



PRIFYSGOL CYMRU

Y Drindod Dewi Sant

UNIVERSITY OF WALES

Trinity Saint David

SWANSEA - ABERTAWE

UNIVERSITY OF WALES: TRINITY SAINT DAVID
PRIFYSGOL CYMRU: Y DRINDOD DEWI SANT

School of Applied Computing

Big Data ACUA6002

MODULE CODE: ACUA6002

TITLE: Big Data

LEVEL: 6

CREDITS: 20

AIM(S)

- To provide students with a comprehensive and systematic understanding of the concepts underlying the collection, storage and analysis of Big Data. The student's technical skills will be extended to allow them to use both data analysis and data mining tools in order to extract meaningful information from large datasets and present it in an effective manner.

LEARNING OUTCOMES

Upon the successful completion of this module, the student should be able to demonstrate the ability to:

1. Critically discuss the concepts behind, sources of and problems related to Big Data
2. Design, build and professionally document a small data warehouse suitable for use with Big Data
3. Perform data analysis and reporting on complex data within a Big Data data warehouse
4. Demonstrate a creative use of appropriate data mining and data visualisation techniques within a Big Data context.

INDICATIVE CONTENT

- Origins and sources of Big Data; volume, velocity, variety and ambiguity problems
- Big Data architectures and technologies; Hadoop, NoSQL and others
- Data warehousing for Big Data; traditional and modern architectures; workload management; emerging technologies
- Information management; Big Data lifecycle; semantic layers and the semantic web
- Data-driven architectures; master data management and metadata management
- Analysis and visualisation of Big Data; data mining, text mining, data modelling.

LEARNING AND TEACHING STRATEGY

Essential concepts and principles will be delivered formally in lectures supported by concurrent practical laboratory exercises, which will allow students to deploy appropriate theory, practices and tools. Case-study driven examples will be explored in detail to accentuate best-practice and practically explore issue resolution.

The allocation of teaching to deliver the module is:

Activity type	Hours	Percentage
Scheduled learning	60	30%
Independent learning	140	70%
Placement learning	0	
TOTAL	200	100%

ASSESSMENT

Description of each assessment component and its weighting, together with an example for new forms of assessment:

Assessment Component 1 – Coursework – 70%

Software development portfolio. Design and produce professional-standard reports based upon an example Big Data data warehouse using appropriate tools and techniques. Select and apply appropriate data mining processes to Big Data within a given problem domain and produce effective visualisations of the results.

Exemplar: 'The data provided is an indicative sample of the type of data collected by an online local news provider. Using this data you should do the following:

- Design and develop an appropriate data warehouse structure to store the data provided. You should justify all of your design decisions.
- Upload the data into your data warehouse, documenting the ETL process in a professional manner.
- Design, develop and critically evaluate reports that can be run to retrieve and display the following data:
 - All stories on a user-defined topic
 - All stories within a user-defined geographical area
 - A timeline showing the development of a complex story over a period of at least a week
 - Stories that are related to a user-defined story by location, content or other features.'

Assessment Component 2 – Written - 30%

2-hour exam. An open book exam based upon a pre-defined Big Data case study.

Exemplar: 'A case study to describe the data collection process of an online local news provider.

- Discuss the different types of data that the company will need to capture, process and store
- Critically appraise the tools and technologies available to the company and recommend which combination of these the company should use.
- Identify any potential weaknesses within the proposed system and suggest creative solutions to help ameliorate these.'

Assessment Summary

Activity type	Percentage
Written	30%
Coursework	70%
Practical	0%
TOTAL	100%

BIBLIOGRAPHY

Essential

Krishnan, K. (2013) *Data Warehousing in the Age of Big Data*. Elsevier.

Du, H (2010), *Data Mining Techniques and Applications – An introduction*. CEngage learning

Background

Foreman, J. W. (2013) *Data Smart: Using Data Science to Transform Information into Insight*. Wiley.

Provost, F. and Fawcett, T. (2013) *Data Science for Business*. O'Reilly Media.

Kimball, R., and Ross, M. (2013), *The Data Warehouse Toolkit (3rd Ed.)* Wiley.

Schutt, R. and O'Neill, C. (Eds.) (2013) *Doing Data Science*. O'Reilly Media.

Smolan, R. and Erwit, J. (2012), *The Human Face of Big Data*. Sterling.

Tuffery, S. (2011), *Data Mining and Statistics for Decision Making*. Wiley.

Witten, I.H., Frank, E. and Hall, M.A. (2011), *Data Mining – Practical Machine Learning Tools and Techniques*. Morgan Kaufmann

Websites

<http://www.kimballgroup.com>

tkdd.acm.org

<http://www.computer.org/tkde> (IEEE)