# IST 722 Syllabus - Data Warehousing

### Spring 2019

Section: M001 | Class No:41730 | Credits: 3.0   
Thursdays 5:00 PM – 7:45PM Hinds Hall 011

## Course Information

### Instructor

Paul D. Taber  
Adjunct Professor  
206 Hinds Hall

pdtaber@syr.edu

Office Hours: TBA

### Course Description

This course provides concepts, principles, and tools for designing, implementing, and using Data Warehouses. More specifically, we introduce database constructs such as Operational Data Store (ODS), Data Warehouse, and Data Mart, as well as their components. We study the differences between Ralf Kimball’s and Bill Inmon's approaches, roles and responsibilities in the design and implementation of a Data Warehouse, project management guidelines and techniques, requirements gathering, dimensional modeling, Extract Transform and Load (ETL) architecture, specification and data loading, master and reference data management, integration approaches (ETL, EII, EAI), analytical reporting concepts, data governance and recent trends in the data warehouse domain.

The course will leverage a portfolio of SQL Server tools that include SQL Server DBMS, SQL Server Integration Services (SSIS), SQL Server Reporting Services (SSRS) and SQL Server Analysis Service (SSAS) to provide hands-on experience in implementing a reporting solution through assignments, lab exercises and projects.

### Learning Outcomes

Taking this course will provide the following learning outcomes:

#### 1. Technical Knowledge

You will gain technical knowledge and comprehension about data warehouses. You will develop the ability to apply these technologies to solve information problems at the individual and organizational levels. After completing this course, you will be able to:

* Describe various database constructs - ODS, Data Warehouse, Data Mart
* Describe the components of a data warehouse
* Differentiate between Ralf Kimball’s and Bill Inmon's approaches
* Describe various integration approaches - ETL, EII, EAI
* Describe a Master Data Management (MDM) solution
* Create database objects using popular database management system products
* Design and implement data warehouse and business intelligence components

#### 2. Management of Solution Development

You will gain knowledge and comprehension of the disciplines used in the development of data warehouse solutions. You will develop the ability to apply these disciplines in developing solutions for certain organizational and business problems. After completing this course, you will be able to:

* Define the roles and responsibilities in the design and development of data warehouses
* Differentiate various requirements gathering and dimensional modeling techniques
* Define project management guidelines

#### 3. Management of Information Technology

You will be able to integrate technical and solution development concepts with the principles of data governance, strategic alignment and information analysis. You will be able to apply these concepts in the analysis of complex management case studies and problems. After completing this course, you will be able to:

* Describe the data governance concepts
* List some of the recent trends in Data Warehouse

### Office Hours

You are welcome to stop by during office hours without an appointment.

Should you need to make an appointment to see me outside of office hours, you must request via email and include the following:

* State your reason for the appointment.
* Include 3 times in which you can meet.
* Make sure 2 times are on different days.

I will respond to you within 24 hours.

### Textbooks

There are two required textbooks for this course:

* W.H. Inmon, Claudia Imhoff, Ryan Sousa. **Corporate Information Factory.** Wiley.   
  ISBN: 0-471-39961-2
* Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker. **The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems.** Wiley. ISBN: 978-0-470-14977-5

You can probably get by on earlier editions of the textbooks, but there might be a mismatch in the textbook chapters from the reading list. Consider yourself warned.

## Academic Expectations

### Expectations at a glance

* You are expected to submit 5 homework.
* You will take 2 exams.
* You will participate in 10 classes by doing pre-class research.
* Attendance is mandatory.
* You will submit one group project with poster presentation, which will be vetted at mid-term.
* All graded work should be individual effort (except the group project).
* You agree to our class honor code.
* Your enrollment in the course implies your agreement to the terms of the syllabus.

### Grades

This table outlines each method by which you will be evaluated in this course. The due dates, where determined, are posted on this syllabus and to Blackboard.

| **Type of Activity** | **Quantity** | **Notes** | **Points Each** | **Points Total** |
| --- | --- | --- | --- | --- |
| Homework | 5 | All 5 are required.  Due dates on Course Calendar and in Blackboard. | 2 | 10 |
| Participation Questions | 10 | Prepare outside class, assessed In Class.  You must be in attendance and come prepared. | 2 | 20 |
| Exams | 2 | In Class. No make ups. Dates Posted.  Exam will take up 30 minutes of class time. | 20 | 40 |
| Midterm Group Project | 1 | Presentation of your selection and design of 4 business processes. Details outlined below. | 10 | 10 |
| Final Group Project | 1 | Final Design, Implementation, Poster, Presentation and working demo of 2 of your 4 business processes. | 20 | 20 |
| **Total Points** | | | | **100** |

