SUPPLEMENT

2020 EDITION

CBSE Class XII

INFORMATICS PRACTICES

With

Python

As per 2020-21 Syllabus

Web Support
at
sultan-chand.com/ws/ipp12

- Presentation on Python Pandas
- Sample Papers
- Model Test Papers
- Program Codes
- Projects
- Practical File
- Django Installation Guide
- Viva Voce
- Supplement

PREETI ARORA

You can download this Supplement
Free of Cost

SCS
sultan chand

SUPPLEMENT

(as per 2020–21 Curriculum)

Dear Reader

The aim of this Supplement to **Informatics Practices with Python** for Class XII is to apprise you of the changes effected in the CBSE syllabus for 2020-21 and provide them in textbook format.

In fact, the syllabus for 2020-21 has witnessed large-scale changes with five chapters (**NumPy**, **Introduction to Software Engineering, Agile Methods and Practical Aspects of Software Engineering, Web Development with Django** and **Interface Python with SQL**) done away with, four chapters completely refurbished and one new chapter added, all of which have been duly taken care of in this Supplement. Besides, chapter-wise Case-based/Source-based Integrated Questions, which have been introduced for the first time in the Class XII curriculum, too have been given at the end of the chapters.

To ensure that your book is complete in all respects once this Supplement is in your hands, an elaborate **Comparative Analysis of Class XII Old and New CBSE Curriculum** has been given right in the beginning to give you a clear-cut idea about the chapters modified or fresh topics included in the 2020-21 syllabus vis-à-vis the old one. The Supplement includes a new chapter **Computer Networks** besides additions/changes as per the syllabus, which have been added to the respective chapters along with requisite additions in exercises.

To keep up with our commitment to contribute to the growth of education and not put any undue financial burden on the reader, the printed copy of this Supplement has been nominally priced.

Happy learning!

COMPARATIVE ANALYSIS OF CLASS XII OLD AND NEW CBSE CURRICULUM

INFORMATICS PRACTICES - 065

UNIT	OLD CURRICULUM (2019-20)	NEW CURRICULUM (2020-21)	CHANGES
I: DATA HANDLING USING PANDAS AND DATA VISUALIZATION	30 Marks	30 Marks	No change in the Weightage
	DATA HANDLING (DH-2)	DATA HANDLING USING PANDAS AND DATA VISUALIZATION	Entire unit has been renamed
	PYTHON PANDAS Advanced operations on Data Frames: pivoting, sorting, and aggregation • Descriptive statistics: min, max, mode, mean, count, sum, median, quartile, var • Create a histogram, and quantiles. • Function application: pipe, apply, aggregation (group by), transform, and apply map. • Reindexing, and altering labels.	DATA HANDLING USING PANDAS –I Introduction to Python libraries- Pandas, Matplotlib. Data structures in Pandas - Series and Data Frames. Series: Creation of Series from – ndarray, dictionary, scalar value; mathematical operations; Head and Tail functions; Selection, Indexing and Slicing. Data Frames: creation - from dictionary of Series, list of dictionaries, Text/CSV files; display; iteration	DH USING PANDAS – I All the topics from class XI previous curriculum have been added to this unit
	NUMPY	DATA HANDLING USING PANDAS –II	DH USING PANDAS –II
	1D array, 2D array, Arrays: slices, joins, and subsets; Arithmetic operations on 2D arrays; Covariance, correlation and linear regression	Operations on rows and columns: add, select, delete, rename; Head and Tail functions; Indexing using Labels, Boolean Indexing; Joining, Merging and Concatenation. Importing/Exporting Data between CSV files and Data Frames.	New topic for CSV files has been added NUMPY has been completely removed (PART OF CLASS XI SYLLABUS IN 2020- 21)
		DATA HANDLING USING PANDAS –II Descriptive Statistics: max, min, count, sum, mean, median, mode, quartile, Standard deviation, variance. DataFrame operations: Aggregation, group by, Sorting, Deleting and Renaming Index, Pivoting. Handling missing values – dropping and filling. Importing/Exporting Data between MySQL database and Pandas.	DH USING PANDAS –II No changes NEWLY ADDED TOPIC: Importing/Exporting Data between MySQL database and Pandas

	PLOTTING WITH PYPLOT Plot bar graphs, histograms, frequency polygons, box plots, and scatter plots.	**DATA VISUALIZATION** Purpose of plotting; drawing and saving following types of plots using Matplotlib – **line plot**, bar graph, histogram, pie chart, frequency polygon, box plot and scatter plot. Customizing plots: color, style (dashed, dotted), width; adding label, title, and legend in plots.	**Newly added topic- LINE PLOTS**
II: Software Engg and Agile Methods	**BASIC SOFTWARE ENGINEERING**	**15 marks**	**The entire topic has been removed**
III: Database Query using SQL	**DATA MANAGEMENT (DM-2) – renamed** **MARKS-15**	**DATABASE QUERY USING SQL** **MARKS-25**	**Increase in the weightage by 10 marks BUT UNIT 3 (2019-20) Renamed as UNIT 2**
	Write a minimal Django based web application that parses a GET and POST request, and writes the fields to a file – flat file and CSV file. Interface Python with an SQL database SQL commands: aggregation functions, having, and group by, order by.	**Math functions: POWER (), ROUND (), MOD ().** **Text functions: UCASE ()/UPPER (), LCASE ()/ LOWER (), MID ()/SUBSTRING ()/SUBSTR (), LENGTH (), LEFT (), RIGHT (), INSTR (), LTRIM (), RTRIM (), TRIM ().** **Date Functions: NOW (), DATE (), MONTH (), MONTHNAME (), YEAR (), DAY (), DAYNAME ().** Aggregate Functions: MAX (), MIN (), AVG (), SUM (), COUNT (); using COUNT (*). Querying and manipulating data using Group by, Having, Order by. **Operations on Relations - Union, Intersection, Minus, Cartesian Product, JOIN**	**DJANGO REMOVED** **INTERFACE PYTHON WITH AN SQL DATABASE REMOVED.** **NEWLY ADDED:** **Math functions: POWER (), ROUND (), MOD ().** **Text functions: UCASE ()/ UPPER (), LCASE ()/LOWER (), MID ()/SUBSTRING ()/ SUBSTR (), LENGTH (), LEFT (), RIGHT (), INSTR (), LTRIM (), RTRIM (), TRIM ().** **Date Functions: NOW (), DATE (), MONTH (), MONTHNAME (), YEAR (), DAY (), DAYNAME ().** **Operations on Relations: Union, Intersection, Minus, Cartesian Product, JOIN**

III: Computer Networks	**NO SUCH TOPIC**	**Introduction to computer networks**	**New Unit added carrying the weightage of 7 marks**
		Introduction to networks, Types of network: LAN, MAN, WAN. **Network Devices: modem, hub, switch, repeater, router, gateway** **Network Topologies: Star, Bus, Tree, Mesh. Introduction to Internet, URL, WWW and its applications- Web, email, Chat, VoIP. Website: Introduction, difference between a website and webpage, static vs dynamic web page, web server and hosting of a website.** **Web Browsers: Introduction, commonly used browsers, browser settings, add-ons and plug-ins, cookies.**	
IV: Societal Impacts	**SOCIETY, LAW AND ETHICS MARKS - 10**	**Renamed to SOCIETAL IMPACTS MARKS- 8**	**Weightage reduced by 2 marks**
	Intellectual property rights, plagiarism, digital rights management, and licensing (Creative Commons, GPL and Apache), open source, open data, privacy. Privacy laws, fraud; cybercrime- phishing, illegal downloads, child pornography, scams; cyber forensics, IT Act, 2000. Technology and society: understanding of societal issues and cultural changes induced by technology. E-waste management: proper disposal of used electronic gadgets. Identity theft, unique ids, and biometrics. Gender and disability issues while teaching and using computers. Role of new media in society: online campaigns, crowd sourcing, smart mobs Issues with the internet: internet as an echo chamber, net neutrality, internet addiction; Case studies - Arab Spring, Wiki Leaks, Bit coin.	**Digital footprint, net and communication etiquettes, data protection, intellectual property rights (IPR), plagiarism, licensing and copyright, free and open source software (FOSS), cybercrime and cyber laws, hacking, phishing, cyber bullying, overview of Indian IT Act.** **E-waste: hazards and management.** **Awareness about health concerns related to the usage of technology.**	• **Identity theft, unique ids, and biometrics.** • **Gender and disability issues while teaching and using computers.** • **Role of new media in society: online campaigns, crowd sourcing, smart mobs** • **Issues with the internet: internet as an echo chamber, net neutrality, internet addiction** • **Case studies - Arab Spring, WikiLeaks, Bit coin** • **Technology and society: understanding of societal issues and cultural changes induced by technology**
RED - Removed Topics	**GREEN – Detailed added topics**	**PURPLE – No Change**	**BLUE - Added Topics**

CONTENTS

Data Visualization using Pyplot (Additions)

2.15 PIE PLOT/CHART

A pie plot is a circular plot divided into slices to show numerical proportion. Pie plots are widely used in the business world. However, many experts recommend avoiding them. The main reason is that it is difficult to compare the different sections of a given pie chart. Also, it is difficult to compare data across multiple pie charts. In many cases, they can be **replaced** by a **bar chart**.

To make a pie chart with Matplotlib, we can use the **plt.pie()** function.

Pie charts show a slice of time or any object we are referring to. A pie graph/pie chart is a specialized graph used in statistics. The independent variable is plotted around a circle.

Pie charts show proportions and percentages between categories by dividing a circle into proportional segments/parts. Each arc length represents a proportion of each category, while the full circle represents the total sum of all the data, equal to 100%.

CTM: Pie charts are circular representations divided into sectors (also called wedges). The arc length of each sector is proportional to the quantity we are describing. It is an effective way of representing information when we are interested mainly in comparing a wedge with the whole pie, instead of comparing wedges with each other.

Practical Implementation–26

To plot a pie chart for the popular languages among the students.

prog_pie2.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_pie...

File Edit Format Run Options Window Help

```
import matplotlib.pyplot as plt

# Data to plot
labels = 'Python', 'C++', 'Ruby', 'Java'
sizes = [215, 130, 245, 210]
colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue']
explode = (0.1, 0, 0, 0)  # explode 1st slice

# Plot
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        shadow=True, startangle=140)

plt.axis('equal')
plt.show()
```

Ln: 17 Col: 0

In the context of the above program for plotting a pie chart, we have used certain keyword arguments for specific purposes and to customize pie charts as defined below:

- **explode:** If specified, it is a list/array of the same length as that of sizes. Each value specifies the radius fraction with which to offset the wedge from the centre of the pie.
- **colors:** This is a list of Matplotlib colours, cyclically used to colour the wedges.
- **labels, labeldistance:** This is a list of labels, one as per each value allocated to size attribute. labeldistance is the radial distance at which the labels are drawn.
- **shadow:** This draws a shadow for wedges or pie.
- **startangle:** Startangle describes the angle at which each slice of the pie is aligned, which is 140° in the above program. Starting at an angle of 140°, the first slice is aligned, followed by the rest of the slices being aligned at the same angle but in an anti-clockwise direction.

Practical Implementation–27

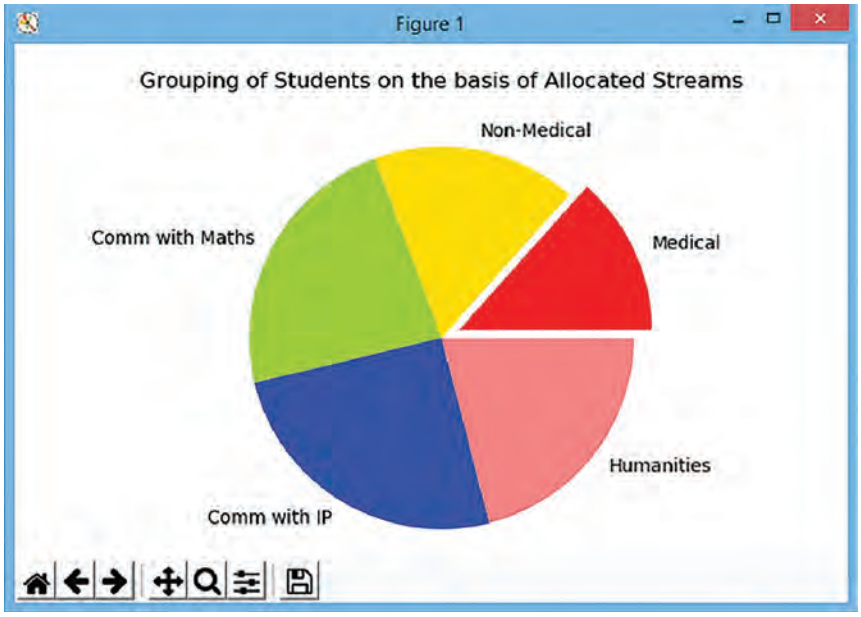
To plot a pie plot for the total number of students enrolled in a particular stream in a school.

prog_pie1.py - C:\Users\preeti\AppData\Local\Programs\Python\Python37-32\prog_pie1.py (3.7.0)

File Edit Format Run Options Window Help

```
import matplotlib.pyplot as plt
# Plot data
stream =['Medical','Non-Medical','Comm with Maths','Comm with IP','Humanities']
no_students = [32,41,55,60,50]
colors = ['red', 'gold', 'yellowgreen', 'blue', 'lightcoral']
# explode 1st slice
explode = (0.1, 0, 0, 0,0)
# Plot
plt.pie(no_students, explode=explode, labels=stream, colors=colors)
plt.title("Grouping of Students on the basis of Allocated Streams" )
plt.show()
```

Ln: 15 Col: 0

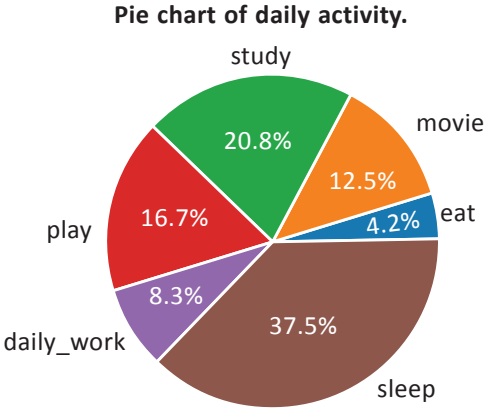


MEMORY BYTES

- Pie charts are circular representations divided into sectors (also called wedges). The arc length of each sector is proportional to the quantity we are describing. It is an effective way of representing information when we are interested mainly in comparing a wedge with the whole pie instead of comparing wedges with each other.

SOLVED QUESTIONS

39. Consider two lists inputted by a user—one list of daily activities and another list that shows the time taken daily to do each activity. Draw a pie chart to depict the time taken for daily activities as shown:



Ans.
```
import matplotlib.pyplot as plt
A=['eat', 'movie', 'study', 'play', 'daily_work', 'sleep']
T=[1,3,5,4,2,9]
plt.pie(T, labels=A, autopct='%1.1f%%')
plt.title('Pie chart of daily activity.')
plt.show()
```

40. Write a Python program to create a pie chart of gold medal achievements of five most successful countries in the 2016 Summer Olympics. Fetch the data from medal.csv

Sample data: (medal.csv)

country	gold_medal
United States	46
Great Britain	27
China	26
Russia	19
Germany	17

Ans.
```
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv('medal.csv')
country_data = df["country"]
medal_data = df["gold_medal"]
colors = ["brown", "red", "green", "orange", "blue"]
explode = (0.1, 0, 0, 0, 0)
plt.pie(medal_data, labels=country_data, explode=explode,
        colors=colors, autopct='%1.1f%%', shadow=True, startangle=140)
plt.title("Gold medal achievements of five most
          successful\n"+"countries in 2016 Summer Olympics")
plt.show()
```

UNSOLVED QUESTIONS

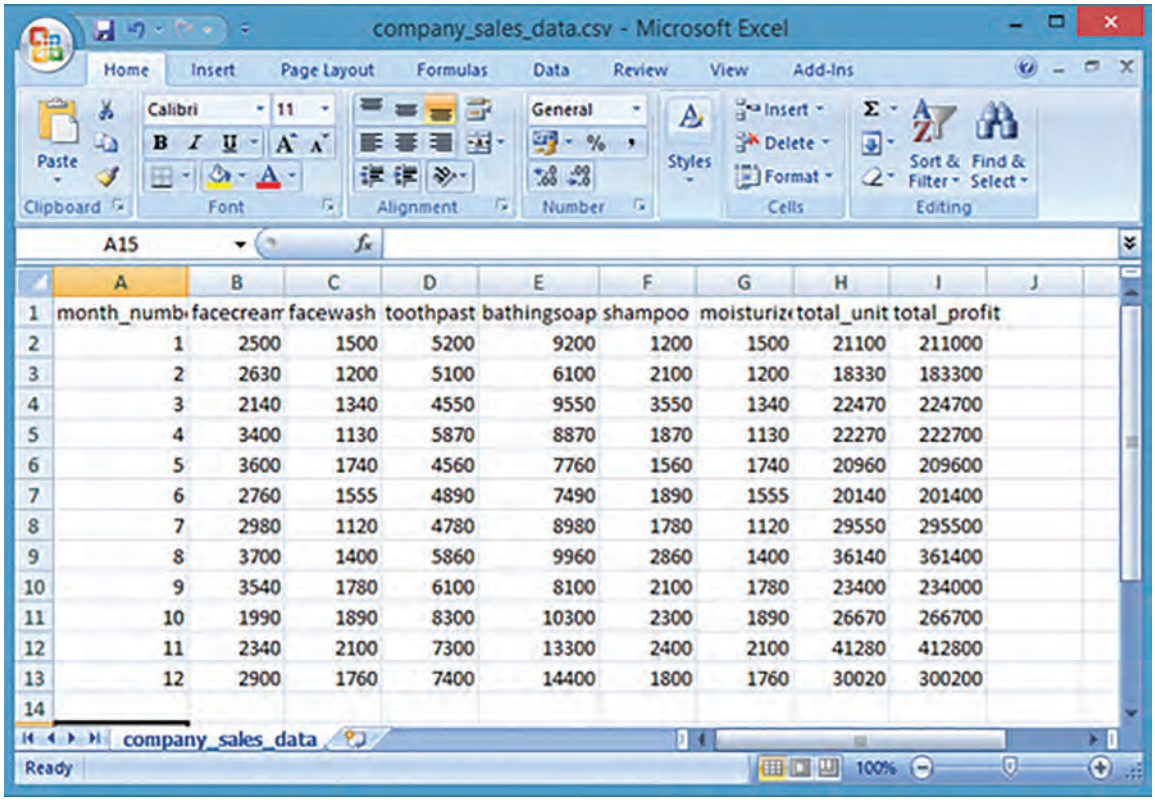
37. How is a pie chart different from a bar graph?

38. Write a Python program to create a pie chart with the title of the Stream and percentage of Students. Make multiple wedges of the pie.
Sample data:
Stream : Science, Commerce, Humanities, Vocational, FMM
Strengths: 29%, 30%, 21%, 13%, 7%

39. Plot a pie chart of a class test of 40 students based on random sets of marks obtained by the students (MM=100).

CASE-BASED/SOURCE-BASED INTEGRATED QUESTIONS

1. Hindustan Departmental Stores sell items of daily use such as shampoo, soap and much more.
They record the entire sale and purchase of goods month-wise so as to get a proper analysis of profit or loss in their business transactions.
Following is the csv file containing the "Company Sales Data".


company_sales_data.csv - Microsoft Excel

	A	B	C	D	E	F	G	H	I
1	month_numb	facecream	facewash	toothpast	bathingsoap	shampoo	moisturize	total_unit	total_profit
2	1	2500	1500	5200	9200	1200	1500	21100	211000
3	2	2630	1200	5100	6100	2100	1200	18330	183300
4	3	2140	1340	4550	9550	3550	1340	22470	224700
5	4	3400	1130	5870	8870	1870	1130	22270	222700
6	5	3600	1740	4560	7760	1560	1740	20960	209600
7	6	2760	1555	4890	7490	1890	1555	20140	201400
8	7	2980	1120	4780	8980	1780	1120	29550	295500
9	8	3700	1400	5860	9960	2860	1400	36140	361400
10	9	3540	1780	6100	8100	2100	1780	23400	234000
11	10	1990	1890	8300	10300	2300	1890	26670	266700
12	11	2340	2100	7300	13300	2400	2100	41280	412800
13	12	2900	1760	7400	14400	1800	1760	30020	300200
14									

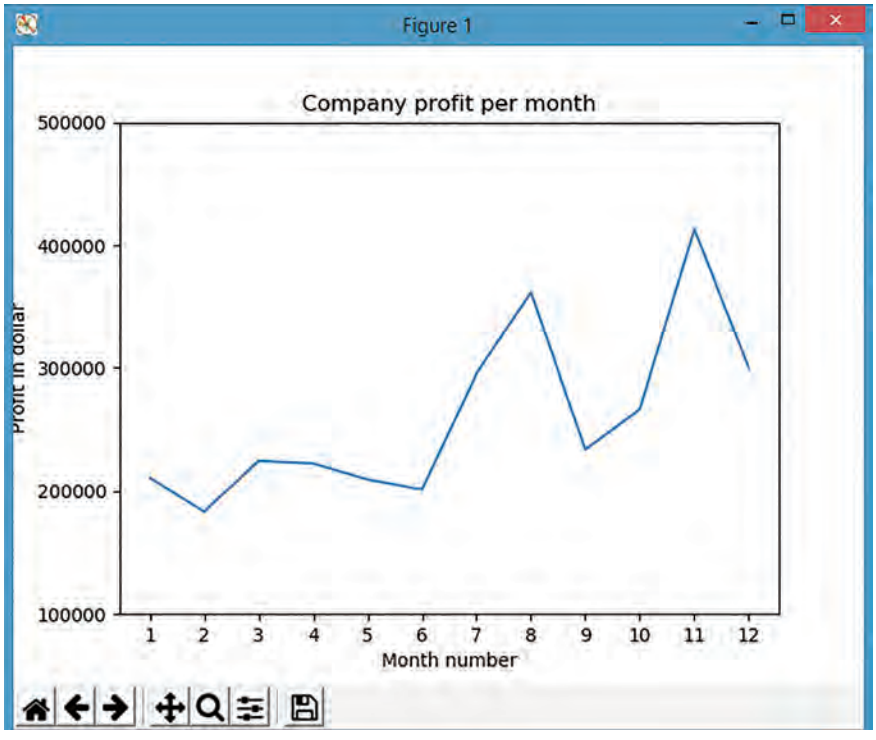
Read the total profit of all months and show it using a line plot. Total profit data has been provided for each month. Generated line plot must include the following properties:

- X label name = Month Number
- Y label name = Total profit

Ans.

```
prog_company_sales.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_company_sales.py (3.7....
File  Edit  Format  Run  Options  Window  Help
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("D:/company_sales_data.csv")
profitList = df ['total_profit'].tolist() #converting dataframe into a list
monthList  = df ['month_number'].tolist()
plt.plot(monthList, profitList, label = 'Month-wise Profit data of last year')
plt.xlabel('Month number')
plt.ylabel('Profit in dollar')
plt.xticks(monthList)
plt.title('Company profit per month')
plt.yticks([100000, 200000, 300000, 400000, 500000])
plt.show()
                                                                    Ln: 17  Col: 0
```



2. NXP Labs is a leading supplier of embedded controllers with a strong legacy in both the industrial and consumer market. It manufactures microcontrollers on a large scale and exports largely to European nations. The format of their operations is described as:

The cost breakdown for a manufactured item, like a microcontroller, can be divided into four cost categories: engineering (including design), manufacturing (including raw materials), sales (including marketing) and profit.

These cost categories applied to a $9.00 microcontroller:

- Engineering $1.35
- Manufacturing $3.60
- Sales $2.25
- Profit $1.80

Develop a pie chart using matplotlib that describes the cost breakdown of building a microcontroller.

Ans.

prog_micro_controller.py - C:/Users/preeti/AppData/Local/Programs/Python/Pytho...

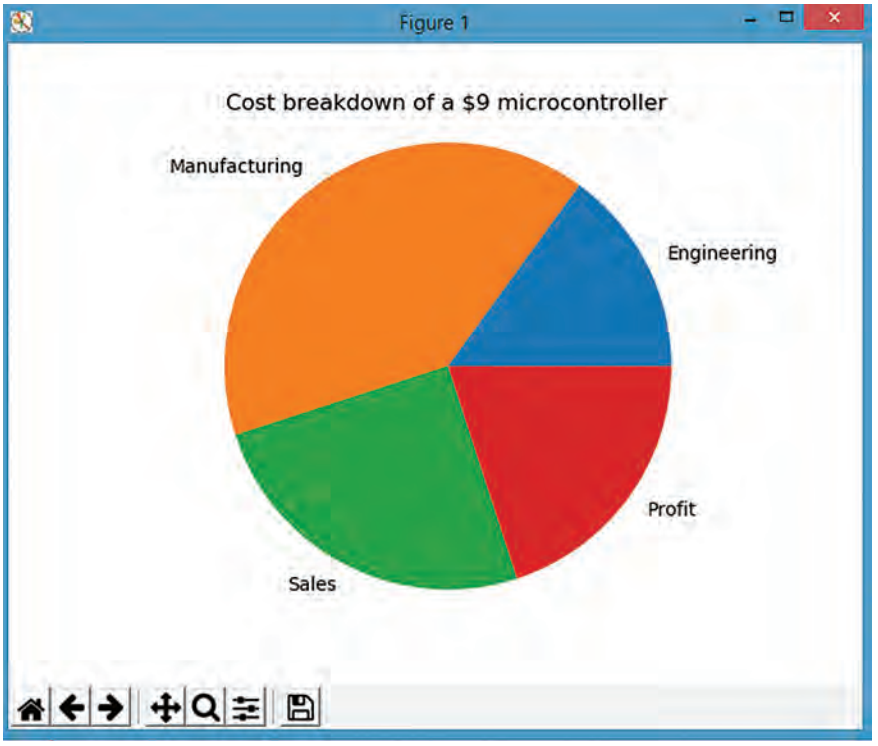
File Edit Format Run Options Window Help

```
#pie chart that describes the cost breakdown of
#building a microcontroller
import matplotlib.pyplot as plt

sizes = [1.35, 3.60, 2.25, 1.80]
labels = 'Engineering', 'Manufacturing', 'Sales', 'Profit'

plt.pie(sizes,
        labels = labels)
plt.title('Cost breakdown of a $9 microcontroller')
plt.axis('equal')
plt.show()
```

Ln: 16 Col: 0



Python Pandas (Additions)
(Data Handling using Pandas)

In Practical Implementation-24, we have learnt to rename the only given column in dataframe. But what happens when we have to rename specific column of a dataframe?

In Pandas, this is done by using the function rename() as exhibited in the Practical Implementation below.

Practical Implementation–60

Create a dataframe student with columns as name of the student, student's marks in subjects IP and BST. Also rename the column name as Nm and IP as Informatics Practices.

```
prog_renme_df.py - C:\Users\preeti\AppData\Local\Programs\Python\Python37-32\prog_renme_df.py (3.7.0)
File  Edit  Format  Run  Options  Window  Help
#Renaming columns in a dataframe
import pandas as pd
s=[['Rinku',79,72],['Ritu',75,73],['Ajay',80,76]]
print('Series Generated As:')
print(s)
df=pd.DataFrame(s,columns=['Name','IP','BST'])
print(df)
df.rename(columns={'Name':'Nm','IP':'Informatics Practices'},inplace=True)
print(df)
                                                                    Ln: 14  Col: 0
```

```
>>>
 RESTART: C:\Users\preeti\AppData\Local\Programs\Python\
f.py
Series Generated As:
[['Rinku', 79, 72], ['Ritu', 75, 73], ['Ajay', 80, 76]]
    Name  IP  BST
0  Rinku  79   72
1   Ritu  75   73
2   Ajay  80   76
      Nm  Informatics Practices  BST
0  Rinku                    79   72
1   Ritu                    75   73
2   Ajay                    80   76
```

As is evident from the above code, rename() method has been used to rename a column in Pandas dataframe.

Another way to change column names in Pandas is to use **rename** function. Using **rename** to change column names is a much better way. One can change names of specific columns easily and not all column names need to be changed.

One of the biggest advantages of using **rename** function is that we can use **rename** to change as many column names as we want. Its syntax is:

```
df.rename(columns=d, inplace=True)
```

Where d is a dictionary and the keys are the columns you want to change. The values are the new names for these columns. Also, **inplace=True** is given as the attribute of rename() to change column names in place. Thus, as shown in the output window, the column names of student dataframe, Name and IP have been renamed as Nm and Informatics Practices respectively.

3.23 ITERATIONS IN DATAFRAME—ITERROWS

Sometimes we need to perform iteration on a complete dataframe, *i.e.*, accessing and retrieving each record one by one in a dataframe. In such cases, it is difficult to write a code to access the values separately. Therefore, it is necessary to perform iteration on dataframe which can be done using any of the two methods:

- <DFObject>.iterrows()—It represents dataframe row-wise, record by record.
- <DFObject>.iteritems()—It represents dataframe column-wise.

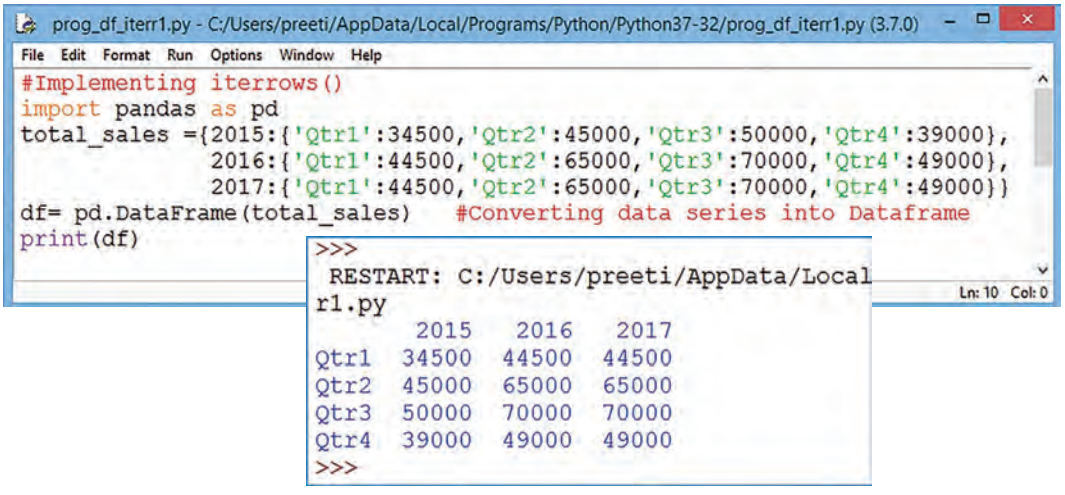
Let us see how the above two methods work. Consider three series for yearly sales of ABC Ltd.

2015—Qtr1: 34500, Qtr2: 45000, Qtr3: 50000, Qtr4: 39000

2016—Qtr1: 44500, Qtr2: 65000, Qtr3: 70000, Qtr4: 49000

2017—Qtr1: 54500, Qtr2: 42000, Qtr3: 40000, Qtr4: 89000

The first step is to represent these series into a dataframe and then to perform iteration (repetition) for accessing and displaying each record one by one.



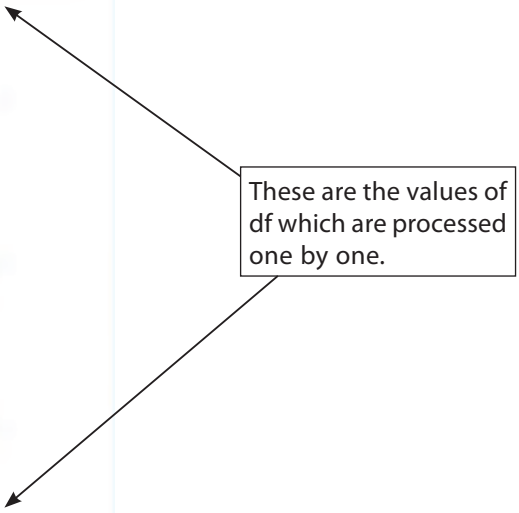
Using iterrows():

The first step has been completed by creating a dataframe from the quarterly sales series. The next step is to display the record in the created dataframe one by one by adding the following code in the previous code:

```
for (row,rowSeries) in df.iterrows():
    print("RowIndex :",row)
    print("Containing :")
    print(rowSeries)
```

This code on execution shall display the records of the dataframe one by one.

```
r1.py
RowIndex : Qtr1
Containing :
2015    34500
2016    44500
2017    44500
Name: Qtr1, dtype: int64
RowIndex : Qtr2
Containing :
2015    45000
2016    65000
2017    65000
Name: Qtr2, dtype: int64
RowIndex : Qtr3
Containing :
2015    50000
2016    70000
2017    70000
Name: Qtr3, dtype: int64
RowIndex : Qtr4
Containing :
2015    39000
2016    49000
2017    49000
Name: Qtr4, dtype: int64
>>>
```



Using iteritems():

This method shall display the data from the dataframe column-wise. After the creation of the dataframe, which we have done in the above example, write the code for displaying the series column-wise as shown below:

```
for (col,colSeries) in df.iteritems():  #displaying column-wise data
    print("Column Index :",col)
    print("Containing :")
    print(colSeries)
```

```
>>>
 RESTART: C:/Users/preeti/AppData/Local/
ise.py
Column Index : 2015
Containing :
Qtr1    34500
Qtr2    45000
Qtr3    50000
Qtr4    39000
Name: 2015, dtype: int64
Column Index : 2016
Containing :
Qtr1    44500
Qtr2    65000
Qtr3    70000
Qtr4    49000
Name: 2016, dtype: int64
Column Index : 2017
Containing :
Qtr1    44500
Qtr2    65000
Qtr3    70000
Qtr4    49000
Name: 2017, dtype: int64
>>>
```

Practical Implementation–61

Write a program to iterate over a dataframe containing names and marks, then calculate grades as per marks (as per the following criteria) and add them to the grade column:

Marks >= 90	Grade A+
Marks 70 – 90	Grade A
Marks 60 – 70	Grade B
Marks 50 – 60	Grade C
Marks 40 – 50	Grade D
Marks < 40	Grade F

```
import pandas as pd
import numpy as np
names=pd.Series(['Sanjeev','Rajeev','Sanjay','Abhay'])
marks=pd.Series([76,86,55,54])
stud={'Name':names,'Marks':marks}
df=pd.DataFrame(stud,columns=['Name','Marks'])
df['Grade']=np.NaN        #this will add NaN to all records of dataframe
print("Initial values in DataFrame")
print(df)
for (col,colSeries) in df.iteritems():
    length=len(colSeries)
    if col=='Marks':
        lstMrks=[]
        for row in range(length):
            mrks=colSeries[row]
            if mrks>=90:
                lstMrks.append('A+')
            elif mrks>=70:
                lstMrks.append('A')
            elif mrks>=60:
                lstMrks.append('B')
            elif mrks>=50:
                lstMrks.append('C')
            elif mrks>=40:
                lstMrks.append('D')
            else:
                lstMrks.append('F')
df['Grade']=lstMrks
print("\n\nDataFrame after calculation of Grades")
print(df)
```

```
Initial values in DataFrame
      Name  Marks  Grade
0  Sanjeev     76    NaN
1   Rajeev     86    NaN
2   Sanjay     55    NaN
3    Abhay     54    NaN


DataFrame after calculation of Grades
      Name  Marks Grade
0  Sanjeev     76     A
1   Rajeev     86     A
2   Sanjay     55     C
3    Abhay     54     C
```

head() and tail() function

The **head()** function is used to get the first n rows.

This function returns the first n rows for the object based on position. It is useful for quickly testing if your object has the right type of data in it.

Syntax:

```
DataFrame.head(n)
```

Here, n is the no. of rows to be extracted.

The **tail()** function is used to get the last n rows.

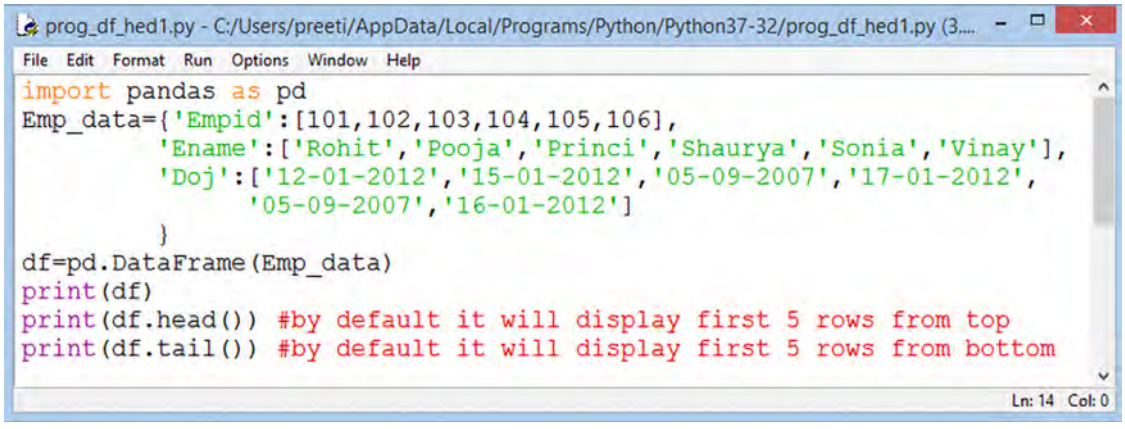
This function returns last n rows from the object based on position. It is useful for quickly verifying data, for example, after sorting or appending rows.

Syntax:

```
DataFrame.tail(n)
```

Practical Implementation–62

Illustrate head() and tail() methods in an Employee dataframe.




```
>>>
 RESTART: C:/Users/preeti/AppData/
.py
   Empid    Ename         Doj
0    101    Rohit  12-01-2012
1    102    Pooja  15-01-2012
2    103   Princi  05-09-2007
3    104  Shaurya  17-01-2012
4    105    Sonia  05-09-2007
5    106    Vinay  16-01-2012
   Empid    Ename         Doj
0    101    Rohit  12-01-2012
1    102    Pooja  15-01-2012
2    103   Princi  05-09-2007
3    104  Shaurya  17-01-2012
4    105    Sonia  05-09-2007
   Empid    Ename         Doj
1    102    Pooja  15-01-2012
2    103   Princi  05-09-2007
3    104  Shaurya  17-01-2012
4    105    Sonia  05-09-2007
5    106    Vinay  16-01-2012
```

head() displays first 5 rows

tail() displays last 5 rows

In the above program, head() and tail() function returns 5 rows from top and bottom respectively; since no argument has been given to these functions, the default function displays 5 rows.

