CSc 484
Database Management Systems

Ken Gamradt

Spring 2024

Introduction

Database

- A **database** is an organized collection of data stored and accessed electronically
- Small databases can be stored on a file system
- Large databases are hosted on computer clusters or cloud storage
- The design of databases spans formal techniques and practical considerations including
  - Data modeling, efficient data representation and storage
  - Query languages
  - Security
  - Privacy of sensitive data
  - Distributed computing issues including supporting concurrent access and fault tolerance

Database

- A **database management system** (**DBMS**) is the software that interacts with
  - end users
  - applications
  - the database itself to capture and analyze the data
- The DBMS software additionally encompasses the core facilities provided to administer the database
- The sum total of the database, the DBMS and the associated applications can be referred to as a **database system**
- Often the term "database" is also used loosely to refer to any of the DBMS
  - the database system
  - an application associated with the database

Database

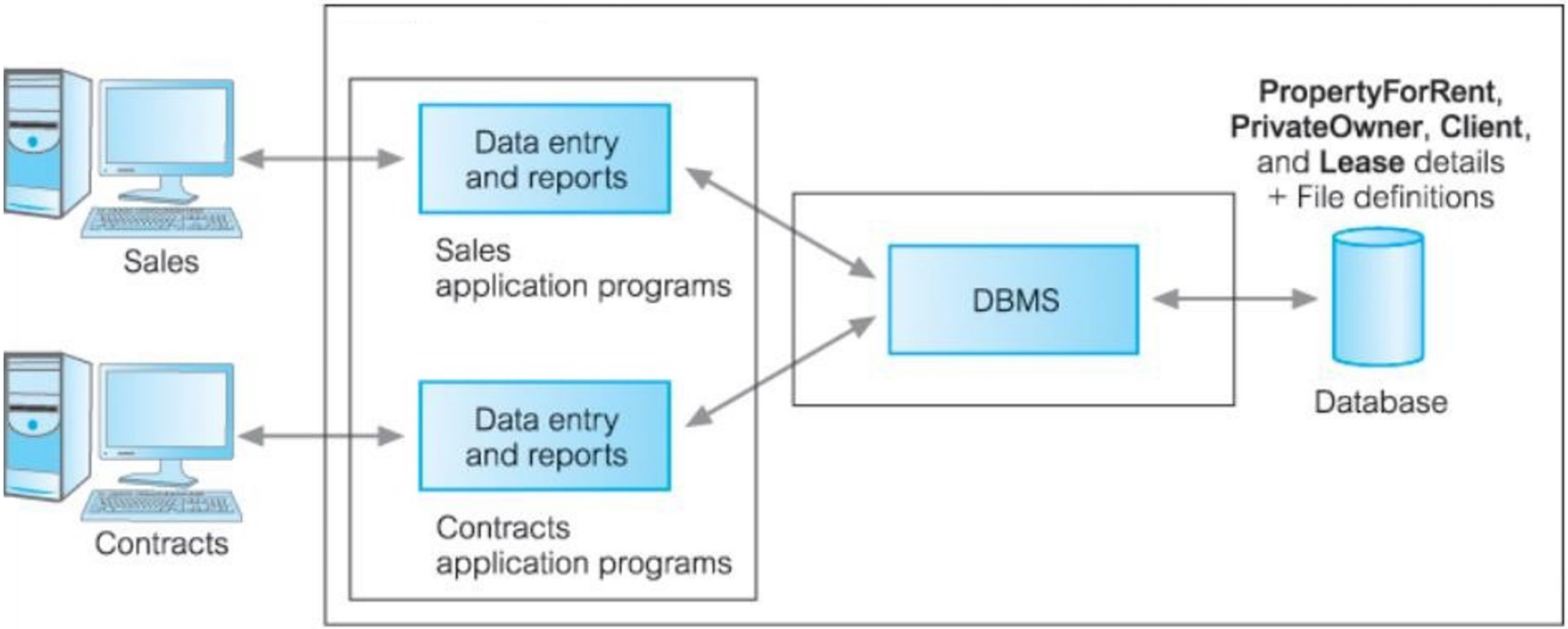
- Computer scientists may classify database management systems according to the database models that they support
- Relational databases became dominant in the 1980s
  - These model data as rows and columns in a series of tables
  - The vast majority use SQL for writing and querying data
- Non-relational databases became popular in the 2000s – 2010s
  - Collectively referred to as NoSQL
  - They use different query languages

