EBU5602 Database

Dr Na Yao

Welcome to Database

- Ask when you don't understand anything during lecture
  - No questions too "silly"!
- Use discussion forum on QMPlus for other time
- Office hour

What this course is about

- Database
- Database Management System
- Data modelling
- Using query language (SQL) to define, create, maintain and control access to the database
- Transaction management
- Semi-structured data/NoSQL

Content

- Week 1
  - Introduction
  - Relational model
  - Relational Algebra
  - Entity-relationship(ER) modelling
- Week 2
  - EER modelling
  - ER to Relational Model mapping
  - SQL
  - Database design
  - Normalization

Content

- Week 3
  - Advanced Normalization
  - Transaction management
- Week 4
  - Distributed DBMS
  - XML
  - NoSQL

Organisation of the module

- Lectures
- Lab: Six 2-hour lab sessions (starting from second teaching week)
- Assessment
  - Coursework: 10%
  - Mid-term test: 10%
  - Exam: 80%
  - Quizzes: beneficial to check your own learning, not assessed

Organisation of the module

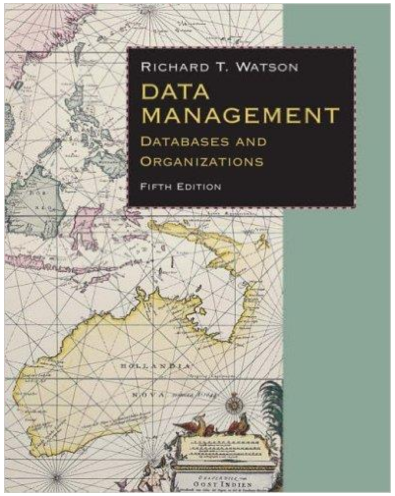
- TAs: TBA
  - You can seek help from the TAs during lab-lectures and labs
- Modules Reps: 2 students
  - General communication with lecturers
  - SSLC meetings
  - Please contact me after the lecture if you are interested!

- Textbook:

  Database systems : a practical approach to design, implementation, and management, 6th edition, Thomas M. Connolly, Carolyn E. Begg. ISBN 9781292061184 *Available from QMUL library as e-book http://www.library.qmul.ac.uk/*

- Recommended reading:

  Data Management: Databases & Organizations, 5th Edition, Richard T. Watson, ISBN 978-0-471-71536-8



Database

- Daily examples?

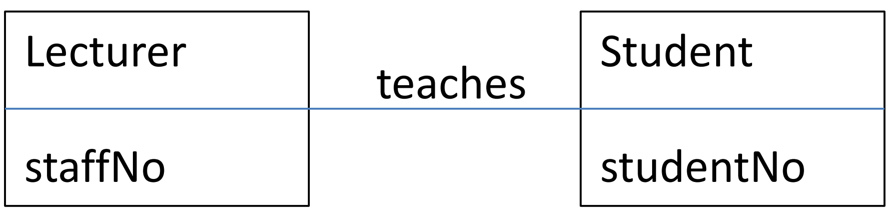
- Data: facts and statistics collected together for reference or analysis (Oxford dictionary)

- **Database**: a shared collection of logically related data (and a description of this data), designed to meet the information needs of an organization.

- System catalogue (metadata) provides description of data to enable program–data independence.

“Original database” by shinichi, flickr

- Logically related data comprises entities, attributes, and relationships of an organization's information.
- Entity: a distinct object in the organization that is to be represented in the database.
- E.g. person, place, thing, etc
- Attribute: a property that describes some aspect of the object that we wish to record.
- Relationship: an association between entities.



