

COURSE OUTLINE

GenAI Coach Program for Data Engineering — Build, Operationalize, and Teach Practical GenAI Solutions

A hands-on program for data engineers, architects, and technical leaders

DURATION

FORMAT

5 days

Instructor-led, lab-driven, project-based

DELIVERY

TRAINER

Online Instructor Led Workshop · Onsite Program for your Teams

Muhammed Shakir

Why this course exists

This program turns generative AI from an emerging idea into a practical data-engineering capability: participants learn to design, implement, operationalize, and teach GenAI solutions across the data pipeline.

By the end, participants will understand how GenAI supports ingestion, transformation, data quality, governance, querying, and analysis—and will have applied it in a working business-focused project that they can explain and coach others to adopt.

What makes this different

Many GenAI programs focus on models or isolated demonstrations. **This is a data-engineering workshop: every important idea is connected to a pipeline, platform, governance decision, or team adoption challenge that participants can inspect, change, and test.** You will:

- **Learn GenAI in data-engineering language** — connect models, prompts, embeddings, vector data-bases, and agents to familiar concepts such as ETL jobs, schemas, data quality checks, repositories, and operating controls.
- **Work across the full data lifecycle** — apply GenAI to ingestion, transformation, schema design, anonymization, querying, anomaly detection, enrichment, and data-lakehouse workflows.
- **Build for responsible adoption** — combine cloud tools and coding assistants with validation, security, compliance, human review, and operational safeguards.

- **Become a coach, not only an implementer** — practice explaining value, identifying adoption opportunities, running workshops, and transferring GenAI knowledge to data-engineering teams.

Every concept is taught against a realistic data workflow, so participants leave with evidence of what works, what fails, and how to guide responsible adoption.

Skills You'll Walk Away With

- Explain the practical foundations of generative AI, LLMs, embeddings, vector databases, and AI assistants in data-engineering contexts.
- Identify high-value GenAI use cases across data ingestion, transformation, quality, governance, querying, and analysis.
- Use GenAI to accelerate data-engineering tasks such as schema design, code generation, ETL development, repository maintenance, and query creation.
- Design prompts and validation workflows for structured and unstructured data processing.
- Implement GenAI-assisted data quality, validation, anonymization, enrichment, anomaly detection, and ETL transformation workflows.
- Integrate GenAI capabilities with cloud and data platforms such as AWS, Google Cloud, Microsoft Azure, Snowflake, and Databricks.
- Design secure, compliant, and responsible GenAI workflows for data pipelines and data-lakehouse environments.
- Operationalize GenAI solutions with deployment, CI/CD, monitoring, governance, and maintenance practices.
- Coach data-engineering teams through GenAI adoption using workshops, demonstrations, communication, and feedback.

Who should attend

This program is for technical professionals who want to move from experimenting with GenAI to applying and coaching its use across data-engineering environments.

- **Data engineers** — applying GenAI to ingestion, transformation, quality, querying, and pipeline maintenance.
- **Data architects** — designing future-ready data architectures and selecting appropriate GenAI patterns and platforms.
- **Analytics and platform engineers** — integrating GenAI with warehouses, lakehouses, cloud services, and analytics workflows.
- **Technical leads and engineering managers** — identifying adoption opportunities and guiding responsible implementation.
- **GenAI champions and coaches** — helping teams understand, adopt, and govern GenAI in daily data work.

Prerequisites: comfort with data-engineering concepts, SQL, ETL or ELT workflows, cloud or data-platform fundamentals, and basic programming. Deep machine-learning expertise or advanced mathematics is not required; concepts are introduced in plain English before technical terminology.

Course Outline

Day 1 — Establish the GenAI Foundation and Discover Use Cases

Build the language and mental models needed to make sound GenAI choices across the data lifecycle.

- Generative AI, LLMs, embeddings, vector databases, and AI assistants
- How GenAI differs from traditional automation, machine learning, and rule-based data processing
- How data engineers leverage GenAI across ingestion, transformation, quality, governance, and analysis
- Common GenAI tools and libraries, including Hugging Face Transformers, TensorFlow, and PyTorch
- Future-ready data architectures and the role of GenAI within them
- GenAI for database and data-warehouse schema design
- Generative data creation, augmentation, and test-data preparation
- Infrastructure setup and configuration assistance
- Data anonymization and sensitive-data handling
- GenAI for database querying and data exploration
- Use-case prioritization by business value, data readiness, risk, and implementation effort
- *Hands-on:* Map a familiar data workflow to suitable GenAI capabilities, then prioritize opportunities using value, feasibility, and governance criteria.

Day 2 — Design Prompts and Accelerate Data Engineering

Turn natural-language requests into repeatable workflows while using AI coding tools responsibly.

