# DATA

## The Core of the Digital Universe

# What is Data?

**Key Points:**

- Data = raw facts, numbers, or observations without context.
- It’s uninterpreted information — once processed → becomes information.
- Every digital system operates on data — sensors, websites, social media, AI, everything.

**Example:**

- “35°C” = data
- “Today’s temperature is 35°C, hotter than average” = information.

**Data → Information → Knowledge → Intelligence.**

# Characteristics of Data

- **Volume –** massive amounts generated every second.
- **Velocity –** the speed at which it’s created.
- **Variety –** text, video, image, audio, logs, sensor data.
- **Veracity –** how accurate/reliable it is.
- **Value –** usefulness after processing.

# Types of Data

- **Structured Data**
  - Highly organized; stored in tables or databases.
  - Rows/columns, easy to search using SQL.
  - Example: Employee database.
- **Unstructured Data**
  - No fixed format; difficult to store in traditional databases.
  - Examples: Emails, videos, chat messages, sensor readings.
- **Semi-Structured Data**
  - **Mix of both — has structure but not fixed schema.**
  - **Example: JSON, XML, API data.**

# Data Formats & Representations

- **Numeric**: integers, floats (sales, temperature).
- **Textual**: documents, transcripts, emails.
- **Visual**: images, videos.
- **Audio**: speech, recordings.
- **Coded**: JSON, XML, CSV.

# How Data is Generated & Flows

- **Generation** – sensors, users, applications.
- **Collection** – APIs, logs, transactions.
- **Storage** – databases, files, data lakes.
- **Processing** – cleaning, filtering, organizing.
- **Consumption** – analytics, AI, dashboards.

***Generate → Store → Process → Use → Refine.***

# Data Models & Structures

- **Relational (Tabular) –** used for structured data.
- **Hierarchical / Tree-based –** XML, file systems**.**
- **Graph-based –** entities connected via relationships (foundation for knowledge graphs).
- **Vector-based –** numerical representation used in AI (text/image embeddings).

# Modern Data Concepts

- **Vector Data**
  - Converts complex data (text, images) into numbers that capture meaning.
  - Used in AI for similarity search, NLP, and semantic understanding.
  - Example: “king - man + woman = queen” → numeric relationship.
- **Knowledge Graphs**
  - Web of connected data — entities (nodes) and relationships (edges).
  - Enables contextual understanding (e.g., Google’s Knowledge Graph).
- **Example:**
  - Sachin Tendulkar → plays_for → India
  - India → part_of → Asia

# Data Quality & Governance

- **Accuracy** → correct values.
- **Consistency** → same data across systems.
- **Completeness** → no missing pieces.
- **Security** → controlled access.

***“Good decisions need good data.”***

# The Evolution of Data Thinking

- **Traditional Era:** manual records, limited sources.
- **Digital Era:** databases, online systems.
- **Big Data Era**: massive unstructured data, real-time.
- **Intelligent Era:** vector + graph data powering AI.

***Data is not just numbers — it's the language of all digital systems.***

***Understanding its structure, flow, and foundation for analytics, AI, and beyond.***

# Follow Us

YouTube CodeNetra

WhatsApp