Data Processing

Introduction to Data Processing

Data processing refers to the collection and manipulation of raw data to produce meaningful information. It is a systematic series of actions used to convert raw data into a useful format, typically involving stages such as input, processing, output, and storage. This chapter delves into the essential concepts of data processing, various system types, stages, activities, and the role of data representation.

Key Concepts in Data Processing

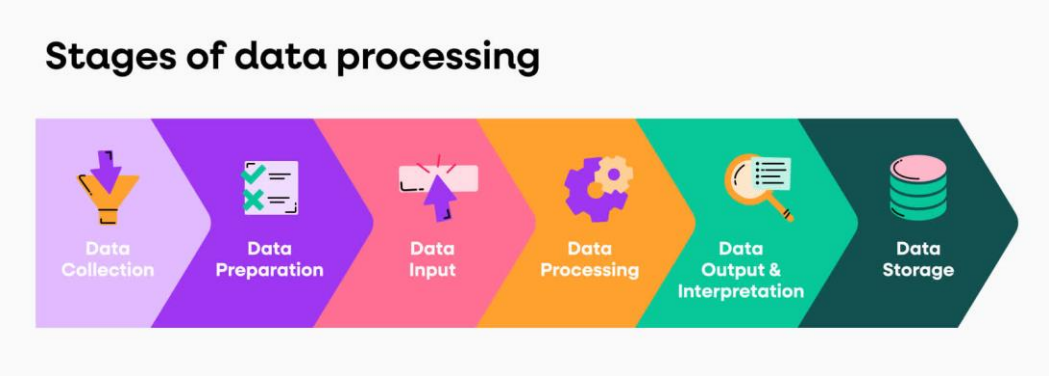
1. **Data**: Raw facts or figures that by themselves have no meaning. Data can take various forms, including numbers, text, images, and sounds.
2. **Information**: When data is processed in a meaningful way, it becomes information. Information is data that has been organized and interpreted to provide value.
3. **Data Processing Cycle**: The sequence of steps involved in transforming raw data into meaningful information.

Stages of Data Processing

The **data processing cycle** consists of the following steps:

1. **Data Collection (Input)**:
    - The first step involves gathering raw data from various sources, such as forms, sensors, or surveys.
    - The data must be accurate, complete, and reliable.
2. **Data Preparation**:
    - In this stage, the data is cleaned and organized to remove errors and ensure that it is ready for processing.
    - Data validation, sorting, and coding are part of this stage.
3. **Data Processing (Transformation)**:
    - The core step where the collected data is manipulated using various techniques or algorithms.
    - This stage involves calculations, comparisons, data sorting, filtering, or applying statistical methods to extract insights.
4. **Data Storage**:
    - Once processed, the data is stored for future use or further processing.
    - This could involve databases, cloud storage, or other storage systems.
5. **Data Output (Information Distribution)**:
    - The final step involves presenting the processed data in a meaningful format, such as reports, graphs, or charts, that can be used for decision-making.

- The output is delivered to end users or stakeholders through different channels, such as digital displays, printouts, or dashboards.



Types of Data Processing Systems

1. **Manual Data Processing**:
   - Involves human intervention at each step of data processing.
   - Suitable for small-scale tasks but time-consuming and prone to errors.
2. **Mechanical Data Processing**:
   - Involves the use of mechanical devices, such as calculators or typewriters, to process data.
   - It was a significant improvement over manual processing but is now outdated.
3. **Electronic Data Processing (EDP)**:
   - The most common type of data processing, where computers process the data automatically.
   - EDP systems use software and hardware to handle large volumes of data quickly and accurately.

Types of Data Processing Systems

1. **Batch Processing**:
   - **Definition**: In this system, data is collected over a period of time and processed as a batch at a scheduled time. There is no

interaction with the user during processing.
    - **Example**: Processing payroll at the end of each month for employees or generating monthly bank statements.
2. **Online Processing (Real-Time Processing)**:
    - **Definition**: Data is processed immediately as it is entered, often used in systems where fast, up-to-date information is critical. It requires continuous user interaction.
    - **Example**: ATM transactions or online flight booking systems where actions are processed in real-time.
3. **Distributed Processing**:
    - **Definition**: Involves using multiple computers across a network to process data. The processing tasks are distributed to different machines, allowing for more efficient resource use and faster processing.
    - **Example**: Cloud computing services such as Google Cloud or Amazon Web Services (AWS) distribute tasks to multiple servers.
4. **Multiprocessing**:
    - **Definition**: Uses multiple processors within a single system to perform tasks simultaneously. Each processor handles a portion of the overall task, making processing faster.
    - **Example**: Supercomputers or modern high-performance servers that handle complex scientific computations.
5. **Time-Sharing Processing**:
    - **Definition**: In this system, multiple users can access a single system simultaneously by sharing processor time. Each task is given a small amount of time, which creates the illusion that all tasks are being handled simultaneously.
    - **Example**: Multi-user systems like mainframe computers used by large organizations.
6. **Interactive Processing**:
    - **Definition**: Users interact with the system directly, and processing is done based on user inputs. The system responds to each user action in real-time or near real-time.
    - **Example**: Online transaction processing systems (OLTP) like e-commerce websites where users input details, and the system processes them immediately.
7. **Parallel Processing**:
    - **Definition**: Similar to multiprocessing but on a larger scale. Multiple processors or cores are used to run processes concurrently, often in systems designed for heavy computational tasks.
    - **Example**: Weather forecasting systems that use parallel processing to predict weather patterns by analyzing multiple data points at once.
8. **Cloud-Based Processing**:
    - **Definition**: Relies on cloud services to handle data processing. Data is processed remotely on a cloud provider’s infrastructure, providing flexibility and scalability.
    - **Example**: Google Docs or Dropbox, where users' data is stored and processed on cloud servers.
9. **Edge Computing**:

