

25 YEARS ANNIVERSARY  
SCT

ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

# IT3090E - Databases

## Chapter 1: Introduction

Muriel VISANI

[murielv@soict.hust.edu.vn](mailto:murielv@soict.hust.edu.vn)

# Contents

- **Chapter 1: Introduction**
- Chapter 2: Relational databases
- Chapter 3: Relational algebra
- Chapter 4: Structured Query Language (SQL)
- Chapter 5: Database Design
- Chapter 6: Indexing
- Chapter 7: Query processing and optimization
- Chapter 8: Constraints, rules and triggers
- Chapter 9: Security
- *(Optional) Chapter 10: Transactions: concurrency and recovery*

# Learning points

- 1. Course overview
- 2. Basic concepts on database
- 3. Data management

# Learning objectives

- Upon completion of this lesson, students will be able to:
  - Recall the concepts of database, DBMS, data model, file system.
  - Identify the characteristics of database and file system approach in data management
  - Know about the different types of database approaches available nowadays

# Keywords and descriptions

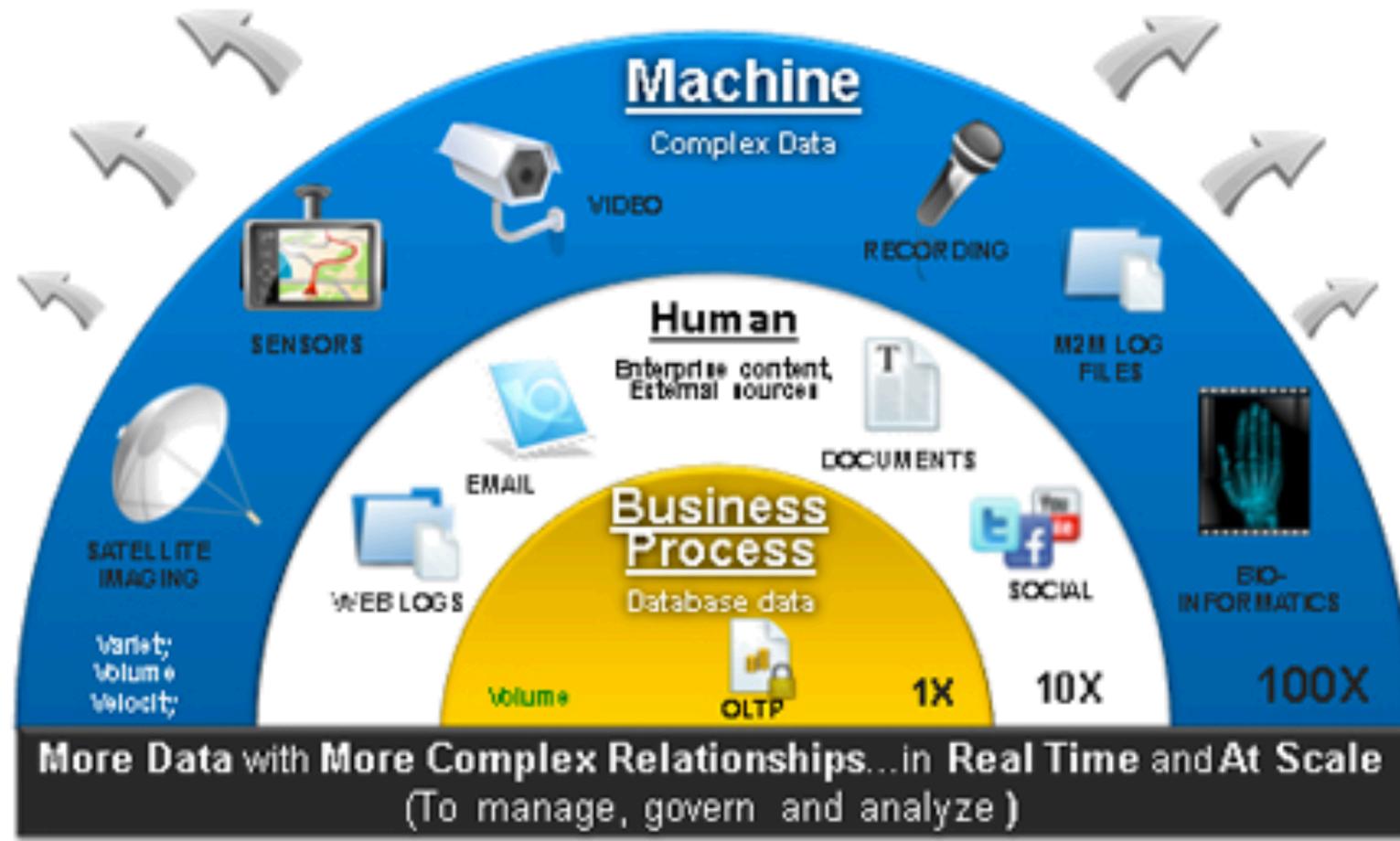
Keyword	Description
Database	A shared collection of related data designed to meet the information needs of an organization
DBMS	A software to facilitate the creation, maintenance and querying of a database
Data model	A set of concepts used to describe the structure of a database: data types, relationships, constraints, semantics, ...
Schema	A data structure organized according to the overall design of the database
Instance	The data itself (in the context of the data model and database schema)

# 1. Introduction

- Major research field with long history (since the creation of computers)
- More than 80% applications use databases
  - Including, of course, web applications and web-pages with dynamic contents (agoda.com...)
- Hot jobs in startups, big enterprises
- Massive Industry: Oracle, IBM, Microsoft, Google, AWS

