Qualification Specification:

OCN NI Level 5 Certificate in Data Engineering Processing Techniques

- Qualification No: 610/6494/4

Version: 1.0

1. Specification Updates

Key changes have been listed below:

Section	Detail of change	Version and date of Issue
Specification	Newly developed qualification	V1.0 July 2025

2. Contents

3. Introduction to Open College Network Northern Ireland (OCN NI)

The Open College Network Northern Ireland (OCN NI) is a UK recognised awarding organisation based in Northern Ireland. We are regulated by CCEA Regulation to develop and award regulated professional and technical (vocational) qualifications from Entry Level up to and including Level 5 across all sector areas. In addition, OCN NI is also regulated by Ofqual to award qualifications in England.

OCN NI is also an educational charity that advances education by developing nationally recognised qualifications and recognising the achievements of learners. We work with centres such as Further Education Colleges, Private Training Organisations, Voluntary & Community Organisations, Schools, SME's and Public Sector bodies to provide learners with opportunities to progress into further learning and/or employment. OCN NI's Strategic Plan can be found on the OCN NI website www.ocnni.org.uk.

For further information on OCN NI qualifications or to contact us, you can visit our website at www.ocnni.org.uk. The website should provide you with details about our qualifications, courses, contact information, and any other relevant information you may need.

OCN NI Contact Details

Open College Network Northern Ireland
Sirius House
10 Heron Road
Belfast
BT3 9LE

Phone: 028 90 463990
Website: **www.ocnni.org.uk**
Email: info@ocnni.org.uk

4. About this Specification

This specification details OCN NI's specific requirements for the delivery and assessment of the **OCN NI Level 5 Certificate in Data Engineering Processing Techniques .**

This specification will provide guidelines for centres to ensure the effective and correct delivery of this qualification. OCN NI qualification specifications are based on research and engagement with the practitioner community to ensure they provide appropriate skills and knowledge for learners.

The qualification specification will detail the following aspects of the OCN NI Level 5 Certificate in Data Engineering Processing Techniques

- **Qualification Features**: this includes the key characteristics and features of this qualification, such as its intended audience, purpose, and credit value.

- **Centre Requirements:** this details the prerequisites and obligations that centres must fulfil to be eligible to deliver and assess this qualification. These include guidelines on staff qualifications, resources, and required procedures.

- **Structure and Content:** this details the structure and content of the qualification including units, and any specific content that learners will be required to study.

- **Assessment Requirements:** this details assessment criteria and assessment methods for this qualification, ensuring that summative assessment approaches are clear.

- **Quality Assurance:** the quality and consistency of delivery and assessment of this qualification are of paramount importance to OCN NI. The mandatory quality assurance arrangements including processes for internal and external quality assurance that all centres offering this qualification must adhere to are detailed.

- **Administration:** guidance on the administrative aspects of delivering this qualification, including registration, certification and record-keeping.

- Reference to other handbooks and policies as appropriate to the qualifications.

It is important to note that OCN NI will communicate any significant updates or changes to this specification in writing to our centres. Additionally, we will make these changes available on our official website at www.ocnni.org.uk.

To stay current, please refer to the online version of this specification as it is the most authoritative and up-to-date publication. Be aware that downloaded and printed copies may not reflect the latest revisions.

4.1 Additional Support

OCN NI offers a comprehensive range of support services designed to assist centres in meeting the delivery and quality assurance requirements of OCN NI qualifications. These services include:

- **Learner Assessment Booklets**: These booklets are created to assist learners in demonstrating the fulfilment of assessment criteria and organising the quality assurance prerequisites for each individual unit.

- **Specimen Assessment Materials**: These booklets are created to assist learners in demonstrating the fulfilment of assessment criteria and organising the quality assurance prerequisites for each individual unit.

- **Qualification Support Pack**: A support pack has been developed to support centres in the delivery of this qualification. The pack includes planning and assessment templates, guides to best practice, etc.

- **Professional Development for Educators**: OCN NI provides opportunities for professional development tailored to meet the various needs of practitioners and quality assurance staff. Centres can join our training sessions, available in both face-to-face and online formats, or explore a wealth of training materials by visiting www.ocnni.org.uk

- **OCN NI Subject Advisors**: Our team of subject advisors offers vital information and support to centres. They provide guidance on specification details, non-exam assessment advice, updates on resource developments, and various training opportunities. They actively engage with subject communities through an array of networks to facilitate the exchange of ideas and expertise, to support practitioners to provide quality education programs to learners.

All centres can access information, support and guidance to support the delivery and quality assurance of this qualification by contacting their designated Business Development Advisor or by contacting us on Contact Us | OCN NI

5. About this Qualification

5.1 Qualification Regulation Information

OCN NI Level 5 Certificate in Data Engineering Processing Techniques
Qualification Number: 610/6494/4

Operational start date: 15/09/2025
Review date: 14/09/2030

The qualification's operational start and end dates define the regulated qualification's lifecycle. The operational end date is the final date for learner registration, while learners have until the certificate end date to complete the qualification and receive their certificates.

It is important to note that all OCN NI regulated qualifications are listed on the Register of Regulated Qualifications (RQF), which can be found at Ofqual Register. This register is maintained by Ofqual in England and CCEA Regulation in Northern Ireland. It contains information about qualifications that are regulated and accredited. It is a key resource for learners, employers, and educational institutions to verify the status and recognition of qualifications.

Centres must adhere to administrative guidelines diligently, with special attention to the fact that fees, registration, and certification end dates for the qualification may be subject to changes. It is a centre's responsibility to make itself aware of updates on any modifications to ensure compliance with the latest requirements. OCN NI provides centres with timely updates through various channels including website, newsletters and through this specification. Information on qualification fees can be found on the Centre Login section of the OCN NI website www.ocnni.org.uk .

5.2 Sector Subject Area

A subject sector area is a specific category used to classify academic and vocational qualifications. Subject sector areas are part of the educational and qualifications framework to organise and categorise qualifications. The sector subject for this qualification is:

Subject Area: 6.1 ICT for practitioners

NOS:
TECDT80851 Design Data Processing Systems
TECDT80841 Build and Implement data Processing Systems
TECDT130141B Develop Data Models
TECIS805301 Assist in Developing and Validating Machine Learning Solutions

5.3 Grading

Grading for this qualification is pass/fail.

5.4 Qualification's Aim and Objectives

Qualification's Aim

The aim of the OCN NI Level 5 Certificate in Data Engineering Processing Techniques is to provide individuals with the knowledge and skills to carry out data engineering activities including using relevant tools, and strategies to collect, store, prepare, analyse, and visualise data for use in analytics and machine learning (ML) applications.

Qualification's Objectives

The objectives of the OCN NI Level 5 Certificate in Data Engineering Processing Techniques are to enable learners to understand:

- the role and importance of data science in data-driven organisations
- how the elements of data influence decisions about the infrastructure of a data pipeline
- how to evaluate data pipelines by using cloud services and select appropriate data storage options
- how to process different types of data
- how to work with machine learning pipelines

5.5 Target Learners

This qualification is targeted at individuals who are currently in or wish to progress into data engineering roles which focus on using cloud-based databases.

5.6 Entry Requirements

The learners should be at least 18 years of age and have a minimum level 3 qualification.

5.7 Progression

Progression from OCN NI Level 5 Certificate in Data Engineering Processing Techniques is to higher level qualifications in the areas of information technology, data analytics, information technology project management and/or systems analysis and design or into employment in these areas.

5.8 Delivery Language

This qualification is exclusively available in English. If there is a desire to offer this qualification in Welsh or Irish (Gaeilge), we encourage you to get in touch with OCN NI. They will assess the demand for such provisions and, if feasible, provide the qualification in the requested language as appropriate.

6. Centre Requirements for Delivering this Qualification

6.1 Centre Recognition

New and existing OCN NI recognised centres must apply for and be granted approval to deliver this qualification prior to the commencement of delivery.

6.2 Qualification Approval

Once a centre has successfully undergone the Centre Recognition process, it becomes eligible to apply for qualification approval. The centre's capability to meet and sustain the qualification criteria will be assessed. Throughout the qualification approval process, OCN NI will aim to ensure that:

- centres possess suitable physical resources (e.g., equipment, IT, learning materials, teaching rooms) to support qualification delivery and assessment
- centre staff involved in the assessment process have relevant expertise and/or occupational experience
- robust systems are in place for ensuring ongoing professional development for staff delivering the qualifications
- centres have appropriate health and safety policies concerning learner equipment use
- qualification delivery by centres complies with current equality and diversity legislation and regulations

6.3 Centre Staffing

To offer this qualification centres are mandated to establish the following roles as a minimum, although a single staff member may serve in more than one capacity*:

- Centre contact
- Programme Co-ordinator
- Assessor
- Internal Quality Assurance (IQA)

*Note: An individual cannot serve as an IQA for their own assessments.

6.4 Tutor Requirements

Tutors responsible for delivering this qualification are expected to possess a high degree of occupational competency. They should meet the following criteria:

- **Occupational Competency:** Tutors should demonstrate a clear understanding of the subject matter, including up-to-date knowledge. They should also have a minimum of one year's relevant experience in this area. This competence should enable them to effectively impart knowledge and practical skills to learners.

- **Qualifications:** Tutors should hold qualifications at a level that is at least one level higher than the qualification they are teaching. This ensures that they have the necessary academic foundation to provide in-depth guidance and support to learners.

These requirements collectively ensure that learners receive instruction from highly qualified and experienced instructors, thereby enhancing the quality and effectiveness of their educational experience.

6.5 Assessor Requirements

The assessment of this qualification takes place within the centre and is subjected to OCN NI's rigorous quality assurance procedures. The achievement of individual units is based on the criteria defined in each unit.

Assessors play a pivotal role in ensuring the validity and fairness of assessments. They are required to meet the following criteria:

- **Occupational Competency:** Assessors should possess a high degree of occupational competency in the relevant subject matter. This expertise enables them to accurately evaluate and measure a learner's knowledge and skills. Additionally, they should hold qualifications at a level that is at least one level higher than the qualification they are assessing, ensuring their in-depth understanding of the subject matter.

- **Assessment Expertise:** Assessors should have direct or related experience in the field of assessment. This includes knowledge of best practices in designing, conducting, and grading assessments. Their expertise ensures that assessments are both fair and valid.

- **Assessors Qualification:** Assessors should hold or be currently undertaking a recognised assessor's qualification; or must have attended the OCN NI Assessment Training.

- **Comprehensive Assessment Oversight:** Assessors are responsible for evaluating all assessment tasks and activities comprehensively. They must thoroughly review and assess each element to ensure a fair and accurate representation of a learner's skills and knowledge.

These rigorous requirements uphold the quality and integrity of the qualification's assessment process, ensuring that learners receive a fair and reliable evaluation of their competencies.

6.6 Internal Quality Assurer Requirements

The Internal Quality Assurer plays a crucial role in the centre's internal quality assurance processes. The centre must designate a skilled and trained IQA who assumes the role of an internal quality monitor responsible for verifying the delivery and assessment of the qualifications.

The Internal Quality Assurer for this qualification must meet the following criteria:

- **IQA Expertise:** IQA should have direct or related experience in the field of internal assurance and have at least one year's occupational experience in the areas they are internally quality assuring. This includes knowledge of best practices in designing, conducting, and grading assessments. Their expertise ensures that assessments are both fair and valid.
- **IQA Qualification:** IQA should hold or be currently undertaking a recognised IQA qualification or must have attended the OCN NI IQA Training.
- **Thorough Evaluation of Assessment Tasks and Activities:** IQAs are tasked with conducting in-depth reviews and assessments of all assessment tasks and activities. Their responsibility is to ensure a comprehensive and meticulous oversight of each element to guarantee a just and precise reflection of a learner's abilities and knowledge and to ensure that all assessment and quality assurance requirements are fulfilled.

