Qualification Specification

Qualification summary

Qualification title	**NCFE Level 5 Diploma: Data Engineer**
Ofqual qualification number (QN)	610/5972/9
Guided learning hours (GLH)	540
Total qualification time (TQT)	1200
Credit value	120
Minimum age	18
Qualification purpose	This qualification is designed for learners who want to upskill or retrain within the digital sector. It is also suitable for learners who want to further their studies in the digital sector. This higher technical qualification (HTQ) will give learners the skills, knowledge and behaviours to meet specific employer needs and industry requirements.
Grading	Pass/merit/distinction
Assessment method	Internally assessed and externally quality assured portfolio of evidence, including task-based assessments.
Work/industry placement experience	Work/industry placement experience is not required.
Apprenticeship/ Occupational standards	Data engineer (OCC1386) This HTQ content has been aligned with the Data Engineer occupational standard. This HTQ is designed to be delivered as a stand-alone qualification which is an alternative to an apprenticeship. It does not form part of an apprenticeship.
Regulation information	This is a regulated qualification. The regulated number for this qualification is 610/5972/9.
Funding	This qualification may be eligible for funding. For further guidance on funding, please contact your local funding provider.

Contents

Section 1: introduction

Centres must ensure they are using the most recent version of the Qualification Specification on the NCFE website.

A higher technical qualification (HTQ) is a prestigious, kite-marked qualification aimed at meeting employers' needs and increasing learner engagement in level 4 or 5 technical education. This HTQ content has been aligned with the Data Engineer apprenticeship standard.

This qualification aims to:

- provide the knowledge, skills and behaviours that are needed to enter occupations across the country
- be understood and recognised as high quality by employers and so have national labour market currency
- give learners confidence that those qualifications are recognised by employers and are perceived to be a credible, prestigious and distinct pathway.

Aims and objectives

This qualification aims to:

- focus on the study of data engineering within the digital sector
- offer breadth and depth of study, incorporating a key core of knowledge
- provide opportunities to acquire a number of practical and technical skills.

The objectives of this qualification are to provide learners with knowledge, skills and behaviours related to the following areas:

- data engineering principles and foundations
- data architecture and management
- data pipelines and automation
- software development for data products
- data analytics, business intelligence and incident management
- professional practice and continuous improvement in data engineering.

Support Handbook

This Qualification Specification must be used alongside the mandatory Support Handbook, which can be found on the NCFE website. This contains additional supporting information to help with planning, delivery and assessment.

This Qualification Specification contains all the qualification-specific information you will need that is not covered in the Support Handbook.

Guidance for entry and registration

This qualification is designed for learners who want to advance their career within data engineering.

It may also be useful to those with suitable industry experience, learners studying other related qualifications within the digital sector, or for those studying qualifications in the following area:

- data analyst.

Registration is at the discretion of the centre in accordance with equality legislation and should be made on the NCFE Portal.

There are no specific prior skills/knowledge a learner must have for this qualification. However, learners may find it helpful if they have already achieved a level 4 qualification.

Centres are responsible for ensuring that all learners are capable of achieving the units and/or learning outcomes (LOs) and complying with the relevant literacy, numeracy, and health and safety requirements.

Learners registered on this qualification should not undertake another qualification at the same level, or with the same/a similar title, as duplication of learning may affect funding eligibility.

Achieving this qualification

To be awarded this qualification, learners are required to successfully achieve **120 credits overall, which is a minimum of a pass** in each of the **6** graded mandatory units.

Please refer to the list of units in appendix A or the unit summaries in section 2 for further information.

To achieve this qualification, learners must successfully demonstrate their achievement of all LOs of the units as detailed in this Qualification Specification. A partial certificate may be requested for learners who do not achieve the full qualification but have achieved at least one whole unit; partial achievement certificate fees can be found in the Fees and Pricing document on the NCFE website.

Progression

Learners who achieve this qualification could progress to the following:

- employment:
  - data engineer
  - machine learning engineer
  - data architect
  - ETL developer
- further education:
  - related apprenticeships
- higher education:
  - data science degree.

Resource requirements

There are no mandatory resource requirements for this qualification, but centres must ensure learners have access to suitable resources to enable them to cover all the appropriate LOs.

Realistic work environment (RWE) recommendation

The assessment of competence-based criteria should ideally be conducted within the workplace. However, in instances where this is not feasible, learners can be assessed in a realistic work environment (RWE) designed to replicate real work settings.

It is essential for organisations utilising an RWE to ensure it accurately reflects current and authentic work environments. By doing so, employers can be confident that competence demonstrated by a learner in an RWE will be translated into successful performance in employment.

In establishing an RWE, the following factors should be considered.

The work situation being represented is relevant to the competence requirements being assessed:

- The work situation should closely resemble the relevant setting.
- Equipment and resources that replicate the work situation must be current and available for use to ensure that assessment requirements can be met.
- Time constraints, resource access and information availability should mirror real conditions.

The learner's work activities reflect those found in the work environment being represented, for example:

- Interaction with colleagues and others should reflect expected communication approaches.
- Tasks performed must be completed to an acceptable timescale.
- Learners must be able to achieve a realistic volume of work as would be expected in the work situation being represented.
- Learners operate professionally with clear understanding of their work activities and responsibilities.
- Feedback from colleagues and others (for example, customers, service users) is maintained and acted upon.
- Account must be taken of any legislation, regulations or standard procedures that would be followed in the workplace.

How the qualification is assessed

Assessment is the process of measuring a learner's skill, knowledge and understanding against the standards set in a qualification.

This qualification is internally assessed and externally quality assured.

The assessment consists of one component:

- an internally assessed portfolio of evidence, which is assessed by centre staff and externally quality assured by NCFE (internal quality assurance must still be completed by the centre as usual).

Learners must be successful in this component to gain the Level 5 Diploma: Data Engineer.

Learners who are not successful can resubmit work within the registration period; however, a charge may apply in cases where additional external quality assurance visits are required.

All the evidence generated by the learner will be assessed against the standards expected of a level 5 learner for each LO.

Unless otherwise stated in this specification, all learners taking this qualification must be assessed in English and all assessment evidence presented for external quality assurance must be in English.