### Grading Scale

I use the following grading scale for translating your points earned in the class into a letter grade for submission to the registrar.

| Total Points Earned | Registrar Grade |
| --- | --- |
| 95 - 100 | A |
| 90 - 94 | A - |
| 85 - 89 | B + |
| 80 - 84 | B |
| 75 - 79 | B - |
| 70 - 74 | C + |
| 65 - 69 | C |
| 60 - 64 | C - |
| 50 - 59 | D |
| 0 - 49 | F |

**NOTE:** The grade of A represents true mastery of the material and is reserved for those students who demonstrate this throughout all aspects of their coursework (Exams, Participation, Homework, and Project).

### Homework

Homework are out of class assignments which demonstrate that you can apply the concepts we learn in class to a real-world scenario.

* Homework is based on a practice activity in Data Warehousing. There are 5 homework.
* These are individual assignments and are subject to our class honor code. Work alone.
* The due dates set are posted on the course calendar and in blackboard.
* Homework are due before the start of class on the day they are due.

### Participation Questions

Participation is based on the assigned readings. For each assigned reading there are a handful of ***class participation*** ***questions*** you must research then answer individually. You bring your printed answers to class, turn them in, and we use them to initiate a discussion at the start of that class topic.

* Each week you will be assigned class preparation questions based on the topic and readings.
* There are 10 classes in which your participation will be graded.
* You must be in attendance to receive full credit. You cannot participate without being there!
* Your class preparation questions will be collected at the beginning of class.
* We will start class with a discussion based on the class preparation questions.

### Exams

Exams serve as an instrument to gauge whether or not students are keeping up with the required readings and grasping important terminology, concepts, and their application:

* There are two exams. They are posted in Blackboard and on the class schedule.
* Exams are timed you will have 30 minutes to complete the exam.
* Exams are issued in the beginning of class. Show up on time or you risk having less time.
* Exams are closed-book.
* Exams consist of multiple choice and short answer questions.
* Exam individual effort. Anyone caught cheating on an exam will be have to appear before the office of academic integrity.
* Please be quiet and courteous to others during the exam period.

### Due Dates & Late Work

Students expect their professors to return graded work back within a reasonable time frame. Late submission of student work makes it difficult for me to do that, and therefore I have strict policies regarding the due dates and the submission of work:

* Homework are due before the start of class on the date provided.
* Submissions (in part or whole) arriving after that time are marked late, so make sure to get it in with plenty of time to spare.
* You have until 11:59 PM the next day to submit late work. After that it will not be accepted.
* Late work is penalized 50% of the value after it is graded
* Accidentally turning in the wrong assignment or incomplete work is not an excuse. Whatever you submit will be graded as-is, and any work submitted after the deadline constitutes a late assignment, so double-check each submission. Get it right the first time, please.

### Group Project

The group project is your chance to apply what you have learned in the course towards completing a design and implementation of dimensional models based on a case study. Not only will you be expected to design and implement a solution, but also analyze the data for metrics, key performance indicators, and data insights, then present those findings in a poster presentation. The deliverable will consist of the following artifacts:

* **Design Document.** Your team will produce a comprehensive document outlining the requirements and data warehouse design for the case study.
* **Implementation.** Your team will produce an implementation of the design, including ROLAP, MOLAP and ETL.
* **Poster.** Your team should create a 2x3 poster which tells the story of your project or a Powerpoint presentation.
* **Presentation.** Your team give an oral presentation of your findings which coincide with your poster.
* **Demonstration.** Your team will provide a working demonstration of your dimensional models, to be explored / browsed by visitors to your poster session.

The group project is graded in two parts. A midterm group project and final group project. The same project team will work on both projects.

#### Midterm Group Project

For the midterm group project you are required to produce a portion of your design document and give a presentation on your findings to date. Specific criteria:

* Decide on 4 business processes your group will model.
* Implement the design document up to and including the High-Level Design. (see Final project for
* Give a brief 10 minute presentation, in class, regarding your rationalization of chosen business processes and their design.
* Emphasize how this has been a group effort to date.