Database Management Systems (DBMS)

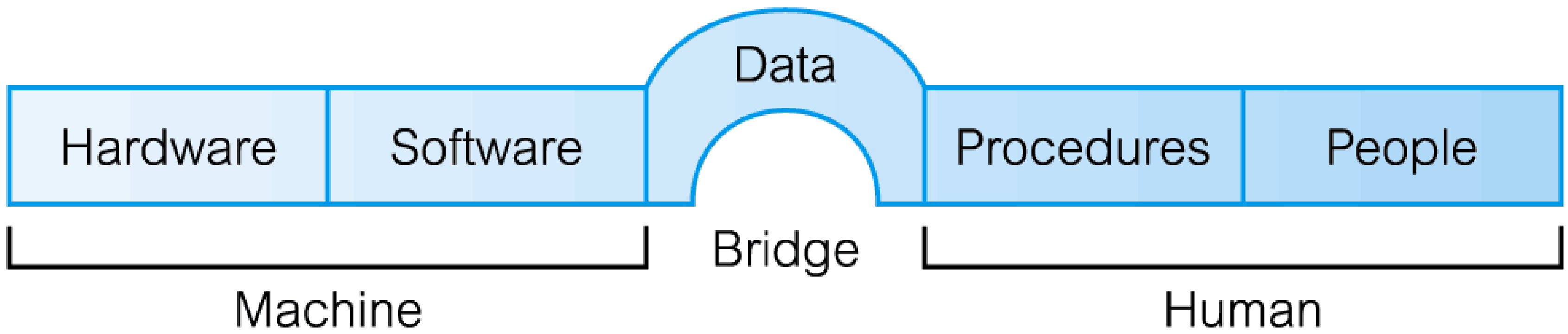
- A software system that enables users to define, create, maintain, and control access to the database.

- (Database) application program: a computer program that interacts with database by issuing an appropriate request (SQL statement) to the DBMS.

Key concepts

- Data model
  - A model is a representation of ‘real world’ objects and events, and their associations.
- Schema vs Data
  - The description of the database is the **database schema**.
  - Data is the actual information stored in the database.
- Data definition language (DDL)
- Data manipulation language (DML)