Similarly, to display first 2 rows, we can use head(2) and to returns last 2 rows, we can use tail(2) and to return 3rd to 4th row, we can write df[2:5]

```
prog_df_hed1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_df_hed1.py (3....
File  Edit  Format  Run  Options  Window  Help
import pandas as pd
Emp_data={'Empid':[101,102,103,104,105,106],
          'Ename':['Rohit','Pooja','Princi','Shaurya','Sonia','Vinay'],
          'Doj':['12-01-2012','15-01-2012','05-09-2007','17-01-2012',
                 '05-09-2007','16-01-2012']
          }
df=pd.DataFrame(Emp_data)
print(df)
print(df.head(2))
print(df.tail(2))
print(df[2:5])
                                                                    Ln: 14  Col: 0
```

```
>>>
 RESTART: C:/Users/preeti/AppData/
.py
   Empid    Ename          Doj
0    101    Rohit   12-01-2012
1    102    Pooja   15-01-2012
2    103   Princi   05-09-2007
3    104  Shaurya   17-01-2012
4    105    Sonia   05-09-2007
5    106    Vinay   16-01-2012
   Empid  Ename          Doj
0    101  Rohit   12-01-2012
1    102  Pooja   15-01-2012
   Empid  Ename          Doj
4    105  Sonia   05-09-2007
5    106  Vinay   16-01-2012
   Empid    Ename          Doj
2    103   Princi   05-09-2007
3    104  Shaurya   17-01-2012
4    105    Sonia   05-09-2007
```

head(2) displays first 2 rows

head(2) displays last 2 rows

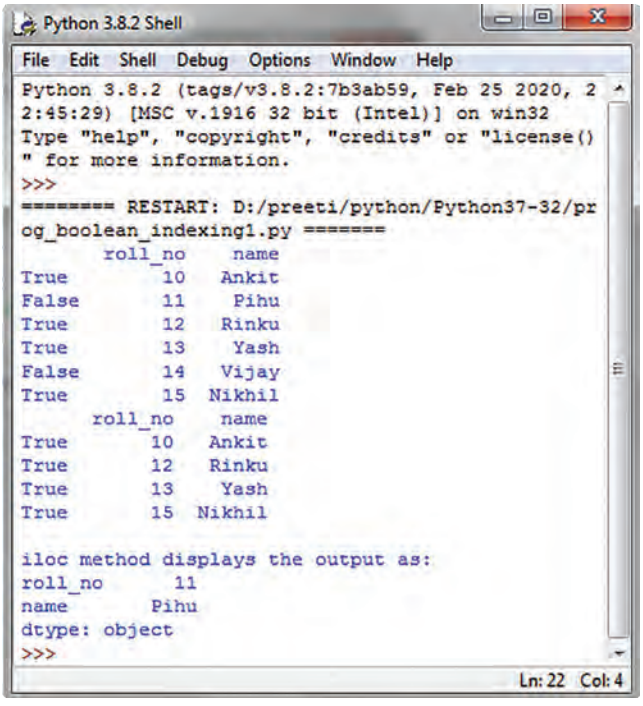
df[2:5] display 2nd to 4th row

3.24 BOOLEAN INDEXING

Boolean indexing is a type of indexing which uses actual values of the data in the dataframe, *i.e.*, using Boolean vector. In order to access a dataframe with a boolean index, we have to create a dataframe in which index of the dataframe contains a boolean value, that is "True" or "False".

Practical Implementation–63

```
prog_boolean_indexing1.py - D:/preeti/python/Python37-32/prog_boolean_in...
File  Edit  Format  Run  Options  Window  Help
#Boolean indexing using data frame
import pandas as pd
d1 = {'roll_no': [10,11,12,13,14,15],
      'name':['Ankit','Pihu','Rinku','Yash','Vijay','Nikhil']}
df1 =pd.DataFrame(d1, index= [True,False,True,True,False,True])
print(df1)
print(df1.loc[True])
print()
print("iloc method displays the output as:")
print(df1.iloc[1])
                                                                    Ln: 14  Col: 0
```



In the above program, a dataframe is created which has a boolean value as its index. This boolean index of the dataframe can be accessed by a user using the functions .loc[] and .iloc[]. The argument passed to these functions is a boolean value, *i.e.*, True or False. print(df1) statement shall print the dataframe with boolean index.

Executing the next statement, df1.loc(True), shall return the dataframe with index value as True. It must be remembered that while using the iloc function, only the integer value is passed as the argument and displays the row with index as 1 since 1 is passed as the argument to iloc.

3.25 CONCATENATION IN DATAFRAME

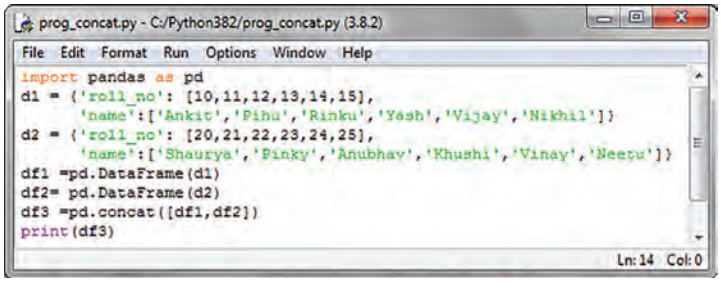
concat() function is used in Pandas for performing concatenation operations along an axis while performing optional set logic (union or intersection) of the indexes (if any) on the other axes. Thus, the concat() function performs the concatenation operation along an axis. The syntax for concat() is:

```
pd.concat(objs,axis=0,join='outer',join_axes=None,ignore_index=False)
```

- **objs –** This is a sequence or mapping of Series, dataframe, or panel objects.
- **axis –** {0, 1, ...}, default 0. This is the axis to concatenate along.
- **join –** {'inner', 'outer'}, default 'outer'. Outer for union and inner for intersection.
- **ignore_index –** boolean, default False. If True, do not use the index values on the concatenation axis. The resulting axis will be labeled 0, ..., n–1.
- **join_axes –** This is the list of Index objects. Specific indexes to use for the other (n–1) axes instead of performing inner/outer set logic.

Practical Implementation–64

Program to concatenate two dataframes.



```
>>>
====================== RESTART: C:/Python382
prog_concat.py ======================
   roll_no     name
0       10    Ankit
1       11     Pihu
2       12    Rinku
3       13     Yash
4       14    Vijay
5       15   Nikhil
0       20  Shaurya
1       21    Pinky
2       22  Anubhav
3       23   Khushi
4       24    Vinay
5       25    Neetu
```

The above program combines/concatenates two dataframes using function concat(), which takes the two dataframes to be combined as the arguments and returns the resultant dataframe, df3, containing the contents of both the dataframes, df1 and df2. The index of the resultant dataframe shall consist of the index values taken from each dataframe respectively.

You can take the final index as per the sequence and not as per the row labels. The above code is to be modified a bit, which is given in the next implementation.

Practical Implementation–65

Program to concatenate two dataframes without taking row labels (Modification of Practical Implementation–64)

prog_concat.py - C:/Python382/prog_concat.py (3.8.2)

File Edit Format Run Options Window Help

```
import pandas as pd
d1 = {'roll_no': [10,11,12,13,14,15],
      'name':['Ankit','Pihu','Rinku','Yash','Vijay','Nikhil']}
d2 = {'roll_no': [20,21,22,23,24,25],
      'name':['Shaurya','Pinky','Anubhav','Khushi','Vinay','Neetu']}
df1 =pd.DataFrame(d1)
df2= pd.DataFrame(d2)
df3 =pd.concat([df1,df2],ignore_index=True)
print(df3)
```

Ln: 12 Col: 0

```
>>>
=== RESTART: C:/Python382/prog_c
   roll_no     name
0       10    Ankit
1       11     Pihu
2       12    Rinku
3       13     Yash
4       14    Vijay
5       15   Nikhil
6       20  Shaurya
7       21    Pinky
8       22  Anubhav
9       23   Khushi
10      24    Vinay
11      25    Neetu
```

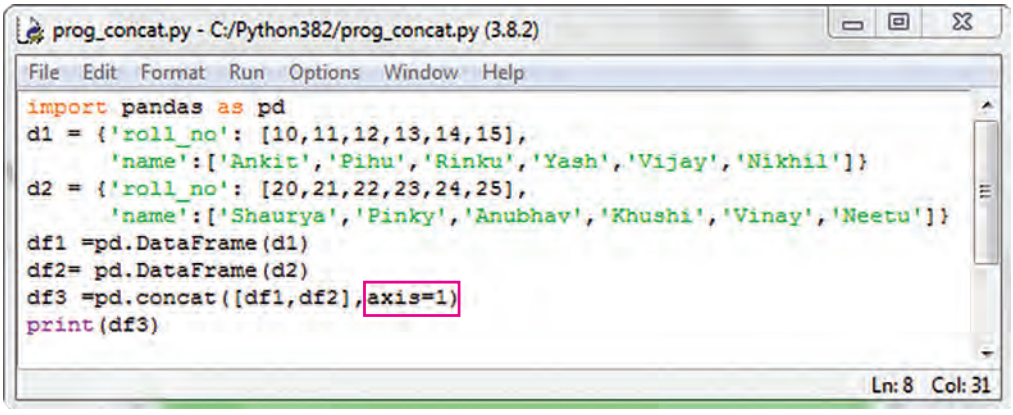
Row labels are adjusted automatically

The above program is the modification of the previous program. Here we have taken "ignore_index" argument and set its value as True. This will ignore the row labels from both the dataframes and the resultant dataframe shall adjust the index automatically as per the join operation while using the concat() function.

Till now, the dataframes have been concatenated along the rows. Alternatively we can concatenate dataframes along columns as well.

Practical Implementation–66

Program to concatenate two dataframes along columns.




```
>>>
=== RESTART: C:/Python382/prog_concat.py ===
   roll_no    name  roll_no     name
0       10   Ankit       20  Shaurya
1       11    Pihu       21    Pinky
2       12   Rinku       22  Anubhav
3       13    Yash       23   Khushi
4       14   Vijay       24    Vinay
5       15  Nikhil       25    Neetu
```

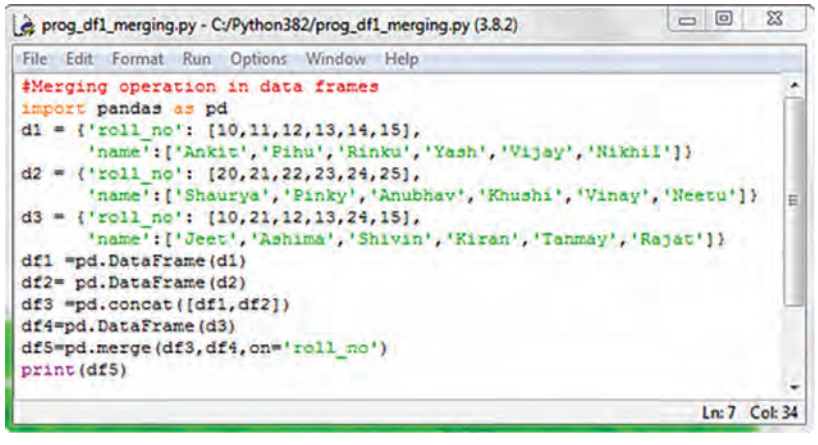
The above code concatenates the two dataframes along columns. To do this, the axis parameter is set as 1; **(axis=1)** means along column or, we can say, column-wise. Hence, the output obtained above is along the columns of the two dataframes.

3.26 MERGE OPERATION IN DATAFRAME

Pandas provides a single function, merge(), as the entry point for all standard database join operations between dataframe objects. We will first discuss merge() function.

Practical Implementation–67

Program to perform merging of dataframes.



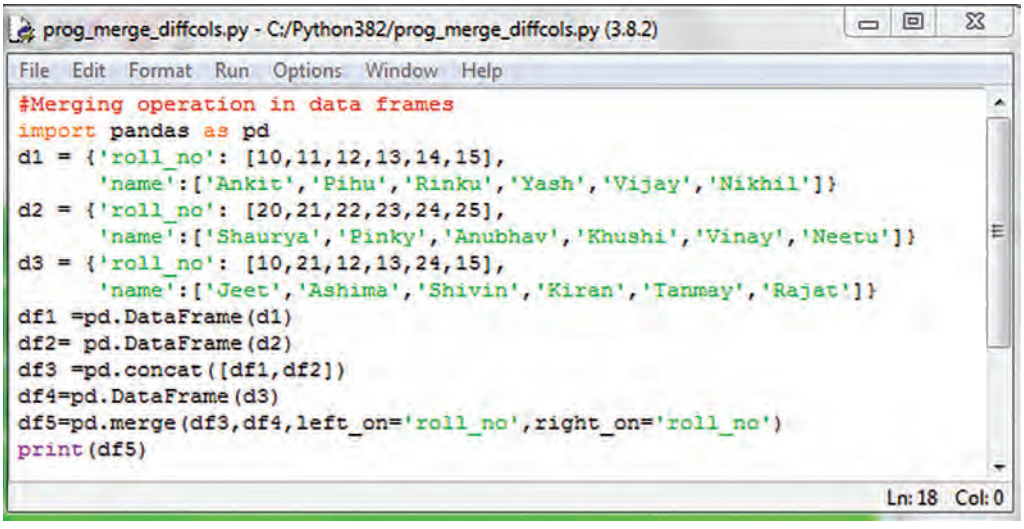
```
>>>
===== RESTART: C:/Python382/prog_df1_merging.py
   roll_no  name_x  name_y
0       10   Ankit    Jeet
1       12   Rinku  Shivin
2       13    Yash   Kiran
3       15  Nikhil   Rajat
4       21   Pinky  Ashima
5       24   Vinay  Tanmay
```

The above program shall generate the output consisting of the rows which are common between the two dataframes for the corresponding column values, which is "roll_no" in the above case.

This was the situation when the merging is done on the basis of a common column. But apart from this, it can also be done on the basis of different columns in the dataframes, which we will now see in the next implementation.

Practical Implementation–68

Program to perform merging of dataframes with uncommon columns and different names.




```
>>>
==== RESTART: C:/Python382/prog_merge_diffcols.py
   roll_no  name_x  name_y
0       10   Ankit    Jeet
1       12   Rinku  Shivin
2       13    Yash   Kiran
3       15  Nikhil   Rajat
4       21   Pinky  Ashima
5       24   Vinay  Tanmay
```

In the above program, the columns on which the dataframes are to be merged have different names. In such a case of merging, the argument "left_on" is to be specified for the left dataframe and "right_on" for the right dataframe name and hence the output.

3.27 CSV FILE

CSV (Comma Separated Values) is a simple **file format** used to store tabular data, such as a spreadsheet or database. A CSV file stores tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format.

For working CSV files in Python, there is an in-built module called **csv**. Files of this format are generally used to exchange data, usually when there is a large amount, between different applications.

3.28 DATA TRANSFER BETWEEN DATAFRAMES AND .CSV FILE

CSV format is a kind of tabular data separated by comma and is stored in the form of plain text.

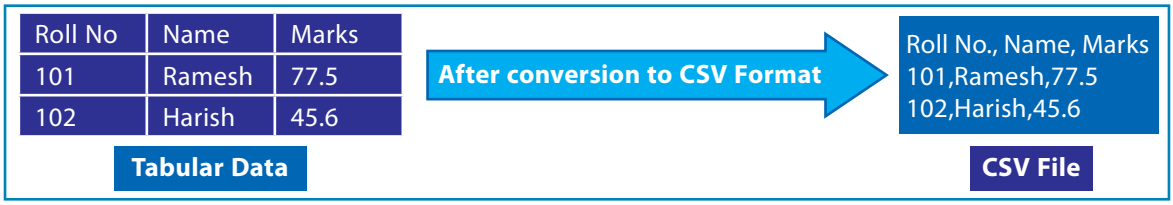


Fig. 3.6: Spreadsheet data vs CSV data

In CSV format:

- Each row of the table is stored in one row.
- The field-values of a row are stored together with comma after every field value.

Advantages of CSV format:

- A simple, compact and ubiquitous format for data storage.
- A common format for data interchange.
- It can be opened in popular spreadsheet packages like MS Excel, Open Office-Calc, etc.
- Nearly all spreadsheets and databases support import/export to CSV format.

CTM: CSV is a simple file format used to store **tabular** data, such as a **spreadsheet** or **database**.

3.28.1 Creating and Reading CSV File

A CSV is a **text file**, so it can be created and edited using any **text editor**. More frequently, however, a CSV file is created by exporting a spreadsheet or database in the program that created it.

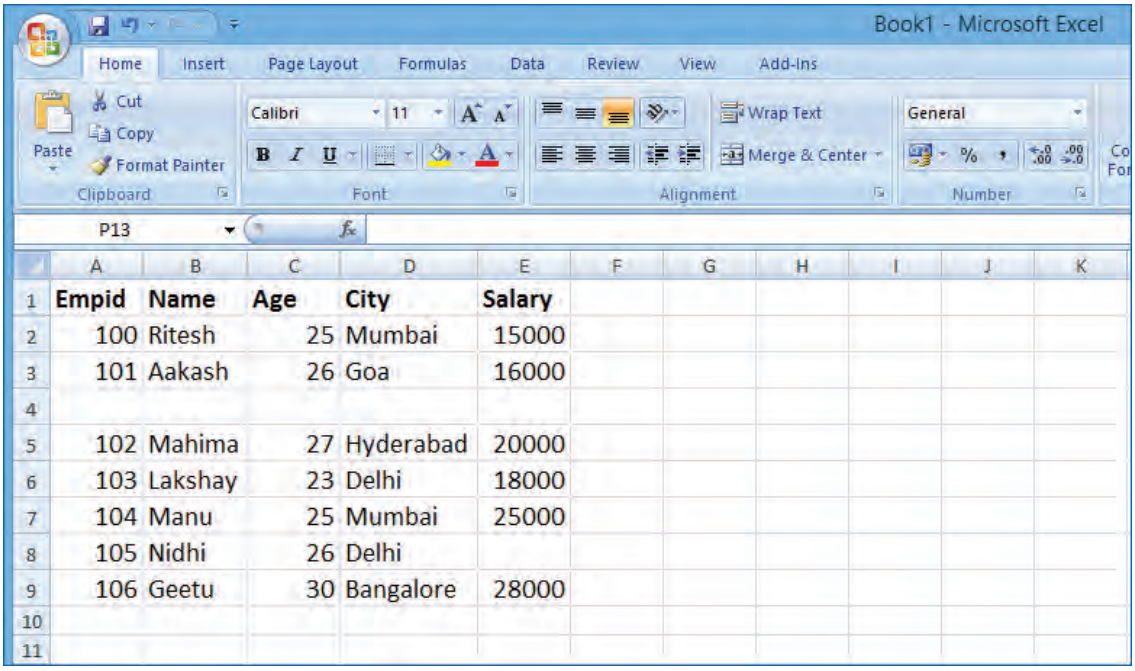
All CSV files follow a standard format, *i.e.*, each column is separated by a delimiter (such as a comma, semicolon, space or a tab) and each new line indicates a new row.

Let us create a CSV file using Microsoft Excel on the basis of "**Employee**" table.

Table: Employee table

Empid	Name	Age	City	Salary
100	Ritesh	25	Mumbai	15000
101	Aakash	26	Goa	16000
102	Mahima	27	Hyderabad	20000
103	Lakshay	23	Delhi	18000
104	Manu	25	Mumbai	25000
105	Nidhi	26	Delhi	
106	Geetu	30	Bangalore	28000

1. Launch Microsoft Excel.
2. Type the data given in the above Table in the Excel sheet (Fig. 3.7). You will also notice that some cell values are missing to represent missing values (NaN) in Pandas dataframe.



	A	B	C	D	E
1	Empid	Name	Age	City	Salary
2	100	Ritesh	25	Mumbai	15000
3	101	Aakash	26	Goa	16000
4					
5	102	Mahima	27	Hyderabad	20000
6	103	Lakshay	23	Delhi	18000
7	104	Manu	25	Mumbai	25000
8	105	Nidhi	26	Delhi	
9	106	Geetu	30	Bangalore	28000
10					
11					

Fig. 3.7: Microsoft Excel Worksheet for Employee

3. Save the file with a proper name by clicking File -> Save or Save As or press Ctrl + S to open the Save As window as shown in Fig. 3.8.
4. Type the name of the file as Employee and select file type as **CSV (Comma delimited) (*.csv)** from the drop-down arrow (Fig. 3.8).
5. Click on **Save** button. Excel will ask for confirmation to select CSV format.
6. Click on **OK** as shown in Fig. 3.9.

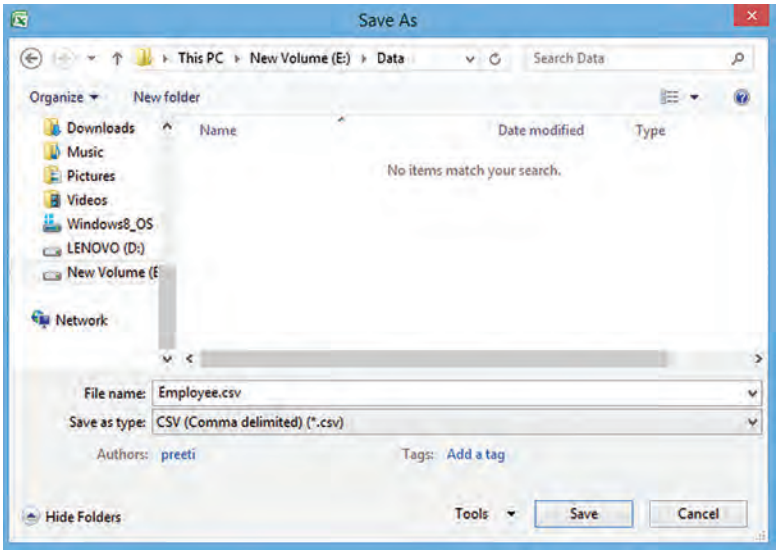


Fig. 3.8: Saving the Worksheet in CSV Format

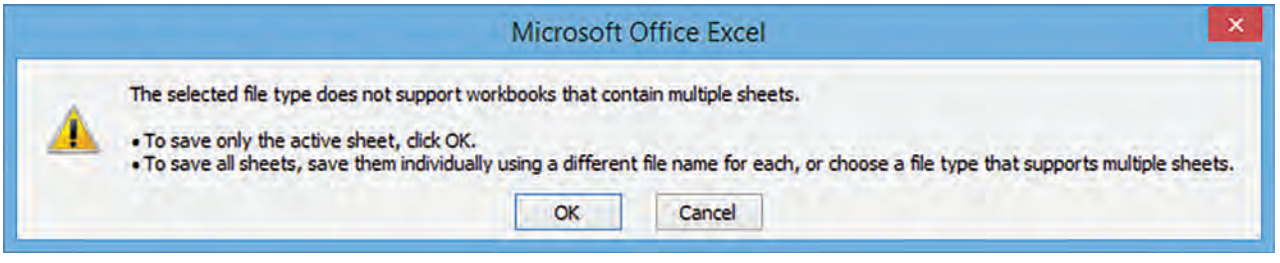


Fig. 3.9: Seeking permission to save in CSV format

7. It will display a dialog box for asking permission to keep comma as delimiter for CSV file (Fig. 3.10).

Fig. 3.10: Setting delimiter for CSV format

8. Lastly, click on **Yes** to retain and save the Excel file in CSV format.

   To view this CSV file, open any Text Editor (Notepad preferably) and explore the folder containing Employee.csv file. (In this case, the path for CSV file is: "E:\Data\Employee.csv").

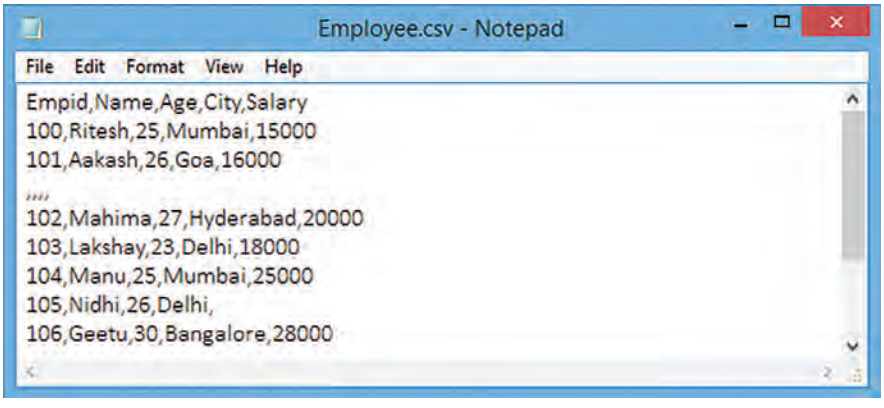


Fig. 3.11: CSV file contents using Notepad

If you open the file in a Notepad editor, you will observe that each column is separated by a comma (,) delimiter and each new line indicates a new row/record (Fig. 3.11).

3.28.2 Reading from a CSV File to Dataframe

After creating a simple "Employee" CSV file, it can be read using **read_csv()** function in Pandas once you know the path of your file. The read_csv() function loads the data in a Pandas dataframe.

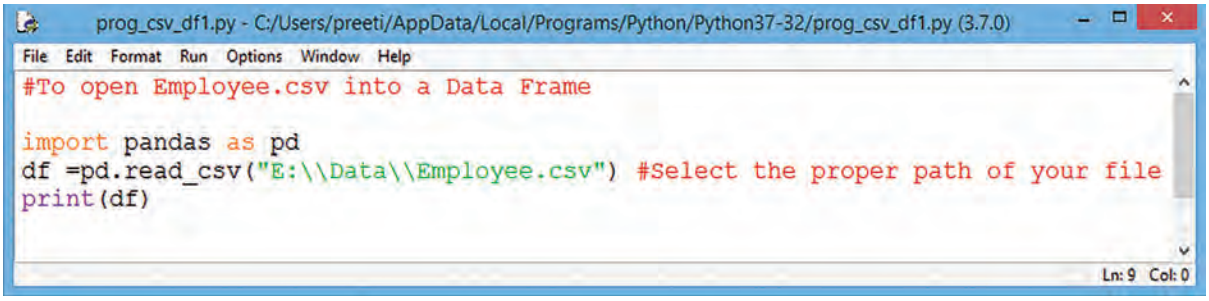
We know that multiple dataframes have multiple data types, *i.e.*, some columns are integers or numbers, some are float, strings or dates, etc. But CSV file treats all these data types as characters only. However, Pandas interprets these data types specifically when loading the data. *For example*, if a column contains only numbers, Pandas will set that column's data type to an integer or float.

Syntax for read_csv() method is:

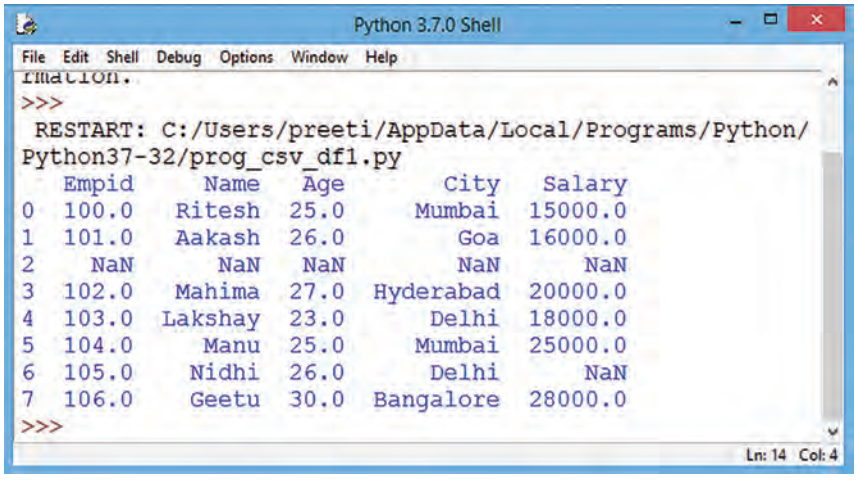
```
import pandas as pd
<df>=pd.read_csv(<FilePath>)
```

Practical Implementation–69

To create and open "Employee.csv" file using Pandas.



The first line imports the Pandas module. The read_csv method loads the data in a Pandas dataframe 'df'. pd.read_csv("path") shall fetch the data from csv file and display all records at the command prompt as shown in the output window.



One thing to be remembered is that the missing values from the CSV file shall be treated as NaN (Not a Number) in Pandas dataframe.

Practical Implementation–70

To display the shape (number of rows and columns) of the CSV file.

We can see the total number of rows (records) and columns (fields) present in the table with the help of **shape** command.

```
   Empid     Name   Age       City   Salary
0  100.0   Ritesh  25.0     Mumbai  15000.0
1  101.0   Aakash  26.0        Goa  16000.0
2    NaN      NaN   NaN        NaN      NaN
3  102.0   Mahima  27.0  Hyderabad  20000.0
4  103.0  Lakshay  23.0      Delhi  18000.0
5  104.0    Manu   25.0     Mumbai  25000.0
6  105.0    Nidhi  26.0      Delhi      NaN
7  106.0    Geetu  30.0  Bangalore  28000.0
>>> df.shape
(8, 5)
```

```
>>> row,column =df.shape
>>> row
8
>>> column
5
```

In the above case, we have directly displayed the row count and column count at Python shell prompt by giving the command as df.shape. We can also display it using variables.

The read_csv() method automatically takes the first row of the CSV file and assigns it as the dataframe header. After the creation of dataframe from a CSV file, you can perform all the dataframes operations on it.

☞ **Reading CSV file with specific/selected columns**

While working with large tables in CSV format, there can be several columns contained in it. But you may require selective columns to be read into a dataframe. This can be done by using "**usecols**" attribute or option along with read_csv() method. *For example*, in the case of "Employee" table, you have to access Name, Age and Salary of employees. This can be done by giving the command as:

```
>>> df = pd.read_csv("E:\\Data\\Employee.csv",
                     usecols = ['Name','Age','Salary'])
>>> df
```

Practical Implementation–71

To display name, age and salary from Employee.csv.

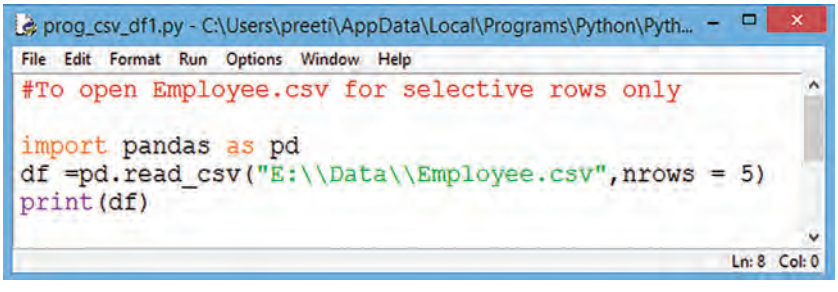
```
>>> df = pd.read_csv("E:\\Data\\Employee.csv",usecols =['Name','Age','Salary'])
>>> df
      Name   Age   Salary
0   Ritesh  25.0  15000.0
1   Aakash  26.0  16000.0
2      NaN   NaN      NaN
3   Mahima  27.0  20000.0
4  Lakshay  23.0  18000.0
5     Manu  25.0  25000.0
6    Nidhi  26.0      NaN
7    Geetu  30.0  28000.0
>>>
```

☞ **Reading CSV file with specific/selected rows**

Like columns, there can be thousands of records in a CSV file. You can display selective records/rows or selective lines using "**nrows**" option or attribute used with read_csv() method. This can be done by giving the command as:

Practical Implementation–72

To display only 5 records from Employee.csv.



In the above code, we have given 5 as the value to '**nrows**' attribute used with read_csv. **Nrows** means number of rows. In the above example, 5 represents the first five records, even empty records containing NaN values, excluding headers.

Hence, the following output shall be obtained.

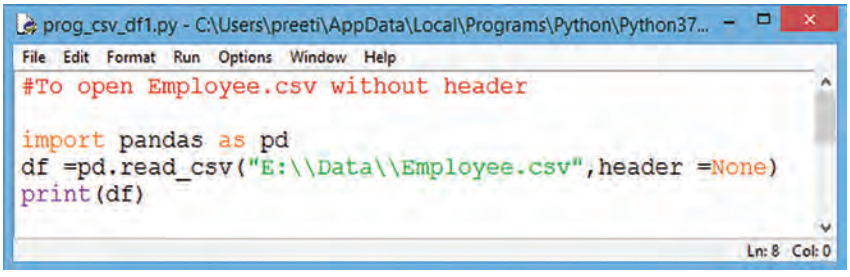
```
>>>
 RESTART: C:\Users\preeti\AppData\Local\Programs\
Python\Python37-32\prog_csv_df1.py
   Empid     Name   Age       City   Salary
0  100.0   Ritesh  25.0     Mumbai  15000.0
1  101.0   Aakash  26.0        Goa  16000.0
2    NaN      NaN   NaN        NaN      NaN
3  102.0   Mahima  27.0  Hyderabad  20000.0
4  103.0  Lakshay  23.0      Delhi  18000.0
>>>
```

☞ **Reading CSV file without header**

If you do not want to display the first row as the header for dataframe using Employee table, then this can be done by specifying **None** argument for '**header**' option or '**skiprows**' option using read_csv() method.

Practical Implementation–73

To display records without header



In the above example, None is the argument passed to 'header' attribute; as a result, the resultant dataframe will contain the header information as the first row, shown in the output below:

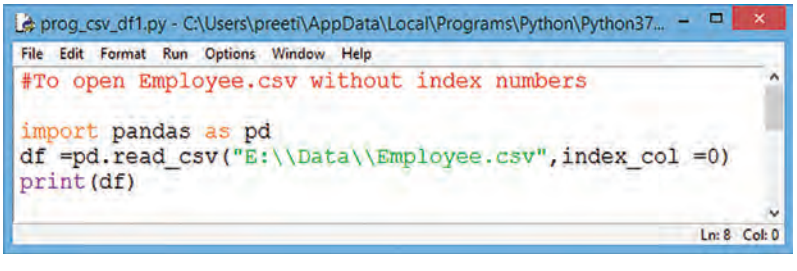
```
 RESTART: C:\Users\preeti\AppData\Local\Programs\
Python\Python37-32\prog_csv_df1.py
       0        1    2          3       4
0  Empid     Name  Age       City  Salary
1    100   Ritesh   25     Mumbai   15000
2    101   Aakash   26        Goa   16000
3    NaN      NaN  NaN        NaN     NaN
4    102   Mahima   27  Hyderabad   20000
5    103  Lakshay   23      Delhi   18000
6    104     Manu   25     Mumbai   25000
7    105    Nidhi   26      Delhi     NaN
8    106    Geetu   30  Bangalore   28000
>>>
```

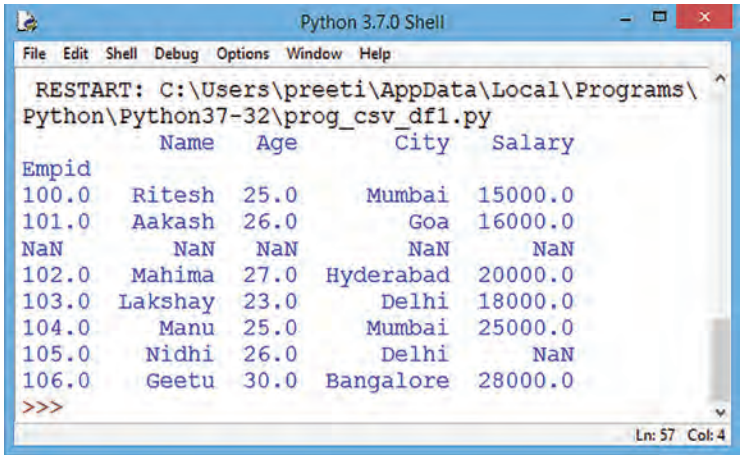
☞ **Reading CSV file without index**

You can also read and load the records into the dataframe without displaying their respective index numbers, which are displayed by default, by specifying the attribute **index_col = 0** using the read_csv() method.

Practical Implementation–74

To display records without index numbers.





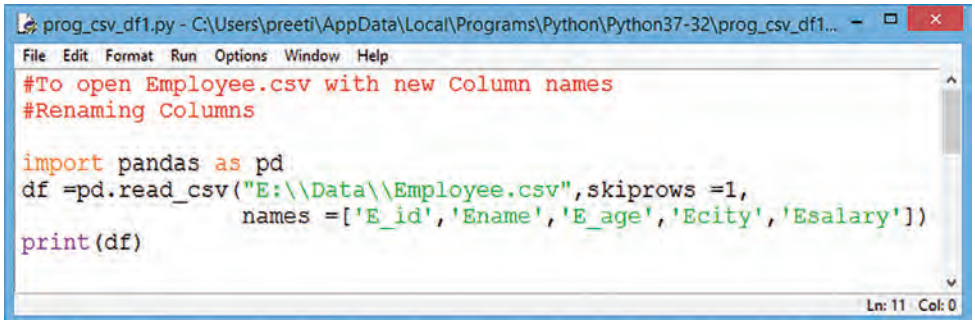
As shown in the above output window, there are no index numbers displayed along with records. Now Empid will be treated as the first column instead of indexes.

☞ **Reading CSV file with new column names**

You can read and load data from CSV file into a dataframe with new column names, *i.e.*, you can rename the columns while reading the .csv file. If the header exists, you have to skip it using **skiprows** option along with **names** option for renaming the columns.

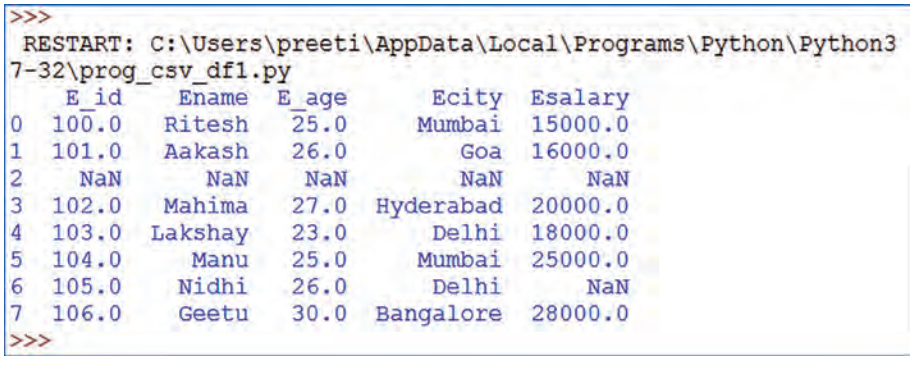
Practical Implementation–75

To display Employee file with new column names.



In the above code, we have given the option skiprows = 1 which will omit the default column names from the CSV file. Names option holds the new column names to be displayed while loading the CSV file into the dataframe.

Hence, the following output shall be displayed:

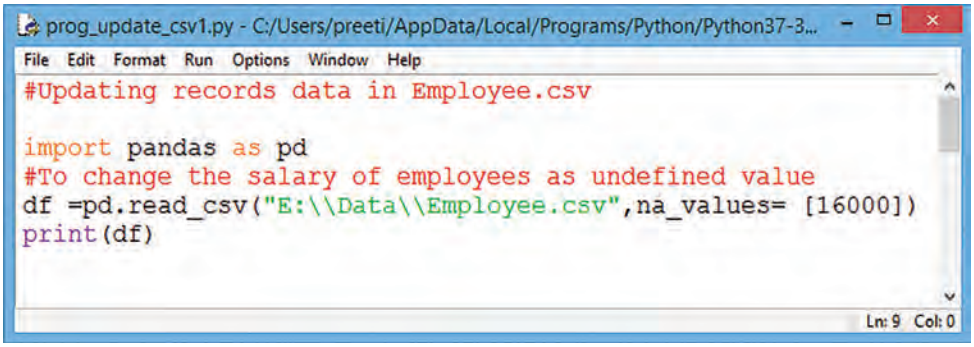


3.29 UPDATING/MODIFYING CONTENTS IN A CSV FILE

In the above section, we learnt how to change column name. Similarly, we can modify or update row data as well. Suppose we want to replace the salary of the employee whose salary is 16000 with NaN value. This can be done by using **na_values** option along with read_csv method for the respective salary as 16000.