7. Qualification Structure

7.1 Qualification Purpose

The OCN NI Level 5 Certificate in Data Engineering Processing Techniques is designed to equip learners with the advanced knowledge and practical skills to perform data engineering activities including using relevant tools, and strategies to collect, store, prepare, analyse, and visualise data for use in analytics and machine learning (ML) applications. Learners will also gain an understanding of the role and importance of data science in data-driven organisations and how the elements of data influence decisions about the infrastructure of a data pipeline and be able to evaluate data pipelines and storage options including an emphasis on machine learning applications.

7.2 Qualification Level

In the context of the OCN NI Level 5 Certificate in Data Engineering Processing Techniques it is essential to understand the significance of qualification levels, as they play a pivotal role in assessing the depth and complexity of knowledge and skills required for successful attainment. This qualification aligns with Level 5, which signifies a advanced level of difficulty and intricacy. It's important to note that qualification levels in the educational framework range from Level 1 to Level 8, complemented by three 'entry' levels, namely Entry 1 to Entry 3.

7.3 Qualification Size

Total Qualification Time (TQT)

This represents the total amount of time a learner is expected to spend to complete the qualification successfully. It includes both guided learning hours (GLH) and independent study or additional learning time.

Guided Learning Hours (GLH)

These are the hours of guided instruction and teaching provided to learners. This may include classroom instruction, tutorials, or other forms of structured learning.

OCN NI Level 5 Certificate in Data Engineering Processing Techniques	
Total Qualification Time (TQT):	200 hours
Total Credits Required:	20 credits
Guided Learning Hours (GLH):	120 hours

7.4 How to Achieve the Qualification

To achieve the **OCN NI Level 5 Certificate in Data Engineering Processing Techniques** learners must complete one mandatory unit – 20 credits.

8. Assessment Structure

This qualification is assessed through internal assessment and each unit is accompanied by specific assessment criteria that define the requirements for achievement.

8.1 Assessment Guidance: Portfolio

The portfolio for this qualification is designed to provide a comprehensive view of a learner's skills and knowledge. It is a holistic collection of evidence that may include a single piece of evidence that satisfies multiple assessment criteria. There is no requirement for learners to maintain separate evidence for each assessment criterion.

When learners are creating their portfolio, they should refer to the assessment criteria to understand the evidence required. Explanations of command words/verbs used in the assessment criteria can be found in Appendix 1 of this document.

It is essential that the evidence in the portfolio reflects the application of skills in real-world situations. Learners should ensure that they provide multiple examples or references whenever the assessment criteria require it.

8.2 Understanding the Units

The units outlined in this specification establish clear assessment expectations. They serve as a valuable guide for conducting assessments and ensuring quality assurance efficiently. Each unit within this specification follows a consistent structure. This section explains the operational framework of these units. It is imperative that all educators, assessors, Internal Quality Assurers, and other personnel overseeing the qualification review and familiarise themselves with this section to ensure a comprehensive understanding of how these units function.

- **Title:** The title will reflect the content of the unit and should be clear and concise.
- **Level:** A unit can have one of six RQF levels: Entry, One, Two, Three, Four or Five. All units within this qualification are Level 5.
- **Credit Value:** This describes the number of credits ascribed to a unit. It identifies the number of credits a learner is awarded upon successful achievement of the unit. One credit is awarded for the learning outcomes which a learner, on average, might reasonably be expected to achieve in a notional 10 hours of learning.
- **Learning Outcome:** A coherent set of measurable achievements.
- **Assessment Criteria:** These enable a judgement to be made about whether or not, and how well, the students have achieved the learning outcomes.
- **Assessment Guidance and Methods:** These detail the different assessment methods within the unit that may be used.
- **Unit Content:** This provides indicative content to assist in teaching and learning.
- **Scope:** This provides possible teaching content.

9. Qualification Summary by Unit

OCN NI Level 5 Certificate in Data Engineering Processing Techniques
In order to achieve the OCN NI Level 5 Certificate in Data Engineering Processing Techniques the learner must successfully complete one unit.

Total Qualification Time (TQT) for this qualification: 200 hours
Guided Learning Hours (GLH) for this qualification: 120 hours

Unit Reference Number	OCN NI Unit Code	Unit Title	Credit Value	GLH	Level
F/651/7712	CBG804	Data Engineering Processing Techniques	20	120	Five

10. Unit Content

10.1 Data Engineering Processing Techniques

Title	Data Engineering Processing Techniques
Level	Five
Credit Value	20
Guided Learning Hours (GLH)	120
OCN NI Unit Code	CBG804
Unit Reference No	F/651/7712
Learn Direct Code	CL1
Unit purpose and aim(s): This unit will enable the learner to understand data engineering including tools and strategies that are used to collect, store, prepare, analyse, and visualise data for use in analytics and machine learning (ML) applications.	

Learning Outcomes	Assessment Criteria
1. Understand the role and importance of data science in data-driven organisations and application of data strategies.	1.1. Evaluate the role and importance of data science in data-driven organisations in relation to: a) data-driven decisions b) data pipeline and the infrastructure to implement data driven decisions c) the role of data engineers in data-driven organisations 1.2. Summarise and evaluate current data strategies and their application including: a) unifying b) innovating
2. Understand how the elements of data influence decisions about the infrastructure of a data pipeline.	2.1. Evaluate the five 'V's' of data including: a) volume b) velocity c) variety – data types and data sources d) veracity e) value 2.2. Evaluate methods to improve data veracity and value for given industry applications and activities including: a) data transformation b) timestamping c) compliance and governance 2.3. Analyse pipeline planning for given industry applications (such as using sales transaction data, similar data not in a standard format, collecting data).
3. Be able to evaluate data pipeline by using cloud services.	3.1. Evaluate a given developed architected framework in relation to: a) Use of the framework b) evolution of data architectures c) current data architecture in the cloud d) current data architecture pipeline ingestion and storage e) current data architecture pipeline processing and consumption f) streamlining analytics pipelines 3.2. Demonstrate the use of a lab to query data using cloud-based tools.

