H8BGD: Programming for Big Data

Module Code:	H8BGD
Long Title	Programming for Big Data APPROVED
Title	Programming for Big Data
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	5
Module Coordinator:	EUGENE O'LOUGHLIN
Module Author:	Margarete Silva
Departments:	
Specifications of the qualifications and experience required of staff	

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 big data
LO2	Assess the challenges associated with processing big data datasets and compare and contrast programming for big data vis-à-vis programming for conventional datasets
LO3	Formulate and compose data flow and software documentation including flowchart, commenting and use-case diagram generation
LO4	Develop practical skills using a professional tool/language of data analytics (e.g. Python, R)

Dependencies

Module Recommendations

21358	H8BGD	Programming for Big Data

Co-requisite Modules

No Co-requisite modules listed

Entry requirements	

H8BGD: Programming for Big Data

Module Content & Assessment

Indicative Content

1. Introduction to Data Programming (50%)
Algorithm design Program I/O Data types and data structures Program control and process models Programming constructs Programming types (imperative, declarative, functional, logic) Programming languages for data analytics (e.g., R, Python) Developing programs for data processing activities (e.g., data extraction, cleaning, merging, aggregation, analysis, reporting)

2. Big Data Programming (50%)
Challenges associated with programming for big data Parallelism for computational processes Storage and compute locality Distributed computing Utilisation of cloud computing platforms for big data processing Distributed programming paradigms Distributed programming environments (e.g., Hadoop/HBase) MapReduce algorithm design Big data programming tools and languages (e.g., Pig, Hive)

Learning Environment
Learning will take place in both a classroom and computer laboratory environment with access to IT resources. Learners will have access to library resources, both physical & electronic and to faculty outside of the classroom where required. Module materials will be placed on Moodle, the College's virtual learning environment. Labs The labs will concentrate on implementing programs and manipulating data for analysis, and how best to implement the theory learned during the module.

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Practical (0260)	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4
Non-Marked:	No		

Assessment Description:
Assessment will be through a series of continuous assessment practical assignments given throughout the semester. Sample assessment: create a Python program that computes a company's inventory and returns the stock for a product requested by the user.

Assessment Type:	Project	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4
Non-Marked:	No		

Assessment Description:
Learners will be assessed through a project with both practical and research elements. Sample project: You are required to carry out a series of analyses of two datasets utilising appropriate programming languages and programming environments. For each of the chosen datasets you are required to compile a report of the analysis (circa 3,000 words for the report)

No End of Module Assessment

No Workplace Assessment

Reassessment Requirement

Repeat examination
Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.

H8BGD: Programming for Big Data

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	No Description	2	Every Week	2.00
Tutorial	No Description	1	Every Week	1.00
Independent Learning	No Description	7.5	Once per semester	0.63
			Total Weekly Contact Hours	3.00

Workload: Part Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	No Description	2	Every Week	2.00
Tutorial	No Description	1	Every Week	1.00
Independent Learning	No Description	89	Once per semester	7.42
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

Paul Teetor. R Cookbook, O'Reilly Media, [ISBN: 0596809158].

Tom White. Hadoop: The Definitive Guide, O'Reilly Media, [ISBN: 1449311520].

Stinerock, R.. (2018), Statistics with R: A Beginner's Guide, 1. Sage.

Supplementary Book Resources

Thomas H. Cormen... [et al.]. (2009), Introduction to algorithms, MIT Press, Cambridge, Mass., [ISBN: 0262033844.].

Donald Miner, Adam Shook. MapReduce Design Patterns, O'Reilly Media, [ISBN: 1449327176.].

This module does not have any article/paper resources

Other Resources

[Website], MIT Open Courseware videolectures.net. http://videolectures.net/mit6046jf05_int roduction_algorithms/.

[Website], Cloudera University. http://university.cloudera.com/onlineres ources/hadoopecosystem.html.

[Website], MIT Open Courseware. http://ocw.mit.edu/courses/electrical-en gineering-and-computer-science/6-00sc-in troduction-to-computer-science-and-progr amming-spring-2011/index.htm.

[Website], Andrew M. Raim. (2013), Introduction to Distributed Computing with pbdR at the UMBC, http://userpages.umbc.edu/~gobbert/paper s/pbdRtara2013.pdf

[Book], Wes McKinney. (2012), Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly.

[Book], Anand Rajaraman, Jeffrey David Ullman. (2014), Mining of Massive Datasets, Cambridge University Press.

Discussion Note:	