viraj
Name: name, dtype: object
>>>
```

As seen from the output above, after grouping the records on the basis of age, the maximum value for Name gets displayed on the basis of maximum age in each group.

3.21.5 transform()

This function is used to modify values of a dataframe. Transform is an operation used in conjunction with groupby. It is used as per the following given pattern:

Dataframe → grouping → aggregate function on each group value → then transform that value in each group value.

For example, consider the runsscored dataframe; transform() can be expressed as:

DATA OF DATAFRAME

	name	age	weight	height	runsscored
0	vishal	15	51	5.1	55
1	anil	16	48	5.2	25
2	mayur	15	49	5.1	71
3	viraj	17	51	5.3	53
4	mahesh	16	48	5.1	51
0	126				
1	76				
2	126				
3	53				
4	76				

In the above example, sum of score of each age group is applied over in order of age. Let us implement transform() in Pandas.

Practical Implementation-53

To implement transform() based on runsscored dataframe.

prog_transform1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-...

File Edit Format Run Options Window Help

```
#Illustrating transform() in Pandas

from collections import OrderedDict
from pandas import DataFrame
import pandas as pd
import numpy as np

table=OrderedDict((
("name",['vishal','anil','mayur','viraj','mahesh']),
('age',[15,16,15,17,16]),
('weight',[51,48,49,51,48]),
('height',[5.1,5.2,5.1,5.3,5.1]),
('runsscored',[55,25,71,53,51])
))
d=DataFrame(table)
print("DATAOFDATAFRAME")
print(d)
print(d.groupby('age')["runsscored"].transform('sum'))
```

Ln: 22 Col: 0

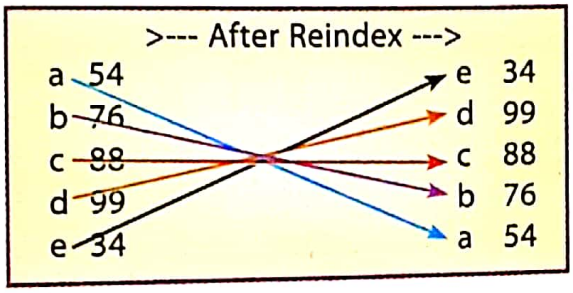
```
>>>
 RESTART: C:/Users/preeti/AppData/Local/Programs
rm1.py
DATAOFDATAFRAME
      name  age  weight  height  runsscored
0   vishal   15      51     5.1          55
1     anil   16      48     5.2          25
2    mayur   15      49     5.1          71
3    viraj   17      51     5.3          53
4   mahesh   16      48     5.1          51
0     126
1      76
2     126
3      53
4      76
Name: runsscored, dtype: int64
>>>
```

One thing to be noted

3.22 REINDEXING AND ALTERING LABELS

Reindexing is a fundamental operation over Pandas series or dataframe. It is a process that makes the data in a series/dataframe conform to a set of labels. It is used by Pandas to perform much of the alignment process. Reindexing in Python Pandas, or we can say changing the order of the rows and columns in Python Pandas dataframe, or changing the order of data of series object is possible with the help of reindex() function.

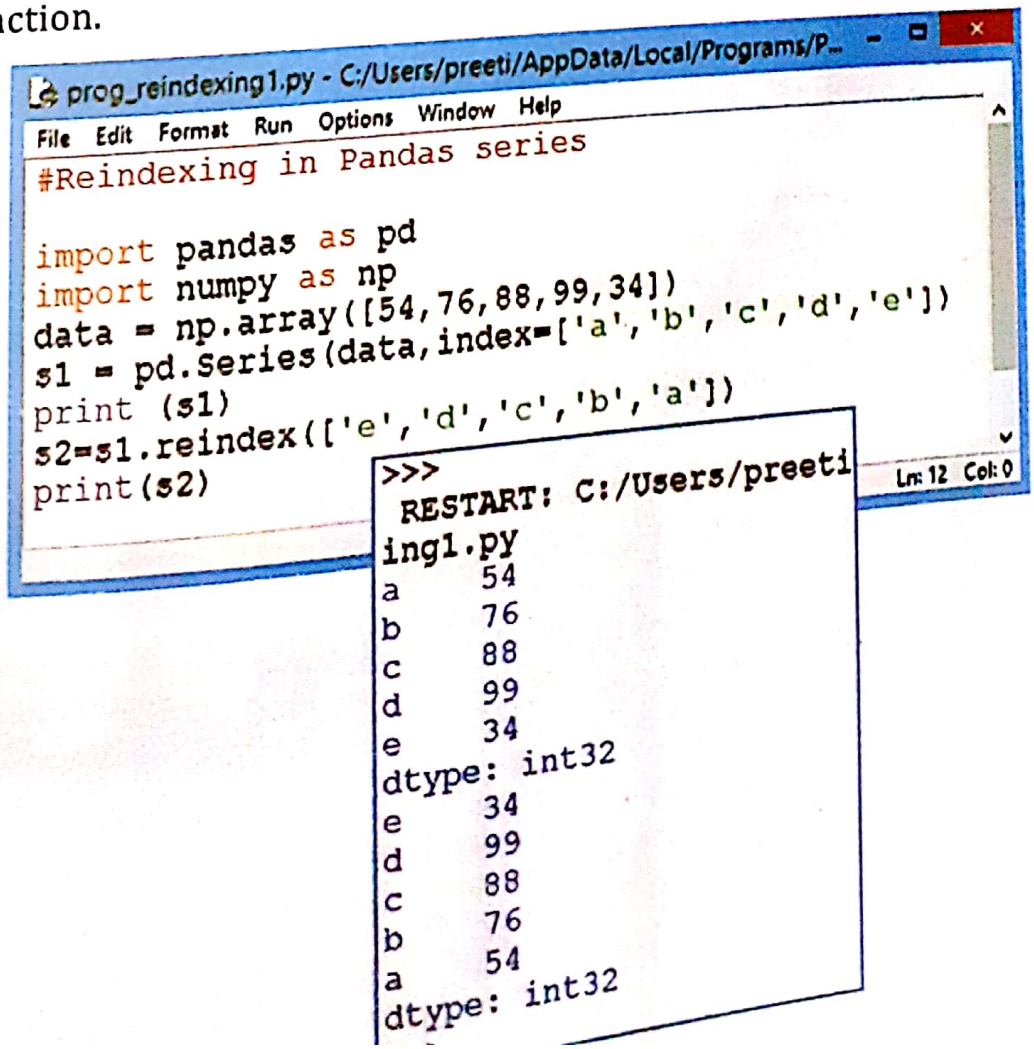
For example, given below is a series for reindexing having first column as label (as index) and second column for value.



Practical Implementation–54

To illustrate reindexing in Pandas series by changing the labels for indices.

The program given below creates a Pandas series with some numeric values, then it is indexed with a, b, c, d, e labels, then again this index is changed to e, d, c, b, a respectively with the help of reindex() function.



MEMORY BYTES

- A series type object is a Pandas data structure that represents a one-dimensional array-like object containing an array of data (of any NumPy supported data type) and an associated array of data labels, called its index.
- A series type object can be created using panda.Series() where data may be specified as a sequence of any type or ndarrays.
- A slice object is created from Series object using a syntax of <Object>[start : end : step].
- Series objects are value-mutable but size-immutable objects.
- The Python Pandas fillna() function is used to fill values with missing values.
- The flexible comparison function always produces the result in Boolean form, *i.e.*, True or False in respective index.
- Dataframe is a two-dimensional array with heterogeneous data, usually represented in tabular format.
- Dataframe has two different index—column-index and row-index.
- The Python Pandas fillna() function is used to fill values with missing values.
- When two dataframe column values are matched with each other, it is called matching.
- Series objects are value-mutable but size-immutable objects.
- The Python Pandas fillna() function is used to fill values with missing values.
- When two dataframe column values are matched with each other, it is called matching.
- Quantiles are points in a distribution that relate to the rank order of values in that distribution. The quantile of a value is the fraction of observations less than or equal to 0.5 by definition.
- We can apply functions to the whole dataframe or subsets of dataframes.
- Data pivoting is a summarizing technique to rearrange the columns and rows in a report, so as to view data from different perspectives.
- Pandas library makes available two functions for pivoting—pivot() and pivot_table() functions.
- With pivot(), if there are multiple entries for a column value pointing to the same index (row), it leads to an error.
- pivot() method creates a new table whose rows and columns are unique.
- pivot() method is used to pivot without aggregation.
- pivot_table() pivots the data by aggregating it, thus it can work with duplicate entries.
- Commonly-used aggregate functions are count(), sum(), mean(), median(), min(), max(), var(), mode().
- A histogram is a plot that lets us discover and show the underlying frequency distribution (shape) of a set of continuous data.
- Pandas' hist() functions lets us create histograms.
- The piping of functions through pipe() basically means the chaining of functions in the order they are executed.
- applymap() applies the function on each individual data element of the dataframe.
- The apply() function manipulates the function argument on individual columns/rows only if the passed function is a Series function. If you pass a single value function, then apply() will behave like applymap().
- The groupby() function rearranges data into groups based on some criteria and stores the rearranged data in a new groupby object. We can apply aggregate functions on a groupby object using agg().
- The aggregation function agg() returns a reduced version of the data by producing one summary result per group.
- transform() returns the transformed version of the summary data by repeating rows for the group to take it some shape on the basis of full data and, thus, the result of transform() can be combined with the dataframe easily.
- The index/column-labels of a dataframe can be changed/reindexed using rename(), reindex() functions.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.

(a) ..... is the most popular open-source Python library used for doing data analysis.
(b) In order to work with Pandas in Python, you need to ..... library in your Python environment.
(c) The two basic and universally-popular data structures of Pandas are ..... and .....
(d) A ..... is a Pandas data structure that represents a one-dimensional array-like object of indexed data.
(e) To create a series object, ... method is used.
(f) To create an empty series object, Series() method is used with ..... .
(g) Missing data in Pandas series and dataframes can be filled with a ..... value.
(h) Dataframe has ..... indices.
(i) Selecting a subset from a dataframe requires ..... function.
(j) ..... methods help in understanding and analyzing the behaviour of data.
(k) ..... is the process of turning the values of a dataset (or a subset of it) into one single value.
(l) ..... function is used to create histograms in Python Pandas.
(m) The ..... function rearranges data into groups based on some criteria.
(n) ..... function is used for finding the highest value from a given set of values or column of a dataframe or a series.
(o) ..... is the process of rearranging the data from rows and columns.
(p) ..... function is used to add all of the values in a particular column of a dataframe.
(q) Passing ..... argument skips the missing values by default.
(r) ..... function calculates the most occurrence of each element along the axis selected.
(s) ..... function in Python Pandas is used to calculate the median or middle value of a given set of numbers.
(t) ..... in statistics are values that divide data into quarters.
(u) The two functions available in Python Pandas for pivoting dataframe are ..... and .....

Answers: (a) Pandas (b) import pandas (c) Series, DataFrame
(d) Series (e) Series() (f) no parameter
(g) NaN-Not a Number (h) two (i) loc
(j) Statistical (k) Data Aggregation (l) Pandas.hist()
(m) groupby (n) max() (o) Pivoting
(p) sum() (q) skipna=True (r) mode()
(s) median() (t) Quartiles (u) pivot(),pivot_table()

(r) True (s) True (t) False (u) True

3. Multiple Choice Questions (MCQs)

(a) Which of the following commands is used to install Pandas?
(i) pip install python-pandas
(ii) pip install pandas
(iii) python install python
(iv) python install pandas

(b) A two-dimensional labelled array that is ordered collection of columns to store heterogeneous data types is:
(i) Series (ii) NumPy array (iii) DataFrame (iv) Panel

(c) In a dataframe, axis-0 is for
(i) Columns
(ii) Rows
(iii) Rows and Columns both
(iv) None of these

(d) Which of the following statements is false?
(i) Dataframe is size-mutable
(ii) Dataframe is values-mutable
(iii) Dataframe is immutable
(iv) Dataframe is capable to hold multiple types of data

(e) Which attribute of dataframe is used to perform the transpose operation on a dataframe?
(i) T (ii) Ndim (iii) Empty (iv) shape

(f) Which attribute of dataframe is used to retrieve its shape?
(i) T (ii) Ndim (iii) Empty (iv) shape

(g) Which attribute of dataframe is used to get number of axis?
(i) T (ii) Ndim (iii) Empty (iv) shape

(h) Which method is used to access vertical subset of a dataframe?
(i) iterrows() (ii) iteritems() (iii) mod() (iv) median()

(i) Which method is used to access horizontal subset of a dataframe?
(i) iterrows() (ii) iteritems() (iii) mod() (iv) median()

(j) Which function is used to find most often appeared value from a set of numbers?
(i) mean() (ii) median() (iii) mode() (iv) count()

(k) Which function is used to change the order of existing indices/labels?
(i) shape() (ii) type() (iii) series() (iv) reindex()

(l) Which function is used to find most often appeared value from a set of numbers?
(i) mean() (ii) median() (iii) mode() (iv) count()

(m) Which function is used to find middle number from a set of numbers?
(i) mean() (ii) median() (iii) mode() (iv) count()

(n) Which function is used to find average value from a set of numbers?
(i) mean() (ii) median() (iii) mode() (iv) avg()

(o) Which of the following statements is true?
(i) Quantiles are points in a distribution that relate to the rank order of the values in their distribution.
(ii) Pivoting technique rearranges data from rows and columns, by possibly aggregating data from multiple sources, in a report form so that data can be viewed in different perspectives.
(iii) Sorting refers to arranging data in a particular order.
(iv) All of these

(p) Which function is used to rename the existing column or index?
(i) std() (ii) hist() (iii) groupby() (iv) rename()

(q) Which plot allows us to discover and show the underlying frequency distribution (shape) of a set of continuous data?
(i) Histogram (ii) Bar (iii) Frequency Polygon (iv) Box plot

(r) Which method applies the function on each individual data element of the dataframe?
(i) apply() (ii) applymap() (iii) hist() (iv) pivot()

Answers: (a) (ii) (b) (iii) (c) (ii) (d) (iii) (e) (i) (f) (iv) (g) (ii)
(h) (ii) (i) (i) (j) (iii) (k) (iv) (l) (iii) (m) (ii) (n) (i)
(o) (iv) (p) (iv) (q) (i) (r) (ii)

SOLVED QUESTIONS

1. What is a Pandas Series?

Ans. Series is a one-dimensional labelled array capable of holding data of any type (integer, string, float, Python objects, etc.).

2. In pandas, S is a series with the following result:

```
S = pd.Series([5, 10, 15, 20, 25])
```

The series object is automatically indexed as 0, 1, 2, 3, 4. Write a statement to assign the series as a, b, c, d, e explicitly.

Ans.
```
S.index = ['a' , 'b', 'c', 'd', 'e']
```

3. What is dataframe?

Ans. Dataframe is a two-dimensional array with heterogeneous data, usually represented in the tabular format. The data is represented in rows and columns. Each column represents an attribute and each row represents a person.

4. Dictionary S_marks contains the following data:

```
S_marks = {'name':['Rashmi', 'Harsh', 'Ganesh', 'Priya', 'Vivek'],
           'Grade':['A1', 'A2', 'B1', 'A1', 'B2']}
```

Write a statement to create dataframe called df. Assume that Pandas has been imported as pd.

Ans.
```
df = pd.DataFrame(S_marks)
```

5. How to check if a DataFrame df has any missing values?

Ans.
```
df.isnull().values.any()
```

6. A dictionary Grade contains the following data:

```
Grade = {'Name':['Rashmi', 'Harsh', 'Ganesh', 'Priya', 'Vivek', 'Anita',
                 'Karthik'],
         'Grade':['A1', 'A2', 'B1', 'A1', 'B2', 'A2', 'A1']}
```

Write statements for the following:

(a) Create a dataframe called Gr.
(b) Find the output of Gr.iloc[0:5] and Gr[0:5]
(c) Add a column called Percentage with following data:
[92, 89, None, 95, 68, None, 93]
(d) Rearrange the columns as Name, Percentage and Grade.
(e) Drop the column (*i.e.*, Grade) by name.
(f) Delete the 3rd and 5th rows.
(g) What does the following will do?
(i) `Gr.drop(0,axis = 0)`
(ii) `Gr.drop(0,axis = "index")`
(iii) `Gr.drop([0,1,2,3],axis = 0)`

Ans. (a)
```
Gr = pd.DataFrame(Grade)
```

(b) Output for both the commands is the same:

```
     Name  Grade
0  Rashmi     A1
1   Harsh     A2
2  Ganesh     B1
3   Priya     A1
4   Vivek     B2
```

```
(c) Gr["Percentage"] = [92, 89, None, 95, 68, None, 93]
(d) Gr = Gr[['Name', 'Percentage', 'Grade']]
(e) Gr.drop('Grade', axis=1)
(f) Gr.drop([2, 4])
```

(g) (i) First row will be dropped. (ii) First row will be dropped.
(iii) First four rows will be dropped.

7. Given a Series that stores the area of some states in km². Write code to find out the biggest and the smallest three areas from the given Series. Given Series have been created like this:

```
Ser1 = pd.Series([34567,890,450,67892,34677,78902,256711,678291,637632,
                  25723, 2367, 11789, 345, 256517])
```

Ans.
```
import pandas as pd
Ser1 = pd.Series([34567,890,450,67892,34677,78902,256711,678291,637632,
                  25723,2367,11789,345,256517])
print("Top 3 biggest areas are:")
print(Ser1.sort_values().tail(3))
print("3 smallest areas are:")
print(Ser1.sort_values().head(3))
```

An alternative code for the above problem would be:

```
import pandas as pd
Ser1 = pd.Series([34567,890,450,67892,34677,78902,256711,678291,637632,
                  25723,2367,11789,345,256517])
print("Top 3 biggest areas are:")
print(Ser1.sort_values(ascending = False).head(3))
print("3 smallest areas are:")
print(Ser1.sort_values(ascending=False).tail(3))
```

Output:

```
Top 3 biggest areas are:
6   256711
8   637632
7   678291
dtype: int64
3 smallest areas are:
12  345
2   450
1   890
dtype: int64
```

8. Given are two objects, a lst object namely lst1 and a Series object namely ser1, both are having similar values, *i.e.*, 2,4,6,8. Find out the output produced by following statements:

(i) `print(lst1*2)` (ii) `print(ser1*2)`

Ans. The output produced by the given statement (i) will be:

```
[2, 4, 6, 8, 2, 4, 6, 8]
```

The output produced by the given statement (ii) will be:

```
0   4
1   8
2   12
3   16
```

Where 0, 1, 2, 3 are the indexes.

9. From the series of areas (given earlier that stores areas of states in km²), find out the areas that are more than 50000 km².

Ans.
```
import pandas as pd
Ser1 = pd.Series([34567,890,450,67892,34677,78902,25617,678291,637632,
                  25723,2367,11789,345,256517])
print(Ser1[Ser1>50000])
```

Output:

```
3     67892
5     78902
7    678291
8    637632
13   256517
dtype: int64
```

10. Write a Python code to create an empty dataframe.

Ans.
```
import pandas as pd
df = pd.DataFrame()
print(df)
```

11. How does dataframe object specify indexes to its data rows?

Ans. It has two indexes—a row index(axis=0) and a column index(axis=1).

12. Write a code statement to list the value of a cell in the 5th row and "item" column from a dataframe "sales"

Ans.
```
sales.item[5]
```

13. Hitesh wants to display the last four rows of the dataframe df and has written the following code :
```
df.tail()
```
But last 5 rows are being displayed. Identify the error and rewrite the correct code so that last 4 rows are displayed. [CBSE Sample Paper 2020]

Ans.
```
df.tail(4)
```

14. Write the command using Insert() function to add a new column in the last place(3rd place) named "Salary" from the list Sal=[10000,15000,20000] in an existing dataframe named EMP already having 2 columns.

Ans.
```
EMP.insert(loc=3, column="Salary", value=Sal)
```

15. Write a Python code to create a dataframe with appropriate headings from the list given below:
```
['S101', 'Amy', 70], ['S102', 'Bandhi', 69], ['S104', 'Cathy', 75],
['S105','Gundaho', 82]
```

Ans.
```
import pandas as pd
# initialize list of lists
data = [['S101','Amy',70],['S102','Bandhi',69],['S104','Cathy',75],
        ['S105','Gundaho',82]]
# Create the pandas DataFrame
df = pd.DataFrame(data, columns = ['ID', 'Name', 'Marks'])
# print dataframe
print(df)
```

16. Write a small Python code to create a dataframe with headings(a and b) From the list given below :
[[1,2],[3,4],[5,6],[7,8]]

Ans.
```
import pandas as pd
df = pd.DataFrame([[1, 2], [3, 4]], columns = ['a','b'])
df2 = pd.DataFrame([[5, 6], [7, 8]], columns = ['a','b'])
df = df.append(df2)
```

17. Who is the main author of Pandas?

Ans. Wes McKinney

18. Write the output for the python statement based on the dataframe cricket provided below.

	Name	Age	Score
0	Sachin	26	87
1	Dhoni	25	67
2	Virat	25	89
3	Rohit	24	55
4	Shikhar	31	47

(i) `df['Age'].quantile([0.25 , 0.5 , 0.75])`

Ans.

0.25	25.0
0.50	25.0
0.75	26.0

(ii) `df[['Age','Score']].applymap(np.mean))`

Ans.

Age	Score	
0	26.0	87.0
1	25.0	67.0
2	25.0	89.0
3	24.0	55.0
4	31.0	47.0

(iii) `df[['Age','Score']].apply(np.mean)`

Ans.

Age	26.2
Score	69.0

(iv) `df.pipe(np.mean)`

Ans.

Age	26.2
Score	69.0

19. Write a Python statement using the above dataframe Student to get a new dataframe with name taken as index and implementing pivoting.

Ans. `df.pivot(index='Name',columns='Subjects').fillna(' ')`

20. Explain the dataframe function transform.

Ans. The transform function transforms the aggregate data by repeating summary result for each row in the group and making the result having same shape as that of the original data. *For example,*

`df1.groupby(by='Tutor')['Classes'].transform(np.mean)`

21. Weight measurement for 16 small orders of French fries (in grams) is given:

78	72	69	81	63	67	65	75	79	74	71	83	71	79	80	69

weight=[78,72,69,81,63,67,65,75,79,74,71,83,71,79,80,69] #common

(i) Create a simple histogram from the above data.

Ans. `plt.hist(weight,bins=16)`

(ii) Create a horizontal histogram from the above data.

Ans. `plt.hist(weight,bins=16, orientation="horizontal")`

(iii) Create a step type histogram for the same.

Ans. `plt.hist(weight,histtype='step')`

22. Write the Python command to create a histogram.

(i) from column having numeric data only in the given dataframe.

Ans. `dataframe.hist()`

(ii) from the specific column (say Category) only

Ans. `dataframe.hist(column='Category')`

23. Write the code fragment to sort the given dataframe (say df and consist of students' details) in descending order of AdmNo.

Ans. `df.sort_values('AdmNo',ascending=False)`

24. Consider the following Python code and write the output for statement S1. [CBSE Sample Paper 2020]

```
import pandas as pd
K=pd.series([2,4,6,8,10,12,14])
K.quantile([0.50,0.75]) ---------- S1
```

Ans.
```
0.50    8.0
0.75   11.0
```

25. Assume following data is stored in dataframe named as df1. Write the commands to:

(i) find total sales per state

(ii) find total sales per employee

(iii) find total sales both employee-wise and state-wise

(iv) find mean, median and min sale state-wise

(v) find maximum sale by individual

Name of Employee	Sales	Quarter	State
RSahay	125600	1	Delhi
George	235600	1	Tamil Nadu
JayaPriya	213400	1	Kerala
ManilaSahai	189000	1	Haryana
RymaSen	456000	1	West Bengal
ManilaSahai	172000	2	Haryana
JayaPriya	201400	2	Kerala

Ans. (i)
```
pv1=pd.pivot_table(dfN,index=['State'],values=['Sales'],aggfunc=np.sum
```

(ii)
```
pv1=pd.pivot_table(dfN,index=['Name of Employee'], values=['Sales'],
aggfunc=np.sum)
```

(iii)
```
pv1=pd.pivot_table(dfN,index=['Name of Employee', 'State'],
values=['Sales'], aggfunc=np.sum)
```

(iv)
```
pv1=pd.pivot_table(dfN,index=['State'],values=['Sales'],aggfunc=[np.
mean,np.min,np.max])
```

(v)
```
pv1=pd.pivot_table(dfN,index=['Name of Employee'],
values=['Sales'],aggfunc=np.max)
```

26. What is Pivoting? Name any two functions of Pandas which support pivoting.

Ans. Pivoting means to use unique values from specified index/columns to form apex of the resulting dataframe pivot() and pivot_table() methods are two functions of Pandas which support pivoting.

27. What is a quartile? How is it related to quantile? How do you generate it in Pandas?

Ans. Quartiles Q1, Q2 and Q3 are three points that divide a distribution into four parts. The four quantiles are called quartiles. With a quantile, the values of the distribution do not need to be multiples of 0.25 while quartiles are multiples of 0.25. In Pandas, we generate these with the function quantile().

28. Consider the following dataframe and answer the questions given below:

```
import pandas as pd
df = pd.DataFrame({"Quarter1":[2000, 4000, 5000, 4400, 10000],
                   "Quarter2":[5800, 2500, 5400, 3000, 2900],
                   "Quarter3":[20000, 16000, 7000, 3600, 8200],
                   "Quarter4":[1400, 3700, 1700, 2000, 6000]})
```

(i) Write the code to find mean value from above dataframe df over the index and column axis. (Skip NaN value)

(ii) Use sum() function to find the sum of all the values over the index axis.

(iii) Find the median of the dataframe df.

Ans. (i)
```
print(df.mean(axis = 1, skipna = True))
print(df.mean(axis = 0, skipna = True))
```

(ii)
```
print(df.sum(axis = 1, skipna = True))
```

(iii)
```
print(df.median())
```

29. **Given a dataframe df1 as shown below:**

City	MaxTemp	MinTemp	Rainfall
Delhi	40	32	24.1
Bengaluru	31	25	36.2
Chennai	35	27	40.8
Mumbai	29	21	35.2
Kolkata	39	23	41.8

(i) Write command to compute sum of every column of the dataframe.
(ii) Write command to compute mean of column **Rainfall**.
(iii) Write command to compute average **MaxTemp**, **Rainfall** for first 5 rows.

Ans. (i) `df1.sum()`
(ii) `df1['Rainfall'].mean()`
(iii) `df1.loc[:11, 'MaxTemp':'Rainfall'].mean()`

30. Give the output for the following code:

```
import pandas as pd
data = [{'a': 1, 'b': 2},{'a1': 5, 'b1': 10, 'c1': 20}]
#With two column indices, values same as dictionary keys
df1 = pd.DataFrame(data, index=['first', 'second'],
                   columns=['a', 'b'])
#With two column indices with one index with another name
df2 = pd.DataFrame(data, index=['first', 'second'],
                   columns=['a', 'b1'])
print(df1)
print(df2)
```

Ans.

	a	b
first	1.0	2.0
second	NaN	NaN

	a	b1
first	1.0	NaN
second	NaN	10.0

31. Given the dataset for weather forecast:

4 Introduction to Software Engineering

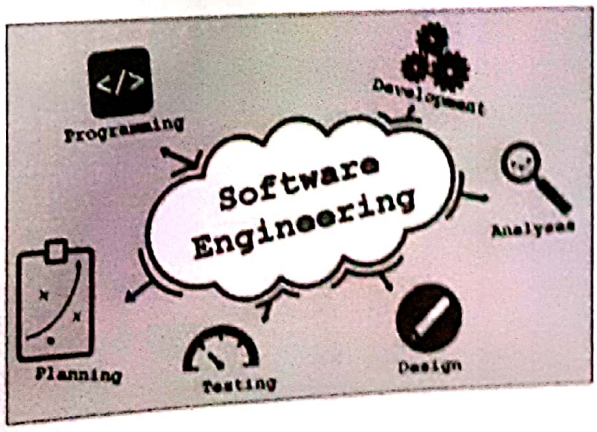
4.1 INTRODUCTION

Computer software are touching all aspects of human life. We encounter software systems in our day-to-day life in a variety of applications such as in business domain, engineering and scientific applications. They may be simple or complex. They may be meant for the internal working of an organization or used in large networks such as railway reservation system, which is one of the most successfully implemented software systems.

Developing a large and complex software application is very challenging in terms of cost and time taken for its completion, which even runs the risk of failure. It, therefore, requires proper planning and a systematic approach to developing a successful and cost-effective application. This is where software engineering plays a significant role as it revolves around building, maintaining and evolving software systems. It includes all activities to manage the project, develop tools, methods and theories that support software production.

This chapter focuses on the basic concepts of software engineering and related topics such as software processes in software development, delivery models, so on and so forth. Let us begin by first discussing what software engineering is all about.

4.2 WHAT IS SOFTWARE ENGINEERING



In simple words, software engineering is a field of engineering for designing and writing programs for computers or other electronic devices. A software engineer, or programmer, creates software (or changes the existing one) and compiles software using methods that can improve it. Today, software surrounds us everywhere—in domestic appliances, communication systems, transportation and in businesses. Software comes in different shapes and sizes—from the program in a mobile phone to the design of a new automobile.

A programmer's main task is to write the source code that makes a program run, whereas a software engineer is responsible for designing, developing and implementing software solutions.

Also, software engineering covers a wider aspect and larger domain as compared to programming. It is about methods, tools and techniques used for developing software. Programming is a component/part of software engineering.

CTM: Software engineering is a branch of engineering where the main focus is on the development of software product using well-defined scientific principles, procedures and methods.

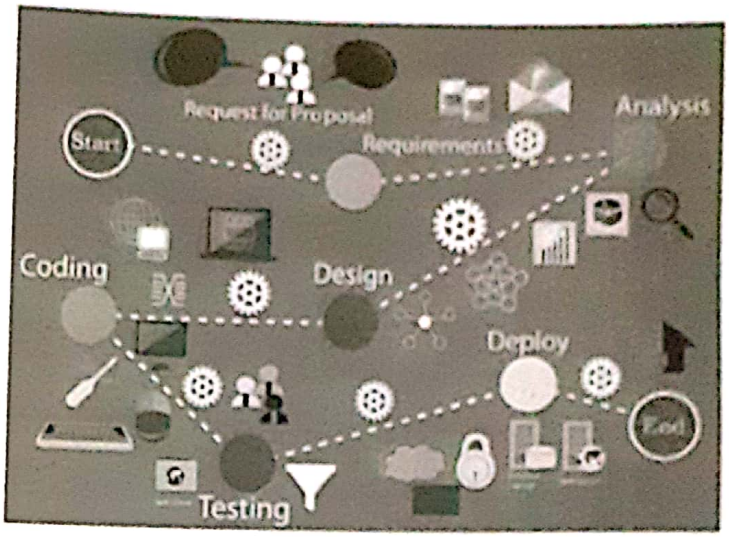
Thus, software engineering is the process of analyzing user needs, designing, constructing and testing end-use applications that will satisfy the needs of the user through software programming language(s). As compared to simple programming, software engineering is used for larger and more complex software systems which are used as critical systems for business or for an organization.

4.2.1 Need for Software Engineering

The need for software engineering arises because of the ever-changing user requirements and environment on which the software works.

- **Large software:** It is easier to build a wall than a house or building. Likewise, as the size of the software becomes large, engineering has to step in to give it a scientific process.
- **Scalability:** If the software process were not based on scientific and engineering concepts, it would be easier to recreate new software than to scale an existing one.
- **To reduce software cost:** As hardware industry develops, large-scale manufacturing lowers the price of computer and electronic hardware. But the cost of software remains high if proper process is not adopted. Thus, through software engineering, programmers plan everything and do away with all those things that are not required. In turn, the cost of software production decreases as compared to a software that does not follow software engineering approach.
- **Dynamic nature:** The ever-growing and adaptive nature of the software hugely depends upon the environment in which the user works. If the nature of the software is always changing, new improvements need to be done in the existing one. This is where software engineering plays an important role.
- **Quality management:** A better software development process provides an improved and quality software product.
- **Reduces complexity:** Large software are always complex and difficult to develop. Software engineering reduces the complexity of any project. Software engineering divides big problems into several small ones and then starts solving each problem one by one. All these small problems are solved independently.
- **Reliable software:** Software should be reliable, which means if you have delivered a software, it should work for at least its given time span or subscription. And if any bugs creep into the software, then the company should fix them. Since testing and maintenance is provided in software engineering, there is no need to worry about its reliability.

4.3 SOFTWARE PROCESS



A software process is a **structured set of activities** required to develop a software system. A software process is a flow chart for developing a software product, which comprises different phases such as gathering requirements, analyzing those requirements, scheduling development phases, checking the developments, implementing changes, etc., and this is carried out till the delivery of the final software product to the post-delivery service methods and more. It is also described as software development process or software development life cycle.

Software process defines a kind of structure that paves the way for the development of a software product. However, there is no ideal software process. It depends upon the product, the organization where it is being developed, and the user for whom it is going to be developed.

A software process (also known as software methodology) comprises a set of related activities that leads to the production of the software. These activities may involve development of the software from scratch or modifying an existing system.

CTM: Software process is the way in which we produce a software. It is also identified as Software Development Life Cycle (SDLC).

4.4 SOFTWARE PROCESS ACTIVITIES

Software process activities are a set of conceptually related activities that are systematically carried out and eventually end up in the production of a software product. This may involve the development of software from the scratch. There are different software processes but all must include four major activities. The four fundamental activities of all software processes are as follows:

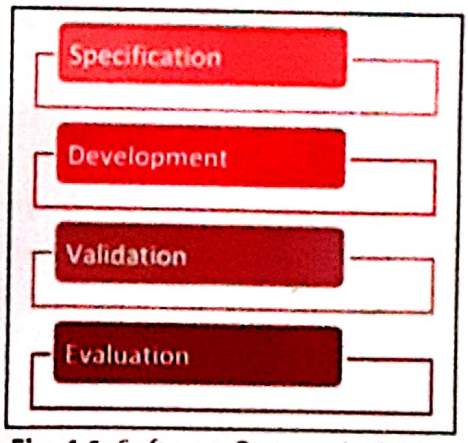


Fig. 4.1: Software Process Activities

- **Software specification:** This phase defines the functionality of the software and constraints involved for developing a product as per the user's requirements and specifications. It describes what the software should do.
- **Software design and development:** This activity constitutes the designing and development of software to meet the user specification. It constitutes programming involved in the creation of a new software.
- **Software validation:** This activity ensures that the software must conform to its specification and meets customer needs.
- **Software evaluation:** The software must be developed in such a manner that it can be modified to meet the changing customer needs.

4.4.1 Software Specification

Software specification, also known as requirements engineering, is the process of understanding and defining what services are required from the system and identifying the constraints in the system's operation and development.

It aims at finding what services are required and what constraints are there to these services and ensures that the software will meet user expectations.

Requirements engineering is a particularly critical stage of the software process as errors at this stage inevitably lead to problems in the system design and implementation later. The requirements engineering process aims at producing an agreed requirements document that specifies the system requirements.

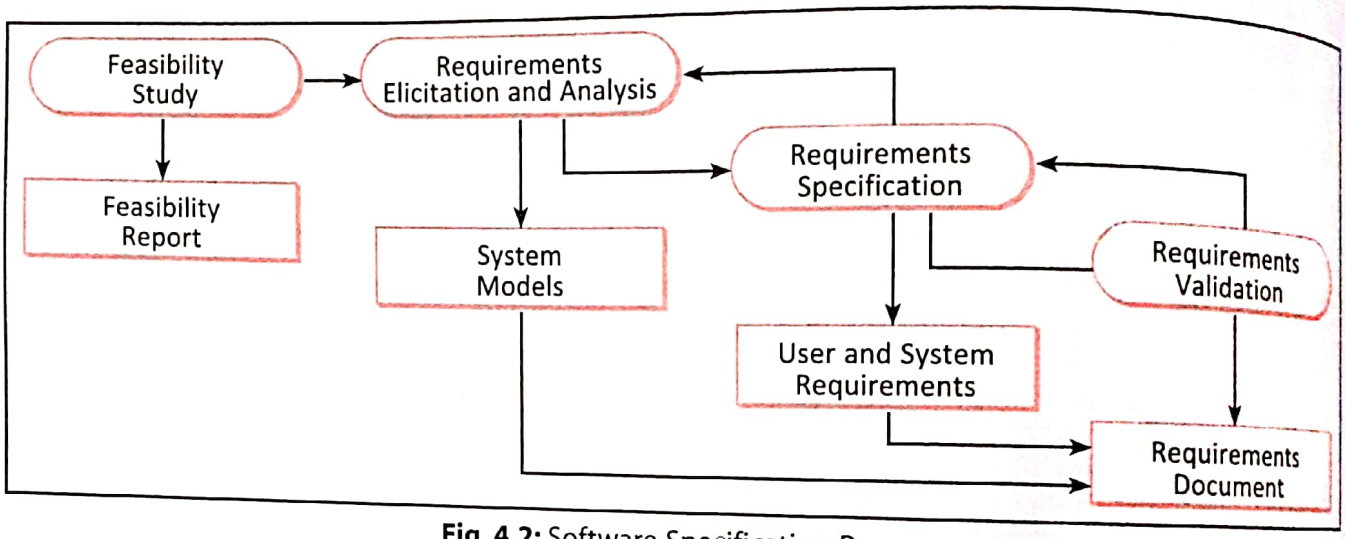


Fig. 4.2: Software Specification Process

Software specification is carried out in four phases:

1. **Feasibility study:** Feasibility study aims at providing an estimate which defines whether the identified user needs can be satisfied or not using the current software and hardware technologies. The study considers whether the proposed system will be cost-effective from a business point of view or not and if it can be developed within the existing budgetary constraints.
2. **Requirements elicitation and analysis:** In this process, system requirements are analyzed through observation of existing systems, discussions with potential users and buyers. This may involve the development of one or more system models and prototypes. These help the system developer understand the system to be specified.
3. **Requirements specification:** Requirements specification is the activity of gathering information through analysis to define a set of user requirements. User requirements are abstract statements of the system requirements for the customer and end-user of the system; system requirements are a more detailed description of the functionality to be provided.
4. **Requirements validation:** This phase checks for consistency and completeness of user requirements. It checks whether all the user requirements have been met or not. During this process, any errors that are discovered are rectified.

4.4.2 Software Design and Development

The implementation stage of software development is the process of converting a system specification into an executable system. It always involves processes of software design and programming.

A software design is a description of the structure of the software to be implemented, the data models and structures used by the system, the interfaces between system components and the algorithms used.

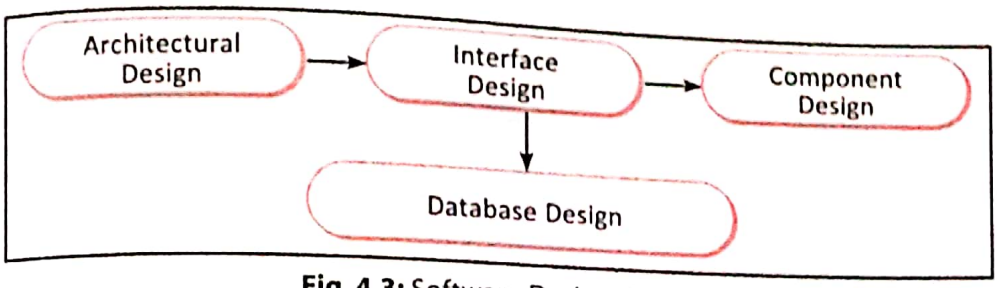


Fig. 4.3: Software Design Process

As is evident from Fig. 4.3, the stages of the design process are sequential. In fact, design process activities are interleaved or interlinked. This leads to feedback from one stage to another in all design processes, and consequent modification is carried out. This phase comprises three main sub-activities along with their specifications at every level:

- **Architectural design**, where the software engineer identifies the overall structure of the system, the principal components (sometimes called sub-systems or modules), relationship between these components and how they are distributed.
- **Interface design**, where the software engineer designs and develops the interfaces required between system components.
- **Component design**, where the software engineer takes each system component and designs the way it will operate.
- **Database design**, where the software engineer designs the system data structures and how these are to be represented in a database. The work here depends on whether an existing database is to be reused or not or a new database is to be created.
- **Programming and implementation**, where on the basis of defined algorithm, database design and determined data structure, the programming is done. This program code is tested for various values of input with expected output and is finally implemented.

4.4.3 Software Validation (Testing)

Software validation, precisely termed as verification and validation (V&V), is intended to show that a system conforms to its specifications and meets the expectations of the customer.

It ensures and validates that the software works as per the proposed design and specifications. For this, testing is performed at three different stages as shown in Fig. 4.4.

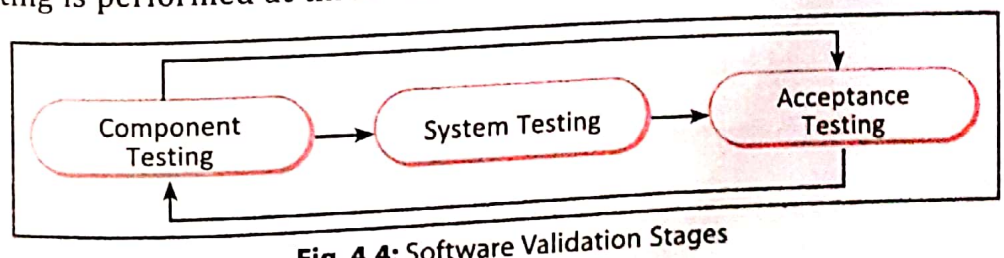


Fig. 4.4: Software Validation Stages

- **Development (or Component) testing:** Components are tested by the team developing the system. Each component is tested independently, without other system components.
- **System testing:** This type of testing involves finding errors that result from interaction between components. However, the components are designed, developed and tested simultaneously. System components are integrated to create a complete system which is again tested together.

- **Acceptance testing:** This is the final stage in the testing process before the system is accepted for operational use. The system is tested with the data supplied by the system customer rather than using the dummy (simulated) data.

4.4.4 Software Evolution/Evaluation

Software evolution means software maintenance. It refers to the process of developing a software initially, then repeatedly updating it for various reasons. The software must be developed in such a manner that it can meet the changing customer needs.

This is the final phase after a software is successfully tested and implemented. The main objective of software evolution is to ensure the reliability and flexibility of the system. The costs of software maintenance are often several times the initial development costs of the software. It caters to the changing needs and requirements of the users such as revised tax slabs or GST rules, etc.

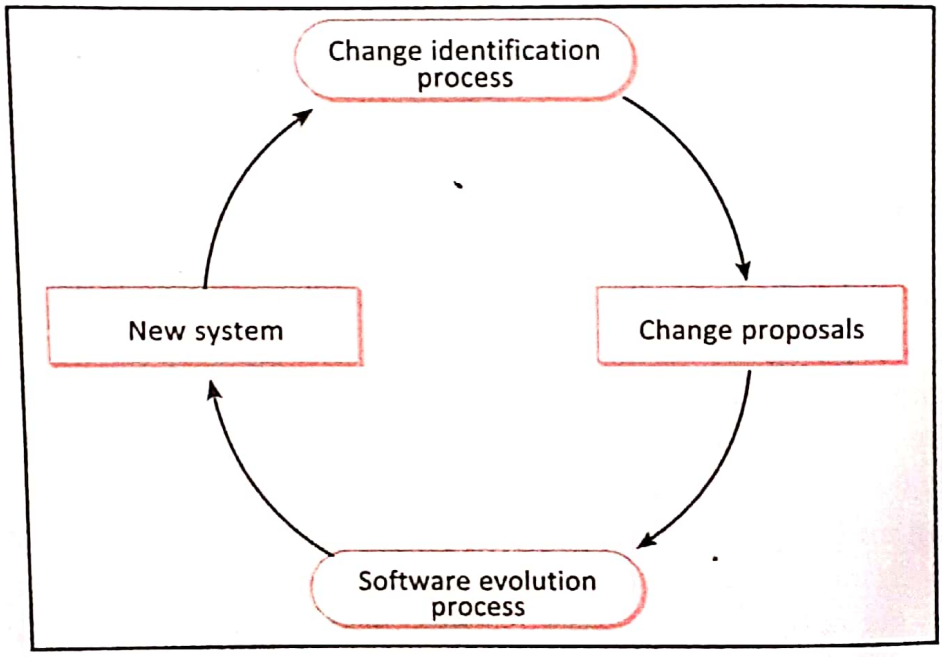


Fig. 4.5: Software Evolution Steps

Software evolution aims at evolving the software at every stage so that it keeps on performing reliably and is flexible to incorporate changes. The most suitable example of software evolution process can be seen nowadays in the form of advanced features added by GMAIL, such as scheduling delivery of an email, increasing the email recall time, sending self-destructing emails, protecting mails through SMS passcode option for end-to-end encryption while sending mails. All these are part of the software evolution process which has been built on the basis of customer's requirements and feedback.

4.5 SOFTWARE PROCESS MODELS

In the previous topic, you have learnt about software processes. Software processes are the ***activities*** involved in producing a software system. Based on this, software process models are ***abstract representations*** of these processes.

CTM: A software process model is an abstract representation of a software process.

Each software process model represents a process from a particular perspective, and thus provides only partial information about that process. In other words, we can say that a software process activity model shows the activities and their sequence but may not show the roles of the people involved in these activities. The widely-used generic software process models are classified into three types:

1. The Waterfall model
2. Evolutionary model
3. Component-based model

These models are often used together, especially for large system development. In case of modern-day large software system, some of the best features of the above three models are used together.

4.5.1 Waterfall Model

The Waterfall model is a sequential approach where each fundamental activity of a process is represented as a separate phase and arranged in a linear order. The Waterfall model is also known as a linear-sequential life cycle model. It is easy to use. In a Waterfall model, you must plan and schedule all activities before starting work on them (plan-driven process). This model consists of a number of phases and each phase must be completed before the next phase begins and there must not be any overlapping in the phases.

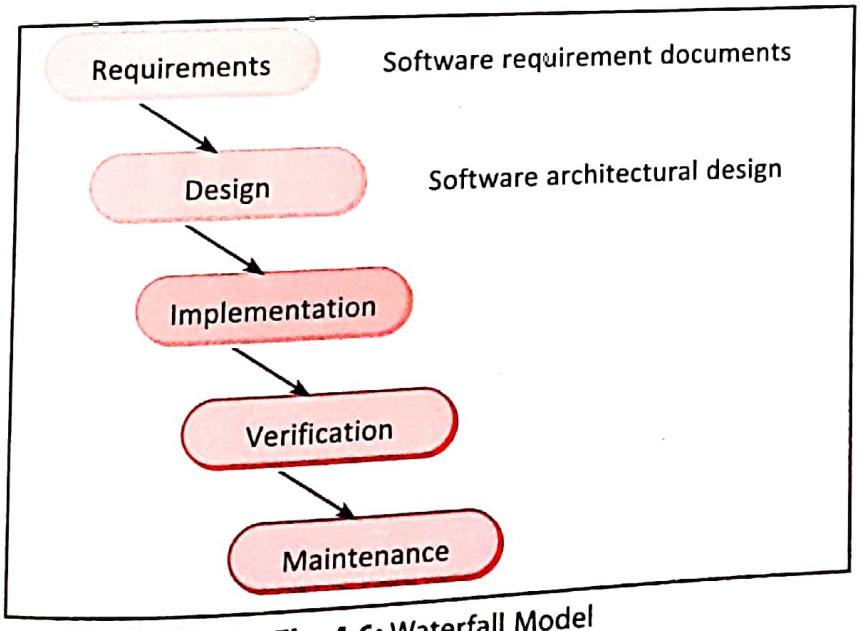


Fig. 4.6: Waterfall Model

The sequential phases in Waterfall model are:

1. **Requirement analysis/specifications:** This is the foremost and the foundation of any development process where all the requirements are taken into account. Requirements specification document must be prepared based on all possible requirements of the software. The outcome of this phase is a document that describes what (not how) a software is supposed to do.
2. **System design:** This phase involves defining the overall system architecture based on the requirements recorded in requirement specification document. Much of the software success depends on this phase.

3. **Implementation and unit testing:** Based on system design, small programs as units are developed and later integrated to work like a software. In this phase, each component is tested along with its development to ensure that it meets the specifications.
4. **Verification:** It comprises integration and system testing. This phase involves testing of the software as a complete system to check whether it fulfils user requirements. After successful testing, the software system is delivered to the customer.
5. **Maintenance:** The last and final stage of Waterfall model is maintenance. In this phase, the system is delivered to the customer. Some issues that may arise at client environment to be resolved are part of the maintenance. Maintenance is an ongoing process and contributes to the regular upgrade of the system.

CTM: The Waterfall model is a sequential and linear model for software development. It comprises several phases in downward direction representing a waterfall. The output of one phase is the input for the next phase.

The Waterfall model is the most appropriate to be used in situations with:

- Well-defined, clear and fixed requirements.
- Stable problem definition. Each phase is clearly understood and laid out.
- Not much changes are needed to be done once the specifications are defined.
- Sequential flow from one phase to the next phase.
- No ambiguous (confusing) requirements.
- The project is small.

However, in certain situations, where a large software system is involved, or in high-risk projects, this model is not suitable and fails to deliver the desired results.

Advantages of Waterfall Model

- Simple, easy to understand and follow
- Easy to arrange and manage the task(s)
- Highly structured, therefore, good for beginners
- Clearly defined stages
- Easy to manage

After specification is complete, low customer involvement is required.

Disadvantages of Waterfall Model

- Inflexible—can't adapt to changes in requirements
- Poor model for long-term projects
- Estimated time and cost for each phase can't be defined well in advance during the development process.
- No working software is developed till the last phase
- Difficult to measure the progress in phases
- Not suitable for complex tasks such as handling big data on social media

V-Model: A Variation of Waterfall model

In the above topic, you have learned about the Waterfall model. The V-model is an extension of the Waterfall model. Usually, this model is pronounced as "Vee" model. This model is quite different from the Waterfall model because, in every phase of it, there is a related testing stage associated. The V-model of SDLC carries out its execution in a sequential manner. The structure it follows takes the shape of letter V. This model is also popularly termed as a Verification and Validation model. Here, each phase has to be finished before beginning the next phase. A sequential design progression is followed like that of the Waterfall model.

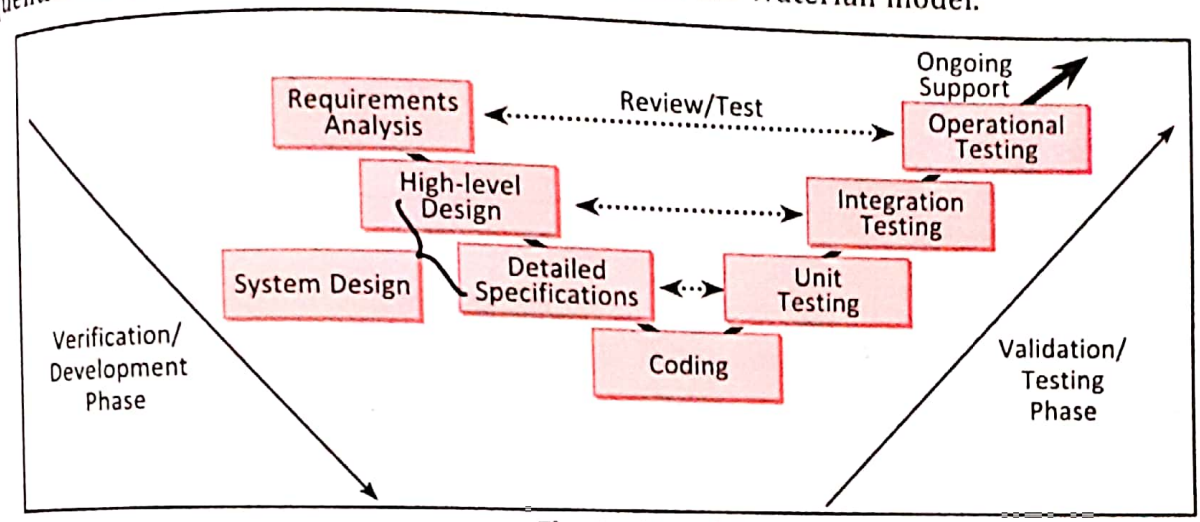


Fig. 4.7: V-Model

V-model is appropriate for use in situations where:

- The requirement is well defined and not ambiguous.
- Acceptance criteria are well defined.
- Project is short-to-medium in size.
- Technology and tools used are not dynamic.

The advantages and limitations of V-model are as follows:

ADVANTAGES	LIMITATIONS
Development and progress is highly organized and systematic.	Not suitable for bigger and complex projects.
Works well for smaller to medium-sized projects.	Not suitable if the requirements are not consistent.
Testing starts from the beginning, so ambiguities are identified from the beginning.	No working software is produced in the intermediate stage.
Easy to manage as each phase has well-defined objectives and goals.	No provision for doing risk analysis, so uncertainty and risks remain.

4.5.2 Evolutionary Model

The Evolutionary model divides the development cycle into smaller 'Incremental Waterfall Model' in which users are able to get access to the product at the end of each cycle. The users provide product feedback for planning to the development team for changing the product, plans or process.

These incremental cycles are typically of two or four weeks' duration and continue till the product is fully developed.

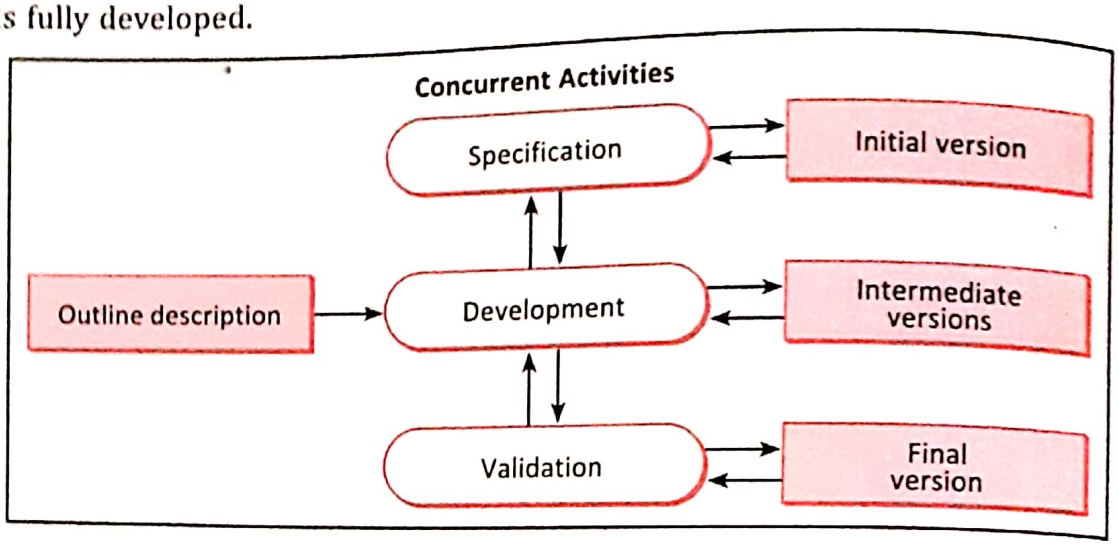


Fig. 4.8: Evolutionary Model

Also, in Evolutionary model, the software development phases involved, such as software specification, development, validation and testing, are inter-related on the basis of customer feedback provided across every stage/activity.

This model is useful for projects using new technology that is not well understood. This is also used for complex projects where all functionality must be delivered at one time, but the requirements are unstable or not well understood at the beginning.

Advantages of Evolutionary Model

- **Error reduction:** The chances of errors get considerably reduced because tests are carried out at each incremental cycle and as a result the risk of failure is reduced.
- **User satisfaction:** Users are given the freedom to experiment with the partially developed system so that they can get complete understanding of the system being developed and suggest changes and modifications wherever required. Therefore, the user is an essential component throughout the development process.
- **High quality:** Quality is maintained following a thorough testing at each and every phase.
- **Low risk:** There is significant reduction in risk as a version is implemented.
- **Reduction in cost:** It reduces the cost by providing structured and disciplined steps.

Disadvantages of Evolutionary Model

- **Multiple versions:** The developer has to make a table of different versions developed.
- **Difficult to divide software:** It is difficult to "divide" the software and the problems in several versions.
- **Uncertain customer needs:** A confused user has uncertainty over his requirements.
- **Time and cost:** Though this model reduces "time and cost", if requirement is not gathered correctly, it may increase overheads in terms of time and cost later.

Following are the evolutionary process models:

(a) Prototyping model

(b) Spiral model

(c) Concurrent development model

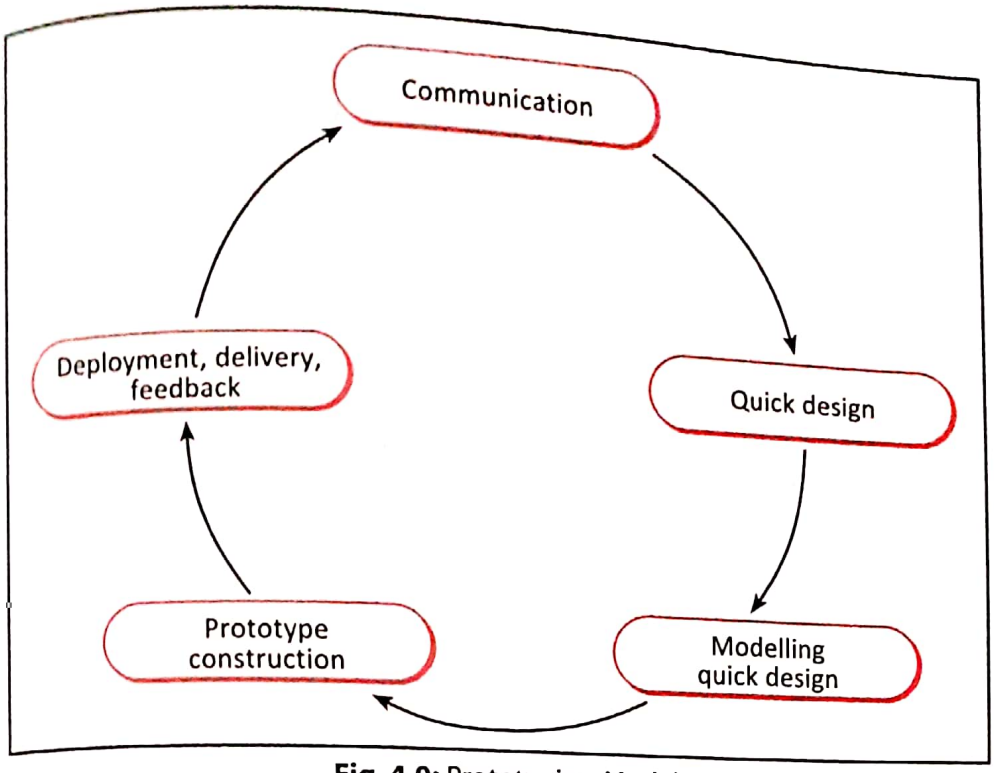


Fig. 4.9: Prototyping Model

Prototyping model: It is a software working model of limited functionality. In this model, working programs are quickly produced.

The Prototyping model consists of the following five stages:

(i) **Communication:** It takes place between the developer and the customer to discuss the overall objectives of the software.

(ii) **Quick design:** This stage involves developing a quick design of known requirements.

(iii) **Modelling quick design:** This stage is essential to get a clear idea about the development of a software.

(iv) **Construction of prototype:** It is evaluated by the customer itself.

(v) **Deployment, delivery, feedback** of developed software solution.

Situations where it is most appropriate to use Prototyping model are:

- When the desired system needs to have a lot of interaction with the end-users.
- Typically, online systems, web interfaces where there is a very high amount of interaction with end-users, are best suited for Prototype model.
- In Prototype model, the end-users constantly work with the system and provide a feedback which is incorporated into the system.

Effect of designing a prototype on overall cost

With an actual prototype in hand, the intent and purpose of a product can be effectively communicated. This helps the product design team work towards eradicating design and functional flaws at an early stage. A working prototype can also serve as an effective tool to keep the project on schedule and avoid costly delays.

From the viewpoint of product development process, the impact of developing a prototype on the product life cycle is that it eliminates redundant or unnecessary features at the earlier stages of development. This can help vastly reduce the development and testing time

over the course of a product's life. If the approach is systematic and correct, prototyping can reduce operating costs and the element of risk associated with the launch of a new software product. Fewer flaws up front lead to an improved end-product as the product conforms to design goals and meets user expectations.

Advantages of the Prototyping Model

Following are the benefits of using a Prototyping model:

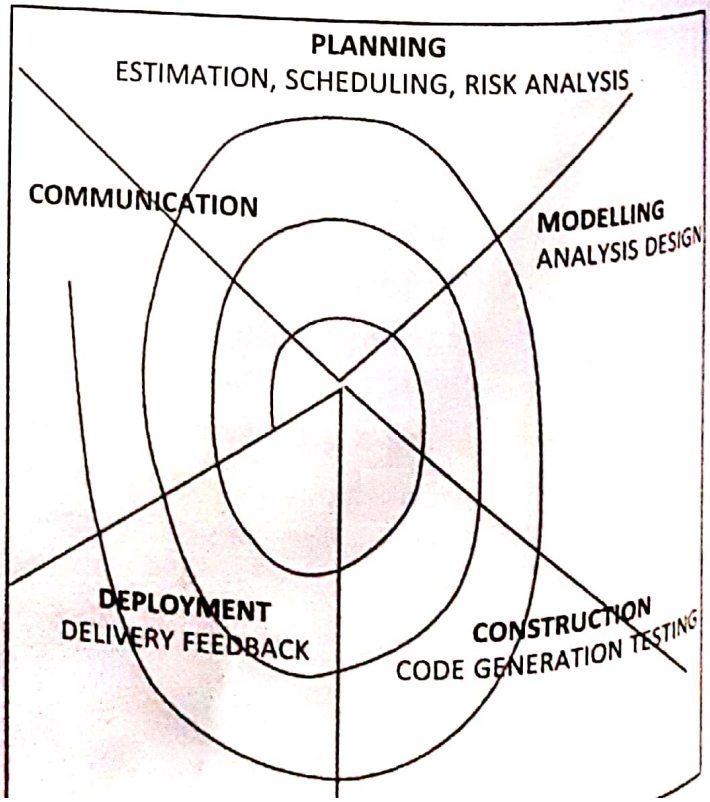
- Users are actively involved in development. Therefore, errors can be detected in the initial stage of the software development process.
- Missing functionality can be identified, which helps reduce the risk of failure as prototyping is also considered a risk-reduction activity.
- It helps team members to communicate effectively.
- There is almost negligible chance of software rejection.
- Quicker user feedback helps achieve better software development solutions.
- It allows the client to compare if the software code matches software specification.
- It is a straightforward model, so easy to understand.
- The prototype helps gain a better understanding of the customer's needs.
- Prototypes can be altered and even discarded.

At the same time, it poses few limitations as well, which are as follows:

Limitations of the Prototyping Model

- Prototyping is a slow and time-consuming process.
- It is very difficult for software developers to accommodate all the changes demanded by the clients.
- The client may lose interest in the final product when he or she is not happy with the initial prototype.
- Developers who want to build prototypes quickly may end up building sub-standard development solutions.
- The cost of developing a prototype is a total waste as the prototype is ultimately thrown away.
- Prototyping may encourage excessive change requests.

Spiral model: It is a combination of prototype and sequential model or Waterfall model. All activities are done in one iteration. Software development is done in phases like planning, modelling, construction, deployment, communication in spiral/continuous manner. These phases are repeated till a fully functional software is developed.



Situations to use spiral model:

- When large and high-budget projects are being implemented.
- When risk assessment is very critical.
- In situations when requirements are not very clearly defined.
- When requirements are vague and even complex.
- If the organization does not have much experience with the domain.
- When ample time is available.

Concurrent development model: This model, also called Concurrent process model, defines a series of events that will trigger transition from state to state for each software engineering activity. During the early stages of design, an inconsistency in analysis model is procured. This will generate the event analysis model correction, which will trigger the analysis activity from the done state to the awaiting changes state.

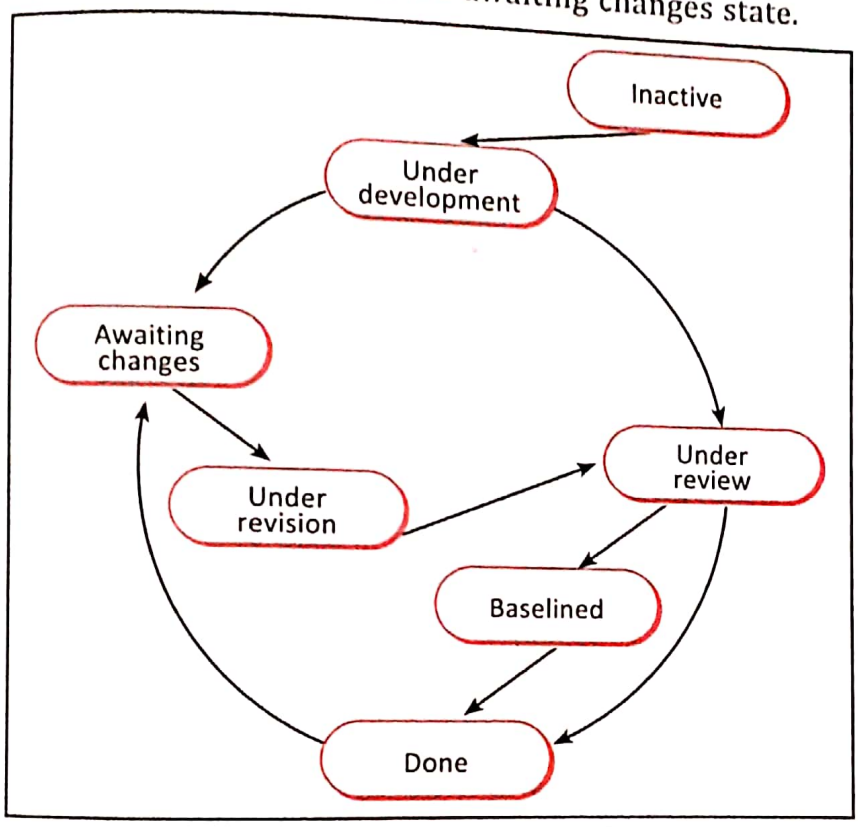


Fig. 4.11: Concurrent Development Model

In the above figure each block represents the state of software engineering activity.

This model is suitable for the development of **client/server applications.**

Situations to use Concurrent development model:

- When incremental release is to be determined by the project team.
- When there is an immediate need of feedback after testing.
- When more features are required to be added later on in the project.
- When there is a need for continuous testing.

Advantages of the Concurrent development model

- This model is applicable to all types of software development processes.
- It is easy to understand and use.
- It gives immediate feedback from testing.
- It provides an accurate picture of the current state of a project.

Disadvantages of the Concurrent development model

- It needs better communication between the team members. This may not be achieved all the time.
- It requires to remember the status of different activities.

4.5.3 Component-based Model

Component-based model is based on the concept of reusability. Software products are built by identifying and reusing components that already exist, and these components are developed through a number of different sources that can be written in several different programming languages. Software development goes through various steps like feasibility study, system design, component identification, component integration, system testing and system release.

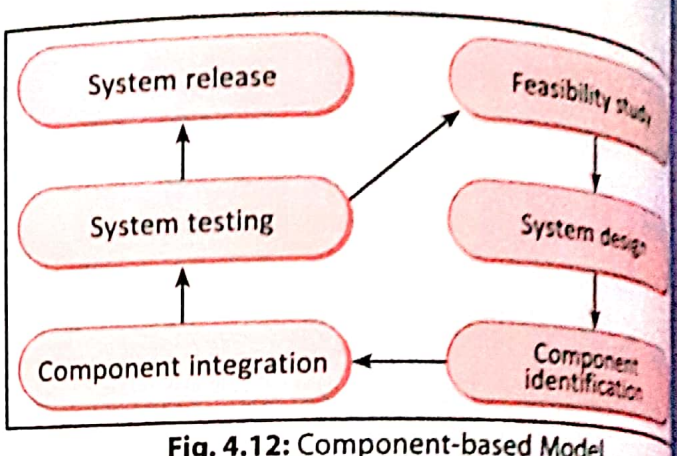


Fig. 4.12: Component-based Model

CTM: A component or more precisely called a software component is a software element that conforms to a software model. They are simple, self-contained pieces of functionality that you can combine in any fashion to solve a problem or set of problems and can be independently deployed and composed without modification according to a composition standard.

Situations where Component-based Model should be used:

- When emphasis is given on buying components and not on building components.
- When the software structure is lower in coupling.
- When the development process is evolutionary, *i.e.*, undergoing transitions and modifications continuously and is concurrent in nature.
- When the focus is on assembling software components and not on developing them.

Advantages of Component-based model

- It reduces the cost and risk of software development.
- It reduces the amount of software to be developed since the reusable components are available before hand.
- Faster delivery of software.

Disadvantages of Component-based model

- Requirement changes affect software development.
- Control over system evolution is lost.
- May lead to an intermediate system that does not meet user specifications in toto.

4.6 DELIVERY MODELS

No system is designed and developed in one stroke. It requires constant modifications as per the changing user requirements. Software delivery is the process of getting a final software product to the market or to the client. Software delivery performs the following actions:

- Finding and tracking client requirements.
- Technology for software development and deployment.
- Testing and validation of software for quality.

The software process activities and delivery models should provide flexibility to modify the developed systems by repeating software process activities at regular intervals as per user changing needs and requirements. This repetitive process is described as **process iteration**. The delivery models that perform these iterative sequence of performing updations on the system under development are also known as **process iteration models**.

A software delivery model is a term widely used in IT industry. It is defined as a way of project delivery based on the location of the client/user. The choice of a delivery model can affect the success of the entire project.

CTM: The delivery model that allows iteratively updation made in the systems iteratively, as per the required changes, is also termed as process iteration model.

Software delivery model is classified into two types:

1. Incremental Delivery Model
2. Spiral Delivery Model

Let us discuss each of them in detail.

4.6.1 Incremental Delivery Model

Incremental delivery model is a process of software development where requirements are broken down into multiple stand-alone modules of software development cycle. Incremental development is done in steps from analysis design, implementation, testing/verification, maintenance.

Incremental delivery model refers to the practice of repeatedly delivering a system to client environment (or to the marketplace) in a series of expanding capabilities. Incremental deliveries are often between three and twelve months.

Incremental delivery model is often confused with iterative development model. A six-month delivery cycle could be composed of 10 short iterations or increments. The results of each iteration are not delivered to the marketplace, but the results of an incremental delivery are delivered to the client.

Incremental delivery model is a development and delivery model that combines the advantages and salient features of two major process models, *viz*, waterfall model and evolutionary model.

The major phases involved in this model consist of:

1. **Requirement Analysis:** Requirement and specification of the software to be developed are collected.
2. **Design:** Some high-end functions are designed during this stage.
3. **Code:** Programming the code is carried out during this phase.
4. **Testing:** Once the system is deployed, it goes through the testing phase.

Incremental delivery model is most appropriate in the following situations:

- This model can be used when the requirements of the complete system are clearly defined and understood. However, some details can evolve with time.
- A new technology is being used.
- Resources with needed skill set are not available.
- There are some high-risk features and goals.

Advantages of Incremental model:

- Generates working software quickly during the software life cycle.
- This model is more flexible—cheaper to change scope and requirements.

- It is easier to test and debug during a smaller iteration.
- In this model customer can respond to each built.
- Lowers initial delivery cost.

Disadvantages of Incremental model:

- Needs good planning and design.
- Needs a clear and complete definition of the whole system before it can be broken down and built incrementally.
- Total cost is higher than **waterfall.**

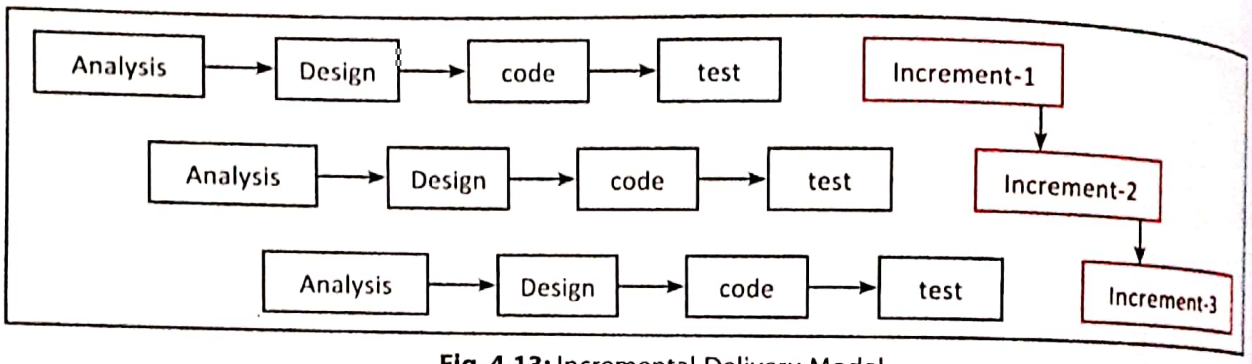


Fig. 4.13: Incremental Delivery Model

4.6.2 Spiral Delivery Model

The spiral model is similar to the incremental model, with more emphasis on risk analysis. The spiral model has four phases or quadrants: Planning, Risk Analysis, Engineering and Evaluation. A software project repeatedly passes through these phases in iterations (called Spirals in this model). During baseline spiral, starting in the planning phase, requirements are gathered and risk is assessed. Each subsequent spiral builds on the baseline spiral.

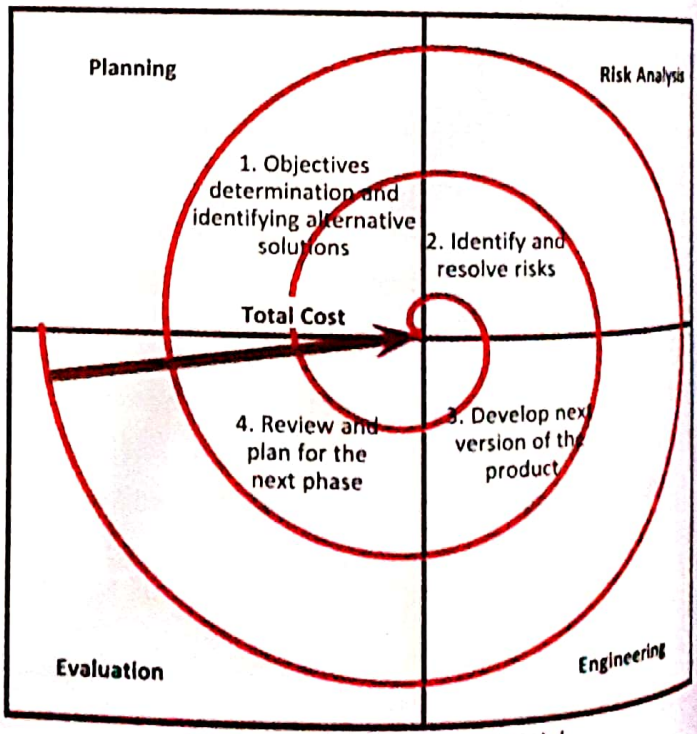


Fig. 4.14: Spiral Delivery Model

Spiral delivery model can be pretty costly to use and doesn't work well for small projects. It's a risk-driven model which means that the overall success of a project highly depends on the risk analysis phase. Risk analysis requires reviewing and analyzing the project from time to time.

The strong side of this model is that there's a possibility to add some additional functionality during the last stages of software product development. Software is repeatedly delivered after each refinement.

The radial dimension of the model represents the cumulative costs. Each path around the spiral is indicative of increased costs.

Each loop of the spiral from X-axis clockwise through 360° represents one phase. One phase is split roughly into four sectors of major activities/quadrants (Fig. 4.14):

- **Planning (First Quadrant):** In this phase, requirements are gathered in terms of 'SRS' that is 'System Requirement Specifications' along with the associated risks involved. It involves determination of objectives, alternatives and constraints.
- **Risk Analysis (Second Quadrant):** In the risk analysis phase, a process is undertaken to identify risk and alternate solutions. A prototype is produced at the end of the risk analysis phase. If any risk is found during the risk analysis then alternate solutions are suggested and implemented.
- **Development and Validation (Third Quadrant):** In this phase product/software is developed, along with testing at the end of the phase. Hence, in this phase, development and testing of the final product is done.
- **Assessment and Evaluation (Fourth Quadrant):** This phase deals with Customer evaluation. It allows the customer to evaluate the output of the project to date before the project continues to the next spiral.

Advantages of Spiral model

- High amount of risk analysis, hence, avoidance of risk is enhanced.
- Good for large and mission-critical projects.
- Strong approval and documentation control.
- Additional functionality can be added at a later date.
- Software is produced early in the software life cycle.

Disadvantages of Spiral model

- Can be a costly model to use.
- Risk analysis requires highly specific expertise.
- Project's success is highly dependent on the risk analysis phase.
- Doesn't work well for smaller projects.

MEMORY BYTES

- The term software engineering is made of two words, software and engineering.
- **Software** is more than just a program code. A program is an executable code which serves some computational purpose.
- **Engineering** is all about developing products, using well-defined, scientific principles and methods.
- The process of developing a software product using software engineering principles and methods is referred to as **software evolution**.
- Waterfall model is the simplest model of software development paradigm. It says that all the phases of SDLC will function one after another in a linear manner.
- In Waterfall model, when the first phase is finished, only then the second phase will start and so on.
- The Iterative model leads the software development process in iterations.
- The evolutionary model is a RAD model where an initial software implementation is rapidly developed from very abstract specifications.
- The Incremental delivery model is a development and delivery model that combines the advantages and salient features of two major process models, *viz.* waterfall model and evolutionary model.
- Component-based model are based on the concept of reusability.
- Spiral model is a combination of both—iterative model and one of the SDLC models.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.

(a) ..... is a field of engineering for designing and writing programs for computers or other electronic devices.

(b) The very first foundation phase of any development process where all the requirements are gathered is known as ..... .

(c) A software ..... is an executable code, which serves some computational purpose.

(d) The input that helps to develop a new system is called ..... .

(e) The model that ensures that all the activities required must be planned and scheduled before starting actual work is known as ..... model.

(f) The activity that performs cost estimation and determines if the proposed system will be cost-effective or not is called ..... .

(g) An ongoing phase which involves correcting and rectifying errors and providing new functionalities is known as ..... phase.

(h) An SDLC model which helps in rapid software implementation is known as ..... model.

(i) The Delivery Model that iteratively updates systems for changes is known as ..... model.

(j) A Delivery Model designed, implemented and tested as a series of software builds until the product is finished is known as ..... model.

(k) The Traditional Waterfall Model ensures that the requirements are fully determined before the ..... phase commences.

(l) Incremental Model is a combination of Waterfall Model and ..... model.

(m) Spiral Model is a type of ..... model.

(n) In Spiral Model activities are divided into ..... quadrants.

Answers: (a) Software engineering (b) Requirements Specification (c) program
(d) System Requirement (e) Waterfall (f) Feasibility Study
(g) Maintenance (h) Evolutionary (i) Process Iteration
(j) Incremental (k) Design (l) Evolutionary
(m) Delivery (n) four

2. State whether the following statements are True or False.

(a) Software Engineering is a structured systematic approach for the design, development and maintenance of the software system.

(b) Software specification mainly deals with the price of the software.

(c) The complexity of a software increases exponentially with increase in their size.

(d) A fault detected at a later stage in software development, costs much higher as compared to early fault detection.

(e) Fault detected at a later stage during software development requires only code correction but documentation remains the same.

(f) SDLC stands for Software Development Life Cycle.

(g) The Waterfall model is a linear and sequential approach of software development.

(h) Waterfall model gives idea of estimated time and cost for each phase of development process.

(i) Waterfall models can handle systems where requirements keep on changing depending upon the target users.

(j) The V-model is a variation of Waterfall model in which process steps are bent upwards after implementation phase.

(k) Incremental delivery model combines the strengths of Waterfall and Evolutionary model.

(l) Waterfall model is more flexible than Incremental delivery model.

(m) The Spiral model combines the iterative nature of Prototyping with systematic aspects of Evolutionary model.

(n) The process of redoing the system as per changed requirements is called Process iteration.

Answers: (a) True (b) False (c) True (d) True (e) False (f) True
(g) True (h) False (i) False (j) False (k) True (l) False
(m) False (n) True

3. Multiple Choice Questions (MCQs)

(a) Which of the following is not a part of software engineering?
(i) Specification Gathering (ii) Design and Implementation
(iii) Publicity (iv) Validation and Verification

(b) SDLC stands for:
(i) System Development Life Cycle (ii) Software Design Life Cycle
(iii) System Development Life Cycle (iv) Software Development Life Cycle

(c) Requirement engineering process includes which of these steps?
(i) Feasibility study (ii) Requirement Gathering
(iii) Software Requirement specification (iv) All mentioned above

(d) If requirements are easily understandable and defined then which model is best suited?
(i) Spiral model (ii) Waterfall model
(iii) Prototyping model (iv) None of the above

(e) A model that is a demo implementation of the system:
(i) Waterfall (ii) Prototype (iii) Incremental (iv) None

(f) Identify the disadvantage of Spiral Model:
(i) Doesn't work well for smaller projects
(ii) High amount of risk analysis
(iii) Strong approval and documentation control
(iv) Additional Functionality can be added at a later date

(g) Which among the following is not a software process model?
(i) Component-based Model (ii) Waterfall Model
(iii) Evolutionary Model (iv) Relational Model

(h) Which among the following is not an advantage of Component-based Model?
(i) Reduces the amount of software to be deployed.
(ii) Results in reduced cost and risks.
(iii) Leads to faster delivery.
(iv) Easy to understand model.

(i) The process of reading the system as per changed requirements is known as:
(i) Process iteration (ii) SDLC
(iii) Prototype (iv) Software Component

(j) The model that leads to faster delivery of the software is referred to as:
(i) Component-based Model (ii) Spiral Model
(iii) Evolutionary Model (iv) Waterfall Model

(k) Which of the following models is suitable for projects with high technical risks and where timeline is aggressive?
(i) Waterfall (ii) Component-based (iii) Evolutionary (iv) Spiral

(l) For scheduling a project following activities are performed. Arrange them in the proper order:
1. Divide time into work-units.
2. Break down the project tasks into smaller, manageable form.
3. Estimate the time frame required for each task.
4. Find out various tasks and correlate them.
(i) 2, 4, 3, 1 (ii) 1, 2, 3, 4 (iii) 4, 2, 3, 1 (iv) 1, 3, 2, 4

(m) Which of the following models cannot cater to high-risk projects?
(i) Waterfall (ii) Evolutionary (iii) Component-based (iv) Spiral

(n) Complexity of a software increases ..... with the increase in the size of software.
(i) linearly (ii) exponentially (iii) twice (iv) inversely

Answers: (a) (iii) (b) (iii) (c) (iv) (d) (ii) (e) (ii) (f) (i)
(g) (iv) (h) (iv) (i) (i) (j) (i) (k) (iii) (l) (i)
(m) (i) (n) (ii)

SOLVED QUESTIONS

1. What is meant by Software Engineering?

Ans. Software engineering is an engineering branch associated with the development of software products using well-defined scientific principles, methods and procedures. The outcome of software engineering is an efficient and reliable software product.

2. Why does the need for Software Engineering arise?

Ans. Software engineering means employing engineering methodology for the design and development of software systems to ensure that—

(i) the software conforms to the specifications and is error-free.
(ii) the software is delivered on time.
(iii) the software is scalable and adaptable.
(iv) the software cost remains within the budget.

3. What is software process?

Ans. Software process is defined as a structured set of activities that is required to develop the software system.

4. What are software process activities?

Ans. The fundamental activities that are common to all software processes are known as process activities. Generally, there are four process activities:

(i) Software Specification. This activity is responsible for defining:
  (a) The main functionalities of the software (software specifications) as per the customer's requirements and expectations.
  (b) Constraints on its operation.
(ii) Software Design and Implementation. This activity is responsible for:
  (a) The design of the proposed software as per software specifications.
  (b) Programming as per the design.
(iii) Software Verification and Validation. This activity phase is responsible for ensuring that—
  (a) The software conforms to all specifications, *i.e.*, no functionality is left out.
  (b) The software works as per the proposed design.
(iv) Software Evolution. This activity ensures two important things:
  (a) The developed software meets the customer requirements.
  (b) The software design ensures adaptability and scalability, *i.e.*, it must evolve to meet the changing customer needs.

5. What are software requirements?

Ans. Software requirements are functional descriptions of a proposed software system. These include the description of the target system, its functionalities and features.

6. What are the various phases of SDLC?

Ans. The generic phases of SDLC are: Requirement Gathering, System Analysis and Design, Coding, Testing and Implementation. These phases depend upon the model we choose to develop a software.

7. What is feasibility study?

Ans. Feasibility study finds out how practical and beneficial the software project development will prove to the organization. A feasibility study is carried out to know the economic, technical and operational feasibility of the project.

It pertains to cost estimation of the proposed software and determines if the proposed system will be cost-effective from the business point of view and whether it can be developed within the existing budgetary constraints.

8. What are functional and non-functional requirements?

Ans. Functional requirements are functional features which are expected by the users from the proposed software product.

Non-functional requirements are related to security performance, look and feel for the user interface

9. What are the reasons that Waterfall models fail to deliver?

Ans. The reasons behind the inability of Waterfall models in certain projects are:

(i) Real projects rarely follow sequential flow. Iterations are made in indirect manner.
(ii) It is mostly difficult for customers or target users to state all requirements explicitly.

10. What are the objectives of Requirement Analysis phase?

Ans. The objectives of Requirement Analysis phase are:

(i) It describes what the customer requires.

(ii) It establishes a basis for creation of software design.

(iii) It defines a set of requirements that can be validated once the software design is built.

11. Which software model is best suitable for client server application? [CBSE Sample Paper 2020]

Ans. Concurrent Process model.

12. Explain Waterfall model. What are its advantages and disadvantages?

Ans. In Waterfall software process model, the project is divided into a sequence of well-defined phases. One phase is completed before the next starts. There is a feedback loop between adjacent phases.

Advantages:

(i) Simplicity

(ii) Lining up the resources with appropriate skills is easy.

Disadvantages:

(i) Highly impractical for most projects.

(ii) Phases are tightly coupled.

13. What are the advantages of Spiral model?

Ans. The main advantage of spiral model is that it is realistic and typifies most software development products/ projects. It combines the best features of most of the earlier models. It strikes a good balance mechanism for easy problem identification and correction, while not missing out proactive problem prevention.

14. What are the main drawbacks of Spiral model?

Ans. The drawbacks of spiral model are:

(i) It is based on customer communication. If the communication is not proper, then the software product that gets developed will not be up to the mark.

(ii) It demands considerable risk assessment. If the risk assessment is done properly, then only the successful product can be obtained.

15. What is meant by Verification?

Ans. Verification means that a set of activities that are carried out confirms that the software correctly implements the specific functionality.

16. Write down any one situation where waterfall software process can be used. Also mention one advantage and one disadvantage of Waterfall.

Ans. Situations to use/apply Waterfall model

(i) When project is small.

(ii) When problem is static.

(iii) Clear and fixed requirements. Stable problem definition.

(iv) Technology is static.

Advantage:

Simple and easy to understand.

Disadvantage:

No working software till the last phase.

17. What is meant by Validation?

Ans. Validation means a set of activities which ensure that the software built is meeting customer requirements.

18. What does design phase/process do?

Ans. Design process is a sequence of steps carried, through which the requirements are translated into a system or software model.

19. Explain the principles which play a major role in the development of software.

Ans. A set of rules which is made on the basis of a number of observations made by the user is called principles of software development. The software development principles are as follows:

(i) **Quality on preference:** Quality is the first objective of software development process. A low level of quality is not accepted by the user or client. The definition for the quality product is described as the product developed according to customer wants along with the best response on a given time with a lot of capacity. The aim of the developer should be to make a unique design software.

(ii) **Possibility of high quality software:** Scope for higher quality software is always there. For getting high quality, many techniques, design inspections on time and hiring the best employee for this can be adopted.

(iii) **On-time delivery:** The better way is to send the sample to the customer so that he can use that software and give feedback. On the basis of the feedback, the developer can take charge and develop the software according to the need of the customer.

(iv) **Fix the problem:** Before formulating solution to a problem, the problem is to be considered in all respects as per the customer needs. Once the problem has been identified, every expected and alternate solution on the basis of time, cost and risk factor is to be tried and developed.

(v) **Valuation of alternate design:** When the requirement of the customer is clearly understood and defined, algorithms are developed and designing is done to build the software, and a best-suited alternative in terms of safety, alteration and functionality is chosen.

(vi) **Selection of appropriate process model:** Several software development processes like waterfall, prototype and spiral model exist. The choice of correct and most appropriate model is to be made as per the user's requirement on the basis of need of software, area of application and, most importantly, risk factor.

(vii) **Minimum gap between client and developer:** According to Dijkstra's algorithm, intellectual gap may be described as the distance between the real-world problem and a computerized solution to the problem. If maintenance of the software is easy, then the gap will be low. With the help of an object-oriented design and the most suitable method of development, it becomes easy to define the structure of the software which will be closer to real-world scenario.

(viii) **Design for change:** The technique or method which is used in the development process of software should have the capacity to change.

(ix) **Reduce tricks:** We should try to reduce those programs and codes which make compaction in that process.

(x) **Management:** As compared to good technology, good management is the best option because it helps the employees motivate themselves for doing their duties in a proper way. Management planning cannot be done for a long period because its base is time and situation which can change any time. Despite lack of resources, a person who has deep knowledge of management can provide a positive result in every case.

UNSOLVED QUESTIONS

5 Agile Methods and Practical Aspects of Software Engineering

5.1 INTRODUCTION

In the previous chapter, we have discussed about the concept of software engineering and various software models and methodologies. In earlier days, the traditional methods such as Waterfall model were very popular to complete a project. But nowadays, developers face various problems while using it to develop a software. The major difficulties lie in the advent of modern technology and changing customer needs at every level during the development of a project.

Of late, software systems are becoming more complicated than ever before in order to meet the enormous data and software requirements of social networking sites such as Facebook, Twitter, YouTube, etc. The conventional or traditional models are incapable of handling such large and complex systems, result in high costs and are more time-consuming.

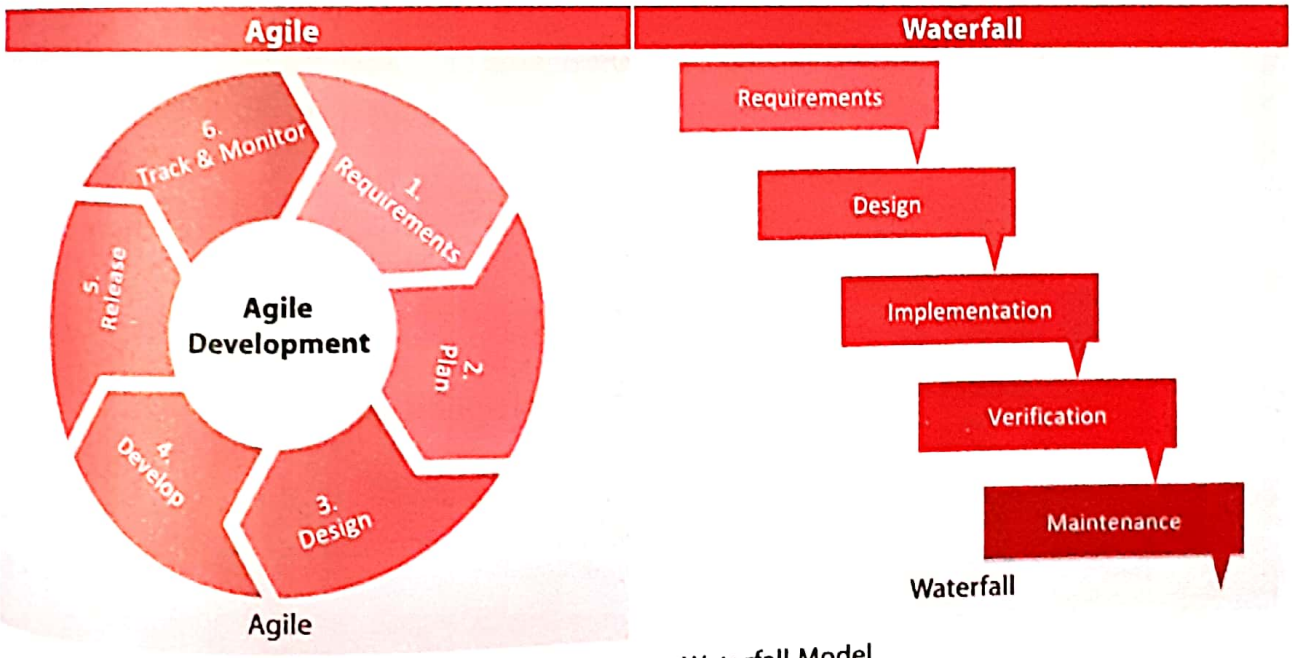


Fig. 5.1: Agile Model vs Waterfall Model

- Continuous cycles
- Small, high-functioning, collaborative teams
- Multiple methodologies
- Flexible/continuous evolution
- Customer involvement

- Sequential/linear stages
- Upfront planning and in-depth documentation
- Contract negotiation
- Best for simple, unchanging projects
- Close project manager involvement

Agile Values *(from the Agile Manifesto)*

The **Agile Manifesto** establishes a set of ***four values*** that are ***people-centric*** and ***result-driven***:

Individuals and Interactions	*Over*	Processes and Tools
Working Software	*Over*	Comprehensive Documentation
Responding to Change	*Over*	Following a Plan
Customer Collaboration	*Over*	Contract Negotiation

That is, while there is value in the items on the right, we value the items on the left more.

The above-defined agile manifesto can be elaborated as:

1. **Individuals and interactions:** In agile development, self-organization and motivation are important, as are interactions like co-location and pair programming.
2. **Working software:** Working software is more useful and important than just presenting documents to clients in meetings.
3. **Customer collaboration:** Requirements cannot be fully identified at the beginning of the software development cycle. Therefore, continuous customer or stakeholder involvement is very important.
4. **Responding to change:** Agile methods are focused on quick responses to change and continuous development.

The above-defined **agile** manifesto is based on **twelve principles**:

1. The highest priority is customer satisfaction by early and continuous delivery of valuable software.
2. Welcomes changing requirements, even in late development.
3. Working software is delivered frequently (in weeks rather than months).
4. Close daily cooperation between business people and developers.
5. Projects are built around motivated individuals who should be trusted.
6. Face-to-face conversation is the best form of communication (co-location) within a development team.
7. Working software is the principal or primary measure of progress.
8. Sustainable development should be maintained at a constant pace.

Pair programming has proven to be an efficient method. It involves two people that produce a higher quality of code as compared to the one developed by individual efforts.

Advantages of Pair Programming

The use of pair programming offers the following advantages or benefits:

1. **Increased Code Quality:** Pairs of developers are more likely to catch problems with code quality while working together. While the driver is focused on writing the code, the navigator is free to think about the style and function. Code quality standards tend to be followed more consistently when developers work together in pairs.
2. **Robust Solutions:** There are many ways to solve problems with software. Developing a software involves selecting the best data structures, algorithms, and class hierarchies, among many other decisions. When developers pair, each of these choices can benefit from the experience of both the developers and the dialogue they have with each other while making those choices.
3. **Developer Communications:** Solo development may be great for focus, but keeping developers informed of each other's work is a problem. Code reviews can reduce this problem but pairing prevents it. Developers that pair are equally familiar with the code that they write and the decisions that went into it. Teams that swap pairing partners regularly maintain a high degree of communication between all developers on a team.
4. **Knowledge Transfer and Mentoring:** Pair programming can be an excellent way for junior programmers to learn rapidly from more senior developers. Juniors and interns experience a sort of "full-immersion" in software development when pairing. They learn the entire complex process, not just by watching, but by participating.
5. **More Efficient:** The common thinking is that it slows down the project completion time because you are effectively putting two programmers to develop a single program instead of having them work independently on two different programs. But it turns out to be more efficient in terms of number of lines of code, better clarity, and systematic approach and less coding errors.
6. **Fewer Coding Mistakes:** Because there is another programmer looking over your work, it results in better code. It allows the driver to remain focused on the code being written while the other attends to external matters or interruption.
7. **Resilient Flow:** Pairing flow is more resilient (flexible) towards interruptions as one programmer deals with errors or interrupts while the other continues working.
8. **Team Cohesion:** People get to know each other more quickly in pair programming. Pair programming may encourage team jellying.

Challenges to Pair Programming

Despite the above-described benefits, pair programming suffers from some bottlenecks:

1. In pair programming, you cannot sit back and self-evaluate your own code.
2. One of the pair may stop being actively engaged.
3. The driver needs to "program aloud". Programming silently reduces the benefit.
4. It costs more man-hours to produce the same features. Balance must be maintained between quality of code and increased coding cost.

5. A "watch the master" phenomenon may arise when an experienced and a novice (beginner or learner) programmer pair up. The novice member may become the observer with the experienced member completing most of the coding.
6. When two experienced users pair up, a "developer's ego" phenomenon may arise, with each member trying to push his/her own ideas.

5.4 SCRUM

Scrum is an agile development method which concentrates on how to manage tasks within a team-based development environment. Scrum believes in empowering the development team and recommends working in small teams.

Scrum is a framework that helps teams work together. Much like a rugby team (where it gets its name) training for the big game, Scrum encourages teams to learn through experience, self-organize while working on a problem, and reflect on their wins and losses to continuously improve.

Process flow of Scrum Methodologies can be described as:

- Each iteration of a scrum is known as Sprint.
- Sprints are periods of time when software development is actually done.
- A sprint usually lasts from one week to one month to complete an item from the product backlog.
  - ♦ Product backlog is a document/list where all details are entered to get the end product.
  - ♦ During each sprint, top items of product backlog are selected and developed.
  - ♦ Team works on the specified sprint backlog.
  - ♦ Team checks the work on daily basis.
- Each sprint ends with a sprint review.
- Team delivers product functionality at the end of the sprint.
- Sprints continue until the project deadline or the project budget is spent.

.4.1 Scrum Team—Roles and Responsibilities

he Scrum team chiefly consists of three roles: **he Scrum Master, Product Owner and the evelopment Team.**

nyone outside the core team doesn't have ıy direct influence over the team. Each of ıese roles in the Scrum has a very clear set of sponsibilities which we will discuss in detail ıw. Under this section, let us focus on the tributes of the Scrum Team as a whole and e ideal team size.

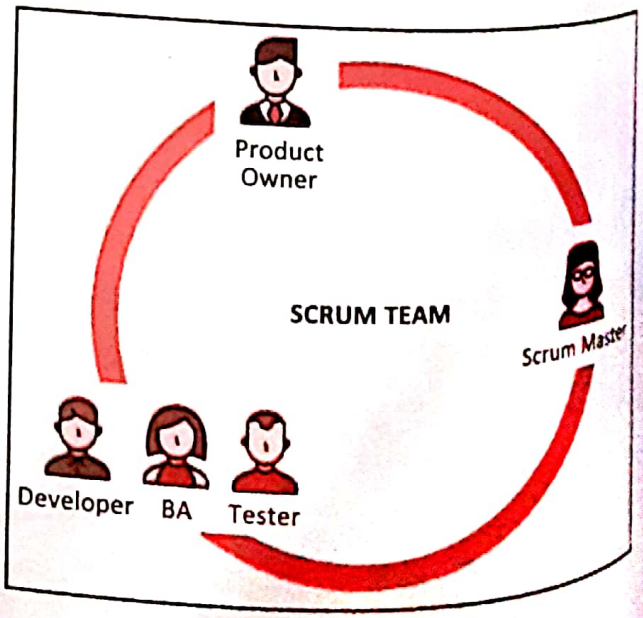


Fig. 5.3: Scrum Team

Scrum Team Size

The recommended Development Team size in Scrum is 6+/– 3, *i.e.*, from 3 to 9 members which does not include the Scrum Master and the Product Owner. Let us discuss each of these roles in detail.

There are three members in a basic Scrum team:

1. **Scrum Master:** Scrum Master is the person who is responsible for facilitating/coaching the Development Team and the Product Owner to work on the day-to-day development activities. He is the one who ensures that the team understands the Scrum Values and Principles and is able to practise them. At the same time, Scrum Master also assures that the Team feels enthusiastic about agile in order to achieve the best out of the framework. Scrum Master also helps and supports the team to become self-organized.

   He is the one responsible for setting up the team, sprint meeting and removes obstacles to progress.

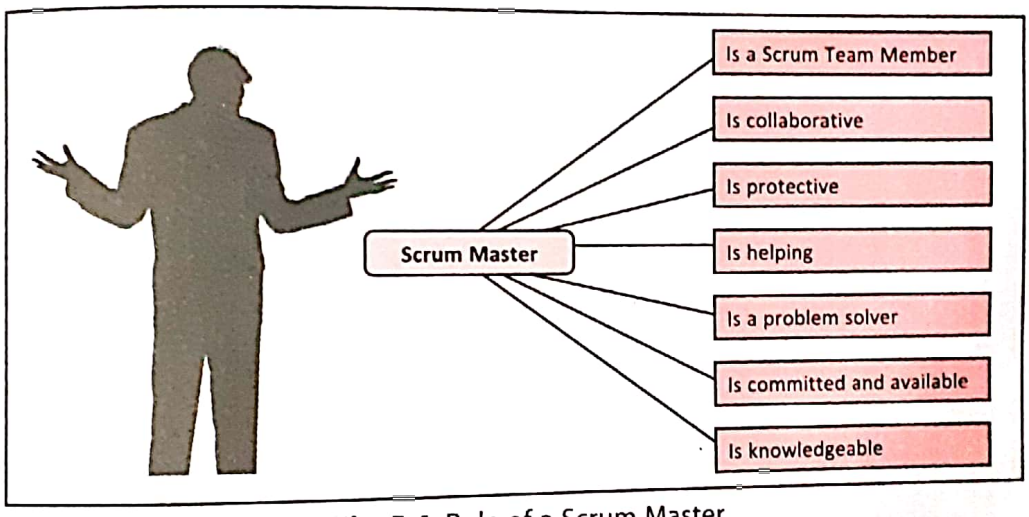


Fig. 5.4: Role of a Scrum Master

2. **Product Owner:** Product Owner is the voice of the customer/stakeholders and, hence, is responsible for bridging the gap between the development team and stakeholders. It's the Product Owner who creates product backlog and is responsible for the delivery of the functionality at each iteration.
3. **The Development Team:** Another very important part of the Scrum Team is the Development Team. The Development Team is comprised of developers proficient in their own area of expertise. It organizes the work to complete the sprint or cycle.

5.4.2 Scrum Events

Scrum Process Framework can be viewed by means of a sequence of events. The Scrum events are time-boxed events. This means that in a project, every Scrum event has a predefined maximum duration. These events enable transparency on the project progress to all who are involved in the project.

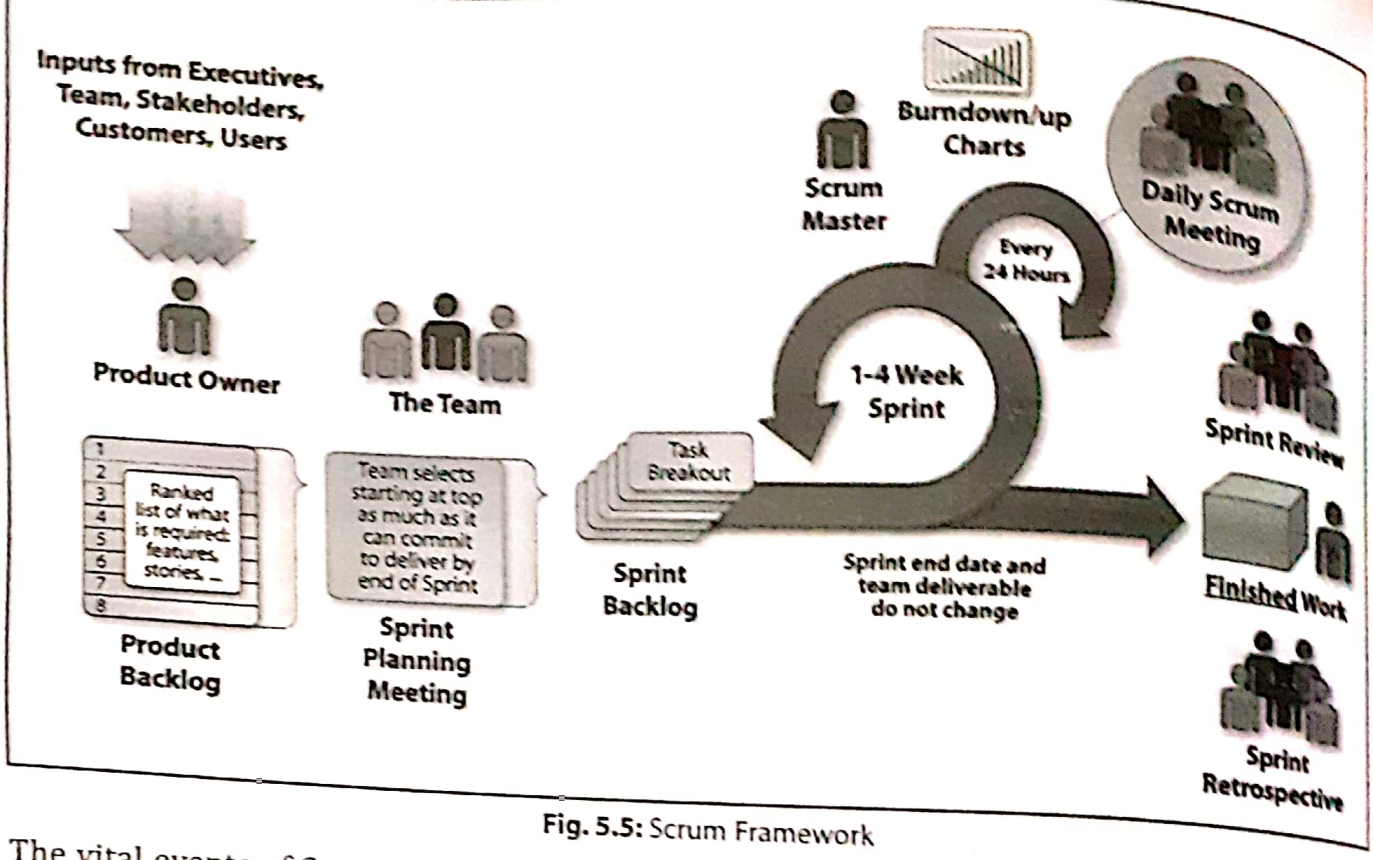


Fig. 5.5: Scrum Framework

The vital events of Scrum are:

- The Sprint
- Sprint Planning
- Daily Scrum Meetings
- The Sprint Review
- The Sprint Retrospective

➢ **The Sprint:** The main part of Scrum is a Sprint, where a usable and potentially releasable product increment is created. Sprints can be of one week to one month in length.

➢ **Sprint Planning:** Team decides what to work in current period. A time-boxed event occurs at the beginning of a sprint where the team determines the product backlog items they will work on during that sprint.

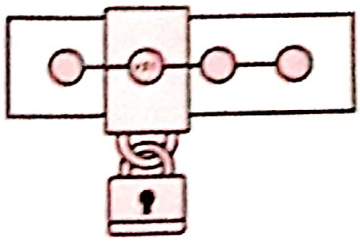
➢ **Daily Scrum Meetings:** The Development Team meets for 5 to 15 minutes daily to inspect progress towards the goal and to synchronize its activities.

➢ **Sprint Review:** A time-bound activity in which the team reviews the task completed as per the Backlog.

➢ **Sprint Retrospective:** A time-boxed event for providing an opportunity for the Scrum team to inspect itself and create a plan for improvements to be enacted during the next sprint. The team discusses the right and wrong development and how to improve.

5.5 VERSION CONTROL SYSTEM

Version Control System (VCS) can be considered as a kind of database. It helps us save a snapshot of the complete project at any point of time. Through it, project files can be tracked along with the details such as who made the change and why the changes were made. Later on, if it is required to take a look at an older snapshot/version, Version Control System shows how exactly it differs from the previous one.



Version Control Systems are a category of software tools that help a software team manage changes to source code over time. Version control software keeps track of every modification to the code in a special kind of database. If a mistake is made, developers can turn back the clock and compare earlier versions of the code to help fix the mistake while minimizing disruption to all team members.

When our project is tracked by Version Control System, any addition/deletion/modification in files of our project will be automatically detected and recorded by it. Version Control System is also known as:

- Source Control Management System
- Revision Control System
- Configuration Management System

5.5.1 Significance of Using Version Control System

Software industry is one which is never stagnating. Software developers working in teams are continually writing new source code and changing the existing one. The code for a project, app or software component is typically organized in a folder structure or "file tree". One developer in the team may be working on a new feature while another may be fixing an unrelated bug by changing the code; each developer may make their changes in several parts of the file tree.

Version control helps the teams solve these kinds of problems, tracking every individual change by each contributor and helping prevent concurrent work from coming into conflict with each other. Changes made in one part of the software can be incompatible with those made by another developer working at the same time. This problem should be discovered and solved in an orderly manner without blocking the work of the rest of the team. Further, in all software development, any change can introduce new bugs on its own and new software can't be trusted until it's tested. So, testing and development proceed together until a new version is ready. Thus, the success of Version Control System is attributed to its salient features described as follows:

Features of Version Control System

- It maintains separate track record for each team member of the project.
- It is easy to compare and merge codes of different branches.
- It is easy to trace changes in the code to find the version that introduced a bug.
- It is simple to compare versions to resolve conflicts in the code during merging.
- It reverts changes made to the code to any state from its history.

5.5.2 Types of Version Control System

Version Control System is classified into two types:

1. Centralized Version Control System
2. Distributed Version Control System

1. **Centralized Version Control System**

Centralized Version Control System (CVCS) (such as CVS, Subversion and Perforce) has a single server that contains all the versioned files, and a number of clients that check out files from that central place. For many years, this has been the standard for version control.

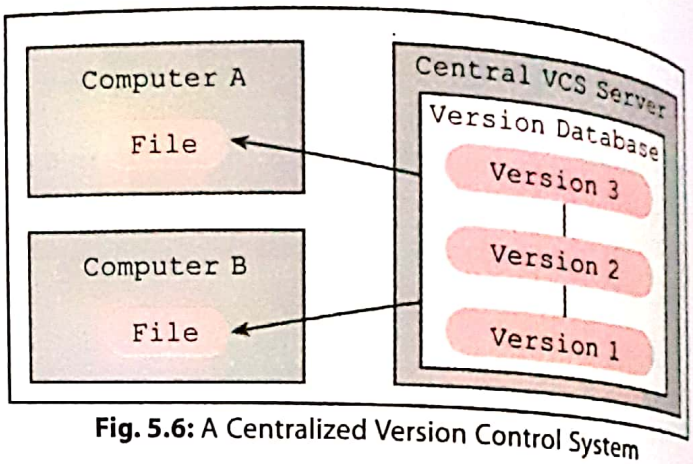


Fig. 5.6: A Centralized Version Control System

This set-up is more advantageous than the local databases or local set-up:

- Everyone in the team knows to a certain degree what everyone else on the project is doing.
- Administrators have complete control over who can do what, and it's far easier to administer a CVCS than to deal with local databases on every client.

However, this set-up has its downsides too.

- In this system, the single point of failure that the centralized server represents is that if the server goes down for an hour, then during that hour nobody can collaborate at all or save their versioned changes to anything they're working on.
- In case the central database or its hard disk becomes unstable or gets corrupted, and proper backups haven't been kept, all the data may get lost, including the entire history of the project except whatever single snapshots people happen to have on their local machines.

2. **Distributed Version Control System**

The limitations posed by Centralized Version Control System have been overcome in Distributed Version Control System. In a DVCS (such as GIT, Mercurial, Bazaar or Darcs), clients don't just check out the latest snapshot of the files; rather, they fully view and access the repository, including its full history. Thus, if any server crashes or stops working, and these systems were collaborating via that server, any of the client repositories can be copied back up to the server to restore it. Every clone takes a full backup of

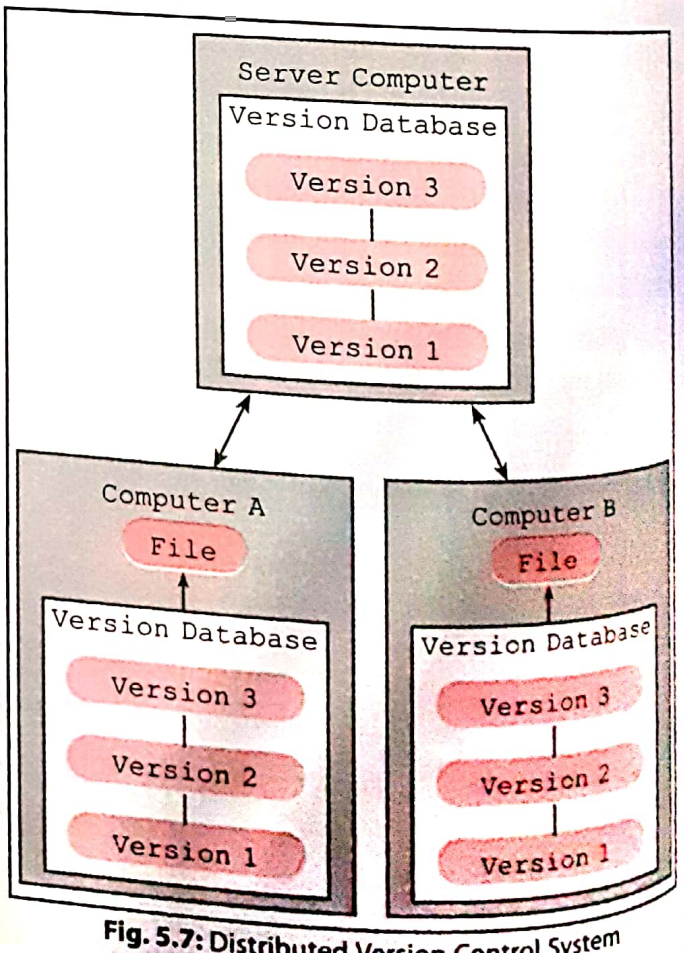


Fig. 5.7: Distributed Version Control System

