

# Level 5 Data Engineer.

The Level 5 Data Engineer apprenticeship equips your organisations with the ability to build and optimise automated data systems and pipelines, ensuring data quality, cataloguing, cleaning, validation, and thorough documentation. Learners will integrate, support, and manage data across stand-alone, distributed, and cloud-based platforms, providing efficient and sustainable data storage solutions. They will also gain the skills to identify and evaluate opportunities for data acquisition and enrichment, select and use appropriate tools to process data in various formats and modes, and adapt to legacy systems as needed.

Our apprenticeship programme covers how to analyse requirements and support the implementation of prototypes or proof-of-concept data products, and maintain data solutions as evolving products. Ultimately, this programme equips learners with the knowledge and expertise in data standards, quality, project management, and service reliability, enabling them to support business users in deriving actionable insights. Through practical application and a real-world portfolio, learners will develop the competencies required to ensure reliable, scalable, and future-proofed data services.

## Who's it for?

This programme is designed for organisations that are seeking to develop the technical expertise and professional skills required to build, optimise, and maintain data systems and pipelines. It is ideal for individuals looking to gain hands-on experience in data engineering, covering data integration, storage solutions, governance, and automation within distributed and cloud-based environments.

This programme is well-suited for those aiming to refine their problem-solving abilities, enhance their proficiency with data engineering tools, and stay current with industry advancements to drive efficient and scalable data solutions in real-world settings. Typical job roles for this apprenticeship include Data Engineer, Data Architect, Data/Cloud Migration and Integration Specialist.

### Business impacts

- ✓ Maximise the value of business data by creating efficient data pipelines and models
- ✓ Respond quicker than competitors and with more agility with in-house prototypes and data solutions
- ✓ Better decision-making and strategic planning
- ✓ Reduce operational costs with optimised data processes and infrastructures
- ✓ Build scalable data operations with cloud computing integration
- ✓ Confidently manage data securely and in compliance with regulations

### Built for performance

- ✓ Programmes designed and delivered by industry experts
- ✓ Dedicated Performance Coach, qualified in both coaching and their specialist subject area
- ✓ Backup from a multidisciplinary performance team
- ✓ Market-leading online live learning experience
- ✓ 24/7 access to programme materials, enrichment resources, study support and specialist insight via our virtual learning platform
- ✓ Mentoring, networking and peer support through BPP Community, including our Student Ambassador Network
- ✓ Learning pathways built using a 'stretch and challenge' model by design, meaning each learner is pushed to their maximum abilities
- ✓ Progression pathways that can take you from entry level to specialist expert
- ✓ Dedicated Functional Skills support if required



# Programme overview.

**Apprenticeship standard:** Data Engineer

**Cost:** £19,000

**Duration:** 19 months

## Entry requirements

As a minimum learners will need to have:

- A minimum of 720 hours of technology-related work experience (IT, Software or Engineering), or
- An A-Level in Computer Science or equivalent and have been in a technical role for three months or longer, or
- A Level 3 or 4 Data/Computing/Engineering apprenticeship

For learners that do not have GCSE English and/or maths at grades 9 to 4 (A\* to C):

- Learners aged 16-18 years must study and pass Functional Skills English and/or maths as part of the apprenticeship programme
- Learners aged 19 or above on the day they start the programme do not need to study or pass Functional Skills English and/or maths, unless required by their employer

## Prepare for the challenges of tomorrow

### Get ahead of evolving workforce skills needs

Every BPP apprentice has access to our exclusive Emerging Skills programme.

Comprised of three bespoke courses, the programme combines expertise from BPP, Microsoft and xUnlocked to give learners essential knowledge and skills in the rapidly emerging areas of AI and sustainability.

- ✓ Available to all learners at no extra cost
- ✓ Accessible anytime, anywhere via our virtual learning platform
- ✓ Self-paced learning to fit into any busy schedule



### Programme contents

#### Generative AI Fundamentals (Four modules)

Developed by BPP's expert data scientists, this course offers an introduction to working with Generative AI effectively, safely and ethically.

