STA 2102 — INFORMATION TECHNOLOGY FOR STATISTICS

Name: Daniel Kioma
Registration Number: SCM222-0784/2025

Topic 1: Introduction to Information Technology & Computers

1. Key Definitions

Information Technology (IT)
Information Technology refers to the integration of computer systems, software, networks, and digital tools to manage information efficiently. It encompasses everything involved in collecting, processing, storing, transmitting, securing, and analyzing data. In modern society, IT is the backbone of decision-making, automation, and innovation.

Data
Data are raw facts and observations collected from various sources. They may appear as numbers, text, audio, images, or measurements. On their own, data have no inherent meaning until processed.

Information
Information is processed and interpreted data. It is data that has been organized, structured, or analyzed in a way that adds

relevance, meaning, and value—allowing individuals or organizations to make informed decisions.

Information System (IS)

An information system is a coordinated framework consisting of hardware, software, data, people, procedures, and networks. Its purpose is to collect, store, process, and distribute information to support operations, planning, and decision-making.

2. Components of an Information System

1. Hardware

Physical devices used in computing — e.g., computers, servers, storage devices, scanners, sensors.

2. Software

Applications and programs that instruct hardware what to do. Includes system software (operating systems) and application software (e.g., Excel, SPSS).

3. Data

The raw materials of an information system. Well-organized data enables accurate analysis and reliable information output.

4. People

Users ranging from data entry clerks to system administrators, analysts,statisticians,
and decision-makers. People drive the system's objectives.

5. Procedures

Rules, policies, and step-by-step instructions governing how technology and data should be used safely and effectively.

6. Networks

Communication systems that connect computers and enable data sharing, remote access, and internet-based services.

3. Classification of Computers

Supercomputers

The fastest and most powerful machines. Used for complex scientific tasks such as climate modeling, genetic research, nuclear simulations, and large-scale statistical computations.

Mainframe Computers

Very large systems used by banks, government agencies, and airlines. They handle massive transaction processing, high-security operations, and simultaneous users.

Minicomputers / Servers

Serve medium to large organizations. They manage network resources, host databases, websites, and provide shared computing power to multiple users.

Microcomputers (Personal Computers)

Desktop and laptop computers commonly used for personal and professional tasks such as office applications, data analysis, programming, and internet access.

Mobile Computers

Smartphones, tablets, and handheld devices. They offer portability, internet connectivity, and powerful applications used even in field data collection.

4. Importance of IT to Statistics

Information Technology has transformed the field of Statistics by enabling:

- **Efficient Data Entry & Validation:**

Tools ensure accurate data capture and reduce human error.

- **Data Cleaning & Transformation:**

Software automates the removal of inconsistencies, missing values, and outliers, ensuring high-quality datasets.

- **Advanced Statistical Analysis:**

Programs like Excel, SPSS, R, Python, SAS allow statisticians to perform complex computations, modeling, forecasting, and machine learning.

- **Data Storage & Retrieval:**

Databases and cloud storage allow easy access, scalability, and secure management of vast datasets.

- **Visualization & Reporting:**

Graphs, dashboards, and automated reporting tools help present insights clearly for decision-makers.

- **Handling Big Data:**

Modern IT tools process extremely large datasets that would be impossible to analyze manually.

Topic 2: FUNDAMENTALS OF COMPUTER OPERATIONS — Enhanced Explanation

Computers entire operation is built around a simple and universal model known as the IPOS Cycle — Input, Processing, Output, and Storage. Everything a computer does fits into these four actions.

