# IST722: Syracuse University Graduate Course Syllabus

## Course Title

IST722: Data Warehouse

## Course Information

### Meeting Times

| Section | Day | Time | Location |
| --- | --- | --- | --- |
| TBD | TBD | TBD | TBD |

### Instructor

| Key | Value |
| --- | --- |
| **Name** | Your Name |
| **Title** | Your Title |
| **Office** | Your Office |
| **Email** | Your Email |
| **Office Hours** | Your Office Hours |

### Course Description

Introduction to concepts of business intelligence (BI) and the practice/techniques in building a BI solution. Focuses are on how to use data warehouses as a BI solution to make better organizational decisions.

### Additional Course Description

This course provides concepts, principles, and tools for designing, implementing, and using Data Warehouses. The course explains the business rationale of a data warehouse, its role in analytics and the common data architectures found in data warehouses. Throughout the course we will learn several technical and managerial concepts and principles such as dimensional data stores, one big table, master data management, data quality, operational data stores, data governance, staging data, project management, extract-transform-load, business intelligence, and semantic metadata.

Students will acquire hands on experience planning, designing, and building a data warehouse along with the business intelligence on top of it using cloud-based solutions. IST659, specifically a solid understanding of SQL DDL/DML, indexes, complex query writing, and query plans is required.

### Prerequisite / Co-requisite:

IST659 Data Administration Concepts and Database Management

### Audience:

This course is appropriate for students seeking careers or with an interest in business analytics, data engineering, information systems/IT or data science.

### Credits:

3

### 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

1. 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

1. 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

### Course Fees and/or Costs

None

### Required Textbooks and Supplies

All required readings are offered online or through the SU library there is no need to purchase these.

* Kimball, R., & Ross, M. (2013). The data warehouse toolkit : the definitive guide to dimensional modeling (3rd ed.). John Wiley & Sons, Inc.
* Rainardi, V. (2008). Building a data warehouse with examples in SQL Server. Apress.

## Academic Expectations

### Attendance and Participation

You are expected to attend and participate in every class session. If you must miss class, you are responsible for making up the work and catching up on what you missed. Please do not expect a pre-recorded lecture or instructor support if you miss class. A suggestion is to partner with a classmate to share notes from class.

In each class there will be frequent opportunities to participate. These include sharing thoughts, opinions and ideas, answering challenge questions, and reporting our team progress. Everyone is expected to participate. Please be respectful of your classmates’ time and perspectives by affording them the opportunity to participate.

### Assignments

Assignments are your opportunity to put into practice the concepts learned in the course. For this reason, assignments are a formative assessment. You will need to use the skills you’ve acquired from the assignments to complete the team project successfully, and this is ultimately when your skills will be evaluated.

For each assignment you must turn in a completed assignment template. The template contains your reflection on the assignment including what you learned and where you struggled into addition to documented evidence of your attempt of the assignments’ completion. Together, these represent a satisfactory submission. In addition, each assignment must be on time.

There are 7 assignments total, and your overall assignment grade of 10 points is based on the following table:

| Number of Assignments Satisfactory and on Time | Points Awarded |
| --- | --- |
| 7 | 10 |
| 6 | 9 |
| 5 | 7 |
| 4 | 5 |
| 3 | 3 |
| 2 | 2 |
| 1 | 1 |
| 0 | 0 |

### Exams

There are 3 exams in the course each exam is worth 30 points each. Your highest 2 exam scores will factor in to your final grade of 60 points. This means you can miss an exam without a penalty to your final grade in the course.

* Exams will be issued on the exam date during class time.
* There are no make up exam dates.
* Exams are closed book.
* Exam mode (on paper or on-line) will be determined by your instructor.
* Exam formats (multiple choice, essay, fill-in-the-blank) will be determined by your instructor.
* Units covered by each exam are posted on the course schedule. This is the emphasis of the exam, however the exams should be considered naturally cumulative.

### Team Project

You are expected to complete a **Team Project**, which demonstrates your team’s ability to design and implement a data warehouse / business intelligence solution as part of a functional team, drawing upon what you learned in the course. There will be time allocated each week so that your group can plan the next steps in your group project.

Specific project rules:

* Your instructor will assign teams.
* Teams will consist of 3 students. Teams of 2 students only happen when the class size is not divisible by 3.
* Your instructor will assign the case study for your team to to work on.
* Minimum work requirement: You team is expected to implement one star schema per student in your group.

