

Liverpool John Moores University

Title: DATA WAREHOUSING AND MINING
Status: Definitive
Code: **5023DACOMP** (125367)
Version Start Date: 01-08-2021

Owning School/Faculty: Computer Science and Mathematics
Teaching School/Faculty: Computer Science and Mathematics

Team	Leader
Glyn Hughes	Y

Academic Level: FHEQ5 **Credit Value:** 20 **Total Delivered Hours:** 55
Total Learning Hours: 200 **Private Study:** 145

Delivery Options

Course typically offered: Semester 2

Component	Contact Hours
Lecture	22
Practical	22
Tutorial	11

Grading Basis: 40 %

Assessment Details

Category	Short Description	Description	Weighting (%)	Exam Duration
Artefacts	AS1	Database Development	40	
Technology	AS2	Data Warehouse Development	60	

Aims

*To develop an awareness of data warehouse architectures and the development, populating with appropriate data (ETL) and uses of data warehouses
To develop facility in the process, tools and techniques of knowledge discovery in data warehouses*

Learning Outcomes

After completing the module the student should be able to:

- 1 Demonstrate knowledge of the structure, management and maintenance of data warehousing systems.
- 2 Distinguish, evaluate and make sound judgments about the effectiveness of alternative data warehousing architectures and their appropriateness for specific data problems.
- 3 Design an effective application of knowledge discovery in data warehouses
- 4 Design and demonstrate the effective use of data mining techniques.

Learning Outcomes of Assessments

The assessment item list is assessed via the learning outcomes listed:

Database Development	1	2
Data Warehouse Development	3	4

Outline Syllabus

Data Warehousing Concepts

- Contrasting Online Transaction processing and Data Warehousing Environments*
- Workload*
- Data modifications*
- Schema design*
- Typical operations*
- Historical data*

Components of a Data Warehouse Environment

- ETL Solution*
- OLAP/OLAP Cube engine*
- Client Analysis tools*
- Etc.*

Characteristics of Data Warehouses

- Subject oriented*
- Integrated*
- Nonvolatile*
- Time Variant*

Data Warehouse Architectures

- Data Warehouse (basic)*
- Data Warehouse (with Staging Area)*
- Data Warehouse (with Staging Area and Data Marts)*

Cross Industry Standard Process for Data mining (Crisp-DM)

- Business understanding*
- Data understanding*
- Data preparation*
- Modelling*
- Evaluation*
- Deployment*

Knowledge Discovery in Databases

Selection

Preprocessing

Transformation

Data Mining

Classification

Regression

Clustering

Summarization

Dependency modelling

Interpretation/Evaluation

Learning Activities

Learning activities include lectures and tutorials where students are encouraged to ask questions / discuss scenarios and supported labs where students are encouraged to put theory gained through lectures and tutorials into practice. Directed reading against appropriate industry and research sources further reinforces learning.

This module will have online practical.

Notes

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. Although it usually contains historical data derived from transaction data, data from other sources can be included. The data warehouse environment also includes an extraction, transportation, transformation and loading (ETL solution), an online analytical processing (OLAP) engine, client analysis tools etc. to manage the process of gathering data and delivering it to business users. Knowledge discovery in databases is the process of identifying valid, novel, potentially useful, understandable patterns in data. The extraction of knowledge from raw data is achieved by applying data mining methods, although data mining is just one step in a broader process.