

UNIT-I COMPUTATIONAL THINKING AND PROGRAMMING

Chapter-1 : Data Structure

Revision Notes

Stacks

- A stack is a linear data structure implemented in LIFO (Last In First Out) manner.
- Insertions and deletions both operations occur only at only one end known as TOP.
- Removing data from stack is called POP operations.
- Adding data into stack is called PUSH operations.
- It is a dynamic data structure *i.e.* stack can grow and shrink.
- Peep refers to inspecting the value at stack's top without removing it.
- Overflow refers to condition when one tries to PUSH an item in stack which is full already.
- Underflow refers to condition when one tries to POP an item from an empty stack.
- Postfix notation refers when operator is placed after its operands *e.g.* $ab +$.
- Prefix notation refers when operator is placed before its operands *e.g.* $+ ab$.

Know the Terms

- **Array** : It is defined as a set of contiguous storage or collection of similar data types.
- **Push operations**: The process of putting a new data element onto stack is known as push operation.
- **Pop operation**: Accessing the content while removing it from the stack is known as pop operation.
- **Underflow**: An error condition that occurs when an item is called for removing from the stack, but the stack is empty.
- **Overflow**: It happens when we try to push one more item onto our stack than it can actually hold.
- **Nested lists**: When one or more elements of list is another list, it is called a nested list.
- **Peep**: It refers to inspecting the value at stack's top without removing it.



UNIT-II COMPUTER NETWORKS

Chapter-2 : Data Communication & Network Topologies

Revision Notes

Communication Technologies

➤ Computer Network

A computer network is a group of two or more interconnected computer systems. You can establish a network connection using either cable or wireless media.

Every network involves hardware and software that connects computers and tools.

➤ **Evolution of Networking**

- (A) **ARPANET**: In 1969, US govt. formed an agency named ARPANET (Advanced Research Projects Agency Network) to connect computers at various universities and defense agencies. The main objective of ARPANET was to develop a network that could continue to function efficiently even in the event of a nuclear attack.
- (B) **NSFNET**: The term “NSFNET” refers to a program of coordinated, evolving projects sponsored by the National Science Foundation that was initiated in 1985 to support and promote advanced networking among U.S. research and education institutions.
- (C) **INTERNET (INTER-connection NETwork)** : The Internet is a worldwide network of computer networks.
- (D) **Interspace**: It is a software that allows multiple users in a client server environment to communicate with each other to send and receive data of various types such as data files, videophone, audio and textual data. Interspace gives the most exceptional type from of communication available on the Internet today.

➤ **Switching Techniques**: These are used for transmitting data across networks.

- (A) **Circuit switching**: In the circuit switching technique, first the complete end to end transmission path between the source and the destination computers is established and then the message is transmitted through the path. The main advantage of this technique is guaranteed delivery of the message which is mainly used for voice communication.
- (B) **Message Switching** : In the message switching technique, no physical path is established between sender and receiver in advance. This technique follows the store and forward mechanism.
- (C) **Packet Switching** : In this technique this message is broken into small data packets. In this switching technique, fixed size of packet can be transmitted across the network.

➤ **Comparison between the various switching techniques:**

Criteria	Circuit switching	Message Switching	Packet Switching
Path established in advance	Yes	No	No
Store and forward technique	No	Yes	Yes
Message follows multiple routes	No	Yes	Yes

➤ **Concept of Communication**

Data can be any text, image, audio, video, and multimedia files. Communication is an act of sending or receiving data.

Thus, data communication refers to the exchange of data between two or more networks or connected devices. These devices must be capable of sending and receiving data over a communication medium.

Examples of such devices include personal computers, mobile phones, laptops, etc.

➤ **Components of Data Communication**

Sender, receiver, communication medium, the message to be communicated, and certain rules called protocols to be followed during communication. The communication media is also called transmission media.

The role of these five components in data communication is as follows:

- (A) **Sender**: A sender is a computer or any such device which is capable of sending data over a network.
- (B) **Receiver**: A receiver is a computer or any such device which is capable of receiving data from the network. In computer communication, the sender and receiver are known as nodes in a network.
- (C) **Message**: It is the data or information that needs to be exchanged between the sender and the receiver.
- (D) **Communication media**: It is the path through which the message travels between source and destination. It is also called medium or link which is either wired or wireless
- (E) **Protocols**: It is a set of rules that need to be followed by the communicating parties in order to have successful and reliable data communication.

➤ **Data communication Terminologies**

- (A) **Concept of Channel** : A data channel is the medium used to carry information or data from one point to another.
- (B) **Band**: It is the unit to measure the data transmission speed. It is equivalent to bps (bits per second).
- (C) **Band-width**: The maximum volume of data that can be transferred over any communication channel at a given point of time is known as the bandwidth. In analog systems, it is measured in hertz (Hz) and in digital systems, it is measured in bits per second (Bps).
- (D) **Data Transfer Rate**: The amount of data transferred per second by the communication channel from one point to another is known as the data transfer rate. It is measured in bits per second (bps), bytes per second (Bps).

$$1 \text{ kbps} = 2^{10} \text{ Bps}$$

$$1 \text{ MBps} = 2^{20} \text{ Bps}$$

$$1 \text{ GBps} = 2^{30} \text{ Bps}$$

$$1 \text{ TBps} = 2^{40} \text{ Bps}$$

➤ **Transmission Media**

Transmission media of a network refers to the connecting media used in the network. It can be broadly defined as anything that can carry information from a source to destination.

