Course Syllabus

Course Information

Course:	ACCT / MIS 6309 Business Data Warehousing
Term:	Spring 2017
Section:	501
Meets:	Friday, 7:00 pm to 9:45 pm, JSOM 1.107
Section:	502
Meets:	Monday & Wednesday, 5:30 pm to 6:45 pm, JSOM 2.106

Professor Contact Information

Instructor:	Kevin R. Crook
Email:	Kevin.Crook@utdallas.edu
Office Location:	JSOM 3.604
Office Hours:	Monday and Wednesday: 4:15 pm to 5:15 pm Friday, 5:45 pm to 6:45 pm
Teaching Assistant:	Roopa Kadavil
Email:	rxk160130@utdallas.edu
Office Location:	JSOM 2.604
Office Hours:	Monday and Wednesday, 10 am to 11 am Tuesday and Thursday, 1 pm to 2 pm Friday, 1:30 pm to 2:30 pm

Course Pre-requisites, Co-requisites, and/or Other Restrictions

While this course has no pre-requisites nor co-requisites, students without prior Information Technology exposure and/or database exposure will face a steeper learning curve.

Course Description

This course provides the student with in depth knowledge of Data Warehousing principles, Data Warehouse techniques, and Business Intelligence systems. The course introduces the topics of Data Warehouse design, Extract-Transform-Load (ETL), Data Cubes, and Data Marts. Students will create Business Intelligence using Data Warehouses with several OLAP and analytical tools.

This course section counts towards SAP Certification. The primary SAP products used in this course will be SAP BusinessObjects and Hadoop.

Student Learning Objectives/Outcomes

- Students will be able to describe architecture and methods for storage and provision of enterprise data.
- Students will develop competency in query development and essential business intelligence reporting.
- Students will demonstrate competency in data modeling, including dimensional modeling.
- Students will learn steps involved in development of an enterprise data warehousing solution and at least one BI tool from end-to-end perspective.

Required Textbooks and Materials

Academic Materials in PDF format:

Instructor will provide academic materials in PDF format for students to download. Materials will include slides and exercises based on peer reviewed academic papers in the fields of Relational Database Theory, Data Modeling using Entity-Relationship Diagrams, Dimensional Theory of Data, Data Warehousing Methodology, Dimensional Modeling, Cube Theory of Data, OLAP Cubes, ETL (Extract, Transform, & Load), Data Cleansing, Data Exploration, Master Data Management, Business Intelligence, Data Visualization, MPP (Massively Parallel Processing) Architectures for Data Warehousing / Data Marts, MapReduce algorithms and processing, Quantum Computing, Click Stream Analytics, and Social Media Analytics.

Cloud Access:

For the Hadoop exercises, we will be using the Amazon Web Services (AWS) Cloud. The service has a trial period and gives students a free quota to use. Students will have to pay for additional usage over the free quota.

Laptop:

Students are required to have a Windows 10 (64-bit) laptop, legitimately licensed copy, which they need to bring to class. Some of the software may run on older versions of Windows, but it will be at your own risk and responsibility. The software will run on 4 GiB of RAM, but more memory is better. If you have a Mac, please be warned that when using dual boot or Parallels students have encountered major issues.

Software:

This course will be heavily using software tools. All of these tools are free to use via open source, community editions, or academic editions.

- ERWin
  - Tool for Data Modeling using Entity Relationship Diagrams
  - Instructor will provide a download link in eLearning
  - Community Edition – they are transitioning to an Academic Edition

- OpenRefine
  - Tool for Data Cleansing / Data Exploration
  - http://openrefine.org
  - Open Source

- Tableau
  - Tool for Data Visualization
  - http://www.tableau.com/academic
  - Academic Edition

- SAP BusinessObjects
  - Tool for building a complex data semantic layer for Business Intelligence
  - Instructor will provide a download in eLearning
  - Academic Edition

- Hadoop Ecosystem
  - MPP (Massively Parallel Processing) platform for general use for anything that needs a distributed file system and massively parallel processing, such as a Data Mart, Click Stream Analytics, Social Media Analytics, etc.
  - Instructor will provide a virtual appliance in the cloud which students can clone and run in the cloud, or download and run in VMWare if they have 4+ cores and 16 GiB+
  - Open Source

Textbooks:

For the textbooks required or recommended for this course, many of these textbooks are available online for free from the UTD library, **http://www.utdallas.edu/library.** However, the library places a limit on the number of students who can view the textbooks simultaneously. Many of the textbooks are also available from Safari Books, **safaribooksonline.com**.

The following textbooks are **required** for this course:

Hadoop Essentials, Shiva Achari, Packt Publishing

Learning Tableau 10, Joshua N. Milligan, Packt Publishing
(please make sure this is for version 10, there is an older version of this book)

For students without a working knowledge of SQL, one of the following textbooks is **recommended**:

Sam's Teach Yourself SQL (one of the two versions of this book):
10 Minutes a Day (less than 4 hours to go through)
24 Hours

For students without a working knowledge of data modeling or ERWin, the following textbook is **recommended**:

Data Modeling Made Simple with CA ERWin,
Donna Burbank and Steve Hoberman,
Technics Publications

The following textbooks are **optional**, and are mentioned for students who want to learn more in depth on subjects covered in this course:

Dimensional Modeling:
The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modelling, Ralph Kimball and Margy Ross, Wiley

