# Designing and Implementing a Data Warehouse

## MET CS 689

## BLENDED FORMAT – Fall 2020

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.

### COURSE DESCRIPTION

This course provides the student with the ability to analyze, design, and implement a data warehouse. The student will gain important foundational skills in applying database analytical functions and implementing extract-transform-load processes. From this point, we cover the modeling and implementation techniques for dimensional data warehouses, star/snowflake schemas, OLAP, and data lakes. The course also introduces Big Data concepts and technologies, including entity resolution in unstructured data and one or more massive-parallelism platforms. Students will do 6 two-week-long assignments and one final project.

### INSTRUCTOR

Mary E. Letourneau, Lecturer

maryleto@bu.edu

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 a DBA in the US division of a global corporation. 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.

Office hours: by appointment

### PREREQUISITES

 MET CS 579 or MET CS 669

 MET CS 521 or MET CS 520

### MATERIALS

Required Books:

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: 978-1-118-53080-1

Krishnan, Krish. *Data Warehousing in the Age of Big Data* , 1st ed., Krish Krishnan. Waltham, MA: Morgan Kaufmann, 2013. ISBN: 978-0-12-405891-0.

McKinney, Wes. *Python for Data Analysis*. Second Edition. Sebastopol, CA: O’Reilly Media, 2013. ISBN-13: 978-1-491-95766-0.

### COURSEWARE

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

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

Jupyter Notebook:

Tableau:

### CLASS RESOURCES

This course will provide students with the following resources:

• Access to Software with Free or Academic Licenses

• Access to Microsoft Azure data warehousing functionality

• Large-scale datasets suitable for warehousing

Recommended minimum system requirements:

• Intel-based

• i5 Core or equivalent

• 8 GB RAM

• 10 GB free disk space (if external, USB 3 or faster)

### CLASS MEETINGS, LECTURES AND ASSIGNMENTS

|  |  |  |
| --- | --- | --- |
| **Week** | **Description** | **Due / On** |
| 1: Sep 2 – Sep 8 | Lecture 01: Introduction | Sep 2 |
|  | Reading: Module 1 | Sep 9 |
|  | Reading: Kimball/Ross Chapter 1 | Sep 9 |
|  | Reading: Krishnan Chapter 6 | Sep 9 |
|  | Optional reading: McKinney Chapter 1 | Sep 9 |
|  | Install tools | Sep 9 |
| 2: Sep 9 – Sep 15 | Lecture 02: Analytic Functions | Sep 9 |
|  | Assignment 1 | Sep 16 |
|  | Term Project submission – Project description and plan | Sep 16 |
|  | Quiz 1 | Sep 16 |
| 3: Sep 16 – Sep 22 | Lecture 03: Extract and Transform | Sep 16 |
|  | Reading: Module 2 | Sep 23 |
|  | Reading: Kimball/Ross Ch 19 & 20 | Sep 23 |
|  | Reading: Krishnan Ch 7 | Sep 23 |
|  | Optional reading: McKinney Ch 6 & 7 | Sep 23 |
| 4: Sep 23 – Sep 29 | Lecture 04: Load and Verification | Sep 23 |
|  | Assignment 2 | Sep 30 |
|  | Quiz 2 | Sep 30 |
| 5: Sep 30 – Oct 6 | Lecture 05: Dimensional Data Modeling | Sep 30 |
|  | Reading: Module 3 | Oct 7 |
|  | Reading: Kimball/Ross Ch 2 & 18 | Oct 7 |
|  | Reading: Krishnan Ch11 | Oct 7 |
| 6: Oct 7 – Oct 13 | Lecture 06: Time, Bitemporality, Slowly-Changing Dimensions | Oct 7 |
|  | Assignment 3 | Oct 14 |
|  | Quiz 3 | Oct 14 |
| 7: Oct 14 – Oct 20 | Lecture 07: Big Data Approaches to Modelling | Oct 14 |
|  | Reading: Module 4 | Oct 21 |
|  | Reading: Krishnan Ch 12 & 13 | Oct 21 |
| 8: Oct 21 – Oct 27 | Lecture 08: Reporting | Oct 21 |
|  | Assignment 4 | Oct 28 |
|  | Quiz 4 | Oct 28 |
| 9: Oct 28 – Nov 3 | Lecture 09: Forwarding Data to Further Stores and Uses | Oct 28 |
|  | Reading: Module 5 | Nov 4 |
|  | Reading: Krishnan Ch 2, 3, 4 & 9 | Nov 4 |
| 10: Nov 4 – Nov 10 | Lecture 10: Dealing with Velocity, Volume, Variability | Nov 4 |
|  | Assignment 5 | Nov 11 |
|  | Quiz 5 | Nov 11 |
| 11: Nov 11 – Nov 17 | Lecture 11: Alternative Storage for Big Data | Nov 11 |
|  | Reading: Module 6 | Nov 18 |
|  | Reading: Krishnan Ch 8 | Nov 18 |
| 12: Nov 18 – Dec 2 | Lecture 12: Performance Analysis and Tuning for Data Warehousing and Big Data | Nov 18 |
| (No class Nov 25) | Assignment 6 | Dec 2 |
|  | Quiz 6 | Dec 2 |
| 13: Dec 3 – Dec 8 | Lecture 13: Course Wrap-Up and Final Exam Preparation | Dec 2 |
|  | Term Project | Dec 2 |
| 14: Dec 9 – Dec 15 | Study Period |  |
| Dec 16 | Final Exam |  |

