Overview of Database Systems

PDBM 1.1–1.5

Dr. Chris Mayfield

Department of Computer Science
James Madison University

Jan 18, 2022

Ice breaker

Discuss in small groups:

1. What is a database?
2. What are common examples?
3. Why should we study databases?

What is a database?

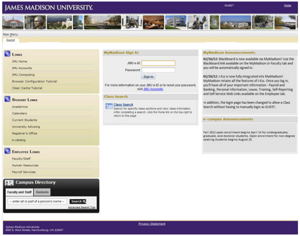
In essence:

- a collection of information, that
- exists over a long period of time

Managed by a DBMS

1. Support storage of large amounts
2. Allow users to specify a schema
3. Give users ability to query data
4. Enable durability and recovery
5. Control concurrent access to data

What are common examples?



MyMadison Sign In

For more information on your JMU e-ID or to reset your password, visit JMU Accounts.

Class Search

Class Search
Search for specific class sections and view class information. After completing a search, click the Home link on the top-right to return to this page.

MyMadison Announcements

02/16/12: Blackboard is now available via MyMadison! Use the Blackboard link available on the MyMadison or Faculty tab and you will be automatically signed in.

01/26/12: J-Ess is now fully integrated into MyMadison! MyMadison retains all the features of J-Ess. Once you log in, you'll have all of your important information - Payroll and Banking, Personal Information, Leave, Training, Self-Reporting and Self Service Web Links available on the Employee tab.

In addition, the login page has been changed to allow a Class Search without having to manually login as GUEST.

e-campus Announcements

Fall 2012 open enrollment begins April 14 for undergraduate, graduate, and doctoral students. Open enrollment for non-degree seeking students begins August 20.

James Madison University
800 S. Main Street, Harrisonburg, VA 22807

And many others!

Why study databases?

Academic

- Databases involve many aspects of computer science
- Well-established and very active area of research
  - Multiple Turing awards in databases

Business

- Everybody needs databases → lots of money to be made

Programmer

- Many applications involve using and accessing databases

Student

- Databases are so cool!
- Google/etc will hire me!
- Need those last 3 credits!

Database management systems

Commercial

Microsoft®
SQL Server®

Open Source

PostgreSQL

Why use a DBMS?

Manage

- Store and process large amounts of data

Organize

- Give structure (i.e., schema) to the data

Query

- Extract interesting/relevant information

Data Independence

"The ability to change the organization of the database itself without changing the application software." (Brookshear 12/e)

Features of a DBMS

Support massive amounts of data

- Far too big for main memory (GB / TB / PB)
- “Recent” trend: databases run on single computers

Persistent storage

- Applications update, query, manipulate data
- Data continues to live long after apps finish

Efficient and convenient access

- Do not search entire database to answer a query
- Tools for users to create and query the data

Concurrent, and atomic access

- Allow multiple users to access database simultaneously
- Provide some guarantee of reliability against failures

Transaction processing

Database operations are grouped into transactions

Transactions should meet ACID requirements:

- Atomicity: All-or-nothing execution of transactions.
- Consistency: Should NOT violate DB's constraints.
    - If it does, it needs to be rolled back
- Isolation: Each transaction must appear to be executed as if no other transaction is executing at the same time.
    - Changes become visible only after committed
- Durability: Any change a transaction makes to the database should persist and not be lost.

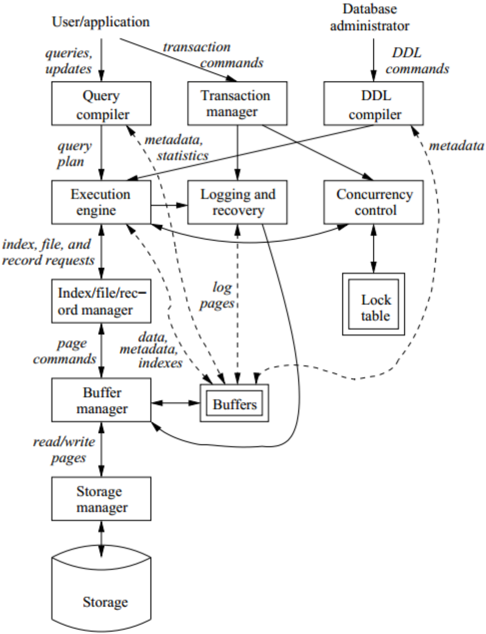


Figure 1.1: Database management system components

A brief history of DBMSs

The earliest DBMSs (1960s) evolved from file systems

- Navigational and hierarchical
- User programmed queries by walking from node to node