- **Definition**: Data processing occurs close to the location where the data is generated rather than at a centralized processing center or cloud. This reduces latency and speeds up the processing of data.
- **Example**: Smart devices in IoT (Internet of Things) systems, such as self-driving cars, where data must be processed locally in real-time.

10. **Mobile Data Processing**:
    - **Definition**: Refers to data processing done on mobile devices, such as smartphones or tablets. These systems often work in conjunction with cloud services but perform a significant amount of processing locally.
    - **Example**: Mobile banking apps that allow users to process transactions and access information directly from their phones.
11. **Network Data Processing**:
    - **Definition**: Data processing that takes place across a network, leveraging multiple systems to collaborate on a single processing task. This is common in systems where network resources are pooled together for efficiency.
    - **Example**: Distributed networks like blockchain, where data processing is decentralized across multiple nodes.
12. **Real-Time Embedded Systems**:
    - **Definition**: Systems that process data in real-time, embedded within devices that control hardware directly. These systems are usually designed for specific tasks and need to operate without delay.
    - **Example**: Embedded systems in devices like pacemakers, automotive control systems, or industrial robots.

Data Processing Activities

The following are the common activities involved in data processing:

1. **Data Collection**:
    - Gathering raw data from various sources such as sensors, surveys, or databases.
2. **Data Entry**:
    - Inputting the raw data into a computer system, often through data entry forms, spreadsheets, or direct input devices.
3. **Data Validation**:
    - Ensuring the data collected is accurate and consistent with set standards.
    - Example: Validating the format of email addresses or ensuring that numeric fields only contain numbers.
4. **Data Sorting**:
    - Organizing data into a particular order or sequence, such as ascending or descending.
5. **Data Filtering**:
    - Removing unwanted or unnecessary data to focus only on relevant information.
6. **Data Summarization**:
    - Reducing detailed data into a simpler, more concise form.

- Example: Summarizing monthly sales data into quarterly reports.

7. **Data Aggregation**:
   - Combining multiple data points into a single value or set of values.
   - Example: Calculating the average score from test results.

Data Representation

Data representation refers to the formats in which data is stored, processed, and transmitted. Some common forms of data representation include:

1. **Binary Representation**:
   - Computers store and process data in binary (0s and 1s). All types of data, including text, images, and sounds, are represented using binary numbers.
2. **Text Representation**:
   - Text data is typically represented using character encoding systems like ASCII (American Standard Code for Information Interchange) or Unicode.
3. **Image Representation**:
   - Images are stored as a collection of pixels, each represented by binary values indicating colors or shades.
4. **Data Formats**:
   - Data can be represented in different formats, such as CSV (Comma Separated Values), XML (Extensible Markup Language), or JSON (JavaScript Object Notation), depending on the type and purpose of the data.

Dec	Symbol	Binary	Dec	Symbol	Binary
65	A	0100 0001	83	S	0101 0011
66	B	0100 0010	84	T	0101 0100
67	C	0100 0011	85	U	0101 0101
68	D	0100 0100	86	V	0101 0110
69	E	0100 0101	87	W	0101 0111
70	F	0100 0110	88	X	0101 1000
71	G	0100 0111	89	Y	0101 1001
72	H	0100 1000	90	Z	0101 1010
73	I	0100 1001	91	[	0101 1011
74	J	0100 1010	92	\	0101 1100
75	K	0100 1011	93	]	0101 1101
76	L	0100 1100	94	^	0101 1110
77	M	0100 1101	95	_	0101 1111
78	N	0100 1110	96	`	0110 0000
79	O	0100 1111	97	a	0110 0001
80	P	0101 0000	98	b	0110 0010
81	Q	0101 0001	99	c	0110 0011
82	R	0101 0010	100	d	0110 0100

Database Concepts

1. **Database**:
   - A structured collection of data that can be easily accessed, managed, and updated. Databases store data in tables, consisting of rows and columns.
2. **Database Management System (DBMS)**:
   - Software that facilitates the creation, management, and manipulation of databases.
   - Examples include MySQL, Oracle, and Microsoft SQL Server.
3. **Relational Database**:
   - A type of database that organizes data into related tables using keys (primary and foreign keys).

- Relational databases are the most common type of database system in use today.

4. **Data Query**:
    - A query is a request to retrieve or manipulate data stored in a database.
    - SQL (Structured Query Language) is the most widely used language for querying databases.

Data processing plays a critical role in converting raw data into meaningful and actionable information. The stages of data processing, various system types, activities, and concepts such as data representation and database management form the foundation for handling data in today's digital world. Understanding these principles is crucial for effective data management, decision-making, and technological advancements.