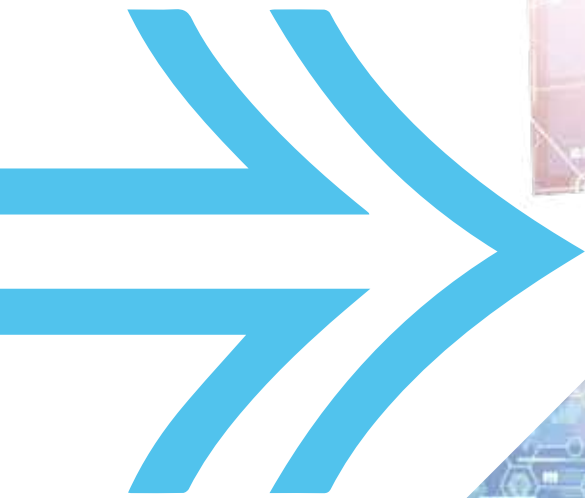




Apprenticeship Programme Guide

DATA ENGINEER

LEVEL 5 APPRENTICESHIP



QA.com

UNLEASH TRANSFORMATIVE INSIGHTS

Data Engineer Level 5 Apprenticeship

QA is the UK's leading data, digital and tech apprenticeship provider.

In partnership with our clients, we train apprentices – delivering economic and societal impact by launching and accelerating careers.

Unmatched Data Expertise

30+

years of data experience

37,000+

learners trained in data

98%

highest tech apprenticeship pass rate

Unrivalled Learning Excellence



Over 1000 hands-on labs



Gold Award Learning Tech Awards



Partnerships with major tech vendors

CONTENTS

Role Profile	5
Job Role Suitability	6
Entry Requirements	8
Develop the Best Data Engineer	9
The Learner's Journey	11
Modules	13
Gateway, EPA and Qualifications	18
Learning Outcomes	20
How is the EPA graded?	21



ROLE PROFILE

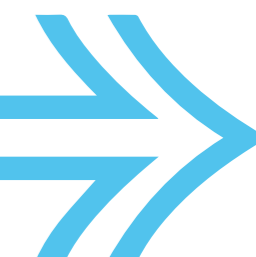
Data Engineer

Data Engineers are pivotal in tackling the critical challenges related to data management, processing, and analytics. Their importance has surged due to the escalating volume and complexity of data, coupled with the growing need for real-time insights and advanced analytics.

As organisations increasingly pivot towards data-driven decision-making and embrace transformative technologies like Machine Learning (ML) and Artificial Intelligence (AI), aligning intricate data infrastructures with strategic business objectives has become more imperative than ever.

They need:

- Attention to detail
- Strong maths and programming skills
- Effective communication, teamwork, and task/time management
- Ability to present information for technical and non-technical audiences



JOB ROLE SUITABILITY

Employers need to evaluate whether a candidate (a new hire or existing employee) is in a suitable job role to successfully complete their programme.

This checklist is designed to assist you in determining whether an apprentice will be able to demonstrate the necessary Data Engineer duties during their programme.

Checklist

- 1 Build and optimise automated data systems and pipelines considering data quality, description, cataloguing, data cleaning, validation, technical documentation and requirements.
- 2 Integrate, support and manage data using standalone, distributed and cloud-based platforms. To ensure efficient, sustainable and effective provision of data storage solutions.
- 3 Support the identification and evaluation of opportunities for data acquisition and data enrichment.
- 4 Select and use appropriate tools to process data in any format, such as structured and unstructured data and in any mode of delivery, such as streaming or batching. Adapt to legacy systems as required.
- 5 Ensure resilience is built into data products against business continuity and disaster recovery plans, and document change management to limit service outages. Support and respond to incidents through the application of technology and service management best practice including configuration, change and incident management.
- 6 Analyse requirements, research scope and options and present recommendations for solutions to stakeholders.
- 7 Support the implementation of prototype or proof-of-concept data products within a production environment.
- 8 Maintain data solutions as continually evolving products, to service the organisation, user or client requirements. Collaborate with technical support teams and stakeholders from implementation to management.
- 9 Working within compliance and contribute towards data governance, organisational policies, standards, and guidelines for data engineering.
- 10 Monitor the data system to meet service requirements to enable solutions such as data analysis, dashboards, data products, pipelines, and storage solutions.
- 11 Keep up to date with engineering developments to advance own skills and knowledge.