1. The IPOS Cycle (The Foundation of All Computing)

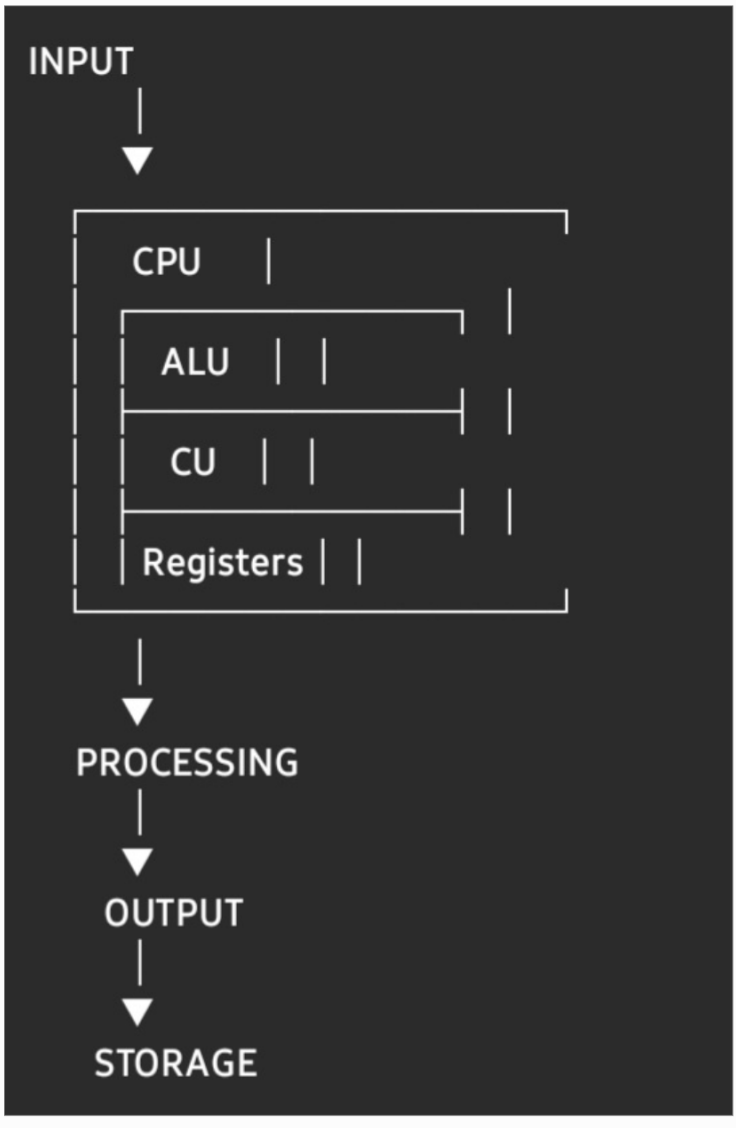
Input: This is how raw data enters the system — from a keyboard, mouse, scanner, microphone, etc.

Processing: The CPU transforms this data into useful information by carrying out instructions.

Output: Results are then shown to the user through screens, printers, or speakers.

Storage: Useful data is saved permanently on devices like SSDs or HDDs for future use.

Insight:
Every app, game, or system — from calculators to artificial intelligence — follows IPOS. It is the “heartbeat” of all computers.



2. The Central Processing Unit (CPU) — The Brain of Digital Life

The CPU is often called the brain of the computer because it handles all thinking tasks. It works through three tightly coordinated components:

a) Arithmetic and Logic Unit (ALU)
This part performs:

Arithmetic operations: addition, subtraction, multiplication, division

Logical operations: AND, OR, NOT, comparisons (e.g., >, <, =)
The ALU is what makes decisions and calculations possible.

b) Control Unit (CU)

Think of it as the CPU's "traffic controller." It:

Fetches instructions from memory

Decodes them

Sends signals to direct every other part of the computer
The CU ensures operations happen in the correct order — no confusion, no clashes.

c) Registers

Registers are ultra-fast storage cells inside the CPU. They temporarily hold the exact data and instructions the CPU needs right now.

Key registers include:

PC (Program Counter): Keeps track of the next instruction

IR (Instruction Register): Stores the current instruction

MAR (Memory Address Register): Holds the memory location being accessed

MDR (Memory Data Register): Holds the actual data being moved

ACC (Accumulator): Stores intermediate results

Insight:
Registers are faster than RAM and cache. They are the CPU's "instant-access toolbox."

3. The Fetch–Decode–Execute Cycle — The Pulse of the CPU

Every task a computer performs — from opening an app to solving equations — is done by repeating this cycle millions or billions of times per second.

Step 1: Fetch

PC sends the address to MAR

Memory sends the instruction to MDR

MDR transfers it to IR

PC moves to the next instruction

Step 2: Decode

The Control Unit analyzes the instruction and prepares the required components.

Step 3: Execute

The ALU performs the operation. Results may be stored in ACC, memory, or output devices.

Insight:
This cycle is continuous. Even while you're reading this, your device is running this cycle at extreme speeds.