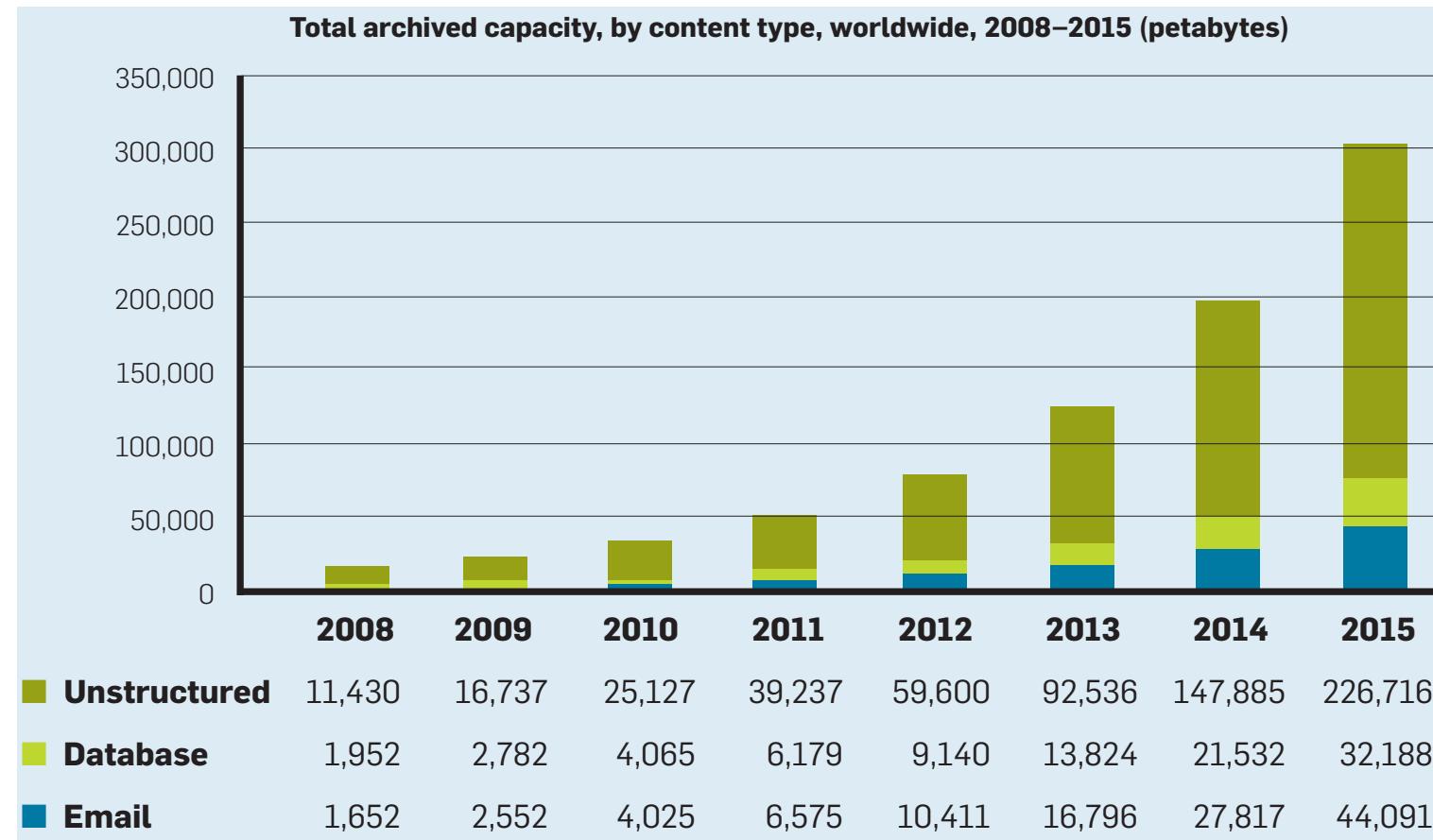
# 1. Introduction

## ➤ What type of data in our digital universe?



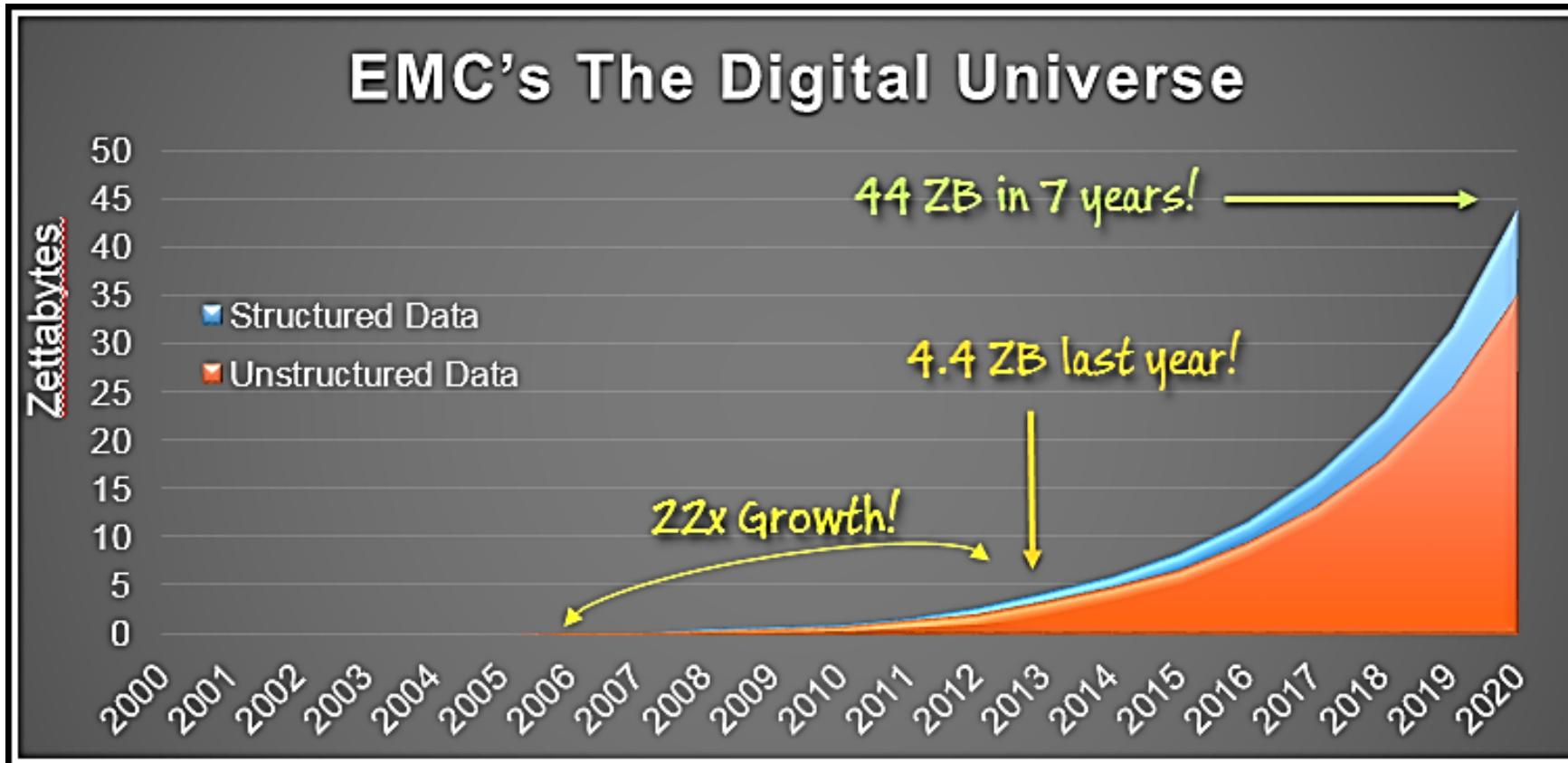
# 1. Introduction

## ➤ How big is our digital universe?



