



ISM 6208:

Data Warehousing

CRN: 58155, Section: 020, Credit Hours: 3

CRN: 53321, Section: 021, Credit Hours: 3

Muma College of Business

School of Information Systems and Management

COURSE SYLLABUS

Semester: Fall 2023

Class Meeting Days: Monday

Class Meeting Time: 5:00 – 8:30 PM

Class Meeting Location: TBA

Instructor: Don Berndt

Office Location: CIS 2073

Office Hours: W 7-8 PM (Online)

Office Phone: (813) 974-6769

Email: dberndt@usf.edu

I. Welcome!

Welcome to the data warehousing class (ISM 6208). Data science is an exciting area and major theme of our graduate programs. This is clearly a multi-disciplinary field, but effective data management is certainly a core skill. This course focuses on data warehousing and visualization techniques, including dimensional modeling and data cubes, analytic SQL for query writing, data visualization (with tools like Tableau) and data warehouse administration.

This course includes significant hands-on experience, so you will apply what you learn. The discussions and assignments will also reinforce important competencies such as creative thinking, teamwork and digital skills (see the nationally-standardized career readiness competencies at <http://www.nacweb.org/career-readiness/competencies/career-readiness-defined/> for details).

II. University Course Description

This course is designed for the MS graduate student and interested MBA students. The course covers the rapidly emerging data warehousing and data mining technologies that are likely to play a strategic role in business organizations.

You can find the official description at <http://ugs.usf.edu/course-inventory/>.

III. Course Prerequisites

The Advanced Database Management (ISM 6218) course or similar two-semester database course sequence (and/or significant professional experience).

IV. Course Purpose

This course is designed for the MS graduate student and interested MBA students. The course covers the rapidly emerging data warehousing and data mining technologies that are likely to play a strategic role in business organizations. Topics include the differences between operational and analytical database systems, dimensional modeling (data cubes) and star schemas, data warehouse performance issues, data quality, the data warehouse development life cycle, data warehouse navigation, and a brief overview of selected data mining techniques. The Oracle database system will be used to illustrate many of the concepts covered in class, as well as providing a platform for hands-on projects. As a prerequisite, students should have had at least two courses covering relational database systems (usually including ISM 6218: Advanced Database Systems), or significant work experience.

The class focuses on a combination of theory and practical database use, which will be illustrated primarily through the Oracle DBMS software. The Oracle database engine is freely available for educational use from Oracle Corporation. Therefore, students will have access to many of the sophisticated tools at home and on mobile platforms. In addition, access to local database servers will be provided, so all hands-on projects can be pursued in an industrial quality environment. Though the primary database engine will be the Oracle DBMS, several cloud-based data warehouse platforms will be covered to highlight common features, as well as alternative approaches.

V. How to Succeed in this Course

First, please visit Canvas often and monitor all course announcements. Canvas will be the primary means of communicating, as well as organizing and presenting the course content. For each week, review the most important material identified in “Start Here” pages. Our in-class time is designed to further and deepen the knowledge, concepts, and skills from the readings – not just repeat the information. We will also use hands-on laboratory activities to reinforce and apply skills. USF also suggests many study tips, so have a look at the resources below.

Study skills of successful students: This one-page handout details what the study skills behaviors of top students look like: <http://bit.ly/successfulstudentbehaviors>. This 18-page PDF provides 101 individual tips for effective study skills and note-taking: <http://bit.ly/studyskillstips>

Best practices for hybrid courses: Successful students should follow these tips to succeed in classes where they switch between remote and in-person attendance:
<https://www.usf.edu/atle/documents/student-best-practices-hybrid.pdf>

Canvas: See the Community Guides at <https://community.canvaslms.com/t5/Student/gh-p/student>.

Micrsoft Teams: See best the practices video for Teams at <https://youtu.be/q6zqaSMtnig>. For some written guides, see the student introduction (<https://www.usf.edu/atle/documents/student-guide-microsoft-teams.pdf>) and etiquette (<http://www.usf.edu/atle/documents/teams-etiquette.pdf>) documents.

VI. Course Topics

The course topics are divided into seven major parts, including hands-on coverage of data warehouse design and implementation techniques.

Part N (0) – Introducing the Data Warehouse: The first meetings provide an overview of data warehousing and data mining technologies. The differences between operational and analytic databases are highlighted, along with matters of scale (data marts and data warehouses). Important presentation technologies, including online analytic processing (OLAP) tools, geographic information systems and data mining techniques are briefly introduced.