4. Memory Hierarchy — Speed vs. Capacity

Computers use a layered memory structure:

1. Registers — fastest, smallest

2. Cache (L1, L2, L3) — very fast, small

3. RAM — fast, medium-sized

4. Secondary Storage (SSD/HDD) — slowest, largest

Insight:
The hierarchy ensures the CPU always has quick access to the data it needs without being slowed down by larger, slower memory.

5. The Operating System (OS) — The Manager of Everything

The OS is the system's "big boss." It manages:

- Memory allocation.
- Processor scheduling.
- File organization.
- Security and user authentication.
- Communication with devices (printers, cameras, keyboards)
- Interrupt handling.
- User interface.

Insight:
Without an OS, computers would be like a body without a nervous system — powerful but unable to coordinate.

6. Interrupts — Handling Events Instantly

An interrupt is a special signal demanding the CPU's immediate attention.

Process:

1. CPU pauses its current task

2. Saves its state

3. Executes the Interrupt Service Routine (ISR)

4. Returns to the previous task

Examples:

- Keyboard keystroke
- Mouse movement
- System errors
- Incoming notifications

Insight:
Interrupts allow computers to multitask smoothly and react quickly to user actions.

TOPIC 3: COMPUTER HARDWARE

Computer hardware refers to all the physical, touchable components of a computer system. These components work together to allow input, processing, storage, and output of information. Understanding them builds a strong foundation for anyone studying IT, statistics, or data science.

1. INPUT DEVICES

Input devices allow users to send data and commands into the computer.

Keyboard: The primary text-entry device; follows standard layouts like QWERTY. Essential for coding, data entry, and statistical software.

Mouse: A pointing device used for navigating graphical interfaces. Enables quick selection and manipulation of on-screen elements.

Scanner: Converts physical documents or images into digital formats—useful for data capture, archiving, and OCR (Optical Character Recognition).

Barcode Reader: Uses laser or imaging technology to read coded information. Widely used in supermarkets, warehouses, and research requiring fast data collection.

Joystick: A directional control device mainly used for gaming and simulations, including statistical modeling in virtual environments.

Touch Screen: Combines input and output. Users directly interact with the display—common in smartphones, ATMs, and modern laptops.

2. OUTPUT DEVICES

These devices allow the computer to communicate results or processed information back to the user.

Monitor: The main screen for visual output; ranges from LCD to LED and high-resolution panels for graphics and data visualization.

Printer: Produces hard copies of documents, reports, and graphs.

Types include inkjet, laser, and thermal printers.

Speakers: Output audio signals. Useful for multimedia, alerts, or educational tools.

Plotters: Specialized printers that draw high-precision diagrams, maps, and engineering designs—important for technical and scientific fields.

Projectors: Display computer output to a larger audience, useful in teaching, presentations, and collaborative data analysis.

3. STORAGE DEVICES

Storage devices preserve data either temporarily or permanently.

A. Primary Storage (directly accessible by the CPU)

RAM (Random Access Memory): Volatile memory used for temporary data storage while a program is running. More RAM = faster multitasking and processing.

ROM (Read-Only Memory): Non-volatile memory that stores essential instructions for startup (firmware).

Cache Memory: Extremely fast, small memory that stores frequently accessed data to speed up CPU operations.

B. Secondary Storage (long-term data storage)

Hard Disk Drive (HDD): Magnetic storage; large capacity but slower due to moving parts.

Solid State Drive (SSD): Flash-based storage; faster, more durable, and silent compared to HDD.

Flash Disk: Portable USB-based storage device for easy data transfer.

Memory Card: Compact flash storage used in phones, cameras, and portable devices.

4. MOTHERBOARD COMPONENTS

The motherboard is the central circuit board that connects all hardware components, allowing communication between them.

CPU Socket: The slot where the processor is installed; determines which CPU types the motherboard supports.

RAM Slots: Hold memory modules used for active processing tasks.

