TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

School of Information Management

MBUA 512 DATABASES AND ANALYTICS

Trimester 2, 2016

COURSE OUTLINE

Prescription

This course provides databases and analytics knowledge for business analysts to function effectively. The databases component covers the fundamentals of relational databases, relational database modelling, and SQL database queries using enterprise database. The analytics component covers data extraction, visualisation and predictive analytics. Workshops enable students to obtain practical experience.

Course Learning Objectives

1. Understand the key concepts in database management
2. Design and implement an effective database on an enterprise system
3. Write effective database queries
4. Apply analytics in solving business problems
5. Interpret analytics results for decision making

Course Content

The course content is divided among 4 modules:

1. Understanding database management
2. Designing and implementing databases
3. Implementing databases and queries
4. Analytics and sense making

Please note that slight variations might be made to this schedule as the trimester progress. Changes will be communicated in class and through Blackboard if necessary.

Topics	Topic	Readings
Module 1: Understanding database management		
1	Introduction to database Management; • Linkages with business & systems analysis, and process design • The Database Environment & Development Process	*Foundations:* MDM Chapter 1 Storey, V. C., Thompson, C. B., & Ram, S. (1995). Understanding database design expertise. Data & Knowledge Engineering, 16(2), 97-124.

Module 2: Designing and modelling databases		
2	Database design and modelling • Class diagrams • E-R modelling	*Foundations:* MDM Chapter 2-5 *Optional:* De Lucia, Andrea, et al. "An experimental comparison of ER and UML class diagrams for data modelling." Empirical Software Engineering 15.5 (2010): 455-492. Eichelberger, H., & Schmid, K. (2009). Guidelines on the aesthetic quality of UML class diagrams. Information and Software Technology, 51(12), 1686-1698. Kesh, S. (1995). Evaluating the quality of entity relationship models. Information and Software Technology, 37(12), 681-689. Bolloju, N., & Leung, F. S. (2006). Assisting novice analysts in developing quality conceptual models with UML. Communications of the ACM, 49(7), 108-112. Genero, M., Poels, G., & Piattini, M. (2008). Defining and validating metrics for assessing the understandability of entity–relationship diagrams. Data & Knowledge Engineering, 64(3), 534-557.
3	Database design and modelling • Logical database design and the relational model • Integrity constraints • Transforming E-R models into relations • Normalization • Physical database design • Designing Fields • Denormalizing and partitioning data • Indexes • Designing for query performance	*Foundations:* MDM Chapter 2-5 *Optional:* Moody, D. L., & Shanks, G. G. (2003). Improving the quality of data models: empirical validation of a quality management framework. Information systems, 28(6), 619-650. Jajodia, S., & Ng, P. A. (1984). Translation of entity-relationship diagrams into relational structures. Journal of Systems and Software, 4(2), 123-133. Yeh, D., Li, Y., & Chu, W. (2008). Extracting entity-relationship diagram from a table-based legacy database. Journal of Systems and Software, 81(5), 764-771. Stonebraker, M., Pavlo, A., Taft, R., & Brodie, M. L. (2014, March). Enterprise database applications and the cloud: A difficult road ahead. In Cloud Engineering (IC2E), 2014 IEEE International Conference on (pp. 1-6). IEEE.
Module 3: Implementing databases and queries		
4	Introduction to SQL Programming • Creating a database • Processing single tables • SELECT/ INSERT / UPDATE	*Foundations:* MDM Chapter 6 http://www.w3schools.com/sql/
5	Advanced SQL Programming • Processing multiple tables (joins) • Data dictionary • Triggers & Routines	*Foundations:* MDM Chapter 7 http://www.w3schools.com/sql/