### Fall 2020 COVID-19 Policies

Classroom Rotations –

Classrooms on campus have new capacities that follow guidelines issued by state and local health and government authorities related to COVID-19 and physical distancing. Before the beginning of the class, and throughout the semester, I will be reaching out to students who have indicated that they want to attend the classroom in-person. Our classroom can hold up to 15 students, which currently supports the expected enrollment. If enrollment increases, I may need to institute rotations of students to come to class on campus alternate weeks. You will be asked to attend remotely on the week that you have rotated out the classroom.

Compliance –

All students returning to campus will be required, through a digital agreement, to commit to a set of [Health Commitments and Expectations](http://www.bu.edu/dos/policies/lifebook/covid-19-policies-for-students/) including face coverings, symptom attestation, testing, contact tracing, quarantine, and isolation. The agreement makes clear that compliance is a condition of being a member of our on-campus community.

You have a critical role to play in minimizing transmission of COVID-19 within the University community, so the University is requiring that you make your own health and safety commitments. Additionally, if you will be attending this class in person, you will be asked to show your [Healthway](https://www.bu.edu/healthway/) badge on your mobile device to the instructor in the classroom prior to starting class, and wear your face mask over your mouth and nose at all times. If you do not comply with these rules you will be asked to leave the classroom. If you refuse to leave the class, the instructor will inform the class that they will not proceed with instruction until you leave the room. If you still refuse to leave the room, the instructor will dismiss the class and will contact the academic Dean’s office for follow up.

### Boston University is committed to offering the best learning environment for you, but to succeed, we need your help. We all must be responsible and respectful. If you do not want to follow these guidelines, you must participate in class remotely, so that you do not put your classmates or others at undue risk. We are counting on all members of our community to be courteous and collegial, whether they are with classmates and colleagues on campus, in the classroom, or engaging with us remotely, as we work together this fall semester.

### CLASS POLICIES

#### Attendance & Absences –

Students are expected to attend all classes or notify the instructor for an excuse with good reason three hours before class. After two unexcused absences the student forfeits all class participation credit.

#### Assignment Completion & Late Work –

All assignments will be submitted through Blackboard, and all quizzes and examinations will be administered through Blackboard. Students may receive up to a 36-hour extension without penalty, on a single assignment or assessment, by notifying the instructor 24 hours before that assignment or assessment is due, giving reason. Other extensions will be granted at the instructor’s discretion based on student circumstances. Students value quick feedback on assignments and assessments. In order to be able to provide this feedback in a timely fashion, no extensions will be longer than 36 hours.

The instructor will apply late penalties at his or her discretion, up to and including forfeiture of grade on any assignment. The instructor may apply additional penalties for repeated seeking of extensions or other late submission of work.

#### Academic Conduct Code –

WRITE IT, OR CITE IT!