	3.3. Demonstrate how to evaluate the security of data architecture by considering: a) a cloud security review b) security of analytic workloads c) machine learning security
4. Be able to select a data storage option that matches the requirements and constraints of a given data analytics use case.	4.1. Explain and evaluate how to store data using current data architecture by considering: a) data lake storage and data warehouse storage b) purpose built databases c) storage to support a pipeline d) storage security 4.2. Determine using a lab how to store and analyse data
5. Be able to implement the steps to process structured, semi structured, unstructured data formats in a data pipeline built in the cloud.	5.1. Explain the big data processing concepts including: a) Apache Hadoop b) Spark 5.2. Determine using a LAB and cloud provider of choice, how to process log files in relation to big data processing. E.g. EMR 5.3. Determine using a LAB and cloud provider of choice, how to update dynamic data in place. E.g. Hudi through Apache.
6. Be able to differentiate the characteristics of an ML pipeline and its specific processing steps and analyse data using Cloud tools that are appropriate to a given use case.	6.1. Summarise the key characteristics of Machine Learning (ML) pipelines including: a) concepts b) lifecycle c) framing the ML problem for a business goal d) collecting data e) labelling to train data using targets f) preprocessing data and feature engineering 6.2. Determine using a LAB within a cloud provider of choice to analyse and visualize data using cloud-based tools. E.g. IoT Analytics, Quicksight, Firehose, OpenSearch services and dashboards.

Assessment Guidance

NOS:
TECDT80851 Design Data Processing Systems
TECDT80841 Build and Implement Data Processing Systems
TECDT130141B Develop Data Models
TECIS805301 Assist in Developing and Validating Machine Learning Solutions

The following assessment method/s may be used to ensure all learning outcomes and assessment criteria are fully covered.

Assessment Method	Definition	Possible Content
Portfolio of evidence	A collection of documents containing work undertaken to be assessed as evidence to meet required skills outcomes OR A collection of documents containing work that shows the learner's progression through the course	Learner notes/written work Learner log/diary Peer notes Record of observation Record of discussion

Practical demonstration/assignment	A practical demonstration of a skill/situation selected by the tutor or by learners, to enable learners to practise and apply skills and knowledge	Record of observation Learner notes/written work Learner log
Coursework	Research or projects that count towards a learner's final outcome and demonstrate the skills and/or knowledge gained throughout the course	Record of observation Learner notes/written work Tutor notes/record Learner log/diary
E-assessment	The use of information technology to assess learners' work	Electronic portfolio E-tests

Learning Outcome	Unit title: Data Engineering Processing Techniques
1. Understand the role and importance of data science in data-driven organisations and application of data strategies.	**Scope** **Teaching will cover:** • **Data-Driven Decisions** • Organizations are investing heavily in data and analytics to become more data-driven. • Examples of personal and organizational decision-making using data. • Distinction between data analytics and AI/ML. • Example scenarios: Identifying pictures of dogs, customer relationship management. • Types of analytics: Descriptive, diagnostic, predictive, prescriptive. • Increasing volume of data and decreasing barriers to analysis. • **Data Pipeline – Infrastructure for Data-Driven Decisions** • Basic layers: Collect data, store and process data, build something useful with data. • Design infrastructure by working backwards from the decision to the data. • Layers: Data sources, ingestion, storage, processing, analysis & visualization, predictions and decisions. • Actions on data: Discover, clean, normalize, enrich. • Iterative processing through the pipeline. • **Role of the Data Engineer in Data-Driven Organizations** • Questions for data engineers: Data availability, storage, format, security, transfer mechanisms, data volume, update frequency, speed. • Questions for data scientists: Data insights, evaluation, visualization, tools, frameworks, AI/ML models. • Key takeaways: Collaboration between data engineers and data scientists, iterative design based on data needs. • **Modern Data Strategies** • Three-pronged strategy: Modernize, unify, innovate. • Modernize: Move to cloud-based services, use purpose-built tools, build loosely coupled pipelines. • Unify: Create a single source of truth, democratize access, use data lakes, simplify governance. • Innovate: Use AI/ML for new insights, proactive decision-making, democratize ML usage.
2. Understand how the elements of data influence decisions about the infrastructure of a data pipeline.	**Scope** **Teaching will cover:** • **Common Data Pipeline Questions** • Data engineer: Data availability, storage, format, security, transfer mechanisms, data volume, update frequency, speed. • Data scientist: Data insights, evaluation, visualization, tools, frameworks, AI/ML models. • **The Five Vs of Data** • **Volume**: Size of dataset, new data generation. • **Velocity**: Frequency of new data generation and ingestion. • **Variety**: Types and formats of data, sources. • **Veracity**: Accuracy, precision, trustworthiness of data. • **Value**: Insights derived from data. • **Strategies for Data Value** • Confirm data meets needs. • Evaluate the feasibility of acquiring data. • Match pipeline design to data. • Balance throughput and cost. • Focus on business needs. • Catalog data and metadata. • Implement governance.