6	Advanced SQL Programming • Advanced querying • Recent developments in SQL	*Foundations:* MDM Chapter 7 http://www.w3schools.com/sql/ Lajus, J. and Mühleisen, H., 2014, June. Efficient data management and statistics with zero-copy integration. In Proceedings of the 26th International Conference on Scientific and Statistical Database Management (p. 12). ACM.
Module 4: Analytics and sense making		
7	Knowledge Discovery in Databases • Data warehouse architectures • Data warehouse design & modelling • ETL process • Cubes • Implementation	*Foundations:* MDM Chapter 9 Frawley, W.J., Piatetsky-Shapiro, G. and Matheus, C.J., 1992. Knowledge discovery in databases: An overview. AI magazine, 13(3), p.57. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., 1996. From data mining to knowledge discovery in databases. AI magazine, 17(3), p.37. *Optional:* Moody, D. L., & Kortink, M. A. (2003). From ER Models to Dimensional Models Part II: Advanced Design Issues. Journal of Data Warehousing, 8, 20-29. Corral, K., Schuff, D., & Louis, R. D. S. (2006). The impact of alternative diagrams on the accuracy of recall: A comparison of star-schema diagrams and entity-relationship diagrams. Decision Support Systems, 42(1), 450-468. Serrano, M., Trujillo, J., Calero, C., & Piattini, M. (2007). Metrics for data warehouse conceptual models understandability. Information and Software Technology, 49(8), 851-870. Winter, R., Klesse, M. (2010). Organizing Data Warehouse Service Providers. Business Intelligence Journal, 14(3), 31-39.

8	Analytics I • Descriptive analytics (OLAP, Visualization)	*Foundations:* MDM Chapter 11 Frawley, W.J., Piatetsky-Shapiro, G. and Matheus, C.J., 1992. Knowledge discovery in databases: An overview. AI magazine, 13(3), p.57. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., 1996. From data mining to knowledge discovery in databases. AI magazine, 17(3), p.37. https://datafloq.com/read/12-algorithms-every-data-scientist-should-know/2024#datascience https://aitsdmclub.wordpress.com/2015/08/13/data-science-cheat-sheets-python-r-mysql-sql-spark-hadoop-hive-machine-learning-django/ http://www.datasciencecentral.com/group/tutorials/forum/topics/cheat-sheet-data-visualization-with-r *Optional:* McAfee, A., Brynjolfsson, E. (2012). Big Data: The Management Revolution. Harvard Business Review, 90(10), 61-68. Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. Communications of the ACM, 54(8), 88-98. Behrangi, M. R., Fattolahi, A., & Watson, H. J. (2007). Determining Information Requirements for BI Applications. Business Intelligence Journal, 12(3), 24. Brohman, M. K., & Watson, H. J. (2006). Maximizing the Return on OLAP and Data Mining Analysts. Business Intelligence Journal, 11(3), 30. Ross, J.W., Beath, C.M., Quaadgras, A. (2013). You May Not Need Big Data After All. Harvard Business Review, 91(12), 90-98.
9	Database application development, data integration, interoperability • Client/Server architectures • Web integration, XML, REST • APIs, Mash-ups, Web scraping • RDF and Linked Data • Open data	*Foundations:* MDM Chapter 8 & 10 http://linkeddatabook.com
10	Further data management approaches • Types of NoSQL DB • MongoDB	*Foundations:* MDM Chapter 13 Menninger, D. (2010). Understanding the analytical database market. 15(1), 49-55.

11	Analytics II • Predictive analytics (data mining) • Analytics information requirements • Deriving business value from analytics	MDM Chapter 11 https://datafloq.com/read/12-algorithms-every-data-scientist-should-know/2024#datascience https://aitsdmclub.wordpress.com/2015/08/13/data-science-cheat-sheets-python-r-mysql-sql-spark-hadoop-hive-machine-learning-django/ http://www.datasciencecentral.com/group/tutorials/forum/topics/cheat-sheet-data-visualization-with-r
12	Data management • Data quality, data dictionary • Data governance • Data security & backup/recovery • Scaling databases • Dataspaces	*Foundations:* MDM Chapter 10 & 12 Franklin, M., Halevy, A. and Maier, D., 2005. From databases to dataspaces: a new abstraction for information management. ACM Sigmod Record, 34(4), pp.27-33. Pipino, L.L., Lee, Y.W. and Wang, R.Y., 2002. Data quality assessment. Communications of the ACM, 45(4), pp.211-218. *Optional:* Redman, T.C. (2013). Data's Credibility Problem. Harvard Business Review, 91(12), 84-88. Lee, Y., Madnick, S., Wang, R., Wang, F., & Zhang, H. (2014). A Cubic Framework for the Chief Data Officer: Succeeding in a World of Big Data. MIS Quarterly Executive, 13(1). Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of management information systems, 12(4), 5-33. Fernández-Medina, E., & Piattini, M. (2005). Designing secure databases. Information and Software Technology, 47(7), 463-477.

Trimester Dates

From Monday 25th July – Friday 4th November

Withdrawal from Course

1. Your fees will be refunded if you withdraw from this course on or before Friday 5th August 2016.

2. The standard last date for withdrawal from this course is Friday 7th October 2016. After this date, students forced to withdraw by circumstances beyond their control must apply for permission on an '*Application for Associate Dean's Permission to Withdraw Late'* including supporting documentation. The application form is available from either of the Faculty's Student Customer Service Desks or online.