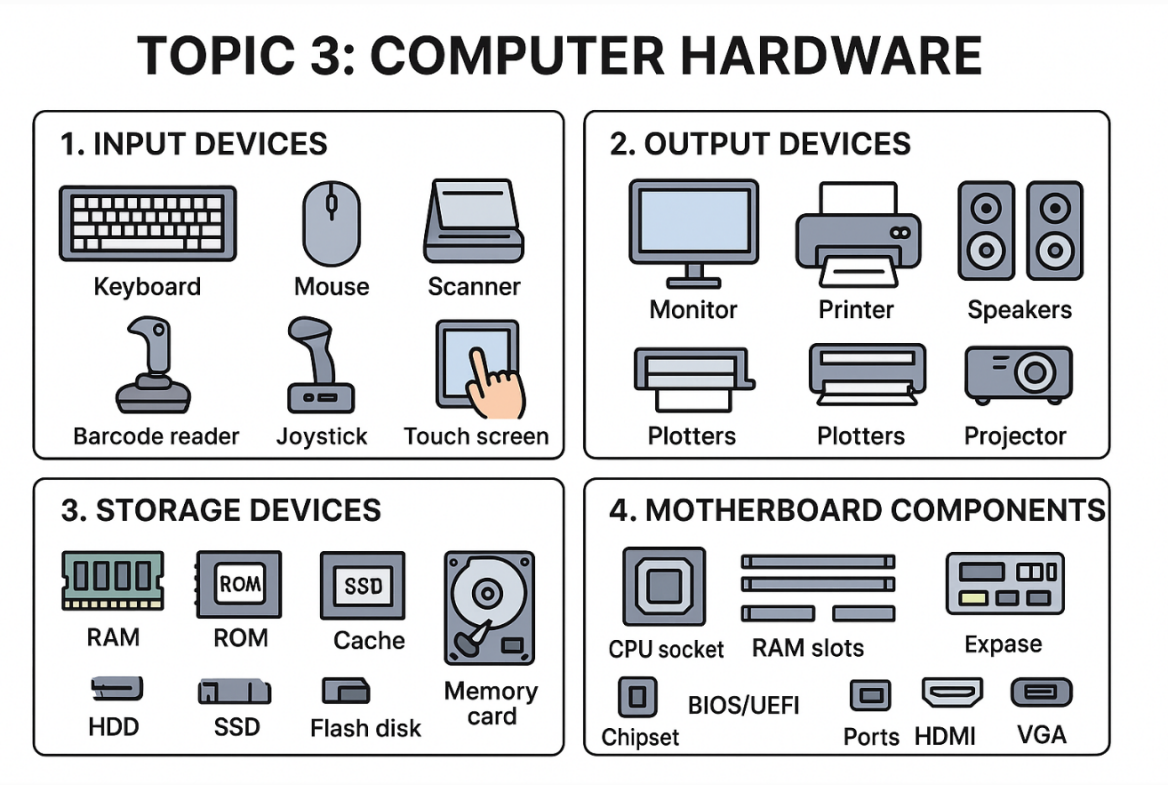
Expansion Slots (PCIe): Allow addition of extra components like graphics cards, network cards, or storage controllers.

BIOS/UEFI: Firmware that initializes hardware during startup and provides basic system control before the OS loads.

Chipset: The communication controller that manages data flow between the CPU, RAM, storage, and peripherals.

Ports (USB, HDMI, VGA, etc.): External connection points for devices such as keyboards, displays, or storage drives.

Power Connectors: Supply electrical power from the power supply unit to the motherboard and its components.

