

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 two-week-long 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

Andrew D Wolfe, Jr

Lecturer
Department of Computer Science
Boston University Metropolitan
College
808 Commonwealth Avenue, Rm
225
Boston, MA 02215

Email: awolfe@bu.edu
Phone: 617-358-1984



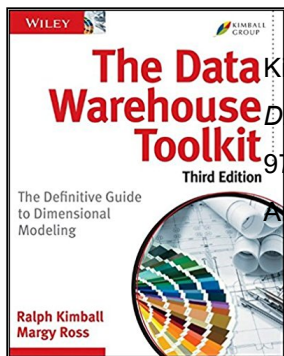
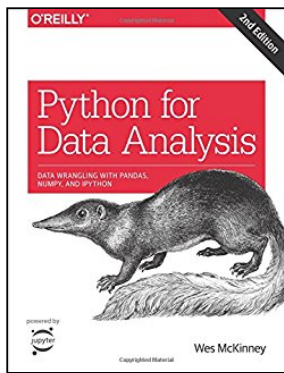
Greetings to Students and worldwide members of the Boston University Community!

I am your instructor, Andrew Wolfe. I have been a database developer my entire career, and first used Oracle in 1984 - that's Oracle 3 for those of you keeping score at home. I have delivered four commercial products, various consulting projects and several in-house implementations on Oracle, and others on SQL Server, Ingres, and PostgreSQL to name a few. I worked at Oracle for about eight years as a Database Architect. Also, I am myself a 2005 MSCS graduate of MET; I wrote a Master's Thesis on modeling information lifecycle in relational databases. I began online facilitating for MET a few years later, then began teaching face-to-face and eLive courses for MET. I am the instructor of record for the database security course, CS 674. Over this time I have absolutely loved teaching students like you and passing along the benefits of my hands-on experience. In 2016, teaching became too personally important to resist any more, so I sought and received a full-time Lecturer position with the MET Computer Science department. I teach courses relating to - you guessed it! - database and security.

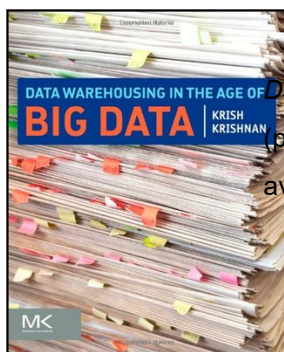
Materials

Required Book

McKinney, Wes. *Python for Data Analysis*. Sebastopol, CA: O'Reilly Media, 2013. ISBN-13: 978-1491957660 ISBN-10: 1491957662 An e-book is available at Vitalsource.com. An e-book is available through Amazon.



Kimball, Ralph and Ross, Margy. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd Edition. Indianapolis, IN: John Wiley & Sons, 2013. ISBN-13: 978-1118530801 An e-book is available at Vitalsource.com. An e-book is available through Amazon.



Data Warehousing in the Age of Big Data, 1st ed., Krish Krishnan. ISBNs: 9780124058910 (paperback), 9780124059207 (eBook). Waltham, MA: Morgan Kaufmann, 2013. eBook available - Amazon Kindle

Online Tutorials

Python <https://docs.python.org/2/tutorial/>

Python Pandas library: <http://pandas.pydata.org/pandas-docs/stable/tutorials.html>

Vertica: <https://my.vertica.com/docs/5.1.6/HTML/index.htm - 8871.htm>

Analytical functions in Vertica <https://my.vertica.com/docs/5.1.6/HTML/index.htm - 10955.htm>

Microsoft OLAP: [https://technet.microsoft.com/en-us/library/ms170208\(v=sql.100\).aspx](https://technet.microsoft.com/en-us/library/ms170208(v=sql.100).aspx)

Hadoop: <http://hadoop.apache.org/docs/r2.7.3/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>

Virtual Machines (Clusters)

- Windows. Installed: Microsoft SQL Server 2012; Microsoft SQL Server OLAP 2012; Microsoft Integration Services; Microsoft Excel; ER/win (Community Edition).

- Linux. Hadoop 4-node cluster.
- Linux. Vertica 4-node cluster.

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 is 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 [Collections](#) to access eBooks and eJournals directly.

If you have questions about library resources, go to [Ask a Librarian: Help & FAQs](#) to email the library or use the live-chat feature.

To locate course eReserves, go to [Reserves](#).

