MATS UNIVERSITY

NAAC GRADE A+ ACCREDITED UNIVERSITY

MATS CENTRE FOR OPEN & DISTANCE EDUCATION

Computer Application

Bachelor of Science (B.Sc.)
Semester - 4

SELF LEARNING MATERIAL

SEC409

MATS UNIVERSITY

Computer Application
CODE: ODL/MSS/BSCB/409

COURSE DEVELOPMENT EXPERT COMMITTEE

1. Prof. (Dr.) Vishwaprakash Roy, School of Sciences, MATS University, Raipur, Chhattisgarh
2. Dr. Prashant Mundeja, Professor, School of Sciences, MATS University, Raipur, Chhattisgarh
3. Dr. Sandhyarani Panda, Professor, School of Sciences, MATS University, Raipur, Chhattisgarh
4. Mr. Y. C. Rao, Company Secretary, Godavari Group, Raipur, Chhattisgarh

COURSE COORDINATOR

Dr.Prashant Mundeja, Professor, School of Sciences, MATS University, Raipur, Chhattisgarh

COURSE /BLOCK PREPARATION

Dr. Meghna Shrivastava, Associate Professor, School of Sciences, MATS University, Raipur, Chhattisgarh

March, 2025

FIRST EDITION :
ISBN: 978-81-987917-5-7

Printed & Published on behalf of MATS University, Village-Gullu, Aarang, Raipur by Mr. Meghanadhudu Katabathuni, Facilities & Operations, MATS University, Raipur (C.G.)

Printed at: The Digital Press, Krishna Complex, Raipur-492001(Chhattisgarh)

MODULE INTRODUCTION

Course has four modules. Each module is divided into individual units. Under this theme we have covered the following topics:

Module 1: Computer Basics

Module 2: Introduction To the Internet

Module 3: MS OFFICE

Module 4: Database

Module 5: Use Of Computers in Biological Sciences

This book discuss about Computer application course, which aim to equip individuals with the skills and knowledge to use computers effectively and develop software applications. These courses cover topics like programming, databases, web development, and more This book is designed to help you think about the topic of the particular MODULE.

We suggest you do all the activities in the MODULEs, even those which you find relatively easy. This will reinforce your earlier learning.

MODULE 1

COMPUTER BASICS

Objectives:

- To know about Computers
- To learn about various generations of computer
- To understand the basic operations of computers
- To know the components and their functions.
- To know about booting of a computer

UNIT 1.1

Computer Basics History

Computers are seen everywhere around us, in all spheres of life, in the field of education, research, travel and tourism, weather forecasting, social networking, e-commerce etc. Computers have now become an indispensable part of our lives. Computers have revolutionized our lives with their accuracy and speed of performing a job, it is truly remarkable. Today, no organization can function without a computer. In fact, various organizations have become paperless. Computers have evolved over the years from a simple calculating device to high speed portable computers.

The growth of computer industry started with the need for performing fast calculations. The manual method of computing was slow and prone to errors. So, attempts were made to develop fast calculating devices, the journey started from the first known calculating device (Abacus) which has led us today to an extremely high speed calculating devices.

Generations

First Generation of Computers (1940-1956)

The first generation of electronic computers appeared during World War II, motivated bythe needs of the military for fast mathematical calculations, especially in the domains of ballistics and cryptography. They used vacuum tubes for processing and magnetic drums for memory storage. Vacuum tubes were bulky, produced large amounts of heat, consumed lots of electrical power, and failed frequently, so these machines were complicated and temperamental. The ENIAC (Electronic Numerical Integrator and Computer) is often thought to have been the first general-purpose electronic computer was invented at the University of Pennsylvania in 1945. ENIAC would fill a room and weigh 30 tons, yet it could crunch complex mathematical problems in seconds problems that would take people hours or days to solve. It made use of some 17,468 vacuum tubes, 7,200 crystal diodes, 1,500 relays, 70,000 resistance and 10,000 capacitance parts, as well as an estimated 5 million joints soldered by hand. Programming these earlycomputers was incrediblycomplicated.Antiquated machines Engineers had to literally re-wire the machine's circuits to change what it could do, a process that might take days, possibly weeks. All input and output was done using punch cards and paper tape, and programming languages in their most primitive incarnations. These earlycomputers

could onlyunderstand machine language, which is represented in binary code (0s and 1s). Some notable first-generation computers were the UNIVAC I (UniversalAutomatic Computer), the first commercial computer produced in the US. The UNIVAC I was delivered to the United States Census Bureau in 1951, one of the first steps in extending computing technology beyond military and scientific research institutions.

Second Generation Computers (1956-1963)

The second generation of computers was another important period of the evolution of computers during this time; vacuum tubes were replaced with transistors. Transistors, which were invented in 1947 by John Bardeen, Walter Brattain and William Shockleyat Bell Labs, represented a quantum leap in electronic technology. These small semiconductor devices cost much less to manufacture, were much more reliable, generated very little heat, and used only a fraction of the power required by vacuum tubes. This period also saw the emergence of the high-level programming language, which significantlysimplified computer programming. The first widely used high-level programming language, FORTRAN (Formula Translation), developed by IBM in 1957, allowed scientists and engineers to write more complex programs with relative ease. COBOL (Common Business-Oriented Language), designed in 1959, was the standard language for business and administrative computing. More sophisticated than tape storage, it was also much more reliable than magnetic drum memory. Similarly, storage technologies, such as magnetic tape and magnetic disks, also progressed, enabling larger data storage capacities and faster access. Computers such as the IBM 1401 and IBM 7090 grew popular in business and scientific communities, showcasing the increasing flexibility and dependability of computing technology. These second-generation computers were much smaller than their predecessors, filling rooms instead of buildings. They were more energy-efficient, more reliable and they could calculate much more quickly. Transistor technology allowed for more complex processing tasks and set the stage for the eventual shrinking of computer hardware.

The Third-Generation of Computers (1964-1971)

The third generation of computers was characterized by the advent of integrated circuits (ICs), which significantly reduced the size of electronic components. Jack Kilby with Texas Instruments and Robert

Noyce at Fairchild Semiconductor independentlyinvented integrated circuit technology in the late 1950s, integrating multiple electronic components on a silicon chip. Integrated circuits enabled further miniaturization, faster processing, and less power consumption. Computer hardware shrank, gained in power and dropped drastically in price. The IBM System/360, released in 1964, is a classic example of this generation's technological advancement. It was a family of compatible computer models that could be upgraded without fully replacing existing hardware, an revolutionary concept for its time. Data is used to train various components based on the updates released in the operating systems in the coming years, which allow more dynamic management of resources and multi-programming capabilities. One of the paradigms fundamental to the development of modern computing was the implementation of time-sharing, which allowed multiple users to interact with the computer all at once. BASIC (Beginner'sAll-purpose Symbolic Instruction Code) was created at Dartmouth College in 1964, making programming more widely accessible to students and non-specialists, as programming languages grew. Semiconductor memory started taking over for magnetic core memory, ramping up computation fastness and reliability even more. Interactive computing emerged as users interacted directly with their computers using keyboards and monitors rather than punch cards and paper tape input of previous generations.

The fourth generation of computers (1971-Present)

The microprocessor, a single integrated circuit that contains the complete central processing unit (CPU), defines the fourth generation of computers. As the first commercial microprocessor, the Intel 4004 was launched in 1971. This was the breakthrough that made personal computers possible, moving computing from the province of the specialist, the institution, to a personal tool. The launch of the Apple II by Apple in 1977 and IBM's Personal Computer (PC) in 1981 were seminal events in computing history. These machines brought computing technology in many homes and small businesses, democratizing access to digital technology. The Microsoft Disk Operating System (MS-DOS) and then Microsoft Windows — they offered standardized operating systems that gave computers a more user-friendly face. GUIs revolutionized the concept of interacting with your computer. In the 1970s, Xerox Alto was developed and later the Apple Macintosh was the first to have a mouse driven interface that replaced complicated text based command systems. This innovation

allowed computers to be used by people without specialized technical training. This era saw significant developments in networking technologies. The ARPANET, an early variant of the internet, was created in 1969, and TCP/IP communication protocol was finalized in 1983. The world became even more interconnected when Tim Berners-Lee invented the World Wide Web in 1989, using it to enable the sharing of information between computers all over the world, paving the way for the modern age of the Internet. Microprocessor technology continued improving at an exponential rate (Moore's Law). This principle implied that suits of regions in the mainframe world would be translated to microchip as law, and the number of transistors able to squeeze into a microchip would double approximatelyeverytwo years, while cost would decrease byhalf. While progress in that direction has stalled in recent years, this rule led us to some astonishing leaps in computational speed for many decades.

The Fifth Generation and Beyond (Emerging Technologies)

The Fourth Generation is there a while, but what most people do not know is that, we are already in the Fifth Generation of Computing, which is characterized bya high degree of integration and the emergence of artificial intelligence, quantum computing and neural networks. These new technologies have the potential to revolutionize computational capabilities like never before. Artificial Intelligence (AI), as well as machine learning, have achieved incredible breakthroughs, boasting systems capable of intricate pattern recognition, natural language processing, and independent decision-making. Neural networks modeled on the structures of the human brain can now recognize images, play complex games, and perform many other tasks at superhuman levels.

Sixth Generation Computing

In the Sixth Generation, computers could be defined as the era of intelligent computers, based on Artificial Neural Networks. One of the most dramatic changes in the sixth generation will be the explosive growth of Wide Area Networking. Natural Language Processing (NLP) is a component of Artificial Intelligence (AI). It provides the ability to develop the computer program to understand human language.

We all know what a computer is? It is an electronic device that processes the input according to the set of instructions provided to it and gives the desired output at a very fast rate. Computers are very versatile as they do a lot of different tasks such as storing data, weather

forecasting, booking airlines, railway or movie tickets and even playing games.

Data: Data is defined as an un- processed collection of raw facts, suitable for communication, interpretation or processing.

For example, 134, 16 'Kavitha', 'C' are data. This will not give any meaningful message.

Information: Information is a collection of facts from which conclusions may be drawn. In simple words we can say that data is the raw facts that is processed to give meaningful, ordered or structured information. For example Kavitha is 16 years old. This information is about Kavitha and conveys some meaning. This conversion of data into information is called data processing.

"A Computer is an electronic device that takes raw data (unprocessed) as an input from the user and processes it under the control of a set of instructions (called program), produces a result (output), and saves it for future use."

SUMMARY: History of Computers

The **history of computers** spans several centuries, evolving from simple calculating devices to today's advanced digital systems. The development is typically divided into **five generations**, each marked by major technological advances.

- **Pre-Computer Era**
  - **Abacus** (c. 3000 BCE): First mechanical calculator.
  - **Pascaline** by Blaise Pascal (1642): First mechanical adding machine.
  - **Analytical Engine** by Charles Babbage (1830s): The first concept of a programmable computer.
  - **Ada Lovelace**: Recognized as the first computer programmer.

- **Generations of Computers**
  1. **First Generation (1940–1956)**
     - Technology: Vacuum tubes
     - Examples: ENIAC, UNIVAC

- Characteristics: Very large, consumed lots of power, slow, used machine language.

2. **Second Generation (1956–1963)**
   - Technology: Transistors
   - Examples: IBM 1401
   - Characteristics: Smaller, more reliable, used assembly language.
3. **Third Generation (1964–1971)**
   - Technology: Integrated Circuits (ICs)
   - Examples: IBM 360 series
   - Characteristics: Faster, more efficient, used high-level programming languages.
4. **Fourth Generation (1971–present)**
   - Technology: Microprocessors
   - Examples: Personal computers (PCs), Laptops
   - Characteristics: Small, affordable, powerful, GUI-based.
5. **Fifth Generation (Present and Beyond)**
   - Technology: Artificial Intelligence, Quantum Computing
   - Focus: Machine learning, natural language processing, robotics.

✅ 5 Multiple Choice Questions (MCQs)

1. **Who is considered the father of the computer?**
   a) Alan Turing
   b) Bill Gates
   c) Charles Babbage
   d) Steve Jobs
   ✔ **Answer: c) Charles Babbage**
2. **Which generation of computers introduced microprocessors?**
   a) First

b) Second
c) Third
d) Fourth

✔ **Answer: d) Fourth**

3. **The Analytical Engine was designed by:**
   a) Ada Lovelace
   b) Blaise Pascal
   c) Charles Babbage
   d) John von Neumann

   ✔ **Answer: c) Charles Babbage**

4. **The main component of the first generation of computers was:**
   a) Transistor
   b) Vacuum tube
   c) Microprocessor
   d) IC

   ✔ **Answer: b) Vacuum tube**

5. **In which generation did AI become a focus of computer development?**
   a) Third
   b) Fourth
   c) Fifth
   d) Second

   ✔ **Answer: c) Fifth**

Short Answer Type Questions

1. What is the contribution of Charles Babbage to computing?
2. Name the technologies used in the first and second generations of computers.
3. Why is Ada Lovelace significant in computer history?

Long Answer Type Questions

1. Describe the five generations of computers and highlight their key technologies and features.
2. Explain the contributions of early computing pioneers like Charles Babbage, Ada Lovelace, and Alan Turing.
3. Compare the characteristics of the first and fourth generations of computers.

UNIT 1.2

Components of a Computer

The computer is the combination of hardware and software. Hardware is the physical component of a computer like motherboard, memory devices, monitor, keyboard etc., while software is the set of programs or instructions. Both hardware and software together make the computer system to function.

Figure 1.2.1: Computer

Let us first have a look at the functional components of a computer. Every task given to a computer follows an Input- Process- Output Cycle (IPO cycle). It needs certain input, processes that input and produces the desired output. The input unit takes the input, the central processing unit does the processing of data and the output unit produces the output. The memory unit holds the data and instructions during the processing.

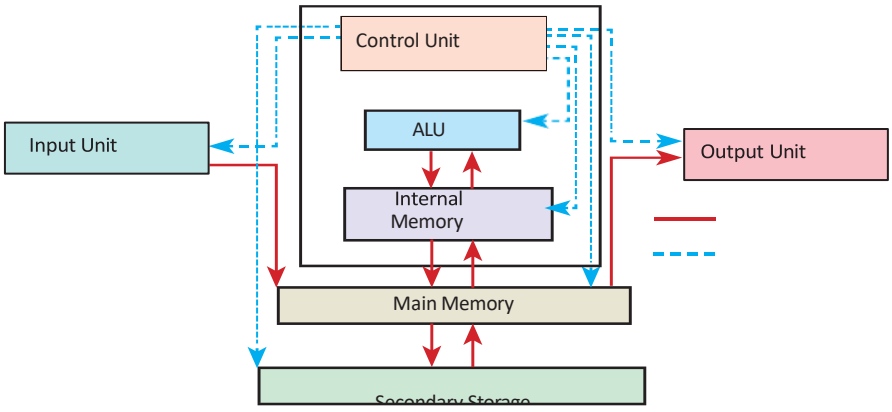


Figure 1.2.2 components of a computer

Input Unit

Input unit is used to feed any form of data to the computer, which can be stored in the memory unit for further processing. Example: Keyboard, mouse, etc.

Central Processing Unit

CPU is the major component which interprets and executes software instructions. It also control the operation of all other components such as memory, input and output units. It accepts binary data as input, process the data according to the instructions and provide the result as output Central Processing Unit

The CPU has three components which are Control unit, Arithmetic and logic unit (ALU) and Memory unit.

Arithmetic and Logic Unit

The ALU is a part of the CPU where various computing functions are performed on data. The ALU performs arithmetic operations such as addition, subtraction, multiplication, division and logical operations. The result of an operation is stored in internal memory of CPU. The logical operations of ALU promote the decision-making ability of a computer.

Control Unit

The control unit controls the flow of data between the CPU, memory and I/O devices. It also controls the entire operation of a computer.

Output Unit

An Output Unit is any hardware component that conveys information to users in an understandable form. Example: Monitor, Printer etc.

Memory Unit

The Memory Unit is of two types which are primary memory and secondary memory. The primary memory is used to temporarily store the programs and data when the instructions are ready to execute. The secondary memory is used to store the data permanently.

The Primary Memory is volatile, that is, the content is lost when the power supply is switched off. The Random Access Memory (RAM) is an example of a main memory. The Secondary memory is non volatile, that is, the content is available even after the power supply is switched

off. Hard disk, CD-ROM and DVD ROM are examples of secondary memory.

Input and Output Devices

Input Devices:

(1) Keyboard: Keyboard (wired / wireless, virtual) is the most common input device used today. The individual keys for letters, numbers and special characters are collectively known as character keys. This keyboard layout is derived from the keyboard of original typewriter. The data and instructions are given as input to the computer by typing on the keyboard. Apart from alphabet and numeric keys, it also has Function keys for performing different functions. There are different set of keys available in the keyboard such as character keys, modifier keys, system and GUI keys, enter and editing keys, function keys, navigation keys, numeric keypad and lock keys.

Figure 1.2.3 Keyboard

(2) **Mouse:** Mouse (wired/wireless) is a pointing device used to control the movement of the cursor on the display screen. It can be used to select icons, menus, command buttons or activate something on a computer. Some mouse actions are move, click, double click, right click, drag and drop. Different types of mouse available are: Mechanical Mouse, Optical, Laser Mouse, Air Mouse, 3D Mouse, Tactile Mouse, Ergonomic Mouse and Gaming Mouse.

MOST COMMONLY USED TYPES OF MOUSE			
SN	**Type of Mouse**	**Mechanism**	**Developed and Introduced**
1	**Mechanical Mouse** 	• A small ball is kept inside and touches the pad through a hole at the bottom of the mouse. • When the mouse is moved, the ball rolls. • This movement of the ball is converted into signals and sent to the computer.	Telefunken, German Company, 02/10/1968
2	**Optical Mouse** 	• Measures the motion and acceleration of pointer. • It uses light source instead of ball to judge the motion of the pointer. • Optical mouse has three buttons. • Optical mouse is less sensitive towards surface.	• In 1988, Richard Lyon, Steve Krish independently invented different versions of Optical Mouse.
3	**Laser Mouse** 	• Measures the motion and acceleration of pointer. • Laser Mouse uses Laser Light. • Laser Mouse is highly sensitive and able to work on any hard surface.	

Table 1.2.1 Commonly used Mouse

(3) **Scanner:** Scanners are used to enter the information directly into the computer's memory. This device works like a Xerox machine. The scanner converts any type of printed or written information including photographs into a digital format, which can be manipulated by the computer.

(4) **Fingerprint Scanner:** Finger print Scanner is a fingerprint recognition device used for computer security, equipped with the fingerprint recognition feature that uses biometric technology. Fingerprint Reader / Scanner is a very safe and convenient device for security instead of using passwords, which is vulnerable to fraud and is hard to remember.

(5) **Track Ball:** Track ball is similar to the upside- down design of the mouse. The user moves the ball directly, while the device itself remains stationary. The user spins the ball in various directions to navigate the screen movements.

(6) **Retinal Scanner:** This performs a retinal scan which is a biometric technique that uses unique patterns on a person's retinal blood vessels.

(7) **Light Pen:** A light pen is a pointing device shaped like a pen and is connected to a monitor. The tip of the light pen contains a light-sensitive element which detects the light from the screen enabling the computer to identify the location of the pen on the screen. Light pens have the advantage of 'drawing' directly onto the screen, but this becomes hard to use, and is also not accurate.

(8) **Optical Character Reader:** It is a device which detects characters printed or written on a paper with OCR, a user can scan a page from a book. The Computer will recognize the characters in the page as letters and punctuation marks and stores. The Scanned document can be edited using a wordprocessor.

(9) **Bar Code / QR Code Reader:** A Bar code is a pattern printed in lines of different thickness. The Bar code reader scans the information on the bar codes transmits to the Computer for further processing. The system gives fast and error free entry of information into the computer. **QR** (Quick response) Code: The QR code is the two dimension bar code which can be read by a camera and processed to interpret the image.

(10) **Voice Input Systems:** Microphone serves as a voice Input device. It captures the voice data and send it to the Computer. Using the microphone along with speech recognition software can offer a completely new approach to input information into the Computer.

(11) **Digital Camera:** It captures images / videos directly in the digital form. It uses a CCD (Charge Coupled Device) electronic chip. When

light falls on the chip through the lens, it converts light rays into digital format.

(12) **Touch Screen:** A touch screen is a display device that allows the user to interact with a computer by using the finger. It can be quite useful as an alternative to a mouse or keyboard for navigating a Graphical User Interface (GUI). Touch screens are used on a wide variety of devices such as computers, laptops, monitors, smart phones, tablets, cash registers and information kiosks. Some touch screens use a grid of infrared beams to sense the presence of a finger instead of utilizing touch-sensitive input.

(13) **Keyer :** A Keyer is a device for signaling by hand, by way of pressing one or more switches. Modern keyers have a large number of switches but not as many as a full size keyboard. Typically, this number is between 4 and 50. A keyer differs from a keyboard, which has "no board", but the keys are arranged in a cluster.

Output Devices:

(1) Monitor: Monitor is the most commonly used output device to display the information. It looks like a TV. Pictures on a monitor are formed with picture elements called PIXELS. Monitors may either be Monochrome which display text or images in Black and White or can be color, which display results in multiple colors. There are many types of monitors available such as CRT (Cathode Ray Tube), LCD (Liquid Crystal Display) and LED (Light Emitting Diodes). The monitor works with the VGA (Video Graphics Array) card. The video graphics card helps the keyboard to communicate with the screen. It acts as an interface between the computer and display monitor. Usually the recent motherboards incorporate built-in video card. The first computer monitor was part of the Xerox Alto computer system, which was released on March 1, 1973.

(2) Plotter: Plotter is an output device that is used to produce graphical output on papers. It uses single color or multi color pens to draw pictures.

(3) Printers: Printers are used to print the information on papers. Printers are divided into two main categories:

- Impact Printers
- Non Impact printers

Impact Printers These printers print with striking of hammers or pins on ribbon. These printers can print on multi-part (using carbon

papers) by using mechanical pressure. For example, Dot Matrix printers and Line matrix printers are impact printers.

A Dot matrix printer that prints using a fixed number of pins or wires. Each dot is produced by a tiny metal rod, also called a “wire” or “pin”, which works by the power of a tiny electromagnet or solenoid, either directly or through a set of small levers. It generally prints one line of text at a time. The printing speed of these printers varies from 30 to 1550 CPS (Character Per Second).

Line matrix printers use a fixed print head for printing. Basically, it prints a page-wide line of dots. But it builds up a line of text by printing lines of dots. Line printers are capable of printing much more than 1000 Lines Per Minute, resulting in thousands of pages per hour. These printers also uses mechanical pressure to print on multi-part (using carbon papers).

Non-Impact Printers

These printers do not use striking mechanism for printing. They use electrostatic or laser technology. Quality and speed of these printers are better than Impact printers. For example, Laser printers and Inkjet printers are non-impact printers.

Laser Printers

Laser printers mostly work with similar technology used by photocopiers. It makes a laser beam scan back and forth across a drum inside the printer, building up a pattern. It can produce very good quality of graphic images. One of the chief characteristics of laser printer is their resolution – how many Dots per inch(DPI). The available resolution range around 1200 dpi. Approximately it can print 100 pages per minute(PPM)

Inkjet Printers

Inkjet Printers use colour cartridges which combined Magenta, Yellow and Cyan inks to create color tones. A black cartridge is also used for monochrome output. Inkjet printers work by spraying ionised ink at a sheet of paper. The speed of Inkjet printers generaly range from 1-20 PPM (Page Per Minute). They use the technology of firing ink by heating it so that it explodes towards the paper in bubbles or by using piezoelectricity in which tiny electric currents controlled by electronic circuits are used inside the printer to spread ink in jet speed. An Inkjet printer can spread millions of dots of ink at the paper every single second.

Speakers: Speakers produce voice output (audio). Using speaker along with speech synthesize software, the computer can provide voice output. This has become very common in places like airlines, schools, banks, railway stations, etc.

Multimedia Projectors

Multimedia projectors are used to produce computer output on a big screen. These are used to display presentations in meeting halls or in classrooms**.**

SUMMARY: Components of a Computer

A computer is made up of several interconnected components that work together to process data and perform tasks. These are broadly classified into hardware and software components.

- **Main Components of a Computer:**
  1. Input Devices Devices that allow users to enter data into the computer.
     Examples: Keyboard, Mouse, Scanner, Microphone.
  2. Output Devices Devices that display or produce the result of computer processes.
     Examples: Monitor, Printer, Speakers.
  3. Central Processing Unit (CPU) Known as the "brain" of the computer. It consists of:
     - ALU (Arithmetic Logic Unit): Performs calculations and logic operations.
     - CU (Control Unit): Directs the operations of the computer.
     - Registers: Small memory areas used for quick data access.
  4. Memory / Storage
     - Primary Memory: RAM (temporary) and ROM (permanent).
     - Secondary Storage: Hard drives, SSDs, USB drives—store data permanently.
  5. Motherboard
     The main circuit board that connects all components.
  6. Software *(not hardware but essential)*
     Set of instructions that tell the hardware what to do. Includes operating systems and applications.

✅ **5 Multiple Choice Questions (MCQs)**

1. Which of the following is considered the "brain" of the computer?

a) RAM
b) Hard Drive
c) CPU
d) Monitor

✔ Answer: c) CPU

2. What does RAM stand for?
   a) Read After Memory
   b) Random Access Memory
   c) Real Access Memory
   d) Rapid Action Machine

   ✔ Answer: b) Random Access Memory

3. Which of these is an output device?
   a) Keyboard
   b) Mouse
   c) Printer
   d) Scanner

   ✔ Answer: c) Printer

4. Which component performs arithmetic and logical operations?
   a) ROM
   b) ALU
   c) CU
   d) HDD

   ✔ Answer: b) ALU

5. Hard drives and SSDs are examples of:
   a) Input devices
   b) Output devices
   c) Primary memory
   d) Secondary storage

   ✔ Answer: d) Secondary storage

✍️ 3 Short Answer Type Questions

1. What is the function of the CPU in a computer?
2. Differentiate between RAM and ROM.
3. Name two input and two output devices.

📝 3 Long Answer Type Questions

1. Explain the main components of a computer and describe their functions.
2. What is the role of memory in a computer system? Differentiate between primary and secondary memory with examples.
3. Describe the architecture and function of the CPU, including its parts: ALU and Control Unit.

UNIT 1.3
Introduction to Computer Networks

Essentially, a computer network is a complex combination of technology, systems and devices that allow the transfer of data, resources and communication across distances and scales. Not just that, these networks became the backbone of our digital world as they enable everything from basic file transfers between personal computers to complex global telecommunication networks composed of billions of devices around the world. The evolution of computer networks traces back to the early days of computing when researchers and institutions sought ways to share computational resources and information more efficiently. Computers were independently operating machines, and data transfer was done physically via punch cards or magnetic tapes. The limitations led to the idea of networking where the computers would communicate and share resources instantly and automatically. We can categorize networks based on geographical coverage and scale, but there are three major network types which we have to know about: Local Area Networks (LAN), Metropolitan Area Networks (MAN) and Wide Area Networks (WAN). These networks can be further classified into various types of networks based on their geographical and technological limits, serving varied communication and resource sharing needs in various spaces.

Local Area Networks (LANs)

Local Area Networks (LANs) are the most intimate and concentrated type of computer networking. These networks generally span a small geographic area (within a single building, office, home, school campus, or small group of adjacent buildings). LANs are known for high data transfer speeds and low latency, making them ideally suited for local communication among computer devices. A Local Area Network (LAN) is a network that connects computers and devices within a limited geographical area, like a home, school, or office building, and aims to facilitate communication, resource sharing, and data exchange between these devices with high efficiency. For example, in a traditional office setting, a LAN might be used to connect desktop computers, servers, printers, and other network-connected devices, enabling employees to share files, access centralized databases, and collaborate effectively. Another application of LANs can be seen in home networks in which a variety of devices like smart phones, computers, smart home devices, and entertainment devices can interact and share internet connectivity. Now a days there are many technologies available for local area networks. Ethernet is the most

common standard for wired LANs and provides high-speed data transmission using twisted-pair copper cables or fiber-optic connections. With Ethernet, one can achieve high data transfer rates from 10 Mbps (Ethernet) to 100 Gbps (100 Gigabit Ethernet), making Ethernet unrivaled in terms of the local area network. Wireless local area networks, or WLANs, have also become hugely popular as Wi-Fi technology has made it possible for devices to communicate wirelesslyand be mobile within the coverage range of the WLAN. There are few major elements that are involved in the architecture of a LAN. Network interface cards (NICs) serve as the physical connectors for individual devices, while network switches serve as central communication hubs, forwarding data packets to devices on the network based on their destination addresses. Routers are gateway devices between the LAN and external networks, such as the internet, and manage traffic flow. Together, these elements form a powerful and efficient foundation for local communications. Resource sharing is one of the major benefits of LANs. File servers: In a centralized location, file servers can store and distribute documents; databases can be accessed by multiple users at the same time; network printers can be shared across an entire organization. In a shared resource model like this, developers gain vastly improved efficiency, lower individual device costs, and a more organized collaborative work process. LAN Design Considerations Security is a major consideration in LAN design and implementation. Firewall Configuration Firewalls create a controlled barrier between local networks and external traffic, filtering out potentially harmful packets of data while allowing legitimate traffic to pass through. Access Control Lists (ACLs) ACLs define who or what can access a resource and what operations they can perform, providing a layer of protection for local networks. Encryption Protocols You can encrypt data in transit using encryption protocols, ensuring data confidentiality and integrity.

