

# PROPOSAL TO INTRODUCE A NEW COURSE

(formerly known as subject)

## 1. COURSE DETAILS

1.1 **Course ID** COMP9318

1.2 **Course name - Long**  
Data Warehousing and Data Mining

1.3 **Course name - Abbreviated**  
Data Warehousing and Data Mining

1.4 **Course Authority** **ext/email**  
Wei Wang 57162/weiw@cse.unsw.edu.au

<b>Proposers</b>	<b>ext/email</b>
Wei Wang	57162/weiw@cse.unsw.edu.au
Mike Bain	56935/mike@cse.unsw.edu.au
Xuemin Lin	56493/lxue@cse.unsw.edu.au
Ashesh Mahidadia	56556/ashesh@cse.unsw.edu.au
John Shepherd	56494/jas@cse.unsw.edu.au

1.5 **Organisational Unit responsible for course**

**School: Computer Science and Engineering**

**Faculty: Engineering**

Academic Group Code (Faculty): ENG

Academic Organisation Code (Owner): COMPSC

## 1.6 Justification of Proposal

The proposed course will focus on an important application area - data warehousing and data mining. Data warehousing and data mining is a highly inter-disciplinary area representing a confluence of several disciplines, such as database, machine learning, statistics, algorithms, neural network, signal processing and data visualization. It is in great demand in many applications, including business intelligent system (BI), customer relationship management system (CRM), financial market analysis, bioinformatics, Internet and World Wide Web management and analysis, etc. This course will provide the foundations of effective data analysis in those applications, and the students are expected to gain the basic skills to process those modern data analyses. This course will cover the *principles, algorithms, implementations, and applications* of data warehouse and data mining.

Currently, there are no courses in the campus covering the topics listed above yet.

## 1.7 Consultation Process

The Head of School, AI and Database groups.

1.8	Units of credit (UOC)	Session/s offered	Hours Per Week
	6	2	3

1.9 **Pre-requisites:** COMP3311 or (COMP9311 and (COMP9024 or COMP2011))  
**Co-requisites:** N/A



2001.

- D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, MIT press, 2001.
- Ian H. Witten, Eibe Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 1999.
- U. Fayyad, G. Grinstein, and A. Wiese, *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufman, 2001.
- M. Garofalakis, J. Gehrke, and R. Rastogi, *Query and Mining Data Streams: You only get one look*, Tutorial Notes in VLDB2002.
- Selected papers from the conference proceedings of SIGMOD, VLDB, SIGKDD, FOCS, SODA, ICALP, ICDE, etc.

Proposed syllabus:

Week	Description
1	<b>Introduction:</b> <ul style="list-style-type: none"> <li>• Admin</li> <li>• Overview (“what is the data mining problem”)</li> </ul>
2	<b>Data warehouse and OLAP 1:</b> <ul style="list-style-type: none"> <li>• The multi-dimensional data model, physical data model (ROLAP/MOLAP/HOLAP) and architecture</li> <li>• OLAP Queries (its difference with OLTP queries)</li> </ul>
3	<b>Data warehouse and OLAP 2:</b> <ul style="list-style-type: none"> <li>• Query processing techniques: data cube, materialized views, bitmap index, star-join, etc.</li> <li>• Integration with data mining</li> </ul>
4	<b>Data Preparation:</b> <ul style="list-style-type: none"> <li>• Data cleansing, integration and transformation</li> <li>• Data reduction (PCA), discretization</li> <li>• Attribute relevance, class comparisons, summary statistics, graphical summarisation</li> </ul>
5	<b>Association Rule Mining:</b> ( <i>might take an extra week</i> ) <ul style="list-style-type: none"> <li>• Transaction databases</li> <li>• Association rules, support, confidence and lift</li> <li>• Association rule mining (Apriori)</li> <li>• Association rule mining (Pattern-growth)</li> <li>• Mining Maximal and closed frequent itemsets</li> </ul>
6	<b>Classification and Prediction 1:</b> <ul style="list-style-type: none"> <li>• Precision and recall; over-fitting; training set and test set, (cross-validation)</li> <li>• Evaluating data mining – hypothesis, comparing algorithms, statistical tests, significance</li> <li>• Decision trees (Learning algorithms, rule extraction, integration with DW)</li> </ul>
7	<b>Classification and Prediction 2:</b> <ul style="list-style-type: none"> <li>• An introduction to Naïve Bayes</li> <li>• (An introduction to BP NN)</li> </ul>
8	<b>Classification and Prediction 3:</b> <ul style="list-style-type: none"> <li>• Regression (linear, log-linear, non-linear, etc.)</li> </ul>
9	<b>Cluster Analysis 1:</b> <ul style="list-style-type: none"> <li>• Introduction and categorisation</li> <li>• Partitioning-based: k-Mean</li> <li>• Hierarchical: BIRCH/CURE/link-based</li> <li>• Density-based: DBSCAN/OPTICS</li> </ul>
10	<b>Cluster Analysis 2:</b> <ul style="list-style-type: none"> <li>• Grid-based: CLIQUE</li> <li>• Model-based: SOM</li> <li>• Outlier analysis</li> <li>• Curse of dimensionality</li> </ul>

