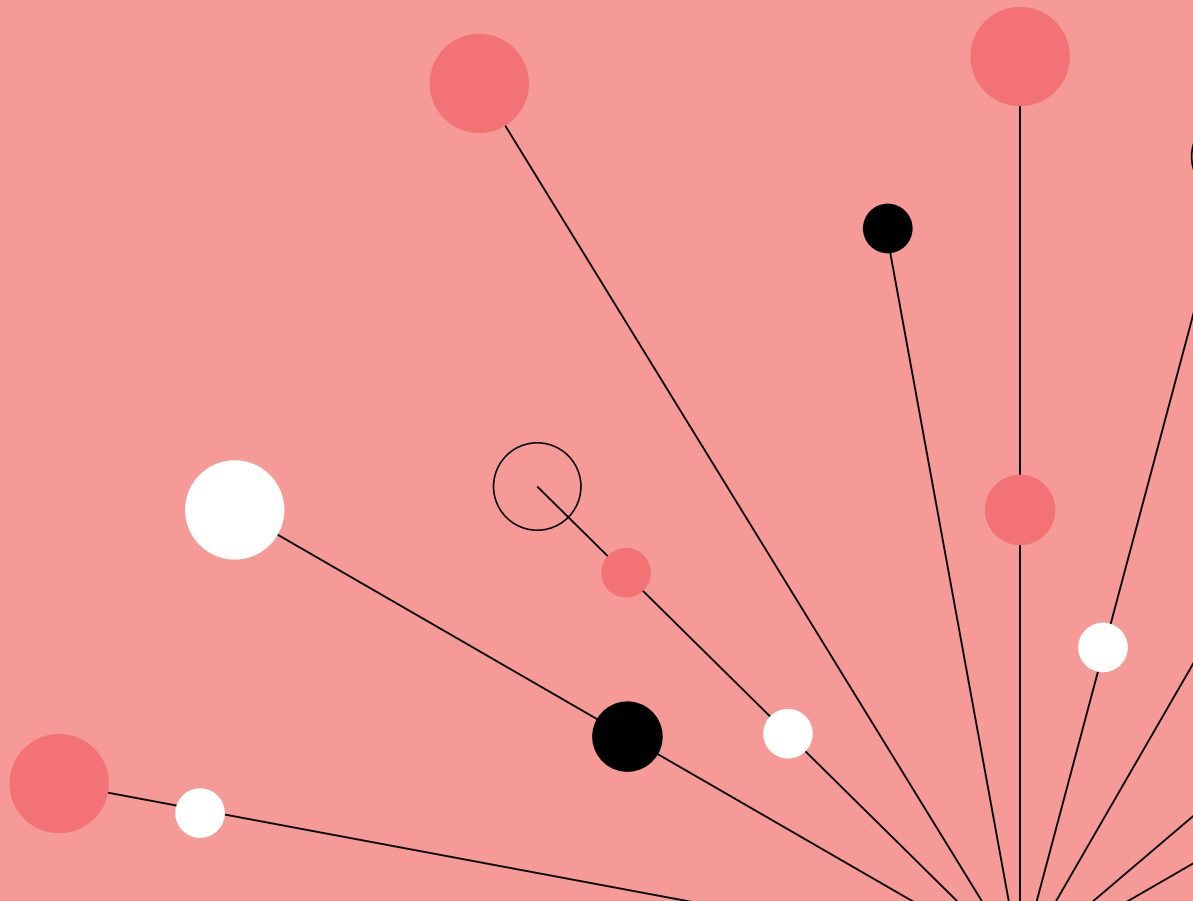
QA | databricks

Apprenticeships England

Databricks Engineer

Level 5 Apprenticeship

Programme Guide



QA.com

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:

40+
Years of training experience

1,000+
AI, cloud & coding hands-on labs

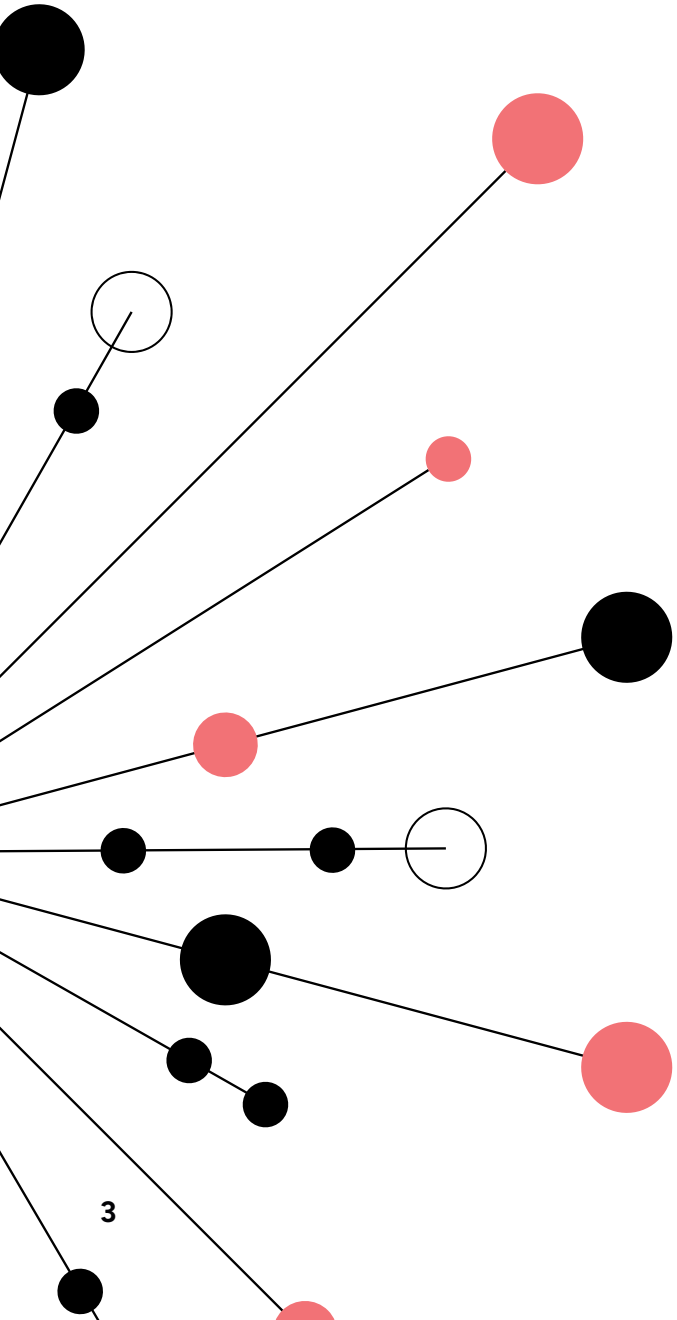
50,000+
Careers launched & accelerated

£500m+
Levy funds 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!