- (A) **Twisted Pair Cable** : It consists of two identical 1mm thick copper wires insulated and twisted together. The twisted pair cables are twisted in order to reduce crosstalk and electromagnetic induction.

Advantages:

- It is easy to install and maintain.
- It is inexpensive.

Disadvantages:

- It is incapable to carry a signal over long distance without the use of repeaters.
- Due to low bandwidth, these are unsuitable for broadband applications.

- (B) **Coaxial Cable** : It consists of a solid wire core surrounded by one or more coil or braided wire shields, each separated from the other by some kind of plastic insulator. It is mostly used in the cable wires.

Advantages:

- Data transmission rate is better than twisted pair cables.
- It provides a cheap means of transporting multi-channel television signals around metropolitan areas.

Disadvantages:

- Expensive than twisted pair cables.
- Difficult to manage and reconfigure.

- (C) **Optical fiber** : It consists of thin glass fibres that can carry information in the form of visible light.

Advantages:

- Transmit data over long distance with high security.
- Data transmission speed is high.
- Provide better noise immunity.
- Bandwidth is up to 10 Gbps.

Disadvantages:

- Expensive as compared to other guided media.
- Need special care while installation.

- (D) **Infrared**: The infrared light transmits data through the air and can propagate within a room, but will not penetrate walls. It is a secure medium of signal transmission. The infrared transmission has become common in TV remote, automotive lift doors, wireless speakers etc.

Advantages:

- Power consumption is used.
- Circuitry cost is less.
- Secure mode of transmission.

Disadvantages:

- Limited to a short range.
- Can be blocked by common materials like walls, people, plants etc.

(E) **Radio wave** : It is an electromagnetic wave with a wavelength between 0.5 cm to 30,000m. The transmission making use of radio frequencies is termed as radio-wave transmission.

Advantages:

- Radio wave transmission offers mobility.
- It is cheaper than laying cables and fibers.
- It offers ease of communication over difficult terrain.

Disadvantages:

- Radio wave communication is insecure communication.
- Radio wave propagation is susceptible to weather effects like rains, thunder storms etc.

(F) **Microwave** : The microwave transmission is a line of sight transmission. Microwave signals travel at a higher frequency than radio waves and are popularly used for transmitting data over long distances.

Advantages:

- It is cheaper than laying cable or fiber.
- It has the ability to communicate over oceans.

Disadvantages:

- Microwave communication is an insecure communication.
- Signals from antenna may split up and get transmitted in different way to different antenna which leads to reduce in signal strength.
- Microwave propagation is susceptible to weather effects like rains, thunder storms etc.
- Bandwidth allocation is extremely limited in case of microwaves.

(G) **Satellite link**: The satellite transmission is also a kind of line of sight transmission that is used to transmit signals throughout the world.

Advantages:

- Area covered is quite large.
- No line of sight restrictions such as natural mountains, tall buildings, towers etc.
- Earth station which receives the signals can be fixed in position or relatively mobile.

Disadvantages:

- Very expensive as compared to other transmission mediums.
- Installation is extremely complex.
- Signals sent to the stations can be tampered by external interference.

➤ Network Devices

These are units that mediate data in a computer network and are also called network equipments. Some of them are following:

- **Modem**: A MODEM (Modulator Demodulator) is an electronic device that enables a computer to transmit data over telephone lines. There are two types of modems namely, internal modem and external modem.
- **RJ45 Connector**: The RJ-45 (Registered Jack) connectors are the plug in devices used in the networking and telecommunication applications. They are used primarily for connecting LANs, particularly Ethernet.
- **Ethernet card**: It is a hardware device that helps in connection of nodes within a network.
- **Hub**: It is a hardware device used to connect several computers together. Hubs can be either active or passive. Hubs usually can support 8, 12 or 24 RJ45 ports.
- **Switch** : A switch (switching hub) is a network device which is used to interconnect computers or devices in a network. It filters and forwards data packets across a network. The main difference between a hub and a switch is that hub replicates what is received on one port onto all the other ports while switch keeps a record of the MAC addresses of the devices attached to it.
- **Router** : It is a hardware device which is designed to take incoming packets, analyse packets, moving and converting packets to the another network interface, dropping the packets, directing packets to the appropriate locations etc.

- **Gateway:** It is a device that connects dissimilar networks.
- **Repeater :** It is a network device that amplifies and restores signals for long distance transmission.
- **Wi-Fi card :** Connects to your laptop either in your USB port or wider card slot. This card generally is geared to a particular Wi-Fi network, so to use it you must be in range of a wireless internet signal dedicated to that network.

