

Designing and Implementing a Data Warehouse

MET CS 689 B1 for Fall 2017

Charles River Campus at MCS room B25

Math and Computer Science Building, 111 Cummington Mall

Tuesdays, 18:00-20:45

Andrew D. Wolfe, Jr.

awolfe@bu.edu

Office hours: Tuesdays 2:00-4:00 at Fuller 225, or by appointment

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.

Prerequisites

MET CS 579 or MET CS 669

MET CS 521 or MET CS 520

Books

The Data Warehouse Toolkit, 3rd ed., Kimball and Ross. ISBN: 9781118530801.

Indianapolis: Wiley, 2013.

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

Python for Data Analytics, 1st ed., McKinney. ISBN: 9781449319793. Sebastopol, CA: O'Reilly, 2012. (2d edition 2017 pending)

Courseware

Courseware will be *Blackboard*. URL is https://onlinecampus.bu.edu/webapps/blackboard/content/launchLink.jsp?course_id=42751_1&toc_id=610416_1

Class Meetings, Lectures & Assignments

Lectures, Readings, and Assignments are subject to change, and will be announced in class as needed within a reasonable time frame.

Date	module	Contents
2017-09-05	1	Lecture 1: Introduction [Lecture]
2017-09-12	1	Krishnan ch 6 [Reading] Kimball/Ross ch 1 [Reading] McKinney ch 1 [Reading] Lecture 2: Analytical Functions [Lecture] TBD analytical functions text [Reading]
2017-09-19	1	Lab: Analytical functions for typical business intelligence questions [Lab]
2017-09-19	2	McKinney ch 6, 7 [Reading] Lecture 3: Extract and Transform [Lecture] Kimball/Ross ch 19,20 [Reading]
2017-09-26	2	Lecture 4: Load and Verification [Lecture] Krishnan ch 7 [Reading]
2017-10-03	2	ETL Lab [Lab]
2017-10-03	3	Kimball/Ross Ch 2, 18 [Reading] Krishnan ch 11 [Reading] Lecture 5: 'Legacy' Dimensional Modeling [Lecture]
2017-10-10		COLUMBUS DAY RESCHEDULED MONDAY
2017-10-17	3	Lecture 6: Big Data approaches to data modeling [Lecture]
2017-10-24		PROFESSOR ABSENT - NO CLASS
2017-10-31	3	Dimensional data modeling lab [Lab]
2017-10-31	4	Lecture 7: Reporting [Lecture] Krishnan ch 12, 13 [Reading]
2017-11-07	4	Lecture 8: Forwarding data to further stores and uses [Lecture]
2017-11-14	4	Midterm Exam [Exam] Business reporting with data warehouses [Lab]
2017-11-14	5	Lecture 9: Dealing with Velocity, Volume, Variability [Lecture]
2017-11-21	5	Krishnan ch 2-4, 9 [Reading] Lecture 10: Alternative Storage for Big Data [Lecture]

2017-11-28	6	Lecture 11: ETL Performance Analysis and Tuning [Lecture]
2017-12-05	5	Entity Resolution using multiple unstructured datasets [Lab]
2017-12-05	6	TBD [Reading] Lecture 12: Tuning Reports and Ad Hoc exploration [Lecture]
2017-12-12	7	Lecture 13: Course Wrap-Up [Lecture]
2017-12-19	6	Data warehouse tuning lab [Lab]
2017-12-19	7	Student Term Project Presentations [Assignment] Term Project [Project] Lecture 14: Final Exam Preparation [Lecture]

Class Resources

This course will provide students with the following resources:

- Access to Microsoft Azure data warehousing functionality
- Access to Hadoop cluster computing resources
- Large-scale datasets suitable for warehousing

Class Policies

Attendance & Absences –

by course delivery mode:

- Face-to-Face:** 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.
- eLive:** Students are expected to attend all plenary sessions and to attend at least five classes remotely and participate during the class. Non-attendance of a plenary session for any reason may be grounds for forfeiting class participation credit. Non-attendance via remote may be excused for good reason by notifying instructor three hours in advance. Failure to attend remotely may be grounds for forfeiting class participation credit and may incur additional penalty.
- Online:** In accord with existing practice, no specific attendance is required of online students. Students are expected to participate in the module 4 online discussion in lieu of class participation.

Assignment Completion & Late Work –

All assignments will be submitted through Blackboard, and all quizzes and examinations will be administered through Blackboard. Students may receive a

36-hour extension without penalty, on a single assignment or assessment, by notifying instructor 36 hours before that assignment or assessment is due, giving reason. Other extensions will be granted at the instructor's discretion based on student circumstances. No access to take a quiz/assessment will be allowed 5 days after its original due date. 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

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

Component	Weight
Quizzes (6)	12%
Assignments (6) – see note	24%
Term Project	20%
Midterm Exam	14%
Final Exam	25%
Participation/ Online Discussion	5%

Note on Lab assignments: Labs will be graded on a minimum-completion/"mastery" basis. Simple completion of specified laboratory tasks will earn the student a grade of 95 (out of 100). To earn a higher grade, the student must demonstrate mastery of the task with additional work, for example:

- Variations in lab tasks that add substantially to the skills demonstrated
- Correct and insightful explanations of tasks in the lab
- Wrap-up essay effectively tying together the tasks of the lab, identifying goals and achievements in the lab work