Components of DBMS Environment

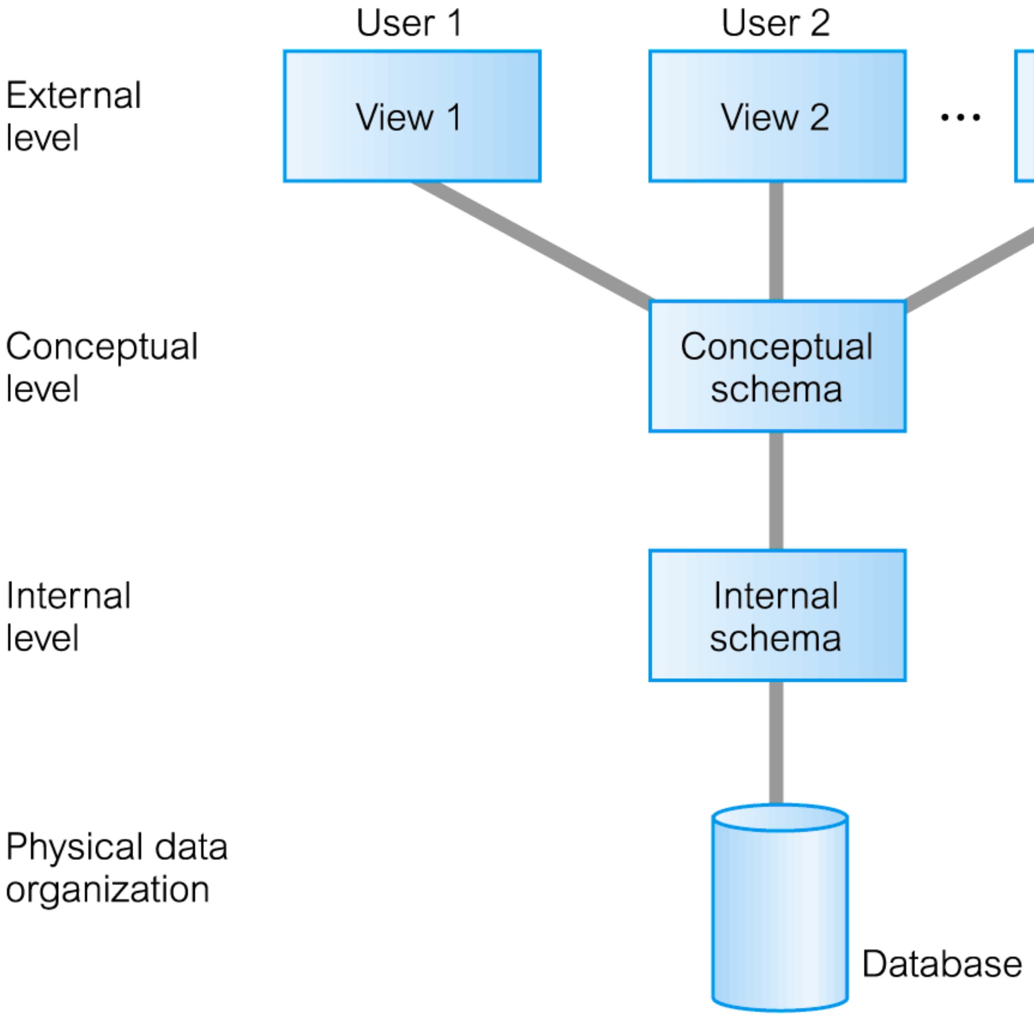


Roles in the Database Environment

- Data Administrator (DA)
- Database Administrator (DBA)
- Database Designers (Logical and Physical)
- Application Developers
- End Users (naive and sophisticated)

Data abstraction

The three-level ANSI-SPARC architecture

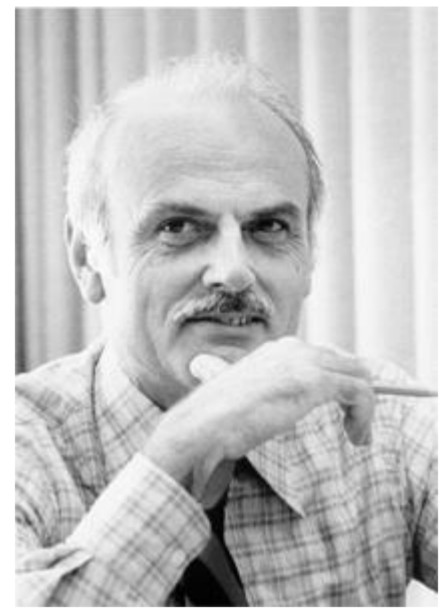


The three-level ANSI-SPARC architecture

- The objective of the three-level architecture is to separate each user’s view of the database from the way the database is physically represented.
  - Users should be able to access same data but with customized view. Each user should be able to change the view without affecting other users.
  - Users should not have to deal with physical database storage details (users’ interaction with database should be independent of storage considerations).
  - Users’ views should be unaffected by changes to the physical aspects of storage.
  - Internal structure of the database should be unaffected by changes to the physical aspects of storage.
  - The DBA should be able to change the conceptual structure of the database with minimum affect to users’ views.

History of DBMS (1)

