

BIS3214

Data Warehousing and Business Intelligence

Module Handbook

Autumn/Winter term – Sept start (24 week module)

2014/15

Joanna Loveday

School of Science and Technology

**Information in alternative formats**

This handbook can be found online at: [*http://unihub.mdx.ac.uk*](http://unihub.mdx.ac.uk)

If you have a disability which makes navigating the website difficult and you would like to receive information in an alternative format,

We can supply sections from this publication as:

* a Word document with enlarged type — sent by email or supplied on a CD or memory stick
* printed copy with enlarged type
* printed copy on non-white paper
* as Braille

Other formats may be possible. We will do our best to respond promptly. To help us, please be as specific as you can about the information you require and include details of your disability

**Disclaimer**

The material in this handbook is as accurate as possible at the date of production however you will be informed of any major changes in a timely manner.

**Other Documents**

Your module handbook should be read and used alongside your programme handbook and the information available to all students on UniHub including the Academic Regulations and Student Charter

# Contents

Contents 3

Module Summary/Introduction 4

Introduction 4

The module teaching team 4

Contacting the Module Leader 4

UniHelp 5

KIS Teaching and Learning Data 5

Module Aims 5

Learning Outcomes 5

Assessment Scheme 6

Assessment Weighting 6

Syllabus 6

Reading Materials 7

Core Texts 7

Additional texts 7

Online Resources 7

Study hours outside class contact 8

Brief Guide to Web-based Module Material 8

Coursework 9

Details of Coursework 9

Deadline for Submission of Coursework 9

Coursework – Part 1 9

Where to submit 12

Group Coursework 12

Electronic Receipt of Coursework 12

Feedback to students on coursework 12

Teaching Plan 13

Useful Information 15

MyUniHub 15

Attendance Requirements 15

Academic Dishonesty 15

Plagiarism 15

Appeals 16

Examples of all Typical/Previous Examination Papers 16

# Module Summary/Introduction

## Introduction

Large scale data management is now the central focus of most industrial data processing applications and the driving force behind Business Intelligence. This module covers the concepts and theories of managing very large data, namely data warehouse and deals with the concepts and theories of modelling and implementation of large volume data storage, preserving data quality, data preparation for dimensional modelling. The module also explores technologies and algorithms for handling a large amount of data, i.e., data cleansing, data segmentation and data transformation, to discover knowledge from data warehouse.

This module presents state of the art technologies and tools that support managers in decision-making process. The module is designed to give an insight on data modelling, dimensional modelling, online analytical transaction processing, decision-making and management of very large datasets. Some of the most widely used contemporary methods enabling modern decision support systems will be evaluated in this module.  In particular, the module explores theories in the domain of management decision making, for example, linear models, utility and decision-making under uncertainty.

The module provides students with an opportunity to explore, review and examine how large scale data can be managed in industry to enhance the decision making power of managers. The module also equips students with sufficient understanding of the current theory of decision making and its applications that address modern and future business needs using business intelligence.

## The module teaching team

Joanna Loveday – Module Leader

## Contacting the Module Leader

You can contact your module leader in the following ways:

**Office Hours – Thursday 1:00 to 3:00 Room No: T131**

**Email**  j.loveday@mdx.ac.uk

Telephone 020 8411 6613

MyUniHub pages <https://myunihub.mdx.ac.uk>

It is not necessary to book an appointment to see Joanna Loveday during the above office hours you just need to drop by.

In the first instance problems should be dealt with by talking to your lecturer after the lecture. Queries concerning course content are particularly suitable for the MyUniHub Discussion Board.

## UniHelp

On the Hendon campus, UniHelp is located on the Ground Floor of the Sheppard Library

Office Hours: Monday to Friday: 08.30 – 21.30

Saturday and Sunday: 11.00 – 18.00

## KIS Teaching and Learning Data

## Module Aims

Developing a Data warehouse for decision support and strategic planning has now become a most desirable objective for business organisations. A data warehouse, incorporating business intelligence, enables business organisations to discover new patterns and new phenomena in data in order to obtain analytical insight of business challenges. The aim of this module is to develop an appreciation and understanding of the techniques and approaches used to develop and deploy a data warehouse for purposes of business intelligence. The module will examine data quality and data transformation methods and practices, in addition to data warehousing modelling techniques, in order to build a quality data warehouse. Intelligent database and data mining theories and methods that aid knowledge discovery are introduced, as well as information retrieval, data presentation and pattern recognition techniques that facilitate and support business intelligence applications.