ENTRY REQUIREMENTS

The entry requirements for this programme are as follows.

Standard entry (any of the below qualifications):

- Two A levels in one or more similar subject(s)
- Level 3 apprenticeship in a similar subject
- International Baccalaureate at level 3 in a similar subject
- BTEC Extended Diploma in a similar subject

OR

- Work experience (2-3 years) in a similar subject-related role

Foundation experience with programming languages (such as Python) required.

English Language and Maths requirements:

- GCSE Grade 4 (C) or above (or equivalent e.g. level 2 in functional skills)

DEVELOP THE BEST DATA ENGINEER

We support learners to discover new skills, practise them in secure environments, and then apply them in the workplace – directly benefiting their organisation. This is achieved through immersive and accessible learning by integrating human expertise with technology.

The DEL5 Apprenticeship embraces a guided discovery approach to learning – focusing on learners and their contexts. This empowers apprentices to take ownership of their learning journey through the “Discover, Practise, Apply” model.

Our data coaches and learner service team provide tailored support for a diverse range of needs, fostering individual growth and success.

This is enhanced by self-paced training through Cloud Academy. Dive into specialisms with 40,000+ course hours, 1,000+ cloud and coding labs, 100+ AI labs and courses, along with the opportunity to earn over 60 certifications.



DISCOVER

Learn new skills



PRACTISE

Hone new skills

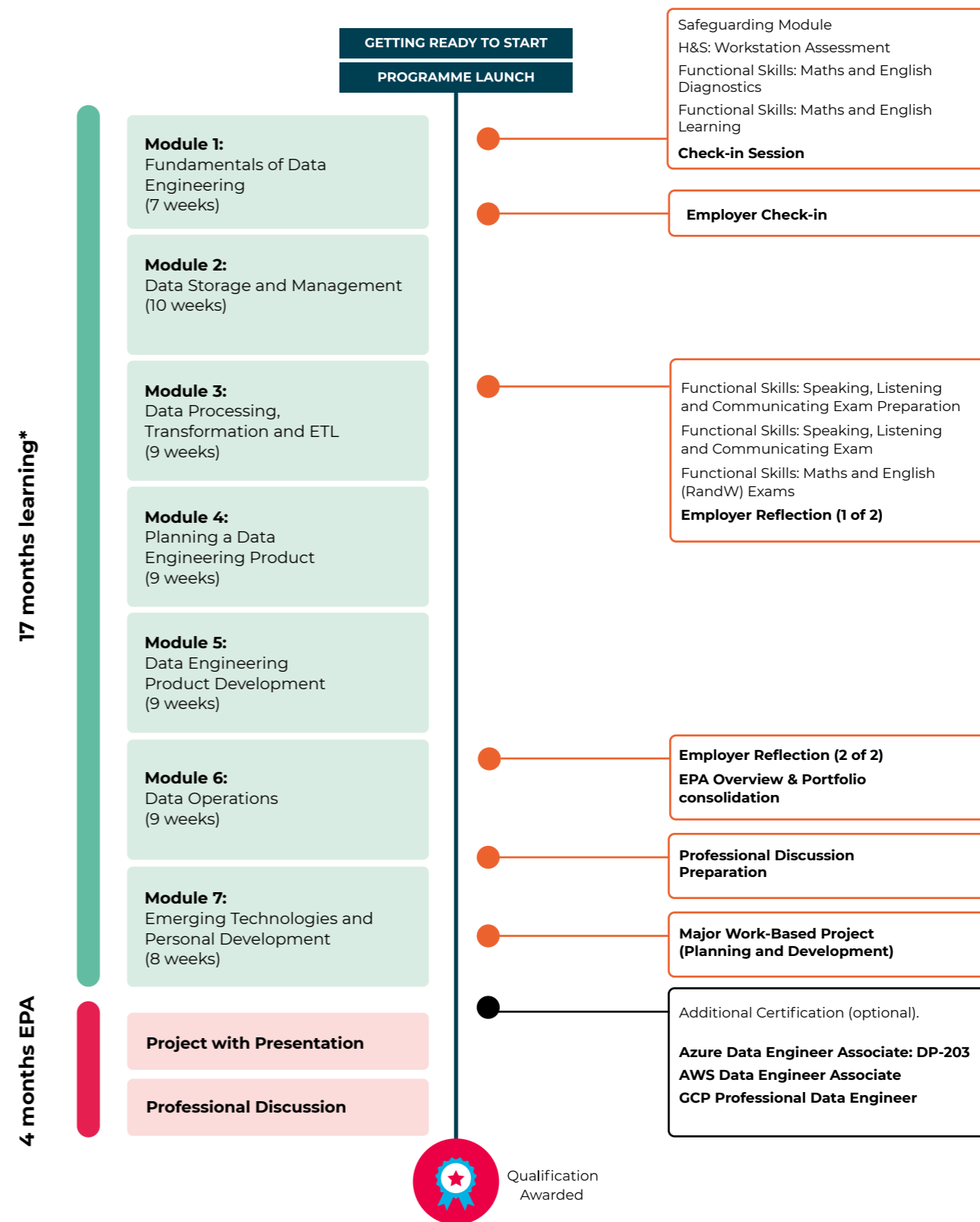


APPLY

Take new skills back into the workplace

THE LEARNER'S JOURNEY

Typical programme duration: 17 months (+4 months End-Point Assessment)



* Live sessions: 23 days

Level 2 functional skills, English and Maths must be passed as part of the programme (if not already) and certificates presented, prior to taking the end-point assessment.

GETTING STARTED MODULE

The programme launch is a short session introducing learners to the apprenticeship outline, structure, and assessment.

This features a day 1 learning activity, familiarising apprentices with the role of a data engineer.

Programme Launch

This module will be delivered virtually – covering:

- Overview of programme, workflow & structure
- Assessment, certification, & qualifications
- Setting clear expectations for the programme
- Time commitment & calendar planning
- Bud, Cloud Academy & technology requirements