Metropolitan Area Networks (MAN)

Metropolitan Area Networks (MANs) are positioned uniquely between local area networks (LANs) and wide area networks (WANs). As the name implies MANs cover much larger areas than LANs typically consisting of entire cities or large campus environments. These provide high-bandwidth connections among multiple buildings, organizations, or institutional sites in a metropolitan area. The focus of a Metropolitan area network is to establish a fast and reliable communication infrastructure connecting various local networks within a defined area. MAN technologies are mainly developed and used by government agencies, educational institutions, large corporations with multiple

facilities within a city and telecommunications providers. LANs being interconnected by MANs allow for more extensive and sophisticated communication and resource-sharing capabilities. The different types of MAN typicallyutilize high-speed Transmission Technology such as fiber optic cables, microwave transmission systems and wireless communication. These technologies support data transmission speeds that are many order of magnitude greater than traditional LAN technologies, normally within 100 Mbps to several Gbps. MANs offer a wide bandwidth and lower latency that makes them suitable for applications involving large scale data transfer and real-time communications. One example of a MAN is the city-provided fibre-optic network that links different municipal institutions such as government services, education, medicine and research. These networks enable seamless communication, shared resources and form the backbone of critical public service infrastructure. In environmental networks, MAN technologies are employed by telecommunications companies to ensure high- speed internet and other communication services across all urban areas. Advanced routing protocols optimize the selection of communication paths, balance loads, and provide redundancy to ensure reliable communication even if certain segments of the network are disrupted. This advanced system enables MANs to ensure optimal uptime and performance across various city landscapes. MAN security issues are more complicated than LAN security issues because The MAN covers a larger area, and multiple networks are interconnected. Far-reaching security approaches incorporate various protective layers that may consist of scrambled correspondence channels, progressed firewall setups, intrusion detecting and prevention frameworks, and inflexible get to control components. The correct use of network segmentation and virtual private network (VPN) technologies are extremely important in preserving the integrity and confidentiality of metropolitan network communications.

Wide Area Networks (WANs)

Wide Area Networks are the most extensive and complicated class of computer networking, covering large areas and linking numerous networks over complete regions, nations, and even continents. WANs act as a worldwide communication backbone, sharing and transmitting data on an unparalleled level. The most common and broadest WAN is the internet itself, which links millions of networks and billions of devices across the globe. WANs create one system for communication that is not limited by geography. In contrast to LANs and MANs, which are limited to geographic conditions, WANs utilize advanced telecommunication technologies and protocols to enable connections over long distances. The networks allow businesses to keep in touch,

share resources, and carry out their business activities anywhere in the world. WANs technology stack is very huge and very heterogeneous. Submarine fiber-optic cables, satellite communications, microwave links, and terrestrial communication lines are common transmission mediums utilized by telecommunications companies and internet service providers. These technologies complement each other to provide a resilient, fault-tolerant worldwide communications network that can process huge amounts of data with incredible reliability. As one of the main WAN protocols, the Internet Protocol (IP) offers basic communication services for WAN. Routing protocols such as Border Gateway Protocol (BGP) facilitate intelligent packet forwarding, ensuring that data reaches its destination through the most efficient path available. This dynamic routing ability is vital to sustain the resilience and performance of global network infrastructures. Wide Area Network connectivity is mostly achieved through aviable means. There are dedicated means of point-to-point communication. Packet-switching approach means, breaking data into packets to be transferred through intermediate nodes dynamically, allowing the network to be efficiently used by more users at a time thus reducing transmission costs because packets follow different paths. VPNs (Virtual Private Networks) reduces overhead and creates encrypted tunnels for secure communications over public networks and infrastructure enabling organizations to create a private communication channel. WANs are widely used by enterprise organizations to keep all offices, data centers, and remote workers interconnected. This helps maintain consistent communication infrastructure as multinational corporations rely on these networks for coordinating global operations and sharing critical business information. The advent of cloud computing and the evolution of distributed computing models have only increased the importance of WANs, with these technologies allowing users access to centralized computational resources from almost any corner of the globe. Wide area network performance optimization requires complex traffic management and ssl>I QoS mechanisms. By deploying sophisticated algorithms, network administrators can identify and prioritize traffic that is crucial to system performance, allocate sufficient bandwidth for important applications, and prevent bottlenecks that can lead to reduced efficiency and higher costs. The vocabulary of the internet comprises a plethora of protocols that govern this process, employing compression techniques, caching mechanisms, and intelligent routing protocols to ensure efficiency and low latency over the long distances involved. Security is a major concern in Wide Area Network design and implementation. Its vast scale and many points of interconnection leave it with plenty of potential weak spots.

Enterprise security defenses include strong encryption protocols, advanced firewall settings, intrusion detection and prevention systems, and consistent monitoring of the network. The rise of zero- trust security models focuses on constant authentication and reduced levels of trust across the different boundaries of the network.

SUMMARY: Introduction to Computer Networks

A **computer network** is a system in which multiple computers and devices are connected together to share resources, exchange files, and communicate. These resources include data, printers, internet access, and applications.

- **Types of Computer Networks:**
  1. **LAN (Local Area Network)** – Covers a small geographic area like a home, school, or office.
  2. **MAN (Metropolitan Area Network)** – Covers a city or a large campus.
  3. **WAN (Wide Area Network)** – Covers large geographical areas (e.g., the internet).

- **Basic Components:**
  - **Nodes** – Devices like computers, printers, or servers in the network.
  - **Switches & Hubs** – Help connect multiple devices in a LAN.
  - **Routers** – Connect different networks and direct data between them.
  - **Transmission Media** – Wired (Ethernet cables) or wireless (Wi-Fi, Bluetooth).

- **Benefits of Networking:**
  - Resource sharing (e.g., printers, files)
  - Centralized data management
  - Faster communication (emails, chats)
  - Remote access and collaboration

Multiple Choice Questions (MCQs)

1. **What is a computer network?**
   a) A computer with multiple monitors
   b) A group of software programs
   c) A group of interconnected computers
   d) A storage device for data
   ✔ **Answer: c) A group of interconnected computers**

2. **Which device is used to connect different networks?**
   a) Switch
   b) Hub
   c) Router
   d) Modem
   ✔ **Answer: c) Router**
3. **Which type of network is used within a building or a small office?**
   a) WAN
   b) LAN
   c) MAN
   d) PAN
   ✔ **Answer: b) LAN**
4. **Wi-Fi is an example of:**
   a) Wired transmission
   b) Optical transmission
   c) Wireless transmission
   d) None of the above
   ✔ **Answer: c) Wireless transmission**
5. **Which of the following is NOT an advantage of computer networks?**
   a) Resource sharing
   b) Centralized data management
   c) Increased hardware costs
   d) Improved communication
   ✔ **Answer: c) Increased hardware costs**

Short Answer Type Questions

1. What is the purpose of a computer network?
2. Differentiate between LAN and WAN.
3. Name two types of transmission media used in networks.

Long Answer Type Questions

1. Explain the basic components of a computer network and their functions.
2. Describe different types of computer networks with suitable examples.
3. Discuss the advantages and disadvantages of using a computer network.

UNIT 1.4

Network topologies

Network Topologies: A Comprehensive Analysis

The basic structural arrangement of how computer networks are organized and connected is represented by Network topologies. Such setup specifies the exact way that the sites develop communication, share assets, and connect and it serves as the basic plan for physical & logical network designing. For network architects, system administrators, and technology professionals who aim to streamline communication infrastructure for improved performance, knowing network topologies is vital for ensuring reliable and efficient data transmission in intricate tech ecosystems. Network topology does not just refer to the interconnection of devices; it is also based on elaborate design principles that affect several aspects of a network reliability, scalability, performance, and resilience. An in-depth analysis of the different types of network topologies allows technology professionals to take strategic decisions regarding network designs, infrastructure investment choices, and implementation of technologies across the enterprise.

Bus topology is one of the most basic forms of network topology and also one of the oldest, consisting of a single linear communication line to which all network devices are directly connected to a single central cable called the network bus or backbone. In this topology, all nodes in the network share a common communication line, transmitting data in sequence down the line of the main cable. Signals are sent by the station to multiple devices at once, but since only the intended recipient uses a specific data packet, this makes it an uncomplicated but potentially difficult communication mechanism. In the case of a bus topology, the building of the architecture is simpler, which is one of the main advantages of it, thus, it appeals more for home networks or early deployments. A linear topology design enables easy cable installation and lower infrastructure requirements for instance, reducing initial setup complexity and total implementation costs. The provision of these devices, with no major disruption to the already developed network, was patented and provided flexibility, especially about network technology. But bus topology has significant disadvantages as well. As more devices are connected, the performance of the network can greatly be impacted because of possible signal collisions and intention for bandwidth. This leads to single point of failure vulnerabilities in which a break in the center cable can render the entire network communication system faulty. As the number of devices and the lengths of the cables

increase, the degradation of the signal becomes more significant, meaning that this technology has inherent scalability limitations, which limit the topologies where it can be effectively deployed, due to the use of devices.

Star Topology

Star topology is an improved and widely used topology that consists of a central network that connects all network nodes. Such architecture establishes a tree communication model where the main node serves as the first communication intermediary between all devices to connect; it manages the data transmissions by routing them to connected devices. Instead, each node in the network connects to the centralized hub through a point-to-point connection, allowing for a more structured and controlled data exchange process. Star topology offers several advantages including simplicity of design, ease of installation and configuration, centralized management, or higher scalability and fault isolation. Since network nodes are independent systems, if one node fails or needs maintenance, the rest of the network infrastructure will not be impaired and carries on with its normal operations. It enables much easier network monitoring/ diagnostics/ performance management from a common place, making it much easier for administrators to design good controls and quickly identify communications issues and performance bottlenecks. Star topology offers significant performance advantages over bus topology in terms of minimizing data transmission conflicts and optimizing bandwidth usage. This dedicated point-to-point connection minimizes the likelihood of collisions and interference, which also contributes to more reliable and predictable communication in comparison to other topologies. Moreover, star topology allows for easy expansion of the network, as new devices can be added by simply connecting them to the central hub without disrupting the existing network infrastructure. While Star topology has many benefits, it is not without potential vulnerabilities. Central hub is a single point of failure, if the central device fails the communication network will fail. This intrinsic design constraint is what requires you to implement redundant hub systems or alternate communication channels to maintain your availability and resilience on the network.

Ring Topology

Open-ended addition: “In this design, devices are the ‘vertices’ and the endpoints of each link are the ‘edges’.” In ring topology, data is transmitted from device to device in a circumnavigate manner, with each device functioning as a repeater to relay data packets to the next

device until the intended recipient is reached. The circular communication mechanism in ring topology provides unique data transmission characteristics, distinguishing this topology from other networking configurations. Data is transmitted in a manner where each node in the ring acknowledges receipt of a packet before forwarding it further along the path, allowing the packet to travel in a single direction throughout the entire ring. The nodes of the network are capable of reconstructing and re-transmitting signals, which greatly increases the range of network communication, while preserving the quality of the signal across potentially very large distances within the network. This ensures that the signals barelyget distorted or lag Instead; methodical routing of data packets takes place with the structured communication pathway. The ring topologywas usuallyimplemented using special hardware network interface cards along with communication protocols that could facilitate the transmission of data around the loop. Closely related, its practical realization can be observed in early token-ring networks during the 1980s and early 1990s. In a token-ring network, they use a special data packet (called the token) that continuously moves around the network and is responsible for determining the right of the node to transmit data. On the other hand, ring topology has some significant drawbacks that have led to its decline in use in present-day network architecture. Because messages can only be sent from one node to the next node, the sequential communication method can be a potential performance bottleneck, depending on the ability of the nodes to receive packets, process packets, and transmit packets. While all devices are connected in a loop and it is desirable to maintain continuity, simplicity of connection ensures that a single node failure can disrupt the entire network's communication route, introducing reliability concerns as the complexity of nodes in the topology increases, leading to the undesirability of ring topology in large, mission-critical network setups.

Mesh Topology

Mesh Topology: Mesh topology is a complex network topology where every node in the network has a direct point-to-point connection with every other node in the network. This approach provides diverse, redundant communication paths at many points to form a super solid and resilient networking architecture. And partial mesh topology connects some nodes to all other nodes while the other nodes have varying interconnectivity, based on the requirement and design constraints of the network. This approach provides the best reliability and fault tolerance with mesh topology. Network communication can dynamically adjust around nodes and connections through many

available communication pathways, enabling continued operational capabilities despite individual failures. Its redundancy makes mesh topology especially suitable for critical infrastructure, telecommunications networks, and applications that demand high network availability and least interruption potential for communication. A mesh topology is costly to implement, and every new node requires numerous network interface connections to be made. The quadratic growth in connection complexity with the number of network nodes imposes a significant scalability barrier. Thus, with increasing network size, full mesh topology become less and less feasible, therefore, most implementations use partial mesh topology, in which, just enough paths provide redundancy without making the infrastructure unmanageable. Wireless mesh networks are modern applications of the principles of mesh topology, generally in distributed communication systems including urban wireless networks, the internet of things (IoT), and community network deployments. One area is related to dynamic routing algorithms that allow nodes to dynamically discover and maintain optimal pathways with minimal overhead on the mobile unit, which will enable support for flexible, self-healing network infrastructures that adapt to their environment.

Tree Topology

A tree topology is a type of network topology that connects multiple star networks with a large bus backbone to establish a hierarchical tree structure. In this topology, a network node is structured in a parent-child relation with the root node being the main communication node from which the subsequent branch nodes derive. Each branch having potential to split into more sub-branches generates multiple sub-networks hence making a tree type structure allowing rich complex communication topology. Tree topology, which is composed of multiple sub networks and point-to-point links, is highly suitable for large enterprise networks, educational institutions, and distributed computing environments because the logical organization and segmented nature of a tree topology leads to reduced complexity during network design. Tree topology offers fine-tuned network control, ease of problem resolution, and optimized communication routing by introducing several layers of interconnectivity. Depending on the needs of the organization, segment can be isolated or connected to one another, granting significant flexibility in design and implementation of networks. Its architectural framework of tree topology also supports systematic and large size scaling of the network as it permits easy attachment of new branches without disturbing the current structure of the network. This allows them to create a modular approach to network

design, where each branch can represent a specific departmental network, geographic location, or functional unit, matching the complexity of the organizational structure. With this setting, network admits could create specific access controls, segment network resources, and regulate communication channels with a level of precision previously unattainable on such a scale. Tree topology, on the other hand, retains the intrinsic risks linked to hierarchical systems. Additionally, the root node is a single point of failure, as when the primary hub fails, network branches may get disconnected. In the domain of mission-critical environments, this leads to the need for redundant root nodes or backup communication methods to implement network reliability and maintain the ability to spam with operational effectiveness.

Hybrid Topology

Hybrid topology is a high-level topology that encourages various topologies to be combined together as per requirements for constructing the best-fit network topology for any project. Hybrid configurations combine different topology characteristics, allowing you to use the strengths of an individual topology type while avoiding any limitations associated with the topology. Such an adaptable architectural design empowers technology practitioners to tailor network architectures that seamlessly cater to multifaceted operational requirements and technical limitations. With hybrid topology, organizations can divide their network infrastructure into various operational zones, each of which consists of topology used in the most suitable topology configuration. For example, a corporate network may use star topology for high-level business departments, mesh topology for important infrastructure units and tree topology for loose filing branches. The modular structure allows for performance to be fine-tuned, reliability to be improved, and the management of the network to be more granular. Implementing hybrid topologies demands a considerable level of sophisticated network design skill, as well as the ability to use modern routing technologies that can effectively bridge disparate network segments. Hybrid infrastructure provides flexibility to the system, but network administrators need to consider communication requirements, performance levels, and technical limitations to reduce the complexity of the network. Communication continuity and the management of intricate interconnections between various topology segments are preserved with more advanced routing protocols and intelligent network management tools. The major benefits of hybrid topology are improved flexibility, increased scalability, and the ability to optimize network performance in different operating environments. Employing

the Right Topology Characteristics In order to fine- tune their networks, organizations can select the topological characteristics that are most beneficial for analysis and related services in different segments.

SUMMARY: Network Topology

Network topology refers to the **physical or logical layout** of a computer network — that is, how different devices (nodes) are connected and how data flows between them. Understanding topology is essential for designing efficient and reliable networks.

- **Types of Network Topologies:**

1. **Bus Topology**
    - All devices are connected to a single backbone cable.
    - Easy to install but can become slow with high traffic.
2. **Star Topology**
    - All devices connect to a central hub or switch.
    - Easy to manage; if the central hub fails, the network goes down.
3. **Ring Topology**
    - Each device is connected to two others, forming a circle.
    - Data travels in one direction; failure in one device can affect the whole network.
4. **Mesh Topology**
    - Every device is connected to every other device.
    - Very reliable but expensive due to high cabling.
5. **Tree Topology**
    - A hybrid of star and bus topologies.
    - Used in larger networks like schools and universities.
6. **Hybrid Topology**
    - A combination of two or more topologies to suit specific needs.

Multiple Choice Questions (MCQs)

1. **Which topology connects each device to a central hub?**
   a) Ring
   b) Bus
   c) Star
   d) Mesh
   ✔ **Answer: c) Star**

2. **In which topology do data packets travel in a circular fashion?**
   a) Mesh
   b) Ring
   c) Bus
   d) Star
   ✔ **Answer: b) Ring**

3. **Which network topology is the most fault-tolerant?**
   a) Star
   b) Ring
   c) Mesh
   d) Bus
   ✔ **Answer: c) Mesh**

4. **Which topology is cost-effective for small networks but slow with heavy traffic?**
   a) Star
   b) Bus
   c) Ring
   d) Tree
   ✔ **Answer: b) Bus**

5. **Tree topology is a combination of which two topologies?**
   a) Star and Ring
   b) Bus and Mesh
   c) Star and Bus
   d) Ring and Mesh
   ✔ **Answer: c) Star and Bus**

Short Answer Type Questions

1. What is network topology?
2. Give one advantage and one disadvantage of mesh topology.

3. Why is star topology widely used in modern networks?

Long Answer Type Questions

1. Compare and contrast bus, ring, and star topologies in terms of structure, advantages, and disadvantages.
2. Explain mesh and hybrid topologies with diagrams and use cases.
3. Describe the importance of choosing the right topology in network design. Include factors like cost, reliability, and scalability.

References

1. Tanenbaum, A. S., Wetherall, D. J., & Feamster, N. (2021). *Computer Networks* (6th ed.). Pearson Education.
2. Rajaraman, V., & Adabala, Neeharika. (2014). *Fundamentals of Computers* (6th ed.). PHI Learning.
3. Kundu, Sudakshina. (2025). *Fundamentals of Computer Networks* (2nd Revised ed.). PHI Learning.
4. Ceruzzi, P. E. (2012). *Computing: A Concise History* (The MIT Press Essential Knowledge series). The MIT Press.
5. "Network Topology: The Physical and Logical Structure of a Network connection Between Model and Nodes" by Engr Wilson Kurt. (2019). Independently Published.

MODULE 2

INTRODUCTION TO THE INTERNET

Objective:

- To understand the different applications and protocols of the Internet, including its advantages and threats.

UNIT 2.1

Internet and Its Applications

It's well known that the Internet is one of the most revolutionary technological advances in human history, transforming how people communicate, work, learn, live and interact with their environment like no other technology before it. Fundamentally, What Is Internet? The Internet is a worldwide interconnection of computer networks that communicate with each other via standardized protocols, with the Internet Protocol (IP) and the Transmission Control Protocol (TCP) being the most widely used protocols. This complex framework of digital interactions transitioned from a basic research and military communication medium to an all- encompassing global structure that permeates almost every facet of contemporary life. The Internet began as ARPANET, a project of the United States Department of Defense's Advanced Research Projects Agency in the late 1960s to create a resilient communication network capable of surviving nuclear conflict. This experiment aimed to develop a distributed messaging system that would still work despite damage to parts of the network. [Tweet "The Internet has its roots back to the 1960s and has come a long way since then."] Since its inception, the Internet has radically transformed human communication, commerce, education, entertainment, and social relationships. It has democratized access to knowledge, opened up new avenues for economic activity, and connected people across distances in a way that was unthinkable mere decades ago. Its capacity to transport an abundance of data instantaneously and continuously has transformed the lives of people and firms and even how human being learns and interrelate.

Email

Historical Development of Email

Electronic mail (email) is among the most important and lasting applications of Internet technology. The mechanism for sending electronic mail between computers was established in the 1970s, with the first mail sent onARPANET in 1971, though its origins date back to 1960s computer networking. For starters, the first network email system was invented by Ray Tomlinson, a computer engineer who chose to use the "@" symbol to separate the name of a user from the host name of the computer they were using. Early email systems were only available to researchers, academics, and government workers. But with the growth of computer networks laying the foundation for personal computing in the late 1980s and 1990s, email rapidly evolved

from a niche means of communication to a way of correspondence widely adopted. The development of user-friendly email interfaces, improved access to the Internet, and the rise of commercial email services such as Hotmail and Yahoo Mail measured this proliferation.

The infrastructure of email itself is technical

Email, at its core, is a complex web of protocols and servers that allows people to send and receive messages across a network of interconnected computers around the world. SMTP is used to send emails and POP (Post Office Protocol)/IMAP (Internet Message Access Protocol) is used to receive and store emails. To facilitate email communication, various protocols have been standardized, allowing emails to be sent and received across different email services and platforms, ensuring a seamless communication experience. An email is sent and passes through several different servers which will act as a relay point to direct the message to its final spark. It incorporates several technical complexities including DNS (Domain Name System) lookups to locate the recipient's email server, authentication methods to protect against spam and unauthorized entries, and complicated routing algorithms that figure out the optimal route to send the message. Most modern email systems use advanced encryption mechanisms to ensure the confidentiality and integrity of transmitted messages.

Features and Functionalities

Modern email systems are about much more than just sending text messages. Rich capabilities have turned email into an all-in-one communication and collaboration platform. This allows users to attach a range of items such as documents, images, videos, and audio recordings. Rich text allows for complex visual formatting, while built-in calendaring and scheduling tools facilitate direct arranging of meetings and events. Today, we can rely on spam-filter mechanisms that rely on machine learning algorithms build-in in the vast majority of email platforms. Advanced security features, such as two-factor authentication, end-to- end encryption, and automatic virus scanning offer users an extra layer of protection against cyber threats. Moreover, many email services provide generous amounts of cloud storage, which means entire piles of correspondence and attached documents can be archived and retrieved from several devices.

Applications in Professional and Personal Settings

It has been a boon to the business world, becoming the default method of communication, killing off letters, and in doing so, drastically slashing costs and speeding up the time between sending a message and

receiving a response. Email versatility: Organizations use it for internal communication, client correspondence, project management, marketing campaigns, and formal documentation. Instant information sharing, collaboration across geographical barriers, and comprehensive communication audit trails have completely transformed workplace productivity. At a more personal level, email is so much more than just basic messaging. It serves as a digital identity, allowing users to register accounts, recover passwords, and communicate with others. Email is a basic authentication and communication mechanism used by social networks, online services, and digital platforms. At best, personal email accounts have become digital filing cabinets, with everything from personal correspondence to helpful digital receipts and notifications.

Saga and the Evolving Landscape

Much as it is widely used though, email presents several problems for the modern digital world. The rising amount of junk email, or spam, remains a serious problem. Identity theft is another major security concern, where attackers try to steal your personal information using fraudulent emails. The emergence of alternative communication tools such as instant messaging and social media are also posing challenges to email's supremacy in various segments of communication. In the face of these challenges, email providers are constantly innovating, introducing advanced machine learning algorithms for spam detection, implementing more sophisticated encryption technologies, and building out additional securityfeatures. The future of email is probably a more integrated, intuitive and intelligent one, in which your email systems are capable of automatically sorting, prioritizing and better handling communications.

Video conferencing

Historical Evolution

The concept of video conferencing originated in the mid-20th century around the same time telecommunications companies were running experiments to find ways of transmitting visual communications over long distances. Early video conferencing systems were expensive to build and deploy and were mostly for large companies and government offices. They were also only available on dedicated communication lines with specialized equipment and were therefore out of reach to most organizations and individuals. The real revolution in video-conferencing came with the spread of fast internet—and advances in digital compression algorithms. Finally, more affordable, easy to use video conferencing applications were deployed in late 1990s to early

2000s. Companies such as Skype, founded in 2003, helped democratize video communication by offering free or low-cost services that could operate on standard personal computers.

Technological Infrastructure

Contemporary video conferencing rests on intricate hierarchies of technological infrastructures that facilitate synchronicity of audio and visual communication over the global networks. WebRTC (Web Real-Time Communication) technologies play a vital role in building smooth, browser-based video comms servers. There are multiple interconnected systems in the technical backbone of video conferencing. These components form a layered framework, encompassing foundational protocols for session initiation, codecs for audio and video interpolation, and routing schemas to facilitate the efficient delivery of media streams. Dynamic adaptive bitrate technologies significantly alter video and audio quality as per the available network circumstances, providing the user the best possible experience.

Professional Applications

Video conferencing has revolutionized communication and collaboration in professional settings. Transitioning to remote work, which was already in the works before the global pandemic, came barreling down, gained unprecedented speed with the introduction of video conferencing technologies. Now businesses can hold meetings, interviews, training sessions, and collaborative projects with participants located anywhere on the globe, eliminating travel costs and dramatically increasing operational flexibility. Video conferencing is used by multinational corporations to provide continuous vocal communication at different geographic locations. As an example of this trend, technical support teams assist their clients and monitor their software in real time through video platforms, and educational institutions conduct their lessons remotely. For instance, video conferencing is used by healthcare providers for telemedicine consultations, offering vital access to specialized medical expertise to patients in remote locales that they could not otherwise receive.

Professional Applications

Video conferencing has revolutionized communication and collaboration in professional settings. Transitioning to remote work, which was already in the works before the global pandemic, came barreling down, gained unprecedented speed with the introduction of video conferencing technologies. Now businesses can hold meetings,

interviews, training sessions, and collaborative projects with participants located anywhere on the globe, eliminating travel costs and dramatically increasing operational flexibility. Video conferencing is used by multinational corporations to provide continuous vocalcommunication at different geographic locations. As an example of this trend, technical support teams assist their clients and monitor their software in real time through video platforms, and educational institutions conduct their lessons remotely. For instance, video conferencing is used by healthcare providers for telemedicine consultations, offering vital access to specialized medical expertise to patients in remote locales that they could not otherwise receive.

Personal and Social Uses

Video conferencing has also changed the dynamics of personal communication on a large scale as well as professional. Families who live far apart from one another can keep in touch via frequent video chats. These technologies enable international journeys of friendship and romance that are unthinkable in previous decades. Video calling features have become integrated into social platforms, thereby normalizing visual communication as a standard interaction method. Video conferencing, through the likes of Zoom, became a lifeline for social interaction during global distress such as the COVID-19 pandemic. Birthday parties, religious events, educational classes, and social events quickly moved online. The second aspect reflected the resilience and adaptability of the technology in facilitating social connections, during adverse conditions.