## Learning Outcomes

On completion of this module the successful student will be able to:

**Knowledge**

1. Analyse the industrial data resource architecture, management process for information resource integration and the process of establishing data relationship in order to build data warehouse that assists management in decision making process
2. Reflect on recent data discovery strategies in order to recommend an appropriate method for incorporating business intelligence in industrial information resources
3. Review the recent business intelligence tools to carry out critical evaluation on methodologies and technologies available for information retrieval, data presentation, data visualisation and pattern recognition

**Skills**

1. Identify data cubes and multi-dimensionality in business information repository and apply the models to develop a quality data warehouse
2. Apply contemporary business intelligence technologies in order to develop intelligent information resources that enable users to view data patterns by deploying various tools

## Assessment Scheme

The grade for the module is based on 100% coursework. There will be two parts to the coursework, coursework 1 (40%) and coursework 2 (60%).

The coursework of this module involves individual and group work, and gives students the opportunity to apply and examine database concepts and theories and to evaluate some of the traditional and intelligent methods in a practical application. Coursework 1 will assess outcomes 1 and 4; Coursework 2 will assess outcomes 2, 3 and 5.

**Students must pass both assessed components of a module individually, (i.e. coursework 1 and 2) in order to pass the module overall. Failure in one of the components will result in a failure of the module.**

## Assessment Weighting

**Coursework : 100%**

**Students must pass both assessed components of a module individually, coursework 1 and 2, in order to pass the module overall. Failure in one of the components will result in a failure of the module.**

## Syllabus

**This module covers two main area, Data Warehouse design and Business Intelligence.**

**See Teaching Plan for further details.**

## Reading Materials

### Core Texts

1. Data Mining: Concepts and Techniques, Third Edition (2011), Jiawei Han, Micheline Kamber, Jian Pei, The Morgan Kaufmann Series in Data Management System. ISBN 13: 978-0-12-381479-1.
2. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling by [Ralph Kimball](http://www.amazon.co.uk/s/ref=rdr_ext_aut?_encoding=UTF8&index=books&field-author=Ralph%20Kimball), [Margy Ross](http://www.amazon.co.uk/s/ref=rdr_ext_aut?_encoding=UTF8&index=books&field-author=Margy%20Ross),John Wiley & Sons; 2nd edition (2002) ISBN: 0-471-20024-7
3. Building the Data Warehouse by William H. Inmon, John Wiley & Sons; 4th edition (2005) ISBN -13: 978-0-7645-9944-6
4. Dimensional Data Warehousing with MySQL (A Tutorial)[Kindle Edition, ~ £7] by[Djoni Darmawikarta](http://www.amazon.co.uk/s/277-3998021-9794011?_encoding=UTF8&search-alias=digital-text&field-author=Djoni%20Darmawikarta) ISBN-13: 978-0-975218-2-0

### Additional texts

1. Oracle Data Warehousing and Business Intelligence Solutions, By Robert Stackowiak, Joseph Rayman, Rick Greenwald, illustrated edition, John Wiley and Sons, 2006, ISBN 0471919217, 9780471919216
2. Data mining for business intelligence: concepts, techniques, and applications in Microsoft Office Excel with XLMiner, By Galit Shmueli, Nitin R. Patel, Peter C. Bruce, Contributor Galit Shmueli, Nitin R. Patel, Peter C. Bruce, Edition: illustrated

Published by Wiley-Interscience, 2006, ISBN 0470084855, 9780470084854

1. Chen, Z. Intelligent Data Warehousing: From Data Preparation to Data Mining, CRC Press, 2002, ISBN: 0849312043
2. Data Warehouse Design, Modern Principles and Methodologies, McGrawhill, Matteo Golfarelli, Stefano Rizzi, 2009, ISBN: 978-0-07-161039
3. Successful Business Intelligence: Secrets to Making BI a Killer App, By Cindi Howson, edition: illustrated, Published by McGraw-Hill Professional, 2007, ISBN 0071498516, 9780071498517

