

Title (Units): **COMP 7130 Advanced Topics in Knowledge and Data Engineering (3, 2, 1)**

Course Aims: To learn the state-of-the-art knowledge and data engineering concepts, processes and systems. To study the recent literature on the data-driven applications from a range of different areas, including social networks, mobile and spatial data, and transactional data.

Prerequisite: Research Postgraduate Student Standing

Course Intended Learning Outcomes (CILOs):

Upon successful completion of this course, students should be able to:

No.	Course Intended Learning Outcomes (CILOs)
	Knowledge
1	Explain relational databases, data warehouses, key-value stores and ecosystems of distributed database systems
2	Describe data acquisition and modeling techniques
3	Describe the batch and stream processing
4	Describe data access, federation and change capture
	Skill
5	Implement selected techniques learnt to build data-driven applications
	Attitude
6	Develop an informed perspective on the importance of knowledge and data engineering

Calendar Description: This is an advanced course that will not only let students learn the state-of-the-art knowledge data engineering concepts, processes and systems but also cover the recent literature on the applications of knowledge and data engineering techniques to problems from a range of different areas, including social networks, mobile and spatial data, and transactional data.

Teaching and Learning Activities (TLAs):

CILOs	TLAs
1-4	Students will learn the basic concepts and fundamental principles in lectures.
1-5	Examples on how to solve problems will be demonstrated in tutorials to help students have a deeper understanding of the teaching materials. Students will apply the knowledge via programming assignments and/or mini project.
2-6	Students will work on programming assignments to enhance the understanding of learning principles, and gain hands-on experience through conducting a mini project.

Assessment:

No.	Assessment Methods	Weighting	CILOs to be addressed	Remarks
1	Written assignment	30%	1-4	Written assignments will be used to evaluate how well students have learned the basic concepts and processes to apply knowledge to solve data engineering problems.
2	Projects	30%	5-6	Individual or group projects will be used to evaluate how well students have learned to implement data-driven applications.
3	Examination	40%	1-6	Examination will be used to evaluate students' overall understanding of various knowledge and data engineering techniques including their strengths and limitations, and assess their ability to use them to solve problems.

Rubrics:

Excellent (A)	<ul style="list-style-type: none"> • Achieve the six CILOs, with strong evidence of having achieved the last CILO, demonstrating a good mastery of both the theoretical and practical aspects of the knowledge and skills in the selected topics • Able to develop correct solutions to problems in data engineering • Demonstrate a thorough understanding and solid knowledge of the selected topics • Able to apply a variety of techniques and relevant knowledge for solving problems in knowledge and data engineering
Good (B)	<ul style="list-style-type: none"> • Achieve most of the six CILOs, with evidence of having achieved the last CILO, demonstrating a good understanding of the knowledge and skills in the selected topics • Able to develop correct solutions to problems in knowledge and data engineering • Demonstrate a competent level of knowledge of the selected topics • Ability to make use of appropriate techniques and knowledge and apply them to familiar problems in knowledge and data engineering
Satisfactory (C)	<ul style="list-style-type: none"> • Achieve some of the six CILOs, demonstrating a basic level of understanding of the knowledge and skills in the selected topics • Able to provide acceptable solutions to problems in knowledge and data engineering • Demonstrate an adequate level of knowledge of the selected topics • Ability to make use of some techniques and knowledge and apply them to familiar situations in knowledge and data engineering
Fail (F)	<ul style="list-style-type: none"> • Achieve none of the six CILOs, with little understanding of the associated concepts and underlying methodologies in the selected topics • Unable to provide solutions to simple problems in knowledge and data engineering • Knowledge of selected topics falling below the basic minimum level • Unable to apply techniques and knowledge to situations or problems in knowledge and data engineering

Learning Outcomes and Weighting:

Content	CILO No.
I. Some advanced topics in knowledge and data engineering in A. Systems: relational databases, data warehouses, key-value stores and distributed database systems B. Data engineering processes C. Query processing D. Data integration E. Database applications	1-6

References:

T Raghu Ramakrishnan and Johannes Gehrke. Database Management Systems, McGraw-Hill, 3rd Edition, August, 2002.

Joseph M. Hellerstein and Michael Stonebraker. Readings in Database Systems, The MIT Press, 4th Edition, January, 2005.

Xin Huang, Laks V.S. Lakshmanan, and Jianliang Xu. Community Search over Big Graphs, Morgan & Claypool Publishers, August, 2019.

Sherif Sakr and Mohamed Gaber. Large Scale and Big Data: Processing and Management, Auerbach Publications, 1st Edition, June, 2014.

Martin Kleppmann. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. O'Reilly Media, 1st Edition, April 2017.

Additional reading materials such as research papers and chapters will be provided.

Course Content in Outline:

Topic

- I. Some of the following topics will be covered:
 - A. Database systems
 - 1. Relational databases
 - 2. Data warehouses
 - 3. Key-value stores
 - 4. Distributed database systems
 - B. Data engineering processes
 - 1. Data acquisition
 - 2. Data modelling
 - C. Query processing
 - 1. Batch vs stream processing
 - D. Data integration
 - 1. Data access
 - 2. Data federation
 - 3. Change capture
 - E. Database applications
 - 1. Social networks
 - 2. Mobile and spatial data
 - 3. Transactional data