Internal assessment

We have created three sample assessments for the internally assessed units, which include a holistic, unit, and partial assessment, which can be found within three separate sample assessment materials documents on the NCFE website. These tasks are not mandatory. You can contextualise these tasks to suit the needs of your learners to help them build up their portfolio of evidence. The tasks have been designed to cover some of the LOs from 4 units and provide opportunities for stretch and challenge. For further information about contextualising the tasks, please contact the Provider Development team.

Each learner must create a portfolio of evidence generated from appropriate assessment tasks to demonstrate achievement of all the LOs associated with each unit. The assessment tasks should allow the learner to respond to a real-life situation that they may face when in employment. On completion of each unit, learners must declare that the work produced is their own and the assessor must countersign this.

If a centre needs to create their own internal assessment tasks, there are four essential elements in the production of successful centre-based assessment tasks; these are:

- ensuring the assessment tasks are meaningful with clear, assessable outcomes
- appropriate coverage of the content, LOs or assessment criteria (AC)
- having a valid and engaging context or scenario
- including sufficient opportunities for stretch and challenge for higher attainers.

External quality assurance

Summatively assessed and internally quality assured grades for completed units must be submitted via the NCFE Portal, prior to an external quality assurance review taking place. Following the external quality assurance review, the unit grades will either be accepted and banked by your external quality assurer (EQA) or, if they disagree with the grades, they will be rejected. More detailed guidance on this process and what to do if your grades are rejected can be found in the Support Handbook and on the NCFE website.

Enquiries about results

All enquiries relating to learners' results must be submitted in line with our Enquiries about Results and Assessment Decisions Policy, which is available on the NCFE website.

Not yet achieved grade

A result that does not achieve a pass grade will be graded as a not yet achieved grade. Learners may have the opportunity to resit.

Grading information

Each unit of the qualification is graded using a structure of not yet achieved, pass, merit, distinction.

Grading internally assessed units

The grading criteria for each unit have been included in the Qualification Specification. Grading criteria have been written for each LO in a unit. Assessors must be confident that, as a minimum, all LOs have been evidenced and met by the learner. Assessors must make a judgement on the evidence produced by the learner to determine the grading decision for the unit.

Once assessors are confident that all the pass criteria have been met, they can move on to decide if the merit criteria have been met. If the assessor is confident that all the merit criteria have been met, they can decide if the distinction criteria have been met. As the grading criteria build up from the previous grade's criteria, the evidence must meet 100% of the grade's criteria to be awarded that grade for the unit.

If the learner has insufficient evidence to meet the pass criteria, a grade of not yet achieved must be awarded for the unit.

Centres must then submit each unit grade via the NCFE Portal. The grades submitted will be checked and confirmed through the external quality assurance process. This is known as 'banking' units. Once a learner's grade has been banked, they are permitted one opportunity to revise and redraft their work; more detail on this process can be found in the Support Handbook.

The internal assessment component is based on performance of open-ended tasks that are assessed holistically against the grading criteria to achieve a grade. Each unit of the qualification is internally assessed and will be allocated a weighting based on the guided learning hours (GLH) and a score based on the holistic grade.

All of the assessment points need to be evidenced in the learner's portfolio, but the grade awarded is based on the standard of work for the LO as a whole. This allows for increased professional judgement on the part of the assessor in terms of the learner's overall level of performance against the LOs.

Awarding the final grade

To achieve the qualification, learners must have achieved a pass in all units.

The calculation of the overall qualification grade is based on the learner's overall performance across all of the units. Learners are awarded their grade based on the points allocated for each grade, across all units. The table below shows the amount of points awarded for each credit, per unit:

Grade	Points per credit
Pass	1
Merit	3
Distinction	5

This means that if a learner gains a pass in a unit of 20 credits, they will receive 20 points.

If they then gained a merit in a unit of 20 credits, they would receive 60 points.

If they then gained distinction in their remaining units, totalling 80 credits, they would receive 400 points.

This would give a total of 480 points, which would then be used to calculate the overall grade, using the table below.

The table below shows the overall total points required for each grade boundary:

Grade	Points score
Not achieved	0 to 119
Pass	120 to 239
Merit	240 to 479
Distinction	480+

The final grade for the qualification is based on a structure of not yet achieved, pass, merit, distinction and will be issued to the centre by NCFE upon the centre claiming the learner's certificate on the NCFE Portal.

For further information on assessment, please refer to the User Guide to the External Quality Assurance Report, which can be found on the NCFE website.

NCFE does not anticipate any changes to our aggregation methods or any overall grade thresholds; however, there may be exceptional circumstances in which it is necessary to do so to secure the maintenance of standards over time. Therefore, overall grade thresholds published within this Qualification Specification may be subject to change.

Records of grades achieved for the NCFE Level 5 Diploma: Data Engineer (610/5972/9)

Grades achieved		Credits per unit	Distinction		Merit		Pass		Points/ grade
Regulated unit number	Unit title	Credits per unit	Points per credit	Points	Points per credit	Points	Points per credit	Points	
K/651/6932	Data engineering principles and foundations	20	5	100	3	60	1	20	
L/651/6933	Data architecture and management	15	5	75	3	45	1	15	
M/651/6934	Data pipelines and automation	25	5	125	3	75	1	25	
R/651/6935	Software development for data products	25	5	125	3	75	1	25	
T/651/6936	Data analytics, business intelligence and incident management	20	5	100	3	60	1	20	
Y/651/6937	Professional practice and continuous improvement in data engineering	15	5	75	3	45	1	15	
							Total points		

Section 2: unit content and assessment guidance

This section provides details of the structure and content of this qualification.

Within learners' portfolios, a range of evidence types are acceptable if all learning outcomes (LOs) are covered, and if the evidence generated can be internally and externally quality assured. For approval of methods of internal assessment other than portfolio building, please contact your external quality assurer (EQA).

The explanation of terms explains how the terms used in the unit content are applied to this qualification. This can be found in section 3.

Behavioural framework

Embedded within higher technical qualifications (HTQs) is the opportunity for learners to develop behaviours relevant to their chosen discipline, in line with the qualification's knowledge and skills.

The following table identifies opportunities to demonstrate the behaviours – embedded within the knowledge and skills – that will be assessed as part of this HTQ. Learners may also naturally demonstrate these behaviours elsewhere, beyond the listing below. All listed behaviours are subject to assessment.

