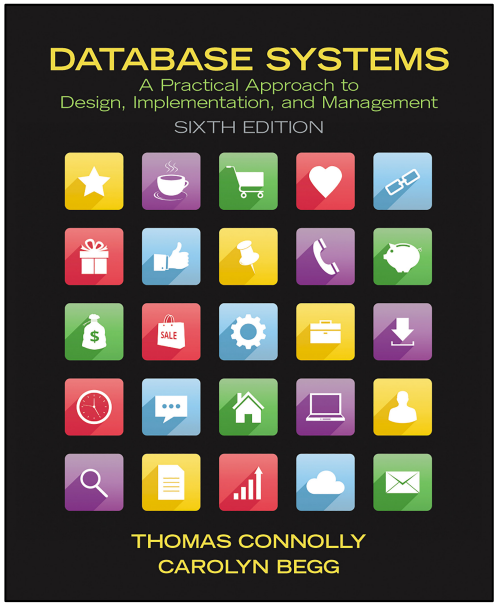
Database introduction

Topic 1
Lesson 1 – Database Basics

Chapter 1 Connolly & Begg

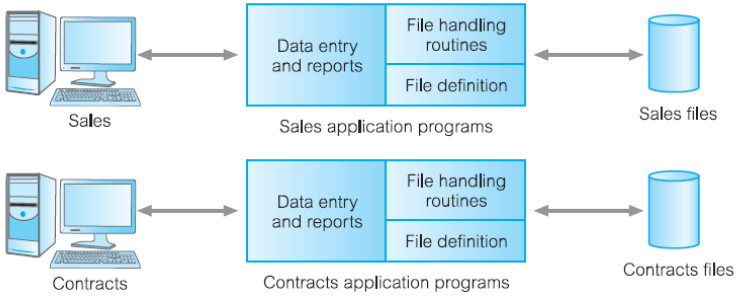


Storing data

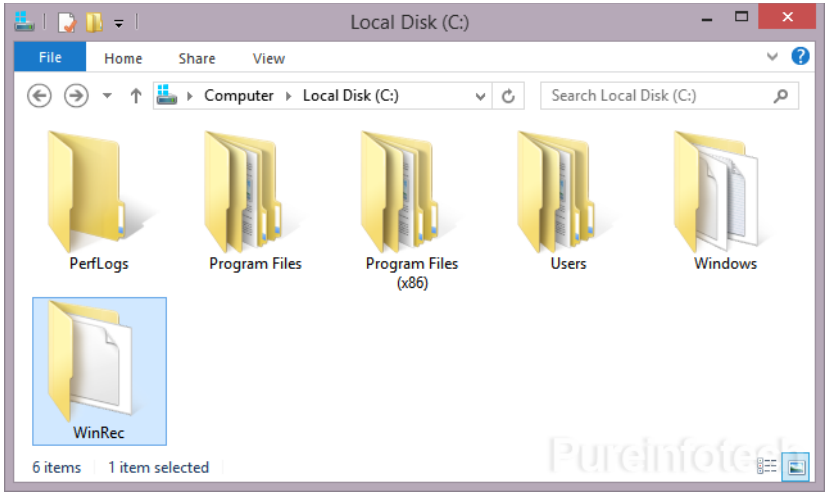
Do I need a database to store data?

File based approach

Of course not ! Many corporations have written applications to access data from data files directly.



As user of computers, we all have data in files on our laptops



Limitations to the file-based approach

What happens when one department wants a copy of the same piece of data?

Limitations to file-based approach

How do I add additional data into the file?