Features and Functionalities

Modern video conferencing technologies come with a wide range of capabilities that goes well beyond simple audio visual interaction. However, features like screen sharing enable participants to collaborate in real time on documents, presentations, and creative projects. Tools such as virtual whiteboards and background blur and virtual background technologies allow interactive brainstorming sessions and the option to display a more professional image while maintaining privacy. Artificial Intelligence is integrated into advanced platforms to upgrade the user experience. Real-time translation of languages, automatic transcription, and also meeting summation tools are now becoming commonplace. Machine learning algorithms in noise cancellation technologies block out ambient noise for a better communication experience. Certain platforms even include emotion recognition and participant engagement tracking, enabling them to monitor the pulse of the meeting.

Security and Privacy Implications

Since video conferencing has become very common, security and privacy issues have started to come forward as an important concern. Significant cases of unauthorized meeting interruptions and data breaches have led platforms to build stronger security mechanisms. End-to-end encryption, waiting rooms, meeting passwords, and user authentication are now standard security features. Organizations should choose video conferencing platforms that are compliant with the appropriate data protection laws. In contrast, sectors like healthcare and finance have stringent confidentiality demands that necessitate platforms with sophisticated security setups and compliance with certain regulatory

benchmarks. Even now, the latest in cyber security is ensuring that video conferencing technologies are addressing user privacy and data protection.

Chatting

Chat is one of the most popular services on the Internet that enables users to talk to each other in real time. It is an instant messaging platform that allows two or more users to communicate in real-time, either via text, voice or video. Chatting is used for many purposes like Personal conversations, professional communication, customer service, online video game and also in virtual learning. Online chat takes many forms: instant messaging (IM) apps, chat rooms, social media messaging services and so on. Instant messaging applications like WhatsApp, Telegram and Facebook Messenger enable their users to send and receive text messages, images, videos, documents and voice notes. Some chat apps support group chats, letting users chat at the same time. Voice and video chatting functions as found in apps like Zoom, Skype and Google Meet help create a more personal and impactful conversation. Another method of chatting was online chat rooms, where users would join certain groups based on their interests and engage in conversations. These spaces are typically used for professional networking, online communities, and learning. These are anonymous chat services where people are randomly connected to a stranger on the web, and they can do so with anonymity. Though texting has revolutionized communication, makingiteasier and faster compared to traditional communication methods like emails or phone calls. Chat services are used by businesses for customer support where the customers get to solve their queries in real time. AI- powered chatbots are becoming widely adopted to automate responses and provide instant assistance to users. While chatting certainly has its advantages, it does

also come with risks, including privacy issues, cyber bullying, and scamming. However, chat services users have to take care to provide personal information in their context and take care to use secure chat services. Now, chatting is an essential component of digital communication and offers an effective means of connecting with people worldwide.

Blogs

Blogs are online diaries/as Sojourn At Henry Hay in July 2003 The Blog (a website with regular updates and comment is where: "A blog is the same as the reality, where individuals are observed (business or organization) need to write articles or opinions (story/news update) chronologically. It lets people reach out to a worldwide audience, share their ideas, experience and expertise. The term "blog" originates from the word "weblog," which was used to describe a type of digital journal in which users logged details about their daily life or interests. Over the years, blogging has grown into a helpful resource for education, marketing, entertainment, and social awareness. Blogs maycover various subjects, including technology, travel, fashion, health, politics, business, and personal experiences. Now, some blogs are more niche, they focus on digital marketing, software development, fitness tips, booking review, etc. There are various blogs, these include personal blogs (which can be both personal or professional blogs) or corporate blogs (which are blog posts on company websites). 1. Personal Blogs: It covers daily life, experiences, opinions, or hobbies. Businesses create corporate blogs to engage with customers, promote their products, and share news about their industry. Professional blogs are typicallymaintained by experts in a specific field who write about research and information regarding their specialized topics. Ablog commonly consists of a title, content, images, videos, and space for comments from the readers. The vast majority of blogs are built and maintained using blogging platforms like Word Press, Blogger, Medium, and Tumblr. They offer tools to create, edit, and publish blog posts with ease. Some bloggers monetize their blogs using advertisements, sponsored content, and affiliate marketing, so blogging can be actively used as a source of income. Because one of the most advantages of blogging is open a platform for free expression and knowledge sharing. It empowers authors to connect with a worldwide readership, bypassing intermediate publishing channels. But to be successful at blogging, you also need to be consistent, creative, and a good writer so that readers come back and new ones find you. The emergence of social media has shaped blogging patterns, as bloggers have combined their articles with social media like as Instagram,

Twitter, and YouTube to connect with a wider audience. Even with the rise of video content and social media, blogs are still a fundamental element of the Internet. All of these blogging platforms play an important role in the Internet by providing yet another source of information, dissemination of news, entertainment, or professional education that continues to enrich the worlds of digital knowledge.

SUMMARY: Internet and Its Applications

The Internet is a global network that connects millions of computers and devices, allowing users to communicate, share information, and access vast resources. It functions through interconnected networks using standard protocols like TCP/IP.

The Internet supports a wide range of applications that are essential for daily life, education, business, entertainment, and communication.

- **Common Applications of the Internet:**

1. Email (Electronic Mail) – Sending and receiving messages instantly.
2. Web Browsing – Accessing websites and online content using browsers like Chrome or Firefox.
3. Social Media – Platforms like Facebook, Instagram, and X (Twitter) for interaction and sharing.
4. E-commerce – Buying and selling goods/services online (e.g., Amazon, Flipkart).
5. E-learning – Online education via platforms like Coursera, Khan Academy, or Google Classroom.
6. Entertainment – Streaming videos, music, games (e.g., YouTube, Netflix, Spotify).
7. Video Conferencing – Communication tools like Zoom, Google Meet, Skype.

Multiple Choice Questions (MCQs)

1. Which protocol is used to connect devices on the Internet?
   a) HTTP
   b) FTP
   c) TCP/IP
   d) SMTP
   ✔ Answer: c) TCP/IP

2. Which of the following is an example of e-commerce?
   a) Gmail
   b) WhatsApp
   c) Amazon
   d) Google Docs
   ✔ Answer: c) Amazon

3. Which application is used to access websites on the Internet?
   a) Media Player
   b) Web Browser
   c) Spreadsheet
   d) Text Editor
   ✔ Answer: b) Web Browser

4. Which of these is a social media platform?
   a) Google Drive
   b) Instagram
   c) Excel
   d) Zoom
   ✔ Answer: b) Instagram

5. What is the full form of HTTP?
   a) Hyper Tool Transfer Protocol
   b) Hyper Text Transfer Protocol
   c) High Transfer Text Protocol
   d) Hyper Text Trading Protocol
   ✔ **Answer: b) Hyper Text Transfer Protocol**

Short Answer Type Questions

1. What is the Internet?
2. Mention any two applications of the Internet in education.
3. What is the use of web browsers?

Long Answer Type Questions

1. Explain any five major applications of the Internet with examples.
2. Discuss the impact of the Internet on education, business, and communication.
3. Describe how the Internet works, including the role of IP addresses, web browsers, and servers.

UNIT 2.2

Internet protocols (FTP, HTTP)

Introduction to Internet Protocols

The development of internet protocols used to facilitate the communication between devices is at the very heart of the evolution of computer networking, being the broad set of rules which govern the communication of data on the global networks. These communication specifications outline the standards, formats, and procedures for data exchange between various computing devices, facilitating seamless and standardized communication across the intricate landscape of digital connectivity. There are several protocols that form the backbone of the internet communication, among those, the most widely used protocols that played a significant role in molding the modern digital communication and information sharing are File Transfer Protocol (FTP) and Hypertext Transfer Protocol (HTTP).

An Overview of File Transfer Protocol (FTP)

FTP (File Transfer Protocol) is one of the oldest and most basic protocols intended for transferring files between computers on a network. FTP, short for File Transfer Protocol, was a technologydesigned to facilitate file transfers Copy to Clipboardse between computers over a network, first developed back when the internet was first being pieced together in the 1970s. It adopts client-server architecture, with one machine functioning as the client that generates requests for file transfer, while the other machine performs the role of the server that responds to those requests and executes the actual file transfer. FTP has a unique two-connection architecture that makes it different from many other network protocols. This mechanism is split into two correspondences of TCP a control connection and a data connection. The control connection (port 21) is responsible for sending commands and replies between the client and server, while the data connection (port 20) transfers the files. Separating file transfer from file system replication enhances file transfer capabilities byallowing features like resuming interrupted transfers, advanced queuing systems, transaction rollback, and a more resilient transfer architecture. Not wasting any time, the FTP also offers various transfer modes that fit several sorts of network surrounding and move necessities. The most frequently employed transfer methods are: ASCII mode, which is the best-suited for text files and handles required character translations across systems; and Binary mode, which guarantees precise transfer of non-text files like images, executables,

and compressed files. MIME is delivering support for transport modes, which indicates that given protocol is designed such that it satisfies various diverse file transfer requirements within heterogeneous computing environments. Traditional FTP implementations have long been a significant concern of security. The protocol as initially released had data and authentication credentials sent in plaintext; this meant that any third parties could intercept the information and use it for their own ends. Today there are secure versions of the protocol, such as FTPS (FTP_Over_SSL/TLS) and SFTP (SSH_FTP). Using upgrade versions, these are secure; this is done over the encrypted data transmission plus a more robust authentication mechanism, hence keeping the confidentiality of transferring files over the network plus protecting against various attacks, such as snooping, traffic analysis, etc.

HTTP — the Protocol for the Web

HTTP, the Hypertext Transfer Protocol, was a ground-breaking protocol that changed the way information was transmitted and accessed on the World Wide Web. HTTP was created in 1989 as a protocol for sharing hypertext information files between different computers on the Internet as part of Tim Berners-Lee's World Wide Web project. HTTP has a broader purpose compared to FTP, which is specifically aimed at file transfer; in contrast, HTTP can be used to retrieve and transmit many types of resources, such as web pages, images, videos, and other multimedia content. Request-Response If you adapt to a client-server interaction model, you will find that HTTP is based on a request-response. HTTP (Hyper Text Transfer Protocol) – is a type of communication protocol that is used in the world wide web to convey requests and transfer information over the internet, its basic command When a user enters any web address or clicks on a hyperlink it makes an (HTTP) request to the web server, which is the computer that the resource is hosted on. It then sends this request to the server, which processes the request and returns a response consisting of the requested content and some metadata that describes the response. This humble little mechanism has evolved into the transport layer communication protocol that underlies the vast majority of the web today, enabling everything from basic HTML page navigation to full-fledged web applications and service-oriented architectures. The HTTP protocol itself defines a complete list of methods that determine what type of action to be performed on a resource. Some of the most used methods are GET (to get a resource), POST (to submit data to be processed), PUT (to update an existing resource), DELETE (to delete a resource), and HEAD (to get info about a resource). In a way, each of these methods has its own role in the interaction between clients and

servers, giving them their versatility and extensibility to adapt to various web communication needs. Different HTTP version prototypes have also undergone a significant evolution wave from HTTP/0.9 to HTTP/1.1 and HTTP/2 boost â†". HTTP/1.1 also introduced the concept of persistent connections, enabling multiple requests to be sent on a single TCP connection, thus enhancing network efficiency. Although HTTP/2 did not reduce the 2 round trips, it did improve performance by including some other features (e.g. multiplexing, header compression, server push, etc.), resulting in quicker and more efficient web communication. This has been further advanced by HTTP/3, an evolving standard utilizing the QUIC transport protocol that uses UDP instead of TCP**.**

SUMMARY: Network Protocols

Network protocols are a set of rules and conventions that determine how data is transmitted and received across networks. They ensure reliable, secure, and standardized communication between devices in a network. Each protocol serves a specific purpose, such as addressing, routing, error checking, or encryption.

- **Common Network Protocols:**

1. **TCP/IP (Transmission Control Protocol / Internet Protocol)**
    - Foundation of the Internet; handles data transmission and addressing.
2. **HTTP/HTTPS (HyperText Transfer Protocol / Secure)**
    - Used for accessing web pages. HTTPS is the secure version.
3. **FTP (File Transfer Protocol)**
    - Transfers files between computers over a network.
4. **SMTP (Simple Mail Transfer Protocol)**
    - Sends emails from one server to another.
5. **DNS (Domain Name System)**
    - Translates domain names (e.g., www.example.com) into IP addresses.
6. **IP (Internet Protocol)**

- Handles addressing and routing of packets across networks.

Multiple Choice Questions (MCQs)

1. Which protocol is used to send emails?
   a) FTP
   b) HTTP
   c) SMTP
   d) IP
   ✔ Answer: c) SMTP

2. What does HTTP stand for?
   a) High Transfer Text Protocol
   b) Hyper Tool Transfer Protocol
   c) Hyper Text Transfer Protocol
   d) Hyper Terminal Text Protocol
   ✔ Answer: c) Hyper Text Transfer Protocol

3. Which protocol is responsible for translating domain names to IP addresses?
   a) SMTP
   b) DNS
   c) IP
   d) TCP
   ✔ Answer: b) DNS

4. Which protocol ensures reliable data transfer between two devices?
   a) IP
   b) FTP
   c) TCP
   d) DNS
   ✔ Answer: c) TCP

5. What is the full form of FTP?
   a) File Text Protocol
   b) File Transfer Protocol
   c) Fast Transfer Program
   d) File Transmission Provider
   ✔ Answer: b) File Transfer Protocol

Short Answer Type Questions

1. What is the function of TCP/IP in a network?

2. Differentiate between HTTP and HTTPS.
3. Why is DNS important in networking?

Long Answer Type Questions

1. Explain the functions of the following network protocols: HTTP, FTP, SMTP, and DNS.
2. Discuss how TCP/IP works in sending data from one computer to another. Include roles of both TCP and IP.
3. Describe the importance of network protocols in ensuring secure and reliable communication. Provide examples.

UNIT 2.3

Website, search engines

Websites and Search Engines: A Comprehensive Exploration

Websites and search engines are the two main pillars at the foundation of the internet digital estate. These great marvels of technology have reshaped what the human race can do in terms of communication, education, and commerce and data interaction. Websites are the content containers of the Internet, ranging from individual blogs to huge corporate sites, and search engines are the powerful tools that help people find their way through this ocean of information. Websites are the core pillars of the internet, acting as the virtual spaces where information, services, and interactive experiences are delivered to users all over the world.

Websites are essentially a group of interlinked web pages that are served on web servers and retrieved by web browsers. This statement has changed significantly the world over where websites today are not just simple pages anymore but provide fully operational platforms to elaborate from, considering their versatility whether on a business, family or personal, fun purpose. As time goes on, websites have become more sophisticated in their architectural structure. Modern websites are built on a front-end using HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), and JavaScript, which provide multi- layered and interactive experiences. HTML forms the core structure and content, CSS controls its appearance and layout, and JavaScript adds interactivity and functionality. Use ‘Data storage in database servers’ for more These 3 technologies allow developers to build anything from static informative pages on a website to complex web apps which can do almost anything that an installed software on the user’s desktop computer can do. Based on their purpose and functionality, websites can be of various types. Because they can showcase a kind of digital portfolio or blog, personal websites are commonly used by people who want to share their thoughts, experiences and artistic creations. Corporate Website; A corporate website is the website of a business/organization that contains information about its services/products/brand. E-commerce sites are websites that sell goods online, allowing you to make a purchase and giving you essential product details. Educational Websites include sources of learning and online courses. Media Websites News, Entertainment & Multimedia. All the aforementioned websites have different design paradigms and technical integrations that need to be considered in order to achieve the intended purpose. User Experience

of websites would have been one of the most important aspects of today`s web design. That is how User interface (UI) and user experience (UX) design became separate disciplines, dedicated to the design of digital spaces that are intuitive, accessible and encourage engagement. Responsive design is now the default approach that allows a website to smoothly adapt even to different devices and screen sizes like desktop computers, smart phones, and tablets. With the ever-expanding usage of mobile internet, this adaptability is vital in modern society. Search engines are the fundamental navigational architecture of the internet, serving as complex information retrieval systems at enable users to locate relevant information in an ever-expanding digital ecosystem. Powered by complex algorithms, these technologies crawl, index, and rank Web Pages to provide users with the most relevant results related to their questions. The search engine itself is not the end goal; it is simply a way of inferring user intent to offer the highest quality info at their fingertips. How search engines work Search engines operate over millions of complex processes. Web crawlers (also called spiders or bots) are programs that roam the web, finding and indexing new pages. These programs move from one page to another by following links, and they also save information they collect about each website. This data is then processed and indexed, which creates a vast database that can be rapidly referenced when a user submits a query.

This is simply because ranking algorithms keep the search engine functioning. Translating search terms into high-qualityresults usingcomplex mathematical models that take multiple factors into consideration. Classic ranking factors comprise relevance of keywords, authority of the website, quality of backlinks, freshness of the content, and user engagement/ interaction metrics. the algorithms behind modern day search engines such as Google have grown to become progressively more complex, understanding context and the intent of a user and the semantic relationships between words and concepts. Today, Google is by far the biggest search engine in the world, with billions of searches processed every day and with driving the search technology standards in the industry. The PageRank algorithm developed by Larry Page and Sergey Brin was a game changer for search, because it meant that we could rank a webpage not just by the stated relevancy to a search term, but also by the authority of that webpage (based on the quantity and quality of links to that page). It was a radical change in the process of returning search results, which was no longer just the appearance of keywords in web texts. Along with them, other magnate search engines have played a major role in

elevating the quality of the digital ecosystem. Microsoft's Bing is an alternative to Google and is tightly woven into Microsoft's suite of products. Yahoo, a pioneer internet portal, remains a provider of search functionality and content aggregation. As an alternative that allows you to start from scratch without tracking by the search engine it would for sure take on the hunt for, DuckDuckGo is a search engine that doesn't know

you because it doesn't track users. The technical architecture that powers search engines is mind-bogglinglycomplex. Huge data centers with thousands of servers working together handle, store and process terabytes of internet data. These centers utilize advanced cooling systems, redundant power supplies and multiple networking technologies all designed to maintain round the clock operation. We must remember how computationally expensive even the basic functionality of the internet is, with search engines processing petabytes of data every second. In the digital world, website securityis a growing worry. Websites are increasingly being used to handle sensitive user information in addition to becoming virtual meeting grounds for vital business and personal interactions and security measures must be strong. The use of SSL (Secure Sockets Layer) encryption has now become so common that search providers like Google favor websites with HTTPS protocols. To keep up with user confidence and data stability, different strategies of cyber security such as defeat against cross-site scripting and SQL infusion along with various other susceptible variables are extremely critical. Search engine optimization (SEO) has become an essential component for websites wanting to rank in search engines. SEO or Search Engine Optimization is a technical and strategic approach to improving a website's ranking and visibility. Some of these include developing a strong website structure, optimizing good content and ensuring fast loading speed on the website, ensuring mobile responsive design and building a network of quality backlinks. SEO is always changing Search engines edit algorithms consistently, and SEO experts need to change as well. Clearly, the relationship between websites and search engines is symbiotic, though complex as well. For websites, they are highly dependent on search engines for traffic and visibility, while search engines are dependent on websites to provide useful and valuable content to users. Because of this interdependence, web design and search technology have been in a constant state of evolution. Structured data markup like Schema org, enabling websites to share richer information with search engines, resulting in rich search results that feature more relevant context and imagery. Mobile Internet usage has

changed website design and search engine functionality radically. Desktop websites are almost obsolete now as mobile devices have taken the lead and that's why responsive designs are the thumb rule now. Google and other search engines are using mobile-first indexing, means they are taking the mobile version of your website to rank and index. That represents an important change in the way people consume the internet; today, a Smartphone or a tablet is the typical browser choice. Both website functionality and search engine technologies have started incorporating artificial intelligence and machine learning. Through advancements in natural language processing, search engines can now more accurately interpret user intent and context, providing more sophisticated, intelligent search results rather than basic keyword matching. With AI, websites are leveraging chatbots, personalization, and dynamic content adaptation to provide a better user journey. One notable characteristic of the website and search engine landscape around the world is regional homogeneity variation. Although Google rules the world some regions have different search engines they prefer. Baidu has a large share in China, Yandex is a major player in Russia, and Naver is South Korea's top search engine. In these cases, the search engine may contain features and algorithms designed for local language, culture, and user behaviors.

Emphasizing user privacy has grown more and more relevant to the world of website and search engine interactivity. Data Collection Practices Transparency and consent have become paramount Consumer awareness about data collection practices by corporations and organizations has caused a boom in privacy- oriented technologies and the promulgations of regulations for consumer protection like the EU's GDPR. In response, seo search engines and websites are embracing transparency in data usage, allowing users to control their data, and strengthening their data protection measures. Websites and search engines show serious and widespread economic impact and the volume of website oriented business is one of the very powerful income. The digital advertisement became a bilion dollar industry with search engines providing suitable and advanced targeting option and ad serving for advertisers. Change E-commerce websites have revolutionised retail, enabling businesses to tap into global markets and consumers to access products and services from anywhere across the globe. The world of digital economy is here to stay, and we cannot remove the websites, search technologies, the new capabilities enabled through these technologies, and their potential impact. New technologies will continue to change websites and search experiences. Natural language processing andAI-based tools have made voice

searches more powerful than ever. AR and VR technologies are paving way for more immersive web experiences. Technologies such as block chain and the decentralized web promise to revolutionize the infrastructure of the modern web, ensuring user privacy and giving users greater control over their own data. Here are a few possible impacts on websites and search engines; In a nutshell, websites and search engines will be much more personalized, intelligent, and integrated in the future. So the web you see will be tailored to your preferences and context as AI becomes smarter and smarter. We will find these websites, applications and (search) interfaces becoming increasingly indistinguishable from each other and working seamlessly together. Accessibility is still a major factor in modern web design and search technologies. By using the right features, websites can be designed to be used by people with different disabilities, including: screen reader compatibility, keyboard navigation, and appropriate color contrast. This recognizes the need for more inclusive digital experiences which is leading search engines to assess and prioritize web content that is accessible

The technical know-how needed to build and maintain websites has grown more specialized. Web developers are required to learn several programming languages, understand dynamic frameworks and libraries, keep pace with fast- evolving technologies. Front-end developers handle the user interface and experience, back-end developers take care of server-side logic and databases, and full-stack developers operate on both sides. Search Algorithms are continually developing, this means the understanding of context, intent, and content quality is becoming increasingly more sophisticated. Models can now examine both web content with impressive nuance as well as signals of expertise, authoritativeness and trustworthiness. This evolution continues the fight against misinformation and provides more incentive to ensure that quality, reliable content appears higher in search results. The global character of websites and search engines has far-reaching implications for culture and language. Languages no longer matter much as technologies are built upon each other for greater capabilities, where translation is utilized for better search scenarios. Thanks to such new capabilities (and tools) websites can now be easily translated and localized, enabling businesses and content creators to reach international audiences more efficiently. While in the website and search engine ecosystem cyber security is an ever-growing challenge. With the growing interconnectedness and complexity of digital platforms, estimated vulnerabilities and errors are key factors to protect against. This means that websites should use strong security measures,

and search engines are constantly improving their ability to assess and filter potentially harmful or low standard material. The combinations of websites and search engines and other digital technologies make for new forms of communication, commerce/other forms of information sharing. The digital landscape continues to evolve, from artificial intelligence-driven personalization to immersive augmented reality experiences, promising increasingly sophisticated and user-centric web experiences.

SUMMARY: Websites and Search Engines

A **website** is a collection of related web pages located under a single domain name (e.g., www.example.com) and accessible over the Internet. Websites are built using web technologies like HTML, CSS, and JavaScript, and can serve various purposes—informational, commercial, educational, or entertainment.

A **search engine** is a software system designed to search information on the World Wide Web. It uses **web crawlers** to index websites and helps users find relevant information based on keywords.

- **Types of Websites:**
  - **Static Website** – Content remains fixed (e.g., basic company info sites).
  - **Dynamic Website** – Content updates based on user interaction (e.g., e-commerce, news).
  - **E-commerce Websites** – Online stores like Amazon or Flipkart.
  - **Educational Websites** – Used for learning, e.g., Khan Academy, Coursera.

- **Popular Search Engines:**
  - **Google**
  - **Bing**
  - **Yahoo**
  - **DuckDuckGo**
  - **Baidu (China)**

- **Basic Functions of a Search Engine:**

1. **Crawling** – Automatically browsing the web to find new pages.
2. **Indexing** – Storing and organizing web content.
3. **Ranking** – Sorting results based on relevance.
4. **Displaying Results** – Showing links and descriptions to users.

Multiple Choice Questions (MCQs)

1. **Which of the following is a search engine?**
   a) Facebook
   b) Instagram
   c) Google
   d) Wikipedia
   ✔ **Answer: c) Google**

2. **A website is a collection of:**
   a) Documents
   b) Images
   c) Web pages
   d) Search results
   ✔ **Answer: c) Web pages**

3. **Which language is commonly used to create the structure of a website?**
   a) SQL
   b) HTML
   c) Python
   d) C++
   ✔ **Answer: b) HTML**

4. **What is the function of a web crawler?**
   a) To create websites
   b) To rank websites manually
   c) To search for and index web pages
   d) To send emails
   ✔ **Answer: c) To search for and index web pages**

5. **Which of the following is a dynamic website?**
   a) Wikipedia
   b) Amazon

c) HTML tutorial page
d) A static image gallery

✔ **Answer: b) Amazon**

Short Answer Type Questions

1. What is a website?
2. Name two popular search engines..
3. What is the role of a search engine?

Long Answer Type Questions

1. Explain how search engines work, including the steps of crawling, indexing, and ranking.
2. Differentiate between static and dynamic websites with examples.
3. Discuss the importance of websites and search engines in modern education and business.

References

- Comer, D.E. (2018). *"The Internet Book: Everything You Need to Know About Computer Networking and How the Internet Works"* (6th ed.). CRC Press, Part I, pp. 1-80
- Deitel, P.J., Deitel, H.M., & Deitel, A. (2012). *"Internet & World Wide Web: How to Program"* (5th ed.). Prentice Hall, Chapter 1, pp. 2-40
- Forouzan, B.A. (2013). *"Data Communications and Networking"* (5th ed.). McGraw-Hill, Chapter 26 & 27, covering FTP and HTTP.
- Kurose, J.F., & Ross, K.W. (2021). *"Computer Networking: A Top-Down Approach"* (8th ed.). Pearson Education, Chapter 2, sections on Application Layer protocols (FTP, HTTP).
- Niederst Robbins, J. (2018). *"Learning Web Design: A Beginner's Guide to HTML, CSS, JavaScript, and Web Graphics"* (5th ed.). O'Reilly Media, Chapter 1, pp. 3-25
- Lewandowski, D. (2019). *"Web Search: Public Searching of the Internet"* (3rd ed.). Springer, Chapter 2, pp. 15-40

MODULE 3

MS OFFICE

Objective:

To introduce the key applications of MS Office, including MS Word, MS PowerPoint, and MS Excel, and provide hands-on knowledge on how to use them effectively.

MICROSOFT OFFICE OVERVIEW

Microsoft Office 2000 is the most efficient suite of applications for document creation, communication and business information analysis. For many functions, the business platform has evolved from paper to the Web. Microsoft Office 2000 extends desktop productivity to the web, streamlining the way you work and making it easier to share, access and analyze information so you get better results. Office 2000 offers a multitude of new features. Of particular importance for this release are the features that affect the entire suite. These Office-wide, or shared features hold the key to the new realm of functionality enabled by Office 2000. Office 2000 offers a new Web-productivity work style that integrates core productivity tools with the Web to streamline the process of sharing information and working with others. It makes it easier to use an organization's intranet to access vital business information and provides innovative analysis tools that help users make better, timelier business decisions. Office 2000 delivers new levels of resiliency and intelligence, enabling users and organizations to get up and running quickly, stay working and achieve great results with fewer resources.

DESIGN GOALS OF MS-OFFICE:

1. **COMMON USER INTERFACE:**

- While learning one application of the suite you get to learn the operational basics of the other applications, while maintaining some uniqueness in the applications.
- Consistency in MS-Office applications is in the form of :
  i. Tool –Bars
  ii. Menus
  iii. Dialog Boxes
  iv. Customizable features and operational features are similar too.

2. QUICK ACCESS TO OTHER APPLICATIONS

- The MS-Office provides the Microsoft Office Short cut Bar which is used for the following:

i. Create a new file based on templates and wizards

ii. Opening existing files and automatically launching the related applications

iii. Add tasks, make appointments, record tasks and add contacts and journal entries.

iv. Create a new Outlook Message.

v. Switch between and launch Microsoft Office Applications.

