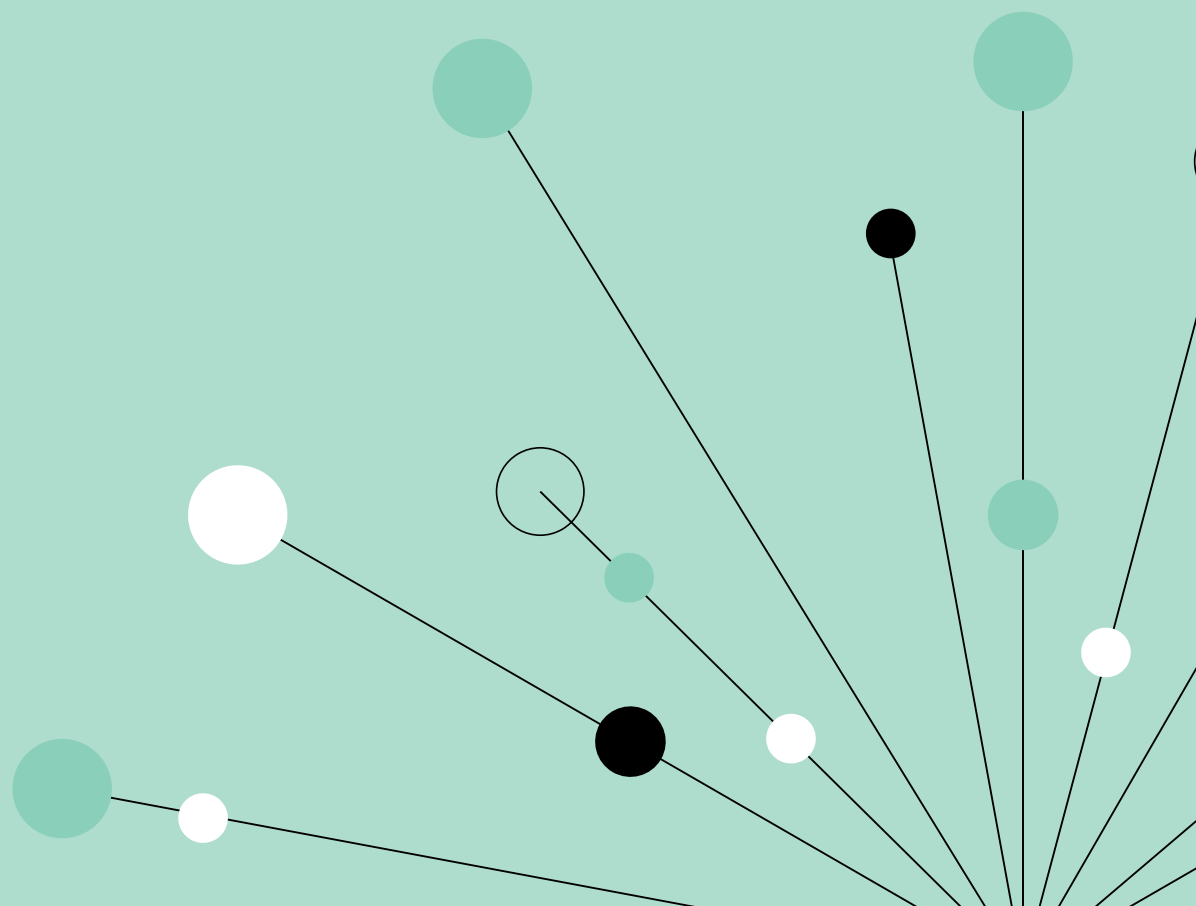


# Data Engineer

Level 5 Apprenticeship

Programme Guide





## Why QA?

Endorsed by 4,000+ global clients, we are the leader in applied and cohort-based learning academies.

Today's biggest technological shifts are shaped by AI, cloud, and data.

In every technology revolution, there are winners and losers – and teams with applied skills make all the difference. We believe you can't change an organisation unless you change the capabilities of its people and ensure human and machine intelligence work together.