➤ Network Topologies

Topology refers to the way in which the workstations attached to the network are interconnected.

Some common network topologies are as follows:

Bus Topology: It uses a common single cable to connect all the workstations . Each computer performs its task of sending messages without the help of the central server. However, only one workstation can transmit a message at a particular time in the bus topology.

Advantages:

- Easy to connect and install.
- Involves a low cost and installation time.
- can be easily extended.

Disadvantages:

- The entire network shuts down if there is a failure in the central cable.
- Only a single message can travel at a particular time.
- Difficult to troubleshoot an error.

Star Topology: It is based on a central node which acts as a hub. A star topology is common in home networks where all the computers connect to the single central computer, using a hub.

Advantages:

- Easy to troubleshoot.
- A single node failure does not affect the entire network.
- Fault detection and removal of faulty parts is easier.
- In case a workstation fails, the network is not affected.

Disadvantages:

- Difficult to expand.
- Longer cable is required.
- The cost of the hub and the longer cables make it expensive over others.
- In case hub fails, the entire network fails.

Tree Topology : It combines the characteristics of the linear bus and star topologies. It consists of groups of star configured workstation connected to a bus backbone cable.

Advantages:

- Eliminates network congestion.
- The network can be easily extended.
- Faulty nodes can be easily isolated from the rest of the network.

Disadvantages:

- Uses large cable length.
- Requires a large amount of hardware components and hence is expensive.
- Installation and reconfiguration is very difficult.

➤ Network Types

- **LAN(Local Area Network) :** It is a network that is confined to a relatively small area. It is generally limited to a geographic area such as writing lab, school or building. It is generally privately owned network over a distance not more than 5 km.
- **MAN (Metropolitan Area Network):** It is a network that covers a group of nearby corporate offices in a city and might be either private or public.
- **WAN (Wide Area Network):** These are the networks spread over large distances, say across countries or even continents through cabling or satellite uplinks.

- **PAN(Personal Area Network):** It is a computer network organized around an individual person. It generally covers a range of less than 10 meters. PAN can be constructed with cables or wirelessly.

➤ **Network Protocol**

A protocol means the rules that are applicable for a network. It defines the standardized format for data packets, techniques for detecting and correcting errors and so on. A protocol is a formal description of message formats and the rules that two or more machines must follow to exchange those messages.

Types of protocols are:

- **Hypertext Transfer Protocol (HTTP) :** It is a communication protocol for the transfer of information on the Internet and world wide web. HTTP is a request/response standard between a client and a server. A client is the end-user while, the server is the website.
- **FTP (File Transfer Protocol):** It is the simplest and most secure way to exchange files over the Internet. The objectives of FTP are:
 1. To promote sharing of Files (computer programs and/or data).
 2. To encourage indirect or implicit use of remote computers.
 3. To shield a user from variations in file storage systems among different hosts.
 4. To transfer data reliably and efficiently.
- **TCP/IP (Transmission Control Protocol/ Internet Protocol):** TCP is responsible for verifying the correct delivery of data from client to server. Data can't be lost in the intermediate network. TCP adds support to detect errors or lost data and to trigger retransmission until the data is correctly and completely received.
IP is responsible for moving packet of data from node to node. IP forwards each packet based on a four byte destination address (the IP number). The Internet authorities assign ranges of numbers to different organisations. The organisations assign groups of their numbers to departments. IP operates on gateway machines that moves data from department to organization, then to region and then around the world.
- **SMTP(Simple Mail Transfer Protocol):** It is a standard protocol for email services on TCP/IP network that provides ability to send and receive e-mail.
- **POP3 (Post Office Protocol Version3):** It is a message access protocol which allows client to fetch an e-mail from remote mail server.
- **Telnet** is a network protocol used to virtually access a computer and to provide a two way, collaborative and text based communication channel between two machines.
- **PPP (Point - to - Point Protocol):** It is a communication protocol of the data link layer that is used to transmit multiprotocol data between two directly connected (point-to-point) computers. It is a byte - oriented protocol that is widely used in broadband communications having heavy loads and high speeds.
- **HTTPS (Hypertext transfer protocol secure):** It is the secure version of HTTP, which is the primary protocol used to send data between a web browser and a website. HTTPS is encrypted in order to increase security of data transfer.
- **VoIP (Voice over Internet Protocol):** It is a proven technology that lets anyone place phone calls over an internet connection. With the rise of broadband, VoIP has become the definitive choice of phone service for consumers and businesses alike.

