25BIO802   **Advanced Database Administration for Life Sciences**   **2 1 0 3**

Course description: This doctoral-level course provides a comprehensive understanding of advanced database administration techniques with a focus on life sciences applications. It is designed to equip students with the knowledge and hands-on skills required to effectively design, manage, and optimize biological databases tailored to biological datasets.

Learning Objectives: The PhD level course titled High-Level Database Administration for Life Sciences with a focus on how to create and administer the databases for the life science domain. This course will familiarize the student with deep understanding of the processes involved in creating different type of databases, managing the query language, and administering the database with optimal performance.

Pedagogy:

1. Fundamental knowledge on database systems
2. Basic programming skills
3. Basic knowledge of Operating systems and Networking
4. o

Course Outline:
Module 1: Relational Model and Query Languages

Introduction to the relational data model: structure of relations, and attributes, Relational algebra, Relational calculus, SQL operators and functions, Data definition language, Data manipulation language, Data control language, Embedded SQL, and Dynamic SQL, Creation and management of database objects, Subqueries, Join Queries, Advanced SQL functions, Stored procedure and functions

Module 2: Database Design

Normalization, database anomalies, functional dependencies, candidate keys, super keys, non-loss decomposition, dependency preservation, first normal form, second normal form, third normal form, Boyce-Codd normal form, multivalued dependencies, fourth normal form, join dependencies, fifth normal form, denormalization

Module 3: Transaction Processing

Database transactions, transaction states, ACID properties, transaction recovery, two-phase commit, commit points, serializability, concurrency control, binary locking protocol, two-phase locking protocol, deadlock, starvation, transaction timestamp

Module 4: Database Administration for SQL in Biological Databases

Installation and configuration of biological database, User and role management, SQL-based data storage for biological datasets, schema design for genomic and proteomic data, indexing strategies for large-scale biological data, user access control and privileges, data security and backup, performance tuning for biological queries, transaction logging, data integrity in biological databases, use of views and stored procedures, normalization challenges in bioinformatics data, integration of SQL with biological data formats (FASTA, GFF, GenBank), case studies on biological databases (NCBI, Ensembl, UniProt).

Database Labs:

Practical 1: Relational Database Design for Genomic Data

Practical 2: Design and Implement a Normalized Relational Database for Genomic Data.

Practical 3 Administering a Biological SQL Database with Indexing and Optimization

References:

1. A. Silberschatz, H.F. Korth & S. Sudarshan: *Database System Concepts*, TMH, 1997.
2. C.J. Date: *An Introduction to Database Systems* 7th Ed., Addison Wesley, Indian Edition, 2000.
3. Elmasri & Navathe: *Fundamentals of Database Systems/Oracle 9i Programming* 5th Ed., Pearson, 2009.
4. Kreines, D. C., & Laskey, B.: *Oracle Database Administration: The Essential Reference*. O'Reilly Media, 1999.

Course Outcomes :

CO1: To develop advanced skills in database administration

CO2: To apply database management techniques on biological data

CO3: To analyse and design normalized databases using functional dependencies and various normal forms.

CO4: To demonstrate knowledge of transaction processing, concurrency control, and recovery techniques to ensure database integrity.

CO5: To establish automation, backup strategies, and cloud/database hosting options