### Success in numbers:

---

**35+**

Years of training  
experience

**1,000+**

AI, cloud & coding  
hands-on labs

**40,000+**

Careers launched  
& accelerated

**£500M+**

Levy spend  
invested

**24 hours**

Feedback time for  
submissions

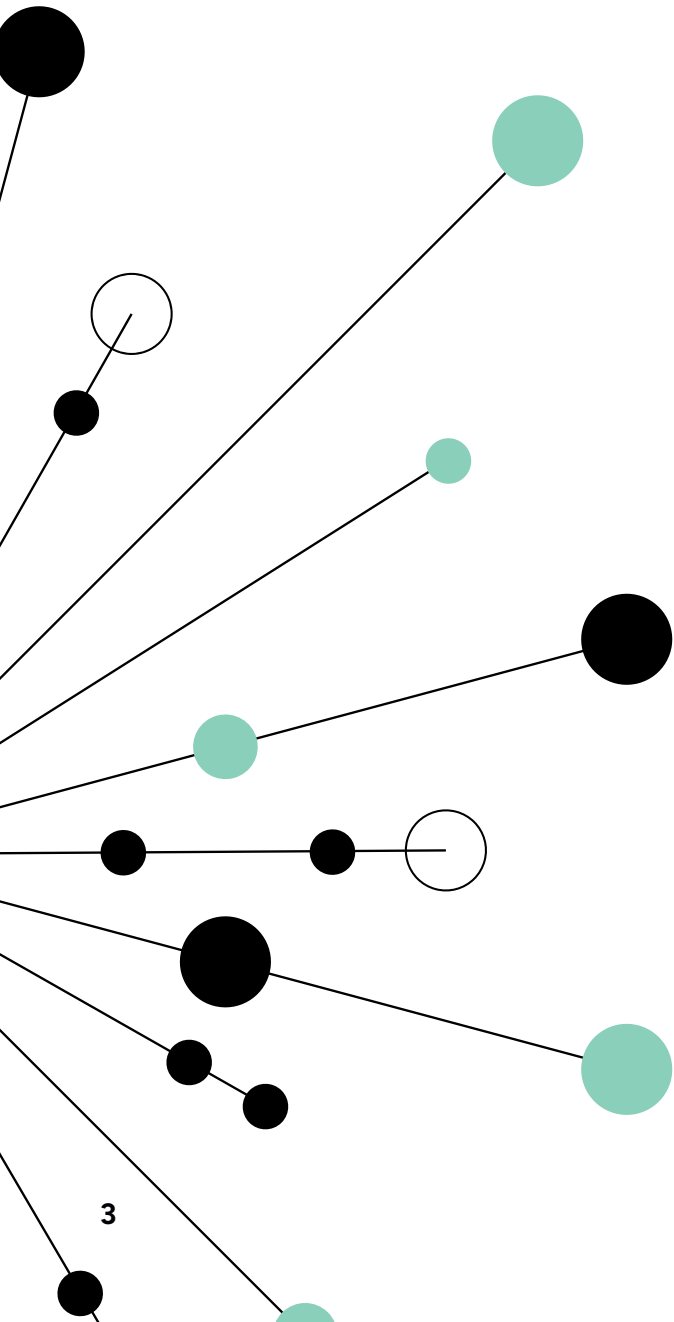
**<1 minute**

Response time to  
learner queries



**Ready to explore how QA can support you?**

Let's dive in!



# Contents

Creating Change	04
Digital by Design	05
Programme Overview	06
Learner Journey	07
Modules	08
Tools and Technologies	11
End-Point-Assessment	12

# Creating Change

Data engineering is the bridge between raw data and actionable insights.

This programme equips your organisation with the critical skills needed to build the infrastructure and systems that transform data into valuable information for downstream activities and strategic decision-making.

Our apprenticeships drive business results by empowering organisations to apply skills consistently at speed and scale.