#### Introduction to Sustainability (Six modules)

Developed in partnership with sustainability experts, xUnlocked, this course builds fundamental knowledge on sustainability and sustainable working practices.

#### Microsoft AI and Security Essentials (Seven modules)

This official Microsoft learning pathway is comprised of a core pathway on AI Essentials, followed by either an AI Fundamentals pathway or Security pathway. Dedicated Microsoft experts guide learners to gain Microsoft digital badges as they progress, with the option to gain a recognised Microsoft Certification on completion.



# AI skills, *for all*.

Embedded AI training modules for every data and tech programme. Empowering every learner to drive AI-centric transformations within their business.

- ✓ Available to all learners at no extra cost
- ✓ Bespoke to BPP, developed by our expert data scientists
- ✓ Self-guided online learning to fit into any busy schedule
- ✓ Accessible anytime, anywhere via our virtual learning platform



## Knowledge and skills gained

Focusing on practical application of technical and non-technical AI skills, the modules explore AI's capacity to optimise structured interactions, align governance frameworks, and deploy scalable solutions, with a significant focus on ethical considerations and operational efficiency.

- ✓ Ability to design and implement complex prompts for diverse use cases
- ✓ Understanding and application of HITL techniques and fact-checking principles
- ✓ Ability to conduct prompt A/B testing
- ✓ Understanding of AI governance frameworks and compliance requirements
- ✓ Ability to adapt AI messaging for different stakeholders
- ✓ Ability to assess AI ecosystems
- ✓ Understanding of API-driven generative AI (GenAI) benefits



### Tools

Large language models (LLMs)  
(for example Copilot or Google Gemini)  
No-code AI automation platforms  
(no prior coding knowledge required)

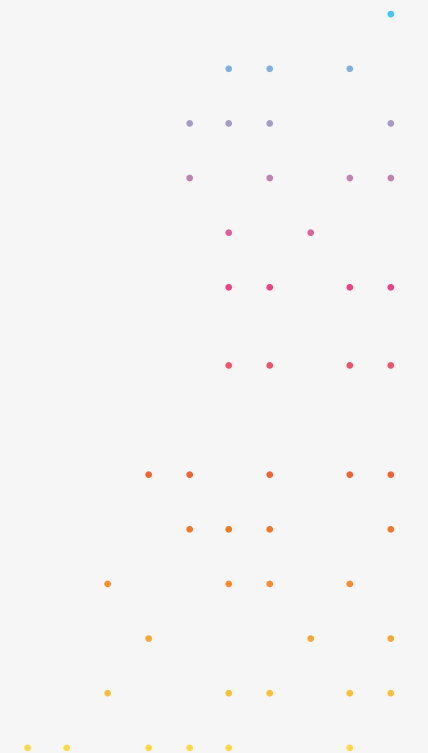


## Optional masterclasses

 Live online sessions available every month

Example topics include:

- Data Leadership
- Ethical Hacking and Cyber Security
- Discovering and Analysing Market Trends
- Sustainable Technology and Green Computing
- Responsible AI
- Setting AI Strategy
- Emerging Landscapes – AI
- Quantum Computing Fundamentals

