



The banner features the Databricks logo at the top left. The main title 'DATABRICKS DATA ENGINEERING WITH AI & RAG' is prominently displayed in white and yellow text. Below the title, a horizontal bar lists technologies: SQL, Python, PySpark, Delta Lake, GenAI, RAG, and AI Agents. A central blue bar contains the text 'BUILD. ANALYZE. AUTOMATE. INNOVATE.' followed by 'End-to-End Data Engineering with Databricks, GenAI & RAG for Real-World Projects'. Below this, five icons represent 'DATA ENGINEERING', 'DATA ANALYTICS', 'AI & GENERATIVE AI', 'RAG SOLUTIONS', and 'AI AGENTS & AUTOMATION'. On the right, a portrait of Mr. Sai Phanindra Tholeti is shown, with a callout box stating 'LEARN BUILD DEPLOY BECOME JOB READY!'. At the bottom, a red bar includes the 'SQL SCHOOL' logo, '120+ MNC Clients', 'ISO 9001:2015 Certified', 'MSME Registered', and contact information: 'Call Today for FREE DEMO +91 99514 40801 www.sqlschool.com'.

databricks

DATABRICKS DATA ENGINEERING WITH AI & RAG

SQL | Python | PySpark | Delta Lake
GenAI | RAG | AI Agents

BUILD. ANALYZE. AUTOMATE. INNOVATE.

End-to-End Data Engineering with Databricks, GenAI & RAG for Real-World Projects

DATA ENGINEERING | DATA ANALYTICS | AI & GENERATIVE AI | RAG SOLUTIONS | AI AGENTS & AUTOMATION

SQL SCHOOL
Premium Quality Training

120+ MNC Clients

ISO 9001:2015 Certified

MSME Registered

Mr. Sai Phanindra Tholeti
Chief Mentor & DataStack Expert
20+ Years of Experience

LEARN BUILD DEPLOY BECOME JOB READY!

Call Today for FREE DEMO
+91 99514 40801
www.sqlschool.com

Databricks Data Engineer with GenAI, Agents & RAG

Master Databricks with AI-powered data pipelines, enterprise RAG applications and AI agents.

At SQL School, our focus goes beyond completing a syllabus. You will learn through hands-on labs, production-style scenarios, end-to-end projects, assignments, interview preparation, and certification guidance designed to build practical **AI Enabled Databricks Data Engineering skills**.

- ✓ Practical Skill Development
- ✓ Real-Time Project Guidance
- ✓ Resume & Interview Preparation
- ✓ Certification Guidance

Pre-Requisites: Participant should have knowledge on Databricks Data Engineering Concepts

Duration: 4 Weeks

Trainer: Mr. Sai Phanindra Tholeti

Trainer Profile: [linkedin.com/in/saiphanindra/](https://www.linkedin.com/in/saiphanindra/)

Detailed Course Content

Ch 1: AI Concepts

- ✓ Artificial Intelligence, Machine Learning and Deep Learning
- ✓ Generative AI concepts and applications
- ✓ Traditional AI vs Generative AI
- ✓ Generative AI use cases in Data Engineering

Ch 2: Large Language Models & Prompt Engineering

- ✓ Introduction to LLMs and foundation models
- ✓ Tokens, context windows and inference
- ✓ Prompt engineering techniques
- ✓ Structured outputs and JSON responses
- ✓ Introduction to embeddings and similarity search

Ch 3: Generative AI with Databricks

- ✓ Databricks AI ecosystem overview
- ✓ Databricks AI Playground
- ✓ Foundation Model APIs
- ✓ Model serving endpoints
- ✓ Accessing LLMs using Python

Ch 4: Databricks AI Functions

- ✓ Introduction to Databricks AI functions
- ✓ Using `ai_query()` in SQL
- ✓ Calling foundation models from SQL
- ✓ AI-powered text processing
- ✓ Structured response generation

Ch 5: Intelligent Data Processing

- ✓ Text classification using AI
- ✓ Sentiment analysis
- ✓ AI-based information extraction
- ✓ Data enrichment using LLMs
- ✓ Working with customer reviews and feedback

Ch 6: Unstructured Data Engineering

- ✓ Working with PDF, JSON and text files
- ✓ Ingesting unstructured data
- ✓ Document parsing and preprocessing
- ✓ PySpark pipelines for AI-ready datasets
- ✓ Storing processed data in Delta Lake

