



National University of Engineering (UNI)
School of Computer Science
Syllabus 2027-I

1. COURSE

CS272. Databases II (Mandatory)

2. GENERAL INFORMATION

2.1 Course	: CS272. Databases II
2.2 Semester	: 5 th Semester.
2.3 Credits	: 3
2.4 Hours	: 1 HT; 4 HP;
2.5 Duration of the period	: 16 weeks
2.6 Type of course	: Mandatory
2.7 Learning modality	: Face to face
2.8 Prerequisites	: CS271. Data Management. (4 th Sem)

3. PROFESSORS

Meetings after coordination with the professor

4. INTRODUCTION TO THE COURSE

This course is a continuation of Data Management I, focusing on the internal mechanisms of database management systems (DBMS) and modern database technologies. It covers query optimization, indexing structures, and transactional integrity in depth. Additionally, it introduces students to non-relational models (NoSQL) and distributed architectures, which are essential for handling large-scale data in contemporary cloud environments.

5. GOALS

- Understand the internal storage structures and indexing techniques.
- Analyze and optimize query execution plans.
- Comprehend advanced transaction protocols and recovery mechanisms.
- Evaluate and implement solutions using NoSQL database models.
- Understand the principles of distributed and parallel databases.

6. COMPETENCES

AG-C08) Problem Analysis: Identifies, formulates, and analyzes complex computing problems. (Usage)

AG-C12) Uses systemic approaches to select, develop, apply, integrate, and manage secure computing technologies to achieve user objectives. (Usage)

7. TOPICS

Unit 1: Query Processing (12 hours)	
Competences Expected: AG-C08,AG-C12	
Topics	Learning Outcomes
• Page structures • Index structures: – B+ trees – Hash indices: static and dynamic – Index creation in SQL • File structures: – Heap files – Hash files • Algorithms for query operators: – External Sorting – Selection – Projection; with and without duplicate elimination – Natural Joins: Nested loop, Sort-merge, Hash join – Analysis of algorithm efficiency • Query transformations • Query optimization: – Access paths – Query plan construction – Selectivity estimation – Index-only plans • Parallel Query Processing (e.g., parallel scan, parallel join, parallel aggregation) • Database tuning/performance: – Index selection – Impact of indices on query performance – Denormalization	• Describe the purpose and organization of both B+ tree and hash index structures [Describe] • Compose an SQL command to create an index (any kind) [Compose] • Specify the steps for the various query operator algorithms: external sorting, projection with duplicate elimination, sort-merge join, hash-join, block nested-loop join [Analyze] • Derive the run-time (in I/O requests) for each of the above algorithms [Analyze] • Transform a query in relational algebra to its equivalent appropriate for a left-deep, pipelined execution [Redesign] • Compute selectivity estimates for a given selection and/or join operation [Compute] • Describe how to modify an index structure to facilitate an index-only operation for a given relation [Describe] • For a given scenario decide on which indices to support for the efficient execution of a set of queries [Assess] • Describe how DBMSs leverage parallelism to speed up query processing by dividing the work across multiple processors or nodes [Describe]
Readings : [Silberschatz2019], [Ramakrishnan2003]	

Unit 2: DBMS Internals (12 hours)	
Competences Expected: AG-C08,AG-C12	
Topics	Learning Outcomes
• DB Buffer Management • Transaction Management: – Isolation Levels – ACID – Serializability – Distributed Transactions • Concurrency Control: – 2-Phase Locking – Deadlocks handling strategies – Quorum-based consistency models • Recovery Manager: – Relation with Buffer Manager • Concurrency Control: – Optimistic concurrency control – Timestamp concurrency control • Recovery Manager: – Write-Ahead logging – ARIES recovery system (Analysis, REDO, UNDO)	• Describe how a DBMS manages its Buffer Pool [Describe] • Describe the four properties for a correct transaction manager [Describe] • Outline the principle of serializability [Outline]
Readings : [Silberschatz2019], [GarciaMolina2008]	