Problems with the file-based approach

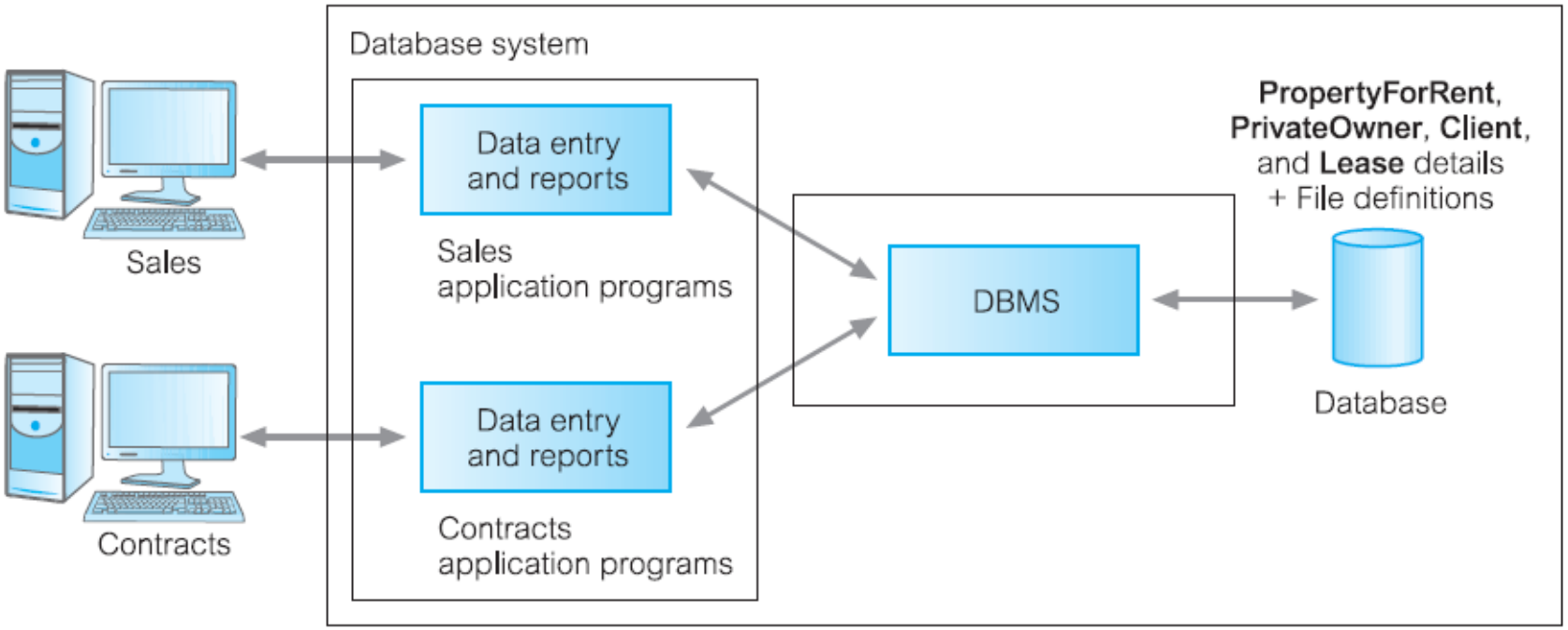
Definition of data was embedded into the application programs, rather than being stored separately and independently.

No control over access and manipulation of data beyond that imposed by application programs.

Example of the file-based approach

```
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()
```

Database approach



Example of the database approach

SELECT name, number FROM phone;

Simpler than writing an application program, at least for the average data consumer

Issue with the database approach

In the file-based approach, applications only had access to their relevant database, but in the database approach I pool all my data together

If I pool all my data together into a database – I might have data that should not been shared with some users.

How can I solve this problem?

Database View

- Allows each user to have his or her own view of the database.
- A view is essentially some subset of the database.
- View mechanism is imperative if we want to secure the data

Benefits?

Discussion 5 W's: What

What is a database?

5W's: 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.

What is a database management system?

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.

Redundancy in an RDB

- A relational data model does not eliminate data redundancy. It limits or controls data redundancy
- Redundancy is used to represent a relationship between two objects in a relational database.

Discussion 5W's: Why

Why use a database?

Let's take some time to list the benefits of a database

Why use a database?

