Computer Science By SAMPAT LILER

These complete notes have been made for class 12th board computer science exam.

7. Understanding Data

1. Introduction to Data

Data is a fundamental component of information processing. It consists of raw facts, numbers, symbols, or observations that do not carry any specific meaning on their own. However, when data is structured and analyzed, it transforms into valuable information that aids decision-making. Data is generated through various means, such as online transactions, social media interactions, business records, and scientific research.

Organizations and individuals use data to understand patterns, predict trends, and optimize operations. The significance of data has grown with advancements in technology, leading to an era of data-driven decision-making. Fields such as Artificial Intelligence (AI), Machine Learning (ML), and Big Data Analytics heavily rely on vast datasets for training models and improving accuracy.

Examples:

A college applicant examines historical placement records, faculty credentials, and student testimonials to decide which institution to apply for.

A financial institution maintains extensive transaction logs to monitor customer spending habits, detect fraudulent activities, and personalize banking services.

---

2. Importance of Data

Data plays a crucial role in decision-making across multiple industries. Governments use census data to formulate policies, businesses analyze consumer data to improve services, and scientists utilize data for research and innovation. Without proper data analysis, decisions can be flawed or misleading.

Data-driven strategies allow businesses to understand customer behavior, predict market trends, and implement targeted marketing campaigns. Real-time data analysis in healthcare helps doctors diagnose diseases early, while meteorologists analyze weather data to predict storms and natural disasters.

Examples:

Meteorological departments analyze atmospheric and satellite data to predict natural calamities such as cyclones, hurricanes, and heatwaves.

Airlines use dynamic pricing models, adjusting ticket costs based on factors like demand, seasonality, and booking time.

---

3. Types of Data

Data can be classified based on its structure and organization. Understanding different types of data is essential for efficient processing and analysis.

A. Structured Data

Structured data is highly organized and follows a predefined format, such as databases, spreadsheets, and tables. Each piece of structured data is categorized into fields and records, making it easy to search, sort, and analyze. Relational databases like MySQL and PostgreSQL store structured data in tabular form with relationships between entities.

Examples:

A hospital's patient management system contains structured records, including Name, Age, Gender, Diagnosis, and Treatment Plan.

An e-commerce website maintains a database of products with attributes like Product ID, Name, Price, and Availability.

B. Unstructured Data

Unstructured data lacks a predefined format and is not easily categorized in traditional databases. This type of data includes multimedia content, emails, text documents, and social media posts. Due to its complexity, specialized tools such as Natural Language Processing (NLP) and Computer Vision are used to analyze unstructured data.

Examples:

Social media platforms process vast amounts of unstructured data, including posts, images, and comments.

Video surveillance systems generate large amounts of unstructured video footage that require AI-powered analysis for object detection and facial recognition.