Practical Implementation–76

To modify the salary of employee earning 16000 with NaN value.




```
>>>
 RESTART: C:/Users/preeti/AppData/Local/Programs/Python/Python3
7-32/prog_update_csv1.py
   Empid     Name   Age       City   Salary
0  100.0   Ritesh  25.0     Mumbai  15000.0
1  101.0   Aakash  26.0        Goa      NaN
2    NaN      NaN   NaN        NaN      NaN
3  102.0   Mahima  27.0  Hyderabad  20000.0
4  103.0  Lakshay  23.0      Delhi  18000.0
5  104.0     Manu  25.0     Mumbai  25000.0
6  105.0    Nidhi  26.0      Delhi      NaN
7  106.0    Geetu  30.0  Bangalore  28000.0
>>>
```

As shown in the above output window, the salary of employee Aakash has been modified from 16000 to NaN.

```
>>>
 RESTART: C:\Users\preeti\AppData\Local\Programs\Python\Python3
7-32\prog_csv_df1.py
    E_id    Ename  E_age      Ecity  Esalary
0  100.0   Ritesh   25.0     Mumbai  15000.0
1  101.0   Aakash   26.0        Goa  16000.0
2    NaN      NaN   NaN        NaN      NaN
3  102.0   Mahima   27.0  Hyderabad  20000.0
4  103.0  Lakshay   23.0      Delhi  18000.0
5  104.0     Manu   25.0     Mumbai  25000.0
6  105.0    Nidhi   26.0      Delhi      NaN
7  106.0    Geetu   30.0  Bangalore  28000.0
>>>
```

3.30 WRITING A CSV FILE WITH DEFAULT INDEX

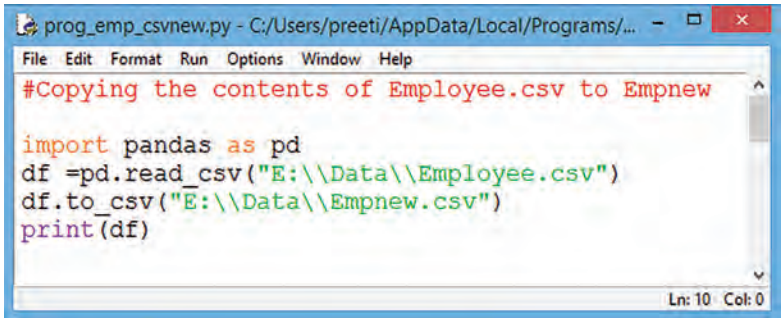
To create a **CSV** file from a dataframe, the **to_csv()** method is used. We can do this either by transferring the records directly to the CSV file or by copying the contents of the original CSV file to another file.

☞ **Copying Employee.csv to Empnew.csv.**

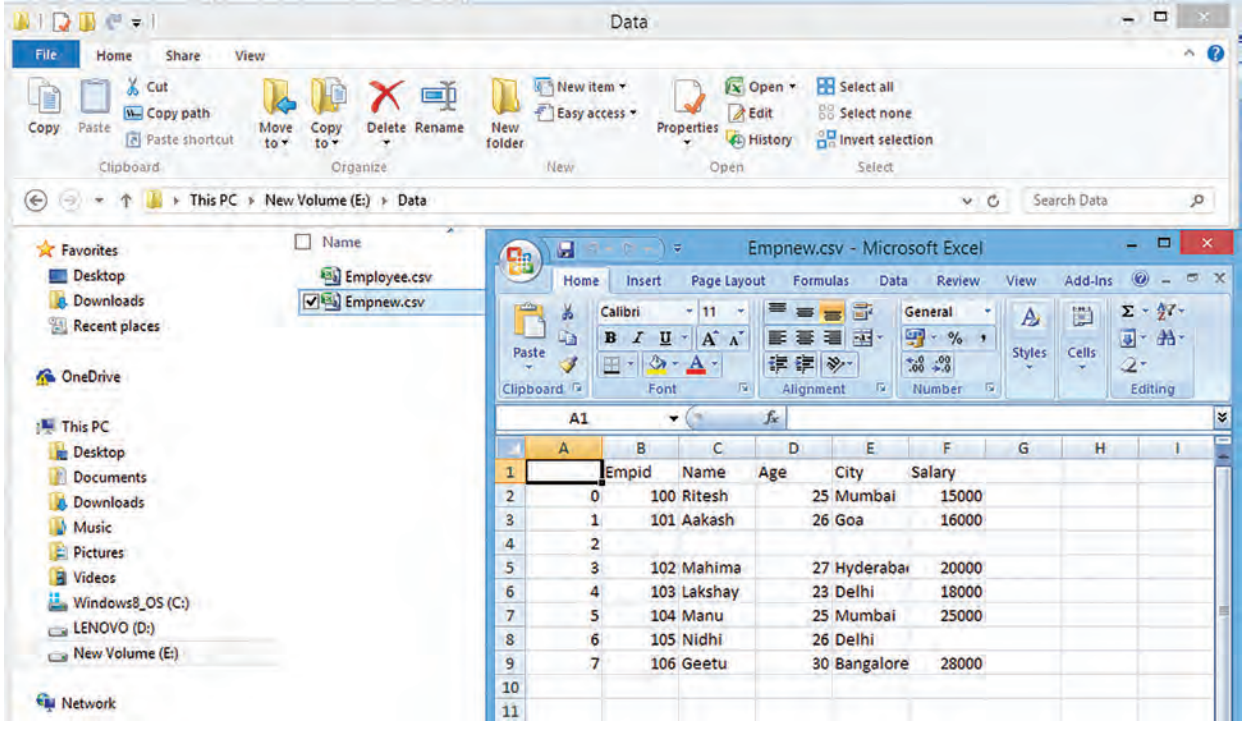
We can copy/write the data, *i.e.*, create a duplicate copy of Employee.csv as Empnew.csv.

Practical Implementation–77

To create a new CSV file by copying the contents of Employee.csv.



Upon executing the above commands, Empnew.csv (duplicate) file shall be created containing the same contents as Employee.csv with default index values. You can browse the folder to see the contents of the newly-created file Empnew.csv as shown below:



As shown in the above screenshot, the contents of the duplicate file Empnew are the same as that of Employee.csv. Creating a duplicate file allows the user to change the data or add new data to the already existing Employee.csv file without affecting the original data. In such a case, Pandas provides us with the option of creating a duplicate file with the previous data using the above mentioned command.

☞ Saving Dataframe as CSV file

This is the direct approach of creating a CSV file by first creating a dataframe and then loading it into a CSV file format.

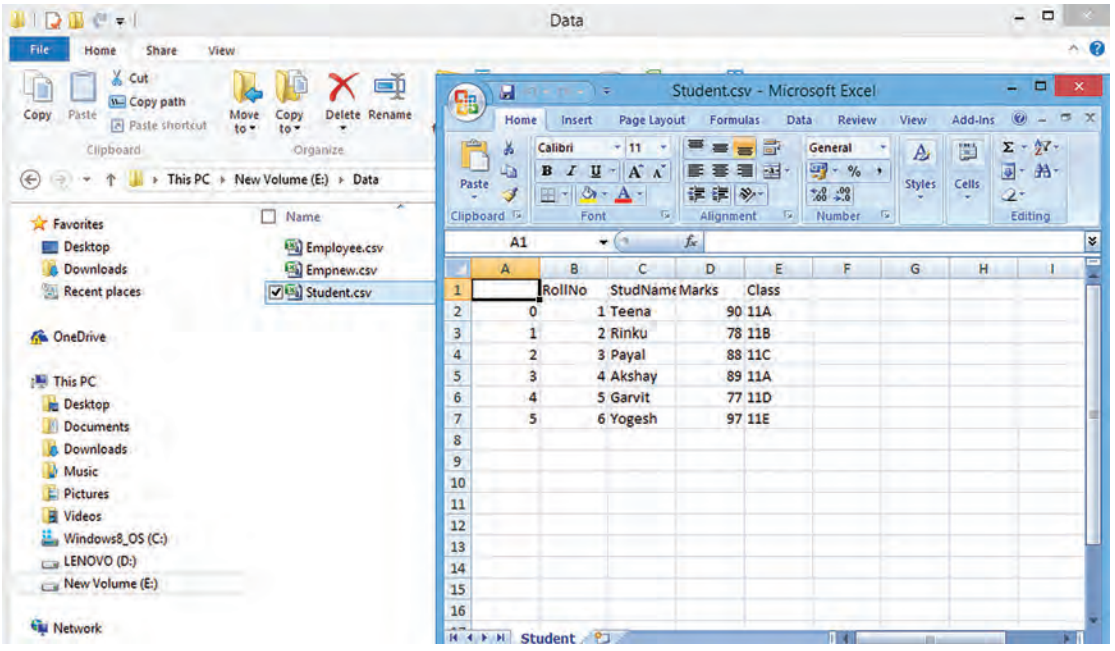
Practical Implementation–78

To create a student CSV file from dataframe.

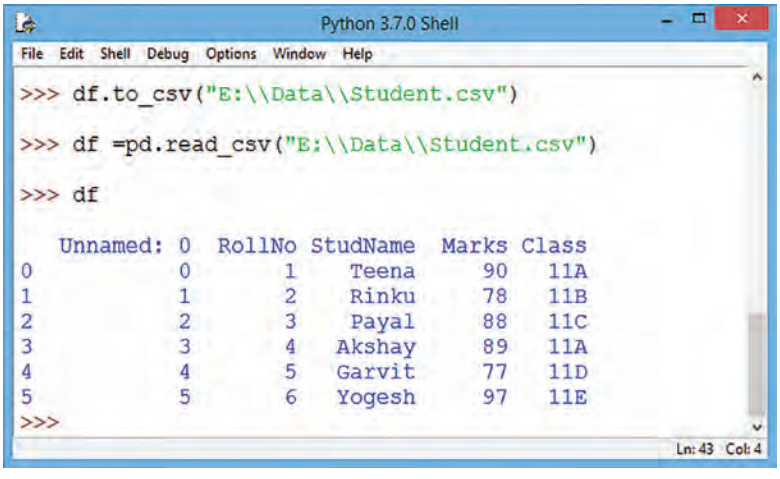
```
>>> Student ={'RollNo':[1,2,3,4,5,6],
              'StudName':['Teena','Rinku','Payal','Akshay','Garvit','Yogesh'],
              'Marks':[90,78,88,89,77,97],
              'Class':['11A','11B','11C','11A','11D','11E']}
```

```
>>> import pandas as pd
>>> df =pd.DataFrame(Student,columns =['RollNo','StudName','Marks','Class'])
>>> df.to_csv("E:\\Data\\Student.csv")
```

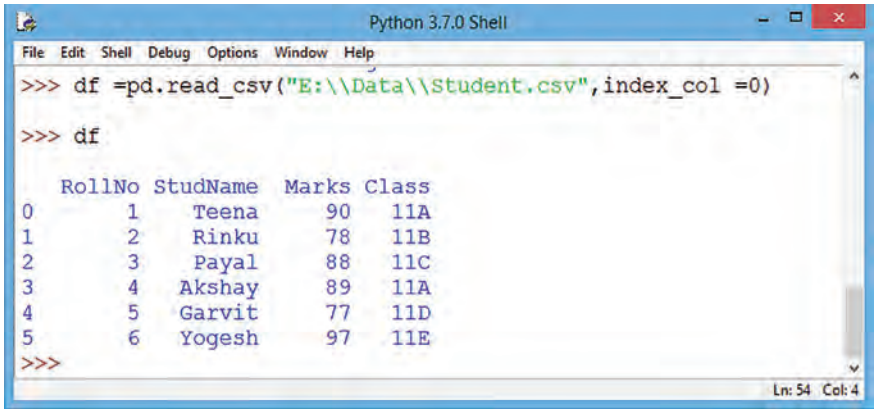
This will create Student.csv file in the Data folder. If you open this file in a spreadsheet like MS Excel, you will get the Student data in the form of rows/records and columns as shown below:



Our next step is to read the Student.csv into a dataframe using read_csv() method.



In the above output window, **Unnamed: 0** column gets displayed automatically along with the index values. To avoid this column, use the attribute **index_col =0** with read_csv() method as shown:

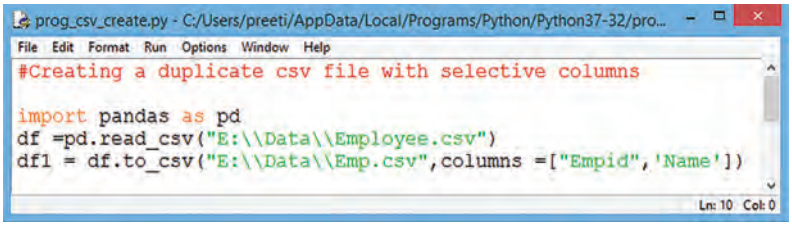


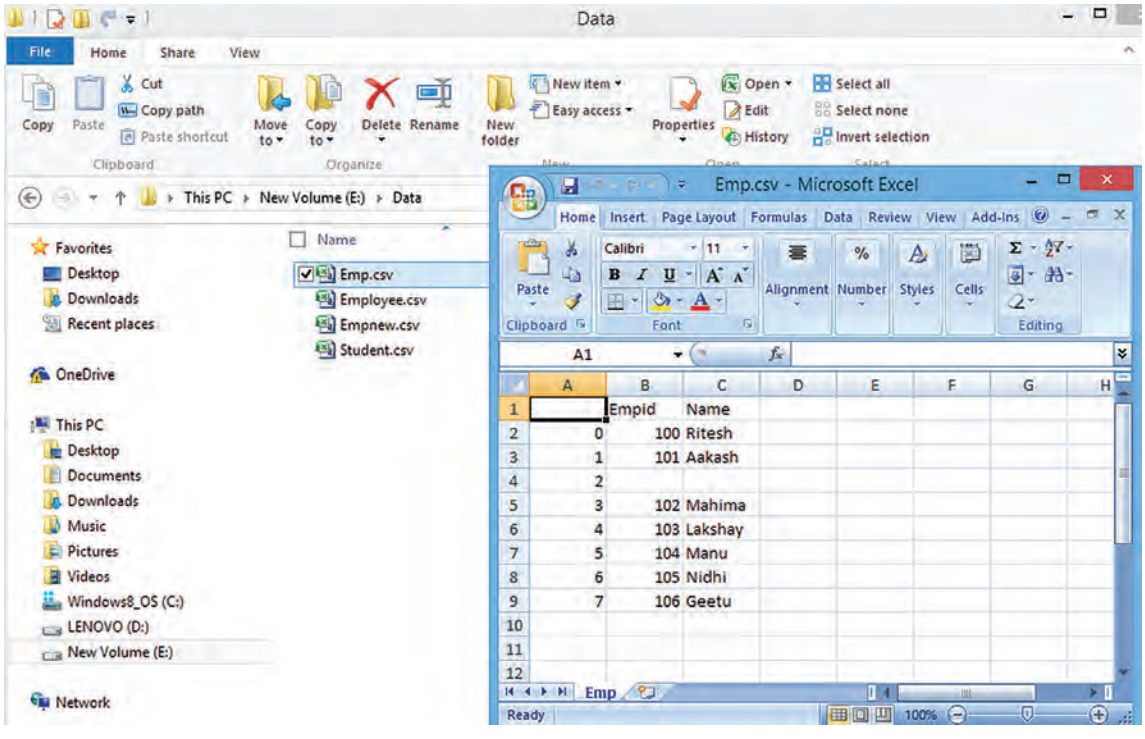
3.31 COPYING FIELDS INTO A NEW FILE

In reference to section 3.30, we have created Employee.csv file with complete set of columns and rows. In certain situations, it is required to create a duplicate file containing only the selected fields.

Practical Implementation–79

To create a duplicate file for Employee.csv containing Empid and Employee name.





Learning Tip: Ensure that field names are the same in both the files (case sensitive).

3.32 TRANSFERRING DATA BETWEEN CSV FILES/SQL DATABASES

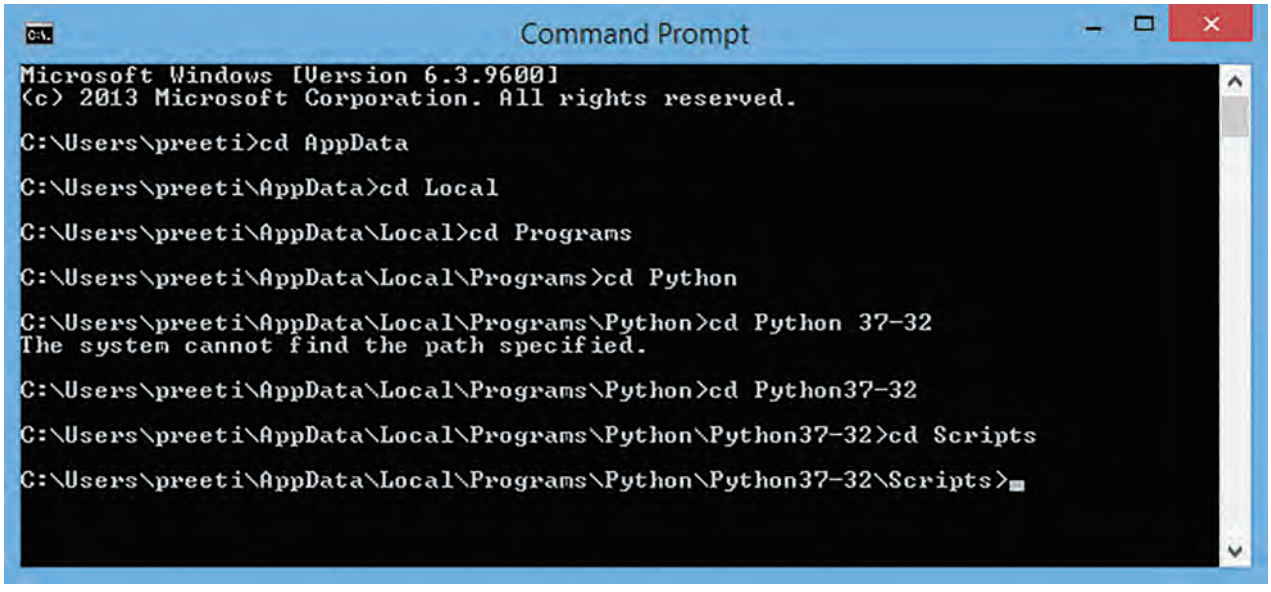
Pandas provides the flexibility to read the contents of a CSV file and insert that data into a database using SQL. SQL is a database where we can easily store information and can later use it for multiple purposes. Before we start working with Pandas and SQL, to access the SQL databases as Pandas dataframe, we need to install the necessary application and required connections of the databases in Python Pandas. We will be using MySQL to implement SQL processing.

Before establishing the connection between Pandas and SQL, we have to install any of the following packages in a virtual environment:

- **MySQLdb** – MySQLdb is the Python module to work with MySQL databases. It is one of the most commonly used Python packages for MySQL.
- **MySQL-connector-python** – This package contains the mysql.connector module, which is entirely written in Python.
- **PyMySQL** – This package contains PyMySQL module. It has been designed as a replacement for MySQLdb.

We shall be using mysql-connector for establishing the Python-SQL connectivity.

How to install this connector with Python has been explained step-wise as follows:



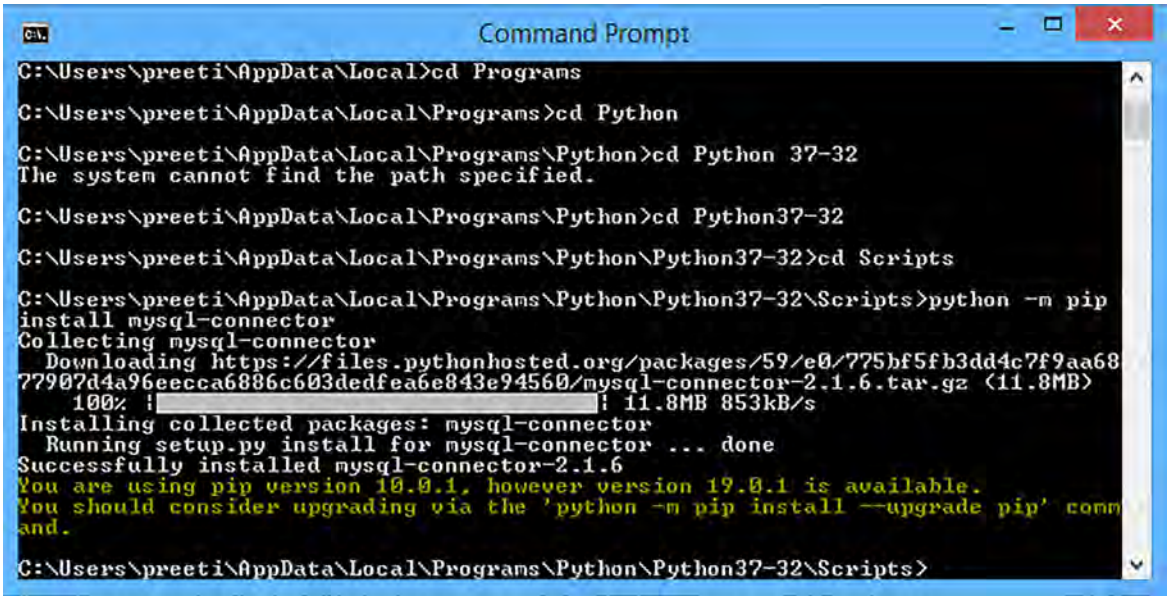
Step 1: To connect Python to MySQL, we have to install mysql-connector using 'pip' command on the command prompt (cmd).

NOTE: Do ensure that mysql connector is installed in the same folder for Python and it should take the path as:

> cd C:\Users\your name\AppData\Local\Programs\Python\Python37-32\Scripts

Step 2: Once you have set the path, type the command as:

> cd C:\Users\your name\AppData\Local\Programs\Python\Python37-32\Scripts \python–m pip install mysql-connector



Mysql-connector shall be downloaded and installed on your system. Once done, we need to check whether it has been properly installed or not.

Step 3: To do this, type import mysql-connector, using Python shell. If no error message gets displayed, then this signifies that the driver has been successfully installed.

```
>>> import mysql.connector
>>>
```

In a nutshell, three things are to be kept in mind for the successful installation of mysql connector:

- Download Python 3.x and then install it.
- Download MySQL API, exe file will be downloaded; install it.
- Install MySQL-Python Connector
- Now connect MySQL Server using Python.

CTM: MySQL Connector Python requires Python to be in the system's PATH. Installation fails if it doesn't find Python.

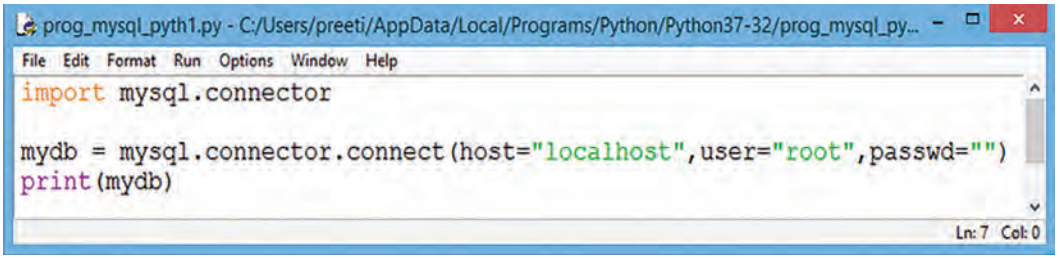
After the installation gets completed, a virtual environment gets created to start working with real databases.

3.33 ESTABLISHING CONNECTION

After completing the first step of installing mysql connector, the next step is using MySQL to establish a connection to the database that you wish to use. Once the connection is successfully established, connection object gets created. Type the following command for establishing the MySQL-Python connectivity:

Consider a 'School' database containing 'Student' table in it. The contents of Student table are as shown:

Rollno	Name	age	city	marks
1	Tarun	23	Mumbai	398
2	Pooja	21	Chail	390
3	Radhika	18	Shimla	388
4	Sonia	24	Goa	300
5	Vinay	25	Pune	410
10	Shaurya	15	Delhi	345



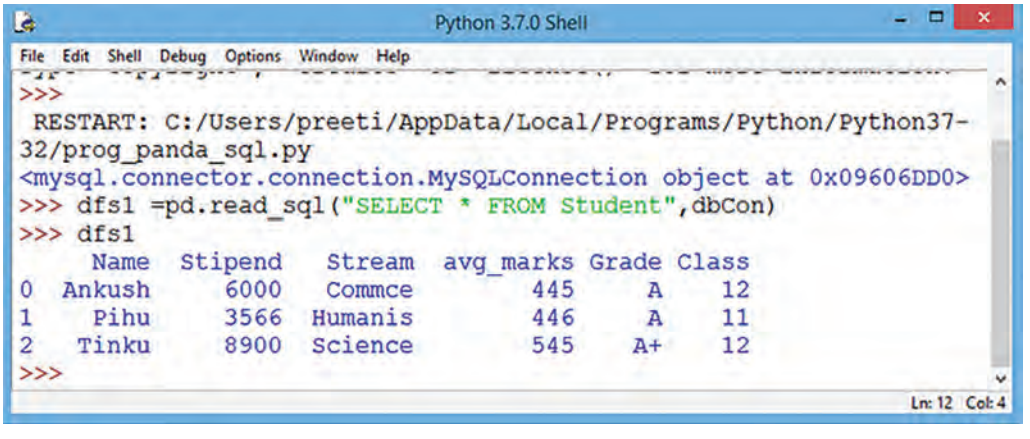
3.34 READING MySQL TABLE AS DATAFRAME

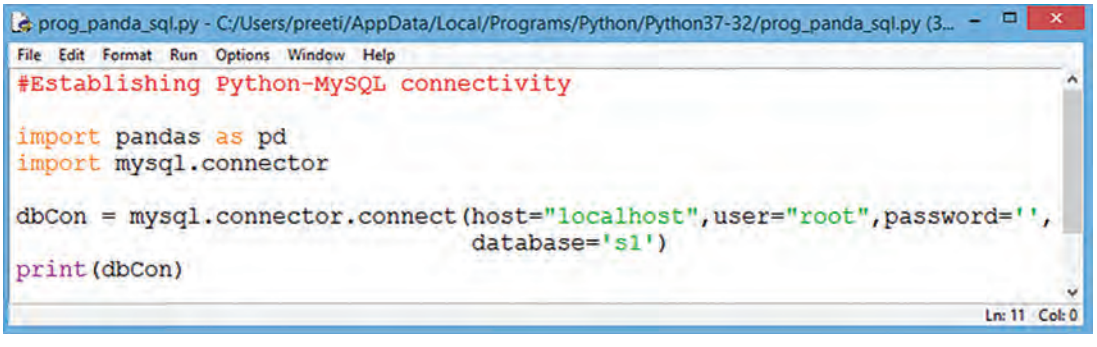
Like CSV file, a database table can be read as a dataframe using pandas **read_sql()** method. The syntax is:

```
pandas.read_sql(sql command, connection_obj)
```

read_sql() method returns a Pandas dataframe object. The frame will have the default-naming scheme where the rows start from zero and get incremented for each row. The columns will be named after the column names of the MySQL database table. The above code shall retrieve all the rows from student table and shall store these details in the dataframe dfs1, which is then displayed as shown in the output.

Start the Python IDLE and type the following command at the shell prompt:





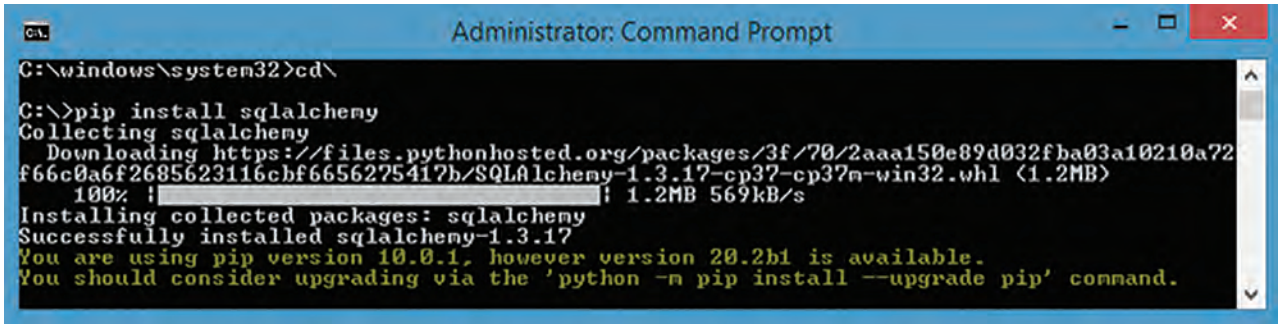
3.34.1 Reading Selective Columns from MySQL Table as Dataframe

The above program displays all the records from a MySQL table. In the same manner, selective columns can also be displayed.

This is done by using and importing a library, sqlalchemy.

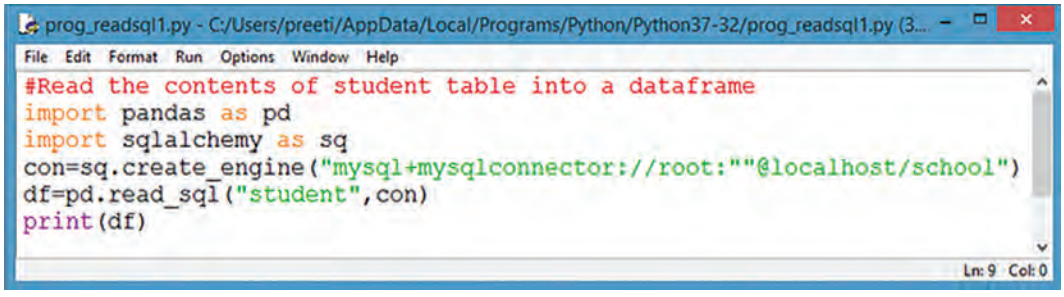
We can read and write data from sql to a dataframe and vice versa using this library sqlalchemy. Thus, the first thing to do is to install it using the command as:

C:\pip install sqlalchemy



Practical Implementation–80

Reading the contents from "student" table.



read_sql() method is used to read the contents from the student table. The first thing to do is to import libraries, pandas and sqlalchemy. The format of connection string is:

mysql+mysqlconnector://username:password@host:port/database_name

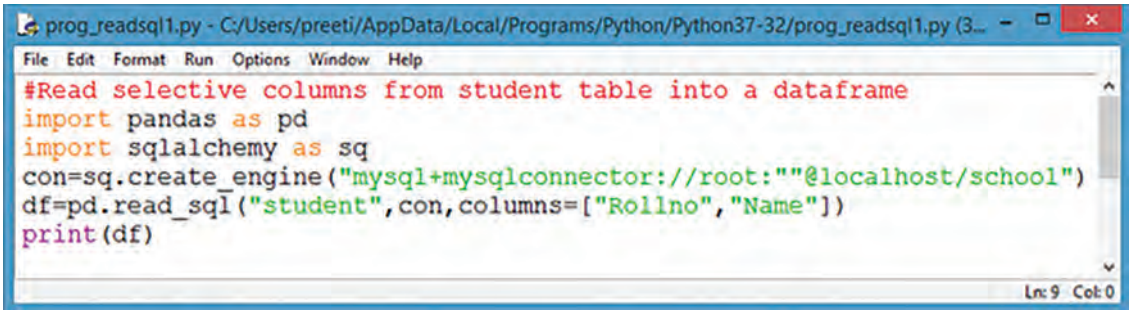
For the above program, the password is " "(blank) and the name of the database is "school". To connect with SqlAlChemy, you use the **create_engine()** function to create an engine object from database URL. You only need to create the engine once per database you are connecting to. The name of the table (student, in the above case) is passed as the argument to read_sql() method. On execution of the above code, all the records from the student table are displayed.

```
>>>
 RESTART: C:/Users/preeti/AppData/Local/Programs/
1.py
   Rollno     Name  age    area  marks
0       2     Pooja   21   Chail    390
1       3   Radhika   18  Shimla    388
2       4    Sonia    24     Goa    300
3       5    Vinay    28    Pune    410
4      10  Shaurya    15   Delhi    345
```

In the above implementation, all the records are being displayed. You can display selective columns as well from the sql database into a dataframe.

Practical Implementation–81

Reading selective records from sql table student.

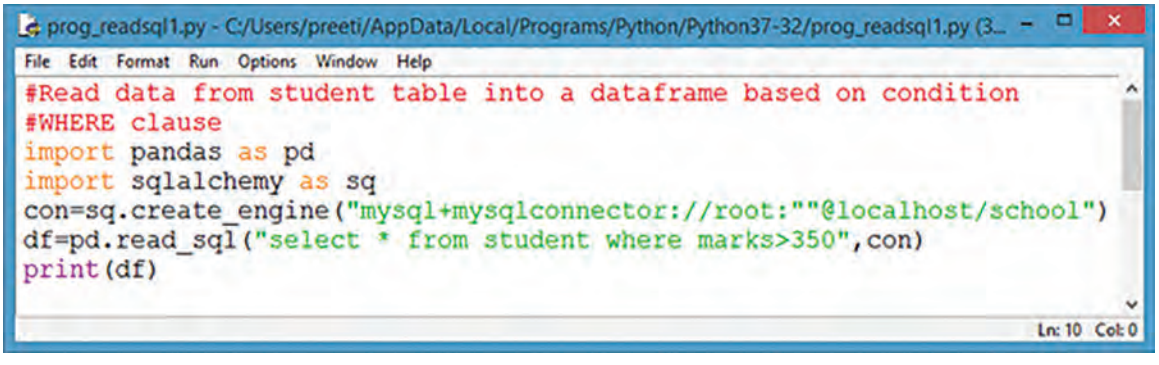


```
prog_readsql1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_readsql1.py (3...
File  Edit  Format  Run  Options  Window  Help
#Read selective columns from student table into a dataframe
import pandas as pd
import sqlalchemy as sq
con=sq.create_engine("mysql+mysqlconnector://root:""@localhost/school")
df=pd.read_sql("student",con,columns=["Rollno","Name"])
print(df)
                                                                    Ln: 9  Col: 0
```

```
>>>
 RESTART: C:/Users/preeti/AppData/I
1.py
   Rollno     Name
0       2    Pooja
1       3  Radhika
2       4    Sonia
3       5    Vinay
4      10  Shaurya
```

Practical Implementation–82

Reading selective records from sql table student using the WHERE clause. To display the records of all the students, who have secured more than 350 marks, from student table.



```
prog_readsql1.py - C:/Users/preeti/AppData/Local/Programs/Python/Python37-32/prog_readsql1.py (3...
File  Edit  Format  Run  Options  Window  Help
#Read data from student table into a dataframe based on condition
#WHERE clause
import pandas as pd
import sqlalchemy as sq
con=sq.create_engine("mysql+mysqlconnector://root:""@localhost/school")
df=pd.read_sql("select * from student where marks>350",con)
print(df)
                                                                    Ln: 10  Col: 0
```

```
>>>
 RESTART: C:/Users/preeti/AppData/Local/
1.py
   Rollno     Name  age    area  marks
0       2    Pooja   21   Chail    390
1       3  Radhika   18  Shimla    388
2       5    Vinay   28    Pune    410
```

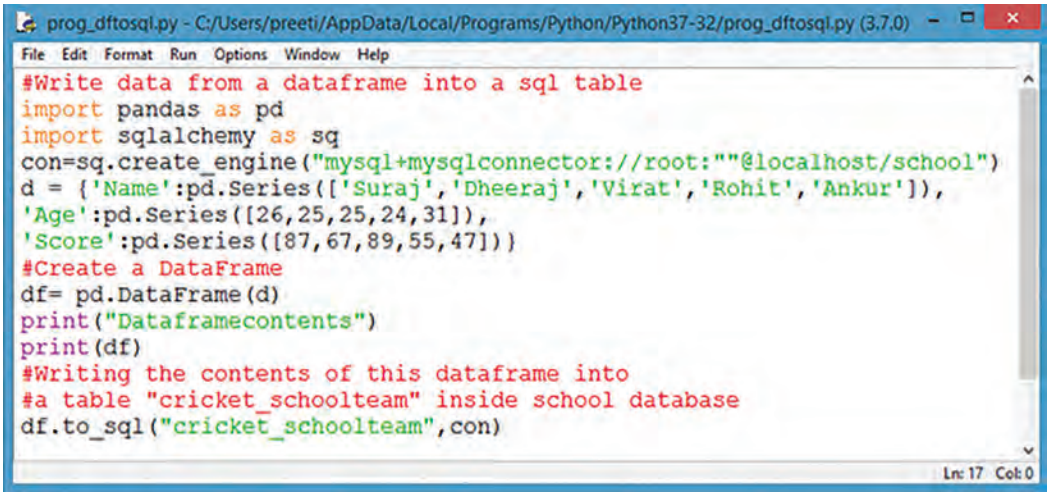
The above program involves creating a dataframe based on the result of a query (Select statement). Hence the contents show the students who have scored more than 350 marks.

3.35 WRITING TO A MySQL TABLE AS DATAFRAME

You can even write the contents of a dataframe into an sql table. .to_sql() method is used to write data from a dataframe into an sql table. Invoke to_sql() method on the Pandas dataframe instance and specify the table name and database connection. This creates a table in MySQL database server and populates it with the data from the Pandas dataframe.

Practical Implementation–83

Writing data into an sql table using dataframe.




```
>>>
 RESTART: C:/Users/preeti/AppData
.py
Dataframecontents
      Name  Age  Score
0    Suraj   26     87
1  Dheeraj   25     67
2    Virat   25     89
3    Rohit   24     55
4    Ankur   31     47
>>>
```

In the above program, the dataframe contents are written into the MySQL table "cricket_schoolteam". .to_sql() method writes the contents of the dataframe into a MySQL table and upon execution shall create a new table containing the school cricket team details furnished through a dataframe. It will display the cursor only since no display statement has been given. Hence, cursor shall be blinking at the shell prompt which indicates the successful creation of this new table.

This can be seen by opening the listing of tables in MySQL by giving the command as:

SHOW TABLES;

```
mysql> show tables;
+--------------------+
| Tables_in_school   |
+--------------------+
| cricket_schoolteam |
| emp                |
| student            |
| student1           |
+--------------------+
4 rows in set (0.01 sec)

mysql> _
```

It can be seen from the above output window that "cricket_schoolteam" table has been created through Pandas dataframe inside school database.

To view the contents of this table, the command is given as:

SELECT * from cricket_schoolteam;

```
mysql> select * from cricket_schoolteam;
+-------+---------+------+-------+
| index | Name    | Age  | Score |
+-------+---------+------+-------+
|     0 | Suraj   |   26 |    87 |
|     1 | Dheeraj |   25 |    67 |
|     2 | Virat   |   25 |    89 |
|     3 | Rohit   |   24 |    55 |
|     4 | Ankur   |   31 |    47 |
+-------+---------+------+-------+
5 rows in set (0.00 sec)
```

As is evident from the output displayed, the contents of the dataframe have been written into the MySQL table, cricket_schoolteam.

This is how we write into an sql table using Pandas dataframes.

Now, the contents displayed as the output contain all the columns from the cricket_schoolteam table along with the index created by default.

This index can be ignored and avoided during displaying of the table contents by giving the argument "index" as False in .to_sql() statement. Index is set to True by default, as shown in the next implementation.

Practical Implementation–84

Writing a dataframe into an SQL table without index values.

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

File Edit Format Run Options Window Help

```
#Write data from a dataframe into a sql table
#without index values
import pandas as pd
import sqlalchemy as sq
con=sq.create_engine("mysql+mysqlconnector://root:""@localhost/school")
d = {'Name':pd.Series(['Suraj','Dheeraj','Virat','Rohit','Ankur']),
'Age':pd.Series([26,25,25,24,31]),
'Score':pd.Series([87,67,89,55,47])}
#Create a DataFrame
df= pd.DataFrame(d)
df.to_sql("cricket_schoolteam",con,index=False)
```

Ln: 14 Col: 0

```
mysql> select * from cricket_schoolteam
    -> ;
+---------+------+-------+
| Name    | Age  | Score |
+---------+------+-------+
| Suraj   |   26 |    87 |
| Dheeraj |   25 |    67 |
| Virat   |   25 |    89 |
| Rohit   |   24 |    55 |
| Ankur   |   31 |    47 |
+---------+------+-------+
5 rows in set (0.00 sec)
```

As observed from the output, the index column is removed and the rest of the columns are being displayed.

