

# Unit II — Data Science, Big Data & Cloud Computing

Emerging Technologies for Business | BMB IT 02

## Unit II: Data Science, Big Data and Cloud Computing

**Syllabus:** Understanding data, information, data value chain and types; the data pipeline — collection, cleaning, storage, analysis, curation and visualization; Big Data — key concepts and business use cases; role of data science in business analytics and decision making; Cloud Computing — fundamentals, service models (IaaS, PaaS, SaaS), deployment models; cloud platforms for data storage, management and scalable analytics.

### Exam Blueprint (2025-26 paper)

#### YE QUESTION BAAR-BAAR AAYA HAI

Section B ke **paanch me se chaar questions** is unit se the — Data Pipeline, Big Data vs Database, Cloud service models & AWS, aur Data Science in decision making. Plus Section A me "Define cloud computing." **Ye unit sabse zyada marks deta hai.**

| Topic                           | 2025-26 paper | Marks |
|---------------------------------|---------------|-------|
| Data pipeline — 3 stages        | Section B     | 7     |
| Big Data vs Database            | Section B     | 7     |
| Cloud service models + AWS      | Section B     | 7     |
| Data Science in decision making | Section B     | 7     |
| Define cloud computing          | Section A     | 2     |

## 1. Understanding Data and Information

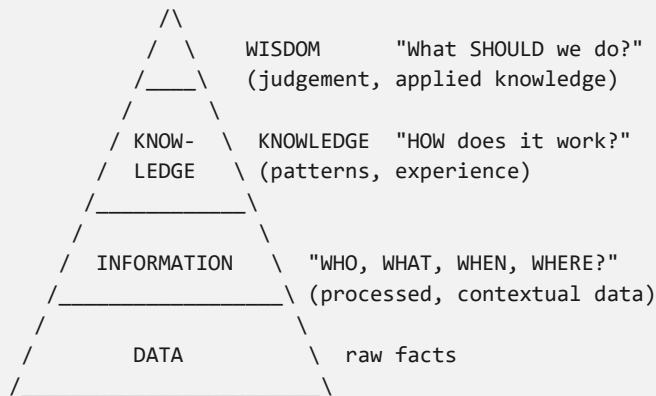
#### DEFINITION - RATTA MAARO

**Data** — raw, unprocessed facts and figures with no meaning by themselves. *Example:* , , .

#### DEFINITION - RATTA MAARO

**Information** — data that has been **processed, organised and given context**, making it meaningful and useful. *Example:* "Sales in Mumbai were 45 units on 12-05-2024."

## 1.1 The DIKW Pyramid



| Level              | Meaning                      | Example                             |
|--------------------|------------------------------|-------------------------------------|
| <b>Data</b>        | Raw facts                    | Sales figure: 45                    |
| <b>Information</b> | Processed and contextual     | Mumbai sold 45 units in May         |
| <b>Knowledge</b>   | Patterns and understanding   | Mumbai sales rise every summer      |
| <b>Wisdom</b>      | Applied judgement for action | Increase Mumbai stock before summer |

## 1.2 Types of Data

| Type                   | Description                                    | Examples                                           | Share of enterprise data |
|------------------------|------------------------------------------------|----------------------------------------------------|--------------------------|
| <b>Structured</b>      | Organised in a fixed schema — rows and columns | Relational database tables, spreadsheets           | ~20%                     |
| <b>Semi-structured</b> | Has tags/markers but no rigid schema           | JSON, XML, email, log files                        | small                    |
| <b>Unstructured</b>    | No predefined model                            | Text documents, images, audio, video, social posts | ~80%                     |

### Other classifications:

- **Qualitative** (descriptive, categorical) vs **Quantitative** (numeric, measurable)
- **Internal** (from the firm's own systems) vs **External** (market, social media, government)
- **Primary** (collected first-hand) vs **Secondary** (already collected by others)
- **Real-time / streaming** vs **Batch / historical**
- **Big Data** vs **Small Data**

## 1.3 The Data Value Chain

### DEFINITION - RATTA MAARO

The **data value chain** describes the sequence of activities that turn raw data into **business value**, with value added at each stage.

DATA ACQUISITION (collect, → DATA ANALYSIS (explore, → DATA CURATION (clean, → DATA STORAGE (organise, → DATA USAGE (decisions,