

Course: MIS 6060/4900

Title: Advanced Database and Big Data Management

Semester: Winter 2018

Instructor: Vijayan Sugumaran, Office: EH 442, Phone: (248) 370-4649

Email: sugumara@oakland.edu

Home page: <http://www.sba.oakland.edu/faculty/sugumara/>

Class Time: Wednesday 6:30 – 9:20pm, 212 EH

Office Hours: Wednesday 1:00 pm- 2:30 pm and by appointment.

Texts: *Required:*

(G.V. Post – DBMS)

Database Management Systems: Designing and Building Business Applications, by Gerald V. Post, Copyright 2011. (<http://www.jerrypost.com/Books>)

(G.V. Post – DM)

Data Mining Applications: Analyzing Business Data

by Gerald V. Post, Copyright 2010. (<http://www.jerrypost.com/Books>)

Beginning SQL Server 2012 For Developers, by Robin Dewson, 3rd Edition, Apress, ISBN-13: 978-1-4302-3750-1. (Available on our course Moodle page)

(MS DWT)

The Microsoft Data Warehouse Toolkit – With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset, Second Edition, by Joy Mundy, Warren Thornthwaite, and Ralph Kimball, Wiley Publishing, Inc., 2011. ISBN: 978-0-470-64038-8. (Available on our course Moodle page)

Cloudera Academic Partnership (CAP)

Training material available through CAP – Lecture (L) materials will be made available on the course Moodle page

Optional:

Data Warehousing Fundamentals for IT Professionals, Second Edition, by Paulraj Ponniah, John Wiley & Sons, Inc., 2010, ISBN: 978-0-470-46207-2.

Pre-requisite: MIS 5140 – Introduction to Database Data Warehouse

This course assumes that you have mastered the contents of the prerequisite course and use that knowledge in assignments, exams, and projects. Please review the materials from the prerequisite course if necessary.

Description: With the need for processing vast amount of data, the role of advanced data management systems in an organization has become increasingly important. The intent of this course is to expose students to advanced concepts in DBMS technology as well as how to manage and process large amounts of data to generate useful information for decision making. “Big Data Analytics” is

receiving a lot of attention these days and organizations are investing heavily in acquiring the necessary infrastructure and skilled workers (Data Scientists) to leverage the vast amount of operational and external data to gain competitive advantage. Business intelligence technologies such as data warehousing, online analytic processing, data mining, XML data processing, and data semantics have matured and become main stream in generating valuable control and decision-support business intelligence. This course will focus on not only the basic concepts in database administration, data warehousing and data mining, but also the tools and techniques used to generate business intelligence within an organizational context. Students will get hands-on experience by working with SQL Server and other open source tools such as Hadoop. We will use the Cloudera distribution of Hadoop and other related tools. The final project is intended to put to practice the theoretical knowledge gained during the course.

Objectives: After completing this course, students will have a working knowledge of:

- Database administration and Distributed Databases
- File organization techniques and deadlock management
- Designing and developing data warehouses and data marts for an organization
- Issues related to data extraction and preparation
- Building business intelligence portals
- Issues in Big Data analytics and parallel data processing
- Hadoop and related big data technologies such as MapReduce, Pig, Hive, and Impala
- Present/Demonstrate a complete data warehouse/data mining solution to peers

Grading:	Midterm	30%
	Assignments	25%
	Final Project	30%
	Project Presentation & class participation	15%

Grading Scale:

Considered “A”s	Considered “C”s	Considered “D”s
4.0 100.% - 98.60	2.9 79.59 – 78.60	1.9 69.59 – 68.60
3.9 98.59 – 96.60	2.8 78.59 – 77.60	1.8 68.59 – 67.60
3.8 96.59 – 94.60	2.7 77.59 – 76.60	1.7 67.59 – 66.60
3.7 94.59 – 92.60	2.6 76.59 – 75.60	1.6 66.59 – 65.60
3.6 92.59 – 90.60	2.5 75.59 – 74.60	1.5 65.59 – 64.60
	2.4 74.59 – 73.60	1.4 64.59 – 63.60
Considered “B”s	2.3 73.59 – 72.60	1.3 63.59 – 62.60
3.5 90.59 – 88.60	2.2 72.59 – 71.60	1.2 62.59 – 61.60
3.4 88.59 – 86.60	2.1 71.59 – 70.60	1.1 61.59 – 60.60
3.3 86.59 – 84.60	2.0 70.59 – 69.60	1.0 60.59 – 59.60
3.2 84.59 – 82.60		
3.1 82.59 – 80.60		
3.0 80.59 – 79.60		

(I do not use a curve to assign final grade)

ntative Schedule

<u>Day</u>	<u>Topics</u>	<u>Chapters</u>
Part I – Data Warehousing Fundamentals		
Jan. 3	SQL Server and Basics of Data Warehousing	G.V. Post-DBMS, Ch 9
Jan. 10	Creating Relational Data Warehouse and Master Data Management	MS DWT – Ch. 5, 6
Jan. 17	Designing and Developing ETL System	MS DWT – Ch. 7
Jan. 24	Core Analysis Services OLAP Database	MS DWT – Ch. 8
Part II – Developing Business Intelligence Applications		
Jan. 31	Building BI Applications in Reporting Services	MS DWT, Ch. 10
Feb. 7	The BI Portal and SharePoint	MS DWT, Ch. 12
Feb. 14	Midterm Exam	
Feb. 28	Incorporating Data Mining	MS DWT, Ch. 13 G.V. Post – DM, Ch.
9		
Part III – Big Data Analytics		
Mar. 7	Big Data Management, Hadoop Ecosystem	Handouts (L1 & L2)
Mar. 14	MapReduce, Writing MapReduce Programs in Java	Handouts (L3 & L5)
Mar. 21	Hadoop tools for data acquisition (Pig)	Handouts (L8 & L9)
Mar. 28	Introduction to Hive	Handouts (L11 & L12)
Apr. 4	Introduction to Impala	Handouts (L13)
Apr 11	Final Project Presentation	
Apr. 18	Final Project Report Due by Midnight – To be Turned in Moodle	