MERCY MUMO

SCM 222-1290/2025

INFORMATION TECHNOLOGY FOR STATISTICS SUMMARY NOTES

Definition of information technology

information technology (IT) is "the study, design, development, application, implementation,

support or management of computer-based information systems.

Definition of statistics

Statistics in Information Technology (IT) can be defined as the science of collecting, analyzing, interpreting, and presenting data to support decision-making, optimize systems, and uncover patterns within technological environment**.**

Role of IT in Different Fields

1. Business & Commerce

Automates business processes (e.g., payroll, inventory).

Enhances communication through email, video conferencing.

Supports decision-making using MIS, data analytics.

Enables e-commerce (online shopping, online banking).

Improves productivity through software tools (Excel, ERP).

2. Education

E-learning platforms (Google Classroom, Moodle).

Online libraries and digital content.

Computer-based testing and assessments.

Virtual classrooms and video lectures.

Enhances research using databases and statistical tools.

3. Healthcare

Electronic medical records (EMR) for efficient patient data management.

Telemedicine & remote diagnosis.

Computer-aided surgeries & medical imaging.

Hospital Management Systems (HMS) for administration.

Health data analytics for disease prediction.

4. Banking & Finance

Online and mobile banking.

ATM services & digital payments.

Fraud detection using AI & monitoring systems.

Financial modeling and forecasting.

Electronic fund transfers (EFT, RTGS, M-Pesa)

5. Government (E-Government)

Digital ID systems & e-citizen services.

Online tax filing, license renewal, public records.

Improved transparency and reduced corruption.

Faster delivery of government services.

6. Agriculture

Precision farming using sensors & IoT.

Weather forecasting to aid planting/harvesting.

Market price information for farmers.

Drones for monitoring farms.

Automated irrigation systems.

7. Transportation & Logistics

GPS systems for navigation and tracking.

Traffic management systems.

Online ticketing for buses, trains, flights.

Supply chain management and route optimization.

Self-driving car technologies

8. Communication

Instant messaging apps (WhatsApp, Messenger).

Video calls & teleconferencing.

Social media platforms.

VoIP services (Skype, Zoom).

Global connectivity without boundaries.

9. Entertainment

Video streaming (Netflix, YouTube).

Digital music platforms.

Video games & VR/AR.

Animation and special effects in movies.

Online content creation.

10. Science & Research

Data analysis using statistical software.

Simulation and modeling.

Storage and sharing of huge datasets.

High-performance computing (supercomputers).

Online research publications.

11. Security & Defense

Cybersecurity systems for protection.

Surveillance systems (CCTV, drones).

Military communication networks.

AI-based threat detection.

Secure data encryption.

12. Manufacturing ⚙

Robotics and automation.

Inventory and supply chain systems.

Quality control using sensors.

Classification of Computers.

Computer: Definition

A computer is an electronic device that can be programmed to manipulate symbols.

1. Functions of a Computer

A computer performs four main functions in the data processing cycle:

Input: Accepts data and instructions from input devices (keyboard, mouse, scanner).

Processing: CPU (with ALU + Control Unit) processes data according to instructions.

Storage: Saves data temporarily (RAM) or permanently (hard drives, SSDs).

Output: Produces results via output devices (monitor, printer, speakers).

Supporting functions:

Communication: Enables networking and data exchange.

Control: Directs and coordinates all operations.

2. Advantages of Computers

Speed: Perform millions of calculations per second.

Accuracy: High precision, minimal errors compared to manual work.

Storage: Store vast amounts of data compactly.

Automation: Automates repetitive tasks, saving time.

Connectivity: Facilitates communication and access to information worldwide.

Versatility: Used in education, business, healthcare, engineering, entertainment, etc.

Decision Support: Helps in data analysis and informed decision-making.

3. Disadvantages of Computers

Cost: High purchase and maintenance expenses.

Dependency: Over-reliance can reduce human skills.

Security Risks: Vulnerable to hacking, viruses, and data theft.

Unemployment: Automation may replace certain jobs.

Health Issues: Prolonged use can cause eye strain, posture problems.