Database-management Systems

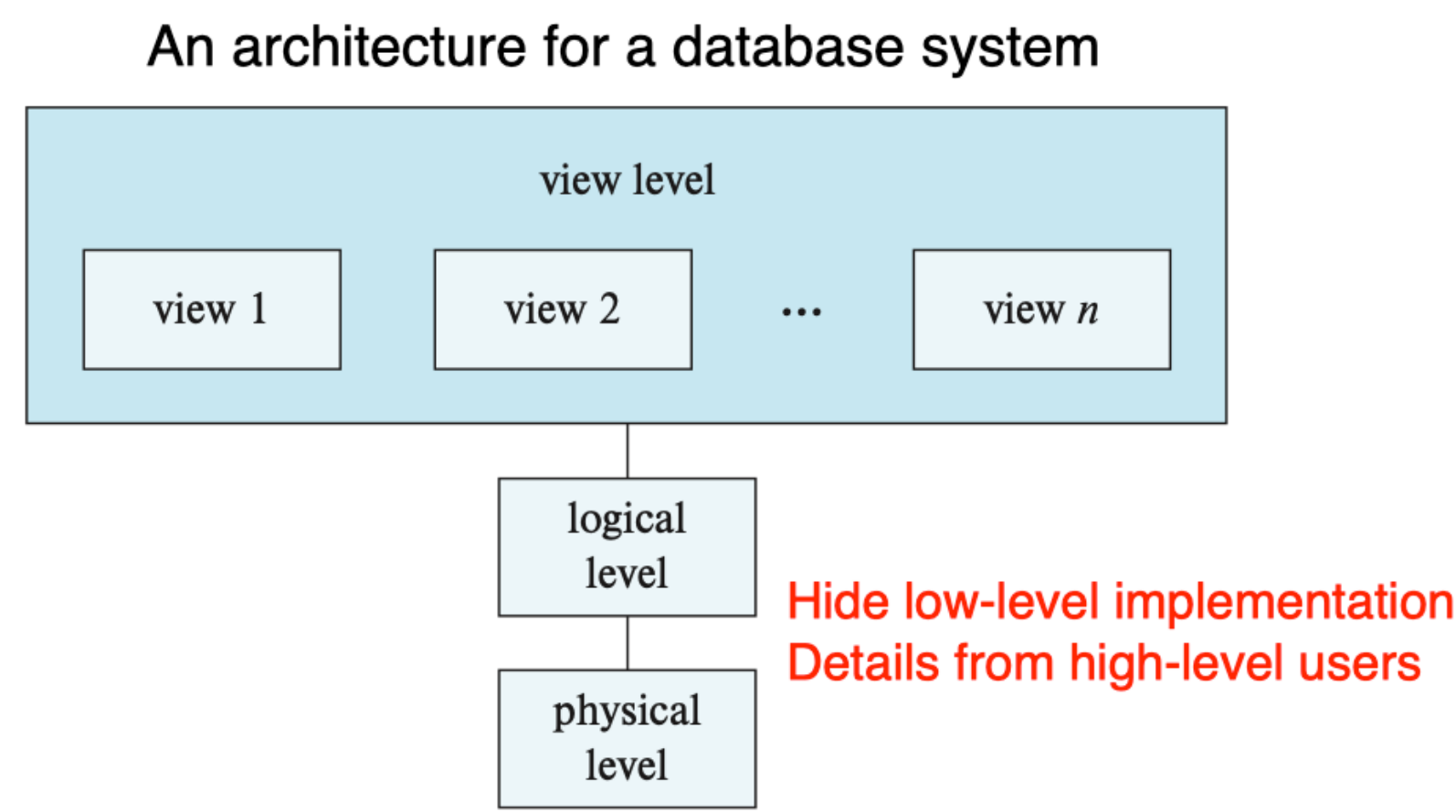
- Database management system (DBMS)
  - A software system that enables users to define, create, maintain, and control access to the database
  - Oracle, MS SQL Server, PostgreSQL, MySQL, SQLite, …

- (Database) Application Programs
  - A computer program that interacts with the database by issuing an appropriate request to the DBMS
    - Typically, an SQL statement

Application Program Connected To Database



View of Data


An architecture for a database system
view level
view 1
view 2
...
view n
logical level
physical level
Hide low-level implementation
Details from high-level users