all the data. Apart from this, Distributed Version Control System works well with several remote sites or repositories they can work with, so you can collaborate with different groups of people in different ways simultaneously within the same project. This allows you to set up several types of workflows that aren't possible in a centralized system.

5.5.3 Why Use a Version Control System

There are many benefits of using a Version Control System for your projects.

1. **Collaboration:** With a VCS, everybody on the team is able to work absolutely freely on any file at any time. The VCS will later allow you to merge all the changes into a common version. There's no question where the latest version of a file or the whole project is. It's in a common, central place, where you work in collaboration with your teammates.
2. **Storing Versions (Properly):** Saving a version of your project after making changes is an essential habit. But without a VCS, this becomes tedious and confusing. A Version Control System acknowledges that there is only one project. Therefore, there's only the one version on your disk that you're currently working on. Everything else—all the past versions and variants—is neatly packed up inside the VCS. When you need it, you can request any version at any time and you'll have a snapshot of the complete project right at hand.
3. **Restoring Previous Versions:** If at any point of time, you realize that the changes you made to the existing software prove to be garbage, you can simply undo them in a few clicks.
4. **Understanding What Happened:** Every time you save a new version of your project, your VCS requires you to provide a short description of what was changed. This helps you understand how your project evolved between versions.
5. **Backup:** A VCS like GIT can act as a backup; every team member has a complete version of the project on his disk including the project's complete history. In case of any hardware crash or central server breakdown (and your backup drives fail), all you need for recovery is one of your teammates' local GIT repository.

Some popular version control systems are GIT, CVS, SVN, Assembla, Mercurial and Bazaar.

5.6 GIT—A DISTRIBUTED VERSION CONTROL SYSTEM

GIT is currently the most popular Distributed Version Control System.

It has its origin in the Linux kernel development and it was founded in 2005 by Linus Torvalds. Nowadays, it is being used by many popular open source projects like Android, Eclipse developer teams, as well as many commercial organizations.

Basically, it was written in the programming language C, but later on GIT has been re-implemented in other languages, *e.g.*, Java, Ruby and Python.

Advantages of Using GIT

- **Performance:** GIT has good performance among other VCS. Committing, branching, merging all are optimized for a better performance than other systems.
- **Security:** It secures our codes. GIT handles security with cryptographic method SHA-1.
- **Branching Model:** We can have multiple local branches which are independent of each other. So there is less friction, **context switching** (switch back and forth to new commit, code and back).
- **Staging Area:** GIT has an intermediate stage called "index" or "staging area" where commits can be formatted and modified before being completed.

- **Distributed:** Distributed means that the repository or the complete code base is mirrored onto the developer's system so that he only can work on it.
- **Open Source:** Being an open source, it invites developers from all over the world to contribute to the software and make it more powerful.

Disadvantages of Using GIT

- GIT is less preferred for handling extremely large files or frequently changing binary files.
- GIT does not support 'commits' across multiple branches or tags.

5.7 BUSINESS USE-CASE DIAGRAM

Business use-case diagram describes how the business is being used by its customers; the activities that directly concern the customer.

An important part of the Unified Modelling Language (UML) is the facilities for drawing use-case diagrams. Use-cases are used during the analysis phase of a project to identify and partition system functionality. They separate the system into actors and use-cases. Actors represent roles that are played by users of the system. Those users can be humans, other computers, pieces of hardware, or even other software systems. The only criterion is that they must be external to the part of the system being partitioned into use-cases.

➢ **What is use-case**

Use-case means how a business system interacts with its environment. It represents the activities that are performed by the users of the system.

➢ **What is an actor**

An actor is a user or outside system which interacts with the system being designed in order to obtain some result from that interaction.

Actor ⟶ Person, machine, information system

Therefore, we can say that a use-case diagram is a representation of a user's interaction with the system where relationship is shown between the user or actor and the different use-cases in which the user is involved.

In UML standard, both business use-case as well as business actor are not defined, so either we need to use some UML tool supporting those or create our own business modelling stereotypes.

CTM: Use-cases are structured outline or template for the description of user requirements modelled in a structured language like English.

➢ **Categories of Business Use-Cases**

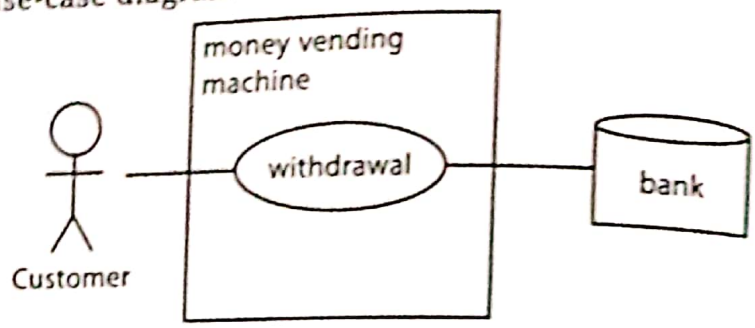
Based on the activities in a business, there are three categories of business use-cases:

- First, activities which are commercially important, known as business processes.
- Second, activities which are not commercially important, but to be performed anyhow to make the business work. *For example*, security, administration, cleaning, etc.
- Third, management work. It defines the type of work that affects business use-cases management.

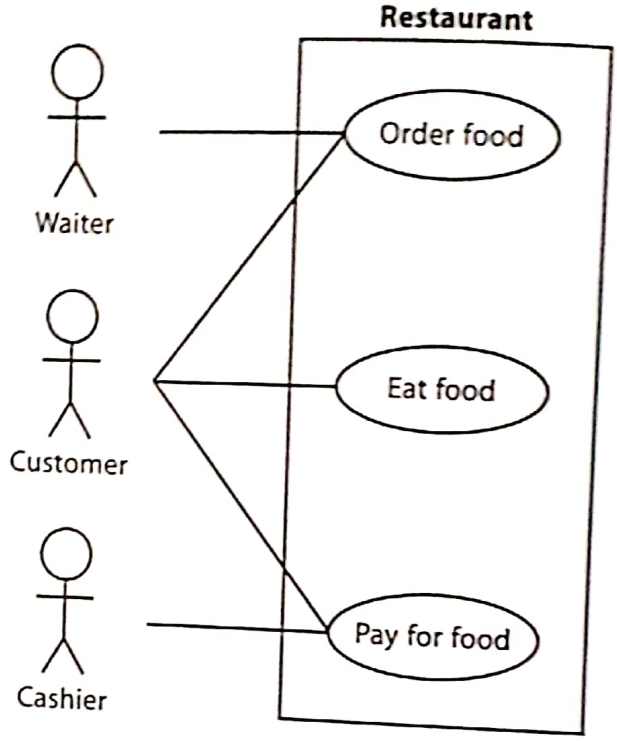
➢ **Business Use-Cases Always Related to Business Actors**

Every business use case should have a relationship to or from a business actor. If business use-*case model has* business use-cases which no one requests or uses, it displays a warning that *something is wrong* with the model.

Example 1: Draw a use-case diagram for bank ATM.



Example 2: Draw a Business use-case diagram for Restaurant Management System.



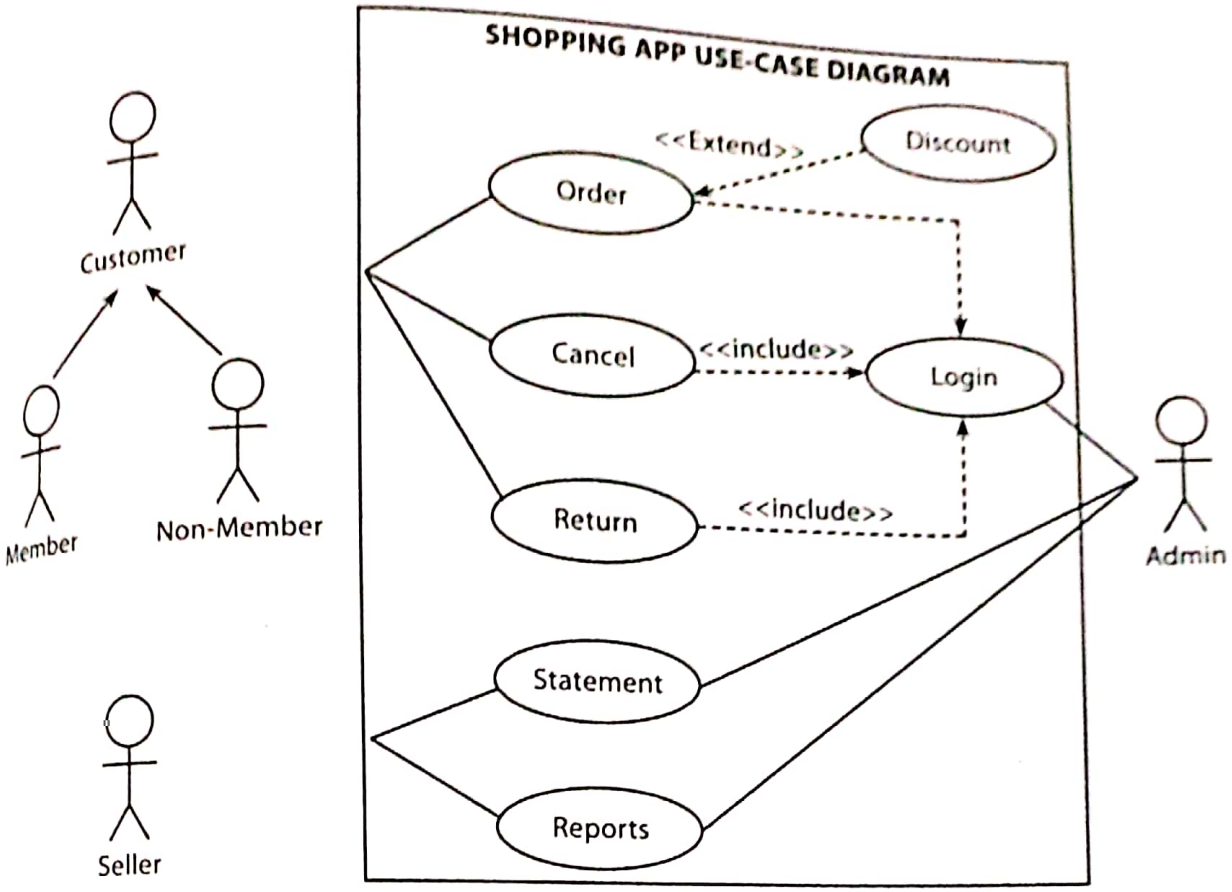
5.7.3 Use-Case Diagram of a Software System—"Shopping app" Use-Case Diagram

For drawing the use-case diagram for **Shopping app** software system, following steps are to be followed:

Steps

1. First draw actors on the left side like customer (then generalize with member and non-member) and seller.
2. Then draw use-cases like order, cancel, return for customer and stmt (statements) and reports for seller.
3. Assign connector ( ) between actor and use-cases.
4. Make use-case login, because order, cancel or return is possible after login; that's why it is necessary (so mark it as include).
5. Draw discount as extended as it is optional.
6. Draw admin as actor and relate with login, stmt and reports.
7. Assign system boundary and name it as shopping app use-case diagram.

Note: Design of use-case diagram designed by two or more persons may vary because creativity of different persons varies.

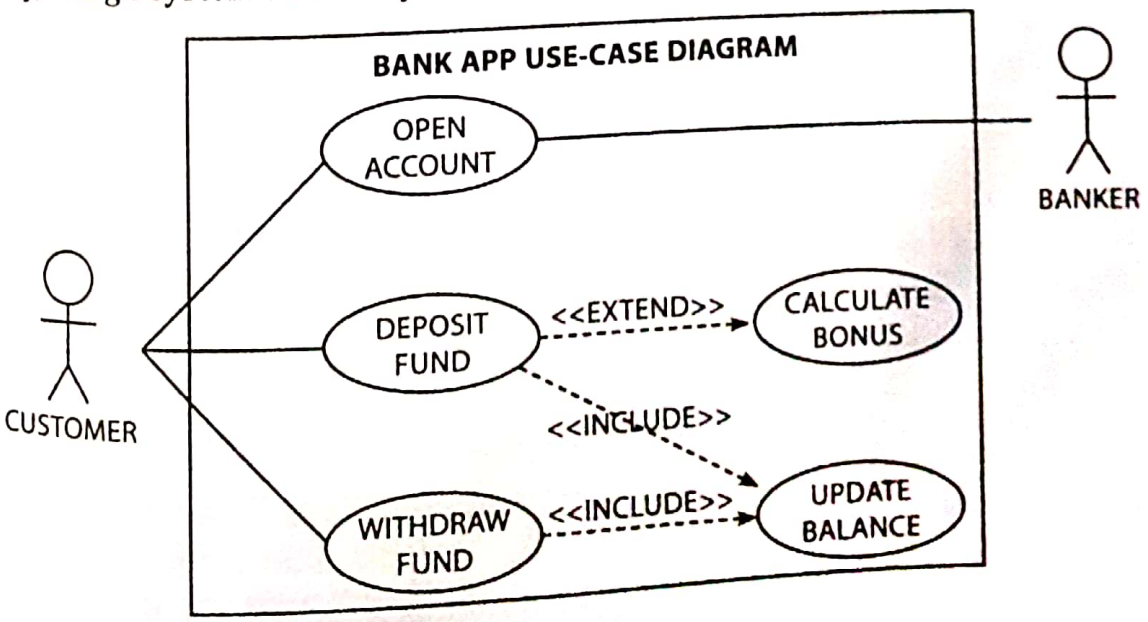


5.7.4 Use-Case Diagram of a Software System—"Banking app" Use-Case Diagram

For drawing the use-case diagram for Banking app software system, following steps are to be followed:

Steps

1. First draw the customer as actor.
2. Then draw use-cases open account, deposit fund, withdraw fund.
3. Next, connect use-cases with the customer.
4. Then draw use-case update bonus as <<include>> (because it is necessary) from deposit fund and withdraw fund use-cases and calculate bonus as optional (<<extend>>).
5. Draw the banker and connect it with open account (because for opening an account there must be a banker).
6. Finally, assign system boundary and assign name as bank app use-case diagram.



MEMORY BYTES

- **Agile software development** is a set of principles for software development in which requirements and solutions evolve through collaboration between self-organizing, cross-functional teams.
- Pair programming is an agile method or an agile software development technique where two programmers work together at a single workstation/computer. One programmer is the driver, who has control over the mouse and keyboard to write the code, while the other serves as the navigator, who reviews each line of that code that the other (driver) is writing.
- The two roles in pair programming are that of a **driver** and a **navigator**.
- Scrum is an agile development method which concentrates on how to manage tasks within a team-based development environment.
- All agile methods adhere to a set of rules described as **Agile Manifesto**, which are well-defined guidelines for the agile methods.
- Each iteration of a Scrum is known as sprint.
- Sprints are periods of time when software development is actually done.
- Product backlog is a document/list where all details are entered to get the end product.
- Version control systems are a category of software tools that help a software team manage changes to source code over time.
- There are two types of version control system—Centralized Version Control System and Distributed Version Control System.
- GIT is a popular distributed version control system.
- The use-case diagram is a formal way of representing how a business system interacts with its environment by illustrating the activities that are performed by the users of the system.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.

(a) ..... is a set of principles for software development in which requirements and solutions evolve through collaboration between self-organizing and cross-functional teams.

(b) All agile methods adhere to a set of rules described as ....., which are well-defined guidelines for the agile methods.

(c) ..... is a conceptual framework for software engineering that promotes development iterations throughout the life cycle of the project.

(d) Pair programming is an ..... practice.

(e) Pair programming allows ..... individuals to work in a pair to develop software.

(f) In pair programming ..... is responsible for actual coding.

(g) ..... organizes software development team to achieve a common goal.

(h) ..... are periods of time when software development is actually done.

(i) ..... is a set of software tools that helps the software team to manage changes to source code.

(j) A ..... is a database of all the edits of the software produced.

(k) GIT is a ..... system.

(l) ..... are a formal way of representation of interaction of business system with its environment.

(m) ..... defines system at implementation level.

(n) ..... is a person or things which are outside the system and are involved in a task.

(o) ..... is the linking line joining the actor and the tasks, *i.e.*, use case.

Answers: (a) Agile software development
(b) Agile manifesto
(c) Agile software development
(d) agile
(e) two
(f) driver
(g) Scrum
(h) Sprint
(i) Version control system
(j) repository
(k) version control
(l) Use-case diagrams
(m) System Use-case diagram
(n) Actor
(o) Communication/Association

2. State whether the following statements are True or False.

(a) Agile software development refers to the software work for speedy and flexible output.
(b) In pair programming, both programmers keep changing their roles and bring solidity to the team.
(c) Reducing the pair to one decreases work efficiency by ten times.
(d) Scrum does not organize software developers as a team to reach a common goal.
(e) Scrum teams are expected to depend on others outside the team.
(f) Version control system helps a software team to manage changes to source code over certain time.
(g) Version control system does not support partial edits.
(h) Update means updating the central repository instead of local working copy.
(i) Pair programming is a routine of software development where only one person can code or think.
(j) Scrum is known as the strategy that organizes software developers as a team to achieve a common goal.
(k) Use-case diagrams are a representation of interaction of business system with its environment.
(l) Commit and update commands only modify the working of local repository.
(m) "CRUD" rule is followed to create use-case diagrams.
(n) SCRUM allows individual team members to be specialists.
(o) The version control systems are a specific, specialized set of software tools.

Answers: (a) True (b) True (c) True (d) False (e) False (f) True
(g) False (h) False (i) False (j) True (k) True (l) False
(m) True (n) True (o) True

3. Multiple Choice Questions (MCQs)

(a) Which of the following statements does not hold true for Agile?
(i) People-focused
(ii) Communication-oriented
(iii) Not flexible
(iv) Speedy

(b) Pair programming is the practice of software development where:
(i) Four people work in a group
(ii) Two people work in a group
(iii) Three people work in a group
(iv) No groups are made

(c) Which of the below mentioned rules is not applicable for pair programming?
(i) Switching role between driver and navigator
(ii) Pairs do not change frequently
(iii) Both parties contribute to solution
(iv) Both (ii) and (iii)

(d) Which of the following is not a member of the basic Scrum team?
(i) The product owner
(ii) The development teams
(iii) Driver
(iv) The scrum master

(e) Which of the following is not an advantage of pair programming?
(i) Collective code ownership
(ii) Better code
(iii) Both (i) and (ii)
(iv) More interruption

(f) How many events take place in scrum?
(i) 4
(ii) 3
(iii) 1
(iv) 2

(g) Which of the following does not contribute to version control system?
(i) Partial edit
(ii) Work integration
(iii) Save multiple versions of software
(iv) Doesn't help to change the source code

(h) A work copy is defined as:
(i) personal copy of all files
(ii) personal copy of partial files
(iii) personal copy of files given to only one developer
(iv) both (i) and (ii)

(i) Repository stores the change of:
(i) Partial edit
(ii) complete edit
(iii) do not store changes
(iv) Both (ii) and (iii)

(j) Repository is also known as:
(i) REPE
(ii) REPO
(iii) Both (i) and (ii)
(iv) None of the above

(k) How many central repositories are there in version control system?
(i) 1
(ii) 2
(iii) Many
(iv) Both (i) and (ii)

(l) Which of the following is not correct about GIT?
(i) Auto backup
(ii) Allows offline repository access
(iii) Not an open source
(iv) Maintains complete history of changes

(m) A use-case diagram represents:
(i) Business system
(ii) Structure of business
(iii) Business system integration with its environment
(iv) All of the above

(n) Which of the following are the elements of use-case diagrams?
(i) Actor
(ii) Use case
(iii) Both (i) and (ii)
(iv) Driver

(o) Which of the following describes the interaction between an actor and a business system?
(i) Business use-case system
(ii) Use-case diagram
(iii) Structure of business
(iv) GIT Architecture

Answers: (a) (iii) (b) (ii) (c) (ii) (d) (iii) (e) (iv) (f) (i)
(g) (iv) (h) (i) (i) (ii) (j) (ii) (k) (i) (l) (iii)
(m) (iv) (n) (iii) (o) (i)

SOLVED QUESTIONS

1. What is agile software development?

Ans. **Agile** software development is a group of software development methodologies based on iterative and incremental development, where requirements and solutions evolve through collaboration between self-organizing, cross-functional teams.

2. Define the terms:
(i) Agile Methods
(ii) Agile Process

Ans. (i) **Agile Methods:**
(a) Methods to overcome perceived and actual weakness in conventional software engineering
(b) To accommodate changes in environment, requirements and use cases
(ii) **Agile Process:** Focus on Team Structures, Team Communications, Rapid Delivery of software and de-emphasis on importance of intermediate product

3. What is agile testing?

Ans. Agile testing is a practice that a quality assurance (QA) manager follows in a dynamic environment where testing requirements keep changing according to customer needs. It is done parallel to the development activity where the testing team receives frequent small codes from the development team for testing.

4. Define the roles in Scrum.

Ans. There are mainly three roles that a Scrum team has:
(i) **Project Owner:** Has the responsibility of managing the product backlog. Works with end-users and customers and provides proper requirement to the team to build the proper product.

(ii) **Scrum Master:** Works with Scrum team to make sure that each sprint gets complete on time. Scrum master ensures proper work flow to the team.

(iii) **Scrum Team:** Each member in the team should be self-organized, dedicated and responsible for high quality of the work.

5. What is Product backlog and Sprint backlog?

Ans. Product backlog is maintained by the project owner which contains every feature and requirement of the product.

Sprint backlog can be treated as a subset of product backlog which contains features and requirements related to that particular sprint only.

6. Explain the difference between traditional Waterfall model and agile testing.

Ans. Agile testing is done parallel to the development activity whereas in traditional Waterfall model, testing is done at the end of the development.

As done in parallel, agile testing is done on small features whereas in Waterfall model, testing is done on the whole application.

7. Explain pair programming and its benefits.

Ans. Pair programming is a technique in which two programmers work as a team in which one programmer writes code while the other one reviews that code. They both can switch their roles.

Benefits:

(i) **Improved code quality:** As second partner reviews the code simultaneously, it reduces the chances of mistakes.

(ii) **Knowledge transfer is easy:** One experienced partner can teach the other partner the techniques and codes.

8. Name some agile quality strategies.

Ans. Some Agile quality strategies are:

(i) Re-factoring (ii) Small feedback cycles

(iii) Dynamic code analysis (iv) Iteration

9. Briefly explain Agile Manifesto and its principles.

Ans. This is the basic theory which most of the agile/Scrum roles aspirants should follow.

According to Agile Manifesto:

(i) Working Software should be demonstrated at regular intervals

(ii) Individuals & interaction—self-organization, self-motivation should be encouraged

(iii) Customer collaboration

(iv) Welcoming change at any point in time in the project

10. Write down any one benefit of pair programming.

Ans. Improved code quality: As second partner reviews the code simultaneously, it reduces the chances of mistake.

11. In the Scrum process, a ScrumMaster differs from a traditional project manager. Justify the statement.

Ans. (i) The ScrumMaster is the servant leader to the Product Owner, Development Team and Organization with no hierarchical authority over the team but rather more of a facilitator, the ScrumMaster ensures that the team adheres to Scrum theory, practices, and rules.

(ii) The ScrumMaster protects the team by doing anything possible to help the team perform at the highest level.

12. Gunveen, Marshy and Aloha are three developers working on an exciting new app, and the launch day is just a day away. Gunveen creates an unmanaged package and saves it in Aloha's folder. Marshy also writes a new piece of code and saves it in Aloha's folder. What could go wrong on the day of the launch? Explain and also mention how version control can help teams in this scenario. **[CBSE Sample Paper 2020]**

Ans. (i) The team members are not working in a systematic way and they are not saving the versions of their work. Changes made in one part of the software can be incompatible with those made by another developer working at the same time.

(ii) Version control exists to solve these problems, and it's within easy reach for every developer. Version control helps teams solve these kinds of problems, tracking every individual change by each contributor and helping prevent concurrent work from conflicting.

(iii) Further, in all software development, any change can introduce new bugs on its own and new software can't be trusted until it's tested. So testing and development proceed together until a new version is ready.

13. What is the importance of daily stand-up meeting?

Ans. Daily stand-up meeting is essential for any team in which:

(i) Team discusses about how much work has been completed.

(ii) Plans to resolve technical issues are deliberated.

(iii) Steps that need to be taken to complete the projects, etc., are discussed.

14. What is the duration of a Scrum sprint?

Ans. Generally, the duration of a Scrum sprint (Scrum cycle) depends upon the size of project and team working on it. The team size may vary from 3-9 members. In general, a scrum script/sprint completes in 3-4 weeks. Thus, on an average, the duration of a scrum sprint (scrum cycle) is 4 weeks.

15. Are there any drawbacks of agile model? If yes, explain.

Ans. Yes, there are some drawbacks of the agile model. Some of them are as follows:

(i) It is not easy to make a prediction about the effort required to complete a task. It becomes more problematic in case of large projects as it becomes difficult to predict the total effort required.

(ii) Sometimes, it is not possible to properly focus on the design and documentation of the project.

(iii) In case the requirements of the client are not understood properly, the final project will not meet the customer requirements. Thus, it will lead to the customer dissatisfaction.

(iv) Only a leader who has considerable experience in agile methodologies is capable of taking important decisions. The team members with little or no experience are not involved in decision-making; thus they don't get a chance to advance their knowledge.

16. What are the responsibilities of a Scrum Master?

Ans. Key responsibilities of a Scrum Master involve:

(i) Tracking and monitoring

(ii) Understanding requirements properly

(iii) Working to reach the project goal

(iv) Processing checking master and quality master

(v) Protecting the team from detachments

(vi) Improving the performance of the team

(vii) Leading the meetings and resolving issues

(viii) Resolving conflicts and impediments

(ix) Communicating and reporting

17. What is a use-case?

Ans. Each use-case describes a unit of complete and useful functionality that business or system provides to its users. It also describes how external user interacts with a system to achieve a desired result. Some examples of use-cases: Hire employee, Buy ticket, Place order, Deposit funds.

18. What is actor in use-case diagrams?

Ans. Actor represents some group (called "role") of external customers or users of the business or software system. Actors have some needs and require specific services from the system. Some examples of actors are: Customer, Student and Passenger. Actor could be not only human but also another system, business, or device.

19. What is the difference between use-case diagram and use-case?

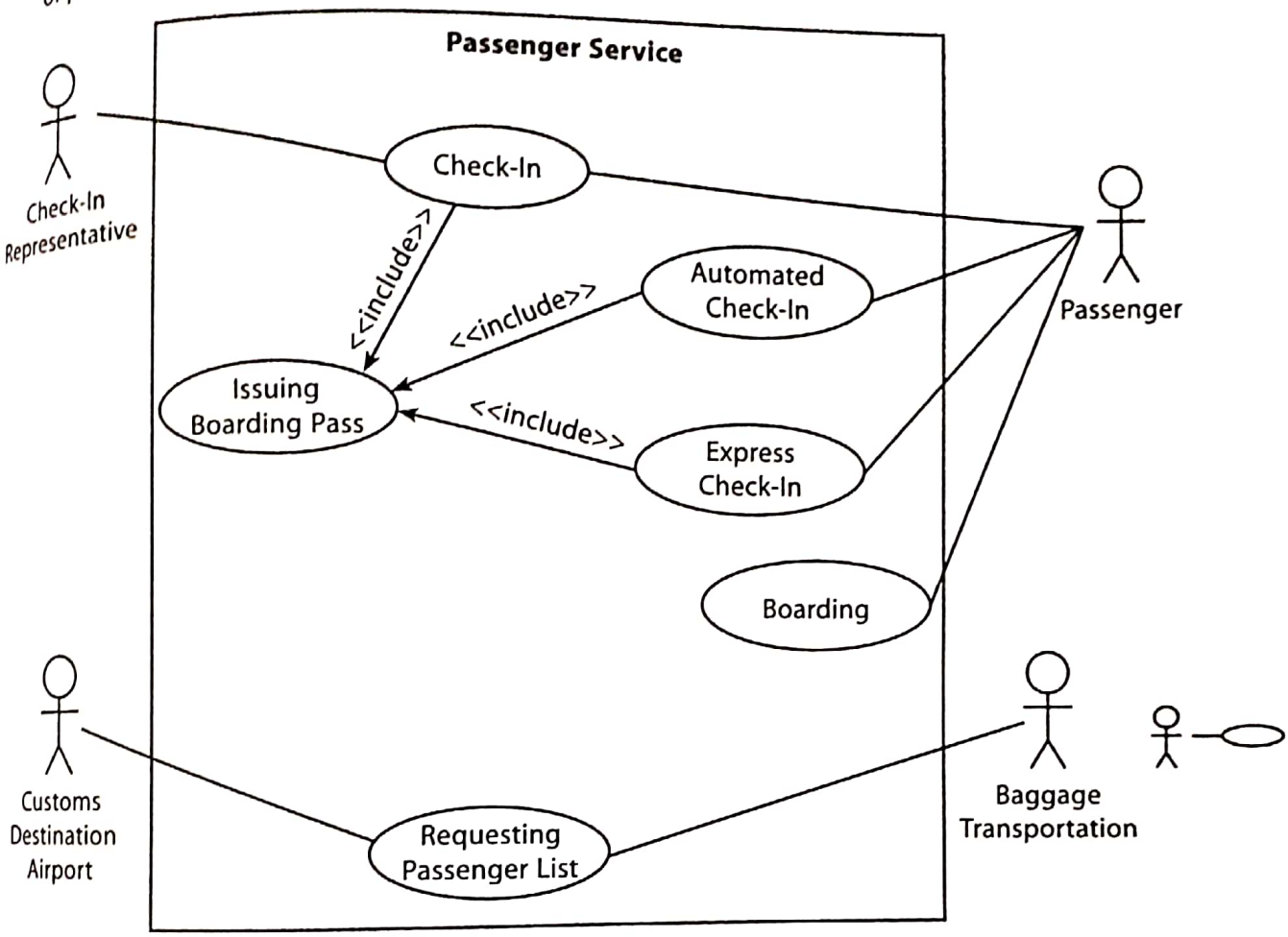
Ans. Use-case diagram shows business or system, its external users, and use-cases applicable to the system. Use case represents one specific goal or need of the user from the system.

20. Draw a use-case diagram for airport check-in and security screening.

Ans. Airport check-in and security screening business model

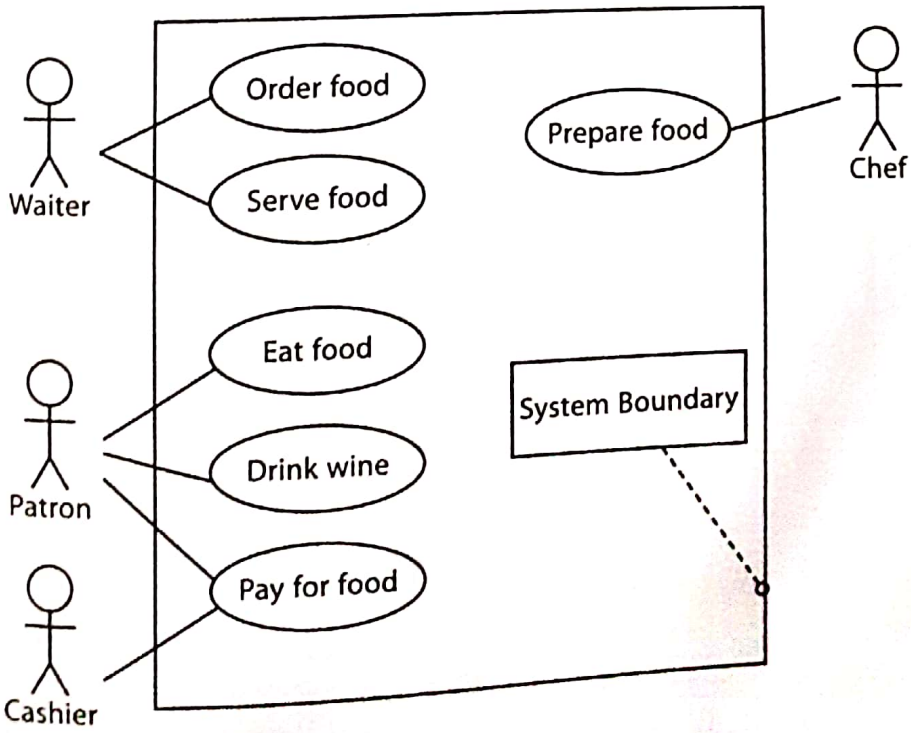
Purpose: An example of a business use-case diagram for airport check-in and security screening.

Summary: Business use-cases are Individual Check-In, Group Check-In (for groups of tourists), Security Screening, etc., representing business functions or processes taking place at an airport and serving needs of passengers.



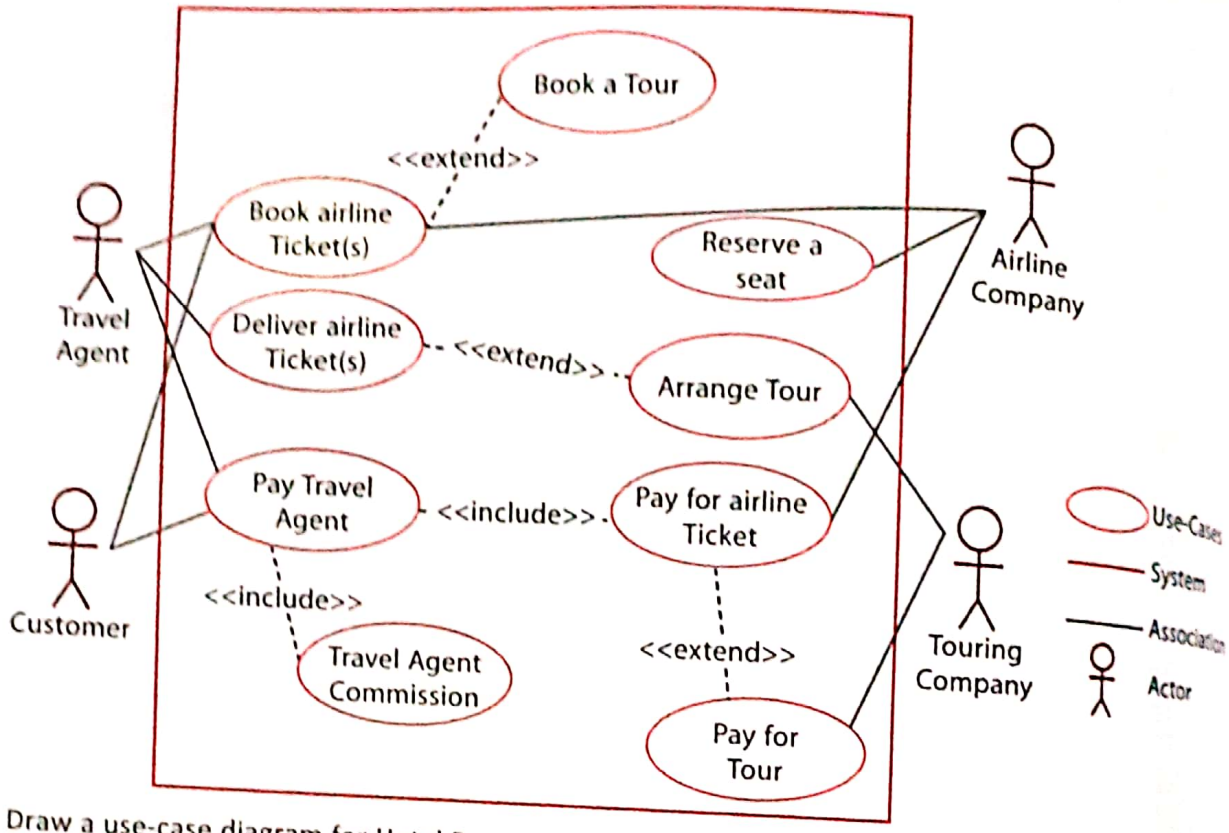
21. Draw a use-case diagram for Restaurant Management System.

Ans.



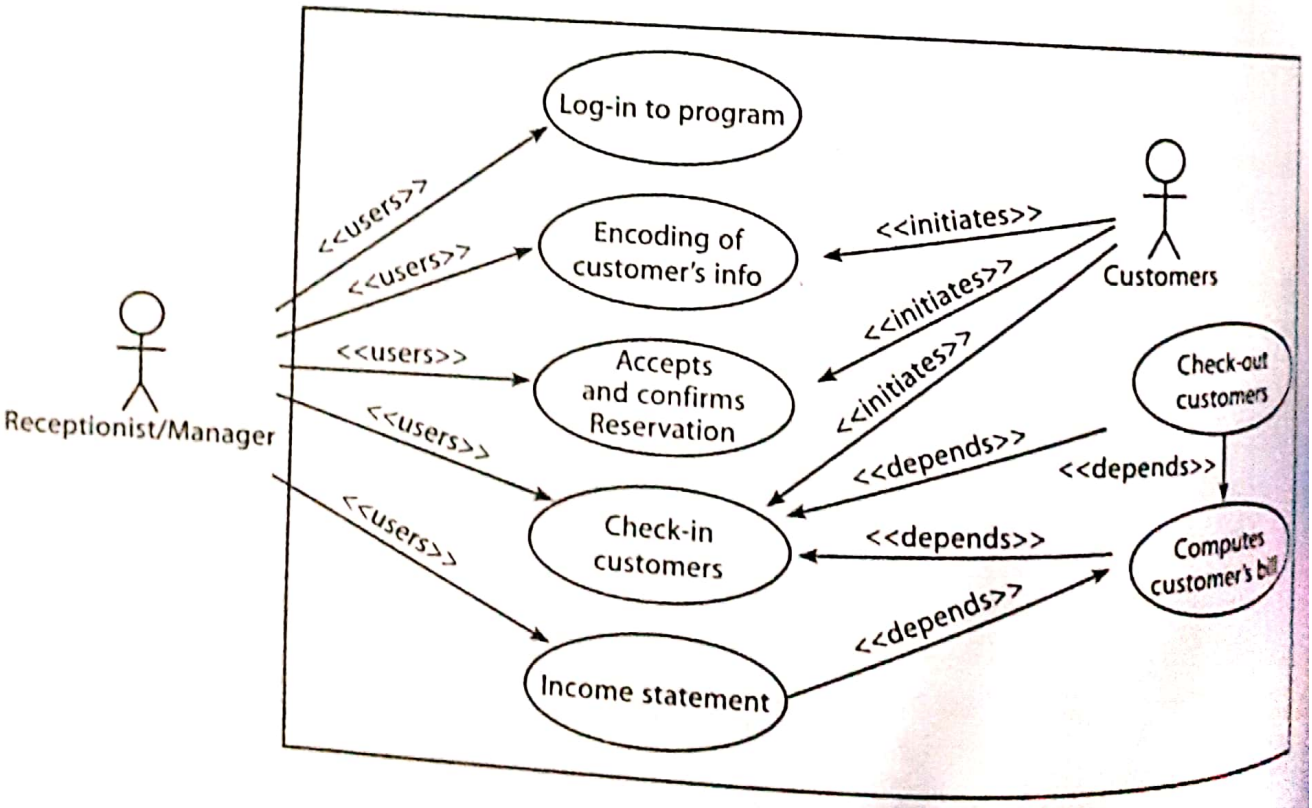
22. Draw a use-case diagram for a Travel agency.

Ans.



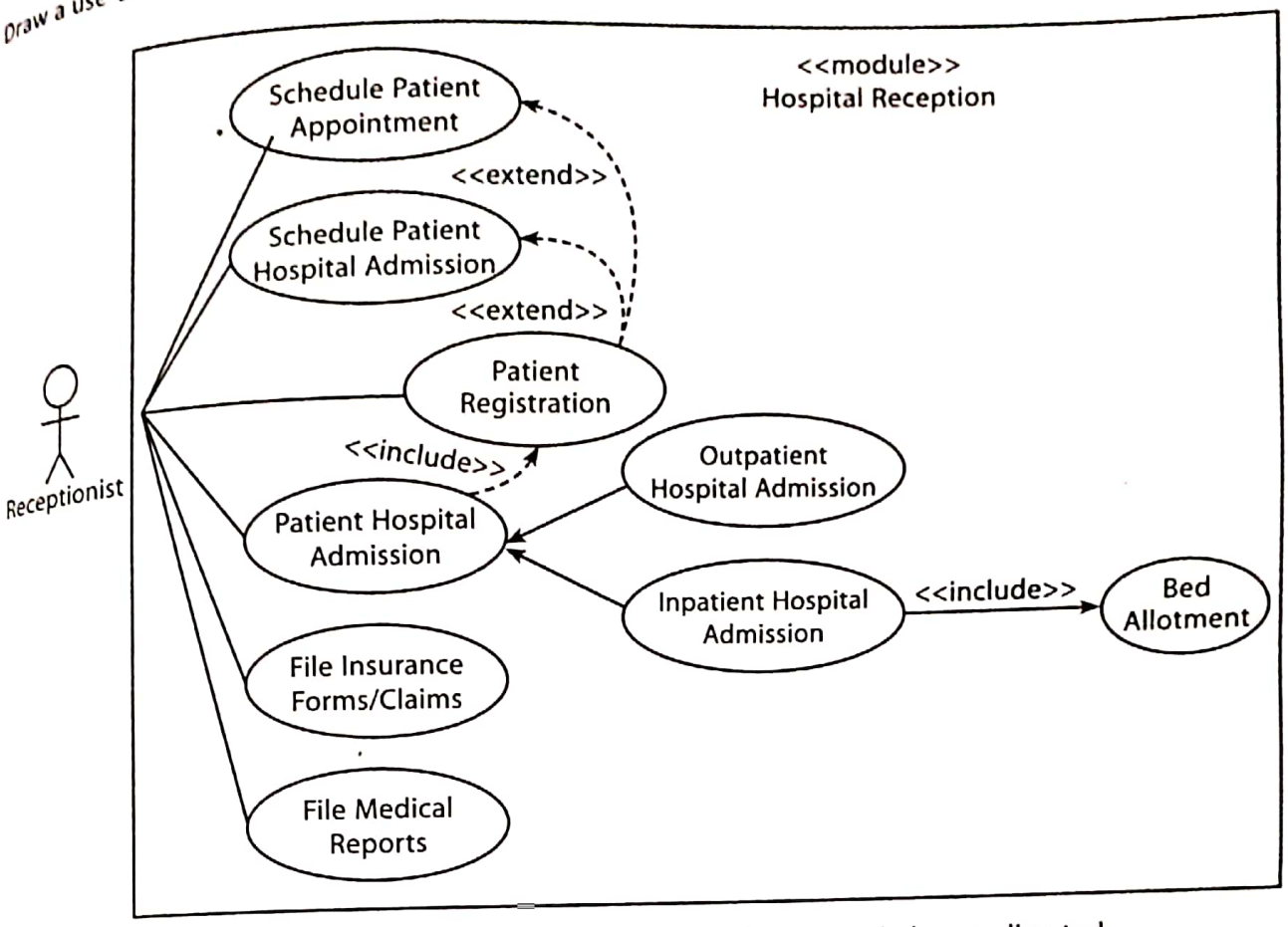
23. Draw a use-case diagram for Hotel Reservation System.

Ans.



24. Draw a use-case diagram for Hospital Management System.

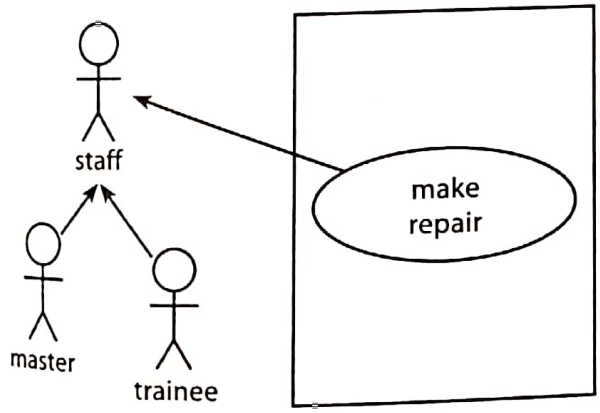
Ans.



25. Draw a use-case diagram and identify the actors for the situations given below as directed:
[CBSE Sample Paper 2020]

Ans. (i) A repair can be made by a master, a trainee or any other repair shop employee.

(ii) Consider an ATM system. Identify at least three different actors that interact with this system.

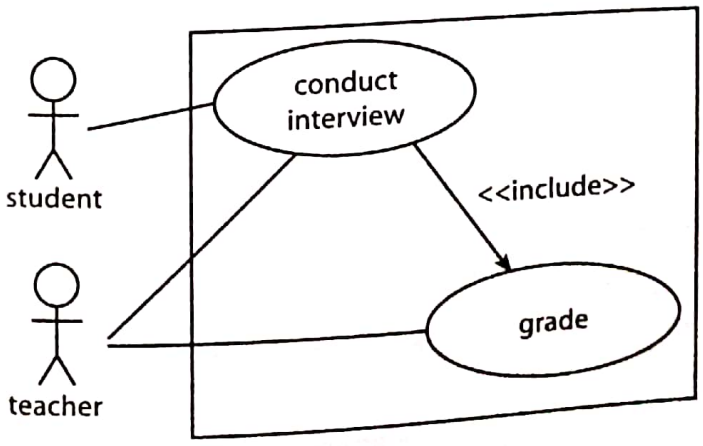


Actors: Master, Trainee

An actor is any entity (user or system) that interacts with the system of interest. For an ATM, this includes:

- Bank Customer
- ATM Maintainer
- Central Bank Computer

26. Look at the following use-case diagrams and write the actors and the situation depicted by the use-case diagrams :



A web framework provides the following four basic things—

- URL (Uniform Resource Locator) mapping
- Database manipulation
- Templating
- Security measures

A few examples of web framework commonly used nowadays are:

CLIENT SIDE	SERVER SIDE
FLASK (PYTHON)	ASP.NET Core
EXPRESS (Java/Script)	Zend Framework
Ruby on Rails (Ruby)	Django
Mojolicious (Perl)	

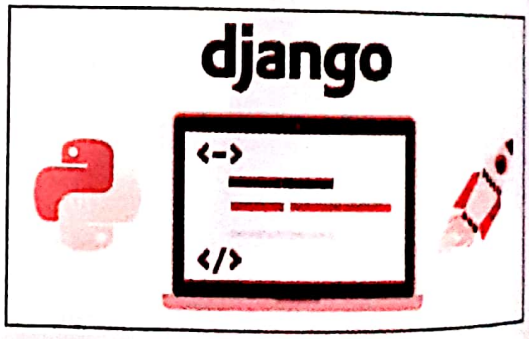
CTM: A web framework is a set of components that helps us develop websites faster and easier.

6.3 WHAT IS DJANGO

Django is pronounced as JANG-oh. It is a high-level Python web framework that encourages rapid development and clean, simple design. It is a web application development framework. Django helps eliminate repetitive tasks making the development process an easy and time-saving experience. It is free and open source.

It was named after the famous guitarist Django Reinhardt. It was developed by Adrian Holovaty and Simon Willison at the *Lawrence Journal-World* newspaper for efficient development in Python. It was open sourced in 2005 and its first version was released on September 3, 2008.

It follows the principle of "Don't Repeat Yourself", which means keeping the code simple and non-repeating. Django is also a high level, MVT architect, which stands for Model View Template, which we will discuss in detail in the following sections.



A similar set of components can be seen each time we build a new website, such as user authentication (signing up, signing in, signing out), process forms, upload files, etc.

Django is a framework that gives us ready-made components to use. It helps in eliminating repetitive tasks, building and maintaining web applications and, hence, makes the development process easy and time-saving as it involves less coding.

One more significant feature of Django is **reusability** as once an application is developed, it can be used in many websites.

CTM: Django is a web development framework that assists in building and maintaining quality web applications.

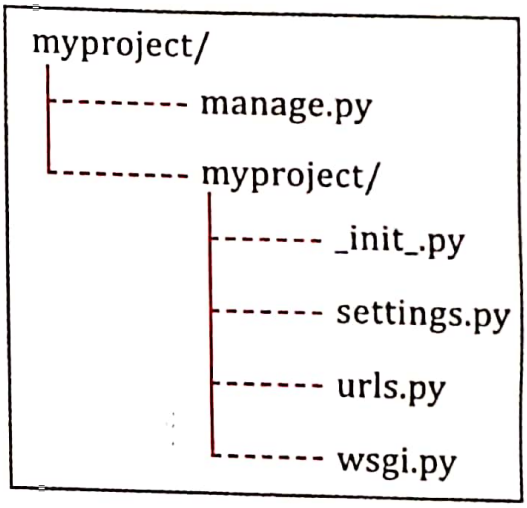
Note: Django is a registered trademark of the Django Software Foundation and is licensed under BSD Licence.

7. Again type **cd newproject**

After this, the command prompt shall be—

C:\myproject\newproject\newproject

8. Type dir to view all the sub-folders created inside this application.



The "myproject" folder is just a project container; it actually contains two elements—

(a) **manage.py:** This file is a kind of your project local django-admin for interacting with your project via command line (start the development server, sync db...).

(b) **The "newproject" subfolder:** This folder is the actual Python package of your project. It contains four files—

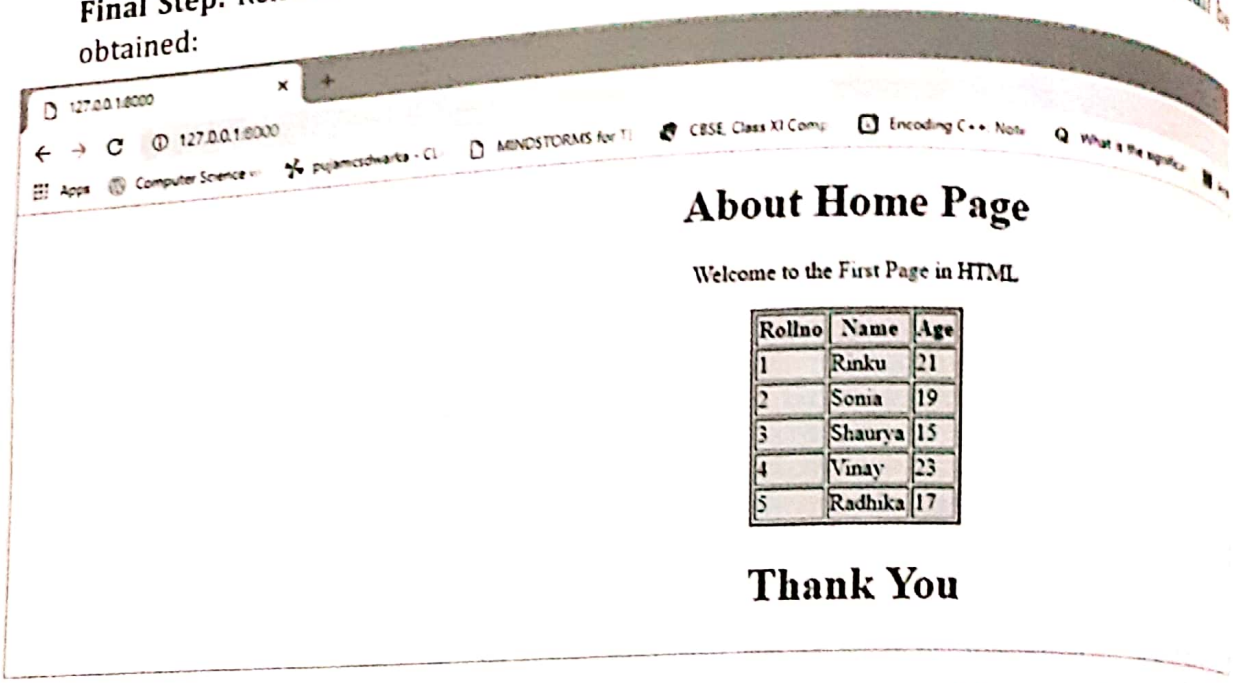
(i) **__init__.py:** Just for Python, treat this folder as package.

(ii) **settings.py:** As the name indicates, it refers to your project settings. It manages the settings of your project.

(iii) **urls.py:** All links of your project and the function to call.

(iv) **wsgi.py:** If you need to deploy your project over WSGI. (The Web Server Gateway Interface (WSGI) is a simple calling convention for web servers to forward requests to web applications or frameworks written in the Python programming language). It serves as an entry point for WSGI compatible web servers.

Final Step: Refresh the home page of your web browser. The following output shall be obtained:



In the above application, we have created Django application comprising of tables in HTML. Here, we will discuss basic HTML tags and tags required for creating tables.

Basic HTML Structure

Any HTML document contains:

(a) Information/text that is to be displayed by the web browser.

(b) Editing and formatting information that tells the web browser about the layout of the web page to be displayed.

When an HTML document is opened in a web browser, the browser interprets these elements and displays the HTML document as a web page. The relationship between an HTML document and a web page can be explained thus: An HTML document can be compared to a text file where all the HTML elements contained inside this document can be compared as its respective pages. In turn, an HTML document shown by a web browser is called a web page.

Any HTML document, in general, contains at least four elements — HTML, HEAD, TITLE and BODY. These elements are specified by the following respective tags:

1. <HTML> . . . </HTML>
2. <HEAD> . . . </HEAD>
3. <TITLE>. . . </TITLE>
4. <BODY> . . . </BODY>

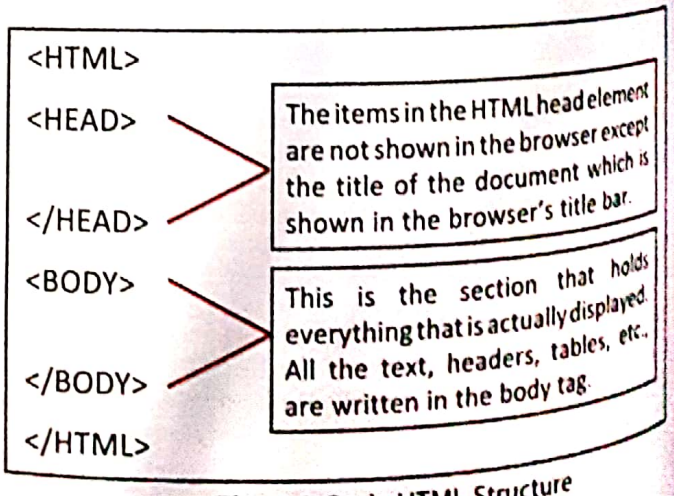


Fig. 6.4: Basic HTML Structure

Every HTML tag is accompanied by some attributes. An attribute is a keyword associated with a tag which provides extra information or some specific purpose to a tag. **Attributes are always used in the opening tag. The value or name assigned to an attribute is always enclosed in quotes.**

6.10 GET AND POST METHODS

GET and POST are the only HTTP methods used when dealing with forms.

Django's login form is returned using the POST method in which the browser bundles up the form data, encodes it for transmission, sends it to the server, and then receives back its response.

6.10.1 Difference between GET and POST Methods

- GET bundles the submitted data into a string and uses it to compose a URL. The URL contains the address where the data must be sent, as well as the data keys and values. We can see this in action if we do a search in the Django documentation, which will produce a URL of the form *https://docs.djangoproject.com/search/?q=forms&release=1*.
- Any request that could be used to change the state of the system, *for example*, a request that makes changes in the database, should use POST. GET should be used only for requests that do not affect the state of the system.

- GET would also be unsuitable for a password form because the password would appear in the URL and, thus, also in browser history and server logs, all in plain text. Neither would it be suitable for large quantities of data or for binary data such as an image. POST, coupled with other protections like Django's CSRF protection, offers more control over access.
- On the other hand, GET is suitable for things like a web search form because the URLs that represent a GET request can easily be bookmarked, shared or resubmitted.

Django handles three distinct parts of the work involved in forms:

- Preparing and restructuring data to make it ready for rendering
- Creating HTML forms for the data
- Receiving and processing submitted forms and data from the client

6.11 WORKING WITH FLAT FILES AND CSV FILES

Files are classified into two types—Flat files and CSV files.

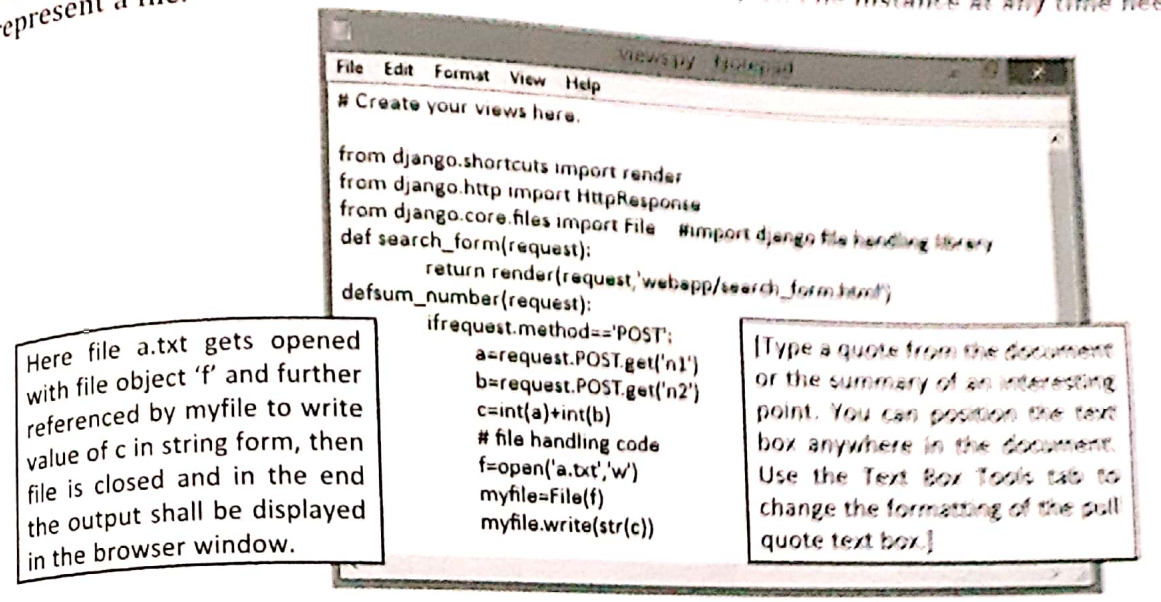
A flat file is a file that stores data in a plain text file. Each line of the text file holds one record, with fields separated by delimiters, such as commas or tabs. While it uses a simple structure, a flat file database cannot contain multiple tables like a relational database can.

On the other hand, a Comma Separated Values (CSV) file is a plain text file that contains a list of data. These files are often used for exchanging data between different applications. *For example*, databases and contact managers often support CSV files.

These files may sometimes be called Character Separated Values or Comma Delimited files. They mostly use the comma character to separate (or delimit) data, but sometimes use other characters like semicolons. The idea is that you can export complex data from one application to a CSV file and then import the data in that CSV file into another application. We will now discuss how these files are handled in Django.

6.11.1 Write the Fields to a Flat File

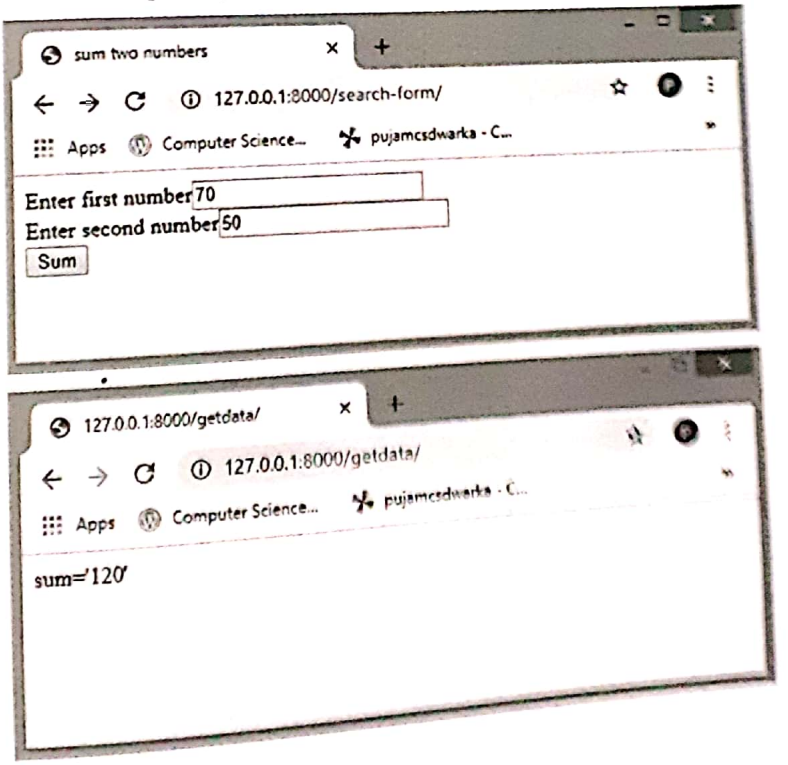
Django uses django.core.files internally through the file object. File instance at any time needs to represent a file.



1. Open the command prompt and move to c:\test1\myproject
2. Type command—python manage.py runserver
3. Open any browser and type the url as *http://127.0.0.1:8000/search-form/*

   It will display two text boxes, after entering a value and click on sum button. Result will be stored in a file a.txt(@C:\test1\myproject) as well as displayed over browser.

   We can also view the output by opening a.txt file also.



6.11.2 Write the Fields to a CSV File

Python comes with an inbuilt library for handling CSV files named CSV. While working with Django, the CSV module file creation capability operates on file-like objects. It also treats HttpResponse objects like a file object. Let us modify views.py file for CSV file creation as shown below:

views.py - Notepad

File Edit Format View Help

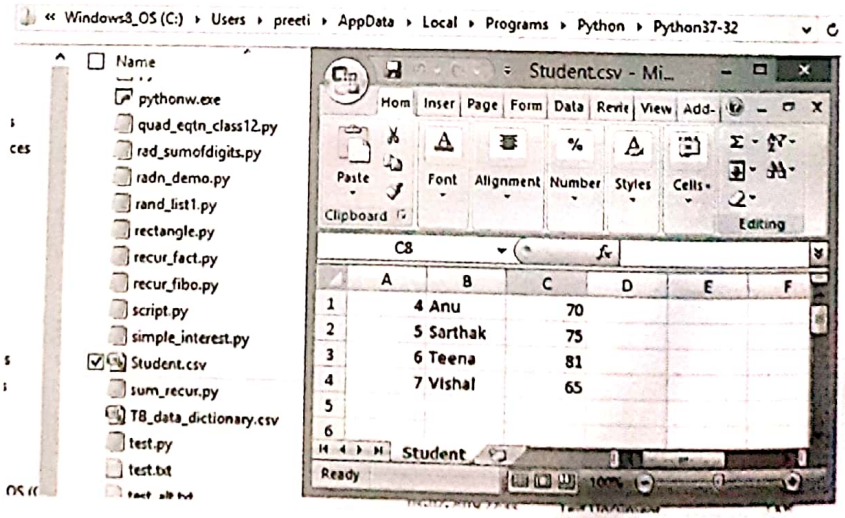
```
from django.shortcuts import render
from django.http import HttpResponse
import csv
def search_form(request):
        return render(request,'webapp/search_form.html')
def sum_number(request):
        if request.method=='POST':
                a=request.POST.get('n1')
                b=request.POST.get('n2')
                c=int(a)+int(b)
                #write csv file
                response=HttpResponse(content_type='text/csv')
                response['Content-Disposition']='attachment; filename="mycsvfile.csv"'
                writer=csv.writer(response)
                writer.writerow([a,b,c])
                return response
```

In the above program, following points are to be noted:

- The response gets a special MIME type-text/csv. This tells browsers that the document is a CSV file, rather than an HTML file.
- The response gets an additional Content-Disposition header, which contains the name of the CSV file.
- Hooking into the CSV-generation APP is easy—just pass response as the first argument to csv.writer. The csv.writer function expects a file-like object and HttpResponse objects.
- For each row in your CSV file, call writer.writerow, passing it an iterable object such as a list or tuple.

6.11.3 Read the Fields from a CSV File

For reading a CSV file, reader() function of CSV module is required to be given. But to read CSV, the first thing we need is a CSV file itself. So in your project, create a new file named Student.**csv** with the relevant data in it.



Create a file in Python using Python reader object with the following code:

```
import csv
with open("Student.csv", 'r') as csvfile:
reader = csv.reader(csvfile)
for row in reader:
 print(row)
```

```
#Reading the contents from a csv file
import csv

with open("Student.csv","r") as csvfile:
    reader = csv.reader(csvfile)

    for row in reader:
        print(row)
```

```
Python 3.7.0 Shell
File Edit Shell Debug Options Window Help
>>>
 RESTART: C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_read_CSVdjango.py
['4', 'Anu', '70']
['5', 'Sarthak', '75']
['6', 'Teena', '81']
['7', 'Vishal', '65']
>>>
Ln: 9 Col: 4
```

Reading the contents from Student.csv

In a Nutshell

1. Install django (**pip install django**).
2. Go to any folder in CMD, type **django-admin startproject <projectName> (*e.g.*, django-admin startproject myproject)**.
3. Enter myproject folder by typing **CD myproject** in CMD.
4. Create web application by typing **django-admin startproject <appName> (*e.g.*, django-admin startproject myapp)**.
5. Using File Explorer, go to myapp folder, create **urls.py** file (as in Slide 13) and modify **views.py** file (as in slide 12) to create responses.
6. Using File Explorer, go to myproject folder, edit **urls.py** file (as in Slide 14) to add the pattern for our web application **myapp**.
7. Using File Explorer, go to myproject folder, edit **settings.py** file (as in Slide 15) to install our web application **myapp**.
8. In CMD, browse to myproject directory and type **python manage.py runserver** to deploy our server and leave it open (if you didn't close command prompt after Step 4, then you are good to go).
9. Go to web browser and enter url(s) to get the responses from Django server.

MEMORY BYTES

- Django is a high-level Python web development framework.
- It is free and open source.
- A web framework is a set of components that helps develop websites faster and more easily.
- A similar set of components can be seen each time we build a new website, *e.g.*, user authentication (signing up, signing in, signing out), forms, upload files, etc.
- Django is a framework that gives you readymade components to use.
- It helps in building and maintaining web applications.
- Django comes with a lightweight web server for developing and testing applications.

- **django-admin startproject mysite** command will create our project (files are created in the specified directory which you can view in explorer).
- **python manage.py startserver** command helps check whether our server is running successfully or not.
- Model in Django project refers to the data management component(s).
- **Views** determine what content is displayed on a given page.
- **URLConfs** determine where that content is going.
- An HTTP GET Request refers to the method of retrieving or extracting information from the given server using a given URL on the web.
- An HTTP POST Request refers to the method of sending the data/information to the server using HTML forms through online processing.
- A Django project refers to a complete application and an app is described as a submodule of the project that caters to some specific functionality.
- The mapping between URL paths to the views of Django project are called URL Confs (URL configurations). The views in Django project are the Python functions.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.
   (a) Whenever a web client has to access a web page, it makes a ..... request and sends URL of the web page.
   (b) Django web framework is managed and maintained by an independent and non-profit organization named ..... .
   (c) PIP stands for ..... .
   (d) ..... is used to install Django command.
   (e) ..... command is used to create project in Django.
   (f) ..... command is used to run Django server.
   (g) The IP address of local host is ..... .
   (h) The default port of Django server is ..... .
   (i) An app is a ..... server to one portion of project.
   (j) ..... file must be present/created in the root folder to treat it like a package or subpackage.
   (k) ..... folder of a Django application holds the changes in models.py and updates database accordingly.

Answers: (a) GET
(b) Django Software Foundation
(c) pip Installs Packages Or pip Installs Python
(d) pip install django
(e) Django_admin start project name
(f) Python manage .py run server
(g) 127.0.01
(h) 8000
(i) submodule
(j) _init_.py
(k) migrations

2. State whether the following statements are True or False.
   (a) An HTTP Get request uses URL to send information to web server.
   (b) The default IP address of local host is 127.0.01 and default port number for Django is 800.
   (c) Isolated Python environments take care of inter-dependencies and allow users to develop different applications using common global set of libraries.
   (d) An app refers to the entire application but a project refers to the submodule.
   (e) A folder or subfolder will be treated as package or subpackage if it contains fill_init_.py .
   (f) Manage.py is used to manage the project, perform administrative tasks and run built-in server.
   (g) The software architecture used by Django is model view template.
   (h) View is used to define business logic and it accesses model data and redirects it to the template.
   (i) Every model class represents a table in databases and is always a subclass model.
   (j) The process of linking URL and displaying actual template is called URL routing.

Answers: (a) True (b) True (c) False (d) False (e) True (f) True
(g) True (h) True (i) True (j) True

3. Multiple Choice Questions (MCQs)

(a) Which of the following is the HTTP request method:
(i) HTML (ii) FTP (iii) TELNET (iv) POST

(b) What is the full form of HTTP?
(i) High Trent Transfer Process (ii) Hyper Text Telnet Protocol
(iii) Hyper Text Transfer Protocol (iv) Homogenous Text Transfer Protocol

(c) The command to run django server is:
(i) python manage.py run (ii) python manage.py runserver
(iii) python manage.py server (iv) python manage.py execute

(d) The IP address of localhost is
(i) 127.0.0.1 (ii) 127.0.0.0 (iii) 127.1.1.1 (iv) 192.168.0.1

(e) What is django?
(i) A web framework to create dynamic application.
(ii) A framework to create desktop application.
(iii) A framework to create mobile application.
(iv) A library to perform matrix operations.

(f) Django architecture consists of:
(i) Models (ii) Views (iii) Templates (iv) All of the above

(g) Which of the following is a django configuration file:
(i) Urls.py (ii) Settings.py (iii) Wsgi.py (iv) Init.py

(h) Which command do we use to create a project?
(i) python manage.py startproject <project-name>
(ii) django-admin startapp < project-name>
(iii) python manage.py startapp <project-name>
(iv) django-admin startproject <project-name>

(i) Which file is used to store location information in django?
(i) urls.py (ii) settings.py (iii) wsgi.py (iv) init.py

(j) Which file is used to manage django projects and perform administrative tasks?
(i) urls.py (ii) settings.py (iii) manage.py (iv) init.py

(k) What is the full form of CSV?
(i) Comma slashed file (ii) Comma separated ~~file~~ Values
(iii) Common source file (iv) Cumulative sum file

(l) Which file is used to define database model?
(i) urls.py (ii) database.py (iii) models.py (iv) init.py

Answers: (a) (iv) (b) (iii) (c) (ii) (d) (i) (e) (i) (f) (iv)
(g) (ii) (h) (iv) (i) (i) (j) (iii) (k) (ii) (l) (iii)

SOLVED QUESTIONS

1. What is the command to install Django?
Ans. pip install django

2. Write the command to create a project.
Ans. django-admin startproject <projectname>

3. How do you start a server?
Ans. python manage.py runserver

4. Write the command to create a new web application.
Ans. django-admin startapp <appname>
For example: in order to create a new app named "users" in an existing project, the command will be
python manage.py startapp users.

5. What are ^ and $ symbols in the regular expressions?
Ans. ^ means beginning of url and $ specifies end of the url.

6. Which file is edited to install the app?

Ans. settings.py

7. Why do we require virtual environment?

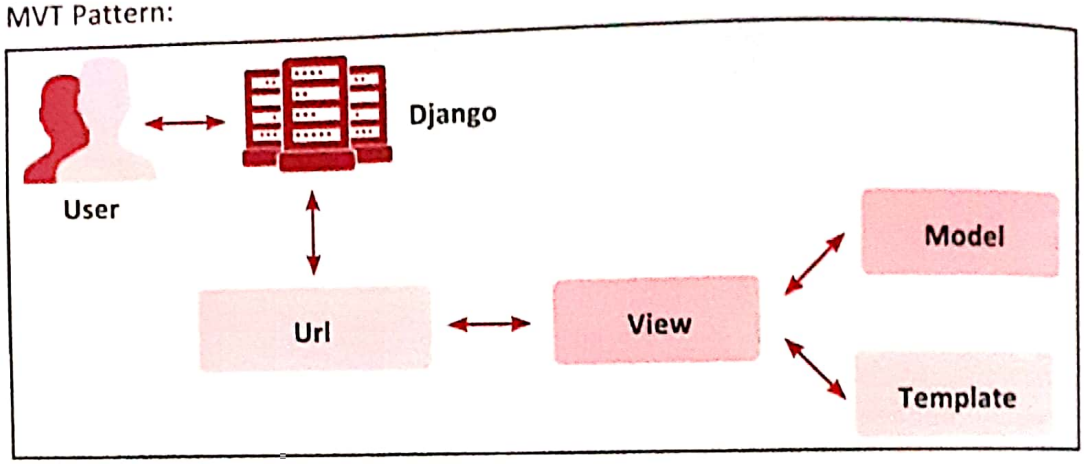
Ans. We create django project with a virtual environment to make it portable, *i.e.*, to allow it to run on any other system without installing libraries and packages required for the said project.

8. What do you understand by Django?

Ans. Django is a free and open source web application framework, written in Python. Django is named after Django Reinhardt, one of the best Jazz guitarists of all time. Django was mainly developed to handle the intensive design of the newsroom. We can even build high-quality web applications using it. It adheres to the DRY principle and focuses on automating as much as possible.

9. Discuss the Django architecture.

Ans. Django MVT Pattern:



Django Architecture

The developer provides the Model, the View and the Template, then just maps it to a URL and Django does the magic to serve it to the user.

10. Explain how you can set up Database in Django.

Ans. We can use the command edit mysite/setting.py; it is a normal Python module with module level representing Django settings.

Django uses SQLite by default; it is easy for Django users and as such doesn't require any other type of installation. In case you are using any other database, other than SQLite, then you have to change the 'default' settings for Django to match with your selected database connection settings.

11. Explain how you can write a View in Django.

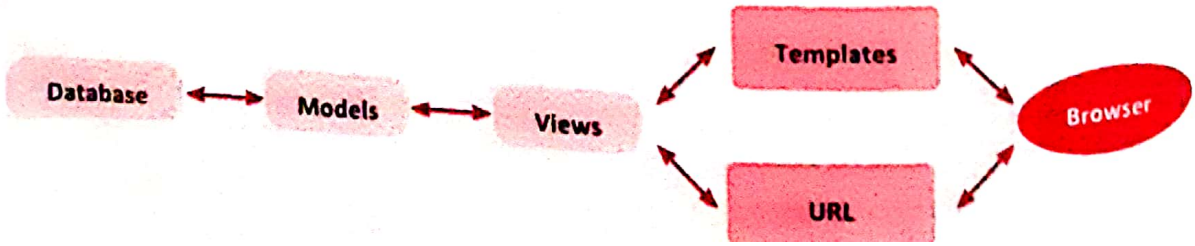
Ans. We can write a View in Django in the following way:

```
(i)   from django.http import HttpResponse
(ii)  import datetime
(iii) def Current_datetime(request):
(iv)  now = datetime.datetime.now()
(v)   html = "<html><body>It is now %s</body></html>" % now
(vi)  return HttpResponse(html)
```

Returns the current date and time as an HTML document.

12. Mention what the Django template consists of.

Ans. The template is a simple text file. It can create any text-based format like XML, CSV, HTML, etc. A template contains variables that get replaced with values when the template is evaluated and tags (% tag %) that control the logic of the template.



13. Specify the architecture of Django.

Ans. Django follows MVC-MVT architecture. MVT stands for Model View Template design pattern which is a bit different from MVC (Model View Controller) pattern.

14. How does Django work?

Ans. Django can be broken into several components:

- **Models.py file:** This file defines our data model by extending our single code line into full database tables and adds a pre-built administration section to manage content.
- **Urls.py file:** It uses a habitual expression to confine URL patterns for processing.
- **Views.py file:** It is the main part of Django. The actual processing happens in View.

15. Name the features available in Django web framework.

Ans. Features available in Django web framework are:

- Admin Interface (CRUD)
- Form handling
- Session, user management, role-based permissions
- Testing Framework
- Templating
- Internationalization
- Object-relational mapping (ORM)

16. List the differences between the GET and POST methods while working with forms.

Ans.

GET Method	POST Method
• GET requests can be cached • GET requests remain in the browser history • GET requests can be bookmarked • GET requests have length restrictions • GET requests are only used to request data (not modifiable)	• POST requests are never cached • POST requests do not remain in the browser history • POST requests cannot be bookmarked • POST requests have no restrictions on data length • URL query string is encoded, so it cannot be used for malicious purposes

17. What are the advantages of using Django for web development?

Ans. The several advantages of using Django for web development are:

- Auto-generated web admin is provided by Django to make website administration easy.
- Pre-packaged API is also available for common user tasks.
- Business logic can be separated from the HTML using this framework.
- You can even divide the code modules into logical groups so as to make it flexible for changing.
- Template system is being provided so as to define HTML template for your web page to avoid code duplication.

18. What is the process of creating a project in Django?

Ans. To start a project in Django, use the command $django-admin.py and then use the following command:

```
Project
_init_.py
manage.py
settings.py
urls.py
```

19. How will you define Django programming language?

Ans. Django is a web framework in Python to develop a web application. Django is a free and open-source web application framework, written in Python. Django makes it easier to build better web applications quickly and with lesser code.

20. Develop a Django-based application that reads every line from a flat file using split() method. Each word read from the line is appended as html text. At the same time, the program fetches input in three vars a, b and c from the user. Write these values to a flat (.txt) file.

Ans. #Make changes in views.py as shown in the code given below for write/read operation

```
from django.shortcuts import render
from django.http import HttpResponse
from django.core.files import File
def search_form(request):
    return render(request,'webapp/search_form.html')
def sum_number(request):
```

```
if request.method=='POST':
 a=request.POST.get('n1')
 b=request.POST.get('n2')
 c=int(a)*int(b)
 f=open('a.txt','w')
 myfile=File(f)
 myfile.write(str(a)+" ")
 myfile.write(str(b)+" ")
 myfile.write(str(c)+" ")
 myfile.close()
 f=open('a.txt','r')
 t=''
 for text in f.readlines():
   for word in text.split():
     t=t+word+'<br>'
 myfile.close()
 return HttpResponse(t)
```

Write the value of a, b and c on a .txt file separated by space through file object's write method.

Read each line from the file through outer for loop and each word is separated in inner for loop with the help of split method. Each word is appended as html text in 't' variable along with <br> tag

UNSOLVED QUESTIONS

1. What is Django?
2. Write a code to create a directory "myproject" for a new project.
3. What will be the structure of the above folder "myproject"?
4. Assume you are in your project folder "myproject". Write a code to create a web application "myapp"
5. Write a code to create a simple view in myapp to say "welcome to my app!".
6. Write the use of GET and POST HTTP requests in Django.
7. Develop a web-based application using Django to calculate the sum of two numbers.
8. What are the advantages of using Django?
9. Write steps to install Django on Windows platform.
10. Write steps to create a web project.
11. Explain the project structure of a Django project.
12. Write steps to create Django application. Also explain the application structure.
13. What is Admin Interface for Django Web application?
14. What is the use of Django POST method? Write an example code using POST method.
15. What is the use of Django GET method? Write an example code using GET method.
16. What are CSV files? List its properties.
17. What is the difference between parse a GET and parse a POST request in Django application?
18. How can you create the Flat file and CSV files?
19. Write the codes to create a web page in Django; the background colour of the page is BLUE and content colour RED.
20. What do you understand by web framework?
21. What does MVC-MVT stand for?
22. When do we use GET and when do we use POST method in form?
23. What are the components of Django?
24. Develop a web application which prompts two numbers and displays the sum of these two numbers after clicking the sum button.
25. Develop a web application which prompts principal amount, rate and time and displays simple interest.
26. Develop a web server application for a survey/poll and capture the response in CSV file.

Interface Python with SQL

7.1 INTRODUCTION

Databases play a vital role in the efficient working of an organization. From a construction firm to a stock exchange, every organization depends on large databases. These are essentially collections of tables and are connected to each other through fields or, more precisely, columns/attributes. These database systems support SQL, the Structured Query Language, which is used to create, access and manipulate data. SQL is used to access data, and also to create and exploit the relationship between stored data. Additionally, these databases support database normalization rules for avoiding redundancy of data. The Python programming language has powerful features for database programming. Python supports various databases like MySQL, Oracle, Sybase, PostgreSQL, etc. Python also supports Data Definition Language (DDL), Data Manipulation Language (DML) and Data Query Statements. For database programming, the Python DB-API is a widely-used module that provides a database application programming interface. It's a standard for database interfaces.

Most Python database interfaces adhere to this standard.

We can choose the right database for our application. Python Database API supports a wide range of database servers such as—

- GadFly
- mSQL
- MySQL
- PostgreSQL
- Microsoft SQL Server 2000
- Informix
- Interbase
- Oracle
- Sybase

7.4.2 Steps for Creating Database Connectivity Applications

There are mainly eight steps that must be followed in order to create a database connectivity application.

Step 1: Start Python.

Step 2: Import packages required for establishing connectivity.

Step 3: Create Database.

Step 4: Open and establish a connection to database.

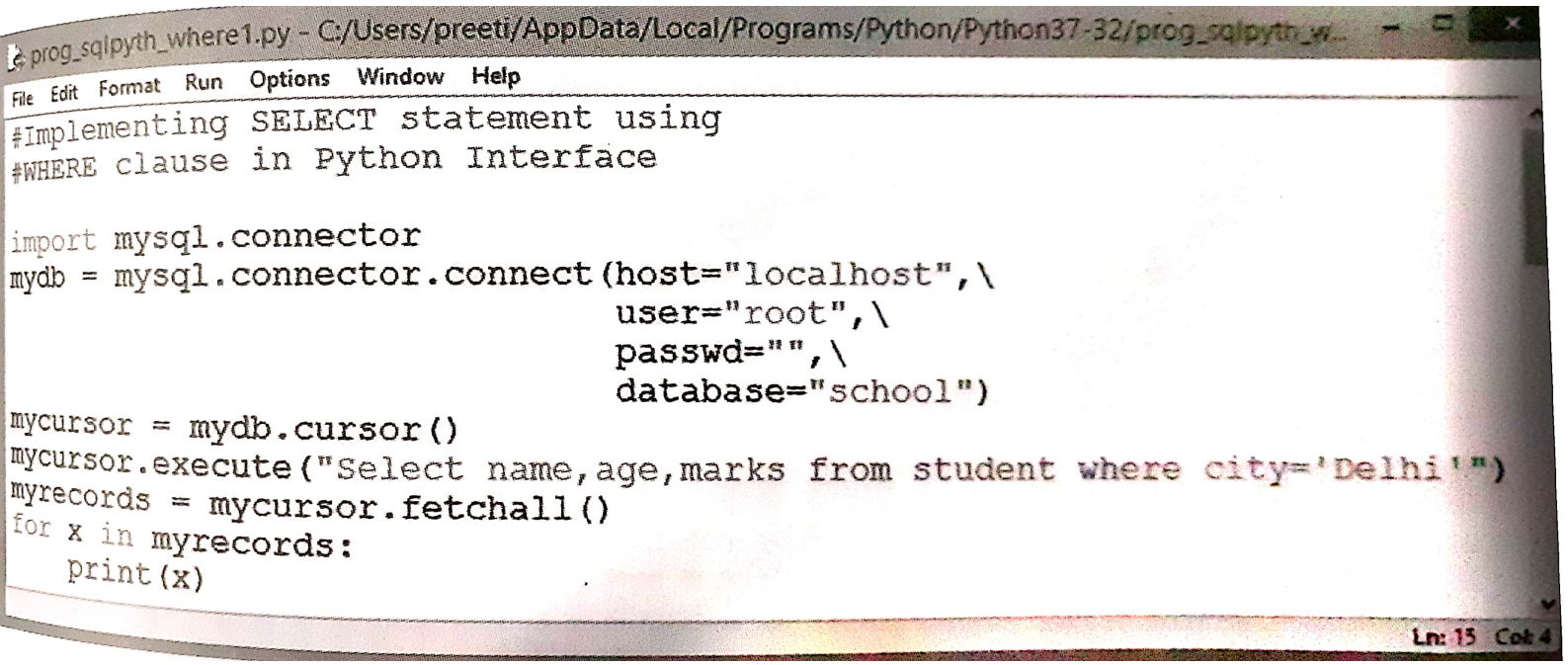
Step 5: Create a cursor object or instance.

Step 6: Execute a query.

Step 7: Extract data from result set.

Step 8: Clean up the environment.

We will now discuss all these steps in detail. The first three steps have already been discussed in the previous section. The next step to be discussed in succession is establishing connection.



7.9 OPERATIONS ON A TABLE IN A NUTSHELL

Following is the program for performing all the operations on a table 'student' through a menu-driven program.

prog_sqlpyth_complete.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_sqlpyth_complete.py (3.7.0)

File Edit Format Run Options Window Help

```
#Menu-driven program to demonstrate four major operations
#performed on a table through MySQL-Python connectivity
def menu():
    c='y'
    while (c=='y'):
        print ("1. add record")
        print ("2. update record ")
        print ("3. delete record")
        print("4. display records")
        print("5. Exiting")
        choice=int(input("Enter your choice: "))
        if choice == 1:
            adddata()
        elif choice== 2:
            updatedata()
        elif choice== 3:
            deldata()
        elif choice== 4:
            fetchdata()
        elif choice == 5:
            print("Exiting")
            break
    else:
        print("wrong input")
c=input("Do you want to continue or not: ")
def fetchdata():
    import mysql.connector
    try:
        db = mysql.connector.connect(host="localhost",user="root",password='',database='sl')
        cursor = db.cursor()
        cursor.execute("SELECT * FROM student" )
        results = cursor.fetchall()
        for x in results:
            print(x)
    except:
        print ("Error: unable to fetch data")

def adddata():
    import mysql.connector
    db = mysql.connector.connect(host='localhost',user='root',password='',database='sl')
    cursor = db.cursor()
    cursor.execute("INSERT INTO student VALUES('Ritu',4000,'Science',345,'B','11')")
    cursor.execute("INSERT INTO student VALUES('Ankush',6000,'Commce',445,'A','12')")
    cursor.execute("INSERT INTO student VALUES('Pihu',3566,'Humanis',446,'A','11')")
    cursor.execute("INSERT INTO student VALUES('Tinku',8900,'Science',545,'A+','12')")
    db.commit()
    print("Records added")

def updatedata():
    import mysql.connector
    try:
        db = mysql.connector.connect(host="localhost",user="root",password='',database='sl')
        cursor = db.cursor()
        sql = ("Update student set stipend=5000 where name='Ritu'")
        cursor.execute(sql)
        print("Record Updated")
        db.commit()
    except Exception as e:
        print (e)
```

```
def deldata():
    import mysql.connector
    db = mysql.connector.connect(host="localhost",user="root",password='',database='s1')
    cursor = db.cursor()
    sql = "delete from student where name='Ritu'"
    cursor.execute(sql)
    print("Record Deleted")
    db.commit()

menu()
```

```
File Edit Shell Debug Options Window Help
Python 3.7.0 (v3.7.0:1bf9cc5093, Jun 27 2018, 04:06:47) [MSC v.19
14 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
 RESTART: C:/Users/preeti/AppData/Local/Programs/Python/Python37-
32/prog_sqlpyth_complete.py
Do you want to continue or not: y
1. add record
2. update record
3. delete record
4. display records
5. Exiting
Enter your choice: 1
Records added
1. add record
2. update record
3. delete record
4. display records
5. Exiting
Enter your choice: 4
('Ritu', 4000, 'Science', 345, 'B', '11')
('Ankush', 6000, 'Commce', 445, 'A', '12')
('Pihu', 3566, 'Humanis', 446, 'A', '11')
('Tinku', 8900, 'Science', 545, 'A+', '12')
1. add record
2. update record
3. delete record
4. display records
5. Exiting
Enter your choice: 2
Record Updated
1. add record
2. update record
3. delete record
4. display records
5. Exiting
Enter your choice: 4
('Ritu', 5000, 'Science', 345, 'B', '11')
('Ankush', 6000, 'Commce', 445, 'A', '12')
('Pihu', 3566, 'Humanis', 446, 'A', '11')
('Tinku', 8900, 'Science', 545, 'A+', '12')
1. add record
2. update record
3. delete record
4. display records
5. Exiting
Enter your choice: 3
Record Deleted
1. add record
2. update record
3. delete record
4. display records
5. Exiting
Enter your choice: 4
('Ankush', 6000, 'Commce', 445, 'A', '12')
('Pihu', 3566, 'Humanis', 446, 'A', '11')
('Tinku', 8900, 'Science', 545, 'A+', '12')
```

```
1. add record
2. update record
3. delete record
4. display records
5. Exiting
Enter your choice: 5
Exiting
>>>
```

MEMORY BYTES

- We use Python-MySQL Connector to connect MySQL.
- **mysql.connector.connect()** method of Python-MySQL Connector is used with the required parameters to connect MySQL.
- Python-MySQL Connector requires Python to be in the system's PATH. Installation fails if it doesn't find Python.
- Install Python-MySQL Connector using pip command.
- MySQLdb is an interface for connecting to a MySQL database server from Python. It implements the Python Database API and is built on top of the MySQL C API.
- 'module_name.connect' is used to make a connection to the database that you wish to use.
- A database connection object controls the connection to the database. It represents a unique session with a database connected from within a script/program.
- In order to put our new connection to good use, we need to create a cursor object. It gives us the ability to have multiple separate working environment through the same connection to the database.
- A database cursor is a special control structure that facilitates row-by-row processing and accessing of records in the result set, *i.e.*, the set of records retrieved according to the query.
- The result set refers to a logical set of records that are fetched from the database by executing an SQL query and made available to the application program.
- Once a Database Connection is established, we are ready to create tables using execute() method of the created cursor.
- Read operation on any table means to fetch some useful information from the table.
- connect() method is used for establishing database connection, cursor() to create a cursor and execute() to execute an SQL query.
- In order to fetch records from a result set, you can use fetchone(), fetchmany() and fetchall() methods.
- Use fetchall() method to fetch multiple values from a database table.
- fetchone() fetches the next row of a query result set. A result set is an object that is returned when a cursor object is used to query a table.
- rowcount is a read-only attribute and returns the number of rows that were affected by an execute() method.
- UPDATE Operation on any database means to update one or more records, which are already available in the table.
- For INSERT, DELETE, and UPDATE queries, COMMIT() method should be executed with the connection object.
- To disconnect Database Connection, use close() method.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.
   (a) ..... methods will return only one row from the result set in the form of a tuple containing a record.
   (b) ..... method returns the result set in the form of tuples containing the records or rows returned by the sql table.
   (c) A ..... refers to a logical set of records that is fetched from the database.
   (d) A session between the application program and the database is called ..... .
   (e) A ..... is a special control structure that facilitates the row-by-row processing of records in the result set.

(f) A ..... query is used to check if data has been added to the table or not.

(g) To connect MySQL database ..... script is used.

(h) The ..... is a property of cursor object that returns the number of rows allowed from the cursor so far.

(i) ..... package can be imported in place of mysql-connector.

(j) A ..... is a collection of organized information that can be easily used, managed, updated, and they are classified according to their organizational approach.

Answers: (a) fetchone() (b) fetchall() (c) ResultSet
(d) Connection (e) database cursor (f) SELECT
(g) Python (h) rowcount (i) MySQLdb
(j) Database

2. State whether the following statements are True or False.

(a) Read operation on any table means to fetch some useful information from the table.

(b) Use fetchall() method to retrieve only one value from a database table.

(c) Rowcount is a read-only attribute.

(d) To disconnect database connection, use connect() method.

(e) Update statement is used to insert data into the table.

(f) A resultset is an object that is returned when a cursor object is used to query a table.

(g) After establishing connection, execute() method is used.

(h) fetchone() fetches the next row of a query result set.

(i) DB-API is the standard for Python's database interface.

(j) connect() method does not require any parameters.

Answers: (a) True (b) False (c) True (d) False (e) False (f) True
(g) True (h) True (i) True (j) False

3. Multiple Choice Questions (MCQs)

(a) Which connector is used for linking the database with Python code?
(i) MySQL-connector (ii) YesSQL: connector
(iii) PostSQL: connector (iv) None of the above

(b) To open a connector to MySQL database, which statement is used to connect with MySQL?
(i) connector (ii) connect (iii) password (iv) username

(c) Which software is used for coding the Python language?
(i) Idea (ii) IDLE (iii) Python Script (iv) Connector

(d) Which command is used for counting the number of rows in a database?
(i) row (ii) rowcount (iii) count() (iv) row_count

(e) Name the method which is used for displaying only one resultset.
(i) fetchmany (ii) fetchno (iii) fetchall (iv) fetchone

(f) To execute all the rows from the result set, which method is used?
(i) fetchall (ii) fetchone
(iii) fetchmany (iv) None of the above

(g) Pick the correct username used for logging in database.
(i) root (ii) local (iii) directory (iv) host

(h) Name the host name used for signing in the database.
(i) localhost (ii) localpost
(iii) localcost (iv) None of the above

(i) The command used for modifying the records is:
(i) update (ii) add
(iii) updateall (iv) None of the above

(j) Which command is used for cleaning up the environment?
(i) my.close (ii) is.close (iii) con.close (iv) mycon.close

Answers: (a) (i) (b) (ii) (c) (ii) (d) (ii) (e) (iv) (f) (i)
(g) (i) (h) (i) (i) (i) (j) (iv)

SOLVED QUESTIONS

1. Which file do we import in Python script to establish Python-MySQL connectivity?

Ans. mysql connector or mysqldb.

2. What is the significance of using connect()?

Ans. connect() function is used to connect or establish a connection with MySQL database.

3. What is the role of execute()?

Ans. The role of execute() is execution of queries which are MySQL queries in this chapter along with Python interface.

4. What is database connectivity?

Ans. Database connectivity refers to connection and communication between an application and a database system.

5. What are the features of MySQL?

Ans. MySQL provides cross-platform support, a wide range of interfaces for application programming, and has many stored procedures like triggers and cursors that help in managing the database.

6. What is a connection? What is its role?

Ans. A connection (represented by a connection object) is the session between an application program and the database. To establish connectivity with the database and to access it, a connection object is a must.

7. Which command is used to verify whether the python application is connected to MySQL database or not? [CBSE Sample Paper 2020]

Ans. is_connected()

8. What are the basic steps to connect with MySQL using table 'Members'?

Ans. import MySQLdb

```
conn = MySQLdb.connect(host=host, user=user, passwd=passwd, db=db)
cursor = conn.cursor()
cursor.execute('SELECT COUNT(MemberID) as count FROM Members WHERE ID = 1')
row = cursor.fetchone()
conn.close()
print(row)
```

9. What is meant by result set?

Ans. The result set refers to a logical set of records that are fetched from the database by executing an SQL query and made available to the application program.

10. What is Python's database interface known as?

Ans. DB-API is the standard for Python's database interface.

11. What does database DB-API include?

Ans. Using Python structure, DB-API provides standard and support for working with databases. The API working consists of the following steps:

(i) Bring in the API module
(ii) Obtain database connection
(iii) Issue SQL statements and then store procedures
(iv) Close the connection

12. Explain the benefits of Python Database Programming.

Ans. (i) Programming in Python is considerably simple and efficient as compared to other languages. Likewise, Database Programming is much easier using Python-MySQL connectivity.
(ii) Python database is portable and the program is also portable, which is an advantage in terms of portability.
(iii) Python supports SQL cursors.
(iv) It also supports Relational Database Systems.
(v) The API of Python for the database is also compatible with other databases.
(vi) It is platform-independent.

13. How can we implement MySQL Database?

Ans. To use MySQL database using Python, we need to first install it on our machine and then type the script given below to implement MySQL in the program:

import mysql-connector

OR import MySQLdb

14. What will be the generated query string?

