



Database

Lesson 1. Introduction to databases



Viet-Trung Chan
ls.hust.edu.vn/~trungtv/

The background of the slide is red with a pattern of overlapping orange-red diamonds of varying sizes, creating a textured look.

Communication channel & course materials

- <https://www.facebook.com/groups/trungtv.students/>

Learning map

Sequence	Title
1	Introduction to databases
2	Relational Databases
3	Relational Algebra
4	Structured Query Language – Part 1
5	Structured Query Language – Part 2
6	Constraints and Triggers
7	Entity Relationship Model
8	Functional Dependency
9	Normalization
10	Storage - Indexing
11	Query Processing
12	Transaction Management – Part 1
13	Transaction Management – Part 2

Outline

- Course overview
- Basic concepts on database
- Data management

Learning objective

- 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

Definitions

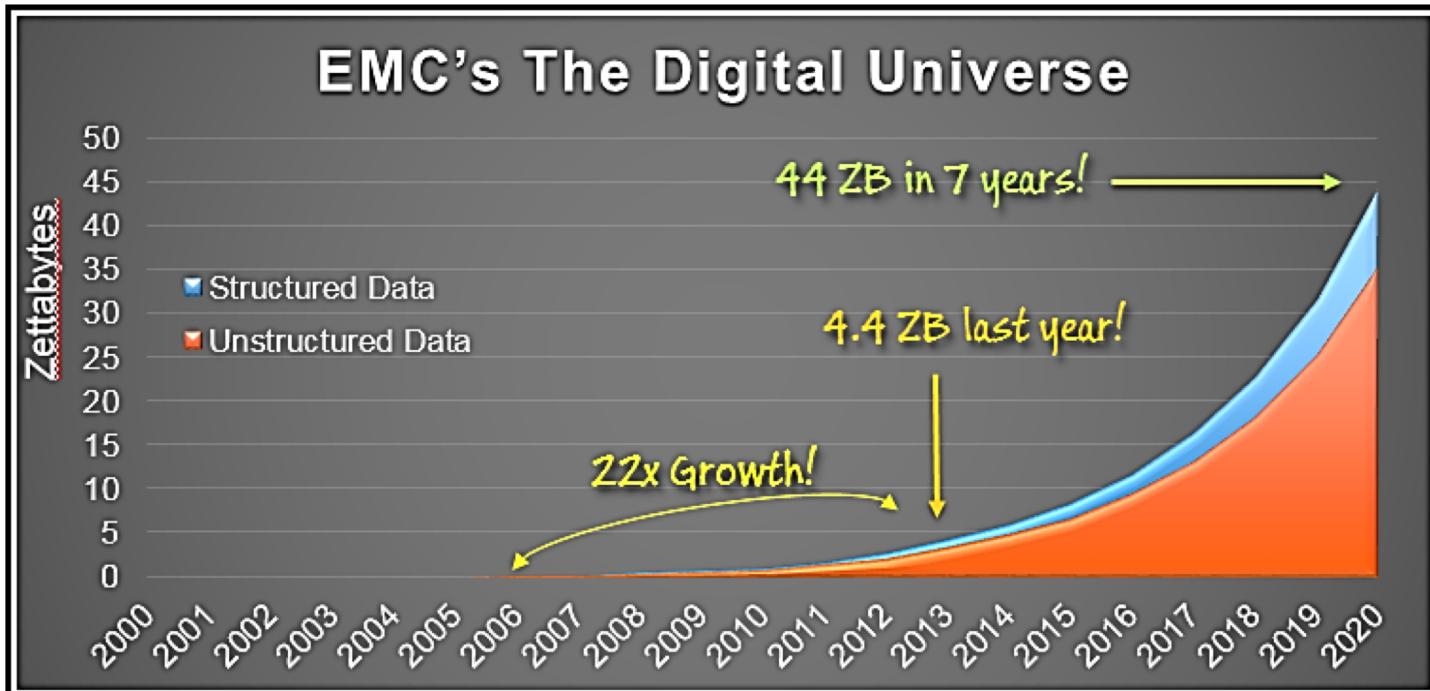
Keyword	Description
Database	A shared collection of related data designed to meet the information needs of an organization
DBMS	A software to facilitate the creation and maintenance of a database
Data model	A set of concepts used to describe the structure of a database: data types, relationships, constraints, semantics, ...
Schema	A data structure fulfilled all features of the parts of the real world which is of interest to the users
Instance	The data itself (in the context of data model and database schema)