	• **Volume and Velocity** • Impact on pipeline layers: ingestion, storage, processing, analysis & visualization. • Examples: Streaming ingestion (clickstream data), batch ingestion (sales transaction data). • Storage decisions: Long-term reporting, short-term fast access. • Processing decisions: Big data processing, streaming analytics. • Analysis/visualization decisions: Historical analysis, real-time IoT data. • **Variety – Data Types** • **Structured**: Rows and columns, well-defined schema (e.g., relational database tables). • **Semistructured**: Elements and attributes, self-describing structure (e.g., CSV, JSON, XML). • **Unstructured**: Files, no predefined structure (e.g., images, movies, clickstream data). • **Variety – Data Sources** • **On-premises databases or file stores**: Controlled by organization, often structured. • **Public datasets**: Aggregated data, often semi structured. • **Events, IoT devices, sensors**: Continual data generation, time-based component. • **Veracity and Value** • Importance of data integrity across pipeline layers. • Examples of data issues: Dated information, missing data, lack of lineage, ambiguity, statistical bias. • Cleaning and transforming data: Define clean, trace errors, change thoughtfully, retain auditable data. • Better approach for analytics: Immutable data, timestamped records, raw data capture. • **Activities to Improve Veracity and Value** • Evaluating data veracity: Questions for data engineers and data scientists. • Best practices for cleaning data: Define clean, trace errors, change thoughtfully, retain auditable data. • Transformation example: Converting field values to common format. • Saving aggregated data vs. raw data records: Benefits of immutable data. • **Maintaining Data Integrity and Consistency** • Secure pipeline layers. • Grant least privilege access. • Implement best practices for data integrity. • Keep audit trails. • Implement data compliance and governance processes. • Maintain a single source of truth. • **Activity: Planning Your Pipeline** • Select a scenario and document ideas about planning a pipeline using the 5 Vs of data. • Use detailed instructions and a worksheet provided.
3. Be able to evaluate data pipeline by using cloud services.	**Scope** **Teaching will cover:** • **Architected Framework and Lenses** • **Well-Architected Framework pillars**: Operational Excellence, Security, Reliability, Performance Efficiency, Cost Optimization, Sustainability. • **Well-Architected Lenses**: Extend guidance to specific domains (E.g. could be Data Analytics Lens, ML Lens). • **Activity**: E.g. Using the Data Analytics Lens to identify cloud best practices for data pipelines. • **Evolution of Data Architectures** • **Application architecture evolution**: Mainframe, client-server, internet 3-tier, cloud-based microservices.

		• **Data stores evolution**: Hierarchical databases, relational databases, nonrelational databases, data lakes, purpose-built cloud data stores. • **Data architectures evolution**: Data warehouses, OLTP vs. OLAP databases, big data systems, lambda architecture and streaming solutions. • **Modern Data Architecture on cloud based systems for example Azure AWS Oracle** • **Key design considerations**: Scalable data lake, performant and cost-effective components, seamless data movement, unified governance. • **Components**: Data lake, data warehousing, log analytics, big data processing, relational databases, nonrelational databases, machine learning. • **Related Cloud services**: E.g. S3, Redshift, OpenSearch Service, EMR, Aurora, DynamoDB, SageMaker. Azure / Oracle also. • **Data movement types**: Outside in, inside out, around the perimeter. • **Governance and discoverability**: E.g. Glue, Lake Formation. • **Modern Data Architecture Pipeline: Ingestion and Storage** • **Ingestion layer**: E.g. Could Match AWS or Azure services to data source characteristics (AppFlow, DMS, DataSync, Kinesis Data Streams, Data Firehose). • **Storage layer**: Durable, scalable storage (E.g. S3, Redshift), metadata catalogs (E.g. Glue Data Catalog, Lake Formation). • **Storage zones**: Landing, raw, trusted, curated. • **Modern Data Architecture Pipeline: Processing and Consumption** • **Processing layer**: Transforms data into a consumable state (SQL-based ELT, big data processing, near real-time ETL). • **Related cloud services**: E.g. Athena, Redshift, QuickSight, SageMaker. • **Streaming Analytics Pipeline** • **Example architecture**: Data sources, ingestion and producers (CloudWatch Events), stream storage (Kinesis Data Streams), stream processing and consumers (E.g. AWS Managed Service for Apache Flink), analysis and visualization (E.g. OpenSearch Service), downstream destinations (E.g. S3, Redshift).
4.	Be able to select a data storage option that matches the requirements and constraints of a given data analytics use case.	**Scope** **Teaching will cover:** • **Module Objectives** • Define storage types in modern data architecture. • Distinguish between different data storage types. • Select appropriate data storage options. • Implement secure storage practices for cloud-based data. • **Storage in the Modern Data Architecture** • Basics of cloud storage and its role in data pipelines. • General related cloud services and their roles in modern data architecture. • **Types of Cloud Storage** • **Block Storage**: Low-latency, high-performance, similar to local storage (e.g. EBS). • **File Storage**: Highly scalable, ideal for content repositories (e.g., EFS). • **Object Storage**: Stores unstructured, semi structured, or structured data, highly scalable (e.g. S3). • **Data Lake Storage** • Centralized repository for structured and unstructured data. • Built using for e.g. S3. • Supports in-place transformation and querying. • Various classes for different access needs and cost considerations (e.g., S3 Standard, S3 Glacier). • **Data Warehouse Storage** • Centralized repositories for structured and semistructured data. • Example: Redshift. • Redshift Spectrum for querying data in data lakes. • **Purpose-Built Databases** • Importance of selecting the right database for application architecture.