# 1. Introduction

## ➤ How big is our digital universe?



Source: <http://hadoopsie.com/is-hadoop-better-than-sql-server/>

# 1. Introduction

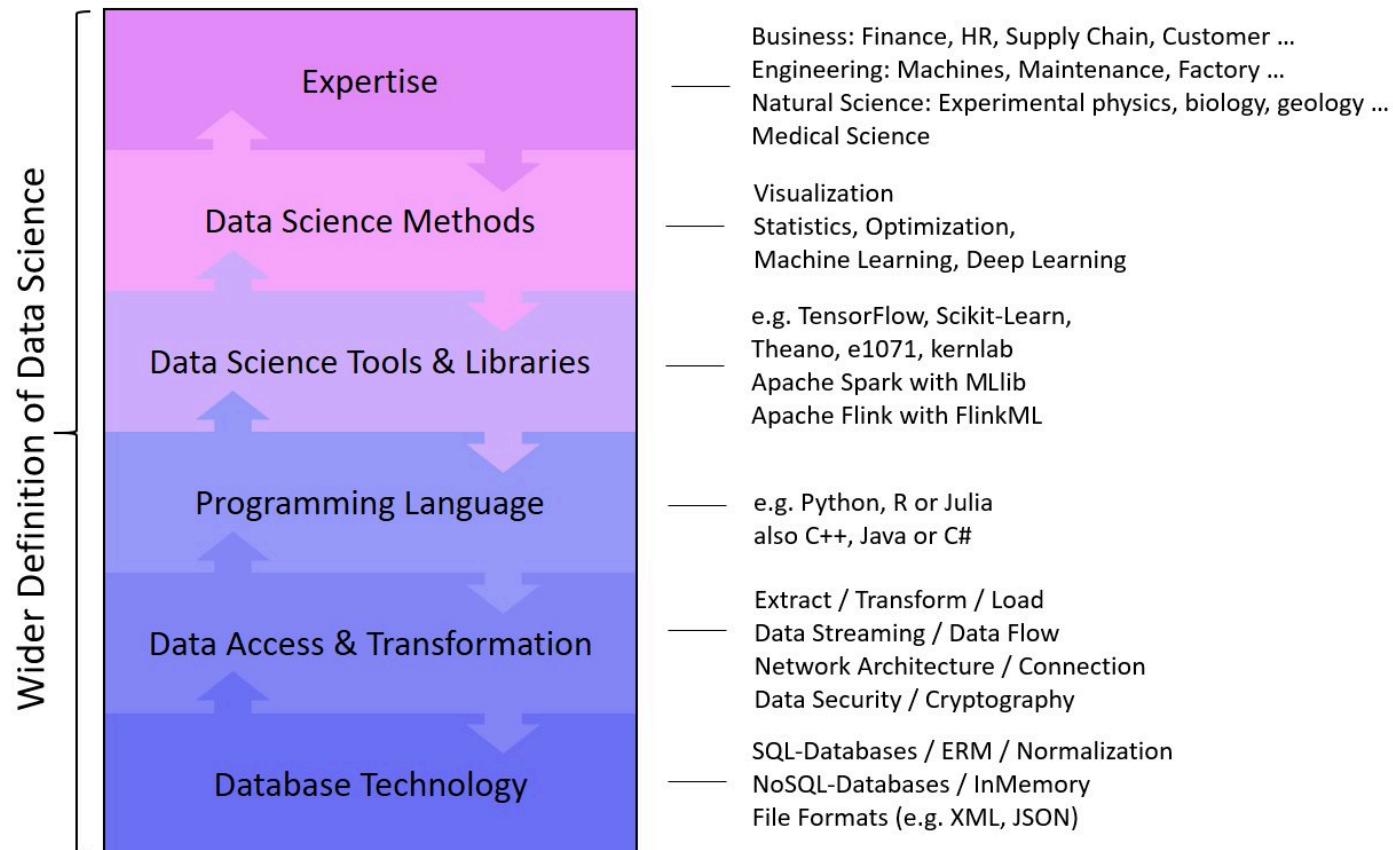
- Data is the new oil of the 21st century