1. Introduction

- Major research field with long history (since the begining of computer)
- 90% applications use databases
- Hot jobs in startups, big coorporates
- Massive Industry: Oracle, IBM, Microsoft, Google, AWS

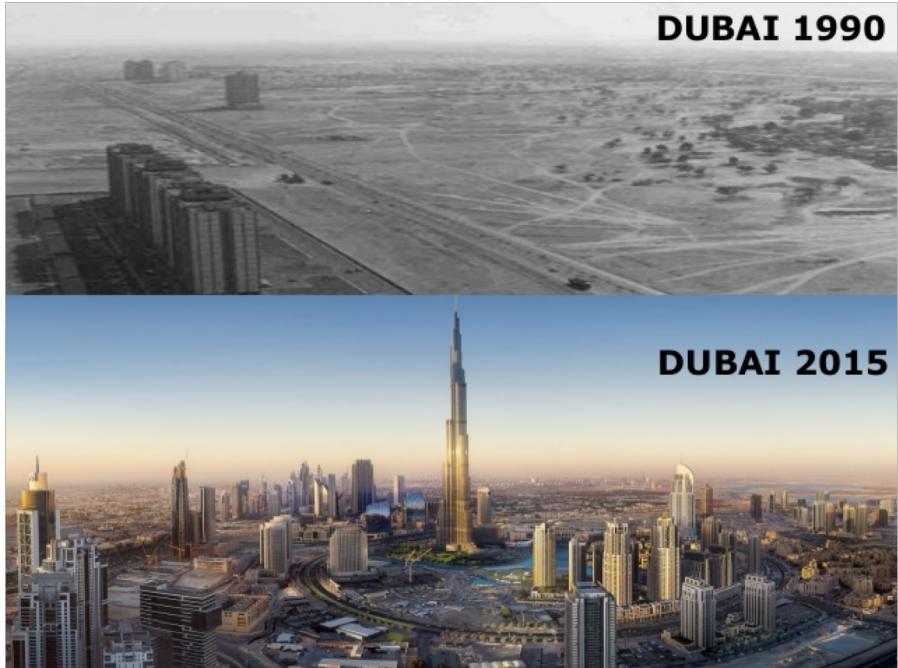
1. Introduction

- How big is our digital universe?



1. Introduction

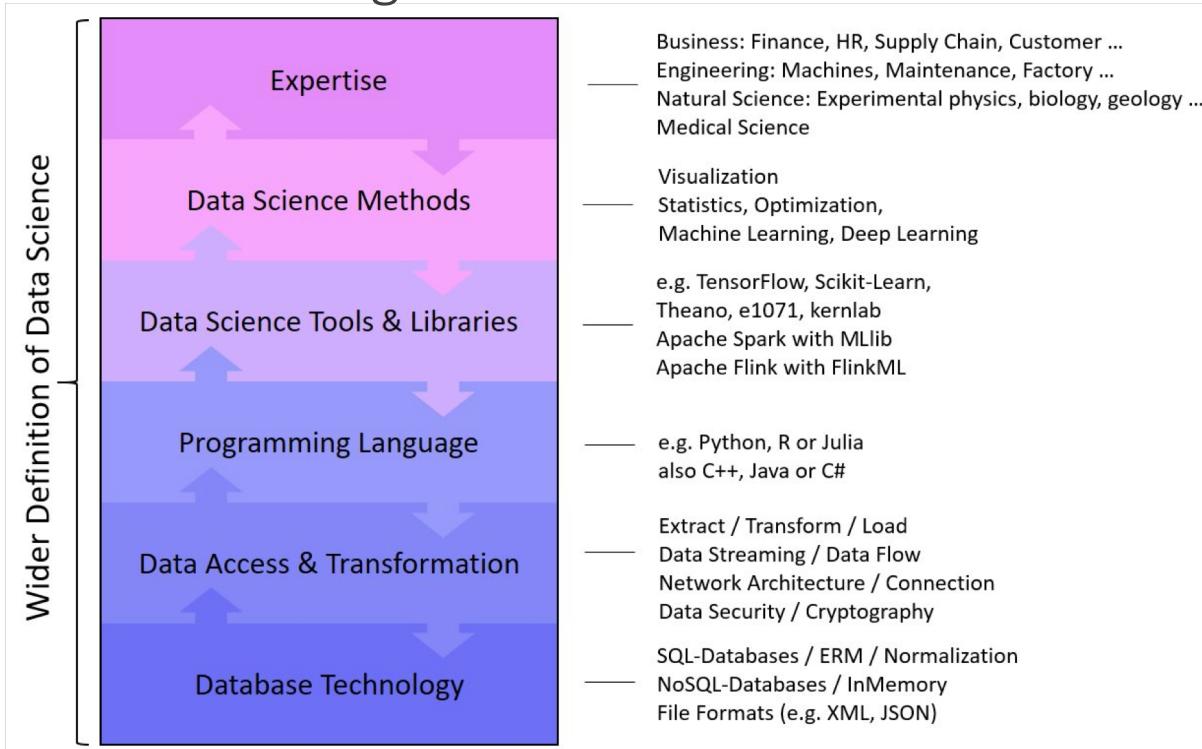
- Data is the new oil of the 21st century