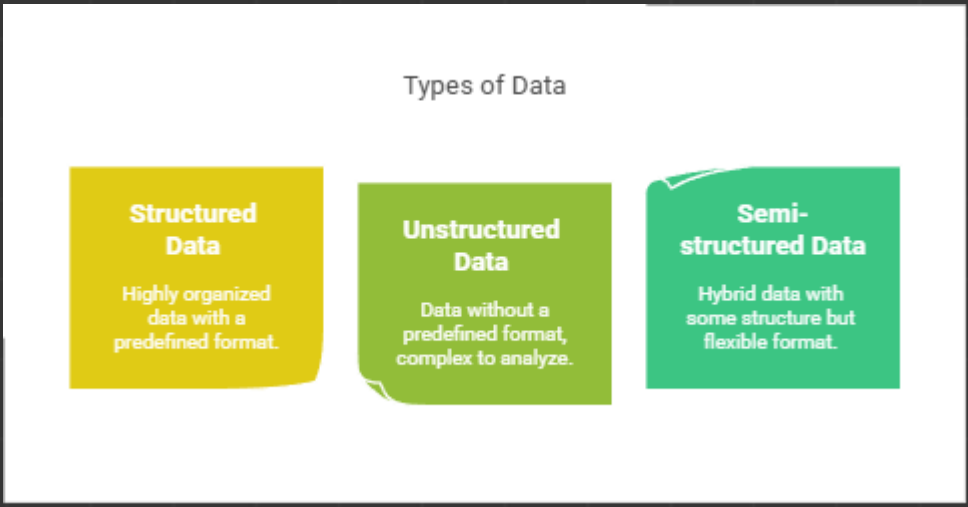
C. Semi-structured Data

Semi-structured data is a hybrid between structured and unstructured data. It has an underlying structure but does not follow a rigid schema. Examples include XML files, JSON objects, and log files. This type of data is widely used in web applications, APIs, and big data processing.

Examples:

A JSON file storing customer details in an e-commerce application, containing fields like Name, Address, and Purchase History.

System log files in a cloud-based service that store timestamped error messages and event logs.



---

4. Data Collection

Data collection involves gathering relevant information from different sources for analysis and decision-making. The quality of collected data determines the accuracy of insights derived. There are two main types of data collection methods:

Primary Data Collection: Data is collected directly from original sources through surveys, interviews, experiments, or observations.

Secondary Data Collection: Data is obtained from pre-existing records, research papers, government databases, and online sources.

Examples:

A market research firm conducts online surveys and focus group discussions to collect primary data on consumer preferences.

A grocery store records sales transactions using barcode scanners, creating a digital sales log that serves as a valuable dataset for trend analysis.

---

5. Data Storage

Once data is collected, it must be stored securely for future retrieval and analysis. Data storage options include:

Local Storage: Hard drives, SSDs, USB drives, and personal computers.

Cloud Storage: Remote servers such as Google Drive, Amazon S3, and Dropbox.

Databases: SQL and NoSQL databases that allow structured and unstructured data storage.

Examples:

Universities store student academic records in cloud-based database systems for easy access and backup.

Mobile applications store user preferences and browsing history on cloud servers to improve user experience.

---

6. Data Processing

Data processing transforms raw data into meaningful insights. The data processing cycle includes:

Input: Data is collected and stored.

Processing: Algorithms and statistical methods analyze the data.

Output: Meaningful insights are extracted and visualized.

Examples:

A bank's fraud detection system processes transaction data in real-time to identify suspicious activities.

An online shopping platform processes customer reviews and ratings to recommend products.

---

7. Statistical Techniques for Data Processing

A. Measures of Central Tendency

Mean: The arithmetic average of a set of values.

Example: If the scores of five students in a test are 78, 85, 90, 92, and 88, the mean is calculated as:

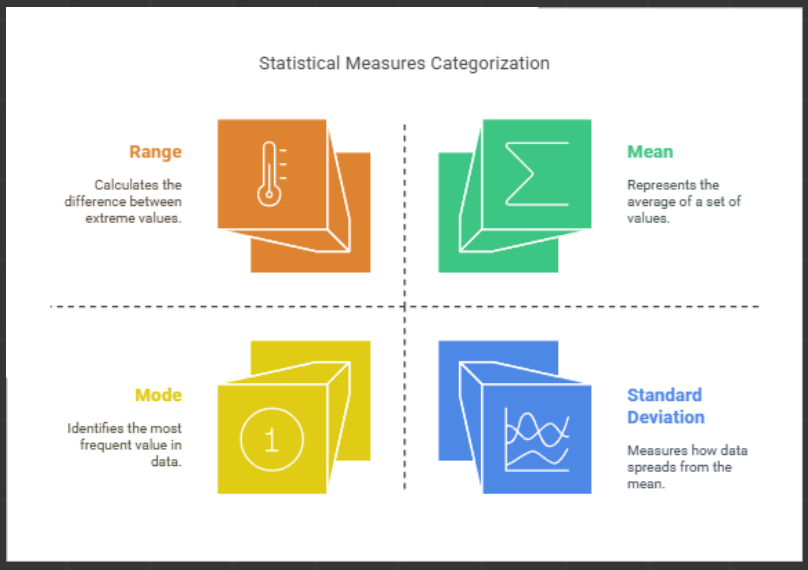
(78 + 85 + 90 + 92 + 88) / 5 = 86.6

Median: The middle value when data is sorted.

