NEW YORK UNIVERSITY
COMPUTER SCIENCE DEPARTMENT
COURANT INSTITUTE OF MATHEMATICAL SCIENCES

Database Systems

Summer 2021  Jean-Claude FRANCHITTI
G22.2433-001  Thu. 6:00 - 8:20 p.m.
==========================================

COURSE DESCRIPTION

The course has a balanced theory-practice focus, covering essential topics, from established database technologies to recent trends, like post-relational databases, Big Data, NoSQL, Cloud storage-as-a-service on (AWS, Azure Cloud, GCP, IBM Cloud, etc.), and more. With respect the established database technologies, the course covers the fundamental study material required to understand how databases work (especially relational ones, such as Access, MySQL, Microsoft SQL Server, Oracle Database, and IBM DB2). The course also delves into explaining how to write SQL queries, and covers the various issues that need to be addressed in designing database systems, implementing them, and using them. The course also outlines how to setup end-to-end data management infrastructure for data science applications. Material covered includes: enterprise data modeling using Entity Relationship Diagrams and UML, deriving a relational model implementation from Entity Relationship Diagrams and UML, Relational Algebra, SQL as a Data Definition Language and as a Data Manipulation Language, maintaining integrity of a database system, normalization (normal forms), physical design and query optimization, recovery and concurrency, on-line analytical processing, data warehouses, object relational databases, and unstructured databases. As part of the course, students will use mainstream commercial software (e.g., CA Erwin, MySQL Workbench, Toad, MySQL, Oracle Database, IBM DB2, Microsoft SQL Server, MongoDB, Neo4j, Cloud storage infrastructures, etc.) and will work on practical exercises and projects.
SYLLABUS

The following syllabus is subject to change:

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
</table>
| 1    | • Introduction to database systems  
|      |   - Fundamental concepts  
|      |   - Architecture and Classification  
|      |   - Organizational aspects of data mgmt.  
|      | • Data mgmt. roadmap  
| 2    | • Relational model and relational database constraints  
|      | • Legacy databases  
|      | • Relational databases  
| 3    | • Enterprise data modeling using the entity-relationship model  
|      | • Conceptual data modeling using EER and UML  
| 4    | • Practical relational database design  
|      | • Data Warehousing and Business Intelligence  
|      | • Data Lakes and Data Swamps  
| 5    | • Relational algebra, relational calculus, and SQL  
|      | • Extended relational databases  
|      | **Midterm**  
| 6    | • Standard Query Language (SQL) features  
|      | • Object databases and object persistence  
|      | • Object databases  
|      | • XML databases  
|      | • Unstructured data and NoSQL databases  
|      | • Cloud data management services  
| 7    | • Functional dependencies and Normalization  
|      | • Data integration  
|      | • Data quality  
|      | • Data governance  
|      | • Big Data and Big Data tools  
|      | • Analytics  
| 8    | • Physical file organization  
|      | • Indexing  
|      | • Physical database design  
|      | • Physical database organization  
|      | • Query execution concepts  
|      | • Database programming techniques  
|      | • Database APIs  
| 9    | • Transaction management  

|   | • Concurrency control  
|   | • Data distribution  
|   | • Distributed transaction mgmt.  
|   | • Distributed and parallel processing  
| 10 | • Recovery  
|   | • Hadoop framework  
|   | • Spark framework  
|   | • Massively Parallel Processing (MPP) databases  
| 11 | • Data Warehouses and OLAP  
|   | • Data management for machine and deep learning  
| 12 | • Case Studies: Data Science end-to-end infrastructure setup for recommendation application  
|   | Final Exam |