Ch 7: AI Data Pipelines & Governance

- ✓ Delta Lake for AI datasets
- ✓ Unity Catalog for data governance

- ✓ Incremental data processing
- ✓ Embedding generation and refresh pipelines
- ✓ Automating AI-based ETL workflows

Ch 8: Introduction to RAG

- ✓ What is Retrieval-Augmented Generation?
- ✓ Limitations of traditional LLM applications
- ✓ RAG vs fine-tuning
- ✓ RAG architecture and components
- ✓ Enterprise RAG use cases

Ch 9: Document Processing for RAG

- ✓ PDF and document ingestion
- ✓ Extracting text from documents
- ✓ Document parsing and cleaning
- ✓ Text chunking strategies
- ✓ Preparing documents for vector search

Ch 10: Vector Embeddings

- ✓ Understanding text embeddings
- ✓ Embedding models
- ✓ Generating vector embeddings
- ✓ Vector similarity concepts
- ✓ Storing document embeddings

Ch 11: Databricks AI Search

- ✓ Introduction to Databricks AI Search (Vector Search)
- ✓ Vector indexes and index creation
- ✓ Delta Sync indexes
- ✓ Similarity search queries
- ✓ Retrieving relevant document chunks

Ch 12: Building RAG Applications

- ✓ Integrating vector retrieval with LLMs
- ✓ Context retrieval and prompt augmentation
- ✓ Generating context-aware responses
- ✓ Building a PDF-based AI chatbot
- ✓ Source citations and hallucination reduction

Ch 13: Advanced Retrieval Techniques

- ✓ Semantic search vs keyword search
- ✓ Hybrid search techniques
- ✓ Reranking retrieved documents
- ✓ Metadata-based filtering
- ✓ Chunk-size optimization

Ch 14: Intelligent RAG Pipelines

- ✓ Query rewriting
- ✓ Multi-query retrieval
- ✓ Structured RAG with SQL data
- ✓ Combining Delta tables and document retrieval
- ✓ Building hybrid structured/unstructured RAG applications

Ch 15: RAG Performance Optimization

- ✓ Evaluating retrieval relevance
- ✓ Answer groundedness
- ✓ Improving retrieval accuracy
- ✓ Reducing hallucinations
- ✓ Latency and token

Real-Time Project (E-Commerce Domain)

Project Title

Retail Sales Analytics Platform using Databricks (Medallion Architecture)

Project Overview

Students will build a production-style data platform using Lakehouse Architecture (Bronze, Silver, and Gold), implementing modern ETL practices, data quality validation, incremental processing, orchestration, optimization, and reporting.

Business Scenario

A multinational retail company receives daily data from multiple operational systems. The company wants to build a centralized analytics platform to:

- Consolidate data from different business domains
- Improve data quality and consistency with AI
- Track sales performance across regions
- Analyse customer purchasing behaviour
- Monitor inventory levels with AI
- Build AI Agents, RAG Artifacts
- Generate business KPIs for decision-makers
- Deliver Power BI dashboards with near real-time insights

Source Systems

Students will work with multiple datasets including:

- Customer Master
- Product Master
- Orders
- Inventory
- Region Master

Resume Highlights

After completing this project, you can confidently showcase experience in:

- PySpark & Spark SQL with AI

- Delta Lake with AI
- Medallion Architecture with LLMs
- ETL Pipeline Development with RAG
- Incremental Data Processing with Genie AI

Why Choose This Training?

- 📁 End-to-End Real-Time Projects & Solutions
- 📊 100% Hands-on Practical Sessions
- ⚡ Production-Ready ETL Pipelines
- 🎯 Interview Preparation
- 📅 Assignments & Practice Labs
- ★ 100% Practical • No Unnecessary Theory
- 🔗 Learn Like You're Working in a Real Company

BUILD REAL SKILLS. WORK ON REAL PROJECTS. BECOME JOB READY.



- 🚀 Ready to Become an AI based Databricks Expert?
- ✅ Join Our Next Batch. Reach us today!
- 📞 Call/WhatsApp: +91 9951440801
- 🌐 www.sqlschool.com

Wish you all the best!