Source: <https://www.economist.com>

# 1. Introduction

- Data science knowledge stack



Source: <https://data-science-blog.com>

## 2. Basic concepts

- 2.1. Data
- 2.2. Database
- 2.3. Data model vs. schema vs. instance
- 2.4. Database management system (DBMS)
- 2.5. Database environment
- 2.6. Database users

## 2.1. Data

- Definitions

<b>Wikipedia</b>	Data is any sequence of one or more symbols given meaning by specific act(s) of interpretation.
<b>Businessdictionary.com</b>	Information in raw or unorganized form (e.g. alphabets, numbers, or symbols) that refer to, or represent, conditions, ideas, or objects. Data is limitless and present everywhere in the universe

- E.g. A specific student data: ID, Name, Age, Gender, Address,...

## 2.2. Database

- Definitions

<b>Wikipedia</b>	Database is a shared collection of related data designed to meet the information needs of an organization
<b>Intro to CS</b>	A database is a collection of information that is organized so that it can be easily accessed, managed and updated

## 2.2. Database

- Logically coherent
- Internally consistent
- Designed for a specific set of purposes
- Independent of a specific software
- Representation of the real world
  - Entities (e.g., Students, Courses)
  - Relationships (e.g., Tam is enrolled in C++)

## 2.2. Database

### [Example] A course management system

#### Entities

- Students
- Courses
- Teachers

#### Relationships

- Students take in some courses
- Course are given by some teachers

## 2.3. Model vs. Schema vs. Instance

<b>Data Model</b>	<ul style="list-style-type: none"><li>• Set of concepts used to describe the structure of a database: data types, relationships, constraints, semantics,</li><li>• Tool for data abstraction</li></ul>
<b>Schema</b>	<ul style="list-style-type: none"><li>• A data structure organized according to the overall design of the database</li></ul>
<b>Instance</b>	<ul style="list-style-type: none"><li>• Data itself</li></ul>

## 2.3. Model vs. Schema vs. Instance

**Data Model**

```
type <type_name> = record  
    <field_name> : <data_type>;  
    <field_name> : <data_type>;  
    ...  
end;
```

**Schema**

```
type student = record  
    ID : string;  
    fullName: string;  
    Birthday: date;  
    Address: string ;  
    Class: string;  
end;
```

**Instance**

( « Stud001 », « Nguyen », 1/4/1983, « 1 Dai Co Viet », « 1F  
VN K50 »)

## 2.4. Database Management System (DBMS)

- Definitions

Wikipedia	A software to facilitate the creation and maintenance of a database
Techtarget	The DBMS provides users and programmers with a systematic way to create, retrieve, update and manage data



## 2.4. Database Management System (DBMS)

“More than 80 % of real world computer applications  
are associated with databases”

\* Korth & Silberschatz. Database System Concepts.



## 2.4. Database Management System (DBMS)

- Defining ~ specifying types of data
- Constructing ~ storing & populating (inserting)
- Manipulating ~ querying, deleting/updating, reporting

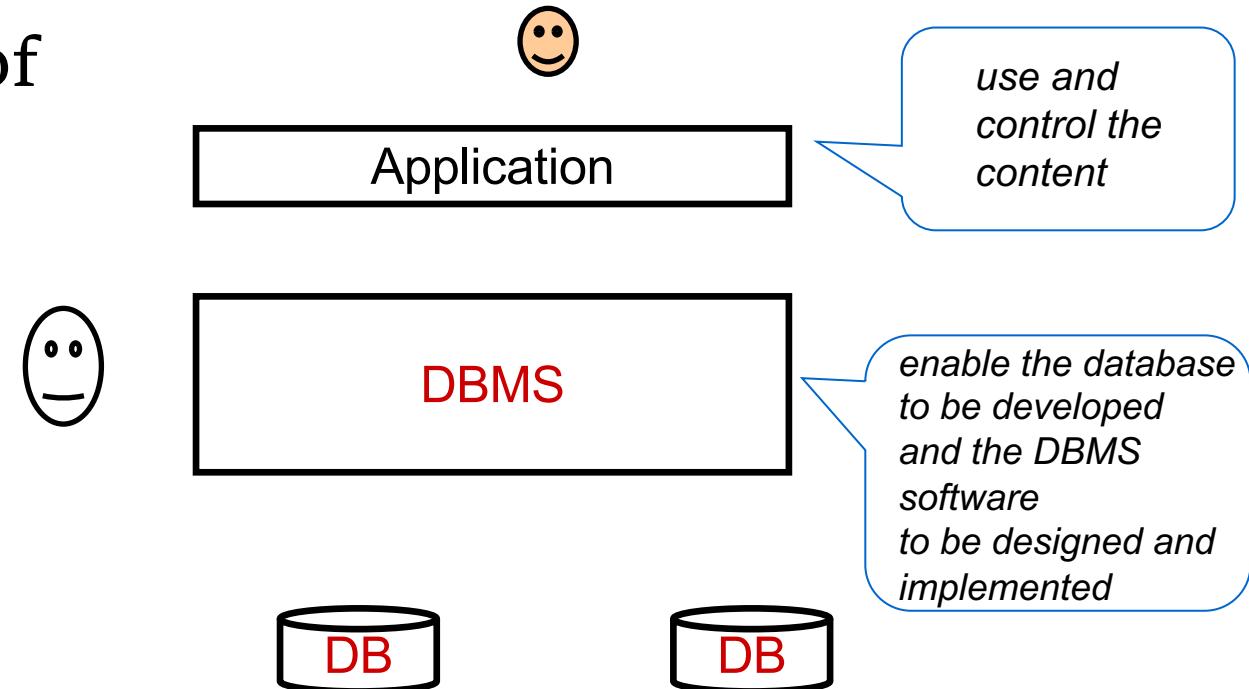


## 2.4. Database Management System (DBMS)

- Serves as intermediate between the data and the users
- The user does not need to know how the data is:
  - Stored physically in the machine
  - Queried internally
  - Modified without the risk of conflicts with other users
- ... But the user still can get to know about it, if he/she wants
- Provides tools to interact with data in "natural" language

## 2.5. Database Environment

- A database environment is a collective system of components that regulates the management, the use of data, and the data itself
- Hardware
- Software
- Data
- Users
- Procedures/Manuals



## 2.6. Database Users

- There are three main types of database users

### 1. Database administrators

- Authorize/regulate access to the database for other users
- Co-ordinate and monitor the database's use
- Acquire software, and hardware resources, control its use and monitor the efficiency of operations.

## 2.6. Database Users

### 2. Database Designers

- define the content, the structure, the constraints, and functions or transactions against the database.
- communicate with the end-users and understand their needs.

### 3. End-users

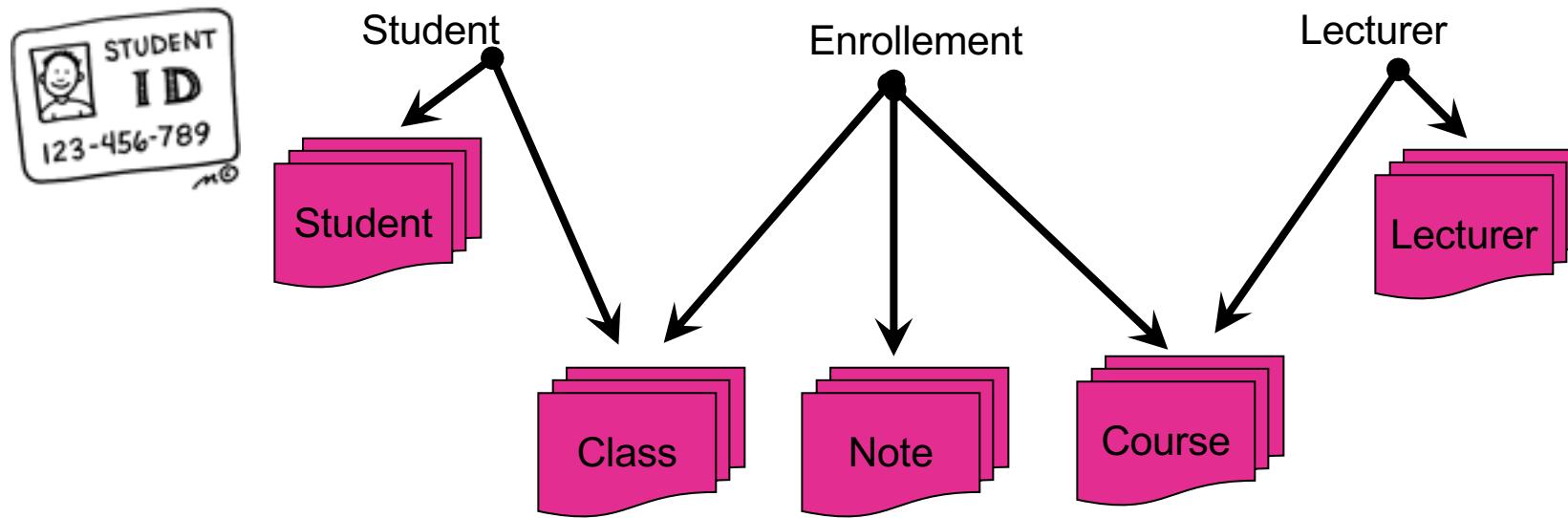
- use the data for queries, reports and some of them actually update the database content
- Different types of end-users
  - Casual end users
  - Naive users
  - Advanced end users

### 3. Data management

- 3.1. File management system approach
- 3.2. Database management system approach

### 3.1. File management system approach

- For instance, using Excel files...