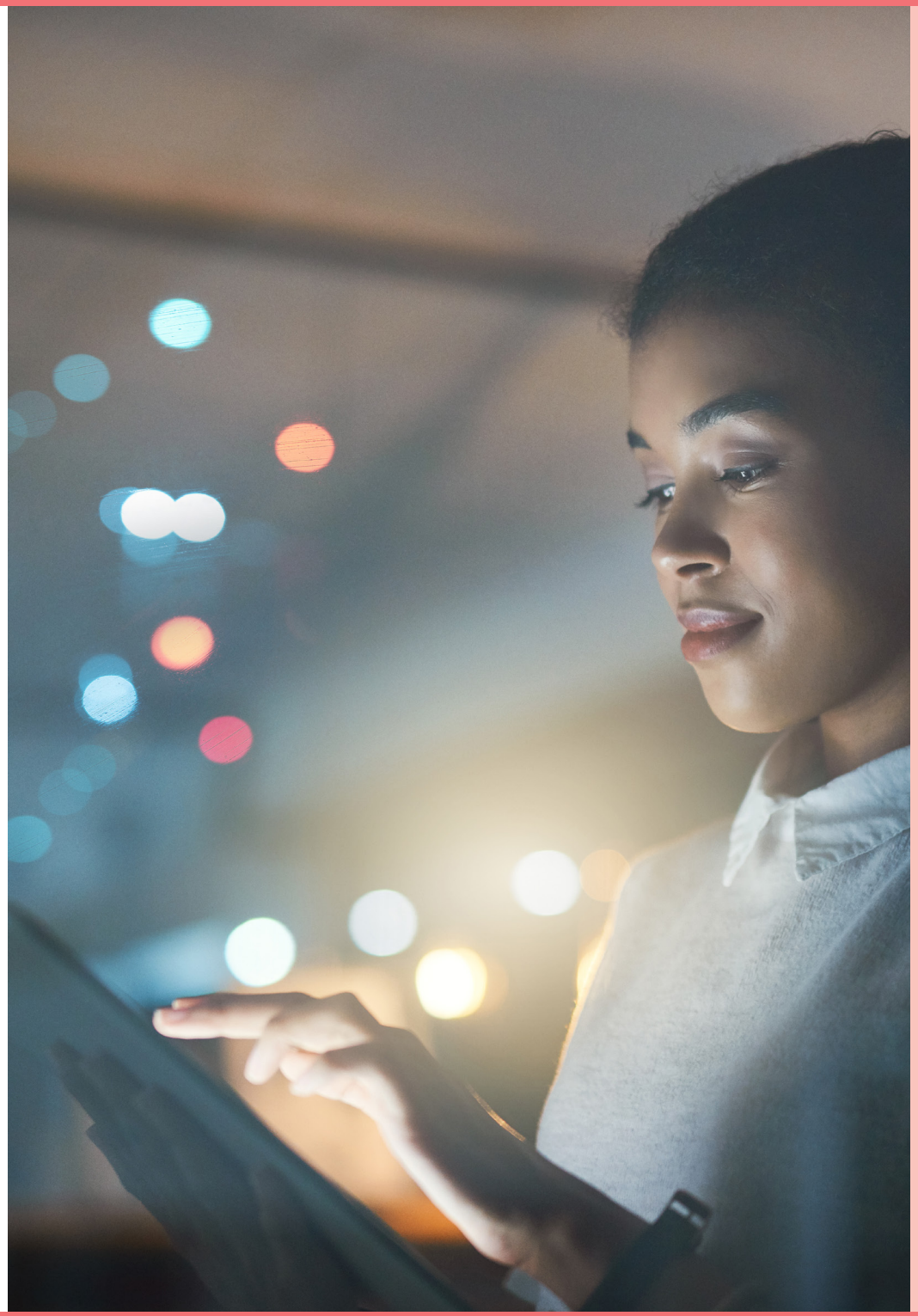
Contents

Creating Change

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

This programme equips your organisation with the essential skills to leverage the Databricks Data Intelligence Platform, turning data into actionable insights to drive strategic decision-making and downstream success.

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



EMEA-Authorised Training Partner

Level Up with the First Databricks Apprenticeship.

Empower your data community

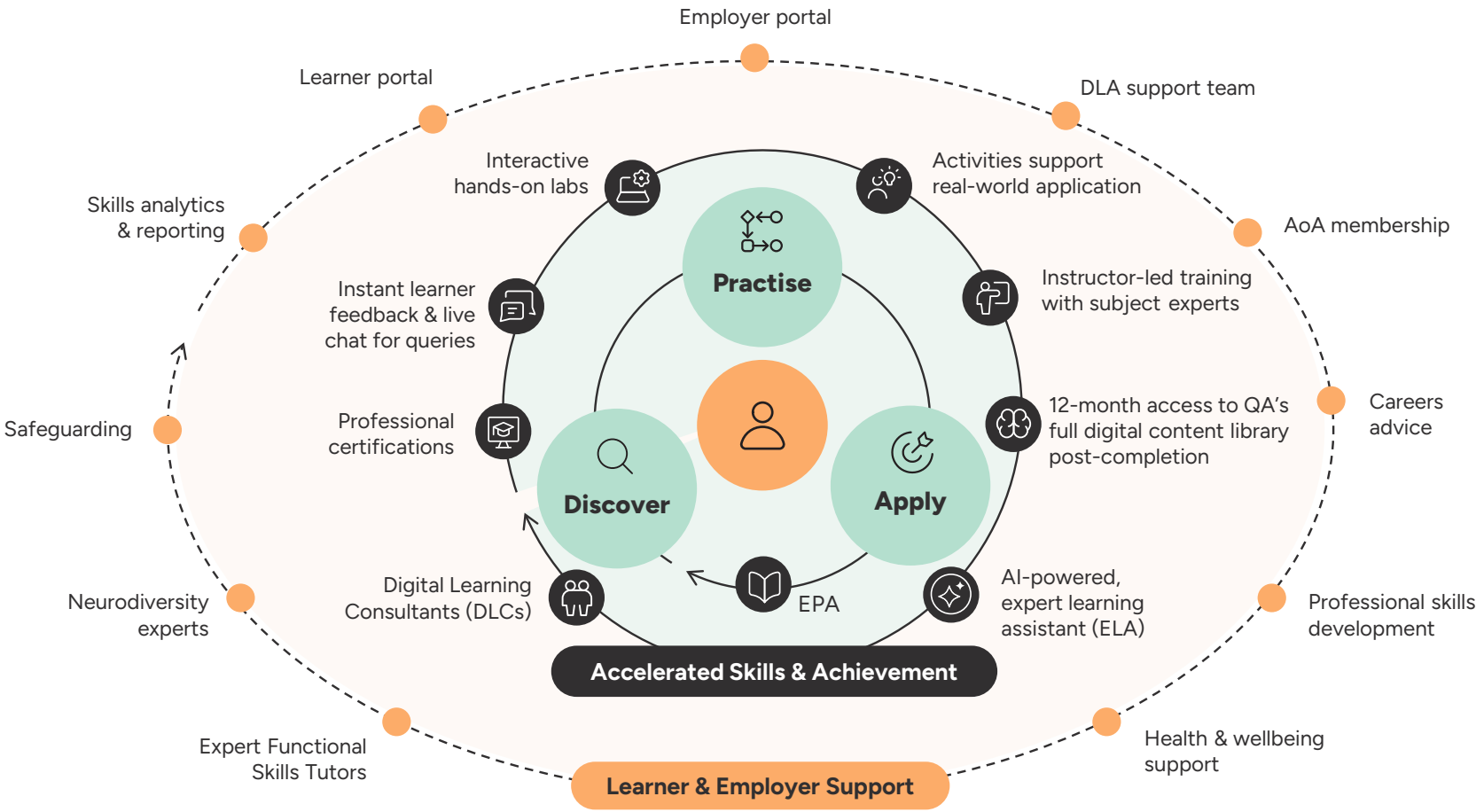
Graduate from a cost-centre to profit-centre.