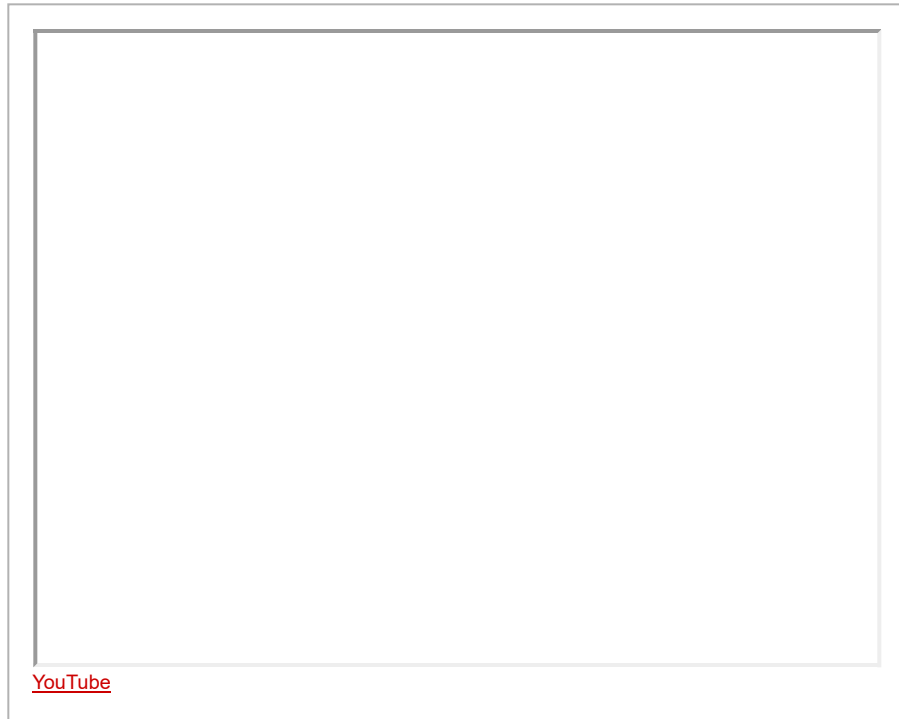
Please note that you are not to post attachments of the required or other readings in the water cooler or other areas of the course, as it is an infringement on copyright laws and department policy. All students have access to the library system and will need to develop research skills that include how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



Please Note

The SMARTHINKING service can be used for Boston University online class work only. Use of this service for personal purposes or for anything other than Boston University online class work will result in deactivation of your SMARTHINKING account.

Study Guide

Module 1 Study Guide and Deliverables

Readings:

- Module 1 online content
- Kimball/Ross ch 1
- Krishnan ch 6
- McKinney ch 1
- [Molinaro, SQL Cookbook, chapters 10, appendix A](#)
- [Window Functions in SQL](#)

Discussions: Please post your introduction as soon

as possible. Discussion 1 postings end
Tuesday, March 20 at 6:00 AM ET

Assignments: 1A: Due Saturday, March 17 at 6:00
AM ET
1B: Due Tuesday, March 20 at 6:00
AM ET

Assessments: Review Quiz 1

Live • Tuesday, March 13 8:00-9:30
Classrooms: PM ET
• Thursday, March 15 7:30-9:00
PM ET

Module 2 Study Guide and Deliverables

Readings:

- Module 2 online content
- Kimball/Ross ch 19, 20
- Krishnan ch 7
- McKinney ch 6, 7

Discussions: Discussion 2 postings end Tuesday,
March 27 at 6:00 AM ET

Assignments: 2A: Due Saturday, March 24 at 6:00
AM ET
2B: Due Tuesday, March 27 at 6:00
AM ET

Assessments: Review Quiz 2

Live • Thursday, March 22 7:30-9:00
Classrooms: PM ET

Module 3 Study Guide and Deliverables

Readings:

- Module 3 online content
- Kimball/Ross Ch 2, 18

- Krishnan ch 11

Discussions: Discussion 3 postings end Tuesday, April 3 at 6:00 AM ET

Assignments: Lab 3: Dimensional data modeling
due Saturday, March 31 at 6:00 AM ET

Assessments: Review Quiz 3

Live • Thursday, March 29 7:30-9:00

Classrooms: PM ET

Module 4 Study Guide and Deliverables

Readings:

