INTERNATIONAL QUALIFICATIONS

AND ASSESSMENT CENTRE (IQAC)

«Approved»

2025 y. «16» 01

Programme	**Level 6 Diploma in Data Science**		
Unit Number/ Unit Title	UNIT 3 CLOUD COMPUTING AND DATA ENGINEERING		
Cohort Code:	L06CCDE-U3		
Unit Level	Level 6		
Total GLH	Total qualification time 200/ Total Guided learning hours 90/ Self-guided learning hours 110		
Credits	20 CATS/ 10 ECTS		
Lecturer			
Start Date		**End Date**	

Unit Aims	This unit explores cloud-based data processing environments and DevOps workflows in scalable data engineering, covering platforms such as AWS, Azure, and GCP. Students will examine cloud-based data processing environments and the principles of DevOps workflows within scalable data engineering. The unit will cover major platforms like AWS, Azure, and Google Cloud Platform (GCP), focusing on their tools and services for data storage, processing, and analytics. By the end of the course, students will be equipped to design and implement efficient data pipelines and understand best practices for integrating DevOps in data engineering processes.
Differentiation Strategies *(e.g. planned activities or support for individual learners according to their needs)*	The total number of students to be in the lesson is approximately 20. This is a multicultural group of students predominantly between the ages of 24 – 45, with numerous ethnic, gender, and creed background. These are UK academic level 5 students; hence it is assumed that they have practical, theoretical, or technological knowledge and understanding of a subject or field of work to find ways forward in broadly defined, complex contexts. These students must be able to

	generate information, evaluate, synthesise the use information from a variety of sources. Various approaches to addressing the various identified students needs will be adopted throughout the lesson. Such will include:- 1. Progressive tasks 2. Digital resources 3. Verbal support 4. Variable outcomes 5. Collaborative learning 6. Ongoing assessment 7. Flexible-pace learning
Equality & Diversity	Variety of teaching techniques will be employed to ensure that the needs of each individual learner are met.
Safeguarding & Prevent	Safeguarding policies and the Prevent duty are strictly observed to ensure the safety, well-being, and inclusivity of all students and staff.
Health & Safety	SIRM H&S policies will be maintained.
Learning Resources	**Teaching and Learning Materials** • Amazon Web Services. (2022). AWS Certified Solutions Architect Study Guide. • Mahmoud, R. (2020). Cloud Computing: Concepts, Technology & Architecture. Prentice Hall. • GCP & Azure documentation. • Turnbull, J. (2018). The Docker Book.

Learning Outcome	Assessment Criteria
LO1. 1. Describe the architecture of cloud platforms.	Written Report: 1.1 Compare IaaS, PaaS, and SaaS models. 1.2 Identify key services (e.g., S3, EC2, Azure Blob).
LO2. 2. Deploy cloud-based data storage and compute services.	Lab Assignment: 2.1 Implement a cloud pipeline for data ingestion. 2.2 Launch compute instances for data analysis.
LO3. 3. Apply DevOps principles to data science workflows.	Group Project: 3.1 Build CI/CD pipelines using tools like GitHub Actions. 3.2 Containerise workflows using Docker.
LO4. 4. Manage data in distributed storage systems.	Case Study: 4.1 Compare file systems (e.g., HDFS vs S3). 4.2 Ensure scalability and security in data storage.

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Cloud Computing Models**	**Cloud Computing Models** IaaS vs. PaaS vs. SaaS: Use cases and trade-offs	LO1: Cloud Architecture Fundamentals	
2.	**AWS Core Services**	**AWS Core Services** EC2, S3, Lambda, RDS – Architecture and pricing	LO1: Cloud Architecture Fundamentals	
3.	**Azure Services**	**Azure Services** Blob Storage, VMs, Functions, Cosmos DB	LO1: Cloud Architecture Fundamentals	
4.	**GCP Services**	**GCP Services** Compute Engine, BigQuery, Cloud Functions	LO1: Cloud Architecture Fundamentals	
5.	**Multi-Cloud Strategie**	**Multi-Cloud Strategie** Hybrid cloud, interoperability challenges	LO1: Cloud Architecture Fundamentals	
6.	**Cloud Storage Options**	**Cloud Storage Options** Object (S3), block (EBS), file (EFS) storage	LO2: Data Storage & Compute Deployment	
7.	**Data Ingestion Pipelines**	**Data Ingestion Pipelines** AWS Kinesis/GCP PubSub for streaming data	LO2: Data Storage & Compute Deployment	
8.	Half-Term Exam	- Review of LO1 topics - Practice questions and mock assessment - **Half-term assessment** based on LO1 (theory)	LO1 LO2	
9.	**Batch Processing**	**Batch Processing** AWS Glue/Azure Data Factory for ETL	LO2: Data Storage & Compute Deployment	

10.	**Serverless Compute**	**Serverless Compute** AWS Lambda/Google Cloud Functions	LO2: Data Storage & Compute Deployment	
11.	**Cost Optimization**	**Cost Optimization** Spot instances, auto-scaling, monitoring tools	LO2: Data Storage & Compute Deployment	
12.	**CI/CD Fundamentals**	**CI/CD Fundamentals** GitHub Actions, Jenkins, GitLab CI	LO3: DevOps for Data Science	
13.	**Infrastructure as Code (IaC)**	**Infrastructure as Code (IaC)** Terraform, AWS CloudFormation	LO3: DevOps for Data Science	
14.	Final Exam Preparation & Review	- Comprehensive review of all learning outcomes - Practice questions and revision of key topics		
15.	Final Exam	- **Final-term assessment** covering all learning outcomes (theory and practical elements)		
16.	Feedback & Reflection	- Review of final exam - Individual feedback on performance - Reflective discussion on key learning points		
17.	**Containerization Basics**	**Containerization Basics** Docker images, Dockerfiles, registries	LO3: DevOps for Data Science	
18.	**Orchestration with Kubernetes**	**Orchestration with Kubernetes** Pods, deployments, services	LO3: DevOps for Data Science	
19.	**MLOps Workflows**	**MLOps Workflows** Model versioning, testing, deployment	LO3: DevOps for Data Science	

20.	**Distributed File Systems**	**Distributed File Systems** HDFS vs. S3 vs. GCS architecture	LO4: Distributed Data Management	
21.	**Data Lakes vs. Warehouses**	**Data Lakes vs. Warehouses** Delta Lake, Snowflake, Redshift	LO4: Distributed Data Management	
22.	**Partitioning Strategies**	**Partitioning Strategies** Time-based, hash, range partitioning	LO4: Distributed Data Management	
23.	Half-Term Exam	**Project** End-to-end cloud data pipeline deployment		
24.	**Security & Compliance**	**Security & Compliance** Encryption, IAM, GDPR/HIPAA compliance	LO4: Distributed Data Management	
25.	**Performance Optimization**	**Performance Optimization** Caching, query optimization, indexing	LO4: Distributed Data Management	
26.	**Real-World Case Study**	**Real-World Case Study** Netflix's data pipeline architecture	LO5: Capstone & Emerging Trends	
27.	**Edge Computing**	**Edge Computing** IoT data processing with AWS Greengrass	LO5: Capstone & Emerging Trends	
28.	**Data Mesh Architecture** and **Ethics in Cloud Data**	**Data Mesh Architecture** Domain-oriented decentralized data **Ethics in Cloud Data** Data sovereignty, environmental impact	LO5: Capstone & Emerging Trends	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4	LO1, LO2, LO3, LO4	
30.	Final Exam		LO1, LO2, LO3, LO4	