Curtin Business School
School of Information Systems

Curtin
University of Technology

Data Mining and Advanced Database Management 601

Unit Outline Semester 2, 2005

Unit Index Number	307773
Hours per Week	4
Weekly Tuition Pattern	2 hours lecture, and 2 hours laboratory.
Credit Value	25 credits
Pre-requisites	Designing Database Systems 502 or equivalent.
Alternate requisites	Knowledge of relational database concepts. Expertise in the use of SQL. Ability to design a relational database schema.
Unit Controller	Amit Rudra
Unit Controller's office	408-3047C
Unit Controller's email	Amit.Rudra@cbs.curtin.edu.au
Contact Telephone Numbers	9266 7055
School Telephone & Fax Numbers	9266 7685 (tel) or 9266 3076 (fax)
Website for Unit Materials	http://elearn.cbs.curtin.edu.au
Student Consultation Times	Available from the school office

Important Notice

The Unit Outline (this document) gives the student important information about the unit, aims, outcomes, syllabus, materials, timetable, program and assessment.

Note that important general information including policies, expectations, examinations, copyright, referencing, academic misconduct, assistance for communication skills, guidelines for assessment and CV development is available in the CBS Student Handbook and can be downloaded at http://www.cbs.curtin.edu.au/business/student-resources/student-handbook-2005.

July 2005
CRICOS provider code 00301J

UNIT INFORMATION

1.   Aims

This unit will deal with the effects of database architecture on the performance of applications using relational and dimensional schemas. It is particularly concerned with designing dimensional schemas for data warehousing applications and the use of SQL and data mining tools in such applications.

2.   Unit Outcomes:

1. Comprehend architecture, design and planning for database tables.
2. Understand performance issues while designing/analysing queries.
3. Gain knowledge of data mining techniques for analysing data from a database.
4. Understand the issues involved and advantages of using a data warehouse.
5. Comprehend design of database access for electronic commerce.
6. To understand and affect security aspects of databases. *(New)*

2.1   Learning Outcomes

On successful completion of this unit, students should be able to:

1. Understand how the database architecture may affect the performance of queries and transactions; and maintain security.
2. Be able to design, create and secure schemas suitable for data warehousing and data mining applications – in particular, fact and dimension tables.
3. Be able to implement these schemas with appropriate data structures including tables, views, materialised views, indexes, partitioned structures, large objects and constraints.
4. Be able to populate these structures with data using SQL and appropriate database utilities and verify the integrity of the data.
5. Be able to execute analytical queries against these data structures.
6. Use information supplied by the query optimizer to tune the queries for performance with high volume data.
7. Understand data mining techniques and be aware of data mining tools, sources and security issues of mining sensitive information.
8. Be able to develop simple web access procedures for data manipulation and database access auditing.

2.2   Professional Skills Outcomes

Employers worldwide want graduates who have developed effective work, life, and professional skills. The following skills are relevant to business graduates and are part of the CBS professional skills program.

On successful completion of this unit, students should be able to:

- Apply appropriate problem solving processes, arguments, critical and creative thinking to database applications using dimensional and relational schemas.
- Identify some database performance problems and analyse their causes.
- Create innovative solutions and report and present arguments to professional peers on data warehousing and data mining applications.

Look for the professional skill icon in your assessment tasks and in the semester program to find out their mark allocation and when they will be taught.

3.   Syllabus

Architecture of a DBMS. Designing constraints and structures for database access – in particular for data warehousing and data mining applications. Loading data and establishing data integrity. Optimization techniques. Mining a database for useful information. Designing database for electronic commerce. Security and ethical issues *(new)*.