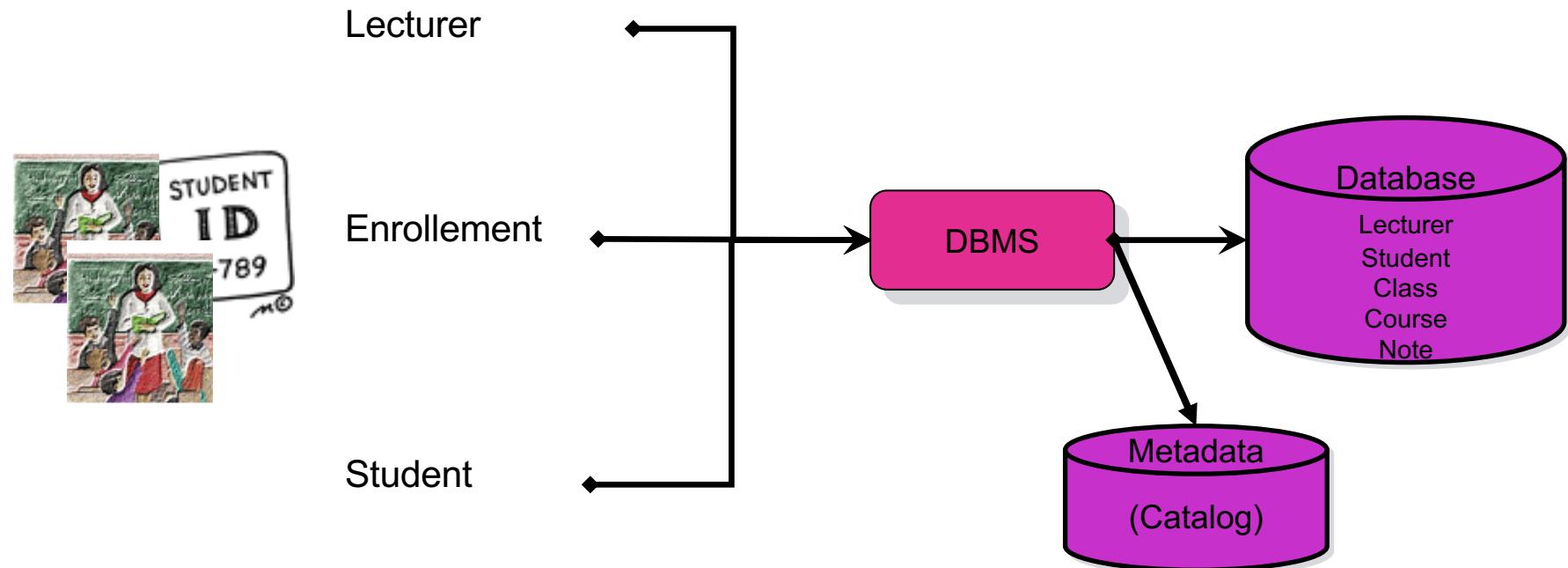


### 3.1. File management system approach

- Limitations

- Uncontrolled redundancy
- Inconsistent data
- Inflexibility
- Limited data sharing
- Poor enforcement of standards
- Low programmer productivity
- Excessive program maintenance
- Excessive data maintenance

## 3.2. Database approach



## 3.2. Database approach

- Controlled redundancy: consistency of data & integrity constraints
- Integration of data: self-contained & represents semantics of application
- Data and operation sharing: multiple interfaces
- Flexibility: data independence, data accessibility, reduced program maintenance
- Services & Controls
  - Security & privacy controls
  - backup & recovery
  - enforcement of standards
- Ease of application development

## 3.2. Database approach

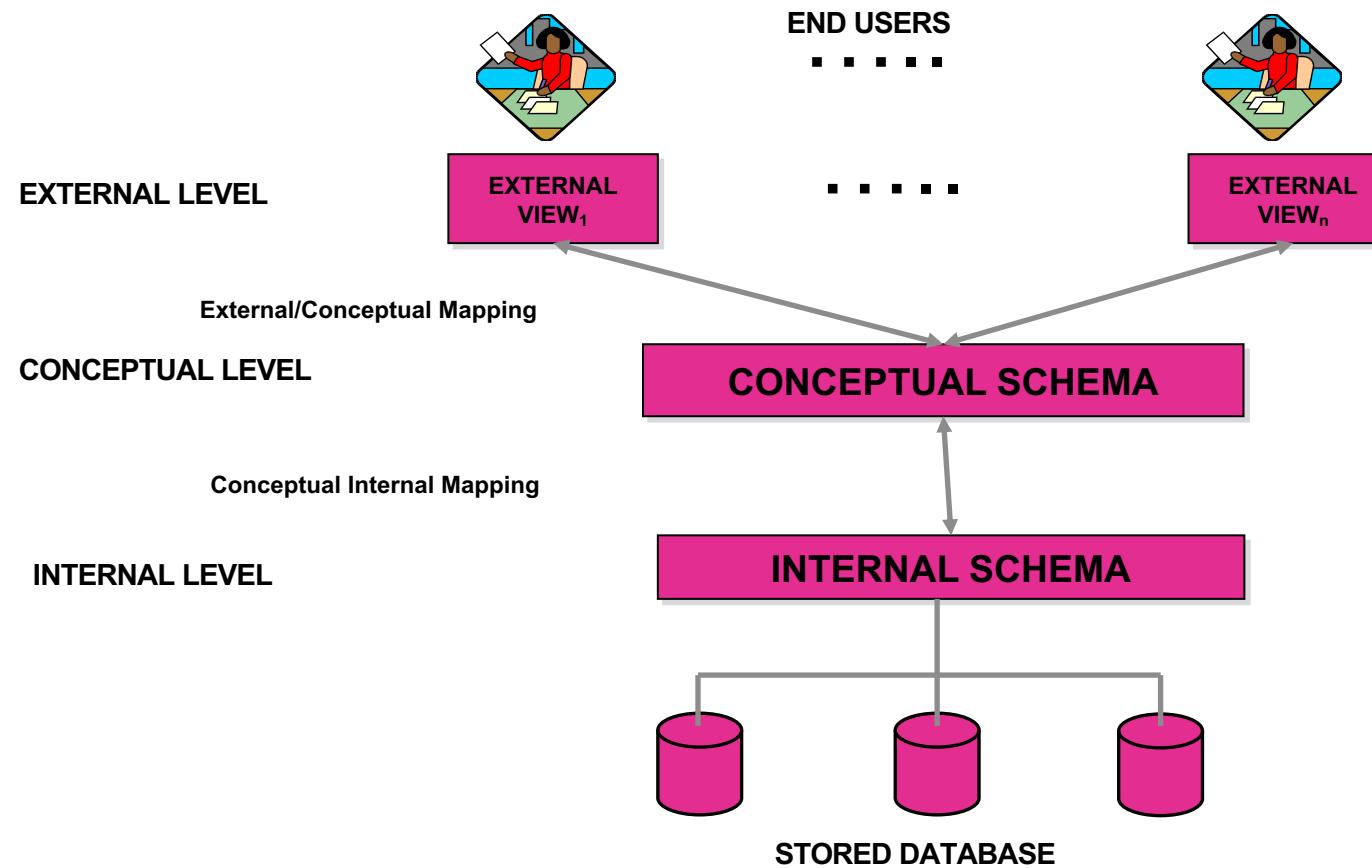
- Characteristics of Database Approach
  - Self-describing
    - Catalog (or meta-data) stores the description of the database
    - Allow the DBMS software to work with different DBs
  - Data Abstraction:
    - Data model used to hide storage details
    - Present the users with a **conceptual** view of the DB
      - Most users only see the **logical** way of storing the data, not the **physical** storage
  - Sharing of data
    - Support multiple views of a DB
    - Allow concurrent access on a DB (e.g. using On-Line Transactional Approach)

