

# C S 84B: DISTRIBUTED DATABASES

## Foothill College Course Outline of Record

Heading	Value
<b>Effective Term:</b>	Summer 2021
<b>Units:</b>	4.5
<b>Hours:</b>	4 lecture, 2 laboratory per week (72 total per quarter)
<b>Advisory:</b>	C S 31A or equivalent.
<b>Degree &amp; Credit Status:</b>	Degree-Applicable Credit Course
<b>Foothill GE:</b>	Non-GE
<b>Transferable:</b>	CSU
<b>Grade Type:</b>	Letter Grade (Request for Pass/No Pass)
<b>Repeatability:</b>	Not Repeatable

## Student Learning Outcomes

- Design a distributed database with implementation strategies to maintain transaction and concurrency control
- Develop query processing and optimization strategies for an existing distributed database design
- Develop data replication and integration plans for an existing distributed database design

## Description

An introduction to distributed data management including distributed database design, implementation techniques including concurrency control, query processing and optimization, data replication, integration and peer-to-peer systems. Distributed database solutions are also presented, including data management systems for Cloud computing.

## Course Objectives

The student will be able to:

- Design a distributed database
- Design implementation techniques for transactions and concurrency control
- Design query processing and optimization strategies
- Design data replication and integration plans
- Design a schema mapping for a peer-to-peer system
- Evaluate database management solutions for distributed databases including solutions for Cloud applications

## Course Content

- Distributed Database Design
  - What is a distributed database system?
- Promises and Complications of Distribution
- Design Issues
- Architecture
- Top-Down Design Process
- Transaction Implementation Techniques
  - Definition of a transaction
  - Properties and types of transactions
- Architecture Revisited for Transaction Management
- Concurrency Control

- Serializability Theory
- Locking-Based Concurrency Control Mechanisms
  - Timestamp-based concurrency control mechanisms
  - Optimistic concurrency control
- Deadlock Management
- Query Processing and Optimization
  - Issues in multi-database query processing
  - Multi-database query processing architecture
  - Query optimization and execution
  - Timestamp-based concurrency control mechanisms
  - Optimistic concurrency control
  - Deadlock management
- Data Replication
  - What is a replicated database?
  - Consistency of replicated databases
  - Update management strategies
  - Replication protocols
- Data Integration
  - Bottom-up design methodology
  - Schema matching
  - Schema integration
  - Schema mapping
- Peer-to-Peer Systems
  - Infrastructure
  - Querying over P2P systems
- Distributed Database Solutions
  - Hadoop
  - Map-Reduce and Pig
  - Publish/subscribe systems
- Data Management in the Cloud
- Cloud Architectures
- Data Management Systems for Cloud Computing
  - BigTable
  - Map-Reduce
  - PNUTS

## Lab Content

- Create a distributed database design given predefined set of requirements
- Evaluate transaction implementation techniques including the appropriate transaction locking given a set of business constraints
- Fill in gaps or inconsistencies in a set of concurrency control rules for an existing distributed database design
- Evaluate a query processing and optimization plan for an existing distributed database design
- Complete a data replication plan given a partial plan
- Prepare a schema mapping for a given incomplete data integration plan
- Evaluate a given schema mapping for a peer-to-peer system
- Recommend one open source solution given a real-world scenario
- Compare and contrast two data management systems for Cloud computing

## Special Facilities and/or Equipment

- A website or course management system with an assignment posting component (through which all lab assignments are to be submitted) and a forum component where students can discuss course material and receive help from the instructor). This applies to all sections, including on-campus (i.e., face-to-face) offerings.
- When taught via Foothill Global Access on the Internet, the college will provide a fully functional and maintained course management system through which the instructor and students can interact.

C. When taught via Foothill Global Access on the Internet, students must have currently existing email accounts and ongoing access to computers with internet capabilities.

## Method(s) of Evaluation

Exams and quizzes

Distributed database design assignments

Project including implementation design for a distributed database for a real-world scenario

3. A description of the database including data type, valid data ranges, constraints and keys.

4. A help page for users of the web application. This may be an FAQ or user manual style help page.

## Discipline(s)

Computer Science

## Method(s) of Instruction

Lectures

Online labs (including sections meeting face-to-face/on-campus), consisting of:

1. An assignment webpage located on a college-hosted course management system or other department-approved internet environment. Here, the students will review the specification of each assignment and submit their completed lab work
2. A discussion webpage located on a college-hosted course management system or other department-approved internet environment. Here, students can request assistance from the instructor and interact publicly with other class members
3. Detailed review of assignment which includes model solutions and specific comments on the student submissions
4. In-person or online discussion, which engages students and instructor in an ongoing dialog, pertaining to all aspects of database management systems

When course is taught fully online:

1. Instructor-authored lecture materials, handouts, syllabus, assignments, tests, and other relevant course material will be delivered through a college-hosted course management system or other department-approved internet environment

## Representative Text(s) and Other Materials

Ozsu, M. Tamer, and Patrick Valduriez. [Principles of Distributed Database Systems, 3rd ed.](#) 2011.

Klepman, Martin. [Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems.](#) 2017.

## Types and/or Examples of Required Reading, Writing, and Outside of Class Assignments

A. Reading:

1. Textbook assigned reading averaging 30 pages per week.
2. Reading online resources as directed by instructor through links pertinent to databases.
3. Reading library and reference material directed by instructor through course handouts.

B. Writing:

1. Technical prose documentation, that supports and describes the database-driven web application assignments that are submitted for grades. The document would include the following aspects of the database application:
  2. A description of the web application including functional and data requirements.