During the day 1 learning activity, learners will explore Data Engineering with an introduction to:

- Cleansing data
- Executing a research activity
- Reflecting on the work completed

TECHNICAL MODULES

These modules provide in-depth knowledge and skills essential for a Data Engineer. Following each module, apprentices will apply their newly acquired learning to their ongoing projects at work.

Module 1: Fundamentals of Data Engineering

This module delves into the growing importance of data engineering and starts building the skills needed to identify risks and tackle real-world data challenges. It introduces learners to the core concepts, principles, practices, and tools used to work with large, complex data sets and stakeholders in an efficient and ethical way.

Topics covered:

- Fundamentals of Data Engineering
- Data Types
- Data Sources
- Data Structures
- Data Formats of Serialisation Techniques
- Data Compression
- Data Modelling
- Normalization & Denormalization
- Data Quality
- Data Storage
- Data Engineering Lifecycles
- Ethical Practices in Data Management
- Data Handling & Secure Data Management
- Data Engineering Tools & Applications
- Descriptive, Predictive, and Prescriptive Analytics

Live session: 2 days

Module duration: 7 weeks

Module 2: Data Storage and Management

This module explores data storage – the cornerstone for managing data that meets standards of accessibility, scalability, and security. This includes RDBMS, such as: SQL Fundamentals, SQL Server, NoSQL databases, data modelling, distributed systems, and cloud deployment.

Hands-on labs provide learners with practical experience in in database design, access management, issue communication, and querying data using SQL.

Topics covered:

- Relational Database Management Systems (RDBMS)
- My SQL
- SQL Fundamentals
- SQL Joins
- Database Design & Modelling
- No SQL
- Distributed File System
- Horizontal & Vertical Partitioning
- Sharding
- Data Replication & Backup Strategies
- Data Architecture and Cloud Platforms
- Querying & Manipulating Data

Live session: 3 days

Module duration: 10 weeks

Module 3: Data Processing, Transformation and ETL

This module covers processes that prepare raw data for analysis, reporting, or other downstream applications. It introduces techniques to optimise data ingestion, such as enhancing data quality and managing complex data ecosystems.

Learners explore Python tools and libraries, ETL/ETL, CI/CD, pipeline orchestration, cloud solutions, and use hands-on labs to participate in exercises for data cleansing.