Obsolescence: Hardware/software become outdated quickly.

No Human Judgment: Computers cannot replace human creativity or ethical reasoning

Characteristics of computers

It responds to a specific set of instructions in a well-defined manner.

(It can execute a prerecorded list of instructions[A program).

It can quickly store and retrieve large amounts of data.

components of a computer Components:

☞**Central processing unit (CPU):** The heart of the computer, this is the component that

actually executes instructions organized in programs ("software") which tell the

computer what to do.

☞**Memory(fast, expensive, short-term memory):** Enables a computer to store, at least

temporarily, data, programs, and intermediate results.

☞**Massstoragedevice (slower, cheaper, long-term memory):** Allows a computer to

permanently retain large amounts of data and programs between jobs. Common mass

storage devices include disk drives and tape drives.

☞ **Inputdevice**: Usually a keyboard and mouse, the input device is the conduit through

which data and instructions enter a computer.

☞**Outputdevice**: A display screen, printer, or other devices that let you see what the

computer has accomplished.

Data vs Information- Differences in meaning

Data are plain facts. When data are processed, organized, structured or presented in a given

context so as to make them useful, they are called Information

when data is interpreted and processed to determine its true meaning, it becomes useful and can be called Information.

Bit

In computing**, a bit can be defined as a variable or computed quantity that can have only two**

possible values. These two values are often interpreted as binary digits and are usually denoted by the numerical digits 0 and 1. The two values can also be interpreted as logical values (true/false, yes/no), algebraic signs (+/−), activation states (on/off), or any other two-valued attribute.

Byte

In most computer systems, **a byte is a unit of data that is eight bit long.** A byte is the unit most

computers use **to represent a character such as a letter, number, or typographic symbol (for**

example, "g", "5", or "?").

4 bytes constitute a word

ANALOG SIGNALS

Analog signals are a representation of time varying quantities in a continuous signal.

Advantages

Closer to natural signals (speech, sound, light).

Can represent very fine variations (high resolution).

Simple to generate and process with basic equipment.

Disadvantages

Prone to noise, distortion, and signal degradation.

Difficult to store and transmit over long distances.

Copies lose quality each time.

Equipment can be bulky and less flexible.

DIGITAL SIGNALS

A digital signal is a way of transmitting data that converts the data to discrete values,

usually based on the binary code that computer systems work upon, which consists of packets of information coded as strings of ones and zeros.

advantanges

Less noise interference, more reliable over long distances.

Easy to store, compress, and transmit.

Supports error detection and correction.

Copies can be made indefinitely without quality loss.

Enables modern computing, networking, and multimedia applications.

Disadvantanges

Requires more bandwidth for transmission.

Conversion from analog to digital may lose detail (quantization error).

More complex and expensive equipment.

Artificial reproduction of natural signals (not exact).

Properties of Digital vs Analog signals

Digital information has certain properties that distinguish it from analog communication

methods. These include

ŋ Y Ś Ö d B 🕐 **Synchronization**– digital communication uses specific synchronization sequences for

determining synchronization.

🕐 **Language**–digital communications requires a language which should be possessed by both

sender and receiver and should specify meaning of symbol sequences.

🕐 **Errors**–disturbances in analog communication causes errors in actual intended

communication but disturbances in digital communication does not cause errors enabling

error free communication. Errors should be able to substitute, insert or delete symbols to be

expressed.

🕐 **Copying–**analogcommunication copies are quality wise not as good as their originals

while

due to error free digital communication, copies can be made indefinitely.

🕐 **Granularity**– for a continuously variable analog value to be represented in digital form there occur quantization error which is difference in actual analog value and digital representation and this property of digital communication is known as granularity.

Applications of Analog Signals

Audio systems – microphones, speakers, vinyl records, amplifiers.

Radio broadcasting – AM/FM radio transmission.

Traditional telephony – landline voice communication.

Old television systems – analog TV signals.

Measurement instruments – thermometers, pressure gauges, sensors.

Industrial control systems – analog circuits, regulators, amplifiers.

Applications of Digital Signals

Telecommunications – mobile phones, VoIP, fiber-optic communication.

Computing – all computer operations, data storage, and processing.

Multimedia – CDs, DVDs, MP3s, streaming platforms.

Medical imaging – MRI, CT scans, digital X-rays.

Internet & networking – data packets, secure communication.

Modern television – digital TV, cable, satellite systems.

Security systems – encryption, authentication, digital forensics.

Digital signal processing (DSP) – filtering, compression, enhancement in engineering and science.

OUTPUT UNIT

Output devices accept binary data from the computer- decodes it into original form and

supplies this result to the outside world. E.g. Monitors, Printer, Video terminals (provides

both input & output functions), graphic displays etc.

MEMORY UNIT

The memory unit stores program and data. There are two classes of memory devices :- Primary memory and Secondary memory.

Primary memory (Main memory)

- Contains a large number of semiconductor cells each capable of storing one bit of information.
- These cells are processed in group of fixed size called words containing 'n' bits. The main memory is organized such that the contents of one word can be stored or retrieved in one basic operation.
- For accessing data, a distinct address is associated with each word location.
- Data and programs must being the primary memory for execution.
- Number of bits in each word is called the word length and it may vary from 16 to 64 bits.
- Fast memory
- Expensive
- Time required to access one word is called Memory Access Time- 10nS to 100nS. This time is fixed and independent of the location.

Secondary storage

They are used when large amount of data have to be stored (also when frequent access is not necessary)

CENTRAL PROCESSING UNIT (CPU)

It consists of Arithmetic and Logic Unit and Control Unit.

Control Unit

- The operations of all the units are coordinated by the control unity (act as the nerve center that sends control signal to other units)

- Timing signal that governs the I/O transfers are generated by the Control Unit.
- Synchronization signals are also generated by the Control Unit
- Byselecting, interpreting and executing the program instructions the control unit is able to maintain order and direct the operation of the entire system

INTERNAL ORGANIZATION OF PROCESSOR

Processor contains a number of registers used for temporary storage of data other than ALU and Control circuitry.

Instruction Register (IR)– holds the instruction that is currently being executed.

Program Counter (PC)– It contains the address of the instruction currently being executed.

General Purpose Registers (R 0 to R n-1)– Facilitates communication with the main memory.

Memory Address Register (MAR)– holds the address of the location to or from which data are to be transferred

Memory Data Register (MDR)– contains the data to be written into or read out of the address location.

System Software

System software is a program that manages and supports the computer resources and operations of a computer system while it executes various tasks such as processing data and information, controlling hardware components, and allowing users to use application software.

systems software functions as a bridge between computer system hardware and the application software.

System software is made up of many control programs**, including the operating system, communications software and database manager.**

Systems software consists of three kinds of programs. The system management

programs, system support programs, and system development programs.

SYSTEM MANAGEMENT PROGRAMS

These are programs that manage the application software, computer hardware, and data

resources of the computer system. **These programs include operating systems, operating**

environment programs, database management programs, and telecommunications monitor

programs.

SYSTEM SUPPORT PROGRAMMES

These are the programs that help the operations and management of a computer system. They provide a variety of support services to let the computer hardware and other system programs run efficiently. **The major system support programs are system utility programs, system performance monitor programs, and system security monitor programs (virus checking programs).**

Operating Systems

An operating system is a collection of integrated computer programs that provide recurring services to other programs or to the user of a computer. **These services consist of disk and file management, memory management, and device management**. In other words, it manages CPU operations, input/output activities, storage resources, diverse support services, and controls various devices.

Operating System Functions

- **Resource Management:** An operating system manages a collection of computer hardware resources by using a variety of programs. It manages computer system resources, including its CPU, primary memory, virtual memory, secondary storage devices, input/output peripherals, and other devices.
- **Task Management**:The function of the operating system that controls the running of many tasks. It manages one program or many programs within a computer system simultaneously.

- **File management:** This is a function that manages data files. An operating system contains file management programs that provide the ability to create, delete, enter, change, ask, and access of files of data. They also produce reports on a file.
- **User Interface:** It is a function of an operating system that allows users to interact with a computer. A user interface program may include a combination of menus, screen design, keyboard commands. A well-designed user interface is essential for an operating system to be popular. Because of the function, users can load programs, access files, and accomplish other tasks.

