

Syllabus

This is a single, concatenated file, suitable for printing or saving as a PDF for offline viewing. Please note that some animations or images may not work.

Description

This [module](#) is also available as a concatenated page, suitable for printing or saving as a PDF for offline viewing.

MET CS 689

Designing and Implementing a Data Warehouse

This course surveys state-of-the art technologies in DW and Big Data, and provides students with the engineering skills required to evaluate, implement, and scale a modern data warehouse using commercially available and open source software. It describes logical, physical and semantical foundation of modern DW infrastructure. Students will create a cube using OLAP and implement decision support benchmarks on Hadoop/Spark vs Vertica database. Students will do 6 assignments and one final project.

Technical Notes

The table of contents expands and contracts (+/- sign) and may conceal some pages. To avoid missing content pages, you are advised to use the next/previous page icons in the top right corner of the learning modules.

This course requires you to access files such as word documents, PDFs, and/or media files. These files may open in your browser or be downloaded as files, depending on the settings of your browser.

Learning Objectives

By successfully completing this course you will be able to:

- Apply standard SQL Analytical functions for business intelligence reporting
- Implement Extract, Transform, and Load for large volumes of varied data using Python
- Perform dimensional data modeling
- Implement business intelligence reporting on data warehouses
- Perform entity resolution among unstructured datasets
- Tune data warehouses for predefined and ad hoc queries
- Perform elementary storage, retrieval, and analysis with a Big Data tool

Instructor

Mary E. Letourneau

Lecturer

Computer Science Department

Metropolitan College

Boston University

maryleto@bu.edu

Greetings to Students and worldwide members of the Boston University Community!

I am your instructor, Mary E. Letourneau. I have worked in the computer industry for over 30 years, starting with chip design and including consulting, programming, teaching, and for the last 12 years databases. I am currently employed as the Director of Information Systems. I earned my M.S. in Computer Information Systems from BU MET in 2015 and have been facilitating and/or teaching part-time for Boston University almost every semester since.

Materials

Required Book

McKinney, W. (2017). *Python for Data Analysis* (2nd ed.). O'Reilly Media. ISBN 9781491957660.



An e-book is available at Vitalsource.com. An e-book is available through Amazon. An e-book is available for free through the BU library. An e-book is available at the [BU bookstore](#).

Kimball, R. & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley & Sons. ISBN 9781118530801.

An e-book is available at Vitalsource.com. An e-book is available through Amazon. An e-book is available for free through the BU library. An e-book is available through the [BU bookstore](#).

Krishnan, K. (2013). *Data Warehousing in the Age of Big Data*. ELSEVIER. ISBN 9780124058910. An e-book is available at Vitalsource.com. An e-book is available through Amazon.

An e-book is available for free through the BU library. An e-book is available through the [BU bookstore](#).

Boston University Library Information

Boston University has created a set of videos to help orient you to the online resources at your disposal. An introduction to the series is below:

met_ode_library_14_sp1_00_intro video cannot be displayed here

All of the videos in the series are available on the [Online Library Resources](#) page, which is also accessible from the Campus Bookmarks section of your Online Campus Dashboard. Please feel free to make use of them.

As Boston University students, you have full access to the BU Library. From any computer, you can gain access to anything at the library that is electronically formatted. To connect to the library, use the link <http://www.bu.edu/library>.

You may use the library's content whether you are connected through your online course or not, by confirming your status as a BU community member using your Kerberos password.

Once in the library system, you can use the links under "Resources" and "Collections" to find databases, eJournals, and eBooks, as well as search the library by subject. Some other useful links follow:

Go to <http://www.bu.edu/library/research/collections> to access eBooks and eJournals directly.

If you have questions about library resources, go to <http://www.bu.edu/library/help/ask-a-librarian> to email the library or use the live-chat feature.

To locate course eReserves, go to <http://www.bu.edu/library/services/reserves>.