B1: Acts proactively and takes accountability adapting positively to changing work priorities, ensuring deadlines are met.
B2: Works collaboratively with stakeholders and colleagues, developing strong working relationships to achieve common goals. Support an inclusive culture and treat technical and non-technical colleagues and stakeholders with respect.
B3: Quality focus that promotes continuous improvement utilising peer review techniques, innovation and creativity to the data system development process to improve processes and address business challenges.
B4: Takes personal responsibility towards net zero and prioritises environmental sustainability outcomes in how they carry out the duties of their role.
B5: Use initiative and innovation to problem solve and trouble shoot, providing creative solutions.
B6: Keeps abreast of developments in emerging, contemporary and advanced technologies to optimise sustainable data products and services.

Unit	Behaviours					
	B1	B2	B3	B4	B5	B6
01: Data engineering principles and foundations	N/A	LO2	N/A	N/A	N/A	N/A
02: Data architecture and management	LO3 LO5	N/A	N/A	N/A	N/A	N/A
03: Data pipelines and automation	N/A	N/A	N/A	N/A	LO3	N/A
04: Software development for data products	LO3	N/A	LO5	LO2	N/A	LO3
05: Data analytics, business intelligence and incident management	N/A	N/A	N/A	N/A	N/A	N/A
06: Professional practice and continuous improvement in data engineering	N/A	LO4	LO4	LO1	N/A	LO3

Unit 01 Data engineering principles and foundations (K/651/6932)

Unit summary				
This unit covers core data engineering concepts, including data structures, storage, integration, and governance. Learners will explore structured and unstructured data, database normalisation, and regulatory compliance.				
Assessment				
Internally assessed unit				
Mandatory	**Graded P/M/D**	**Level 5**	**20 credits**	**90 GLH**

Learning outcomes (LOs)	Mandatory teaching content
1. Explore the principles of data	**Knowledge**: Principles of data: • lawfulness, fairness, and transparency • purpose limitation • data minimisation • accuracy • storage limitation • integrity and confidentiality • accountability. Types of datasets: • open and public • administrative • research. Approaches to gaining data through research. The value of external data sources for enriching internal data. How direct data acquisition can enhance business operations.
2. Explore the concepts of data governance	**Knowledge**: Appropriate legislation and how it is applied to the safe use of data: • UK General Data Protection Regulation (GDPR)/Data Protection Act 2018 • Digital Economy Act 2017 • Computer Misuse Act 1990. Considerations for regulatory requirements: • data and information security standards:

Learning outcomes (LOs)	Mandatory teaching content

	o NIST
	    o PCI DSS.
	
	The use of data governance frameworks (for example, government led (UK), industry recognised).
	
	Features of policies and procedures relating to data governance:
	
	• compliance with regulatory requirements
	• data privacy
	• data anonymisation
	• data ownership
	• data lineage
	• ethical usage
	• metadata
	• security
	• quality control:
	    o completeness
	    o timeliness
	    o consistency.
	
	Skills:
	
	Apply data solutions to address common organisational requirements.
	
	Work collaboratively with appropriate stakeholders to meet requirements.

3. Explore and apply the principles of data normalisation and redundancy in relational databases	**Knowledge**:
	
	Principles of data normalisation:
	
	• Codd's 12 rules
	• normal form (NF):
	    o 1NF – eliminating duplicate records
	    o 2NF – eliminating partial dependencies
	    o 3NF – eliminating transitive dependencies
	    o BCNF (Boyce-Codd normal form) – a stricter version of 3NF
	    o 4NF – removing multi-valued dependencies
	    o 5NF – eliminating join dependencies.
	
	Features of relational database management systems (RDBMSs).
	
	Advantages of data normalisation to databases with regards to key concepts:
	
	• data protection
	• data redundancy
	• inconsistent dependency.

Learning outcomes (LOs)	Mandatory teaching content
	Skills: Query, manipulate and organise data to meet requirements, using industry-standard tools and programming languages (for example, SQL and Python). Manage and implement database access controls to meet business requirements: • access levels (for example, resource, administrator) • types of access (for example, guest, standard) • access control methods (for example, MAC, RBAC). Implement automated validation checks.
4. Explore approaches to data integration	**Knowledge**: Approaches to data integration: • manual integration • API (application programming interface) integration • third-party integrations • direct integration. How integration of disparate data sources supports business outcomes.
5. Explore the unique features and functions of different data formats	**Knowledge**: Characteristics of different data formats. Data formats: • structured • semi-structured • unstructured. How different data formats can be stored: • data stores: o relational database (SQL) o non-relational database (NoSQL) • distributed systems:

Learning outcomes (LOs)	Mandatory teaching content
	o distributed file systems (for example, Apache Hadoop, Amazon S3) o distributed database o distributed NoSQL o distributed object storage. **Skills**: Apply a range of data storage solutions and distributed file systems to implement a system that meets specified requirements. Use data systems securely to meet requirements, in compliance with relevant legislation, policies and procedures.
6. Explore common data development frameworks and architectures	**Knowledge**: Processes involved in developing the data product to function within the organisation. Characteristics of common architectures: • cloud • on-premises • monolithic • modular. Features and uses of common frameworks (for example, Apache Spark, DBT, Apache Airflow).

Grading criteria

Learning outcomes (LOs)	Pass	Merit	Distinction
LO1: Explore the principles of data	**P1**: describe key principles of data management **P2**: explain how external data sources can affect business operations	**M1**: analyse different types of external data sources	**D1**: evaluate the use of direct data acquisition in business operations
LO2: Explore the concepts of data governance	**P3**: describe core data governance concepts **P4**: identify regulatory and legislative standards influencing data management within an organisation **P5**: evidence stakeholder collaboration to satisfy specific requirements	**M2**: analyse the role of data governance within organisations	**D2**: evaluate how data governance can impact organisational policies

Learning outcomes (LOs)	Pass	Merit	Distinction
LO3: Explore and apply the principles of data normalisation and redundancy in relational databases	**P6**: apply the principles of database normalisation to reduce data redundancy **P7**: perform database queries and validation checks considering user access control	**M3**: analyse how normalisation supports key concepts of database performance and the removal of duplicate data	**D3**: evaluate the effectiveness of access controls and validation checks when maintaining a secure operational database
LO4: Explore approaches to data integration	**P8**: describe different approaches to integrate data	**M4**: analyse approaches to the integration of disparate data sources and the need for this to support business outcomes	
LO5: Explore the unique features and functions of different data formats	**P9**: explain characteristics of common data and storage formats **P10**: evidence the use of different types of secure data stores, such as SQL or similar	**M5**: compare how different data storage solutions meet the requirements of various data formats and uses	
LO6: Explore common data development frameworks and architectures	**P11**: describe data development frameworks and their integration within different architectural models	**M6**: analyse how organisational infrastructure affects the selection and implementation of data frameworks and architectures	**D4**: evaluate how the choice of development frameworks and architecture impacts scalability, performance, and cost-effectiveness of a data solution

