MIS 430: Advanced Database and Data Warehousing: Design and Implementation

Instructor
Don Jernigan
Contact: e-mail: djerniga@gmu.edu; Cell: (703) 740-6903;
Lync/ Skype: don.jernigan@cleverspeck.com
Office Hours: By appointment

Course Materials
Textbook:
There is no textbook for this class. You are required to take class notes.
Lecture slides and other reference materials will be made available using blackboard.

Software:
Microsoft SQL Server 2012 Developer Edition
This is widely used commercially to support an organization’s data management requirements in form of database, data warehouse, and business intelligence.

You are required to download and install the SQL Server Developer Edition 2012 from the MSDN Academic Alliance.
Detailed instructions will be provided on how to install on your own laptop. You are required to bring your laptop to class for in-class hands-on instruction and in-class exercises.

Minimum requirement for laptop is:
• 2GB of RAM (recommended 4GB) and
• 10 GB of free disk space (recommended 20 GB).

Additional Data Modeling software will be used in the course.

Course Description
Data management is critical to efficient and effective operation of all modern organizations. Businesses collect large amount of data as part of the daily operations, it is vital to ensure that databases be suitably designed to ensure data quality as well as speed of operations required for storage as well as retrieval of this information. In the first part of this course, students will learn implementation designs and database administration policies that are used for optimizing databases for business operations. As part of this we will cover some advanced SQL for implementation of database elements such as views, constraints, and triggers as well SQL for retrieval of data using complex queries.

Traditionally, organizations undertook data collection and storage with the objective of assisting day to day operations, such as order processing, inventory management, payroll, etc. This was achieved through use of databases that supported a variety of business applications. However, these databases could not provide insight required for strategic decision making. For e.g., helping managers in deciding where to open the next store? Which customer segment to target for which types of promotion? Furthermore, unlike day-to-day operations, strategic decision making requires access to large amounts of historical information. Additionally, most of these queries are ad-hoc queries and involve large number of attributes. These criterions make transactional databases ineffectual in providing dedicated support for such decision making. Data warehousing focuses on providing just this support. Data warehousing refers not just to the design and storage of historical information but identifies the entire infrastructure involved in enabling the necessary decision making. In the second part of the course, students will learn the design principles for a data warehouse, and utilize these to create suitable data warehouse designs that meet the business requirements. Students will also learn hands-on skills and implement a data warehouse to support the business intelligence requirements.

Learning Objectives
This course aims to provide a comprehensive understanding of database management and data warehousing, with specific focus on:

✓ Physical implementation of database optimized to serve business requirements.
✓ Apply the understanding to implement database objects such as views, indexes, triggers, and check constraints.
✓ Apply the SQL for both database implementations as well as information retrieval.
✓ Modeling and implementation of database using Data Modeling software.
✓ Understanding the difference in the role of transaction databases and data warehouses, their objectives and how it influences the process of design and implementation of each
✓ Architecture of data warehouses, and associated decision support systems
✓ Design approaches for data warehousing, with specific focus on relational data warehouse design
✓ Implementation and operationalization of data warehouses
✓ Evaluating and tuning performance of data warehouses
✓ Use of data warehouses for Reporting, and OLAP
✓ Creating business intelligence reports
✓ Develop extract, transform, and load processes

Grading
Students must be officially registered in this course to receive a grade. It is the sole responsibility of the student to verify their own registration status. Specifically, you will not receive a grade if your name does not appear on the official class list. (Don’t wait until the end of the semester to be surprised.) Registration problems should be directed to either the SOM Office of Student Services or the Registrar’s Office. Grading for the course will be based on total points earned by the end of the course. Final course letter grade assignments will be as follows:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>93% - 100%</td>
</tr>
<tr>
<td>A-</td>
<td>90% - 93%</td>
</tr>
<tr>
<td>B+</td>
<td>87% - 89.99%</td>
</tr>
<tr>
<td>B</td>
<td>83% - 87.99%</td>
</tr>
<tr>
<td>B-</td>
<td>80% - 82.99%</td>
</tr>
<tr>
<td>C+</td>
<td>76% - 79.99%</td>
</tr>
<tr>
<td>C</td>
<td>70% - 75.99%</td>
</tr>
<tr>
<td>D</td>
<td>64% - 69.99%</td>
</tr>
<tr>
<td>F</td>
<td>Below 64%</td>
</tr>
</tbody>
</table>

Mid Term Exam = 25%
Final Exam = 35%
Quizzes = 20%
Assignments = 20%
Total = 100%

Attendance
Class Attendance is your responsibility, and you are solely responsible for all assignments, material presented/provided and announcements made in class.

Make up Exams/Quizzes
Barring extenuating circumstances no exceptions will be made for absence. Final exam due date/time will not be rescheduled. Adequate proof needs provided to prove extenuating circumstances. Work related time conflict does not constitute extenuating circumstance. If you have any work related time conflict, you need to resolve it in advance.

Learning Goals
Learning goals for the SOM Undergraduate Programs
a. Our students will be competent in their discipline.
b. Our students will be aware of the uses of technology in business.
c. Our students will be effective communicators.
d. Our students will have an interdisciplinary perspective.
e. Our students will be knowledgeable about global business and trade.
f. Our students will recognize the importance of ethical decisions.
g. Our students will be knowledgeable about the legal environment of business.
h. Our students will be knowledgeable about team dynamics and the characteristics of effective teams.
i. Our students will understand the value of diversity and the importance of managing diversity in the context of business.
j. Our students will be critical thinkers.

