Module Specification

Big Data [TSI]

Version: 2021-22, v1.0, 26 Oct 2021

Contents

Part 1: Information

Module title: Big Data [TSI]

Module code: UFCEE1-12-M

Level: Level 7

For implementation from: 2021-22

UWE credit rating: 12

ECTS credit rating: 6

Faculty: Faculty of Environment & Technology

Department: FET Dept of Computer Sci & Creative Tech

Partner institutions: Transport and Telecommunication Institute

Delivery locations: Transport and Telecommunication Institute Latvia

Field: Computer Science and Creative Technologies

Module type: Standard

Pre-requisites: None

Excluded combinations: None

Co-requisites: None

Continuing professional development: No

Professional, statutory or regulatory body requirements: None

Part 2: Description

Overview: This module will provide students an understanding of concepts and techniques which underpin the capture, management and interrogation of large data sets.

Features: Not applicable

Educational aims: The aim of this module is to provide students with an understanding of technologies, challenges and trends within the subject area of Big Data and Cloud Computing

Outline syllabus: Data Storage and Retrieval:
Importance of data for business.
Understand the difference between data, information and knowledge.
Traditional ways to store and retrieve data.
Big Data challenges and opportunities.

Introduction to Big Data:
Defining Big Data: Sources of Big Data. The four dimensions of Big Data - volume, velocity, variety, veracity. Introducing storage and MapReduce.
Business application of Big Data: Big Data applications/examples in business.
Delivering business benefit from Big Data. Establishing the business importance of Big Data.
Addressing the challenge of extracting useful data/knowledge.
Integrating Big Data with traditional data.

SQL Databases vs. NoSQL Databases:
Understand the growing amounts of data.
The relational database management systems (RDBMS).
Capabilities of traditional RDBMSs.
Overview of Structured Query Languages (e.g. SQL).
Introduction to NoSQL databases.
Understanding the difference between a relational DBMS and a NoSQL database.
Identifying the need to employ a NoSQL DB.

Storing Big Data:
Analysing data characteristics: Selecting data sources for analysis.
Introduction of selected Big Data stores from the following list: Hadoop, Cassandra, Amazon S3, BigTable, etc.

Achieving Data Quality:

Introduction to data quality.
Why is data quality a business problem?
Problems when data is not “fit for purpose”.
Preparing data.
Ways to improve data quality.
Understand ETL - Extract, Transform, Load procedures to improve Data Quality.

Knowledge-based Information Retrieval:
Introduction to knowledge-based information retrieval.
Use for ontologies for knowledge modelling.
Learn how to build an ontology to link knowledge with data.
Using ontologies for information retrieval – case study.
Machine learning for knowledge acquisition: Introduction to machine learning and pattern recognition. Capabilities of different modelling, analysis and algorithmic techniques.

Big Data and Cloud Computing (technology, challenges and trends):
Cost of storing Big Data.
Is cloud computing a solution?
Issues: privacy and trust.
Future of Big Data and cloud computing.
Future research trends in Big Data.

Part 3: Teaching and learning methods

Teaching and learning methods: The module is delivered through weekly lectures and weekly tutorial sessions. Each lecture will direct the course and introduce the new ideas and skills required. Then small group tutorial sessions will enable each student to carry out the study and research exercises described in the associated work-sheet under the guidance of a Tutor.

Module Learning outcomes: On successful completion of this module students will achieve the following learning outcomes.

MO1 Understand the importance of data for business applications and the difference between data, information and knowledge in terms of their uses

MO2 Understand the challenges in storage and retrieval of small and large amounts of data, and the difference between SQL and NoSQL databases

MO3 Apply problem solving skills necessary for identifying the organizational needs to employ a SQL or NoSQL DB

MO4 Understand the four dimensions of Big Data i.e. volume, velocity, variety, veracity, which are important challenges the delivery of business benefits from Big Data

MO5 Be able to apply problem-solving skills to address the challenge of extracting useful data and application of data quality checks

MO6 Master various ways to improve data quality by understanding why data quality is a business problem

MO7 Apply knowledge modelling skills to generate ontologies to define domain knowledge and relationships between entities, and use them for information retrieval purposes

MO8 Demonstrate knowledge of Big Data management using Cloud computing and associated privacy and trust issues

Hours to be allocated: 120

Contact hours:

Independent study/self-guided study = 112 hours

Face-to-face learning = 48 hours

Total = 160

Reading list: The reading list for this module can be accessed at readinglists.uwe.ac.uk via the following link https://rl.talis.com/3/uwe/lists/55CF09B9-8FC4-6BF1-4426-90839BFCA98B.html?lang=en-gb&login=1

Part 4: Assessment

Assessment strategy: The assessment is split into two parts.

The Component A - Report.
This element involves students investigating and solving a business related cloud adoption problem based on given requirements, proposing a solution and preparing implementation specifications. The actual assignment topics are carefully chosen to demonstrate some basic principles, which are especially significant to the course.

Component A - Oral presentation/Viva
This element will consist of questions related to course work produced by the student which should test the students understanding of the fundamental concepts presented in the course work as well as their understanding and ability to apply those concepts and ideas to real-life scenarios (case studies).

There will be opportunities for formative assessment in the form of regular in-class presentations of research/implementation completed as part of tutorial work completed and subsequent group discussions.

Assessment components:

Report - Component A (First Sit)
Description: Written Report (3,000 words)
Weighting: 75 %
Final assessment: No
Group work: No
Learning outcomes tested: MO3, MO5, MO6, MO7, MO8

Presentation - Component A (First Sit)
Description: Report/Viva around the submitted report
Weighting: 25 %
Final assessment: Yes
Group work: No
Learning outcomes tested: MO1, MO2, MO4, MO5, MO6, MO8

Report - Component A (Resit)
Description: Written Report (3000 words)
Weighting: 75 %
Final assessment: No
Group work: No
Learning outcomes tested: MO3, MO5, MO6, MO7, MO8

Presentation - Component A (Resit)
Description: Presentation/Viva of report
Weighting: 25 %
Final assessment: Yes
Group work: No
Learning outcomes tested: MO1, MO2, MO4, MO5, MO6, MO8

Part 5: Contributes towards

This module contributes towards the following programmes of study:

Computer Science (Data Analytics and Artificial Intelligence) {Double Degree} [Feb][FT][TSI][2yrs] MSc 2021-22