3. SHARING DATA ACROSS APPLICATIONS

- Microsoft Office Provides several means of sharing data between applications:

i. Copying – copies the data from the source application to the target applications using the clipboard.

ii. Linking-links the data from the source document to the target document and saves with the source document.

iii. Embedding- embeds the data from the source document to the target document and saves with the source document.

- Microsoft Office extends the data sharing beyond application integration by providing workgroup integration with the Microsoft Outlook.
    - Users can mail documents, spreadsheets, presentations and datafiles from within the source applications.

4. PROVIDING A COMMON LANGUAGE:

- Providing the common language has been a more challenging goal from Microsoft Office. It provides a common macro programming language for all the applications –Visual Basic for the Applications.

COMPONENTS OF MICTROSOFT OFFICE:

- **MS-WORD**
- **MS-EXCEL**
- **MS-POWER-POINT**

UNIT 3.1

MS WORD

Ms-word is a powerful word processor that allows you to create :

- Memos
- Fax coversheets
- Web pages
- Reports
- Mailing labels
- Brochures
- Tables
- And many other professional and business applications.

Ms-word provides easy graphics handling, calculation of the data tables , ability to create a mailing list, list sorting and efficient file management.

Major enhancements of the Word are as follows:

a) AUTO SUMMARIZE FEATURE – automatically summarizes the key points in the document. Word determines the most important sentences and gives a custom summary based on the analysis.

b) AUTO COMPLETE FEATURE- automatically offers suggestions to complete the word or the phrase that has been typed partially. To accept suggestions ,press the Enter Key and the Word automatically replaces the partially typed word with the complete word. Word automatically completes the current date, a day of the week, a month other than the current one, your name and the company name and the AutoText entries. Word automatically completes the current date, a day of the week, a month other than the current one, your name and the company name and the AutoText entries.

c) AUTOMATIC GRAMMAR CHECKING-marks the incorrect grammar with a green wavy line as you type.

d) LETTER WIZARD –helps you format and enter the key information for the letters to ensure that they are consistent and professional. It lets you write quickly and easily and also to add to your letter.

e) OFFICE ASSISTANT – uses *IntelliSense natural-language technology.* The assistant anticipates the kind of help you require and suggests the Help topics on the work that you are doing. This office assistant provides the visual examples and the step –by –step instructions for the specific tasks.

f) SMART SPELLING FEATURE-

- Recognizes your name, your organization's name and the professional names of varying ethnicity.
- Recognizes your writing pattern and does not mark some patterns as errors in the document.
- Ignores the Internet and the file addresses as error in the spellings.

g) NATURAL LANGUAGE grammar checker- offers improved syntactical analysis , better rewrite suggestions and user friendly grammar styles. The Spelling and the grammar checking combination facility – eliminates the separate dialog boxes and provides the interface that lets you proofread the document online.

h) HYPERLINKS FEATURES- links to the Microsoft Outlook, HTML or the other files on any internal and external Web site or file server.

SUMMARY: MS Word

Microsoft Word (MS Word) is a **word processing software** developed by Microsoft. It is part of the **Microsoft Office suite** and is used to create, edit, format, and print text documents. MS Word is widely used in offices, schools, and homes for tasks such as writing letters, creating reports, designing resumes, and more.

- **Key Features of MS Word:**

- **Text Formatting** – Change font, size, style, color, and alignment.
- **Page Layout** – Set margins, page size, orientation, and columns.
- **Tables** – Insert and format tables to organize data.
- **Images & Graphics** – Insert pictures, shapes, charts, and SmartArt.
- **Spelling & Grammar Check** – Automatically detect and suggest corrections.
- **Templates** – Pre-designed layouts for letters, resumes, reports, etc.
- **Mail Merge** – Combine documents with databases for mass emailing or printing.
- **Save & Export** – Save in different formats like .docx, .pdf, etc.

Multiple Choice Questions (MCQs)

1. **Which file extension is used for saving Word documents by default?**
   a) .xls
   b) .ppt
   c) .docx
   d) .txt
   ✔ **Answer: c) .docx**

2. **Which feature in MS Word checks for spelling errors?**
   a) Font
   b) Spell Check
   c) Format Painter
   d) Header
   ✔ **Answer: b) Spell Check**

3. **What is the use of the "Mail Merge" feature?**
   a) Sending emails
   b) Combining text and data for mass printing
   c) Designing tables
   d) Formatting text
   ✔ **Answer: b) Combining text and data for mass printing**

4. **Which tab contains the 'Page Orientation' option?**
   a) Insert
   b) Design
   c) Layout
   d) Review
   ✔ **Answer: c) Layout**

5. **Which of the following is used to repeat the last action in MS Word?**
   a) Ctrl + Z
   b) Ctrl + Y
   c) Ctrl + S
   d) Ctrl + P
   ✔ **Answer: b) Ctrl + Y**

Short Answer Type Questions

1. What is MS Word used for?
2. Name any two features of MS Word.
3. What is the purpose of the ‘Header and Footer’ in MS Word?

Long Answer Type Questions

1. Explain the steps to create and format a basic document in MS Word.
2. What is Mail Merge? Describe its importance and how to use it in MS Word.
3. Discuss the various formatting options available in MS Word and their uses in document preparation.

UNIT 3.2

MS-EXCEL:

MS-Excel is a spreadsheet package. When you start excel, a blank workbook appears in the document window. The workbook is the main document using excel for storing and manipulating the data.

A workbook has individual worksheets each of consisting of data. Each work sheet is made up of 256 columns and 65,536 rows.

The FEATURES OF EXCEL:

a) The multiple Undo feature can Undo up to the last 16 actions.

b) When you quit Microsoft Excel with multiple files open, you get a YES to ALL option. You can choose this option to save all the files before exiting, instead of being prompted to close each open file.

c) Conditional Formats dynamically apply a different font style , pattern and the border to the cells whose values fall outside or with in the limit specified by you.. this lets you quickly spot areas of interest without reading through tables of values.

d) The **Hyperlinks Feature** helps you to create hyperlinks that connect to other office files on the system, your network. A hyperlink can be text in the cell , a graphic or you can write a formula that creates a hyperlink.

e) The **Web Queries** features allow you to create and run the queries to retrieve data available on the World Wide Web.

f) The **Internet assistant wizard** steps you through the process of saving the worksheet data and the charts in the HTML format. You can save the data and the chart as a complete new Web Page or add them to an existing Web Page.

g) The new **Share Workbook** feature lets multiple users open a workbook on the network and edit the document simultaneously.

h) **CellTips** and the **ScrollTips** automatically display the comments added to cell.

i) The worksheet has expanded to include 65,536 rows and you can type up to 32,000 characters in a cell.

j) **Natural Language** formulas allow you to create formulas that use row and the column headers instead of the range references.

k) The **Auditing** and the **Validation** facility allow you to circle the invalid data and to see at a glance all the entries that don't meet your validation rules.

l) The enhanced **Get External Data** features enable you to query Access and the other databases either on the system or a network or the Internet or intranet resources.

SUMMARY: MS Excel

Microsoft Excel is a **spreadsheet application** developed by Microsoft. It is widely used for data entry, calculation, analysis, and visualization. Excel organizes data into **rows and columns**, and allows users to perform **mathematical functions**, create **charts**, and use **formulas** for dynamic computations.

Excel is essential in fields like business, finance, education, and research due to its powerful data-handling and analytical capabilities.

- **Key Features of MS Excel:**
  - **Cells, Rows, and Columns** – Basic structure for organizing data.
  - **Formulas and Functions** – For calculations (e.g., =SUM(), =AVERAGE()).
  - **Charts and Graphs** – Visual representation of data (e.g., bar, pie, line charts).
  - **Sorting and Filtering** – For organizing and finding data quickly.
  - **Pivot Tables** – Summarize large data sets efficiently.
  - **Conditional Formatting** – Highlights data based on conditions.
  - **Data Validation** – Restricts types of data entered in cells.

Multiple Choice Questions (MCQs)

1. Which of the following is the default file extension for an Excel workbook?
   a) .docx
   b) .xlsx
   c) .pptx
   d) .txt
   ✔ Answer: b) .xlsx

2. Which symbol is used to start a formula in Excel?
   a) =
   b) +
   c) $
   d) @
   ✔ Answer: a) =

3. Which function is used to add a range of numbers in Excel?
   a) =ADD()
   b) =TOTAL()
   c) =SUM()
   d) =PLUS()
   ✔ Answer: c) =SUM()

4. What is the intersection of a row and a column called?
   a) Table
   b) Chart
   c) Cell
   d) Formula
   ✔ Answer: c) Cell

5. Which feature helps to quickly fill in values based on a pattern?
   a) Filter
   b) AutoSum
   c) Flash Fill
   d) Data Validation
   ✔ Answer: c) Flash Fill

Short Answer Type Questions

1. What is MS Excel used for?
2. Name any two built-in functions in Excel.
3. What is a cell in Excel

Long Answer Type Questions

1. Explain the structure of an Excel spreadsheet. Describe the role of rows, columns, and cells.
2. What are formulas and functions in MS Excel? Give examples and explain how they are used.

References

- Lambert, J. (2021). *"Microsoft Word 365 Step by Step"* (Microsoft Press), relevant introductory chapters covering basic features and document creation.
- Gookin, D. (2019). *"Word 2019 For Dummies"* (Wiley), Part 1, covering the basics of using Word.
- EC Wuyong. (2022). *"PowerPoint Essentials"*. Relevant introductory sections on creating and editing presentations. CCI Learning (2018). *"Microsoft PowerPoint 2016 Exam 77-729 Certification Guide"*.
- Walkenbach, J. (2019). *"Microsoft Excel 2019 Bible"* (Wiley), Part 1, covering Excel fundamentals and worksheet basics.
- Alexander, M., & Kusleika, R. (2018). *"Excel 2016 Power Programming with VBA"* (Wiley), introductory chapters on the Excel interface and basic operations

MODULE 4

DATABASE

Objective:

To introduce the concepts of database management systems (DBMS), their types, and functions, as well as an overview of computer graphics and their application in databases.

UNIT 4.1

Database Management Systems

Introduction to Database Management System

As the name suggests, the database management system consists of two parts. They are:

1. Database and
2. Management System

What is a Database?

To find out what database is, we have to start from data, which is the basic building block of any DBMS.

Data: Facts, figures, statistics etc. having no particular meaning (e.g. 1, ABC, 19 etc)

Record: Collection of related data items, e.g. in the above example the three data items had no meaning. But if we organize them in the following way, then they collectively represent meaningful information.

	Name	Age
1	ABC	19

Table or **Relation**: Collection of related records

	Name	Age
1	ABC	19
2	DEF	22
3	XYZ	28

The columns of this relation are called **Fields**, **Attributes** or **Domains**. The rows are called **Tuples** or **Records**.

Database: Collection of related relations. Consider the following collection of tables:

T1

Roll	Name	Age
1	ABC	19
2	DEF	22
3	XYZ	28

T2

Roll	Address
1	KOL
2	DEL
3	MUM

T3

Roll	Year
1	I
2	II
3	I

T4

Year	Hostel
I	H1
II	H2

We now have a collection of 4 tables. They can be called a "related collection" because we can clearly find out that there are some common attributes existing in a selected pair of tables. Because of these common attributes we may combine the data of two or more tables together to find out the complete details of a student. Questions like "Which hostel does the youngest student live in?" can be answered now, although ***Age*** and ***Hostel*** attributes are in different tables.

A database in a DBMS could be viewed by lots of different people with different responsibilities

For example, within a company there are different departments, as well as customers, who each need to see different kinds of data. Each employee in the company will have different levels of access to the database with their own customized **front-end** application.

In a database, data is organized strictly in row and column format. The rows are called **Tuple** or **Record**. The data items within one row may belong to different data types. On the other hand, the columns are often called **Domain** or **Attribute**. All the data items within a single attribute are of the same data type.

What is Management System?

A **database-management system** (DBMS) is a collection of interrelated data and a set of programs to access those data. This is a collection of related data with an implicit meaning and hence is a database. The collection of data, usually referred to as the **database**, contains information relevant to an enterprise. The primary goal of a DBMS is to provide a way to store and retrieve database information

that is both *convenient* and *efficient*. By **data,** we mean known facts that can be recorded and that have implicit meaning.

The management system is important because without the existence of some kind of rules and regulations it is not possible to maintain the database. We have to select the particular attributes which should be included in a particular table; the common attributes to create relationship between two tables; if a new record has to be inserted or deleted then which tables should have to be handled etc. These issues must be resolved by having some kind of rules to follow in order to maintain the integrity of the database.

Database systems are designed to manage large bodies of information. Management of data involves both defining structures for storage of information and providing mechanisms for the manipulation of information. In addition, the database system must ensure the safety of the information stored, despite system crashes or attempts at unauthorized access. If data are to be shared among several users, the system must avoid possible anomalous results.

Because information is so important in most organizations, computer scientists have developed a large body of concepts and techniques for managing data. These concepts and technique form the focus of this book. This chapter briefly introduces the principles of database systems.

Database Management System (DBMS) and Its Applications:

A Database management system is a computerized record-keeping system. It is a repository or a container for collection of computerized data files. The overall purpose of DBMS is to allow he users to define, store, retrieve and update the information contained in the database on demand. Information can be anything that is of significance to an individual or organization.

Databases touch all aspects of our lives. Some of the major areas of application are as follows-

1. Banking
2. Airlines
3. Universities
4. Manufacturing and selling
5. Human resources

Enterprise Information

- *Sales*: For customer, product, and purchase information.

- *Accounting*: For payments, receipts, account balances, assets and other accounting information.
- *Human resources*: For information about employees, salaries, payroll taxes, and benefits, and for generation of paychecks.
- *Manufacturing*: For management of the supply chain and for tracking production of items in factories, inventories of items inwarehouses and stores, and orders for items.

Online retailers: For sales data noted above plus online order tracking, generation of recommendation lists, and maintenance of online product evaluations.

Banking and Finance

- *Banking*: For customer information, accounts, loans, and banking transactions.
- *Credit card transactions*: For purchases on credit cards and generation of monthly statements.
- *Finance*: For storing information about holdings, sales, and purchases of financial instruments such as stocks and bonds; also for storing real-time market data to enable online trading by customers and automated trading by the firm.
  - *Universities*: For student information, course registrations, and grades (in addition to standard enterprise information such as human resources and accounting).
  - *Airlines*: For reservations and schedule information. Airlines were among the first to use databases in a geographically distributed manner.
  - *Telecommunication*: For keeping records of calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about the communication networks.

Purpose of Database Systems

Database systems arose in response to early methods of computerized management of commercial data. As an example of such methods, typical of the 1960s, consider part of a university organization that, among other data, keeps information about all instructors, students, departments, and course offerings. One way to keep the information on a computer is to store it in operating system files. To allow users to manipulate the information, the system has a number of application programs that manipulate the files, including programs to:

- ✓ Add new students, instructors, and courses
- ✓ Register students for courses and generate class rosters
- ✓ Assign grades to students, compute grade point averages (GPA), and generate transcripts

System programmers wrote these application programs to meet the needs of the university.

New application programs are added to the system as the need arises. For example, suppose that a university decides to create a new major (say, computer science).As a result, the university creates a new department and creates new permanent files (or adds information to existing files) to record information about all the instructors in the department, students in that major, course offerings, degree requirements, etc. The university may have to write new application programs to deal with rules specific to the new major. New application programs may also have to be written to handle new rules in the university. Thus, as time goes by, the system acquires more files and more application programs.

This typical **file-processing system** is supported by a conventional operating system. The system stores permanent records in various files, and it needs different application programs to extract records from, and add records to, the appropriate files. Before database management systems (DBMSs) were introduced, organizations usually stored information in such systems. Keeping organizational information in a file- processing system has a number of major disadvantages.

Data redundancy and inconsistency. Since different programmers create the files and application programs over a long period, the various files are likely to have different structures and the programs may be written in several programming languages. Moreover, the same information may be duplicated in several places (files). For example, if a student has a double major (say, music and mathematics) the address and telephone number of that student may appear in a file that consists of student records of students in the Music department and in a file that consists of student records of students in the Mathematics department. This redundancy leads to higher storage and access cost. In addition, it may lead to **data inconsistency**; that is, the various copies of the same data may no longer agree. For example, a changed student address may be reflected in the Music department records but not elsewhere in the system.

Difficulty in accessing data. Suppose that one of the university clerks needs to find out the names of all students who live within a particular postal-code area. The clerk asks the data-processing department to generate such a list. Because the designers of the original system did not anticipate this request, there is no application program on hand to

meet it. There is, however, an application program to generate the list of *all* students.

The university clerk has now two choices: either obtain the list of all students and extract the needed information manually or ask a programmer to write the necessary application program. Both alternatives are obviously unsatisfactory. Suppose that such a program is written, and that, several days later, the same clerk needs to trim that list to include only those students who have taken at least 60 credit hours. As expected, a program to generate such a list does not exist. Again, the clerk has the preceding two options, neither of which is satisfactory. The point here is that conventional file-processing environments do not allow needed data to be retrieved in a convenient and efficient manner. More responsive data-retrieval systems are required for general use.

Data isolation. Because data are scattered in various files, and files may be in different formats, writing new application programs to retrieve the appropriate data is difficult.

Integrity problems. The data values stored in the database must satisfy certain types of **consistency constraints**. Suppose the university maintains an account for each department, and records the balance amount in each account. Suppose also that the university requires that the account balance of a department may never fall below zero. Developers enforce these constraints in the system by adding appropriate code in the various application programs. However, when new constraints are added, it is difficult to change the programs to enforce them. The problem is compounded when constraints involve several data items from different files.

Atomicity problems. A computer system, like any other device, is subject to failure. In many applications, it is crucial that, if a failure occurs, the data be restored to the consistent state that existed prior to the failure.

Concurrent-access anomalies. For the sake of overall performance of the system and faster response, many systems allow multiple users to update the data simultaneously. Indeed, today, the largest Internet retailers may have millions of accesses per day to their data by shoppers. In such an environment, interaction of concurrent updates is possible and may result in inconsistent data. Consider department *A*, with an account balance of $10,000. If two department clerks debit the account balance (by say $500 and $100, respectively) of department *A* at almost exactly the same time, the result of the concurrent executions may leave the budget in an incorrect (or inconsistent) state. Suppose that the programs executing on behalf of each withdrawal read the old balance, reduce that value by the amount being withdrawn, and write

the result back. If the two programs run concurrently, they may both read the value $10,000, and write back $9500 and $9900, respectively. Depending on which one writes the value last, the account balance of department *A* may contain either $9500 or $9900, rather than the correct value of $9400. To guard against this possibility, the system must maintain some form of supervision.

But supervision is difficult to provide because data may be accessed by many different application programs that have not been coordinated previously.

Security problems. Not every user of the database system should be able to access all the data. For example, in a university, payroll personnel need to see only that part of the database that has financial information. They do not need access to information about academic records. But, since application programs are added to the file-processing system in an ad hoc manner, enforcing such security constraints is difficult.

These difficulties, among others, prompted the development of database systems. In what follows, we shall see the concepts and algorithms that enable database systems to solve the problems with file-processing systems.

Advantages of DBMS

Controlling of Redundancy: Data redundancy refers to the duplication of data (i.e storing same data multiple times). In a database system, by having a centralized database and centralized control of data by the DBA the unnecessary duplication of data is avoided. It also eliminates the extra time for processing the large volume of data. It results in saving the storage space.

Improved Data Sharing : DBMS allows a user to share the data in any number of application programs**.**

Data Integrity : Integrity means that the data in the database is accurate. Centralized control of the data helps in permitting the administrator to define integrity constraints to the data in the database. For example: in customer database we can can enforce an integrity that it must accept the customer only from Noida and Meerut city.

Security : Having complete authority over the operational data, enables the DBA in ensuring that the only mean of access to the database is through proper channels. The DBA can define authorization checks to be carried out whenever access to sensitive data is attempted**.**

Data Consistency : By eliminating data redundancy, we greatly reduce the opportunities for inconsistency. For example: is a customer address is stored only once, we cannot have disagreement on the stored values.

Also updating data values is greatly simplified when each value is stored in one place only. Finally, we avoid the wasted storage that results from redundant data storage**.**

Efficient Data Access: In a database system, the data is managed by the DBMS and all access to the data is through the DBMS providing a key to effective data processing

Enforcements of Standards: With the centralized of data, DBA can establish and enforce the data standards which may include the naming conventions, data quality standards etc.

Data Independence: In a database system, the database management system provides the interface between the application programs and the data. When changes are made to the data representation, the meta data obtained by the DBMS is changed but the DBMS is continues to provide the data to application program in the previously used way. The DBMs handles the task of transformation of data wherever necessary.

Reduced Application Development and Maintenance Time: DBMS supports many important functions that are common to many applications, accessing data stored in the DBMS, which facilitates the quick development of application.

Disadvantages of DBMS

1) It is bit complex. Since it supports multiple functionality to give the user the best, the underlying software has become complex. The designers and developers should have thorough knowledge about the software to get the most out of it.
2) Because of its complexity and functionality, it uses large amount of memory. It also needs large memory to run efficiently.
3) DBMS system works on the centralized system, i.e.; all the users from all over the world access this database. Hence any failure of the DBMS, will impact all the users.
4) DBMS is generalized software, i.e.; it is written work on the entire systems rather specific one. Hence some of the application will run slow.

Data Models

Underlying the structure of a database is the data model: a collection of conceptual tools for describing data, data relationships, data semantics, and consistency constraints. A data model provides a way to describe the design of a database at the physical, logical, and view levels

The data models can be classified into four different categories:

- **Relational Model**. The relational model uses a collection of tables to represent both data and the relationships among those data. Each table has multiple columns, and each column has a unique name. Tables are also known as **relations**. The relational model is an example of a record-based model.

  Record-based models are so named because the database is structured in fixed-format records of several types. Each table contains records of a particular type. Each record type defines a fixed number of fields, or attributes. The columns of the table correspond to the attributes of the record type. The relational data model is the most widely used data model, and a vast majority of current database systems are based on the relational model.

  Entity-Relationship Model. The entity-relationship (E-R) data model uses a collection of basic objects, called

  entities, and *relationships* among these objects.

  An entity is a "thing" or "object" in the real world that is distinguishable from other objects. The entity- relationship model is widely used in database design.

  Object-Based Data Model. Object-oriented programming (especially in Java, C++, or C#) has become the dominant software-development methodology. This led to the development of an object-oriented data model that can be seen as extending the E-R model with notions of encapsulation, methods (functions), and object identity. The object-relational data model combines features of the object-oriented data model and relational data model.

  Semi-structured Data Model. The semi-structured data model permits the specification of data where individual data items of the same type may have different sets of attributes. This is in contrast to the data models mentioned earlier, where every data item of a particular type must have the same set of attributes. The **Extensible Markup Language (XML)** is widely used to represent semi-structured data.

  Historically, the **network data model** and the **hierarchical data model** preceded the relational data model. These models were tied closely to the underlying implementation, and complicated the task of modeling data. As a result they are used little now, except in old database code that is still in service in some places.

Database Languages

A database system provides a **data-definition language** to specify the database

schema and a **data-manipulation language** to express database queries and updates. In practice, the data- definition and data-manipulation languages are not two separate languages; instead they simply form parts of a single database language, such as the widely used SQL language.

Data-Manipulation Language

A **data-manipulation language (DML)** is a language that enables users to access or manipulate data as organized by the appropriate data model. The types of access are:

- Retrieval of information stored in the database
- Insertion of new information into the database
- Deletion of information from the database
- Modification of information stored in the database

There are basically two types:

- **Procedural DMLs** require a user to specify *what* data are needed and *how* to get those data.
- **Declarative DMLs** (also referred to as nonprocedural DMLs) require a user to specify *what* data are needed

without specifying how to get those data.

Declarative DMLs are usually easier to learn and use than are procedural DMLs. However, since a user does not have to specify how to get the data, the database system has to figure out an efficient means of accessing data. A **query** is a statement requesting the retrieval of information. The portion of a DML that involves information retrieval is called a **query language**. Although technically incorrect, it is common practice to use the terms *query language* and *data-manipulation language* synonymously.

Data-Definition Language (DDL)

We specify a database schema by a set of definitions expressed by a special language called a **data-definition language** (**DDL**). The DDL is also used to specify additional properties of the data.

We specify the storage structure and access methods used by the database system by a set of statements in a special type of DDL called a **data storage and definition** language. These statements define the implementation details of the database schemas, which are usually hidden from the users.

The data values stored in the database must satisfy certain **consistency constraints**.

For example, suppose the university requires that the account balance of a department must never be negative. The DDL provides facilities to specify such constraints. The database system checks these constraints every time the database is updated. In general, a constraint can be an arbitrary predicate pertaining to the database. However, arbitrary predicates may be costly to test. Thus, database systems implement integrity constraints that can be tested with minimal overhead.

- **Domain Constraints**. A domain of possible values must be associated with every attribute (for example, integer types, character types, date/time types). Declaring an attribute to be of a particular domain acts as a constraint on the values that it can take. Domain constraints are the most elementary form of integrity constraint. They are tested easily by the system whenever a new data item is entered into the database.

- **Referential Integrity**. There are cases where we wish to ensure that a value that appears in one relation for a given set of attributes also appears in a certain set of attributes in another relation (referential integrity). For example, the department listed for each course must be one that actually exists. More precisely, the *dept name* value in a *course* record must appear in the *dept name* attribute of some record of the *department* relation.

Database modifications can cause violations of referential integrity. When a referential-integrity constraint is violated, the normal procedure is to reject the action that caused the violation.

- **Assertions**. An assertion is any condition that the database must always satisfy. Domain constraints and referential-integrity constraints are special forms of assertions. However, there are many constraints that we cannot express by using only these special forms. For example, "Every department must have at least five courses offered every semester" must be expressed as an assertion. When an assertion is created, the system tests it for validity. If the assertion is valid, then any future modification to the database is allowed only if it does not cause that assertion to be violated.

- **Authorization**. We may want to differentiate among the users as far as the type of access they are permitted on various data values in the database. These differentiations are expressed in terms of **authorization**, the most common being: **read authorization**, which allows reading, but not modification, of data; **insert authorization**, which allows insertion of new data, but not modification of existing data; **update authorization**, which allows modification, but not deletion, of data; and **delete authorization**, which allows deletion of data. We may assign the user all, none, or a combination of these types of authorization.

- The DDL, just like any other programming language, gets as input some instructions (statements) and generates some output. The output of the DDL is placed in the **data dictionary**,which contains **metadata**–that is, data about data. The data dictionary is considered to be a special type of table that can only be accessed and updated by the database system itself (not a regular user). The database system consults the data dictionary before reading or modifying actual data.

Data Dictionary

We can define a data dictionary as a DBMS component that stores the definition of data characteristics and relationships. You may recall that such "data about data" were labeled metadata. The DBMS data dictionary provides the DBMS with its self describing characteristic. In effect, the data dictionary resembles and X-ray of the company's entire data set, and is a crucial element in the data administration function.

The two main types of data dictionary exist, integrated and stand alone. An integrated data dictionary is included with the DBMS. For example, all relational DBMSs include a built in data dictionary or system catalog that is frequently accessed and updated by the RDBMS. Other DBMSs especially older types, do not have a built in data dictionary instead the DBA may use third party stand alone data dictionary systems.

Data dictionaries can also be classified as active or passive. An active data dictionary is automatically updated by the DBMS with every database access, thereby keeping its access information up-to-date. A passive data dictionary is not updated automatically and usually requires a batch process to be run. Data dictionary access information is normally used by the DBMS for query optimization purpose.

The data dictionary's main function is to store the description of all objects that interact with the database. Integrated data dictionaries tend to limit their metadata to the data managed by the DBMS. Stand alone data dictionary systems are more usually more flexible and allow the DBA to describe and manage all the organization's data, whether or not they are computerized. Whatever the data dictionary's format, its existence provides database designers and end users with a much improved ability to communicate. In addition, the data dictionary is the tool that helps the DBA to resolve data conflicts.

Although, there is no standard format for the information stored in the data dictionary several features are common. For example, the data dictionary typically stores descriptions of all:

- Data elements that are define in all tables of all databases. Specifically the data dictionary stores the name, datatypes, display formats, internal storage formats, and validation rules.

The data dictionary tells where an element is used, by whom it is used and so on.