However, when you will execute the code, it will display the error message as the table already exists.

Thus, in order to overwrite the existing table with the one without the index values, you have to give "if_exists" argument with .to_sql() method as shown in the implementation below.

Practical Implementation–85

To overwrite an existing SQL table by using suitable argument with overwriting of the contents.

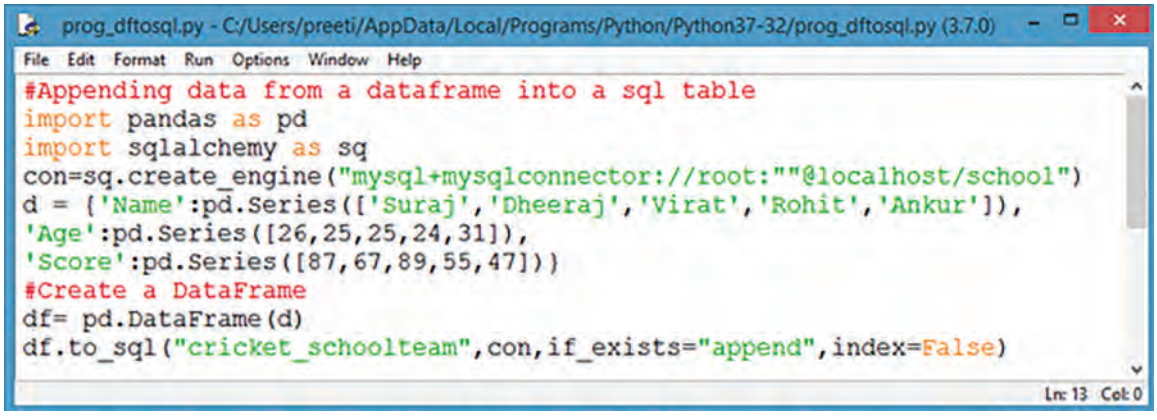

```
mysql> select * from cricket_schoolteam;
+---------+-------+-------+
| Name    | Age   | Score |
+---------+-------+-------+
| Suraj   |    26 |    87 |
| Dheeraj |    25 |    67 |
| Virat   |    25 |    89 |
| Rohit   |    24 |    55 |
| Ankur   |    31 |    47 |
+---------+-------+-------+
5 rows in set (0.00 sec)
```

Thus, if_exist argument shall overwrite the contents of the cricket_schoolteam table, excluding index column and overwriting the table, as "replace" has been given as the parameter to if_exist.

We can also add (append) the records by replacing parameter replace with append with if_exist argument.

Practical Implementation–86

Program to append the records of the dataframe into the MySQL table.



```
mysql> select * from cricket_schoolteam;
+---------+------+-------+
| Name    | Age  | Score |
+---------+------+-------+
| Suraj   |   26 |    87 |
| Dheeraj |   25 |    67 |
| Virat   |   25 |    89 |
| Rohit   |   24 |    55 |
| Ankur   |   31 |    47 |
| Suraj   |   26 |    87 |
| Dheeraj |   25 |    67 |
| Virat   |   25 |    89 |
| Rohit   |   24 |    55 |
| Ankur   |   31 |    47 |
+---------+------+-------+
10 rows in set (0.00 sec)
```

As shown in the output window, another set of records of the same dataframe has been added (appended) in the table cricket_schoolteam.

MEMORY BYTES

- head() returns the first 'n' rows of a dataframe.
- The default number of elements to display is five for head () and tail(), but you may pass a custom number.
- tail() returns the last 'n' rows of a dataframe
- Pandas provides a single function, merge(), as the entry point for all standard database join operations between dataframe objects.
- The acronym CSV is short for Comma Separated Values and refers to tabular data saved as plain text where data values are separated by commas.
- You can use pandas.read_csv() function to read data from a CSV file in your dataframe.
- The <DF> .to_csv() function saves the data of dataframe on a CSV file.
- The function read_csv() is Pandas' function and to_csv() is dataframe structure's function.
- The mysql-connector library can be used to interact with SQL databases from within Python.
- You can use pandas.read_sql() function to read data from an SQL table in your dataframe.
- The <DF>.to_sql()function saves the data of dataframe in an SQL table.
- We are using MySQL Connector Python to connect MySQL.
- mysql.connector.connect() method of MySQL Connector Python is used with required parameters to connect MySQL.
- MySQL Connector Python requires Python to be in the system's PATH. Installation fails if it doesn't find Python.
- Install MySQL Connector Python using pip command.
- Install sqlalchemy using pip command to read and write a dataframe into an sql table and vice versa.
- mysqlconnector 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.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.

(v) ..... function is used to read data from a CSV file in your dataframe.

(w) ..... function saves the data of dataframe on a CSV file.

(x) ..... function is used to read the contents from an sql database into a dataframe.

(y) ..... function is used to write the contents to an sql database from a dataframe.

Answers: (v) pandas.read_csv(); (w) to_csv() (x) read_sql()
(y) .to_sql()

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

(v) CSV refers to tabular data saved as plain text where data values are separated by commas.

(w) read_csv() method automatically takes the last row of the CSV file.

(x) Apart from the available dataset, pivoting can be successfully implemented by importing a .csv file as well.

Answers: (v) True (w) False (x) True

3. Multiple Choice Questions (MCQs)

(s) CSV stands for:

(i) Comma Separated Values (ii) Comma Separated Variables

(iii) Column Separated Values (iv) Column Separated Variables

(t) Which of the following commands is used to install sqlalchemy?

(i) pip install sqlchemy (ii) pip install sql-chemy

(iii) pip install sqlalchemy (iv) pip install sql:chemy

Answers: (s) (i) (t) (iii)

SOLVED QUESTIONS

33. Hitesh wants to display the last four rows of the dataframe df and has written the following code:

```
df.tail()
```

However, the last 5 rows are being displayed. Identify the error and rewrite the correct code so that only the last 4 rows get displayed. [CBSE Sample Paper 2020]

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

34. Name the functions you can use to iterate over dataframes.

Ans. iterrows( ) and iteritems( ).

35. What is the basic difference between iterrows( ) and iteritems( )?

Ans. <DF>.iteritems( ) iterates over vertical subsets in the form of (col-index, Series) pairs while <DF>.iterrows( ) iterates over horizontal subsets in the form of (row-index,Series) pairs.

36. Consider the dataframe df given as under:

	Emp_id	Name	Gender	Designation	City	Mobile
0	E01	Ritika	F	Manager	Delhi	9867585858
1	E02	Geet	M	Programmer	Mumbai	9988879990
2	E03	Rahul	M	IT Officer	Ahmedabad	7866557789
3	E04	Rijul	M	Analyst	Lucknow	9017363636
4	E05	Anita	F	Programmer	Bangalore	9818765544

Write a program to print only the Name, Gender and Designation for all rows.

Ans.
```
import pandas as pd
import numpy as np
emp = [['E01',"Ritika",'F','Manager',"Delhi","9867585858"],
       ['E02',"Geet",'F','Programmer',"Mumbai","9988879990"]]
df = pd.DataFrame(emp,columns=['Emp_id','Name','Gender','Designation',
                               'City','Mobile'])
print(df)
for i, row in df.iterrows():
    print(row["Name"],'\t',
        row["Gender"],'\t',
        row["Designation"])
```

37. Write a program to iterate over a dataframe containing names and marks, which then calculates grades as per marks (as per guidelines below) and adds them to the grade column

Marks >= 90	grade A+;	Marks 50-60	grade C;
Marks 70-90	grade A;	Marks 40-50	grade D;
Marks 60-70	grade B;	Marks <40	grade F;

Ans.
```
import pandas as pd
import numpy as np
Name = pd.Series(['Jiten', 'Manu', 'Ritika', 'Ajay'])
Marks = pd.Series([76.0, 56.0, 91.0, 67.0])
Stud = {'Name': Name, 'Marks': Marks}
df1 = pd.DataFrame(Stud, columns = ['Name', 'Marks'])
df1['Grade'] = np.NaN              #This will add NaN values to column Grade
print("Initial values in dataframe")
print(df1)
for (col,colSeries) in df1.iteritems():
  length = len(colSeries)          #number of entries in colSeries
  if col == 'Marks':
    lstMrks = []                   #initialize empty list
    for row in range(length):
      mrks = colSeries[row]
      if mrks >= 90:
        lstMrks.append('A+')       #grade appended to list lstMrks
      elif mrks >=70 and mrks <90:
        lstMrks.append('A')        #grade appended to list lstMrks
      elif  mrks >= 60 and mrks <70:
        lstMrks.append('B')        #grade appended to list lstMrks
      elif  mrks >= 50 and mrks <60:
        lstMrks.append('C')        #grade appended to list lstMrks
      elif  mrks >=40 and mrks <50:
        lstMrks.append('D')        #grade appended to list lstMrks
      else:
        lstMrks.append('F')        #grade appended to list lstMrks
df1['Grade'] = lstMrks             #This will put values of list lstMrks
                                    that contains computed grades to the
                                    column Grade.
print("\nDataFrame after calculating grades")
print(df1)
```

38. Write a Pandas program to get first n records of a DataFrame with three columns as col1, col2 and col3.

Sample Output:

```
Original DataFrame
   col1  col2  col3
0     1     4     7
1     2     5     5
2     3     6     8
3     4     9    12
4     7     5     1
5    11     0    11
First 3 rows of the said DataFrame:
   col1  col2  col3
0     1     4     7
1     2     5     5
2     3     6     8
```

Ans.
```
import pandas as pd
d = {'col1': [1, 2, 3, 4, 7, 11],
     'col2': [4, 5, 6, 9, 5, 0],
     'col3': [7, 5, 8, 12, 1,11]}
df = pd.DataFrame(d)
print("Original DataFrame")
print(df)
print("\nFirst 3 rows of the said DataFrame:")
df1 = df.head(3)
print(df1)
```

39. Write a Pandas program to get last n records of a DataFrame.

Sample Output:

```
Original DataFrame
   col1  col2  col3
0     1     4     7
1     2     5     5
2     3     6     8
3     4     9    12
4     7     5     1
5    11     0    11
Last 3 rows of the said DataFrame:
   col1  col2  col3
3     4     9    12
4     7     5     1
5    11     0    11
```

Ans.
```
import pandas as pd
d = {'col1': [1, 2, 3, 4, 7, 11],
     'col2': [4, 5, 6, 9, 5, 0],
     'col3': [7, 5, 8, 12, 1,11]}
df = pd.DataFrame(d)
print("Original DataFrame")
print(df)
print("\nLast 3 rows of the said DataFrame:")
df1 = df.tail(3)
print(df1)
```

40. What are the advantages of CSV file formats?

Ans. Advantages:

1. A simple, compact and ubiquitous format for data storage.
2. A common format for data interchange.
3. It can be opened in popular spreadsheet packages like MS Excel, Cal, etc.
4. Nearly all spreadsheets and databases support import/export to CSV format.

41. By default, read_csv( ) uses the values of the first row as column headers in dataframes. Which argument will you give to ensure that the top/first row's data is used as data and not as column headers?

Ans. header = none

For example:

```
mdf = pd.read_csv("Mydata.csv", header=none)
```

42. Explain briefly the CSV format of storing files.

Ans. The acronym CSV stands for Comma Separated Values and refers to tabular data saved as plain text where data values are separated by commas. In CSV format:

- Each row of the table is stored in one row, *i.e.*, the number of rows in a CSV file is equal to the number of rows in the table (or sheet or database table, etc.).
- The field values of a row are stored together with commas after every field value. However, after the last field's value, no comma is given; just the end of line.

43. Write a program that reads from a CSV file where the separator character is '$'. Read only the first 5 rows in your dataframe. Give column headings as ItemName, Quantity, Price. Make sure to read the first row as data and not as column headers.

Ans.
```
import pandas as pd
df = pd.read_csv("data.csv", sep="$",
                 names=["ItemName","Quantity","Price"],
                 header=None, nrows=5)
print(df)
```

44. You want to read data from a CSV file in a dataframe but you want to provide your own column names to the dataframe. What additional argument would you specify in read_csv( )?

Ans. Names, *for example:*
```
import pandas as pd
df = pd.read_csv("Employee.csv", usecols=['Name','Age','Salary'])
print(df)
```

Output:
```
      Name    Age    Salary
0   Ritesh   25.0   15000.0
1   Aakash   26.0   16000.0
2      NaN    NaN       NaN
3   Mahima   27.0   20000.0
4  Lakshay   23.0   18000.0
5     Manu   25.0   25000.0
6    Nidhi   26.0       NaN
7    Geetu   30.0   28000.0
```

45. Which argument would you give to read_csv( ) if you only want to read the top 5 rows of data?

Ans. nrows, *for example:*
```
import pandas as pd
df = pd.read_csv("Employee.csv", nrows=5)
print(df)
```

46. Write a command to store data of dataframe mdf into a CSV file Mydata.csv, with separate character as "@".

Ans.
```
mdf.to_csv("Mydata.csv", sep="@")
```

47. WAP to read details such as item, sales made in a dataframe and then store this data in a CSV file.

Ans.
```
import pandas as pd
d = {'Fridge':[12], 'Cooker':[5], 'Juicer':[15], 'Iron':[11]}
df = pd.DataFrame(d)
print(df)
df.to_csv("file.csv")
```

48. WAP to read data from a CSV file where separator character is "@". Make sure that the top row is used as data, not as column headers.

Ans.
```
import pandas as pd
df = pd.read_csv('read.csv', sep='@', header=None)
print(df)
```

49. Why do you need connection to an SQL database in order to get data from a table?

Ans. A connection works like a channel connecting Python and the database file through which data can travel to and from SQL database.

50. Write a program that lists only those records from SQL table Stud that have marks in the range 50-60.

Note. Table: Stud table is in MySQL database.

Ans.
```
import pandas as pd
import mysql-connector as conn
conn = mysql.connector.connect("new.db")
query = "select * from Stud where marks>=50 and marks <=60;"
df1 = pd.read_sql(query, conn)
```

The output produced by the above code is:
```
   RollNo      name   marks
0       1   Radhika    95.5
1       2   Shaurya    90.5
```

51. If query is a string storing an SQL statement, write statements so that the data is fetched based on query from SQL database Mydata.db.

Ans.
```
df = pd.read_sql("select * from stud", conn)
```

52. WAP to input the details of students through Python and put those values into student table of Mydata.db.

Ans.
```
import pandas as pd
import mysql-connector as conn
dict = {'Name':['Rinku','Rohit','Vikram'],
        'Marks':[56,78,89]}
df = pd.DataFrame(dict)
print(df)
conn = mysql.connector.connect(host="localhost",user="root",
                               password='',database="Mydata.db")
df.to_sql("Class",conn)
```

UNSOLVED QUESTIONS

43. Name the function to iterate over a DataFrame horizontally.

44. Name the function to iterate over a DataFrame vertically.

45. How do you iterate over a DataFrame? Explain with the help of code snippet.

46. Write a program to iterate and print a DataFrame one column at a time and print only first three columns.

47. Write a program to iterate and print a DataFrame one row at a time and print only first five rows.

48. Create a DataFrame from two series – Name and Grade, Name and Marks of five students.
(a) Display the first three records from student dataframe.
(b) Display the last two records from student dataframe.

49. Create a DataFrame of dictionary consisting of Name, Sub1, Sub2, Sub3, Sub4, Sub5 of five students.
(a) Display the DataFrame.
(b) Display the first 5 rows and bottom 3 rows of student dataframe.

50. What is CSV file?

51. What is the use of nrows?

52. How can we create CSV file? Explain with an example.

53. Create a CSV file with default index.

54. What do you mean by commit?

55. What is the purpose of writing "cursor.execute()"?

56. Write a program that reads students marks from a result CSV and displays percentage of each student.

57. Write the name of function to store data from a dataframe into a CSV file.

58. How can we import specific columns from a CSV file?

59. What are the advantages of CSV file formats?

60. What all libraries do you require in order to bring data from a CSV file into a dataframe?

61. You want to read data from a CSV file in a dataframe but you want to provide your own column names to the dataframe. What additional argument would you specify in read_csv()?

62. By default, read_csv() uses the value of first row as column headers in dataframes. Which argument will you give to ensure that the top/first row's data is used as data and not as column headers?

63. Which argument would you give to read.csv() if you only want to read top 10 rows of data?

64. Write a program to create two dataframes with the following data:

df1

Emp_code	Name
110	Taksh
112	Jeet Arora
114	Shubham Jain

df2

Emp_code	Name	Salary
110	Taksh	45000
112	Jeet Arora	56000
114	Shubham Jain	55000

Store these two dataframes as two separate table files inside the same database.

65. Why do you need connection to an SQL database in order to get data from a table?

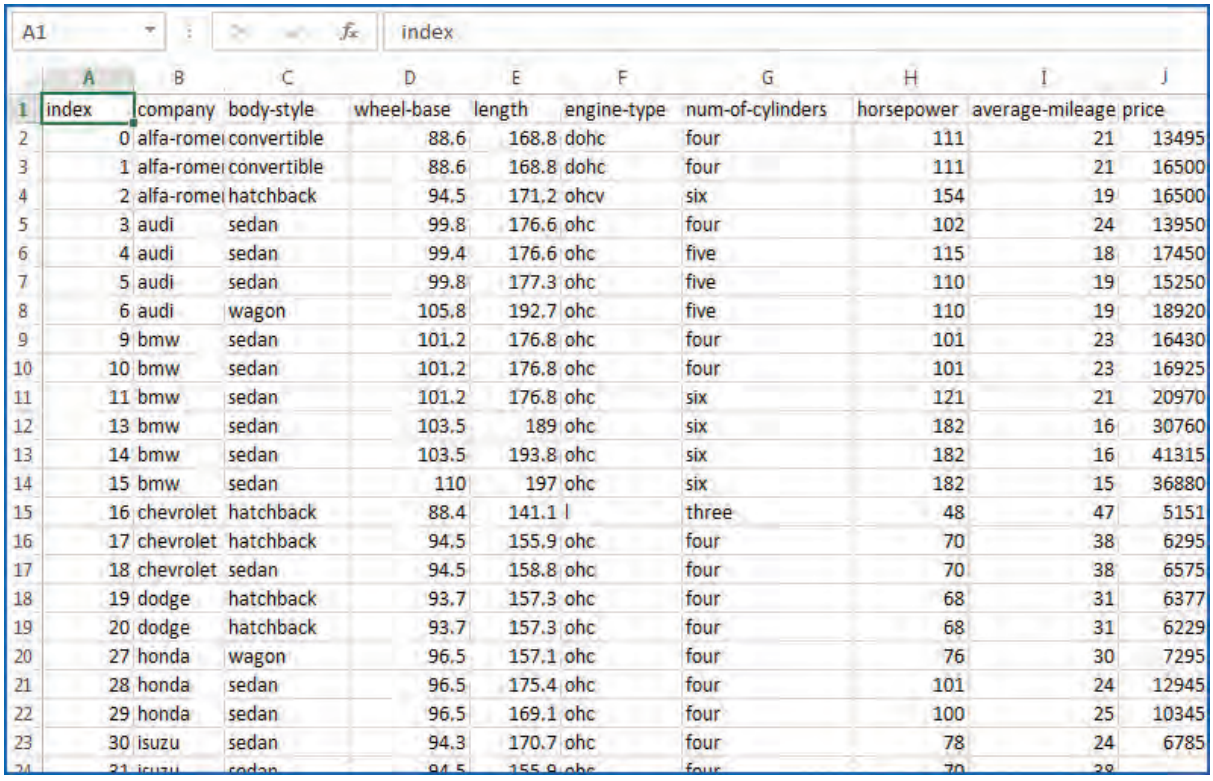
66. What all libraries do you require in order to interact with databases (and dataframe) from within Python?

CASE-BASED/SOURCE-BASED INTEGRATED QUESTIONS

1. Competent Automobiles is a company that is a large manufacturer of automobiles. It manufactures cars of different makes. Consider the given automobile dataset for performing various operations using Pandas dataframes and CSV files.

This Automobile dataset has different characteristics of an automobile such as body-style, wheel-base, engine-type, price, mileage, horsepower, etc.

Automobile dataset



A1 : index

	A	B	C	D	E	F	G	H	I	J
1	index	company	body-style	wheel-base	length	engine-type	num-of-cylinders	horsepower	average-mileage	price
2	0	alfa-rome	convertible	88.6	168.8	dohc	four	111	21	13495
3	1	alfa-rome	convertible	88.6	168.8	dohc	four	111	21	16500
4	2	alfa-rome	hatchback	94.5	171.2	ohcv	six	154	19	16500
5	3	audi	sedan	99.8	176.6	ohc	four	102	24	13950
6	4	audi	sedan	99.4	176.6	ohc	five	115	18	17450
7	5	audi	sedan	99.8	177.3	ohc	five	110	19	15250
8	6	audi	wagon	105.8	192.7	ohc	five	110	19	18920
9	9	bmw	sedan	101.2	176.8	ohc	four	101	23	16430
10	10	bmw	sedan	101.2	176.8	ohc	four	101	23	16925
11	11	bmw	sedan	101.2	176.8	ohc	six	121	21	20970
12	13	bmw	sedan	103.5	189	ohc	six	182	16	30760
13	14	bmw	sedan	103.5	193.8	ohc	six	182	16	41315
14	15	bmw	sedan	110	197	ohc	six	182	15	36880
15	16	chevrolet	hatchback	88.4	141.1	l	three	48	47	5151
16	17	chevrolet	hatchback	94.5	155.9	ohc	four	70	38	6295
17	18	chevrolet	sedan	94.5	158.8	ohc	four	70	38	6575
18	19	dodge	hatchback	93.7	157.3	ohc	four	68	31	6377
19	20	dodge	hatchback	93.7	157.3	ohc	four	68	31	6229
20	27	honda	wagon	96.5	157.1	ohc	four	76	30	7295
21	28	honda	sedan	96.5	175.4	ohc	four	101	24	12945
22	29	honda	sedan	96.5	169.1	ohc	four	100	25	10345
23	30	isuzu	sedan	94.3	170.7	ohc	four	78	24	6785
24	31	isuzu	sedan	94.5	155.9	ohc	four	70	38	

From the given dataset, print the first and last five records of automobiles of different makes.

Ans. Printing first five rows:

```
import pandas as pd
df = pd.read_csv("D:/preeti/automobiledataset/Automobile_data.csv")
df.head(5)
```

Printing last five rows:

```
import pandas as pd
df = pd.read_csv("D:/preeti/automobiledataset/Automobile_data.csv")
df.tail(5)
```

2. Create two dataframes using the following two Dictionaries. Merge the two dataframes and append the second dataframe as a new column to the first dataframe on the basis of the manufacturing company's name.

```
Car_Price = {'Company': ['Toyota', 'Honda', 'BMW', 'Audi'],
             'Price': [23845, 17995, 135925 , 71400]}
car_Horsepower = {'Company': ['Toyota', 'Honda', 'BMW', 'Audi'],
                  'horsepower': [141, 80, 182 , 160]}
```

Expected Output:

	Company	Price	horsepower
0	Toyota	23845	141
1	Honda	17995	80
2	BMW	135925	182
3	Audi	71400	160

Python Pandas merges two dataframes and appends the new dataframe as a new column.

Ans.
```
import pandas as pd
Car_Price = {'Company': ['Toyota', 'Honda', 'BMW', 'Audi'],
             'Price': [23845, 17995, 135925 , 71400]}
carPriceDf = pd.DataFrame.from_dict(Car_Price)

car_Horsepower = {'Company': ['Toyota', 'Honda', 'BMW', 'Audi'],
                  'horsepower': [141, 80, 182, 160]}
carsHorsepowerDf = pd.DataFrame.from_dict(car_Horsepower)

carsDf = pd.merge(carPriceDf, carsHorsepowerDf, on="Company")
print(carsDf)
```

3. Consider the same scenario as described in Question 1.

Competent Automobiles is a company that is a large manufacturer of automobiles. It manufactures cars of different makes. Consider the given automobile dataset for performing various operations using Pandas dataframes and CSV files.

This Automobile dataset has different characteristics of an automobile such as body-style, wheel-base, engine-type, price, mileage, horsepower, etc.

Automobile dataset

A1 | index

	A	B	C	D	E	F	G	H	I	J
1	index	company	body-style	wheel-base	length	engine-type	num-of-cylinders	horsepower	average-mileage	price
2	0	alfa-romei	convertible	88.6	168.8	dohc	four	111	21	13495
3	1	alfa-romei	convertible	88.6	168.8	dohc	four	111	21	16500
4	2	alfa-romei	hatchback	94.5	171.2	ohcv	six	154	19	16500
5	3	audi	sedan	99.8	176.6	ohc	four	102	24	13950
6	4	audi	sedan	99.4	176.6	ohc	five	115	18	17450
7	5	audi	sedan	99.8	177.3	ohc	five	110	19	15250
8	6	audi	wagon	105.8	192.7	ohc	five	110	19	18920
9	9	bmw	sedan	101.2	176.8	ohc	four	101	23	16430
10	10	bmw	sedan	101.2	176.8	ohc	four	101	23	16925
11	11	bmw	sedan	101.2	176.8	ohc	six	121	21	20970
12	13	bmw	sedan	103.5	189	ohc	six	182	16	30760
13	14	bmw	sedan	103.5	193.8	ohc	six	182	16	41315
14	15	bmw	sedan	110	197	ohc	six	182	15	36880
15	16	chevrolet	hatchback	88.4	141.1	l	three	48	47	5151
16	17	chevrolet	hatchback	94.5	155.9	ohc	four	70	38	6295
17	18	chevrolet	sedan	94.5	158.8	ohc	four	70	38	6575
18	19	dodge	hatchback	93.7	157.3	ohc	four	68	31	6377
19	20	dodge	hatchback	93.7	157.3	ohc	four	68	31	6229
20	27	honda	wagon	96.5	157.1	ohc	four	76	30	7295
21	28	honda	sedan	96.5	175.4	ohc	four	101	24	12945
22	29	honda	sedan	96.5	169.1	ohc	four	100	25	10345
23	30	isuzu	sedan	94.3	170.7	ohc	four	78	24	6785
24	31	isuzu	sedan	94.5	155.9	ohc	four	70	38	

(a) Write the statement in Pandas to print the most expensive car's company name and price.

Expected Output:

	company	price
35	mercedes-benz	45400.0

Ans.
```
import pandas as pd
df = pd.read_csv("Automobile_data.csv ")
df = df[['company','price']][df.price==df['price'].max()]
print(df)
```

(b) Write a program to print details of all Toyota cars.

Ans.
```
import pandas as pd
df = pd.read_csv("Automobile_data.csv")
car_Manufacturers = df.groupby('company')
toyotaDf = car_Manufacturers.get_group('toyota')
print(toyotaDf)
```

(c) Write a program to find each company's highest priced car.

Ans.
```
import pandas as pd
df = pd.read_csv("Automobile_data.csv ")
car_Manufacturers = df.groupby('company')
priceDf = car_Manufacturers['company','price'].max()
print(priceDf)
```

(d) Write a program to sort all cars on the basis of the Price column.

Ans.
```
import pandas as pd
carsDf = pd.read_csv("Automobile_data.csv ")
carsDf = carsDf.sort_values(by=['price', 'horsepower'], ascending=False)
print(carsDf)
```

5 Computer Networks
(As per latest Syllabus)

5.1 INTRODUCTION

The greatest breakthrough in technology and communication over the past 20 years has been the development and advancement of the computer network. From emailing a friend, to online bill payment, to downloading data from the internet, to e-commerce, networking has made our world much smaller and forever changed the way we communicate.

Network provides salient features which have made our life easy and comfortable, be it sending an email, withdrawing money from an ATM machine, online railway or airline reservation, or sharing audio and video files. Apart from these, the most extensively-used feature is the Print command sent from a computer to get a printout from a printer attached to some other computer. All this involves a network.

It is the network that connects various computers to each other and handles a large volume of data.

Fig. 5.1: A Computer Network

5.2 COMPUTER NETWORK—A BRIEF OVERVIEW

Several devices connected to each other for reliable communication/transfer of data constitute a network. A network can consist of a computer, a fax machine, a printer, a camera, a cell phone, etc. A collection of interconnected computers is called a **Computer Network**. Two computers or devices are said to be interconnected if they are capable of sharing and exchanging information with each other by following a protocol (set of rules).

CTM: A computer network is a collection of interconnected computers and other devices to share data and other resources (hardware and software resources).

5.2.1 Advantages of Computer Networks

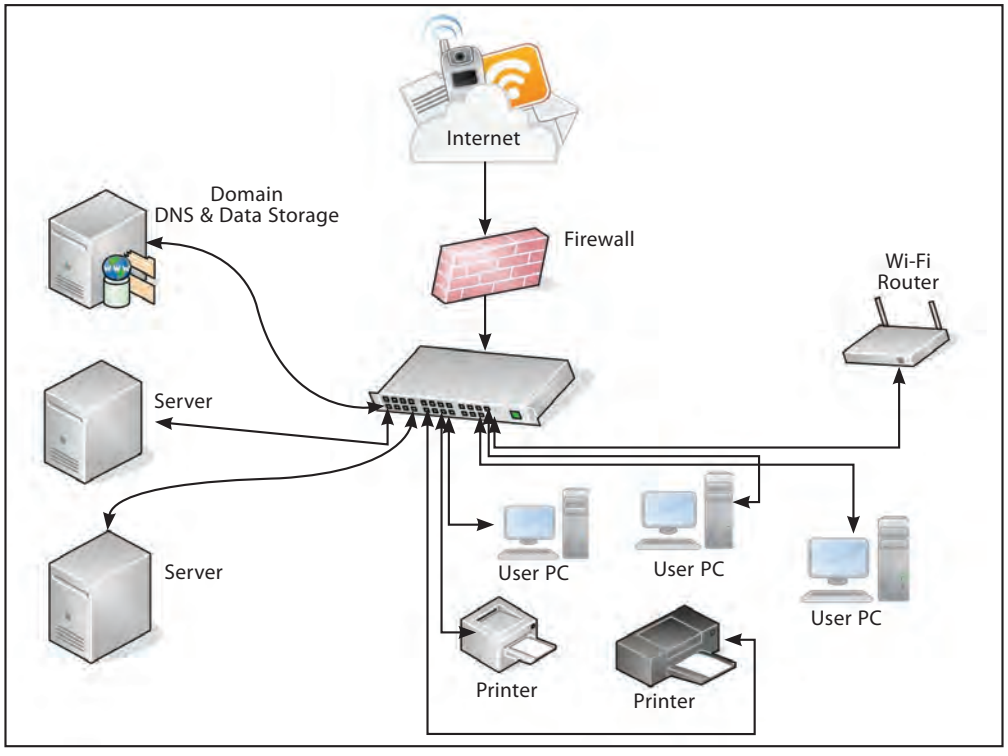


Fig. 5.2: The Network Diagram

Let us now discuss why networks are essential. Are there any advantages of networked computers over stand-alone machines? Yes, networked systems are far better. A network uses a distributed processing system in which a task is divided among several devices which are interconnected with each other. Therefore, instead of a single computer being responsible for completing the entire task, all the interconnected computers are responsible for completing the task assigned to them. This leads to better performance with high processing speed.

Networks have several advantages which are described below:

(a) **Resource Sharing:** The primary use of a network is to share among users programs/ applications, data and peripheral devices connected to the network, irrespective of their physical location. You must have noticed in your networked computer labs that when a print command is given on one computer, the document is printed by a printer attached to some other system. This allows printing of documents by several users and, hence, the printer is shared by multiple users on the network. Other resources like hard disk, DVD drive, scanner, etc., can also be shared on a computer network. *For example*, sharing database, audio and video files, antivirus software, application software, printers and scanners, etc.

(b) **Improved Communication:** A computer network enables fast, reliable and secure communication between users. It saves time and offers easy communication methods. *For example*, in an organization, managers work at different locations to make financial reports. While working on a network, any change made by one manager on his/her computer can easily be seen by other managers and employees. Thus, a network allows managers to easily update information. This increases their efficiency and allows them to complete their work quickly.

(c) **Reduced Communication Cost:** Sharing resources also reduces communication cost. Using public networks, we can send a large quantity of data at a low cost. Internet and mobile networks are playing a very important role in sending and receiving text, image, audio and video data at a low cost.

(d) **Reliability of Data:** Reliability means backing up of data, *i.e.*, data can be copied and stored on multiple computers. In a network system, all computers are connected to each

other. Thus, the information or message which is shared by each device is stored on their respective workstations (computers). If, due to some reason (hardware crash, etc.), the data gets corrupted and, thus, becomes unavailable on one computer, a copy of the same data can be accessed from another workstation for future use. This leads to smooth functioning and further processing without disruption.

(e) **Central Storage of Data:** Files can be stored on a central node (the file server) that can be shared and made available to each and every user in an organization. With centralized processing, data is stored and retrieved from a single central location. Thus, there is no duplication of data and almost no data redundancy.

5.3 EVOLUTION OF NETWORK

The network did not evolve in a single day; rather, it took decades to become more powerful, efficient and reliable. The network has passed through several stages which are described below:

- **ARPANET (Advanced Research Project Agency Network):** ARPANET, which was jointly designed and named by the Advanced Research Projects Agency (ARPA) and US Department of Defence (DoD), was the first network and came into existence in 1969. It was a project that connected a handful of computers at different universities and US DoD for sharing of data and messages and playing long-distance games, and socializing with people to share their views.
- **NSFNET (National Science Federation Network):** In the mid-80's, another federal agency, NSFNET (National Science Federation Network), created a new network which was more efficient than ARPANET. Its main aim was to use network only for academic research and not for any private business activity. Later, many private companies combined their own private networks with ARPANET and NSFNET to make a more capable and broad network—the internet. It is the internet that links two or more networks to make a large network for sharing information and messages.

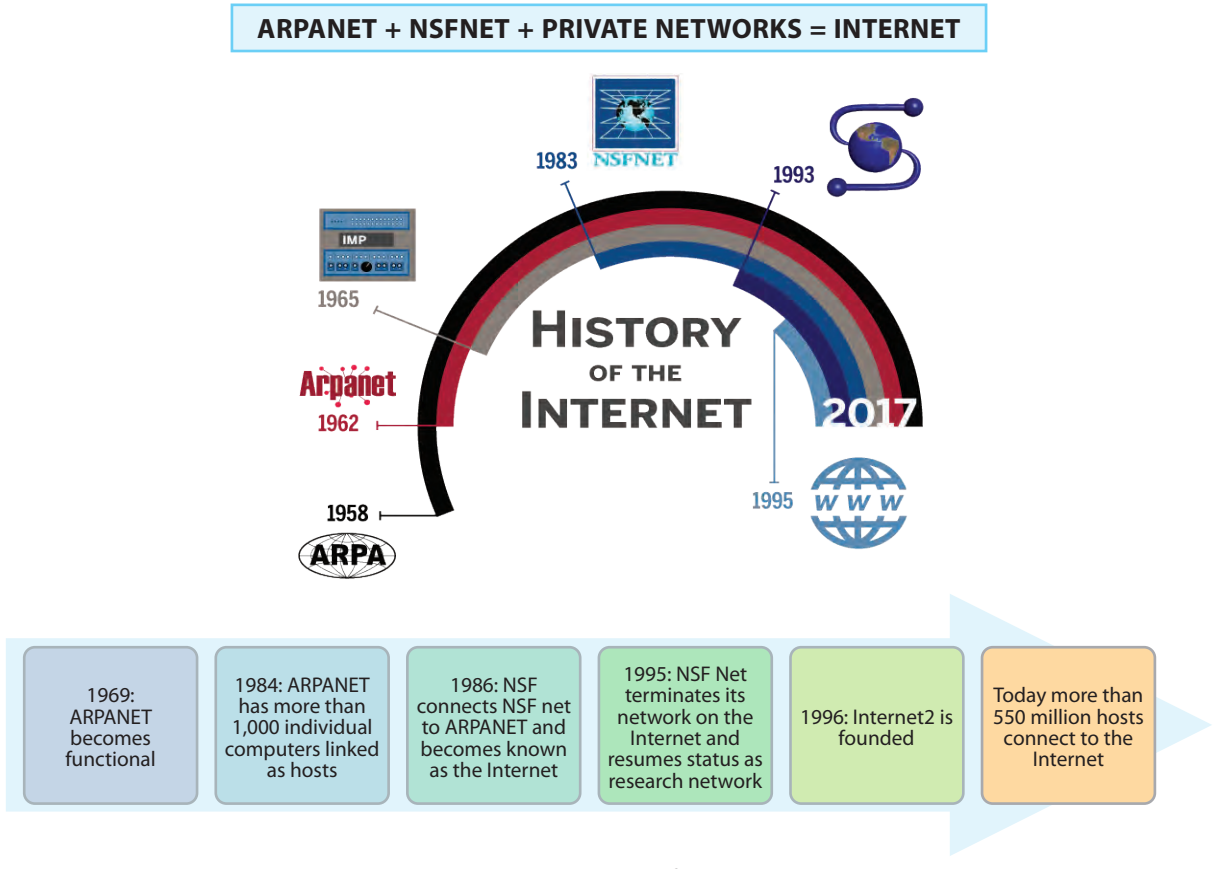


Fig. 5.3: Evolution of Internet

- **Internet:** In the 1990's, internet, which is a network of networks, came into existence. The internet has evolved from ARPANET. The computers are connected through World Wide Web that comprises a large network and shares a common communication protocol (Transmission Control Protocol-Internet Protocol, TCP/IP). It allows computers of different types to exchange information and is known as internet. Millions of domestic, business and government networks are connected to each other for the purpose of sharing files, data, email, etc. Most of the computers are not connected directly to the internet. Instead, they are connected to smaller networks which are further connected to a backbone network through gateways.

CTM: Network of networks makes the internet.

- **Interspace:** Interspace is a software that allows multiple users in a client-server environment to communicate with each other by sending and receiving data of various types such as data files, video, audio and textual data in a 3-D environment. It facilitates online real-time exchange of data. Interspace is the most advanced term of communication available on the internet today.

5.4 HOW DOES INTERNET WORK

One of the greatest things about the internet is that nobody really owns it. It is a global collection of networks, both big and small. These networks connect together in many different ways to form the single entity that we know as internet. In fact, the very name comes from this idea of interconnected networks.

Since its beginning in 1969, the internet has grown from four host computer systems to tens of thousands. However, just because nobody owns the internet does not mean that it is not monitored and maintained. The Internet Society, a non-profit group established in 1992, oversees the formation of the policies and protocols that define how we use and interact with the internet.