- Module 4 online content
- Krishnan ch 12, 13

Discussions: Discussion 4 postings end Tuesday, April 10 at 6:00 AM ET

Assignments: Lab 4: Business reporting with data warehouses
due Tuesday, April 10 at 6:00 AM ET

Assessments: Review Quiz 4

Live • Thursday, April 5 7:30-9:00 PM

Classrooms: ET

Module 5 Study Guide and Deliverables

Readings:

- Module 5 online content
- Krishnan ch 2-4, 9

Discussions: Discussion 5 postings end Tuesday, April 17 at 6:00 AM ET

Assignments: Lab 5: Big Data Workshop
due Tuesday, April 17 at 6:00 AM ET

Assessments: Review Quiz 5

Live • Thursday, April 12 7:30-9:00

Classrooms: PM ET

Module 6 Study Guide and Deliverables

Readings:

- Module 6 online content
- Krishnan ch 8
- [10 Tips to Improve ETL Performance](#)
- [Optimization and Tuning in Data Warehouses](#)

Discussions: Discussion 6 postings end Tuesday, April 24 at 6:00 AM ET

Assignments: Assignment 6 due Tuesday, April 24 at 6:00 AM ET

Assessments: Review Quiz 6

Live • Thursday, April 19 7:30-9:00

Classrooms: PM ET

Important: Final Exam

You will be responsible for setting up your own appointment with an approved proctoring option. This exam will be three hours in length and will cover material from the entire course. Further information about the testing centers will be forthcoming from the exam coordinator.

Final Exam Details

The Final Exam is a proctored exam available from **Wednesday, April 25 at 8:00 AM ET to Saturday, April 28 at 11:59 PM ET**. The Computer Science department requires that all final exams be proctored.

You will receive a technical support hotline number before the start of the exam. Please bring this number with you to the exam.

Grading Information

Item	Title	Git?	Type	Weight
Discussion 1	Replacement for "TOP" or "LIMIT"	FALSE	Discussion	2.0%
Discussion 2	Extract, Load, Transform - ETL or ELT?	FALSE	Discussion	2.0%
Discussion 3	What Slowly Changing Dimension is Best?	FALSE	Discussion	2.0%
Discussion 4	Beautiful Lies	FALSE	Discussion	2.0%
Discussion 5	Time is Money	FALSE	Discussion	2.0%
Discussion 6	Why Can't You Just Open the Door and Look?	FALSE	Discussion	2.0%
Final Exam	CS 689 Final Exam	FALSE	Exam	15.0%
Assignment 1-A	Software and Appliance Installations	FALSE	Lab	2.0%
Assignment 1-B	Analytical/Windowed Functions	TRUE	Lab	10.0%
Assignment 2-A	Python and GitHub	TRUE	Lab	2.0%
Assignment 2-B	Python Operations with Extract, Transform, and Load	TRUE	Lab	10.0%
Assignment 3	Dimensional Data Modeling	TRUE	Lab	10.0%
Assignment 4	Reporting	TRUE	Lab	10.0%
Assignment 5	Big Data	TRUE	Lab	10.0%
Assignment 6	Tuning	TRUE	Lab	10.0%
Review Quiz 1	Analytical/Windowed Functions	FALSE	Quiz	0.0%

Review Quiz 2-A	ETL	FALSE	Quiz	0.0%
Review Quiz 2-B	Python	FALSE	Quiz	0.0%
Review Quiz 3	Dimensional Data Modeling	FALSE	Quiz	0.0%
Review Quiz 4	Reporting and Forwarding Data	FALSE	Quiz	0.0%
Review Quiz 5	Big data fundamentals	FALSE	Quiz	0.0%
Review Quiz 6	Managing performance of business intelligence systems	FALSE	Quiz	0.0%
Assignment T-1	Term Project Scoping	TRUE	Term Project	1.5%
Term Project	Final Term Project	TRUE	Term Project	7.5%
				100.0%

Discussion Board Grading

Discussion grading is based on the expectation of three posts by the student: one original post (76 points), that is, starting a new discussion thread, one reply on a thread started by another student (10 points), and one other reply posting (10 points). The original post should be at least three paragraphs that present and support an opinion or perspective, and defends it. Reply posts are judged by their contribution to or challenge of the other posters on the thread. While politeness and affirmation of classmates is appropriate, these will not add to the grade. A student with strong posts meeting these criteria will earn a 96 for that discussion. Additional points can be earned by doing additional postings, citing well-referenced sources, or having particularly effective discussion.

