Course Syllabus

Course Information

Course Number	CSCI 5803
Title	Data Warehousing
College	School of the Sciences
Department	Division of Computer Science
Term	Spring 2024
Format	Online
Meeting Days	Monday, Wednesdays
Meeting Time	2:30 - 3:50 PM
Meeting Room	Online
Meeting Link	https://twu-edu.zoom.us/j/8362252819
Weeks	17
Number of Meetings	15
Hours per Week	3
Credit Hours	3

Instructors Information

Instructor	Islam Akef Ebeid
Contact Number	(940) 898-2165
E-mail	iebeid@twu.edu
Office Days	Mondays, Wednesdays, Fridays
Office Hours	4:00 - 5:00 PM
Mode	drop by, appointment outside these hours
Office	MCL 412
Zoom	https://twu-edu.zoom.us/j/8362252819

Course Description and Prerequisites

CSCI 4513. Data Warehousing. Concepts, principles, and tools for designing, implementing, and using data warehouses. Includes constructs such as operational data store (ODS), data warehouse, data mart, and their components. Roles and responsibilities in designing and implementing a data warehouse; management guidelines and techniques; requirements gathering; dimensional modeling; Extract, Transform, and Load (ETL) architecture; data management; security; analytical reporting concepts; and recent trends in the data warehouse domain. Prerequisite: CSCI 3423. Three lecture hours a week. Credit: Three hours.

CSCI 5803. Data Warehousing. Design, implementation, and management of data warehouse systems and their applications; requirements for gathering data for data warehousing; data warehouse architecture; dimensional model design for data warehousing; physical database design for data warehousing; extracting, transforming, and loading strategies; design and development of intelligence applications for decision support; and expansion and support of a data warehouse. Prerequisite: CSCI 5203. Three lecture hours a week. Credit: Three hours.

Course Objectives and Student Learning Outcomes

This course unlocks a skillset related to permanent data storage and management. Data warehousing and analytics are crucial for any organization. Data as an asset is a concept that has been growing in most organizations, emphasizing that the real value in any organization is its information. This course will review data warehousing fundamentals, such as data modeling beyond the relational data model, the star schema, data stores, data marts, reporting tools, and advanced analytics.

The concept of data warehousing has changed significantly in the past decade. Previously, it has been associated with business analytics and a special set of skills and topics like the ones mentioned above. In addition to a special kind of nonrelational data, a model called a data store and associated data marts. Today, data warehousing is associated with what is known as data lakes. That is how to integrate multiple heterogeneous data sources into one data store. More design patterns have emerged due to the industry developing more efficient ways to store, manage, query, and analyze large heterogeneous data sources. Thus, you might see additional topics not mentioned in the original course description in this course.

The student learning outcomes are:

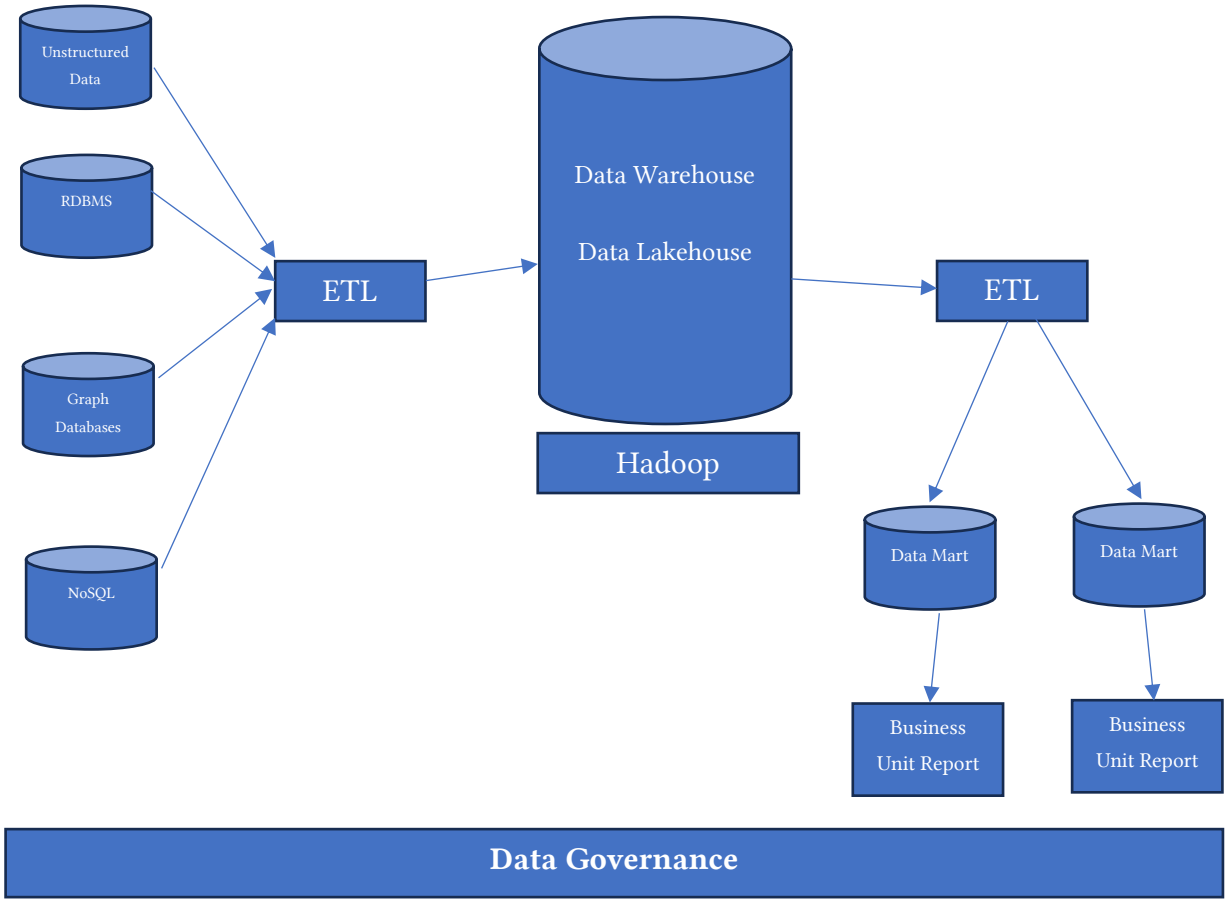
1- Correctly select and evaluate the technology needed for data warehouses.
2- Comprehensively formulate a complex data warehouse design.
3- Correctly assess an organization for warehouse efficacy in strategic decision-making.
4- Correctly manage data warehouse operational, technical, and user metadata.
5- Effectively design data extraction, transformation, and load (ETL) strategies for data integration.
6- Create an effective data warehouse design and reflect on alternative design methodologies and goals.
7- Correctly design analytics reports and OLTP processes.
8- Effectively plan data warehouse reporting tools, including query and online analytical processing OLAP.
9- Accurately explain the role of security in data access and data integration processes.

Course Format

The course will be divided into a lecture component focusing on the suggested books plus additional sources. The practical part concentrates on more hands-on matters in data engineering, data storage, databases, and data warehousing. The course evaluation will be done through assignments, reading quizzes, the final exam, and a final project.

Class split time is split over two weekly sessions. The session on Mondays will be dedicated to the lecture component. At the same time, the session on Wednesdays will be focused on the practical component. On Mondays, the instructor will divide the session into a lecture presentation for 30 to 45 minutes. A short break will follow this, then a reading quiz and a discussion of the readings for this week for about 3o to 45 minutes. Assignments and quizzes will be due the following week. The quizzes will be straightforward from the readings. The quizzes will consist of 5 to 10 multiple-choice questions and will be conducted in class or taken home. Assignments will vary from coding to database implementation to writing reviews.

Course Philosophy



Modern data warehousing focuses on staging and integrating multiple data sources, especially data lakes. This is done via specialized extract transform and load operations via scripting languages such as Python. The integration usually occurs on a distributed system such as Hadoop. Following is another extract transform and load step to standardize the integrated data and produce reports for specific business units. This process is governed by certain practices to ensure data security and quality, known as data governance. This course will closely follow the workflow representing those practices and study each step. The course will start with understanding different data types, from structured to unstructured. We will learn to perform another extract, transform, and load (ETL) operations using Python. We will then review different data models and database management systems, including advanced topics in relational databases, NoSQL databases, and graph databases. We will then go over distributed storage systems like Hadoop. Following, we will learn how to implement and query a data warehouse and the data models involved. Then, we will learn how to create data marts from the data stores in the data warehouse and perform analytics and reports. Following we will understand data

mining and AI in the context of data warehousing. And finally, we will learn the best practices in data governance.

Course Material

We will be following some of the chapters from the following textbooks. You don't have to buy any of those textbooks. They are available online, or I will provide you with some printed material.

Textbooks

Code	Title	Author	Link
[PYTHONMK]	Python for Data Analysis	Wes McKinney	https://wesmckinney.com/book/
[PYTHONTP]	Think Python 2nd Edition	Allen Downey	https://greenteapress.com/thinkpython2/thinkpython2.pdf
[DB2WATTS]	Database Design – 2nd Edition	Adrienne Watt	https://opentextbc.ca/dbdesign01/
[DLAPM]	Data Lakehouse in Action	Pradeep Menon	https://www.packtpub.com/product/data-lakehouse-in-action/9781801815932
[DDIA]	Designing Data-Intensive Applications	Martin Kleppmann	https://www.oreilly.com/library/view/designing-data-intensive-applications/9781491903063/

Software

Name	Download
MySQL	https://dev.mysql.com/downloads/installer/
MySQL Workbench	https://dev.mysql.com/downloads/workbench/
Microsoft SQL Server	https://www.microsoft.com/en-us/sql-server/sql-server-downloads
Microsoft Power BI	https://powerbi.microsoft.com/en-us/downloads/
Microsoft Azure	https://azure.microsoft.com/en-us/pricing/offers/ms-azr-0170p
Python	https://www.anaconda.com/download
PyCharm	https://www.jetbrains.com/pycharm/download/?section=windows
DataGrip	https://www.jetbrains.com/datagrip/download/#section=windows

Course Schedule

Week	Dates	Title	Lecture	Lab	Assignment	Reading	Quiz	Due
1	Week of January 15	Introduction	Syllabus, Introduction to data storage, types and analytics	Surveys	Extra Credit: Introductions Survey	Syllabus	Python programming skill level	January 17
2	Week of January 22	Data Modeling	Database Management Systems and Entity-Relationship Modeling	Data types and data preprocessing in Python	No assignment	[DB2WATTS] Chapters 6 to 10	Quiz 1	January 29
3	Week of January 29	Data modeling for data analytics	Enhanced ER Diagrams, Object-Relational Mapping, Analytics, and SQL	Designing an ER diagram for a problematic dataset	Assignment 1: Write a Python program to read a CSV using built-in libraries and Pandas	[DB2WATTS] Chapters 11 to 16	Quiz 2	February 5
4	Week of February 5	Introduction to data engineering	[DLAPM] Chapter 1: The Evolution of Data Analytics	Designing an ER diagram for a problematic dataset	Assignment 2: Use MySQL to load the VideoFlick database into MySQL. Query the VideoFlick Database using SQL	[DDIA] Chapter 1 (Pages 3-25)	Quiz 3	February 12

5	Week of February 12	Other types of database architectures	NoSQL Databases	Exploring CouchDB	Assignment 3: Write two pages on the different types of database technology in today's market. You can search on Google. Please don't ask ChatGPT or Bard.	[DDIA] Chapter 2 (Pages 27-48)	Quiz 4	February 19
6	Week of February 19	Other types of database architectures	NoSQL Databases	Example of a Graph Database from Neo4j	Assignment 4: Load a graph database into Neo4j	[DDIA] Chapter 2 (Pages 49-67)	Quiz 5	February 26
7	Week of February 26	Introduction to Data Warehousing	[DLAPM] Chapter 2: The Data Lakehouse Architecture	Create a Microsoft Azure student account. Follow the following video. https://www.youtube.com/watch?v=sUODhP4bDrI	Assignment 5: Download and turn in the data warehouse model along with a brief description of what you did.	[DDIA] Chapter 3 (Pages 69-90)	Quiz 6	March 4
8	Week of March 4	Extract, Transform, and Load	[DLAPM] Chapter 3: Ingesting and Processing Data in a Data Lakehouse	Advanced data preprocessing in Python with Pandas	Assignment 6: Write a script to load a datafile into a relational table	[DDIA] Chapter 3 (Pages 90-103)	Quiz 7	March 18
9	Week of March 11	Spring break	Spring break	Spring break	Spring break	Spring break	Spring break	Spring break
10	Week of March 18	Storing data in a data Lakehouse	[DLAPM] Chapter 4: Storing and serving data in a data warehouse	[DLAPM] Chapter 8: Part 1.	Assignment 7: Finish your data Lakehouse implementation at home.	[DDIA] Chapter 4	Quiz 8	March 25
11	Week of March 25	Artificial Intelligence and Data Mining	[DLAPM] Chapter 5: Deriving Insights from a Data Lakehouse	[DLAPM] Chapter 8: Part 2.	Assignment 8: Build data insights from your data Lakehouse	[DDIA] Chapter 7 Part 1	Quiz 9	April 1
12	Week of April 1	Data Governance Applied to Data Lakehouse	[DLAPM] Chapter 6: Data Governance	[DLAPM] Chapter 8: Part 3.	Assignment 9: Turn in your Data Lakehouse as a report with descriptions.	[DDIA] Chapter 7 Part 2	Quiz 10	April 15
13	Week of April 8	Data Security in Data Lakehouse	[DLAPM] Chapter 7: Data Security	[DLAPM] Chapter 8: Part 4.	Assignment 10	No reading	No quiz	April 15
14	Week of April 15	Scaling Data Lakehouse	[DLAPM] Chapter 9 Scaling the Data Lakehouse Architecture	Practical data governance	No assignments	No readings	No quiz	April 22

15	Week of April 22	The Future of Data Engineering	[DDIA] Chapter 12 The Future of Data Systems	Practical matters in scaling	Course evaluation survey, Introductions survey	No readings	No quiz	April 29
16	Week of April 29	Course Recap	Course Recap	Course Recap	Course Recap	Course Recap	Course Recap	May 6
17	Week of May 6	Final Exam	Final Exam	Final Exam	Final Exam	Final Exam	Final Exam	Final Exam

Grade Policy

Grade Structure

Criteria	Number of Occurrences per Week	Points per Occurrence	Total Points	Percentage of Total
Reading Quizzes	10/16	30	300	30%
Homework Assignments	10/16	40	400	40%
Final Exam	1/16	100	100	10%
Final Project	4/16	100	400	20%
Total	16/16	-	1000	100%

Grade Bracket

A	90%-100%	Excellent work
B	80%-90%	Good work
C	70%-80%	Satisfactory work
D	50%-70%	Needs work
F	Below 50%	Failure to show competence

Academic Resources

How to succeed in this class

There are specific steps you take to achieve a good grade first; if you are having trouble finishing an in-class assignment, that will NOT automatically result in a bad grade. A good grade could be achieved if you showed effort and explained your thought process despite having an incorrect result. Complete all your assignments to the best of your abilities, and please ask questions if things need clarification. See your errors and mistakes as opportunities to learn more. I will provide you with feedback. Sometimes, the input will sound like criticism. Not everyone likes that. Please understand that the feedback comes from my intention to ensure you know the content and build the necessary skills**.**

Technology

Please let me know if you need a computer or a laptop for this class. If you need any assistance with technology, please reach out to the IT Solutions (https://twu.edu/technology/)

Library Services

Please don't buy textbooks or software that you might need for this class before checking in at the library first. https://twu.edu/library/. If you have any questions about the library or how to find a resource, please get in touch with me or Susan Whitmer (swhitmer@twu.edu)

Food Security

Minerva's Market is in The Student Union at Hubbard Hall, Room 1203
Social Work Food Pantry is in the Old Main Building, Room 406

Mental Health
If you need help with any issue that is affecting your academic performance, please refer to: https://twu.edu/student-health-services/mental-health/

or

https://twu.edu/counseling/

If you need immediate help, please go directly to Jones Hall Room 269 (M-F 8 AM to 6 PM) or call the Crisis Line: (940) 898-4357

Writing
If you have difficulty communicating in written English language, please let the instructor know, and please refer to the following:

https://twu.edu/write-site/

Tutoring
If you would like additional help for the class or any other classes, please notify the instructor and refer to:

https://catalog.twu.edu/graduate/services-available-students/tutoring-centers/

University Policies
For general university policies, please refer to https://web.saumag.edu/academics.

Disability Access Policy Statement
Texas Woman's University strives to make all learning experiences accessible. If you anticipate or experience academic barriers based on your disability (e.g., mental health conditions, learning disabilities, chronic medical conditions, etc.), please register with Disability Services for Students (DSS) to establish reasonable academic accommodations. After registration with DSS, please get in touch with me to discuss how to implement your accommodation.

DSS contact information: DSS website (https://twu.edu/disability-services/); dss@twu.edu; 940-898-3835; CFO Ste. 106.

If you have any questions regarding disability, please reach out to me or Nadaya Cross (ncross1@twu.edu)

Title IX: Sexual Violence Education
TWU is committed to providing a safe learning environment free of all forms of sexual misconduct, including sexual harassment, sexual assault, domestic violence, dating violence, and stalking. Federal laws (Title IX and the Violence Against Women Act) and TWU policies prohibit discrimination based on sex and therefore prohibit sexual misconduct. As students, if you or someone you know is experiencing sexual harassment, relationship violence, stalking, or sexual assault, there are campus resources available to provide support and assistance. Alleged sexual misconduct can be non-confidentially reported to the Title IX Coordinator at the Report an Incident website

(https://twu.edu/civility/report-an-incident/) or at (940) 898-2968. Additionally, please be aware that under Title IX of the Education Amendments of 1972, all employees must disclose information about such misconduct to the Title IX Office. Students who wish to speak to a confidential employee who does not have this reporting responsibility can contact TWU Counseling and Psychological Services at (940) 898-3801 for the Denton Campus, (214) 689-6655 for the Dallas Campus, and (713) 794-2059 for the Houston Campus.

Title IX: Pregnant Students
Title IX is a federal law that requires schools that receive federal funds to provide reasonable accommodation to students who are pregnant or have pregnancy-related conditions. This includes pregnancy, pre-natal doctor appointments, childbirth, false pregnancy, miscarriage, termination of pregnancy, or recovery from any of these conditions. Students needing academic accommodations due to pregnancy-related conditions should complete the Pregnancy Accommodation form (https://twu.edu/pregnancy-accommodation-form/) to coordinate educational needs.

Academic Integrity
Honesty in completing assignments is essential to the mission of the University and the development of the personal integrity of students. In submitting graded assignments, students affirm that they have neither given nor received unauthorized assistance and abided by all other provisions of the Academic Integrity Policy and the Student Code of Conduct as found on the TWU website and in the TWU Student Handbook. Cheating, plagiarism, collusion, dual submission of a paper, or other academic dishonesty will not be tolerated. It will result in appropriate sanctions, including failing an assignment, failing the class, being removed from an educational program, or being suspended or expelled. Allegations of academic dishonesty in this course may be reported to the Office of Civility and Community Standards. The specific disciplinary process for academic dishonesty is in the TWU Student Code of Conduct (https://public.powerdms.com/TWU1/documents/1745742) and Academic Integrity Academic Integrity Policy (https://public.powerdms.com/TWU1/documents/1748544). For details on avoiding plagiarism, review the library Tutorial: Avoiding Plagiarism (https://libguides.twu.edu/c.php?g=270163&p=1803990).

To ensure the integrity of the academic process, Texas Woman's University vigorously affirms the importance of academic honesty as defined by the Academic Integrity Policy and the TWU Student Code of Conduct. Therefore, Texas Woman's University faculty members may use Turnitin to compare a student's work with multiple sources to detect and prevent plagiarism. It then reports a similarity percentage and provides links to those specific sources. The tool itself does not determine whether a paper has been plagiarized. Instead, that judgment must be made by the individual faculty member. Some of the required assignments in this course may be checked for plagiarism using Turnitin.com.

Attendance Policy
Consistent attendance is vital to academic success and is expected of all students. Grades are determined by academic performance, and instructors may give students written notice that attendance related to specific classroom activities is required. Absence does not exempt students from academic requirements. Even if documented, excessive absences may result in a student's failing the course. Excused absences are within the purview of the instructor. Students must consult with instructors regarding make-up work.

Departmental Policies

Grading Policy
https://docs.google.com/document/d/1eeTJG916awbljyMG6zIOsuak2U_ozbCN/edit?usp=drive_link

AI Usage Policy
https://docs.google.com/document/d/1rPm6TjS8FRFTLbWD9ERbbBySfXCcHg9j/edit?usp=drive_link

Academic Honesty
https://docs.google.com/document/d/1n2yJvdRV2BzUb0QezPLeZwJyUJp_eE3Z/edit?usp=drive_link

Instructor Policies

Code of Conduct
The instructor classifies behavior as positive, negative, and neutral. Positive behavior is direct and healthy engagement with the instructor, the material, and colleagues. Positive behavior is crucial for fostering a growth-oriented learning environment. Neutral behavior is where the student is disengaged from the learning environment, the material, the instructor, and the colleagues for personal reasons. Negative behavior is behavior in bad faith. That is behavior that seeks to disrupt the learning environment actively. This includes toxic behaviors such as disrespect, acting out, drama, ridicule, retaliation for perceived slights, crossing people's limits, and verbal harassment. Any negative behavior in class towards the material, the instructor, or your colleagues will be met with zero tolerance. You will be asked to leave if negative behavior occurs in class or on Zoom. If you do not comply, public safety will be informed, or you will be dropped from the online meeting. If negative behavior occurs in an email message, the message will be ignored and reported. Please don't come to class in a negative state of mind. You may come to class if you are in a neutral state of mind and don't want to engage. You can also inform the instructor not to engage with you if that's the case. Or you can be excused to leave.

Late Assignments
Late work is not accepted unless there is a compelling reason. If this is the case, please communicate promptly.

Academic Dishonesty
I rely on Turnitin to gauge the level of plagiarism in your work. This applies to writing and coding. The above 30% similarity is too much, and I consider this plagiarism. Using internet resources is allowed with restrictions that the instructor will mention during the classroom. Yet please don't copy and paste code or answers for any questions.

Holidays
The instructor will follow the federally and state-recognized holiday schedule by the University, which can be found here:

https://twu.edu/media/documents/registrar/Calendar-at-a-Glance-2022-2023.pdf

However, if you need special accommodation for religious or other types of holidays that you observe, please let the instructor know beforehand.

Diversity
The instructor is committed to diversity, inclusion, and equality in the classroom and by the university policies, regardless of cultural background, country of origin, religion, race, ethnicity, and sexual orientation. Please let the instructor know how you would like to be addressed. During the first lecture, the instructor will ask the students about their names, pronouns, and other forms of addressing they want to be referred to. Please notify the instructor if that changed or if you were addressed mistakenly.

Office Hours
Generally, email is the best way to reach me. You may drop by my office anytime within the specified hours if you need help. However, it is preferred that you email first because multiple people might come at the same time. If that's the case, I will meet with people on a first come, first served basis. I will also have to limit the sessions to 15 minutes if people are waiting. If two or more students come simultaneously, it will be in the order of the last name. Please come prepared to office hours with questions.

Conflict
In case of conflict between students in the classroom, the instructor will act as a mediator until proper university authorities or public safety are notified.