



INFO 531/ISTA 431: Data Warehousing and Analytics in the Cloud

Course Description

Data Warehousing and Analytics In the Cloud will utilize concepts, frameworks, and best practices for designing a cloud-based data warehousing solution and explore how to use analytical tools to perform an analysis of data. The first few sections of the course will provide an understanding of data warehousing. The next few sections will provide an overview of the field of cloud computing, and its main concepts by utilizing cloud computing platforms. The last few sections of the course will examine data preparation, modeling, and data analytics using Python-based machine learning libraries and Pandas DataFrame.

Course Prerequisites

Familiarity with database concepts and basic SQL query syntax.

Instructor and Contact Information

Instructor: Nayem Rahman, PhD

Email: nrahman1@arizona.edu

Cell Phone: 503-317-9940 [Please send me a text before calling]

Office hour: By appointment

D2L Course URL: <https://d2l.arizona.edu/d2l/home/1202776>

Course Format and Teaching Methods

The instructor will supply the course content to the students in various formats, including slides, reading materials, and explanatory videos by module. The general course structure is organized in terms of module materials, assessments, weekly discussion forums, homework assignments, and final projects (graduate students only). Module materials include weekly reading, lecture slides, explanatory videos, and hands-on exercises provided by the instructor and organized in modules. To succeed in this course, 3 hours of study time per hour of the lecture are required, including reading textbooks, working on assigned works, and participating in discussion forums.

Expected Learning Outcomes

By the end of the semester students will be able to:

- Analyze various cloud programming models and apply them to solve problems on the cloud
- Design tables and indexes to efficiently distribute data in tables across many nodes
- Implement data warehousing techniques on-premise and on cloud platforms
- Write queries and evaluate query plans to optimize query performance
- Use machine learning libraries in Python
- Perform extensive data manipulation exercises using Python libraries and Pandas functions
- Use a BI application to perform data analysis and visualization tasks

Absence and Class Participation Policy

The UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at the university website. Actively participating in the course every week by viewing lectures, responding to discussions, and participating in other course events are vital to the learning process. Absences may affect a student's final course grade. If you anticipate being absent, are unexpectedly absent for an extended period of time, or are unable to participate in class activities, please contact me as soon as possible. *To request a disability-related accommodation to this attendance policy, please contact the Disability Resource Center at (520) 621-3268.* If you are experiencing unexpected barriers to your success in your courses, the Dean of Students Office is a central support resource for all students and may be helpful. The Dean of Students Office is located in the Robert L. Nugent Building, room 100, or call 520-621-7057.

Makeup Policy for Students Who Register Late

Students who register late for the course will be afforded extended due dates for the first two weeks of class. After the initial two weeks, all due dates will be reinforced.

Course Communications

All class-wide course communication will happen within the course D2L shell. Personal communication, one-on-one communications, will be sent to student UA email addresses.

Required Texts or Readings

The following book is the main sources we will use in this course:

- Zburivsky, Danil (2021). Designing Cloud Data Platforms, Manning.

Recommended Texts or Readings

All course materials will be supplied through the D2L shell. The following books are the main sources we will use heavily in this course:

- Inmon, Bill. Building the Data Lakehouse, Technics Publications, 2021.
- Klint, Lars. Microsoft Azure in Action, Manning.
- Foster, Ian. Cloud Computing for Science and Engineering, MIT Press.
- How, Matt. The Modern Data Warehouse in Azure, Apress.
- Inmon, W.H. (2005). Building the Data Warehouse. John Wiley & Sons, Inc. ISBN-13: 978-0764599446
- Grippa, Vinicius M., and Kuzmichev, Sergey (2021). Learning MySQL: Get a Handle on Your Data. O'Reilly Media. ISBN-13: 978-1492085928
- Chen, Daniel Y. (2018). Pandas for Everyone: Python Data Analysis, Pearson Addison-Wesley.