11	<b>Data Mining Applications 1:</b> <ul style="list-style-type: none"> <li>• Mining sequence data</li> <li>• Text mining (document/email classification/routing, etc.) or/and Web mining (content, log, usage)</li> </ul>
12	<b>Data Mining Applications 2:</b> <ul style="list-style-type: none"> <li>• Mining data streams</li> <li>• Mining biological data</li> </ul>
13	<b>Other Issues in Data Mining</b> <ul style="list-style-type: none"> <li>• Advanced data mining techniques (SVM, rough set, etc.)</li> <li>• Privacy issues (privacy preserving mining)</li> <li>• Summarizing DM techniques from other perspectives (e.g., from a Machine Learning perspective)</li> </ul>
14	Review and catch-up

### 1.21 Industrial experience component

N/A

## 2. RESOURCE STATEMENT

### 2.1 Enrolments

Estimated or proposed enrolments for the next three years.

2005: 100

2006: 150

2007: 200

### 2.2 Resource Requirements

#### Staffing Requirements:

Hours per week

Full-time Academic Staff: 5 hours/week (one semester only)

Part-time Teaching Staff: 6 hours/week.

General Staff: Covered by standard support of the Computer Support Group

**Field Costs:** N/A

**Studio/Laboratory Requirements:** N/A

**Materials Requirements:** N/A

**Equipment Costs:** N/A

**Computing Requirements:** Use of existing facilities in the school

**Library Requirements:** Standard textbooks for reference, journal articles and conference proceedings

**Capital Funds Requirements:** N/A

**2.3 Servicing Implications:**

N/A

**2.4 Teaching Arrangements:**

YES

NO            X

**2.5 Alternative Delivery Arrangements:**

N/A

**2.6 Details of Tuition Fees:**

Proposed fee: Standard Engineering course fee scale applies.

\$            for non-award enrolment (local)

\$            for non-award enrolment (international)

\$            for course which forms part of full fee-paying program (for local students)

\$            for course which forms part of full fee-paying program (for international students)

### 3. AUTHORISATION

#### 3.1 Principal Librarian's Endorsement

I have examined the Library needs related to the above proposal and certify that existing Library holdings, staffing, services and accommodation are adequate / inadequate (delete one) to cover the demands that are inherent in it.

Appropriate arrangements for the use of digitised material to support this course have been made by the Course Authority with the Principal Librarian.

Further Comments:

Principal Librarian  
/ /2004

#### 3.2 Head of School's Approval

I have examined the resource implications of the above proposal in regard to staff, space, materials, equipment, capital funds, and computing, and certify that the School can cover the demands that are inherent in it.

Further Comments:

Head of School  
/ /2004

#### 3.3 Dean's Approval

I have examined the resource implications of the above proposal in regard to staff, space, materials, equipment, capital funds, and computing, and certify that:

- 3.3.1 (i) the proposal involves no additional resources. (A statement from the Head of School explaining how this can be achieved must be provided); or
- (ii) the proposal involves additional resources and it is proposed to redeploy existing resources within the faculty. (A statement from the Head of School explaining how this will be achieved must be provided); or
- (iii) the proposal involves additional resources to be obtained as set out below; or

(iv) the additional resources essential to bring the proposal into effect cannot be found within resources available to the faculty.

3.3.2 a fee will / will not (delete one) be charged for this course. If a fee is to be charged the Dean certifies as follows:

I have ensured that the Vice-Chancellor has been advised of the proposed fee arrangements, and note that approval of fee arrangements is needed before a new course can be implemented.

3.3.3 the proposal conforms to the University's commitment to Equal Opportunity in Education.

Statement from Head of School on Source of Additional Resources and/or Further Comments:
--

Dean  
/ /2004