DTS207TC Database Development and Design

Lecture 1
Introduction & Review of 106

Di Zhang, Autumn 2025

Learning Outcomes

- Second round of learning: advanced content of database
  - Analysis
    - Demonstrate an understanding of advanced SQL topics.
    - State the main concepts in data warehousing and data mining.
  - Application
    - Illustrate the issues related to Web technologies and DBMS and XML as a semi-structured data representation formalism.
    - Identify the principles underlying object-relational models.
    - Identify and apply the principles underpinning transaction management within DBMS.

More content beyond...

- Extensions for Computer Science:
  - Hadoop
  - Physical Storage
  - Information Retrieval (Google's search engine, the origin of Big Data)

Textbook

- Textbook:
  - https://core.xjtlu.edu.cn/mod/folder/view.php?id=7005
- Video:
  - https://search.bilibili.com/all?vt=16822812&keyword=Database%20System%20Concepts&from_source=webtop_search&spm_id_from=333.1007&search_source=5
  - https://www.youtube.com/results?search_query=database+system+concepts

*Comparison of AI-related Disciplines

- Math: Conjecture & Proof (Example)
- Physics: Modeling & Experiments (infeasible for many other disciplines, such as finance)
- Statistics: Induction, with mathematical guarantees (with definition of distributions)
- **Computer**: Construction, without mathematical guarantees (formal definition of large software can be slow and costy)
  - Scalability: Brute force prevails over skill (一力降十会)
- Deep Learning: Construction + Induction, without solid mathematical guarantees

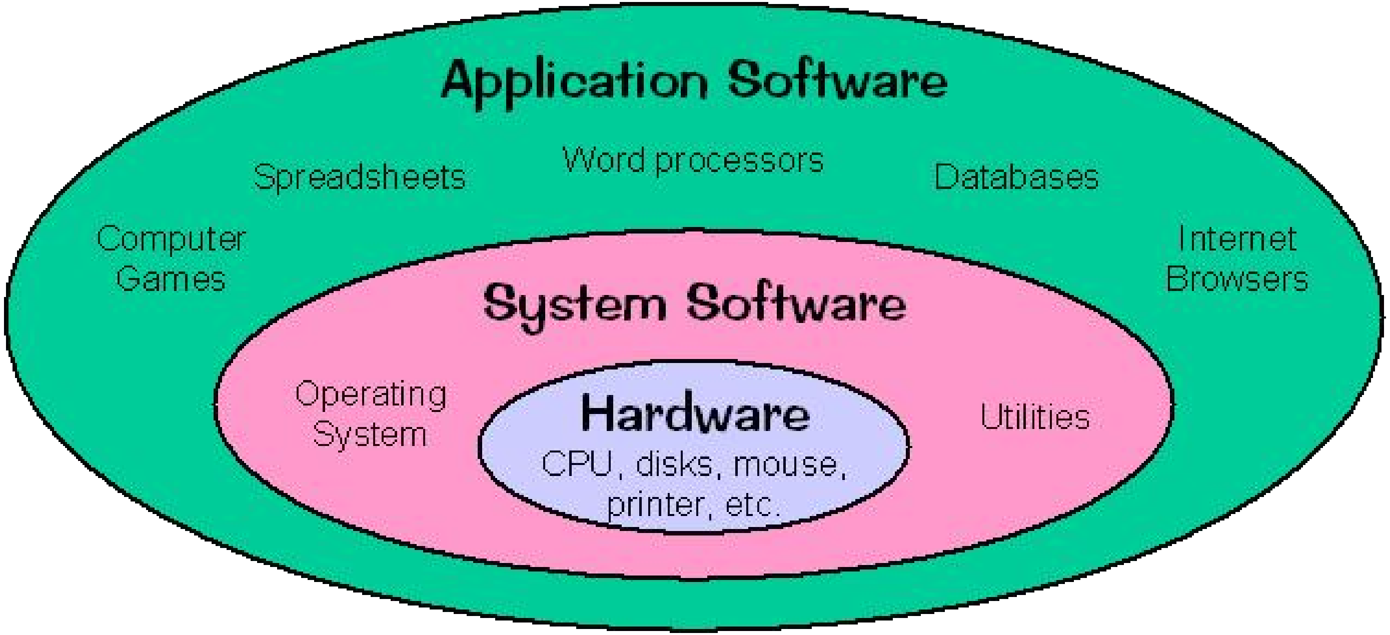
Computer Science

- Construct like building blocks...

