Data Parsing and Document Structure (Simple Notes)

1. What is Data Parsing?

Data Parsing is the process of reading raw data (like PDF, CSV, or text files) and converting it into a structured format that computers can understand.

Example:
Raw Text: CHAPTER 1 - Introduction to AI
After Parsing: Section = CHAPTER 1, Content = Introduction to AI

2. What is Document Structure?

Document Structure refers to the organization of a document. It includes headings, subheadings, paragraphs, and sections.

Example Structure:
• CHAPTER 1
• 1.1 Introduction
• 1.2 Applications

3. Relationship Between Them

Step 1: The document already has structure (headings and sections).
Step 2: Data parsing detects and extracts that structure.
Step 3: The structured data is stored in JSON or database format.

In short: Document Structure is how a document is organized. Data Parsing is the process of understanding and extracting that organization.