Topics covered:

- Python for Data Processing & Transformation
- Data Quality & Cleansing Techniques
- Data Integration
- Batch vs Real-time Processing
- Introduction to Data Integration Patterns & Architecture
- ETL and ELT Architecture & Tools
- On-demand Cloud Computing Platforms
- Data Engineering in the Cloud
- Data Lineage & Pipeline Orchestration using Azure
- CI/CD ETL Processes
- Building Complete Data Pipelines

Live session: 4 days

Module duration: 9 weeks

Module 4: Planning a Data Engineering Product

This module examines the process of developing data products. It starts with defining business needs and evaluating requirements, then shifts to agile frameworks for user requirements and data product design. It underlines best practises in software development and gap analysis in existing tools.

Learners plan through testing data quality, automation, and risk analysis – with considerations on cost, security, scalability, net-zero, governance, compliance, and strategy.

Topics covered:

- Best Practices in Software Development
- Software Development Lifecycle
- Introduction to Agile and DevOps
- Containerisation
- Data Product Tools & Technologies
- Sustainable Data Product Design
- Evaluating Organisational Requirements
- Costing
- Risk Management
- Root Cause Analysis
- Version Control
- Communication & Documentation

Live session: 4 days

Module duration: 9 weeks



Module 5: Data Engineering Product Development

This module covers the fundamentals to successfully build and test data products. Learners are familiarised with sustainable development practices and methods for continuous improvement to create data products with longevity that drive business outcomes.

It delves into structured data extraction, integration platforms, interfaces, data ingestion optimisation, automation of pipelines, testing, security, and prototype evaluation.

Topics covered:

- Software Engineering Principles
- Programming for Data Products
- Structured Data Extraction
- Data Integration Platforms
- Optimisation of Data Ingestion
- Automation of Data Pipelines
- Data Development Frameworks
- Prototyping, Testing & Debugging
- Technical Documentation
- Technical Debt Management

Live session: 4 days

Module duration: 9 weeks

Module 6: Data Operations

This module covers data operations – the foundation for streamlining the flow of data and promoting a culture of continuous improvement in analytics.

It introduces learners to best practices, technology service management, incident response, troubleshooting, release methods, monitoring, data pipeline management, data product evaluation, quality assurance, root cause investigation, forecasting tools, machine learning model development, and data governance.

Topics covered:

- Data Pipeline Deployment & Management
- Optimisation & Automation
- Forecasting & Monitoring Tools
- Troubleshooting & Incident Response
- Analysis & Root Cause Investigation
- Problem Management
- Business Continuity Operations
- Data Product Evaluation, Development & Continuous Improvement
- Quality Assurance
- Presenting a Data Product to Stakeholders

Live session: 4 days

Module duration: 9 weeks

Module 7: Emerging Technologies and Personal Development

This module immerses learners into the future of Data Engineering – it explores the latest technologies and strategies revolutionising data management and analysis.

Learners are empowered to advance their personal development by embracing innovative, data-driven approaches and insights.

Topics covered:

- Feature Engineering for Machine Learning
- Predictive Modelling
- Visualisation & Reporting
- Monitoring & Managing Machine Learning Pipelines
- Generative AI Solutions
- Large Language Models (LLM)
- IoT & Edge Computing

Live session: 2 days

Module duration: 8 weeks

GATEWAY, EPA AND QUALIFICATIONS

Gateway and End-Point Assessment

This final stage gets learners ready to go through the “gateway”. The apprenticeship gateway is an internal QA process – ensuring that your learners’ work is prepared for assessment by the End-Point Assessment Organisation (EPAO). This serves to increase their chances of success.

At this pre-gateway stage learners:

- Submit a project brief
- Consolidate & submit their portfolio
- Complete mock professional discussions to prepare for EPA

In addition, learners are required to complete the following:

- KSBs (Knowledge, Skills, Behaviours) clearly evidenced
- Functional skills exams (except exempt apprentices)

Once learners have met all the above criteria, they go through the gateway. When approved, it typically takes 4 months from gateway to achievement.