Utility software.

Utility software is system software designed to help analyze, configure, optimize or maintain a computer. A single piece of utility software is usually called a utility or tool.

Utility software usually focuses on how the computer infrastructure (including the computer hardware, operating system, application software and data storage) operates.

Types of utilities

- **Anti-virus utilities** scan for computer viruses.
- **Backup utilities** can make a copy of all information stored on a disk, and restore either the entire disk (e.g. in an event of disk failure) or selected files (e.g. in an event of accidental deletion).
- **Datacompression utilities** output a shorter stream or a smaller file when provided with a stream or file.
- **Diskcheckers** can scan operating hard drive.
- **Diskcleaners** can find files that are unnecessary to computer operation, or take up considerable amounts of space. Disk cleaner helps the user to decide what to delete when their hard disk is full.

- **Diskcompression** utilities can transparently compress/uncompress the contents of a disk, increasing the capacity of the disk.
- **Diskdefragmenters** can detect computer files whose contents are broken across several locations on the hard disk, and move the fragments to one location to increase efficiency.
- **Diskpartitions** can divide an individual drive into multiple logical drives, each with its ownfile system which can be mounted by the operating system and treated as an individual drive.
- **Disk space analyzers** for the visualization of disk space usage by getting the size for each folder (including sub folders) & files in folder or drive. showing the distribution of the used space.
- **Diskstorage utilities**
- **Archive utilities** output a stream or a single file when provided with a directory or a set of files. Archive utilities, unlike archive suites, usually do not include compression or encryption capabilities. Some archive utilities may even have a separate un-archive utility for the reverse operation.
- **File managers** provide a convenient method of performing routine data management tasks, such as deleting, renaming, cataloging, uncataloging, moving, copying, merging, generating and modifying data sets.
- **Cryptographic utilities** encrypt and decrypt streams and files.
- **Hexeditors** directly modify the text or data of a file. These files could be data or an actual program.
- **Memorytesters** check for memory failures.

Network utilities analyze the computer's network connectivity, configure network settings, check data transfer or log events.

Registry cleaners clean and optimize the Windows registry by removing old registry

keys that are no longer in use.

Screensavers were desired to prevent phosphor burn-in on CRT and plasma computer monitors by blanking the screen or filling it with moving images or patterns when the computer is not in use. Contemporary screensavers are used primarily for entertainment or security.

System monitors for monitoring resources and performance in a computer system.

System profilers provide detailed information about the software installed and hardware attached to the computer

Software: A set of instructions/programs that tell the computer how to perform tasks.

Works with hardware (physical components) to process data into information.

Two main categories: System Software and Application Software.

2. Types of Software

System Software

Controls and manages hardware resources.

Provides a platform for application software.

Examples:

Operating Systems (OS): Windows, Linux, macOS.

Utility Programs: Antivirus, disk management, backup tools.

Device Drivers: Enable communication between hardware and OS.

Application Software

Programs designed for specific user tasks.

Examples:

Word processors (MS Word).

Spreadsheets (Excel).

Database systems (Oracle, MySQL).

Web browsers (Chrome, Edge).

Multimedia tools (Photoshop, VLC).

- Middleware

Acts as a bridge between applications and operating systems.

Example: Database middleware, API connectors.

3. Characteristics of Software

Intangible (cannot be touched).

Developed, not manufactured.

Can be reused and updated.

Subject to bugs/errors.

Requires compatibility with hardware.

4. Advantages of Software

Automates tasks and increases efficiency.

Enhances productivity (e.g., office applications).

Provides user-friendly interfaces.

Enables communication (email, internet apps).

Supports innovation (AI, data analysis, simulations).

5. Disadvantages of Software

Can contain bugs or vulnerabilities.

Needs regular updates and maintenance.

May require licensing (costly).

Dependent on hardware compatibility.

Risk of malware/viruses.

6. Examples of Software Categories

System Software: Windows OS, Linux, macOS.

Application Software: MS Office, Photoshop, Zoom.