### Online Resources

1. **Oracle9*i* Data Warehousing Guide, Release 2 (9.2):** <http://download.oracle.com/docs/cd/B10500_01/server.920/a96520/toc.htm>
2. **Data Warehousing Tutorial from 1 Key Data:** <http://www.1keydata.com/datawarehousing/datawarehouse.html>

## Study hours outside class contact

The study hours for each credit point are 10 hours. For a 30-credit module this would equate to 300 hours. Therefore, if a module has time-tabled activities i.e. lecture/seminar/lab, of 3 hours per week for a 24 week period (1 hrs Lecture and 2 hrs Seminar/Lab activities; total of 72 hours), then the out of class study commitment per module, in 24 week period, students have to study 228 hours.

## Brief Guide to Web-based Module Material

There is a module web page available on MyUniHub (**My Learning Portlet)** that contains:

* the module handbook
* lecture slides, which become available on the day of each lecture

# Teaching Plan

|  |  |  |
| --- | --- | --- |
| Lecture sessions in week no: | Title | Content |
| 1 | Module Introduction  Introduction to DBMS | Module Teaching Plan  Introduction Database approach: 3-schema architecture External, Conceptual & Internal schemas DBMS components, Characteristics  Brief history  Extending database capabilities |
| 2 | Data Modelling | Data Modelling, Example Database Application (COMPANY) ER model concepts, Relationships & Relationship Types, Structural constraints, Recursive relationships, Subtype, ER diagram, Transformation ER model, EER and UML for data modelling |
| 3 | Normalisation | Introduction Functional dependency, Direct, Indirect, Partial dependency  Normalization 1NF and dependency problems 2NF – solves partial dependency 3NF – solves indirect dependency BCNF – well-normalized relations |
| 4 | Introduction to Data Warehouse. | Introduction to Data warehouse What Is a Data Warehouse?  Differences between Operational Database Systems and Data Warehouses |
| 5 | Multidimensional Data Modelling | Multidimensional Data Model  Data Cubes  Stars, Snowflakes, and Fact Constellations: Schemas for Multidimensional Databases  Examples  Measures: Their Categorization and Computation OLAP Operations in the Multidimensional Data Model |
| 6 | Data Warehouse Architecture | Data Warehouse Architecture  Steps for the Design and Construction of Data Warehouses  A Three-Tier Data Warehouse Architecture   Data Warehouse Back-End Tools and Utilities  Types of OLAP Servers: ROLAP versus MOLAP versus HOLAP |
| 7 | OLAP | What is OLAP OLAP Operations SQL for OLAP |
| 8 | ETL | Review of DW Concepts ETL Example |
| 9 | Dimension Tables | Create Dimension Dimension Table Surrogate Keys Monster/small/junk Dimension Slowly change Dimension |
| 10 | Coursework Feedback |  |
| 11 | Data Warehouse Implementation | Data Warehouse Implementation  Efficient Computation of Data Cubes Indexing OLAP Data  Efficient Processing of OLAP Queries |
| 12 | Data Warehouse Implementation  Efficient Computation of Data Cubes Indexing OLAP Data  Efficient Processing of OLAP Queries  From Data Warehousing to Data Mining |
| 13 | Introduction to Business Intelligence | Basic Concepts  Statistical concepts  Algorithms |
| 14 | Introduction to Business Intelligence |
| 15 | Association and Correlations | Mining Frequent Patterns, Associations, and Correlations Market Basket Analysis: A Motivating Example  Frequent Itemsets, Closed Itemsets, and Association Rules  Frequent Pattern Mining: A Road Map |
| 16 | Efficient and Scalable Frequent Itemset Mining Methods  The Apriori Algorithm: Finding Frequent Itemsets  Improving the Efficiency of Apriori  Mining Various Kinds of Association Rules  Mining Multilevel Association Rules  Mining Multidimensional Association Rules from Relational Databases and Data Warehouses |
| 17 | Classification and Prediction | Classification and Prediction  Decision Tree Induction  Tree Pruning |
| 18 | Bayesian Classification  Associative Classification: Classification by Association Rule Analysis  k-Nearest-Neighbour Classifiers  Other Classification Methods -Genetic Algorithms  Prediction  Linear Regression |
| 19 | Cluster Analysis | Cluster Analysis,What Is Cluster Analysis?  Types of Data in Cluster Analysis  Interval-Scaled Variables, Binary Variables, Categorical, Ordinal, and Ratio-Scaled Variables, Variables of Mixed Types, Vector Objects  A Categorization of Major Clustering |
| 20 | Coursework Feedback |  |
| 21 | Other Data Mining Areas | Other Data Mining Areas  Mining Stream, Time-Series, and Sequence Data  Graph Mining, Social Network Analysis, and Multi-relational Data Mining  Mining Object, Spatial, Multimedia, Text, and Web Data |
| 22 | Big Data |  |
| 23 | Viva |  |
| 24 | Viva |  |