## 3.2. Database approach

- Characteristics of the approach
  - Persistence
    - store data on secondary storage
  - Retrieval
    - a declarative query language
    - a procedural database programming language
  - Performance
    - retrieve and store data quickly
    - Able to deal with (possibly) large volumes of data
      - For instance, the DBMS PostgreSQL enables to store tables of size up to 16TB

## 3.2. Database approach

- 3-tier Schema Model (ANSI-SPARC Architecture)



## 3.2. Database approach

- Internal schema:
  - lowest level of data abstraction
  - keeps information about the physical storage of the data on the disk
  - tells us what data is stored in the database and how
  - It never deals with the physical devices. Instead, internal schema views a physical device as a collection of physical pages
  - Most database users never view the internal schema (no need)

## 3.2. Database approach

- Conceptual (logical) schema:
  - Defines all database entities, their attributes, and their relationships
  - Security and integrity information
  - In the conceptual level, the data available to a user must be contained in or derivable from the physical level
  - Database administrators, database designers and advanced end-users usually deal with the conceptual schema

## 3.2. Database approach

- External views:
  - External level is nearest to the user
  - Each external view is intended for a certain group/profile of end-users (e.g. cashier, manager...)
  - Describes the segment of the database which is needed for this user group and hides the remaining details from the database from the specific user group
  - Casual and naïve end-users usually only deal with external views

## 3.2. Database approach

- There exist different database approaches:
  - **Relational Approach**
    - Tables of data linked together by common data
    - SQL language is generally used to define/manipulate/query the data
      - SELECT name from Authors
    - The most widely used approach until nowadays

Authors		
id	name	surname
Onid3n7	J.K.	Rowling

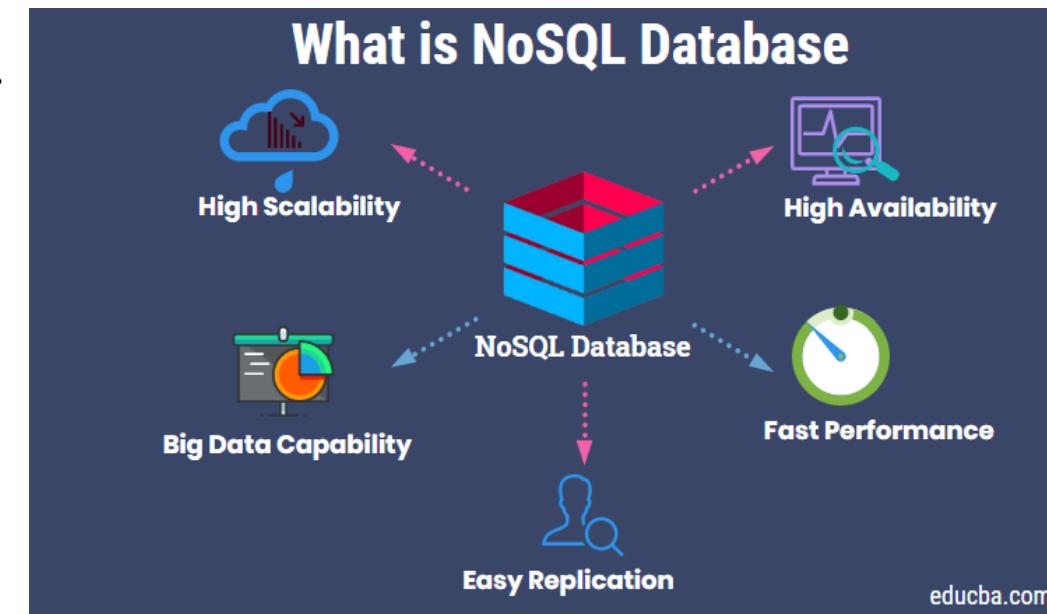
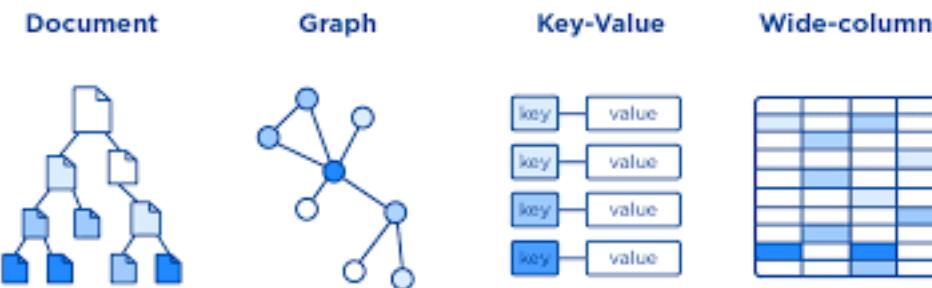
Books			
id	title	isbn	author id
2mpi4wr9	Harry Potter 1	12347809	Onid3n7