During this time, learners:

- Complete their professional discussion, which is underpinned by their portfolio
- Complete their project report and presentation with questioning

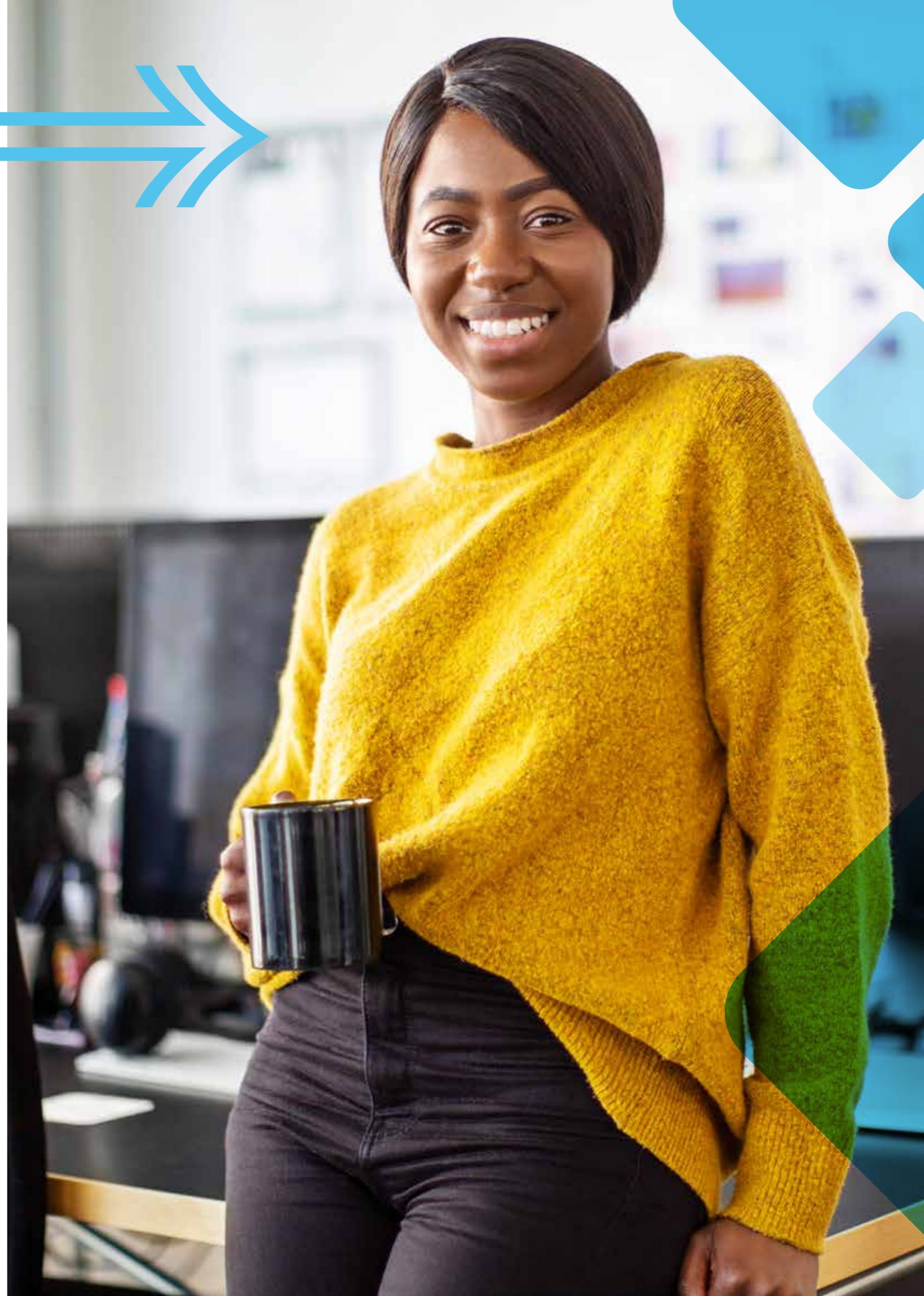
EPA prep 4 weeks + EPA period 4 months

Qualifications



On top of achieving their Data Engineer L5 apprenticeship award, learners have the option to complete:

- Azure Data Engineer Associate: DP-203
- AWS Data Engineer Associate
- GCP Professional Data Engineer