#### Project deliverables

Your project deliverables will be:

* A document (or website) which contains the project details and documentation.
* Code: SQL used to build the DW, perform the ETL and the BI you built.
* Video: A video presentation, demo and reflection no more than 10 minutes in length.

Your team should decide how best to accomplish this. Git repository, cloud file share, etc…

#### Outline of Project Artifacts

1. Project Document:
   1. Assemble a Project Charter and team members
   2. Devise a project plan
   3. Outline functional requirements based on your activity of profiling data in the case study.
   4. Overview which business processes you will model from those functional requirements and explain their business value.
   5. Assign primary roles to your team members
2. High-level dimensional modeling
   1. Bus matrix
   2. Outline any issues, questions, or doubts you have
3. Detail-level dimensional modeling
   1. Fully completed dimensions and facts table designs in the worksheet
   2. Identified sources for your source to target map (high level just shows source to target Db’s and schemas)
   3. Establish Naming conventions and project standards
4. Data Warehouse Implementation
   1. Following your conventions / standards
   2. Adopts techniques learned in the course, such as use of staging and enterprise bus, OBT.
5. Initial Loads Data Pipeline (ETL/ELT) done:
   1. Source to target maps (can be generated from ETL tool).
   2. Source code for Data pipeline.
   3. Documentation for Data Pipelines (can be generated)
   4. Other ETL Documentation as needed: Such as data quality, master data or survivorship rules you may have used.
6. Business Intelligence:
   1. BI Documentation: Briefly explain the goal of your analytics and what type of BI it is.
   2. BI Dashboard or Application in Power BI or Tableau.
   3. Other BI Documentation as needed: Use of KPI’s or other semantic metadata you may have used. Explanation of applications of drill-across or drill-through for example.
7. Presentation and Demo
   1. Presentation and demo cannot exceed 10 minutes total.
   2. Brief overview of your project plan and goals (keep it at an executive level)
   3. A quick demo of your BI Application, no more than 5 minutes, emphasizing how it achieved b.
   4. A brief reflection on the experience, what you’d do if you had more time, etc.
   5. Should be a recorded video with shareable link.

#### Project Milestones

Milestones are your opportunity to show the class your progress to date and to solicit feedback from your instructor. The due dates of the milestones are listed on the course schedule. For each milestone you will provide a presentation to the class updating your professor and your classmates as to your progress. What should be covered in your presentation?

* Brief overview / reminder of your project and modeled business processes.
* Progress so far (varies based on the milestone, see below)
* Challenges and issues you have.
* What are your next steps?

For every milestone, your group must turn in your presentation slide deck before class. This makes presentation more organized on milestone days as I can pre-download the presentations before class.

* Milestone 01 – Bus Matrix.
* Milestone 02 – Detail Dimensional Model and implementation for at least 1 business process.
* Milestone 03 – BI demonstration of business processes.

Notes:

* You should be prepared to answer questions from the audience. Keep your presentation organized and short. If you go over time you will be cut off and miss out on instructor feedback!
* Please note milestones are not graded. You will only get verbal feedback. I suggest someone in your team take notes during the presentation.
* While your milestones are not graded, your effort in completing milestones does factor into your project grade.
* You do not need to turn in any project deliverables until the end, but its recommended to work on your project document as you complete the milestones.

#### Grading Rubric for Project

Here are some guidelines for completing the project. Answering “yes” to all of these implies that your project has the potential to earn the highest marks possible.

* Did you meet every milestone? Were you prepared to present? Was the feedback provided taken into consideration?
* Did you model the proper number of business processes / star schemas based on the number of members in your group?
* Can the value of the business processed you selected be justified?
* Did you implement your business processes in the data warehouse as a star schema for each? with an initial load of data? Following enterprise bus technical architecture?
* Did you implement meaningful, actionable business intelligence for those business processes?
* Does the data warehouse your team implemented clearly exhibit the 4 characteristics of a data warehouse?
* Do your deliverables code, video and document check off all the requirements in the outline of project artifacts? Is it easy for a grader to find them and determine the requirements were me?
* Does your demo confirm you satisfied the functional requirements of the business processes you implemented?

