

Data Engineering Pipelines for AI

COURSE OVERVIEW

Data Engineering Pipelines for AI provides a practical, hands-on introduction to designing, building, and managing the data pipelines that power modern AI systems. The course focuses on how to collect, clean, transform, and orchestrate large-scale data flows using contemporary tools and architectures, ensuring data is reliable, accessible, and optimized for machine learning. Participants will learn how to build end-to-end pipelines that integrate batch and real-time processing, support model training and deployment, and maintain high performance, governance, and scalability across enterprise AI environments.

WHO SHOULD ATTEND?

This course is ideal for data engineers, machine learning engineers, cloud engineers, ETL developers, BI engineers, software developers, and analytics professionals responsible for building or maintaining data pipelines. It also suits technical managers, AI product teams, and IT professionals seeking to strengthen their understanding of scalable data architectures and the data foundations required to support enterprise AI initiatives.

COURSE OUTCOMES

Delegates will gain the skills and knowledge to:

- Design scalable data ingestion pipelines that integrate databases, APIs, and streaming sources.
- Build reliable transformation workflows that clean, validate, and prepare data for AI models.
- Architect optimized data lake and warehouse solutions tailored for machine learning workloads.
- Automate end-to-end MLOps pipelines connecting data processing, model training, and deployment.
- Implement strong data quality and governance frameworks for reliability, traceability, and compliance.
- Optimize pipeline performance and cost efficiency across batch and real-time production environments.

KEY COURSE HIGHLIGHTS

At the end of the course, you will understand;

- How to design and build AI-optimized data pipelines for ML and deep learning.
- Techniques for data ingestion, preparation, and feature engineering for AI workloads.
- Orchestration and automation of data workflows using tools like Apache Airflow and Kubeflow.
- Deployment and monitoring of AI models within data pipelines.
- Management of real-time and batch data processing in cloud and hybrid environments.
- Governance, compliance, and ethical considerations in AI data engineering.
- Integration of AI pipelines with business intelligence and advanced analytics systems.

All our courses are dual-certificate courses. At the end of the training, the delegates will receive two certificates.

1. A GTC end-of-course certificate
2. Continuing Professional Development (CPD) Certificate of completion with earned credits awarded