

COMP3420 Advanced Databases and Data Mining

2010-S1 Basic Course Information

1 Course Description

The course deals with data analysis on very large volume of data to help business making strategic decisions. Specifically, it considers the issues related to the design and implementation of data warehouses, multidimensional data analysis, and fundamental data mining techniques and applications built on data warehouses.

2 Course Syllabus

This course examines the design of data warehouses and their use for data mining. Topics may include: star and snowflake data models for data warehouses, architectures of data warehouse, fact tables and datacube computation, materialized view computation and maintenance, OLAP (On-Line Analytical Processing), data preprocessing (data extraction, data cleaning, data reduction, data transformation, and data integration), data mining concepts and techniques including associated rules, clustering, etc.

3 Prerequisites

COMP1100 or COMP2720; COMP2400; 6 units of 2000-level IT courses; and 6 units of 1000-level MATH/STAT courses.

4 Lectures, Laboratories and Consultations

The course consists of thirty-hour lectures, *six* tutorials/labs (two hours each) which run together. The detailed arrangement is as follows.

Two lectures are offered per week. They are on Mondays and Tuesdays. The lectures are scheduled at

- 9:00am – 10:00am on Mondays, JD101 (Building 27)
- 1:00pm – 3:00pm on Tuesdays, PHYS T (Building 39a)

Six tut/labs will be offered for the course, which are run on Weeks 3, 5, 7, 9, 11, and 13, and there are four groups. The times for the tutorials and labs are scheduled at

- 10:00am – 12:00pm on Tuesdays, CSIT N112 (Tutor:)
- 5:00pm – 7:00pm on Tuesdays, CSIT N112 (Tutor:)
- 3:00pm – 5:00pm on Wednesdays, CSIT N114 (Tutor:)
- 11:00am – 1:00pm on Thursdays, CSIT N112 (Tutor:)

The consulting times are as follows.

- 11:00am – 12:00pm on Mondays, CSIT N334
- 11:00am – 12:00pm on Fridays, CSIT N334

5 Textbook and Reference Books

The following text book will be used for this course:

Data Mining: Concepts and Techniques
by Jiawei Han and Micheline Kamber
Morgan Kaufmann Publishers, 2nd Edition, March 2006, ISBN 1-55860-901-6.

The following reference book is recommended for this course:

Fundamentals of Database Systems
by Elmasri and Navathe
Addison-Wesley, 5th Edition, 2007.

6 Assessment

There will be TWO assignments and ONE final exam. Following the assessment policy of departmental courses, the assessment will be based on a 40:60 weighting for assignments (A/100) and the final examination (E/100), and the final marks will be calculated by the following formula:

$$Total = 0.4A + 0.6E$$

Note that final marks are moderated in departmental examiners' meetings and may be scaled as a result of this moderation.

The department's policy on plagiarism will be enforced.

7 Staff

The lecturers are Dr Weifa Liang and Dr Peter Christen, CSIT Building, phone: 6125-3019 (Liang), 6125-5690 (Christen). Course email: comp3420@cs.anu.edu.au. The manager of this course is Dr Liang.