Unit 3: Distributed Databases/Cloud Computing (12 hours)	
Competences Expected: AG-C08,AG-C12	
Topics	Learning Outcomes
• Distributed DBMS: – Distributed data storage – Distributed query processing – Distributed transaction model – Homogeneous and heterogeneous solutions – Client-server distributed databases • Parallel DBMS: – Parallel DBMS architectures: shared memory, shared disk, shared nothing – Speedup and scale-up, e.g., use of the MapReduce processing model – Data replication and weak consistency models	• Describe the key components of a distributed DBMS, including distributed data storage, query processing, and transaction management [Describe] • Analyze the tradeoffs between parallel DBMS architectures: shared memory, shared disk, and shared nothing [Analyze] • Describe data replication strategies and weak consistency models in distributed database systems [Describe]
Readings : [Silberschatz2019], [Ozsu2020]	

Unit 4: NoSQL Systems (12 hours)	
Competences Expected: AG-C08,AG-C12	
Topics	Learning Outcomes
• Why NoSQL? (e.g., Impedance mismatch between Application [CRUD] and RDBMS) • Key-Value and Document data model • Storage systems (e.g., Key-Value systems, Data Lakes) • Distribution Models (Sharding and Replication) • Graph Databases • Consistency Models (Update and Read, Quorum consistency, CAP theorem) • Processing model (e.g., Map-Reduce, multi-stage map-reduce, incremental map-reduce) • Case Studies: Cloud storage system (e.g., S3); Graph databases; "When not to use NoSQL"	• Develop a use case for the use of NoSQL over RDBMS [Create] • Describe the defining characteristics behind Key-Value and Document-based data models [Describe]
Readings : [Sadalage2012], [kleppmann2017designing]	

Unit 5: Semi-structured and Unstructured Databases (12 hours)	
Competences Expected: AG-C08,AG-C12	
Topics	Learning Outcomes
• Vectorized unstructured data (text, video, audio, etc.) and vector storage: – TF-IDF Vectorizer with ngram – Word2Vec – Array database or array data type handling • Semi-structured databases (e.g., JSON): – Storage: * Encoding and compression of nested data types – Indexing: * Btree, skip index, Bloom filter * Inverted index and bitmap compression * Space filling curve indexing for semi-structured geo-data – Query processing for OLTP and OLAP use cases: * Insert, Select, update/delete tradeoffs * Case studies on Postgres/JSON, MongoDB, and Snowflake/JSON	• Describe vectorized representations of unstructured data (e.g., TF-IDF, Word2Vec) and their storage models [Describe] • Apply indexing techniques (e.g., B-tree, inverted index, Bloom filter) to semi-structured data stored in JSON format [Apply] • Analyze query processing tradeoffs for semi-structured databases in OLTP and OLAP use cases [Analyze]
Readings : [kleppmann2017designing], [manning2008introduction], [mikolov2013efficient], [agrawal2023toward], [mongodb_documentation]	

8. WORKPLAN

8.1 Methodology

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

8.2 Theory Sessions

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

8.3 Practical Sessions

The practical sessions are held in class where a series of exercises and/or practical concepts are developed through problem solving, problem solving, specific exercises and/or in application contexts.

9. EVALUATION SYSTEM

Theory Sessions:

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

Lab Sessions:

In order to verify their competences, several activities including active learning and roleplay will be developed during lab sessions.

Oral Presentations:

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

Reading:

Throughout the course different readings are provided, which are evaluated. The average of the notes in the readings is considered as the mark of a qualified practice. The use of our virtual campus allows each student to access the course information, and interact outside the classroom with the teacher and with the other students.

The final grade NF depends on the following grades:

- PE : Partial Exam.
- FE : Final Exam.
- P : Permanent Evaluation (Average of practical work and graded assignments).

The final grade NF is calculated as follows:

$$NF = \frac{PE + FE + P}{3}$$

10. BASIC BIBLIOGRAPHY