Unit 02 Data architecture and management (L/651/6933)

Unit summary
In this unit, learners will gain an understanding of data warehousing, cloud storage, governance, security, and business continuity. This unit also covers frameworks for data quality, ethical considerations, and sustainability in data management.
Assessment
Internally assessed unit

Mandatory	Graded P/M/D	Level 5	15 credits	70 GLH

Learning outcomes (LOs)	Mandatory teaching content

1. Explore and analyse the inherent risks of data collection and usage	**Knowledge**:
	
	Risks associated with data collection:
	
	• privacy and security risks
	• non-compliance issues
	• ethical concerns
	• misuse of data
	• data quality issues:
	   o incomplete data
	   o inaccurate data
	   o inconsistent data.
	
	The potential impact of poor data quality on organisations.

2. Explore frameworks for data quality and how contributing dimensions support reliability	**Knowledge**:
	
	Components of data quality frameworks:
	
	• governance
	• profiling
	• quality dimensions:
	   o accuracy
	   o completeness
	   o consistency
	   o timeliness
	   o accessibility
	   o uniqueness
	• cleansing
	• validation
	• continuous monitoring and reporting:
	   o validity.
	
	Features of common data quality frameworks (for example, DAMA UK, TDQM).
	
	How to resolve identified risks to ensure data quality and to mitigate against future issues.

Learning outcomes (LOs)	Mandatory teaching content
	Skills: Monitor and analyse data quality metrics to meet operational requirements when processing data, considering: • quality • accuracy • reliability.

3. Develop a data product that complies with regulatory requirements	**Knowledge**:
	
	Key considerations when developing a data product:
	
	• system design
	• compliance and metrics (for example, regulatory requirements)
	• resource requirements
	• scalability.
	
	Skills:
	
	Apply data product design methods to meet requirements:
	
	• collation of information and requirements
	• creating an entity relationship diagram (ERD)
	• testing and evaluation
	• refinements:
	  ○ user requirements
	  ○ business requirements:
	    ▪ cost
	    ▪ resourcing
	    ▪ accessibility.

4. Explore the principles of data warehousing	**Knowledge**: The features of the ETL (extract, transform, load) process, and their role in preparing data for analysis. How to use an ETL pipeline to connect disparate data sources to meet business requirements. Core principles of data warehousing: • Kimball's dimensional data modelling • Inmon's data warehouse model. The principles for design and implementation: • ETL process • schema design (for example, star, snowflake) • separation of operational and analytical systems

Learning outcomes (LOs)	Mandatory teaching content
	• data quality and consistency • performance and scalability • security and access control. The role of data lakes, data warehouses and data marts in meeting business requirements. **Skills**: Manage data manipulation tasks, applying appropriate tools and techniques (for example, filtering, joining sorting), ensuring data integrity.
5. Analyse relevant policies and strategies to ensure business continuity when implementing data solutions	**Knowledge**: Features of key business continuity strategies (for example, system monitoring and surveillance), and their role in ensuring organisational resilience and the recovery of important data and services. The role of policies to support business continuity and mitigate against negative impact in relation to data provision: • business continuity • data backup • disaster recovery • incident reporting. **Skills**: Apply shell script (for example, Bash) to monitor availability, log downtime, and communicate findings to stakeholders through automated messages.

Grading criteria

Learning outcomes (LOs)	Pass	Merit	Distinction
LO1: Explore and analyse the inherent risks of data collection and usage	**P1**: describe common risks associated with data collection and usage	**M1**: analyse the effects of poor data quality on organisational operations and compliance	
LO2: Explore frameworks for data quality and how contributing dimensions support reliability	**P2**: describe key components and quality dimensions of a data quality framework	**M2**: analyse the benefits and limitations of implementing a recognised data quality framework within an organisation	
	P3: identify and monitor data quality metrics in line with operational requirements		

Learning outcomes (LOs)	Pass	Merit	Distinction
LO3: Develop a data product that complies with regulatory requirements	**P4**: design a compliant data product to meet user requirements **P5**: develop the data product in line with regulatory requirements	**M3**: evaluate the data product to ensure compliance and user requirements	**D1**: refine the data product to ensure user requirements
LO4: Explore the principles of data warehousing	**P6**: apply data cleaning, validation, and manipulation techniques during ETL processes	**M4**: compare the differences between data lakes, warehouses, and marts in meeting diverse data storage and retrieval needs	**D2**: evaluate the effectiveness of the ETL process in preparing data for business analysis
LO5: Analyse relevant policies and strategies to ensure business continuity when implementing data solutions	**P7**: describe backup, disaster recovery, and monitoring policies essential for ensuring business continuity in data provision **P8**: communicate system downtime and access issues effectively to stakeholders using automated messages	**M5**: analyse the importance of business continuity strategies to ensure systems are resilient and recoverable	**D3**: evaluate the effectiveness of business continuity plans and monitoring frameworks in maintaining secure, resilient data operations

Unit 03 Data pipelines and automation (M/651/6934)

Unit summary
This unit focuses on data ingestion, ETL processes, automation, and cloud-based pipeline development. Learners will apply batch, real-time, and on-demand processing methodologies to optimise data workflows.

Assessment
Internally assessed unit

Mandatory	Graded P/M/D	Level 5	25 credits	110 GLH

Learning outcomes (LOs)	Mandatory teaching content
1. Explore approaches used for handling and storing data	**Knowledge**: The considerations and use of data handling and storage approaches. Considerations: • relational and non-relational storage • cloud and on-premises solutions • structured versus unstructured data handling. Handling approaches: • ETL (extract, transform, load) • ELT (extract, load, transform) • API (application programming interface)-based transfer • manual transfer. Storage approaches: • data warehousing • data lakes • cloud storage and backups • archive and retention. Processes and considerations involved in supporting the movement of data during cloud migration: • migration planning • data transfer methods • tools and platforms • security and compliance • testing. The considerations of different data transfer methods (for example, batch, streaming, on-demand) to meet business requirements.

