Carnegie Mellon University, Heinz College
Data Warehousing
Spring 2016 (95-797)

Course
  Course Number/Section  95-797
  Course Title           Data Warehousing
  Course Location       HbH 236
  Dates and Time        Tue/Thu 9:00am-10:20am

Professor Contact Information
  Professor             Beibei Li (beibeili@andrew.cmu.edu)
  Office Location       HbH 3026
  Office Hours          Tue 6:00pm-7:00pm (or by individual appointment)

TA Contact Information
  TA Location           TBD
  Office Hours          Tue/Wed 3-5pm

Textbook

Other Material
Other course material will consist of handouts distributed via Blackboard and in-class exercises. Students should check the class Blackboard site for course material prior to each class.

Prerequisites and Requirements
  Course Prerequisite: Database Management (95-703)
  Software Requirement: Students MUST have a laptop with Microsoft SQL Server 2012 with Analysis Services and Integration Services installed. If you need to install/upgrade the software, Heinz Computing Services has the media and instructions.
  In prior years students have had significant difficulty getting this software installed. Please start the process of installing this software on your laptops during the first week of class.

Course Description
This course will introduce you to the major activities involved in a data warehousing project. The class will begin with an in-depth review of baseline data warehouse principles and concepts. Once the basic principles have been established, the remainder of the class will be built around a group data warehouse project. The project will begin with your group gathering requirements and developing a data warehouse design. Once the design is complete you will build a prototype data warehouse containing the necessary structures within your database and populating them with source data. This will require you to develop
the table definitions, extract/transformation/load (ETL) logic, and example report definitions. I intend this
class to be a hands-on example of a simple data warehouse implementation.

**Breakdown of Grades**

The breakdown of grades is as follows:
- Quizzes (2) 30%
- Project Requirements & Detailed Design 15%
- Project Presentation 15%
- Final Project 30%
- Class Participation 10%

This course cannot be taken Pass/Fail or for Audit credit because of the large percentage of grades
that are dependent on the group project.

*Quizzes*: There will be two in-class quizzes focused on the key principles of data warehousing covered in
class. They will consist of true/false, multiple choice, short answer or similar-type questions. The planned
dates for the quizzes are on the course outline below, but are subject to change based on our progress in
class. (See the schedule below for specific dates.) If they change it will be announced in advance. If you
are unable to attend a class where a quiz is scheduled, you must discuss alternate arrangements with me
IN ADVANCE, or you will receive a zero for the quiz. Alternate arrangements are not guaranteed,
however, and will be made solely at my discretion based on the individual student’s circumstances.

*Project Requirements & Detailed Design Document*: Each group will be required to submit a
Requirements and Detailed Design document for their project. This document is similar to a deliverable
you would provide to a client before you began building a data warehouse system. It outlines the high
level aspects of the data warehouse, the goals it will server, and covers key design considerations.

*Project Presentations*: During the final week, your project group will give a presentation of your project
to the class. Each group member is required to give a portion of the presentation, and should be familiar
with the subject matter. The scheduled date for the in-class presentations is on the class outline schedule
below. More details on the presentation will be provided as the course progresses.

*Course Project*: The focus of this course will be the creation of a data warehouse prototype around a
specified topic. The project will be due during the final week of class. As your project will be the key
focus of this course, all group members should be able to speak intelligently about the focus of your
group’s project and the subject it is covering at any point after the project begins. Failure to do so when
asked may negatively impact your evaluation.

*Peer Evaluation*: Each student will submit a peer evaluation for each of their team members as part of the
course project. This peer evaluation will cover key aspects of the project and will be factored in to the
individual course project score. The peer evaluation form will be provided with the final project
assignment and will cover the team efforts for both presentations and the course project.

*Participation*: Participation points will be earned by being prepared for and actively participating in class
discussions and submitting non-graded assignments and tasks on time.

**A Note on Collaboration and Plagiarism**

The presentations and project for this class will be group assignments so, obviously, you may collaborate
on those within your group. However, please be cautious of plagiarism. While I expect you to do outside
research, please be sure to incorporate your findings into your work in a proper manner. Quote and
reference other’s work correctly when using it in your own. Please refer to the University's plagiarism and cheating policy. Results of instances of plagiarism or cheating range from receiving no credit for an assignment to failure in the course.

Course Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Notes / Work Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/12</td>
<td>Course Introduction</td>
<td>Syllabus</td>
</tr>
<tr>
<td></td>
<td>1/14</td>
<td>- Data Warehousing Basics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Basic Dimensional Modeling</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Exercise – Dimensional Modeling</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1/19</td>
<td>Advanced Dimensional Modeling</td>
<td>Project Team Assignments</td>
</tr>
<tr>
<td></td>
<td>1/21</td>
<td>- Slowly Changing Dimensions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Exercise – Advanced Modeling</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1/26</td>
<td>Quiz 1</td>
<td>Course project and source data approval</td>
</tr>
<tr>
<td></td>
<td>1/28</td>
<td>- Other Design Techniques</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Data Warehouse Requirements</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- ETL Overview</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Microsoft SSIS Lab</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2/2</td>
<td>Advanced SQL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/4</td>
<td>- OLAP vs. OLTP</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Visio Demo</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Exercise - Multidimensional Example</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Microsoft SSAS Lab</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2/9</td>
<td>Mid-Term Group Presentation</td>
<td>Project Detailed Design Due</td>
</tr>
<tr>
<td></td>
<td>2/11</td>
<td>- Advanced Topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Team Meetings &amp; Working Sessions</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2/16</td>
<td>Quiz 2</td>
<td>Guest Lecture</td>
</tr>
<tr>
<td></td>
<td>2/18</td>
<td>- Advanced Topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Team Project Time in Class</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2/23</td>
<td>Final Group Project Presentation</td>
<td>Final Presentation</td>
</tr>
<tr>
<td></td>
<td>2/25</td>
<td>- Final Projects Due</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Exam week</td>
<td>No Final Exam</td>
<td>Final Project</td>
</tr>
</tbody>
</table>

Notes
- The schedule shown above is tentative and subject to change.
- You should bring your laptop to class for when a lab is planned to fully participate in the lab.