

MODULE DESCRIPTOR

Module Title

Big Data Systems

Reference	CM3153	Version	1
Created	December 2023	SCQF Level	SCQF 9
Approved	April 2024	SCQF Points	15
Amended		ECTS Points	7.5

Aims of Module

This module aims to introduce students to state-of-the-art tools and analytics techniques for Big Data tasks, including NoSQL data stores, and modern parallel computation methodologies.

Learning Outcomes for Module

On completion of this module, students are expected to be able to:

- 1 Discuss different components of the Big Data ecosystem.
- 2 Manipulate different types of NoSQL data stores.
- 3 Assess the suitability of a given data store for a given problem.
- 4 Demonstrate the use of a parallel computation framework to extract information from Big Data.
- 5 Assemble a scalable data solution using a Big Data computation framework.

Indicative Module Content

Big Data ecosystems. Three V's - volume, velocity, veracity. Key-value store, document store, graph database, relational database. Schema migration. Apache Hadoop, MapReduce, Spark. Issues in Big Data ethics, law and security.

Module Delivery

This module is based on lectures supplemented with laboratory sessions, where appropriate software is applied to varied learning tasks and practical exercises.

Indicative Student Workload

	Full Time	Part Time
Contact Hours	30	N/A
Non-Contact Hours	120	N/A
Placement/Work-Based Learning Experience [Notional] Hours	N/A	N/A
TOTAL	150	N/A
<i>Actual Placement hours for professional, statutory or regulatory body</i>		

ASSESSMENT PLAN

If a major/minor model is used and box is ticked, % weightings below are indicative only.

Component 1

Type: Coursework Weighting: 100% Outcomes Assessed: 1, 2, 3, 4, 5
 Description: This coursework will consist of a big data development exercise.

MODULE PERFORMANCE DESCRIPTOR**Explanatory Text**

The calculation of the overall grade for this module is based on 100% weighting of component 1. An overall minimum grade D is required to pass the module.

Module Grade	Minimum Requirements to achieve Module Grade:
A	The student must achieve an A in C1.
B	The student must achieve a B in C1.
C	The student must achieve a C in C1.
D	The student must achieve a D in C1.
E	The student must achieve an E in C1.
F	The student must achieve an F in C1.
NS	Non-submission of work by published deadline or non-attendance for examination

Module Requirements

Prerequisites for Module	None.
Corequisites for module	None.
Precluded Modules	None.

INDICATIVE BIBLIOGRAPHY

- 1 HARRISON, G., 2015. Next Generation Databases: NoSQL, NewSQL, and Big Data. Apress.
- 2 LESKOVEC, J., ANAND, R. and ULLMAN, J.D., 2015. Mining of massive datasets. Cambridge University Press.
- 3 BERMAN, J., 2018. Principles and practice of big data: preparing, sharing, and analyzing complex information. 2nd ed. London: Academic Press.
- 4 MISHRA, R., 2018. PySpark recipes: a problem-solution approach with PySpark2. United States: Apress.
- 5 WIKTORSKI, T., 2019. Data-intensive systems principles and fundamentals using Hadoop and Spark. Cham: Springer.