University of Engineering and Technology
School of Computer Science
Syllabus of Course – Academic Period 2017-I

1. **Code and Name:** CS272. Databases II
2. **Credits:** 3
3. **Hours of theory and Lab:** 1 HT; 4 HP;
4. **Professor(s)**
   Meetings after coordination with the professor
5. **Bibliography**

[Bur04] Donald K. Burleson. *Physical Database Design Using Oracle*. CRC Press, 2004.

[Cel05] Joe Celko. *Joe Celko's SQL Programming Style*. Elsevier, 2005.

[Dat05] C.J. Date. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Elsevier, 2005.

[M T99] Patrick Valduriez M. Tamer Ozsu. *Principles of Distributed Database Systems, Second Edition*. Prentice Hall, 1999.

[Pet98] Julita Vassileva Peter Brusilovsky Alfred Kobsa. *Adaptive Hypertext and Hypermedia, First Edition*. Springer, 1998.

[Phi97] Eric Newcomer Philip A. Bernstein. *Principles of Transaction Processing, First Edition*. Morgan Kaufmann, 1997.

[Ram04] Shamkant B. Navathe Ramez Elmasri. *Fundamentals of Database Systems, Fourth Edition*. Addison Wesley, 2004.

6. **Information about the course**

   (a) **Brief description about the course** Information Management (IM) plays a leading role in almost every area where computers are used. This area includes the capture, digitization, representation, organization, transformation and presentation of information; Algorithms to improve the efficiency and effectiveness of access and update of stored information, data modeling and abstraction, and physical file storage techniques.

   It also covers information security, privacy, integrity and protection in a shared environment. Students need to be able to develop conceptual and physical data models, determine which IM methods and techniques are appropriate for a given problem, and be able to select and implement an appropriate IM solution that reflects all applicable constraints, including scalability and Usability.

   (b) **Prerrequisites:** CS271. Bases de Datos I. (4^{to} Sem)

   (c) **Type of Course:** Mandatory

7. **Competences**

   - To make the student understand the different applications that the databases have, in the different areas of knowledge.
   - Show appropriate ways of storing information based on their various approaches and their subsequent retrieval of information.

8. **Contribution to Outcomes**

b) An ability to design and conduct experiments, as well as to analyze and interpret data. (**Assessment**)

i) An ability to use the techniques, skills, and modern computing tools necessary for computing practice. (**Usage**)

j) Apply the mathematical basis, principles of algorithms and the theory of Computer Science in the modeling and design of computational systems in such a way as to demonstrate understanding of the equilibrium points involved in the chosen option. (**Usage**)

9. Competences (IEEE)

C1. An intellectual understanding and the ability to apply mathematical foundations and computer science theory.⇒ **Outcome b**

C7. Being able to apply the software engineering principles and technologies to ensure that software implementations are robust, reliable, and appropriate for their intended audience.⇒ **Outcome j**

CS4. Deploy appropriate theory, practices, and tools for the specification, design, implementation, and maintenance as well as the evaluation of computer-based systems.⇒ **Outcome i**

10. List of topics

1. Physical Database Design
2. Transaction Processing
3. Information Storage and Retrieval
4. Distributed Databases

11. Methodology and Evaluation
Methodology:

Theory Sessions:
The development of the theoretical sessions is focused on the student, through his active participation, solving problems related to the course with the individual contributions and discussing real cases of the industry. The students will develop throughout the course a project of application of the tools received in a company.

Lab Sessions:
Practical sessions are held in the laboratory. Laboratory practices are performed in teams to strengthen their communication. At the beginning of each laboratory the development of the practice is explained and at the end the main conclusions of the activity in group form are highlighted.

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 the UTEC Online virtual campus allows each student to access the course information, and interact outside the classroom with the teacher and with the other students.
Evaluation System:

.

12. Content

Unit 1: Physical Database Design (10)

Competences Expected: C1

Learning Outcomes	Topics
• Explain the concepts of records, record types, and files, as well as the different techniques for placing file records on disk [Usage] • Give examples of the application of primary, secondary, and clustering indexes [Usage] • Distinguish between a non-dense index and a dense index [Usage] • Implement dynamic multilevel indexes using B-trees [Usage] • Explain the theory and application of internal and external hashing techniques [Usage] • Use hashing to facilitate dynamic file expansion [Usage] • Describe the relationships among hashing, compression, and efficient database searches [Usage] • Evaluate costs and benefits of various hashing schemes [Usage] • Explain how physical database design affects database transaction efficiency [Usage]	• Storage and file structure • Indexed files • Hashed files • Signature files • B-trees • Files with dense index • Files with variable length records • Database efficiency and tuning

Readings : [Bur04], [Dat05], [Cel05]

Unit 2: Transaction Processing (12)

Competences Expected: C1

Learning Outcomes	Topics
• Create a transaction by embedding SQL into an application program [Usage] • Explain the concept of implicit commits [Usage] • Describe the issues specific to efficient transaction execution [Usage] • Explain when and why rollback is needed and how logging assures proper rollback [Usage] • Explain the effect of different isolation levels on the concurrency control mechanisms [Usage] • Choose the proper isolation level for implementing a specified transaction protocol [Usage] • Identify appropriate transaction boundaries in application programs [Usage]	• Transactions • Failure and recovery • Concurrency control • Interaction of transaction management with storage, especially buffering

Readings : [Phi97], [Ram04]

Unit 3: Information Storage and Retrieval (10)

Competences Expected: C1

Learning Outcomes	Topics
• Explain basic information storage and retrieval concepts [Usage] • Describe what issues are specific to efficient information retrieval [Usage] • Give applications of alternative search strategies and explain why the particular search strategy is appropriate for the application [Usage] • Design and implement a small to medium size information storage and retrieval system, or digital library [Usage] • Describe some of the technical solutions to the problems related to archiving and preserving information in a digital library [Usage]	• Documents, electronic publishing, markup, and markup languages • Tries, inverted files, PAT trees, signature files, indexing • Morphological analysis, stemming, phrases, stop lists • Term frequency distributions, uncertainty, fuzziness, weighting • Vector space, probabilistic, logical, and advanced models • Information needs, relevance, evaluation, effectiveness • Thesauri, ontologies, classification and categorization, metadata • Bibliographic information, bibliometrics, citations • Routing and (community) filtering • Multimedia search, information seeking behavior, user modeling, feedback • Information summarization and visualization • Faceted search (e.g., using citations, keywords, classification schemes) • Digital libraries • Digitization, storage, interchange, digital objects, composites, and packages • Metadata and cataloging • Naming, repositories, archives • Archiving and preservation, integrity • Spaces (conceptual, geographical, 2/3D, VR) • Architectures (agents, buses, wrappers/mediators), interoperability • Services (searching, linking, browsing, and so forth) • Intellectual property rights management, privacy, and protection (watermarking)

Readings : [Pet98], [Ram04]

Unit 4: Distributed Databases (36)	
Competences Expected: C1	
Learning Outcomes	**Topics**
• Explain the techniques used for data fragmentation, replication, and allocation during the distributed database design process [Usage] • Evaluate simple strategies for executing a distributed query to select the strategy that minimizes the amount of data transfer [Usage] • Explain how the two-phase commit protocol is used to deal with committing a transaction that accesses databases stored on multiple nodes [Usage] • Describe distributed concurrency control based on the distinguished copy techniques and the voting method [Usage] • Describe the three levels of software in the client-server model [Usage]	• 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
Readings : [M T99], [Dat05]	