Assignments and Examinations

Readings: All readings will be provided within the course. Students will be expected to read both formal and informal articles, books, and documents.

Written Assignments

Assignments will be posted on D2L before the start of each week.

Grading Scale and Policies

Grade Distribution for this Course:

Undergraduate		Grade Explanation
A	90-100%	Exemplary work
B	80-89%	Expectations were exceeded
C	70-79%	Expectations were met
D	60-69%	Expectations were partially met
E	≤ 59%	Expectations were not met

Late Work Policy: As a rule, work will not be accepted late except in case of documented emergency or illness. You may petition the professor in writing for an exception if you feel you have a compelling reason for turning work in late.

Dispute of Grade Policy: To dispute a grade on a quiz, project, or exam, please email the instructor within one week and provide a clear explanation of where the discrepancy exists.

Course Calendar/ Schedule

Tentative Schedule (Subject to change)

	Week of	Topics	Readings
1	01/11	Decision Support Systems (DSS) Introduction to Data Warehousing	Inmon (2005)
2	01/18	SQL Basics; ETL Metadata Models	Journal Articles/Technical Papers
3	01/25	Data Warehousing Governance Best Practices; Performance Optimization in Data Warehouses	Journal Articles/Technical Papers
4	02/01	Introduction Cloud Computing	Zburivsky (2021)
5	02/08	Cloud data platform architecture	Zburivsky (2021)
6	02/15	Data Platform and Data Warehouses	Zburivsky (2021)
7	02/22	Cloud Data Storage Models	Zburivsky (2021)
8	03/01	Cloud Data Warehousing	Zburivsky (2021)
XX	03/08	SPRING BREAK - March 4-12, 2023	None
9	03/15	Data Platform in Cloud	Zburivsky (2021)
10	03/22	Incremental Refresh in Data Warehousing	Journal Articles/Technical Papers
11	03/29	Temporal Data Update Methodologies	Journal Articles/Technical Papers
12	04/05	Data Collection & Storage Using Python	Chen (2018)
13	04/12	Data Processing Using Python	Chen (2018)
14	04/19	Data Analytics and Machine Learning I	Chen (2018)
15	04/26	Data Analytics and Machine Learning II	Chen (2018)
16	05/03	Final Project	

Final Project

There is a final project in place of a final exam for this class. In groups, you will work collaboratively to design and implement a data warehousing or machine learning project.

Assessment and Grade Distribution

- Discussion Posts: 10%
- Assignments: 75%
- Final Project: 15%

Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed.

Threatening Behavior Policy

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See <http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students>.

Accessibility and Accommodations

At the University of Arizona, we strive to make learning experiences as accessible as possible. If you anticipate or experience barriers based on disability or pregnancy, please contact the Disability Resource Center (520-621-3268, <https://drc.arizona.edu/>) to establish reasonable accommodations.

Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog.

The University Libraries have some excellent tips for avoiding plagiarism, available at <http://new.library.arizona.edu/research/citing/plagiarism>.

Selling class notes and/or other course materials to other students or to a third party for resale is not permitted without the instructor's express written consent. Violations to this and other course rules are subject to the Code of Academic Integrity and may result in course sanctions. Additionally, students who use D2L or UA e-mail to sell or buy these copyrighted materials are subject to Code of Conduct Violations for misuse of student e-mail addresses. This conduct may also constitute copyright infringement.

UA Nondiscrimination and Anti-harassment Policy

The University is committed to creating and maintaining an environment free of discrimination; see <http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy>

Our classroom is a place where everyone is encouraged to express well-formed opinions and their reasons for those opinions. We also want to create a tolerant and open environment where such opinions can be expressed without resorting to bullying or discrimination of others.

Additional Resources for Students

UA Academic policies and procedures are available at <http://catalog.arizona.edu/policies>

Confidentiality of Student Records

<http://www.registrar.arizona.edu/personal-information/family-educational-rights-and-privacy-act-1974-ferpa?topic=ferpa>

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.