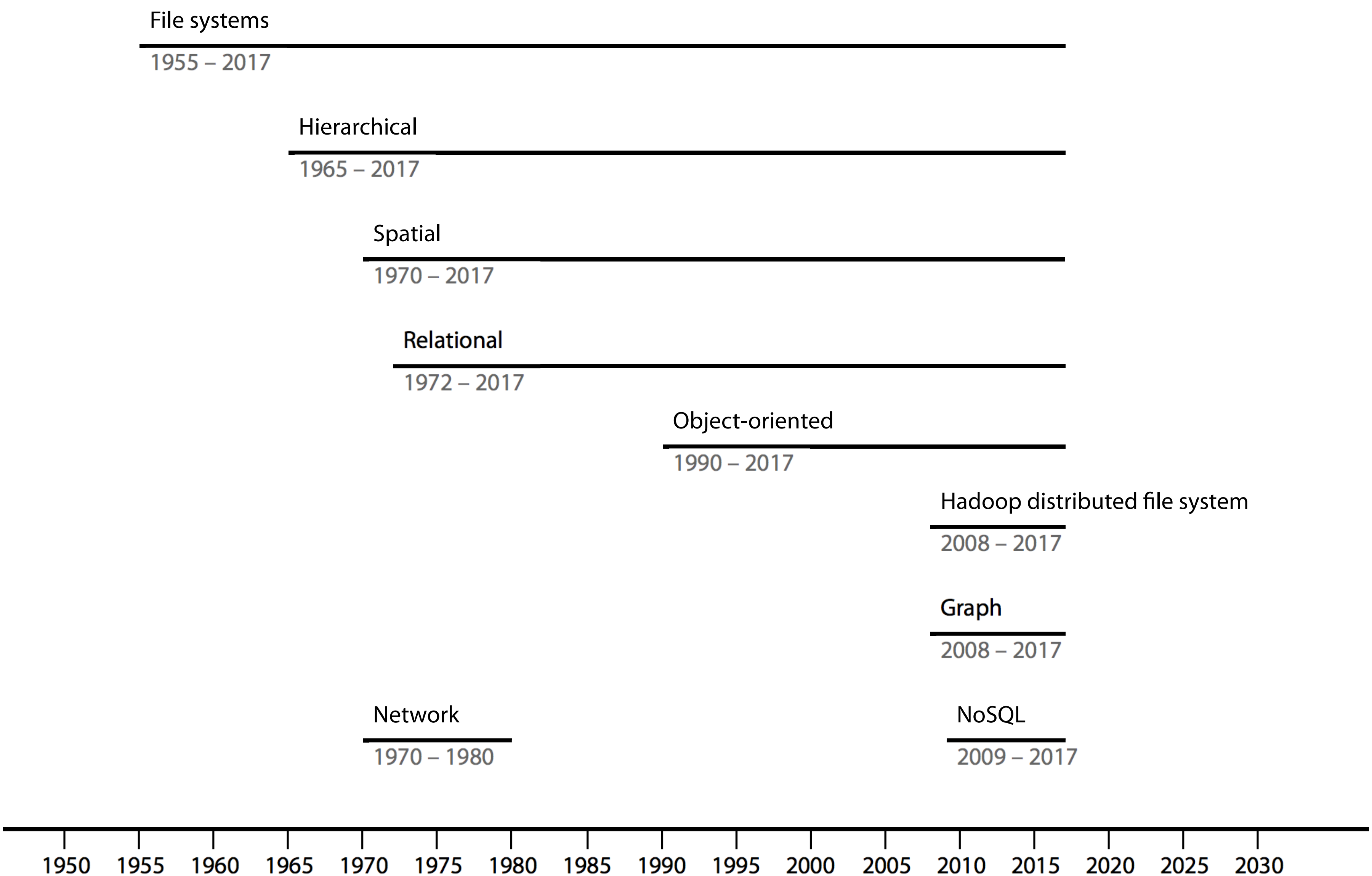
- Early 60s
  - Charles Bachman (Turing Award 1973)
  - Integrated Data Store
  - Network Data Model
  - CODASYL
- Late 60s
  - IBM
  - Information Management Systems (IMS)
  - Hierarchichal Data Model
- 70s
  - Edgar Codd, IBM (Turing Award 1981)
  - Relational Data Model



Edgar F. Codd - Wikipedia

History of DBMS (2)

- 80s
  - Relational Data Model
  - SQL
  - Transaction Management (James Gray, Turing Award 1999)
- Now
  - Object-oriented Data Model
  - Data warehouse and data mining
  - Accessing databases through the web/internet
  - Multimedia data
  - Text data (information retrieval)
  - Structure of the data (XML)


File systems
1955 – 2017
Hierarchical
1965 – 2017
Spatial
1970 – 2017
Relational
1972 – 2017
Object-oriented
1990 – 2017
Hadoop distributed file system
2008 – 2017
Graph
2008 – 2017
Network
1970 – 1980
NoSQL
2009 – 2017
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030

Advantage of DBMSs

- Control of data redundancy
- Data consistency
- More information from the same amount of data
- Sharing of data
- Improved data integrity
- Improved security
- Enforcement of standards
- Economy of scale
- Balance conflicting requirements
- Improved data accessibility and responsiveness
- Increased productivity
- Improved maintenance through data independence
- Increased concurrency
- Improved backup and recovery services

Disadvantages of DBMSs

- Complexity
- Size
- Cost of DBMS
- Additional hardware costs
- Cost of conversion
- Performance
- Higher impact of a failure