Example: If the sorted salaries of employees are 30K, 40K, 50K, 60K, 70K, the median is 50K.

Mode: The most frequently occurring value in a dataset.

Example: In the dataset {4, 5, 6, 6, 7, 8, 6, 9}, the mode is 6.



B. Measures of Variability

Range: The difference between the highest and lowest values.

Example: If the highest temperature in a week is 38°C and the lowest is 24°C, the range is 14°C.

Standard Deviation: A measure of data dispersion from the mean.

Example: If the heights of five people are 160 cm, 165 cm, 170 cm, 175 cm, and 180 cm, the standard deviation helps understand how much each height deviates from the average height.

8. Database Concepts

Introduction

Data management is essential for efficiently handling large amounts of information. Traditional manual record-keeping methods, such as maintaining physical registers, are inefficient and error-prone.

For example, in a school, attendance records of 50 students across 26 working days require 1,300 manual entries every month. Such large-scale data management becomes tedious and prone to inconsistency, loss, and errors.

Limitations of Manual Record-Keeping:

1. **Data Duplication & Inconsistency** – Manual entry often leads to errors in student names and roll numbers.
2. **Data Loss & Damage** – Registers can be misplaced or destroyed.
3. **Limited Access & Modification** – Searching and updating information manually is time-consuming.

Example 1: Consider a hospital maintaining patient records in files. Misplacing a file can lead to loss of critical medical history.

Example 2: A library with manual logs might record incorrect return dates, leading to disputes over fines.

File System

A **file system** is a method for storing and managing data in files on a computer. Each file can contain different types of information, such as text, images, or program code.

Limitations of File Systems:

(A) **Difficulty in Access** – Accessing data requires separate application programs.

(B) **Data Redundancy** – Same data appears in multiple files, wasting storage.

(C) **Data Inconsistency** – Duplicate data can lead to mismatches when updating.

(D) **Data Isolation** – Related files are not linked, making retrieval harder.

(E) **Data Dependence** – Changing file structure requires modifying programs.

(F) **Limited Data Sharing** – Users cannot set specific access permissions.

Example 1: A company using Excel sheets to store employee details faces difficulty in searching and updating records across multiple files.

Example 2: A retail store using separate files for inventory and billing struggles to keep stock levels updated.

Database Management System (DBMS)

A **DBMS** overcomes file system limitations by organizing data into a structured database. It allows users to create, store, manage, and retrieve data efficiently.

Advantages of DBMS:

1. **Data Integrity & Accuracy** – Reduces inconsistency by enforcing constraints.
2. **Efficient Data Retrieval** – Structured querying speeds up searches.
3. **Concurrent Access & Security** – Multiple users can access different data subsets.
4. **Data Independence** – Database structure changes do not affect application programs.

Example 1: A university uses MySQL to manage student records, courses, and faculty data centrally. **Example 2:** A banking system stores customer details in a secure Oracle database to prevent unauthorized access.

DBMS vs. File System (Comparison Table)

Feature	File System	DBMS
Data Storage	Individual files on disk	Organized tables in databases
Data Redundancy	High (duplicate entries)	Low (controlled constraints)
Security	Weak access control	Strong access control
Querying	Manual searches required	SQL queries for quick access
Multi-user Access	Limited	Concurrent multi-user support

Relational Data Model

The **Relational Model** organizes data in tables (relations) consisting of rows (tuples) and columns (attributes). Tables can be linked using keys.

Key Properties:

1. Attributes (Columns)

Definition:

Attributes are the **columns** in a table that define the type of data stored in a database. Each attribute represents a specific characteristic or property of the entity stored in the table.

Example:

Consider a table **Students** that stores student information. The attributes (columns) could be:

Student_ID	Name	Age	Course
101	Aman	18	B.Tech
102	Ravi	19	BCA
103	Priya	20	B.Sc

Here, **Student_ID, Name, Age, and Course** are attributes because they define different characteristics of a student.