Learning outcomes (LOs)	Mandatory teaching content
2. Explore the application of data transfer from one location to another	**Knowledge**: Features and use cases of data ingestion frameworks (for example, Apache NiFi). Different data ingestion modes (batch, streaming, and on-demand) and their application for data processing. Appropriate tools and technologies used for streaming, batching, and on-demand data transfer between systems (for example, Apache Kafka, ETL platforms, REST APIs). The 7 Rs of cloud migration as a strategy for moving applications and data to the cloud. Types and implementation of migration strategies (for example, big bang, trickle, zero downtime) to meet requirements. **Skills**: Use different techniques to optimise data ingestion processes using appropriate data ingestion frameworks: • batch • streaming • on-demand.
3. Explore deployment methods for data pipelines and automated processes	**Knowledge**: Functions and features of pipeline deployment methods (for example, manual, scheduled, serverless) to ensure data flows reliably, securely, and efficiently. How version control and rollback planning are critical in data pipeline development and deployment: • collaboration • tracking • automated testing • recover from failure. Techniques and tools to effectively identify and troubleshoot issues within data processing pipelines: • monitoring and logging tools (for example, AWS CloudWatch) • debugging methodologies (for example, unit testing) • performance tuning techniques (for example, parallel processing).

Learning outcomes (LOs)	Mandatory teaching content
	Skills: Apply appropriate techniques and tools to identify and troubleshoot issues with data processing pipelines.
4. Explore the types and uses of common data engineering tools and applications	**Knowledge**: Features and application of data engineering tools used to design, build, manage, and optimise data pipelines and infrastructure: • data ingestion tools (for example, Apache NiFi) • data transformation tools (for example, Talend) • data storage tools (for example, Snowflake) • orchestration and workflow management tools (for example, Prefect) • data monitoring (for example, Monte Carlo). Common programming languages used in data engineering: • Python • SQL • Scala • Java • Bash/Shell • R • Golang. The role of YAML as a configuration and metadata format in data engineering. How data engineering tools, programming languages, and configuration formats are applied within an organisation to design, build, automate, and manage data pipelines and infrastructure. **Skills**: Automate batch, real-time, and on-demand data pipelines utilising programming languages and data integration platforms with graphical user interfaces or infrastructure-as-code tools. Develop and apply appropriate processes and algorithms to support extraction of structured data from unstructured data.

Grading criteria

Learning outcomes (LOs)	Pass	Merit	Distinction
LO1: Explore approaches used for handling and storing data	**P1**: describe different approaches for data handling and storage		

Learning outcomes (LOs)	Pass	Merit	Distinction
LO2: Explore the application of data transfer from one location to another	**P2**: apply appropriate ingestion frameworks to optimise a data ingestion process **P3**: apply appropriate transfer techniques to move data securely and efficiently during cloud migration	**M1**: critically analyse different cloud migration strategies and how they impact data transfer performance and continuity	**D1**: evaluate tools and technologies used to transfer data between systems
LO3: Explore deployment methods for data pipelines and automated processes	**P4**: describe different types of deployment processes for new data pipelines **P5**: describe the functions and features of data pipelines **P6**: identify and troubleshoot issues within data pipelines using appropriate tools and methods	**M2**: analyse how version control and rollback planning contribute to secure and effective pipeline development and deployment	
LO4: Explore the types and uses of common data engineering tools and applications	**P7**: describe a range of data engineering tools and their roles in supporting data pipeline infrastructure **P8**: automate a basic data pipeline using appropriate scripting languages	**M3**: critically analyse the benefits of integrating data engineering tools into a data pipeline to enhance operational efficiency	**D2**: evaluate how the choice of tools and automation strategies supports long-term scalability, maintenance, and security

Unit 04 Software development for data products (R/651/6935)

Unit summary				
This unit covers software engineering principles in data development, debugging, version control, and best practices for creating scalable and maintainable data solutions.				
Assessment				
Internally assessed unit				
Mandatory	**Graded P/M/D**	**Level 5**	**25 credits**	**110 GLH**

Learning outcomes (LOs)	Mandatory teaching content
1. Explore the principles of software development for data products	**Knowledge**: The principles of software development and how to apply them to data products: • version control • designing for modularity, reusability, and maintainability • code readability and documentation • debugging, logging, and traceability • automated testing and CI/CD (continuous integration/continuous deployment) • privacy by design. The role of development (DEV), staging (STG), and production (PRD) environments, and how each supports the software development lifecycle. The use of best practice approaches (for example, DRY) to enhance code quality, reusability, and testability within data-centric environments. **Skills**: Produce and maintain technical documentation for the data product containing the required technical and non-technical information for stakeholders. Contribute to and apply best practice approaches in software development principles when working with stakeholders.
2. Analyse how to cost and build a system to align with organisational sustainability responsibilities	**Knowledge**: Considerations when costing a system build: • requirements analysis (for example, features) • materials (for example, hardware, software) • staffing and operational costs (for example, development) • cost estimation methods (for example, return on investment (ROI)) • cloud versus on-premises.

Learning outcomes (LOs)	Mandatory teaching content
	Factors to consider when undertaking and implementing a system build: • architecture design (for example, modular) • technologies (for example, cloud services) • development environment (for example, CI/CD) • testing. How system design and implementation decisions can be aligned with organisational sustainability strategies (for example, E-waste reduction). **Skills**: Design a secure data product to meet business and user requirements.
3. Explore the appropriate use of technology and service management in relation to data products	**Knowledge**: How configuration management practices maintain accurate and up-to-date information on IT assets and systems to support effective service delivery and change control. The role and use of change management processes to control, assess, and authorise system modifications. Techniques to manage and respond to IT incidents by prioritising, diagnosing, and resolving service disruptions. How industry-standard tools (for example, service desk) support service management activities. How service management practices align with recognised frameworks and standards. **Skills**: Review use and configuration of existing tools and techniques to identify potential implementation issues to support business change or transitions.
4. Explore the processes for evaluating and implementing prototypes within a production environment	**Knowledge**: How to evaluate prototypes against defined criteria (for example, functionality, usability) to determine readiness for implementation. Developmental process to refine prototypes into production-ready solutions. How release strategies (for example, phased) are used to support the effective implementation of prototypes.

Learning outcomes (LOs)	Mandatory teaching content
	Skills: Evaluate prototype data products by assessing their cost–benefit value, compatibility with organisational data infrastructure, and the availability of open-source tools and online support.