Programming Software: Compilers, interpreters, IDEs (e.g., Visual Studio, Eclipse).

System Software = manages hardware.

Application Software = helps users perform tasks.

Programming Software = helps developers build applications.

DATA AND DATA TYPES

1. Definition of Data

Data: Raw facts, figures, or symbols without context (e.g., numbers, text, images).

Information: Processed/organized data that carries meaning and is useful for decision-making.

Key Point: Data → processed → becomes Information.

Data Type: A classification that specifies which kind of value a variable can hold and what operations can be performed on it.Major Categories of Data Types

Primitive (Basic) Data Types

Integer (int): Whole numbers (e.g., 5, -12).

Float/Double: Numbers with decimals (e.g., 3.14, -0.001).

Character (char): Single letter, digit, or symbol (e.g., 'A', '7').

Boolean: Logical values (True/False, 1/0).

Composite Data Types

Array: Collection of elements of the same type (e.g., list of integers).

String: Sequence of characters (e.g., "Hello").

Record/Struct: Grouping of different data types under one name.

Abstract Data Types (ADTs)

List: Ordered collection of elements.

Stack: Last-In-First-Out (LIFO) structure.

Queue: First-In-First-Out (FIFO) structure.

Tree/Graph: Hierarchical or networked data representation.

Specialized Data Types

Date/Time: Represents calendar dates and timestamps.

Pointer: Stores memory addresses.

Object (OOP): Encapsulation of data + methods.

3. Properties of Data Types

Size: Amount of memory required (e.g., int = 4 bytes).

Range: Minimum and maximum values allowed.

Operations: Defines valid operations (e.g., addition for integers, concatenation for strings).

Compatibility: Determines if data can be converted/cast into another type.

A data file is a collection of related records stored together in a computer system.

It acts as a container for data, enabling storage, retrieval, and manipulation.

Examples: Word documents, spreadsheets, databases, log files, multimedia files.

2. Types of Data Files

Text Files: Store plain text (e.g., .txt, .csv).

Human-readable.

Binary Files :Store data in machine-readable format (e.g., images, executables).

Not easily readable by humans.

Database Files: Structured storage for records (e.g., .db, SQL files).

Used in relational databases.

Multimedia Files: Store audio, video, and images (.mp3, .mp4, .jpg).

Log Files: System-generated files that record events or activities.

3. Properties of Data Files

Organized: Groups related information together.

Accessible: Can be retrieved by programs or users.

Persistent: Stored permanently until deleted.

Structured/Unstructured: May follow a format (database) or be free-form (text).

Portable: Can be transferred across systems.

4. Advantages

Easy storage and retrieval of large amounts of data.

Supports data sharing across applications.

Enables backup and recovery.

Facilitates analysis and decision-making.

Can be structured for efficiency (databases).

5. Disadvantages

Vulnerable to corruption or loss if not backed up.

May require specific software to access (e.g., database files).

Large files can consume storage space.

Security risks if files are not protected.

Unstructured files may be harder to analyze.

Data files = organized collections of data.

Types: text, binary, database, multimedia, log.Bit = smallest unit (0/1).

Byte = 8 bits, represents a character.

Multiples: KB, MB, GB, TB.

Applications:

Bits

for signals, bytes for storage & files.Bits:

Digital signals (on/off states).

Boolean logic (true/false).

Network transmission (bit rate).

Bytes:

File sizes (documents, images, videos).

Memory capacity (RAM, storage).

Character encoding (ASCII, Unicode).

Multiples of Bytes

Kilobyte (KB): 1,024 bytes.

Megabyte (MB): 1,024 KB ≈ 1 million bytes.

Gigabyte (GB): 1,024 MB ≈ 1 billion bytes.

Terabyte (TB): 1,024 GB ≈ 1 trillion bytes.

Petabyte (PB): 1,024 TB.

(Note: Some manufacturers use decimal multiples, e.g., 1 MB = 1,000,000 bytes.Here are clear, student-friendly notes on constructing random and sequential data files in computing.

Constructing Random and Sequential Data Files

Data files are structured ways of storing information on secondary storage (like hard disks, flash drives). The way records are arranged affects how fast they can be accessed.