For the following benefits:

- 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

Why use a database?

Even more benefits:

- Balance conflicting requirements
- Improved data accessibility and responsiveness
- Increased productivity
- Improved maintenance through data independence
- Increased concurrency
- Improved backup and recovery services

Any drawbacks to using a database?

Database disadvantages

- Complexity
- Size
- Cost of DBMS
- Additional hardware costs
- Cost of converting your data to a database
- Performance can degrade since time increases size of data
- Higher impact of a failure

5 W's: What role can a database user play?

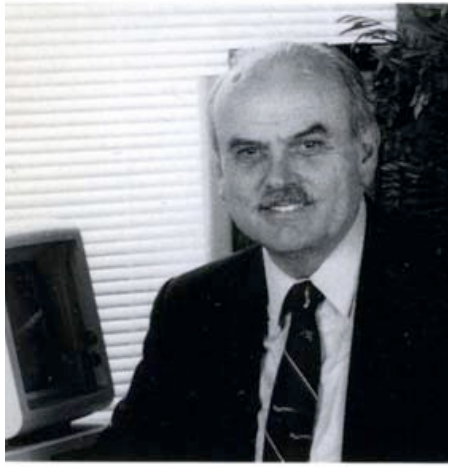
Different views into a database allow different users to accomplish very different tasks on the data and the database.

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

Discussion: 5W's: Who

Who are the RDB pioneers?

E. Codd
IBM
relational model



Michael Stonebraker
U. C. Berkeley
Ingres DB

Chris Date
IBM
System R

Who are the database corporate players?

Work in small groups .

5 Minutes, task: identify corporate providers of databases and the type of database provided.

Feel free to use the Internet to identify them.

5 W's: Who are the corporate players?

Open Source

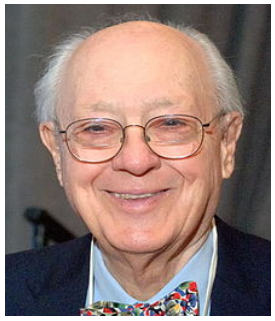
RDB Companies

Computer Companies

5 W's: When did RDBs originate?

- Codd's paper: 'A relational model of data for large shared data banks' (1970)
  - Linked the representation and operations on data to special type of set – a relation
- Date @ IBM starts development of R
  - Boyce & Chamberlin develop SEQUEL based on relational algebra
- Stonebraker, UC Berkeley develops Ingres
  - Stonebraker & Wong develop QUEL based on relational calculus

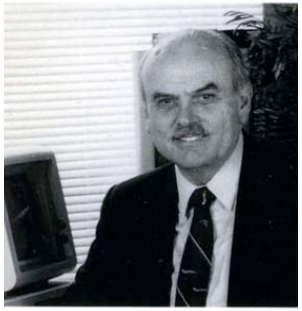
Bachman



Bachman vs. Codd debate
1974 @ ACM SIGMOD
the "Great Debate" Panel

Integrated Data Services (IDS) developed at GE by Charlie Bachman in the 1960s

Codd



5 W's: Where can you find databases?

5 W's: Where can you find databases?

Everywhere..

