AI for Data Engineering - Build the Data Pipelines That Power AI Systems

Course Overview

Artificial intelligence (AI) doesn't work without data. Before AI models can generate insights, make predictions, or automate decisions, data must be collected, cleaned, secured, and delivered reliably. The **AI in Data Engineering** course prepares learners to build the **data pipelines that make AI possible**.

This course focuses on **practical, job-ready data engineering skills** used in modern cloud environments. Learners work with real tools and realistic scenarios to move data from raw sources to analytics- and AI-ready formats. Rather than focusing on theory alone, the course emphasizes **hands-on labs, applied problem solving, and an end-to-end project**.

By the end of the course, learners will have designed and demonstrated a complete **data engineering pipeline with AI integration**, suitable for inclusion in a professional portfolio.

Who This Course Is For

This course is designed for learners who want practical experience building modern data systems, including:

- Aspiring or junior data engineers
- Data analysts seeking to move into engineering roles
- Software developers working with data-driven systems
- Technical professionals transitioning into AI-related roles

A basic familiarity with Python and data concepts is helpful, but the course is structured to guide learners step-by-step through increasingly complex workflows.

What You Will Learn

By completing this course, you will be able to:

- Explain the role of data engineering in AI-enabled systems
- Design reliable data pipelines from ingestion to analytics
- Work confidently with cloud-based data infrastructure
- Inspect, clean, and validate raw datasets
- Store and query data for analytics and AI use cases
- Build automated, event-driven data workflows
- Orchestrate multi-step pipelines with error handling
- Demonstrate an end-to-end pipeline that supports AI inference

How the Course Works

- **Duration:** 8 weeks (24 total instructional hours)

- **Format:** Instructor-led sessions with guided labs
- **Learning Style:** Applied, skills-based, and project-focused

Each week includes:

- A focused lecture introducing core concepts
- A hands-on lab using industry-standard tools
- A short knowledge check or take-home assignment

From **Week 4 through Week 8**, learners work on a cumulative **Capstone Project** that integrates skills from across the course.

Weekly Course Outline

Week	Topics	Overview	Focus Areas
1	Foundations & the Data Lifecycle	Learn what data engineering is, why it matters for AI, and how data moves from raw sources to usable formats. Set up a secure cloud-based development environment and run your first complete data pipeline.	• Data lifecycle • Data quality • Environment setup
2	Data Pipelines & Processing Patterns	Explore how data pipelines work in practice. Learn the difference between ETL and ELT, batch and streaming processing, and build an automated pipeline using Python.	• Pipeline design • Transformation logic • Reliability
3	Cloud Fundamentals	Work with cloud storage and identity systems. Learn how to securely store data, manage access, and control costs in a cloud environment.	• Cloud storage • Access control • Cost awareness
4	Security, Architecture & Capstone Proposal	Learn how to design secure data systems and translate requirements into architecture. Begin the Capstone Project by proposing an end-to-end data pipeline.	• Secure architecture • System design • Project planning
5	Serverless Processing	Build event-driven pipelines that respond automatically to new data using serverless technologies.	• Automation • Scalability • Event-driven design
6	Relational Databases	Work with structured data using relational databases and Structured Quey Language (SQL). Learn how databases support analytics and AI feature storage.	• SQL • Structured data • Analytics integration
7	Workflow Orchestration	Learn how to coordinate complex, multi-step data workflows with retries, logic, and error handling.	• Orchestration • Reliability • Observability

Week	Topics	Overview	Focus Areas
8	Data Lakes & Capstone Completion	Explore modern data lake architectures and complete the Capstone Project. Present a working, end-to-end data pipeline with analytics or AI output.	• Data lakes • Integration • Demonstration

Capstone Project: End-to-End Data Engineering with AI

The **Capstone Project** is the core applied component of the course. Learners design and build a complete data engineering pipeline that:

- Ingests data automatically
- Stores raw and processed data securely
- Cleans and prepares data for analysis
- Produces analytics or AI-driven output

Learners are **not required to build a complex custom AI model**. The emphasis is on demonstrating that the **data pipeline works reliably end-to-end** and enables AI or analytics use cases.

The capstone is designed to be:

- Practical and job-relevant
- Portfolio-ready
- A clear demonstration of applied data engineering skills

Assessment & Completion

This course uses a **Complete / Incomplete** model focused on demonstrated skills rather than exams.

To successfully complete the course, learners must:

- Participate in weekly labs and activities
- Complete required assignments and knowledge checks
- Contribute to and present the Capstone Project

Completion indicates readiness to work with real-world data engineering systems.

What You'll Leave With

By the end of the course, you will have:

- Hands-on experience with modern data engineering tools
- A clear understanding of how data supports AI systems
- An end-to-end data pipeline you can explain and demonstrate
- A practical project suitable for a professional portfolio