Data dependence in applications

Topic 2
Lesson 3 – Software and hardware architectures

Chapter 1,2

DATABASE SYSTEMS
A Practical Approach to
Design, Implementation, and Management
SIXTH EDITION

THOMAS CONNOLLY
CAROLYN BEGG
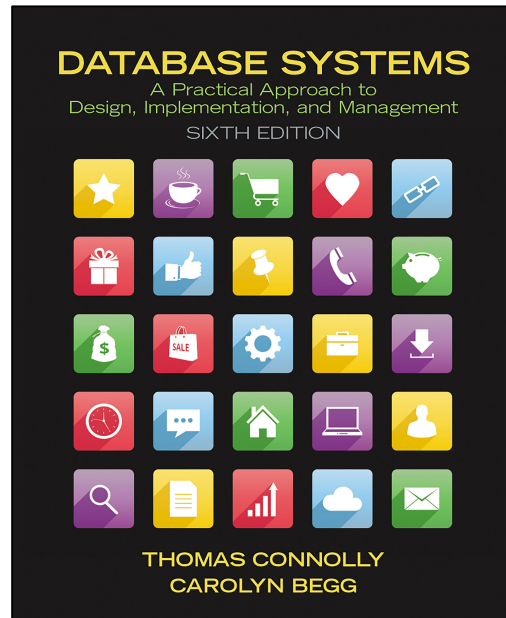
Sections 1.2, 1.3, 2.15, 2.2.1, 2.2.2

Example of data dependence

```python
in_file = open("phonevalues.txt",
                     "rt")
while True:
    in_line = in_file.readline()
    if not in_line:
        break
    in_line = in_line[:-1]
    name, number = in_line.split(",")
    numbers[name] = number
in_file.close()
```

Goal of a database

Data Independence: immunity of a data application to changes in the data

What is a database?

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

System catalog (metadata) provides description of data to enable program's data independence.

Logically related data comprises entities, attributes, and relationships of an organization's information.
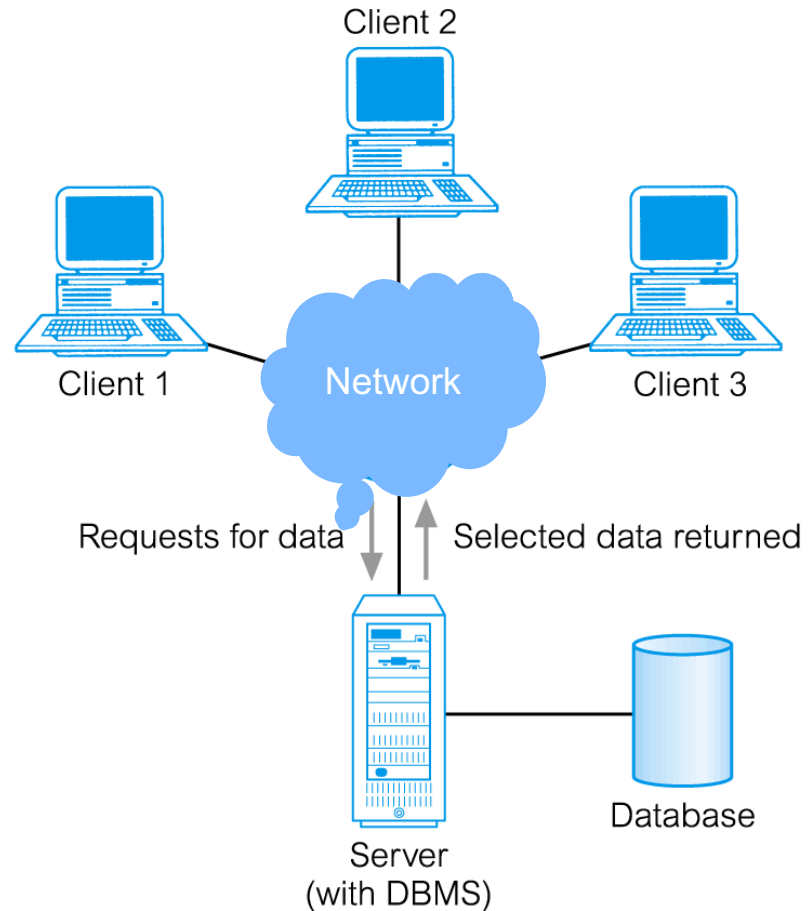
System catalog

- Repository of information (metadata) describing the data in the database. It is updated by DDL commands
- The system catalog is one of the fundamental components of DBMS.
- The system catalog typically stores:
  - names, types, and sizes of data items;
  - constraints on the data;
  - names of authorized users;
  - data items accessible by a user and the type of access granted to a user;
  - usage statistics.