4.   Materials

4.1   *Texts, Manuals, References, Other Resources*

4.2   *Texts*

Loney, K. and Koch, G., (2002), Oracle 9i The Complete Reference, McGraw-Hill/Osborne

4.3   *References*

In addition to the set textbook(s) the following references will be useful:

Oracle 9i/Oracle 10g Online References.
Han, J., Kamber, M. (2001), Data Mining – Concepts and Techniques, Morgan Kaufmann Publishers.
Kyte, T. (2001), Expert One-on-One Oracle, Apress.
Abbey, M., Corey, M., Abramson, I., (2002), Oracle 9i A Beginner's Guide, McGraw-Hill/Osborne.
Hobbs, L., Hillson, S., Lawande, A., (2003), Oracle 9iR2 Data Warehousing, Digital Press
McDonald, C., et al, (2004), Mastering Oracle PL/SQL, Apress.
Hotka, D., et al, (1999), Oracle Development Unleashed, SAMS.
Afyouni, H. (2006), Database Security and Auditing – *Protecting Data Integrity and Accessibility*, Thompson – Course Technology.

Most of these references are available in the Library. (Use the latest edition of any text).

5.   Timetable of Classes

1 x two-hour lecture and 1 x two-hour laboratory per week.

5.1   *Lectures*

The purpose of the lecture is to describe some of the principles in the unit and the application of those principles. A lecture does not cover all there is to learn about the area. You will need to read and study further those areas covered in the lecture. Similarly, it is not possible to cover in the lecture all the unit material contained in the Study Guide. However, all material in the Study Guide, whether covered in the lecture or not, forms the subject matter of the unit and is examinable in the tests and the final examination. The laboratory work will relate to and expand the lecture material and is essential in understanding the topics and in developing skills in the techniques used.

5.2   *Tutorials*

There are no tutorials in this unit.

5.3 Timetable of Classes

CODE	TYPE	DAY	TIME	ROOM
Sem 001	**Lecture**	Tuesday	17.30-19.30	204:234
Lab 001	Laboratory	Tuesday	19.30 - 21.30	402:231
Lab 002	Laboratory	Wednesday	17.30-19.30	402:231
Lab 003	Laboratory (Resv.)	Tuesday	15.00-17.00	402:231

6. Program

Week	Begin Date	Lecture	Pre-readings	Laboratory	Assessment Due
1	1 Aug	Architecture of a DBMS-Oracle Review of SQL and data warehousing.	Chapters 1-15. CD-ROM: SQL Ref., Oracle Concepts Guide	Review of SQL. Creating tables and applying constraints.	
2	8 Aug	Database organization and structures: indexes, tablespaces. Space calculations. Keys and constraints. Loading data.	Chapter 18. SQL Ref. Concepts Guide, CD-ROM: Utilities Guide.	Loading tables, conventional and direct methods.	
3	15 Aug	Data warehouse schemas – fact and dimension tables. Structures for data warehousing schema.	Online papers-ACM & IEEE from Gecko. Oracle manuals. References	Dimensional schemas. Fact and dimension tables. Structured tables. Indexes.	
4	22 Aug	Optimisation, access path selection. Optimiser hints. Materialised views.	Chapters 38, 40. SQL Ref.	Optimisation. Partitioned tables. Materialised views.	Labs 1-4 – Friday, 27 Aug.
5	29 Aug	*Tuition Free Week*			
6	5 Sept	The SQL analytical functions. Rollup, grouping, windowing. Optimisation of the analytic functions.	References, Online manuals	Optimisation using materialised views, dimensions and query rewrite.	
7	12 Sept	Large objects. Trees. PL/SQL: Procedures, functions and packages; cursors.	Ch. 27-29, 40. Concepts Guide, Online Ref.	SQL analytic function exercises.	Mid-Semester Test – Tue. 13 Sept.
8	19 Sept	Web applications using PL/SQL	Ch. 27-29, 39. Concepts Guide, Online Ref.	Large objects, trees. PL/SQL.	Labs 5-7 – Friday, 23 Sep.
9	26 Sept	*Tuition Free Week*			
10	3 Oct	Database security	References.	Web applications using PL/SQL.	
11	10 Oct	Data warehousing and introduction to data mining.	Oracle online, References	Data mining tools.	
12	17 Oct	Data mining tools.	References	Data mining examples.	Group assignment presentations
13	24 Oct	Data mining sources.	References	Data mining examples.	Group assignment presentations
14	31 Oct	Revision Unit feedback.			Assignment – Friday, 4 Nov.
15	7 Nov	Study week			
16	14 Nov	Exam week 1			
17	21 Nov	Exam week 2			

7.   Assessment

7.1   Summary

To pass this unit you must:
 a) Receive an overall grade of 5 or above and a mark greater than or equal to 50, AND
 b) Submit all assessment activities, AND
 c) Pass the exam.