➤ **Wireless Technologies**

- **GSM (Global System for Mobile Communication):** It is leading digital cellular system. In covered areas, cell phone users can buy one phone that will work anywhere, where the standard is supported. It uses narrow band TDMA, which allows eight simultaneous calls on same radio Frequency.
- **CDMA (Code Division Multiple Access):** It is a digital cellular technology that uses spread spectrum techniques. CDMA does not assign a specific frequency to each user. Instead, every channel uses the full available spectrum.
- **WLL(Wireless in Local Loop):** WLL is a system that connects subscribers to the public switched telephone network using radio signals as a substitute for other connecting media.
- **GPRS (General Packet Radio System):** It is a third-generation step towards internet access. GPRS is also known as GSM-IP that is a Global-System Mobile Communications Internet Protocol as it keeps the users of this system online, allows to make voice calls, and access internet on-the-go. Even Time-Division Multiple Access (TDMA) users benefit from this system as it provides packet radio access.
- **Wi-Fi** stands for wireless fidelity that uses radio waves to provide wireless high speed internet access.
- **Wimax** stands for world wide Interoperability for Microwave access, designed for creating metropolitan area network. It is similar to the wifi standard, but supports a greater range of coverage.

Mobile Telecommunication and Web Services

➤ Web Services

- **WWW** : The world wide web or W3 or simply the web is a collection of linked documents or pages, stored on millions of computers and distributed across the Internet.
- **HTML (HyperText Markup Language)** : It is a computer language that describes the structure and behavior of a web page. This language is used to create web pages.
- **XML (Extensible Markup Language)**: It is a meta language that helps to describe the markup language.
- **Domain Name** : It is a unique name that identifies a particular website and represents the name of the server where the web pages reside.
- **URL(Uniform Resource Locator)** : It is a means to locate resources such as web pages on the Internet. URL is also a method to address the web pages on the Internet. There are two types of URL namely, Absolute URL and Relative URL.
- **Website** : A collection of related web pages stored on a web server is known as a website.
- **Web browser**: A software application that enables to browse, search and collect information from the web is known as web browser.
- **Web server** : The web pages on the Internet are stored on the computers that are connected to the Internet. These computers are known as web servers.
- **Web hosting**: Web hosting or website hosting is the service to host, store and maintain websites on the world wide web.
- **Wireless/ Mobile Technologies** : The G in wireless networks refer to the generation of the underlying wireless network technology.
- **1G(First Generation)** : 1G networks (NMT, C-Nets, AMPS, TACS) are considered to be first analog cellular systems, which started in early 1980s. There were radio telephone systems even before that. 1G networks were conceived and designed purely for voice calls with almost no consideration of data services (with the possible exception of built in modems in some headsets).
- **2G(Second Generation)** : 2G networks (GSM, CDMA one, D-AMPS) are the first digital cellular systems launched in early 1990s, offering improved sound quality, better security and higher total capacity. GSM supports circuit switched data (CSD), allowing users to place dial-up data calls digitally, so that the networks switching station receives actual ones and zeroes rather than the screen of an analog modem.
- **3G (Third Generation)** : 3G networks (UMTS FDD and TDD, CDMA 2000 3x, TD-SCDMA, Arib WCDMS, EDGE, IMT-2000 DECT) are newer cellular networks that have data rates of 334 Kbits and more. The UN's International Telecommunications Union IMT-2000 standard requires stationary speeds of 2Mbps and mobile speeds of 334 Kbps for a true 3G.
- **4G (Fourth Generation)** : 4G technology refers to the fourth generation of mobile phone communication standards. LTE and WiMAX are marketed as parts of this generation, even though they fall short of the actual standard.



Know the Terms

- **Backbone** : Referring to the Internet, a central network that provides a pathway for other networks to communicate.
- **Bridge** : A data link layer device that limits traffic between two network segments by filtering the data between them based on hardware addresses.
- **Broadband** : A transmission system capable of carrying many channels of communication simultaneously by modulating them on one of several carrier frequencies.
- **Congestion** : Traffic in excess of network capacity.
- **Data Link** : The physical connection between two devices such as Ethernet, LocalTalk or Token Ring that is capable of carrying information in the service or networking protocols such as AppleTalk, TCP/IP or XNS.
- **Destination Address**: Address of network device that is receiving data.
- **Gbps** : Gigabits per second

