| Course code | Course Name | L-T-P -Credits | Year of Introduction |
|---|---|---|---|
| CS208 | Principles of Database Design | 2-1-0-3 | 2016 |

**Pre-requisite:** CS205 Data structures

**Course Objectives**

1. To impart the basic understanding of the theory and applications of database management systems.
2. To give basic level understanding of internals of database systems.
3. To expose to some of the recent trends in databases.

**Syllabus:**

Types of data, database and DBMS, Languages and users. Software Architecture, E-R and Extended E-R Modelling, Relational Model – concepts and languages, relational algebra and tuple relational calculus, SQL, views, assertions and triggers, HLL interfaces, relational db design, FDs and normal forms, Secondary storage organization, indexing and hashing, query optimization, concurrent transaction processing and recovery principles, recent topics.

**Expected outcome**.

Students will be able to:

1. define, explain and illustrate the fundamental concepts of databases.
2. construct an Entity-Relationship (E-R) model from specifications and to perform the transformation of the conceptual model into corresponding logical data structures.
3. model and design a relational database following the design principles.
4. develop queries for relational database in the context of practical applications
5. define, explain and illustrate fundamental principles of data organization, query optimization and concurrent transaction processing.
6. appreciate the latest trends in databases.

**Text Books:**

1. Elmasri R. and S. Navathe, *Database Systems: Models, Languages, Design andApplication Programming*, Pearson Education, 2013.
2. Sliberschatz A., H. F. Korth and S. Sudarshan, *Database System Concepts*, 6/e, McGraw Hill, 2011.

**References:**

1. Powers S., *Practical RDF*, O'Reilly Media, 2003.
2. Plunkett T., B. Macdonald, *et al.*, *Oracle Big Data Hand Book*, Oracle Press, 2013.

**Course Plan**

| Module | Contents | Hours (42) | Sem.ExamMarks |
|---|---|---|---|
| I | **Introduction:** Data: structured, semi-structured and unstructured data**,** Concept & Overview of DBMS, Data Models, Database Languages, Database Administrator, Database Users, Three Schema architecture of DBMS. Database architectures and classification. (Reading: ElmasriNavathe, Ch. 1 and 2. Additional Reading: Silbershatz, Korth, Ch. 1) **Entity-Relationship Model:** Basic concepts, Design Issues, Mapping Constraints, | 06 | 15% |

| | | | |
|---|---|---|---|
| | Keys, Entity-Relationship Diagram, Weak Entity Sets, Relationships of degree greater than 2 (Reading: ElmasriNavathe, Ch. 7.1-7.8) | | |
| **II** | **Relational Model:** Structure of relational Databases, Integrity Constraints, synthesizing ER diagram to relational schema (Reading: ElmasriNavathe, Ch. 3 and 8.1, Additional Reading: Silbershatz, Korth, Ch. 2.1-2.4) **Database Languages:** Concept of DDL and DML relational algebra (Reading: Silbershatz, Korth, Ch 2.5-2.6 and 6.1-6.2, ElmasriNavathe, Ch. 6.1-6.5) | **06** | **15%** |
| **FIRST INTERNAL EXAM** | | | |
| **III** | **Structured Query Language (SQL**): Basic SQL Structure, examples, Set operations, Aggregate Functions, nested sub-queries (Reading: ElmasriNavathe, Ch. 4 and 5.1) **Views, assertions and triggers** (Reading: ElmasriNavathe, Ch. 5.2-5.3, Silbershatz, Korth Ch. 5.3). **Functions, Procedures and HLL interfaces** (Reading: Silbershatz, Korth Ch. 5.1-5.2). | **07** | **15%** |
| **IV** | **Relational Database Design:** Different anomalies in designing a database, normalization, functional dependency (FD), Armstrong's Axioms, closures, Equivalence of FDs, minimal Cover (proofs not required). Normalization using functional dependencies, INF, 2NF, 3NF and BCNF, lossless and dependency preserving decompositions (Reading: Elmasri and Navathe, Ch. 14.1-14.5, 15.1-15.2. Additional Reading: Silbershatz, Korth Ch. 8.1-8.5) | **07** | **15%** |
| **SECOND INTERNAL EXAM** | | | |
| **V** | **Physical Data Organization**: index structures, primary, secondary and clustering indices, Single level and Multi-level indexing, B-Trees and B+-Trees (basic structure only, algorithms not needed), Indexing on multiple keys (Reading Elmasri and Navathe, Ch. 17.1-17.4) **Query Optimization**: algorithms for relational algebra operations, heuristics-based query optimization, Cost-based query optimization (Reading Elmasri and Navathe, Ch. 18.1-18.3, 18.6-18.8) | **08** | **20%** |
| **VI** | **Transaction Processing Concepts:** overview of concurrency control and recovery acid properties, serial and concurrent schedules, conflict serializability. Two-phase locking, failure classification, storage structure, stable storage, log based recovery, deferred database | **08** | **20%** |

| | | | |
|---|---|---|---|
| | modification, check-pointing, (Reading Elmasri and Navathe, Ch. 20.1-20.5 (except 20.5.4-20.5.5) , Silbershatz, Korth Ch. 15.1 (except 15.1.4-15.1.5), Ch. 16.1 – 16.5) R**ecent topics (preliminary ideas only):** Semantic Web and RDF(Reading: Powers Ch.1, 2), GIS, biological databases (Reading: Elmasri and Navathe Ch. 23.3-23.4) Big Data (Reading: Plunkett and Macdonald, Ch. 1, 2) | | |
| **END SEMESTER EXAM** | | | |

**Question Paper Pattern:**

1. There will be *five* parts in the question paper – A, B, C, D, E
2. Part A
   a. Total marks : 12
   b. *Four* questions each having 3 marks, uniformly covering module I and II; All *four* questions have to be answered.
3. Part B
   a. Total marks : 18
   b. *Three* questions each having 9 marks, uniformly covering module I and II; T*wo* questions have to be answered. Each question can have a maximum of three subparts
4. Part C
   a. Total marks : 12
   b. *Four* questions each having 3 marks, uniformly covering module III and IV; All *four* questions have to be answered.
5. Part D
   a. Total marks : 18
   b. *Three* questions each having 9 marks, uniformly covering module III and IV; T*wo* questions have to be answered. Each question can have a maximum of three subparts
6. Part E
   a. Total Marks: 40
   b. *Six* questions each carrying 10 marks, uniformly covering modules V and VI; *four* questions have to be answered.
   c. A question can have a maximum of three sub-parts.
7. There should be at least 60% analytical/numerical/design questions.