CS457/CS557 - Data Base Processing

Course Description

CS457 provides an introduction to databases and data models. It will introduce students to the foundations of data base systems and cover basic topics such as relational algebra, the evolution of data models, and techniques of query optimization and processing. Students will have an opportunity to develop their skills using technologies such as SQLITE, Postgres, MySQL, MongoDB and Redis.

The principal goals will be:

- Learn to effectively use SQL to query and maintain databases.

- Design data relational database.

- Learn the basic alogorithms used in database implementations.

- Understand the characteristics of various classes of databases, including NoSQL databases.

- Learn the JSON format and how to query Mongo databases.

Course Outline

We will try to follow the same pattern each week (calendar permitting). On Monday I will discuss a topic for for the week. Discuss the new ideas for the week and work through a few basic examples. On Wednesday we will work though examples together (on the `cs.instate.edu` server), and I will give you an assignment which will be due to the following Monday. By Friday you should have had to time to think about the assignment, so I will set aside time to answer questions, and to work through some more difficult examples.

- Week 1 - Class protocols, course objectives, using Linux.

    - Summary
        * We will discuss the database concept, what capabilities we require in a data base, and a few different types of databases.
        * There will be a problem set on relational algebra.
    - Overview of Data Models and Databases
        * Relational Model - our Topic for the first half of the course
            · Relations
            · Relational Algebra
            · Examples of Relational Databases: SQLite, Postgres, DB2, Oracle, Sybase, MySQL, MariaDB
        * Other Database Models
            · Hierarchical Model

- · Network Model
- · Semantic Model
- · Document Model
- · Object Oriented Model
- · Semi-structured Models

- • Week 2 - SQLite: the Basics
  - – Summary
    - * We will learn to use SQLite3, and begin learning SQL.
    - * There will be a programming assignment that involves reading a data file into an SQLite table and doing some SELECT queries.
  - – What is SQL?
    - * SQL is each of the following: DML, DDL, DCL, DQL, TCL
  - – What is SQLite?
  - – How to Use SQLite
    - * Databases and Tables
      - · Creating tables
      - · Importing data
  - – Basic queries
    - * CREATE, INSERT, SELECT
  - – Modifying tables
    - * INSERT, UPDATE, DELETE

- • Week 3 - Complex Queries
  - – Summary
    - * We will finish a first pass over the SQL language.
    - * There will be a programming assignment involving multiple tables.
  - – The SQL Language
    - * Data Definition Language: CREATE, DROP
    - * Data Query Language: SELECT
    - * Data Manipulation Language: DELETE, INSERT, UPDATE
    - * Data Control Language: GRANT, REVOKE
    - * Transaction Control Language: ROLLBACK, COMMIT, SAVEPOINT
  - – SQLite Syntax Diagrams

- • Week 4 - SQL Practice and Triggers
  - – Summary
    - * Triggers will be introduced. Triggers are one way to turn a static database into a dynamic entity.
    - * The week 3 assignment will be expanded using triggers.
  - – Triggers
    - * Functions that are automatically executed in response to a database event.

* DML triggers.
  * DDL triggers.
  * Other events.

- Week 5 - Efficiency: Data Base Design

  – Summary

    * Design databases that maximize integrity and minimize redundancy
    * Begin the database design project.

  – Define the goals for the database design.

    * Online Transaction Processing
    * Decision Support Systems

  – Normalization

    * Example of Normalization Process
    * Anomalies: Insertion, Deletion, Update
    * Extensibility

- Week 6 - Building a Working Data Base

  – Summary

    * Apply Normalization Process to our design.
    * Complete the database design process

- Week 7 - Database Internals

  – Summary

    * Study some of the algorithms and data structures used in databases.
    * The project for the week will involve implementing a tree data structure in C.

  – Algorithms and Trees

    * B Trees
    * B+ Trees
    * Tries
    * Hash tables
    * Hash buckets
    * Heap files

- Week 9 - Relational Client Server Databases

  – Summary

    * Study the new issues related to the client/server model.
    * Implement an earlier project in a client/server environment.

  – Brief discussion of MYSQL, MariaDB, Oracle, Postgres, etc.

  – Concurrency issues.

    * Consistency.
    * Lost data.

  – Examples and Exercises MariaDB or Postgres.

- Week 10 - Data Exchange Formats - XML and JSON

    - Summary
        * A brief review of Javascriptm, followed by a study of JSON Data structures and files.
        * The assignment will involve parsing JSON in a high level language.
    - Using JSON
        * Javascript Examples
        * Python Examples

- Week 11 - NoSQL: Mongo

    - Summary
        * Introduction to the MongoDB and the Mongo Shell
        * The project for the week will involve doing a list of query exercises on a database created by the instructor.
    - Mongo, JSON and Javascript
        * The Mongo Shell
        * MongoDB CRUD operations.
        * Sharding and an introduction to big data.

- Weeks 12 and 13 - Mastering Mongo Syntax

    - Summary
        * Learn to use the aggregation pipeline and the Map-Reduce perspective.
        * The project for the week will involve complex, multi-stage queries.
    - MongoDB
        * Map-Reduce
        * Aggregation pipeline
        * Operations

- Week 14 and 15 - More on Big Data

    - Summary
        * A brief introduction to Data Mining.
        * There will be a programming assignment on a big data problem.
    - The Stream Data Model
        * Sub-log-space counting.
        * Exotic hashing methods.
        * Detecting duplicates.
        * Heavy hitters.

Prerequisites. Students need to have completed CS202 or CS500, or received permission of the instructor.

Requirements. Assignments must be done on time.

Time: MWF 1:00

Room: Online

Professor: G. Exoo

Email: `cs457@cs.indstate.edu`

Office: Online

Office Hours: ???

Grading:

Class Participation	20%
Weekly assignments	70%
Final Exam	10%

Grading Scale:

92	-	100	A
90	-	91	A-
88	-	89	B+
82	-	87	B
80	-	81	B-
78	-	79	C+
72	-	77	C
70	-	71	C-
68	-	69	D+
60	-	67	D
0	-	59	F

Important Web Links

Class Web Page

CS457 on cs.indstate.edu

Academic Integrity Policy

Student Guide

It's On BLUE.

Indiana State University fosters a campus free of sexual misconduct including sexual harassment, sexual violence, intimate partner violence, and stalking and/or any form of sex or gender discrimination. If you disclose a potential violation of the sexual misconduct policy I will need to notify the Title IX Coordinator. Students who have experienced sexual misconduct are encouraged to contact confidential resources listed below. To make a report to the Title IX Coordinator, visit the Equal Opportunity and Title IX website.

Confidential Resources:

The ISU Student Counseling Center
HMSU 7th Floor
812-237-3939.

Campus Ministries

For more information on your rights and available resources.

All things related to Covid-19

The ISU Victim Advocate: Leah Reynolds

HMSU Room 813

812-237-3829 (office)

812-243-7272

leah.reynolds@indstate.edu