- **GHz** : Giga Hertz
- **KBPS** : A unit of measure used to describe the rate of data transmission equal to 1000 bits per second.
- **MBPS** : A unit of measure used to describe the rate of data transmission equal to one million bits per second.
- **Packet** : A discrete chunk of communication in a per defined format.
- **Session** : An on-going relationship between two computing devices involving the allocation of resources and sustained data flow.
- **Token Ring** : A ring topology network that uses token passing for MAC.
- **Transport Protocol** : The protocol layer of the OSI-7 layer model that is concerned with management of data flow between source and destination.
- **Adware** : A general term used for software that invades your computer in the form of persistent pop-up ads.
- **Anti-virus** : A security program that can run on a computer or mobile device and protects you by identifying and stopping the spread of malware on your system.
- **ASP(Active Server Pages)** : It is a technology developed by Microsoft for the purpose of creating dynamic pages that have a .asp extension. such pages utilize VBScript of Jscript code.
- **E-mail (Electronic mail)** : It refers to electronic messages that are sent over communications networks on the Internet. E-mail is sent and retrieved with various protocols including SMTP, POP3 and IMAP.
- **Encryption** : The process by which data is converted into a secret code from its original form via some algorithm. The data in its original form can then only be read by those who can reverse the encryption. Encryption prevents unauthorized reading of data.
- **Hacker** : A general term used for anyone who spends time poking into computers and operating systems, trying to discover their vulnerabilities.
- **Host** : A computer that is accessed by a user from a remote location, also any computer on a network.
- **IP Address** : It is a unique identifier for every computer or device on a network.
- **Java Script** : An object oriented high level language used for web development. JavaScript was originally created by Netscape. It can be used to interact with the user as well as create dynamic content such as moving images and clocks.
- **Phishing** : It is a social engineering technique where cyber attackers attempt to fool you into taking an action in response to an Email. Phishing was a term originally used to describe a specific attack scenario.
- **Scripting Language** : It is a language used to write scripts that are usually executed in a web browser or a web server. Scripting languages are often used to add functionality to web pages such as interaction with the user, animations and different menus. Code written in scripting language is not compiled, rather interpreted. Some scripting languages include JavaScript, VBScript, PHP and Python.
- **Spyware** : A type of malware that is designed to spy on the victim's activities, capturing sensitive data such as the person's passwords, online shopping and screen, contents.



UNIT III: Database Management

Chapter-3 : Database Concepts



Revision Notes

Communication Technologies

➤ Database Concepts and Needs:

- A database is tabular organization of data. Each table comprises of rows (records) and columns (attributes). Each record contains values for the corresponding attributes. The values of the attributes for a record are interrelated. For example, different cases have different values for the same specifications (length, color, engine capacity, etc).

- The database oriented approach supports multiple views of the same data. For example, a clerk may only be able to see his details, whereas the manager can view the details of all the clerks working under him.
 - In database oriented approach, we store the common data in one table and access it from the required tables. Thus, the redundancy of data decreases.
 - Multiple views of the same database may exist for different users. This is defined in the view level of abstraction.
 - The logical level of abstraction defines the type of data that is stored in the database and the relationship between them.
 - The design of the database is known as the database schema.
 - The instance of the database is the data contained by it at that particular moment.
 - The database administrator has the total control of the database and is responsible for setting up and maintaining the database.
 - A data model is the methodology used by a particular DBMS to organize and access the data.
 - Hierarchical, Network and Relational model are the three popular data models. However, the relational model is more widely used.
- **Hierarchical Model**
- The hierarchical model was developed by IBM in 1968.
 - The data is organized in a tree structure where the nodes represent the records and the branches of the tree represent the fields.
 - Since the data is organized in a tree structure, the parent node has the links to its child nodes. This is called (PCR) parent child relation.
 - If we want to search a record, we have to traverse the tree from the root through all its parent node to reach the specific record. Thus, searching for a record is very time consuming.
 - The hashing function is used to locate the root.
 - SYSTEM 2000 is an example of hierarchical database.
- **Network Model**
- Data is represented by collection of records and relationships among data are represented by links.
 - Recording of relationship in the network model is implemented by using pointers.
 - Recording of relationship implementation is very complex since pointers are used. It supports many-to-many relationship and simplified searching of record, since a record has many access paths.
 - DBTG CODASYL was the first network database.
- **Relational Model**
- The Relational model, organizes data in the form of independent tables (consisting of rows and columns) that are related to each other.
 - A table consists of a number of rows (records/tuples) and columns (attributes). Each record contains values for the attributes.
 - The **degree** of the table denotes the number of columns.
 - **Cardinality** is defined as the number of rows in a table.
 - A domain in the relational model is said to be atomic, if it consists of indivisible units. For example, name is not atomic since it can be divided into first name and last name.
 - E.F. Codd laid down 12 rules (known as Codd's 12 rules), that outline the minimum functionality of a RDBMS. A RDBMS must comply with at least 6 rules.
 - A super key is a type of attribute that collectively identifies an entity in an entire set. For example, the bank account number is a super key in the bank accounts table.
 - A candidate key (also known as primary key) is the smallest subset of the super key for which there does not exist a proper subset that is a super key.
 - Out of the multiple candidate keys, only one is selected to be the primary key and the remaining are alternate keys.
 - A foreign key is the primary key of a table that is placed into a related table to represent one-to-many relationship among these tables.

- A primary key is a set of one or more attributes that can uniquely identify the relation.
- Primary key uniquely identifies the records in the table.



Know the Terms

- **DBMS:** It stands for Database Management System that enables users to define, create and maintain the database and provides controlled access to this database.
- **DBA:** DBA is Database Administrator that has the central control over the system.
- **RDBMS:** RDBMS stands for Relational Database Management System.
- **Meta Data:** It means data about data i.e., a logical description of the structure of a data.
- **File Processing System:** It is a collection of new data files stored in the hard drive of a system.



Chapter-4 : Structured Query Language (SQL)



Revision Notes

Communication Technologies

1. Structured Query Language (SQL)

- When a user wants to get some information from a database file, he can issue a query.
- A query is a user-request to retrieve data or information with a certain condition.
- SQL is a query language that allows user to specify the conditions, (instead of algorithms).