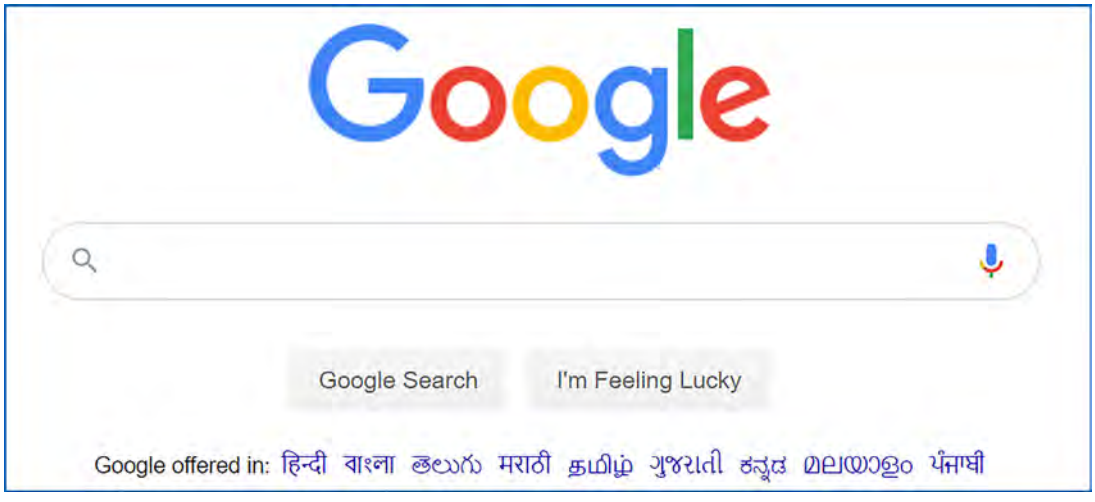


Fig. 5.4(a): Working of the internet

Before we learn about the basic underlying structure of the internet, *e.g.*, domain name servers, network access points and backbones, we first need to understand how our computer connects to others.

Every computer that is connected to the internet is part of a network, even the one in our home. *For example*, we may use a modem and dial a local number to connect to an **Internet Service Provider** (ISP). At work, a computer may be part of a **Local Area Network** (LAN), but it most likely still connects to the internet using an ISP that the company has contracted with.

When it connects to the ISP, it becomes part of their network. The ISP may then connect to a larger network and become part of that network. The internet is simply a network of networks.

Most large communication companies have their own dedicated backbones connecting various regions. In each region, the company has a **Point of Presence** (POP). The POP is a place for local users to access the company's network, often through a local phone number or dedicated line. The amazing thing here is that there is no overall controlling network. Instead, there are several high-level networks connecting to each other through **Network Access Points** or NAPs.

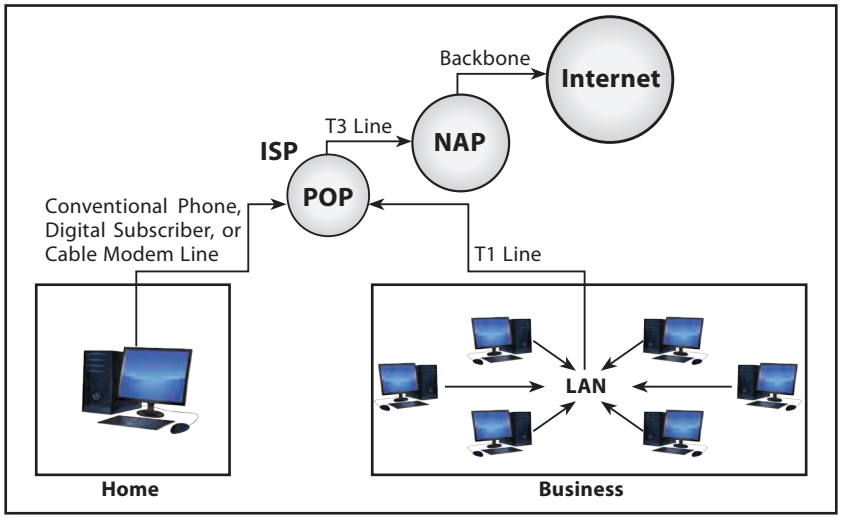


Fig. 5.4(b): POP and NAP

Gateway: Gateway is a device that connects dissimilar networks. A backbone is a central interconnecting structure that connects one or more networks just like the trunk of a tree.

At the source computer, the message to be sent is broken down into small parts called packets. Each packet is given a serial number, *e.g.*, 1, 2, 3. All these packets are sent to the destination computer. The destination computer receives the packets in random order (10 may come before 1). The packets are reassembled in the order of their number and message is restored.

How it functions smoothly: Every computer connected to the internet uses the same set of rules for communication. A set of rules is called protocol. Communication protocol used by internet is TCP/IP. The TCP (Transmission Control Protocol) part is responsible for dividing the message into packets on the source computer and reassembling them at the destination computer. The IP (Internet Protocol) is responsible for handling the address of the destination computer so that the packet is sent to its proper destination.

5.4.1 Elementary Terminology of Networks

1. **Nodes (Workstations):** The term node refers to computers that are attached to a network and are seeking to share resources.
2. **Server:** A computer that facilitates the sharing of data, software and hardware resources on the network.
3. **Network Interface Unit (NIU) (MAC Address):** A network interface unit is an interpreter that helps in establishing communication between the server and the client.
4. **IP Address:** Every machine on a TCP bar/IP Network has a unique identifying number called an IP Address.
5. **Domain Name:** It is a way to identify and locate the computers connected to the internet. It must be unique.

5.5 TYPES OF NETWORKS

A computer network may be small or big depending upon the number of computers and other network devices linked together. Thus, networks vary in size, complexity and geographical spread. A computer network can be on a table, in a room, building, city, country, across continents or around the world.

On the basis of geographical spread, networks may be classified as:

1. PAN
2. LAN
3. MAN
4. WAN

5.5.1 Personal Area Network (PAN)

PANs are small networks used to establish communication between a computer and other handheld devices in the proximity of up to 10 metres using wired USB connectivity or wireless systems like Bluetooth or Infrared. PANs are used to connect computers, laptops, mobiles and other IT-enabled devices to transfer files including emails, digital photos, audio and video, etc. The Bluetooth technology implements PAN. PAN may include wireless computer keyboard and mouse, Bluetooth-enabled headphones, wireless printers and TV remotes.

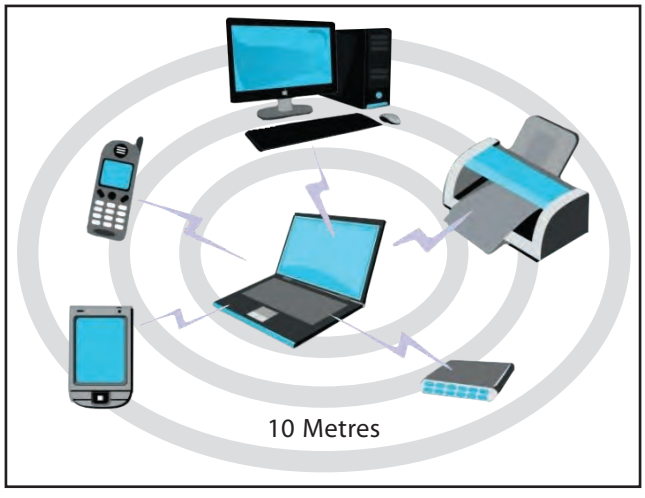


Fig. 5.5: Personal Area Network (PAN)

CTM: The network that belongs to a single person or user is known as PAN.

5.5.2 Local Area Network (LAN)

LAN is a privately owned computer network covering a small geographical area (small physical area), like a home, office or a building such as a school. It can cover an area spread over a few metres to a radius of a few kilometres.

Occasionally, a LAN can span a group of nearby buildings. In addition to operating in a limited space, a LAN is owned, controlled and managed by a single person or organization. A LAN can be set up using wired media (UTP cables, coaxial cables, etc.) or wireless media (Infrared, Radio waves).

If a LAN is set up using unguided media, it is known as WLAN (wireless LAN). The key purpose of a LAN is to share resources. LAN users can share data, programs, printer, disk, modem, etc.

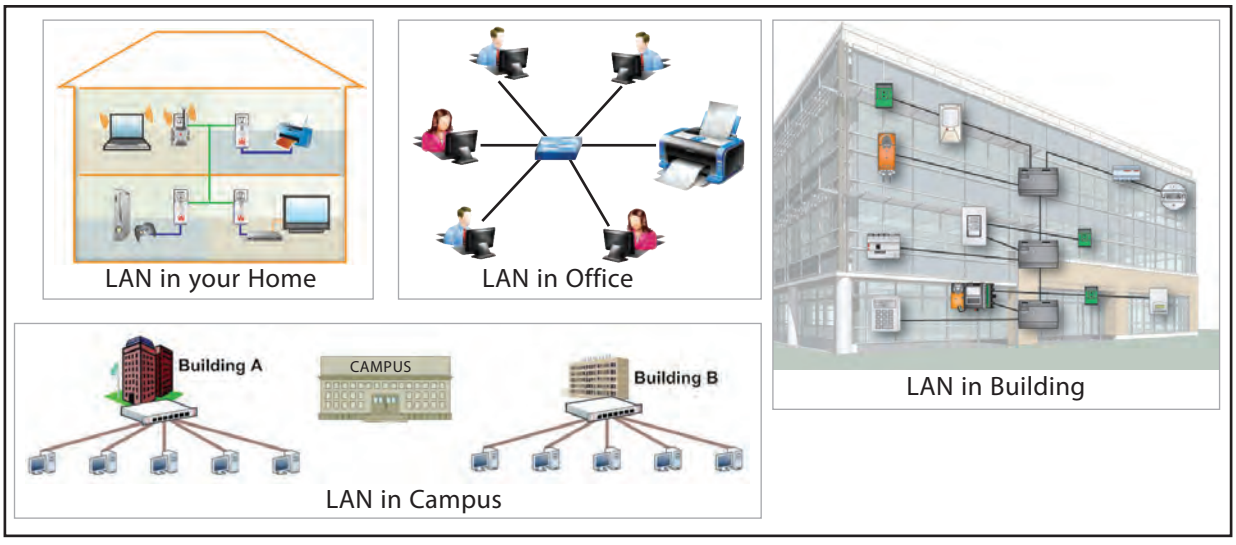


Fig. 5.6: Local Area Network (LAN)

Data transfer rate speed over a Local Area Network can vary from 10 mbps to 1 gbps.

5.5.3 Metropolitan Area Network (MAN)

MAN is larger than a LAN and can cover a city and its surrounding areas. A MAN usually interconnects a number of LANs and individual computers. It also shares the computing resources among users. All types of communication media (guided and unguided) are used to set up a MAN. A MAN is typically owned and operated by a single entity such as a government body or a large corporation. A good example of MAN is the interconnected offices of a Multinational Corporation (MNC) or cable television networks available in the whole city.

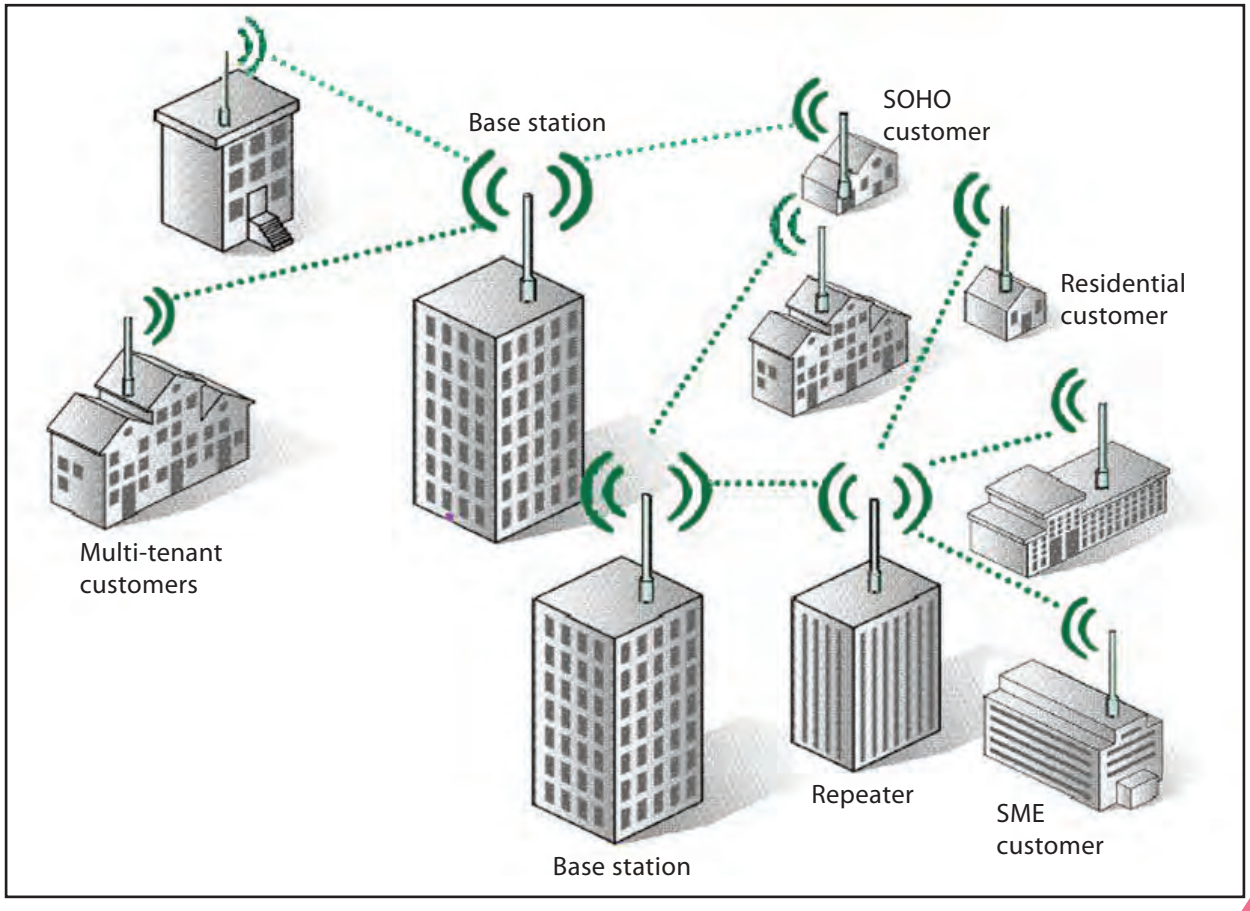


Fig. 5.7: Metropolitan Area Network (MAN)

5.5.4 Wide Area Network (WAN)

A WAN is a telecommunication network. This type of network spreads over a large geographical area across countries and continents. WANs are generally used to interconnect several other types of networks such as LANs, MANs, etc. They facilitate fast and efficient exchange of information at high speed and low cost. A WAN uses common carriers like satellite systems, telephone lines, etc.

A WAN can cover an area with a radius spanning hundreds of kilometres. A network of ATMs, banks, government offices, international organizations' offices, etc., spread over a country, continent or covering many continents are examples of WAN.

All types of communication media (guided and unguided) are used to set up a WAN. The best known example of a WAN is the internet. The internet is the largest WAN spanning the entire planet.

CTM: A WAN interconnects all the computers across the world.

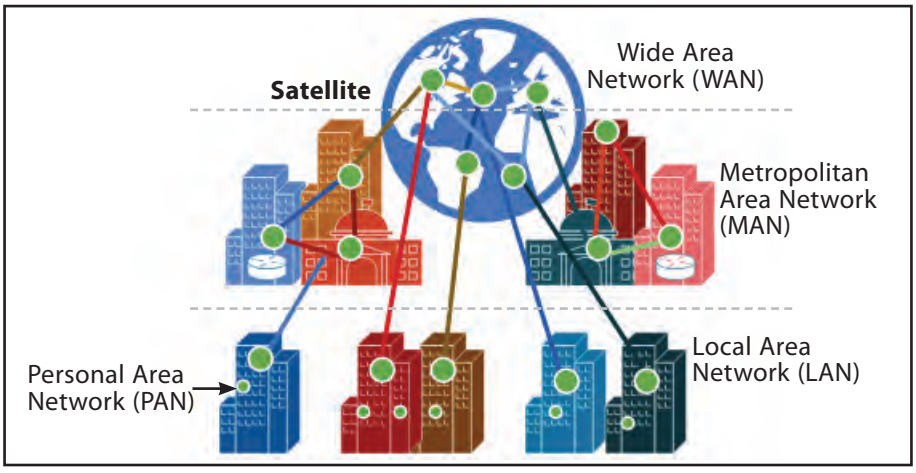


Fig. 5.8: Wide Area Network (WAN)

The following table summarizes the characteristics of PANs, LANs, MANs and WANs.

Parameter	PAN	LAN	MAN	WAN
Area covered	Small area (up to 10 m radius)	A building or campus (up to 10 km)	A city (up to 100 km radius)	Entire country, continent or globe
Networking cost	Negligible	Inexpensive	Expensive	Very expensive
Transmission speed	High speed	High speed	Moderate speed	Low speed
Error rate	Lowest	Lowest	Moderate	Highest
Network devices used	WLAN, USB Dongle	LAN/WLAN, Hub/ Switch, Repeater, Modem	Router, Gateway	Router, Gateway
Technology/ media used	Infrared, Bluetooth	Ethernet, Wi-Fi	Optical fibre, Radio wave, Microwave	Microwave Satellite

CTM: LAN and WAN are the two primary and best-known categories of area networks; the others have emerged with technological advances.

Comparing LAN and WAN

As we have seen, computer networks can be classified into LAN, MAN and WAN categories on the basis of their geographical domains. A WAN extends over a large geographical area, such as states or countries. A LAN is confined to relatively smaller areas, such as an office, a building, etc. A MAN usually covers an entire city. It uses the LAN technology. The most common example of MAN is the cable television network. Thus, the basic points of difference between LAN and WAN are as follows:

1. The distance between the nodes in a LAN connection is limited to a specific range. The upper limit is approximately 10 kilometres and the lower limit is one metre. On the other hand, WANs are spread across thousands of kilometres in different countries or regions.
2. LANs operate between speeds of 1 mega bit per second (mbps) and 10 mbps while WANs operate at speeds of less than 1 mbps. To achieve speeds of several hundred mbps, it is advisable to use the optical fibre technology.
3. The error rate in LANs is lower than that in WANs because of the short distances involved in LANs. The error rate in LANs is approximately one thousand times less than that in WANs.
4. As LANs are limited by distance, an entire LAN is usually under the control of a single entity, such as an organization. On the other hand, WANs are usually operated and controlled by multiple organizations.

Thus, we can say that in comparison to WANs, LANs cover a limited area but they operate with high speed and low error rates.

5.6 NETWORK DEVICES

1. **Modem:** A MODEM (Modulator DEModulator) is an electronic device that enables a computer to transmit data over telephone lines. It is a device used to convert digital signals into analog signals and vice versa. There are two types of modems, namely internal modem and external modem.
2. **RJ-45 Connector:** RJ-45 is a standard type of connector for network cables. The RJ-45 (Registered Jack) connectors are the plug-in devices used in networking and telecommunication applications. They are used primarily for connecting LANs, particularly Ethernet.

CTM: RJ-45 is a short term for Registered Jack-45. It is an eight-wire connector used to connect computers on LANs, especially Ethernets.

3. **Ethernet Card:** It is a hardware device that helps in the connection of nodes within a network. Ethernet card is also known as a network card, network adapter or NIC (network interface card). It is a card that allows computers to communicate over a computer network. On Ethernet card, a physical address of each communicating computer is mentioned. Physical address is known as MAC address.
4. **Hub:** It is multi-port and unintelligent network device which simply transfers data from one port of the network to another. A hub is a hardware device used to connect several computers together with different ports. When the packet reaches one port, it is copied to all other ports of the hub without changing the destination address in the frame. Rather, it simply copies the data to all of the nodes connected to the hub.

   Hubs can be either active or passive. Hubs can usually support 8, 12 or 24 RJ-45 ports.

Fig. 5.9: Hub

But the problem with hub is that it is not an intelligent device. It shares bandwidth with all the attached devices and broadcasts the data, *i.e.*, sends the data frames to all the connected nodes as it does not remember devices/computers connected to it. Also, it cannot filter the data and causes unnecessary traffic jams.

A hub can both send as well as receive information but only one task at a time. However, a hub is an inexpensive way to connect multiple nodes/devices to network.

CTM: Hub is a device used to connect several computers with each other.

5. **Switch:** A switch (switching hub) is a network device which is used to interconnect computers or devices on a network. It filters and forwards data packets across a network. It is also a multi-port device but with some intelligence and so the data packets received from one port of network are refreshed and delivered to the other port of the network. The main difference between hub and switch is that hub replicates what it receives on one port onto all the other ports, while switch keeps a record of the MAC addresses of the devices attached to it.

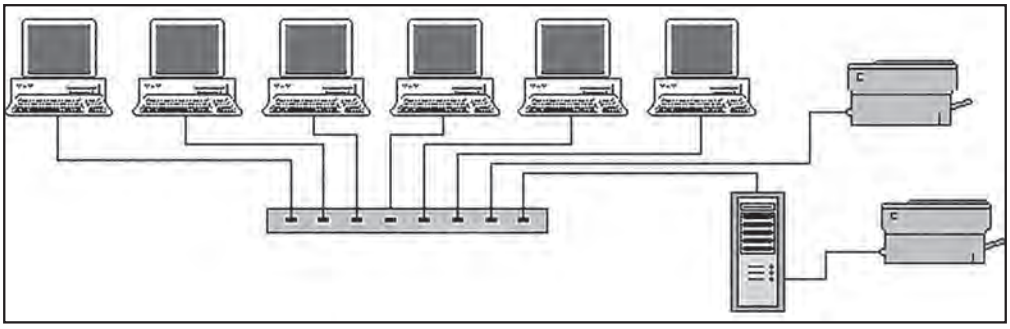


Fig. 5.10: Switch

CTM: A switch is a device that transmits data to the computers in a LAN.

6. **Bridge:** A bridge is a device that works on the physical layer as well as on data link layer. A network bridge connects multiple network segments at the data link layer (layer 2) of the OSI model. Bridges relay frames between two originally separate segments. When a frame enters a bridge, the bridge not only regenerates the signal but also checks the physical address of the destination and forwards the new copy only to that port.

An important advantage of using a bridge is that it is a smarter hub as it can filter network traffic on the basis of the MAC addresses.

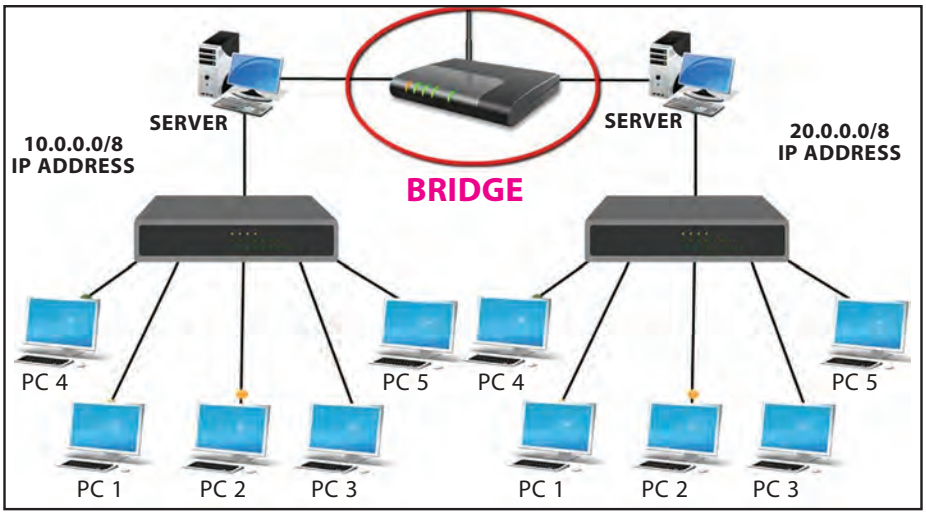


Fig. 5.11: Ethernet Bridge

CTM: A bridge is a device that links two segments together of the original network.

7. **Gateway:** A gateway is a device that connects dissimilar networks. In internet, several networks are communicating with each other and each network has a different configuration. In order to make reliable communication, there must be a device that helps in communicating.

   Gateway is a device which establishes an intelligent connection between a local area network and external networks with completely different structures.

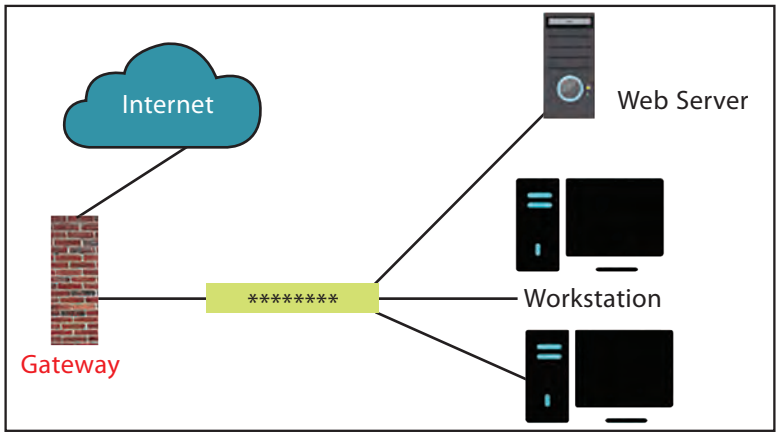


Fig. 5.12: Gateway

CTM: A gateway is a device that connects dissimilar networks.

8. **Repeater:** A repeater is a device that operates only on the physical layer of the OSI model. As a signal travels a fixed distance, before attenuation of the signal, a repeater is used which amplifies and restores signals for long-distance transmission. A repeater is an electronic device that receives a signal before it becomes too weak and regenerates the original signal. Also, it is a two-port network device that strengthens the signal intensity and connects two identical networks. In most twisted pair Ethernet configurations, repeaters are required for cable runs longer than 100 metres. A repeater does not change the functionality of the network; instead, it makes the signal strong before it degrades.

   Repeaters are also extensively used in broadcasting where they are termed as translators or boosters.

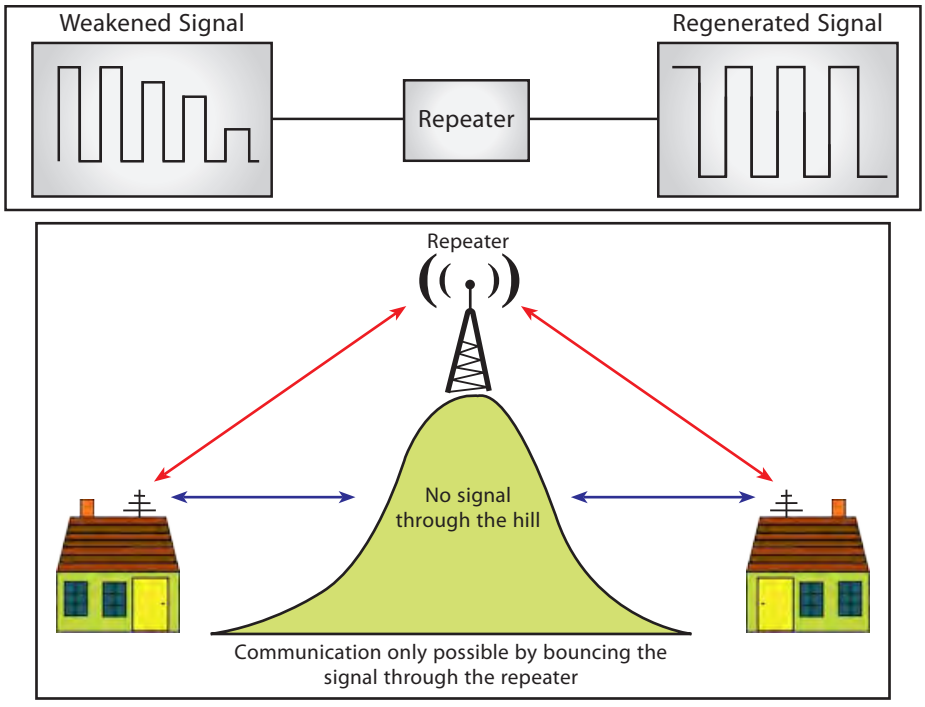


Fig. 5.13: Working of a Repeater

CTM: Repeater is a device that amplifies a signal that is transmitted across the network so that the signal is received in the same way as it is sent.

9. **Router:** A router is a networking device that forwards data packets from the source machine to the destination machine by using the shortest path. Routers are used at the network layer, which is the third layer of the OSI model.

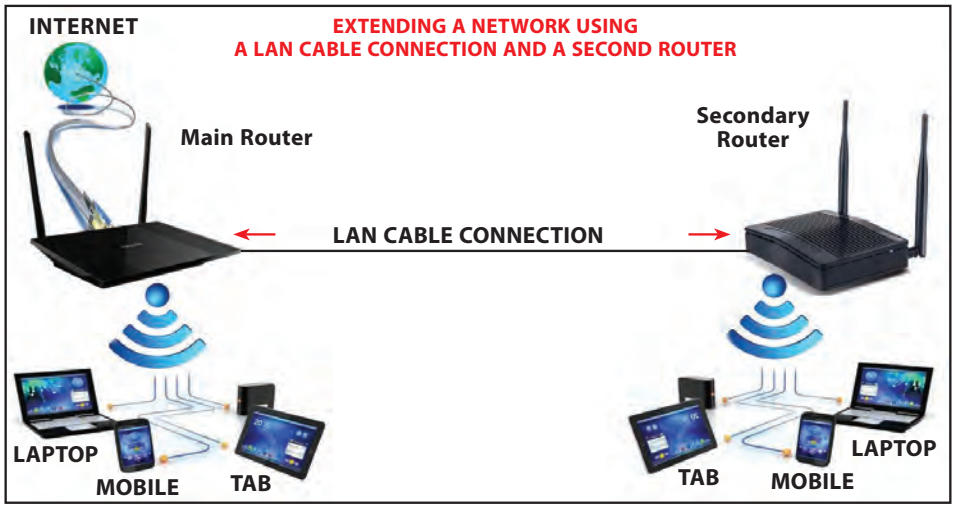


Fig. 5.14: Router

CTM: A router is a networking device that helps in forwarding packets from one machine to another.

10. **Wi-Fi Card:** A Wi-Fi card is either an internal or external Local Area Network adapter with a built-in wireless radio and antenna. A Wi-Fi card is used in a desktop computer that enables a user to establish an internet connection. Wi-Fi cards are known as wireless fidelity cards as they allow the user to set up connection without any wire. Wireless Fidelity (Wi-Fi) cards are widely used in notebook computers due to their highly portable nature. The most common Wi-Fi cards used in desktop computers are PCI-Express Wi-Fi cards made to fit the PCI-Express card slots on the motherboard.

5.7 NETWORK TOPOLOGIES

Topology is a way of connecting devices with each other either physically or logically. Two or more devices make a link and two or more links form a topology. It is basically a geometrical representation of how a network is laid out.

CTM: Topology is a way of connecting several devices with each other on a network.

Types of Topologies

Basically, there are five types of topologies and each topology has some advantages and disadvantages.

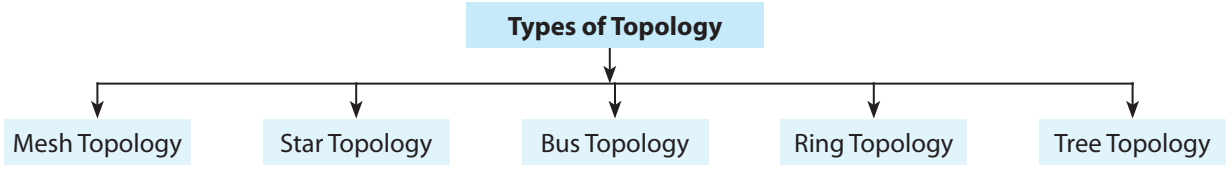


Fig. 5.15: Classification of Network Topologies

5.7.1 Mesh Topology

In mesh topology, each computer is connected with the other computer. There is a point-to-point link between each dedicated node (workstation). In this type of topology, the link carries traffic only between the two connected devices. A fully connected mesh network has **n(n–1)/2** links, where n is the total number of connecting nodes.

For example, if there are five computers and each is connected with the other one, then there will be 5(5–1)/2=10 links.

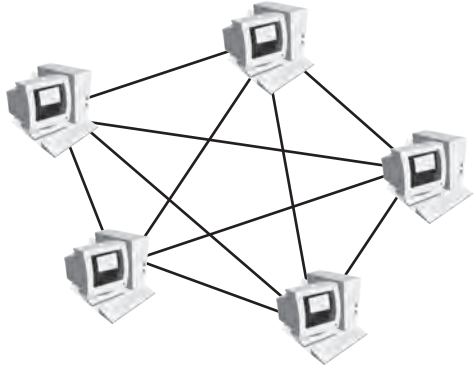


Fig. 5.16: Mesh Topology

Advantages of Mesh Topology

(a) Each communicating device carries its own data through its own dedicated link, hence eliminating traffic problems.

(b) A mesh topology is robust. If one link becomes unusable, it does not affect the entire system.

(c) Expansion and modification in topology can be done without disrupting other nodes.

(d) There is the advantage of privacy or security of data. When every message travels along a dedicated line, only the intended recipient sees it. Physical boundaries prevent other users from gaining access to messages.

(e) Point-to-point links make fault identification and fault isolation easy. Traffic can be routed to avoid links with suspected problems. This facility enables the network manager to discover the precise location of the fault and aids in finding its cause and solution.

Disadvantages of Mesh Topology

(a) Mesh topology is highly expensive to set up and involves high maintenance costs because of the amount of cabling and the number of I/O ports required.

(b) The hardware required to connect each link (I/O ports and cable) can be prohibitively expensive.

(c) Set-up and maintenance of this topology is very difficult. Even administration of the network is tough.

CTM: When there is a point-to-point link between each computer in a network, it forms mesh topology.

5.7.2 Star Topology

In star topology, each communicating device is connected to a central controller called ***hub***. ***(A hub is a hardware device used to connect several computers together.)*** Unlike mesh topology, the devices in star topology send and receive data indirectly; the data passes to and from the hub. If the data is required to be sent from one computer and received by another computer, then this can be accomplished only by the central controller as each data is first sent to the hub, which then relays that data to the destination computer.

The most popular and widely used **LAN** technology, **Ethernet**, currently operates in Star Topology.

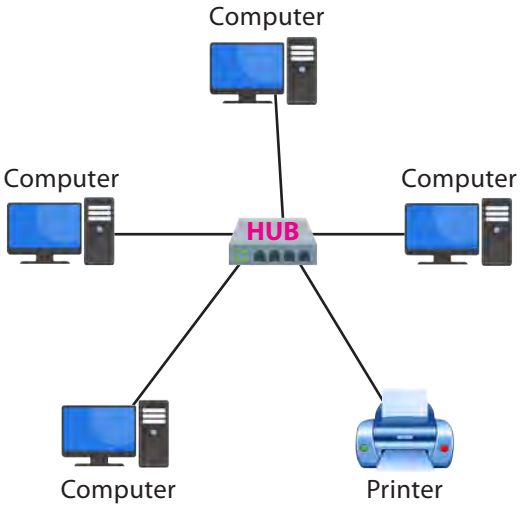


Fig. 5.17: Star Topology

Advantages of Star Topology

(a) Fewer wires are required as compared to mesh topology which, thus, reduces the number of input/output ports.

(b) Installation and maintenance of network is easy and takes less time.

(c) It is easy to detect faults in this network as all computers are dependent on the central hub.

This means that any problem which makes the network non-functioning can be traced to the central hub.

(d) The rate of data transfer is fast as all the data packets or messages are transferred through central hub.

(e) As the nodes are not connected to each other, any problem in one node does not hamper the performance of other nodes in the network.

(f) Removal or addition of any node in star topology can take place easily without affecting the entire performance of the network.

Disadvantages of Star Topology

(a) Extra hardware is required for installation of central controller known as hub.

(b) All nodes of star topology are dependent on central hub and, therefore, any problem in the hub makes the entire network shut down.

(c) The performance of the entire network is directly dependent on the performance of the hub. If the server is slow, it will cause the entire network to slow down.

(d) More cabling is required in star topology as compared to any other topology (ring, bus, tree) as all nodes are directly connected to a central hub.

5.7.3 Bus Topology

Bus topology is a multipoint configuration, *i.e.*, several devices are connected to a main long cable which acts as backbone. Nodes are connected by drop lines and taps. A drop line is a connection between the long cable and devices and taps are the connectors that are punctured inside the main cable. The data flows from one end of the cable to the other.

However, as the signal travels a long distance, it becomes weaker and weaker.

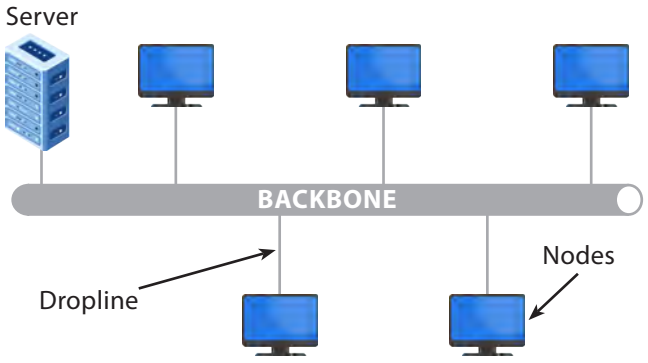


Fig. 5.18: Bus Topology

Therefore, there should be a limited number of nodes connected to a line. Ethernet is a common example of bus topology.

Advantages of Bus Topology

(a) Nodes can be connected or removed easily from bus network.

(b) It requires less cable length than a star topology.

(c) Bus network is easy to implement and can be extended up to a certain limit.

(d) It works well for small networks.

Disadvantages of Bus Topology

(a) If there is a fault or break in the main cable, the entire network shuts down.

(b) Terminators are required at both ends of the backbone cable.

(c) Fault isolation is difficult to detect if the entire network shuts down.

(d) When the network is required in more than one building, bus network cannot be used.

(e) The signal becomes weaker if number of nodes becomes large.

(f) Performance degradation occurs with the increased number of nodes.

(g) Collision of data can take place because several nodes can transmit data to each other at one time.

CTM: There is a main cable which is connected to several workstations through taps. Collision of data can take place in bus topology.

5.7.4 Ring Topology

In ring topology, each node is connected to two other nodes on either side of it, forming a ring network. It shows the line configuration in which each node is connected to one predecessor node and one successor node. Signal is transmitted only in one direction along the entire ring in a circular fashion. In ring topology, each device is incorporated with a repeater to strengthen the signal as a signal passes through all nodes in the entire network. When the data is transmitted from one node to its recipient node, then the intermediate node regenerates the signal and passes the signal to the destined node.

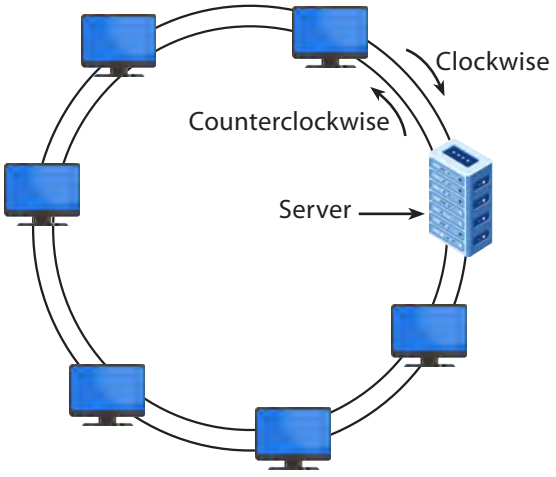


Fig. 5.19: Ring Topology

Token Ring is an example of ring topology.

Advantages of Ring Topology

(a) A central server is not required in ring topology as the data is passed between two nodes which then passes through the entire network.

(b) The data is transmitted in one direction only and, hence, the transmission rate increases.

(c) The adding or removing of network nodes is easy as the process requires changing only two connections.

(d) The configuration makes it easy to identify faults in network nodes.

(e) In this topology, each node transmits the data to its next node in a ring.

(f) It is relatively cheaper as compared to star topology.

Disadvantages of Ring Topology

(a) If there is a fault in a single node, it can cause the entire network to fail.

(b) The movement or changes made to network nodes affect the entire network's performance.

(c) Transmission speed becomes slower with an increase in the number of nodes.

(d) If there is a fault or break in a cable to which all other nodes are connected, the entire network shuts down.

(e) For proper communication between each node, it is required that each computer must be turned on.

CTM: In ring topology, each workstation is connected with the predecessor node as well as with the successor node and, thus, forms a ring. Data is transmitted only in one direction.

5.7.5 Tree Topology

In tree topology, all or some of the devices are connected to the central hub, called an active hub, and some of the devices are connected to the secondary hub, which may be an active hub or passive hub. An active hub contains

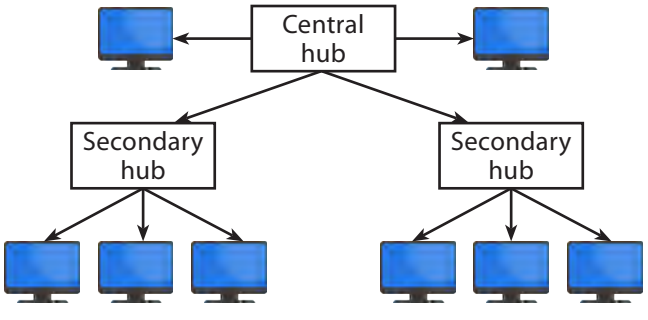


Fig. 5.20: Tree Topology

the repeater that regenerates the signal when it becomes weaker with longer distances. A passive hub simply provides a connection between all other connecting nodes.

Advantages of Tree Topology

(a) The tree topology is useful in cases where a star or bus cannot be implemented individually.

(b) It is most suited in networking multiple departments of a university or corporation, where each unit (star segment) functions separately, and is also connected with the main node (root node).

(c) The advantages of centralization that are achieved in a star topology are inherited by the individual star segments in a tree network.

(d) Each star segment gets a dedicated link from the central bus. Thus, failing of one segment does not affect the rest of the network.

(e) Fault identification is easy.

(f) The network can be expanded by the addition of secondary nodes. Thus, scalability is achieved.

Disadvantages of Tree Topology

(a) As multiple segments are connected to a central bus, the network depends heavily on the bus. Its failure affects the entire network.

(b) Owing to its size and complexity, maintenance is not easy and costs are high. Also, configuration is difficult in comparison to other topologies.

(c) Though it is scalable, the number of nodes that can be added depends on the capacity of the central bus and on the cable type.

CTM: In tree topology, the main central hub is connected to the secondary hub through which all devices are connected with each other. Tree topology is just like a parent-child relationship.

The decision to select a particular topology for a network does not take place in isolation as the topology determines the type of media and access methods that would be used on the network. Therefore, several factors are taken into consideration before selecting a topology which are as follows:

- **Cost:** Cost signifies that the network should be cost-effective. For this, it is required to minimize the installation cost. This can be done by selecting an economical transmission medium (that is, wires) for data exchange between the network nodes. The network cost can also be minimized by reducing the distances between the nodes.
- **Flexibility:** Flexibility signifies that the network must be flexible enough, *i.e.*, it should be easy to reconfigure. Reconfiguring a network means to rearrange the existing nodes or add new ones to the network.
- **Reliability:** It refers to the degree of trust that can be placed on a network. Like any other system, a network can also encounter failure. A network failure can occur due to the following causes:
    1. When an individual node stops working.
    2. When the entire network fails. This is caused by a more serious fault that stops the working of an individual node.

5.8 INTRODUCTION TO INTERNET AND WEB SERVICES

Internet is described as a network of networks, an example of Wide Area Network (WAN). With the advent of internet, it has become possible to access almost any information, communicate with anyone across the world, and carry out several internet-based tasks.

All these tasks can be executed by connecting a computer to the internet, which is also called **going online**.

➢ How internet works

Our computers or smartphones get linked to the internet using phone line/Mobile ISP (Internet Service Provider) and usually become a part of LAN. This LAN is further connected to ISP using a high speed phone line like T1 line (1.5mbps), whereas a normal phone line or modem typically handles 30,000 to 50,000 bits per second.

ISPs then connect to larger ISPs which maintain fibre optic backbones for an entire region. Backbones around the world are connected through fibre optic lines, undersea cables or satellite links. In this manner, every computer on the internet is connected to each other for communication and sharing of resources.

Thus, internet is a classification of WAN, working with the help of various networking devices and protocols (especially TCP/IP) to forward data from source to destination devices without worrying about the constraints of dissimilar devices and architecture.

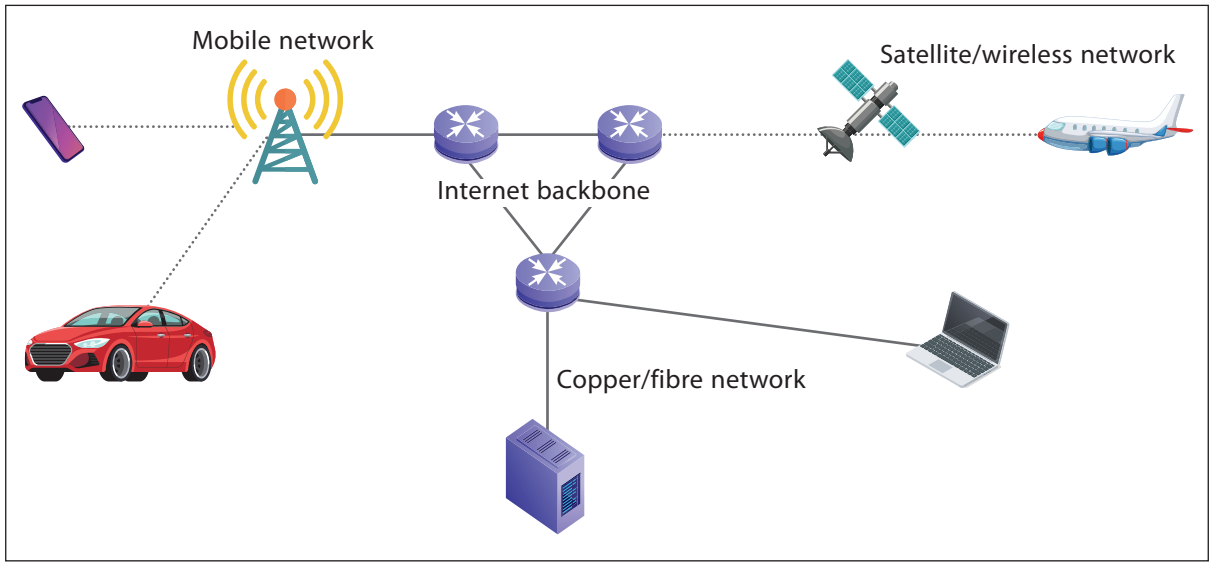


Fig. 5.21: Communication across internet

Internet offers several important and extensively used features or services which are described as follows:

5.8.1 WWW (World Wide Web)/Web

People usually think that the internet and the World Wide Web (WWW) are the same entities. However, they are closely linked but quite different in their overall working and concept.

The internet is a huge network of computers all connected together. The World Wide Web ('www' or 'web' for short) is a collection of webpages found on this network of computers. Our web browser uses the internet to access the web.

The World Wide Web is a way of exchanging information between computers on the internet through a huge collection of multimedia documents and elements. World Wide Web was created by Timothy Berners Lee in 1989 at CERN in Geneva. World Wide Web came into existence as a proposal by him to allow researchers to work together effectively and efficiently at CERN.

WWW/Web is an information service that can be used for sending and receiving information over the internet through interlinked hypertext documents. Web pages may contain text, images, videos and other multimedia components as well as web navigation features consisting of hyperlinks. The documents are formatted in a markup language called HTML (Hyper Text Markup Language) that provides links to other documents as well as graphics, audio and video files. The World Wide Web is based on client-server architecture where a client sends a request and the server processes that request and sends responses. A WWW client is called a web browser and a WWW server is called a web server.

CTM: WWW is a set of programs and protocols that allows the user to create and display multimedia web pages and is linked to the internet.

5.8.2 Domain Names

To communicate over the internet, we can use IP addresses. But it is not possible to remember the IP address of a particular website or computer every time. Domain names make it easier to resolve IP addresses into names, *for example*, cbse.nic.in, google.com, meritnation.com, etc. It is the system which assigns names to some computers (web servers) and maintains a database of these names and corresponding IP addresses. Domain names are used in URLs to identify particular web servers, *for example*, in the URL https://www.cbse.nic.in/welcome.htm, the domain name is cbse.nic.in.

A domain name consists of the following parts.

1. Top-level domain name or primary domain name, and
2. Sub-domain name(s).

For example,

In the domain name cbse.nic.in:

in is the primary domain name

nic is the sub-domain of in

cbse is the sub-domain of nic.

The top-level domains are categorized into following domain names:

Generic Domain Names

- ·com – Commercial business
- ·edu – Educational institutions
- ·gov – Government agencies
- ·mil – Military
- ·net – Network organizations
- ·org – Organizations (non-profit)

Country-specific Domain Names

.in – India

·au – Australia

·ca – Canada

.ch – China

.nz – New Zealand

.pk – Pakistan

.jp – Japan

.us – United States of America

5.8.3 URL

URL stands for uniform resource locator that helps in locating a particular website or a web page, *for example*, http://www.cbse.nic.in/academics.html is a URL for a specific website. In this URL, 'http' stands for hypertext transfer protocol, and 'www.cbse.nic.in' indicates the IP address or the domain name where the source is located. 'academics.html' specifies the name of the specified html document on the website of CBSE.

CTM: URL stands for uniform resource locator that stores the address of a web page.

5.8.4 Email (Electronic mail)

Short for electronic mail, e-mail or email is information stored on a computer that is exchanged between two users over a network. It is a fast and efficient way to communicate with multiple users at the same time and is the cheapest and fastest method to send files and other information across the network. The basic purpose of using email is to send and receive messages. It also allows a user to save messages in a file for future retrieval, print mail messages, reply, and attach a file along with the mail message to be sent from sender to receiver. The salient features of email are:

- Automatic/default reply to messages.
- Auto-forward and redirection of messages.
- Facility to send copies of a message to many people.
- Automatic filing and retrieval of messages.
- Addresses can be stored in an address book and retrieved instantly.
- Notification if a message cannot be delivered.

5.8.5 Chat

Chatting is the other method of internet conversation. It enables people connected anywhere on the internet to join in live discussions. Chat sessions allow many users to join in the same free-form conversation, usually centred on a discussion topic. The examples of chatting software are MSN Messenger, Yahoo Messenger, IRC (Internet Relay Chat), Pidgin, etc.

5.8.6 VoIP (Voice over Internet Protocol)

Voice over Internet Protocol (VoIP) is a technology that allows us to make voice calls using a broadband internet connection instead of a regular (or analog) phone line. VoIP services convert our voice into a digital signal that travels over the internet. If we are calling a regular phone number, the signal is converted to a regular telephone signal before it reaches the destination. VoIP can allow us to make a call directly from a computer or a special VoIP phone. In addition, wireless "hotspots" in locations such as airports, parks and cafes allow us to connect to the internet and may enable us to use VoIP service wirelessly.

Advantages of VoIP

(a) Less cost

(b) Accessibility

(c) Flexibility

(d) Better voice quality

(e) Extra/Less expensive features

Disadvantages of VoIP

(a) Reliable internet connection required

(b) Power outages/Emergencies

(c) Latency

Common examples of VoIP apps are Nextiva, Aircall, Skype, WhatsApp, Google Hangouts, Viber and Facebook Messenger.

5.8.7 Website

A website is a collection of various web pages, images, video, audio or other kinds of digital assets that are hosted on one or several web servers. The first page of a website is known as home page where all the links related to other documents are displayed. The web pages of a website are written using HTML and the information is transferred over the internet through HTTP protocol. The HTML documents consist of several hyperlinks that are accessed through HTTP protocol. Examples of various websites are: cbse.nic.in, google.com, amazon.in, etc.

CTM: A website is a collection of several web pages which are related to each other through hyperlinks.

5.8.8 Web Page

A web page is an electronic document/page designed using HTML. It displays information in textual or graphical form. Traversal from one web page to another web page is possible through hyperlinks.

A web page can be classified into two types:

- **Static web page:** A web page which displays same kind of information whenever a user visits it is known as a static web page. A static web page generally has .htm or .html as extension.
- **Dynamic web page:** An interactive web page is a dynamic web page. A dynamic web page uses scripting languages to display changing content on the web page. Such a page generally has .php, .asp, or .jsp as extension.

5.8.9 Web Server

A web server is a server that stores web pages and when a web client sends any request to a server, the server responds to the request and displays the requested web pages. A web server is a program that runs on a computer connected to the internet. Web server waits for a request, finds the documents and generates information, if required, and sends it back to the browser that requested for it. A single web server may support multiple websites, or a single website may be hosted on several linked or mirrored web servers.

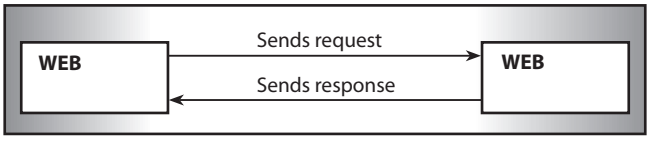


Fig. 5.22: Sending and Receiving Request

Some popular web servers are: Apache web server, Netscape enterprise web server, Microsoft internet information server, etc.

CTM: A web server stores web documents and responds to the requests made by web browsers.

5.8.10 Web Hosting

Web hosting is a service which is provided by companies to its clients to allow them to construct their own websites which are accessible to the internet users via World Wide Web. Such companies are known as web hosts. These companies provide space on a web server they own for use by their clients as well as provide internet connectivity. The websites which are constructed display information for their organization in the form of web pages. The host may also provide an interface or control panel for managing the web server so as to add news and events related to their organization or for uploading some information which may be valuable for the internet users. A client can also use control panel for installing scripts as well as other modules and service applications like email. *webhostingsitesindia.co.in* is one of the top domain name registration and web hosting companies in India. It is the only hosting company which provides support in regional languages.

CTM: Web hosting is a service that is provided by the company to users to create web-based applications.

5.9 WEB BROWSER

It is a software that helps in accessing web pages and, thus, is also called web client. It helps the user to navigate through the World Wide Web and display web pages. Some popular web browsers are: Mozilla Firefox, Opera, AOL, Webkit, Iceweasel, etc.

CTM: A web browser is a WWW client that navigates through the WWW and displays web pages.

The primary function of a web browser is to render HTML code (the code used to design or "markup" web pages). Each time a web browser loads a web page received from web server, it processes the HTML code, which may include text, links, and references to images and other items, such as cascading style sheets and JavaScript functions. The browser processes these items, then renders them in the browser window.

The major functions carried out by a web browser are:

- Send and receive internet resources
- Access web pages, render and display them
- Select and save our favourite pages
- Print documents
- Keep records of our activity
- Store information in the cloud
- Install applications

5.9.1 Commonly-used Web Browsers

Some of the most popularly used web browsers are explained in brief as follows:

Google Chrome

Google Inc. has developed Google Chrome web browser. It is an open and free source application. It runs on different operating systems, like Microsoft Windows, Mac OS X, Chrome OS, Linux, Android and iOS. It was launched in 2008 and since then has become the most popular web browser across the world. Google Chrome is very strong in its application performance and JavaScript processing speeds. It allows users to create local desktop shortcuts which will open the desired web page quickly and easily.

Mozilla Firefox

This web browser is a free and open source application developed by the Mozilla Corporation and the Mozilla Foundation in 2002. It works on Microsoft Windows, Mac OS and Linux operating systems. Firefox features a pop-up blocker, anti-phishing and anti-malware warnings.

Apple Safari

This web browser is free and closed source, developed by Apple Inc. It works on OS X, iOS, and Microsoft Windows operating systems. It was launched in 2003 and works on webkit to render graphics and run JavaScript. Safari is faster and more energy efficient than other browsers.

Internet Explorer

This web browser was developed by Microsoft Corporation and integrated into the Microsoft Windows operating system in all its versions. It was launched in 1995 and was the most popular web browser until it was overtaken by Google Chrome in 2011. It supports ad-ons, improved security and power saving features.

Opera

Opera is a web browser developed by the company Opera Software. It is compatible mainly with Microsoft Windows and Mac OS X operating systems, although it also works, in older versions, on Linux. It was launched in 1995. It provides integrated ad blocker and sidebar extension for multitasking.

After discussing the various web browsers available, we will now discuss how to manage web browser settings.

Google Chrome Browser Settings:

1. No More Notification Requests – chrome://settings/content/notifications
2. Get Around Ad Blocker Blocking – chrome://settings/content/javascript
3. Font and Sizing – chrome://settings/fonts
4. Review Passwords – chrome://settings/passwords
5. Customize Startup Pages – chrome://settings/onStartup
6. Send a Do Not Track Request – chrome://settings/privacy
7. Set Flash to Ask First – chrome://settings/content
8. Mic and Camera Access – chrome://settings/content
9. Send Reports to Google – chrome://settings/syncSetup

   Resetting Chrome to Default Settings:

   If Chrome has been uninstalled or got corrupted and needs to be reset, then the following simple steps are to be followed:

   Scroll to the bottom of settings and click Advanced. Then:

   - On Mac, Chromebook, or Linux: Under Reset Settings, click Restore settings to their original defaults > Reset Settings.
   - On Windows: Under Reset and cleanup, click Reset Settings > Reset Settings.

Mozilla Firefox Browser Settings:

1. Homepage, Font and Color, Downloads – //option/general
2. Search Bar and Search Engine – //option/search
3. Forms, Passwords, History, Cookies, Security – //option/privacy and security
4. Firefox Account – //option/firefox account

   Resetting Mozilla Firefox to Default Settings:

   - Click the menu icon, then click on "Help". Click on Firefox's main menu button, represented by three horizontal lines.
   - Click "Troubleshooting Information".
   - Click on "Refresh Firefox".
   - Confirm.
   - Click "Finish".

Internet Explorer Browser Settings:

1. Homepage, Browsing history, Appearance (font color) – //tools/internet options/general
2. Security Level for Zones (internet, local internet, trusted sites, restricted sites) – //tools/internet options/security
3. Privacy, Pop-up Blocker – //tools/internet options/privacy
4. Parental Control, Autocomplete – //tools/internet options/content
5. Setup Internet Connections – //tools/internet options/connections
6. Default Web Browser, Add-ons – //tools/internet options/programs
7. Accessibility, Browsing, Multimedia, Security – //tools/internet options/advanced/settings
8. Resetting Internet Explorer to Default Settings – //tools/internet options/advanced/Reset

5.9.2 Add-on Extensions

Add-ons are tools which get integrated into our browser. They are similar to regular apps or programs, but only run when the browser runs. Add-ons can allow the viewing of certain types of web content, such as Microsoft's Silverlight necessary for Netflix movies.

How Add-ons are installed

There are two key ways in which add-ons can be installed—through an external installer and through the browser's own add-on service. The add-on service is the most reliable way of installing an add-on, with the browser service providing a relative "vetting" process for the general safety of the add-on. Outside programs can also install add-ons in web browser as part of its separate installation process. Microsoft Office, *for example*, may place an add-on which speeds up the in-browser opening of Office documents.

How to remove Add-ons

Some add-ons, particularly the ones installed outside the browser, create an entry in the "Programs" portion of the control panel. These can be removed in the same way as we uninstall any other program. Many add-ons, however, can only be removed through the browser's add-on manager.

5.9.3 Plug-ins

A plug-in is a piece of software that acts as an add-on to a web browser and gives the browser additional functionality. Plug-ins can allow a web browser to display additional content it was not originally designed to display. An example of a plug-in is the free Macromedia Flash Player, a plug-in that allows the web browser to display animations using the Flash format.

Most plug-ins are available as free downloads. To install the plug-in, we visit the website of the plug-in's developer and click on a link that will download the installer onto our system. We can save the installer to an easy-to-find location such as the Desktop or a specific folder we have created to organize all of our downloads. Once we have downloaded the installer, we can run/open it and follow the prompts to install the plug-in on our system.

Difference between Add-ons and Plug-ins

Plug-in is a complete program whereas add-on is not a program. *For example*, Flash, a plug-in made by Adobe, is required to play a video in flash player. Also, Java is a plug-in made by Sun Microsystem which is used to run programs based on Java. Plug-ins are not bound to browsers only. Flash can be installed on computers to play flash files. Similarly, Java can be installed to run Java files.

On the other hand, an add-on is not a complete program. It is used to add a particular functionality to a browser. If we want to install an add-on on a different working environment, say, WER operating system, we cannot do it. Thus, add-ons have a certain limit with compatible applications only.

5.9.4 Cookies

Cookies are small bits of data stored as text files on a browser. Websites use those small bits of data to keep track of users and enable user-specific applications.

What's in a Cookie

Each cookie is effectively a small lookup table containing pairs of (key, data) values. Once the cookie has been read by the code on the server or client computer, the data can be retrieved and used to customize the web page appropriately.

Suppose we want to have a counter for a webpage to access it three times only; this value can be stored over the user's computer. At first attempt, it will be 1, then 2 and then 3, but for the next attempt, the server will not allow to access that page.

When are Cookies created

Writing data to a cookie is usually done when a new webpage is loaded with code to create cookies.

Why are Cookies used

Cookies are the easiest way to carry information from one session on a website to another, or between sessions on related websites, without having to burden a server machine with massive amounts of data storage. Storing data on the server without using cookies would also be problematic because it would be difficult to retrieve a particular user's information without requiring a login on each visit to the website.

How long does a Cookie last

The time of expiry of a cookie can be set through using server site scripting language when the cookie is created. By default, the cookie is destroyed when the current browser window is closed, but it can be made to persist for an arbitrary length of time after that.

How secure are Cookies

Cookies are not a threat to privacy since they can only be used to store information that the user has volunteered or that the web server already has.

5.10 SETTING UP A COMPUTER NETWORK—AN EXAMPLE

The network functioning is based on Client-Server architecture which requires effective and efficient network design. It defines how clients are connected to server machine(s) on a network. The most important rule or methodology for network (LAN) design is the 80:20 rule.

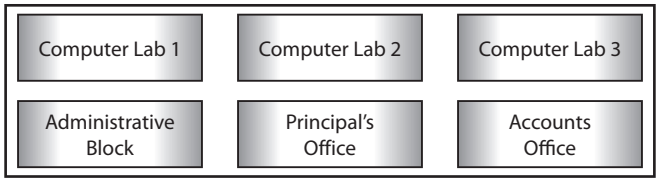
The 80:20 Thumb Rule

This thumb rule states that in a well-organized and designed network, 80 percent of the traffic on a given network segment is local (*i.e.*, destined for a destination system in the same workgroup), and not more than 20 per cent of the network traffic should move across a backbone. The backbone in a network violating the 80:20 rule leads to network congestion and traffic jams.

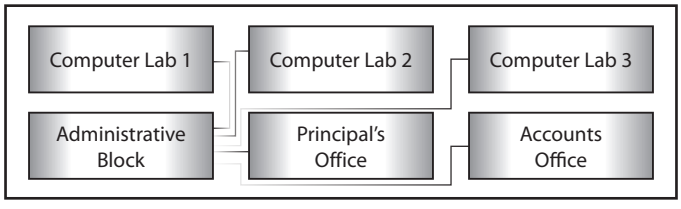
Keeping the above significant rule in mind, let us take an example to understand how to go about LAN design.

An educational society (say XYZ Educational Society), with its head office in Chennai (Tamil Nadu) and schools in various parts of the globe, is setting up a new senior secondary school, 'SF School', in Bahadurgarh (Haryana). The 'SF School' will have 3 computer labs with 30 computers in each lab, one Accounts office with three computers, one Administrative block with five computers, and a Principal's office with one computer.

Let us see how a computer network can be set up in the school. First of all, we can draw a rough sketch of the school with computers at various locations as follows:



1. Independent LANs can be set up in each of the following buildings: Computer Lab1, Computer Lab2, Computer Lab3, Administrative Block and Accounts Office.
2. These LANs can be set up in STAR topology using UTP cable (economical, reliable and easily available).
3. For this, one switch (with suitable number of ports) will be required in each of these buildings. More than one switch can be used in computer labs if a switch with more than 30 ports is not available.
4. Two internet connections (broadband for high speed) can be procured in Administrative Office. Two connections should be procured from two different ISPs so that:
   (a) Internet connection in Administrative office can be kept separate from the other Computer labs where students may do a lot of experimentation.
   (b) If one internet connection is not working, the other can be used in case of urgent requirements.
5. These buildings can then be connected as follows:



6. This interconnection will ensure that each building is directly connected to Administrative block.
7. This way, internet connection will be available in each building irrespective of the status of the other building.
8. Server (if any) of the school may be placed in the Administrative block so that it remains safe (physically) and a firewall can be set up so that the whole network remains safe from any kind of virus or intrusion attacks.

   There is no need to put in any extra efforts or expenses to link the school to its head office. This can be taken care of using the internet connections.

Tips to solve technical questions based on Networking

Where Server should be placed: Server should be placed in the building where the number of computers is maximum.

1. **Suggest a suitable cable layout of connection:** A suitable cable layout can be in the following two ways:

   (a) **On the basis of server:** First, the location of the server is found out. Server should be placed in that building where the number of computers is maximum (according to the 80:20 rule). After finding the server position, each building distance is compared with the server building directly or indirectly (taking other building(s) in between). The shortest distance is counted whether it is directly or indirectly calculated.

   (b) **On the basis of distance from each building:** The distance between each building is compared to all other buildings, either directly or indirectly. The shortest distance is calculated whether it is direct or through some other building.

2. Where the following devices should be placed

Server	: Large number of computers in the building
HUB/Switch	: Each building
Modem	: In the server room
Repeater	: It is used if the distances are higher than 70m. It regenerates data and voice signals.
Router	: When one LAN is required to be connected to the other LAN
Best Layout	: Star (from Server), BUS topology
Best Cable	: Twisted Pair, Ethernet Cable, Coaxial cable (when distance is in metres); For large distances—Fibre optics cable.
Best connecting technique	: In hilly regions, radio waves should be used and city-to-city, state-to-state satellite should be used.

MEMORY BYTES

- ARPANET was the first internet followed by NSFNET and other small networks.
- A gateway is a device that connects dissimilar networks.
- A backbone is a central interconnecting device that connects two or more computers.
- Topology is a way of connecting computers physically or logically.
- Star topology uses a central hub where each computer indirectly connects with the other computer on the network.
- A set of rules that governs internet is called protocol.
- TCP protocol is responsible for sequential arrangement of packets.
- IP protocol is responsible for fragmentation of data into packets and sends those packets in random order.
- FTP protocol is used to share files across networks.
- TELNET is a remote login where a user can login on another user's system.
- HTTP is used for displaying web pages.

- Web browser is an application program that helps in opening web pages.
- The first page of any website is known as home page.
- Communication media is a transmission media for transmitting data across the network.
- Guided media is also known as wired media while unguided media is also known as wireless media.
- LAN, MAN, WAN and PAN are the four types of networks.
- Viruses are malicious programs that can damage files, disks, file allocation table, etc.
- Spams are unsolicited mails that eat up the disk space.
- Hub refers to a networking component which acts as a convergence point of a network allowing the transfer of data packets.
- Switch refers to a device which filters and forwards data packets across the network.
- Web hosting service is a type of internet hosting service that allows individuals and organizations to host their own website and users with online systems to store information such as images, videos, etc.
- A data channel is the medium used to carry information or data from one point to another.

OBJECTIVE TYPE QUESTIONS

1. Fill in the blanks.

(a) Through ..... you can establish contact with anyone in the world.

(b) The main function of ..... is to divide the message or data into packets of a definite size on the source computer.

(c) ..... refers to wireless fidelity which enables us to connect to the ISP without any cable.

(d) ..... is a software that enables us to access the internet and explore websites.

(e) Web page constitutes the ..... .

(f) A ..... is someone with a strong interest in how things work, who likes to create and modify things for their own enjoyment.

(g) A computer ..... is a small software program that spreads from one computer to another and interferes with the normal functioning of computer.

(h) Electronic junk mail or junk newsgroup postings are known as ..... .

(i) Digital signature meets the need for ..... and integrity.

(j) The first network that planted the seed of internet was ..... .

(k) The protocol used for internet is ..... .

(l) A device used to connect dissimilar networks is called..... .

(m) ..... is responsible for handling the address of the destination computer so that each packet is delivered to its proper destination.

(n) Tricking people through authentic-looking emails or websites is called ..... .

(o) A program designed to replicate and eat up a computer's storage is called ..... .

(p) A digital document issued to a site by a certification authority of the internet is called a ..... .

(q) To connect computers located in a specific building or campus is known as ..... .

(r) Wi-Fi, infrared and Bluetooth are examples of ..... .

(s) Interspace is a ..... .

(t) A server that provides its services to other workstations on a network is a ..... .

(u) The technique of switching in which data is divided into smaller fragments is called ..... .

(v) ..... is a dedicated link between the sender and the receiver.

(w) ..... is the measuring unit of speed at which the data transfer takes place.

(x) All the computers are connected with each other in an unorganized manner in topology ..... .

(y) In ....., all computers share equivalent responsibility for processing data.

Answers: (a) Internet (b) TCP (c) Wi-Fi
(d) Browser (e) World Wide Web (f) Hacker
(g) Virus (h) SPAM (i) Authentication
(j) ARPANET (k) TCP/IP (l) Gateway
(m) IP (n) Hacking (o) WORM
(p) Digital certificate (q) LAN
(r) Communication Mediums (s) Network (t) Dedicated server
(u) Packet switching (v) Circuit switching (w) bits/second
(x) Mesh (y) Peer-to-peer network

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

(a) A set of rules that governs internet is called protocol.
(b) A repeater handles different protocols.
(c) A hub is known as an intelligent device on the network.
(d) A location on a net server is called a website.
(e) A document that uses HTTP is called a web page.
(f) A switch is a device used to segment networks into sub-networks or subnets.
(g) Email helps us to send and receive messages through video conferencing.
(h) The degeneration of a signal over a distance on a network is called attenuation.
(i) Coaxial cable possesses higher tensile strength than optical fibre.
(j) When two entities are communicating and do not want a third party to listen, this situation is defined as secure communication.

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

3. Multiple Choice Questions (MCQs)

(a) A computer network:
(i) Is a collection of hardware components and computers
(ii) Is interconnected by communication channels
(iii) Allows sharing of resources and information
(iv) All of the above

(b) What is a firewall in computer network?
(i) The physical boundary of network
(ii) An operating system of computer network
(iii) A system designed to prevent unauthorized access
(iv) A web browsing software

(c) What is the use of Bridge in the network?
(i) To connect LANs (ii) To separate LANs
(iii) To control network speed (iv) All of the above

(d) Each IP packet must contain:
(i) Only Source address (ii) Only Destination address
(iii) Source and Destination address (iv) Source or Destination address

(e) Which of these is not a communication channel?
(i) Satellite (ii) Microwave (iii) Radio wave (iv) Wi-Fi

(f) MAN stands for ..... .
(i) Metropolitan Area Network (ii) Main Area Network
(iii) Metropolitan Access Network (iv) Metro Access Network

(g) Which of these is not an example of unguided media?
(i) Optical Fibre Cable (ii) Radio wave
(iii) Bluetooth (iv) Satellite

(h) In which topology are all the nodes connected through a single Coaxial cable?
(i) Star (ii) Tree (iii) Bus (iv) Ring

(i) Which of the following is the smallest network?
(i) WAN (ii) MAN (iii) LAN (iv) Wi-Fi

(j) Which protocol is used for the transfer of hypertext content over the web?
(i) HTML (ii) HTTP (iii) TCP/IP (iv) FTP

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

SOLVED QUESTIONS

1. Define a network. What is its need?

Ans. A network is an interconnected collection of autonomous computers that can share and exchange information.

Need for networking:

(a) *Resource sharing:* Resources are shared by all computers over the network for effective utilization.
(b) *File sharing:* A file in a network can be accessed from anywhere.

2. Write **two** advantages and **two** disadvantages of network.

Ans. *Advantages of network:*

(a) We can share resources such as printers and scanners.
(b) We can share data and access files from any computer.

Disadvantages of network:

(a) If there is any problem in the server, then no communication can take place.
(b) Network faults can cause loss of data.
(c) If there is no privacy mechanism used, then the entire network can be accessed by an unauthorized person.

3. What is ARPANET? What is NSFNET?

Ans. ARPANET (Advanced Research Project Agency Network) is a project sponsored by US Department of Defence. NSFNET, developed by the National Science Foundation, was a high capacity network and strictly used for academic and engineering research.

4. What are the various types of networks?

Ans. A network is an interconnection of several nodes through some communication media with the goal of sharing data, files and resources. There are three types of networks:

(a) Local Area Network (LAN)
(b) Metropolitan Area Network (MAN)
(c) Wide Area Network (WAN)

5. Name the various layers of coaxial cable.

Ans. Coaxial cable consists of the following layers:

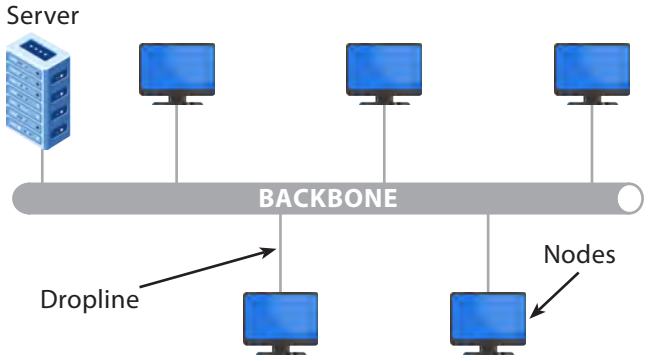
(a) A metallic rod-shaped inner conductor
(b) An insulator covering the rod
(c) A metallic outer conductor called shield
(d) An insulator covering the shield
(e) A plastic cover

6. What is a spam mail? [CBSE D 2015]

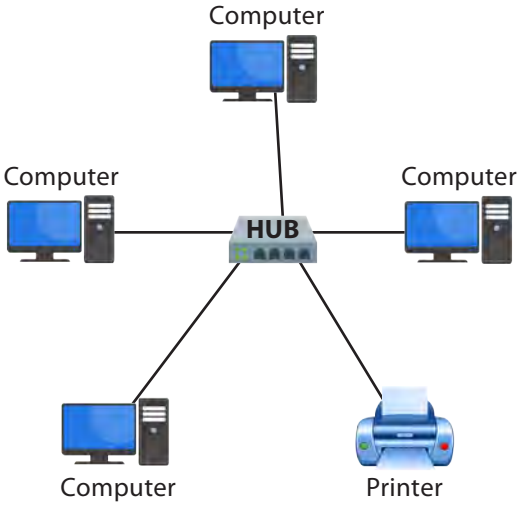
Ans. Spam is the abuse of electronic messaging systems (including most broadcast media, digital delivery systems) to send unsolicited bulk messages indiscriminately.

7. Illustrate the layout for connecting five computers in a Bus and a Star topology of Networks.

Ans. **Bus Topology**



Star topology



8. In networking, what is WAN? How is it different from LAN? [CBSE D 2011]

Ans. Internet is an example of WAN (Wide Area Network). Most WANs exist to connect LANs that are not in the same geographical area. WAN is different from LAN due to its network range. WAN is for connecting computers anywhere in the world without any geographical limitation whereas LAN is confined within a range of 100m to 500m.

9. What is meant by topology? Name some popular topologies.

Ans. Topology is the arrangement by which computers are connected with each other, either physically or logically.

The popular topologies are:

(a) Bus or Linear Topology
(b) Ring Topology
(c) Star Topology
(d) Tree Topology

10. What is TCP/IP?

Ans. TCP/IP (Transmission Control Protocol/Internet Protocol) is a protocol for communication between computers used as a standard for transmitting data over networks and is the basis for standard internet protocols. It is also responsible for assembling packets at the receiver's side.

11. Define web browser and web server.

Ans. **Web Browser:** A web browser is a software which is used for displaying the content on web page(s). It is used by the client to view websites. Examples of web browser—Google Chrome, Firefox, Internet Explorer, Safari, Opera, etc.

Web Server: A web server is a software which entertains the request(s) made by a web browser. A web server has different ports to handle different requests from web browser, like generally FTP request is handled at Port 110 and HTTP request is handled at Port 80. Example of web server is Apache.

12. What is web hosting?

Ans. Web hosting is a means of hosting web server applications on a computer system through which electronic content on the internet is readily available to any web-browser client.

13. What is hacking?

Ans. Hacking is a process of accessing a computer system or network without knowing the access authorization credential of that system. Hacking can be illegal or ethical depending on the intention of the hacker.

14. What are cookies?

Ans. Cookies are messages that a web server transmits to a web browser so that the web server can keep track of the user's activity on a specific website. Cookies are saved in the form of text files in the client computer.

15. What is Web 2.0?

Ans. Web 2.0 is a concept that takes the network as a platform for information sharing, interoperability, user-centred design, and collaboration on the internet or World Wide Web. A Web 2.0 site allows users to interact and collaborate with each other. Examples of Web 2.0 include social networking sites, Facebook, Google+, Twitter, etc.

16. Give one advantage of bus topology of network. Also state how four computers can be connected with each other using star topology of network.

Ans. In bus topology, the workstations can easily be extended or removed. In star topology, four computers can be connected with each other through a server.

17. What is the difference between Email and Chat? [CBSE D 2014]

Ans. (a) Chat is a type of software while Email is a protocol.
(b) Chat requires the permission of both parties while Email does not.
(c) Chat is typically software dependent while Email is not.
(d) Chat needs accounts on the same provider while Email does not.

18. What are VoIP?

Ans. VoIP are communication protocols and transmission technologies for delivery of voice communication and multimedia sessions over Internet Protocol (IP) networks, such as the internet. Also, we can say that VoIP are IP technology, internet telephony and broadband telephony.

19. Define the following data communicating devices:

(a) Repeater (b) Bridge
(c) Router (d) Gateway

Ans. (a) **Repeater:** It is a device that amplifies and restores the signal before it gets degraded and transmits the original signal back to the destination. A repeater is a regenerator and not an amplifier.

(b) **Bridge:** A bridge is a device designed to connect two LAN segments. The purpose of a bridge is to filter traffic on a LAN. Bridge relays frames between two originally separate segments. When a frame enters a bridge, the bridge not only regenerates the signal but also checks the physical address of the destination and forwards the new copy only to that port.

(c) **Router:** Routers operate in the physical, data link and network layers of the OSI model. They decide the path a packet should take. A router is a networking device whose software and hardware are usually tailored to the tasks of routing and forwarding data packets across the network.

(d) **Gateway:** A gateway operates on all the seven layers of OSI model. A network gateway is a computer which has internet-working capability of joining together two networks that use different base protocols. Gateway converts one protocol to another and can, therefore, connect two dissimilar networks.

UNSOLVED QUESTIONS

1. What is internet?
2. What is network?
3. What are the various types of topologies?
4. Describe bus topology and star topology.
5. Define the following terms:
   (a) Baud (b) Communication channel
   (c) Hubs (d) Repeaters
6. What is modem? Define the functioning of internal modem and external modem.
7. Expand and explain the following terms:
   (a) PPP (b) POP3
   (c) VoIP (d) IRC
8. Describe the following networking devices:
   (a) Hubs (b) Repeaters
   (c) Routers (d) Bridges
   (e) Gateways
9. What are Wi-Fi cards? Explain.
10. What is https? How does it work?
11. What are hubs? How are active hubs different from passive hubs?
12. What are the facilities provided by the Server in a network environment?
13. Which network is easy to expand?
14. Which device filters the data and which device can handle different protocols?
15. What is a network? What are its goals and applications?
16. What is the role of a switch in a network?
17. Briefly discuss the role of the following devices in the context of networking:
   (a) Repeater (b) Router
   (c) Bridge (d) Gateway
18. When would you prefer (i) hubs over repeaters, (ii) bridges over hubs, and (iii) switches over other networking devices?
19. Write the steps to install add-ons in google chrome web browser.
20. What are cookies? Why are cookies created?

