



COURSE INFORMATION

Course Title:

INFS501 Advanced Database Systems

Semester:

July Session, 2020

CLASS HOURS: Monday through Friday, 180 minutes each day, for three weeks. At the end of each week there will be a one-hour discussion session; the instructor will also be available by appointment.

Discussion Section: 1 hour each week.

Field Trip: Not required.

CREDITS(s): 3

OVERVIEW

This course provides students with the opportunity to learn about various types of Advanced Database Systems. It deepens students' understanding of the database modelling, design, and implementation techniques in various complex data intensive applications. The course covers topics including computing platforms, data mining and data stream, data integration, and data warehousing, data security and privacy, amongst others.

LEARNING OUTCOME

Upon successful completion of this course, the students should have basic expertise in the following areas:

- Understand distributed computing concepts and major platforms.
- Understand the similarity query processing concepts and techniques.
- Understand techniques to manage streaming data.
- Develop insights into issues and challenges in data integration including schema integration, data matching, and use of ontologies.
- Relate your conceptual and practical knowledge in advanced database systems to past lectures as well as emerging trends such as big data.
- Understand the concepts, modelling, and implementation details of data warehousing.
- Develop a broad range of skills and knowledge essential to critically identify and properly address issues in complex data management and analysis.

LEARNING RESOURCES

Suggested textbooks:

- [LRU] Mining of Massive Datasets. Jure Leskovec, Anand Rajaraman, Jeff Ullman. 2nd edition. Cambridge University Press.
- [JPT] Multidimensional Databases and Data Warehousing, Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen. Morgan & Claypool Publishers. 2010.
- [HDI] Principles of Data Integration. Alon Halevy, AnHai Doan, Zachary Ives. 1st edition. Elsevier. 2012.
- Research papers from SIGMOD, VLDB, ICDE, etc.

Disclaimer: Course schedule is subject to change and you will be responsible for abiding by any such changes.



WEEKLY SCHEDULE				
Week	Day	Topic	Reading	
1	1	Distributed Computing Platforms - Map Reduce - Algorithms Using MapReduce - Spark	Chapter 2 of [JRU]	
	2	Similarity Queries - Similarity and distance functions - Deduplicating documents - Introduction to Locality Sensitive Hashing (LSH)	Chapter 3 of [JRU]	
	3	Data Streams - Data stream model - Sampling data in a stream - Filtering streams - Counting distinct elements in a stream	Chapter 4 of [JRU]	
	4	Schema Mapping and Integration - Schema matching - Schema mapping Quiz 1	Chapter 5 of [HDI]	
	5	Data Matching - Problem definition - Rule-based matching - Probabilistic matching Assignment 1 is due	Chapter 7 of [HDI]	
2	1	Mid-semester Examination		
	2	Data Matching - Learning-based matching - Collective matching - Blocking	Chapter 7 of [HDI]	
	3	Ontology and Knowledge Representation - Knowledge representation - Description Logics - Semantic Web	Chapter 12 of [HDI]	
	4	Data Warehouse Fundamentals - OLAP and OLTP - Data warehouse modelling - Data Cubes Quiz 2	Chapter 2 of [JPT]	
	5	Data Warehouse Systems - Querying a data warehouse - Query languages	Chapter 2 of [JPT]	
3	1	Data Warehouse Algorithms - Materialized views - Indexing Assignment 2 is due	Chapter 4 of [JPT]	

Disclaimer: Course schedule is subject to change and you will be responsible for abiding by any such changes.



2	Data Warehouse Algorithms - Join processing - Data cube computation	Chapter 4 of [JPT]
3	Data Security and Privacy - Security measure in databases - Attack and defence - k-anonymity, differential privacy	
4	Revision	
5	Final Examination	

ASSESSMENT

Assessment Task	Score Percentage
Attendance and Participation	10%
Quiz 1	5%
Quiz 2	5%
Assignment 1	10%
Assignment 2	10%
Mid-semester Examination	20%
Final Examination	40%
Total	100%

DETAILS ON GRADE COMPONENTS

1. Attendance and Participation

Students are required to attend all classes and participate actively. Students should treat their classroom obligations as they would treat any serious professional engagement. Your participation grade will be based on the instructor's assessment of how well you contribute to classroom dynamics relative to your class peers.

NB: In case of an absence, the student is responsible for the materials and assignments for that class; it is the student's responsibility to inform the instructor regarding absences and assignments that are missed. **Unexcused absence from three or more scheduled class sessions will be grounds for failure in this course.** If you do have to miss class due to a personal emergency, please let the instructor know as soon as possible. Such emergencies will be dealt with on a case-by-case basis.

Participation grades will be based on **quality** (in-class performance that reflects intellectual depth, insightfulness, and contribution to class learning) and **quantity** (consistency and regularity of performance).

Accordingly, **you are expected to read the related chapters before participating the class.** Be prepared to be called upon to "open" a class discussion by presenting your full analysis and thoughts on the assigned topic at the start of class, or to be asked through a "cold call" for comment during the discussion.

The grading of class participation is necessarily a subjective exercise. However, high-quality comments have one or more of the following characteristics: **(1) insightfulness, (2) appropriate application of course concepts, and (3) advancement of the in-class discussion at hand.**

Disclaimer: Course schedule is subject to change and you will be responsible for abiding by any such changes.



2. Quizzes

Quizzes will happen on the first and second weeks to review and reinforce what has been learned in the previous classes.

3. Assignments

To enhance the learning process, students have to submit two assignments. Each assignment weighs 10% of the total marks.

Assignment 1 covers Divide-and-conquer, Dynamic programming and Greedy algorithm. It is due on Friday of the first week.

Assignment 2 covers graph theories and graph algorithms. It is due on Monday of the third week.

4. Mid-semester Examination

The mid-semester examination will be held on Monday of the second week. It will cover the contents of the first week including but not limited to Divide-and-conquer, Randomised algorithms, Dynamic programming and Greedy algorithms.

5. Final Examination

An in-class final examination will be administered at the end of the course on Friday of the third week. Details of contents, exam format, etc. will be announced in class well before the examination dates.

6. Course Grading

Upon completion of this course, you receive a final grade. A final grade is a letter grade that carries with it a numerical value, as outlined below.

Grade	Mark
A	80-100
B	70-79
C	60-69
D	50-59
E	0-49

To pass this subject, students are required to obtain Grade B or above in order to satisfy all the intended learning outcomes.

CLASSROOM ETHICS & COURSE POLICIES

Being respectful of others' opinions, values and culture

Cell phones are only permitted when the usage is related to the course. Absolutely **NO TEXTING** during class will be tolerated. If you have an emergency situation and you must be able to be reached, set the phone to vibrate and leave the room immediately when it goes off.

Disclaimer: Course schedule is subject to change and you will be responsible for abiding by any such changes.



Any student with a documented disability needing academic adjustments or accommodations should notify the instructor or the program administrator before the start of the program, so such an arrangement will be made accordingly.

Any student who anticipates a schedule conflict because of religious reasons should make arrangements within two weeks of the start of class.

Academic misconduct

Please follow the guideline of the university policy. Academic dishonesty or misconduct will not be tolerated and may result in disciplinary action including a grade F for the course. Work submitted must be the original work of the student. Original work may include the words and ideas of others, but the source of these words and ideas must be indicated in a manner consistent with an academically recognized form, style, and citation manual. Resubmission of work previously presented in another course is prohibited.