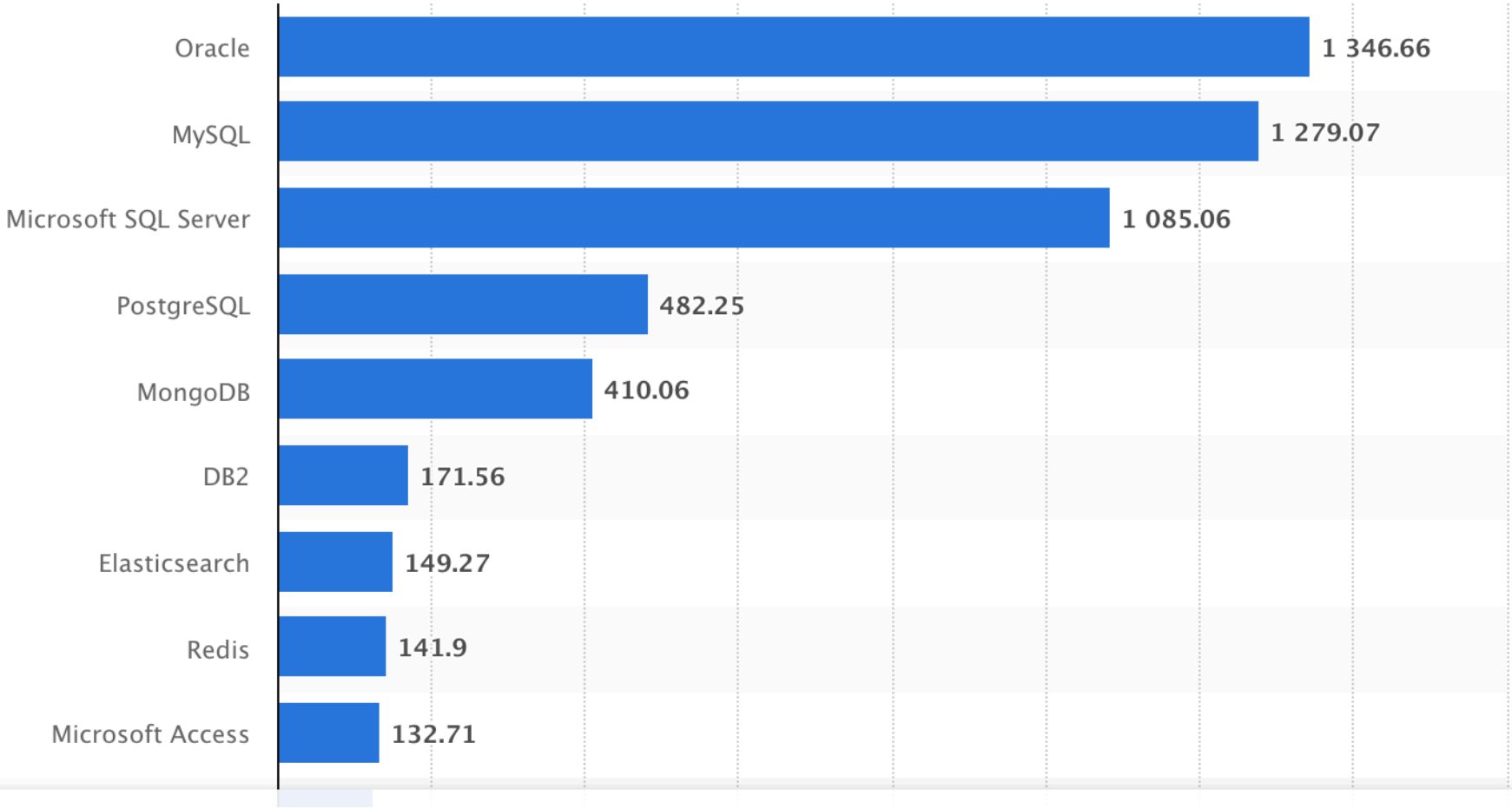
World Data Centre for Climate – 6 Petabytes
National Energy Research Scientific Computing Center (NERSC) – 2.8 petabytes
AT&T – 312 Terabytes
Google – 91 million searches per day
Sprint - 70,000 call record insertions per second

LexisNexis – 3 petabytes on Americans
YouTube – 2,000 petabytes

Amazon - 75 petabytes of data

CIA – all content digitized: growth rate of 100 articles per month
Library of Congress – index for the library: growth rate of 10,000 items per day
Uber – 100 petabytes for an analytical database

DB popularity as of September 2019



https://www.statista.com/statistics/809750/worldwide-popularity-ranking-database-management-systems/

Summary

In this module you learned:

- The definition of a database and a database management system
- The file-based approach and its limitations
- The database approach and its limitations
- The different user roles for a database