- Tables define in all databases. For example, the data dictionary is likely to store the name of the table creator, the date of creation access authorizations, the number of columns, and so on.
- Indexes define for each database tables. For each index the DBMS stores at least the index name the attributes used, the location, specific index characteristics and the creation date.
- Define databases: who created each database, the date of creation where the database is located, who the DBA is and so on.
- End users and The Administrators of the data base
- Programs that access the database including screen formats, report formats application formats, SQL queries and so on.
- Access authorization for all users of all databases.
- Relationships among data elements which elements are involved: whether the relationship are mandatory or optional, the connectivity and cardinality and so on.

Database Administrators and Database Users

A primary goal of a database system is to retrieve information from and store new information in the database. People who work with a database can be categorized as database users or database administrators.

Database Users and User Interfaces

There are four different types of database-system users, differentiated by the way they expect to interact with the system. Different types of user interfaces have been designed for the different types of users.

Naive users are unsophisticated users who interact with the system by invoking one of the application programs that have been written previously. For example, a bank teller who needs to transfer $50 from account *A* to account *B* invokes a program called *transfer*. This program asks the teller for the amount of money to be transferred, the account from which the money is to be transferred, and the account to which the money is to be transferred.

Application programmers are computer professionals who write application programs. Application programmers can choose from many tools to develop user interfaces. **Rapid application development (RAD)** tools are tools that enable an application programmer to construct forms and reports without writing a program. There are also

special types of programming languages that combine imperative control structures (for example, for loops, while loops and if-then-else statements) with statements of the data manipulation language. These languages, sometimes called *fourth-generation languages*, often include special features to facilitate the generation of forms and the display of data on the screen. Most major commercial database systems include a fourth generation language.

Sophisticated users interact with the system without writing programs. Instead, they form their requests in a database query language. They submit each such query to a **query processor**, whose function is to break down DML statements into instructions that the storage manager understands. Analysts who submit queries to explore data in the database fall in this category.

Online analytical processing (OLAP) tools simplify analysts' tasks by letting them view summaries of data in different ways. For instance, an analyst can see total sales by region (for example, North, South, East, and West), or by product, or by a combination of region and product (that is, total sales of each product in each region). The tools also permit the analyst to select specific regions, look at data in more detail (for example, sales by city within a region) or look at the data in less detail (for example, aggregate products together by category).

Another class of tools for analysts is **data mining** tools, which help them find certain kinds of patterns in data. **Specialized users** are sophisticated users who write specialized database applications that do not fit into the traditional data-processing framework.

Among these applications are computer-aided design systems, knowledge base and expert systems, systems that store data with complex data types (for example, graphics data and audio data), and environment-modeling systems.

SUMMARY: Introduction to DBMS

A **Database Management System (DBMS)** is software that helps in the storage, retrieval, and management of data in databases. It acts as an interface between the user and the database, ensuring that data is consistently organized and easily accessible.

Key Concepts:

- **Database**: A collection of logically related data.
- **DBMS**: Software used to manage databases efficiently and effectively.
- **Examples**: MySQL, Oracle, PostgreSQL, MS SQL Server, MongoDB.

Functions of DBMS:

- Data storage, retrieval, and update
- User access control and data security
- Backup and recovery
- Data integrity and consistency

Types of DBMS:

1. **Hierarchical DBMS**
2. **Network DBMS**
3. **Relational DBMS (RDBMS)** – Most commonly used
4. **Object-oriented DBMS**

Advantages of DBMS:

- Reduces data redundancy
- Ensures data integrity
- Provides data security
- Facilitates data sharing among users

Disadvantages:

- Cost of hardware and software
- Complexity
- High initial setup cost

✅ **Multiple Choice Questions (MCQs)**

1. **Which of the following is NOT a type of DBMS?**
   A) Hierarchical
   B) Network
   C) Relational
   D) Circular
   → **Answer: D) Circular**

2. **Which of these is a primary function of a DBMS?**
   A) Creating websites
   B) Managing files manually
   C) Providing backup and recovery
   D) Designing processors
   → **Answer: C) Providing backup and recovery**

3. **What does DBMS stand for?**
   A) Digital Basic Management Software
   B) Data Backup and Management Service
   C) Database Management System
   D) Direct Binary Management Source
   → **Answer: C) Database Management System**

4. **Which of the following is an example of a relational DBMS?**
   A) MongoDB
   B) MySQL
   C) XML
   D) JSON
   → **Answer: B) MySQL**

5. **What is one key advantage of using a DBMS?**
   A) Increases data redundancy
   B) Provides low-level programming
   C) Enhances data security
   D) Decreases data integrity
   → **Answer: C) Enhances data security**

✅ Short Answer Type Questions

1. Define a Database Management System (DBMS).
2. Mention any two advantages of using a DBMS.
3. List two examples of DBMS software.

✅ Long Answer Type Questions

1. Explain the key features and advantages of a DBMS.
2. Differentiate between the types of DBMS with examples.
3. Describe the functions of a DBMS in detail.

UNIT 4.2

Database Architecture

The architecture of a database system is greatly influenced by the underlying computer system on which the database system runs. Database systems can be centralized, or client-server, where one server machine executes work on behalf of multiple client machines. Database systems can also be designed to exploit parallel computer architectures. Distributed databases span multiple geographically separated machines.

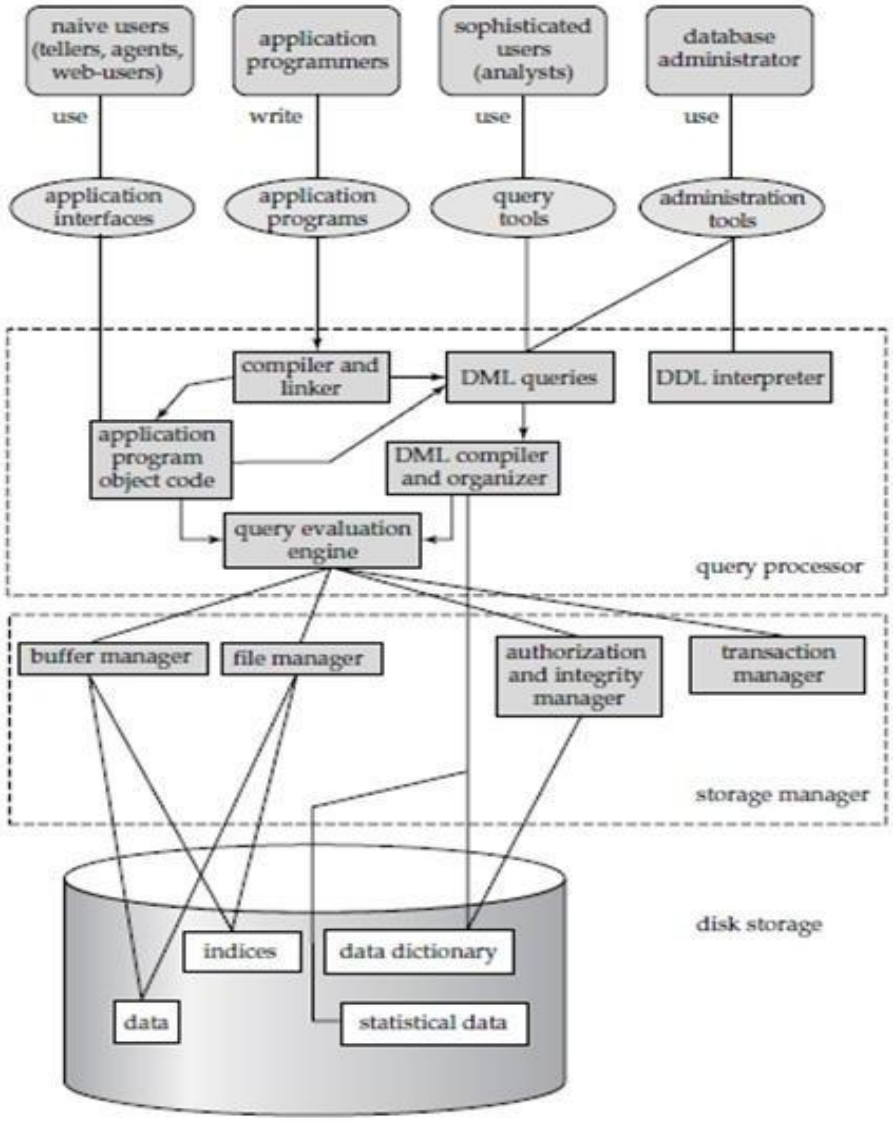


Figure 4.2.1 Database System Architecture

A database system is partitioned into modules that deal with each of the responsibilities of the overall system. The functional components of a database system can be broadly divided into the storage manager and the query processor components. The storage manager is important because databases typically require a large amount of storage space. The query processor is important because it helps the database system simplify and facilitate access to data.

It is the job of the database system to translate updates and queries written in a nonprocedural language, at the logical level, into an efficient sequence of operations at the physical level.

Database applications are usually partitioned into two or three parts, as in Figure . In a two-tier architecture, the application resides at the client machine, where it invokes database system functionality at the server machine through query language statements. Application program interface standards like ODBC and JDBC are used for interaction between the client and the server. In contrast, in a three-tier architecture, the client machine acts as merely a front end and does not contain any direct database calls. Instead, the client end communicates with an application server, usually through a forms interface.

The application server in turn communicates with a database system to access data. The business logic of the application, which says what actions to carry out under what conditions, is embedded in the application server, instead of being distributed across multiple clients. Three-tier applications are more appropriate for large applications, and for applications that run on the WorldWideWeb.

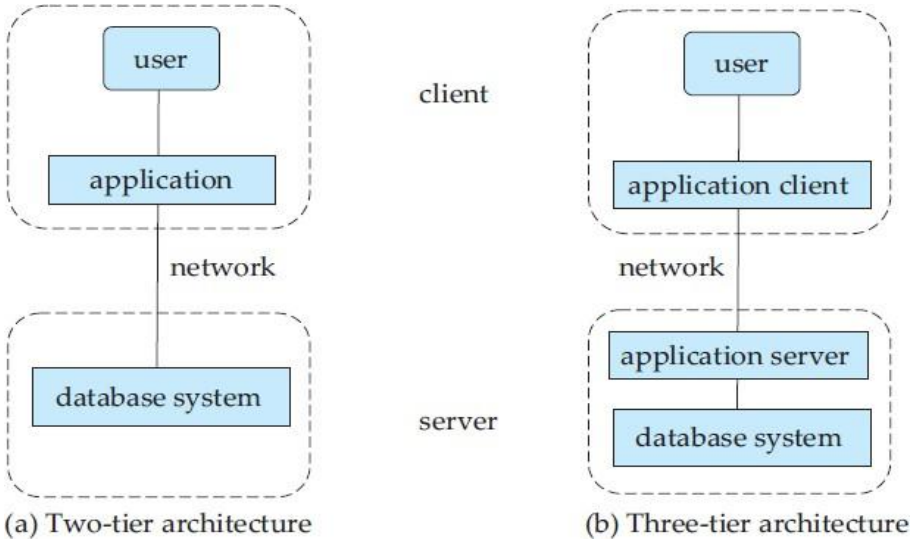


Fig. 4.2.2: Two-tier and three-tier architectures.

- **Query Processor:** The query processor components include **DDL interpreter,** which interprets DDL statements and records the definitions in the data dictionary.
- **DML compiler,** which translates DML statements in a query language into an evaluation plan consisting of low-level instructions that the query evaluation engine understands.

A query can usually be translated into any of a number of alternative evaluation plans that all give the same result. The DML compiler also performs **query optimization**, that is, it picks the lowest cost evaluation plan from among the alternatives.

- **Query evaluation engine,** which executes low-level instructions generated by the DML compiler **DDL interpreter,** which interprets DDL statements and records the definitions in the data dictionary.
- **DML compiler,** which translates DML statements in a query language into an evaluation plan consisting of low-level instructions that the query evaluation engine understands.

A query can usually be translated into any of a number of alternative evaluation plans that all give the same result. The DML compiler also performs **query optimization**, that is, it picks the lowest cost evaluation plan from among the alternatives.

Query evaluation engine, which executes low-level instructions generated by the DML compiler.

A *storage manager* is a program module that provides the interface between the low level data stored in the database and the application programs and queries submitted to the system. The storage manager is responsible for the interaction with the file manager. The raw data are stored on the disk using the file system, which is usually provided by a conventional operating system. The storage manager translates the various DML statements into low-level file-system commands. Thus, the storage manager is responsible for storing, retrieving, and updating data in the database.

The storage manager components include:

- **Authorization and integrity manager**, which tests for the satisfaction of integrity constraints and checks the authority of users to access data.
- **Transaction manager**, which ensures that the database remains in a consistent (correct) state despite system failures, and that concurrent transaction executions proceed without conflicting.
- **File manager**, which manages the allocation of space on disk storage and the data structures used to represent information stored on disk.

Buffer manager, which is responsible for fetching data from disk storage into main memory, and deciding what data to cache in main memory. The buffer manager is a critical part of the database system, since it enables the database to handle data sizes that are much larger than the size of main memory.

Transaction Manager:

A transaction is a collection of operations that performs a single logical function in a database application. Each transaction is a unit of both atomicity and consistency. Thus, we require that transactions do not violate any database-consistency constraints. That is, if the database was consistent when a transaction started, the database must be consistent when the transaction successfully terminates. Transaction - manager ensures that the database remains in a consistent (correct) state despite system failures (e.g., power failures and operating system crashes) and transaction failures.

Conceptual Database Design - Entity Relationship(ER) Modeling

Database Design Techniques

1. ER Modeling (Top down Approach)
2. Normalization (Bottom Up approach)

What is ER Modeling?

A graphical technique for understanding and organizing the data independent of the actual database implementation

We need to be familiar with the following terms to go further**.**

Entity

Any thing that has an independent existence and about which we collect data. It is also known as entity type. In ER modeling, notation for entity is given below.

Entity instance

Entity instance is a particular member of the entity type. Example for entity instance : A particular employee

Regular Entity

An entity which has its own key attribute is a regular entity. Example for regular entity : Employee.

Weak entity

An entity which depends on other entity for its existence and doesn't have any key attribute of its own is a weak entity.

Example for a weak entity : In a parent/child relationship, a parent is considered as a strong entity and the child is a weak entity.

In ER modeling, notation for weak entity is given below

Attributes

Properties/characteristics which describe entities are called attributes. In ER modeling, notation for attribute is given below.

Domain of Attributes

The set of possible values that an attribute can take is called the domain of the attribute. For example, the attribute day may take any value from the set {Monday, Tuesday ... Friday}. Hence this set can be termed as the domain of the attribute day.

Key attribute

The attribute (or combination of attributes) which is unique for every entity instance is called key attributeE.g the employee_id of an employee, pan_card_number of a person etc.If the key attribute consists of two or more attributes in combination, it is called a composite key.

In ER modeling, notation for key attribute is given below.

Simple attribute

If an attribute cannot be divided into simpler components, it is a simple attribute. Example for simple attribute : employee_id of an employee.

Composite attribute

If an attribute can be split into components, it is called a composite attribute.

Example for composite attribute : Name of the employee which can be split into First_name, Middle_name, and Last_name.

Single valued Attributes

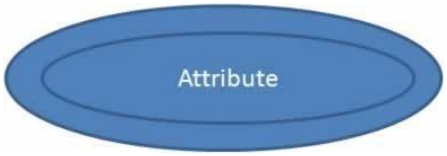
If an attribute can take only a single value for each entity instance, it is a single valued attribute. example for single valued attribute : age of a student. It can take only one value for a particular student.

Multi-valued Attributes

If an attribute can take more than one value for each entity instance, it is a multi-valued attribute. Multi-valued

example for multi valued attribute : telephone number of an employee, a particular employee may have multiple telephone numbers.

In ER modeling, notation for multi-valued attribute is given below.



Stored Attribute

An attribute which need to be stored permanently is a stored attribute Example for stored attribute : name of a student

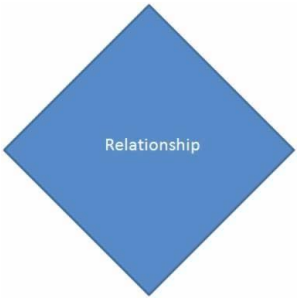
Derived Attribute

An attribute which can be calculated or derived based on other attributes is a derived attribute.

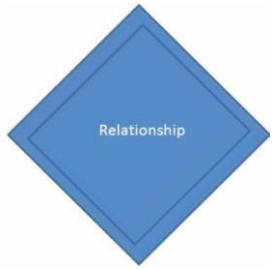
Example for derived attribute : age of employee which can be calculated from date of birth and current date. In ER modeling, notation for derived attribute is given below.

Example : An employee works for an organization. Here "works for" is a relation between the entities employee and organization

In ER modeling, notation for relationship is given below



However in ER Modeling, To connect a weak Entity with others, you should use a weak relationship notation as given below



Degree of a Relationship

Degree of a relationship is the number of entity types involved. The n-ary relationship is the general form for degree n. Special cases are unary, binary, and ternary ,where the degree is 1, 2, and 3, respectively.

Example for unary relationship : An employee ia a manager of another employee Example for binary relationship : An employee works-for department.

Example for ternary relationship : customer purchase item from a shop keeper

Cardinality of a Relationship

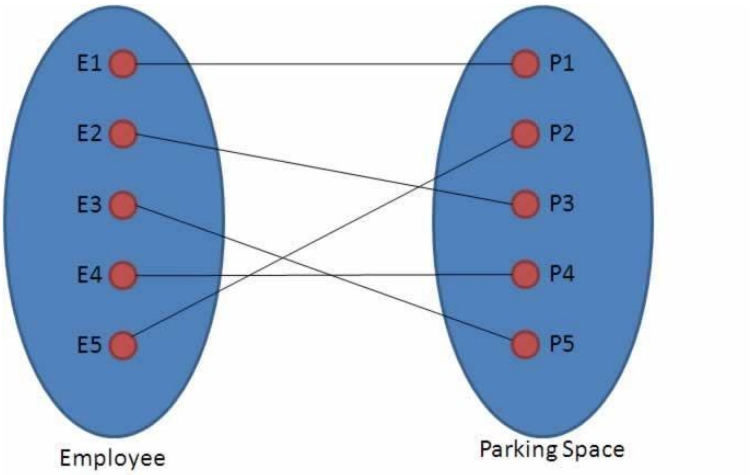
Relationship cardinalities specify how many of each entity type is allowed. Relationships can have four possible connectivities as given below.

1. One to one (1:1) relationship
2. One to many (1:N) relationship
3. Many to one (M:1) relationship
4. Many to many (M:N) relationship

The minimum and maximum values of this connectivity is called the cardinality of the relationship.

Example for Cardinality – One-to-One (1:1)

Employee is assigned with a parking space

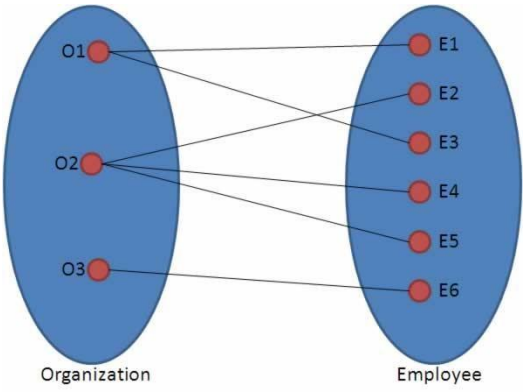


One employee is assigned with only one parking space and one parking space is assigned to only one employee. Hence it is a 1:1 relationship and cardinality is One-To-One (1:1)

In ER modeling, this can be mentioned using notations as given below

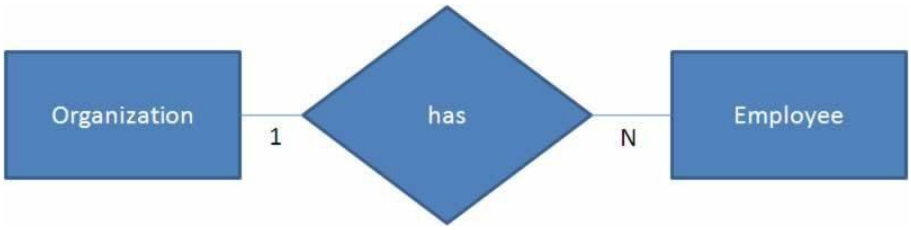
Example for Cardinality – One-to-Many (1:N)

Organization has employees



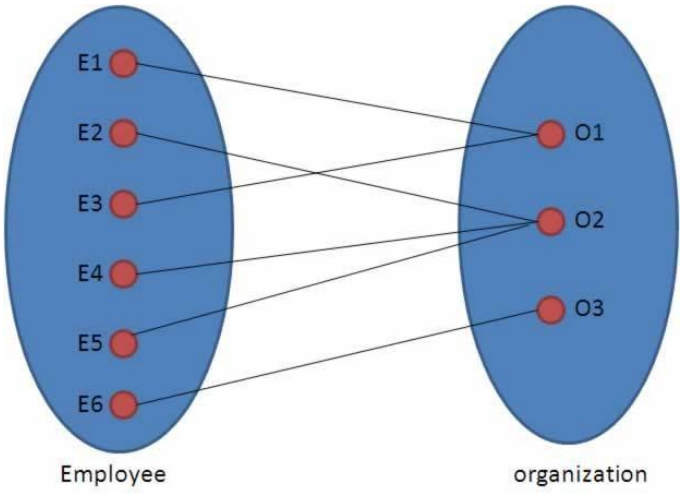
One organization can have many employees , but one employee works in only one organization. Hence it is a 1:N relationship and cardinality is One-To-Many (1:N)

In ER modeling, this can be mentioned using notations as given below



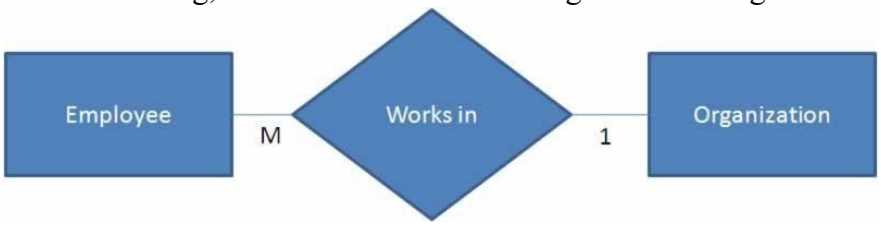
Example for Cardinality – Many-to-One (M :1)

It is the reverse of the One to Many relationship. employee works in organization



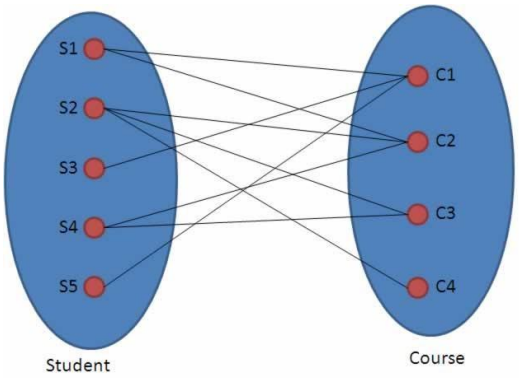
One employee works in only one organization But one organization can have many employees. Hence it is a M:1 relationship and cardinality is Many-to-One (M :1)

In ER modeling, this can be mentioned using notations as given below



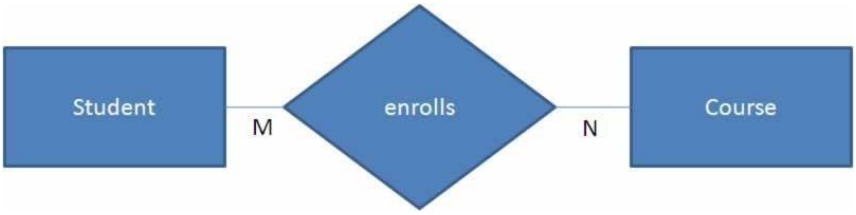
Cardinality – Many-to-Many (M:N)

Students enrolls for courses



One student can enroll for many courses and one course can be enrolled by many students. Hence it is a M:N relationship and cardinality is Many-to-Many (M:N)

In ER modeling, this can be mentioned using notations as given below



Relationship Participation

1. Total

In total participation, every entity instance will be connected through the relationship to another instance of the other participating entity types

2. Partial

Example for relationship participation

Consider the relationship - Employee is head of the department.

Here all employees will not be the head of the department. Only one employee will be the head of the department. In other words, only few instances of employee entity participate in the above relationship. So employee entity's participation is partial in the said relationship.

However each department will be headed by some employee. So department entity's participation is total in the said relationship.

Advantages and Disadvantages of ER Modeling (Merits and Demerits of ER Modeling)

Advantages

1. ER Modeling is simple and easily understandable. It is represented in business users language and it can be understood by non-technical specialist.
2. Intuitive and helps in Physical Database creation.
3. Can be generalized and specialized based on needs.
4. Can help in database design.
5. Gives a higher level description of the system.

Disadvantages

1. Physical design derived from E-R Model may have some amount of ambiguities or inconsistency.

2. Sometime diagrams may lead to misinterpretations

SUMMARY:

E-R Modeling (Entity-Relationship Modeling) is a high-level conceptual data modeling technique used in the design phase of databases. It visually represents the structure of data and the relationships among data using entities, attributes, and relationships.

- **Key Components of E-R Model:**
  - **Entity**: A real-world object or concept that can be identified distinctly (e.g., Student, Employee).
    - **Entity Set**: A collection of similar types of entities.
    - **Types**:
      - Strong Entity (has a key attribute)
      - Weak Entity (depends on another entity)
  - **Attributes**: Properties or characteristics of an entity (e.g., Student has RollNo, Name).
    - **Types**:
      - Simple (atomic)
      - Composite
      - Derived
      - Multivalued
  - **Relationship**: Association among entities (e.g., *Enrolled* between Student and Course).
    - **Degree of Relationship**:
      - Unary (1 entity)
      - Binary (2 entities) – most common
      - Ternary (3 entities)
    - **Cardinality**: One-to-One (1:1), One-to-Many (1:N), Many-to-Many (M:N)
  - **Keys**:
    - **Primary Key**: Uniquely identifies an entity.
    - **Foreign Key**: Used to link related entities.

- **ER Diagram**: A graphical representation of E-R Model using symbols:
    - Rectangle → Entity
    - Ellipse → Attribute
    - Diamond → Relationship
    - Lines → Connect entity and attributes/relationships

✅ Multiple Choice Questions (MCQs)

1. **What does E-R stand for in E-R modeling?**
   A) Entity Report
   B) Entry Relationship
   C) Entity-Relationship
   D) Extended-Relation
   → **Answer: C) Entity-Relationship**

2. **Which symbol is used to represent an entity in an ER diagram?**
   A) Ellipse
   B) Diamond
   C) Rectangle
   D) Triangle
   → **Answer: C) Rectangle**

3. **A weak entity set has which of the following?**
   A) A foreign key only
   B) No attributes
   C) No primary key of its own
   D) A composite key
   → **Answer: C) No primary key of its own**

4. **What kind of attribute can have multiple values?**
   A) Simple
   B) Derived
   C) Multivalued
   D) Composite
   → **Answer: C) Multivalued**

5. **The relationship between Student and Course in a "Registers" relationship is typically:**
   A) One-to-One
   B) Many-to-One
   C) One-to-Many
   D) Many-to-Many
   → **Answer: D) Many-to-Many**

☑ Short Answer Type Questions

1. What is an entity in an E-R model?
2. Define a weak entity and give an example.
3. What is cardinality in E-R modeling?

☑ Long Answer Type Questions

1. Explain the components of an E-R model with suitable examples and a diagram.
2. Differentiate between strong and weak entities.
3. Design an E-R diagram for a library management system

UNIT 4.3

Relational Model

The relational model is today the primary data model for commercial data processing applications. It attained its primary position because of its simplicity, which eases the job of the programmer, compared to earlier data models such as the network model or the hierarchical model. In this, we first study the fundamentals of the relational model. A substantial theory exists for relational databases.

Structure of Relational Databases

A relational database consists of a collection of tables, each of which is assigned a unique name. For example, consider the *instructor* table of Figure:4.3.1 , which stores information about instructors. The table has four column headers: *ID*, *name*, *dept name*, and *salary*. Each row of this table records information about an instructor, consisting of the instructor's *ID*, *name*, *dept name*, and *salary*. Similarly, the *course* table of Figure 4.3.2 stores information about courses, consisting of a *course id*, *title*, *dept name*, and *credits*, for each course. Note that each instructor is identified by the value of the column *ID*, while each course is identified by the value of the column *course id*.

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Figure:4.3.1 **The *instructor* relation (2.1)**

course_id	*title*	*dept_name*	*credits*
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

Figure 4.3.2 The *course* relation (2.2)

Figure 4.3.3 shows a third table, *prereq*, which stores the prerequisite courses for each course. The table has two columns, *course id* and *prereq id*. Each row consists of a pair of course identifiers such that the second course is a prerequisite for the first course.

Thus, a row in the *prereq* table indicates that two courses are *related* in the sense that one course is a prerequisite for the other. As another example, we consider the table *instructor*, a row in the table can be thought of as representing the relationship between a specified *ID* and the corresponding values for *name*,*dept name*, and *salary* values.

course_id	prereq_id
BIO-301	BIO-101
BIO-399	BIO-101
CS-190	CS-101
CS-315	CS-101
CS-319	CS-101
CS-347	CS-101
EE-181	PHY-101

Figure 4.3.3 The *prereq* relation. (2.3)

Thus, in the relational model the term **relation** is used to refer to a table, while the term **tuple** is used to refer to a row. Similarly, the term **attribute** refers to a column of a table.

Examining Figure 4.3.1 , we can see that the relation *instructor* has four attributes:

ID, *name*, *dept name*, and *salary*.

We use the term **relation instance** to refer to a specific instance of a relation, i.e., containing a specific set of rows. The instance of *instructor* shown in Figure 4.3.1 has 12 tuples, corresponding to 12 instructors.

In this topic, we shall be using a number of different relations to illustrate the various concepts underlying the relational data model. These relations represent part of a university. They do not include all the data an actual university database would contain, in order to simplify our presentation.

The order in which tuples appear in a relation is irrelevant, since a relation is a *set* of tuples. Thus, whether the tuples of a relation are listed in sorted order, as in Figure 4.3.1 , or are unsorted, as in Figure 4.3.4, does not matter; the relations in the two figures are the same, since both contain the same set of tuples. For ease of exposition, we will mostly show the relations sorted by their first attribute. For each attribute of a

relation, there is a set of permitted values, called the **domain** of that attribute. Thus, the domain of the *salary* attribute of the *instructor* relation is the set of all possible salary values, while the domain of the *name* attribute is the set of all possible instructor names.

We require that, for all relations *r*, the domains of all attributes of *r* be atomic. A domain is **atomic** if elements of the domain are considered to be indivisible units

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Figure 4.3.4 **Unsorted display of the *instructor* relation.** (2-4)

For example, suppose the table *instructor* had an attribute *phone number*, which can store a set of phone numbers corresponding to the instructor. Then the domain of *phone number* would not be atomic, since an element of the domain is a set of phone numbers, and it has subparts, namely the individual phone numbers

in the set.

The important issue is not what the domain itself is, but rather how we use domain elements in our database. Suppose now that the *phone number* attribute stores a single phone number. Even then, if we split the value from the phone number attribute into a country code, an area code and a local number, we would be treating it as a nonatomic value. If we treat each phone number as a single indivisible unit, then the attribute *phone number* would have an atomic domain.

The **null** value is a special value that signifies that the value is unknown or does not exist. For example, suppose as before that we include the attribute *phone number* in the *instructor* relation. It may be that an instructor does not have a phone number at all, or that the telephone number is unlisted. We would then have to use the null value to signify that the value is unknown or does not exist. We shall see later that null

values cause a number of difficulties when we access or update the database, and thus should be eliminated if at all possible. We shall assume null values are absent initially.

Database Schema

When we talk about a database, we must differentiate between the **database schema**, which is the logical design of the database, and the **database instance**, which is a snapshot of the data in the database at a given instant in time. The concept of a relation corresponds to the programming-language notion of a variable, while the concept of a **relation schema** corresponds to the programming-language notion of type definition.

In general, a relation schema consists of a list of attributes and their corresponding domains. The concept of a relation instance corresponds to the programming-language notion of a value of a variable. The value of a given variable may change with time

dept_name	*building*	*budget*
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

Figure 4.3.5 **The *department* relation**.(2-5)

similarly the contents of a relation instance may change with time as the relation is updated. In contrast, the schema of a relation does not generally change. Although it is important to know the difference between a relation schema and a relation instance, we often use the same name, such as *instructor*, to refer to both the schema and the instance. Where required, we explicitly refer to the schema or to the instance, for example "the *instructor* schema," or "an instance of the *instructor* relation." However, where it is clear whether we mean the schema or the instance, we simply use the relation name.

Consider the *department* relation of Figure 4.3.5. The schema for that relation is

department (*dept name*, *building*, *budget*)

Note that the attribute *dept name* appears in both the *instructor* schema and the *department* schema. This duplication is not a coincidence.

Rather, using common attributes in relation schemas is one way of relating tuples of distinct relations.

For example, suppose we wish to find the information about all the instructors who work in the Watson building. We look first at the *department* relation to find the *dept name* of all the departments housed in Watson. Then, for each such department, we look in the *instructor* relation to find the information about the instructor associated with the corresponding *dept name*.

Let us continue with our university database example. Each course in a university may be offered multiple times, across different semesters, or even within a semester. We need a relation to describe each individual offering, or section, of the class. The schema is

section (*course id*, *sec id*, *semester*, *year*, *building*, *room number*, *time slot id*)

Figure 4.3.6 shows a sample instance of the *section* relation. We need a relation to describe the association between instructors and the class sections that they teach. The relation schema to describe this association is

teachers (*ID*, *course id*, *sec id*, *semester*, *year*)

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

Figure 4.3.6 **The *section* relation.(2-6)**

Figure 4.3.7 shows a sample instance of the *teaches* relation. As you can imagine, there are many more relations maintained in a real university database. In addition to those relations we have listed already, *instructor*, *department*, *course*, *section*, *prereq*, and *teaches*,we use the following relations in this text:

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

Figure 4.3.7 The *teaches* relation. (2-7)

- *student* (*ID*, *name*, *dept name*, *tot cred*)
- *advisor* (*s id*, *i id*)
- *takes* (*ID*, *course id*, *sec id*, *semester*, *year*, *grade*)
- *classroom* (*building*, *room number*, *capacity*)
- *time slot* (*time slot id*, *day*, *start time*, *end time*)

Keys

We must have a way to specify how tuples within a given relation are distinguished. This is expressed in terms of their attributes. That is, the values of the attribute values of a tuple must be such that they can *uniquely identify* the tuple. In other words, no two tuples in a relation are allowed to have exactly the same value for all attributes.

A **superkey** is a set of one or more attributes that, taken collectively, allow us to identify uniquely a tuple in the relation. For example, the *ID* attribute of the relation *instructor* is sufficient to distinguish one instructor tuple from another. Thus, *ID* is a superkey. The *name* attribute of *instructor*, on the other hand, is not a superkey, because several instructors might have the same name. Formally, let R denote the set of attributes in the schema of relation r. If we say that a subset K of R is a *superkey* for r , we are restricting consideration to instances of relations r in which no two distinct tuples have the same values on all attributes in K. That is, if $t1$ and $t2$ are in r and $t1 = t2$, then $t1.K = t2.K$.

A superkey may contain extraneous attributes. For example, the combination of *ID* and *name* is a superkey for the relation *instructor*. If K is a superkey, then so is any superset of K. We are often interested in

superkeys for which no proper subset is a superkey. Such minimal superkeys are called **candidate keys**.

It is possible that several distinct sets of attributes could serve as a candidate key. Suppose that a combination of *name* and *dept name* is sufficient to distinguish among members of the *instructor* relation. Then, both *{ID}* and *{name*, *dept name}* are candidate keys. Although the attributes *ID* and *name* together can distinguish *instructor* tuples, their combination, *{ID*, *name}*, does not form a candidate key, since the attribute *ID* alone is a candidate key.

We shall use the term **primary key** to denote a candidate key that is chosen by the database designer as the principal means of identifying tuples within a relation. A key (whether primary, candidate, or super) is a property of the entire relation, rather than of the individual tuples. Any two individual tuples in the relation are prohibited from having the same value on the key attributes at the same time. The designation of a key represents a constraint in the real-world enterprise being modeled.

Primary keys must be chosen with care. As we noted, the name of a person is obviously not sufficient, because there may be many people with the same name. In the United States, the social-security number attribute of a person would be a candidate key. Since non-U.S. residents usually do not have social-security numbers, international enterprises must generate their own unique identifiers.

An alternative is to use some unique combination of other attributes as a key. The primary key should be chosen such that its attribute values are never, or very rarely, changed. For instance, the address field of a person should not be part of the primary key, since it is likely to change. Social-security numbers, on the other hand, are guaranteed never to change. Unique identifiers generated by enterprises generally do not change, except if two enterprises merge; in such a case the same identifier may have been issued by both enterprises, and a reallocation of identifiers may be required to make sure they are unique It is customary to list the primary key attributes of a relation schema before the other attributes; for example, the *dept name* attribute of *department* is listed first, since it is the primary key. Primary key attributes are also underlined. A relation, say $r1$, may include among its attributes the primary key of another relation, say $r2$. This attribute is called a **foreign key** from $r1$, referencing $r2$.

The relation $r1$ is also called the **referencing relation** of the foreign key dependency, and $r2$ is called the **referenced relation** of the foreign key. For example, the attribute *dept name* in *instructor* is a foreign key from *instructor*, referencing *department*, since *dept name* is the primary key of *department*. In any database instance, given any tuple, say *ta*, from the *instructor* relation, there must be some tuple, say *tb*, in the

department relation such that the value of the *dept name* attribute of *ta* is the same as the value of the primary key, *dept name*, of *tb*. Now consider the *section* and *teaches* relations. It would be reasonable to require that if a section exists for a course, it must be taught by at least one instructor; however, it could possibly be taught by more than one instructor. To enforce this constraint, we would require that if a particular (*course id*, *sec id*, *semester*, *year*) combination appears in *section*, then the same combination must appear in *teaches*. However, this set of values does not form a primary key for *teaches*, since more than one instructor may teach one such section. As a result, we cannot declare a foreign key constraint from *section* to *teaches* (although we can define a foreign key constraint in the other direction, from *teaches* to *section*).

The constraint from *section* to *teaches* is an example of a **referential integrity constraint**; a referential integrity constraint requires that the values appearing in specified attributes of any tuple in the referencing relation also appear in specified attributes of at least one tuple in the referenced relation.

SUMMARY: Relational Model

The **Relational Model** is the most widely used model in database systems. It was proposed by **E. F. Codd** in 1970 and organizes data into **tables (called relations)**, which consist of **rows (tuples)** and **columns (attributes)**.

- **Key Concepts:**
  - **Relation**: A table with rows and columns.
  - **Tuple**: A row in a table, representing a single record.
  - **Attribute**: A column in a table, representing a data field.
  - **Domain**: The set of permissible values for an attribute.
  - **Schema**: The structure of the database (table names, attributes, data types).
  - **Instance**: The actual data in a database at a given time.
- **Types of Keys:**
  - **Primary Key**: Uniquely identifies each tuple.
  - **Candidate Key**: A set of attributes that can uniquely identify a tuple.
  - **Foreign Key**: An attribute in one table that refers to the primary key in another.

- **Super Key**: A set of one or more attributes that uniquely identify a tuple.

🔹 **Integrity Constraints:**

- **Entity Integrity**: Primary key cannot be null.
- **Referential Integrity**: Foreign key must match a primary key in another table.

✅ **Multiple Choice Questions (MCQs)**

1. **Who proposed the Relational Model?**
   A) Charles Babbage
   B) E. F. Codd
   C) Edgar Allan Poe
   D) Dennis Ritchie
   → **Answer: B) E. F. Codd**

2. **In a relational database, a tuple represents:**
   A) A column
   B) A table
   C) A row
   D) A key
   → **Answer: C) A row**

3. **Which key uniquely identifies a record in a relation?**
   A) Super Key
   B) Foreign Key
   C) Candidate Key
   D) Primary Key
   → **Answer: D) Primary Key**

4. **What is a relation in the relational model?**
   A) A mathematical equation
   B) A table
   C) A graph
   D) A data type
   → **Answer: B) A table**

5. **Which constraint ensures that a foreign key must match an existing primary key?**
   A) Domain constraint
   B) Entity integrity
   C) Referential integrity
   D) Tuple constraint
   → **Answer: C) Referential integrity**

Short Answer Type Questions

1. What is a relational model in DBMS?
2. Define primary key and foreign key.
3. What is a domain in the relational model?

Long Answer Type Questions

1. Explain the key components of the relational model with examples.
2. Describe different types of keys used in the relational model.
3. Discuss the integrity constraints in the relational model. Why are they important?

UNIT 4.4

Introduction to computer graphics, color model, graphic file format

Computer graphics is a thrilling, living field, combining artistic expression with technological innovation, ultimately changing the waywe see, create, and interact with visual information in the digital domain. Computer graphics, at a rudimentary level, is the field that pertains to the generation, manipulation and visualization of visual information through computational techniques applied to visual phenomena covering a wide spectrum of domains from entertainment and design through scientific visualization and industrial modeling. In the early days of computing, computer graphics were relatively primitive, with limited display technologies and computational power restricting visual representations to basic line drawings

and Geometric Shapes. Well, this may have seemed ground-breaking back then, but as computer technology advanced exponentially, graphics changed dramatically. The range of technologies in modern computer graphics now includes such things as advanced 3D rendering, photorealistic simulations, real-time interactive environments, and complex visual effects that were previously thought to be impossible. Geometric modeling, namely defining the shape of objects; rendering converting 3D objects to a 2D representation; animation - creating the illusion of movement; visual representation - defining how light interacts in a scene. When we talk about geometric modeling we answer how we describe mathematically an object and a surface, how we describe it and what we expect from shape, structure, and space relationship. It uses a number of mathematical methods, including polygon meshes, vector graphics, and parametric surfaces to model real or fictitious objects in a digital format. Rendering is a key component of computer graphics that converts mathematical models into images that can be viewed. In this, detailed mathematics simulate the behavior of light when hitting surfaces, including reflection, refraction, shadows, and the characteristics of materials. As more advanced rendering techniques like raytracing, radiosity, and global illumination results in more realistic visual representations translating to the behaviour of the physical world itself. In computer graphics, the color models are crucial for how color is represented and manipulated digitally. These models underpin the specification, representation, and transformation of color within a digital context. Common color models are RGB (Red, Green, Blue), CMYK (Cyan, Magenta, Yellow, Key/Black), HSL (Hue, Saturation, Lightness) and HSV (Hue, Saturation, Value) that can be chosen according to computation as well as visual utility. The RGB color model is by far the most common model used for digital displays and computer graphics. Originally created to align with the human

visual processing system, it uses the principle of additive colors to recreate colors by mixing various intensities of red, green, and blue light. In an 8-bit color system, each of the three color channels ranges from 0 to 255 (representing 256 distinct colors), resulting in about 16.7 million possible color combinations. This model of colors is especially useful for creating representations of objects on electronic displays such as monitors, televisions, or mobile devices. On the other hand, CMYK color model is subtractive color model used for printing. CMYK, on the other hand, is subtractive color, adding colored inks to absorb certain wavelengths of light. Cyan, magenta, yellow, and black inks are mixed in different proportions; the combination yields a large variety of colors. This is a more accurate simulation of how color printing works in the real world, and provides accurate color reproduction in physical prints. The HSL and HSV color models provide another way to represent color, based on human-based color properties. These models represent colors with a triplet of values: hue (the color itself), saturation (the intensity of color), and either lightness or value (the brightness). These models are particularly useful for user interfaces in graphic design, image editing, and color selection as they are a more human perceptual model of colors and how people think and talk about colors. Another critical computer graphics concept of color depth, the number of bits used by each pixel to represent color information. Typical color depths are 8-bit (256 colors), 16-bit (65,536 colors), 24-bit (16.7 million colors), and 32-bit (4.3 billion colors). By allowing more subtle gradations in color, high color depths are able to represent them in a manner that is more true to life, leading to more accurate visual representations while minimizing quantization error visible to the human eye commonly recognized as the banding effect.

This field also includes important graphic file formats, which are standardized ways to store visual information for easy transmission between computer systems. These formats determine the encoding, compression, and preservation of image data, optimizing aspects like image quality, file size, and compatibility across various systems and applications. There are too many of them, and they exist because of the myriad needs of other use cases and ecological contexts. Raster file formats are one major type of graphic formats, storing the image as a grid or pixel arrays.

Some of the popular raster formats are JPEG, PNG, GIF and TIFF. JPEG (Joint Photographic Experts Group) is a widely used lossy compression method, well-suited for photographs and complex images where a certain level of quality degradation is acceptable. PNG (Portable Network Graphics) PNG provides lossless compression, supports transparency, and is suitable for web graphics and images with clear edges and text. Vector graphics formats offer an alternative way

to create images using mathematical equations and geometric primitives rather than just grids of pixels. SVG (Scalable Vector Graphics), AI (Adobe Illustrator), and EPS (Encapsulated PostScript) are examples of formats that define images using lines, curves, and shapes instead of pixels. It allows for infinite scaling without quality loss, meaning vector graphics work great for logos, illustrations, and designs that require adjusting the scale. There is also a sub- category for file types you have to give extra attention to mainly due to their prevalence and advanced techniques of compressions (JPEG). JPEG uses discrete cosine transform (DCT) algorithms to compress image data, which was developed in the early 1990s, to significantly reduce the file size while retaining acceptable visual quality. It also has adjustable compression settings that allow users to customize the balance between file size and image quality based on their needs. PNG is a more advanced raster format that can overcome limitations in older formats such as GIF. This versatile format handles full-color, grayscale, and indexed-color images of multiple bit depths.

PNG's lossless compression and alpha channel support allow for such beautiful transparency effects, which is a big advantage. PNG is a preferred format for web designers and digital artists for graphics that require initial details and transparency. SVG, which stands for Scalable Vector Graphics, is a new vector format, based on XML - markup used to describe 2D vector graphics. As an open web standard, SVG allows for resolution independent scaling, making it perfect for responsive web design and dynamic graphic applications. SVG graphics can be embedded directly in HTML, manipulated with CSS and JavaScript, and can be richly interactive in ways that just aren't possible with static raster formats. This brings us to the most rudimentary of file types, the bitmap (BMP) format. BMP files are not compressed, providing high image quality, but also typically use more storage space than compressed formats. The straightforwardness inherent in this format makes it practical for certain cases, such as pixel-level access in a focusing application or compatibility with older systems that demand it.

The TIFF (Tagged Image File Format) is a flexible format that supports all the compression processes and color spaces. TIFF is commonly used in professional photography, publishing, and scientific imaging, allowing for high-quality images with little loss of quality. It provides the versatility of lossy and lossless compression for various imaging requirements. Graphics Interchange Format (GIF) over the years have made a one of a kind name on the web for know images especially animated and low-color-depth images. GIF, limited to 256 color colors, is a fil with lossless compression, allowing frame storage to create simple animations. GIFDespite having limitations due to technology,

GIF is still popular for creating short, looping animations, or memes, that can be shared across various digital platforms. Professional photographers and imaging specialists make use of unprocessed raw image formats that capture the sensor data coming directly from digital cameras.

These formats maintain the greatest amount of image information, making them extremely flexible for post-processing. Photographers had unprecedented control over how their images were developed and how the colors were manipulated, despite occupying a comparatively large amount of storage space. Revolutionary techniques such as ray tracing, which so closely mimics how light interacts with the world around us that it's like bringing reality to life on the screen, have become a crucial part of modern computer graphics. Ray tracing simulates the unique interactions between light rays and scene objects, leading to incredibly realistic reflections, shadows, and surfaces. While early ray tracing was too computationally intensive to be practical for most applications, recent advances in hardware have made it increasingly feasible, bringing real-time ray tracing to video games and interactive applications. Computer graphics intersects with artificial intelligence, openingthe door to procedural generation, automated image enhancement, and intelligent content creation. Today, we have machine learning algorithms that create photoreal images — change backgrounds effortlessly, upscale graphics from lower resolution to higher one, or even generate entirely synthetic images impossible for human eyes.

The ultimate frontier of computer graphics: Virtual and augmented reality technologies push the boundaries of what computer graphics can achieve, creating immersive environments that bridge the gap between digital and physical realities. These technologies use advanced graphics rendering techniques, low-latency processing, and sophisticated color and depth perception to create realistic spatial experiences. To solve this problem, color management systems were developed that allow the color of a given medium to be controlled. These systems create standardized color profiles and translation mechanisms from digital to print, typically striving for colors to appear as consistently as possible alongside varying types of display technology. The future of computer graphics looks to be even more incredible. As quantum computing, advanced neural networks and increasingly sophisticated rendering techniques take hold, they will fundamentally transform the ways we create, perceive and interact with visual information. The evolution of the field has led to the development of everything from hyper-realistic simulations to wholly original forms of artistic expression — and computer graphics have thus increasingly pushed the boundaries of human visual communication. Computer graphics is an area where

mathematics, physics, perceptual psychology, and artistic design intersect. With the development of technology the field will inarguably progress allowing even more powerful means of creativity, communication and computational visual representation.

SUMMARY

- **Introduction to Computer Graphics**

Computer graphics is the branch of computer science that deals with the creation, manipulation, and representation of visual images using computers.

Applications:

- Video games
- Simulations
- Animation and movies
- Graphic design
- CAD (Computer-Aided Design)

Types:

- **2D Graphics** – Images represented using 2D geometry (lines, shapes).
- **3D Graphics** – Use 3D models, rendering, and projection techniques.
- **Interactive Graphics** – User interaction (e.g., GUI, games).

- **Color Models**

A **color model** is a mathematical model describing the way colors can be represented as tuples of numbers.

Common Color Models:

1. **RGB (Red, Green, Blue)**
    - Used in digital screens
    - Additive model
    - Example: (255, 0, 0) = Red
2. **CMYK (Cyan, Magenta, Yellow, Black)**
    - Used in color printing

- Subtractive model

3. **HSV/HSB (Hue, Saturation, Value/Brightness)**
   - Based on human perception
   - Common in graphic design tools
4. **YUV / YCbCr**
   - Used in video and broadcasting systems

🔹 Graphic File Formats

Graphics can be stored using various file formats, which differ based on compression, color depth, transparency, and usage.

Popular File Formats:

- **JPEG (.jpg)**
  - Compressed (lossy)
  - Best for photos
  - No transparency
- **PNG (.png)**
  - Compressed (lossless)
  - Supports transparency
  - Suitable for web graphics
- **GIF (.gif)**
  - Limited to 256 colors
  - Supports animation
  - Lossless
- **BMP (.bmp)**
  - Uncompressed
  - Large file size
  - High quality
- **SVG (.svg)**
  - Vector format
  - Scalable without quality loss

- Used for web and icons

✅ Multiple Choice Questions (MCQs)

1. **Which of the following is an additive color model?**
   A) CMYK
   B) RGB
   C) HSV
   D) YCbCr
   → **Answer: B) RGB**

2. **Which image format supports animation?**
   A) JPEG
   B) BMP
   C) PNG
   D) GIF
   → **Answer: D) GIF**

3. **Which color model is commonly used in printing?**
   A) RGB
   B) HSV
   C) CMYK
   D) YUV
   → **Answer: C) CMYK**

4. **What does JPEG stand for?**
   A) Java Picture Expert Group
   B) Joint Photographic Experts Group
   C) Joint Photo Enhancement Graphics
   D) Java Pixel Encoding Group
   → **Answer: B) Joint Photographic Experts Group**

5. **Which file format is vector-based?**
   A) JPEG
   B) BMP
   C) SVG
   D) GIF
   → **Answer: C) SVG**

✅ Short Answer Type Questions

1. What is the RGB color model used for?
2. Name two file formats that support transparency.
3. Differentiate between raster and vector graphics.

✅ Long Answer Type Questions

1. Explain the different color models used in computer graphics.
2. Describe common graphic file formats and their uses.
3. What is computer graphics? Explain its types and applications.

References

- Elmasri, R., & Navathe, S.B. (2017). *"Fundamentals of Database Systems"* (7th ed.). Pearson Education, Chapter 1, pp. 2-35.
- Connolly, T., & Begg, C. (2015). *"Database Systems: A Practical Approach to Design, Implementation, and Management"* (6th ed.). Pearson Education, [1] Chapter 1, pp. 3-30
- Foley, J.D., Van Dam, A., Feiner, S.K., & Hughes, J.F. (1990). *"Computer Graphics: Principles and Practice"* (2nd ed.). Addison-Wesley, Chapter 1 & relevant sections on color models and file formats.
- Hearn, D., Baker, M.P., & Carithers, W. (2014). *"Computer Graphics with OpenGL"* (4th ed.). Pearson Education, relevant introductory chapters on graphics systems, color, and image formats

MODULE 5

USE OF COMPUTERS IN BIOLOGICAL SCIENCES

Objective:

To explore the applications of computers in the field of biological sciences, with a focus on bioinformatics and its databases.

UNIT 5.1

Use of computer in biological science

The computerization of biological science is among the most transforming technological advances of the late 20th and early 21st centuries. This unprecedented intersection has transformed how biological research is performed, interpreted and understood, birthing new fields and opening up new vistas of scientific inquiry. In molecular biology, ecological studies, and manyother fields, computers have become an integral tools that allow researchers to analyze massive datasets, model complex biological processes, and make breakthroughs that would not have been possible only a few decades previously.

Computers have had the greatest impact on biological science in the field of genomics and molecular biology. Around the turn of the century, as high-throughput sequencing technologies became widely available, many researchers suddenly had the ability to produce vast quantities of genetic data in record time. These large amounts of data, containing billions of DNA base pairs, are managed, stored and performed with computer systems. Bioinformatics, a relatively new field which combines biology, computer science, and information technology, has become instrumental in decoding the vast amount of information contained within genetic material. The advanced algorithms and numerical methods used employ highly complex computational techniques that enable researchers to compare and contrast genetic sequence data, to model proteins, and predict the impact of genetic variation and complex gene expression inall its manifestations. By simulating complex biological processes, computational modeling has fundamentally transformed our understanding of biological systems. Such computational models can capture everything from molecular interactions inside a single cell to the ecosystem-level dynamics of entire biological communities. To date, researchers can simulate genetic mutations, predict protein folding, model cellular metabolic networks, and explore potential biological interactions, all with unprecedented accuracy. These computational methods enable scientists to test hypotheses and explore scenarios that would be cost-prohibitive, impractical, or ethically challenging to investigate with traditional experimental techniques. Computers are powerful instruments in molecular dynamics, helping to model the complex motions and behavior of biological molecules. Powerful molecular simulation software enables researchers to monitor the action of proteins, enzymes, and other biological macromolecules at atomic resolution. Such simulations can show how genetic mutations can alter protein structure, how drugs interact with their cellular targets, and how intricate bimolecular processes unfold at the atomic or molecular scales. Today, between researchers, 3D visualizations of

molecular interactions can be generated in ways that simplywere not possible in the recent past. Computational technologies have also dramatically transformed ecological and environmental biology. Through Geographic Information Systems (GIS) and sophisticated computational modeling, scientists can track and predict complex ecological patterns, calculate species distributions, and determine environmental changes. With unprecedented accuracy, researchers can now analyze satellite imagery, track biodiversity, model climate change impacts, and predict which ecosystems will respond in certain ways. Particularly relevant to global environmental challenges, which include loss of biodiversity, climate change and conservation of ecosystems, these computational tools are necessary to understand the problem and find a solution. One such area in which computational prowess and biological inquiry intersect in particularly alluring ways is systems biology. Bioinformatics systems biology bioinformatics is treating biological systems as complex networks of starred interconnected components, and using computational approaches to analyze and model these interrelated processes (Bryant et al. Systems biology integrates information at molecular, cellular, and organism scales, leading to a systems view of biological systems complexity. Due to this, modeling biological systems is now possible using powerful computer algorithms capable of connecting disparate datasets, normalizing and/or filtering them, detecting emergent properties, and integrating them into a unified model describing their dynamic and interconnected manner.

Computers play an integral part in medical research and pharmacology, drug discovery, personalised medicine, and understanding disease pathways. High- performance computing powers screening of massive libraries of chemical compounds to predict their potential therapeutic effects, and to simulate their interactions with biological targets. An analysis of medical imaging data along with the identification of disease patterns and predictions of individual patient outcomes based on genetic and clinical information are possible with the help of machine learning and artificial intelligence algorithms. These computational techniques are fast-tracking the drug discovery process, particularly in the context of targeted therapies and personalized medicine. One of the most transformative aspects of evolutionary biology has been the computational analysis.