Harness AI and machine learning

Maximise potential of emerging technologies.

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 join instructor-led sessions, practise skills in hands-on, risk-free labs, and collaborate with peers.

Apply

Learnings are applied on the job through work-based activities at key stages, supported and reviewed by specialist DLCs.

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:

databricks

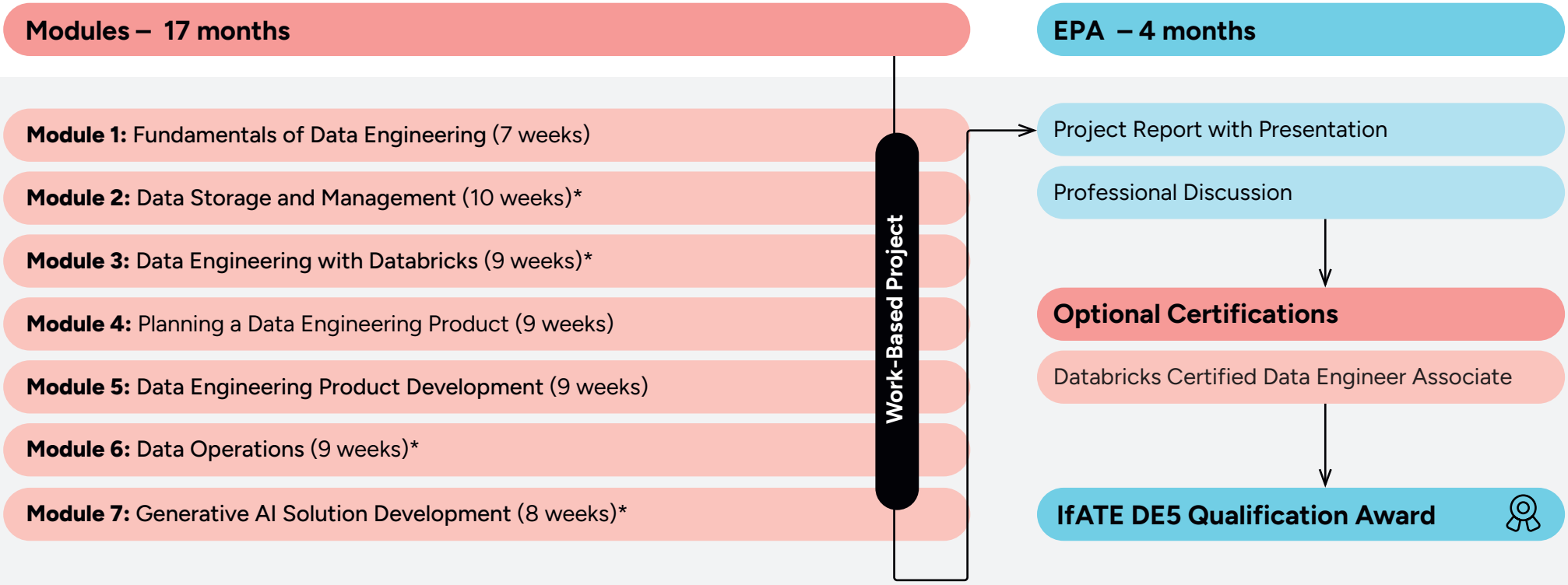
Experience QA's self-paced learning with interactive labs and AI-powered learning assistant.