Who's Who: Roles and Responsibilities

You will meet many BU people in this course and program. Some of these people you will meet online, and some you will communicate with by email and telephone. There are many people behind the scenes, too, including instructional designers, faculty who assist with course preparation, and video and animation specialists.

People in Your Online Course in Addition to Your Fellow Students

Your Facilitator. Our classes are divided into small groups, and each group has its own facilitator. We carefully select and train our facilitators for their expertise in the subject matter and their excellence in teaching. Your

facilitator is responsible for stimulating discussions in pedagogically useful areas, for answering your questions, and for grading homework assignments, discussions, term projects, and any manually graded quiz or final-exam questions. If you ask your facilitator a question by email, you should get a response within 24 hours, and usually faster. If you need a question answered urgently, post your question to one of the urgent help topics, where everyone can see it and answer it.

Your Professor. The professor for your course has primary responsibility for the course. If you have any questions that your facilitator doesn't answer quickly and to your satisfaction, then send your professor an email in the course, with a cc to your facilitator so that your facilitator is aware of your question and your professor's response.

Your Faculty and Student Support Administrator, Jeff Behn. Jeff is here to ensure you have a positive online experience. You will receive emails and announcements from him throughout the semester. Jeff represents Boston University's university services and works for the Office of Distance Education. He prepares students for milestones such as course launch, final exams, and course evaluations. He is a resource to both students and faculty. For example, he can direct your university questions and concerns to the appropriate party. He also handles general questions regarding Online Campus functionality for students, faculty, and facilitators, but he does not provide tech support. He is enrolled in all classes and can be contacted within the course through Online Campus email as it is running. You can also contact him by external email at jeffbehn@bu.edu or call (617) 358-1985.

People Not in Your Online Course

Although you will not normally encounter the following people in your online course, they are central to the program. You may receive emails or phone calls from them, and you should feel free to contact them.

Your Computer Science Department Online Program Coordinator, Peter Mirza. Peter administers the academic aspects of the program, including admissions and registration. You can ask him questions about the program, registration, course offerings, graduation, or any other program-related topic. He can be reached at metcsol@bu.edu or (617) 353-2566.

Your Computer Science Department Program Manager, Kim Richards. Kim is responsible for administering most aspects of the Computer Science Department. You can reach Kim at kimrich@bu.edu or (617) 353-2566.

Andrew Gorlin, Academic Advisor. Reviews requests for transfer credits and waivers. Advises students on which courses to take to meet their career goals. You can reach Andrew at asgorlin@bu.edu, or (617)-353-2566.

Professor Anatoly Temkin, Computer Science Department Chairman. You can reach Professor Temkin at temkin@bu.edu or at 617-353-2566.

Professor Lou T. Chitkushev, Associate Dean for Academic Affairs, Metropolitan College. Dr. Chitkushev is responsible for the academic programs of Metropolitan College. Contact Professor Chitkushev with any issues that you feel have not been addressed adequately. The customary issue-escalation sequence after your course facilitator and course faculty is Professor Temkin, and then Professor Chitkushev.

Professor Tanya Zlateva, Metropolitan College Dean Dr. Zlateva is responsible for the quality of all the academic programs at Boston University Metropolitan College.

Disability Services

In accordance with University policy, every effort will be made to accommodate unique and special needs of students with respect to speech, hearing, vision, or other disabilities. Any student who feels he or she may need an accommodation for a documented disability should contact the [Office of Disability Services](#) at (617) 353-3658 or at access@bu.edu for review and approval of accommodation requests.

Netiquette

The Office of Distance Education has produced a netiquette guide to help you understand the potential impact of your communication style.

Before posting to any discussion forum, sending email, or participating in any course or public area, please consider the following:



Ask Yourself...

- How would I say this in a face-to-face classroom or if writing for a newspaper, public blog, or wiki?
- How would I feel if I were the reader?
- How might my comment impact others?
- Am I being respectful?
- Is this the appropriate area or forum to post what I have to say?