Part I – Designing the Data Warehouse: Key data warehouse design techniques involving dimensional modeling and star schemas, as well as the role of metadata, are studied in detail. This provides the foundation necessary to pursue the in-class and semester projects. A series of case studies are used to illustrate issues throughout the semester.

Part II – Querying the Data Warehouse: The focus of this section is interface design, emphasizing the ability to effectively navigate, pursue ad-hoc queries, and support decision-making tasks in a data warehouse environment.

Part III – Visualizing the Data Warehouse: The focus of this section is data visualization, from reporting to storytelling. Tools like Tableau and PowerBI used for hands-on activities.

Part IV – Mining the Data Warehouse: The focus of this section is using data cubes as input to data mining and machine learning. While these topics are covered in other classes, some examples here showcase the whole lifecycle in data warehousing.

Part V – Administering the Data Warehouse: Data warehouse administration issues such as capacity planning, indexing strategies, partitioning and aggregation, the role of parallelism, and performance tuning issues will be covered. In addition, storage subsystems and related data warehousing infrastructure requirements will be reviewed.

Part VI – Managing the Data: Most of the effort on a project involves the data extraction, transformation, and loading activities required to populate the data warehouse. These techniques, along with data quality issues are highlighted in this section.

Part VII – Managing the Data Warehouse: Managerial issues involving business strategy and project justification are considered based on outside readings and case studies. In addition, the data warehouse development lifecycle and project team requirements are discussed.

Part VIII – Extending the Data Warehouse: Special topics on knowledge discovery in databases, data and text mining techniques, geographic information systems, and other approaches that leverage data warehouses are covered as time permits.

VII. Student Learning Outcomes

By the end of this course, students will be able to:

1. Describe a variety of data warehousing technologies.
2. Contrast analytic databases with online transaction processing databases.
3. Write queries using analytic SQL extensions.
4. Design data cubes using dimensional modeling.
5. Implement data warehouses using SQL data definition statements and integrity constraints.
6. Apply various techniques such as index structures and optimization modes to improve data warehouse query performance.
7. Contrast alternative cloud-based data warehousing technologies with the relational model.

VIII. Required Texts and/or Readings and Course Materials

Required Text: Kimball and Ross, *The Data Warehouse Toolkit*, Third Edition, John Wiley & Sons, 2013. This book provides excellent coverage of dimensional modeling and a collection of case study chapters that review design patterns.

Outside readings will be used to supplement the textbooks.

You should also have access to a fairly current database textbook for review purposes (if necessary). There are also plenty of Oracle DBMS online references (such as *Oracle: The Complete Reference*), documents and tutorials. We will use the Oracle DBMS to implement different techniques for the semester project., though you are free to use other database systems as well.

IX. Supplementary (Optional) Texts and Materials

Oracle maintains an extensive library of online documentation that students will find useful during the course (see <https://docs.oracle.com/en/database/oracle/oracle-database/>).

X. Grading Scale

Grading Scale (%)	
94-100	A
90 – 93	A-
87 – 89	B+
84 – 86	B
80 – 83	B-
77 – 79	C+
74 – 76	C
70 – 73	C-
67 – 69	D+
64 – 66	D
60 – 63	D-
0 – 59	F

XI. Grade Categories and Weights

The course relies on three major types of assessments: 1) four group assignments1) online quizzes (as self-paced checks on core modules), 2) group assignments that apply skills and 3) a final project that requires some creative thinking and the synthesis of skills. These activities are weighted as show below.

Assessment	Percent of Final Grade
Online Quizzes (8)	40%
Assignments (4) – Adjusted based on course duration.	40%
Final Project	15%
Participation and Attendance	5%

XII. Course Assessments

There are several categories of assignments or activities used for assessment, including the following items.

1. **Online quizzes** (8) are part of some core modules (following the introductory module), accounting for **40%** of the grade.
2. Assignments (4) are given in selected modules, with roughly two weeks to complete (for a total of **40%** of the grade). The assignments exercise skills in 1) dimensional modeling, 2) analytic SQL query writing, 3) data visualization and 4) data warehouse performance tuning. These assignments are intended to be group activities. Note: For shorter versions of the course (6-8 weeks), one assignment may be dropped.

3. The **final project (15%)** The final project (also a group effort) focuses on synthesizing all the skills on a project of interest to the students. Typically, this involves identifying some interesting data sets, loading the data (using ETL techniques), designing data cubes, writing analytic queries and/or data visualization.
4. Class **participation (5%)** will be based on contributions in class, discussion boards and/or other collaborative **meetings**. This will along with peer assessments from group members will be used to reward highly engaged participants.