Learning Goals of the Information Systems and Operations Management Program
a. Apply knowledge of information technology and business functions to understand its application in assessing, designing and improving business processes.
b. Develop data organization, storage and processing solutions to support organizational needs for information management. They will also have the option of developing skills in the area of supporting decision making through business intelligence solutions.
c. Use knowledge of computer networks as part of the IT solutions for improving business processes. They will also have option of developing more advanced skills in the areas of network and security.
d. Effectively manage information technology projects.

Understand the overall systems development life cycle and be able to recommend IT system solutions accordingly. They will also have option of learning appropriate development tools to develop prototype of IT solutions for business management.

Disability
All academic accommodations due to disability must be arranged through the Disability Resource Center (DRC). If you are a student with a disability and you require academic accommodations, please contact the DRC at 993-2474. I will cooperate fully with the DRC to accommodate a student’s special needs.
**Honor Code**

GMU students are expected to be familiar with the Honor Code of George Mason University and with its specific application to exams, assignments, and class work required by faculty in the program. If you are in doubt about how the honor system applies to a particular assignment or class, it is your responsibility to clarify the requirements with the professor. Concerns about breaches of the honor system may be discussed with the professor, or with the Associate Dean of the School of Management. More detail on honor code provided on class website.

**Communications**

All communications from me to you will be directed via e-mail. I will only address and reply to all of e-mails from/to your @gmu.edu e-mail address for concerns of privacy and confidentiality. If you use another e-mail account as your primary e-mail, please be sure to forward your gmu e-mail to that account.

Your communications with me – e-mail is the preferred channel of communication, since it ensures your accessibility to me irrespective of where I may be. You should always feel free to send me an e-mail, no matter if it is a question, comment, concern, something interesting you came across (related to class or otherwise), etc. Basically – when in doubt, decide in favor of clicking the “send” button.

**HONOR CODE**

*To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the GMU community and with the desire for greater academic and personal achievement, we, the student members of the University Community have set forth this honor code:

Student members of the George Mason University community pledge not to cheat, plagiarize, steal, or lie in matters related to academic work. The Honor Code of George Mason University deals specifically with cheating and attempted cheating, plagiarism, lying, and stealing.

A. **Cheating encompasses the following:**
   1. The willful giving or receiving of an unauthorized, unfair, dishonest, or unscrupulous advantage in academic work over other students.
   2. The above may be accomplished by any means whatsoever, including but not limited to the following: fraud; duress; deception; theft; trick; talking; signs; gestures; copying from another student; and the unauthorized use of study aids, memoranda, books, data, or other information.
   3. Attempted cheating.

B. **Plagiarism encompasses the following:**
   1. Presenting as one's own the words, the work, or the opinions of someone else without proper acknowledgment.
   2. Borrowing the sequence of ideas, the arrangement of material, or the pattern of thought of someone else without proper acknowledgment.

C. **Lying encompasses the following:**
   The willful and knowledgeable telling of an untruth, as well as any form of deceit, attempted deceit, or fraud in an oral or written statement relating to academic work. This includes but is not limited to the following:
   1. Lying to administration and faculty members.
   2. Falsifying any university document by mutilation, addition, or deletion.
   3. Lying to Honor Committee members and counsels during investigation and hearing. This may constitute a second charge, with the committee members who acted judges during that specific hearing acting as accusers.

D. **Stealing encompasses the following:**
   Taking or appropriating without the permission to do so, and with the to keep or to make use of wrongfully, property belonging to any of the George Mason University community or any property located on the university campus. This includes misuse of university computer resources (see the Responsible Use of Computing Policy section in the "General Policies" chapter). This section is relevant only to academic work and related materials.

*Source: George Mason University Faculty Handbook http://www.gmu.edu/facstaff/handbook/aD.html*
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 26</td>
<td>Syllabus and Introduction to Course SQL Server Components Transactions</td>
</tr>
<tr>
<td>February 2</td>
<td>Indexes Views</td>
</tr>
<tr>
<td>February 9</td>
<td>Constraints Triggers</td>
</tr>
<tr>
<td>February 16</td>
<td>Partitioning Normal Form Modeling</td>
</tr>
<tr>
<td>February 23</td>
<td>Introduction to Data Warehousing</td>
</tr>
<tr>
<td>March 2</td>
<td>Star Schemas Dimensions</td>
</tr>
<tr>
<td>March 16</td>
<td>Snowflake Schemas Facts</td>
</tr>
<tr>
<td>March 23</td>
<td>Hierarchies Aggregations</td>
</tr>
<tr>
<td>March 30</td>
<td>ETL</td>
</tr>
<tr>
<td>April 6</td>
<td>SSAS Cubes</td>
</tr>
<tr>
<td>April 13</td>
<td>Reports and Business Intelligence</td>
</tr>
<tr>
<td>April 20</td>
<td>Data Mining Predictive Analytics</td>
</tr>
<tr>
<td>April 27</td>
<td>Introduction to NoSQL Big Data</td>
</tr>
<tr>
<td>May 4</td>
<td>Review</td>
</tr>
<tr>
<td>May 11</td>
<td>Final Exam (cumulative) 7:30PM – 10:15PM</td>
</tr>
</tbody>
</table>