5. Explore the lifecycle of implementing data solutions within a business	**Knowledge**:
	
	Processes involved in the lifecycle of implementing data solutions within a business:
	
	• ensure that data solutions are scoped accurately and align with business objectives, with consideration for:
	  o business context
	  o stakeholder needs
	  o success criteria
	• evaluation of prototype data solutions to test feasibility, validate technical approaches, and gather feedback to inform further development
	• the building and integration of data pipelines, models, and interfaces in line with design specifications
	• completion of testing and implementation of production deployment strategies
	• evaluation and monitoring of the performance of implemented data solutions.
	
	Skills:
	
	Review and monitor the health and performance of data systems (for example, identifying technical debt) as part of continuous improvement.

Grading criteria

Learning outcomes (LOs)	Pass	Merit	Distinction
LO1: Explore the principles of software development for data products	**P1**: describe key software development principles for data products		
	P2: produce technical documentation based on customer feedback for a data product that meets organisational and both technical and non-technical stakeholder needs		

Learning outcomes (LOs)	Pass	Merit	Distinction
LO2: Analyse how to cost and build a system to align with organisational sustainability responsibilities	**P3**: describe the features of a scalable and efficient system that aligns with organisational strategies for sustainability and security	**M1**: analyse the effectiveness of a system build including pricing, operational costs, and processing requirements	**D1**: justify system design and implementation decisions based on alignment to an organisation's sustainability strategy
LO3: Explore the appropriate use of technology and service management in relation to data products	**P4**: describe configuration and change management practices that are used to maintain accurate information about IT systems	**M2**: analyse how robust change and incident management practices contribute to continuous improvement and service reliability	
	P5: describe service management practices and the use of industry-standard tools to support the resolution of service disruptions	**M3**: analyse how the use and configuration of existing tools and techniques can potentially cause issues during business change and transitions	
LO4: Explore the processes for evaluating and implementing prototypes within a production environment	**P6**: describe the processes for evaluating and implementing prototypes against defined criteria	**M4**: evaluate the strengths, weaknesses, and integration challenges of prototypes when transitioning into production environments	
LO5: Explore the lifecycle of implementing data solutions within a business	**P7**: describe the processes involved in the lifecycle of implementing data solutions	**M5**: critically analyse the role of continuous improvement strategies in maintaining system efficiency, security, and relevance over time	**D2**: evaluate the full lifecycle of implementing data solutions, from initial scoping and prototyping through development, deployment, and continuous improvement
	P8: apply reviewing and monitoring techniques to ensure the health and performance of a data system		

Unit 05 Data analytics, business intelligence and incident management (T/651/6936)

Unit summary				
In this unit, learners will explore how to apply analytics techniques, including descriptive, predictive, and prescriptive analytics. This unit also covers forecasting, root cause analysis, and leveraging data for business insights.				
Assessment				
Internally assessed unit				
Mandatory	**Graded P/M/D**	**Level 5**	**20 credits**	**90 GLH**

Learning outcomes (LOs)	Mandatory teaching content
1. The application of data analysis techniques in industry	**Knowledge**: How to extract value from existing data products. The use of data analytics principles within a business environment: • descriptive data analytics • predictive data analytics • prescriptive data analytics. **Skills**: Analyse the potential opportunities and benefits for businesses in extracting value from existing data products, considering: • cost • environmental impact • ethics/privacy • operational benefits.
2. Analyse the use of on-demand cloud computing services to meet requirements	**Knowledge**: Implications of using local, remote or distributed solutions, considering: • pricing model and cost management • scalability • compliance with security standards • compatibility with existing systems • capability. Considerations for using on-demand cloud computing platforms: • scalability • managed infrastructure • access to latest technologies • upfront costs.

Learning outcomes (LOs)	Mandatory teaching content
3. Explore the processes used to monitor and optimise the performance of a data product	**Knowledge**: How to implement the processes required to monitor and optimise the performance of a data product, considering: • change management and version control • pipeline monitoring and incident management • query performance and resource usage monitoring. **Skills**: Develop and apply simple forecast and monitoring tools.

4. Explore advanced data interrogation techniques	**Knowledge**:
	
	How to implement and manage a root cause investigation in line with standard processes.
	
	Explore types and uses of data interrogation techniques:
	
	• root cause analysis (RCA):
	   o safety-based RCA
	   o process-based RCA
	• data profiling
	• anomaly detection.
	
	The use of different analytical methods to meet requirements (for example, sentiment analysis for marketing, time series for seasonal trends).
	
	Skills:
	
	Identify potential risks within operations and suggest mitigations and resolutions.
	
	Escalate and communicate risks using the appropriate process and level of information.
	
	Investigate and respond to incidents effectively:
	
	• conduct investigation and RCA
	• identify and record potential resolutions
	• communicate outcomes effectively with appropriate stakeholders.

Grading criteria

Learning outcomes (LOs)	Pass	Merit	Distinction
LO1: The application of data analysis techniques in industry	**P1**: compare principles of descriptive, predictive, and prescriptive analytics **P2**: demonstrate the ability to extract value from existing data products	**M1**: critically analyse how current data systems may be used to provide additional business insights	
LO2: Analyse the use of on-demand cloud computing services to meet requirements	**P3**: compare the use of local, remote and distributed solutions with consideration for costing, scalability, compliance and compatibility	**M2**: analyse on-demand cloud platforms in regard to scalability and financial and regulatory risks	**D1**: critically evaluate the financial, compliance, and strategic impacts of selecting different cloud platforms for enterprise data management
LO3: Explore the processes used to monitor and optimise the performance of a data product	**P4**: describe processes for monitoring and optimising data product performance, including change management, pipeline monitoring, and query optimisation **P5**: develop a forecasting or monitoring tool to support incident anticipation and response	**M3**: analyse the developed forecasting/monitoring tool to check its effectiveness	
LO4: Explore advanced data interrogation techniques	**P6**: apply root cause analysis and other data interrogation techniques to investigate and respond to incidents and performance issues **P7**: identify risks, suggest mitigations, and escalate issues appropriately in line with organisational procedures	**M4**: analyse the role of root cause analysis and anomaly detection in strategic decision-making and operational improvement	**D2**: evaluate the use of advanced analytical techniques in driving long-term strategic decisions

Unit 06 Professional practice and continuous improvement in data engineering (Y/651/6937)

Unit summary
This unit focuses on industry best practices, ethical considerations, sustainability, and strategies for keeping up with technological advancements in data engineering. Learners will also develop communication and stakeholder engagement skills.

