Database Systems for Data Science

16:198:527

Much of the world's data resides in databases. The purpose of this course is to introduce relational and NoSQL database concepts with emphasis on both theoretical and practical learning. This course helps students learn and apply knowledge of the SQL language and implementing components of relational and NoSQL database systems (DBMS). Students will create database instances in the cloud for both relational and NoSQL database systems such as MySQL, SQL Server, Amazon Redshift, Google BigQuery and MongoDB. Through a couple of hands-on projects students will practice building and running advanced SQL scripts and Python/Java codes.

This course is suitable for entry level M.Sc. Students who have not taken an undergraduate database course. This course counts as category B for the M.Sc. degree requirement. This course does NOT count as category B for Ph.D. students

Credits:
3

Semester:
Fall

Course Type:
Graduate

Textbooks:

1. Database Systems, N. Jukic, S. Vrbsky, S. Nestorov, Prospect Press, 2017.

2. NoSQL, P. J. Sadalage, M. Fowler, 2013.

Expected Work:
Projects, Homework, Quizzes, Midterm & Final exams.

Projects (40%) *Two semester long projects: for the first project students will take real or simulated data and build a full web application. For the second project they will be given access to large amounts of data and students will apply to it cleaning techniques and then use MongoDB to extract useful information from it. The projects will be evaluated both technically (implementation) and also according to its utility and presentation.*

***Homework* and/or Quizzes** *(10%)*

Midterm and Final exams *(50%)*

Syllabus:

Part I. Basics

Weeks 1 and 2: Overview of Database system design, ER Diagrams, and the Relational Model

Week 3: SQL: Building and Querying a Relational Database

Week 4: Advanced SQL examples

Part II. Applications

Week 5: Application Development Back end: AWS, JDBC, Java Servlets, etc.

Week 6: Internet Applications Client side: JavaScript, JSON, HTML, CSS, etc.

Part III. NoSQL Data Bases

Week 7: Overview of Key-Value Stores (Ex: Amazon Dynamo),

Week 8: Column Oriented Databases (Ex: Amazon Redshift)

Week 9: Document Databases (Ex: Mongo DB)

Part IV. Storage and efficiency

Week 10: Storage, Indexing and Query Evaluation

Week 11: Writing efficient queries

Week 12: Partitioning, cloud storage, and pricing models

Part V. Databases in real life

Week 13: Data Cleansing

Week 14: Data Warehousing

***Learning Goals*:**
Students enrolled in this class

- will be prepared to contribute to a rapidly changing field by acquiring a thorough grounding in the core principles and foundations of relational and NoSQL database systems.

- will acquire a deeper understanding on (elective) topics of more specialized interest, and be able to critically review, assess, and communicate current developments in the field.

- will be prepared for the next step in their careers, for example, by having done a research project (for those headed to Ph. D. program), a programming project (for those going into the software industry), or some sort of business plan (for those going into start-ups).

Course objectives and **how learning outcomes will be assessed**:

Students will be able to use relational databases tools and web development languages to create significant web applications. Students will also be able to use tools to query and analyze large amounts of unstructured data using NoSQL based tools. Evaluation of concepts will be done through a midterm and a final exam; the evaluation of practical competency will be done through homework assignments and projects.

Academic integrity policy:

We take academic integrity quite seriously. Copying answers from any source including published solutions is considered academic dishonesty.

* In case of learning disabilities, please provide verification from the College Coordinator. Also inform us at the beginning of the semester of any planned absences due to participation in professional events.

*Sakai will be used for weekly announcements related to Homework assignments, project progress reports, and final project demo. Students dully registered for the class will get periodic email alerts.

Professors:
Saed Sayad, Antonio Miranda, James Abello

Name:

NetID:

Major:

Gender:

Age:

Why DBDS: