CS 3312: Data Integration and Processing

University of Texas at El Paso
Department of Computer Science
Fall 2026

Instructor Information:
Monika Akbar
Email: makbar [at] utep.edu
Department of Computer Science, UTEP
Office: CCSB 3.0422
Office Hours: Monday (9 to 10 AM), Wednesday (1 to 2 PM), or by appointment.

Course Description:

This course will cover data storage and retrieval for relational and non-relational databases. Topics include data quality and data integration for AI-driven tools and methods, e.g., data cleaning, normalization, reconciliation, and data workflows.

CRN: 17322
Prerequisite(s): CS 2401 with a grade of C or better.
Location: QUIN 206
Days and Time: Tuesdays and Thursdays – 10:30 AM to 11:50 AM

Reference Books: The following are recommended reference books for the course. Additional books and resources may be provided throughout the semester as needed. All recommended books will be available through the UTEP Library.

1. Borthen, J. (2025). *Building Data Integration Solutions*. O'Reilly Media.
2. Reis, J., & Housley, M. (2022). *Fundamentals of Data Engineering: Plan and Build Robust Data Systems*. O'Reilly Media.
3. Doan, A., Halevy, A., & Ives, Z. (2012). *Principles of Data Integration*. Morgan Kaufmann.

Required Materials: Readings, tutorials, software documentation, and other course materials will be provided through Blackboard. Students should have access to a computer on which course software can be installed and a reliable Internet connection.

Grading:

Grades are communicated to students in a timely manner. It is the student's responsibility to keep track of their grades by compiling the grades they receive. The approximate grade distribution is as follows:

Attendance, active class participation, homework	20%
Exams, quizzes	50%
Project deliverables, assignments, and presentations	30%
Total	100%

The nominal percentage-score-to-letter-grade conversion is as follows:

- 90% or higher is an A
- 80-89% is a B

- 70-79% is a C
- 60-69% is a D
- below 60% is an F

These minimums may be lowered without notice but will not be raised.

Course Project: Students will design and implement a prototype data integration and processing workflow using heterogeneous data sources. The project will require students to acquire or ingest data, assess and improve data quality, integrate data across sources, process and transform the data, and produce a documented output suitable for downstream analysis or AI/ML use. Project deliverables may include intermediate design documents, demonstrations, reports, and presentations.

Course Policies and Expectations:

- **Blackboard and communication**
  - Blackboard will be used for announcements, course materials, assignments, and grades. Students should check Blackboard frequently, at least once a day.
  - Please use your UTEP e-mail account when contacting the instructor. Please add [CS3312] prefix in the subject of the email.
- **Academic integrity**
  - **Scholastic Dishonesty:** Any student who commits an act of scholastic dishonesty is subject to discipline. Scholastic dishonesty includes, but is not limited to, cheating, plagiarism, collusion, submission for credit of any work or materials that are attributable to another person. Cheating is copying from the test paper of another student; communicating with another student during a test to be taken individually; giving or seeking aid from another student during a test to be taken individually; possession and/or use of unauthorized materials during tests (i.e., crib notes, class notes, books, etc.); substituting for another person to take a test; falsifying research data, reports, or academic work offered for credit; or paying a person or company to complete coursework (i.e., contract cheating). Plagiarism is using someone's work in your assignments without the proper citations or submitting the same paper or assignment from a different course without direct permission of instructors. To avoid plagiarism, see http://sa.utep.edu/osccr/wp-content/uploads/sites/8/2012/09/AvoidingPlagiarism.pdf. Collusion is unauthorized collaboration with another person in preparing academic assignments. *All suspected violations of academic integrity at The University of Texas at El Paso are reported to the* Office of Student Conduct and Conflict Resolution (OSCCR) *for possible disciplinary action.* Students are expected to comply with student standards of conduct; see HOOP: Student Conduct and Discipline.
- **AI technologies**
  - Use of generative AI or other automated tools is permitted only when explicitly authorized for a course activity or assignment.
  - When authorized, students must follow the instructor's directions for disclosure, citation, and documentation of AI use. Students remain responsible for the accuracy, originality, and understanding of submitted work. If given permission to use any of these tools, students must properly cite and give full credit to the program used upon submission of every relevant assignment. For example, text generated using ChatGPT must be cited:

Chat-GPT(version). Date of query (year/month/day). "Text of your query." Generated using OpenAI. https://chat.openai.com/

- **Attendance and participation**
  - Participation in lectures, class activities, discussions, quizzes, and other assigned learning activities is important for successful completion of the course.
  - Attendance will be taken at every class and will count towards your class participation grade. Students should be on time for all scheduled sessions and attend the entire session.
  - Activities designed for a particular class session or week may not be available for make-up after the deadline unless otherwise specified.
- **Deadlines, late work, and make-up work**
  - Make-up work, exams, or quizzes may be provided for documented emergencies or other circumstances approved by the instructor. Students should contact the instructor as early as possible when a course requirement will be missed.
- **Course materials**
  - Course materials are intended for students enrolled in this course and should not be redistributed publicly without permission.
- **Special accommodations:** Students requesting accommodation based on a disability must register with the UTEP Center for Accommodations and Support Services (CASS). Contact the Center for Accommodations and Support Services at 915-747-5148, or email them at cass@utep.edu, or apply for accommodations online via the CASS portal.
- **Student resources:** UTEP provides a variety of student services and support:
  Technology Resources
  - **Help Desk:** Students experiencing technological challenges (email, Blackboard, software, etc.) can submit a ticket to the UTEP Helpdesk for assistance. Contact the Helpdesk via phone, email, chat, website, or in person if on campus.

  Academic Resources
  - **UTEP Library:** Access a wide range of resources including online, full-text access to thousands of journals and eBooks and librarian assistance for enrolled students.
  - **University Writing Center (UWC):** Submit papers here for assistance with writing style and formatting, ask a tutor for help and explore other writing resources.

  Individual Resources
  - **Military Student Success Center:** Assists personnel in any branch of service to reach their educational goals.
  - **Center for Accommodations and Support Services:** Assists students with ADA-related accommodations for coursework, housing, and internships.
  - **Counseling and Psychological Services:** Provides a variety of counseling services including individual, couples, and group sessions as well as career and disability assessments.

  Note: When in doubt about course requirements, authorized procedures, or the use of tools and resources, contact the instructor.

Learning Outcomes

Level 1: Knowledge and Comprehension.

Level 1 outcomes are those in which the student has been exposed to the terms and concepts at a basic level and can supply basic definitions. The material has been presented only at a superficial level. Upon successful completion of the course, students will be able to:

1a. Describe the impact of data quality on AI algorithms.
1b. Describe and compare data models including structured, semi-structured, and how they can be used for data integration and processing.
1c. Identify current challenges in data integration, in particular, fairness and reliability.
1d. Describe technical solutions to data management challenges.
1e. Describe processes for data collection and processing for AI models.

Level 2: Application and Analysis.

Level 2 outcomes are those in which the student can apply the material in familiar situations, e.g., can work a problem of familiar structure with minor changes in the details. Upon successful completion of the course, students will be able to:

2a. Demonstrate uses of metadata associated with structured and non-structured data.
2b. Use a relational query language (e.g., MySQL) to elicit information from a relational database.
2c. Use a non-relational query language (e.g., SPARQL) to elicit information from a non-relational database.
2d. Use and extend current methods and tools for data processing, with a focus on those used in AI, e.g., for retrieving, cleaning, and pre-processing data.
2e. Demonstrate the ability to design applications that solve real-world problems using data-driven AI methods.

Level 3: Synthesis and Evaluation.

Level 3 outcomes are those in which the student can apply the material in new situations. This is the highest level of mastery. Upon successful completion of the course, students will be able to:

3a. Design a high-level data integration model from a problem statement using standard notation and modeling principles.
3b. Design and implement a data workflow that enables the integration of data from heterogeneous, decoupled data for an AI model applying best practices for fairness and reliability.