### Grading Summary:

| **Type of Activity** | **Quantity** | **Points Each** | **Total Points** |
| --- | --- | --- | --- |
| 1. Assignments | 7 | N/A | 10 |
| 2. Exams | 3 (best 2 count) | 30 | 60 |
| 3. Team Project | 1 | 30 | 30 |
| **Total Points** |  |  |  |

### Grading Scale:

| **Student Achievement** | **Percentage** | **Grade Points** | **Registrar Grade** |
| --- | --- | --- | --- |
| Mastery | 95 - 100 | 4.0 | A |
|  | 90 - 94 | 3.667 | A - |
| Satisfactory | 85 - 89 | 3.333 | B + |
|  | 80 - 84 | 3.0 | B |
| Low Passing | 75 - 79 | 2.667 | B - |
|  | 70 - 74 | 2.333 | C + |
| Unsatisfactory | 65- 69 | 2.0 | C |
|  | 60 - 64 | 1.667 | C - |
|  | 0 - 59 | 0 | F |

## Other Course Policies

* All work is due on the dates provided. No late work is accepted, unless explicitly noted. The reasoning is feedback is time-sensitive.
* No make up exams are permitted. Lowest exam score is dropped.
* Final grades will not be rounded up. 94/100 is an A-, please don’t ask.

### Academic Integrity

* We take academic integrity seriously, and so should you.
* It is our expectation that your work will be 100% representative of your academic abilities.
* If you work with someone else on an Assignment, please disclose that in your reflection.
* Examinations are summative assessments. There should be no outside help or assistance.
* When in doubt as to whether you can use a resource outside those provided in the course, **ask your instructor**.
* All violations of academic integrity will be reported to the AIO office. Proposed grade sanction is F in the course.

## Course Calendar

### Course Schedule

| **Class** | **Date** | **Topic for That Week** (details below) |
| --- | --- | --- |
| 01 | 8/28 | [A - Introduction to Data Warehousing](#class-01) |
|  | 9/4 | NO CLASSES - LABOR DAY |
| 02 | 9/11 | [B - Data Warehouse Tools and Architectures](#class-02) |
| 03 | 9/18 | [C - Planning The Data Warehouse](#class-03) |
| 04 | 9/25 | [D - Dimensional Modeling](#class-04) |
| 05 | 10/2 | [Exam 1 (A,B,C), Project Milestone 1 Presentations](#class-05) |
|  | 10/9 | NO CLASSES - FALL BREAK |
| 06 | 10/16 | [E - Data Warehouse Development](#class-06) |
| 07 | 10/23 | [F - Building Data Pipelines](#class-07) |
| 08 | 10/30 | [G - Business Intelligence](#class-08) |
| 09 | 11/6 | [Exam 2 (D,E,F), Project Milestone 2 Presentations](#class-09) |
| 10 | 11/13 | [H - Incremental Data Pipelines](#class-10) |
|  | 11/20 | NO CLASSES - THANKSGIVING BREAK |
| 11 | 11/27 | [I - Managing the data warehouse](#class-11) |
| 12 | 12/4 | [Exam 3 (G,H,I), J - Trends in Data Warehousing](#class-12) |
| 13 | 12/11 | [Final Project Presentations (Milestone 3)](#class-13) |
| 13 | 12/15 | Final Project Is Due |

### Class 01

### Topic: A - Introduction to Data Warehousing

Readings / Videos:

* Course Syllabus
* Rainardi, V. (2008). Building a data warehouse with examples in SQL Server. Apress, Ch 1.
* What Is a Data Warehouse? (365 Data Science). Www.youtube.com. <https://www.youtube.com/watch?v=AHR_7jFCMeY>

In Class Agenda:

* Go over the course syllabus
* Lecture: What is data warehousing? The 10,000 foot view.
  + What is data warehousing? The 10,000 foot view.
  + Components of the DW
  + Data warehouse vs data lakes vs data lake house
  + Data warehouse vs business intelligence
  + Process-oriented vs reference-oriented data
* Demo: Accessing the external world.
* Discussion: Assignment A

### Class 02

### Topic: B - Data Warehouse Tools and Architectures

Readings / Videos:

* Kimball, R., & Ross, M. (2013). The data warehouse toolkit : the definitive guide to dimensional modeling (3rd ed.). John Wiley & Sons, Inc., Ch 1.
* Let’s Compare the Kimball and Inmon Data Warehouse Architectures. (nullQueries). Www.youtube.com. <https://www.youtube.com/watch?v=Tff34jj_V-0>
* Data Vault vs Traditional Data Warehouse Architectures. (nullQueries). Www.youtube.com. <https://www.youtube.com/watch?v=D914nNWGP6E>

‌ Due before class:

* Assignment A

In Class Agenda:

* Lecture: Data Warehouse Tools and Architectures
  + Components of a technical architecture
  + Historical underpinnings of each architecture
  + Tools of the modern data warehouse
  + Kimball, Inmon, Linstedt
* Discussion: Assignment A
* Project Time:
  + Project group formation and ice breakers to get to know your teammates.
  + Deliverable: Form your Blackboard group, name your team, decide when and how you will meet outside of class.