Writing

When you are writing, please follow these rules:

- **Stay polite and positive in your communications.** You can and should disagree and participate in discussions with vigor; however, when able, be constructive with your comments.
- **Proofread your comments before you post them.** Remember that your comments are permanent.
- **Pay attention to your tone.** Without the benefit of facial expressions and body language your intended tone or the meaning of the message can be misconstrued.
- **Be thoughtful and remember that classmates' experience levels may vary.** You may want to include background information that is not obvious to all readers.
- **Stay on message.** When adding to existing messages, try to maintain the theme of the comments previously posted. If you want to change the topic, simply start another thread rather than disrupt the current conversation.
- **When appropriate, cite sources.** When referencing the work or opinions of others, make sure to use correct citations.

Reading

When you are reading your peers' communication, consider the following:

- **Respect people's privacy.** Don't assume that information shared with you is public; your peers may not want personal information shared. Please check with them before sharing their information.
- **Be forgiving of other students' and instructors' mistakes.** There are many reasons for typos and misinterpretations. Be gracious and forgive other's mistakes or privately point them out politely.
- **If a comment upsets or offends you, reread it and/or take some time before responding.**

Important Note

Don't hesitate to let your instructor or your faculty and student support administrator know if you feel others are inappropriately commenting in any forum.

All Boston University students are required to follow academic and behavioral conduct codes. Failure to comply with these conduct codes may result in disciplinary action.

Registration Information and Important Dates

[View the drop dates for your course.](#)

[Withdraw or drop your course.](#)

- If you are dropping down to zero credits for a semester, please contact your college or academic department.
- **Nonparticipation in your online course does not constitute a withdrawal from the class.**
- If you are unable to drop yourself on student link please contact your college or academic department.

Technical Support

Experiencing issues with BU websites or Blackboard?

It may be a system-wide problem. Check the BU Information Services & Technology (IS&T) [news page](#) for announcements.

Boston University technical support is available via email (ithelp@bu.edu), the [support form](#), and phone (888-243-4596). Please note that the IT Help Center has multiple locations. All locations can be reached through the previously mentioned methods. For IT Help Center hours of operation please visit their [contact page](#). For other times, you may still submit a support request via email, phone, or the support form, but your question won't receive a response until the following day. If you aren't calling, it is highly recommended that you submit your support request via the technical-support form as this provides the IS&T Help Center with the best information in order to resolve your issue as quickly as possible.

Examples of issues you might want to request support for include the following:

- Problems viewing or listening to sound or video files
- Problems accessing internal messages
- Problems viewing or posting comments
- Problems attaching or uploading files for assignments or discussions
- Problems accessing or submitting an assessment

To ensure the fastest possible response, please fill out the online form using the link below:

IT Help Center Support
888-243-4596 or 617-353-4357 or Web
Check your open tickets using BU's ticketing system .

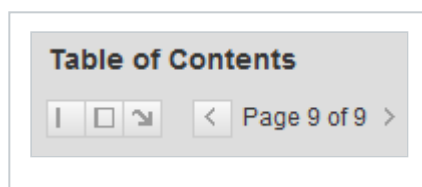
Navigating Courses

For best results when navigating courses, it is recommended that you use the Mozilla [Firefox](#) browser.

The Table of Contents may contain folders. These folders open and close (+ and – signs) and may conceal some pages. To avoid missing content pages, you are advised to use the next- and previous-page buttons (and icons) in the top-right corner of the learning content.

Please also familiarize yourself with the navigation tools, as shown below; these allow you to show and hide both the Course Menu and the Table of Contents on the left. This will be helpful for freeing up screen space when moving through the weekly lecture materials.

Navigation tools for the Table of Contents are shown in the image below:



Clicking on the space between the Course Menu and the Table of Contents allows you to show or hide the Course Menu on the left:



Web Resources/Browser Plug-Ins

To view certain media elements in this course, you will need to have several browser plug-in applications installed on your computer. See the Course Resources page in the syllabus of each individual course for other specific software requirements.

- Check your computer's compatibility by reviewing Blackboard's [System Requirements](#)
- Check your browser settings with Blackboard's [Connection Test](#)
- Download most recent version of [Adobe Flash Player](#)
- Download most recent version of [Adobe Acrobat Reader](#)

How to Clear Your Browser Cache

The IT Help Center recommends that you periodically [clear your browser cache](#) to ensure that you are viewing the most current content, particularly after course or system updates.

This page is also found within the "How to..." section of the [online documentation](#), which contains a list of some of the most common tasks in Blackboard Learn.

Boston University Metropolitan College