Relational DBMS (1970s to now)

- View database in terms of relations or tables
- High-level query and definition languages such as SQL

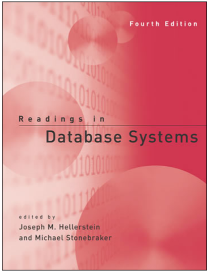
Object-oriented DBMS (1980s)

- Inspired by object-oriented languages
- Object-relational DBMSs

“New” types of data:

- Semi-structured data (XML, JSON)
- Data streams (continuous queries)

Two great overview papers



What Goes Around Comes Around
PDF Link

Anatomy of a Database System
PDF Link

by Michael Stonebraker
and Joseph M. Hellerstein

Today's (yesterday's?) databases

RDBMS = Relational DBMS

- The relational model uses *relations* to structure data
- Separates logical view (externals) from physical view (internals)

`ClassList` relation:

Student	Course	Grade
Hermione Grainger	Potions	A-
Draco Malfoy	Potions	B
Harry Potter	Potions	A
Ron Weasley	Potions	C

Structured query language (SQL) for accessing/modifying data:

```
SELECT student FROM roster WHERE grade >= 'B';
```

SQL vs Python

Declarative programming:

```sql
SELECT student FROM roster WHERE grade >= 'B';
```

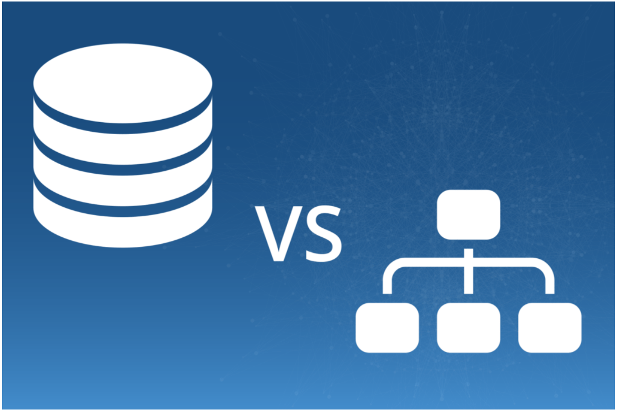
Imperative programming:

```python
import csv

def main():
    data = open("roster.csv")
    data = csv.reader(data)
    for row in data:
        student = row[4]
        grade = row[7]
        if grade >= "B":
            print(student)

if __name__ == "__main__":
    main()
```

Data storage



https://raima.com/database-system-vs-file-system/

Administrivia

Welcome back!

Course logistics

Course home page: https://w3.cs.jmu.edu/mayfiecs/cs374

Find Q&A on Discord: https://discord.com

Homework and grades: https://canvas.jmu.edu

Your TODO List

- Read the Syllabus if you haven't already
- See other items on today's lesson outline
- Will try to form project teams next week

Grade requirements

Assignments 20%

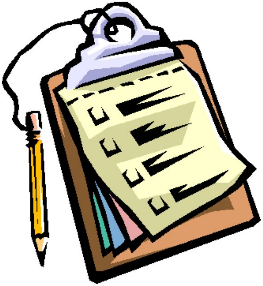
- Written problems
- SQL programming

Group Project 30%

- Use lots of public data
- Design database/queries
- Build a web application

Midterm #1 25%

Midterm #2 25%



Tips for success

Come to class prepared

- Read (parts of) the textbook

Start homework early

- Something due each week

Stop by office hours

- Mon/Wed/Fri 2:20–4:00 PM
- Other times by appointment

HAVE FUN!