### Class 03

### Topic: C - Planning The Data Warehouse

Readings / Videos:

* Rainardi (2008) Ch 3,4
* Kimball (2013) Ch 17

Due before class:

* Half of assignment B

In Class Agenda:

* Lecture: Planning the Data Warehouse
  + User centric vs Data Centric Approaches
  + Functional and non-functional requirements
  + Understanding business processes and reference data
  + Identifying business processes
* Discussion: Assignment B Progress
* Project Time:
  + System Selection Show
  + Deliverable: Profile 3 Business Processes

### Class 04

### Topic: D - Dimensional Modeling

Readings / Videos:

* Kimball (2013) Ch 2
* Is Kimball’s Dimensional Modelling dead in 2022? Is OBT (“one big table”) the way to go? (2022, May 3). <https://www.reddit.com/r/dataengineering/comments/uhohlv/is_kimballs_dimensional_modelling_dead_in_2022_is/>

Due before class:

* Assignment B, Assignment C

In Class Agenda:

* Lecture: Dimensional Modeling
  + Dimensional Model Designs Fact & Dimension Tables
  + Slowly / Rapidly Changing Dimensions
  + Advanced Dimensional Modeling Concepts
* Discussion: Assignments B and C
* Project Time:
  + Bus Matrix, Priority Grid, Milestone 1
  + Deliverable: Milestone 1 readiness

### Class 05

### Topic: Exam 1, Project Milestone 1 Presentations

Readings / Videos:

* None

Due before class:

* Project Milestone 1 Presentations

In Class Agenda:

* Exam 1:
  + Covers Topics A,B,C
  + Issued in class, on paper
* Project:
  + Team selection for presentation order.
  + Project Milestone 1 Presentations & instructor feedback.
  + Strict adherence to the time limit to everyone gets an opportunity.

### Class 06

### Topic: E - Data Warehouse Development

Readings / Videos:

* What Is The Modern Data Stack - Intro To Data Infrastructure Part 1. (Seattle Data Guy). Www.youtube.com. <https://www.youtube.com/watch?v=-ClWgwC0Sbw>
* Tools of The Modern Data Stack. (nullQueries). Www.youtube.com. <https://www.youtube.com/watch?v=KKLw9MdcLDw>

Due before class:

* Assignment D

In Class Agenda:

* Lecture: Data Warehouse Development
  + What is data warehouse development
  + Cloud versus On-Premises data warehouses.
  + Understanding the DW development process
  + Examples of the DW development process
* Discussion: Assignment D
* Project Time:
  + Work on your detailed dimensional modeling.
  + Deliverable: High-level source to target map

### Class 07

### Topic: F - Building Data Pipelines

Readings / Videos:

* ETL vs ELT | Modern Data Architectures. (Kahan Data Solutions). Www.youtube.com. <https://www.youtube.com/watch?v=_Nk0v9qUWk4>
* What Is DBT and Why Is It So Popular - Intro To Data Infrastructure Part 3. (Seattle Data Guy). Www.youtube.com. <https://www.youtube.com/watch?v=8FZZivIfJVo>
* Rainardi (2008) Ch 7,8

Due before class:

* Assignment E

In Class Agenda:

* Lecture: Building Data Pipelines
  + Generic ETL Approaches and their rationale
  + Common ETL Patterns
  + DBT Approach to Data Pipelines
* Discussion: Assignment E
* Project Time:
  + DBT project setup for your project group.
  + Start work on ELT pipelines
  + Deliverable: DBT project group setup

### Class 08

### Topic: G - Business Intelligence

Readings / Videos:

* Kimball, R. (2003, March 20). The Soul of the Data Warehouse, Part 1: Drilling Down. Kimball Group. <https://www.kimballgroup.com/2003/03/the-soul-of-the-data-warehouse-part-one-drilling-down/>
* Kimball, R. (2003, April 5). The Soul of the Data Warehouse, Part 2: Drilling Across. Kimball Group. <https://www.kimballgroup.com/2003/04/the-soul-of-the-data-warehouse-part-two-drilling-across/>
* Enochson, H. (2019). 27 Examples of Key Performance Indicators. OnStrategy. [https://onstrategyhq.com/resources/27-examples-of-key-performance-indicators/(https://onstrategyhq.com/resources/27-examples-of-key-performance-indicators/)]
* Power BI Tutorial in 10 min. (Kevin Stratvert). Www.youtube.com. <https://youtu.be/NNSHu0rkew8>

Due before class:

* Assignment F

In Class Agenda:

* Lecture: Business Intelligence
  + Business Intelligence Explained
  + How we categorize and measure BI
  + Common BI terminology
  + Heuristics for building effective BI
* Discussion: Assignment F
* Project Time:
  + Prepare for Milestone 2
  + Deliverable: Team readiness for Milestone 2

### Class 09

### Topic: Exam 2, Project Milestone 2 Presentations

Readings / Videos:

* None

Due before class:

* Project Milestone 2 Presentations

In Class Agenda:

* Exam 2:
  + Covers Topics D,E,F
  + Issued in class, on paper
* Project:
  + Team selection for presentation order.
  + Project Milestone 2 Presentations & instructor feedback.
  + Strict adherence to the time limit to everyone gets an opportunity.

### Class 10

### Topic: H - Incremental Data Pipelines

Readings / Videos:

* Functional Data Engineering - A Set of Best Practices | Lyft. (n.d.). Www.youtube.com. <https://www.youtube.com/watch?v=4Spo2QRTz1k&t=952s>
* Beauchemin, M. (2018, January 8). Functional Data Engineering — a modern paradigm for batch data processing. Medium. <https://maximebeauchemin.medium.com/functional-data-engineering-a-modern-paradigm-for-batch-data-processing-2327ec32c42a>

Due before class:

* Assignment G

In Class Agenda:

* Lecture: Incremental Data Pipelines
  + Strategies for ETL Incremental Data loads
  + Snap Shotting
  + Type 2 SCD
  + DBT and incremental pipelines
* Discussion: Assignment G
* Project Time:
  + Work on DBT ELT
  + Work on BI Plans
  + Deliverable: Progress reports

### Class 11

### Topic: I - Managing the Data Warehouse

Readings / Videos:

* IBM Technology. (2022). What is Master Data Management. In YouTube. <https://www.youtube.com/watch?v=l83bkKJh1wM>
* IBM Technology. (2022). Data Governance Explained in 5 Minutes. In YouTube. <https://www.youtube.com/watch?v=uPsUjKLHLAg>
* What is Reverse ETL? (Kahan Data Solutions). Www.youtube.com. <https://www.youtube.com/watch?v=DRAGfc5or2Y>
* Frameworks of Data Governance. (nullQueries). Www.youtube.com. <https://www.youtube.com/watch?v=gCwKnaITV8g>
* 4 Native Snowflake Data Quality Checks & Features You Should Know. (2022, April 21). Www.montecarlodata.com. <https://www.montecarlodata.com/blog-snowflake-data-quality-features/>

Due before class:

* None

In Class Agenda:

* Lecture: Managing the Data Warehouse
  + Master Data Management
  + Data Governance
  + Data Quality Approaches
  + The importance of Documentation and Metadata
* Project Time:
  + Work on DBT ELT
  + Work on BI
  + Deliverable: Progress reports

### Class 12

### Topic: J - Trends in Data Warehousing

Readings / Videos:

* What Is a Lakehouse? (2020, January 30). Databricks. <https://www.databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html>
* What Is Apache Iceberg? Features & Benefits. (n.d.). Dremio. Retrieved August 7, 2023, from <https://www.dremio.com/resources/guides/apache-iceberg/>

Due before class:

* None

In Class Agenda:

* Exam 3:
  + Covers Topics G,H,I
  + Issued in class, on paper
* Lecture: Trends in Data Warehousing
  + A selected topic on Data Warehousing Trends: Data Lake Houses
* Project Time:
  + Prepare for final presentations
  + Deliverable: readiness update

### Class 13

### Topic: Exam 3, Final Project Presentations

Readings / Videos:

* None

Due before class:

* Final Project Presentations

In Class Agenda:

* Project:
  + Team selection for presentation order.
  + Final Project Presentations & demos
  + Final instructor feedback for group touch-ups prior to grading.
  + Strict adherence to the time limit to everyone gets an opportunity.