Introduction to Databases

Data Management and File Structure
CENG 351

General Concepts

- Database definition
  - A collection of logically related data
- Data
  - Known facts
  - Types: text, graphics, images, sound, videos
- Database management system (DBMS)
  - Software for defining and managing a database

Database Examples

- Hospital patients
- Literature (published articles in a certain field)
- Student information system
- Library system
- Music shop
- Airline reservation

Example: Library Database

Data about the books in a library, the clients of the library, and the books borrowed.

Books:

- Name
- Author
- ISBN number
- Publisher
- Year published
- …..

Clients

- Name and surname
- Address
- Telephone
- Membership start date
- …..

Borrow

- Book
- Client
- Borrow date
- Due date
- …..

Database Models

- <u>Flat files</u>   '60
- Hierarchical   '60
- Network   '70
- <u>Relational</u>   '80
- Object oriented   '90
- Object relational   '90
- Web enabled   '90

Flat Files

Characteristics:

- Data is stored as records in regular files

- Records usually have a simple structure and fixed number of fields

- For fast access may use indexing of fields in the records

- No mechanisms for relating data between files

Relational Database

Characteristics:

- Data is organized into tables: rows & columns
- Each row represents an instance of an entity
- Each column represents an attribute of an entity
- Relationships between entities are represented by values stored in the columns of the tables (keys)
- Accessible by Standard Query Language (SQL)

Relationships

- Used to connect tables

- Field(s) that have the same value in the related tables are used

- Borrow.ClientName=Client.Name creates relationship between borrow and client

Database Integrity

- If a book is deleted from books table, the record in borrow table is not valid.

- Database Management Systems preserve integrity

- Preserving integrity needs a lot of file access operations

SQL

- Standard Query Language (SQL) is a computer language for accessing and updating database systems.

- SQL statements are used to retrieve and update data in a database.

Updating Data

Syntax for executing queries, updating, inserting, and deleting records.

- SELECT - extracts data from one or more table
- INSERT INTO - inserts new data into a table
- UPDATE - updates data in a table
- DELETE FROM - deletes data from a table

Example

Select all clients who borrowed Database book from the library between 1-1-2012 and 1-1-2013.

SELECT Name

FROM Client

WHERE

Borrow.Name=Client.Name **AND**

Borrow.Borrow_date >='1-1-2012' **AND**

Borrow.Borrow_date <='1-1-2013' **AND**

Borrow.Book='Database'

Running Queries

- Running Queries can take a very long time

- For tables, we create indexes (B+Trees)

- For temporary tables, often hash tables are used

File Syllabus Revised

What will I learn in this course?

You will learn about:

- How the files are stored on hard disks.
- How files are organized for fast access including:
  - Sorting
  - Indexing
  - Hashing
- How efficient file access algorithms can be developed.
- How you can implement and evaluate these algorithms in C

Course Outline (1)

- Introduction:
  - The need for managing large dynamic files
  - Operations on files
  - Definitions
- Disks
  - Disk hardware structure
  - Timing on disk I/O
  - Records and buckets
  - Double buffering

Course Outline (2)

- Sequential files
  - Pile files
  - Sorted sequential files
- Sorting
  - Internal sorting
  - External sorting
- Indexing
  - Linear indexes
  - Tree indexing
  - B-Trees
  - B+Trees
  - Operations on B+Trees

Course Outline (3)

- Hashing
  - Hashing with buckets and chaining
  - Solving collisions in hashing by combining bucketing and chaining
- Introduction to Database Management Systems

Questions?