Assessment				
Internally assessed unit				
Mandatory	**Graded P/M/D**	**Level 5**	**15 credits**	**70 GLH**

Learning outcomes (LOs)	Mandatory teaching content
1. Explore the principles of sustainable data products and environmental, social and governance responsibilities	**Knowledge**: Principles of sustainable data products: • reusable and secure • valuable and trustworthy • discoverable and understandable. The features of an environmental, social and governance (ESG) strategy and implementation considerations: Environmental: • energy and emissions • waste and biodiversity • climate strategy. Social: • societal impact • privacy and trust • customer satisfaction. Governance: • ethics and controls • risk and compliance • transparency and data governance strategy. **Skills**: Identify and apply sustainable solutions to support environmental strategies during product and service delivery, considering: Sustainable solutions: • efficiency pipelines • green infrastructure

Learning outcomes (LOs)	Mandatory teaching content
	• carbon monitoring. Green data engineering: • optimised code • smart storage • resource control.
2. Analyse approaches to continuous improvement and how this can be of benefit to an organisation	**Knowledge**: Approaches to continuous improvement and how they contribute to sustainable data products (for example, lessons learned, capturing best practice). The use of common frameworks for continuous improvement (for example, Lean, Kaizen, Six Sigma). The benefits to organisations of conducting continuous improvement.
3. Explore areas of advancement within data collection, handling and management to support organisational operations	**Knowledge**: Assess strategies employed by organisations to ensure they remain up to date on technical advancements in data science, data engineering and artificial intelligence (AI) (for example, testing new tools via proof of concept (PoC)). **Skills**: Review common operational practices and recommend potential opportunities for innovation. Emerging areas of innovation to consider: • tools • technologies • techniques. Investigate the emergence of new technologies being utilised within data science, data engineering and AI. Identify and apply strategies for continuing professional development (CPD) within the sector.
4. Explore the common methods and techniques used to communicate with stakeholders within organisations	**Knowledge**: Common methods and techniques used to communicate with stakeholders within organisations: • data visualisation • presentation skills/software • collaboration platforms • storytelling

Learning outcomes (LOs)	Mandatory teaching content
	• communication channels. How to communicate with a range of stakeholders, taking accessibility and diversity into consideration: • adapt language to suit audience • provide context • legislative requirements for the sharing of information • timing and frequency • feedback loops. **Skills**: Communicate data product outcomes to support implementation and integration, considering: • clear and structured presentation of information • adapting language and detail tailored to stakeholder requirements • alignment with organisational standards.

Grading criteria

Learning outcomes (LOs)	Pass	Merit	Distinction
LO1: Explore the principles of sustainable data products and environmental, social and governance responsibilities	**P1**: describe the principles of sustainable data products and organisational ESG responsibilities **P2**: identify and apply sustainable data solutions to support environmental strategies	**M1**: analyse the role of ESG strategies in the design, deployment, and maintenance of sustainable data products	**D1**: evaluate green data engineering practices and organisational responsibilities for environmental social governance
LO2: Analyse approaches to continuous improvement and how this can be of benefit to an organisation	**P3**: describe continuous improvement approaches within an organisational setting **P4**: evidence examples of good practice and reflection from a project	**M2**: critically analyse the importance of continuous improvement frameworks within business development	**D2**: evaluate the impact of implementing continuous improvement strategies on organisational resilience, quality, and employee engagement

Learning outcomes (LOs)	Pass	Merit	Distinction
LO3: Explore areas of advancement within data collection, handling and management to support organisational operations	**P5**: identify strategies to remain aware of current advancements in data engineering, data science, and AI relevant to organisational operations	**M3**: analyse how investment in technical advancement strategies supports competitive advantage and operational excellence	
	P6: apply horizon scanning techniques to review potential applications of emerging technologies within the sector		
LO4: Explore the common methods and techniques used to communicate with stakeholders within organisations	**P7**: review common operational practices and recommend potential opportunities for innovation	**M4**: critically analyse the importance of adapting communication styles, channels, and timing to different stakeholder needs to maximise engagement and understanding	
	P8: communicate data product outcomes to support implementation and integration		
	P9: discuss communication methods and techniques when communicating data products to different stakeholders in a diverse and accessible manner	**M5**: analyse and compare a range of communication methods and techniques for communicating data products to stakeholders	

NCFE assessment strategy

The key requirements of the assessment strategies or principles that relate to units in this qualification are summarised below.

The centre must ensure that individuals undertaking assessor or quality assurer roles within the centre conform to the assessment requirements for the unit they are assessing or quality assuring.

Knowledge LOs

- Assessors will need to be both occupationally knowledgeable and qualified to make assessment decisions.
- Internal quality assurers (IQAs) will need to be both occupationally knowledgeable and qualified to make quality assurance decisions.

Competence/skills LOs

- Assessors will need to be both occupationally competent and qualified to make assessment decisions.
- IQAs will need to be both occupationally knowledgeable and qualified to make quality assurance decisions.

The centre with whom the learners are registered will be responsible for making all assessment decisions. Assessors must be **contracted** to work directly with the centre, contributing to all aspects of standardisation. The centre must ensure a process of training is followed, including during induction and quality assurance activities. Occupationally competent and qualified assessors from the centre must use direct observation to assess practical skills-based outcomes.

Section 3: explanation of terms

This table explains how the terms used at **level 5** in the unit content are applied to this qualification (not all verbs are used in this qualification).

Act (as a role model)	Serve as a model in a particular behavioural or social role for another person to emulate.
Adapt (approaches)	Modify, adjust, make suitable for purpose.
Adhere to	Follow, keep, maintain, respect, abide by, give support to (for example, adhere to a strict code of practice).
Analyse	Break down the subject or complex situations into separate parts and examine each part in detail, identify the main issues and show how the main ideas are related to practice and why they are important (reference to current research or theory may support the analysis).
Critically analyse	This is a development of 'analyse' that explores limitations as well as positive aspects of the main ideas in order to form a reasoned opinion.
Apply	Use knowledge, understanding, or skills in a practical context or given scenario to achieve a specified outcome.
Ascertain	Find out for certain.
Assess	Estimate and make a judgement.
Automate	To program or configure a system or process to operate without manual intervention.
Clarify	Explain the information in a clear, concise way showing depth and understanding.
Collaborate (L7)	Work jointly with.
Communicate	To convey information, ideas, or feelings effectively through various means to a specific audience.
Compare	Examine the subjects in detail looking at similarities and differences.
Compare and contrast	Examine the subjects in detail, looking at similarities and differences and distinguish between (identify) striking differences.
Demonstrate	Apply skills in a practical situation and/or show an understanding of the topic.
Describe	Provide an extended range of detailed information about the topic or item in a logical way.
Design	Plan and create a detailed specification or solution for a system, process, or product to meet defined requirements.