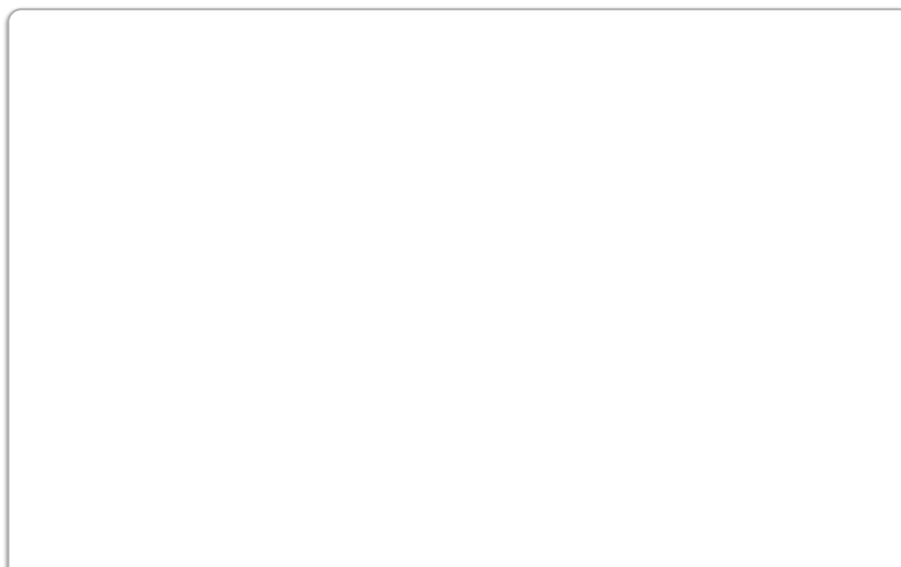
Please note that you are not to post attachments of the required or other readings in the water cooler or other areas of the course, as it is an infringement on copyright laws and department policy. All students have access to the library system and will need to develop research skills that include how to find articles through library systems and databases.

Free Tutoring Service



Free online tutoring with Smarthinking is available to BU online students for the duration of their courses. The tutors do not rewrite assignments, but instead teach students how to improve their skills in the following areas: writing, math, sciences, business, ESL, and Word/Excel/PowerPoint.

You can log in directly to Smarthinking from Online Campus by using the link in the left-hand navigation menu of your course.



Smarthinking Tutoring Overview



[YouTube](#)

Please Note

Smarthinking may be used only for current Boston University online courses and career services. Use of this service for purposes other than current coursework or career services may result in deactivation of your Smarthinking account.

Study Guide

Module 1 Study Guide and Deliverables

Readings:

- Module 1 online content
- Kimball/Ross ch 1
- Optional:
 - McKinney Ch 1
 - [Molinaro, SQL Cookbook](#)
 - Chapter 10:
Working with Ranges
 - Appendix A:
Window Function Refresher

- [Window Functions in SQL](#)

Discussions: Please post your introduction as soon as possible.

Extra Credit Discussion 1 due
Tuesday, March 23, at 6:00 AM ET.

Assignments: Assignments 1A and 1B due Tuesday,
March 23, at 6:00 AM ET

Term project topic and plan
submission due Thursday, March 25 at
6:00 AM ET

Assessments: Quiz 1 due Tuesday, March 23, at 6:00
AM ET

Live Classrooms:

- Wednesday, March 17 from
8:00 - 9:30 PM ET
- Saturday, March 20 from 11:00
AM - 12:30 PM ET

Module 2 Study Guide and Deliverables

Readings:

- Module 2 online content
- Kimball/Ross ch 2, 18
- Krishnan ch 6, 7

Discussions: Extra Credit Discussion 2 due
Tuesday, March 30, at 6:00 AM ET

Assignments: Assignment 2 due Tuesday, March 30,
at 6:00 AM ET

Assessments: Quiz 2 due Tuesday, March 30, at 6:00
AM ET

Live Classrooms:

- Wednesday, March 24 from
8:00 - 9:30 PM ET

- Saturday, March 27 from 11:00 AM - 12:30 PM ET

Module 3 Study Guide and Deliverables

- Readings:**
- Module 3 online content
 - Kimball/Ross Ch 19, 20

Discussions: Extra Credit Discussion 3 due Tuesday, April 6, at 6:00 AM ET

Assignments: Assignment 3 due Tuesday, April 6, at 6:00 AM ET

Term Project update due Thursday April 8 at 6:00 AM ET

Assessments: Quiz 3 due Tuesday, April 6, at 6:00 AM ET

- Live Classrooms:**
- Wednesday, March 31 from 8:00 - 9:30 PM ET
 - Saturday, April 3 from 11:00 AM - 12:30 PM ET