Group assignments and the final project are to be completed by a small group (approximately 3-5 people) and involve the design and implementation of database experiments, as well as query writing, design, and even database administration activities. The application areas may extend one of the example databases design used in class or based on a work-related project (with prior approval).

XIII. Grade Dissemination

The Canvas system will be used throughout the course. Students can access scores and other information at any time using “Grades” in Canvas.

XIV. Course Schedule

This is a tentative course schedule for a 12-week format. Of course, adjustments would be made based on course duration such an 8-week format.

Week	Modules	Assignments
1/12	Part N: Introducing the Data Warehouse	NA
2/12	Part I: Designing the Data Warehouse	Quiz 1: Dimensional Modeling Assignment 1: Dimensional Modeling
3/12	Case Study: Retail Sales Case Study: Inventory	Quiz 2: Retail Sales CS Quiz 3: Inventory CS
4/12	Part II: Analytic SQL	Quiz 3: DBMS Architecture Assignment 2: Query Writing
5/12	Case Study: Procurement Case Study: Order Management	Quiz 4: Procurement CS Quiz 5: Order Management CS
6/12	Part III: Visualizing the Data Warehouse	Quiz 6: Visualization Assignment 3: Visualization
7/12	Case Study: Superstore Data Visualization	NA
8/12	Part IV: Mining the Data Warehouse	Quiz 7: Mining the DW
9/12	Special Topic: Decision Tree Induction	NA
10/12	Part V: Administering the Data Warehouse	Quiz 8: Administering the DW
11/12	Part VI: Managing the DW (Data Quality)	Final Project
12/12	Part VII: Extending the DW (Special Topics)	NA

* Note: The Schedule is subject to revision.

XV. Standard University Policies

Policies about disability access, religious observances, academic grievances, academic integrity and misconduct, academic continuity, food insecurity, and sexual harassment are governed by a central set of policies that apply to all classes at USF. These may be accessed at: <https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx>

XVI. Covid-19 Procedures

All students must comply with university policies and posted signs regarding COVID-19 mitigation measures, including wearing face coverings and maintaining social distancing during in-person classes. Failure to do so may result in dismissal from class, referral to the Office of Student Conduct and Ethical Development, and possible removal from campus.

Additional details are available on the University's Core Syllabus Policy Statements page: <https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx>

XVII. Course Policies

Late Work Policy: Graduate students may be balancing work demands with their studies. Extensions may be possible on a per-assignment basis in advance of a due date. Unexcused late assignments will be assessed a penalty: a half-letter grade if it is 1-2 days late, or a full-letter grade for 3-7 days late. Assignments will not be accepted if overdue by more than seven days.

Grades of "Incomplete": Grades of incomplete are to be avoided if at all possible (especially at the graduate level). Any students that are at risk of not completing the course should reach out to the instructor as soon as possible. The preferred approach is to turn in most of the assignments even in draft form and receive a final grade, which can be adjusted later once the final versions of any missing assignments are submitted. The formal university policies governing an "I" grade are as follows.

For undergraduate courses: An "I" grade may be awarded to a student only when a small portion of the student's work is incomplete and only when the student is otherwise earning a passing grade. The time limit for removing the "I" is to be set by the instructor of the course. For undergraduate students, this time limit may not exceed two academic semesters, whether or not the student is in residence, and/or graduation, whichever comes first. For graduate students, this time limit may not exceed one academic semester. "I" grades not removed by the end of the time limit will be changed to "IF" or "IU," whichever is appropriate.

For graduate courses: An Incomplete grade ("I") is exceptional and granted at the instructor's discretion only when students are unable to complete course requirements

due to illness or other circumstances beyond their control. The course instructor and student must complete and sign the "I" Grade Contract Form that describes the work to be completed, the date it is due, and the grade the student would earn factoring in a zero for all incomplete assignments. The due date can be negotiated and extended by student/instructor as long as it does not exceed two semesters for undergraduate courses and one semester for graduate courses from the original date grades were due for that course. An "I" grade not cleared within the two semesters for undergraduate courses and one semester for graduate courses (including summer semester) will revert to the grade noted on the contract.

Group Work Policy: Everyone must take part in a group project. All members of a group will receive the same score; that is, the project is assessed, and everyone receives this score. In addition, However, every person in the group will provide the instructor with a peer assessment at the end of the semester. The instructor will adjust the total assignment grade based on peer suggestions (by up to 10%).

Email: The Canvas Inbox is the preferred method of communication, but you may use official USF email as a second method (see the email address above). The goal is to respond within 1-2 days to any request.

Canvas: This course will be offered via USF's learning management system (LMS), Canvas. The Canvas Inbox feature is the preferred method of communication. If you need help learning how to perform various tasks related to this course or other courses being offered in Canvas, please view the following videos or consult the Canvas help guides. You may also contact USF's IT department at (813) 974-1222 or help@usf.edu.

XVIII. Course Policies: Student Expectations

Attendance Policy: Students are expected to attend classes. Faculty must inform students of attendance requirements on syllabi. Instructors should accommodate excused absences by making arrangements with students ahead of time (when possible) or by providing a reasonable amount of time to make up missed work.

Course Hero / Chegg Policy: The [USF Policy on Academic Integrity](#) specifies that students may not use websites that enable cheating, such as by uploading or downloading material for this purpose. This does apply specifically to Chegg.com and CourseHero.com – any use of these websites (including uploading proprietary materials) constitutes a violation of the academic integrity policy.

Professionalism Policy: Per university policy and classroom etiquette; mobile phones, iPods, etc. **must be silenced** during all classroom and lab lectures. Those not heeding this rule will be asked to leave the classroom/lab immediately so as to not disrupt the learning environment. Please arrive on time for all class meetings. Students who habitually

disturb the class by talking, arriving late, etc., and have been warned may suffer a reduction in their final class grade.

End of Semester Student Evaluations: All classes at USF make use of an online system for students to provide feedback to the University regarding the course. These surveys will be made available at the end of the semester, and the University will notify you by email when the response window opens. Your participation is highly encouraged and valued.

Food and Drink Policy: Please adhere to the firm policy of no beverages (other than bottled/capped water), food, tobacco products, or like items in the classroom. Your understanding of the necessity for this policy and cooperation will be greatly appreciated. This policy will be strictly enforced.

Turnitin.com: In this course, turnitin.com may be utilized. Turnitin is an automated system which instructors may use to quickly and easily compare each student's assignment with billions of web sites, as well as an enormous database of student papers that grows with each submission. After the assignment is processed, as instructor I receive a report from turnitin.com that states if and how another author's work was used in the assignment. For a more detailed look at this process visit <http://www.turnitin.com>.

XIX. Learning Support and Campus Offices

Tampa Campus

Tutoring Hub

Example: The Tutoring Hub offers free tutoring in [several subjects](#) to USF undergraduates. Appointments are recommended, but not required. For more information, email asctampa@usf.edu

Writing Studio

Example: The Writing Studio is a free resource for USF undergraduate and graduate students. At the Writing Studio, a trained writing consultant will work individually with you, at any point in the writing process from brainstorming to editing. Appointments are recommended, but not required. For more information or to make an appointment, email writingstudio@usf.edu

Counseling Center

Example: The Counseling Center promotes the wellbeing of the campus community by providing culturally sensitive counseling, consultation, prevention, and training that enhances student academic and personal success. Contact information is available [online](#).

Center for Victim Advocacy

Example: The Center for Victim Advocacy empowers survivors of crime, violence, or abuse by promoting the restoration of decision making, by advocating for their rights, and by offering support and resources. Contact information is available [online](#).

St. Petersburg Campus

Student Success Center

Example: The Student Success Center provides free tutoring and writing consultations. Contact information is available [online](#).

Wellness Center

Example: The Wellness Center provides counseling and medical services, as well as prevention programs and victim advocacy. Contact information is available [online](#).

Sarasota-Manatee Campus

Tutoring and Writing Support

Example: Learning Support Services provides free tutoring and writing consultations for a variety of courses and subjects such as, Accounting, Biology, Chemistry, Finance, Math & Statistics, Physics, and Spanish. Make an appointment [online](#).

Counseling and Wellness Center

Example: The Counseling and Wellness Center is a confidential resource where you can talk about incidents of discrimination and harassment, including sexual harassment, gender-based crimes, sexual assault, stalking, and domestic/relationship violence. Call 941-487-4254

Victim Advocate

Example: A Victim Advocate is available 24/7 by calling (941) 504-8599. For assistance leave a message with your phone number and your call will be returned as soon as possible. The Victim Advocate is available to assist victims of crime, sexual assault, and partner violence.

XX. Important Dates to Remember

All the dates and assignments are tentative and can be changed at the discretion of the professor. To confirm these dates, please see the official USF calendar maintained by the Registrar (see <https://www.usf.edu/registrar/calendars/>).