Develop	Identify, build and extend a topic, plan or idea.
Discuss	To examine or talk about a subject in detail, considering different opinions or perspectives.
Distinguish between	Discuss identified differences between more than one item, product, object or activity.
Empower	Equip or supply with an ability; enable or permit.
Enable	Supply with the means, knowledge, or opportunity; make able.
Establish (L5 and L6)	Set up on a permanent basis; get generally accepted; place beyond dispute.
Evaluate	Examine strengths and weaknesses, arguments for and against and/or similarities and differences; judge the evidence from the different perspectives and make a valid conclusion or reasoned judgement; apply current research or theories to support the evaluation when applicable.
Critically evaluate	To make a judgement about the value, validity or effectiveness. The additional word 'critically' includes analysis, reflection and evidence-based reasoning.
Evidence	To provide facts, data, or observations that support a claim, conclusion, or action.
Explain	Apply reasoning to account for how something is or to show understanding of underpinning concepts (responses could include examples to support the reasons).
Explore	Investigate, examine, or discuss a topic, concept, or system in detail to gain a deeper understanding.
Facilitate (L6)	Make easier; assist the progress of.
Formulate (L5, L6 and L7)	Draw together; set forth in a logical way; express in systematic terms or concepts.
Give constructive feedback	Provide commentary that serves to improve or advance; be helpful.
Identify	Ascertain the origin, nature, or definitive characteristics of.
Implement (L5 and L6)	Put into practical effect; carry out.
Initiate	Originate/start a process.
Intervene effectively	Change an outcome.
Investigate	Detailed examination or study; enquire systematically.

Justify	Give a comprehensive explanation of the reasons for actions and/or decisions.
Mentor	Serve as a trusted counsellor or teacher to another person; help others succeed.
Monitor	Maintain regular surveillance.
Negotiate	Discuss with a view to finding an agreed settlement.
Perform	To carry out a task, action, or function as required or specified.
Produce	To create or generate something, such as a document, report, or output, according to given specifications.
Recognise	Acknowledge the validity of.
Recommend	Revisit and judge the merit of; endorse a proposal or course of action; advocate in favour of.
Refine	To improve something by making small changes, making it more precise, effective, or elegant.
Reflect on	Consult with oneself, recognising implications of current practice with a view to changing future practice.
Represent views of	Act as an advocate; speak, plead or argue in favour of.
Research (L5 and L6)	A detailed study of a subject to discover new information or reach a new understanding.
Resolve	Solve; settle; explain.
Review	To examine a subject or process, issue or body of work, with the aim of summarising key points, highlighting strengths and weaknesses and making recommendations or judgements.
Critically review	Revise, debate and judge the merit of.
Review and revise	Revisit, judge the merit of and make recommendations for change.
Secure	Make safe; obtain (information or evidence).
Set objectives (L6)	Identify the outcomes required.
Signpost	Point the way; indicate.
Summarise	Select the main ideas, arguments or facts and present in a precise, concise way.
Support	Strengthen, support or encourage; corroborate; give greater credibility to.
Triangulate (L7)	Identify three aspects to ensure validity.

| Work in partnership | Work in association with two or more individuals (this may include stakeholders, service users and/or carers). |

Section 4: support

Support materials

The following support materials are available to assist with the delivery of this qualification and are available on the NCFE website:

• Qualification Factsheet.

Useful websites

Centres may find the following websites helpful for information, materials and resources to assist with the delivery of this qualification:

• Datasets (Kaggle)
• Open data resources (UK Data Service)
• Dataset Search (Google)
• Earthdata (NASA)
• Find open data (data.gov.uk)

These links are provided as sources of potentially useful information for delivery/learning of this subject area. NCFE does not explicitly endorse these websites or any learning resources available on these websites.

Other support materials

The resources and materials used in the delivery of this qualification must be age-appropriate and due consideration should be given to the wellbeing and safeguarding of learners in line with your institute's safeguarding policy when developing or selecting delivery materials.

Products to support the delivery of this qualification may be available. For more information about these resources and how to access them, please visit the NCFE website.

Reproduction of this document

Reproduction by approved centres is permissible for internal use under the following conditions:

• You may copy and paste any material from this document; however, we do not accept any liability for any incomplete or inaccurate copying and subsequent use of this information.
• The use of PDF versions of our support materials on the NCFE website will ensure that correct and up-to-date information is provided to learners.
• Any photographs in this publication are either our exclusive property or used under licence from a third party:
    o They are protected under copyright law and cannot be reproduced, copied or manipulated in any form.
    o This includes the use of any image or part of an image in individual or group projects and assessment materials.
    o All images have a signed model release.

Contact us

NCFE
Q6
Quorum Park
Benton Lane
Newcastle upon Tyne
NE12 8BT

Tel: 0191 239 8000*
Fax: 0191 239 8001
Email: customersupport@ncfe.org.uk
Website: www.ncfe.org.uk

Appendix A: units

To simplify cross-referencing assessments and quality assurance, we have used a sequential numbering system in this document for each unit.

Knowledge-only units are indicated by a star. If a unit is not marked with a star, it is a skills unit or contains a mix of knowledge and skills.

Mandatory units

Unit number	Regulated unit number	Unit title	Level	Credit	GLH
Unit 01	K/651/6932	Data engineering principles and foundations	5	20	90
Unit 02	L/651/6933	Data architecture and management	5	15	70
Unit 03	M/651/6934	Data pipelines and automation	5	25	110
Unit 04	R/651/6935	Software development for data products	5	25	110
Unit 05	T/651/6936	Data analytics, business intelligence and incident management	5	20	90
Unit 06	Y/651/6937	Professional practice and continuous improvement in data engineering	5	15	70

The units above may be available as stand-alone unit programmes. Please visit the NCFE website for further information.

Change history record

Version	Publication date	Description of change
v1.0	January 2026	First published version