There are two main methods of organizing data in files:

1. Sequential Data Files

A sequential file stores records one after another in a pre-defined order (usually alphabetical, numerical, or chronological).

How they are constructed

Records are written in sequence from the first to the last.

You must sort the records before putting them into the file.

Updating involves:

1. Reading the whole file.

2. Creating a new updated file.

3. Writing the new file in order again.

Characteristics

Simple to create.

Good when data is processed in batches.

Access is slow when searching because you start at the beginning and read sequentially.

Examples

Payroll files

Examination results sorted by index number

Bank monthly statements

Advantages

Easy to construct and maintain

Efficient for reading all records

Uses less storage overhead (no index needed)

Disadvantages

Slow for searching individual records

Updates require rewriting the whole file

2. Random (Direct Access) Data Files

A random file allows data to be stored and retrieved directly using a unique key (e.g., ID number).

Access to records is not sequential.

How they are constructed

Each record is assigned a key field.

A hash function or algorithm is used to calculate the exact storage location.

Records are placed in those calculated positions.

Collisions (two keys generating the same location) are handled using:

Overflow areas

Re-hashing

Chaining

Characteristics

Very fast access

Records can be read or updated directly

Good for systems that require instant responses

Examples

ATM systems

Library management systems

Student database accessed by admission number

Advantages

Very fast for searching and updating

No need to process the whole file

Disadvantages

More complex to design

Requires extra space for overflow handling

More sensitive to file corruption.

Disk Storage Fundamentals

Disk storage is organized into tracks (circular paths), sectors (smallest storage units), clusters (groups of sectors), surfaces (each side of a platter), and comes in different types (HDDs, SSDs, optical, etc.). Practically, you can check storage and file organization using tools like Windows File Explorer, Disk Management, or command-line utilities.

1. Tracks

- Concentric circles on a disk platter where data is magnetically recorded.
- Each platter surface has thousands of tracks.

2. Sectors

- Smallest physical storage unit on a disk.

- Traditionally 512 bytes, though modern drives often use 4 KB sectors.
- Each track is divided into multiple sectors.

3. Clusters

- A cluster is a group of sectors treated as one unit by the file system.
- Example: NTFS often uses 4 KB clusters (8 sectors of 512 bytes).
- Helps manage files efficiently, but small files may waste space due to cluster size.

4. Surfaces

- Each platter has two surfaces (top and bottom) that can store data.
- Read/write heads float above each surface to access tracks and sectors.

5. Types of Storage

- Magnetic HDDs: Use spinning platters, tracks, and sectors.
- SSDs: Use flash memory, no moving parts, faster access.
- Optical Disks (CD/DVD/Blu-ray): Use laser to read/write data.
- Hybrid Drives: Combine HDD and SSD features. Disk Storage Fundamentals

Disk storage is organized into tracks (circular paths), sectors (smallest storage units), clusters (groups of sectors), surfaces (each side of a platter), and comes in different types (HDDs, SSDs, optical, etc.). Practically, you can check storage and file organization using tools like Windows File Explorer, Disk Management, or command-line utilities.

1. Tracks

- Concentric circles on a disk platter where data is magnetically recorded.
- Each platter surface has thousands of tracks.

2. Sectors

- Smallest physical storage unit on a disk.
- Traditionally 512 bytes, though modern drives often use 4 KB sectors.
- Each track is divided into multiple sectors.

3. Clusters

- A cluster is a group of sectors treated as one unit by the file system.

- Example: NTFS often uses 4 KB clusters (8 sectors of 512 bytes).
- Helps manage files efficiently, but small files may waste space due to cluster size.

4. Surfaces

- Each platter has two surfaces (top and bottom) that can store data.
- Read/write heads float above each surface to access tracks and sectors.

5. Types of Storage

- Magnetic HDDs: Use spinning platters, tracks, and sectors.
- SSDs: Use flash memory, no moving parts, faster access.
- Optical Disks (CD/DVD/Blu-ray): Use laser to read/write data.
- Hybrid Drives: Combine HDD and SSD features.