H9DAP: Database and Analytics Programming

Module Code:	H9DAP
Long Title	Database and Analytics Programming APPROVED
Title	Database and Analytics Programming
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	10
Module Coordinator:	Arghir Moldovan
Module Author:	Arghir Moldovan
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	PhD/MSc in a computing or cognate discipline. May have industry experience also.

Learning Outcomes

On successful completion of this module the learner will be able to:

#	Learning Outcome Description
LO1	Analyse, compare, contrast and critically evaluate the characteristics of programming languages, programming environments and database systems commonly utilised for data analytics solution implementation.
LO2	Critically 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	Evaluate tools and techniques for managing the data pipeline and preparing data for further analysis through data wrangling, cleaning, and validation.
LO4	Critically assess methods and practices for software development in order to design and implement data programming requirements.
LO5	Evaluate, design and implement solutions for processing datasets by using key programming patterns and constructs for data analytics, relevant programming languages, and suitable database systems.

Dependencies

Module Recommendations

No recommendations listed

Co-requisite Modules

No Co-requisite modules listed

Entry requirements	A level 8 degree or its equivalent in any discipline

H9DAP: Database and Analytics Programming

Module Content & Assessment

Indicative Content
Module IntroductionIntroduction to Data Programming Overview of programming languages, tools and frameworks for data analytics, and productionalizing tools (e.g., GitHub). Programming types (imperative, declarative, functional, logic, etc.);. Data analytics methodologies;. Algorithm design, Program I/O
Overview of the data programming language Syntax and semantics, expressions and statements, basic data types, conversion and coercion, built in data structures (arrays, matrices, lists, etc.), indexing data structures, program flow control and iteration
Input/Output and Functions Input/output data from structured/semi-structured file formats (csv, xml, json);. Input data from the Internet (e.g., web scraping);. Defining functions;. Lambdas for functional programming
More Advanced Data Operations Dealing with NA values;. Catching exceptions;. Use of support libraries (e.g., Pandas, Numpy, dfply);. Regular Expressions; . Text analytics
Database Programming – Relational Databases Database system technologies;. Programmatically connecting to databases;. Create/Read/Update/Delete (CRUD) Operations;. SQL Optimization, Indexing and Normalization
Database Programming – NoSQL Databases, Data Lakes NoSQL Overview and Data Models; Document Model, Key-Value Model, Column Family, Aggregates, Graph Model, Triple Stores);. NoSQL Data Modelling Concepts;. Query Languages for Data in NoSQL; NoSQL systems
ETL and Data Pipelines Data Cleaning, wrangling and validation Developing programs for data processing activities (e.g., data extraction, cleaning, merging, aggregation, analysis, reporting) . Data wrangling techniques
Data plotting and visualisation Plotting and visualisation principles;. Plotting libraries (e.g., Matplotlib, ggplot);. Dashboard frameworks (e.g., Plotly)
Big Data Programming Challenges associated with programming for big data;. Parallelism for computational processes;. Distributed computing platforms for big data processing
Design patterns Data science patterns;. Design patterns for big data processing
Data streaming Stream input sources, live data stream, window-based transformations, combination of batch and stream computations
Productionalizing Data Analytics Tools, testing, Portable Format for Analytics (PFA),. Integrating machine learning models into production,. Data Security

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Formative Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	Yes		

Assessment Description:
Formative assessment will be provided on the in-class individual or group activities. Feedback will be provided in written or oral format, or on-line through Moodle. In addition, in class discussions will be undertaken as part of the practical approach to learning.

Assessment Type:	Continuous Assessment	**% of total:**	30
Assessment Date:	n/a	**Outcome addressed:**	4,5
Non-Marked:	No		

Assessment Description:
This assessment will consist of practical tasks in the form of an in-class test. This will assess learners' knowledge and competences on programmatically processing and analysing datasets including operations with database connectivity.

Assessment Type:	Project	**% of total:**	70
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	No		

Assessment Description:
The terminal assessment will consist of a project that will evaluate all learning outcomes. Learners will have to identify and carry out a series of analyses of a large dataset (or a collection of large datasets that are somehow related or complement each other), utilising appropriate programming languages, tools and techniques (e.g., data wrangling) for data preparation, programming environments and database systems. The final submission will consist of an academic research paper style report as well as the implemented data analytics artefact.

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 the Database and Analytics Programming 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.

H9DAP: Database and Analytics Programming

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Classroom & Demonstrations (hours)	24	Every Week	24.00
Tutorial	Other hours (Practical/Tutorial)	24	Every Week	24.00
Independent Learning	Independent learning (hours)	202	Every Week	202.00
			Total Weekly Contact Hours	48.00

Module Resources

Recommended Book Resources

Todd Morley. (2019), Data Science Design Patterns, 1st edition. Addison-Wesley Professional, p.512, [ISBN: 9780134000053].

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

Thomas A. Runkler. (2012), Data Analytics, Springer Science & Business Media, p.137, [ISBN: 978-3834825889].

Wes McKinney. (2017), Python for Data Analysis, O'Reilly Media, p.550, [ISBN: 978-1491957660].

Supplementary Book Resources

Paul Teetor. (2011), R Cookbook, "O'Reilly Media, Inc.", p.413, [ISBN: 978-0596809157].

Nathan Marz,James Warren. (2015), Big Data, Manning Publications Company, p.328, [ISBN: 978-1617290343].

Tom White. Hadoop, O'Reilly Media, [ISBN: 9781491901687.].

Donald Miner,Adam Shook. (2016), Mapreduce Design Patterns, O'Reilly Media, p.275, [ISBN: 9781491927922].

This module does not have any article/paper resources

Other Resources

[website], MIT Open Courseware. (2016), Introduction to Computational Thinking and Data Science, https://ocw.mit.edu/courses/electrical-e ngineering-and-computer-science/6-0002-i ntroduction-to-computational-thinking-an d-data-science-fall-2016/

[website], DataCamp, Learn R, Python & Data Science Online, https://www.datacamp.com/

Discussion Note:	