Data Abstraction

- **Physical level**
  - Describe how the data are actually stored
  - Describes complex low-level data structures in detail

Data Abstraction

- **Logical level**
  - Describes what data are stored in the database, and what relationships exist among those data
  - Describe the entire database in terms of a small number of relatively simple structures
  - Physical data independence
    - Changes to physical level should be possible without having to change the logical level
    - The user of the logical level does not need to be aware of the complex physical-level structures
    - One of the most important benefits of using a DBMS
  - Data administrators, who must decide what information to keep in the database, use the logical level

Data Abstraction

- **View level**
  - Describes only part of the entire database
  - The system may provide many views for the same database

Data Models

- A collection of conceptual tools for describing:
  - **Data**
  - **Data relationships**
  - **Data semantics**
  - **Data constraints**
- Purpose of data modeling is to represent data in an understandable way

Data Models

- Can be classified into four different categories:
  - **Relational Model**
    - Use a collection of tables to represent both the data and the relationships among those data
    - Several Fixed-format types of records
    - Most widely used data model
  - **Entity-Relationship Model**
    - Widely used in database design
  - **Semi-structured Data Model**
    - Same type data may have different set of attributes
  - **Object-Based Data Model**

Relational Data Model

The *instructor* table

id	name	dept_name	salary
Filter...	Filter...	Filter...	Filter...
10101	Srinivasan	Comp. Sci.	65000.00
12121	Wu	Finance	90000.00
15151	Mozart	Music	40000.00
22222	Einstein	Physics	95000.00
32343	El Said	History	60000.00
33456	Gold	Physics	87000.00
45565	Katz	Comp. Sci.	75000.00
58583	Califieri	History	62000.00
76543	Singh	Finance	80000.00
76766	Crick	Biology	72000.00
83821	Brandt	Comp. Sci.	92000.00
98345	Kim	Elec. Eng.	80000.00

The *department* table

dept_name	building	budget
Filter...	Filter...	Filter...
Biology	Watson	90000.00
Comp. Sci.	Taylor	100000.00
Elec. Eng.	Taylor	85000.00
Finance	Painter	120000.00
History	Painter	50000.00
Music	Packard	80000.00
Physics	Watson	70000.00

