

## University of the West of Scotland

## Module Descriptor

Session: 2024/2025

<b>Title of Module: Data Engineering</b>			
<b>Code: COMP10XXX</b>	<b>SCQF Level: 10 (Scottish Credit and Qualifications Framework)</b>	<b>Credit Points: 20</b>	<b>ECTS: 10 (European Credit Transfer Scheme)</b>
<b>School:</b>	School of Computing, Engineering and Physical Sciences		
<b>Module Co-ordinator:</b>	TBC		
<b>Summary of Module</b>			
<p>Data engineering has emerged as a critical discipline in modern software system design. With the exponential growth of the amount of data collected and processed, this discipline addresses data challenges ranging from scalability and consistency to efficiency and maintainability of data infrastructure and pipelines. In today's data-rich environment, professionals across various domains, including software engineering, data science, and analytics need to have a comprehensive understanding of data engineering practices. As a role, data engineers are responsible for the design and management of the infrastructure and pipelines that collect, transform, and store data so that it can then be used in downstream tasks by data analysts, data scientists, business intelligence professionals, and other stakeholders.</p> <p>This module aims to equip students with the skills to plan and construct data systems that effectively meet organizational and customer needs. By exploring the data engineering lifecycle, students will gain insights into selecting and integrating diverse technologies to cater to downstream data consumers.</p> <p>The syllabus will include the following topics:</p> <ul style="list-style-type: none"> <li>○ Data Engineering Overview <ul style="list-style-type: none"> <li>○ Definition, Roles, Activities, and Responsibilities</li> <li>○ Introduction to the Data Engineering Lifecycle</li> <li>○ Phases: Generation, Storage, Ingestion, Transformation, Data Serving</li> </ul> </li> <li>○ Data Architecture <ul style="list-style-type: none"> <li>○ Principles</li> <li>○ Concepts: Domains and Services, Distributed Systems, Coupling</li> <li>○ Types: Data Warehouse, Data Lake, Data Mesh</li> </ul> </li> <li>○ Data Engineering Life Cycle <ul style="list-style-type: none"> <li>○ Data Generation Sources: APIs, Files, Databases</li> <li>○ Data Formats: JSON, XML, and Language-Specific Formats</li> <li>○ Data Storage Systems: File, Block, Object, Cache, Streaming Storage</li> <li>○ Data Storage Abstractions: Data Warehouse, Lake, Lakehouse, OLTP, OLAP</li> </ul> </li> </ul>			

- Data Ingestion: Key Considerations (Frequency, Sync/Async, Throughput, Reliability, Security), Batch and Message/Stream Ingestion, Key Methods (APIs, Databases, SSH, SFTP, Queue/Event-Streaming)
- Building Data Pipelines
  - ETL and ELT Overview
  - Extracting Data from Various Sources (NoSQL, SQL DB, APIs)
  - Loading Data into Data Warehouses
  - Transforming and Validating Data
  - Best Practices
- Queries, Modeling, and Transformation
  - Queries: Definition, Lifecycle, Queries on Streaming Data
  - Data Model: Definition, Types, Modeling Streaming Data
  - Transformations: Batch Transformation, Materialized Views, Query Virtualization, Streaming Transforms, Processing
- Data Serving
  - Considerations: Trust, Use Cases, Users, Products
  - Analytics: Business, Operational, Embedded
  - Ways to Serve Data: File Exchange, Database, Streaming
- Security and Privacy
  - People, Processes, and Technology to Enforce Security and Privacy

Throughout the practical sessions of this module, students will develop and deploy components of data pipelines across the data engineering lifecycle into suitable cloud, on-premises, or local systems. They will also explore the usage of various mainstream industry tools that enable them to implement the core principles of data engineering.

This module will work to develop a number of the key 'I am UWS' Graduate Attributes to make those who complete this module:

- Universal: critical thinker; analytical; and research-minded
- Work Ready: motivated; digitally literate; effective communicator; and ambitious
- Successful: innovative; resilient; and transformational

#### Module Delivery Method

Face-To-Face	Blended	Fully Online	HybridC	Hybrid 0	Work-Based Learning
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

*If this module is delivered within the BSc (Hons) IT Software Development Programme the 'Blended' module delivery method applies.*

**See Guidance Note for details.**

#### Campus(es) for Module Delivery

The module will **normally** be offered on the following campuses / or by Distance/Online Learning: (Provided viable student numbers permit) (tick as appropriate)

Paisley:	Ayr:	Dumfries:	Lanarkshire:	London:	Distance/Online Learning:	Other:
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Add name

### Term(s) for Module Delivery

(Provided viable student numbers permit).

Term 1	<input checked="" type="checkbox"/>	Term 2	<input type="checkbox"/>	Term 3	<input type="checkbox"/>
--------	-------------------------------------	--------	--------------------------	--------	--------------------------

### Learning Outcomes: (maximum of 5 statements)

These should take cognisance of the SCQF level descriptors and be at the appropriate level for the module.

At the end of this module the student will be able to:

L1	Demonstrate a detailed understanding of the fundamental principles and concepts of data engineering, including the roles, activities, and responsibilities of a data engineer
L2	Explain the phases of the data engineering lifecycle, such as data generation, storage, ingestion, transformation, and data serving, and critically appraise the key considerations and technologies associated with each phase.
L3	Develop and deploy data pipelines using suitable tools and principles of data engineering given a practical scenario
L4	Critically evaluate and compare different data storage systems, ingestion methods, and data serving techniques, considering factors such as reliability, security, and scalability and analyse the effectiveness of various data engineering approaches.

### Employability Skills and Personal Development Planning (PDP) Skills

<b>SCQF Headings</b>	During completion of this module, there will be an opportunity to achieve core skills in:
Knowledge and Understanding (K and U)	<p>SCQF Level <b>10</b></p> <p>A critical understanding of the principal theories, concepts and principles in the domain of data engineering.</p> <p>Detailed knowledge and understanding of how to create and assess a data pipeline.</p>
Practice: Applied Knowledge and Understanding	<p>SCQF Level <b>10</b></p> <p>Apply knowledge, skills and understanding in executing the development and implementation of data pipelines along the lifecycle of data engineering.</p>

Generic Cognitive skills	<p>SCQF Level <b>10</b></p> <p>Identify problems, analyse results, and interpret common error messages to solve problems in a logical manner</p> <p>Consolidate knowledge, concepts, skills, practices and methodology in the production of data engineering solutions.</p>	
Communication, ICT and Numeracy Skills	<p>SCQF Level <b>10</b></p> <p>This subject area is entirely computer based so ICT skills feature heavily in the practice of the subject area. Communicating effectively and appropriately in commercial style, numerate, written reports produced using standard office ICT equipment and software.</p>	
Autonomy, Accountability and Working with others	<p>SCQF Level <b>10</b></p> <p>Exercise autonomy and initiative to implement independently parts of data engineering solutions at a professional level</p>	
<b>Pre-requisites:</b>	Before undertaking this module the student should have undertaken the following:	
	<b>Module Code:</b>	<b>Module Title:</b>
	<b>Other:</b>	
<b>Co-requisites</b>	<b>Module Code:</b>	<b>Module Title:</b>

\*Indicates that module descriptor is not published.

<b>Learning and Teaching</b>	
<b>In line with current learning and teaching principles, a 20-credit module includes 200 learning hours, normally including a minimum of 36 contact hours and maximum of 48 contact hours.</b>	
<p><b>Learning Activities</b></p> <p>During completion of this module, the learning activities undertaken to achieve the module learning outcomes are stated below:</p>	<p><b>Student Learning Hours</b> (Normally totalling 200 hours): (Note: Learning hours include both contact hours and hours spent on other learning activities)</p>
Lecture/Core Content Delivery	18
Tutorial/Synchronous Support Activity	6
Laboratory/Practical Demonstration/Workshop	24

Independent Study	152
	200 Hours Total

**\*\*Indicative Resources: (eg. Core text, journals, internet access)**

The following materials form essential underpinning for the module content and ultimately for the learning outcomes:

Bryan, K. and Ransome, T. (2023) Cracking the Data Engineering Interview. Packt Publishing Ltd.

Crickard, P. (2020) Data Engineering with Python. Packt Publishing Ltd.