Module 4 Study Guide and Deliverables

- Readings:**
- Module 4 online content
 - Krishnan ch 12, 13

Discussions: Extra Credit Discussion 4 due Tuesday, April 13, at 6:00 AM ET

Assignments: Assignment 4 due Tuesday, April 13, at 6:00 AM ET

Assessments: Quiz 4 due Tuesday, April 13, at 6:00 AM ET

- Live Classrooms:**
- Wednesday, April 7 from 8:00 - 9:30 PM ET
 - Saturday, April 10 from 11:00 AM - 12:30 PM ET

Module 5 Study Guide and Deliverables

- Readings:**
- Module 5 online content
 - Krishnan ch 2, 3, 4, 11
- Discussions:** Extra Credit Discussion 5 due Tuesday, April 20, at 6:00 AM ET
- Assignments:** Assignment 5 due Tuesday, April 20, at 6:00 AM ET
- Assessments:** Quiz 5 due Tuesday, April 20, at 6:00 AM ET

- Live Classrooms:**
- Wednesday, April 14 from 8:00 - 9:30 PM ET
 - Saturday, April 17 from 11:00 AM - 12:30 PM ET

Module 6 Study Guide and Deliverables

- Readings:**
- Module 6 online content
 - Krishnan ch 8, 9
 - Optional:
 - [10 Tips to Improve ETL Performance](#)
 - [Optimization and Tuning in Data Warehouses](#)
- Discussions:** Extra Credit Discussion 6 due Tuesday, April 27, at 6:00 AM ET
- Assignments:**
- Assignment 6 due Tuesday, April 27, at 6:00 AM ET

- Final Term Project due
Tuesday, April 27, at 6:00 AM
ET
- Final Project Presentation
during Week 6

Assessments: Quiz 6 due Tuesday, April 27, at 6:00
AM ET

Live Classrooms:

- Wednesday, April 21 from 8:00 -
9:30 PM ET
- Saturday, April 24 from 11:00
AM - 12:30 PM ET

Final Exam Details

The Final Exam is a proctored exam available from **Wednesday, April 28 at 6:00 AM ET to May 1 at 11:59 PM ET**. The Computer Science department requires that all final exams be administered using an online proctoring service called Examity that you will access via your course in Blackboard. Detailed instructions regarding your proctored exam will be forthcoming from the Assessment Administrator. You will be responsible for scheduling your own appointment within the defined exam window.

Final Exam duration: **three hours**

Grading Information

Item	Title	Type	Weight
Assignment 1-A	Software and Appliance Installations	Lab	2.5%
Assignment 1-B	Analytical/Windowed Functions	Lab	2.5%
Assignment 2	Warehouse Design	Lab	5.0%
Assignment 3	Extract, Transform, and Load	Lab	5.0%
Assignment 4	Presentation of Data	Lab	5.0%

Assignment 5	Big Data	Lab	5.0%
Assignment 6	Tuning	Lab	5.0%
	All Assignments		30.0%
Quiz 1	Analytical/Windowed Functions	Quiz	5.0%
Quiz 2	Dimensional Data Modeling	Quiz	5.0%
Quiz 3	ETL	Quiz	5.0%
Quiz 4	Reporting and Forwarding Data	Quiz	5.0%
Quiz 5	Big Data Fundamentals	Quiz	5.0%
Quiz 6	Managing Performance of Business Intelligence Systems	Quiz	5.0%
	All Quizzes		30.0%
Final Exam	CS 689 Final Exam	Exam	25.0%
Term Project	Final Term Project	Term Project	15.0%
	Grand Total		100.0%

Discussion Board Grading (up to 6 points extra credit, one point per week)

Maximum points will be awarded based on three substantive posts by the student. Two posts should be on two different days and 15 hours apart over the module's one-week period. Posts are judged by their contribution to developing a conversation regarding that week's discussion focus. Early or late postings for that week's discussion are not eligible for extra credit.