```
query = "INSERT INTO books(title, author) VALUES ("C++", "E.Balagurusamy")"
```

Ans. "INSERT INTO books(title, author) VALUES ("C++", "E.Balagurusamy")"

15. Which record shall be inserted in the table "student" after executing the following code:

```
import mysql.connector
db1 = mysql.connector.connect(host ="localhost", user ="root",
                              passwd ="", database ="school")
cursor =db1.cursor()
query = "INSERT INTO books(title, author) VALUES ("C++", "E.Balagurusamy")"
cursor.execute(query)
db1.commit()
```

Ans. 'C++', 'E.Balagurusamy'

16. Which record will get inserted in the table by the following code:

```
import mysql.connector
db1 = mysql.connector.connect(host ="localhost", user ="root",
                              passwd ="", database ="school")
cursor =db1.cursor()
query = "INSERT INTO books(title, author) VALUES ("IP","Dr. Archana Aggarwal")"
cursor.execute(query)
db1.commit()
```

Ans. "IP","Dr. Archana Aggarwal"

17. Consider the table "books" present inside the database "school".

Title	Author
C++	E.Balagurusamy
IP	Dr. Archana Aggarwal
Accountancy	Dr. G.S.Grewal
Mathematics	Gupta-Bansal

On the basis of the above table, what will be the output of the following code:

```
import mysql.connector
db1 = mysql.connector.connect(host ="localhost", user ="root",
                              passwd ="", database ="school")
cursor = db1.cursor()
cursor.execute("SELECT * FROM books")
row = cursor.fetchone()
while row is not NONE:
    print(row)
    row = cursor.fetchone()
```

Ans.

(C++	E.Balagurusamy)
(IP	Dr. Archana Aggarwal)
(Accountancy	Dr. G.S.Grewal)
(Mathematics	Gupta-Bansal)

18. Write a Python database connectivity code that deletes records from table "books" of database school that have title as "Mathematics".

Ans.
```
import mysql.connector
db1 = mysql.connector.connect(host ="localhost", user ="root",
                              passwd ="", database ="school")
cursor =db1.cursor()
query = "DELETE FROM books WHERE name ='Mathematics' "
data =('Mathematics',)
cursor.execute(query.data)
db1.commit() #permanently commit the changes
print("Rows affected:",cursor.rowcount())
db1.close()
```

19. Explain the various database operations one can perform using Python-MySQL connectivity.

Ans. There are various operations that can be performed within a Python program. To deal with these statements, one must have a good knowledge of Database programming and SQL.

Database Operations/ Environment Variables	Description
INSERT	It is an SQL statement used to create a record into a table.
READ	Fetches useful information from the database.
UPDATE	It is used to update already existing or available record(s).
DELETE	It is used to delete records from the database.
ROLLBACK	It works like "undo" which reverts all the changes that you have made.

20. In a school, a database named "school" is created in MySQL whose password is "cbse". Smith is trying to add a new record of a student having details(3, 'Michelle', 'Agartala') in a "student" table.

Ans.
```
import mysql.connector
mydb = mysql.connector.connect(host="localhost", user="root",
                               passwd="cbse", database="school")
mycursor = mydb.cursor()
mycursor.execute("INSERT INTO student values(3,'Michelle', 'Agartala');")
mydb.commit()
```

21. Write a code in Python to read the contents of "number.csv" file consisting of data from a MySQL table and print the data of the table on the screen in tabular form of the table.

Ans.
```
f = open('numbers.csv', 'r')
with f:
    reader = csv.reader(f)
    for row in reader:
        for e in row:
            print(e)
```

UNS...

8.6 AGGREGATE FUNCTIONS & CONDITIONS ON GROUPS (HAVING CLAUSE)

You may use any condition on group, if required. HAVING <condition> clause is used to apply a condition on a group.

mysql> Select Job, Sum(Pay) from EMP Group By Job HAVING Sum(Pay)>=8000;

mysql> Select Job, Sum(Pay) from EMP Group By Job HAVING Avg(Pay)>=7000;

mysql> Select Job, Sum(Pay) from EMP Group By Job HAVING Count(*)>=5;

mysql> Select Job, Min(Pay),Max(Pay), Avg(Pay) from EMP
Group By Job HAVING Sum(Pay)>=8000;

mysql> Select Job, Sum(Pay) from EMP where city='Dehradun'
Group By Job HAVING count(*)>=5;

WHERE Vs HAVING:

Where clause works in respect of the whole table but **Having** clause works on Group only. If Where and Having both are used, then Where will be executed first. WHERE is used to put a condition on individual row of a table, whereas HAVING is used to put a condition on individual group formed by GROUP BY clause in a SELECT statement.

Aggregate Functions & Group (Group by Clause): Other Combinations

Consider the following table Employee with NULL values against the Salary field for some employees.

Employee

Ecode	Ename	Salary	Job	City
E1	Ritu Jain	NULL	Manager	Delhi
E2	Vikas Verma	4500	Executive	Jaipur
E3	Rajat Chaudhary	6000	Clerk	Kanpur
E4	Leena Arora	NULL	Manager	Bengaluru
E5	Shikha Sharma	8000	Accountant	Kanpur

None of the aggregate functions takes NULL into consideration. NULL values are simply ignored by all the aggregate functions as clearly shown in the examples given below:

An Aggregate function may be applied on a column with DISTINCT or * (ALL) symbol. If nothing is given, ALL scope is assumed.

- **Using SUM (<Column>)**

  This function returns the sum of values in the given column or expression.

  mysql> Select Sum(Salary) from Employee;

  mysql> Select Sum(DISTINCT Salary) from Employee;

  mysql> Select Sum (Salary) from Employee where City='Kanpur';

  mysql> Select Sum (Salary) from Employee Group By City;

  mysql> Select Job, Sum(Salary) from Employee Group By Job;

- **Using MIN (<Column>)**

  This function returns the Minimum value in the given column.

  mysql> Select Min(Salary) from Employee;

  mysql> Select Min(Salary) from Employee Group By City;

  mysql> Select Job, Min(Salary) from Employee Group By Job;

- **Using MAX (<Column>)**

  This function returns the Maximum value in the given column.

  mysql> Select Max(Salary) from Employee;

  mysql> Select Max(Salary) from Employee where City='Kanpur';

  mysql> Select Max(Salary) from Employee Group By City;

- **Using AVG (<Column>)**

  This function returns the Average value in the given column.

  mysql> Select AVG(Salary) from Employee;

  mysql> Select AVG(Salary) from Employee Group By City;

- **Using COUNT (<*|Column>)**

  This function returns the number of rows in the given column.

  mysql> Select Count (*) from Employee;

  mysql> Select Count(Salary) from Employee Group By City;

  mysql> Select Count(*), Sum(Salary) from Employee Group By Job;

MEMORY BYTES

- A function is a predefined command that performs some specific operations.
- MySQL provides Aggregate or Group functions which work on a number of values of a column/expression and return a single value as the result.
- **SELECT:** To display row or column information.
- **DISTINCT:** To select different or non-duplicate information.
- **MAX():** To select maximum value of a particular column.
- **MIN():** To select minimum value of a particular column.
- **SUM():** To find total value of a particular column.
- **AVG():** To find average value of a particular column.
- **COUNT():** Number of records in the table.
- Group by clause is used to divide the result in groups on the basis of identical values in a particular field or a group of fields.
- HAVING clause is used in combination with GROUP BY clause.
- Sum() will give the same answer irrespective of the NULL values in the specified column.
- The Having clause is used to specify filtering conditions for groups.
- Group By clause is used whenever aggregate functions by group are required.
- The 'HAVING' clause is used to place conditions on groups created by 'GROUP BY' clause because here the 'WHERE' clause is not useable.
- The DISTINCT keyword is used to remove duplicate values in a particular column.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.
   (a) The ..... function works with data of multiple rows at a time and returns aggregated value
   (b) The ..... clause lets you arrange the result set in the order of single column, multiple columns and custom sort order too.
   (c) To specify filtering condition for groups, the ..... clause is used in MySQL.
   (d) By default, the ORDER BY clause sorts the result set in the ..... order.
   (e) To sort the result set in descending order, ... keyword is used with ORDER BY.
   (f) Two types of MySQL functions are single row function and ..... function.
   (g) The ..... function returns the total number of rows, including duplicates and NULL in a table.
   (h) Group functions are also known as ..... functions.
   (i) In MySQL, ... option causes a group function to consider only unique values of the argument expression.
   (j) The ..... function returns the lowest value from the given column or expression.

   Answers: (a) group/row (b) Order by (c) Having
   (d) ascending (e) DESC (f) multiple row
   (g) count(*) (h) aggregate (i) distinct
   (j) MIN()

2. State whether the following statements are True or False.
   (a) The ORDER BY clause combines all those records that have identical values in a particular field or a group of fields.
   (b) The WHERE clause is used to specify filtering conditions for groups.
   (c) DISTINCT option causes a group function to consider only the unique values of the argument expression.
   (d) By default, ORDER BY clause sorts the result set in descending order.
   (e) COUNT() function ignores duplicate and null values while counting the records.
   (f) The return value of MAX() function is a numeric value.
   (g) Multiple row function is also known as scalar function.
   (h) SUM() function is used to count the total number of records in a table.
   (i) Argument type of AVG() function can be numeric or string data type.
   (j) Group functions can be applied to any numeric values, some text types and DATE values.

   Answers: (a) False (b) False (c) True (d) False (e) True (f) True
   (g) False (h) False (i) False (j) True

3. Multiple Choice Questions (MCQs)
   (a) What will be the order of the data being sorted after the execution of given SQL query?
   SELECT *FROM STUDENT ORDER BY ROLL_NO;
   (i) Custom Sort (ii) Descending
   (iii) Ascending (iv) None of the above
   (b) Which values will not be considered by SQL while executing the following statement?
   SELECT COUNT( ) FROM DEPARTMENT;
   (i) Numeric value (ii) Text value (iii) Null value (iv) Date value
   (c) Which of the following clauses is used to sort the result set?
   (i) SORT BY (ii) GROUP BY (iii) ARRANGE BY (iv) ORDER BY
   (d) Which clause is used in query to place the condition on groups in MySQL?
   (i) Where (ii) Having (iii) Group By (iv) Both (i) & (ii)

(e) If column "Marks" contains the data set{25, 35, 25, 35, 38}, what will be the output after the execution of the given query?

SELECT MARKS(DISTINCT) FROM STUDENTS;

(i) 25,35,25,35,38 (ii) 25,25,35,35,38 (iii) 25,35,38 (iv) 25,25,35,35

(f) If column "Salary" contains the data set{10000,15000,25000,10000,15000}, what will be the output after the execution of the given query?

SELECT SUM(DISTINCT SALARY) FROM EMPLOYEE;

(i) 75000 (ii) 25000 (iii) 10000 (iv) 50000

(g) What SQL statement do we use to display the record of all students whose last name contains 5 letters ending with "A"?

(i) SELECT * FROM STUDENTS WHERE LNAME LIKE '_ _ _ _A';

(ii) SELECT * FROM STUDENTS WHERE LNAME LIKE '_ _ _ _ _';

(iii) SELECT * FROM STUDENTS WHERE LNAME LIKE '????A';

(iv) SELECT * FROM STUDENTS WHERE LNAME LIKE '*A';

(h) What SQL statement do we use to find the total number of records present in the table Product?

(i) SELECT * FROM PRODUCT; (ii) SELECT COUNT(*) FROM PRODUCT;

(iii) SELECT FIND(*) FROM PRODUCT; (iv) SELECT SUM() FROM PRODUCT;

(i) Which of the following functions is not an aggregate function?

(i) Round() (ii) Sum() (iii) Count() (iv) Avg()

(j) Which of the following functions is used to FIND the largest value from the given data in MySQL?

(i) MAX() (ii) MAXIMUM() (iii) LARGEST() (iv) BIG()

Answers: (a) (iii) (b) (iii) (c) (iv) (d) (ii) (e) (iii) (f) (iv)
(g) (i) (h) (ii) (i) (i) (j) (i)

SOLVED QUESTIONS

1. What is SQL?

Ans. SQL stands for Structured Query Language. This language is used to create, manage table and manipulate stored records in a table.

2. What is null value in MySQL?

Ans. If a column in a row has no value, then the column is said to be null.

3. Which keyword eliminates redundant data from a query result?

Ans. DISTINCT

4. Differentiate between WHERE and HAVING clause.

Ans. Where clause is used to select particular rows that satisfy condition whereas having clause is used in connection with the aggregate function, GROUP BY clause. *For example*, select * from Student where marks>75; this statement shall display the records for all the students who have scored more than 75 marks.

On the contrary, the statement – select * from Student group by stream having marks>75; shall display the records of all the students grouped together on the basis of stream but only for those students who have scored marks more than 75.

5. Why is it not allowed to give String and Date type arguments for SUM() and AVG() functions? [HOTS]

Ans. String and dates are not real numbers that we calculate, so sum() or avg() functions are not valid for them.

6. How are NULL values treated by aggregate functions?

Ans. None of the aggregate functions takes NULL into consideration. NULL is simply ignored by all the aggregate functions.

7. There is a column C1 in a table T1. The following two statements: SELECT COUNT(*) FROM T1; and SELECT COUNT(C1) from T1; are giving different outputs. What may be the possible reason? [HOTS]

Ans. There may be a null value.

8. What is the purpose of GROUP BY clause?

Ans. GROUP BY clause is used in a SELECT statement in combination with aggregate functions to group the result based on distinct values in a column.

9. What is HAVING clause?

Ans. HAVING clause is used in combination with GROUP BY clause in a SELECT statement to put condition on groups.

10. Consider a database LOANS with the following table:

Table: LOANS

AccNo	Cust_Name	Loan_Amount	Instalments	Int_Rate	Start_Date	Interest
1	R.K. Gupta	300000	36	12.00	19-07-2009	1200
2	S.P. Sharma	500000	48	10.00	22-03-2008	1800
3	K.P. Jain	300000	36	NULL	08-03-2007	1600
4	M.P. Yadav	800000	60	10.00	06-12-2008	2250
5	S.P. Sinha	200000	36	12.50	03-01-2010	4500
6	P. Sharma	700000	60	12.50	05-06-2008	3500
7	K.S. Dhall	500000	48	NULL	05-03-2008	3800

Answer the following questions.

(a) Display the sum of all Loan Amount whose Interest rate is greater than 10.

Ans. Mysql> Select sum(Loan_Amount) from LOANS Where Interest >10;

(b) Display the Maximum Interest from Loans table.

Ans. Mysql> Select Max(Interest) from LOANS;

(c) Display the count of all loan holders whose name ends with 'Sharma'.

Ans. Mysql> Select Count(*) from LOANS Where Cust_Name Like '%Sharma';

(d) Display the count of all loan holders whose Interest is Null.

Ans. Mysql> Select Count(*) from LOANS Where Interest Is NULL;

(e) Display the Interest-wise details of Loan Account Holders.

Ans. Mysql> Select * from LOANS Group By Interest;

(f) Display the Interest-wise details of Loan Account Holders with at least 10 instalments remaining.

Ans. Mysql> Select * from LOANS Group By Interest Having Instalments>=10;

(g) Display the Interest-wise count of all loan holders whose Instalments due is more than 5 in each group.

Ans. Mysql> Select Count (*) from LOANS Group By Interest Having Instalments>5;

11. Name two Aggregate (Group) functions of SQL. [CBSE D 2016]

Ans. Count()
Max()

12. Consider the table: [CBSE D 2016]

Table: COMPANY

SID	SALES
S101	20000
S103	NULL
S104	10000
2015	15000

What output will be displayed by the following SQL statement?

SELECT AVG(SALES) from company;

Ans. 15000

13. Consider the table 'Hotel' given below: [CBSE D 2016]

Table: HOTEL

EMPID	CATEGORY	SALARY
E101	MANAGER	60000
E102	EXECUTIVE	65000
E103	CLERK	40000
E104	MANAGER	62000
E105	EXECUTIVE	50000
E106	CLERK	35000

Mr. Vinay wanted to display average salary of each category. He entered the following SQL statement. Identify error(s) and rewrite the correct SQL statement.

```
SELECT CATEGORY,SALARY
FROM HOTEL
GROUP BY CATEGORY;
```

Ans. SELECT category, avg(salary) from hotel group by category;

14. (a) Mr. Manav, a database administrator in "Global Educational and Training Institute", has created the following table named "Training" for the upcoming training schedule: [CBSE Sample Paper 2019]

Table: TRAINING

Training_Id	Name	Email_Id	Topic	City	Fee
ND01	Mr. Rajan	raj@gmail.com	Cyber Security	New Delhi	10000
GU01	Ms. Urvashi	urv@yahoo.com	ICT in Education	Gurugram	15000
FD01	Ms. Neena	neena@rediff.com	Cyber Security	Faridabad	12000
ND02	Mr. Vinay	NULL	ICT in Education	New Delhi	13000
GU02	Mr. Naveen	nav@gmail.com	Cyber Security	Gurugram	NULL

Help him in writing SQL query for the following purpose:

i. To count how many female candidates will be attending the training.

Ans. Select count(name) from training where name like 'Ms.%';

ii. To display list of free trainings.

Ans. Select * from training where fee is NULL;

iii. To display all the cities where Cyber Security training is scheduled along with its fee.

Ans. Select city, fee from training where topic = 'Cyber Security';

iv. To add a column feedback with suitable data type.

Ans. Alter table training add feedback varchar(20);

(b) Observe the table named "Training" given above carefully and predict the output of the following queries:

i. Select city from training where topic = "Cyber Security";

Ans.

City
New Delhi
Faridabad
Gurugram

ii. Select count(Training_Id) from training where email_Id like '%gmail%';

Ans. 2

iii. Select AVG (Fee) from training where Topic = "Cyber Security";

Ans. 11000

iv. Select name from training where INSTR ((Email_Id, '@')=0);

Ans. Mr. Vinay

13. Consider the table 'Hotel' given below: [CBSE D 2016]

Table: HOTEL

EMPID	CATEGORY	SALARY
E101	MANAGER	60000
E102	EXECUTIVE	65000
E103	CLERK	40000
E104	MANAGER	62000
E105	EXECUTIVE	50000
E106	CLERK	35000

Mr. Vinay wanted to display average salary of each category. He entered the following SQL statement. Identify error(s) and rewrite the correct SQL statement.

```
SELECT CATEGORY,SALARY
FROM HOTEL
GROUP BY CATEGORY;
```

Ans. SELECT category, avg(salary) from hotel group by category;

14. (a) Mr. Manav, a database administrator in "Global Educational and Training Institute", has created the following table named "Training" for the upcoming training schedule: [CBSE Sample Paper 2019]

Table: TRAINING

Training_Id	Name	Email_Id	Topic	City	Fee
ND01	Mr. Rajan	raj@gmail.com	Cyber Security	New Delhi	10000
GU01	Ms. Urvashi	urv@yahoo.com	ICT in Education	Gurugram	15000
FD01	Ms. Neena	neena@rediff.com	Cyber Security	Faridabad	12000
ND02	Mr. Vinay	NULL	ICT in Education	New Delhi	13000
GU02	Mr. Naveen	nav@gmail.com	Cyber Security	Gurugram	NULL

Help him in writing SQL query for the following purpose:

i. To count how many female candidates will be attending the training.

Ans. Select count(name) from training where name like 'Ms.%';

ii. To display list of free trainings.

Ans. Select * from training where fee is NULL;

iii. To display all the cities where Cyber Security training is scheduled along with its fee.

Ans. Select city, fee from training where topic = 'Cyber Security';

iv. To add a column feedback with suitable data type.

Ans. Alter table training add feedback varchar(20);

(b) Observe the table named "Training" given above carefully and predict the output of the following queries:

i. Select city from training where topic = "Cyber Security";

Ans.

City
New Delhi
Faridabad
Gurugram

ii. Select count(Training_Id) from training where email_id like '%gmail%';

Ans. 2

iii. Select AVG (Fee) from training where Topic = "Cyber Security";

Ans. 11000

iv. Select name from training where INSTR ((Email_Id, '@')=0);

Ans. Mr. Vinay

15. Consider the following tables GAMES and PLAYER. Write SQL commands and output for the following statements.

Table: GAMES

GCode	GameName	Number	PrizeMoney	ScheduledDate
101	Kabaddi	2	5000	23-Jan-2007
102	Badminton	2	12000	12-Dec-2013
103	Table Tennis	4	8000	14-Feb-2014
105	Chess	2	9000	1-Jan-2015
108	Table Tennis	4	25000	19-Mar-2014

Table: PLAYER

PCode	Name	Gcode
1	Ravi Shankar	101
2	Amir	108
3	Jatin	101
4	Shahrukh	103

(a) To display details of those games which have PrizeMoney more than 7000.

Ans. SELECT * from GAMES WHERE prizemoney>7000;

(b) To display the content of the GAMES table in ascending order of ScheduledDate.

Ans. SELECT * from GAMES ORDER BY ScheduledDate;

(c) SELECT COUNT(DISTINCT GameName) FROM GAMES;

Ans. 4

(d) SELECT MAX(ScheduledDate), MIN(ScheduledDate) FROM GAMES;

Ans. 01-Jan-2015 23-Jan-2007

16. Write SQL commands for statements (a) to (c).

Table: FACULTY

F_ID	FNAME	LNAME	HIREDATE	SALARY
102	MANOJ	SABHARWAL	12-10-2012	12000
103	PARVEEN	ARORA	15-12-2014	9000
104	SANJEEV	SHARMA	01-01-2016	14000
105	RASHMI	MALHOTRA	02-04-2000	20000
106	NITIN	SRIVASTAVA	12-12-2013	10000

Table: COURSES

C_ID	F_ID	CNAME	FEES
C21	102	BOOLEAN ALGEBRA	14000
C22	106	COMPUTER NETWORK	20000
C21	104	C++	18000
C24	106	HUMAN BIOLOGY	25000
C25	102	BIOTECH	30000
C26	103	COMPUTER TECH	40000

(a) To display details of those Faculty Members whose salary is greater than 12000.

Ans. SELECT * from faculty where salary> 12000;

(b) To display the details of courses whose fees is in the range of 15000 to 50000 (both values included).

Ans. SELECT * from Courses where fees between 15000 and 50000;

(c) SELECT COUNT (DISTINCT F_ID) from COURSES;

Ans.
COUNT (DISTINCT F_ID)
102
103
104
106

17. Given the table 'Player' with the following columns: [CBSE 2018]

Table: PLAYER

PCODE	POINTS
1	50
2	NULL
3	40

Write the output of the following statements:

(a) SELECT AVG(POINTS) FROM Player;

Ans. AVG(POINTS)

45

(b) Select COUNT(POINTS) FROM Player;

Ans. COUNT(POINTS)

2

18. Kunal has entered the following SQL command on Table 'STUDENT' that has TotalMarks as one of the columns: [CBSE 2017]

SELECT COUNT(*) FROM STUDENT;

The output displayed is 20.

Then, Kunal enters the following command:

SELECT COUNT(*) FROM STUDENT WHERE TotalMarks < 100;

The output displayed is 15.

Then, Kunal enters the following command:

SELECT COUNT(*) FROM STUDENT WHERE TotalMarks >= 100;

He predicts the output of the above query as 5. Do you agree with Kunal?

Give reasons for your answer.

Ans. Yes, Total rows=20, rows with TotalMarks <100 is 15, so remaining rows left are 20-15=5

OR

No, the output of the query may not always be 5 as there may be rows with Total Marks as NULL which would have not been included in either of the two SELECT statements mentioned.

19. What is an ORDER BY clause and GROUP BY clause?

Ans. ORDER BY clause is used to display the result of a query in a specific order (sorted order).

The sorting can be done in ascending or in descending order. However, the actual data in the database is not sorted but only the results of the query are displayed in sorted order. If order is not specified then, by default, the sorting will be performed in ascending order.

For example,