Phylogenetic reconstruction, the process of tracing all biological relationships, is now dominated by complex computations. Researchers can explore genetic data obtained from various species, mount complex evolutionary trees, and investigate the genetic mechanisms involved in species diversification. Toward this end, both have been applied, in concert with machine learning and other computational methods, to make revolutionary advances in our understanding of the history of life

on our planet, uncovering the complex webs of genetic inheritance and evolutionary adaptation. Computational methods have already transformed experimental design and data management in biology. Laboratory information management systems (LIMS) assist researchers in keeping track of experimental protocols, managing collections of samples, and ensuring data integrity. Statistical analysis allows scientists to interpret experimental results with precision, determining their statistical significance and comparison to possible biases. These two generations of computational tools enable biological research processes to be more efficient, reproducible, and rigorous. The advent of big data technology has also expanded the scope of computational capacities in biological science. Cloud computing resources and distributed computing networks enable researchers to manipulate and interpret massive biological datasets that could never have been handled with traditional computing infrastructure. Such technologies allow research collaborations to take place between different institutions as well as between sites separated by geographical locations, accelerating the sharing of computational resources as well as scientific knowledge.

As ML and AI enter the biological research space in these new and different way, it is most certainly going to enable new approaches to data analysis and research methods. Many of these computational methods are capable of recognizing complex patterns in biological data, making predictions about biological phenomena, and generating hypotheses. AI algorithms can predict functions of genes, recognize potentially disease-related genetic variants, and suggest possible therapeutic interventions in fields like genomics. In particular, computational approaches have been pivotal for the understanding of complex biological networks. Network biology involves the use of computational techniques to understand the complex interactions among various biological entities, including genes, proteins, and metabolic pathways. Network Based Analysis–These approaches give us insight into how biological systems are organized, how they respond to perturbations, and how different components interact to maintain cellular homeostasis.

Synthetic biology is yet another exciting frontier where computational technologies are becoming critical. Such advanced form-finding tools enable researchers to algorithmically design biological systems with precision never before possible. With these tools, scientists can engineer synthetic genetic circuits, build organism systems, and create new biotechnological platforms for applications including but not limited to drug development and bioremediation. It has also changed our insights into neurobiology and the brain functioning. The novel combination of advanced neuroimaging techniques and powerful

computational analysis offers new tools for scientists to map brain activity, gain greater insight into neural networks, and explore the complex mechanisms of cognition and neurological disorders. Computational models of neural systems are able to model brain function, predict responses to neuromodulation and provide insight into the rules of information processing in neural systems. Computational Methods for Genetic Diversity and Population Dynamics in Population Genetics and Conservation Biology.

Computational models can also mimic population growth, forecast genetic differences, and inform conservation strategies for threatened organisms. Such techniques are vital to tackle global biodiversity challenges and formulate sustainable conservation measures. Computer technologies have also changed biological education and research training. New forms of learning by using interactive computational tools, simulation software, and online platforms fundamentally transformed the way biological sciences are taught and learned. Three new tools that dramatically improve the way students interact with core biological concepts by way of interactive visualizations, computational modeling exercises and virtual laboratory experiences.

Modern biology was revolutionized with the arrival of computational technologies, which opened up new lines of research and led to interdisciplinary collaboration. Biologists can now collaborate with computer scientists, mathematicians, physicists and engineers to create new ways to conduct research. So much creativity springs from collaborative efforts, leading to breakthroughs and the founding of new scientific disciplines that combine established fields of study. Yes, the growing use of computational technologies in biological science is not without its own challenges. Researchers need to overcome challenges surrounding data privacy, limited computational resources, and any biases that exist in computational models. Reliability, reproducibility and the ethical application of computational methods continue to be an area of concern in the scientific community. Despite these challenges, computer science enters biological science as one of the most important scientific developments of the day. “Computational technologies have changed biological research from a mainly observational and experimental science into a data-centric, predictive and deeply integrative one. With new computational power, the multifaceted complexity and wonder of biological systems will enable an even greater depth of exploration. We are only just beginning to scratch the surface of how computers and biological science can work together to do extraordinary things, and in the coming years, this relationship will continue to grow and flourish.

Computational power is underlying some of the hardest and most impactful problems in science, from untangling the puzzles of how

genes are passed down from parent to child to deciphering complex interactions among ecological networks, driving science forward for a deeper and deeper understanding of the inner workings of life. With this eyes to the future, you can see that the line between biological aspect of research and computational aspect of research is getting blur, leading to a completely different era of science exploration and understanding.

SUMMARY: Use of Computers in Biological Science

Computers play a vital role in modern biological sciences, enabling faster analysis, accurate data storage, simulation, and modeling of complex biological systems.

- **Key Applications:**

1. **Bioinformatics**
    - Use of computers to collect, analyze, and interpret biological data (e.g., DNA sequences, protein structures).
    - Tools: BLAST, GenBank, FASTA
2. **Genomics & Proteomics**
    - Genome sequencing and mapping
    - Protein structure prediction and analysis
3. **Medical Imaging**
    - CT scans, MRIs, and ultrasounds analyzed using computer software
4. **Drug Discovery**
    - Molecular modeling and simulations
    - Virtual screening and docking using computer algorithms
5. **Data Analysis & Statistics**
    - Used in research to analyze large biological datasets (e.g., gene expression)
6. **Simulation & Modeling**
    - Simulating biological processes like cell division, protein folding, etc.
7. **Environmental and Agricultural Applications**
    - Studying ecosystems, crop modeling, genetic engineering

✅ Multiple Choice Questions (MCQs)

1. **Which of the following fields combines biology with computer science?**
   A) Biophysics
   B) Biotechnology
   C) Bioinformatics
   D) Biochemistry
   → **Answer: C) Bioinformatics**

2. **Which tool is commonly used for comparing DNA sequences?**
   A) Photoshop
   B) BLAST
   C) Excel
   D) Oracle
   → **Answer: B) BLAST**

3. **In biological science, computer simulations are often used to:**
   A) Play video games
   B) Grow bacteria
   C) Model biological processes
   D) Create hardware
   → **Answer: C) Model biological processes**

4. **Which application area uses computers to design and test new drugs?**
   A) Ecology
   B) Pharmacology
   C) Evolution
   D) Botany
   → **Answer: B) Pharmacology**

5. **GenBank is an example of a:**
   A) Hardware component
   B) Biological database
   C) File format
   D) Programming language
   → **Answer: B) Biological database**

✅ Short Answer Type Questions

1. What is bioinformatics?

2. Name two computer applications used in biological sciences.

3. How do computers assist in drug discovery?

Long Answer Type Questions

1. Describe the role of computers in bioinformatics and genomic research.
2. Explain the various applications of computers in biological sciences.
3. How has the use of computers improved biological research? Give examples.

Unit 5.2

Importance of Bioinformatics

The most integral part of modern biological sciences is bioinformatics, which is an interdisciplinary field that integrates biology, computer science, mathematics and statistics. This game-changing expertise has transformed our approach to biological data analysis, as it equips researchers with the tools to extract insightful knowledge from the enormous volumes of biological information produced by new generation experience platforms. Given that how processing, analyzing, and interpreting complex biological data has become increasingly important; the relevance of bioinformatics in our world today cannot be overemphasized especially in a scientific world where so much data is generated. The fields of bioinformatics emerged in the light of exponential growth of biological data primarily after the advent of high-throughput sequencing technologies and the completion of the Human Genome Project. With the advances in this area, researchers can now mine and reach conclusions from real biological data sets that were previously impenetrable. Bioinformatics changed how we study biological systems and enable a meaningful and systematic approach to biological questions that were once impossible to address. Bioinformatics is much more than a technical tool. It has transformed the landscape of biology, from an experiment-driven field to one that is now equally driven by experimental and computational approaches. This change not only hastens the speed at which scientific discovery takes place, it has also broadened the set of questions that can be examined. Bioinformatics has revolutionized the field of biological sciences, from decoding genomic sequences to predicting protein structures and functions.

Additionally, bioinformatics has been pivotal in democratizing biological research. The creation of user-friendly tools and databases has enabled researchers who are not computationally inclined to perform sophisticated analyses. This democratization has led to increased collaboration across fields and helped to diversify the contributors to scientific progress. One exciting feature of many bioinformatics resources is their open-access nature, which allows for capacity globalization across teams of scientists and promotes the exchange of knowledge at a pace never seen before. Bioinformatics has played a key role in translating biological knowledge into clinical applications in the fields of healthcare and medicine. This has paved the way for personalized medicine, where therapeutic approaches are tailored to the genetic makeup of the individual. Bioinformatics has led to better diagnostic approaches, targeted therapies, and preventive techniques, as it studies types of genetic variants related to diseases. That knowledge has been particularly revolutionary in the fields of

cancer research, rare genetic disorders, and infectious disease, where insights at the molecular level have driven incredible progress for patients.

Bioinformatics applications have added significantly to the agricultural sector. Researchers have identified genetic markers for desirable traits, such as disease resistance, yield potential, and nutritional quality, by analyzing plant and animal genomes. Breeding done with this field and laboratory knowledge in mind has allowed to the development of high yielding varieties of wheat and maize and quadrupling livestock yield to securing food in a changing climate. With this in mind, bioinformatics has emerged as a vital resource in the cause for sustainable agriculture and global food security. Bioinformatics methods have contributed to environmental management. Metagenomic analyses in the field allowed researchers to characterize microbial communities across diverse ecosystems without the need for classical culturing techniques. This has informed us important aspects of ecosystem functioning and resilience, and has also contributed to new generation of conservation and environmental management strategies. Bioinformatics has also played a role in monitoring and mitigating threats from biological systems due to environmental changes, including the impact of climate change on species and adaptations. Bioinformatics has industrial applications in numerous sectors, including but not customarily reduced to biotechnology, pharmaceuticals in addition to biological processes for data development and laboratory functions. Various bioinformatics tools contribute to the creation of sustainable bioprocesses and new bio-based products via enzyme engineering, metabolic pathway optimization, and synthetic biology approaches. This impacts many areas, such as biofuels, biomaterials, and biochemicals, playing a role in the shift towards a more sustainable and bio-based economy.

With an eye on the future, the Get More use of Bioinformatics as it is expected to grow manifold. The continuous evolution of high-throughput technologies has made possible the production of biological data at an impressive scale and speed, which in turn require and increasingly complex computational methods for their analysis and interpretation. New areas like single-cell genomics, spatial transcriptomics and multi-omics integration offer new opportunities but also challenges for bioinformatics. Machine learning methods applied to bioinformatics will also very likely bring new tools to the biologist's toolkit, and the idea of biological patterns becoming novel news will likely become regular part of the field as new sequencing data comes out with some regularity. Overturning bioinformatics has become an undeniable part of contemporary biological study and uses. It enables everything from fundamental scientific discovery to real-

world applications in healthcare, agriculture, environmental stewardship, and industrial biotechnology. However, it is important to remember that, as biological data continues to expand through institutional growth and greater sophistication, bioinformatics will be increasingly critical in applying that data toward solving some of the greatest challenges facing society today. Bioinformatics tool and methodology development and refinement have been rapid and will continue apace, with increasing accessibility and integration with other technology advances to enable innovation and progress in the biological sciences and beyond.

Components of Bioinformatics

A multidisciplinary field, bioinformatics consists of several important components that together allow for the storage, retrieval, analysis, and interpretation of biological data. The knowledge of these building blocks forms the basis for the applicability of bioinformatics in various fronts by providing the necessary paradigms and methodologies to solve complex questions in biology. These fundamental aspects are key to how bioinformatics combines insights into biological processes with tools from computing to propel the progress of scientific knowledge. One of the vey fundamental elements of bioinformatics are biological databases. These databases collect, arrange and provide access to the large amounts of biological data for research across the globe. GenBank (Benson et al. 2005)), UniProt (Higgins et al. 2015), and the Protein Data Bank (PDB) (Berman et al. 2000) are examples of primary databases that archive raw experimental data such as nucleotide sequences, protein sequences, and molecular structures. These primary data used in secondary databases (e.g. Pfam and KEGG) add value by providing annotations, functional classifications and pathways. For example, systems such as Entrez and SRS allow researchers to navigate these integrated databases seamlessly across different data types, supporting comprehensive analyses of biological data. These databases must be developed and maintained by increasingly complex data management systems that are capable of managing the exponential growth of biological information while preserving data quality, consistency, and accessibility.

Another core aspect of bioinformatics includes the use of sequence analysis, referring to a collection of techniques and procedures to analyze and interpret DNA, RNA, and protein sequences. Pairwise sequence alignment algorithms, including the Needleman-Wunsch and Smith-Waterman algorithms, align sequences to find similarities that may indicate evolutionary relationships or the conservation of function. Instead, multiple sequence alignment generalises this approach to compare many sequences at once, revealing conserved motifs and domains where evolutionary pressure against change often

indicates functional importance. Turn the long format to a wide format | with the help of tools like BLAST and FASTA, there are many sequence databases available to researchers, For this, how can the sequences be found in a long format? Methods that are used to analyze sequence alignments are very important for gene prediction, protein function assignment, and evolutionary studies, and have played a major role in our comprehension of biological systems at the organelle and molecular levels.

Structural bioinformatics is specifically the area dealing with three-dimensional structure of biological macromolecules such as proteins and nucleic acids. This element harnesses computational techniques to anticipate, model and examine molecular architectures — yielding principles about their physical characteristics and biological purposes. Finally, homology modeling relies on the observation that proteins with similar sequence usually have a similar structure, in order to predict the structure for a protein having related protein structures. In contrast, ab initio methods try to predict structures from first principles, taking into account the physicochemical properties of amino acids and nucleotides. Adding to these methods, molecular dynamics simulations allow for modeling how molecules change over time, capturing conformational changes that are key to function. By combining structural information with sequence and functional data, this approach has shed light on previously poorly understood biological phenomena such as protein-protein interactions, enzyme mechanisms, and drug-target binding, providing insights for rational drug design and protein engineering. Bioinformatics tools and methods have revolutionized genomics, study of an organisms complete set of genes. This method involves using sequence data generated by sequencing technologies that is highly fragmented and genome assembly algorithms that can piece together these fragments and generate more complete genomes. Annotation pipelines subsequently screen and categorize genetic elements in these assembled sequences, such as genes, regulatory elements, and repetitive elements. Comparative genomics techniques focus on comparing the genomes of multiple species to identify similarities and differences among them, providing insights into evolutionary relationships, gene conservation, and species-specific adaptations. Population genomics extends this work to the genetic variation present within species, indentifying polymorphisms associated with a trait of interest. The combination of genomic information with systems biological approaches that integrate other sources of biological data leads to more integrated understanding of biological systems and their responses to environmental perturbations.

Transcriptomics, the study of all RNA transcripts produced by the genome, is also an important field within bioinformatics. High-throughput sequencing data is transformed into genomic information through RNA-seq analysis pipelines, which can be used to quantify gene expression levels, detect alternative splicing events and identify novel transcripts. Differential expression analysis is used to identify genes in which the expression level is significantly altered in different conditions and can reveal the underlying mechanisms of diseases, developmental processes and responses to environmental insults. Single-cell transcriptomics takes this a step further by profiling gene expression in individual cells and uncovering cellular heterogeneity in tissues and developmental trajectories. Recent developments in transcriptomics have included the integration of transcriptomic data with genomics and proteomics data, providing a more comprehensive view of gene regulation and cellular function. Bioinformatics plays a glove-in-hand role in analyzing and interpreting data in proteomics, the large-scale study of proteins. Protein identification algorithms compare mass spectrometry data against the sequence of known proteins contained in public databases (such as Uniprot or REFSEQ) to identify proteins present in the sample. Colonised SCFA-treated organoids generated unique quantitative proteome data sets, complementing the transcriptomic analysis and providing extensive coverage of the protein response to SCFA treatment across both conditions. Functional associations between proteins are identified using protein-protein interaction networks generated from experimental and computational approaches, helping to better inform cellular pathways and complexes. Analysis of post-translational modifications reveals chemical modifications that fine-tune protein function, providing an additional layer of complexity to the regulatory mechanisms for protein activity which is beyond the scope of sequence data alone. The proteomic methods aided by bioinformatics have played a critical part in the uncovering of biomarkers, identifying drug targets, and studying disease processes at the protein level.

Systems biology: Systems biology is the integrative branch of bioinformatics which studies biological systems as a whole rather than individual pieces. This integrative strategy links multiple omics data types to build complete models for cellular networks and pathways. Network analysis tools pinpoint modules of interconnected genes or proteins that frequently reflect functional units within the cell. Similarly, pathway enrichment methods identify which biological pathways are overrepresented in a set of differentially expressed genes or proteins, giving insights into the biological processes affected under particular conditions. Metabolic network behavior can be predicted by using flux balance analysis and numerical simulations to find optimal system configurations, extrapolating how changing one part might

influence the entire process. From models of cellular differentiation to disease progression, these system-level analyses provide a more holistic perspective on biological processes − better capturing the interactions between different molecular components rather than focusing on individual components alone. Given the complexity of biological data, machine learning and artificial intelligence have increasingly served as tools within bioinformatics, providing effective techniques for pattern classification and predictive analysis. Supervised learning algorithms trained on labeled datasets can learn to predict several aspects, such as the secondary structure of a protein, the function of a gene, or the susceptibility to a disease based on sequence or structural features. Unsupervised learning algorithms find groupings of data, leading to new classifications of diseases or cellular phenotypes. Particularly, convolutional neural networks and recurrent neural networks have achieved outstanding performance in applications from protein structure prediction, image analysis in the biomedical domain, to complex pattern recognition in multi-omics data. Supervised machine learning models form the basis of predictive analysis for various biomolecular profiles and functional domains, where biological knowledge and advanced computational techniques are the secrets to achieve best results at these brackets.

Statistical methods underpin bioinformatics analyses that are rigorous approaches for hypothesis testing and inference in biological data. Correction for multiple testing adjusting procedures is fundamental challenge when testing several thousands of hypotheses simultaneously, this is common in genomics and proteomics studies. Bayesian methods inherently include prior knowledge into the analysis (which is useful for biological analysis to aid interpretation of new data with existing knowledge). These low-dimensional representations enhance the interpretation of high-dimensional omics data. Employing this statistical groundwork guarantees that any deductions inferred from bioinformatics assessments are resilient and dependable–an essential consideration considering the far-reaching consequences for medical, agricultural, and environmental applications. These components of bioinformatics — biological databases, sequence analysis, structural bioinformatics, genomics, transcriptomics, proteomics, systems biology, machine learning and statistical methods — form the core of a toolbox for probing the molecular mechanisms of life. The emergence of these components is, however, not a zero-sum game, with success in one field driving success across the board. These many diverse components together have not only advanced our fundamental knowledge of biological systems, they have had real world applications in many fields such as medicine, agriculture, and more. These will be necessary for generating insights and driving forward

scientific innovation as the technologies in this domain evolve and biological data becomes rapidly more complex.

Bioinformatics Applications

Bioinformatics has a wide range of applications in various fields of science and practice. These applications utilize the computational tools and techniques of bioinformatics to tackle complex biological questions and challenges, converting raw biological data into meaningful knowledge. Bioinformatics has emerged as a crucial aspect of contemporary biological research and its applications, spanning everything from increasing fundamental scientific knowledge to generating novel approaches to health, agriculture, and industry. Bioinformatics has transformed medicine, revolutionizing the ways we understand disease processes and develop treatment options. One of the leading applications of genomic medicine is to use whole-genome sequencing and complex computational analyses to enhance our understanding of the genetic variants that are linked to disease. Such analyses have uncovered causative mutations for rare Mendelian diseases, genetic risk factors for common ones, and somatic mutations driving cancer evolution. An example of this is the Cancer Genome Atlas (TCGA) project which profiled genomic aberrations in several cancer types, resulting in better classification systems and targeted treatment plans. Pharmacogenomics expands on the aforementioned observations to anticipate patients' variable responses to specific medications using their respective genetic profiles so that healthcare professionals can tailor the selection of specific drugs and their dosages, mitigating negative outcomes. This individualized model of medicine, "precision medicine" shows novel paradigm of "one which fits all " to targeted mechanism of action to give causing in less sever toxicity and side effects.

The use of bioinformatics has revolutionized infectious disease research, especially in the age of emerging and re-emerging pathogens. Genome sequencing and analysis tools for pathogens have become standard methods for characterizing disease-causing agents, virulence factors, and tracking transmission dynamics. Real-time genomic surveillance can help researchers track the evolution of the pathogen during outbreaks, identify new variants with changed properties, and realign public health responses. The COVID-19 pandemic had a profound impact on many aspects of people’s lives, but it also placed bioinformatics front-and-centre in this regard, with global efforts to sequence, analyze, and monitor SARS-CoV-2 genomes, while diagnostic tests and vaccines were developed with astonishing speed, all thanks to targeted collaboration of scientists across many borders. Metagenomics approaches have additionally widened our scope of identification and characterization of pathogens without prior

knowledge or culturing requirements directly from clinical samples, therefore, the possibility for the diagnosis of infectious agents with unknown etiologies. The field of bioinformatics has significantly accelerated the process of drug discovery and development due to reduced time and cost of novel therapeutics reaching the market. In this approach, chemical libraries with up to 1 million compounds are screened using structure-based docking methods to identify candidate drug-like compounds for experimental validation. The design of structure- based drugs apply knowledge of the structures of biomolecules to create new molecules that interact specifically with a target, with optimized binding affinity and selectivity. Network pharmacology, which enables the analysis of complex relationships between drugs and biological networks of multiple targets, allowing for a more complete portrait of drug effects and possible side effects. These numerical approaches are alongside classical experimental methods, allowing for more focused and effective drug discovery. In addition, bioinformatics has supported drug repurposing efforts, where licensed drugs are screened and assessed for new therapeutic indications, thus avoiding the lengthy de novo drug development process.

In the field of agricultural sciences, bioinformatics has emerged as one of the most important tools in crop improvement as well as livestock breeding programs. A further development is marker-assisted selection, where genetic markers associated with desirable phenotypic traits are used to improve the speed of breeding by directing breeding decisions without the need to assess the whole phenotype (Kumar and Sinha 2017). Genomic selection takes this one step further by using all of the genetic markers at once, which allows for prediction of complex traits influenced by many genes. They have been especially useful for traits that are challenging or expensive to assess directly when working with a specifically modified crops, such as drought tolerance or disease resistance. Comparative genomics across diverse plant species has identified conserved genes along with regulatory elements that dictate key agronomic traits and can be harnessed for genetic improvement. Moreover, the development of bioinformatics tools has enabled the characterization of plant-microbe interactions and manipulation of beneficial microorganisms to introduce sustainable crop management strategies. Bioinformatics applications have revolutionized the field of biodiversity and evolutionary biology. Phylogenomics is the reconstruction of the evolutionary relationships among species using the information in entire genomes or transcriptomes, which has provided insights challenging long-standing taxonomic definitions as well as resolving parts of the tree of life with unprecedented resolution. Molecular dating methods allow for estimating the timing of evolutionary events, e.g., species divergence or gene duplication,

forming a temporal framework for evolutionary history reconstruction. Population genomics approaches examine genetic variation between individuals of species revealing their demographic histories, gene flow versus isolation, and selection signatures. In what ways have these methods helped us understand speciation processes, adaptation to changing environments, and conservation priorities for endangered species? Metagenomics has extended this capacity by allowing for the characterization of microbial communities in different environments, revealing a tremendous diversity that might never have been accessed through conventional cultivation-based approaches.

Applications of bioinformatics have proven useful in environmental monitoring and management, including ecological and ecosystem health, sustainability and resilience. Ecological genomics strategies scrutinize the responses of organisms to environmental change at the level of genetic expression, the costs of these responses can be useful harbingers of ecosystem stress in advance of symptoms on the ecosystem fabric. Using eDNA analysis, we can test the presence of potentially invasive or endangered species in an ecosystem by analyzing trace DNA signature in an environmental sample. eDNA analyses permit the non-invasive assessment of biodiversity in the local habitat and monitoring of rare or invasive species. In different environments, including soil and marine ecosystems, bioinformatics-based analyses of metagenomic data have elucidated functional capabilities of microbial communities that impact various biogeochemical cycles and ecosystem services. These methodologies have guided conservation policies, pollution assessments, and restorative initiatives, resulting in improved environmental governance practices as anthropogenic stressors continue to mount. In the industrial field, bioinformatics has enabled the design and optimization of bioprocesses for a wide range of applications including biopharmaceuticals and biofuels. These genome-scale metabolic models are used in metabolic engineering to help predict how genetic changes will alter cellular metabolism, ultimately informing the design of microbial strains with improved production potential. In enzyme engineering, computational methods are performed to elucidate modifications in protein residues that may improve catalytic characteristics, stability or substrate selectivity, leading to the design of better biocatalysts for industrial reactions. Synthetic biology goes a step further and involves the design and construction of new biological parts, devices, and systems, as well as the redesign of existing biological systems not found in nature, thus expanding the functional capacity of biology that can be tapped for industrial processes. Using bioinformatics tools and synthetic biology pipelines, these strategies have shed light on how bioprocesses can be made more sustainable for

the production of biofuels, chemicals, pharmaceuticals, and other commodities, aiding the bio-based economy transformation.

Bioinformatics has also been pivotal in covering our understanding of complex biological systems and phenomena. Systems biology approaches incorporate various omics data to build intricate models of cellular networks, explaining how various constituents impact each other to yield emergent properties at the system level. These models have contributed to our understanding of how cells respond to perturbations, the mechanisms of disease, and potential targets of intervention for therapeutic strategies. Bioinformatics analyses of gene expression dynamics during embryogenesis have provided insights into the molecular mechanisms controlling the final fate of cells and the patterns of tissues (6, 7). Supported by bioinformatics approaches, neuroscience can exploit brain connectivity patterns from experimental data, sequencing data of genes expressed in distinct neuronal cell types and the genetic basis of neurological disorders to further our knowledge of brain function and dysfunction. Bioinformatics today with other most advanced technologies has broadened the horizons of scientific research and applications. Combined with sophisticated computational analyses, technologies for single-cell omics have uncovered unparalleled cellular heterogeneity within tissues, reshaping the paradigm of development, disease, and cellular identity. These emerging technologies, such as spatial transcriptomics and proteomics, allow spatially resolved profiling of both gene expression and protein abundance2,3,6,8. Multi-omics integration methods joint the omics (genomics, transcriptomics, proteomics, metabolomics, etc.) data of the same samples, providing a more complete picture of biological systems than any single omics data. This approach has been especially beneficial in augmenting our knowledge of diseases such as cancer, diabetes, and neurodegenerative disorders, where multiple factors are contributors to the disease process.

Another significant application field includes the training and development of bioinformatics education and resources. By training the next generation of researchers in these interdisciplinary areas, bioinformatics education programs work to ensure that the future workforce is able to take on tomorrow's challenges. Well-established and effective database maintenance and curation efforts preserve and improve the biological data repositories that are indispensable instruments for the scientific community. Continuous advances in algorithms and software development contribute to methodological progress in biological data analysis to improve accuracy, efficiency, and accessibility. Bioinformatics education and resource development efforts play an import role in the larger scientific enterprise by allowing researchers in many other areas of investigation to utilize

bioinformatics approaches in their work without requiring extensive computational expertise. Bioinformatics Applications in Future: Although biological questions will change, and more so will the technologies. Precision medicine approaches seek to incorporate genomic, environmental, and lifestyle information to facilitate a transition to highly personalized health care regimens beyond the current paradigm centered on genetics. Digital health applications are being developed based on bioinformatics analyses of wearable device data and electronic health records, which are expected to enable real-time health monitoring and provide personalized recommendations. As a result, agriculture applications will eventually be one of the biggest drivers of demand for genomics technologies as society struggles to navigate food security challenges in the face of rapid environmental change, focusing on the creation of climate-resilient crops able to withstand extreme weather events and sustainable farming practices that limit environmental impact and increase resource efficiency. Industrial biotechnology will further combine bioinformatics to improve bioprocesses by allowing for more sustainable methods and contributing significantly to the circular bioeconomy. These new applications demonstrate the ongoing evolution of bioinformatics, which continues to play a role in tackling some of the biggest challenges faced by society.