Midterm rubric

| **Performance** | **Criteria** |
| --- | --- |
| Exemplary  10 | * Complete design document up to / including High-Level Design. * Design based on 4 business processes. * Organized presentation hits on all key points. * Presentation kept to 5-10 minutes. * No glaring issues with dimensional model designs. |
| Outstanding 9 | * Only one of the exemplary criteria not met. * This includes one issue with dimensional model design. |
| Satisfactory 8 | * 1-2 exemplary criteria not met. * More than 1 issue with dimensional model design. |
| Unsatisfactory  7 or lower | * 3 or more project criteria are not met. * Poor or missing designs. |

#### Final Group Project

For the final group project, you’ll continue on with your groups’ work at midterm. You are expected to fix issues with your midterm based on feedback, and then choose 2 of your 4 business processes to implement.

The project will be graded as follows:

| **Artifact / Component Criteria** |
| --- |
| **Design Document: Single Electronic Document, with:** |
| Document Organization:   * Table of contents * Cover page, page numbers * Well organized, professional |
| Overview Section:   * Project Scope / Mission * Team Members and Roles * Key Stakeholders |
| Analysis Section:   * Business processes * Bus matrix * Bubble Chart * Attribute List * Issues list |
| High-Level Design Section:   * Detailed bus matrix * Detailed dimensional design worksheet |
| Detailed-Design Section:   * Dimensional Hierarchies * Snowflake model diagrams * ETL Specifications (high-level source to target map * Detailed ETL flow for each source to target |
| Team contribution report |
| **Implementation: Build physical structures in SQL Server** |
| ROLAP Schema in SQL Server:   * Fact Tables * Dimension Tables * Star Schema Diagrams * Consistent with Design * Proper PK / FK use |
| SSIS ETL Code / Packages:   * Populates ROLAP Schema From Sources * Initial Load * Subsequent Loads |
| Analysis Services MOLAP Database:   * Cubes to match Star Schemas / Business Processes * Dimensional Hierarchies configured * Facts / KPI's as appropriate |
| BI Front End (Excel or Other):   * Pivot Tables linked to SSAS database cubes * Dashboards (In Power View or Other) * Well organized and meaningful |
| **Live Demonstration and poster** |
| Presentation   * Explain project and process * Discussion successes and failures * Explain how it was a group effort. * Highlight key insights and factoids (about the data) |
| Poster   * Define Problem * Outline Your Process / Approach * Lessons Learned (about Process) * Key insights and factoids (about the data) |
| Working Interactive Demo   * Demonstrate exploring your visualization * Meaningful, relevant data set * Visitors should be able to explore it without intervention. |

Rubric for Final Project

| **Performance** | **Criteria** |
| --- | --- |
| Exemplary  20 | * Project, report, implementation, poster, presentation and demo are of the highest quality. * Dimensional models and visualizations are useful / meaningful. * Implemented 2 dimensional models. |
| Outstanding 19 | * All project criteria are completed correctly. * Dimensional models and visualizations are useful / meaningful. * Implemented 2 dimensional models. * All Deadlines met. |
| Satisfactory 18-16 | * 1 - 2 project criteria are not completed correctly. * Project is of average quality. |
| Unsatisfactory  15 or lower | * 3 or more project criteria are not completed correctly. * Poor or missing criteria. |

### Group Work

The only group work permitted in this course is the Group Project. All other gradable work (Homework, Exams, Class Preparation Questions) in this course are individual effort. It is assumed that each of you know the difference between simply discussing work with your classmates and working collaboratively. When in doubt please review the class honor code. “I didn’t know” is not an excuse.

### Course Calendar

Here's our schedule of topics to be covered each day throughout the semester. Also included in the calendar are the problem set due dates.

| **Date** | **Unit** | **Topic** | **Reading** | **Due** |
| --- | --- | --- | --- | --- |
| 01/17 | 01 | Introductions / Welcome, Syllabus,  SQL Refresher Course | Materials Online |  |
| 01/24 | 02 | Introduction to Data Warehousing | Inmon Ch. 1-2  Kimball Ch. 1 |  |
| 01/31 | 03 | Data Warehouse Constructs and Components | Inmon Ch. 3-8 | H1: Homework 1 |
| 02/07 | 04 | Project Management & Requirements Gathering | Kimball Ch. 2-3 |  |
| 02/14 | 05 | Introduction to Dimensional Modeling | Kimball Ch. 6-7 |  |
| 02/21 | 06 | Dimensional Modeling Design in SQL | Kimball Ch. 8 | H2: Homework 2 |
| 02/28 | 07 | Technical Architecture | Kimball Ch. 4  Materials Online |  |
| 03/07 |  | Midterm Group Project Design Presentations, Exam 1 |  | MGP: Midterm Group Project, Exam 1 |
|  |  |  |  |  |
| 03/21 | 08 | Introduction to ETL Design | Kimball Ch. 9 | H3: Homework 3 |
| 03/28 | 09 | ETL Development with SSIS | Kimball Ch. 10 |  |
| 04/04 | 10 | Master Data Management & Data Governance / Introduction to Business Intelligence | Kimball Ch. 11  Materials Online | H4: Homework 4 |
| 04/11 | 11 | Business intelligence design and development | Kimball Ch. 12 |  |
| 04/18 | 12 | Big Data In the Data Warehouse | Materials Online | H5: Homework 5 |
| 04/25 |  | Exam 2 - Poster Presentation |  | Exam 2 |
|  |  | Poster Presentations (4/25 – iSchool joint poster session) |  | FGP: Final Group Project Due (4/25) |