CASE-BASED/SOURCE-BASED INTEGRATED QUESTIONS

1. Rovenza Communication International (RCI) is an online corporate training provider company for IT-related courses. The company is setting up their new campus in Kolkata. You, as a network expert, have to study the physical locations of various blocks and the number of computers to be installed. In the planning phase, provide the best possible answer for the queries (a) to (d) raised by them.

   Block-to-Block distance (in Mtrs.):

From	To	Distance
Administrative Block	Finance Block	60
Administrative Block	Faculty Recording Block	120
Finance Block	Faculty Recording Block	70

Expected Computers to be installed in each block:

Block	Computers
Administrative Block	30
Finance Block	20
Faculty Recording Block	100

(a) Suggest the most appropriate block where RCI should plan to install the server.

(b) Suggest the most appropriate block-to-block cable layout to connect all three blocks for efficient communication.

(c) Which type of network out of the following is formed by connecting the computers of these three blocks?

(i) LAN (ii) MAN (iii) WAN

(d) Which wireless channel out of the following should be opted by RCI to connect to students from all over the world?

(i) Infrared (ii) Microwave (iii) Satellite

Ans. (a) Faculty recording block (due to maximum number of computers)

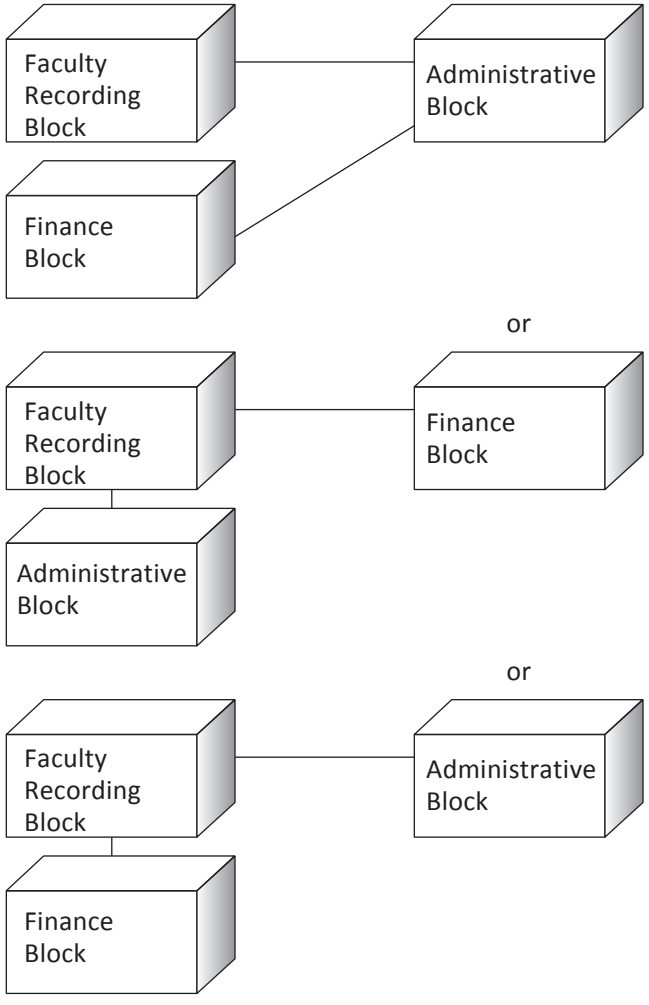
or

Finance (due to shorter distance from the other buildings)

or

Administrative block (due to name or nature of the building)

(b) The different cable layout can be:



(c) (i) LAN

(d) (iii) Satellite

2. University of Correspondence in Allahabad is setting up a network between its different wings. There are 4 wings named Science (S), Journalism (J), Arts (A) and Home Science (H).

Distance between various wings:

Wing A to Wing S	100 m
Wing A to Wing J	200 m
Wing A to Wing H	400 m
Wing S to Wing J	300 m
Wing S to Wing H	100 m
Wing J to Wing H	450 m

Number of Computers:

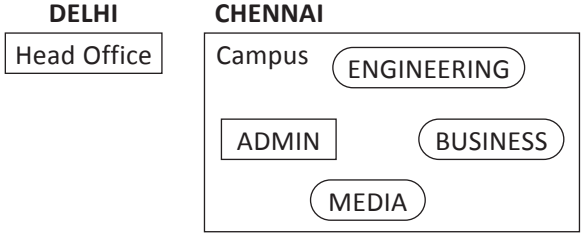
Wing A	150
Wing S	10
Wing J	5
Wing H	50

(a) Suggest the most suitable Topology for networking the computers of all wings.
(b) Name the wing where the Server is to be installed. Justify your answer.
(c) Suggest the placement of Hub/Switch in the network.
(d) Mention how economic technology will provide internet accessibility to all wings.

Ans. (a) Star Topology can be used to network the computers of all wings.
(b) The Server should be installed in Wing A as Wing A has maximum number of computers and installing the server in this wing will help reduce the network traffic.
(c) Hub/Switch will be required in all the wings.
(d) The economic way to provide internet accessibility to all the wings is to use the proxy server at Wing A and connect to the internet through an Ethernet cable using routers and switch.

3. Perfect Edu. Services Ltd. is an educational organization. It is planning to set up its India campus at Chennai with its head office at Delhi. The Chennai campus has four main buildings—ADMIN, ENGINEERING, BUSINESS and MEDIA.

You, as a network expert, have to suggest the best network-related solutions for their problems raised in (a) to (d), keeping in mind the distances between the buildings and other given parameters.



Shortest distance between various buildings:

ADMIN to ENGINEERING	55 m
ADMIN to BUSINESS	90 m
ADMIN to MEDIA	50 m
ENGINEERING to BUSINESS	55 m
ENGINEERING to MEDIA	50 m
BUSINESS to MEDIA	45 m
DELHI Head Office to CHENNAI Campus	2175 m
Number of Computers installed at various buildings is as follows:	
ADMIN	110
ENGINEERING	75
BUSINESS	40
MEDIA	12
DELHI Head Office	20

(a) Suggest the most appropriate location for the server inside the Chennai campus (out of the 4 buildings) to get the best connectivity for maximum number of computers. Justify your answer.

(b) Suggest and draw the cable layout to efficiently connect various buildings within the Chennai campus for connecting the computers.

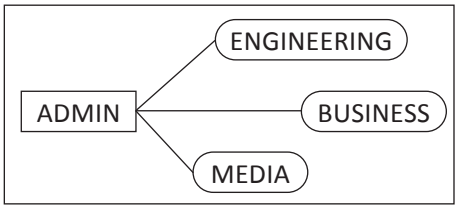
(c) Which hardware device will you suggest to be procured by the company to be installed to protect and control the internet use within the campus?

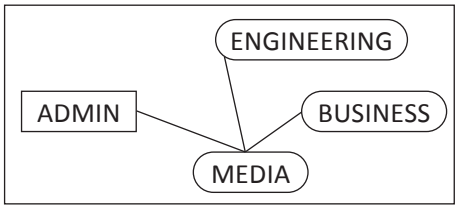
(d) Which of the following will you suggest to establish online face-to-face communication between the people in the Admin Office of the Chennai campus and Delhi Head Office?

(i) Cable TV (ii) Email (iii) Video conferencing (iv) Text Chat

Ans. (a) ADMIN (due to maximum number of computers) or MEDIA (due to shorter distance from other buildings)

(b) Any one of the following:

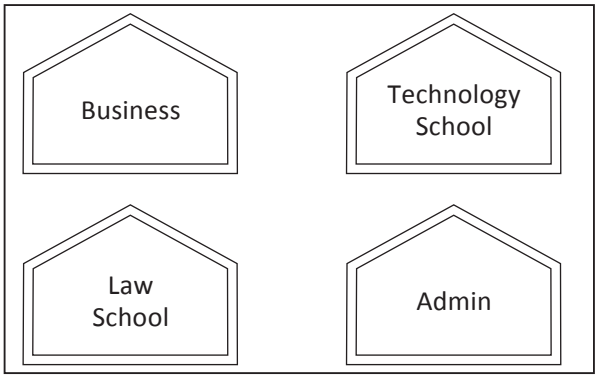




(c) Firewall or Router

(d) (iii) Video conferencing

4. Great Studies University is setting up its academic schools at Sunder Nagar and planning to set up a network. The university has 3 academic schools and one administration centre as shown in the diagram below:



Centre-to-centre distance between various buildings:

Law School to Business School	60 m
Law School to Technology School	90 m
Law School to Admin Centre	115 m
Business School to Technology School	40 m
Business School to Admin Centre	45 m
Technology School to Admin Centre	25 m

Number of computers in each of the Schools/Centre:

Law School	25
Technology School	50
Admin Centre	125
Business School	35

(a) Suggest the most suitable place (*i.e.*, schools/centre) to install the server of this university with a suitable reason.

(b) Suggest an ideal layout for connecting these schools/centre for a wired connectivity.

(c) Which device will you suggest to be placed/installed in each of these schools/centre to efficiently connect all the computers within these schools/centre?

(d) The university is planning to connect its admission office in the closest big city, which is more than 350 km from the university. Which type of network out of LAN, MAN or WAN will be formed? Justify your answer.

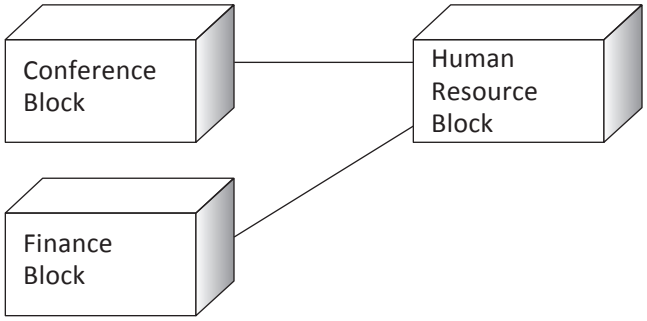
Ans. (a) Admin Centre because it has the maximum number of computers, or Business School because it is closest to all other centres (minimum cable length required).

(b) BUS topology is the most suitable cable layout.

(c) Switch

(d) WAN is the preferred network for this purpose because 350 km is more than the range of LAN and MAN.

5. Tech Up Corporation (TUC) is a professional consultancy company. The company is planning to set up new offices in India with its hub at Hyderabad. As a network adviser, you have to understand their requirements and suggest to them the best available solutions. [CBSE D 2014]



Block-to-Block distance (in Mtrs.):

Block (From)	Block (To)	Distance
Human Resource	Conference	60
Human Resource	Finance	60
Conference	Finance	120

Expected Number of Computers to be installed in each block:

Block	Computers
Human Resource	125
Conference	25
Finance	60

(a) What will be the most appropriate block where TUC should plan to install their server?

(b) Draw a block-to-block cable layout to connect all the buildings in the most appropriate manner for efficient communication.

(c) What will be the best possible connectivity out of the following to connect the new set-up of offices in Bengaluru with its London base office?

(i) Infrared (ii) Satellite Link (iii) Ethernet Cable

(d) Which of the following devices will you suggest to connect each computer in each of the above buildings?

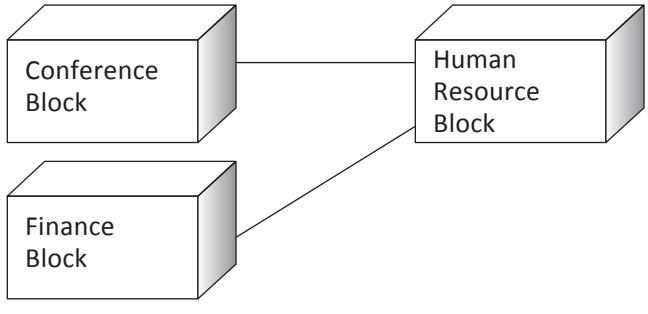
(i) Gateway (ii) Switch (iii) Modem

(e) Write names of any **two** popular Open Source Software which are used as Operating Systems.

(f) Write **any two** important characteristics of cloud computing.

Ans. (a) Human resource will be the most appropriate block where TUC should plan to install the server.

(b)



(c) Satellite Link

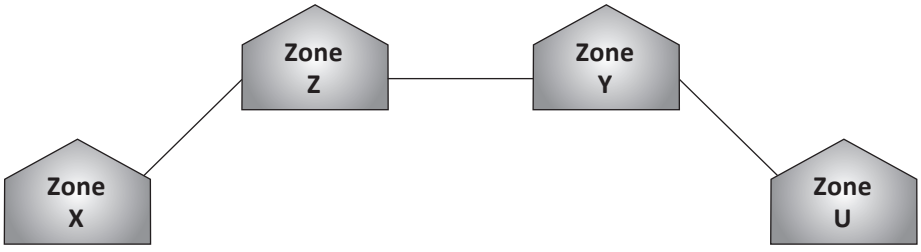
(d) Switch

(e) Linux and Open Solaris

(f) **Cost reduction:** Unlike on-site hosting, the price of deploying applications in the cloud can be less due to lower hardware costs from more effective use of physical resources.

Choice of applications: This allows flexibility for cloud users to experiment and choose the best option for their needs. Cloud computing also allows a business to use, access and pay only for what they use, with a fast implementation time.

6. Sony has set up its Branch at Srinagar for its office and web-based activities. It has four Zones of buildings as shown in the diagram:



Branch-to-branch distance is:

Zone X to Zone Z	40 m
Zone Z to Zone Y	60 m
Zone Y to Zone X	135 m
Zone Y to Zone U	70 m
Zone X to Zone U	165 m
Zone Z to Zone U	80 m

Number of Computers:

Zone X	50
Zone Z	130
Zone Y	40
Zone U	15

(a) Suggest the most suitable cable layout or Networking Topology of connections between the Zones.

(b) Suggest the most suitable place (*i.e.*, Zone) to house the ERP and BI Server of this organization with a suitable reason, with justification.

(c) Suggest the placement of the following devices with justification:

(i) Repeater (ii) Hub/Switch

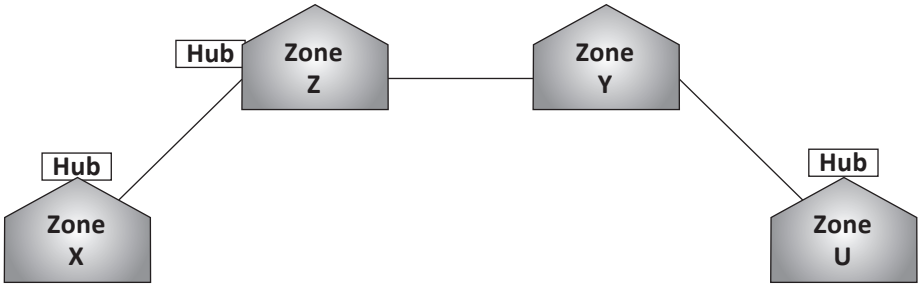
(d) Which is the most economic type of cable for the selected topology?

Ans. (a) Bus Topology

(b) The most suitable place (*i.e.*, Zone) to house the ERP and BI Server is Zone Z as it has the most number of computers; thus, cabling cost will be reduced and most traffic will be local.

(c) **Repeater:** As per the suggested layout, separate repeaters need not be installed as each building/ zone will be having a hub that acts as a repeater.

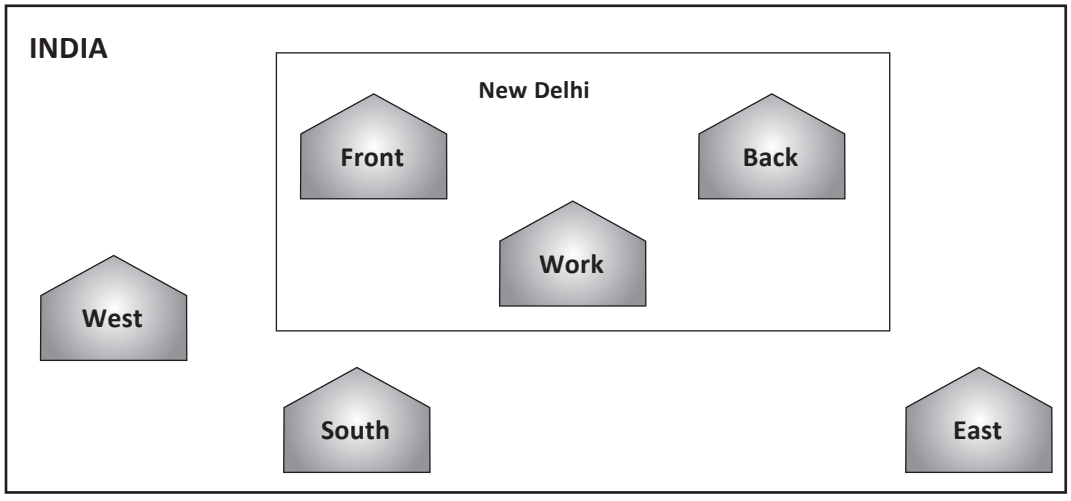
Hub/switch should be placed in each zone.



(d) An economic type of cable is Ethernet or Coaxial cable as it can connect two computers at an economic rate though it provides lesser speed than other expensive methods.

7. Bhartiya Connectivity Association is planning to expand its office in four major cities in India to provide regional IT infrastructure support in the field of Education & Culture. The company has planned to set up their head office in New Delhi at three locations and have named their New Delhi offices as "Front Office", "Back Office" and "Work Office". The company has three more regional offices—"South Office", "East Office" and "West Office"—located in other major cities of India.

A rough layout of the same is as follows:



Approximate distance between these offices as per network survey team is as follows:

Place From	Place To	Distance
Back Office	Front Office	10 km
Back Office	Work Office	70 m
Back Office	East Office	1291 km
Back Office	West Office	790 km
Back Office	South Office	1952 km

In continuation of the above, the company experts have planned to install the following number of computers in each of their offices:

Back Office	100
Front Office	20
Work Office	50
East Office	50
West Office	50
South Office	50

(a) Suggest network type (out of LAN, MAN, WAN) for connecting each of the following sets of their offices:

(i) Back Office and Work Office

(ii) Back Office and South Office

(b) Which device out of the following will you suggest to be produced by the company for connecting all the computers within each of their offices?

(i) Switch/Hub

(ii) Modem

(iii) Telephone

(c) Which of the following communication mediums will you suggest to be procured by the company for connecting their local office units in New Delhi for very effective and fast communication?

(i) Telephone Cable

(ii) Optical Fibre

(iii) Ethernet Cable

(d) Suggest a cable/wiring layout for connecting the company's local office located in New Delhi. Also, suggest an effective method/technology for connecting the company's regional office—"East Office", "West Office" and "South Office"—with offices located in New Delhi.

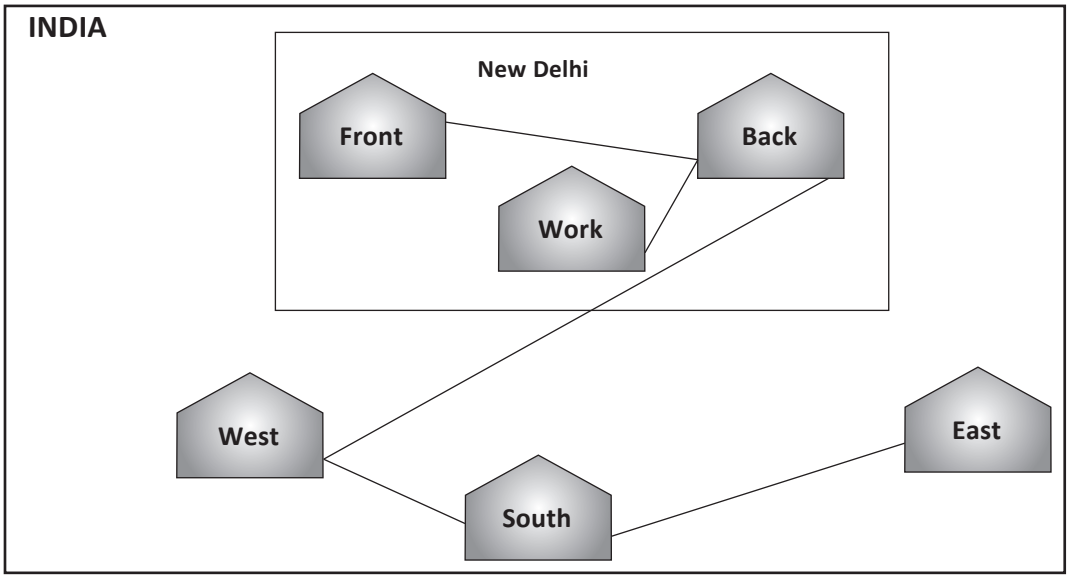
Ans. (a) Back Office and Work Office—MAN

Back Office and South Office—WAN

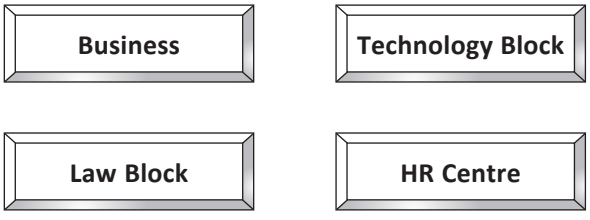
(b) Switch/Hub

(c) Optical Fibre

(d) Cable/Wiring Layout:



8. Quick Learn University is setting up its academic blocks at Prayag Nagar and planning to set up a network. The university has 3 academic blocks and one human resource centre as shown in the diagram given below:



Centre-to-centre distance between various blocks is as follows:

Law block to business block	40 m
Law block to technology block	80 m
Law block to HR centre	105 m
Business block to technology block	30 m
Business block to HR centre	35 m
Technology block to HR centre	15 m

Number of computers in each of the buildings is as follows:

Law block	15
Technology block	40
HR centre	115
Business block	25

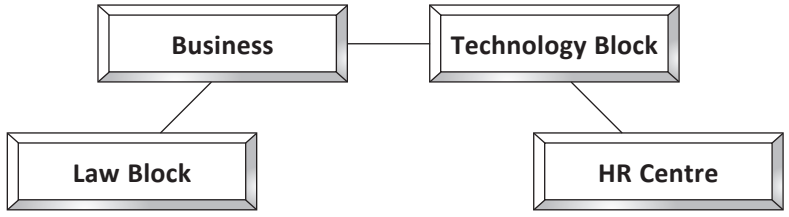
(a) Suggest a cable layout of connection between the blocks.

(b) Suggest the most suitable place to house the server of the organization with suitable reason.

(c) Which device should be placed/installed in each of these blocks to efficiently connect all the computers within these blocks.

(d) The university is planning to link its sales counters situated in various parts of the same city. Which type of network out of LAN, MAN or WAN will be formed?

Ans. (a)



(b) HR centre because it consists of the maximum number of computers to house the server.

(c) Switch should be placed in each of these blocks.

(d) MAN will be formed as Sales counters are in the same city.

6 Societal Impacts
(As per latest Syllabus)

6.1 INTRODUCTION

The digital world is changing at a tremendous pace. New communication technologies open up new possibilities but by using them you also expose yourself and others to various risks. Many people have trouble assessing these risks, especially with regard to the subject of safe digital communication. This is particularly true of people working in regimes with high levels of censorship. However, your data can be used or misused by others as well—governments, companies, or even other persons (sometimes even unintentionally).

Information Technology (IT) plays a central role in commerce, industry, government, education, entertainment and society at large. Its economic and social benefits hardly need

explanation. But like any other technology, IT also has problematic implications and some negative impacts on society. It poses and creates problems related to ethics and, in general, contains three main types of ethical issues: personal privacy, access right and harmful actions. Let us look closely at these issues, exploring in each case the ways in which they affect the public reactions to this technological change.

How to protect yourself, your sources, or your friends? What are the safe routes to take? How do you secure your personal data? This chapter aims to address these issues to help you choose your own 'level' of safety. We will also learn about the laws which can provide us with complete safety while working online. This chapter also discusses computer ethics or standards of conduct

pertaining to computers. It also deals with how computing professionals should make decisions regarding professional and social conduct.

Basically, computer ethics describe the difference between ethical and unethical. *For example*, while it is easy to duplicate copyrighted electronic (or digital) content, computer ethics would suggest that it is wrong to do so without the author's approval. And while it may be possible to access someone's personal information on a computer system, computer ethics would advise that such an action is unethical.

Let us discuss these topics in detail.

6.2 DIGITAL FOOTPRINTS

This is the term used to describe the trail, traces or "footprints" that people leave online. Digital footprints are recorded with each and every online activity we perform—whether it is interactions on social media, chatting and connecting with friends on social media sites, online shopping, locations through Facebook check-ins, etc.

These digital footprints can be compared to the footprints that you leave behind when walking along a beach. Every step you take leaves an impression on the sandy surface that allows another person to see and follow your marks. Digital footprints are also termed as **Digital Tattoos**. However, a beach footprint gets washed away after some time whereas digital footprints stay forever and cannot be done away with. These are classified into two types:

1. Active digital footprint
2. Passive digital footprint

An **active digital footprint** is created when a user intentionally shares their personal information either through social media platforms or websites and apps. Examples of active digital footprints are:

- Sharing of personal information on Facebook, Instagram, Twitter and other social media platforms.
- Working with online forms, such as signing up while logging on to mail accounts to send or receive emails or text messages.
- Accidental/intentional acceptance to install cookies on user's respective devices when prompted through the browser.

A **passive digital footprint** is created when information is collected from the user without their knowledge. Examples of passive digital footprints are:

- Websites that install cookies in user's device without their disclosure.
- Apps and websites that use geo-location to detect user's location.
- Social media news channels and advertisers that use comments, likes and shares of the user to know about their profiles and areas of interest, based on which they serve the advertisements as per the user's requirements and interests.

Digital footprints should be properly managed so as to safeguard your personal information, freedom and to prevent any possibility of incurring financial loss. If proper care is not taken, digital footprints may result in online threats such as scams, identity theft, privacy issues and fake websites. So, one must take due care while working online and especially while doing financial transactions. Therefore, following measures are to be adopted to manage digital footprints:

- Enter name into several search engines.
- Double-check privacy settings; don't trust unsafe sites (sites without https prefix).
- Create strong, memorable passwords.
- Keep all the software up to date.
- Review the mobile usage. Delete unwanted and obsolete (.temp-temporary) files from your device.
- Build reputation through behaviour.

6.3 NET AND COMMUNICATION ETIQUETTES

Netiquette is the short form of "internet etiquette" or communication etiquettes over the internet. Just like etiquette—a code of polite behaviour in society—netiquette is a code of good behaviour while working on the internet. It includes several aspects of the internet, such as social media, email, online chat, web forums, website comments, multiplayer gaming and other types of online communication.

While working online, netiquettes revolve around the general idea of respecting others online. Following are the Do's and Don'ts that one should follow while working on the internet.

Do's	Don'ts
Keep messages and posts brief.	Post inflammatory/offensive comments.
Reread your posts or emails to make sure they say what you intend.	Write in ALL CAPS. It is considered as SHOUTING on the net.
Remember that you leave a digital footprint. So, be careful what you write.	Respond to internet trollers/personal attacks.
Use discretion.	Post private or embarrassing images/comments.
Include a subject line in an email.	Use sarcasm. It might be misinterpreted.
Protect personal information.	Violate copyright laws. Make sure your work is your own or properly cited.
Obey copyright laws.	Exclude people or talk behind their backs.
Stay focused and stick to the topic.	Spam others by sending large number of unsolicited emails.

6.4 DATA PROTECTION

Data protection addresses the practices, safeguards, and binding rules put in place to protect your personal information and ensure that it remains in control while working online. In short, you should be able to decide what information should be shared, who has access to it, for how long and for what reason, and whether rights to modify some of this information should be given or not.

Data protection secures sensitive data—from databases to big data, cloud, file system and other crucial applications. One must ensure that there shouldn't be any kind of unprotected data or data breach as it may contribute to several serious issues:

- Physical data loss
- Attack by a virus, worm or malicious code
- Targeted by hackers
- DDoS(Distributed denial of service) attacks
- Loss of money
- Intellectual property at risk

Thus, it is mandatory to protect data. This can be achieved by taking proper and careful measures while working online in the following ways:

- Through Data Encryption.
- Not sharing private information such as passwords, credit card credentials, etc.
- Not oversharing on social networking sites using public domain.
- Adopting complete security mechanism for protection against possible threats.
- Avoiding opening phishing emails.
- Being careful about using Wi-Fi connections.
- Being alert to impersonators.

6.5 INTELLECTUAL PROPERTY RIGHTS (IPR)

You must have heard the word property; it is generally used to mean a possession or, more specifically, something on 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. 6.1(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. 6.1(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.

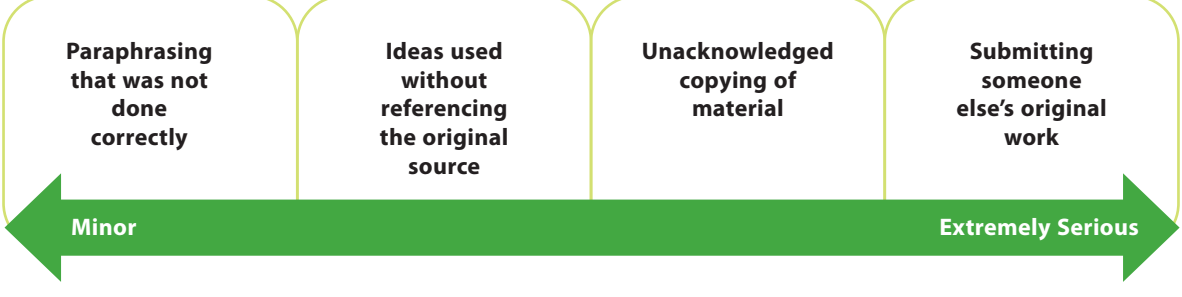
6.6 PLAGIARISM

Plagiarism is copying someone else's work and then passing it off as one's own. It is morally wrong because it is an act of stealing. In other words, it is copying information and not giving the author credit for it.

Fig. 6.2: 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.

6.7 LICENSING AND COPYRIGHTING

Software Licensing is the legal right to run or the privilege given to you by a company to access their application (or program). A common example of licensing that we observe in our daily lives is the licence we purchase for proprietary software such as Windows OS to get them installed on our computer system as shown in Fig. 6.3 (a).

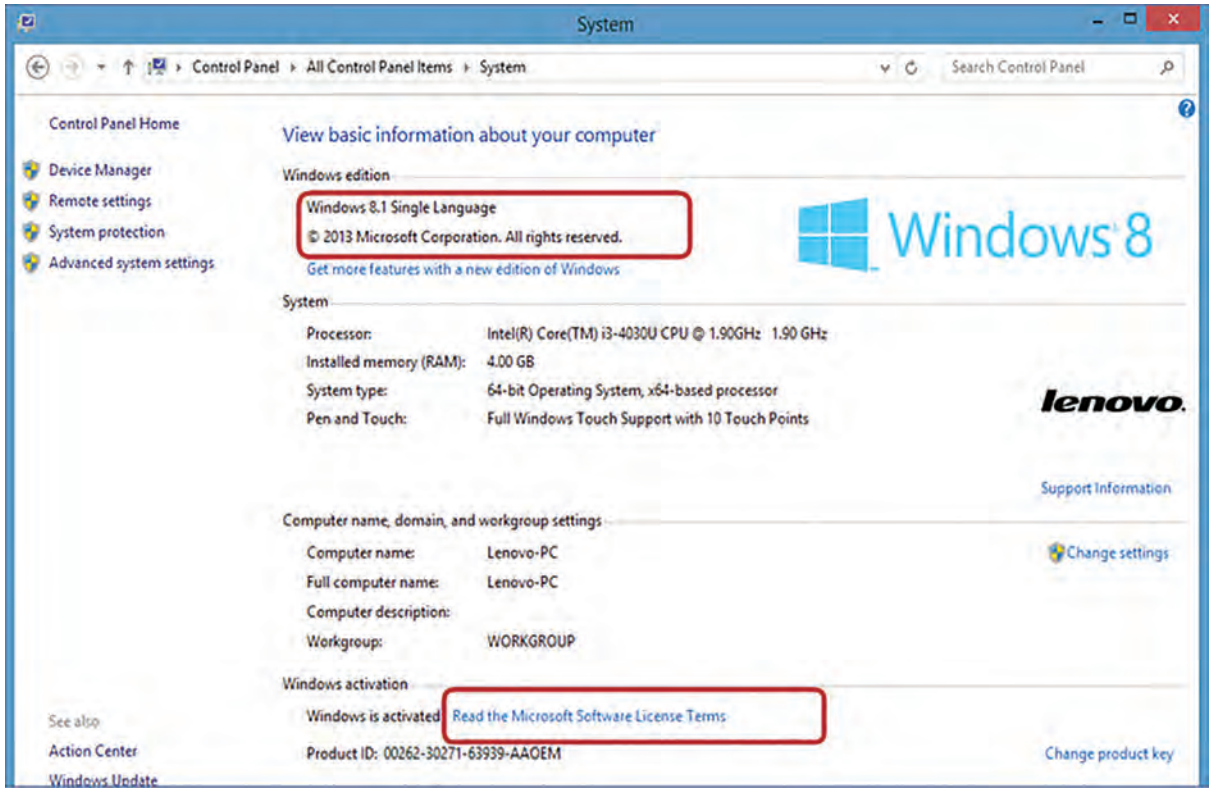


Fig. 6.3(a): Software Licensing

The given figure shows Windows 8 as a copyright of Microsoft, which indicates that this software is a licensed product of Microsoft and has to be legally purchased for use. It comes with a licence agreement which is to be read first and to be agreed upon for the successful installation and usage of the software under the link "**Read the Microsoft Software Licence Terms**".

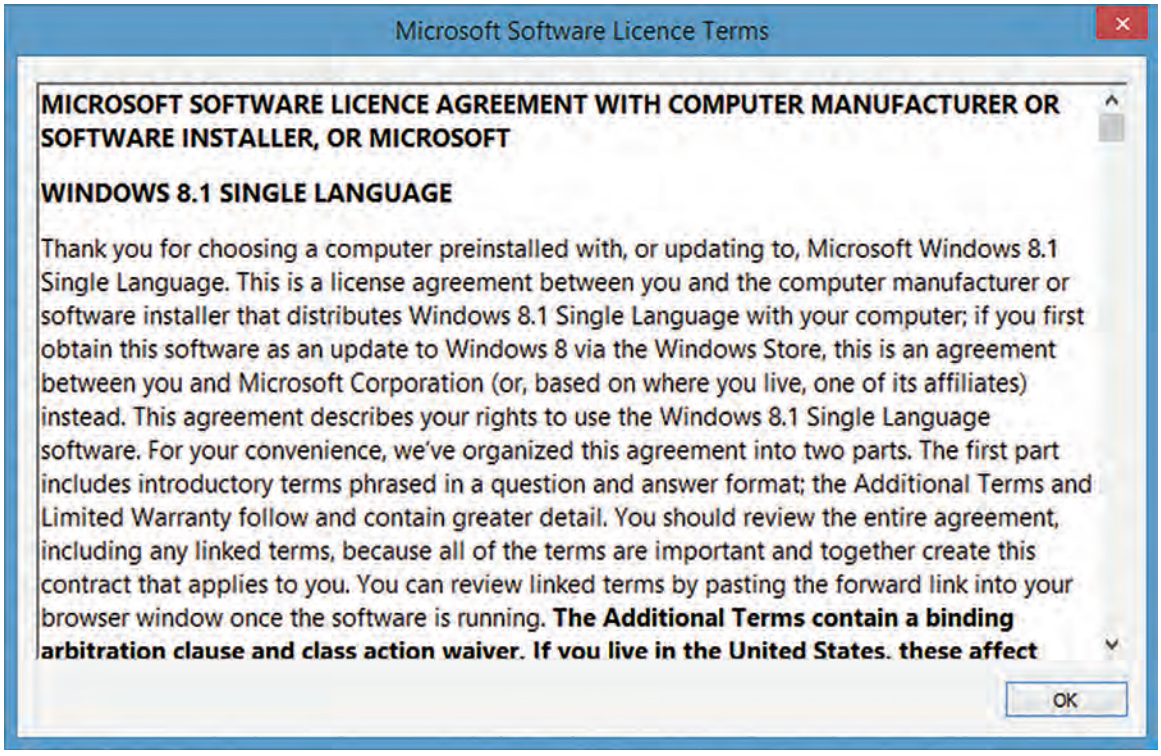


Fig. 6.3(b): A Licensing Agreement

As soon as you click on this link, you get to see a licence agreement (Fig. 6.3 (b)) with terms and conditions. A licence agreement governs the use of licensed software. Licence agreements typically allow the software to run on a limited number of computers and allow copies to be made, though solely for backup purposes.

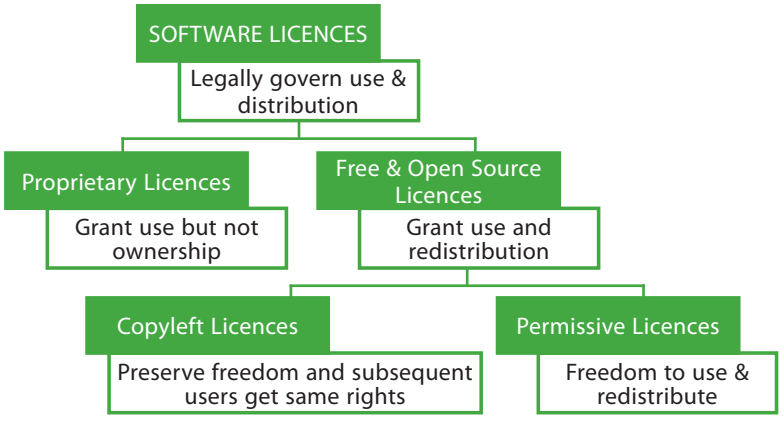
Advantages of using Licensed Software

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 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 as 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 the 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.

OSS is further categorized as Permissive and Copyleft licences. Permissive Licences provide a royalty-free licence to do virtually anything with the source code. They permit using, copying, modifying, merging, publishing, distributing, sublicensing, and/or selling, but distribution can only be made without the source code as source code modifications can lead to permissive licence violation, *for example*, Apache, W3C, etc. In case of Copyleft licences, source code has to be provided. Also, distribution and modification of source code is permitted. Examples include General Public Licence (GPL), Creative Commons Licence, Lesser General Public Licence (LGPL), Mozilla Public Licence (MPL), etc.

Let us discuss important features of OSS licences.

GPL (General Public Licence) has the following characteristics:

1. **Copying the Software:** There's no limit to copying the code. You can copy it on your own server, on your client's server, on your local workstations, anywhere and as many times as you want.
2. **Distribution:** You can distribute it in your thumb or hard drives, you can distribute the code under this licence with a download link on your website, or you can print the code on paper; in fact whatever form of distribution you prefer.

3. **Charge a Fee:** You can charge someone for the software but remember to give them a copy of GNU–GPL, which would tell them that they can get the software free from elsewhere. This also gives you a chance to tell them why you are charging for it.

A **CC (Creative Commons) Licence** has four basic parts:

1. **Accreditation:** Author must be acknowledged as the creator of the work. The work can then be modified, distributed, copied and used otherwise.
2. **Shared with CC:** The work can be modified and distributed but only under CC Licence.
3. **Non-Commercial:** Work can be modified and distributed but not for commercial purposes. The word "commercial" is a bit vague in its meaning since no solid-lined definition is available.
4. **No Derivative Works:** You can copy and distribute the licensed work but you can't modify it in any way or create work based on the origin.

