BHSCDAD: Data Application Development

Module Code:	BHSCDAD
Long Title	Data Application Development APPROVED
Title	Data Application Development
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	5
Module Coordinator:	Arghir Moldovan
Module Author:	Arghir Moldovan
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	MSc and/or PhD degree in computer science or cognate discipline. May have industry experience also. Experience with tools, frameworks and programming languages for data analytics.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
#	**Learning Outcome Description**
LO1	Design algorithms and implement key programming patterns and constructs for data analytics.
LO2	Apply practical skills using a professional tool/language of data analytics (e.g., R, Python).
LO3	Assess the challenges associated with data application development and compare and contrast best practices for the data application development.
LO4	Investigate parallel and distributed computing and write programs for processing datasets in distributed computing and cloud computing environments using relevant programming paradigms and techniques.

Dependencies
Module Recommendations
No recommendations listed
Co-requisite Modules
No Co-requisite modules listed

Entry requirements	See Section 4.2 Entry Procedures and Criteria for the programme.

BHSCDAD: Data Application Development

Module Content & Assessment

Indicative Content

Data Applications Design and Development
Data science methodologies (e.g., KDD, CRISP-DM) Software design and development processes, use-case modelling, flowcharts, data-flow modelling Documentation and reporting

Tools and Frameworks
Tools and frameworks for data applications development (e.g., R Studio, JupyterLab) Programming languages for data analytics (e.g., R, Python) Use of support libraries (e.g., R pacakges, Pandas)

Data Structures and Functions
Data types and data structures for analytics Indexing and working with data structures Creating and working with functions

Extract, Transform, Load
Read/write data from/to different file formats (e.g., csv, xlsx, xml, json) Extract data from the Internet (e.g., connecting to APIs, web scraping) Programmatically connecting to databases, Create/Read/Update/Delete (CRUD) Operations Dealing with missing values Developing programs for data processing activities (e.g., data extraction, cleaning, merging, aggregation, analysis, reporting)

Data Visualisation
Data visualisation principles Data visualisation libraries (e.g., ggplot2) Dashboard frameworks (e.g., R Shiny)

Big Data Programming
Challenges associated with programming for big data Utilisation of cloud computing platforms for big data processing Distributed programming frameworks (e.g., Hadoop, Spark) Distributed programming paradigms (e.g., MapReduce) Design patterns

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Formative Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4
Non-Marked:	Yes		

Assessment Description:
Formative assessment will be provided on the in-class individual or group activities.

Assessment Type:	Practical (0260)	**% of total:**	30
Assessment Date:	Week 8	**Outcome addressed:**	2
Non-Marked:	No		

Assessment Description:
This assessment will consist of a practical in-class test, that will assess learners' competences on programmatically processing and analysing datasets.

Assessment Type:	Project	**% of total:**	70
Assessment Date:	Sem 2 End	**Outcome addressed:**	1,2,3,4
Non-Marked:	No		

Assessment Description:
The terminal assessment will consist of a project with both practical and research elements that will evaluate all learning outcomes. Learners will have to identify and carry out a series of analyses of at least two large datasets that complement each other, utilising appropriate programming languages, tools and techniques.

No End of Module Assessment

No Workplace Assessment

Reassessment Requirement

Coursework Only
This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.

Reassessment Description
The reassessment strategy for this module will consist of a project that will assess all learning outcomes. Students who fail the module will be afforded an opportunity to do the repeat project over the Summer months.

BHSCDAD: Data Application Development

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	No Description	24	Per Semester	2.00
Tutorial	No Description	24	Per Semester	2.00
Independent Learning	No Description	77	Per Semester	6.42
			Total Weekly Contact Hours	4.00

Workload: Part Time

Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	No Description	24	Per Semester	2.00
Tutorial	No Description	24	Per Semester	2.00
Independent Learning	No Description	77	Per Semester	6.42
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

J. D. Long, Paul Teetor. (2019), R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics, 2nd Edition. O'Reilly Media, p.600, [ISBN: 978-1492040682].

Wes McKinney. (2017), Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd Edition. O'Reilly Media, p.550, [ISBN: 978-1491957660].

Supplementary Book Resources

Todd Morley. (2019), Data Science Design Patterns, 1st Edition. Addison-Wesley Professional, p.512, [ISBN: 978-0134000053].

Tom White. (2015), Tom White. Hadoop: the Definitive Guide; Storage and Analysis at Internet Scale, 4th Edition. O'Reilly Media, p.756, [ISBN: 978-1491901632].

Bill Chambers, Matei Zaharia. (2018), Spark: The Definitive Guide: Big Data Processing Made Simple, 1st Edition. O'Reilly Media, p.606, [ISBN: 978-1491912218].

This module does not have any article/paper resources

Other Resources

[Website], DataCamp,
https://www.datacamp.com/

[Website], Andrew M. Raim. (2013), Introduction to Distributed Computing with pbdR,
https://pdfs.semanticscholar.org/e526/f5 7c29b5c60a0483eb544a3c8c705fd3af1c.pdf

Discussion Note:	