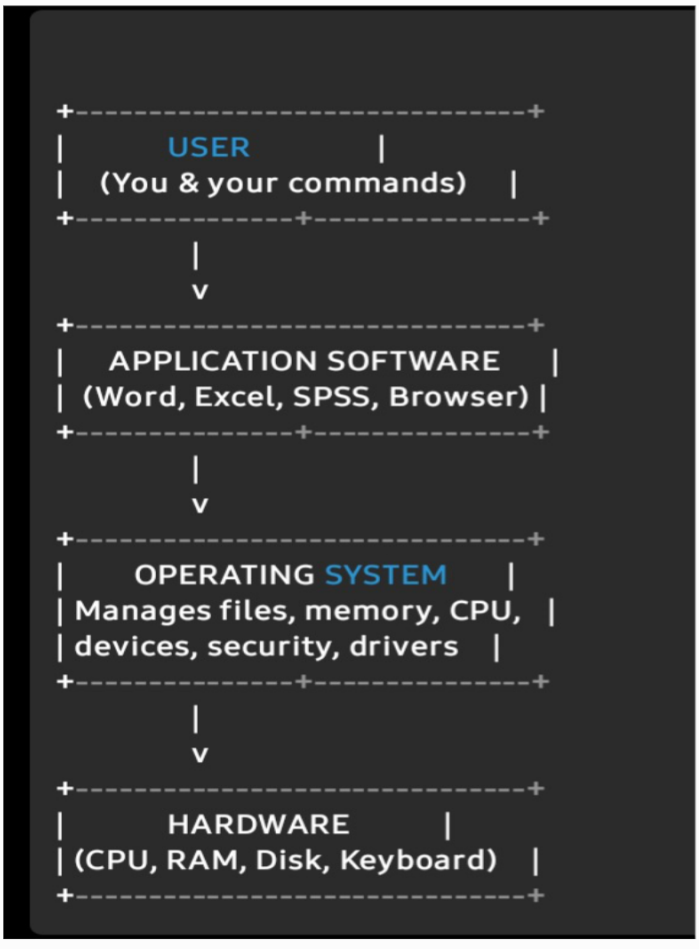


TOPIC 4: COMPUTER SOFTWARE BASICS — With Simple Diagrams

1. SYSTEM SOFTWARE

(a) Operating System (OS)

Diagram: How the OS Connects Users, Applications, and Hardware



Explanation:

The OS sits in the middle and makes sure that the hardware obeys commands from software and the user.
It is the “manager” or “controller” of all operations.

Key OS Functions

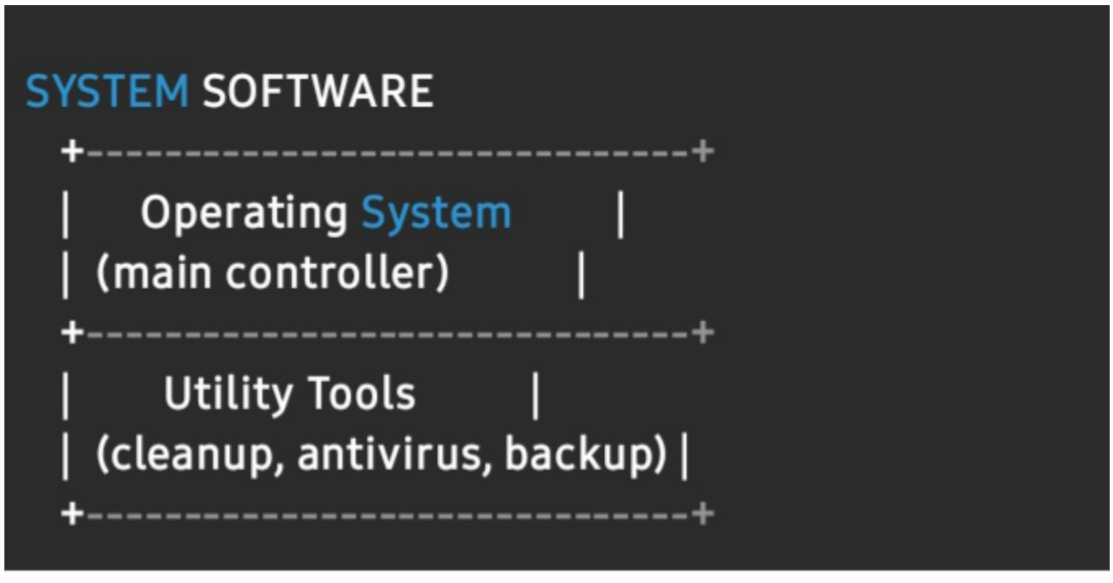
- Hardware Management – keeps all hardware components in sync.
- Memory Allocation – assigns RAM to programs and takes it back when done.
- File Management – create, delete, organize, protect files.
- Device Control – uses drivers to run printers, keyboards, screen, etc.

- GUI / CLI Interface – allows the user to communicate via a screen or command line.

- Security – manages passwords, permissions, user accounts.

(b) Utility Software

Diagram: Where Utilities Fit



Utility software acts like “maintenance tools”:

- Antivirus – protects from malware
- Disk Cleanup – removes junk
- Compression Tools – ZIP/RAR files
- Backup Utilities – save copies of important data

They help the OS keep the system healthy.