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

1. Introduction

- Data science knowledge stack



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

- Definition

Wikipedia

Data is **any sequence of one or more symbols** given meaning by specific act(s) of interpretation.

Businessdictionary.com

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

- Definition

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

2.2 Database [2]

- Logically coherent
- Internally consistent
- Specific purpose
- Representation of the real world
- Entities (e.g., Students, Courses)
- Relationships (e.g., Tam is enrolled in C++)

2.2 Database example

[Example] A course management system

Entities	Relationships
<ul style="list-style-type: none">• Students• Courses• Teachers	<ul style="list-style-type: none">• Students take in some courses• Course are given by some teachers

2.3. Model vs. Schema vs. Instance

Data Model	<ul style="list-style-type: none">• Set of concepts used to describe the structure of a database: data types, relationships, constraints, semantics,• Tool for data abstraction
Schema	<ul style="list-style-type: none">• Data structure fulfilled all features of the parts of the real world which is of interest to the users
Instance	<ul style="list-style-type: none">• Data itself

2.3. Example of Model vs. Schema vs. Instance

Data Model	<pre>type <type_name> = record <field_name> : <data_type>; <field_name> : <data_type>; ... end;</pre>
Schema	<pre>type student = record ID : string; fullName: string; Birthday: date; Address: string ; Class: string; end;</pre>
Instance	(« Stud001 », « Nguyen », 1/4/1983, «1 Dai Co Viet », « 1F VN K50 »)

2.5. Database Management System (DBMS)

- Definition

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



2.5. Database Management System (DBMS)

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

* Korth & Silberschatz. Database System Concepts.



2.5. Database Management System (DBMS) [2]

- Defining ~ specifying types of data
- Constructing ~ storing & populating
- Manipulating ~ querying, updating, reporting



2.6. 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



Application



DBMS



use and control the content

enable the database to be developed and the DBMS software to be designed and implemented

2.7. Database Users

- Database administrators
 - authorize access to the database
 - co-ordinate and monitoring its use
 - acquire software, and hardware resources, controlling its use and monitoring efficiency of operations.

