

2018

CIS 4400 Data Warehousing for Analytics

Royce Kok

CUNY Bernard M Baruch College, royce.kok@baruch.cuny.edu

NYC Tech-in-Residence Corps

rdomanski@sbs.nyc.gov

Follow this and additional works at: https://academicworks.cuny.edu/bb_oers

 Part of the [Computer Sciences Commons](#)

Recommended Citation

Kok, Royce and NYC Tech-in-Residence Corps, "CIS 4400 Data Warehousing for Analytics" (2018). *CUNY Academic Works*.
https://academicworks.cuny.edu/bb_oers/4

This Syllabus is brought to you for free and open access by the Baruch College at CUNY Academic Works. It has been accepted for inclusion in Open Educational Resources by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

**Zicklin School of Business - Baruch College
City University of New York**

CIS 4400 Data Warehousing for Analytics

CIS 4400 – QTRA (66856) – Tuesday/Thursday 7:30 pm – 8:45 pm
Room: WmAnita Newman Vertical Campus 8-135

Course Syllabus

Professor Royce Kok
Office Hours: After lecture on Tuesday / Thursday or by appointment
E-Mail: royce.kok@baruch.cuny.edu (Preferred)
Please put the following in the *Subject* line for any e-mail to me: CIS 4400 followed by the specific subject of your e-mail.
Office: Room 11-270A (extension: 3388)
Course Materials : Course materials (including this syllabus) will be posted on Blackboard

Objectives This advanced course will provide students with an in-depth understanding of the design and implementation of database warehousing and analytics database systems. Specific topics include data warehouse modeling and architecture, the ETL process, administration, security, column-store, streaming and NoSQL databases, and complex event processing. Students develop a complete data warehouse system including implementation of a business intelligence suite.

Topics Include:

- Brief Review of Relational Design and SQL
- Data Warehouse modeling and Architecture
- Extraction, Transformation and Loading (ETL)
- Data Warehouse Technical Architectures
- NoSQL Databases
- Big Data Processing: Hadoop/Spark
- Commercial and Open Source BI Tools
- BI Application Design and Development

Learning Goals Upon successful completion of this course, students will be able to:

- Translate business needs and drivers into IT requirements for business intelligence systems
- Use the supporting technologies and data models for business intelligence including the process of and techniques for transforming business transaction data into appropriate analytic structures
- Explore state-of-the-art solutions for building and managing large data warehouse systems
- Discuss appropriate modeling approaches for a variety of industry specific requirements such as healthcare, banking, insurance, on-line advertising, and others.

- Develop a complete business intelligence system in a team setting using all of the tools and techniques presented during the course.

Upon successful completion of this course, students will have advanced skills to effectively design, develop, implement and manage medium to large-scale data warehouse systems.

- Technology Literacy: Students will master technologies used to develop and deploy data warehouses and analytics systems.
- Knowledge Integration: Students will be able to analyze business requirements across multiple industries and address these requirements with appropriate data warehousing and analytics technologies.
- Written communication: Students will analyze a business and develop and write a business analytics proposal that will be implemented during the semester.
- Oral communication: Students will present their business analytics solution
- Teamwork and Leadership: Students will work in groups to analyze a business and develop and write a business analytics proposal that will be implemented during the semester.
- Ethical Awareness: Students will discuss issues of privacy, customer data collection and management, energy use by data centers, and ethical concerns when collecting, analyzing and presenting analytical data.

**Textbooks /
Materials /
Resources**

- [Business Intelligence Guidebook: From Data Integration to Analytics](#) by Rick Sherman. Elsevier Science & Technology (2014). ISBN: 978-0-12-411461-6 Price: \$60
- Reference to Professor Holowczak's website : <http://faculty.baruch.cuny.edu/rholowczak/classes/4400/>
- Additional course materials will be provided on Blackboard.
- All students *must* obtain and maintain an e-mail account at Baruch and with an alternate Internet e-mail provider. If you use Gmail, Hotmail, Yahoo or other free mail service, be aware of storage limitations. Consider using multiple addresses to ensure my messages are delivered to you.
- Database and DW tools used this semester include [Oracle 12c](#), [Pentaho](#) and Tableau
- Computer Labs: VC 11-125, VC 11-115 (others to be announced)
- Academic Journals (e.g., CACM, IEEE CS) and Trade Magazines (e.g., Information Week, PC Week)

**Course
Content**

In addition to required reading in the textbook, there will be 2 to 3 homework assignments including implementations using the Oracle RDBMS and Pentaho BI Tools. Examinations will consist of a Mid-term exam and a Final exam. Students are expected to spend a significant amount of time outside the classroom meeting in groups and learning to use the Oracle Relational Database Management System and Pentaho development tools.

Prerequisites CIS 3400 Database Management Systems I
 Students must have a firm understanding of topics covered in CIS 3400 including the relational model, E-R diagraming, normalization and SQL. Students who received less than a "B" in CIS 3400, or students who have taken CIS 3400 more than 1 year ago should consult with the instructor prior to continuing on in CIS 4400.