(c) Device Drivers

Diagram: Why Drivers Are Needed
USER --> Application --> OS --> Driver --> Hardware

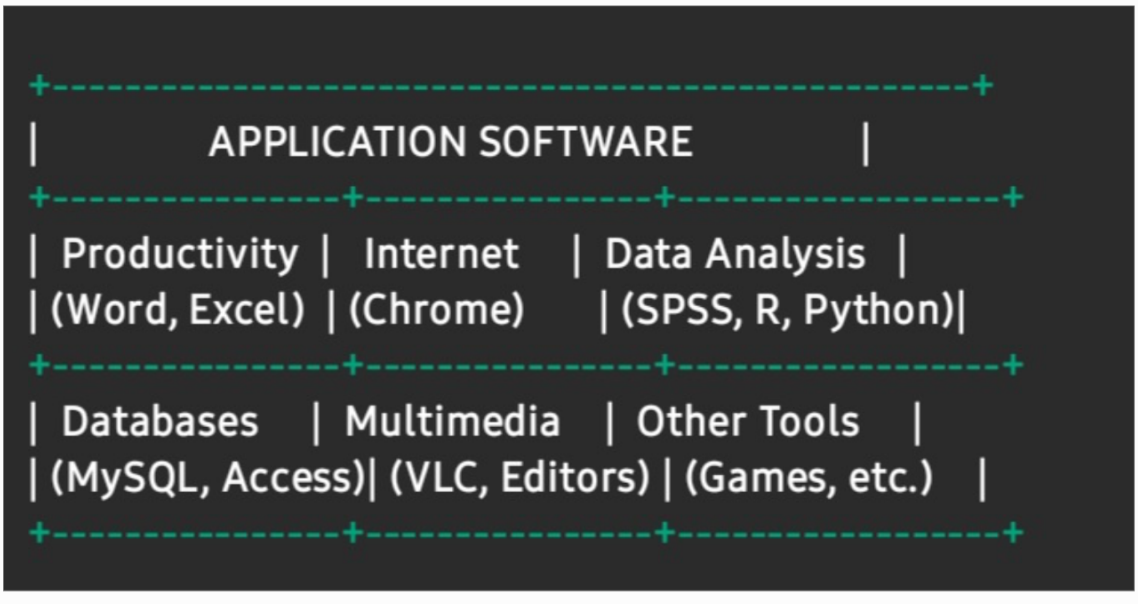
Without drivers, the OS would not understand how to operate devices like:

- printer
- graphics card
- scanner
- Wi-Fi adapter
- USB devices

Drivers act as translators between hardware and the operating system.

2. APPLICATION SOFTWARE

Diagram: Types of Application Software



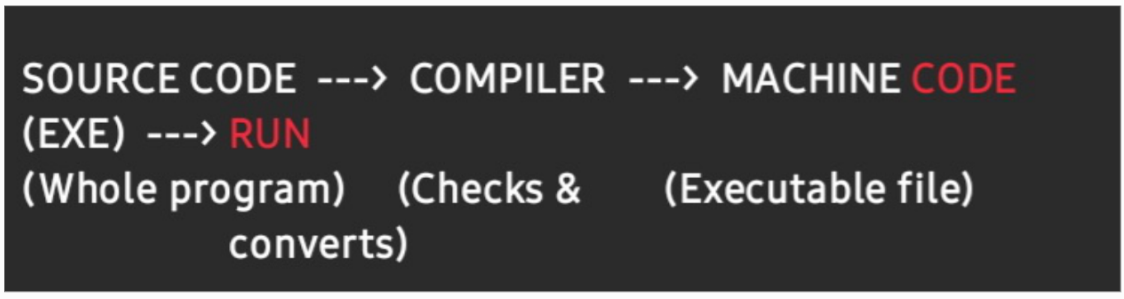
These programs help users complete specific tasks such as writing, analyzing data, browsing the web, or managing databases.

3. Compilers vs Interpreters

Programming languages are written in human-readable form. Computers, however, only understand machine code.