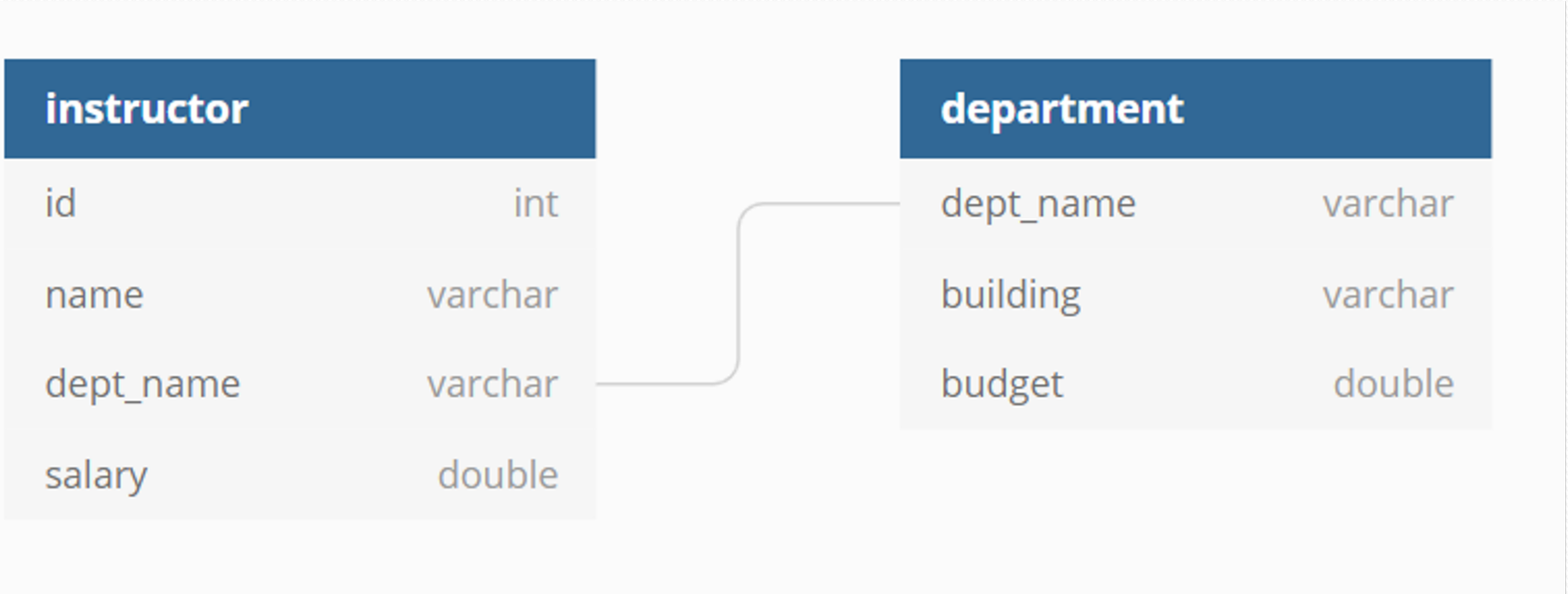
Relational Data Model

- Use a collection of tables to represent both data and the relationships among those data
- Tables are known as relations
- Each table has multiple columns
  - Each column has a unique name
  - Each row represents one piece of information
- Most widely used data model

Relational Model

- **Table**
  - Contains records of a particular type
  - Each record type defines a fixed number of fields, or attributes
- **Columns**
  - Correspond to the attributes of the record type
- **Rows**
  - Each row represent a piece of record, or information

Example of Schema Diagram



Database Language

- **Data-definition language (DDL)**
  - To specify the database schema
- **Data-manipulation language (DML)**
  - To express database queries and updates

- **SQL language**
  - DDL and DML
  - Most widely used
  - Supported by almost all relational database systems

DDL

- Specify the database schema
- Specify the storage structure and access methods used by the database system
  - Data values stored in the database must satisfy certain consistency constraints
  - Different users have different authorization on the database
    - read, insert, update, delete

```
create table department(
    dept_name    char(20),
    building     char(15),
    budget       numeric(12, 2)
);
```

DML

- Enable users to access or manipulate data as organized by the appropriate data model
  - Retrieval of information stored in the database
  - Insertion of new information into the database
  - Deletion of information from the database
  - Modification of information stored in the database

DML

- Two types of DML
  - **Procedural DMLs**
    - Require a user to specify what data are needed and how to get those data
  - **Nonprocedural DMLs** (also referred as **Declarative DMLs**)
    - Require a user to specify what data are needed without specifying how to get those data
    - SQL is the most widely used nonprocedural language
  - It is common practice to use the term query language and data-manipulation language
  - synonymously
    - A query is a statement requesting the retrieval of information

DML

```
select instructor.name
    from instructor
    where instructor.dept_name = 'History';


select instructor.ID, department.dept_name
    from instructor, department
    where instructor.dept_name = department.dept_name and
          department.budget > 95000;
```

Database Users and Administrators

- **Database user**
  - Application programmers
  - Naïve users
  - Sophisticated users

- **Database administrator**
  - Schema definition
  - Storage structure and access-method definition
  - Schema and physical-organization modification
  - Granting of authorization for data access
  - Routine maintenance

History of Database

- **1950s and early 1960s**
  - Data processing using magnetic tapes for storage
    - Tapes provided only sequential access
  - Punched cards for input
- **Late 1960s and 1970s**
  - Hard disks allowed direct access to data
  - Network and hierarchical data models in widespread use
  - Ted Codd defines the relational data model
  - Would win the ACM Turing Award for this work
  - IBM Research begins System R prototype
  - UC Berkeley (Michael Stonebraker) begins Ingres prototype
  - Oracle releases first commercial relational database
  - High-performance (for the era) transaction processing

History of Database

- **1980s**
  - Research relational prototypes evolve into commercial system
  - SQL becomes industrial standard
  - Parallel and distributed database systems
  - Object-oriented database systems
- **1990s**
  - Large decision support and data-mining applications
  - Large multi-terabyte data warehouses
  - Emergence of web commerce

History of Database

- **2000s – 2010s**
  - Big data storage systems
    - Google BigTable, Yahoo PNuts, Amazon
    - "NoSQL" systems
  - Big data analysis:
    - MapReduce
- **2010s**
  - SQL reloaded
    - SQL front end to MapReduce systems
    - Massively parallel database systems
    - Multi-core main-memory databases

Acknowledgements

- WIKIPEDIA
  - https://en.wikipedia.org/wiki/Database