	• Factors: Application workload, data shape, performance, operations burden. • Common use cases and Azure / AWS services (e.g., Aurora, DynamoDB, Neptune). • **Storage in Support of the Pipeline** • Comparison of ETL and ELT pipelines. • Example architectures for both pipeline types. • **Securing Storage** • Built-in security features cloud related services • Access management through policies and encryption options. • Centralized governance with for e.g. Lake Formation or other • Security features of cloud related, for e.g. Redshift. • **Lab: Storing and Analyzing Data** • Build a proof of concept using Azure or AWS Redshift. • Use SQL queries
5. Be able to implement the steps to process structured, semi structured, unstructured data formats in a data pipeline built in the cloud.	**Scope** **Teaching will cover:** • **Big Data Processing Concepts** • Introduction to big data processing. • Iterative data pipeline: ingestion, storage, transformation, processing, and analysis. • **Types of Data Processing** • **Batch Processing**: Processes large amounts of cold data, tolerates structured and unstructured data, examples could include EMR and Apache Hadoop. • **Streaming Processing**: Processes data sequentially in near real-time, examples could include Kinesis Data Streams and Apache Spark Streaming. • **Frameworks for Big Data Processing** • **Batch Processing**: Apache Hadoop MapReduce, Apache Hive, Apache Pig. • **Stream Processing**: Apache Spark, Kinesis, Apache Flink. • **Apache Hadoop** • Open-source, distributed processing framework. • Components: HDFS, YARN, MapReduce, Hadoop Common. • Benefits: Scalability, fault tolerance, flexibility. • Challenges: Stability issues, security concerns. • **Apache Hudi** • Open-source data management framework. • Storage types: Copy on Write (CoW), Merge on Read (MoR). • View options: Read-optimized, incremental, real-time. • **Labs** • **Processing Logs: Could use Azure or AWS.** • **For specific example look at EMR** • Use EMR to process logs stored in an S3 bucket. • Use Apache Hive for data processing. • **Updating Dynamic Data in Place** • Use S3, Athena, AWS Glue, and Apache Hudi for in-place data updates and queries.
6. Be able to differentiate the characteristics of an ML pipeline and its specific processing steps and analyse data using Cloud tools that are appropriate to	**Scope** **Teaching will cover:** • Objectives: Distinguish between labels, features, and samples, describe ML lifecycle phases, frame ML problems, collect and preprocess data, feature engineering, model development and deployment, E.g. AWS or Azure ML infrastructure services, Further examples could be SageMaker features, generative AI. • **Module Objectives** • Distinguish between labels, features, and samples. • Describe phases of the ML lifecycle. • Frame ML problems. • Collect and preprocess data. • Feature engineering. • Model development and deployment.

a given use case.	• ML infrastructure services. • SageMaker features. • Define generative AI. • **ML Concepts** • Comparison of ML to traditional analytics. • Types of ML models: Supervised, unsupervised, reinforcement. • Subcategories of AI: Machine learning, deep learning, generative AI. • **The ML Lifecycle** • Phases: Identify business goal, frame ML problem, process data, develop model, deploy model, monitor model. • Roles involved: Data scientist, data engineer, domain experts, AI/ML architect. • **Framing the ML Problem** • Identify business goal. • Determine features and labels. • Establish metrics and data sourcing strategy. • Evaluate if ML is the correct approach. • **Collecting Data** • Build ingestion pipeline. • Protect data veracity. • Collect enough data for training and testing. • Apply labels to training data. • **Preprocessing Data** • Clean, partition, scale, augment, balance, and format data. • Exploratory data analysis. • **Feature Engineering** • Feature creation and transformation. • Feature extraction and selection. • **Developing a Model** • Training, tuning, and evaluating a model. • Types of accuracy metrics: Accuracy, precision, recall, F1 score. • Scaling and cost management. • **Deploying a Model** • Deploying and monitoring the model in production. • Inference process. • Scaling and cost management. • MLOps approach. • **ML Infrastructure on AWS / Azure / Oracle use only one example!** • Compute, networking, and storage services. • Workflow services: SageMaker, EMR, AWS ParallelCluster, AWS Batch, EKS, ECS. • **SageMaker** • Integrated workbench for ML. • SageMaker Studio, Data Wrangler, Processing, Canvas. • Training and deploying models. • **Working with Generative AI** • Foundation models (FM), large language models (LLM), prompt engineering. • Related cloud services: E.g. SageMaker Jumpstart, Bedrock, Q Developer. • **AI/ML Services on AWS / Azure / Oracle** • **Activity - Labeling with SageMaker Ground Truth** • Label images for training data. • Lab introduction: Analysing and Visualizing Streaming Data with AWS Data Firehose, OpenSearch Service, and OpenSearch Dashboards (Other Azure or Oracle related LABS could also be used)

11. Quality Assurance of Centre Performance

11.1 Internal Quality Assurance

When delivering and assessing this qualification, centres must align with stakeholders' expectations and address learners' needs by implementing a practical and applied programme. Centres have the flexibility to customise programmes to meet local requirements and establish connections with local employers and the broader vocational sector.

The Assessor should work with the Internal Quality Assurer to ensure that the assessment is planned in line with OCN NI requirements. Assessment Plans must be developed and approved by the Internal Quality Assurer prior to the delivery of the qualification.

All units within this qualification must undergo internal assessment. Learners must provide evidence that they have appropriately met all assessment criteria required for that grade.

The assessment format for all units involves a task conducted after the delivery of the unit's content, or part of it, if multiple tasks are used. Tasks may exhibit in various forms, encompassing practical and written types. Please refer to 'OCN NI's Assessment Definitions Guide' for additional details.

A task constitutes a distinct activity completed independently by learners, separated from teaching, practice, exploration, and other activities guided by tutors. Tasks are assigned to learners with a specified start date, completion date, and explicit requirements for the evidence to be produced. Some tasks may include observed practical components and require diverse forms of evidence.

A valid assignment will enable a clear and formal assessment outcome which meets the requirements of the assessment criteria. Assessment decisions are based on the specific assessment criteria given in each unit and set at each grade level. The way in which individual units are written provides a balance of assessment of understanding, practical skills and vocational attributes appropriate to the purpose of qualifications.

It is the Assessor's role to ensure that learners are appropriately prepared for assessment, this begins from induction onwards. Assessors should ensure that learners understand how assessment tasks are used to determine the award of credit, the importance of meeting assessment timelines, and that all learners work must be independently created, where source documents are used this should be appropriately referenced, learners should be aware of what would constitute plagiarism and the possible consequences.

When conducting the assessment, Assessors must ensure they do not provide direct input, instructions or specific feedback which may compromise the authenticity of the work submitted.

Once the Assessor has authenticated the learners work, they must transparently demonstrate the rationale behind their assessment decisions. Once a learner completes all assigned tasks for a unit, the Assessor will allocate a grade for the unit. Refer to the 'Unit Grading Matrix' for additional information on the grading process.