```
SELECT name, city FROM student
ORDER BY name;
```

The above query returns name and city columns of table student sorted by name in ascending or descending order.

For example,

SELECT * FROM student ORDER BY city DESC;

It displays all the records of table student ordered by city in descending order.

GROUP BY clause

The GROUP BY clause can be used in a SELECT statement to collect data across multiple records and group the results by one or more columns.

For example,

SELECT name, COUNT(*) as "Number of employees" FROM student
WHERE marks>350 GROUP BY city;

20. Consider the following tables Product and Client. Write SQL commands for the statements (a) to (c) and give outputs for SQL queries (d) to (f).

Table: PRODUCT

P_ID	Product Name	Manufacturer	Price
TP01	Talcum Powder	LAK	40
FW05	Face Wash	ABC	45
BS01	Bath Soap	ABC	55
SH06	Shampoo	XYZ	120
FW12	Face Wash	XYZ	95

Table: CLIENT

C_ID	Client Name	City	P_ID
01	Cosmetic Shop	Delhi	FW05
06	Total Health	Mumbai	BS01
12	Live Life	Delhi	SH06
15	Pretty Woman	Delhi	FW12
16	Dreams	Bengaluru	TP01

(a) To display the details of those Clients whose city is Delhi.

Ans. Select * from Client where City="Delhi";

(b) To display the details of Products whose Price is in the range of 50 to 100 (both values included).

Ans. Select * from product where Price between 50 and 100;

(c) To display the details of those products whose name ends with 'wash'.

Ans. Select * from product where name like '%wash';

(d) SELECT DISTINCT Address FROM Client;

Ans. (The above question may consist of DISTINCT City. If it is DISTINCT City, the following is the answer.)

City

Delhi
Mumbai
Bengaluru

(e) SELECT Manufacturer, Max(Price), Min(Price), Count(*) FROM Product GROUP BY Manufacturer;

Ans.

Manufacturer	Max(Price)	Min(Price)	Count(*)
LAK	40	40	1
ABC	55	45	2
XYZ	120	95	2

(f) SELECT ProductName, Price * 4 FROM Product;

Ans.

ProductName	Price*4
TalcumPowder	160
FaceWash	180
BathSoap	220
Shampoo	480
Face Wash	380

21. What are aggregate functions?

Ans. Aggregate functions are those functions that return single values from groups of values. There are several group functions such as: SUM(), AVG(), MAX(), MIN(), COUNT(), etc.

22. What is a base table?

Ans. A table from which the values can be derived for other table. *For example*, in case of views, the values are extracted from the base table on which the view depends.

23. What is NULL value?

Ans. A NULL value in a table is a value in a field which is blank. In other words, a NULL value is an unidentified or unavailable value and not equal to zero.

24. What is a NOT NULL Constraint?

Ans. If you do not want a column to have a NULL value, then you need to define such constraint on this column specifying that NULL is now not allowed for that column.

25. (a) Explain the concept of Cartesian product between two tables with the help of example.

[CBSE D 2014]

Note: Answer the questions (b) and (c) on the basis of the following tables SHOP and ACCESSORIES.

Table: SHOP

Id	SName	Area
S001	ABC Computronics	CP
S002	All Infotech Media	GK II
S003	Tech Shoppe	CP
S004	Geek Tenco Soft	Nehru Place
S005	Hitech Tech Store	Nehru Place

Table: ACCESSORIES

No	Name	Price	Id
A01	Mother Board	12000	S01
A02	Hard Disk	5000	S01
A03	Keyboard	500	S02
A04	Mouse	300	S01
A05	Mother Board	13000	S02
A06	Keyboard	400	S03
A07	LCD	6000	S04
T08	LCD	5500	S05
T09	Mouse	350	S05
T10	Hard Disk	4500	S03

Ans. (a) When you join two or more tables without any condition, it is called Cartesian product or Cross Join.

Example – SELECT * FROM SHOP, ACCESSORIES;

(b) Write the SQL queries:
(i) To display Name and Price of all the Accessories in ascending order of their Price.
(ii) To display Id and SName of all Shops located in Nehru Place.
(iii) To display Minimum and Maximum Price of each Name of Accessories.

Ans. (i) SELECT Name, Price FROM ACCESSORIES ORDER BY Price;
(ii) SELECT Id, SName FROM SHOP WHERE Area='Nehru Place';
(iii) SELECT Name, MAX(Price), MIN(Price) FROM ACCESSORIES ;

(c) Write the output of the following SQL commands:
(i) SELECT DISTINCT NAME FROM ACCESSORIES WHERE PRICE>=5000;
(ii) SELECT AREA, COUNT(*) FROM SHOP GROUP BY AREA;
(iii) SELECT COUNT(DISTINCT AREA) FROM SHOP;
(iv) SELECT NAME, PRICE*0.05 AS "DISCOUNT" FROM ACCESSORIES WHERE ID IN ('S02', 'S03');

Ans. (i) Name
Mother Board
Hard Disk
LCD

(ii)

AREA	COUNT
CP	2
GK II	1
Nehru Place	2

(iii) COUNT 3

(iv)

NAME	DISCOUNT
Keyboard	25.00
Mother Board	650.00
Keyboard	20.00
Hard Disk	225.00

26. Consider the following tables CARDEN and CUSTOMER and answer (a) and (b) parts of the question:

Table: CARDEN

Ccode	CarName	Make	Color	Capacity	Charges
501	A-Star	Suzuki	RED	3	14
503	Indigo	Tata	SILVER	3	12
502	Innova	Toyota	WHITE	7	15
509	SX4	Suzuki	SILVER	4	14
510	C Class	Mercedes	RED	4	35

Table: CUSTOMER

CCode	CName	Ccode
1001	HemantSahu	501
1002	Raj Lal	509
1003	Feroza Shah	503
1004	Ketan Dhal	502

(a) Write SQL commands for the following statements:
(i) To display the names of all the silver colored Cars.
(ii) To display name of car, make and capacity of cars in descending order of their sitting capacity.
(iii) To display the highest charges at which a vehicle can be hired from CARDEN.

Ans. (i) SELECT CarName FROM carden WHERE Color LIKE 'Silver';
(ii) SELECT CarName, Make, Capacity FROM carden ORDER BY Capacity Desc;
(iii) SELECT MAX(Charges) FROM carden;

(b) Give the output of the following SQL queries:

(i) SELECT COUNT(DISTINCT Make) FROM CARDEN;

(ii) SELECT MAX(Charges),MIN(Charges) FROM CARDEN;

(iii) SELECT COUNT(*),Make FROM CARDEN;

(iv) SELECT CarName FROM CARDEN WHERE Capacity=4;

Ans. (i) COUNT(DISTINCT Make)
4

(ii)

MAX(Charges)	MIN(Charges)
35	12

(iii)

COUNT(*)	Make
5	Suzuki

(iv) CarName
SX4
C Class

27. Consider the following tables STOCK and DEALERS and answer (a) and (b) parts of this question: [CBSE D 2010]

Table: STOCK

ItemNo	Item	Dcode	Qty	UnitPrice	StockDate
5005	Ball Pen 0.5	102	100	16	31-Mar-10
5003	Ball Pen 0.25	102	150	20	01-Jan-10
5002	Gel Pen Premium	101	125	14	14-Feb-10
5006	Gel Pen Classic	101	200	22	01-Jan-09
5001	Eraser Small	102	210	5	19-Mar-09
5004	Eraser Big	102	60	10	12-Dec-09
5009	Sharpener Classic	103	160	8	23-Jan-09

Table: DEALERS

Dcode	DName
101	Reliable Stationers
103	Classic Plastics
102	Clear Deals

(a) Write SQL commands for the following statements:

(i) To display details of all Items in the Stock table in ascending order of StockDate.

(ii) To display ItemNo and Item name of those items from Stock table whose UnitPrice is more than Rupees 10.

(iii) To display the details of those items whose dealer code (Dcode) is 102 or Quantity in Stock (Qty) is more than 100 from the table Stock.

(iv) To display Maximum UnitPrice of items for each dealer individually as per Dcode from the table Stock.

Ans. (i) SELECT * FROM STOCK ORDER BY StockDate;

(ii) SELECT ItemNo, Item FROM STOCK WHERE UnitPrice>10;

(iii) SELECT * FROM STOCK WHERE Dcode=102 OR Qty>100;

(iv) SELECT Dcode, MAX (UnitPrice) FROM STOCK GROUP BY Dcode;

(b) Give the output of the following SQL queries:

(i) SELECT COUNT(DISTINCT Dcode) FROM Stock;

(ii) SELECT Qty*UnitPrice FROM Stock WHERE ItemNo=5006;

(iii) SELECT MIN(StockDate) FROM Stock;

Ans. (i) Count(DISTINCT Dcode)
3

9.2 INTELLECTUAL PROPERTY RIGHTS

You must have heard the word property; it is generally used to mean a possession or, more specifically, something to which the owner has legal rights.

You might have also encountered the phrase intellectual property. This term has become more commonplace during the past few years, especially in the context of computer ethics. But what exactly does it refer to?

Intellectual property refers to creations of the intellect (hence, the name); inventions, literary and artistic works, symbols, names, images, and designs used in commerce are part of it.

Intellectual property is usually divided into two branches, namely **industrial property** which, broadly speaking, protects inventions and **copyright**, which protects literary and artistic works.

Fig. 9.1(a): Intellectual Property Rights

Intellectual property is divided into two categories:

(a) **Industrial property**, which includes inventions (patents), trademarks, industrial designs, commercial names, designations and geographic indications (location-specific brands), etc.

(b) **Copyright**, which includes literary and artistic works such as novels, poems and plays, films, musical works, artistic works such as drawings, paintings, photographs, sculptures, and architectural designs. Copyright is a legal concept, enacted by most governments, giving the creator of original work exclusive rights to it, usually for a limited period.

Fig. 9.1(b): Copyright

In a nutshell, copyright laws protect intellectual property which includes literary and artistic works such as novels, poems, plays, films, musical works, artistic works such as drawings, paintings, photographs and sculptures, and architectural designs.

But unlike protection of inventions, copyright law protects only the form of expressions of ideas, not the ideas themselves.

9.3 PLAGIARISM

Plagiarism is copying someone else's work and then passing it off as one's own. It is morally bad because it is an act of stealing. In other words, it is copying information and not giving the author credit for it.

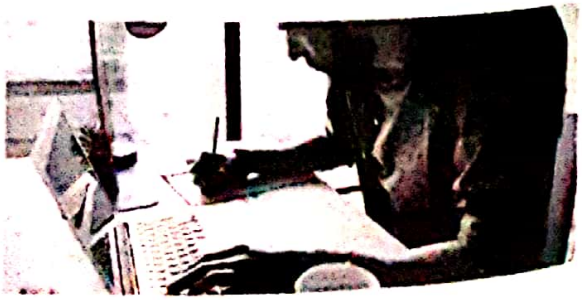
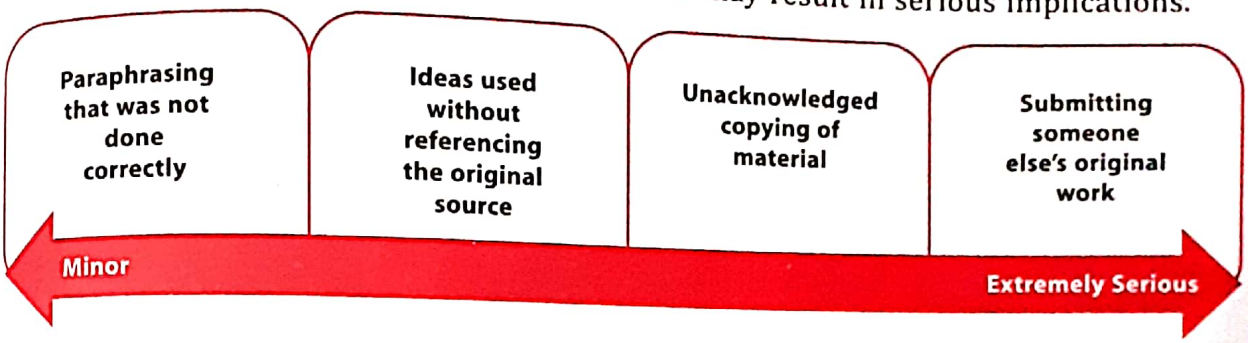


Fig. 9.2: Plagiarism

Copying programs written by other programmers and claiming them as your own could be an act of plagiarism. It involves lying, cheating, theft, and dishonesty.

Plagiarism can be classified as **accidental/unintentional** or **deliberate/intentional**. **Intentional plagiarism** includes copying someone else's work, cutting and pasting blocks of text or any kind of media (audio, video files or movie clips) from electronic sources without documenting and at the same time publishing it on the web without the permission of developers/creators. On the other hand, unintentional/accidental plagiarism involves careless paraphrasing (changing the words or sentence construction of a copied document), quoting text excessively along with poor documentation. Usually, accidental plagiarism cases are less serious in comparison to deliberate plagiarism that may result in serious implications.



How to avoid Plagiarism

Plagiarism is a bad practice and should be avoided by the following simple measures:

1. Use your own words and ideas.
2. Always provide reference or give credit to the source from where you have received your information.
3. If, by any chance, it becomes necessary to use someone's exact words, don't forget to put them in quotes and give credit using in-text citations.
4. As far as possible, try to include the source in your works citing page number.
5. Cite the name of the website, a URL, or the name of authors, and acknowledge them if you have used their work after rearranging the order of a sentence and changing some of the words.
6. Take information in the form of bulleted notes in your own words rather than copying the entire content or complete sentences.
7. Use online tools to check for plagiarism.

9.4 DIGITAL RIGHTS MANAGEMENT

Digital Rights Management (DRM) is a term for access control technologies that are used by hardware manufacturers, publishers, copyright holders and individuals to limit the use of digital content and devices. The term is used to describe any technology that inhibits the use of digital content that is not desired or intended by the content provider.

It attempts to control use of digital media by preventing access, copying or conversion to other formats by the end-users. There are several instances in our day-to-day life when we intend

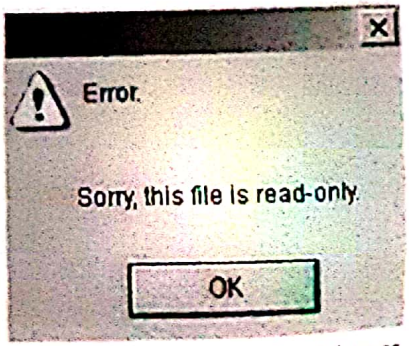


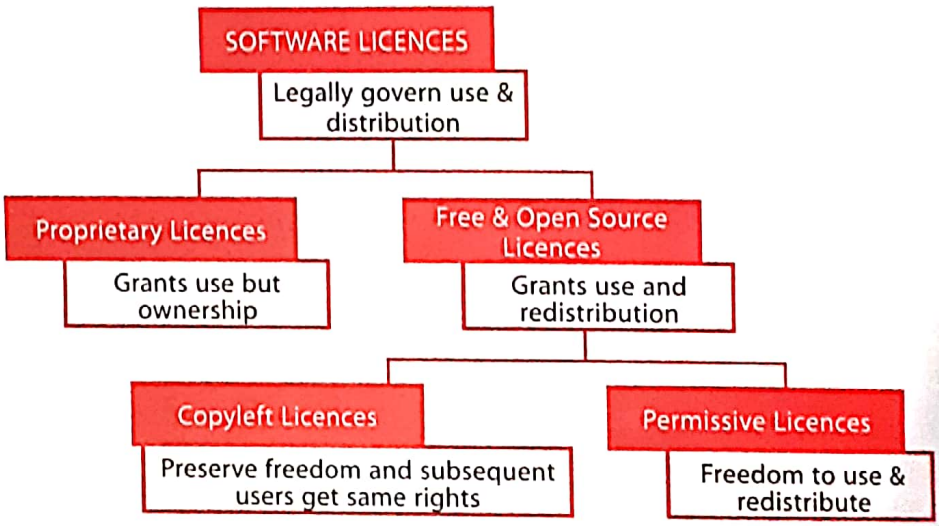
Fig. 9.3(a): Unauthorized Access to Digital Content

Advantages of using Licensed Software

It is always good to use a licensed software as licensing provides advantages while working with these software.

1. By using licensed software, you are able to contribute to the further development of the program you are using. It also helps the economy generate not just computer-related jobs, but opportunities for other people/businesses whose work is much dependent on "*written*" technology.
2. Apart from the legal obligation to use licensed software, it comes with outright support not found in "*pirated*" software.

Thus, licences provide rules and guidelines for others to use your work, *for example*, Open Source Licences help others to contribute to your work or project without seeking special individual permission to do so. Licences can be further classified into two types on the basis of criteria of their use—Proprietary Licences and Free and Open Source Licences.



In **Proprietary Licences**, exclusive rights in the software are retained with the owner/developer/publisher. They reserve all the freedoms and rights to use and distribute these proprietary software. Violation of the proprietary licence amounts to copyright infringement. They restrict inspection, modification of source code and further distribution.

However, freeware (*i.e.*, no-cost software) can be distributed using proprietary licences. Most commercially available software are covered under proprietary licences.

On the other hand, **Free and Open Source Licence** refers to a software that users can safely run, adapt and redistribute without legal restraint, and which emphasizes on freedom.

Open source software is a software with source code that is publicly available under a licence that gives users the right to study, change, and distribute that software and emphasizes on security, cost-saving, and transparency. Hence, free and open source software (FOSS) allows using, copying, studying and modifying the software, and the source code to be openly shared so that people are encouraged to voluntarily improve the design of the software.

CTM: An OSS (open source software) refers to freedom to use, share and/or modify the source code and allow copyrights to other users.

3. **Online social networks:** Risks range from identity theft to online and physical stalking, embarrassment, discrimination and blackmailing.

To provide protection from these online privacy threats, privacy laws were made which prohibit disclosure or misuse of information about private individuals.

Privacy laws refer to laws that deal with regulating, storing and using personally identifiable information of individuals, which can be collected by governments, public or private organizations or other individuals.

Privacy laws are considered within the context of an individual's privacy rights or within reasonable expectation of privacy.

The Information Technology Act, 2000 has two sections relating to privacy:

- Section 43A, which deals with implementation of reasonable security practices for sensitive personal data or information and provides for compensation to a person affected by wrongful loss or wrongful gain.
- Section 72A, which provides for imprisonment for a period of up to 3 years and/or a fine of up to ₹ 5,00,000 to a person who causes wrongful loss or wrongful gain by disclosing personal information about another person while providing services under the terms of lawful contract.

The IT Act, 2000 has provisions that permit the interception, monitoring and decryption of digital communications. It provides for collection and monitoring of traffic data. It allows the government to set a national encryption standard. Projects like the Central Monitoring System, NATGRID, Phone & Internet Interception used by police are some crucial measures taken up by the Government of India for safeguarding privacy laws. We will be discussing IT Act, 2000 in detail in successive sections.

9.8 CYBERCRIME

Cybercrime is defined as a crime in which a computer is the object of the crime (hacking, phishing, spamming) or is used as a tool to commit an offence (child pornography, hate crimes). Cybercriminals may use computer technology to access personal information, business trade secrets or use the internet for exploitative or malicious purposes. Criminals can also use computers for communication and document or data storage. Criminals who perform these illegal activities are often referred to as hackers.

CTM: Cybercrime may also be referred to as computer crime.

Common types of cybercrime include online bank information theft, identity theft, online predatory crimes and unauthorized computer access. More serious crimes like cyber terrorism are also of significant concern. Cybercrime encompasses a wide range of activities but these can generally be broken into two categories:

1. Crimes that target computer networks or devices: These types of crimes include viruses and denial-of-service (DoS) attacks.
2. Crimes that use computer networks to advance other criminal activities: These types of crimes include cyberstalking, phishing and fraud or identity theft.

9.8.1 Phishing

Phishing is an attempt to acquire sensitive information such as usernames, passwords and credit card details (and sometimes, indirectly, money) by masquerading as a trustworthy entity in an electronic communication.

Fig. 9.7(a): Phishing

Phishing is typically carried out by email spoofing or instant messaging and it often directs the users to enter personal information at a fake website, the look and feel of which is identical to the legitimate one, the only difference being the URL of the website in question.

Communications purporting to be from social websites, auction sites, banks, online payment processors or IT administrators are often used to lure victims. Phishing emails may contain links to websites that distribute malware.

The protective measures to be followed against Phishing include:

1. Never open or download a file from an unsolicited email, even from someone you know. (You can call or email the person to double check that it really came from them.)
2. Keep your operating system updated.
3. Use a reputable antivirus program.
4. Enable two-factor authentication whenever available.
5. Confirm the authenticity of a website prior to entering login credentials by looking for a reputable security trust mark.
6. Look for HTTPS in the address bar when you enter any sensitive personal information on a website to make sure your data will be encrypted.

9.8.2 Illegal Downloading

Illegal downloading is obtaining files that you don't have the right to use from the internet. Illegal downloading is the root cause of digital piracy.

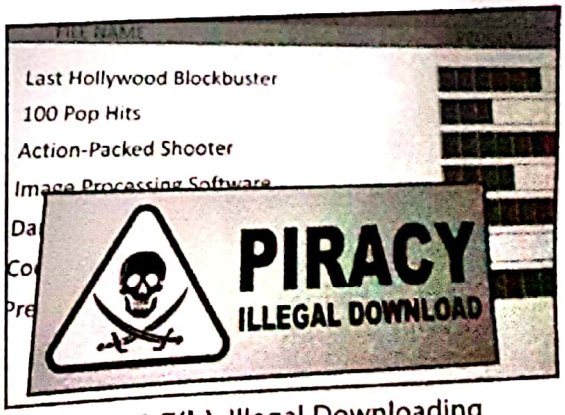


Fig. 9.7(b): Illegal Downloading

Digital piracy involves illegally sharing copyrighted media such as games, music, movies, TV shows and software. It does not matter whether you upload the content to share with others without permission or whether you download it for free or for a bargain price. Anyone involved in this type of operation is breaking the law.

For example, if you are a member of a website that shares music files without permission from the recording labels, you are involved in piracy, even if you pay a fee to the website for downloading them. Another example could be if you are able to access and download movies or TV shows for free or at extremely low prices, which are usually illegal copies, and thus come within the ambit of piracy. To avoid downloading media illegally, obtain it from reputable dealers that have the rights to sell it to you.

Therefore, illegal downloading should be discouraged as it incurs heavy losses to developers/producers and the person who is carrying out this illegal downloading can face punishment or fine due to this unethical act.

9.8.3 Child Pornography

Child pornography is publishing and transmitting obscene material about children in electronic form. In recent years, child pornography has increased due to easy access to the internet and easily available videos on the internet. Child pornography is the most heinous crime which occurs and has led to various other crimes such as sex tourism, sexual abuse of the child, etc.

There are provisions under Cyber law to stop child pornography, leading to punishments like,

- 1st conviction with 2 to 5 years of imprisonment and a fine of ₹ 1 lakh.
- 2nd or subsequent conviction with an imprisonment of up to 7-10 years and also a fine which may extend up to ₹ 10 lakh.

Child pornography laws provide severe penalties for producers and distributors in almost all Western societies, usually including incarceration, with shorter duration of sentences for non-commercial distribution depending on the extent and content of the material distributed. Convictions for possessing child pornography also usually include prison sentences, but those sentences are often converted to probation for first-time offenders.

9.8.4 Cyber Scams and Frauds

The term 'Internet fraud' generally refers to any scheme that uses one or more components of the internet, such as chat rooms, email, message boards, or websites, to present fraudulent solicitations to prospective victims, to conduct fraudulent transactions, or to transmit the proceeds of fraud to financial institutions or to others connected with the scheme.

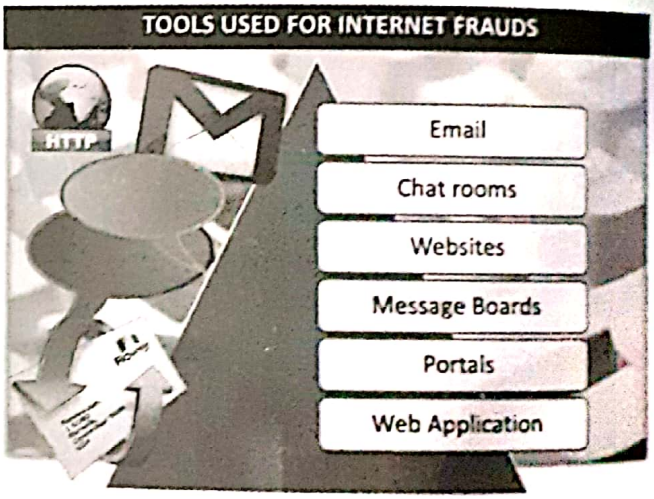


Fig. 9.7(c): Tools used for Internet Frauds

The various cyber frauds that are carried out online across the world are:

- **Credit/Debit Card Fraud**—This involves the use of credit/debit card to obtain money or acquire properties without appropriate authorization. Fraudsters obtain credit/debit card numbers of victims from unsecured web media and use them to carry out illegal or unethical transactions.
- **Non-delivery of Goods/Service Fraud**—This refers to a scam in which people are encouraged to pay for goods and/or service via a web portal, and then nothing is delivered to the buyer.
- **Spoofing/Phishing Scam**—Spoofing is a type of fraud in which a fraudster masquerades as another person by using the other person's identity to transact business and obtain vital information such as bank account numbers, credit card numbers and associated passwords. Phishing is a form of spoofing in which the web page of a particular entity can be duplicated and positioned with URL for the purpose of luring people to divulge vital financial information.

- **Identity Theft**—This entails the use of another person's personal information, without appropriate consent, for the purpose of fraudulent practices. Often personal information may even be leaked online or stolen during web surfing.

Identity theft is a type of fraud that involves using someone else's identity to steal money or gain other benefits. Online identity theft refers to an act of stealing someone else's personal information such as name, login details, etc., and then posing as that person online.

- **Auction Fraud**—In auction frauds, people are encouraged to participate in online auction and when money has been paid for specific items, the fraudster would send either a lower standard item or a counterfeit.

9.8.5 Cyber Forensics

Cyber forensics is an electronic discovery technique used to determine and reveal technical criminal evidence. It often involves electronic data storage extraction for legal purposes.

Although still in its infancy, cyber forensics is gaining traction as a viable way of interpreting evidence. Cyber forensics is also known as computer forensics. The various capabilities of cyber forensics are:

- Recover deleted files
- Find out what external devices have been attached and what users accessed them
- Determine what programs ran
- Recover web pages
- Recover emails and users who read them
- Recover chat logs
- Determine file servers used
- Discover document's hidden history
- Recover phone records and SMS text messages from mobile devices
- Find malware and data collected

Learning Tip: Any criminal act that is facilitated by the use of electronic gadgets such as computer, smartphone, laptop, etc., involving communications or information systems through the internet, is referred to as cybercrime.

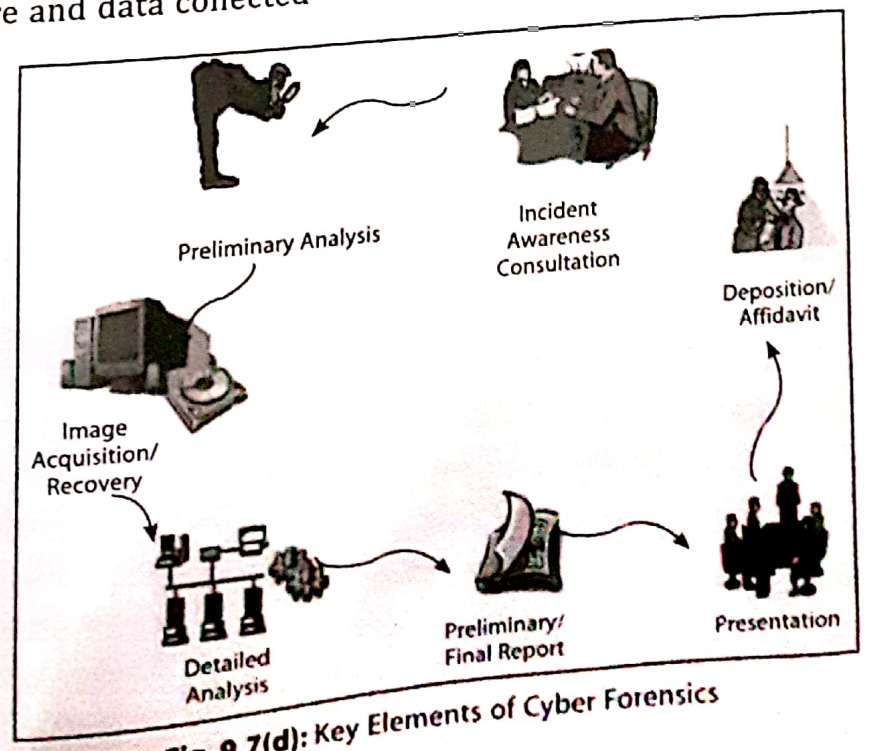


Fig. 9.7(d): Key Elements of Cyber Forensics

9.9 INFORMATION TECHNOLOGY ACT, 2000

The Information Technology Act, 2000 (also known as ITA-2000, or the IT Act) is an Act of the Indian Parliament (No 21 of 2000) notified on 17 October, 2000. It is the primary law in India dealing with cybercrime and electronic commerce.

The original Act contained 94 sections, divided into 13 chapters and 4 schedules. The laws apply to the whole of India. Persons of other nationalities can also be indicted under the law if the crime involves a computer or network located in India.

The IT Act is a set of recent legal enactments, currently existing in India, which provide legal support to computer users against cybercrime.

The cyber police works as a detector to discover cybercrimes. They have the authority in respect of all the offences committed under ITA (The Information Technology Act, 2000), Central Act No. 21 of 2000, or crimes related to Intellectual Property Rights. The cyber laws in India and the provisions for legal action and punishment have been explained in a nutshell in Fig. 9.8.

SECTION	OFFENCE	PENALTY
67A	Publishing images containing sexual acts	Imprisonment up to seven years, and/or fine up to Rs. 10,00,000.
67B	Publishing child porn or predating children online	Imprisonment up to five years and/or fine up to Rs. 10,00,000 on 1st conviction. Imprisonment up to seven years and/or fine up to Rs. 10,00,000 on 2nd conviction.
67C	Failure to maintain records	Imprisonment up to three years and/or fine up to Rs. 2,00,000.
68	Failure/refusal to comply with orders	Imprisonment up to seven years and/or possible fine.
69	Failure/refusal to decrypt data	Imprisonment up to three years and/or fine up to Rs 1,00,000.
70	Securing access or attempting to secure access to a protected system	Imprisonment up to ten years and/or fine.
71	Misrepresentation	Imprisonment up to three years and/or fine up to Rs. 1,00,000.

Fig. 9.8: IT ACT, 2000 (Cyber law) of India

IT Act, 2000—Amendments

A major amendment was made to the IT Act in 2008. It introduced Section 66A which penalized sending of "offensive messages". It also introduced Section 69, which gave authorities the power of "interception or monitoring or decryption of any information through any computer resource". Amendments also contained penalties for child pornography, cyberterrorism and voyeurism. The Act was passed in December 2008 and came into force in October 2009.

Here are some eco-friendly **waste disposal techniques** that you can use to dispose of electronic waste locally:

- **Give Your Electronic Waste to a Certified E-Waste Recycler**

  The positive aspect of e-waste recycling is that you have quite a few recycling options. You need to find an e-waste recycler which is officially certified by the Basel Action Network (BAN). BAN is a non-profit organization of recycling companies dedicated to recycling e-waste in a safe and responsible way. All members have to make a pledge and display their Pledge of Responsible Recycling. So, working alongside a certified recycler means that you don't have to worry about polluting another nation or risk losing your personal details to criminals.

  Precautions to be taken before donating or recycling your electronics:

  - Upgrade your computer instead of simply replacing it.
  - Format all your personal information from your product before discarding it.
  - Take out the batteries from your gadgets before getting rid of them.

- **Sell off Your Outdated Technology**

  As the old saying goes, 'One man's junk is another man's treasure'. This can be applied to helping you get rid of your old electronics. You can tap into online sites like craigslist, eBay or even resort to having a garage sale as this will help you get rid of your outdated electronics as well as earn some money.

- **Donating Your Outdated Technology**

  Old gadgets that you no longer need can be donated as they may be useful to others. Your old computer may be useful to either an NGO or a student. You should ask yourself these two questions before disposing of your old electronics:

  - Is the electronic item working?
  - Does the computer have any of your personal information?

  A lot of organizations and businesses offer electronic donation programs which you can choose from.

- **Visit Civic Institutions**

  Enquire amongst your government, universities and schools for any recycling programs they run as a lot of organizations have started assigning a certain day and place for environmentally conscious citizens to come and drop off their **e-waste.**

- **Give Back to Your Electronic Companies and Drop off Points**

  A lot of electronic companies tend to have an exchange policy whereby they take back your old gadgets when you buy a later version, sometimes offering you a discount on your new purchase. A few recycling companies have set up electronic drop off initiatives along with drop off points for products such as cellphones and tablets after which they are recycled. You can ask your local electronics shops regarding any information about drop off locations.

- **Safeguard Both the Environment and Your Sensitive Information**

  Electronics are an important part of our lives today but the flip side is the e-waste that comes along with it. So make sure to format your electronic devices before disposing them of in a proper manner as the consequences of not doing so can be dangerous.

9.13 GENDER AND DISABILITY ISSUES WHILE TEACHING AND USING COMPUTERS

While there is recognition of the potential of information and communication technologies (ICT) as a tool for the promotion of gender equality and empowerment of women, a "gender divide" has also been identified, reflected in the lower number of women accessing and using ICT as compared to men. Unless this gender divide is specifically addressed, there is a risk that ICT may exacerbate existing inequalities between women and men and create new forms of inequality. If, however, the gender dimensions of ICT—in terms of access and use, capacity-building opportunities, employment and potential for empowerment—are explicitly identified and addressed, ICT can be a powerful catalyst for political and social empowerment of women, and for promotion of gender equality.

Disability Issues

Clinical categories of cognitive disabilities include autism and Down syndrome. Less-severe cognitive conditions include the sub-category of so called learning disabilities, such as dyslexia (reading) and dyscalculia (mathematics). The functional disability perspective ignores the medical and behavioural causes of cognitive disabilities and focuses instead on the abilities and challenges the person with a cognitive disability faces.

Functional cognitive disabilities may involve difficulties or deficits involving:

- Problem-solving
- Attention
- Memory
- Math comprehension
- Visual comprehension
- Reading
- Linguistic (speech)
- Verbal comprehension

The following list shows the benefits that access to ICTs can bring to people within the wide spectrum of learning disabilities. These include:

- **Improved writing**—Standard word processors contain built-in tools for checking grammar, spell-checking and predictive typing. Specialized writing support programs, such as Clicker 5, can aid word recognition and writing through the use of symbols or pictures and speech supported grids.
- **Multimedia**—The use of multimedia such as graphics, sound and video can stimulate and encourage interaction and some degree of learning for people with more profound cognitive impairments, as well as for pre-literate children. Much of the software required to create multimedia is freely available online.
- **Sensory stimulation**—The use of switches, combined with specialized software games, can enable some people with profound and multiple learning difficulties to play basic cause-and-effect games and even develop some basic computer interaction skills. In addition to these computer-based activities, the use of augmentative and alternative communication strategies and devices, particularly those employing symbols, can aid communication for persons with more profound and multiple-learning impairments.

9.14 ROLE OF NEW MEDIA IN SOCIETY

The word 'media' is derived from the word 'medium', signifying mode or carrier. It was first used in respect of books and newspapers, *i.e.*, print media. But with the advent of technology, media has been transformed into new media, which has become as essential as our daily needs. The most popular amongst these is "social media".

New media is an umbrella term which includes different technologies. It incorporates blogs, podcasts, video games, virtual worlds, Wikipedia (online encyclopaedia), other mobile devices, interactive TVs, and even websites and emails.

New media technology has resulted in massive social changes in the behaviour and lifestyle patterns of consumers. New media has shaped modern culture by affecting the way people behave, communicate, learn and conceive of themselves and their world.

New media offers the following advantages:

(a) On-demand availability
(b) Any time availability
(c) Any place availability
(d) Convertible as per device
(e) Multiple forms of expression (audio/video/animation/text, etc.)
(f) Fast-spreading/easily sharable
(g) Interactive (readers can submit their feedback/reviews)

Learning Tip: Forms of new media constitute internet, websites, computer multimedia, social networking, video games and blogs.

In this section, we will discuss the role of new media in society in some specific ways.

CTM: New media means mass communication using IT tools and techniques/digital technologies such as internet.

9.14.1 Online Campaigns

A **campaign** is a planned set of activities that people carry out over a period of time in order to achieve something, such as a social or political change.

A planned, organized online campaign with rotation through a variety of targeted sites is not hard to coordinate, does not cost a great deal of money, and can generate traffic and sales faster than Search Engine Optimization (SEO). It takes a bit of research and a solid knowledge of the target customers, but one can plan and implement a simple online marketing plan in the space of a dedicated day.

An online campaign involves showing ads/ messages/pop-ups online for a specific product/ service or for any specific cause in a specific time period. An online campaign can be used to show ads/messages/pop-ups on multiple websites and in many formats, but they will all advertise the same product/service or idea (*Fig. 9.11*).

Fig. 9.11: Online Campaigns

The entire process of online campaigns constitutes the following steps:

Step 1: Identifying Customer

Step 2: Choosing Targets

Step 3: Budgeting

Step 4: Creating Campaign Content

Step 5: Tracking and Monitoring Campaigns/Ads

CTM: An online campaign is a planned set of activities carried online for a specific purpose or target.

Online campaigns are successful as they play a significant role in influencing masses through social media. Online campaigns offer many benefits such as:

(a) Growth in potential

(b) Reduced expenses

(c) Elegant communication

(d) Better control

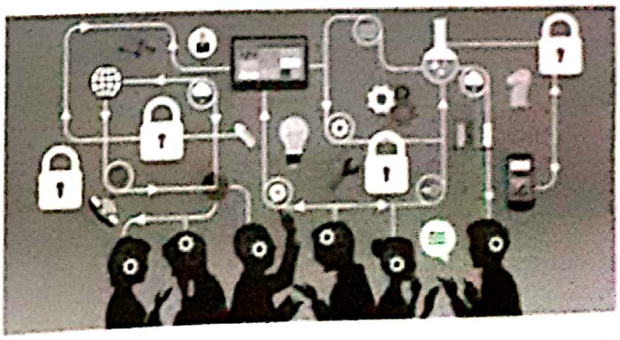
(e) Improved customer service

(f) Competitive advantage

However, online campaigns have certain limitations, such as:

1. Customers usually ignore ads
2. Viewing problems
3. Higher ad rates
4. Consumers easily get distracted
5. Too many options lead to confusion among customers
6. Face-to-face interaction is limited

9.14.2 Crowdsourcing



Crowdsourcing is the practice of turning to a body of people to obtain needed knowledge, goods or services. It is a term used to describe the process of getting work or funding from a large group of people in an online environment. While the idea behind crowdsourcing isn't new, its active use started around 2006. The phrase was initially coined by Jeff Howe when he described a world in which people outside of a company contribute towards a project's success. Beta versions of video games are perfect examples of crowdsourcing, granting players early access to the game to find out bugs before the finished product is released for sale and distribution. Crowdsourcing is a powerful business marketing tool to leverage creativity and resources of its own audience in promoting and growing the company for free.

Businesses, individuals and organizations of all kinds have used this process to solicit ideas and raise money as well as consolidate and promote information. This allows them to engage with a broader spectrum of sources than they would if they had used employees, suppliers and other traditional sources of expertise via conventional routes of engagement.

Benefits and Drawbacks

Crowdsourcing can give entities access to a wider set of expertise at a faster pace and at a lower cost than traditional routes. On the other hand, there is no guarantee that the crowd that the entities reach has the expertise, experience or resources to deliver what they need nor whether they're actually reaching the best sources to deliver the best possible outcome.

Crowdsourcing also has different, if not more, management requirements. Entities that turn to crowdsourcing may need to be more specific about their needs and expectations as well as how they evaluate responses, compensate participation or parse ownership of the final results. Entities may also need to manage thousands, if not tens of thousands or more, of responses to their requests.

Types of Crowdsourcing

There are different types of crowdsourcing.

1. **Crowd Contest:** A crowd contest is a popular type of crowdsourcing. In this case, an organization may use a crowd contest to create something, such as a graphic design.

   The organization may seek multiple qualified designers who will individually submit their best idea or ideas in the hope of being the one winner who will be paid for producing the top design.

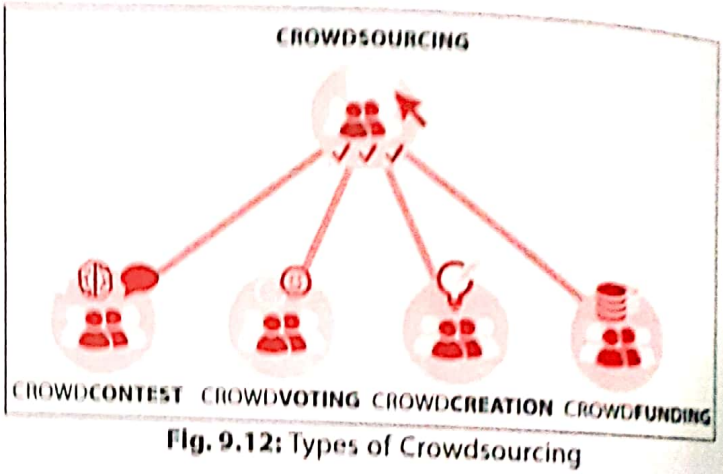


Fig. 9.12: Types of Crowdsourcing

2. **Macrotasking:** The process of macrotasking seeks out many respondents, assigning each one some task or a smaller part within a larger project. In this case, an entity can use crowdsourcing to seek out an individual who can handle a specific project—a type of crowdsourcing called macrotasking. Or the entity can seek out many respondents, assigning each one some task or a smaller part within a larger project. This is called microtasking.

3. **Crowdfunding:** Another subset of crowdsourcing is crowdfunding. It refers to a process in which individuals or non-profits (charitable) ask for money to cover the costs of an identified endeavour (whether it's a dream vacation, a down payment for a house, an artistic endeavour or something altruistic). Entrepreneurs can also seek out funds for their businesses, although investment laws can limit what they're able to request and promise in return.

4. **Crowd Voting:** This method is adopted to determine the best. Entities may opt to use crowd voting to engage stakeholders to determine a course of action. It's as simple as the name sounds: the crowd votes to determine the best. Entities may opt to use crowd voting to engage stakeholders, such as employees or customers, as well as the public to help prioritize ideas or help determine a course of action.

   Some areas of crowdsourcing have spurred specific recognition for the individuals involved. Citizen science, *for example*, recognizes the contributions that individuals make to science when they contribute data or help with reporting and analysis.

5. **Crowd Creation/Smart Mob:** Smart mob refers to a large group of people mobilized by social media to meet in a public place typically for the purpose of social or political protest.

The Battle of Seattle was one of the first examples of smart mobs. On November 30, 1999, protestors lined up the streets of Seattle to protest against the WTO. The protestors were able to communicate with one another by using technological devices such as websites, cellphones, laptops, radios and other similar technologies.

The impact of smart mob technology already appears to be both beneficial and destructive, used by some of its earliest adopters to support democracy and by others to coordinate terrorist attacks. The technologies that are beginning to make smart mobs possible are mobile communication devices and pervasive computing.

Smart mobs sometimes are manipulated by the dispatchers who control the 'mobbing system' (*i.e.*, those who own the contact list and the means to forward instant messages to a group) and are induced to cause distress and aggravation to individuals who have been targeted or singled out for whatever reason. There is a tendency to keep the dynamics of smart mobbing 'covert' and not to discuss such incidents on the internet.

Flash mobs are a specific form of smart mob, originally describing a group of people who assemble suddenly at a public place, do something unusual for a brief period of time, and then quickly disperse. The difference between flash and smart mobs is primarily with regard to their duration: flash mobs disappear quickly, but smart mobs can have a more enduring presence. The term flash mob is claimed to have been inspired by "smart mob".

CTM: A smart mob is an organized group of people who have gathered through the use of social media to protest or lobby for a publicized social or political issue.

However, there are a few downsides of crowdsourcing, such as consumers involved are not employees, which means organizations cannot contain or control them.

9.15 ISSUES WITH THE INTERNET

Over the course of the last two decades, internet has become nearly ubiquitous. From the ages of rare dial-up connections, our relatively reliable 4G network and endless options for free Wi-Fi at public establishments is its own kind of digital paradise. However, the internet isn't perfect, and if we want to build a world where every one has fast, reliable, and affordable internet access, there are a number of problems we'll need to resolve. Although, there are a number of issues with the internet, some of them are affecting us more in today's scenario, which are as under:

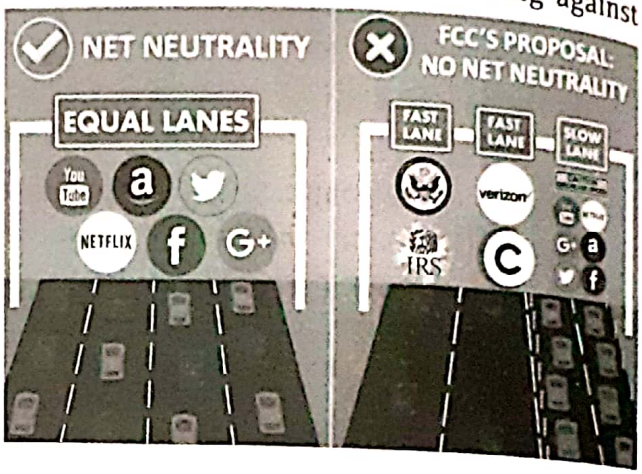
- Net Neutrality
- Internet as an Echo Chamber
- Internet Addiction
- Cloud Attacks
- AI-Built Security Issues
- Botnet Problems
- Lack of Confidence

Amongst the above-mentioned problems, we shall be discussing the most common internet issues, such as net (network) neutrality, internet as an echo chamber, and internet addiction.

9.15.1 Net Neutrality

Net neutrality means Internet Service Providers (ISPs) treat all information on the internet equally in an unbiased manner and do not discriminate against anyone or charge differently by user content, website platform, application, type of attached equipment, or method of communication.

It is based on the principle that individuals should be free to access all content and applications equally, regardless of the source, without Internet Service Providers discriminating against specific online services or websites. In other words, it is the principle that the company that connects you to the internet does not get to control what you do on the internet.



Columbia University law professor Tim Wu coined the term "network neutrality" in a 2003 paper about online discrimination. At the time, some broadband providers, including Comcast, banned home internet users from accessing virtual private network.

The Bush-era FCC (Federal Communications Commission) took a first pass at anti-discrimination rules for the internet in a policy statement in 2005. It prohibited internet service providers from blocking legal content or preventing customers from connecting the devices of their choosing to their internet connections.

Following are the indirect rules of net neutrality:

- No rules preventing blocking of website, services, or content online.
- No rules preventing throttling or slowing down of website or services online.
- No rules preventing paid prioritization where broadband providers give preferential treatment to some websites and services over others.

9.15.2 Internet as an Echo Chamber

Echo chamber is a metaphorical description of a situation in which beliefs are amplified or reinforced by communication and repetition inside a closed system. By visiting an echo chamber, people are able to seek out information which reinforces their existing views. This may increase political and social polarization and extremism. The term is a metaphor based on the acoustic echo chamber, where sounds reverberate in a hollow enclosure.

In simple words, little groups of like-minded members in any web app or social media over the internet are known as echo chambers. In today's hyperconnected world, we're all just a click away from each other. This should be bringing us closer together. But increasingly, we're finding ourselves

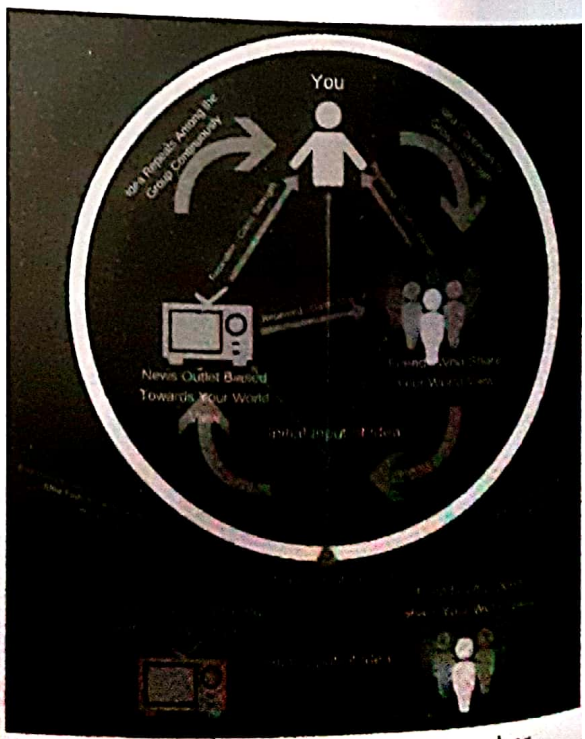


Fig. 9.13: Internet as an Echo Chamber

pulled further apart as our own natures push us into comfortable little groups of like-minded members, or 'echo chambers'.

Facebook, Twitter, Google and other internet services design their algorithms to filter out what they think we don't want to see. They also provide the tools to let us exclude any one who irritates us. The results seal us up in online bubbles of personally tailored content that makes us feel safe and confident and, importantly for these services, keeps us engaged on these sites for long.

9.15.3 Internet Addiction

As the name suggests, its focus is on compulsion with the internet—just as other areas of media addiction can be seen in television addiction, radio addiction, etc.

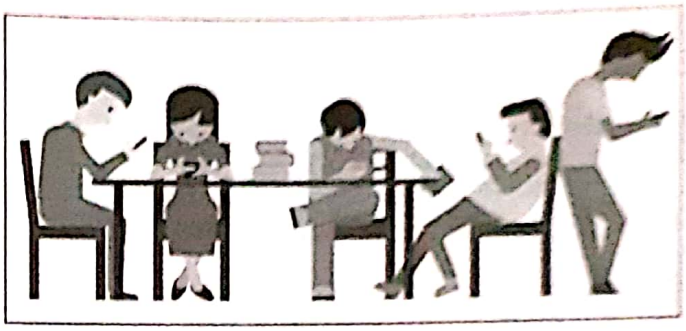


Fig. 9.14: Internet Addiction

Due to the explosion of the digital age, Internet Addiction Disorder has become the top culprit in technology addiction of late. The troubling thing about this disorder is that if you are addicted to it, you are endlessly surrounded by technology.

Thus, internet addiction refers to the compulsive need to spend a lot of time on the internet, to the point where relationships, work and health suffer.

Signs and symptoms of internet addiction: The disorder may present itself in both physical and emotional manifestations. Some of the emotional symptoms of Internet Addiction Disorder may include:

- Depression
- Feeling of guilt
- Feeling of euphoria when using computer
- Isolation
- Defensiveness
- Agitation
- Fear
- Boredom with routine tasks
- Dishonesty
- Anxiety
- Inability to prioritize or keep schedules
- No sense of time
- Avoidance of work
- Mood swings
- Loneliness
- Procrastination

Physical symptoms of internet addiction disorder may include:

- Backache
- Headaches
- Insomnia
- Poor nutrition (failing to eat or eating excessively to avoid being away from the computer)
- Poor personal hygiene (*e.g.*, not bathing to stay online)
- Neck pain
- Dry eyes and other vision problems
- Weight gain or weight loss

What are treatment options

Some of the more common psychological treatments of Internet Addiction Disorder include:

- Individual, group, or family therapy
- Dialectical Behavioural Therapy (DBT)
- Art therapy
- Reality therapy
- Behaviour modification
- Cognitive Behavioural Therapy (CBT)
- Recreation therapy

9.16 ROLE OF NEW MEDIA—CASE STUDIES

In the previous sections, we have discussed the role of new media in society and some issues related to the use of internet. In this section, we are taking up some real-life case studies exhibiting the important and significant role played by the new media, especially the social media and the internet.

9.16.1 Case Study 1: Arab Spring

The term "Arab Spring" was popularized in early 2011 by the Western media when the successful uprising in Tunisia against former leader Zine El Abidine Ben Ali emboldened similar anti-government protests in most Arab countries.

On December 17th, 2010, a Tunisian named Muhammad Bouazizi got into an altercation with a Tunisian officer. It was reported that Bouazizi was selling fruits within the city when the officer approached him and demanded that he stop doing so because he did not have a permit. The two got into an argument and the officer confiscated his fruit cart. The 'Time' magazine then carried a report which read: "It wasn't the first time it had happened but it would be the last. Not satisfied with accepting the 10-dinar fine that Bouazizi tried to pay ($7, the equivalent of a good day's earnings), the police woman allegedly slapped the scrawny young man, spat in his face and insulted his dead father" (Abouzeid, 2011). And as Abouzeid wrote: "What happened afterwards was the defining moment in what would be one of the largest protest movements the region had seen."

However, in July 2012, a report was published by the United States Institute of Peace based on an extensive content analysis of bit.ly links from the uprisings in Tunisia, Egypt, Libya, and Bahrain. The authors came to some conclusions that countered the initial assumption that social media was a causal mechanism in the uprisings. Bit.ly links, or short URLs, are predominantly used in social media such as Twitter.

9.16.2 Case Study 2: WikiLeaks

WikiLeaks, developed by Julian Assange, is a website intended to collect and share confidential information on an international scale. Born in 1971 in Townsville, Australia, Assange used his genius IQ to hack into the databases of many high-profile organizations. In 2006, Assange began work on WikiLeaks and earned the Time magazine's "Person of the Year" title in 2010. Seeking to avoid extradition to Sweden in an alleged assault case, Assange was granted political asylum by Ecuador and was holed up at the country's embassy in London since 2012.

Assange and WikiLeaks hit the headlines again during the summer of 2016 as the U.S. presidential race was narrowing down to two main contestants, Democrat Hillary Clinton and Republican Donald Trump. In early July, WikiLeaks released more than 1,200 emails from Clinton's private server during her tenure as Secretary of State. Later that month, WikiLeaks released an additional round of emails from the Democratic National Committee that hinted at an effort to undermine Clinton's primary opponent.

In April 2019, Assange's asylum was rescinded and he was arrested in London.

9.16.3 Case Study 3: Bitcoins

Bitcoin is a cryptocurrency or a digital currency that uses rules of cryptography for regulation and generation of units of currency. Bitcoin falls under the scope of cryptocurrency and was the first and most valuable digital currency. Bitcoin is a form of electronic cash. It is a decentralized digital currency without a central bank or single administrator.

Bitcoins are completely virtual coins designed to be 'self-contained' for their value, with no need for banks to move and store the money.

Once we own bitcoins, they possess value and trade just as if they were nuggets of gold in our pocket. We can use bitcoins to purchase goods and services online, or we can tuck them away and hope that their value increases over the years. Bitcoins are traded from one personal 'wallet' to another. A wallet is a small personal database.

MEMORY BYTES

- Cyber law is a term used to describe legal issues related to the use of communications technology, particularly "cyberspace", *i.e.*, the internet.
- Intellectual property refers to creation of the intellect (hence, the name); inventions, literary and artistic works, symbols, names, images and designs used in commerce are part of it.
- Digital Rights Management (DRM) is a term for access control technologies that are used by hardware manufacturers, publishers, copyright holders and individuals to limit the use of digital content and devices.
- All software are protected under copyright. Copyright is owned by the developer or publisher.
- FOSS grants user substantial rights and freedoms.
- Licences provide rules and guidelines for others to use your work. Open source licences help others to contribute to your work or project without seeking special individual permission to do so.
- GPL licensed code can't be used for paid and proprietary software.
- All electronic waste is made up of deadly chemicals such as lead, cadmium, beryllium, mercury and brominated flame retardants.
- Identity theft occurs when someone, without your knowledge, acquires a piece of your personal information and uses it to commit a fraud.
- Phishing is an attempt to acquire sensitive information such as username, password, and credit card details, etc.
- Look for HTTPS in the address bar when you enter any sensitive personal information on a website to make sure your data is encrypted.
- An identity verification service is used to ensure that users or customers provide information that is associated with the identity of a real person.
- You should set up privacy settings for a social media site on your own.
- Proper ICT training should be imparted to teachers to solve the problem of gender bias in computer classroom teaching and learning.

- Cyber forensics is also known as computer forensics.
- An online campaign is a planned set of activities carried online for a goal or purpose.
- Crowdsourcing refers to the practice of getting the contribution/funding from the crowd in the form of needed services, ideas or content, mainly from the online community.
- A smart mob is a large group of people mobilized by social media to meet in a public place typically for the purpose of social or political protest.
- Net neutrality is the principle that individuals should be free to access all content and applications equally, regardless of the source, without Internet Service Providers discriminating against specific online services or websites.
- The Arab Spring is the name given to a series of anti-government protests, uprisings, and armed rebellions that originated in Tunisia in December 2010 and spread across the Middle East in early 2011.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.
   (a) Any information about you or created by you that exists in digital form is referred to as ..... .
   (b) Stealing someone's intellectual work and representing it as your own is known as ..... .
   (c) Creative creations of mind such as patents, trademark and copyright are ..... property.
   (d) Any fraudulent business practice that extracts money from an ignorant person is called a(n) ..... .
   (e) Software which usually limit the functionality after a trial period are known as ..... .
   (f) IT Amendment Act, 2008 came into force from the year ..... .
   (g) Online ..... is a theft of personal information in order to commit fraud.
   (h) A ..... is an organized group of people who have gathered through the use of social media to protest for a social or political cause.
   (i) ..... is a specific type of disorder where a person can't strike a balance between their time online and offline.
   (j) The practice of attempting to acquire sensitive information from individuals over the internet by means of deception is called ..... .

   Answers: (a) digital property (b) plagiarism (c) intellectual
   (d) scam (e) shareware (f) 2009
   (g) identity theft (h) smart mob (i) Internet addiction
   (j) phishing

2. State whether the following statements are True or False.
   (a) Public Domain Software is free and can be used with restrictions.
   (b) An important factor of privacy is consumer consent.
   (c) Intellectual Property Rights are the rights of owners to decide how much information/data is to be shared or exchanged.
   (d) Copying and pasting data from internet or other digital resources is ethical.
   (e) E-document becomes legal after it is digitally signed.
   (f) E-waste is very hazardous if not handled carefully.
   (g) The data stored in biometric database can be updated.
   (h) Source code of proprietary software is normally available.
   (i) Free software is same as freeware.
   (j) Never respond to an email or advertisement that claims you have won something.

   Answers: (a) False (b) True (c) True (d) False (e) True (f) True
   (g) False (h) False (i) False (j) True

3. Multiple Choice Questions (MCQs)

(a) Using someone else's twitter handle to post something will be termed as:
(i) Fraud (ii) Identity theft (iii) Online stealing (iv) Violation

(b) Standard security protocol that establishes encrypted links between a web server and a browser is called ..... .
(i) Online safety technology (ii) SDT technology
(iii) Web encryption technology (iv) Secure Sockets Layer Technology

(c) Intellectual Property Rights protect the use of information and ideas that are of:
(i) Ethical Value (ii) Moral Value
(iii) Social Value (iv) Commercial Value

(d) Credit card fraud may include:
(i) Stealing of credit card (ii) Unauthorized and illegal use of credit card
(iii) Stealing information (iv) Phishing

(e) The term "Intellectual Property Rights" covers:
(i) Copyrights (ii) Trademarks (iii) Patents (iv) All of the above

(f) Which of the following is a/an open-source software?
(i) Microsoft Windows (ii) Adobe Photoshop (iii) MySQL (iv) MS Powerpoint

(g) "Rice Bucket Challenge" is an example of:
(i) Crowd Sourcing (ii) Online Campaigns
(iii) Smart Mobs (iv) Internet Addiction

(h) Online personal account, personal websites are examples of:
(i) Digital Wallet (ii) Digital Property (iii) Digital Certificate (iv) Digital Signature

(i) Unsolicited commercial email is known as ..... .
(i) Spam (ii) Malware (iii) Virus (iv) Spyware

(j) Which of the following is not a type of a cybercrime?
(i) Data theft (ii) Forgery
(iii) Damage to data (iv) Installing antivirus for protection

Answers: (a) (ii) (b) (iv) (c) (iv) (d) (ii) (e) (iv) (f) (iii)
(g) (ii) (h) (ii) (i) (i) (j) (iv)

SOLVED QUESTIONS

1. What do you understand by 'Privacy of data'?

Ans. The ethical and legal rights that individuals have with regard to control over discussions and use of their personal information is known as privacy of data.

2. If someone hacks your website, who would you complain to?

Ans. The complaint has to be lodged to the police under the IT Act.

3. What is the importance of cyber law?

Ans. In internet, communication technology uses the means of transferring textual messages, pictures and much more. Each time there may be a number of threats on either the sender's or the receiver's side which create a bridge between networking communication. To sort out these problems, the Internet Security Council took a number of precautions. These predefined rules are called cyber laws or laws of internet.

[CBSE Delhi 2015]

4. Write two applications of cyber law.

Ans. Cyber law encompasses a wide variety of political and legal issues related to the internet and other communication technologies, including intellectual property, privacy, freedom of expression and jurisdiction.

5. Jhilmalini has stolen a credit card. She used that credit card to purchase a laptop. What type of offence has she committed?

Ans. She has committed a fraud.

6. Name the primary law in India dealing with cybercrime and electronic commerce.

Ans. The primary law is Information Technology Act, 2000.

7. Sutapa received an email from her bank stating that there is a problem with her account. The email provides instructions and a link, by clicking on which she can log on to her account and fix the problem. Help Sutapa by telling her the precautions she should take when she receives these type of emails.

Ans. She should check whether it is a valid bank site or not by checking in the url https. It is always better to type the url and then login to the site. She should not click on the link provided in the email.

8. Explain any two ways in which technology can help students with disabilities.

Ans. Different types of ICT tools assist people with learning disabilities to achieve positive outcomes. They are:

- Talking Word processors
- Screen Readers
- Conversion of local language to Braille
- Eye Tracking mouse

9. Explain the role of online social media campaigns, crowd sourcing and smart mobs in society.

Ans. Role of Social Media Campaigns:

- A social media campaign should focus around a singular business goal, whether it is on Facebook or Instagram. Common goals for social media campaigns include:
- Getting feedback from users.
- Building email marketing lists
- Increasing website traffic
- Crowd sourcing is the practice of engaging a 'crowd' or group for a common goal — often innovation, problem solving, or efficiency. It is powered by new technologies, social media and web 2.0. Crowd sourcing can take place on many different levels and across various industries.
- Smart mobs, so named because each person in the group uses technology to receive information on where to go and what to do. This ability to stay on top of current events makes smart mobs extremely effective.

10. Ms Sharmishtha has many electronic gadgets which are not usable due to outdated hardware and software. Help her to find three best ways to dispose the used electronic gadgets.

Ans.
1. Give your electronic waste to a certified e-waste recycler.
2. Donating your outdated technology.
3. Give back to your electronic companies and drop off points.

11. Which of the following come under cybercrime? [HOTS]
1. Theft of a brand new sealed laptop.
2. Access to a bank account for an unauthorized money transaction.
3. Modification in a company's data with unauthorized access.
4. Photocopying a printed report.

Ans. (2) and (3)

12. List three points of network security components. [HOTS]

Ans. The three network security components are:

- Anti-virus and anti-spyware.
- Firewall, to block unauthorized access to your network.
- Intrusion Prevention Systems (IPS) to identify fast-spreading threats such as zero-day or zero-hour attacks.

13. Name the crimes for which cyber laws are enforced strictly in India.

Ans. These are:

(a) cybercrimes,
(b) electronic and digital signatures,
(c) intellectual property, and
(d) data protection and privacy.

14. What are privacy laws?

Ans. Privacy laws refer to laws that deal with regulating, storing and using personally identifiable information of individuals, which can be collected by government, public or private organizations, or other individuals. Privacy laws are considered within the context of an individual's privacy rights or within reasonable expectations of privacy.

15. Explain phishing.

Ans. Phishing is the fraudulent attempt to obtain sensitive information such as usernames, passwords and credit card details (and money), often for malicious reasons, by disguising as a trustworthy entity in an electronic communication. Phishing is typically carried out by email spoofing or instant messaging, and it often directs users to enter personal information at a fake website, the look and feel of which is identical to the legitimate one and the only difference is the URL of the website in question.

16. What is meant by the term cyber forensics?

Ans. Cyber forensics is an electronic discovery technique used to determine and reveal technical criminal evidence. It often involves electronic data storage extraction for legal purposes. Although still in its infancy, cyber forensics is gaining traction as a viable way of interpreting evidence. Cyber forensics is also known as computer forensics.

17. List some of the risks that internet users face.

Ans. The main risks internet users face include:

- Inappropriate or unwanted contact (cyberbullying and spamming)
- Inappropriate or inaccurate content (pornography and hate sites)
- Deceptive or fraudulent commerce (counterfeit and malicious sites)

18. What does a phishing email look like? [HOTS]

Ans. Typically, a phishing email appears to come from a financial institution, a large company, a chain store, a social networking site, or a government agency. The messages try to mimic a legitimate site by using the same or similar colours, logos, fonts and layout. And they often include a link to a legitimate-looking but phony web page that asks you to enter personal information. One tip-off that an email may be phishing is the use of phrases such as "Verify your account" and "Your account will be closed" if you don't provide certain sensitive information such as login name and password. A legitimate business will never ask for such personal information via email. Promising big lottery winnings, prizes or other types of windfall if you pay money upfront is another common phishing scam.

19. What if I have already responded to a phishing email and fear I may be a victim of identity theft? [HOTS]

Ans. Immediately change the password to your online account, and notify the fraud department of the institution that was mimicked. When you receive your credit reports, review them carefully for accounts you don't recognize, which may be signs of identity theft. Consider placing a security freeze on each report, which means that no new credit can be extended without your personal approval.

20. How can we prevent identity thefts?

Ans. Identity theft refers to the acquisition of personal data of the victim and using it for illegal purposes. It is the most common type of fraud that may lead to financial losses and at times may result in criminal action as the victim might be personified.

A few steps to follow in order to prevent identity thefts:

- Ensure a strong and unique password
- Avoid posting confidential information online
- Do not post personal information on social media
- Shop from known and trusted websites only
- Use the latest version of the browsers
- Install advanced malware and spyware tools
- Use specialized security solutions against financial data
- Always update your system and the software
- Protect your social security number
- Download only well-known apps and share limited details

21. What are the different ways in which authentication of a person can be performed?

Ans. The different methods of user identification are:

1. **Password:** This is something that the user should know since they start their activity.
2. **Token:** This is something the user is provided with and should have it to perform an activity.
3. **Biometrics:** This means biological characteristics of a person registered for verification.
4. **OTP:** A one-time pin or password is sent to the user through which they verify their identity.

22. What do you mean by cyber security?

Ans. Cyber security is the combination of best processes and practices to ensure the security of networks, computers, programs, data and information from attack, damage or unauthorized access.

23. What do you understand by the term 'open data'?

Ans. **Open data** is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions of copyright, patents or other mechanisms of control. The goals of the open data movement are similar to those of other "open" movements such as open source, open hardware, open content, open education, open educational resources, open government, open knowledge, open access, open science, and open web.

24. "Open data can be a powerful force for public accountability." Explain.

Ans. "Open data can be a powerful force for public accountability—it can make existing information easier to analyze, process, and combine than ever before, allowing a new level of public scrutiny." Governments that enable public viewing of data can help citizens engage within the governmental sectors and "add value to that data."

25. What are the different types of cybercrimes?

Ans. Common types of cybercrimes include online bank information theft, identity theft, online predatory crimes and unauthorized computer access. More serious crimes like cyberterrorism are also of significant concern.

Cybercrime encompasses a wide range of activities, but these can generally be broken into two categories:

- Crimes that target computer networks or devices. These types of crimes include viruses and denial-of-service (DoS) attacks.
- Crimes that use computer networks to advance other criminal activities. These types of crimes include cyberstalking, phishing and fraud or identity theft.

26. What does cybercrime mean?

Ans. Cybercrime is defined as a crime in which a computer is the object of the crime (hacking, phishing, spamming) or is used as a tool to commit an offence (child pornography, hate crimes, etc.). Cybercriminals may use computer technology to access personal information, business trade secrets, or use the internet for exploitative or malicious purposes.

27. What security measures should be taken against phishing attacks?

Ans. The following security precautions are recommended against phishing:

- Use updated computer security tools, such as anti-virus software, spyware and firewall.
- Never open unknown or suspicious email attachments.
- Never divulge personal information requested by email, such as your name or credit card number.
- Double check the website URL for legitimacy by typing the actual address in your web browser.
- Verify the website's phone number before placing any calls to the phone number provided via email.

28. How is digital piracy related to illegal downloads?

Ans. Digital piracy involves illegally sharing copyrighted media such as games, music, movies, TV shows and software. It does not matter whether you upload the content to share with others without permission or whether you download it for free or for a bargain price.

Illegal downloads from the internet involve violation of copyright laws where users download material, such as music, movies, and other forms of media, without properly purchasing the product or doing so without proper permission of copyright holder. Hence, illegal downloads result in digital piracy.

29. Explain the term 'Web beacons'. [HOTS]

Ans. Pictures in email messages—also called 'Web beacons'—can be adapted to secretly send a message back to the sender. Spammers rely on information returned by these images to locate active email addresses. Images can also contain harmful code and be used to deliver a spammer's message in spite of filters.

30. How can we block Web beacons? [HOTS]

Ans. (a) The best defence against Web beacons is to prevent pictures from downloading until you've had a chance to review the message.

(b) Beware of malware disguised as holiday greetings in email.

(c) Approach links in email, on social networking sites, or in IMs with caution.

31. Explain gender and disability issues while teaching and using computers.

Ans. Studies have shown that in many countries computer use in schools is dominated by men. Female teachers have less regard for their own skills and knowledge than their male counterparts.

Females know less about information technology, enjoy using the computer less than male students, and perceive more problems with software.

32. What are the possible causes that contribute to gender and disability issues? [HOTS]

Ans. Possible causes that contribute to this problem are:

1. Differences in parental support.
2. Restricted access to computers.
3. Fewer female role models and a casual approach towards computer activities in school.

Gender differences are being found both inside and outside the school. This means that both teachers and parents have to be made aware of this before remedial measures are initiated. Same is the case with people with disabilities.

33. What factors contribute to low success rate among people with disabilities in various professions?

Ans. Several factors contribute to low success rate of people with disabilities in professional areas. Some of these are:

- Some students with disabilities experience isolation as a result of not being accepted by their peers.
- They rarely have access to positive role models with disabilities.
- Support systems available in high school cease after graduation, and many students with disabilities lack the determination to prepare for college exams and compete for job opportunities, besides independent living skills necessary to make successful transitions to adulthood.
- Youths with disabilities continue to live with their parents or in other dependent living situations after high school more often than their peers without disabilities; they also engage in fewer social activities.
- Social isolation has a negative effect on personal as well as academic and career success.

34. Describe some common forms of student plagiarism.

Ans. According to 'The Reality and Solution of College Plagiarism' created by the Health Informatics Department of the University of Illinois, Chicago, mainly there are 10 forms of plagiarism that the students commit:

1. Submitting someone else's work as their own.
2. Taking passages from their own previous work without adding citations.
3. Rewriting someone else's work without properly citing sources.
4. Using quotations without citing the source.
5. Interweaving various sources together in the work without citations.
6. Citing some passages, and not all, that should be cited.
7. Melding together cited and uncited sections of the work.
8. Providing proper citations without much changing the structure and language of the borrowed ideas.
9. Citing the source inaccurately.
10. Relying too much on others' works and failing to bring original ideas into the text.

35. What do you understand by Biometric Authentication?

Ans. **Biometrics** is the technical term for body measurements and calculations. It refers to metrics related to human characteristics. Biometric authentication (or realistic authentication) is used in Computer Science as a form of identification and **access control**. It is also used to identify individuals in groups that are under **surveillance**.

36. Describe measures to recycle your e-waste safely.

Ans. The following measures can be adopted to recycle the e-waste safely:

1. **Use a certified e-waste recycler:** Find an e-waste recycler certified by the Basel Action Network (BAN). BAN is a non-profit organization devoted to certifying e-Stewards, recyclers who are committed to safely and responsibly recycling electronics. Members take and demonstrate the Pledge of Responsible Recycling, so working with a certified e-Steward means you don't have to worry that your gadget will become another nation's pollution or a criminal's newest project. BAN's recycler locator will help you find the certified safety and comfort of e-Stewards in your area.
2. **Visit Civic Institutions:** Check with your local government, schools, and universities for additional responsible recycling options. With e-waste becoming such a large problem, government offices and schools are assigning days when citizens can bring unwanted electronics to a designated drop off location.

   Many communities post a calendar that will include recycling days, so check your local paper or visit their website. When you recycle your items locally, you can make the occasion a day trip and a community event. Encourage your neighbours to join you and spread the word about educated e-waste disposal.
3. **Explore Retail Options:** Best Buy, *for example*, isn't certified through BAN's e-Steward program, but they do have an effective recycling program in all of their stores. They claim to use only recyclers which adhere to the highest standards of e-waste processing. Specifically, their website discloses that the e-waste that you bring to their stores will not end up in a foreign country or in any landfill:

   "We partner directly with a short list of qualified, respected recycling companies who ensure all products collected for recycling through Best Buy are handled responsibly. These recycling companies meet our standards, and we encourage them to examine and consider additional third-party standards for responsible practices (such as the EPA R2 and e-Stewards)."

   You can drop off all kinds of e-waste for recycling at Best Buy including:

   - Cell phones
   - GPS devices
   - Paper shredders
   - Laptops
   - TVs
   - Speakers
   - Memory cards
   - Notebooks
   - Power cords
   - DVD players
   - Desktops
4. **Donate your Electronics:** Reusing is always better than recycling. If your electronics still have life left, you can reduce e-waste pollution and share technology with people who wouldn't otherwise have access to it.

37. List the key features that make Biometric systems most authenticated and trusted.

Ans. A number of factors are weighed before selecting a particular biometric for use in a specific application. They include:

- **Universality:** It means that every person using a system should possess the trait.
- **Uniqueness:** This means the trait should be sufficiently different for individuals in the relevant population so that they can be distinguished from one another.
- **Permanence:** It relates to the manner in which a trait varies over time. More specifically, a trait with 'good' permanence will be reasonably invariant over time with respect to the specific matching algorithm.
- **Measurability (collectability):** This relates to the ease of acquisition or measurement of the trait. In addition, acquired data should be in a form that permits subsequent processing and extraction of the relevant feature sets.
- **Performance:** It relates to accuracy, speed and robustness of technology.
- **Acceptability:** This relates to how well individuals in the relevant population accept the technology so that they are willing to have their biometric trait captured and assessed.
- **Circumvention:** It relates to the ease with which a trait might be imitated using an artefact or substitute.

38. How can disability lead to less participation in STEM (Science, Technology, Engineering and Mathematics) areas? [HOTS]

Ans. (i) Individuals with disabilities experience far less career achievement than their non-disabled peers.

(ii) Adults with disabilities are more likely to be unemployed and often experience underemployment, dislike their job status, and work in jobs with low socio-economic status.

(iii) Individuals with disabilities are under-represented in STEM professions, and scientists and engineers with disabilities experience higher unemployment rate than do other scientists and engineers.

UNSOLVED QUESTIONS

1. What do you mean by cybercrime?
2. What is meant by cyber ethics? How are they important to us?
3. What do you mean by identity theft? How can we prevent it?
4. What do you understand by illegal download?
5. What is child pornography?
6. What is phishing?
7. What is cyber law?
8. How is cyber law implemented in India?
9. What is IT Act, 2000 of India?
10. What amendments were introduced in the IT Act, 2008?
11. How do people unknowingly commit cybercrime?
12. What are the different ways to do phishing?
13. What precautions should be taken to avoid phishing?
14. What procedure is followed by the police to track/investigate cybercrime cases?
15. What do you understand by the concept of cyber forensics?
16. What precautions should be taken so that students do not indulge in cybercrime unknowingly?
17. Define e-waste. What are the various methods for effective e-waste management?
18. What is meant by gender discrimination? How does it hamper classroom teaching and learning process?
19. Differentiate between Proprietary Licence and Open Source Licence.
20. How secure is Biometrics Authentication System?
21. What is the relationship between ethics and law?
22. What is the difference between fairness and ethics in law?
23. How do ethical norms differ from the law?
24. What do you mean by privacy law?
25. Which types of gender and disability issues are created while teaching and using computer?
26. List the benefits that ICT can bring about to people with learning disabilities.
27. What do you understand by e-waste management? What are the ways for proper disposal of used electronic gadgets?
28. What is the difference between threat, vulnerability and risk?
29. What is the goal of information security within an organization?
30. What is the difference between encoding, encryption and hashing?
31. How would you implement two-factor authentication for a public facing website?
32. In how many ways can you secure your system from cyberattacks?
33. What is IT Act, 2000? What has been the impact of this Act?

APPENDIX A

CBSE SAMPLE QUESTION PAPER (SOLVED)

CLASS XII

INFORMATICS PRACTICES (065)

TIME: 3 Hrs M.M: 70

General Instructions:

1. All questions are compulsory.
2. Question Paper is divided into 4 sections A, B, C and D.
3. Section A comprises questions (1 and 2)
   - Question 1 comprises Data Handling-2 (DH-2) (Series,Numpy)
   - Question 2 comprises questions from Data Handling-2 (DH-2) (Dataframes and its operations)
4. Section B comprises questions from Basic Software Engineering.
5. Section C comprises questions from Data Management-2 (DM-2)
6. Section D comprises questions from Society, Law and Ethics-2 (SLE-2)

SECTION–A

1. Answer the following questions:

(a) Name one function to create an evenly-spaced ndarray. (1)

Ans. linspace()

(b) Name any two functions that can perform arithmetic operations on ndarray. (1)

Ans. add(), subtract(), multiply() and divide() [any two]

(c) What is data visualization? (1)

Ans. Data visualization refers to the graphical or visual representation of information and data using visual elements like chart, graphs, maps, etc.

Or

What is the role of legends in graph/chart?

Ans. Legends are used to name the patterns in the graph/chart.

(d) Write a Python program to find sum of elements in each row of a two-dimensional array of shape (3,3). (Assume any (3,3) ndarray) (2)

Ans.
```
import numpy as np
A=np.array([[1,1,2],[3,5,8],[13,21,34]])
print(A)
for i in range(3):
    sum1=0
    for j in range(3)):
        sum1=sum1+A[i][j]
    print(sum1)
```

(e) Write a NumPy program to generate five random numbers from the normal distribution. (2)

Ans.
```
import numpy as np
x = np.random.normal(size=5)
print(x)
```

(f) Write a Python statement to split the following array 'Arr' into three different arrays(Arr1,Arr2,Arr3). (2)

Ans. `np.split(Arr,[2,5],axis=1)` or `np.hsplit(Arr,[2,5])`

(g) Explain the following terms of NumPy arrays: (3)

(i) Rank (ii) Shape

(iii) Itemsize

Ans. (i) Rank: The number of axes in an ndarray is called its rank.

(ii) Shape: The number of elements along each axis of an ndarray is called its shape.

(iii) Itemsize: Size of each element of an ndarray in bytes.

Or

Given two ndarrays:

```
A=np.array([1,2,3,4,5])
B= np.array([3,4,6,-1,-3])
```

Write the code to calculate:

(i) Covariance (ii) Correlation

(iii) Regression

Ans. (i) `cov=np.cov(A,B)` (ii) `cor=np.corrcoef(A,B)`

(iii) `np.polyfit(A,B,1)`

2. (a) Which method is used to calculate the average value of the dataset from a dataframe? (1)

(i) iterrows() (ii) iteritems()

(iii) mod() (iv) median()

Ans. (iv)

(b) In Pandas, S is a series with the following data: (1)

```
S=pd.Series([15,10,15,20,25])
```

Find the output of: `S[1:3]`

Ans. 1 10

2 15

Or

What will be the output of the following code?

```
import pandas as pd
import numpy as np
Arr1 = np.array([[1,2,3],[4,5,6]])
df = pd.DataFrame(Arr1, columns = ['num1','num2','num3'])
print(df.iloc[0,2])
```

Ans. 3

(c) State true/false: (1)

The applymap() applies the given function on each individual data element of the dataframe.

Ans. True

(d) Which method is used to change the name of the existing indexes/column label in a dataframe? (1)

Ans. rename()

(e) Explain Sorting in Pandas. What are the two methods of sorting? (2)

Ans. Sorting means arranging the contents in ascending or descending order. Data of dataframe can be sorted according to the values of row and column. By default, sorting is done on row labels in ascending order.

The two methods of sorting in Pandas are:

(i) sort_values()

(ii) sort_index()

(f) Write the syntax to add a column in a dataframe. (2)

Ans. You can add new columns to an already existing dataframe.

Syntax to add or change a column:

```
<dfobject>.<Col_Name>[<row label>]=<new value>
```

(g) A dictionary Grade contains the following: (3)

```
Grade={'Name':['Rashmi','Harsh','Ganesh', 'Priya','Vivek'],
       'Grade':['A1','A2','B1','A1','B2']}
```

Write statements for the following:

(i) Create a dataframe named "Gr".

(ii) Add a column called 'Marks' with the following data: [97,92,95,89,96,82]

(iii) Delete 3rd and 5th rows.

Ans. (i) `Gr = pd.DataFrame(Grade)` (ii) `Gr["Marks"]=[97,92,95,89,96,82]`
(iii) `Gr.drop([2,4])`

(h) From the following ordered set of data: (3)

45, 49, 57, 67, 67, 68, 70, 70, 71, 73, 77, 79, 79, 80, 80, 81

(i) Create a horizontal boxplot (ii) Show means in the boxplot

(iii) Create boxplot without the box

Ans. (i)
```
import matplotlib.pyplot as pl
A=[45, 49, 57, 67, 67, 68, 70, 70, 71, 73, 77, 79, 79, 80, 80, 81]
pl.boxplot(A,vert=False)
```
(ii)
```
import matplotlib.pyplot as pl
A=[45, 49, 57, 67, 67, 68, 70, 70, 71, 73, 77, 79, 79, 80, 80, 81]
pl.boxplot(A, showmeans=True)
```
(iii)
```
import matplotlib.pyplot as pl
A=[45, 49, 57, 67, 67, 68, 70, 70, 71, 73, 77, 79, 79, 80, 80, 81]
pl.boxplot(A, showbox=False)
```

(i) Given the dataframe Weather, do the following: (4)

(i) Write command to compute sum of every column.

(ii) Write command to compute mean of column "Rainfall".

(iii) Write command to compute average of all the columns for last 10 rows only.

(iv) Write command to compute sum of every row.

	mintemp	maxtemp	Rainfall	Humidity
0	20.6	28.9	24.3	30
1	20.7	28.7	26.5	40
2	21	30	24.8	70
3	19	25	25.9	60
4	7.6	19.2	15.4	50
5	8.9	18.9	15.2	48
6	13.4	22.1	14.1	25
7	14.8	24.5	13.6	55
8	20.8	35.4	18.4	18
9	20	34.5	19.5	27
10	21.1	34.5	12.7	52

Ans. (i) `Weather.sum()` (ii) `Weather['Rainfall'].mean()`
(iii) `Weather.loc[:11,:].mean()` (iv) `Weather.sum(axis=1)`

SECTION–B

3. (a) SDLC stands for _____. (1)

(i) System Design Life Cycle (ii) Software Design Life Cycle

(iii) System Development Life Cycle (iv) Software Development Life Cycle

Ans. (iii)

(b) The model that leads to faster delivery of software is referred to as _____. (1)

(i) Component-based model (ii) Spiral model

(iii) Evolutionary model (iv) Waterfall model

Ans. (i)

(c) What is Cohesion? (1)

Ans. Cohesion is a measure that defines the intra-dependability among the elements of the module.

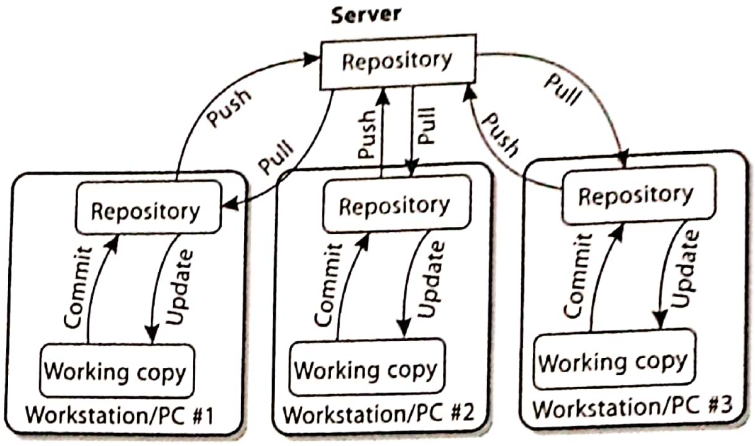
(d) What is verification and validation? (2)

Ans. **Verification:** Verification is a term that refers to a set of activities which ensures that software implements a specific function.

Validation: It refers to a set of activities which ensures that the software has been built according to the needs of clients.

(e) Differentiate between Commit and a Push request on a version control system with diagram. (3)

Ans. Commit refers to updating the file in local repository on a distributed version control system.
Push refers to updating the file in remote repository on a distributed version control system.



Or

What are the objectives of requirement analysis phase?

Ans. Objectives of requirement analysis phase are:

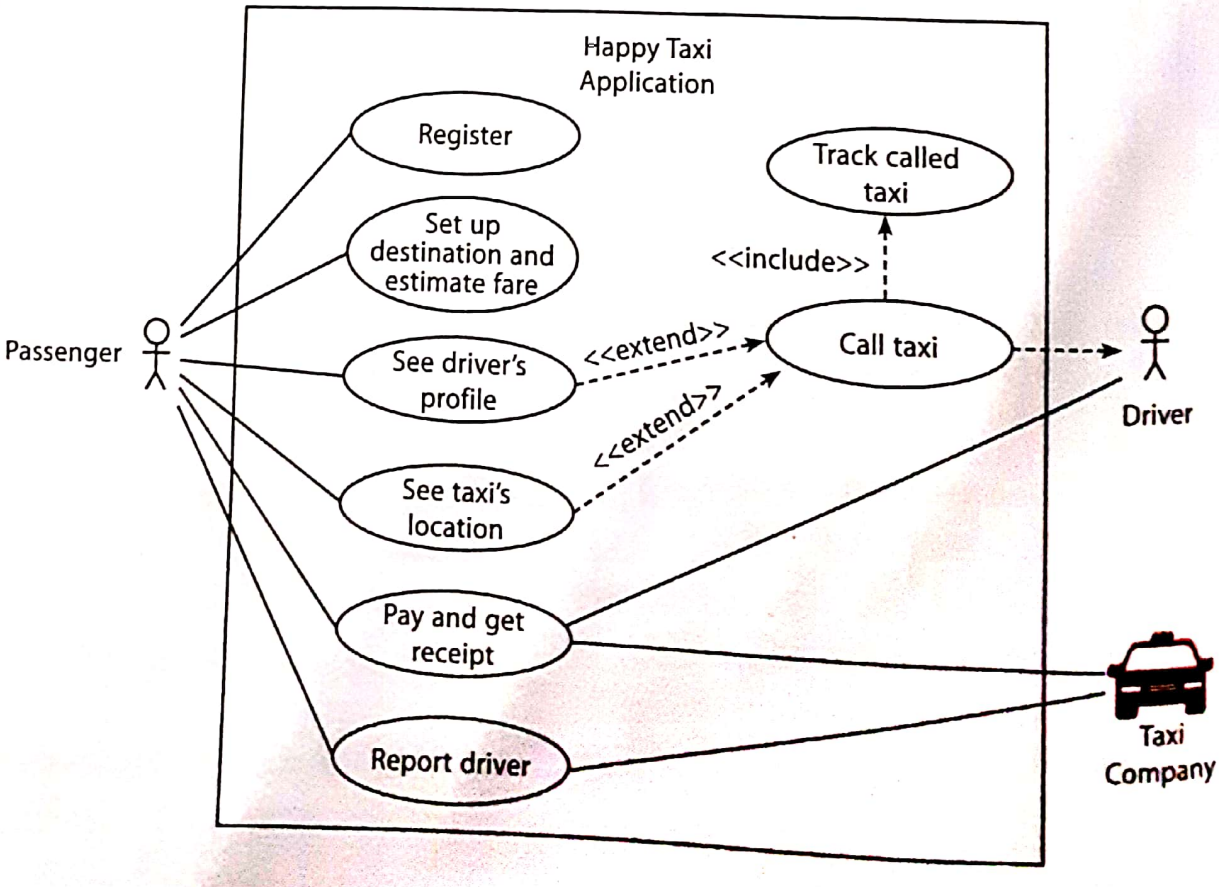
(i) It describes what the customer requires.
(ii) It establishes a basis for creation of software design.
(iii) It defines a set of requirements that can be validated once the software design is built.

(f) Write and explain scrum events in one line. (3)

Ans. (i) Sprints—A time box of one month or less to create saleable product.
(ii) Sprints planning meeting—The work to be performed is planned in this event by entire scrum team.
(iii) Daily scrum meeting—A 15-minute time box event for development team to sync activities and plan for the next 24 hours.
(iv) Sprint review—An informal meeting and presentation of the increment to inspect the increment.

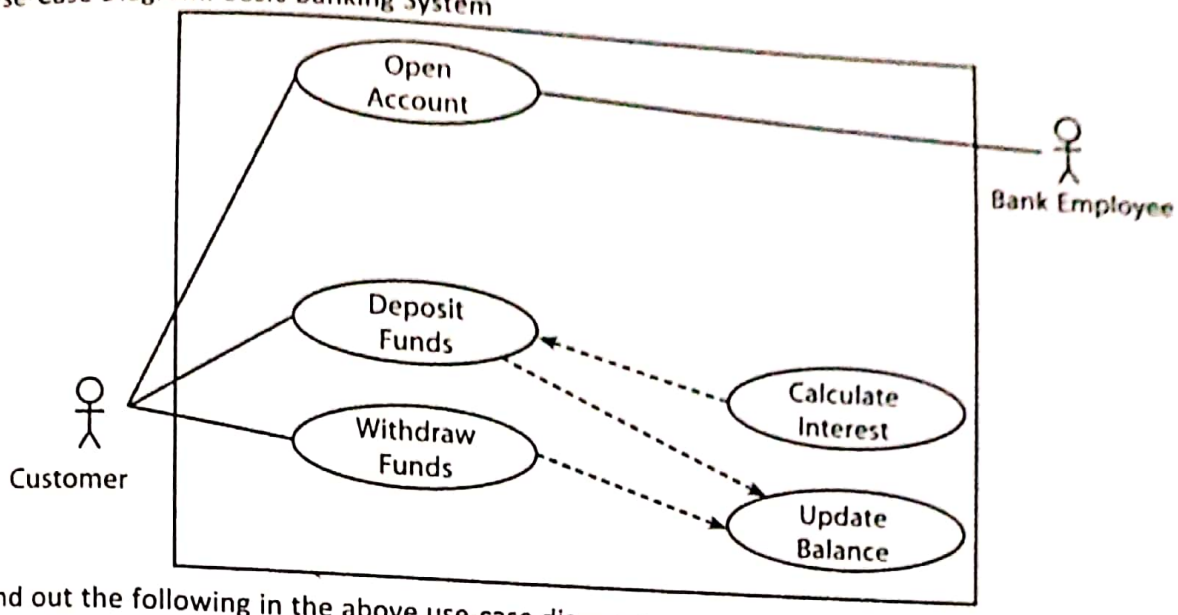
(g) Draw a use-case diagram for a Taxi Booking Application. (4)

Ans.



Or

Use-Case Diagram: Basic Banking System



Find out the following in the above use-case diagram:

(i) Actors (ii) Relation

(iii) Use Cases (iv) Main Flow

Ans. (i) Actors: Bank Employee, Customer

(ii) Relation:

Deposit Funds ←<<extend>>---- Calculate Interest

Deposit Funds ----<<include>> → Update Balance

Withdraw Funds ----<<include>> → Update Balance

(iii) Use Cases: Open Account, Deposit Funds, Calculate Interest, Withdraw Funds, Update Balance

(iv) Main Flow:

A customer can "open account" through Bank Employee.

A customer can "deposit funds" for which interest may or may not be calculated, but the balance must be updated.

A customer can "withdraw funds" for which balance must be updated.

SECTION–C

4. (a) What is Django? **(1)**

Ans. Django is a free and open source web application framework, written in Python.

(b) Is Django a high-level web framework or low-level framework? **(1)**

Ans. Django is a high-level Python web framework which was designed for rapid development and clean, realistic design.

(c) Write command to install mysql-connector. **(1)**

Ans. pip install mysql-connector

(d) The command to run Django server is: **(1)**

(i) python manage.py run

(ii) python manage.py runserver

(iii) python manage.py server

(iv) python manage.py execute

Ans. (ii)

(e) Name the method which is used for displaying only one resultset. **(1)**

(i) fetchmany (ii) fetchone

(iii) fetchall (iv) fetchno

Ans. (ii)

(f) Create a table "GARMENT" with the following specifications: (3)

Name of the column	Data type	Size	Constraint
GCODE	NUMBER	3	Primary key
GNAME	VARCHAR	25	NOT NULL
SIZE	VARCHAR	3	DEFAULT= "L"
COLOR	VARCHAR	12	
PRICE	NUMBER	(7, 2)	

Ans. Create table Garment(Gcode number(3) primary key, Gname Varchar(25) not null, size varchar(3) default= 'L', color varchar(12), price number(7,2))

(g) Write SQL queries for the above table "GARMENT": (3)

(i) Add a column "Dt_of_Sale" of "date" type

(ii) To display name and price of all garments whose price is greater than 1000 and name starts with "T".

(iii) To store current date in "Dt_of_Sale" column whose values in "NULL"

Ans. (i) Alter Table Garment Add Dt_of_Sale Date;

(ii) Select Gname, Price from Garment where price>1000 and Gname like "T%";

(iii) Update Garment set Dt_of_Sale=curdate() where Dt_of_sale is NULL;

(h) Write SQL query for (i) to (iii) on the table "STUDENT" given below: (4)

Stud_No	SName	Stipend	Stream	Class
1001	Karan	400.00	Medical	12B
1002	Divakar	450.00	Commerce	11C
1003	Divya	300.00	Commerce	12C
1004	Arun	350.00	Humanities	12C
1005	Sabina	500.00	Non-medical	11A
1006	John	400.00	Medical	12B
1007	Robert	250.00	Humanities	11A
1008	Rubina	450.00	Non-medical	12A
1009	Vikas	500.00	Non-medical	12A
1010	Mohan	300.00	Commerce	12C

(i) To display the lowest, highest and average stipend of students.

(ii) To display stream and no. of students in each stream.

(iii) To display the details in following order of stream: 'medical', 'commerce', 'humanities', 'non-medical'.

(iv) What will be the output of the following statement:

Select sum(stipend), count(stream) from student group by stream having stream= 'non-medical';

Ans. (i) Select min(stipend),max(stipend),avg(stipend) from Student;

(ii) Select stream, count(stream) from Student group by stream;

(iii) Select * from Student order by Field(stream, 'medical', 'commerce', 'humanities', 'non-medical');

(iv)

Sum(stipend)	count(stream)
1450.00	3

SECTION–D

5. (a) Using someone else's twitter handle to post something will be termed as: (1)

(i) Fraud

(ii) Identity theft

(iii) Online stealing

(iv) Violation

Ans. (ii)

(b) Name any two popular anti-viruses.

Ans. Quick Heal, Avira, McAfee, Avg (any two) (1)

(c) IT ACT, 2000 Section 72 deals with?

Ans. Privacy and confidentiality (1)

(d) Explain Denial-of-Service attack.

Ans. In computing, a **denial-of-service attack (DoS attack)** is a cyber attack in which the perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting services of a host connected to the internet. (2)

(e) Write the advantages of using a licensed software.

Ans. It is always good to use a licensed software as licensing provides advantages while working with these software. (2)

(i) By using a licensed software, we are able to contribute to the further development of the program we are using. It also helps the economy generate not just computer-related jobs, but also opportunities for other people/businesses whose work is dependent on "written" technology.

(ii) Apart from the legal obligation to use licensed software, it comes with the outright support not found in "pirated" software.

(f) What do you mean by e-waste? What is its composition? Write any two benefits of e-waste recycling. (3)

Ans. E-waste means discarded electrical or electronic devices. It contains metallic and non-metallic elements, alloys and compounds like copper, aluminium, gold, etc. Benefits of e-waste recycling are:

(i) It allows for recovery of valuable precious metals.

(ii) It protects public health and water quality.

(iii) It creates jobs.

(iv) It reduces toxic waste

(v) It saves landfill space.

Or

Define Crowdsourcing and write its any four benefits.

Ans. Crowdsourcing refers to the practice of obtaining contribution from the crowd in the form of needed services, ideas or content, mainly from the online community.

Benefits:

(i) Accelerated collaborative innovation

(ii) Sharing of ideas

(iii) Co-creation

(iv) Cost-saving

(v) Increased efficiency

(vi) Engagement of stakeholders

APPENDIX B

MODEL TEST PAPER (UNSOLVED)

CLASS XII
INFORMATICS PRACTICES (065)

TIME: 3 Hrs. **M.M: 70**

SECTION–A

1. Answer the following questions:

(a) Create an ndarray with values ranging from 10 to 49, each spaced with a difference of 3. (1)

(b) What are NumPy arrays? (1)

(c) Which function do we use to create an array with zero values? (1)

(d) Predict the output of the following code fragment: (2)

```
import numpy as np
x = [15,22,3,26,8,34,23,51]
x1,x2,x3 = np.split(x, [3,5])
print(x1,x2,x3)
```

(e) Write a program to find and plot the linear regression line for the two sets of data: [10, 20, 30, 40, 50] [400, 800, 1100, 1700, 2100] (2)

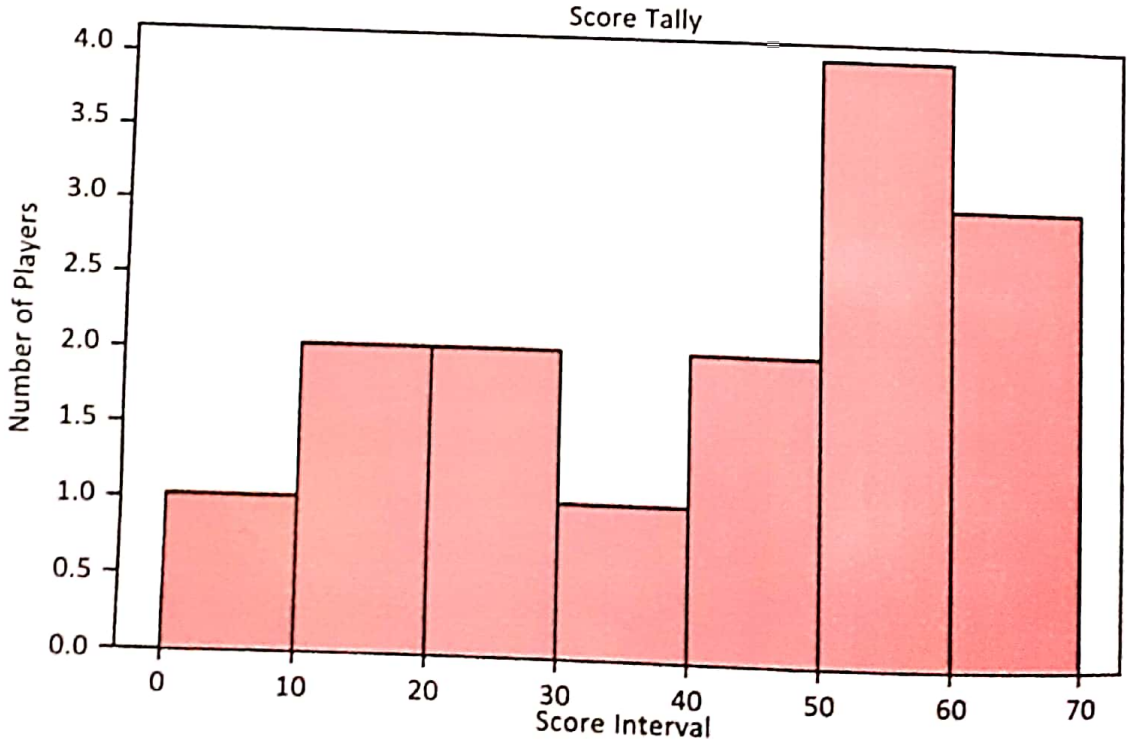
(f) What will be the output of the following code? (2)

```
import numpy as np
arr1=np.arange(18).reshape(3,6)
print(arr1)
```

Or

Write the code to create a 5 X 3 two-dimensional array from the range of integers 1 .....11.

(g) Write the Python code to plot the given histogram with the following dataset: (3)



Score=[8,10,15,25,28,35,47,49,50,63,67,53,57,58,69]

2. (a) What is the use of pipe() function in Python Pandas? (1)

(b) Name any two methods provided by Pandas for reindexing and relabelling. (1)

(c) What is a dataframe? (1)

(d) Write down the syntax of iterating over a dataframe. (1)

(e) Consider the following dataframe: (2)

DataFrame:-df

	YEAR	MONTH	GENDER	BIRTHS
0	2013	1	F	4060
1	2012	1	M	3000
2	2013	2	F	5000
3	2014	2	M	4400
4	2014	1	F	3500
5	2015	3	M	4000
6	2012	2	F	3300

Write the statement and output for the following (refer to the above DataFrame (df)):

(i) Display the total births for each year (using pivot table).

(ii) Display the total number of males and females for each year (using pivot table).

(f) What is the main problem in pivot() with the earlier versions of Python? How can it be resolved? (2)

(g) Write the code to create an ndarray having 6 zeroes in it. Write the statement to change the 3rd and 5th element of this ndarray to 15 and 25 respectively. (3)

(h) Define the following: (3)

(i) Pivoting (ii) Pivot Table

(iii) Data Aggregation

(i) What will be the output of the following code? (4)

```
import pandas as pd
import numpy as np
#Create a Dictionary of series
d = {'Name':pd.Series(['Sachin','Dhoni','Virat','Rohit','Shikhar']),
     'Age':pd.Series([26,25,25,24,31]),
     'Score':pd.Series([87,67,89,55,47])}
#Create a DataFrame
df = pd.DataFrame(d)
print(df)
print(df.count())
print("count age",df[['Age']].count())
print("sum of score",df[['Score']].sum())
print("maximum score",df[['Score']].max())
print("mean age",df[['Age']].mean())
print("mode of age",df[['Age']].mode())
print("median of score",df[['Score']].median())
```

SECTION–B

3. (a) Define the term software process. (1)

(b) Which one is not a software process model? (1)

(i) Waterfall Model (ii) Relational Model

(iii) Evolutionary Model (iv) Component-based Model

(c) What is cohesion? (1)

(d) Explain the terms: (2)

(i) Sprints (ii) Scrum

(e) Define waterfall model and compare it with evolutionary model. (3)

(f) Explain Agile Manifesto with a proper diagram. (3)

(g) Draw a business use-case diagram for Restaurant Management System. (4)

Or

Draw a use-case diagram and identify the actors for the situations (i) to (ii) as directed:

(i) A repair can be made by a master, a trainee or any other repair shop employee.

(ii) Consider an ATM system. Identify at least three different actors that interact with this system.

SECTION–C

4. (a) Is Django a high-level web framework or a low-level framework? (1)

(b) Write the command to install MySQL connector. (1)

(c) Write the difference between % and _ wild card characters with reference to like clause in SQL. (1)

(d) Which of the following is a Django configuration file? (1)

(i) urls.py (ii) settings.py

(iii) wsgi.py (iv) init.py

(e) Write an SQL statement for creating a table LIBRARY having the following attributes: (1)

(i) BOOK_ID – an integer – Primary key (ii) BOOK_TITLE – a string

(iii) QTY – an integer (iv) PRICE – a real number

Or

Shivani, a salesgirl in an outlet, created a table CUSTOMERS. PhoneNumber is one of the columns of this table, which is character type. To find the details of the customer whose PhoneNumber is 356789, she wrote the following MySQL query, which did not give the desired result:

```
Select * from CUSTOMERS
WHERE PhoneNumber = 356789;
```

Help her run the query.

(f) Give the Python statement to (3)

(i) establish a connection with the database.

(ii) create a cursor for the connection.

(iii) execute create table SQL query.

(g) Give any three differences between GET and POST submission methods of HTML form. (3)

(h) Write a Python code to create a table in MySQL and insert 5 rows in it. (4)

Specifications are as follows:

Database name: School **Table name: XIIG_IP**

Structure of the table:

Column Name	Data type
R.No (Primary Key)	Integer
Name	Varchar(25)
Class	Integer
Section	Char(1)
Percentage	Integer

SECTION–D

5. (a) Name any one open-source operating system and open-source browser. (1)

(b) Define identity theft. (1)

(c) Excessive social networking is an example of: (1)

(i) Echo chamber (ii) Crowd sourcing

(iii) Internet addiction (iv) Net neutrality

(d) Differentiate between freeware and shareware. (2)

(e) Why was WikiLeaks in the news? (2)

(f) Explain any three major issues with the internet. (3)

Or

Write notes on the following:

(i) Crowd sourcing (ii) Smart mobs

(iii) Online campaigns