## Class Honor Code

The class honor code represents our commitment to Academic Integrity in this course. I drafted the class honor code to avoid academic negligence - situations where students are unaware that their actions are actually a form of cheating. Our honor code remedies this problem by clearly stating the expectations of Academic Integrity for this course. It states:

1. **All work is my own.** Answers on all student work, assignments (problem sets, projects, papers, homework, etc...) and assessments (quizzes, exams, tests, etc...) are my own individual work (except where collaboration is explicitly permitted).
2. **I will not share answers.** I will not make answers (either my own or the professor's) to work, assignments (problem sets, projects, papers, homework, etc...) and assessments (quizzes, exams, tests, etc...) available to anyone else in or out of class.
3. **I will not misrepresent my ability.** I will not engage in any activity which misrepresents or falsifies my knowledge of the subject matter and therefore improves my grade dishonestly.
4. **I will give credit.** I will always pay attribution to my sources, and not misrepresent the works of others as my own.
5. **I accept the honor code and its consequences.** I understand and accept that that all work I submit is subject to the honor code, and if I violate this honor code I will be brought up on charges of academic fraud.

### Self-Reporting

If you feel may have inadvertently violated our class honor code, please let me know immediately. I prefer this be done prior to your submission of the work in question. Depending on the degree of the violation and the nature of the work, we will discuss the following:

* ways we can remediate your grade accordingly to reflect your actual accomplishments, and
* Means to prevent such activities from occurring in the future.

Under most circumstances, self-reported incidents are kept between teacher and student. They are not filed with the Academic Integrity Office.

### Violations and Sanctions

Violations of the class honor code which are not self-reported will be treated as an intentional act of academic dishonesty. Under these conditions the following will occur, without exception:

* The students in question will be notified by me of the incident.
* A new case will be filed with the University's Office of Academic Integrity, with recommended sanctions appropriate to the violation.
  + On assignments where sources and citations are required, in-appropriate citations or lack of citations will result in a grade of 0 for the assignment.
  + Copying homework, participation questions, or re-using a pre-existing project will result in an F in the course.
  + Cheating on an exam with result in an F in the course.

## University Policies

### Academic Integrity

Syracuse University’s Academic Integrity Policy holds students accountable for the integrity of the work they submit. Students should be familiar with the policy and know that it is their responsibility to learn about course-specific expectations, as well as about university policy. The university policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The policy also prohibits students from submitting the same written work in more than one class without receiving written authorization in advance from both instructors. The presumptive penalty for a first offense by an undergraduate student is course failure, accompanied by a transcript notation indicating that the failure resulted from a violation of Academic Integrity Policy. The standard sanction for a first offense by a graduate student is suspension or expulsion. For more information and the complete policy, see <http://academicintegrity.syr.edu>

### Disability-Related Accommodations

If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), <http://disabilityservices.syr.edu>, located at 804 University Avenue, room 309, or call 315-443-4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue students with documented disabilities “Accommodation Authorization Letters,” as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible.

### Religious Observances Policy

SU religious observances policy, found at <http://supolicies.syr.edu/emp_ben/religious_observance.htm> , recognizes the diversity of faiths represented among the campus community and protects the rights of students, faculty, and staff to observe religious holidays according to their tradition. under the policy, students are provided an opportunity to make up any examination, study, or work requirements that may be missed due to are religious observance provided they notify their instructors before the end of the second week of classes. for fall and spring semesters, an online notification process is available through myslice />StudentServices/Enrollment/MyReligiousObservances from the first day of class until the end of the second week of class.

### Educational Use of Student Work

I intend to use academic work that you complete this semester for educational purposes in this course during this semester as well as in future semesters. Your registration and continued enrollment constitute your permission.