- Data Collection
- Data Storage
- Data Integration
- Data Processing
- Data Analysis Support
- Scalability and Performance

Learner Journey

This Databricks 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 2, 3, 6 & 7 feature official Databricks training.

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:

- Databricks Fundamentals
- Get Started with Databricks for Data Engineering
- Relational Databases, SQL Joins & MySQL
- SQL Fundamentals
- Database Design & Modelling
- NoSQL
- Distributed Systems & Sharding
- Horizontal & Vertical Partitioning

Live Instructor Sessions: 3 Days

03

Module 3:
Data Engineering with Databricks

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

Topics:

- Data Engineering with Databricks
- Data Ingestion with Delta Lake
- Build Data Pipelines with Delta Live Tables
- Deploy Workloads with Databricks Workflows
- Data Quality & Cleansing
- Batch & Real-Time Processing
- Data Integration & Architecture Patterns
- ETL & ELT with SSIS
- CI/CD for Data Pipelines
- Data Management & Governance with Unity Catalog
- Data Lineage & Orchestration
- Cloud Platforms & Data Engineering

Live Instructor Sessions: 4 Days

04

Module 4:
Planning a Data Engineering Product

Examines the processes, methods, and strategic considerations of developing scalable, secure, and sustainable 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:

- Data Extraction & Ingestion Optimisation
- Pipeline Automation & Integration Platforms
- Interfaces & User Requirements
- Security & Testing
- CI/CD & Data Pipelines
- Data Cleansing with Python
- Building Pipelines with Azure DevOps
- Docker for Python Applications
- Version control with GitHub

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

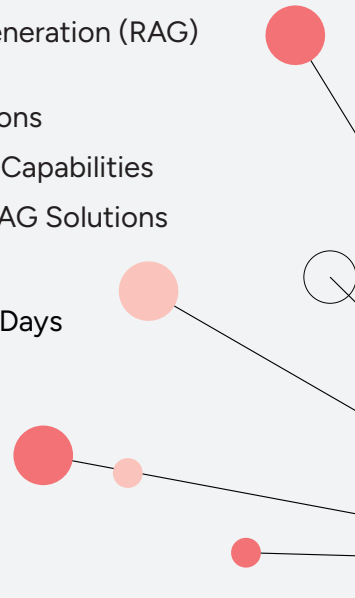
Module 7:
Generative AI Solution Development

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

Topics:

- Introduction to Data Science & Machine Learning
- Generative AI Solution Development
- Data Preparation & Management Best Practices
- Ethical Considerations in Generative AI
- Prompt Engineering & Response Optimisation
- Retrieval-Augmented Generation (RAG) & Vector Databases
- AI Assistants & Applications
- Azure OpenAI & OpenAI Capabilities
- Preparing & Evaluating RAG Solutions

Live Instructor Sessions: 2 Days



Tools and Technologies

Databases

- SQL Server
- NoSQL

Data Warehousing and Processing

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

Cloud Platforms

- Azure
- AWS

Data Engineering Services

- Databricks Data Intelligence Platform
- 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

- Databricks Data Intelligence Platform
- 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

