

Database Design

Course abstract:

Module 1: Introductory Concepts (1 session)

Databases and Information Systems, An example usage context, Database system concepts and architecture.

Module 2: Semantic Database Design (3 sessions)

High-level conceptual modeling, ER Modeling concepts, ER Diagrams, Cardinality constraints, Higher-order relationships, Enhanced ER Model (EER), Weak-entity types, Subclasses and inheritance, Specialization and Generalization, Modeling of UNION types using categories

Module 3: Relational Model, Languages and Systems (7 sessions)

Relational algebra (2 sessions): Relational model concepts, Relational integrity constraints, Update operations on relations, Relational algebra model, ER to relational mapping

SQL (2 sessions): Data definition in SQL, Queries and update statements, Views, Integrity constraints, Specifying indexes, Embedded SQL.

IBM DB2 case study (2 sessions): Architecture of DB2, Data definition and manipulation in DB2.

EER to Relational mapping (1 session)

Module 4: Database design using the relational model (4 sessions)

Functional dependencies (2 sessions): Keys in a relational model, Concept of functional dependencies, Normal forms based on primary keys, Boyce-Codd Normal Forms

Further Dependencies (2 sessions): Multi-values dependencies and fourth normal form, Join dependencies and fifth normal form, Inclusion dependencies, Other dependencies and normal forms

Module 5: Storage and Indexing Structures (6 sessions)

Storage structures (3 sessions): Secondary storage devices, Buffering of blocks, File Organization, Heaps, Sorted Files, Hashing and overflow handling techniques, Dynamic hashing, Extensible hashing, Other file organizations

Indexing methods (3 sessions): Basic terminology, Primary indexes, Clustering index, Secondary index, Multilevel indexes, ISAM, B-trees, B+ trees, inserting and searching algorithms for B+ trees, Other indexing methods

Module 6: Transaction Processing and Concurrency Control (6 sessions)

Transaction Fundamentals (3 sessions): OLTP environments, Concurrency issues, need for transactions, Necessary properties of transactions (ACID properties), Transaction states, serializability, Serial schedules, Conflict serializability, View serializability, Recoverable and non-recoverable schedules, Cascading rollbacks, Cascadeless schedules

Concurrency control (3 sessions): Serialized and non-serialized schedules, Testing for serializability, Locking, Lock compatibility matrix, Locking and serializability, Deadlocks and starvation, Two-phase locking (2PL) protocol, Conservative, strict and rigorous 2PL, 2PL with lock conversions, Timestamp-ordering based protocol, Multi-versioning protocol, Multi-granularity locking, Deadlock prevention protocols, Wait-die and wound-wait schemes, Time-out based schemes, Deadlock recovery, Nested transactions

Module 7: Database recovery techniques (3 sessions)

Recovery concepts, Deferred updates technique, Immediate update technique, Shadow paging, ARIES recovery algorithm

Module 8: Query Processing and Optimization (2 sessions)

Translating SQL into relational algebra, Basic query operations, Heuristics in query optimization, Selectivity and cost estimates in query optimization, Semantic query optimization

Module 9: Database Security and Authorization (1 session)

Discretionary access control, Mandatory access control and multi-level security, Statistical database security

Module 10: Enhanced Data Models for specific applications (2 sessions)

Active database concepts, Temporal databases, Spatial databases, multi-media databases

Module 11: Distributed databases and issues (3 sessions)

Data fragmentation, replication and allocation in distributed databases, Types of distributed database systems Query processing in distributed databases, Concurrency control and recovery in distributed databases

Textbook:

Elmasri, Navathe. Fundamentals of Database Systems (Third Edition), Pearson Education, 2004.