Once the Assessor has completed the assessment process for the task, the assessment decision is recorded formally, and feedback is provided to the learner. The feedback should show the learner the outcome of the assessment decision, how it was determined or where the criteria has been met, it may indicate to the learner why achievement of the assessment criteria has not been met. It must be clear to the learner that this Assessment outcome is subject to verification.

For further information on assessment practice, please see the 'OCN NI Centre Handbook'. Assessment Training is also available and can be booked through the OCN NI Website.

11.2 Internal Quality Assurance

The role of the Internal Quality Assurer is to ensure appropriate internal quality assurance processes are carried out. The Internal Quality Assurer must oversee that assessments are conducted in accordance with relevant OCN NI policies, regulations, and this specification.

The Internal Quality Assurer must ensure assessments are fair, reliable, and uniform, thereby providing a consistent standard for all learners.

Internal Quality Assurers are required to provide constructive feedback to Assessors, identifying areas of strength and those that may require improvement. This feedback contributes to the ongoing professional development of Assessors.

Contributing to the standardisation of assessment practices within the centre is an important function of this role. This entails aligning assessment methods, grading criteria, and decision-making processes to maintain fairness and equity.

Internal Quality Assurers will actively engage in the sampling and monitoring of assessments to ensure the consistency and accuracy of assessment decisions. This process helps identify trends, areas for improvement, and ensures the robustness of the overall assessment system.

For further information on Internal Quality Assurance practice, please see the 'OCN NI Centre Handbook'. Internal Quality Assurance Training is also available and can be booked through the OCN NI Website.

11.3 Documentation

For internal quality assurance processes to be effective, the internal assessment and Internal Quality Assurance team needs to keep effective records.

- The programme must have an assessment and Internal Quality Assurance plan. When producing a plan, they should consider:
    - the time required for training and standardisation activities
    - the time available to undertake teaching and carry out assessment,
    - consider when learners may complete assessments and when quality assurance will take place
    - the completion dates for different assessment tasks
    - the date by which the assignment needs to be internally verified
    - sampling strategies
    - how to manage the assessment and verification of learners' work so that they can be given formal decisions promptly
    - how resubmission opportunities can be scheduled.

The following documents are available from OCN NI and document templates can be found in the Centre Login section of the OCN NI website www.ocnni.org.uk:

- A1 – Learner Assessment Record per Learner
- Learner Authentication Declarations
- Records of any reasonable adjustments applied for and the outcome – please see 'OCN NI's Reasonable Adjustments and Special Consideration Policy' for further information
- M1 Internal Quality Assurance Sample Record
- M2 Feedback to Assessor
- Records of any complaints or appeals

11.4 External Quality Assurance

All OCN NI recognised centres are subject to External Quality Assurance. External quality assurance activities will be conducted to confirm continued compliance with the CCEA Regulation General Conditions of Recognition, OCN NI terms and conditions and the requirements outlined within this qualification specification.

The External Quality Assurer is assigned by OCN NI. The External Quality Assurer will review the delivery and assessment of this qualification. This will include, but is not limited to, the review of a sample of assessment evidence and evidence of the internal quality assurance of assessment and assessment decisions. This will form the basis of the External Quality Assurance report and will help OCN NI determine the centre's risk.

The role of the External Quality Assurer serves as an external overseer of assessment quality, working to uphold consistency, compliance, and continuous improvement within the assessment process. Their role is crucial in ensuring that assessments are valid, reliable, fair, and aligned with the required standards and regulations.

For further information on OCN NI Centre Assessments Standards Scrutiny (CASS) Strategy, please see the OCN NI Centre Handbook.

11.5 Standardisation

As a process, standardisation is designed to ensure consistency and promote good practice in understanding and the application of standards. Standardisation events:

- make qualified statements about the level of consistency in assessment across centres delivering a qualification
- make statements on the standard of evidence that is required to meet the assessment criteria for units in a qualification
- make recommendations on assessment practice
- produce advice and guidance for the assessment of units
- identify good practice in assessment and Internal Quality Assurance

Centres offering this qualification must carry out internal standardisation activities prior to the claim for certification.

Centres offering units of an OCN NI qualification must attend and contribute assessment materials and learner evidence for standardisation events if requested.

OCN NI will notify centres of the nature of sample evidence required for standardisation events (this will include assessment materials, learner evidence and relevant Assessor and Internal Quality Assurer documentation). OCN NI will make standardisation summary reports available and correspond directly with centres regarding event outcomes.

12. Administration

12.1 Registration

A centre must register learners for this qualification within 20 days of commencement of the delivery of the programme.

For further information on learner registration please see the OCN NI Centre Handbook and the QuartzWeb Manual, available through the Centre Login section of the OCN NI website. Administration training is also available and can be booked through www.ocnni.org.uk.

12.2 Certification

Once all internal quality assurance activities have been successfully completed, the centre can claim certification for the learner(s).

Certificates will be issued to centres within 20 working days from completion of a satisfactory external quality assurance activity, if appropriate, alternatively from the submission of an accurate and complete marksheet.

It is the responsibility of the centre to ensure that certificates received from OCN NI are held securely and distributed to learners promptly and securely.

For further information on the uploading of results please see the QuartzWeb Manual for guidance, administration training is also available and can be booked through OCN NI

12.3 Charges

OCN NI publishes all up-to-date qualification fees in its Fees and Invoicing Policy document. Further information can be found on the centre login area of the OCN NI website.

12.4 Equality, Fairness and Inclusion

OCN NI's are committed to ensuring all learners have an equal opportunity to access our qualifications and assessment, and that our qualifications are awarded in a way that is fair to every learner.

OCN NI is committed to making sure that:

- learners with a protected characteristic are not, when they are undertaking one of our qualifications, disadvantaged in comparison to learners who do not share that characteristic

- all learners achieve the recognition they deserve for undertaking a qualification and that this achievement can be compared fairly to the achievement of their peers

For information on reasonable adjustments and special considerations please see the OCN NI Centre Handbook and Reasonable Adjustments and Special Considerations Policy held in the back office of the OCN NI website.