Densmore, J. (2021) Data Pipelines Pocket Reference. O'Reilly Media.

Kleppmann, M. (2017) Designing Data-Intensive Applications. 'O'Reilly Media, Inc.'

Reis, J. and Housley, M. (2022) Fundamentals of Data Engineering. 'O'Reilly Media, Inc.'

(\*N.B. Although reading lists should include current publications, students are advised (particularly for material marked with an asterisk\*) to wait until the start of session for confirmation of the most up-to-date material)

**Attendance and Engagement Requirements**

In line with the [Student Attendance and Engagement Procedure](#): Students are academically engaged if they are regularly attending and participating in timetabled on-campus and online teaching sessions, asynchronous online learning activities, course-related learning resources, and complete assessments and submit these on time.

For the purposes of this module, academic engagement equates to the following:

Attending all timetabled synchronous classes and engagement with asynchronous learning activities and resources.

**Equality and Diversity**

The University's Equality, Diversity and Human Rights Procedure can be accessed at the following link: [UWS Equality, Diversity and Human Rights Code](#).

Aligned with the overall commitment to equality and diversity stated in the Programme Specifications, the module supports equality of opportunity for students from all backgrounds and with different learning needs. Using our VLE, learning materials will be presented electronically in formats that allow flexible access and manipulation of content (part-time and distant learning students should check with their programme leader for any queries). The module complies with University regulations and guidance on inclusive learning and teaching practice. Specialist assistive equipment,

support provision and adjustments to assessment practice will be made in accordance with UWS policy and regulations.

(N.B. Every effort will be made by the University to accommodate any equality and diversity issues brought to the attention of the School)

### Supplemental Information

<b>Divisional Programme Board</b>	Computing
<b>Assessment Results (Pass/Fail)</b>	Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>
<b>School Assessment Board</b>	Business & Applied Computing
<b>Moderator</b>	TBA
<b>External Examiner</b>	A Jindal
<b>Accreditation Details</b>	n/a
<b>Changes/Version Number</b>	1.0

#### Assessment: (also refer to Assessment Outcomes Grids below)

Assessment for this module consists of a written report and a coursework portfolio of practical work.

Assessment 1: Written report analysing a case study/problem, requiring students to describe and propose data engineering solutions aligned with the lifecycle. (40%)

Assessment 2: A coursework assessment that requires a group of students to act as a development team and develop, document, and evaluate a data engineering pipeline on a cloud platform, demonstrating their practical application of data engineering concepts and proficiency in using taught technologies. (60%)

(N.B. (i) **Assessment Outcomes Grids** for the module (one for each component) can be found below which clearly demonstrate how the learning outcomes of the module will be assessed.

(ii) An **indicative schedule** listing approximate times within the academic calendar when assessment is likely to feature will be provided within the Student Module Handbook.)

### Assessment Outcome Grids (See Guidance Note)

Component 1							
Assessment Type (Footnote B.)	Learning Outcome (1)	Learning Outcome (2)	Learning Outcome (3)	Learning Outcome (4)	Learning Outcome (5)	Weighting (%) of Assessment Element	Timetabled Contact Hours
Review/ Article/ Critique/ Paper	x	x				40	-

Component 2							
Assessment Type (Footnote B.)	Learning Outcome (1)	Learning Outcome (2)	Learning Outcome (3)	Learning Outcome (4)	Learning Outcome (5)	Weighting (%) of Assessment Element	Timetabled Contact Hours
Portfolio of practical work			x	x		60	-

<b>Combined Total for All Components</b>						<b>100%</b>	<b>- hours</b>
--	--	--	--	--	--	-------------	----------------

**Change Control:**

What	When	Who
Further guidance on aggregate regulation and application when completing template	16/01/2020	H McLean
Updated contact hours	14/09/21	H McLean
Updated Student Attendance and Engagement Procedure	19/10/2023	C Winter
Updated UWS Equality, Diversity and Human Rights Code	19/10/2023	C Winter
Guidance Note 23-24 provided	12/12/23	D Taylor
General housekeeping to text across sections.	12/12/23	D Taylor

Version Number: MD Template 1 (2023-24)