## Empower your data community

Graduate from a cost-centre to profit-centre

---



## Harness AI and machine learning

Maximise potential of emerging technologies

---

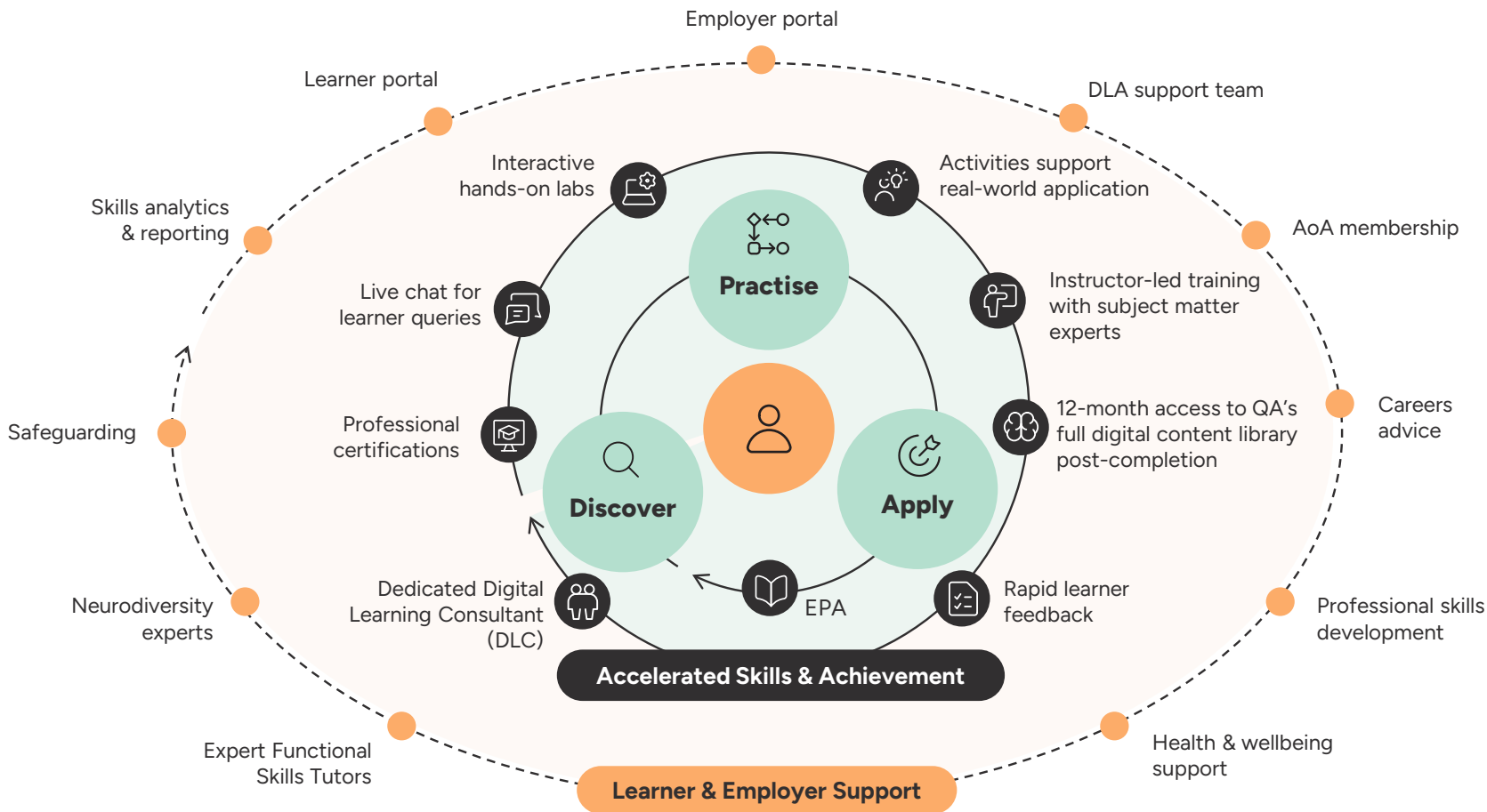


## Unlock time and cost savings

Stem revenue loss by capturing missed insights

# Digital by Design

Our market-leading approach accelerates skill development and achievement through our **Discover, Practise, Apply** methodology, ensuring that both learners and employers are fully supported throughout their programme.



## Discover

Leveraging QA's learning platform, learners follow a development path focused on their job role.



