
CSE 344 Introduction to Data Management

Credits

4.0 (3 hrs lecture, 1 hr section)

Lead Instructor

Dan Suciu

Textbook

- *Database Systems: The Complete Book*, Garcia-Molina

Course Description

Introduces database management systems and writing applications that use such systems; data models (e.g., relational, semi-structured), query languages (e.g., SQL, XQuery), language bindings, conceptual modeling, transactions, security, database tuning, data warehousing, parallelism, and web-data management.

Prerequisites

either CSE 311 or CSE 321.

CE Major Status

Selected Elective

Course Objectives

At the end of this course, students should:

1. Understand basic concepts related to data management such as the notion of data models, physical and logical data independence, query languages, query execution, and cost-based query optimization.
2. Acquire knowledge of the relational data model including the relational algebra, datalog and the relational calculus.
3. Acquire knowledge of the semi-structured data model.
4. Learn several declarative query languages such as SQL, XQuery, and Pig Latin
5. Know how to use a relational database management system and a semi-structured database management system. This includes designing a database, loading data into the database, updating the database, and querying the database.
6. Understand the notion of transaction processing, ACID properties, and different degrees of consistency.

7. Write applications that use a database management system as a back-end. Know how to use transactions when developing such applications.
8. Understand the principles of parallel data processing and know how to use a parallel relational database system or a MapReduce system.
9. Understand the concept of data management as a service in the cloud and gain basic experience with using data management systems hosted in a public cloud such as SQL Azure and Amazon Elastic MapReduce.
10. Understand challenges of data management related to data integration and data cleaning.
11. Gain experience managing a large dataset (e.g., 0.5 TB graph).

ABET Outcomes

- (a) an ability to apply knowledge of mathematics, science, and engineering
- (c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability
- (j) knowledge of contemporary issues
- (k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice

Course Topics

- Introduction to the notion of data models and the relational data model (2 lecture hours)
- The declarative query language SQL (4 lecture hours)
- Relational algebra, relational calculus, and datalog (3 lecture hours)
- Query evaluation and database tuning (2 lecture hours)
- The semi-structured data model: XML, XPath, and XQuery (2 lecture hours)
- Database design (4 lecture hours)
- Transaction processing (3 lecture hours)
- Parallel data processing (4 lecture hours)
- Data management as a service in the cloud (1 lecture hour)
- Data integration and data cleaning (2 lecture hours)