12.5 Retention of Evidence

OCN NI has published guidance for centres on the retention of evidence. Details are provided in the OCN NI Centre Handbook and can be accessed via the OCN NI website.

OCN NI Level 5 Certificate in Data Engineering Processing Techniques
Qualification Number: 610/6494/4

Operational start date: 15/09/2025
Review date: 14/09/2030

Open College Network Northern Ireland (OCN NI)
Sirius House
10 Heron Road
Belfast
BT3 9LE

Phone: 028 90 463990
Email: info@ocnni.org.uk
Web: **www.ocnni.org.uk**

12.6 Appendix 1 - Definition of OCN NI's Assessment Verbs

The following verbs are working definitions of those used in OCN NI assessments with examples of how they can be applied and used in different but equally valid contexts.

Verb	Definition	Example
Analyse	To examine a subject systematically by breaking it into component parts, identifying relationships between elements, and explaining how these parts contribute to the overall meaning or function.	The learner will be expected to break down the subject into its key components or constituent parts to understand its structure. Examine relationships between these different elements to identify patterns, connections, or dependencies. Investigate causes and effects to understand how different factors influence outcomes. Compare and contrast different aspects to highlight similarities, differences, and significance. Evaluate the importance of each component in relation to the whole system or concept. Draw evidence-based conclusions about how the individual parts work together to create the overall meaning, purpose, or function. Consider multiple perspectives to provide a comprehensive understanding of the subject's complexity and nuances.
Demonstrate	To show autonomous mastery of challenging tasks, demonstrating analytical thinking, and providing comprehensive evidence of advanced capabilities rather than simply following basic instructions.	The learner will be expected to independently select appropriate methodologies, execute complex multi-stage processes while adapting to unexpected challenges, critically evaluate their decisions and outcomes, synthesize theoretical knowledge with practical application, and produce detailed documentation that evidences their strategic thinking, problem-solving approach, and professional competency standards throughout the entire activity.
Determine	To systematically investigate complex problems using advanced analytical methods, critically evaluate multiple solutions, and reach well-reasoned conclusions supported by	The learner will be expected to research comprehensive data sources, apply sophisticated analytical frameworks, compare competing methodologies against established criteria, synthesize findings from multiple perspectives, justify their reasoning through evidence-based

	comprehensive evidence and professional judgment	arguments, and formulate definitive conclusions while acknowledging limitations and potential alternatives in their decision-making process.
Explain	Make clear a given subject matter and / or give reasons for the procedure in a given situation or regarding a given subject matter. Set out purposes or reasons to rationalise a response or action.	The learner will be expected to make the subject clear by presenting information in a logical, structured way that others can easily understand. Give reasons for why something happens, exists, or is done in a particular way, using evidence to support your explanations. Set out purposes by identifying the aims, objectives, or goals behind actions, decisions, or processes. Show cause and effect relationships to demonstrate how one factor leads to or influences another. Use appropriate terminology accurately to ensure explanation is precise and professional. Provide step-by-step processes when describing how something works or should be carried out. Justify reasoning by offering logical arguments and evidence to support points. Connect theory to practice by showing how underlying principles apply in real situations. Anticipate questions an audience might have and address potential areas of confusion or misunderstanding.
Evaluate	To make reasoned judgements about the value, effectiveness, or significance of a subject by weighing evidence, assessing strengths and weaknesses, and reaching supported conclusions.	The learner will be expected to establish clear criteria for judgement to ensure evaluation is objective and systematic rather than based on personal preference. Gather relevant evidence from credible sources to support assessment and provide a solid foundation for judgements. Assess strengths and weaknesses by identifying what works well and what doesn't, considering both positive and negative aspects fairly. Weigh evidence carefully by considering the quality, reliability, and relevance of different sources and information. Consider multiple perspectives to ensure evaluation acknowledges different viewpoints and potential biases. Make reasoned judgements based on logical analysis rather than unsupported opinions or assumptions. Measure against standards by comparing the subject to established benchmarks, best practices, or expected outcomes. Draw supported conclusions that are clearly linked to the evidence presented and the

		criteria established. Consider implications of findings for future decisions, actions, or understanding.
Select	To choose and identify the most appropriate items or information from a range of options based on specific criteria, relevance, and requirements.	The learner will be expected to comprehend the criteria and requirements for selection. This involves understanding the specific attributes, qualities, or characteristics that are important for the task. The learner conducts research and gathers a range of potential items or information. This involves identifying sources, collecting data, and compiling a list of options to choose from. The learner evaluates the available options against the selection criteria. This involves comparing and contrasting different items or pieces of information to determine their suitability. The learner makes informed decisions based on their evaluation. This involves using critical thinking and judgment to choose the most appropriate items or information. The learner justifies their selection by explaining the rationale behind their choices. This involves articulating the reasons why the selected items or information meet the criteria and requirements. The learner ensures that the selected items or information are accurate and relevant to the task. This involves verifying the validity and reliability of the chosen options.
Summarise	To present the main points, key findings, or essential information of a subject in a concise format, capturing the most important aspects without unnecessary detail.	The learner will be expected to identify the main points by distinguishing between essential information and supporting details to focus on what is most important. Extract key findings from source material, selecting the most significant results, conclusions, or outcomes that address the core purpose. Eliminate unnecessary detail by removing examples, repetition, and peripheral information that doesn't contribute to the central message. Maintain logical structure by organising the summary in a coherent sequence that reflects the original flow of ideas or arguments. Use the learner's own words where possible to demonstrate understanding while preserving the original meaning and

		intent of the source. Preserve accurate meaning by ensuring condensed version faithfully represents the author's main arguments and conclusions. Apply appropriate length by matching the summary's scope to the requirements while ensuring all essential points are covered. Check for completeness to verify that no critical information has been omitted that would affect understanding of the subject. Maintain objectivity by presenting the information neutrally without adding the learner's own opinions or interpretations.