## Practise

Learners come together for instructor-led training sessions, practising their skills through hands-on labs and sandboxes in a safe environment while collaborating with peers.



## Apply

These practiced learnings are applied on the job through work-based activities at key and sequenced stages, fully supported and reviewed by the specialist DLC team.

# Programme Overview



**Details of standard:** [Data Engineer](#)



**Total cost:** £19,000



**Programme duration:** 21 months



**Live instructor sessions:** 23 days

Delivered in collaboration with our strategic vendor partners:



Experience QA's self-paced learning platform with interactive labs and configurable learning.



Data  
Collection



Data  
Storage



Data  
Integration



Data  
Processing



Data  
Analysis Support



Scalability  
and Performance

# Learner Journey

This Data Engineer programme integrates live and online workshops with self-paced learning, employing a guided discovery approach for individual learner contexts.

Learners are assigned a Digital Learning Consultant (DLC) for personalised coaching and support. These specialists ensure their successful progress, wellbeing, and readiness for assessments.

## Modules – 17 months

**Module 1:** Fundamentals of Data Engineering (7 weeks)

**Module 2:** Data Storage and Management (10 weeks)

**Module 3:** Data Processing, Transformation and ETL (9 weeks)

**Module 4:** Planning a Data Engineering Product (9 weeks)

**Module 5:** Data Engineering Product Development (9 weeks)

**Module 6:** Data Operations (9 weeks)

**Module 7:** Emerging Technologies and Personal Development (8 weeks)

Work-Based Project

## EPA – 4 months

Project Report with Presentation

Professional Discussion

## Optional Certifications

Azure Data Engineer Associate: DP-203

AWS Data Engineer Associate

IfATE DE5 Qualification Award 



# Modules

Following each module, learners apply their newly acquired knowledge and skills to ongoing work projects.

---

## 01

### **Module 1:** Fundamentals of Data Engineering

---

Delves into the importance of data engineering and develops skills to identify risks and solve real-world data challenges.

Covers core concepts, principles, practices, and tools for managing large data sets and collaborating with stakeholders efficiently and ethically.

#### **Topics:**

- Data Types, Sources, Quality, Structures, Compression, Storage, Formats of Serialisation, Modelling, Normalisation & Denormalisation, Handling & Secure Management
- Data Engineering Lifecycle
- Data Engineering Tools & Applications
- Ethical Practices in Data Management
- Descriptive, Predictive & Prescriptive Analytics

**Live Instructor Sessions:** 2 Days



# 02

## **Module 2:** Data Storage and Management

---

Explores data storage, the cornerstone of managing data that meets standards of accessibility, scalability, and security.

### **Topics:**

- Relational Database Management Systems
- MySQL
- SQL Fundamentals
- SQL Joins
- Database Design & Modelling
- NoSQL
- Distributed File System
- Horizontal & Vertical Partitioning
- Sharding
- Data Replication & Backup Strategies
- Data Architecture & Cloud Platforms
- Querying & Manipulating Data

**Live Instructor Sessions:** 3 Days

# 03

## **Module 3:** Data Processing, Transformation and ETL

---

Covers processes that prepare raw data for analysis, reporting, or other downstream activities.

### **Topics:**

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

**Live Instructor Sessions:** 4 Days

# 04

## **Module 4:** Planning a Data Engineering Product

---

Examines the process and methodologies in developing data products.

### **Topics:**

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

**Live Instructor Sessions:** 4 Days

# 05

## **Module 5:** Data Engineering Product Development

---

Covers the fundamentals to successfully build and test data products.

### **Topics:**

- 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 Instructor Sessions:** 4 Days

# 06

## **Module 6:** Data Operations

---

Explores the foundation for streamlining the flow of data and promoting a culture of continuous improvement in analytics.

