

## **Approval: 9<sup>th</sup> Senate Meeting**

Course Name:	Information and Database Systems
Course Number:	CS309
Credits:	3-0-2-4
Prerequisites:	CS 207 Applied Databases Practicum
Intended for:	UG
Distribution:	Compulsory for CSE; CS elective for EE and ME
Semester:	6th

### ***Preamble:***

In line with the philosophy of IIT Mandi curriculum where experiments are done before formal theory is introduced, this course comes after Applied DB practicum. While in Applied DB practicum emphasis is on hands on practice of SQL, this course provides an in-depth view of modern databases, mainly relational databases. The course also introduces the concepts related to information systems in organisational usage and considers the different models of information modeling. Further, in recent times, popular relational database systems like [DB2](#), [SQLServer](#), [Oracle](#), and [Sybase](#) are struggling to handle the massive scale of data introduced by the Web. For example, Facebook absorbs 15 TeraBytes of data each day while eBay maintains a 6.5 PetaByte of data. A new class of database systems is therefore emerging to handle data at this scale. The course offers exposure of the emerging technologies to handle Big Data and lays foundation for advances course on database technologies.

### ***Course Outline:***

The students will be exposed to the core concepts in information and database systems. The focus will be on building the skills of identifying organizational information requirements, modeling them using conceptual data modeling techniques, converting the conceptual data models into relational data models and verifying its structural characteristics with normalization techniques, and implementing and utilizing a relational database. Students will also learn transaction management, ACID properties of transactions, online transaction processing and use of stored procedures and triggers as well as query processing in database management systems. Emerging technologies to handle Big Data like Hadoop will be introduced briefly.

Two mini-projects and one final project in the course will exercise students' knowledge of database design, web programming, and information integration, to build real world applications.

Modules:

1. Introduction (6 hours)
  - a. Information Modeling: background, approaches, information system lifecycle.
  - b. Four information levels: conceptual, logical, physical, external
  - c. Conceptual Schema Design Procedure : facts, constraints, roles, value, set comparison, final checks
2. Relational Database Design (6 hours)
  - a. Overview of ER, Barker notation, mapping from ORM to ER
  - b. Overview of UML, mapping from ORM to UML
  - c. Relational schemas, functional dependencies, Normal forms
3. Data Manipulation with SQL (9 hours)

- a. Relational Algebra
  - b. SQL : Basic operations, Joins, Nested and correlated queries, views, Triggers
  - c. Embedded SQL and database application development
4. Transactions (9 hours)
  - a. ACID properties
  - b. Concurrency Control Techniques
  - c. Recovery Techniques
5. Principles of query processing (3 hours)
  - a. Indexes
  - b. Query plans and operators
  - c. Cost-based query optimization
6. Data storage (3 hours)
  - a. Databases Vs. FileSystems (Google FileSystem, Hadoop Distributed FileSystem)
7. Scalable data processing (6 hours)
  - a. MapReduce and introduction to systems based on MapReduce (Hadoop)
  - b. Introduction to Scalable key-value stores (Amazon Dynamo, Google BigTable, HBase)

***Textbooks:***

1. Information Modeling and Relational Databases, Second Edition ((The Morgan Kaufmann Series in Data Management Systems) by [Terry Halpin](#) and [Tony Morgan](#))

***References:***

1. Fundamentals of Database Systems, 6th edition by Elmasri, Ramez and Navathe, Shamkant
2. Raghu Ramakrishnan and Johannes Gehrke, *Database Management Systems*, 3rd ed, McGraw-Hill
3. Assorted Readings from web resources like Yahoo's Hadoop Tutorial, Google's Big Table etc.