Database approach
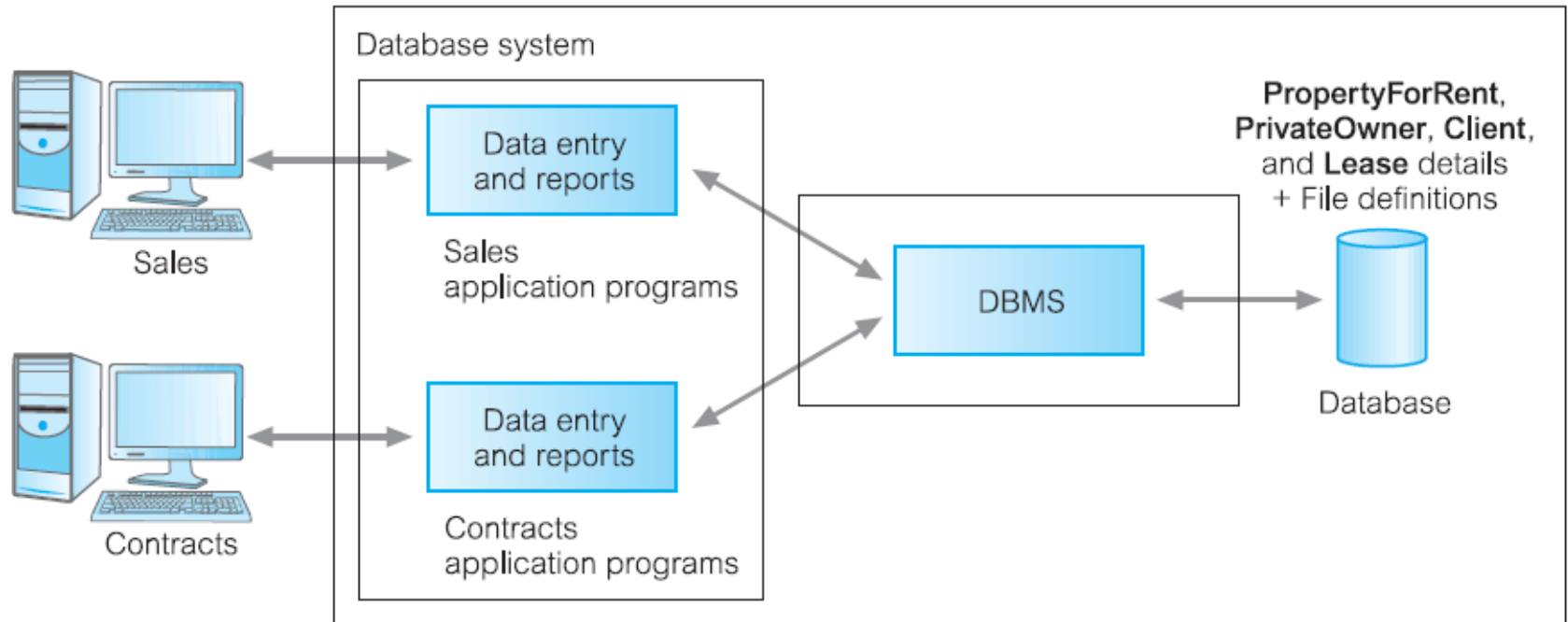
Client-Server Architecture

Portions of the database application deployed on potentially different hardware with a well-defined method for communication

Point to consider…

For relational database systems:

We know that the description of the structure of the data is stored in the system catalog. How does a user define the structure of their data?

How does a user manipulate their data?

Northeastern University
Khoury College of
Computer Sciences

Data sublanguages

DDL or data **definition** language: used to specify the structure of the database schema. The data generated from DDL commands change the system catalog.

DML or the data **manipulation** language is used to read and update the data. The data generated from DML commands changes the data in a user table(s).

Application programmers can embed DDL or DML statements in a high-level programming language such as Python or Java

DDL and DML commands

How does a user define the structure of their data?
DDL Commands

- Create a new table   (CREATE command)
- Update structure of table   (ALTER command)
- Delete a table   (DROP command)

How does a user manipulate their data?
DML Commands

- Create a new tuple (INSERT command)
- Read a tuple   (SELECT command)
- Update a tuple   (UPDATE command)
- Delete a tuple   (DELETE command)

Redundancy in an RDB

- A relational data model does not eliminate data redundancy. It limits or controls the use of data redundancy

- Redundancy is used to represent a relationship between two objects in a relational database.

- A record in a RDB can reference another record in the database – the reference is done by duplicating the name of the referenced object.

Redundancy allowed in a database

Student

sid	name	major
0001	Lucy Liu	**CS**
0002	John Smith	**DS**

Redundancy: Represents a relationship between a student and an available major

Majors

mid	M_name
CS	Computer Science
DS	Data Science
EN	English
PH	Philosophy

Summary

In this module you learned:

- The system catalog and how it supports data independence for database applications
- The use of DDL commands to update the system catalog
- The use of DML commands to manipulate user data