2. Types of SQL commands

- **Data Definition Language (DDL) Commands:** All the commands used to create, modify or delete physical structure of an object like table. *e.g.*, Create, Alter, Drop.
- **Data Manipulation Language (DML) Commands:** All the commands used to modify contents of a table comes under this category. *e.g.*, Insert, Delete, Update
- **Transaction Control Language (TCL) Commands:** These commands are used to control transaction of DML commands. *e.g.*, Commit, Rollback.

3. Basic structure of an SQL query

- **General structure** SELECT, ALL/DISTINCT, *, AS, FROM, WHERE
- **Comparison** IN, BETWEEN, LIKE "% _"
- **Grouping** GROUP BY, HAVING, COUNT(), SUM(), AVG(), MAX(), MIN()
- **Display order** ORDER BY, ASC/DESC
- **Logical operators** AND, OR, NOT
- **Create Database** To create a database, use the CREATE DATABASE statement.
Syntax CREATE DATABASE database_name;
- **Show Database** To display the name of all databases, use SHOW command.
Syntax SHOW DATABASES;
- **USE database** To use the database follow the syntax
USE Database_name;
- **Drop database**
To drop database, DROP DATABASE statement is used.
Syntax DROP DATABASE database_name;
- **Show tables** To show all tables in current database, SHOW TABLES command is used.
Syntax SHOW TABLES;

4. Constraint

- Constraint is a condition applicable on a field or group of fields.
- Two types of constraint:
 - **Column constraint:** Apply only to individual column.
 - **Table constraint:** Apply to group of columns.
 - **Different constraints:**
 - Unique Constraint-Primary Key constraint.
 - Default Constraint-Check constraint.

Applying Constraint

Example: Create a student table with student id, student name, father's name, age, class, address.

```
CREATE TABLE student
sid char (4) PRIMARY KEY,
sname char (20) NOT NULL,
fname char (20),
age number (2) CHECK (age<20),
class char (5) NOT NULL,
address char (50));
```

5. SELECT Command

The SELECT command is a query that is given to produce certain specified information from the database table.

Syntax:

```
SELECT <column-name>,[<column-name>.....]
FROM <table-name>;
```

Example: Write a query to display the name and salary of the employee in emp table.

```
SELECT ename, sal
FROM emp;
```

Variations of SELECT command:**(i) Selecting specific Rows.....WHERE clause****Syntax:**

```
SELECT <column-name> [<column-name>.....]
FROM <table-name>
WHERE <condition>;
```

Example: Display the codes, names and salary of employees who belong to 'Manager' category.

```
SELECT ecode, ename, sal
FROM emp
WHERE job="MANAGER";
```

(ii) Searching for NULL (IS NULL command)

The Null value in a column can be searched for in a table using IS NULL in the WHERE Clause

Syntax:

```
SELECT..... <column-name>, <column-name>.....
FROM <table-name>
WHERE <column-name> IS NULL;
```

Example: Display the codes, names and jobs of employees whose DeptNo is Null.

```
SELECT ecode,ename, job
FROM emp
WHERE DeptNo IS NULL;
```

(iii) IS NOT NULL Command

Example: Display the names and jobs of those employees whose DeptNo is not Null.

```
SELECT ename, job FROM emp
WHERE DeptNo IS NOT NULL;
```

(iv) Sorting Result-ORDER BY Clause

The resulting column can be sorted in ascending and descending order using the ORDER BY clause.

Syntax:

```
SELECT <column-name>, <column-name>.....
FROM <table-name>
WHERE <condition>
ORDER BY <column-name>ASC/DESC;
```

Example:

Display the list of employees in descending order of employee code, who is manager.

```
SELECT * FROM emp
WHERE job="MANAGER"
ORDER BY ecode DESC;
```

(v) Conditions based on a range

SQL provides a BETWEEN operator that defines a range of values that the column value must fall for the condition to become true.

Example:

```
SELECT Roll_no, name From
student WHERE Roll_no BETWEEN 100 AND 103;
```

The given command displays Roll_no and name of those students whose Roll_no lies in the range 100 to 103 (both 100 and 103 are included in the range).

(vi) Conditions based on a list

To specify a list of values, IN operator is used. This operator selects values that match any value in the given list.

Example:

```
SELECT * FROM student WHERE city IN ('Delhi', 'Agra', 'Gwalior');
```

The above command displays all those records whose city is either Delhi or Agra or Gwalior.

(vii) Conditions based on Pattern

SQL provides two wild card characters that are used while comparing the strings with LIKE operator.

- (a) percent (%) matches any string.
- (b) Underscore(_) matches any one character.

Example:

```
SELECT Roll_no, name, city FROM student WHERE Roll_no LIKE "%3";
```

This query displays those records where last digit of Roll_no is 3 and may have any number of characters in front.