	Theory	Practice
Application	Optimization geometry	Machine Learning Computer Graphics Internet ....
Software	Discrete Mathematics: Turing Machine (doodle) + Von Neumann architecture	Network Database Compiler OS
Hardware	Electromagnetism	Digital Circuit (Clock Speed) paper/mechanical/...

System Software

- System software is software designed to provide a platform for other software.



Q: Can you give some examples of system software?

OS:Conceptualize Hardware

Functionaility	Windows/Linux
Process:	Queue/Time Sharing (for realtime), nice command
Isolation	Test and Set-> Mutex -> Threading/Process -> Queue/...
Paralellization	
Storage:	
	Addr./Blocks -- similiar to a city
Memory	
Disk	NTFS/ext4
Interface:	
GUI	
Progamming	cmd/Bash
I/O:	
	Mouse/Keyboard
Import/Export	Monitor

Database:Conceptualize Bits

Functionaility	PostgreSQL
Process: Isolation Paralellization	Transaction
Storage: Memory Disk	Relational Tables
Interface: GUI Progamming	SQL
I/O: Import/Export	Import csv ODBC/Drivers

Survey: How much do you still remember?

- DBMS
- OLTP/OLAP
- DDL/DML/SQL
- E-R/Relational
- 1NF/2NF/3/NF

DBMS

- A super powerful "warehouse manager".

Analog

	warehouse	Database
Physical level	It's like putting things in boxes and cabinets; the physical layer controls how the cabinet is placed.	It controls where and how data resides.
Logical Level	Defines the structure and relationships of data.	Catalog system that tells you the contents and categories of each box.
Logical Level	The view layer is the data that users see. Users do not need to know how the data is stored or the detailed structure of the data. They only need to see the information they need.	Combinable room and door lock settings.

OLTP/OLAP

- OLTP (Online Transaction Processing): Handles daily business operations.
- OLAP (Online Analytical Processing): Analyzes and summarizes data to identify patterns and trends.
- Analogy:
    - The CPU (Central Processing Unit) is like an all-around "transaction processor," adept at handling a wide variety of small tasks at high speed.
    - The GPU (Graphics Processing Unit) is like a professional "data analyst," skilled at handling high-volume, repetitive tasks, particularly graphics and data analysis.

DDL/DML/SQL

- SQL (Structured Query Language) is a language used to "chat" with a database. Just like chatting with a friend, you use SQL to tell the database what you want to do, such as store data, retrieve data, or modify data.

- DDL (Data Definition Language) is a part of SQL specifically used to define and modify database structures. Just like when you build a house, DDL is the tool used to design and modify the blueprints.

  - Specific examples:

    Create a table: CREATE TABLE table_name (field1 type, field2 type);

    Modify a table: ALTER TABLE table_name ADD COLUMN new_column type;

    Drop a table: DROP TABLE table_name;

DDL/DML/SQL

- DML (Data Manipulation Language)
  - DML is part of SQL and is specifically used for data manipulation, such as inserting, querying, updating, and deleting. Just like moving or arranging things around a room, DML is a tool for manipulating data.

*More about SQL

- a DSL
- Declaration-based language: 4th generation language
- Further reading: NoSQL->NewSQL

E-R and Relational

- The E-R model is a graphical tool used to design and describe relationships between data. It helps us understand the structure of data and the connections between them.
  - Its elements include entities, attributes, and relationships.
- The relational model is a model used to store and manage data in a tabular format.
  - Its elements include tables, fields, and foreign keys.

Mapping from E-R to Relational

- The mapping from the E-R model to the relational model isn't unique.
  - Different levels of normalization (such as 1NF, 2NF, and 3NF) result in different table structure designs.

- Reflection: Does the E-R model necessarily map to the relational model?