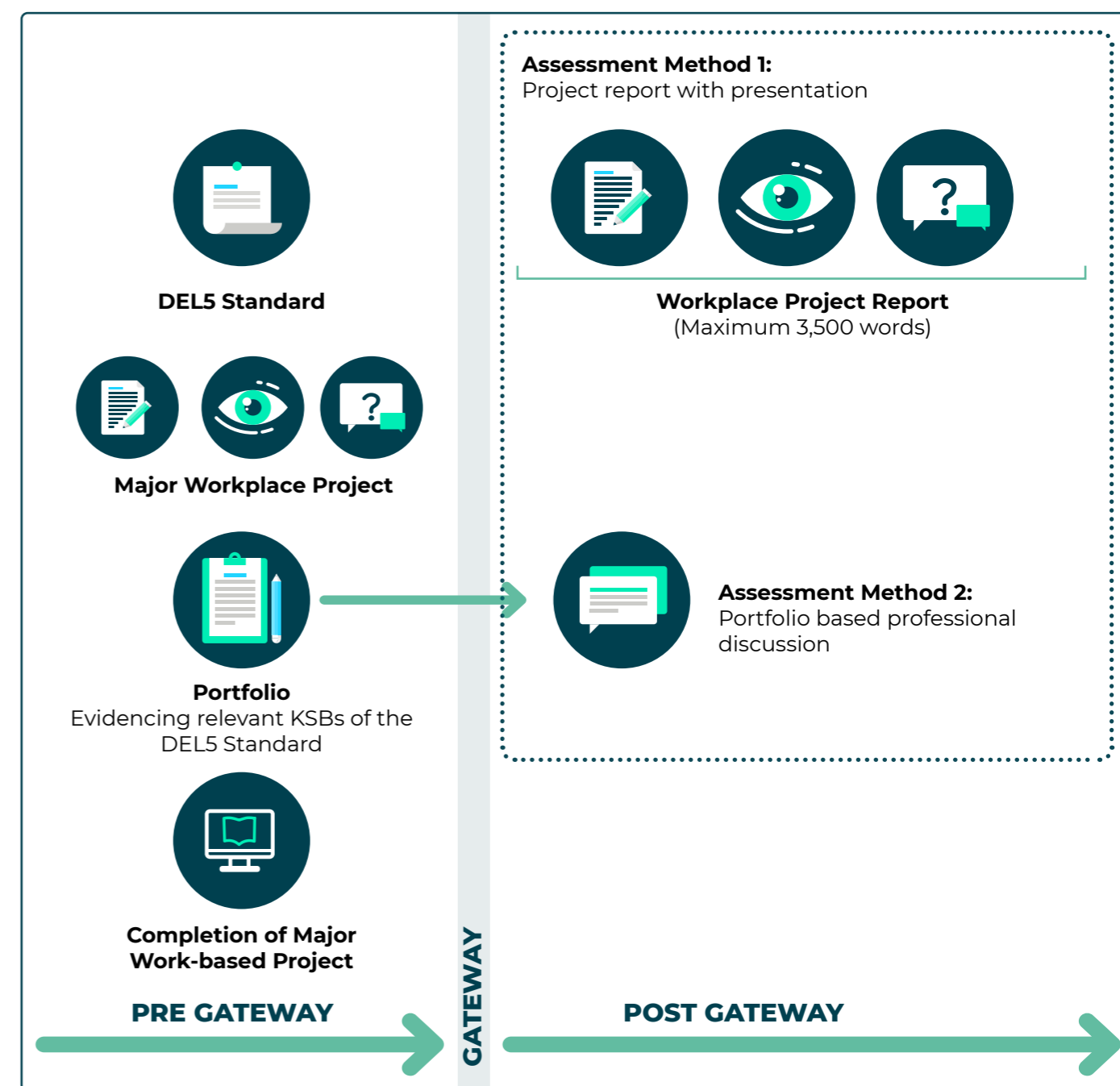
LEARNING OUTCOMES

Apprentices will be assessed on 3 key areas; their ability to convey knowledge, demonstrate practical skills, and exhibit professional workplace behaviour. These are developed throughout the apprentice's learning journey, with the goal of displaying these competencies during their EPA.

These KSBs contribute to rounded development – aligning with IfATE occupational standards that prioritise the significance of both technical and soft skills in the workplace.

Here is the full breakdown of what Data Engineer apprentices will be assessed against: [DE5 Standard](#)

HOW THE EPA IS GRADED



FOR MORE INFORMATION, PLEASE CONTACT

0333 060 7701
qa.com/contact



V1.0 APRIL 2024

This information is correct as of publishing in April 2024