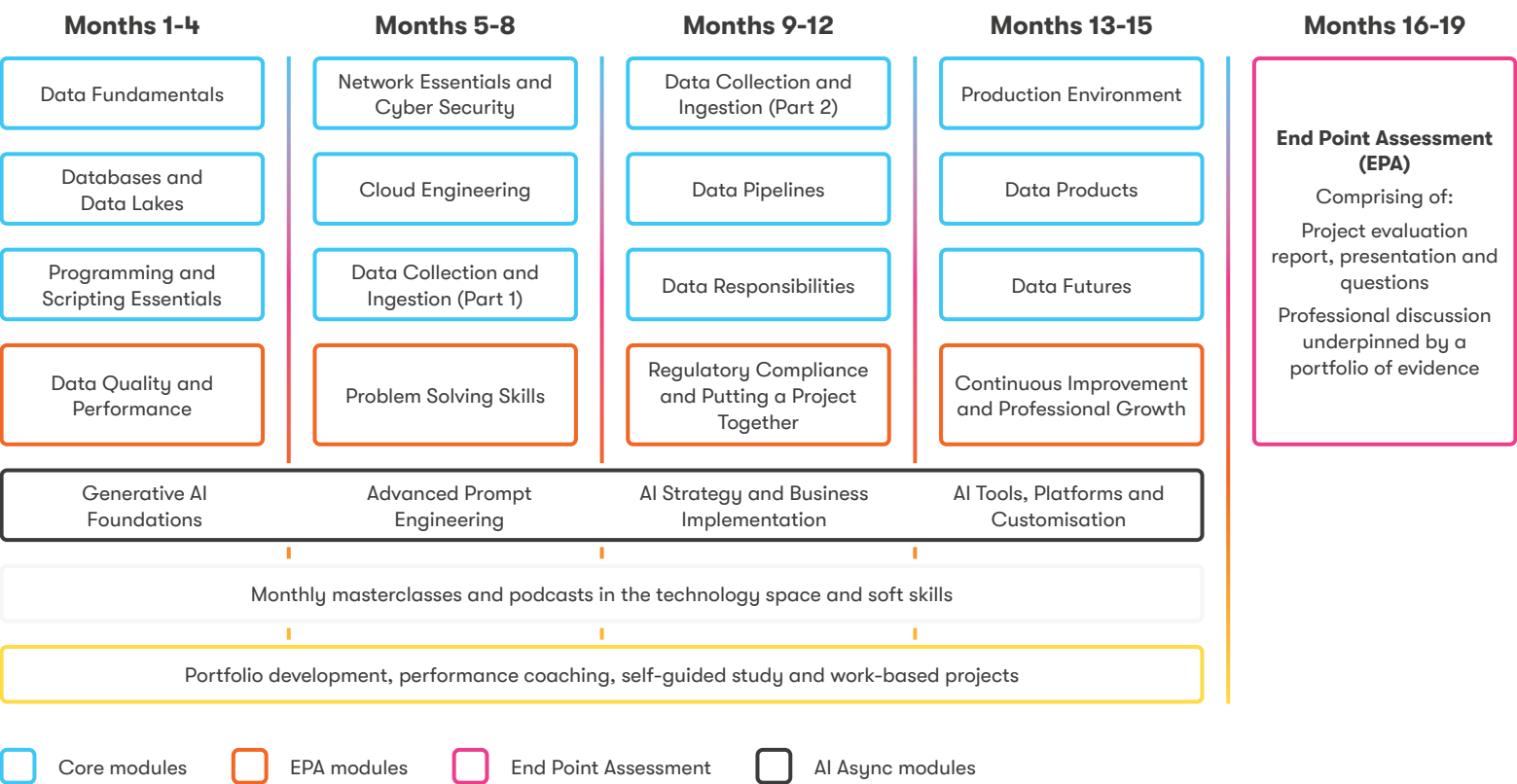


# Level 5 Data Engineer.

## Study mode

Online weekly: flexible learning that fits busy schedules, with two to three hours of online live lectures and seminars.

## 19-month programme (inc. EPA)



## Apprenticeship standard

 Data Engineer

## Delivered by

 BPP

## Time commitment\*\*

-  **15 months** on programme
-  **189 hours** in online live training sessions comprised of:
  - 57 x three-hour** webinars
  - 3 x six-hour** hackathons
-  **3-4 hours** guided self-study, per module, via our virtual learning platform
-  **1 hour** performance coaching session, every month
-  **1 hour** progress review, every eight weeks
-  **6 hours** per week in off-the-job learning, during working hours
-  **1 hour** EPA preparation session
-  **3 months** in End Point Assessment



# Programme modules.

## Data Fundamentals

This foundational module provides learners with a broad introduction to key data concepts, setting the stage for deeper exploration in later modules.

