



## Data Engineering Course Descriptor

Course Title	Data Engineering	Faculty	EDGE Innovation Unit (London)
Course Code	NCHMV6108	Course Leader	Marta Rodriguez-Martinez
Credit points	15	Teaching Period	6 weeks
FHEQ level	6	Date approved	March 2021
Compulsory/Optional	Compulsory		
Prerequisites	None		
Co-requisites	None		

### COURSE SUMMARY

This course will draw on knowledge of Python, SQL, and database management from the previous modules. The course will cover core aspects of data engineering, focusing on designing, building, testing and maintaining data infrastructure. Learners will identify how large-scale processing systems (including cloud-based solutions) can help to drive insight and improve data quality, reliability, and efficiency in their organisation. Learners will expand their knowledge of databases in the context of data lakes and data warehouses. They will gain experience in pipeline design, create automated workflows, and get familiar with techniques for testing, scheduling, and monitoring workflows to achieve performance optimization.

### COURSE AIMS

- Teach learners the distinction between various data storage solutions.
- Train learners to schedule, automate, and monitor data pipelines.
- Train learners to use ETL and ELT processes to build a data warehouse.

### LEARNING OUTCOMES

On successful completion of the course, learners will be able to:

#### KNOWLEDGE AND UNDERSTANDING

- K1c Critically compare traditional and cloud-based warehouse solutions, in terms of their resources, data storage, processing, and security.

- K3c Critically evaluate the uses (or opportunities for the use) of analytical and transactional databases, data lakes, and data warehouses in their organisation..
- K3c Critically evaluate how advanced data engineering solutions can support and enhance data analytics in their organisation.

### **SUBJECT SPECIFIC SKILLS**

- S1c Critically appraise approaches to designing an ETL pipeline to migrate data from multiple sources into a data warehouse, and prepare for further analysis.
- S1c Critically evaluate approaches to configuring and scheduling automated data pipelines.
- S3c Apply appropriate techniques to test a pipeline, and implement data quality checks and monitoring procedures.

### **TRANSFERABLE AND PROFESSIONAL SKILLS**

- T2c Apply problem solving and analytical skills to investigate a problem and find a solution.
- T3c Critically evaluate existing research and extract important information in order to solve a problem.

### **TEACHING AND LEARNING**

The course learning and teaching hours will be structured as follows:

- Seminar = 15 hours
- Applied Seminar = 7.5 hours
- Coaching session = 1 hour
- Academic Drop-ins = 4 hours
- Reading and Assessment Prep = 32.5 hours
- Application of learning = 90 hours

Total = 150 hours

### **ASSESSMENT**

#### **FORMATIVE**

Learners will be formatively assessed during the course through a range of techniques. During Seminars, applied learning seminars and one to one coaching, questioning and quizzing are used as formative assessment. The purpose of the formative assessment is to provide learners with developmental feedback to support their learning.

**SUMMATIVE**

AE	Assessment Type	Weighting	Online submission	Duration	Length
1	Set exercises	60%	Yes	Requiring on average 20-30 hours to complete	2500 words +/- 10%
2	Written assignment	40%	Yes	-	1500 words +/- 10%

The summative assessment will be assessed in accordance with the assessment aims set out in the Programme Specification.

**FEEDBACK**

Learners will receive formal feedback in a variety of ways. For summative assessments, feedback is written and provided within 20 days of submission, and for formative assessment it could include verbal or written feedback. Learners will also have a progress review each quarter with their Multiverse Coach and their Line Manager. These reviews are an opportunity to discuss progress to date both in their role and on the degree programme.

For summative exams, feedback is provided through generic internal examiners' reports, which are posted on the Applied learning platform.

**INDICATIVE READING**

Note: Comprehensive and current reading lists for courses are produced annually in the Course Syllabus or other documentation provided to learners; the indicative reading list provided below is used as part of the approval/modification process only.

**BOOKS**

Buelta, J., 2020. *Python Automation Cookbook* / Buelta, Jaime. 1st ed.

Silverston, L., 2011. *The data model resource book, Volume 1: A library of universal data models for all enterprises*. John Wiley & Sons.

Krishnan, K., 2019. *Building Big Data Applications*. Academic Press.

**INDICATIVE TOPICS**

Learners will study the following topics:

- ETL vs ELT in data warehousing
- Data Lakes and Warehouses
- Design and implement data pipelines
- Traditional and cloud-based solutions
- Monitoring and testing pipelines
- Automated data pipelines

<b>Title: NCHMV6108 Data Engineering Course Descriptor</b>					
<b>Approved by: Academic Board</b>					
Version number	Date approved	Date published	Owner	Location	Proposed next review date
1.0	March 2021	March 2021	Marta Rodriguez-Martinez	Multiverse website	March 2026
<b>Modifications (As per AQF4)</b>					
Version number	Date approved	Date published	Modification (including category number)		