CSC 355 Database Systems

Marcus Schaefer

Databases ?

Database

1. DB models aspects of the real world
   (*miniworld*, *universe of discourse*)

2. Collection of *data*
   - logically coherent $\longrightarrow$ *Information*
   - Meaningful

3. Designed for specific purpose

Uses of Database

- Traditional (Employee, student, product database)
- Online Shopping
- Search Engines
- Data Warehousing (OLAP)
- Data Mining
- Genetic Databases
- Geographical Information Systems

Types of Database

- Traditional (Postgres, Oracle, MySQL)
- Deductive Databases
- Multimedia Databases
- Distributed Databases
- Spatial Databases
- Object-Oriented Databases
- No-SQL Databases

No-SQL Databases

- Key-Value (Riak)
- Columnar (Cassandra, HBase)
- Document (MongoDB)
- Graph (Neo4J)

Sizes of Database

- Personal (1 User), Megabytes
- Workgroup (<25 Users), Megabytes
- Department (25-100 Users), Gigabytes
- Enterprise (100-1000s), Gigabytes
- Internet (> 1000s), Terabytes - Petabytes

How long does it take to find a piece of data in petabytes of data?

Database Management System (DBMS)

Software to

1. Define a database
   (data types, relations, constraints)

2. Construct a database
   (populate database with data)

3. Manipulate database
   (query and update data in database)

Database People

- Database designers
- Application developers
- Database administrators
- Users

Explore
University Sample Database

LastName	FirstName	SID	SSN	Career	Program	City	Started	A
Snowdon	Jonathan	8871	123123123	GRD	INFO-SYS	Springfield	2005	
Winter	Abigail	11035	111111111	GRD	PHD	Chicago	2003	
Patel	Deepa	14662		GRD	COMP-SCI	Evanston	2003	
Starck	Jason	19992	789789789	UGRD	INFO-SYS	Springfield	2003	
Johnson	Peter	32105	123456789	UGRD	COMP-SCI	Chicago	2004	
Patel	Prakash	75234		UGRD	COMP-SCI	Chicago	2001	
Brennigan	Marcus	90421	987654321	UGRD	COMP-GPH	Evanston	2001	
Snowdon	Jennifer	93321	321321321	GRD	COMP-SCI	Springfield	2004	

Explore
University Sample Database

University Sample Database

User Data
- Records, Fields (Columns)
- Data elements

Meta Data
- Data Types
- Relationships
- Constraints
- Indexes

Other Data
- Log Records
- Statistics

Relational Databases
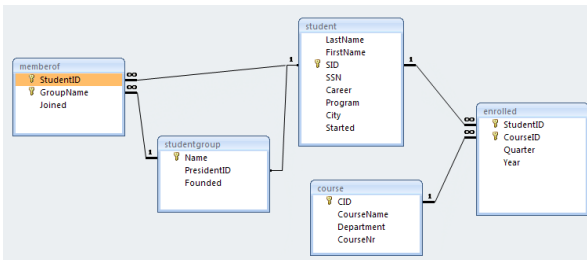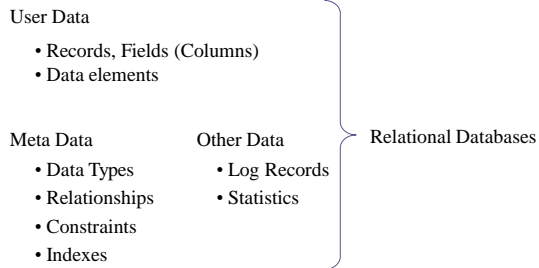
File Processing

File system is backbone of operating system

File system for data storage:
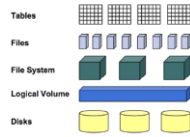
Disadvantages of File Processing

• Program-Data Dependence
• Redundancy (Duplication of Data)
• Limitation on data sharing
• Development time
• Maintenance

Advantages of Databases

• Program-Data Independence
• Control of Data Redundancy
• Data Consistency
• Data Quality (constraints)
• Data Sharing (customized access through views)
• Improved Data Access
• Program Maintenance

DBMS Architecture
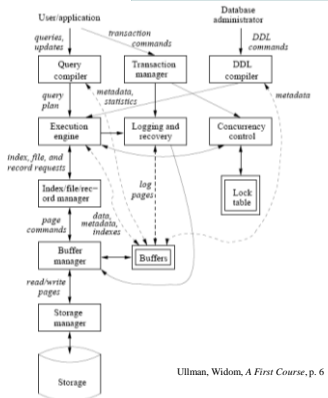

Ullman, Widom, *A First Course*, p. 6

DBMS

Open Source	Proprietary
H2	Access (Microsoft)
MySQL,	DB2 (IBM)
Postgres	Oracle
	SQL Server (Microsoft)
	Sybase (SAP)

Data Modeling

Describe structure of data (relationships, behavior) at different levels of abstraction.

Conceptual/External
  high-level user view

Internal
  Logical: structure of data for DBMS
  Physical: storage details (indexes) for DBMS

Data Models

Conceptual/External
    ER-model (Entity-Relationship)
    UML

Logical
    Relational data model
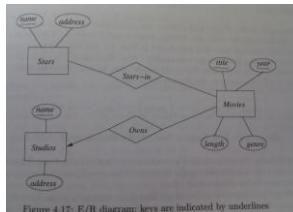   Object data model
    Network data model
    Hierarchical data model

ER-modeling

Describes **entities,** their **relationships,** and **attributes**

Used for designing and analyzing a database

Ullman, Widom, *A First Course*, p. 149

Figure 4.17: E/R diagram; keys are indicated by underlines

Relational Data Model

- Data in tables (extensional representation of relation)

- Models relationship between data in tables

Ullman, Widom, *A First Course*, p. 26

```
Movies(
    title:string,
    year:integer,
    length:integer,
    genre:string,
    studioName:string,
    producerC#:integer
)
MovieStar(
    name:string,
    address:string,
    gender:char,
    birthdate:date
)
StarsIn(
    movieTitle:string,
    movieYear:integer,
    starName:string
)
MovieExec(
    name:string,
    address:string,
    cert#:integer,
    netWorth:integer
)
Studio(
    name:string,
    address:string,
    presC#:integer
)
```

Figure 2.5: Example database schema about movies

Database Languages

DDL: Data definition language
       defines data types, tables
       includes DSL (Data storage language)

DML: Data Manipulation Language
       language for retrieving and manipulating data

Types:
       high-level (nonprocedural, declarative): SQL
       low-level (procedural)

Transactional Processing

Transaction: A group of database operations that should
      appear as a unit to the user.

      Example:
            Transfer $100 from account A to account B.

Requirements on transactions:

      Atomicity
      Consistency
      Isolation
      Durability

Class Outline

Week	
1	Intro to Database systems, Relational Model (Chapters 1/2)
2-4	SQL (with transactions) (Chapter 6)
5	Relational Design: Functional Dependencies and Normalization (Chapter 3)
6	Constraints & Triggers (Chapter 7)
7	Views & Indices (Chapter 8)
8-9	Database Programming (Chapter 9)
10	Advanced Topics: recursive SQL, ORL, semi-structured data, No-SQL