NATIONAL UNIVERSITY
of Computer & Emerging Sciences, Lahore

Department of Computer Science

CS451 – Advance Database Concepts

Spring 2020

Instructor Name: Muhammad Ishaq Raza
Email address: ishaq.raza@nu.edu.pk
Office Location/Number: M-160
Office Hours: Tu & Th 1:30-3:30 PM

TA Name (if any):
Email address:
Office Location/Number:
Office Hours:

Course Information

Program: BS **Credit Hours:** 3 **Type:** Elective
Pre-requisites (if any): CS203 – Database Systems
Course Website (if any):
Class Meeting Time: M, W 11:00 AM – 12:20 PM
Class Venue: CS-05

Course Description/Objectives/Goals:

This course is intended for students who wish to specialize in database management systems or wish to practice the advanced techniques involved in optimization of data storage, database design and queries. The course primarily addresses design and implementation of a database management system. It covers advanced topics like physical storage and access methods, query optimization, transaction processing, concurrency control, and recovery techniques.

Course Textbook

1. Ramez Elmasri, *Fundamentals of Database Systems* (7th Edition)

Additional references and books related to the course:

1. Raghu Ramakrishnan, *Database Management Systems*(3rd Edition)
2. Jefferey D. Ullman, Jennifer Widom , Hector Garcia-Molina, Database System Implementation
3. George Coulouris, *Distributed Systems; Concepts and Design* (3rd Edition)

(Tentative) Grading Criteria

1. Assignments (10%)
2. Quizzes (10%)
3. Paper Report (10%)
4. Class Participation (5%)
5. 2 Midterm Exams (25%)
6. Final Exam (40%)

Grading Scheme: Relative

Course Policies

1. Quizzes may be un-announced.
2. No makeup for missed quiz or assignment.
3. Minimum eligibility to pass this course is to get 50% marks.

Tentative Weekly Schedule

Week	Topics to be covered	Topics Detail	Readings (Textbook)	No of Lec.	Asst.
1-2	Transaction Processing Concepts and Theory	• Issues in Transaction Processing • Why Concurrency Control is Needed • Why Recovery is Needed • Transaction States and Operations, System Log, Commit Point of a Transaction • ACID Properties of Transactions • Characterizing Schedules based on Recoverability • Characterizing Schedules based on Serializability • Transactions Isolation Levels and Possible Violations	Ch 20	3	
2-3	Concurrency Control Techniques	• Two-phase locking techniques for concurrency control • Concurrency control based on timestamp ordering • Multiversion concurrency control techniques • Validation (optimistic) techniques and snapshot isolation concurrency control	Ch 21	3	A1
4-5	Database Recovery Techniques	• Recovery concepts • NO-UNDO/REDO recovery based on deferred update • Recovery techniques based on immediate update • Shadow paging • The ARIES recovery algorithm • Recovery in multidatabase systems • Database backup and recovery from catastrophic failures	Ch 22	3	
5-7	Data Storage, Basic File Structures and Hashing	• Disk storage devices • Buffering of blocks • Placing file records on disk • Operations on files • Files of unordered records (heap files) • Files of ordered records (sorted files) • Hashing techniques: internal hashing, static hashed files (external hashing) • Hashing techniques that allow dynamic file expansion: dynamic, extendible, and linear hashing techniques	Ch 16	4	A2
7-10	Indexing Structures for Files and Physical Database Design	• Type of single-level ordered indexes: primary indexes, clustering indexes, secondary indexes • Multilevel indexes: two level indexes, multilevel indexes • Dynamic multilevel indexes using B-Trees, B^+-Trees and B^*-Trees • Indexes on multiple keys: ordered indexes on multiple attributes, partitioned hashing, grid files • Other types of indexes: hash indexes, bitmap indexes, function-based indexing • Physical database design in relational databases • An overview of database tuning in relational systems	Ch 17	6	A3
10-11	Strategies for Query Processing	• Translating SQL queries into relational algebra and other operators • Algorithms for external sorting • Algorithms for SELECT operation • Algorithms for JOIN operation • Algorithms for PROJECT and SET operations • Implementing aggregate operations and different type of JOINs • Combining operations using pipelining • Parallel algorithms for query processing: parallel database architecture approaches (share-memory, shared-disk, and shared-nothing architectures), operator level parallelism, intra-query parallelism, inter-query parallelism	Ch 18	3	A4
12	Query Optimization	• Query trees and heuristics for query optimization • Choices of query execution plans, nested subquery optimization, materialized views • Use of selectivities in cost-based optimization • Cost components for query optimization • Catalog information used in cost functions • Histograms • Multi-relation queries and JOIN ordering choices (left-deep join tree, right-deep join tree, and bushy join tree) • Adaptive and semantic query optimization	Ch 19	2	
13	Distributed Database Concepts	• Distributed computing systems and Big data technologies • Distributed database concepts; Data fragmentation, replication, and allocation techniques for distributed database design • Overview of transaction management in distributed databases • Types of distributed database systems • Distributed database architectures.	Ch 23	2	
14	NOSQL Databases and Big Data Storage Systems	• Introduction to NOSQL systems • The CAP theorem • Document-based NOSQL systems and MongoDB; NOSQL Key-Value stores; column-based NOSQL systems; NOSQL graph databases and Neo4j	Ch 24	2	