# Useful Information

The School has a student website <https://myunihub.mdx.ac.uk> dedicated to enrolled Science and Technology students, which provides information to support you on your programme of study, including information on the School’s Academic staff and:

* UniHelp opening hours
* Module Review Forms
* Learning Resources: Science and Technology
* Programme Handbooks

#### And other useful information such as

* Library Catalogue

## MyUniHub

Lecturers' contact details can also be found on MyUniHub, the university's online learning environment. This can be accessed from the following url: <https://myunihub.mdx.ac.uk> Within each module you can find 'module information' which displays contact details for the lecturer and other information about the module.

## Attendance Requirements

You should attend all scheduled classes. If you do not do so, you may not be able to demonstrate that you have achieved the Learning Outcomes for the module, and you are at risk of being graded “X” in the module. The definition of the X grade is: “Fail – incomplete without good reason: may not be reassessed.” As a general guide, you need to attend at least 75% of scheduled classes oin order to be able to demonstrate achievement of all Learning Outcomes. On some modules, there may be more specific attendance requirements.

## 

## Academic Dishonesty

Taking unfair advantage in assessment is considered a serious offence by the university, which will take action against any student who contravenes the regulation through negligence, foolishness or deliberate intent.

Academic dishonesty is a corrosive force in the academic life of the university; it jeopardises the quality of education and devalues the degrees and awards of the University.

The full regulations on academic dishonesty are given in the University Regulations, Section F Infringement of assessment regulations - academic misconduct.

## Plagiarism

Plagiarism is one specific form of cheating.

The University Regulation Section F clearly sets out the University’s understanding of plagiarism and the regulations by which you as a student of the University are bound. The key University regulation is F2.3 which defines plagiarism as “The presentation by the student as their own work of a body of material (written, visual or oral) which is wholly or partially the work of another, either in concept or expression, or which is a direct copy.”

Work presented for assessment must be the candidate’s own, or the work of a project group as requested by the tutor. Plagiarism is the representation of another person’s published or unpublished work as the candidate’s own by unacknowledged quotation. It is not an offence if the material is acknowledged by the candidate as the work of another through the accurate use of quotation marks and the provision of detailed references and a full bibliography, although the Assessment Board will not expect work to rely heavily on direct quotations.

In addition, the University Regulations set out the process for investigating allegations of plagiarism and describes the penalties. If you are found guilty, the repercussions are very serious indeed.

You should take steps, therefore, to understand what plagiarism is, how it can be identified and how you can avoid committing it; perhaps most importantly, you should reflect and come to understand why it is to your enormous advantage never to plagiarise because it is in effect cheating yourself and your fellow students).

Full details on the Infringement of assessment regulations - Academic misconduct, can be found in the University Regulations - Section F.

## Appeals

The full regulations on appeals are given in the University Regulations. Section G - Appeal regulations and procedures

## Examples of all Typical/Previous Examination Papers

Please go to the University student portal website <https://myunihub.mdx.ac.uk> for copies of previous examination papers in all subject areas across the University.