- How businesses generate value from data and the critical role of data engineering
- Data project management
- Stakeholder collaboration
- Modern data ecosystems, architectures, governance, and regulatory considerations

## Network Essentials and Cyber Security

Learners will examine network infrastructures and cyber security risks, essential for building secure, distributed data products.

- Core networking concepts
- Cyber threat migration
- Security best practices (including TCP/IP, the OSI model, encryption, and security frameworks)
- Designing secure networked architectures
- Managing security requirements and responding to incidents

## Data Collection and Ingestion (Part 2)

Building on foundational knowledge, this module explores event-driven architectures, focusing on Kafka, pub-sub models, and event streaming.

- Advanced Kafka scenarios (including Azure Event Hubs, SASL authentication, and troubleshooting techniques)
- Python libraries for data collection, automation, and anomaly detection
- Monitoring ingestion services

## Databases and Data Lakes

Learners will explore the structuring, storage, and retrieval of data to ensure reliability and integrity for downstream use. Learners develop expertise in SQL and NoSQL, data normalisation, query optimisation, and database performance monitoring.

- Managing databases and data lakes while understanding data consistency, availability and quality
- Open-source standards
- Regulatory compliance and security

## Cloud Engineering

Learners will develop an advanced understanding of cloud fundamentals, focusing on designing and managing cloud data products using microservices, containers, and orchestration with core Azure services.

- Cloud skills for AWS and Google Cloud
- Cloud deployments
- Serverless computing
- Integration using platforms like Snowflake

## Data Pipelines

Learners will develop advanced skills in data integration, workflow management, and resource scalability, ensuring efficient data flow in both cloud and non-cloud environments like Airflow.

- Data cleansing
- Data enrichment
- Data testing
- De-bugging
- Service-level agreements

## Programming and Scripting Essentials

Learners will develop essential programming and scripting skills for modern data-driven enterprises.

- Understanding programming tools including Python, Linux, Git, Parallel programming with Spark, and data manipulation using Pandas
- Data pipelines
- Error handling
- Unit testing and containerisation with Docker

## Data Collection and Ingestion (Part 1)

Learners will explore data acquisition, focusing on planning, evaluating data quality, and implementing validation checks.

- Key architectural practices, including data staging, APIs, and microservices
- Introduction to core architectural patterns
- Data pipelines
- ETL and ELT processes

## Data Responsibilities

Learners will examine the legal, ethical, and governance aspects of data engineering, focusing on privacy-by-design, compliance, and risk management.

- Roles and responsibilities within data-driven enterprises
- Data stewardship, governance, and security
- Stakeholder engagement and communication skills (including presentation techniques)



# Programme modules.

## Production Environment

Learners will develop the skills to work in production environments, covering Secure DevOps, MLOps, continuous integration, integration testing, logging, and deployment strategies.

- Service management
- Performance monitoring
- Stress testing and documentation

## Data Products

Learners will gain the skills to develop, maintain, and enhance data products, with a focus on AI-enabled solutions. Learners gain expertise in understanding business requirements, troubleshooting, data visualisation and dashboarding.

- Artificial intelligence (AI)
- Machine learning
- Data analytics
- Evaluating product success

## Data Futures

Learners will gain the skills to assess technological implementations, manage innovation, and embrace continuous professional development.

- Data ethics
- Sustainability in technology
- Evaluation of emerging technologies through horizon scanning and scorecards

## EPA: Project, Project Evaluation Report, and Presentation'

- The project will present a typical business task, showcasing real life examples of the learner's everyday work, that will solve a specific problem, recurring issue, or an idea, demonstrating skills and knowledge learnt on programme
- The project evaluation report is a 3,500 word report. It will combine the scope, research, methodology and findings of the project
- The presentation will comprise of an overview of the project, the project scope, a summary of actions taken, and the project outcomes. During the presentation, learners must have access to audio-visual presentation equipment