### **Topics:**

- 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 Instructor Sessions:** 4 Days

# 07

## **Module 7:** Emerging Technologies and Personal Development

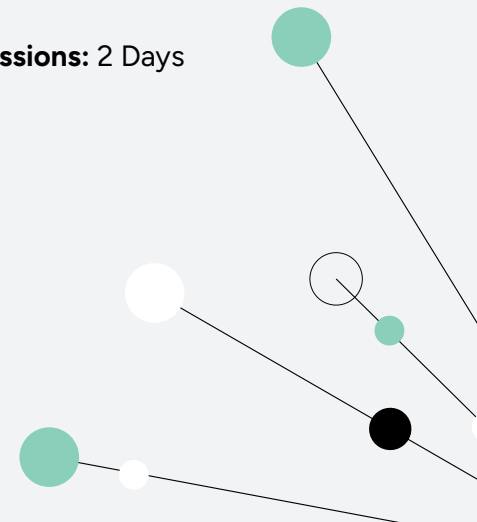
---

Highlights cutting-edge technologies and strategies revolutionising data management and analysis.

### **Topics:**

- 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 Instructor Sessions:** 2 Days



# Tools and Technologies

## Databases

- SQL Server
- NoSQL

## Data Warehousing and Processing

- SQL Server-based Data Warehouses
- Synapse Analytics
- Python
- PySpark
- SQL (Structured Query Language)
- Apache Spark
- Azure Databricks

## Cloud Platforms

- Azure
- AWS

## Data Engineering Services

- Azure Storage
- Azure Synapse
- Azure Data Flows
- Azure Data Factory
- Azure Stream Analytics
- Azure Databricks
- Azure Data Governance
- AWS Storage
- AWS Athena
- AWS Glue
- AWS EMR
- AWS MSK

## Security and Governance

- OpenSSL
- Microsoft Purview
- Identity & Access Management (IAM) Tools
- Data Anonymization Tools

## Business Intelligence and Visualisation

- Power BI

## Version Control

- Git

## Containerisation and Orchestration

- Docker
- Kubernetes

## Development Environments

- Jupyter Notebooks

## Infrastructure as Code

- Terraform

# End-Point-Assessment

We ensure all learners are fully prepared for their End-Point-Assessment (EPA) through our internal gateway process, maximising their success rates.

## Assessment criteria:

### 01

#### Knowledge

Ability to convey knowledge effectively.

### 02

#### Skills

Demonstrate practical skills with confidence.

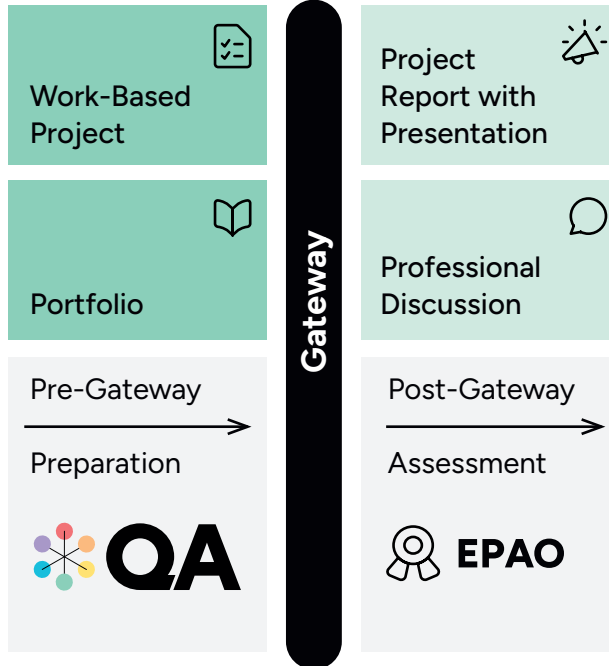
### 03

#### Behaviour

Exhibit professional workplace behaviour.

Explore the detailed assessment criteria within the **Data Engineer standard**.

## EPA process:




**Project Report with Presentation:** Prepare a project report, demonstrate achievements and knowledge, and participate in a Q&A.

**Professional Discussion:** Engage in a formal two-way conversation to showcase knowledge, skills, and behaviours.



# Ready to partner with us?

## Let's talk:

 0113 220 7150

 [qa.com/contact](https://qa.com/contact)

This information is correct as of publishing in August 2024

V1.0 2024

Funded by



Funded by