- Prompt anatomy for data-engineering tasks: role, context, task, constraints, and output format
- Parsing messy data and flattening nested structures
- Prompting for schema inference, transformation logic, and query generation
- Structured outputs, validation rules, and retry strategies
- Using examples and representative data safely without exposing sensitive information
- Measuring quality, consistency, latency, and token or platform cost
- GitHub Copilot and AI-assisted development for data-engineering tasks
- Code generation, explanation, debugging, testing, refactoring, and documentation
- Repository maintenance and administration with GenAI assistance
- Generating SQL, ETL code, pipeline configuration, and infrastructure definitions
- Reviewing AI-generated code for correctness, maintainability, licensing, and security
- Human-in-the-loop development practices for data teams

- *Hands-on:* Build and evaluate prompt variants for schema mapping and transformation-rule generation, then use an AI coding assistant to extend and review a small ETL workflow.

Day 3 — Build GenAI-Assisted Data Quality and ETL Workflows

Combine conventional data-quality controls with GenAI-assisted validation, transformation, enrichment, and explanation.

- GenAI for data-quality rules and validation automation
- Detecting anomalies, inconsistencies, missing values, and schema drift
- Data cleansing, enrichment, classification, and normalization
- Natural-language explanations for data-quality findings
- Complementing, not replacing, deterministic quality checks
- Building feedback loops for continuous improvement
- GenAI for data transformation and ETL optimization
- Natural-language processing for structured and unstructured data
- Document ingestion, cleaning, chunking, metadata, and indexing
- Vector databases for scalable ingestion and lookup
- Retrieval and knowledge-base patterns for data professionals
- Data lineage, reproducibility, and approval boundaries in AI-assisted transformations
- *Hands-on:* Implement a GenAI-assisted data-quality and NLP transformation workflow that validates records, flags anomalies, and records its transformation steps.

Day 4 — Integrate and Operationalize GenAI on Data Platforms

Connect GenAI capabilities to the platforms where data-engineering work already happens.

- Google Cloud Vertex AI with BigQuery
- AWS Glue Studio with CodeWhisperer
- Snowflake with Snowpark and Streamlit
- Databricks AI Assist
- Microsoft Azure Synapse Analytics with Copilot
- API integration, model deployment, credentials, and environment configuration
- Choosing between managed services, libraries, and custom application components
- GenAI for data pipelines and ETL workflow orchestration
- GenAI for data repositories, platform maintenance, and administration
- Generative AI for data mining and analysis
- Data-lakehouse patterns and architecture considerations
- Deployment, scaling, caching, rate limits, and failure recovery
- CI/CD pipelines for GenAI models and data-engineering applications

- *Hands-on:* Implement a platform integration in a cloud sandbox, then package the GenAI-enabled pipeline with configuration, deployment checks, version control, and a CI/CD path.

Day 5 — Govern, Coach, and Demonstrate the Capstone

Protect data and users while establishing the controls that make GenAI adoption trustworthy.

- Sanitizing and validating GenAI inputs and outputs
- Compliance and security for AI-driven data processing
- Sensitive-data handling, anonymization, access control, and least privilege
- Prompt injection, data leakage, unsafe generated code, and dependency risks
- Responsible GenAI principles for data professionals
- Monitoring quality, cost, latency, drift, and operational failures
- Communicating GenAI benefits and limitations to data-engineering teams
- Identifying and fostering adoption opportunities in data processes
- Leading workshops, demonstrations, and training sessions
- Coaching patterns for experimentation, review, and responsible scaling
- Capstone integration review: architecture, data quality, security, operations, and adoption plan
- Presenting GenAI solutions and responding to stakeholder feedback
- *Hands-on:* Create a governance and security checklist, test the project against adversarial inputs, and present its evaluation results, safeguards, operating model, and coaching plan.

Capstone Project — Build a GenAI Solution for Data Engineering

Demonstrate practical skills by completing a GenAI project with a clear business application and a responsible operating plan.

- Develop a GenAI-powered data-quality monitoring tool
- Create an NLP-based ETL transformation pipeline
- Design an anomaly-detection system for large datasets using GenAI
- Document architecture, data flows, validation, security, deployment, monitoring, and team adoption
- *Deliverable:* A working prototype or production-simulated solution with a technical demonstration, evaluation evidence, governance checklist, and coaching presentation.

A note on delivery

This is a five-day build-and-learn program, not a slide-only survey. Each day combines instruction, guided demonstration, individual or pair lab work, review, and troubleshooting. The through-line is the responsible use of GenAI across the data-engineering lifecycle—from foundation and use-case discovery to ETL implementation, cloud integration, operationalization, governance, and team coaching. Participants leave

with working artifacts, implementation patterns, evaluation evidence, and the confidence to help their teams adopt GenAI thoughtfully.