Grading	<input type="checkbox"/> Midterm Exam	15%
	<input type="checkbox"/> Final Exam	25%
	<input type="checkbox"/> Homework and Class participation	30%
	<input type="checkbox"/> Oracle or SQL Server Data Warehouse Project	30%

This is a tentative grading schedule and is subject to change. Exams will *Not* be multiple choice. Homework assignments are due at the beginning of the class period. Late assignments will be graded down 5% per day late.

Prerequisites CIS 3400 Database Management Systems I
 Students *must* have a firm understanding of topics covered in CIS 3400 including the relational model, E-R diagraming, normalization and SQL. Students who received less than a "B" in CIS 3400, or students who have taken CIS 3400 more than 1 year ago should consult with the instructor prior to continuing on in CIS 4400.

Topics / Schedule (Tentative)

The following table gives a tentative lecture schedule for the course.

Week	Topics	Chapter in BI Guidebook
1	Course Introduction and Review of E-R Model Relational Model and SQL	Chapter 8 Notes/Handouts
2	Data Warehouse Project Planning	Chapters 1, 2, 3
3	Data Warehouse Architecture	Chapters 4, 5, 6 and 7
4	Dimensional Modeling	Chapters 9 and 10
5	Dimensional Modeling (Continued)	Chapters 9 and 10
6	Extraction, Transformation and Loading (ETL) and ELT <i>Topic 1 – ETL Subsystems</i>	11 and 12
7	Extraction, Transformation and Loading (ETL) and ELT (Cont.) <i>Topic 2 – ETL System Consideration</i>	11 and 12
8	Extraction, Transformation and Loading (ETL) and ELT (Cont.) <i>Topic 3 – ETL Architecture</i>	11 and 12
9	Review for Mid Term Exam Mid term exam 3/19 or 3/21	
10	Data Warehouse Technical Architecture	Notes/Handouts
11	Clustering and Distributed DBMS CAP Theorem	Notes/Handouts
12	NoSQL Databases	Notes/Handouts
13	Big Data Processing Architectures HADOOP and Spark	Notes/Handouts

14	Web Applications Integration, XML and semi-structured data analytics	Notes/Handouts
15	BI Application Design and Development	13 and 14
16	Final Exam Review	

Please note that this schedule is subject to change. Students are expected to come to class prepared and ready to participate. The associated chapters should be read ahead of time.

Additional Notes

- Baruch College's policy on Academic Honesty, which states, in part:

"Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the college's educational mission and the students' personal and intellectual growth. Baruch students are expected to bear individual responsibility for their work, to learn the rules and definitions that underlie the practice of academic integrity, and to uphold its ideals. Ignorance of the rules is not an acceptable excuse for disobeying them. Any student who attempts to compromise or devalue the academic process will be sanctioned."

Academic sanctions in this class will range from an F on the assignment to an F in this course. A report of suspected academic dishonesty will be sent to the Office of the Dean of Students. Additional information and definitions can be found at http://www.baruch.cuny.edu/academic/academic_honesty.html

- No makeups will be given for missed quizzes or exams.
- The instructor retains all midterm and final exams.
- The final exam must be taken by all students in the time slot posted in the college bulletin. Please make your business and travel plans to accommodate this schedule.
- Grades will not be given out via e-mail under any circumstances.
- If you miss class, it is your responsibility to find out about any announcements or assignments you may have missed.
- Cell phones etc. should be turned off during class and especially during exams.
- In general, the time to let me know about any problems or issues concerning missing class, long term illnesses, job related problems, academic probation, etc. is *before* you have missed a week or two of classes.
- All homework assignments are to be done individually. Students handing in similar work will both receive a 0 and face disciplinary actions.
- The instructor reserves the right to give unannounced quizzes if it appears students are not putting the time in to prepare for class.
- Students are expected to spend time outside of the classroom learning to use Oracle and/or SQL Server. This means you will need to spend time in the computer labs at school or on your own computer systems.
- Assignments may be turned in via e-mail. However, you are strongly advised to send me a sample attachment to ensure I can decode and view it properly. Not all e-mail programs work the same and many students have had problems sending attachments such as MS Word files.

- **Make backups of all of your work!** This includes any assignment and project materials you produce. I reserve the right to ask you to resubmit any assignment at any time.
- There will be programming assignments. If you have never taken a programming course before, you should (at the very least) read up on some notes is on Professor Holowczak's website: holowczak.com/programming-concepts-tutorial-programmers/
- In some portions of the course, we will work with web access to databases. It is important that you understand how a web browser and server interact and how to write some simple HTML including HTML forms. There are many tutorials on the web and you should know how to do this anyhow so dig in...
- You may wish to participate in the Baruch College CIS group on [Facebook](#)