APACHE Licence offers the following salient features:

1. **Rights are Never-ending:** Once the rights under Apache Licence have been granted, you can continue to use them forever. There's no need to renew them.
2. **Worldwide Authority of Rights:** Even if rights are granted to one country, automatically, they are considered as granted to all countries.
3. **Rights for No Fee or Royalty:** Any charge, neither upfront nor per usage, or on any other basis is applicable.
4. **Rights are Irrevocable:** No one can ever tell you that your derivative of the code that was licensed under this licence can't be in use anymore. (A clause in the licence states that if you sue someone over patent infringement on anything under this licence, then your licence is terminated, but that only applies to patented work, and as long as you don't sue anyone over the work, you won't have to worry about it.)

6.7.1 Difference between Licensing and Copyright

Copyright is a type of intellectual property protection and licensing is a kind of risk control measure that can be applied to control copyright and less exposure, so the licensor (copyright owner) can grant permission that usually takes the form of a licensing agreement to use its copyrighted material. This agreement specifies the exact material to be used, purpose of the work and the duration of the licence.

6.8 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.

6.8.1 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.

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 anti-virus 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.

6.8.2 Hacking

Hacking is the process of gaining unauthorized access to a computing device or a group of computer systems. This is done through cracking of passwords and codes which gives access to the systems.

The difference between a hacker and a cracker is that while a cracker breaks the security of computer systems, a hacker likes to explore computer systems and master them.

Types of Hackers

Black Hat Hackers or Crackers are individuals with extraordinary computing skills, resorting to malicious/destructive activities. Black hat hackers use their knowledge and skills for their own personal gains probably by hurting others.

White Hat Hackers are individuals who use their hacking skills for defence. This means that white hat hackers use their knowledge and skills for the common good. Ethical hacking, also known as penetration testing or white hat hacking, involves the same tools, tricks and techniques that hackers use, but with one major difference—ethical hacking is legal.

Grey Hat Hackers are individuals who work both offensively and defensively at different times. Their behaviour cannot be predicted. Sometimes they use their skills for the common good while at other times they use them for personal gains.

How to Prevent Hacking

- Download software from authorized websites only.
- Scan all types of hard drives before running.
- Keep strong passwords.
- Never store or share login information.
- Do not click and open random email attachments.

6.8.3 Cyber Bullying

Cyber bullying is the attack upon an individual or group through the use of electronic means such as instant messaging, social media, email and other forms of online communication with the intent to abuse, intimidate or overpower.

Following acts are considered as cyber bullying:

- Posting any kind of humiliating content about the victim.
- Hacking the victim's account.
- Sending or posting vulgar messages online.
- Threatening to commit acts of violence.
- Stalking by means of calls, messages, etc.
- Threats of child pornography.

Cyber bullying differs from in-person bullying:

- **More difficult to recognize** – Bullying conducted via text or online medium can more easily go unnoticed.
- **More relentless** – Cyber bullying doesn't end at school and can reach the child's home.
- **More enduring** – It leaves a paper trail that can follow both the bully and the victim for years.

Different Types of Cyber Bullying

The various forms of cyber bullying are explained as:

- **Doxing** – Publishing revealing personal information about an individual online, for the purpose of defaming, humiliating or harassing the victim.
- **Harassment** – Posting threatening, hurtful or intimidating messages online, or sending them directly to someone, with the intention of harming that person.

- **Impersonation** – Creating fake accounts or gaining access to a person's real social media accounts and posting things to damage the victim's reputation.
- **Cyberstalking** – Tracking and monitoring a person's online activity and using the internet to stalk or harass an individual.

How to Prevent Cyber Bullying

The victims of cyber bullying are usually young children. So parents must ensure proper vigilance and care towards their children who are hooked to the internet.

- Be aware of the child's online activities.
- Watch for the following signs of cyber bullying in children:
  - Refusal to allow to see what they are doing online
  - Refusal to discuss what they are doing online
  - Sudden, unexplained increase or decrease in online activity
  - Deactivating social media accounts
  - Emotional responses (including sadness, anger, happiness) linked to their device usage
- Adults should also teach children to recognize and be aware of the signs of cyber bullying themselves.

6.9 CYBER LAW AND INFORMATION TECHNOLOGY ACT, 2000

Cyber law is the part of the legal system that deals with cyberspace, internet and legal issues. It covers a broad area, like freedom of expression, access to and utilization of internet, and online security or online privacy. It is commonly known as the law of the web.

What is the importance of Cyber Law

Cyber law plays a very important role in this new approach to technology. It is important as it is concerned with almost all aspects of activities and transactions that take place either on the internet or other communication devices. Whether we are aware of it or not, each action and each reaction in cyberspace has some legal and cyber-legal views.

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.

6.9.1 IT Act Amendments

Information Technology Act is a set of recent legal enactments, currently existing in India, which provide legal support to computer users against cybercrime.

The cyber police work as a detector to discover cybercrimes. They have the authority in respect of all the offences committed under TITA (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. 6.4.

SECTION	OFFENCE	PENALTY
67A	Publishing images containing sexual acts	Imprisonment up to seven years, and/or fine up to ₹ 10,00,000.
67B	Publishing child porn or predating children online	Imprisonment up to five years and/or with fine up to ₹ 10,00,000 on 1st conviction. Imprisonment up to seven years and/or fine up to ₹ 10,00,000 on 2nd conviction.
67C	Failure to maintain records	Imprisonment up to three years and/or fine up to ₹ 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 ₹ 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 with fine up to ₹ 1,00,000.

Fig. 6.4: IT ACT, 2000 (Cyber law) of India

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.

6.10 E-WASTE MANAGEMENT

Whenever an electronic device covers up its working life, or becomes non-usable due to technological advancements or becomes non-functional, it is not used anymore and comes under the category of e-waste or electronic waste. As the technology is changing day by day, more and more electronic devices are becoming non-functional and turning into e-waste. Managing such non-functional electronic devices is termed as e-waste management.

All **electronic waste** is made up of deadly chemicals such as lead, cadmium, beryllium, mercury and brominated flame retardants. Disposing of gadgets and

devices improperly increases the chances of these dangerous chemicals contaminating the soil, polluting the air and leaching into water bodies.

When e-waste is deposited in a landfill, it tends to leach when water passes through it, picking up trace elements. The contaminated landfill water then reaches natural groundwater with increased toxic levels and can be very harmful if it enters any drinking water source. Thus, e-waste if not managed properly causes several problems to the environment as well as to living organisms.

E-waste Hazards

On Environment

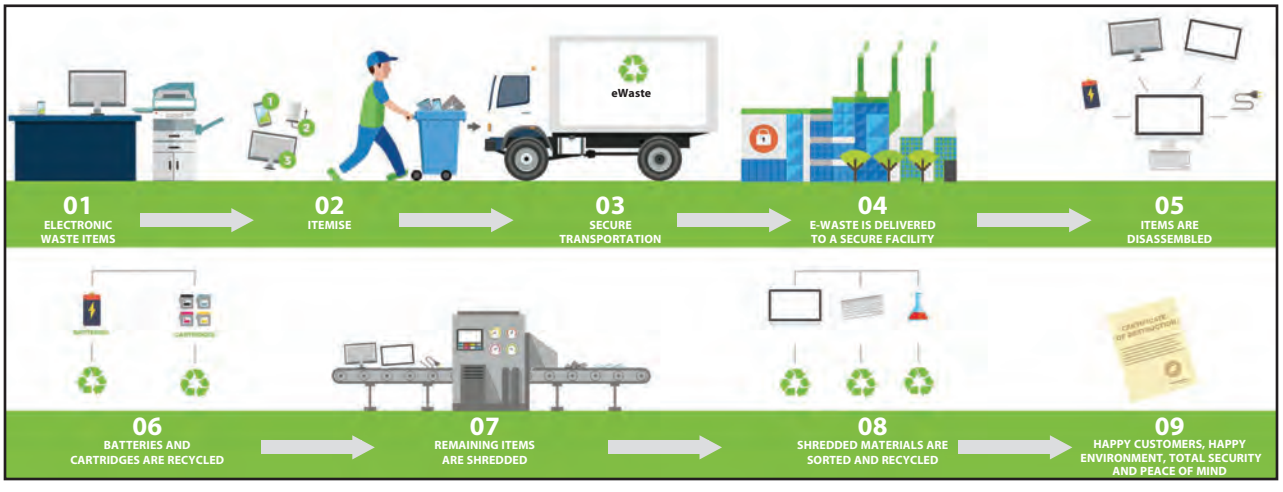
- Acidification of soil
- Air pollution
- Pollution of groundwater
- Landfills with lead and heavy metals

On Human Health

- Lung cancer
- DNA damage
- Asthmatic bronchitis
- Chronic brain damage
- Damage to the heart, liver and spleen

Despite having an eco-friendly approach, recycling usually leads to overseas shipping and dumping the gadgets in pits. Worse still, some recycling companies ship e-waste to Third World countries.

A lot of children in such countries earn their livelihoods by scavenging gold, silver, iron, and copper from the tech waste which is harmful to their health. Countries that are used as dumping grounds usually have high rates of cybercrime as the salvaged hard drives can give criminals direct access to your personal files and information.



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 local 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 or Leave at Drop-off Points**

A lot of electronic companies tend to have an exchange policy whereby they take back your old gadgets when you buy the latest 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 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 disastrous.

6.11 HEALTH CONCERNS RELATED TO OVERUSE OF TECHNOLOGY

Technology can have a large impact on users' mental and physical health. Being overly connected can cause psychological issues such as distraction, narcissism, expectation of instant gratification and even depression. Besides affecting users' mental health, use of technology can also have negative repercussions on physical health causing vision problems, hearing loss and neck strain. Thus, various physical and psychological disorders may crop up due to prolonged and continuous use of technology.

Physical Problems:

1. **Repetitive Strain Injury:** The pain exists even when resting and as a result it becomes very difficult to accomplish even easy and ordinary tasks.
2. **Carpal Tunnel Syndrome:** This is an illness caused by injuries that occur due to force on the median nerve found in the wrist. Its symptoms include tingling in hands and fingers, a feeling of lethargy, sudden pain in wrists and arms, and sometimes even in shoulders, neck and in the body.
3. **Computer Vision Syndrome:** Experts believe that people blink their eyes more frequently while using computers than they do otherwise and that this can cause various eye and vision-related problems.
4. **Radiation:** Computer screens produce radiations of various types. These radiations can cause headaches and inattentiveness.
5. Sleeping disorders and decrease in productivity
6. Loss of attention and stress

Psychological Disorders:

- Fear of technology
- Computer anxiety
- Internet addiction
  - **Ego surfing:** An illness of regularly searching for one's own name on the web and checking what information is available on the net.
  - **Infornography:** The word, derived from pornography and information, describes the state of "trying to soothe hunger for information on the net."
  - **Blog streaking:** A desire to spread information online that shouldn't be known to everybody.
  - **YouTube-Narcissism:** Constantly uploading one's own videos in order to introduce and make themselves known to others.
  - **Google-Stalking:** Trying to get information about all of one's relatives or acquaintances on the web.
  - **Photo lurking:** Looking at the photo albums of others' on the net.
  - **Wikipediholism:** Contributing to the internet encyclopaedia, Wikipedia, sending someone's own writings, and revising the present texts.

In order to avoid these problems, one should learn how to use these technologies in moderation rather than avoiding using them. Some of the users of computer technologies are not even aware of the health-related problems that they face. Therefore, necessary precautions should be taken in this regard by all long-time internet users.

MEMORY BYTES

- Cyber Safety refers to the safe and responsible use of internet to ensure safety and security of personal information and not posing threat to anyone else's information.
- Cookies are small text files—bits of information—left on your computer by websites you have visited which let them 'remember' things about you.
- Confidentiality of information signifies that only authorized users get access to sensitive and crucial information.
- Cyber trolling is an act of cybercrime where a person intentionally starts arguments or upsets others by posting provocative comments.
- A trojan horse is a program which is intended to carry out malicious operations without the knowledge of the user.
- A worm is a computer program which copies itself across a network.
- Phishing is the attempt to acquire sensitive information such as usernames, passwords, 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 will be encrypted.
- You should set up privacy settings for a social media site on your own.
- All software is protected under copyright. Copyright is owned by the developer or publisher.
- FOSS grants a 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 cannot 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.

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 these

(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 the 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, the 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 law of internet.

4. Write two applications of cyber law. [CBSE Delhi 2015]

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. 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)

6. 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.

7. Name the crimes for which cyber laws are enforced strictly in India.

Ans. These are:

(a) cyber crimes, (b) electronic and digital signatures,
(c) intellectual property, and (d) data protection and privacy.

8. 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.

9. 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.

10. 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 type of windfall if you pay money upfront is another common phishing scam.

11. 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.

12. (a) Write down names of any three social networking sites.

Ans. Facebook, Twitter, LinkedIn, Google+ (any three).

(b) What are the merits of social networking?

Ans.
- Lowest cost form of marketing
- Huge potential audience and the possibility of messages going viral
- Offers a closer connection with your clients
- Source of instant feedback

(c) What are the demerits of social networking?

Ans.
- Unreliable information
- Lack of control
- Can be addictive

(d) What is cyber trolling?

Ans. Internet trolls hide behind their computer screens and actively go out of their way to cause trouble on the internet.

(e) Write down the medium/ways of trolling.

Ans. YouTube video comments, blog comments, forums, email, FB, Twitter, Instagram, social networking sites and anonymous ways of networking.

(f) What is cyber stalking?

Ans. Cyber stalking is defined as the unlawful act of harassing a person or collecting an individual's private information using electronic network.

13. 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 changing too much 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.

14. 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 has an effective recycling program in all of its stores. They claim to use only recyclers which adhere to the highest standards of e-waste processing. Specifically, their website discloses that 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.

15. 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 holders. Hence, illegal downloads result in digital piracy.

16. 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 codes and can be used to deliver a spammer's message in spite of filters.

17. How can we block web beacons? [HOTS]

Ans. (a) The best defence against web beacons is to prevent pictures from downloading until you have 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.

UNSOLVED QUESTIONS

1. What do you mean by cybercrime?
2. What is meant by cyber ethics? How are they important for us?
3. What do you understand by illegal download?
4. What is phishing?
5. What is cyber law?
6. How is cyber law implemented in India?
7. What is IT Act, 2000 of India?
8. What amendments were introduced in IT Act in the year 2008?
9. How do people unknowingly commit cybercrime?

10. What are the different methods to do phishing?
11. What precautions should be taken by people to avoid phishing?
12. What procedure is followed by the police to track/investigate cybercrime cases?
13. What precautions should be taken so that students do not indulge in cybercrime unknowingly?
14. Define e-waste. What are the various methods for effective e-waste management?
15. Differentiate between Proprietary Licence and Open Source Licence.
16. How is internet affecting one's daily life? Give one example each of positive and negative aspects of internet.
17. Name any three areas where internet addiction is advantageous.
18. List at least three emotional and physical symptoms of internet addiction.

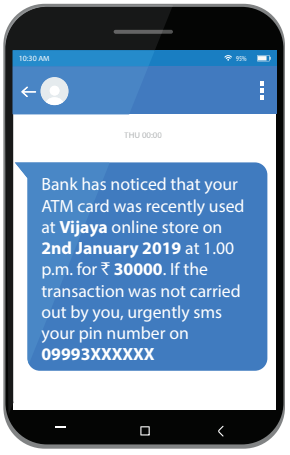
CASE-BASED/SOURCE-BASED INTEGRATED QUESTIONS

1. Jeet has to prepare a project on "Swachh Bharat Abhiyan". He decides to gather information from the internet. He downloads three web pages (webpage 1, webpage 2, webpage 3) containing information on Swachh Bharat Mission. Which of the following steps taken by Jeet is/are an example of plagiarism or copyright infringement. Give justification in support of your answer.
   (a) He read a paragraph on "Swachh Bharat Abhiyan" from webpage 1 and rephrased it in his own words. He finally pasted the rephrased paragraph in his project.
   (b) He downloaded three images of "Swachh Bharat Abhiyan" from webpage 2. He made a collage for his project using these images.
   (c) He downloaded "Swachh Bharat Abhiyan" icon from web page 3 and pasted it on the front page of his project report.

Ans. All of the above steps (a, b and c) taken by Jeet shall be considered as plagiarism or copyright infringement as they are in violation of copyright laws, which state that one should use one's own words and ideas. However, Jeet has copied the ideas along with the images from a website without giving reference to the author or the website.

It is necessary to provide some reference or give credit to the source from where you have received your information. Since he has used information from a website, he should have cited the name of the website, its URL or the name of the author along with proper acknowledgement.

2. Angad received an SMS, shown alongside, from his bank querying a recent transaction. Answer the following questions as what he should do upon receiving this SMS:
   (a) Should he SMS his pin number to the given contact number?
   (b) Should he call the bank helpline number to recheck the validity of the SMS received?



Ans. (a) If he responds by sending his pin number, then there is an absolute chance of hacking of his bank account which can result in a huge financial loss to him.
   (b) He should call the bank helpline number to recheck the validity of the SMS received since this is an act of phishing (cybercrime) which is aimed at stealthily getting information from the user.

8 More on SQL (Additions) (Database Query using SQL)

8.7 AGGREGATE FUNCTIONS IN SQL

SQL provides a large collection of in-built functions, also called library functions, that can be used directly with SQL statements for performing calculations on data. These functions are the Aggregate functions. They summarize the results of a query and return a single value calculated from values in a column instead of providing the listing of all of the rows.

Aggregate functions help to summarize large volumes of data. These functions result in a single value for an entire group or table.

Syntax:

select <function>(column_name) from <table_name>;

While working with standard library functions, the following points must be kept in mind:

(a) The name of the column on which the function is to be executed must be enclosed within parentheses.

(b) Only one column can be specified within one set of parentheses.

(c) To use more than one function with the same select statement, functions are written one after the other separated by a comma.

(d) If the same function is required to be used for more than one column, then again the function name has to be repeated in the select statement.

8.7.1 Mathematical/Numeric Functions

Mathematical functions perform mathematical operations on numeric values. The most commonly used mathematical functions are pow(), mod(), round() and aggregate functions, which have been discussed in the chapter.

Let us discuss these functions first.

☞ **pow():** Returns the argument raised to the specified power. pow() works the same way. Its syntax is:

pow(m,n)

Here, m is the number and n is the power to be raised to number 'm'.

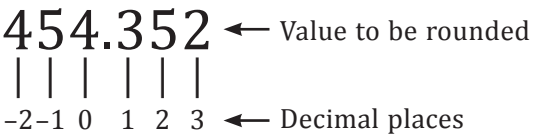
For example, (i) select pow(2,4); Result: 16
(ii) select pow(2,–2); Result: 0.25
(iii) select pow(–2,3); Result: –8
(iv) select pow(2.37,3.45); Result: 19.6282……

☞ **round():** round(x) rounds the argument to the **0** decimal place, whereas round(x, d) rounds the argument to **d** decimal places. The round() function returns a number rounded to a certain number of decimal places.

Syntax: round(column_name, decimals)

column_name – Required (It is the field to be rounded off.)

decimals – Required (Specifies the number of decimals to be returned.)



Decimal place position value is rounded off to the next integer if the next number on the right side is greater than 5 (>=5).

Default decimal place is 0 if nothing is specified.

For example, (i) select round(–1.23); Result: –1
(ii) select round(–1.58); Result: –2
(iii) select round(1.58); Result: 2
(iv) select round(3.798, 1); Result: 3.8
(v) select round(1.298, 0); Result: 1
(vi) select round(23.298, –1); Result: 20

If the second argument is negative, *i.e.*, –1, –2, –3, then it rounds off to the nearest 10, 100, 1000 respectively.

```
mysql> select round(454.352,2);
+--------------------+
| round(454.352,2) |
+--------------------+
|             454.35 |
+--------------------+
1 row in set (0.00 sec)

mysql> select round(454.352,0);
+--------------------+
| round(454.352,0) |
+--------------------+
|                 454 |
+--------------------+
1 row in set (0.00 sec)

mysql> select round(454.352,-1);
+--------------------+
| round(454.352,-1) |
+--------------------+
|                 450 |
+--------------------+
1 row in set (0.00 sec)

mysql> select round(454.352,-2);
+--------------------+
| round(454.352,-2) |
+--------------------+
|                 500 |
+--------------------+
1 row in set (0.00 sec)
```

☞ **truncate():** Truncates the argument to specified number of decimal places.

For example, (i) select truncate(7.29,1); Result: 7.2
(ii) select truncate(27.29,–1); Result: 20

☞ **mod():** The mod() function returns the remainder of one number divided by another. The following is the syntax of the mod() function:

Syntax: mod(dividend, divisor)

dividend – Is a literal number or a numeric expression to divide.

divisor – Is a literal number or a numeric expression by which the dividend is to be divided.

For example, (i) select mod(11, 3); Result: 2
(ii) select mod(10.5, 3); Result: 1.5

◆ **Continued on Book page 8.3, Fig. 8.4 (Aggregate Functions in a Nutshell)**

8.7.2 String Functions

These functions are used to deal with the string type values. The various built-in String library functions are ascii(), lower(), upper(), len(), left(), right(), trim(), ltrim(), rtrim(), etc.

(a) **ascii():** Returns the ASCII code value of a character (leftmost character of string).

Syntax: ascii(character);

For example,

mysql> select ascii('a') from dual; returns 97.

mysql> select ascii('A') from dual; returns 65.

mysql> select ascii('1') from dual; returns 49.

mysql> select ascii('ABC') from dual; returns 65.

- The ASCII value for upper case characters 'A' to 'Z' is 65 to 90.
- The ASCII value for lower case characters 'a' to 'z' is 97 to 122 and for digits '0' to '9', the ASCII value is 48 to 57.

Note: If no table name is specified, then SQL uses Dual table which is a dummy table used for performing operations.

(b) **lower():** Converts character strings data into lower case.

Syntax: lower(string);

For example,

mysql> select lower("INFORMATION TECHNOLOGY");

Returns – information technology

(c) **upper():** Converts character strings data into upper case.

Syntax: upper(string);

For example,

mysql> select upper('information technology');

Returns – INFORMATION TECHNOLOGY

(d) **len():** Returns the length of the character string. It takes spaces between the strings into account for calculating the total length of the string passed as an argument to len().

Syntax: len(string);

For example,

mysql> select len('Information Technology');

Returns – 22

(e) **replace():** Replaces all occurrences of the second string (string2) in the first string (string1) with a third string (string3).

Syntax: replace('string1','string2','string3');

For example,

mysql> select replace('INFORMATION TECHNOLOGY','INFORMATION','LATEST');

Returns – LATEST TECHNOLOGY

Returns NULL if any one of the arguments is NULL.

(f) **left():** Returns leftmost characters from a string, passed as an argument, with the specified number of characters counting from left. left() function is used to retrieve portions of the string.

Syntax: left(string, integer);

For example,

mysql> select left('INFORMATION TECHNOLOGY', 6);

Returns – INFORM

(g) **right():** Returns rightmost characters from a string, passed as an argument, with the specified number of characters counting from right. right() function is used to retrieve portions of the string.

Syntax: right(string, integer);

For example,

select right('STRING FUNCTION', 8)

Returns – FUNCTION

(h) **ltrim():** Returns a string after removing leading spaces/blanks from the left side of the string passed as an argument.

Syntax: ltrim(string);

For example,

mysql> select ltrim(' LIBRARY FUNCTION');

Returns – LIBRARY FUNCTION

(i) **rtrim():** Returns a string after removing trailing spaces/blanks from the right side of the string passed as an argument.

Syntax: rtrim(string);

For example,

mysql> select rtrim('LIBRARY FUNCTION ');

Returns – LIBRARY FUNCTION

CTM: ltrim() and rtrim() functions return the string passed as an argument after removing all the leading and trailing spaces. But they do not delete the spaces in between the strings.

(j) **trim():** Returns a string after removing the spaces in between the string passed as the argument to it.

Syntax: trim(string);

For example,

mysql> select trim('LIBRARY FUNCTION');
Returns – LIBRARYFUNCTION

(k) **reverse():** Returns reverse of an inputted string.

Syntax: reverse(string);

mysql> select reverse('LIBRARY FUNCTION');
Returns – NOITCNUF YRARBIL

(l) **replicate():** Repeats an inputted string for a specified number of times.

Syntax: replicate(string, integer);

mysql> select replicate('FUNCTION', 3);
Returns – FUNCTIONFUNCTIONFUNCTION

(m) **substring():** Returns part of an inputted string. A substring() function retrieves a portion of the given string starting at the specified character (start_index) to the number of characters specified (length).

Syntax: substring(string, start_index, length);

mysql> select substring('STRING FUNCTION', 1, 6);
Returns – STRING

mysql> select substring('STRING FUNCTION', 8, 8);
Returns – FUNCTION

8.7.3 Date/Time Functions

- **curdate():** Returns the current system date.
  For example,
  select curdate(); Result: '2020-06-11'
- **now():** Returns the current date and time.
  For example,
  select now(); Result: '2020-06-11 13:58:11'
- **sysdate():** Returns the time at which the function executes.
  For example,
  select sysdate(); Result: '2020-06-11 13:59:23'
- **date():** Extracts the date part of a date or date-time expression.
  For example,
  select date('2020-06-11 01:02:03'); Result: '2020-06-11'
- **month():** Returns the month from the date passed.
  For example,
  select month('2020-06-11'); Result: 6
- **year():** Returns the year from the inputted date.
  For example,
  select year('2020-06-11'); Result: 2020

- **dayname():** Returns the name of the weekday.
  For example,
  select dayname('2020-06-11'); Result: THURSDAY
- **dayofmonth():** Returns the day of the month (0-31).
  For example,
  select dayofmonth('2020-06-11'); Result: 11
- **dayofweek():** Returns the weekday index of the argument.
  For example,
  select dayofweek('2020-06-11'); Result: 5 (Sunday is counted as 1)
- **dayofyear():** Returns the day of the year (1-366).
  For example,
  select dayofyear('2020-06-11'); Result: 202

8.8 SQL JOINS

An SQL JOIN clause is used to combine rows from two or more tables based on a common field between them. While querying for a join, more than one table is considered in FROM clause. The process/function of combining data from multiple tables is called a JOIN. SQL can extract data from two or even more than two related tables by performing either a physical or virtual join on the tables using WHERE clause.

The types of SQL joins are as follows:

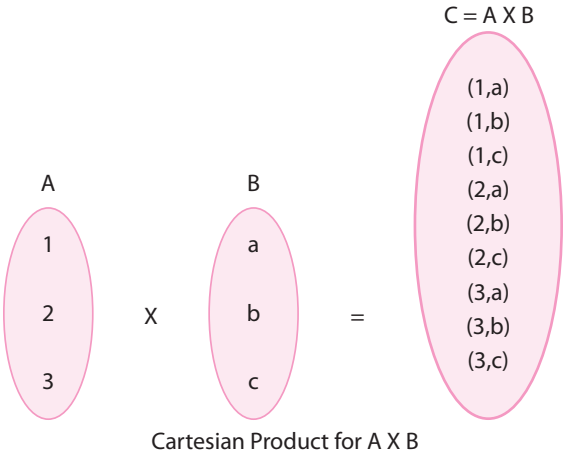
1. Cartesian Product (Cross Product)
2. Equi Join
3. Inner Join
4. Outer Join
5. Self Join
6. Non-Equi Join
7. Natural Join

1. Cartesian Product (Cross Product)

The Cartesian product is also termed as cross product or cross-join. The Cartesian product is a binary operation and is denoted by (X). The degree of the new relation formed is the sum of the degrees of two relations on which Cartesian product is performed. The number of tuples in the new relation is equal to the product of the number of tuples of the two tables on which Cartesian product is performed.

For example,

If A={1,2,3} and B={a,b,c}, find A X B.



Cartesian Product for A X B

Table: student

Rollno	Name
1	Rohan
2	Jaya
3	Teena

Table: games

gameno	gname
10	Football
11	Lawn Tennis

Cartesian product for student X games:

mysql> select Name, gname from student, games;

Therefore, a Cartesian product is formed when no join conditions exist or are invalid. When we perform a Cartesian product between two tables, all the rows in the first table are joined to all the rows in the second table. Using a Cartesian product operation results in a large number of rows as the output, so it is seldom used.

student X games

Rollno	Name
Rohan	Football
Jaya	Football
Teena	Football
Rohan	Lawn Tennis
Jaya	Lawn Tennis
Teena	Lawn Tennis

2. Equi Join

An Equi join is a simple SQL join condition that uses the equal to sign (=) as a comparison operator for defining a relationship between two tables on the basis of a common field.

Syntax for Equi join:

```
select <column1>, <column2>,....
from <table1>, <table2>
where <table1.column1> = <table2.column2>;
```

For example,

Table: student

Rollno	Name
1	Rohan
2	Jaya
3	Teena
4	Diksha

Table: fees

Rollno	Fee
4	4500
2	5500
3	5000

```
mysql> select A.Rollno, Name, fee
       from student A, fees B
       where A.Rollno = B.Rollno;
```

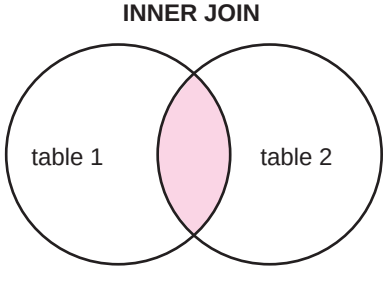
Resultant Table

Rollno	Name	Fee
4	Diksha	4500
2	Jaya	5500
3	Teena	5000

In the given SELECT statement, A and B are the alias names.

3. Inner Join

The Inner join is a classification of Equi join where either of the equivalent queries gives the intersection of two tables, *i.e.*, it returns the rows which are common in both the tables.




```
mysql> select student.Rollno, Name, fee
       from student, fees where student.Rollno = fees.Rollno
       order by student.Rollno;
```

OR

```
mysql> select A.Rollno, Name, fee
       from student A, fees B where A.Rollno = B.Rollno
       order by A.Rollno;
```

Thus, the output for the above command will be:

Resultant Table

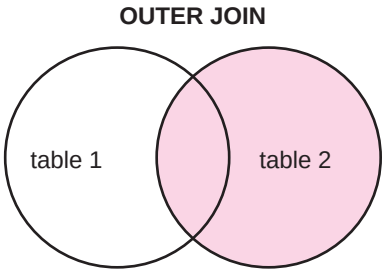
Rollno	Name	Fee
2	Jaya	5500
3	Teena	5000
4	Diksha	4500

4. Outer Join

The Outer join keyword returns all rows from the right table (table 2), with the matching rows in the left table (table 1). The result is NULL in the left side when there is no match.

mysql> select A.Rollno, Name, fee from student A, fees B

where A.Rollno = B.Rollno order by B.fee desc;



Resultant Table

Rollno	Name	Fee
2	Jaya	5500
3	Teena	5000
4	Diksha	4500

Note: Self, Non-equi and Natural join are beyond the scope of this book.

Let us take an example of two tables to explain the concept of MySQL Joins:

Table: Vehicle

CODE	VTYPE	PERKM
101	VOLVO BUS	160
102	AC DELUXE BUS	150
103	ORDINARY BUS	90
105	SUV	40
104	CAR	20

Note:

- PERKM is Freight Charges per kilometre
- VTYPE is Vehicle Type

Table: Travel

NO	NAME	TDATE	KM	CODE	NOP
101	Janish Kin	2015–11–13	200	101	32
103	Vedika Sahai	2016–04–21	100	103	45
105	Tarun Ram	2016–03–23	350	102	42
102	John Fen	2016–02–13	90	102	40
107	Ahmed Khan	2015–01–10	75	104	2
104	Raveena	2016–05–28	80	105	4

Note:

- NO is Traveller Number
- KM is Kilometres travelled
- NOP is number of travellers in vehicle
- TDATE is Travel Date

(a) To display NO, NAME, TDATE from the table TRAVEL in descending order of NO.

Ans. select NO, NAME, TDATE from TRAVEL order by NO DESC;

(b) To display the NAME of all the travellers from the table TRAVEL who are travelling by vehicle with code 101 or 102.

Ans.
```
select NAME from TRAVEL
where CODE='101' OR
CODE='102'; OR
select NAME from TRAVEL
where CODE=101 OR CODE=102;
```

(c) To display the NO and NAME of those travellers from the table TRAVEL who travelled between '2015–12–31' and '2016–04–01'.

Ans.
```
select NO, NAME from TRAVEL
where TDATE >= '20160401' AND TDATE <= '20151231'; OR
select NO, NAME from TRAVEL
where TDATE BETWEEN '20160401' AND '20151231';
```

(d) To display the CODE, NAME, VTYPE from both the tables with distance travelled (km) less than 90 km.

Ans.
```
select A.CODE, NAME, VTYPE
from TRAVEL A, VEHICLE B
where A.CODE=B.CODE AND KM<90;
```

(e) To display the NAME and amount to be paid for vehicle code as 105. Amount to be paid is calculated as the product of KM and PERKM.

Ans.
```
select NAME,
KM*PERKM from TRAVEL A, VEHICLE B
where A.CODE=B.CODE AND A.CODE='105';
```

JOIN Operation using Table alias names

8.9 UNION

The UNION operator is used to combine the result-set of two or more SELECT statements. The Union operation is used to return all the distinct rows selected by either query. For executing Union between two tables, the number of columns selected from each table should be the same. Also, the datatypes of the corresponding columns selected from each table should be the same.

Table: boys

Rollno	Name
1	Rohan
12	Jayant
3	Tinku

Table: girls

Rollno	Name
6	Reema
10	Jaya

mysql> select Name from boys where Rollno < 12
union select Name from girls where Rollno > 6;

The resultant table is:

Name
Rohan
Tinku
Jaya

mysql> select Rollno, Name from boys
union select Rollno, Name from girls order by girls.name;

The resultant table is:

Rollno	Name
10	Jaya
12	Jayant
6	Reema
1	Rohan
3	Tinku

MEMORY BYTES

- The ORDER BY keyword in MySQL is used to sort the result-set in ascending or descending order.
- The ORDER BY keyword sorts the records in ascending order by default. To sort the records in descending order, use the DESC keyword.
- An aggregate function is a function where the values of multiple rows are grouped together as input based on a certain criterion to form a single value of more significant meaning.

SOLVED QUESTIONS

35. How would you display system date as a result of a query?

Ans. Mysql> select curdate();

36. How would you calculate 13*15 in SQL?

Ans. Mysql> select 13*15;

37. Define a function.

Ans. A function is a special type of predefined command set that performs some operations and returns a single value.

38. What will be the output of the following code?
mysql> select concat (concat ('Inform', 'atics'),'Practices');

Ans. Informatics Practices

39. What will be the output of the following code?
mysql> lcase ('INFORMATICS PRACTICES CLASS 11TH');

Ans. informatics practices class 11th

40. What will be the output of the following code?
mysql> select concat (lower ('Class'), upper('xii'));

Ans. classXII

41. If Str="INFORMATICS PRACTICES" and Str1=" FOR CLASS XI", then give the SQL command to display the output string as “INFORMATICS PRACTICES FOR CLASS XI”.

Ans. select concat(Str,Str1);

42. (i) Create and open Database named MYORG.

Ans. create database MYORG;

use MYORG;

(ii) Write a command to display the name of the current month.

Ans. select month(curdate());

(iii) Write commands to display the system date.

Ans. select sysdate();

(iv) Write a query to find out the result of 6^3.

Ans. select pow(6,3);

(v) Write a command to show the tables in the MYORG Database.

Ans. use MYORG;

show tables;

(vi) Add one column State of data type VARCHAR and size 30 to table DEPT.

Table: DEPT

DeptID	DeptName	MgrID	Location
10	SALES	8566	Mumbai
20	PERSONNEL	8698	Delhi
30	ACCOUNTS	8882	Delhi
40	RESEARCH	8839	Bengaluru

Ans. alter table DEPT add(state varchar(30));

43. Write SQL commands for the following on the basis of the given table CLUB.

Table: CLUB

COACH_ID	COACH NAME	AGE	SPORTS	DATA OF APP	PAY	SEX
1	KUKREJA	35	KARATE	1996-03-27	1000	M
2	RAVINA	34	KARATE	1998-01-20	1200	F
3	KARAN	34	SQUASH	1998-02-19	2000	M
4	TARUN	33	BASKETBALL	1998-01-01	1500	M
5	ZUBIN	36	SWIMMING	1998-01-12	750	M
6	KETAKI	36	SWIMMING	1998-02-24	800	F
7	ANKITA	36	SQUASH	1998-02-20	2200	F
8	ZAREEN	37	KARATE	1998-02-22	1100	F
9	KUSH	41	SWIMMING	1998-01-13	900	M
10	SHAILYA	37	BASKETBALL	1998-02-19	1700	M

(a) To show all information about the swimming coaches in the club.

(b) To list names of all coaches with their date of appointment (DATE OF APP) in descending order.

(c) To display a report showing coach name, pay, age and bonus (15% of pay) for all the coaches.

(d) Give the output of the following SQL statements:

(i) select lcase(SPORTS) from CLUB;

(ii) select mod(age,5) from CLUB where SEX='F';

(iii) select pow(3,2) from CLUB where SPORTS='KARATE';

Ans. (a) select * from CLUB where SPORTS= 'SWIMMING';

(b) select COACHNAME DATE OF APP from CLUB order by DATE OF APP desc;

(c) select COACHNAME, PAY, AGE, PAY*0.15 as "BONUS" from CLUB;

(d) (i) lcase(SPORTS)

karate
karate
squash
basketball
swimming
swimming
squash
karate
swimming
basketball

(ii) mod(age,5)

4

1

1

2

(iii) 9

UNSOLVED QUESTIONS

22. Define a function. Why are they useful?

23. Write commands to display the system date.

24. Write a command to display the name of the current month.

25. Write a command to print the day of the week of your birthday in the year 1999.

26. What is the difference between SYSDATE() and NOW() function?

27. Consider two fields—B_date, which stores the birth date, and J_date, which stores the joining date of an employee. Write commands to find out and display the approximate age of an employee as on joining date.

CASE-BASED/SOURCE-BASED INTEGRATED QUESTIONS

ABC Associates has over 2000 employees on its roll and deals with customer support services. Help the company to perform associated operations for calculating the salaries of their employees and to perform relevant analysis on the fetched data from the SQL database/table.

1. How to retrieve records where sal between 1000 to 2000?

Ans. select * from emp where sal>=1000 and sal<2000;

2. Select all records where dept no of both emp and dept table matches.

Ans. select * from emp where exists(select * from dept where emp.deptno=dept.deptno);

3. If there are two tables emp1 and emp2, and both have common records, how can we fetch all the records but common records only once?

Ans. (select * from emp) union (select * from emp1)

4. How to fetch only common records from two tables emp and emp1?

Ans. (select * from emp) intersect (select * from emp1)

5. How can we retrieve all records of emp1 that are not present in emp2?

Ans. (select * from emp) minus (select * from emp1)

6. Count the totalsalary deptno wise where more than 2 employees exist.

Ans.
```
select deptno, sum(sal) as totalsal
from emp
group by deptno
having count(empno) > 2;
```

7. Suppose there is annual salary information provided by emp table. How can we fetch monthly salary of each and every employee?

Ans.
```
select ename,sal/12 as monthlysal from emp;
```

8. Select all records from emp table where deptno=10 or 40.

Ans.
```
select * from emp where deptno=30 or deptno=10;
```

9. Select all records from emp table where deptno=30 and sal>1500.

Ans.
```
select * from emp where deptno=30 and sal>1500;
```

10. Count MGR and their salary in emp table.

Ans.
```
select count(MGR), count(sal) from emp;
```

11. To display all the employees grouped together on the basis of deptno and sal in descending order.

Ans.
```
select ename, deptno, sal from emp order by deptno, sal desc;
```