OpenRefine:
Using OpenRefine, Ruben Verborgh, Packt Publishing

BusinessObjects:
Creating Universes with SAP BusinessObjects, Taha Mahmoud, Packt Publishing

Assignments & Academic Calendar

Week Number Start Date	Topics Covered in Class	Assignments Due
Week 1 January 8 - 14	Introduction to the Course Dimensional Theory of Data Data Warehouse Methodologies	
Week 2 January 15 - 21	Dimensional Modeling Software: ERWin	
Week 3 January 22 - 28	Dimensional Modeling (continued) Software: ERWin	
Week 4 January 29 – February 4	Dimensional Modeling (continued) Software: ERWin	
Week 5 February 5 - 11	Cube Theory of Data OLAP Cubes	
Week 6 February 12 - 18	ETL – Extract, Transform, & Load Data Cleansing Data Exploration Master Data Management Software: OpenRefine	
Week 7 February 19 - 25	Business Intelligence Software: SAP BusinessObjects	

Week 8 February 26 – March 4	Business Intelligence – finish up Exam 1 501 – Friday, 7:00 – 8:15 PM (lecture after exam) 502 – Wednesday, 5:30 – 6:45 PM	
Week 9 March 5 - 11	Data Visualization Software: Tableau	Data Warehouse Design Project Due Friday, March 10, 2017 at 11:59PM
Week 10 March 12 - 18	Spring Break	
Week 11 March 19 - 25	Data Visualization (continued) Software: Tableau	
Week 12 March 26 – April 1	Data Visualization (continued) Software: Tableau	
Week 13 April 2 - 8	MPP – Massively Parallel Processing Architectures for Data Warehousing / Data Marts Quantum Computing (futuristic) Teradata Overview Vertica Overview Hadoop Overview Software: Hadoop (virtual appliance in cloud)	

Week 14 April 9 - 15	Hadoop Ecosystem (HDFS, MapReduce, Pig, Pig Latin, hBase, NoSQL) Hadoop Data Marts using Hive Software: Hadoop (virtual appliance in cloud)	
Week 15 April 16 - 22	Hadoop - Click Stream Analytics Hadoop – Social Media Analytics Software: Hadoop (virtual appliance in cloud)	Data Visualization Project Due Friday, April 21, 2017 at 11:59PM
Week 16 April 23 - 29	Hadoop – finish up Exam 2 501 – Friday, 7:00 – 8:15 PM (lecture after exam) 502 – Wednesday, 5:30 – 6:45 PM **No course assignments may be submitted for credit after 11:59 PM, Sunday, April 30th, 2017**	

Grading Policy

Grading Scheme

Data Warehouse Design Project	25 %
Data Visualization Project	25 %
Exam 1	25 %
Exam 2	25 %
Total	**100 %**

Grading Scale for Letter Grades
(Please note that undergraduate courses allow for a grade of A+, but graduate courses do not.)

93 – 100	A
90 – 92	A-
87 – 89	B+
83 – 86	B
80 – 82	B-
77 – 79	C+
73 – 76	C
70 – 72	C-
67 – 69	D+
63 – 66	D
60 – 62	D-
0 – 59	F

Course & Instructor Policies

- **Professional Conduct** - expected at all times from all students.
  Examples of unprofessional conduct includes, but is not limited to:
  - Trying the "game the system"
  - Coming late to class (without good reason)
  - Leaving early from class (without good reason)
  - Talking in class
  - Working on other subjects while in class
  - Using electronic devices for other than the course materials
- **Make-up Exams** – restricted to documented emergencies with sufficient written proof provided. Make-up exams will be in essay format and content will differ from the regular exams. Since make-up exams will differ in content, no curve fitting may be applied at either the exam level or the semester level.
- **Extra Credit** – no extra credit assignments are available.
- **Class Attendance** – required, except for legitimate emergencies. Please be on-time and remain until class is dismissed.
- **Late Work** – 24 hour grace period to allow for technical problems with assignment submission. After the grace period, reduced by 25 % per 24 hour period. Please do not ask for any extensions or lifting of late penalties, unless it is a documented emergency with sufficient written proof provided.
- **Instructor Response Policy:**
  - For questions about course material, before contacting the instructor, please follow the following protocol:
    - If in class, ask questions in class so everyone can benefit.
    - If between classes, if possible, wait until the next class to ask questions so everyone can benefit.
    - Come to the office hours (instructor's or the TA's) to ask questions.
    - If it's urgent and/or cannot be done using one of the methods above:
      - First, email the TA with the question
      - Allow 24 hours (not including weekends, nor holidays, nor breaks) for the TA to respond. TA will not be available during breaks.
      - If the TA's answer does not satisfy you, please forward the email to instructor (Please do not send emails to instructor unless you have tried to resolve it with the TA first)
      - Instructor may save them until the next class and read and answer them at the start of class so everyone can benefit.
  - For personal questions or issues that only pertain to you, please follow the following protocol:
    - Come to instructor's office hour
    - If it cannot be handled in office hours, please email the instructor
  - Instructor will attempt to reply to all emails within 48 hours, not including weekends, nor holidays, nor breaks. Instructor will not generally be available during breaks.
- eLearning – will be used for class content and any changes to class content.

Comet Creed

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

"As a Comet, I pledge honesty, integrity, and service in all that I do."

UT Dallas Syllabus Policies and Procedures

The information contained in the following link constitutes the University's policies and procedures segment of the course syllabus.

Please go to http://go.utdallas.edu/syllabus-policies for these policies.

The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.