Names and Contact Details

	Staff	Contact details	Room	Office Hours
Course Coordinator & Lecturer	Markus Luczak-Roesch	markus.luczak-roesch@vuw.ac.nz 04 463 5878	RH516	by appointment
Programme Administrator	Kim Hann Chris King	sim-ictgradschool@vuw.ac.nz 04 463 5103	RH521	9am-5pm

Class Times and Room Numbers

Tuesday 9:30 am-11pm 1pm-2:30pm
Location: WIG501 (ICT Graduate School premises, Wigan Street, Level 5, Room 501)

Course Delivery

The course will be centered on practical problem-solving tasks, instead of being lecture-based. The course's content is divided among modules, which are focused on specific problems described in a case or directly by an industry partner. The course is offered over twelve weeks, one day a week. Typically, a day will start with a presentation and a discussion of theoretical elements, followed by a workshop in which students will apply the concepts seen that morning to a simple situation. The afternoon session will follow a similar format. Students are expected to use afternoon workshop to apply the concepts seen in class to their project. This course will develop your ability for team work.

Readings

Refer to readings in the teaching schedule.

Expected Workload

This is a 15-point course. One point equates to approximately 10 hours of work, for a total of 150 hours for the course. You are expected to attend all course sessions, read assigned materials, and contribute to workshop activities. With 12 three-hour classes, a total of 36 in-class hours are required. The remaining 114 hours will be spread over the 12 teaching weeks. The following breakdown estimates the required time for each task, giving you a rough idea of how much time you may need to spend.

- Estimate of time required for assessments: 56 hours
- Application-oriented activities between a morning and afternoon class: 12 hours
- Group project work (feeding into group presentation and technical report): 46 hours

Assessment

Assessment items		Due date	Length	%	CLO(s)
1	One page reflective report on personal experiences with data management and analytics	Friday, 29th July 2016, 23:59	4 hours	10	1, 4, 5
2	Design and implementation of a relational database model for a business use case	Friday, 23rd September 2016, 23:59	16 hours	30	2, 3
3	Group presentation about the extended business use case	Tuesday, 25th October 2016, 9:30	4 hours	10	1,2,3,4,5
4	Technical report about the extended business use case	Friday, 4th November 2016, 23:59	2000 words (32 hours)	50	1,2,3,4,5

The Assessment Handbook will apply to all VUW courses: see http://www.victoria.ac.nz/documents/policy/staff-policy/assessment-handbook.pdf.

If you cannot complete an assignment or sit a test or examination, refer to
www.victoria.ac.nz/home/study/exams-and-assessments/aegrotat

Penalties
The penalty for late submission of work without a prior extension arrangement is a reduction of 10% of the available marks each calendar day, starting from the due date and time, up to 5 days after the due date. At the course coordinator's discretion, work handed in after 5 days may be assessed and feedback provided, but no grade will be assigned.

Extensions
Personal extensions are granted only in special circumstances and supporting evidence such as a medical certificate may be requested by the course coordinator.

Group Work
Students are required to work together in some workshop as well as participate in team activities in the classroom.

Use of Turnitin
Student work provided for assessment in this course may be checked for academic integrity by the electronic search engine http://www.turnitin.com. Turnitin is an on-line plagiarism prevention tool which compares submitted work with a very large database of existing material. At the discretion of the Head of School, handwritten work may be copy-typed by the School and submitted to Turnitin. A copy of submitted materials will be retained on behalf of the University for detection of future plagiarism, but access to the full text of submissions will not be made available to any other party.

Materials and Equipment
Software:
R statistical programming environment, SQLLite

Student feedback
Not applicable as this is the first time the course is offered.

Communication of Additional Information
Additional information or changes will be conveyed by means of in-class announcements, Blackboard, and e-mail. Please ensure that you check these communication channels regularly.

Link to general information
For general information about course-related matters, go to
http://www.victoria.ac.nz/vbs/studenthelp/general-course-information

Note to Students
Your assessed work may also be used for quality assurance purposes, such as to assess the level of achievement of learning objectives as required for accreditation and academic audit. The findings may be used to inform changes aimed at improving the quality of VBS programmes. All material used for such processes will be treated as confidential, and the outcome will not affect your grade for the course.

****