Example:

```
SELECT Roll_no, name, city FROM student WHERE Roll_no LIKE "1_3";
```

This query displays those records whose Roll_no starts with 1 and second letter may be any letter but ends with digit 3.

6. The INSERT command

The tuples are added to relation using INSERT command of SQL.

Syntax:

```
INSERT INTO <table-name>[<column list>]
VALUES (<value1>,<value2>,<value3>.....);
```

Example: Enter a new record in student table.

```
INSERT INTO student (sid,sname,fname,age, class,address)
VALUES("101","Mohan","Pawan",15,"8","Jaipur");
```

Output:

sid	sname	fname	age	class	address
101	Mohan	Pawan	15	8	Jaipur

7. The DELETE command

The delete command removes the tuples from the tables. This command removes the entire row from the table and not the individual field. So, no field argument is needed.

Syntax:

```
DELETE FROM <table-name>
```

```
WHERE <condition>;
```

Example: Delete all the records of employee whose salary is less than 3000.

```
DELETE FROM emp
```

```
WHERE sal<3000;
```

To delete all the record from the table.

Syntax:

```
DELETE FROM<table-name>;
```

8. The UPDATE command

The UPDATE command is used to change some values in existing rows. The UPDATE command specifies the rows to be changed using the WHERE clause, and new data using the SET keyword.

Example: Update the salary of employee to 5000 whose employee code is 1011.

```
UPDATE emp
```

```
SET sal=5000
```

```
WHERE empno=1011;
```

9. The ALTER TABLE command

The ALTER command is used to change the definition of existing table.

(a) To add columns to a table

Syntax

```
ALTER TABLE<table-name> ADD <column name> <data type> <size>;
```

(b) To modify existing columns of a table

Syntax:

```
ALTER TABLE <table-name>
```

```
MODIFY (Column-name newdatatype (newszie));
```

Example: To modify column job of table emp to have new width of 30 characters.

```
ALTER TABLE emp
```

```
MODIFY (job char (30));
```

10. The DROP Command:

The DROP command is used to drop the table from the database. For dropping a table, all the tuples should be deleted first *i.e.*, the table should be empty.

Syntax:

```
DROP TABLE <table-name>;
```

Example: Drop the student table from the database.

```
DROP TABLE student;
```

Some Example:

Ex 1. Write a query on the customers table whose output will exclude all customers with a rating ≤ 100 , unless they are located in Shimla.

```
SELECT * FROM customers WHERE rating >100 OR city ="Shimla";
```

Ex 2. Write a query that selects all orders except those zeros or NULLs in the amount field.

```
SELECT * FROM Orders WHERE amt <>0 AND (amount IS NOT NULL);
```

Ex 3. Write a query that lists customers in descending order of rating. Output the rating field first, followed by the customers name and number.

```
SELECT rating, cust_name, cust_num FROM customers ORDER BY rating DESC;
```

Ex 4. Write a command that puts the following values in their given order, into the salesman table:

cust_name-Manisha, city-Manali, comm.-NULL, cust_num-1901.

INSERT INTO salesman (city, cust_name, comm, cust_num) VALUES ("Manali", "Manisha", NULL, 1901);

Operators in SQL:

The following are the commonly used operators in SQL:

- (i) Arithmetic Operators +, -, *, /, %
- (ii) Relational Operators =, <, >, <=, >=, <>
- (iii) Logical Operators OR, AND, NOT

- Arithmetic operators are used to perform simple arithmetic operations.
- Relational operators are used when two values are to be compared and logical operators are used to connect search conditions in the WHERE clause in SQL.

Other Operators:

- (iv) Range check – between low and high
- (v) List check – in
- (vi) Pattern check – like, not like (% and _ (under score) are used).

11. SQL Functions:

SQL supports functions which can be used to compute and select numeric, character and date columns of a relation. These functions can be applied on a group of rows. The rows are grouped on a common value of a column in the table. These functions return only one value for a group and therefore, they are called aggregate or group functions.

- (i). SUM (): It returns the sum of values of a column of numeric type.
e.g., Select sum (salary) from employee;
- (ii). AVG (): It returns the average of values of a column of numeric type.
e.g., Select avg (salary) from employee;
- (iii). MIN (): It returns the minimum value of the values of a column or a given relation.
e.g., Select min (salary) from employee;
- (iv). MAX (): It returns the maximum value of the values of a column or a given relation.
e.g., Select max (salary) from employee;
- (v). COUNT (): It returns the number of rows in a relation.
e.g., Select count () from employee;*

Join

A join is a query that combines rows from two or more tables. In a join query, more than one tables are listed in FROM clause.

The function of combining data from multiple tables is called joining.

Joins are used when we have to select data from multiple tables is called joining. Joins are used to extract data from two tables, when we need a relationship between certain columns in these tables.

There are different kind of SQL joins:

1. Equi-Join

Equi join is a simple SQL join condition that uses equal sign as a comparison operator.

Syntax

```
SELECT column1, column2, column3
FROM Table1, Table2
WHERE Table1. column1 = Table2. column1;
```

2. Natural Join

The natural join is a type of equi join and it structured in such a way that, columns with same name of associated tables will appear once only.