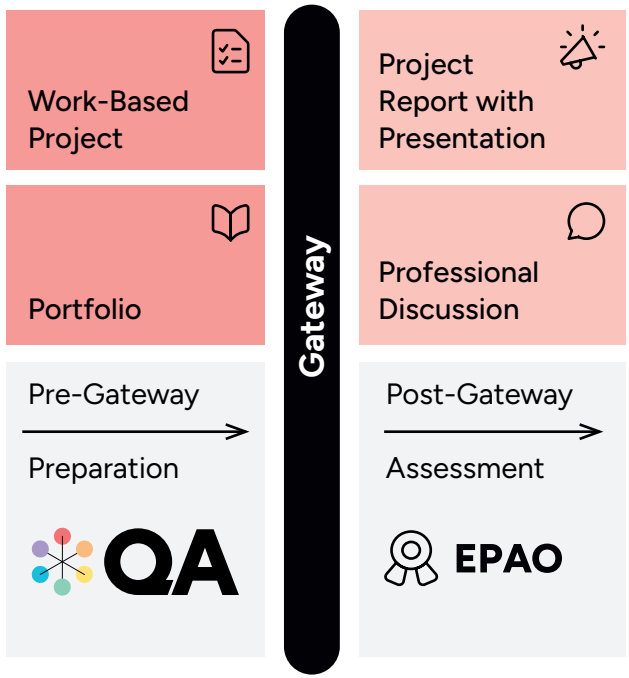
Skills
Demonstrate practical skills with confidence.

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.

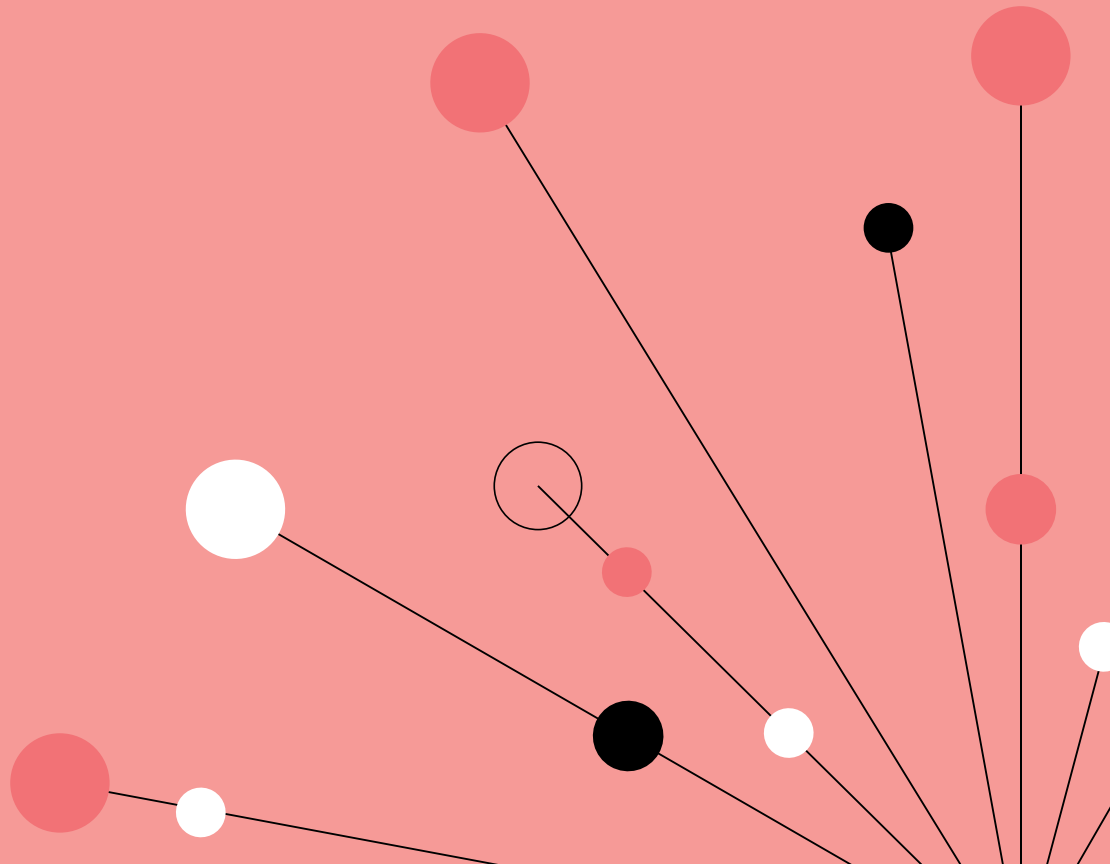
Professional Discussion: Engage in a formal two-way conversation to showcase knowledge, skills, and behaviours.

Ready to partner with us?

Let’s talk:

0113 220 7150

qa.com/contact



This information is correct as of publishing in March 2025

V1.0 2025

Funded by Department for Education