Bioinformatics plays a role from basic science to end-products at public health, agriculture, environmental science, and industry. These applications highlight the diverse and impactful nature of bioinformatics as a driving force in contemporary science and technology. Bioinformatics, as a field, will continue to evolve alongside advances in technology, with new applications and frameworks emerging as biological information grows and computational methods become more powerful, ultimately leading to greater insights and innovations that will help address some of the most pressing challenges faced by humanity and advance our understanding of life in all its complexity. Bioinformatics the interdisciplinary foray of biology with computer science, mathematics, and statistics as an offshoot, is well positioned as a frontier that will continue to fuel scientific progress and technological innovation across numerous areas in future.

Summary

Bioinformatics is an interdisciplinary field that combines biology, computer science, mathematics, and information technology to collect, analyze, and interpret vast amounts of biological data. With the explosion of genomic and proteomic data in modern research, bioinformatics has become essential for managing, analyzing, and visualizing biological information. It plays a crucial role in understanding the structure, function, and evolution of biological

molecules and systems. The main **components of bioinformatics** include **biological databases**, which store DNA, RNA, and protein sequences; **tools and algorithms** for analyzing data (e.g., sequence alignment, structure prediction); and **computational methods** for modeling biological processes and solving complex biological questions.

The **scope of bioinformatics** spans various domains such as **genomics** (study of genomes), **proteomics** (study of proteins), **drug discovery**, **agriculture**, **personalized medicine**, and **evolutionary biology**. It helps in identifying genes, predicting protein structures and functions, analyzing gene expression data, designing new drugs, and understanding genetic disorders. As a result, bioinformatics has become a cornerstone of modern biology and biotechnology, enabling discoveries that were previously impossible due to data limitations.

✅ Multiple Choice Questions (MCQs)

1. **Which of the following best defines bioinformatics?**
   a) Study of bacteria using microscopes
   b) Application of IT to biological data analysis
   c) Manual data recording in biology labs
   d) Engineering of medical devices
   Answer: b) Application of IT to biological data analysis

2. **Which is a major component of bioinformatics?**
   a) Weather forecasting
   b) Biological databases
   c) Food preservation
   d) Animal breeding
   Answer: b) Biological databases

3. **What is the role of algorithms in bioinformatics?**
   a) Cleaning laboratory equipment
   b) Analyzing and processing biological data
   c) Drawing cells
   d) Feeding lab animals
   Answer: b) Analyzing and processing biological data

4. **Which of the following is NOT a part of bioinformatics scope?**
   a) Drug discovery
   b) Protein structure prediction
   c) Soil testing
   d) Genomic data analysis
   Answer: c) Soil testing

5. **Which field contributes to the development of bioinformatics tools and software?**
   a) Chemistry
   b) History
   c) Computer Science
   d) Geography
   Answer: c) Computer Science

Short Answer Questions

1. Define bioinformatics and its significance in biology.
2. Name two key components of bioinformatics.
3. Mention any two areas where bioinformatics is applied**.**

Long Answer Questions

1. Explain the main components of bioinformatics and how they work together.
2. Describe the scope of bioinformatics with examples from medicine, agriculture, and environmental science.
3. Discuss how bioinformatics has revolutionized biological research and data analysis.

Unit 5.3

Bioinformatics Databases

Biological databases are the foundation of contemporary bioinformatics and computational biology, supporting organized storage of the vast biological data produced by scientific research. Such databases act to store, organize and provide access to a rich variety of biological information including but not limited to nucleotide and protein sequences, three-dimensional structure, functional annotations, metabolic pathways and taxonomic classification. These databases have become invaluable resources for scientists all over the world for data sharing, comparative analysis and the discovery of new biological knowledge as biological data has surged to new levels in the last few decades. The earliest biological databases date to the 1960s and 1970s, when the first protein sequence databases were created. The second wave of databases began in the 1980s and 1990s when DNA sequencing technologies improved, leading to the development of specialized nucleotide sequence databases and, driven by high-throughput sequencing and other omics technologies, an explosion in the diversity and specialization of biological databases. There are now hundreds of biological databases that cater to specific research communities or types of data.

There are many biological databases that are used worldwide, the principal ones include: the European Molecular Biology Laboratory (EMBL), the DNA Data Bank of Japan (DDBJ), the National Center for Biotechnology Information (NCBI), Swiss-Prot, and the Protein Data Bank (PDB). These databases constitute fundamental pillars of the biological data infrastructure world, which collectively comprise the International Nucleotide Sequence Database Collaboration (INSDC) and other community-based international data sharing efforts. The databases possess varying elements, advantages, and historic importance in the bioinformatics community.

European Molecular Biology Laboratory (EMBL)

EMBL is the European Molecular Biology Laboratory, a molecular biology research organization with sites in different countries in Europe, but for biological databases it refers to the EMBL nucleotide sequence database, now part of the European Nucleotide Archive (ENA). Founded in 1980, the EMBL database became one of the first global DNA and RNA sequence databases and has grown into a sophisticated data infrastructure operated by the European Bioinformatics Institute (EMBL-EBI) based at the Wellcome Genome Campus in Hinxton, England. All publicly available sequence data (from individual researchers or genome sequencing projects, and from the scientific literature) are collected, maintained and distributed

through the EMBL database. You are at: Home · How It Works · It acts as the main resource for nucleotide sequences of Europe and is a member of the International Nucleotide Sequence Database Collaboration (INSDC) with NCBI's GenBank and DDBJ, which makes sure that the three main houses of nucleotide sequences keep in step with one another.

EMBL data model is hierarchical, which means that it separates sequence records into primary sequences and their features. Every EMBL entry is rich in information, which includes not only the sequence in question, but also taxonomic data, literature references, functional annotations and cross-references to other sequence and structural databases. Data is stored in a common flat file format, known as the EMBL format, which organizes data in a human-readable manner, with 2-letter codes denoting the type of data in each line.

At EMBL-EBI, we offer a suite of tools and services to access and analyze the sequence data. The ENA Browser enables user to query and retrieve sequences with accession numbers, keywords, or sequence similarity. RESTful APIs and FTP services enable programmatic access to the data, allowing large-scale analyses to be automated. EMBL-EBI also provides tools for sequence analysis, including FASTA and BLAST, which compare a sequence against the database. In addition to nucleotide sequences, EMBL-EBI also organizes many other biological databases and resources of diverse data types. These include structural data deposited in the Protein Data Bank, European Nucleotide Archive, the functionally focused Array Express, eukaryotic genome annotation in Ensembl, protein family classification in InterPro and many others. This collection of databases, coupled together, provides researchers with information of a broad scope across many biological types. With respect to biological research, the EMBL database has played a key role as a repository, providing a means to share, standardise and analyse sequence data. It has funded these and many other discoveries across genomic, evolutionary biology and other life science disciplines. As sequencing technologies evolve, EMBL-EBI repeatedly realigns its infrastructure to enable scientists to handle the ever-growing volume of biological data while keeping pace with increasing complexity, thus recoiling as a vital source in the scientific community.

DNA Data Bank of Japan (DDBJ)

The DNA Data Bank of Japan (DDBJ) is the main nucleotide sequence database in Asia and one of the three bases of the International Nucleotide Sequence Database Collaboration (INSDC). The DNA Data Bank of Japan (DDBJ) was born from a necessity in 1986, when it was established at the National Institute of Genetics (NIG) in Mishima, Japan, to complement the Edmund D. Perkins Institute and create an

Asian bulk nucleotide sequence repository that would aid regional biologists and support global biological data infrastructure. DDBJ receives DNA sequences directly from researchers and sequencing projects, most of them based in countries in Asia, but it will take submissions from scientists anywhere in the world. DDBJ participates in the International Nucleotide Sequence Database Collaboration (INSDC) and shares data on a daily basis with partner databases (the European Sequence data at the European Molecular Biology Laboratory-European Nucleotide Archive (EMBL-EBI) and GenBank at the National Center for Biotechnology Information (NCBI) in the USA) at the same time, so that the same aggregate of sequence data can be found at all three databases. This kind of scientific data sharing is one of the most successful examples of international data synchronization.

Sequence records are organized in a structured format similar to that of EMBL and GenBank. Entries include basic identifiers (like accession numbers, which make a sequence unique) and sequence data, taxonomic information, bibliographic references, and feature annotations for biological significance of parts of the sequence. In its traditional flat file format, the DDBJ describes this information in a few different line types, allowing for both human readability and machine parsing. In addition, DDBJ provides a wide range of services, not limited to just storing data. The Nucleotide Sequence Submission System (NSSS) offers both web-based and offline tools for researchers to submit new sequence data to GenBank. Data submitted are validated for quality and consistency before assignment of accession numbers and integration into the database. Information retrieval is also available, such as getentry for accessing individual records by accession numbers, and ARSA (All-Round Sequence Search) for keyword-based searches from various fields. The center also offers a number of analytical tools that make it easier for researchers to analyze the sequence data they generate. These include services for sequence similarity searches (BLAST and FASTA), multiple sequence alignments, as well as specialized resources for analyzing next generation sequencing data. The DDBJ Read Archive (DRA) stores raw sequencing reads from high-throughput sequencing platforms, and the Japanese Genotype-phenotype Archive (JGA) is a secure repository for human genetic variation data, with controlled access to protect the privacy of human subjects.

DDBJ has evolved over the past decades to provide new services that met the needs in genomic research. Metadata describing research projects or biological materials, the BioProject and BioSample databases in DDBJ, respectively, are also important to interpret sequence submissions. Its centers also focused on metagenomic and

environmental sequence data reflecting the genomic research's context. DDBJ enables biological science researchers worldwide to access and use biological data, whilst also offering specialised support to the scientific community of Asia. DDBJ thus not only promotes world genomics research through high data quality, but also through the provision of data to international initiatives to promote bioinformatics. Since the previous update, DDBJ continues improving both data submission services and data management, working not only to provide broad re-deposited data but also services and infrastructure to help to manage new succeeding sequencing technologies as well as the increasing volumes of data. Such infrastructure supports DDBJ to keep services relevant in the biological research ecosystem.

National Center for Biotechnology Information (NLM)

The National Center for Biotechnology Information (NCBI) is among the most extensive and use biological data repository in the world. NCBI was established in 1988, as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH) in United States, which is also formed by congressional legislation to develop information systems for molecular biology and genetics. This federally funded institute grew from a small database provider into a far-reaching, multifactorial bioinformatic resource (one that serves millions of researchers worldwide) under the guidance of its founding director, Dr. David Lipman. GenBank, the most widely used nucleotide sequence database, is at the heart of NCBI's resources and serves as the U.S. Nucleotide Sequence Database node of the International Nucleotide Sequence Database Collaboration (INSDC) (17). GenBank was originally created at Los Alamos National Laboratory and transferred to NCBI management in 1992. It houses publicly available DNA and RNA sequences from a submission by individual laboratories (such as those found in GenBank), large-scale sequencing projects (such as the Human Genome Project), and patent applications. GenBank entries provide detailed details on a sequence, such as its source organism, publications, functional annotations and cross-references to other databases.

Established as the core biological sequence repository, GenBank is now part of a greater data ecosystem hosted at NCBI, which spans across databases with diverse biological data types. Description The RefSeq database [1] a non-redundant, curated database of reference genomes, transcripts, and proteins. Gene — Information on gene loci, containing names, chromosomal positions and phenotypes. The Protein database includes amino acid sequences resulting from translations of coding sequences from GenBank and other sources (e.g., Swiss-Prot). Among NCBI's structural biology resources is the Molecular Modeling database (MMDB), a three-dimensional structure database that has

been prepared from the Protein Data Bank (PDB), emphasizing biological assemblies and structure→sequence relationships. Conserved Domain Database (CDD) determines conserved protein domains, and PubChem is a chemical structure database of small molecules and their biological activities.

In the area of literature resources, NCBI hosts PubMed, the largest biomedical literature database in the world, with over 30 million citations. PubMed Central (PMC) offers this service but adds open access to the full-text of biomedical and life sciences journal literature. These databases of literature interconnect with sequence and structural databases in an interconnected information system. NCBI powers search and analysis tools that enable researchers to explore its vast collections of data. Entrez provides a search interface across all NCBI databases and can help users to discover interdependencies between different types of data. The standard algorithm for sequence similarity searches globally is BLAST (Basic Local Alignment Search Tool) developed by NCBI scientists. Additional Bioinformatics Analysis Tools (e.g., Primer-BLAST: for design of PCR primers, CD-Search: conserved domain search, Genome Data Viewer: genome visualization and analysis, etc.). In anticipation of the burgeoning high-throughput sequencing technologies, NCBI set up specialized databases to host their information such as the Sequence Read Archive (SRA) to store raw sequencing data and the Database of Genotypes and Phenotypes (dbGaP) to store genotype-phenotype association data in a way that balances accessibility to researchers contributing data and protection from privacy breaches for human subjects. The database for research projects and biological sample metadata as part of the BioProject and BioSample databases.

NCBI has additionally created resources for clinical and medical uses. 9; ClinVar: The database assesses the relationship between genomic variants and health; ClinVar: The database records single nucleotide polymorphisms and other genomic configurations) OMIM–the Online Mendelian Inheritance in Man database available due to NCBI–is a comprehensive resource for the relationship of human genes in disease. Educational resources are another major component of NCBI's mission. Title Text (if applicable): Bookshelf ID: NCBI Bookshelf. NCBI also preserves the educational materials of online courses, webinars, and alpha versions of point-and-click training bibliographic utilities that teach bioinformatics through successful examples of its use by the broader scientific community. NCBI has had a tremendous impact on biological and biomedical research. Ans- NCBI has facilitated the acceleration of scientific discovery in multiple areas from fundamental molecular biology to clinical genetics and drug development by creating an integrated information infrastructure.

NCBI continues to innovate and develop new solutions for managing, integrating, and analyzing the widespread biological data in response to the evolving needs of the scientific community.

Swiss-Prot

Swiss-Prot is one of the most respected and highly curated protein sequences databases in the bio-informatics world. Swiss-Prot is a protein sequence database introduced in 1986 by Amos Bairoch at the University of Geneva, Switzerland based on the philosophy of obtaining the most accurate information on as few sequences as possible, thus recognizing the need to manually curate and annotate protein sequences rather than just having vast amounts of data. This method has ensured Swiss-Prot provides an invaluable resource for researchers who need the most accurate protein data. The unique aspect of Swiss-Prot is its manual annotation. Whereas many other biological databases depend heavily on automated annotation, Swiss-Prot entries are manually curated in detail by expert biocurators with expertise in different areas of protein science. These curators review the scientific literature, experimental evidence, and computational predictions to generate accurate and sophisticated annotations for each protein entry. Manual curation is performed on all entries ensuring ultimately superior data quality and reliability for Swiss-Prot.

Swiss-Prot entries provide rich and structured information about a protein. In addition to the amino acid sequence itself, entries also feature recommended protein names, gene names and function descriptions. They include post-translational modifications, domains and sites, subcellular location, tissue specificity, developmental expression and diseases. Entries also include cross-references to many other databases, literature references that support the annotations, and controlled vocabulary terms that ensure consistency and facilitate computational analyses. Suzanne goes on to explain that Swiss-Prot was dramatically reorganized in 2002 into the UniProt (Universal Protein Resource) Consortium through an agreement between the European Bioinformatics Institute (EMBL-EBI), the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR). In this context, Swiss-Prot is the manually annotated part of UniProtKB (the UniProt Knowledgebase), which is supplemented by TrEMBL (Translated EMBL), a database of computationally annotated protein sequences awaiting manual curation. The UniProtKB/Swiss-Prot database uses multiple quality assurance techniques to keep the quality of its data. Entries are subjected to consistency check to detect and rectify any anomaly in annotation. Information for different species is combined into a single entry if it pertains to the same protein, and, where appropriate, information specific to other species of the same protein is clearly noted, minimizing redundancy. Sequences are

reviewed and updated with new evidence as it becomes available, enabling the database to reflect current scientific knowledge.

Swiss-Prot offers a range of tools and interfaces for accessing and analyzing its data. This affords abilities for querying the database by querying up protein or gene names, the accession numbers, by functions, and other properties on the UniProt website itself. It contains visualization tools for protein characteristics, sequences, and structures, which improve the understanding of protein information. Swiss-Prot data can be incorporated into analysis pipelines used by computational biologists via programmatic access through RESTful APIs and FTP downloads. Swiss-Prot has had a profound impact on biological research. Its teasingly edited data have underpinned thousands of studies in areas from structural biology and proteomics to systems biology and drug discovery. It has also played a key role in functional annotation of new proteins, detection of conserved domains and motifs, prediction of protein interactions, and elucidation of the molecular basis of diseases. Swiss-Prot has set essential protein annotation standards and protocols that have impacted the wider bioinformatics community. The use of controlled vocabularies and ontologies like the Gene Ontology terms that are used for the functional annotation have standardizing data in biology, making data representation similar across different databases and research groups.

With the continuing advances in proteomics research, Swiss-Prot is evolving to integrate additional data and annotations. With its enduring emphasis on manual curation and validation, the database now offers comprehensive information derived from high-throughput proteomics studies. Swiss-Prot is a protein sequence database that continues to balance breadth of coverage with depth of annotation, thus providing a critical resource for researchers looking for accurate protein information amid a cyclone of biological data.

Protein Data Bank (PDB)

The Protein Data Bank (PDB) is the worldwide repository for three-dimensional structure data of biological macromolecules, and mainly of proteins and nucleic acids. In 1971, the PDB was founded at Brookhaven National Laboratory representing the first molecular database in the field of biology and a resource now heavily relied on by structural biologists, biochemists, biophysicists, pharmacologists, and drug designers globally. Understanding that three-dimensional structural information yields insights into molecular function not obtainable from sequence data alone: the PDB was born. As experimental techniques such as X-ray crystallography for the determination of structures became more prevalent in the 1960s, the scientific community started to appreciate the necessity of a common

repository to archive and distribute structural coordinates. So you know that the PDB started 7 structures and now it covers more than 180,000 structures due to an exponential increase in structural biology research.

In 2003, control of the PDB was passed to the Worldwide Protein Data Bank (wwPDB), a collaborative group of organizations in the United States (RCSB PDB), Europe (PDBe), Japan (PDBj) and more recently China (CNCPDB). This collaborative governance mindset serves as a universal entry point to structural data but allows for a consistent set of standards for data format, validation, and annotation. The Protein Data Bank (PDB) records experimentally determined structures that can be obtained by different methods, but X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM) are typically the most common techniques. In the database, every structure gets a unique four character long alphanumeric designation called the PDB ID, which is used as a standard reference in scientific papers. For instance, the structure of myoglobin, the first protein structure to be determined, is designated as 1MBN. When the PDB was first created, the underlying data model underwent major changes. The database was originally designed to store atomic coordinates and simple annotation information in a fixed-column format. The more flexible macromolecular Crystallographic Information File (mmCIF) format was introduced in 1997 to cater for the increasing complexity of structural data. Currently, the standard is the PDBx/mmCIF format, which is a generalized model of structural data as well as experimental details, chemical components, and biological annotations.

A PDB entry has much more information than three-dimensional coordinates. This comprises information on the experimental method and conditions, the resolution of the atomic structure, the biological source of the molecule, literature references and functional annotations. The database even tracks ligands, cofactors and other nonpolymer chemical components that interface with the macromolecules, which can be important for drug discovery science. PDB data quality and validation are core concerns. We perform extensive validation checks on every structure submitted to the database to detect any errors or inconsistencies in atomic coordinates, geometric parameters, and experimental data. The wwPDB has created detailed validation reports that present assessments of structure quality to assist users in assessing the reliability of structural information for their intended purpose. The Protein Data Bank offers a number of tools and services to facilitate access to and use of structural data. Access is primarily through the web portals hosted by the wwPDB member organizations, including RCSB. org, PDBe. org, and PDBj. org. The platforms furnish strong search capabilities to users, querying the database by sequence,

structure, function, or experimental parameters. Interactive visualization tools allow users to navigate structures and emphasize functional areas, binding sites, and structural motifs.

We offer bulk distribution of all structural data via FTP services for conducting large scale computational analyses on the entire database. RESTful APIs provide a programmatic interface to integrate the structural data into automation and custom applications. Specific tools have been developed to cater more routine analytical tasks like structural alignment, binding site identification and molecular docking. The PDB has had a revolutionary influence on scientific research. The database has led to innumerable discoveries across fields from basic biochemistry to clinical medicine. An important approach in the current pharmaceutical development is structure based drug design which heavily depends on PDB data to help identify potential drug targets as well as generate candidate molecules optimized for the targets. These computational strategies for predicting protein structure, including AlphaFold and RoseTTAFold, were trained and benchmarked using PDB structures, resulting in impressive progress made in predicting protein folding. Educational materials are another facet of the PDB's mission. For example, the RCSB PDB curates the website PDB-101 incorporates educational resources, tutorials, and curricula for students and educators from all levels. This teaching contributes to building structural literacy within the next generation of scientists and the public.

With experimental methods continuing to evolve, notably the "resolution revolution" in cryo-EM and the development of integrative structural biology approaches, the PDB is faced with new challenges with regards to managing more and more complex structural data. The wwPDB is working to accommodate these new data types through the development of new data formats and validation protocols, but will seek to maintain the absolute free availability of these data while building on long-term efforts to assure data quality.

Integration and Future Directions:

The biological databases described above — EMBL, DDBJ, NCBI, Swiss-Prot and PDB — do not function independently but comprise an integrated ecosystem that together propel life sciences investigation. The integration happens at several layers, including formal data sharing arrangements, through technical cross-references connecting relevant pieces of information stored across databases. This integration allows researchers to seamlessly go from sequence to structure to function and gain a more comprehensive view of biological systems. A formal integration example is the International Nucleotide Sequence Database Collaboration (INSDC) where EMBL, DDBJ and NCBI's GenBank

transmit data each day so nucleotide sequence collections remain up to date and synchronized. The Worldwide Protein Data Bank (wwPDB) similarly promotes the consistent representation of structural data among its member organizations. UniProt as a whole: the UniProt Consortium is a group of agencies that federate the Swiss-Prot and other protein databases in a single framework that sets the standard for accessing protein knowledge worldwide. Technical integration is achieved through cross-references and mapping mechanisms that link identifiers. In fact, each entry in any one database often contains cross-references to related records in other databases — in a sense linking up all of biomedicine in one big web. Swiss-Prot protein sequences, for example, may be associated with the corresponding nucleotide sequences in GenBank, three-dimensional structures in PDB, and literature citations in PubMed. Identifier mapping services (e.g. UniProt [3], NCBI [4]) can facilitate users in tracing relationships between the different classes of biological data.

The integration is further supported through the standardisation of data formats and controlled vocabularies. Biological ontologies such as the Gene Ontology (GO) and Sequence Ontology (SO) serve as standardized terminologies to annotate genes' functions and sequence characteristics in databases. Exchange formats (for example, FASTA for sequences and mmCIF for structures), facilitate transferring data between resources. Data availability is one of the main challenges and opportunities for the future of biological databases. Advancements in high-throughput technologies are leading to an exponential growth in data volume, requiring novel approaches in data storage, management, and analysis. Data Mining for Biology Machine learning and artificial intelligence are being used more and more to derive knowledge from the growing data in biology and to find patterns that will be overlooked by human analysts. Data quality continues to be a major issue, and databases are devising ever more sophisticated validation techniques to ensure reliable data. Finding the sweet spot between automated processing, which is obviously needed for big data, and manual curation, which helps ensure that what those millions of databases point to is accurate, remains a work in progress. A sustainable way forward might lie in hybrid approaches that marry automatic pipelines to targeted expert curation.

We are still in the early days of the integration of diverse data types. Biological databases are increasingly moving beyond classical sequences and structures and adding data from proteomics, metabolomics, transcriptomics, and other omics fields. The integration of omics, the amalgamation of these disparate data types to yield more holistic perspectives of biological systems, is the cutting edge of bioinformatics investigation. With the growth of end-user communities

from specialized bioinformaticians to clinicians, students and researchers with varied backgrounds, accessibility and usability are other challenges. Databases are also creating richer interfaces, improve visualization tools and educational resources for this broader constituency. The future landscape also remains defined by ethical and legal considerations. Secure access frameworks are essential to ensure controlled access to these kinds of data, especially human genetic information where data privacy can be a concern. Open access policies facilitate scientific collaboration and reproducibility but should be weighed against privacy protections and intellectual property considerations.

Cloud computing and distributed data systems provide new paradigms for biological data management. Such methods can help alleviate storage problems and enable compute-intensive analyses on the data in place, minimising the need for large data transfers. Database federations, where local copies are maintained, but where shared standards and interfaces are adopted, may evolve more widely. In the end, the development of biological databases corresponds to the ever-changing landscape of life sciences research itself. As the next generation of researchers gains better insights into biological systems and new experimental methods become available, these digital storehouses will be updated and will remain key components in the effective storage, preservation, and access to the expanding repository of biological information. The continued development of these resources, following the principles of openness, quality, and integration, will continue to be central to advancing biological discovery and its application to medicine, agriculture, and the environment.

Summary

Biological databases are organized collections of biological information that are stored electronically and made accessible for research and analysis. These databases have become a fundamental part of modern biological science and bioinformatics, especially with the massive increase in biological data generated by genome sequencing, proteomics, and other high-throughput techniques. Biological databases store various types of data, such as DNA and RNA sequences, protein structures, gene expression profiles, molecular interactions, and biological pathways. They are essential for understanding biological functions, discovering new genes or proteins, comparing sequences, and identifying disease-associated mutations.

There are different types of biological databases: primary databases, which contain raw data (e.g., GenBank, EMBL); secondary databases, which contain analyzed or derived data (e.g., Swiss-Prot); and

specialized databases, which focus on specific organisms, diseases, or biological processes (e.g., KEGG, PDB). These databases can be accessed through the internet and are used by researchers worldwide to share and retrieve biological information. They also play a crucial role in drug discovery, disease research, functional genomics, and evolutionary studies. In summary, biological databases are the backbone of bioinformatics, enabling data-driven discoveries and collaboration in life sciences**.**

Multiple Choice Questions (MCQs)

1. What is the primary purpose of biological databases?
   a) Performing laboratory experiments
   b) Growing biological samples
   c) Storing and organizing biological information
   d) Training scientists in wet lab techniques
   Answer: c) Storing and organizing biological information

2. Which of the following is a primary biological database?
   a) Swiss-Prot
   b) GenBank
   c) KEGG
   d) PDB
   Answer: b) GenBank

3. What type of data is commonly found in biological databases?
   a) Weather data
   b) Financial statistics
   c) DNA sequences and protein structures
   d) Historical records
   Answer: c) DNA sequences and protein structures

4. Which database is used for storing 3D structures of biological macromolecules?
   a) GenBank
   b) KEGG
   c) PDB (Protein Data Bank)
   d) EMBL
   Answer: c) PDB (Protein Data Bank)

5. Which of the following describes a secondary biological database?
   a) Contains raw sequence data
   b) Focuses on historical data
   c) Contains curated and interpreted data
   d) Stores satellite images
   Answer: c) Contains curated and interpreted data

Short Answer Questions

1. What is a biological database? Give one example.
2. What are primary and secondary biological databases?
3. Why are biological databases important in research?

Long Answer Questions

1. Describe the different types of biological databases and their roles.
2. Explain how biological databases contribute to bioinformatics and scientific discovery.
3. Discuss the significance of biological databases in modern biology and give examples of commonly used databases.

References

- Lesk, A.M. (2019). *"Introduction to Bioinformatics"* (6th ed.). Oxford University Press, Chapter 1, pp. 3-15
- Mount, D.W. (2001). *"Bioinformatics: Sequence and Genome Analysis"* (Cold Spring Harbor Laboratory Press), Chapter 1, pp. 1-10
- Baxevanis, A.D., & Ouellette, B.F.F. (2005). *"Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins"* (3rd ed.). Wiley-Interscience, Chapter 2, pp. 15-40
- Rösling, F. (2008). *"Bioinformatics: From Genomes to Drugs"* (Wiley-VCH), Chapter 1, pp. 1-15
- Alterovitz, G., & Ramoni, M. (Eds.). (2011). *"Translational Bioinformatics: Case Studies and Applications"* (Springer),

MATS UNIVERSITY

MATS CENTER FOR OPEN & DISTANCE EDUCATION

UNIVERSITY CAMPUS : Aarang Kharora Highway, Aarang, Raipur, CG, 493 441

RAIPUR CAMPUS: MATS Tower, Pandri, Raipur, CG, 492 002

T : 0771 4078994, 95, 96, 98 M : 9109951184, 9755199381 Toll Free : 1800 123 819999

eMail : admissions@matsuniversity.ac.in Website : www.matsodl.com