

Lecture 1. Part 1.

Data and Databases

Topics

1. The Importance and Classification of Data
2. Databases as a Component of Data Technologies

What is Data?

Literally: **Data** (from Latin) = “a given,” “a fact.”

ISO, 1996: Forms of information representation that are handled by information systems and their users.

ISO, 2015: A representation of information in a formalized manner suitable for communication, interpretation, or processing, capable of being interpreted in multiple ways.

Data are facts, text, graphics, images, sounds, and analog or digital video segments presented in a form suitable for storage, transmission, and processing.

Why Do We Need Data?

Traditional Business Applications

- Generating reports
- Accounting systems
- Descriptive analytics

Modern Applications

- Identifying relationships between events and outcomes (machine learning)
- Targeted communications
- Pattern recognition (e.g., facial recognition)
- Classification

Why do we need data?

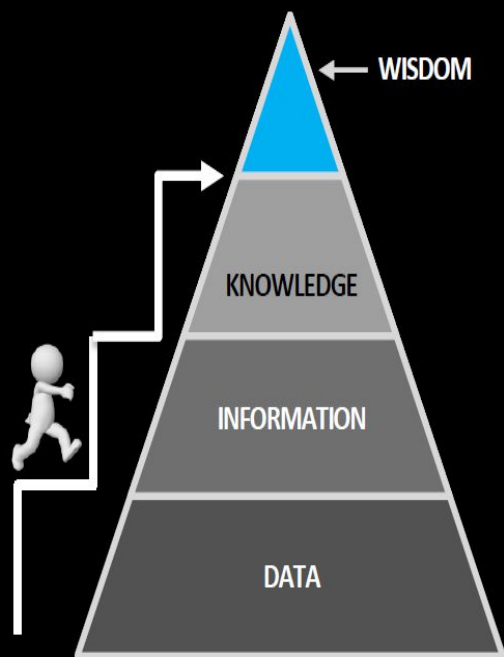
- From **data** **information**
- From **information** **knowledge**
- From **knowledge** **wisdom**

Knowledge is understanding *what* to do and *how* to do it at a given moment.

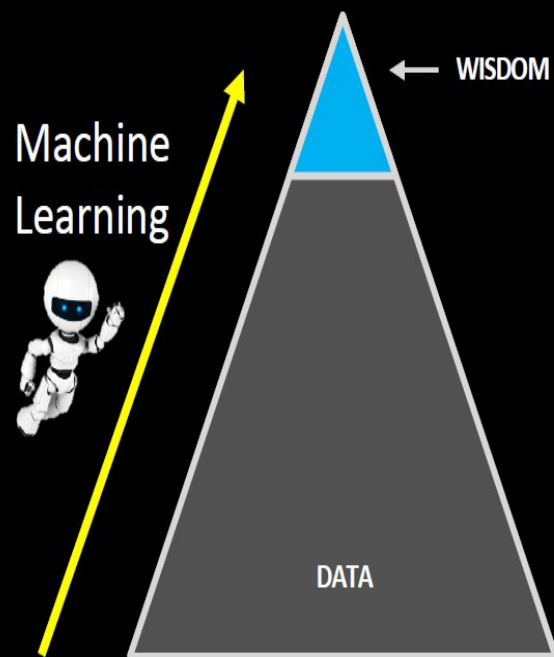
Wisdom is the ability to foresee possible future developments.

Data is a critical asset because knowledge can be automatically derived from it.

До Big Data и Deep Learning



После Big Data и Deep Learning



С ростом объёма данных и появлением технологий Big Data, сократился путь от данных к мудрости за счет развития технологий машинного обучения

Goal: Replacing Humans with Machines and AI

- **Mechanization:** Replacing human *physical labor* with machines.
- **Automation:** Replacing human *managerial or operational tasks* (the machine executes a predefined program or control algorithm).
- **Cybernetization:** Replacing humans in *decision-making* processes.

Data and **databases** are essential enablers of automation and cybernetization.

Data Is the New Oil

Data = raw material for creating value.

Different types of data serve different purposes.



Knowledge from Data Comes from Models and Analytics

Data Mining is the process of discovering meaningful patterns, correlations, and insights from large datasets using techniques from statistics, artificial intelligence, and database systems.



Core Tasks in Data Technologies

Data technologies enable organizations to collect, store, analyze, and act on data effectively.

The fundamental tasks include:

- **Ingest (Load)**
- **Store (Save)**
- **Process**
- **Make Decisions**

1. Ingest (Load)

This step involves acquiring raw data from various sources—such as sensors, user inputs, files, APIs, or external systems—and importing it into a processing or storage environment.

Examples:

- Uploading CSV files
- Streaming real-time sensor data
- Extracting data from web forms or mobile apps

Requires DBMS? **✗ Not necessarily**

Initial data ingestion may occur outside a DBMS (e.g., into data lakes or staging areas). However, loading data into a database for structured use does involve a DBMS.

2. Store (Save)

Data must be persistently and reliably stored for future access, ensuring integrity, security, and efficient retrieval.

Examples:

- Saving customer records
- Archiving transaction logs
- Storing configuration settings

Requires DBMS? ✓ Yes

Structured, scalable, and secure storage of operational data almost always relies on a Database Management System (DBMS)—relational (e.g., PostgreSQL, MySQL) or non-relational (e.g., MongoDB).

3. Process

Processing includes transforming, cleaning, aggregating, enriching, or analyzing data to make it useful for applications or decision-making.

Examples:

- Calculating monthly sales totals
- Removing duplicates or correcting errors
- Joining data from multiple tables

Requires DBMS? ✓ Often

While some processing can happen in application code or analytics platforms, complex queries, joins, and transactions are typically handled by a DBMS—especially in OLTP and OLAP systems.

4. Make decision

Using processed data and derived insights to support or automate human or system decisions.

Examples:

- Approving a loan based on credit score
- Recommending products to users
- Triggering alerts in fraud detection

Requires DBMS? ⚠ Indirectly

Decision-making itself is usually performed by business logic, AI models, or dashboards—but these rely heavily on data retrieved and prepared by a DBMS. Real-time decisions often query databases directly.

What Is the Role of Databases and DBMS?

Data Processing: From Raw Material to Final Product

Databases are tools—means to achieve an end result.

Just as oil fields, drilling rigs, pumps, and refineries are essential in producing petroleum products, databases are critical infrastructure in transforming raw data into valuable information and decisions.

Professions That Require DBMS Knowledge

- Software Developer
- Quality Assurance (QA) Engineer / Tester
- Data Analyst
- Machine Learning Engineer / Data Engineer
- DevOps Engineer
- Project Manager
- Technical Support Specialist

(Note: All these roles often interact with databases—whether to build, test, analyze, deploy, manage, or support data-driven systems.)