No.	Assessment Activity	Percentage %
1	Laboratory and mid-semester test	30%
2	Assignment and presentation	30%
3	Final examination	40%
	Total	100%

7.2   Assessment Details

7.2.1   Laboratory Work and Mid-Semester Test

Selected laboratory exercises and lecture concepts will be assessed to provide feed back about the student's understanding of the material and skills developed relating to the material.

7.2.2   Group Assignment and Presentation

A feasibility study for a data warehousing/data mining application will be developed. It will include the creation of schemas, database objects, queries using analytic functions, optimisation, programs for data cleansing and the creation of dimensional data and appropriate web interfaces and data mining activities and tools.

The group will be required to present a summary of the assignment to the appropriate laboratory session.

The group will be required to determine the proportion of marks assigned to each group member.

Report writing uses similar principles and skills to those used when writing essays; however, the key differences are that reports include a formal summary, a numbering system, and a set of recommendations. Increasingly popular is substituting an *Executive Summary* in place of the *Abstract*. An *Executive Summary* encompasses all the major issues within the report. In business many readers will not read beyond the *Executive Summary,* hence this component must be written carefully to ensure it embodies the intent of the report. Some useful sites for writing different types of reports are as follows:

http://startup.curtin.edu.au/study_skills/report.html - Curtin's Study Skills Report Writing guide site. This is a comprehensive and user-friendly guide.

http://www.arbld.unimelb.edu.au/students/essay/essaytech.html
The University of Melbourne Essay and Report Writing Guide

http://www.auditnet.org/audit_report_writing_guide.htm
An Audit Report Writing Guide

7.2.3   Final Examination

This will cover material developed in the lectures and laboratory exercises.

7.3 Assessment Compliance Information

Due dates will be strictly adhered to. Extensions will be granted only in cases of demonstrated urgent need. It is your responsibility to check the due date.

The Final Exam will be held during the formal examination period. It is the student's responsibility to check the date and time of the Final Exam on the Curtin website.

Official release results for this unit are published on Oasis on the Curtin website .
https://prodweb3.curtin.edu.au/exrprd/results.search

7.4 Penalty for Late Submission of Assessments

Late submissions may be penalised depending on the circumstances.

7.5 Relationship of Assessment Activities to Learning Outcomes and Professional Skills

This table illustrates how the assessment activities relate to the assessment of the learning outcomes and professional skills.

Outcomes	Assessment Activity			
Learning Outcomes	Lab & Test	Assignment & Presentation		Exam
1. Understand how the database architecture may affect the performance of queries and transactions; and maintain security.				
2. Be able to design, create and secure schemas suitable for data warehousing and data mining applications – in particular, fact and dimension tables.				
3. Be able to implement these schemas with appropriate data structures including tables, views, materialised views, indexes, partitioned structures, large objects and constraints.				
4. Be able to populate these structures with data using SQL and appropriate database utilities and verify the integrity of the data.				
5. Be able to execute analytical queries against these data structures.				
6. Use information supplied by the query optimizer to tune the queries for performance with high volume data.				
7. Understand data mining techniques and be aware of data mining tools, sources and security issues of mining sensitive information.				
8. Be able to develop simple web access procedures for data manipulation and database access auditing				
Professional Skills				
1. *(paste from Section 2.2)*				
2.				

8. General Information

Please see the CBS Student Handbook as described on Page 1 of this document for important general information.