Syntax

```
SELECT* FROM Table1
NATURAL JOIN Table2;
```

Know the Terms

- **Attribute:** A set of properties *e.g.*, name, datatype, size, etc., used to characterise the data items of entities. A group of attributes constructs an entity-type (or table), i.e.: all values of a certain column must confirm to the same attributes. Attributes are optionally complemented by constraints.
- **Column:** A set of values of a single table which resides on the same position within its rows.
- **Constraint:** Similar to attributes constraints define rules at a higher level, data items must confirm to. *e.g.*: null, primary and foreign key, uniqueness, default value, user-defined-criteria like STATUS < 10.
- **Database:** A set of tables. Those tables contain user data and the data dictionary.
- **Data Control Language (DCL):** A class of statements which defines the access rights to data, *e.g.*: GRANT, REVOKE.
- **Data Definition Language (DDL):** A class of statements which defines logical and physical design of a database, *e.g.*: CREATE TABLE.
- **Data Manipulation Language (DML):** A class of statements which retrieves and manipulates data, *e.g.*: SELECT, INSERT, UPDATE, DELETE, COMMIT, ROLLBACK.
- **Relational Model:** A database in which inter-table relationships are primarily organized through common data columns which define a one-to-many relationship between a row of the primary key table and one or more rows of the matching foreign key table. Equi-joins relate tables that have matching primary/foreign key values, but other comparisons (relationships) may be defined. Besides describing how the database tables are related, the relational model also defines how the related data can be accessed and manipulated. SQL is the most commonly used relational model database language.
- **Relationship:** A reference between two different or the same entity. References are not implemented as links. They base upon the values of the entities.
- **Row:** One record in a table containing information about one single entity. A row has exactly one value for each of its columns - in accordance with First Normal Form. This value may be NULL.
- **Statement:** A single command which is executed by the DBMS. There are 3 main classes of statements: DML, DDL and DCL.
- **Table (=Relation):** A set of rows of a certain entity-type, i.e. all rows of a certain table have the same structure.
- **Transaction:** A logical unit of work consisting of one or more modifications to the database. The ACID criterium must be achieved. A transaction is either saved by the COMMIT statement or completely cancelled by the ROLLBACK statement.
- **Value:** Implementation of a single data item within a certain column of a certain row.
- **View:** A virtual table containing only its definition and no real data. The definition consists of a query to one or more real tables or views. Queries to the view are processed as queries to the underlying real tables.



Chapter-5 : Interface of Python with SQL Database

Revision Notes

- Python's standard for database interfaces is the Python DB-API
- Python Database API supports a wide range of database servers such as GadFly, mSQL, MySQL, Oracle, Sybase etc.
- You need to download separate DB-API module for each database you need to access.
- DB-API provides a minimal standard for working with databases.
- MySQLdb is an interface for connecting to a MySQL database servers from Python.
- Connect method of MySQLdb interface is used to create a connection object using MySQLdb module.

- A cursor is a Python object that enables you to work with the database. In database terms, the cursor is positioned at a particular location within a table or tables in a database.
- To get a cursor you need to call the cursor method on the database object.
- To save your changes to the database, you must commit the transaction using commit()
- When you are done with the script, close the cursor and then the connection to free up the resources
- The DB-API's include a defined set of exceptions.

Database Object Methods

Method	Description
close()	Closes the connection to the database.
commit()	Commits any pending transaction to the database.
cursor()	Returns a database cursor object through which queries can be executed.
rollback()	Rolls back any pending transaction to the state that existed before the transaction began.

➤ Cursor object Attributes and Methods

Note: c is the cursor object.

Syntax	Description
c.arraysize()	The readable/writable numbers of rows that fetchmany() will return if no size is specified.
c.close()	Close the cursor, c.
c.execute(sql, params)	Executes the SQL query on string sql, replacing each placeholder with the corresponding parameter from the param sequence or mapping
c.executemany(sql, seq-of-params)	Executes the SQL query once for each item in the seq-of-params sequence of sequences or mappings. This method should not be used for operations that create results set (such as SELECT statement)
c.fetchall()	Returns a sequence of all the rows that have not yet been fetched.
c.fetchmany(size)	Returns a sequence of rows (each row itself being a sequence); size defaults to c.arraysize
c.fetchone()	Returns the next row of the query result set as a sequence or None when the results are exhausted. Raises an exception if there is no result set.
c.rowcount()	The read only row count for the last operation (e.g. select, insert, update delete) or -1 if not available or applicable

Know the Terms

- **DB-API:** It is a specification for a common interface to relational databases.
- **Cursor:** It is a Python object that enables you to do the work with the database.
- **Result Set:** It is an object that is returned when a cursor object is used to query a table.
- **Database:** It is a collection of organised information that can easily be used, managed, updated and they are classified according to their organizational approach.
- **MySQLdb:** It is for connecting to a MySQL database server from Python.