## 3.2. Database approach

- There exist different database approaches:

- **Other approaches:**

- NoSQL ("Not only SQL")
  - Covers a variety of database types, based on their data model
  - Store data differently than relational tables.
    - document, graph, key-value, wide-column...
  - Provide more flexible schemas than relational databases
  - Most fitted for big data (huge volumes of data)



## 3.2. Database approach

- There exist different database approaches:
  - Other approaches:
    - XML Documents
      - XML files of data structured according to an XML **schema**, linked to each other by common data
      - Widely used for databases that need numerous exchanges between several users
        - *E.g.* between researchers

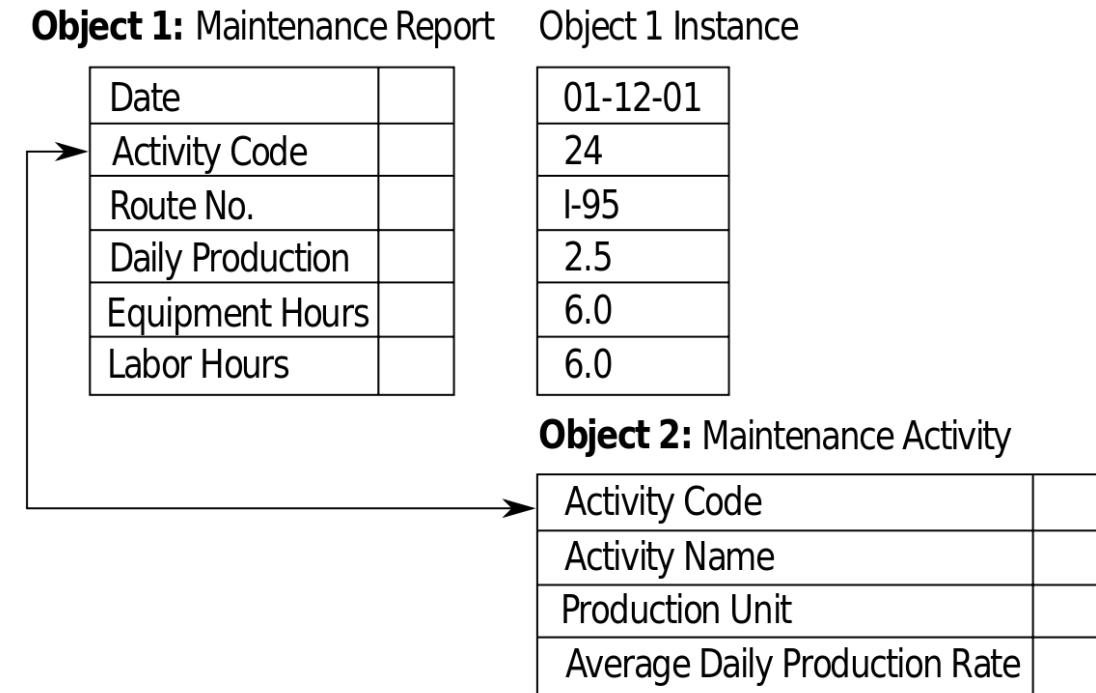
owl/@id	owl/species	owl/name	owl/region
1201	Bubo bubo	Eagle Owl	Eurasia
1202	Strix occidentalis	Spotted Owl	North America

Relational database	XML database
A relational database contains tables.	An XML database contains collections.
A relational table contains records with the same schema.	A collection contains XML documents with the same schema.
A relational record is an unordered list of named values.	An XML document is a tree of nodes.
A SQL query returns an <i>unordered</i> set of records.	An XQuery returns an <i>ordered</i> sequence of nodes.

```
<?xml version="1.0"?>
<birds>
  - <owl id="1201">
    <species>Bubo bubo</species>
    <name>Eagle Owl</name>
    <region>Eurasia</region>
  </owl>
  - <owl id="1202">
    <species>Strix occidentalis</species>
    <name>Spotted Owl</name>
    <region>North America</region>
  </owl>
</birds>
```

## 3.2. Database approach

- There exist different database approaches:
  - **Other approaches:**
    - Object-oriented databases
      - Efficient and effective, but failed to replace relational databases so far



## 3.2. Database approach

- There exist different database approaches:
  - In this course, we will focus mainly on the **relational approach**
  - Most widely used (in companies, on the web...)

# Summary

- Overview
  - Course overview
  - Course objective
  - Motivation for studying databases
- Basic concepts
  - Data
  - Database
  - Data model vs. schema vs. Instance
  - Database management system (DBMS)
  - Database environment
  - Database users
- Data management
  - File management system approach
  - Database management system approach



VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG  
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you for  
your attention!



# Questions