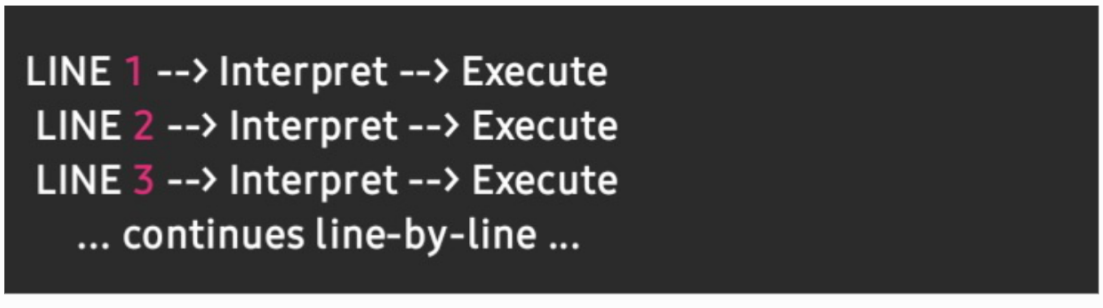
Translators convert code into machine-understandable instructions, and the two major types are:

Diagram 1: Compiler Workflow



- Translates entire program at once
- Executable runs fast
- Errors found after compilation
- Examples: C, C++, Java

Diagram 2: Interpreter Workflow



- Reads and executes each line one at a time
- Slower to run
- Easy for learning because errors appear immediately
- Examples: Python, R, JavaScript

TOPIC 5: DATA AND DATA FILES

1. Data Units

In computing, data is measured in hierarchical units. Each level represents a larger capacity than the one before it.

Bit (b)

The smallest unit of data—can only be a 0 or 1. Think of it as a light switch (ON/OFF).

Byte (B)

8 bits combined. A byte typically stores one character (e.g., “A”, “7”, “$”).

Kilobyte (KB)

1 KB = 1024 bytes. Small text files or short notes often fall in this range.

Megabyte (MB)

1 MB = 1024 KB. Good for storing songs, images, and medium documents.

Gigabyte (GB)

1 GB = 1024 MB. Modern software, videos, and apps are usually measured in GB.

Terabyte (TB)

1 TB = 1024 GB. Very large storage capacity—used for big datasets, servers, and backups.

› Why 1024 and not 1000?

Computers operate in binary, and $1024 = 2^{10}$, which naturally fits the digital system.

Data Unit	Size	Explanation
Bit (b)	Smallest unit	A single 0 or 1 (like an ON/OFF switch)
Byte (B)	8 bits	Often represents one character (A, B, $)
Kilobyte (KB)	1024 bytes	Small text files
Megabyte (MB)	1024 KB	Images, songs
Gigabyte (GB)	1024 MB	Videos, applications
Terabyte (TB)	1024 GB	Large backups, servers
Petabyte (PB)	1024 TB	Massive data archives
Exabyte (EB)	1024 PB	Huge-scale internet data
Quettabyte (QB)	1024 ZB	Largest official unit (introduced 2022)

2. File Organization Types

How data is arranged on storage affects how quickly and efficiently it can be accessed.

a) Sequential File Organization

Records are stored one after the other, in a fixed linear order.
To access a particular record, the computer reads from the beginning until it finds it.
Best suited for:

- Batch processing (e.g., payroll at end of month)
- Processes where data is accessed in a predictable order

Advantages:

- Simple and inexpensive
- Efficient for sequential tasks

Disadvantages:

- Slow when accessing specific records
- Updating or inserting new records is difficult

b) Random/Direct File Organization

Records are stored based on a calculated address (using hashing or indexing).
Provides instant access without scanning previous records.
Perfect for:

- Databases
- Systems requiring frequent updates and queries
- **Advantages:**
- Extremely fast access
- Easy to modify and update data

Disadvantages:

- Requires fixed-size records
- More complex storage management

3. File Formats

Different tasks require different file types. Understanding them helps you choose the most efficient format.

CSV (Comma-Separated Values)

Stores data in rows and columns using commas.
Lightweight, readable, and easy to import into statistical software.
Ideal for data analysis.

TXT (Text File)

Plain text with no formatting.
Good for logs, notes, and simple data.