2.7. Database Users [2]

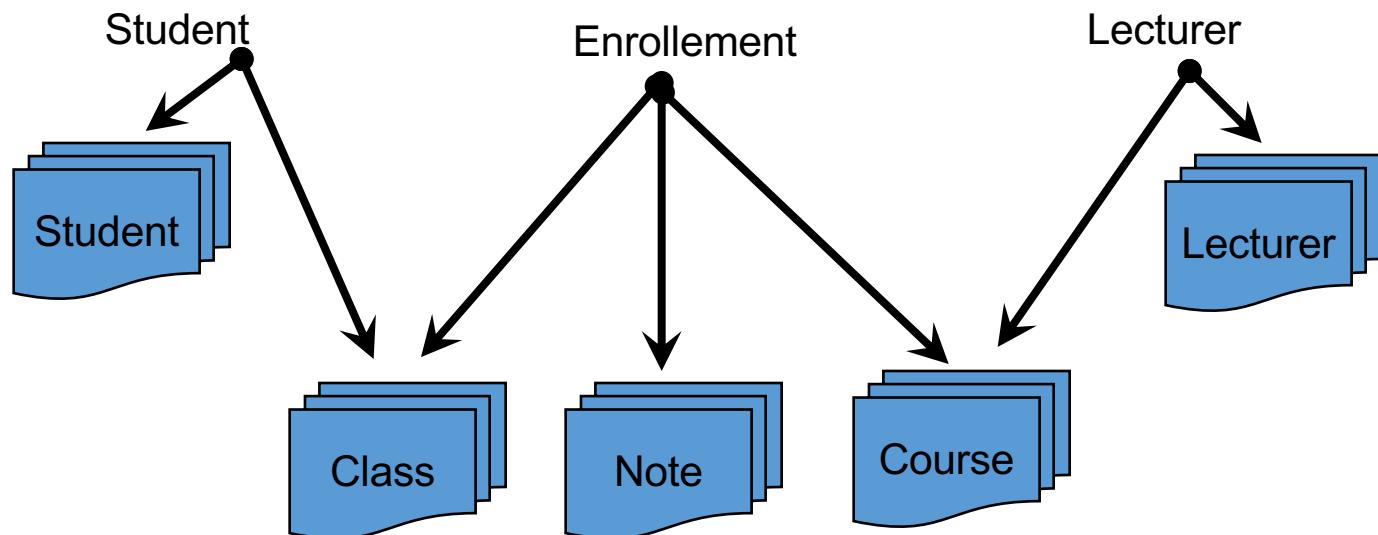
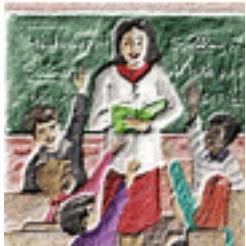
- Database Designers
 - define the content, the structure, the constraints, and functions or transactions against the database.
 - communicate with the end-users and understand their needs.
- End-users
 - use the data for queries, reports and some of them actually update the database content.
 - Casual end users
 - Naive users
 - Sophisticated end users