---

2. Tuples (Rows)

Definition:

A tuple represents a **single record (row)** in a table. Each row contains specific values for all attributes.

Example:

Using the same **Students** table, each row is a **tuple**:

- **(101, Aman, 18, B.Tech)** → 1st tuple (row)
- **(102, Ravi, 19, BCA)** → 2nd tuple (row)
- **(103, Priya, 20, B.Sc)** → 3rd tuple (row)

Each row represents an individual student's record in the table.

---

3. Domains

Definition:

A domain is the **set of allowable values** that an attribute can take. Each attribute in a table has a domain that restricts the type of data that can be stored in it.

Example:

For the **Students** table:

- The domain of the **Student_ID** attribute is **positive integers** (e.g., {101, 102, 103, ...}).
- The domain of the **Age** attribute is **whole numbers between 18 and 25** (e.g., {18, 19, 20, 21, ..., 25}).
- The domain of the **Course** attribute is **a set of predefined course names** (e.g., {B.Tech, BCA, B.Sc, B.Com, ...}).

If an attribute value falls outside its domain (e.g., an age of **150** or a course named **"XYZ"**), it is considered **invalid data.**

---

4. Degree

Definition:

The **degree** of a table refers to the **number of attributes (columns)** present in that table.

Example:

In the **Students** table:

Student_ID	Name	Age	Course

There are **4 attributes** (columns), so the **degree of the table is 4.**

If another attribute, such as **Email**, is added, the degree becomes **5.**

---

5. Cardinality

Definition:

Cardinality refers to the **number of tuples (rows)** in a table.

Example:

In the **Students** table:

Student_ID	Name	Age	Course
101	Aman	18	B.Tech
102	Ravi	19	BCA
103	Priya	20	B.Sc

There are **3 rows (tuples)** in the table, so the **cardinality of the table is 3.**

If we add 2 more students, the cardinality becomes **5.**

Summary Table

Concept	Definition	Example in Students Table
Attributes (Columns)	Defines fields in a table	Student_ID, Name, Age, Course
Tuples (Rows)	Represents records	(101, Aman, 18, B.Tech)
Domains	Set of allowed values for an attribute	Age: {18-25}, Course: {B.Tech, BCA, B.Sc}
Degree	Number of attributes (columns)	4 (Student_ID, Name, Age, Course)
Cardinality	Number of tuples (rows)	3 (number of student records)

Example Table: STUDENT

RollNumber	Name	DateOfBirth	GuardianName	ContactNumber
101	John Doe	2003-05-15	Mark Doe	9876543210
102	Jane Smith	2002-02-28	Lisa Smith	8765432109
103	Ali Khan	2003-08-18	Ahmed Khan	7654321098

Keys in a Relational Database

Keys ensure data uniqueness and establish relationships between tables.

Types of Keys:

- **Candidate Key** – A set of attributes uniquely identifying tuples.
- **Primary Key** – The chosen candidate key ensuring uniqueness.
- **Composite Key** – A primary key consisting of multiple attributes.
- **Foreign Key** – Links a table to another table's primary key.

Example: Keys in Student Database

StudentID (PK)	Name	Class	GuardianID (FK)
201	Alex	12	501
202	Maria	11	502
203	Raj	12	503

Example 1: In a university database, **StudentID** serves as a primary key, while **CourseID** in the **ENROLLMENT** table acts as a foreign key.

Example 2: A travel agency database links **BOOKING** records to customers using **CustomerID** as a foreign key.

👉 Subscribe Youtube Channel - Anvira Education - YouTube

👉 Join Course - Https://Anviraeducation.Com/

👉 Follow Us On Facebook - Https://Www.Facebook.Com/Anviraedu

👉 Follow Us On Instagram - https://www.instagram.com/anvira_edu/

👉 Sampat Sir Instagram - https://www.instagram.com/writersampat/

👉 Join Our Telegram Channel - https://t.me/Anviraeducation20