Please review the Policy on Academic Conduct:

<http://www.bu.edu/met/metropolitan_college_people/student/resources/conduct/code.htm>

Neither the University, nor I, nor your classmates can tolerate plagiarism or other academic misconduct in any formal submission for this class. Please show appropriate respect for all – and for yourself – by expressing your own mastery of the material in your own words, diagrams, programming, etc. You must include references for everything you copy or quote. When you make such inclusions, mark and attribute them clearly and in appropriate academic style. You may not submit any other student’s work as your own, nor may you provide anyone else, in class or outside, with your own work on this class. Contact your instructor with any questions.

#### Grading Criteria

Overview:

Grades of coursework will be applied to the final course grade with the following weights:

|  |  |
| --- | --- |
| **Component** | **Weight** |
| Lab Assignments | 25 % |
| Quizzes | 25 % |
| Class Participation | 5 % |
| Final Project | 20 % |
| Final Exam | 25 % |
| **Total** | **100 %**  |

Lab assignments:

Labs will be graded using the following rubric:

Participation:

Participation includes asking questions, offering insights, sharing experiences, etc. relevant to the material being discussed. As such, participation implies attendance to lectures. Still it is understood that life happens. Let the instructor know as soon as possible if you cannot attend class. Up to two classes can be missed without impacting the participation part of your grade, if notice is provided in advance.

Extra Credit:

Up to 3 points added to final grade (up to ½ point per topic): Every two weeks will have a new discussion topic in the Blackboard to discuss. These topics rarely have a “correct” answer, and can be approached from many perspectives. An “A” grade in this portion of the grade requires substantive content relative to the topic posted on three different days during the two weeks, at least one of these posts must be an original post submitted before the end of the day modules’ first week’s Saturday (see schedule below – all times are Eastern Standard). The purpose is to encourage you to post early in the period and then go back later to read and respond to other students’ posts. (“I agree” or repeating someone else’s post is not substantive; neither is one short sentence considered substantive.) Please be respectful in your posts. Feel free to debate and disagree, but do it with extreme sensitivity.

|  |  |  |  |
| --- | --- | --- | --- |
| Module | Discussion Period Starts | Original Post due | Discussion Period Ends |
| 1 | Wed Sep 2nd 6 AM | Sat Sep 5th midnight | Wed Sep 16th 6 AM |
| 2 | Wed Sep 16th 6 AM | Sat Sep 12th midnight | Wed Sep 30th 6 AM |
| 3 | Wed Sep 30th 6 AM | Sat Oct 3rd midnight | Wed Oct 14th 6 AM |
| 4 | Wed Oct 14th 6 AM | Sat Oct 17th midnight | Wed Oct 28th 6 AM |
| 5 | Wed Oct 28th 6 AM | Sat Oct 31st midnight | Wed Nov 11th 6 AM |
| 6 | Wed Nov 11th 6 AM | Sat Nov 14th midnight | Wed Dec 2nd 6 AM |

##### Term Project:

While this one-semester course provides a solid foundation in data warehouses and big data, it is not exhaustive. The term project is intended to be an opportunity for you to further explore a topic from this course that is of interest to you. You will spend the first few weeks reviewing the topics and selecting one. The remaining weeks will be spent researching materials NOT already part of the curriculum and experimenting. The project submission will be a short report describing your research and sharing your findings, along with any successful code, design, project, etc. created during the experimentation.

Submission of work:

All labs and the term project will be submitted through the Assignments links in the Blackboard. The quizzes and final exam will be done online via the Assessments section of Blackboard. Except for the final exam, all work should be submitted by 6 AM EST the day it is due. If an assignment or quiz will be submitted late let the instructor know as soon as possible, but at least 24 hours beforehand. Up to two assignments can be submitted late without penalty **only** if pre-approved by the instructor. Otherwise, there will be a 5-point penalty for each day an assignment or quiz is late. Assignments and quizzes can be up to three days late before a “0” grade is posted. Quiz and assignment grades cannot be released until either all students have submitted or the late period has expired.

The final exam will occur during regular class time on final exam week.

### CONCLUSION

Ask questions early and often. I check my email frequently throughout the day, including weekends. I do not have an office on-campus, but I can arrange to meet on-campus before or after class, or online any other day of the week through the Blackboard Live Office feature.

This syllabus is subject to change. Announcements of changes will be made as early as possible.