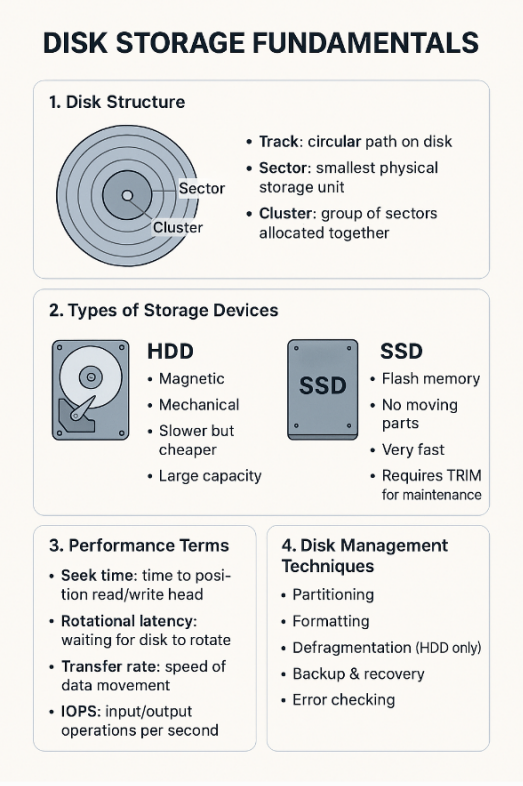
XLSX (Excel Workbook)

Spreadsheet format created by Microsoft Excel.
Supports formulas, charts, and multiple sheets.
Commonly used in data analysis and reporting.

SQL / Database Files

Used to store structured data in databases (MySQL, PostgreSQL, SQLite, etc.).
Supports indexing, queries, and relational operations.
Essential for large datasets and multi-user systems.

TOPIC 6: DISK STORAGE FUNDAMENTALS



1. Disk Structure — How Data Is Physically Organized

Modern storage (especially HDDs) arranges data in a structured pattern, similar to a city layout:

🔹 Tracks

Think of tracks like circular roads on the disk platter.
Each track stores a ring of data.

🔹 Sectors

Sectors are the smallest physical storage units on a disk, usually 512 bytes or 4 KB.
They are like individual houses along a road.

🔹 Clusters

Operating systems group several sectors into clusters.
Why? Because it's faster to read/write grouped sectors together.
A cluster is like a neighborhood made of several houses (sectors).

Why this matters:
If a file is larger than a cluster, it spreads across multiple clusters.
On HDDs, scattered clusters reduce speed. On SSDs, this matters less because there is no mechanical movement.

2. Types of Storage Devices — HDD vs SSD

🌀 HDD (Hard Disk Drive)

A traditional storage device that uses:

- Magnetic platters
- A spinning disk
- A moving read/write head

Pros:
✔ Cheaper
✔ Very large storage capacities

Cons:
✖ Slower (mechanical parts)
✖ More fragile
✖ Generates heat and noise

HDDs are often used for archives, backups, and bulk storage.

⚡ SSD (Solid State Drive)

SSD uses flash memory — similar to your smartphone storage.

Pros:
✔ Very fast
✔ No moving parts
✔ Energy efficient
✔ Quiet and durable

Cons:
✖ More expensive per gigabyte
✖ Requires TRIM to maintain performance (TRIM helps SSDs clean up unused memory blocks)

SSDs are ideal for operating systems, apps, and performance-critical tasks.

3. Performance Terms — How Fast Is Your Disk?

These metrics explain how quickly your storage device can access and transfer data:

⏱️ Seek Time

Time needed for HDD's read/write head to move to the correct track.
Important only for HDDs (SSDs have near-zero seek time).

🔄 Rotational Latency

The waiting time for the desired sector to rotate under the read/write head.
Again, only affects HDDs because SSDs have no moving parts.

🚚 Transfer Rate

Actual speed at which data is read/written.
Measured in MB/s or GB/s.
SSDs have much higher transfer rates.

⚙️ IOPS (Input/Output Operations per Second)

Measures how many read/write operations the device can handle

per second.
SSDs dominate this metric because they can access data instantly.

4. Disk Management Techniques

Proper disk management improves performance, longevity, and data safety.

📌 Partitioning

Dividing a disk into logical sections.
Useful for:

- Installing multiple operating systems
- Separating system files from personal data
- Better data organization

📌 Formatting

Prepares a storage device with a file system (e.g., NTFS, FAT32, exFAT).

📌 Defragmentation (HDD ONLY)

Reorganizes scattered file clusters so the disk head travels less.
SSDs should NOT be defragmented—they don't benefit and it reduces lifespan.

📌 Backup & Recovery

Regular backups protect you from:

- Hardware failure
- Human error
- Malware attacks

📌 Error Checking

Tools like chkdsk (Windows) detect and fix filesystem errors before they cause data loss.