3. Data management

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

3. Data management

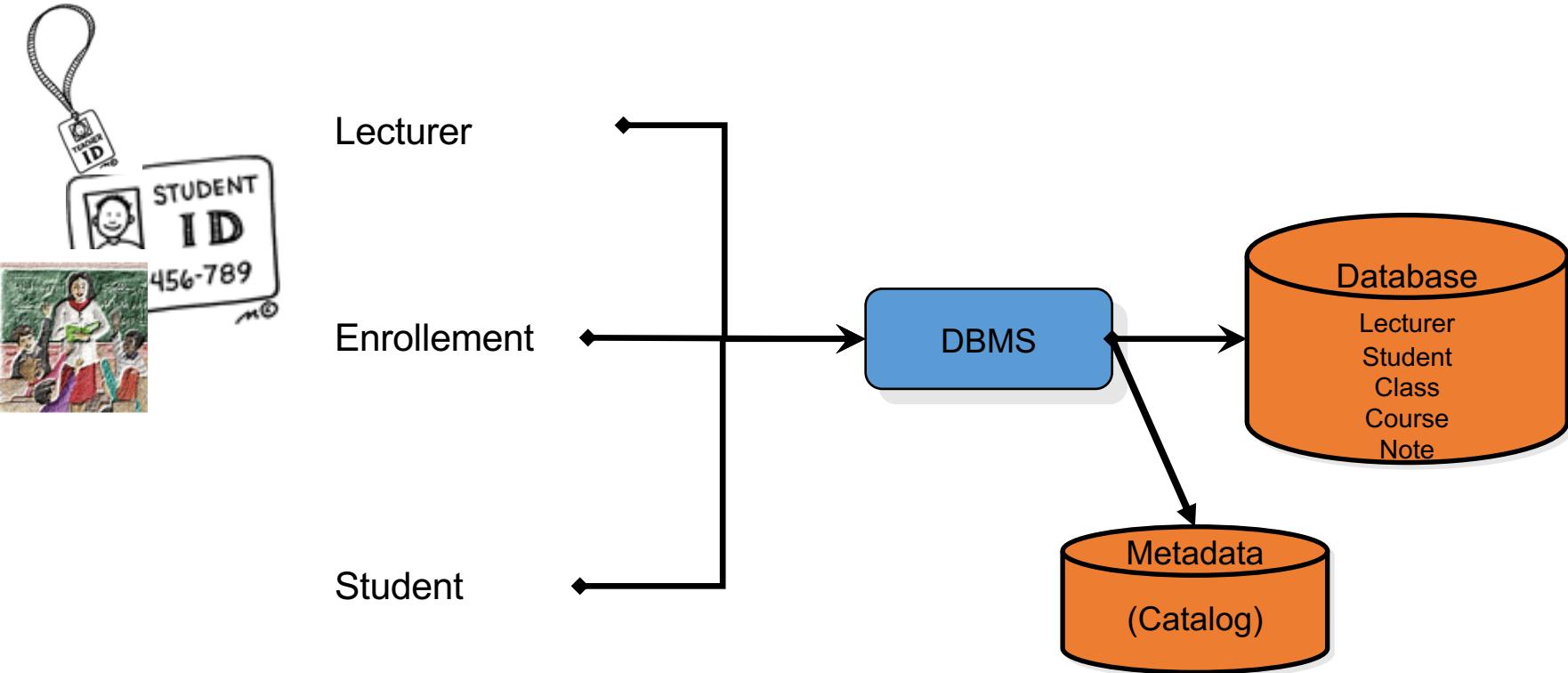
❖ 3.1. File management system approach



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 [2]

- **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**

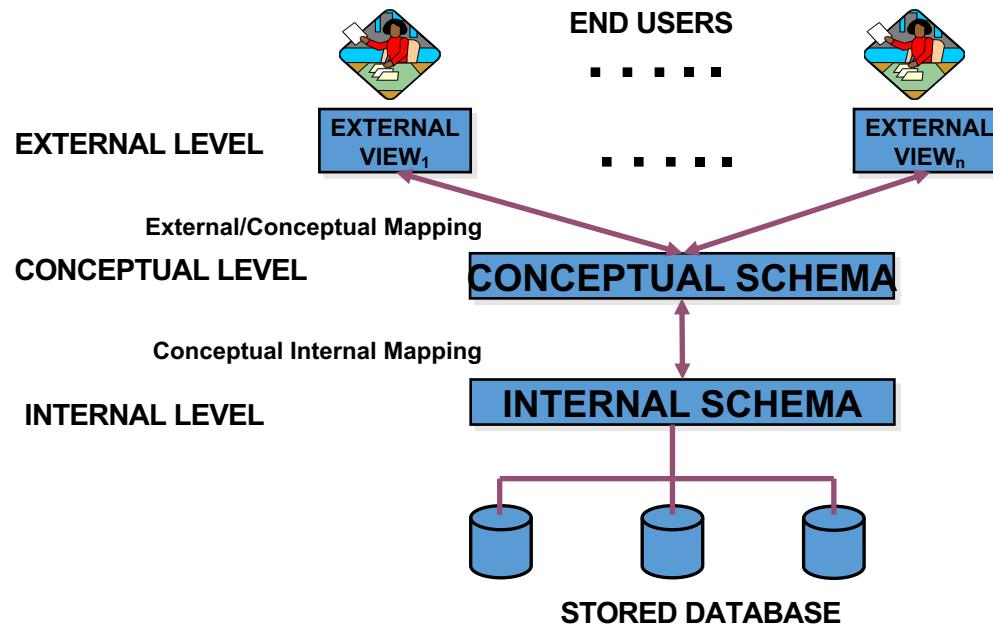
3.2. Database approach [3]

- ***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
- Sharing of data
 - Support multiple view of a DB
 - Allow concurrent access on a DB

3.2. Database approach [3]

- *Characteristics of Database Approach*
- Persistence
 - store data on secondary storage
- Retrieval
 - a declarative query language
 - a procedural database programming language
- Performance
 - retrieve and store data quickly
 - deal with large volume of data

3-tier Schema Model (ANSI-SPARC Architecture)



Remark

- DBMS is
 - more expensive
 - more complex
 - general

Quiz 1.

Quiz Number	1	Quiz Type	OX	Example Select
Question	What is a shared collection of related data designed to meet the information needs of an organization?			
Example	<ul style="list-style-type: none">A.DataB.DatabaseC.Database Management SystemD.Data model			
Answer	B			
Feedback				

Quiz 2.

Quiz Number	2	Quiz Type	OX	Example Select
Question	Which are the roles of Database Management System?			
Example	<ul style="list-style-type: none">A.Defining ~ specifying types of dataB.Constructing ~ storing & populatingC.Manipulating ~ querying, updating, reportingD.Generating ~ issuing data automatically			
Answer	A, B and C			
Feedback				

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



Next lesson: Database models

- Weiss, Mark A. Data Structures and Algorithm Analysis in C++. 4th Edition. Pearson 2014.
- Goodrich and Tamassia. Data Structures and Algorithms in Java, 4th ed. 2006.
- C.H. Papadimitriou. and U.V. Vazirani, Algorithms, McGraw-Hill.