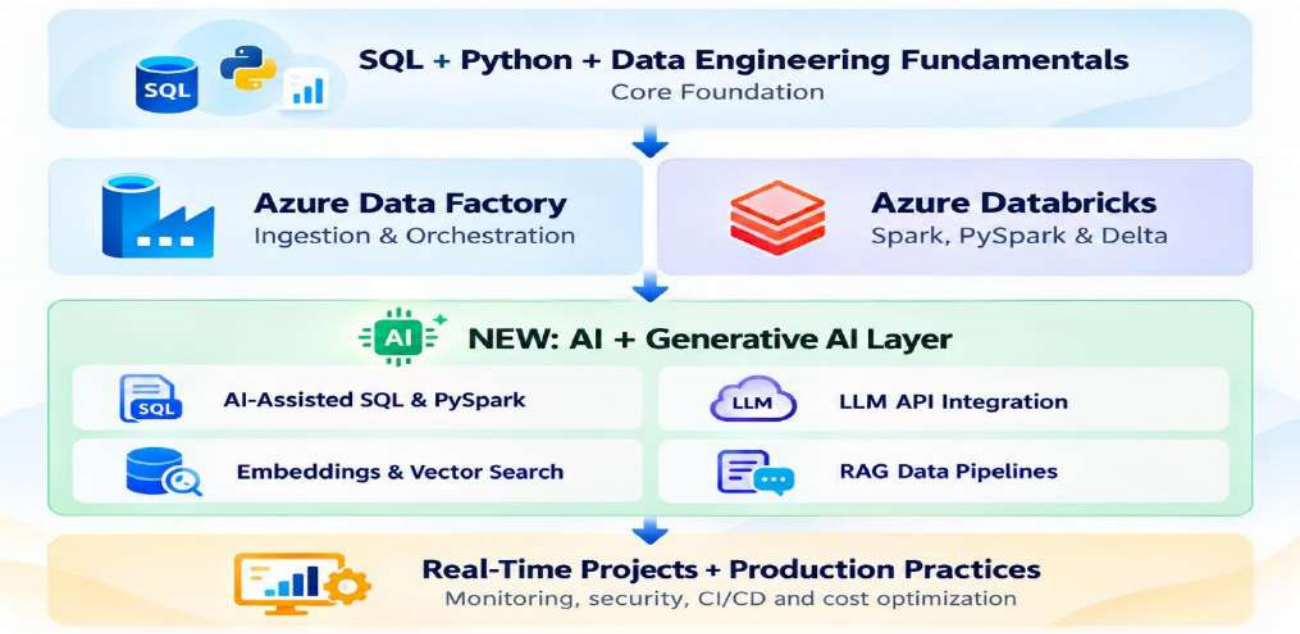
Azure Data Engineering with AI, Generative AI

SQL • Python • Azure Data Factory • Azure Databricks • GenAI • RAG • LLMs



What You Will Learn to Build

This Azure Data Engineer Training with AI is designed to help you understand and implement a **complete end-to-end data engineering solution** using Microsoft Azure & AI, Gen AI Technologies.

Through hands-on training and projects, you will learn how to:

- Ingest data from databases, files, APIs and streaming sources
- Orchestrate data pipelines using **Azure Data Factory (ADF)**
- Store and manage large-scale data using **Azure Data Lake Storage (ADLS)**
- Transform and process data using **Azure Databricks, Spark & PySpark**
- Build **Bronze → Silver → Gold** data architectures
- Implement AI, Gen AI and LLMs with Agentic AI
- Implement RAG Artifacts and AI Agents
- Implement security, monitoring, optimization and CI/CD practices

What are the target Job Roles from this course?

From this 100% Practical, step by step course – you can opt for below job roles:

1. Azure Data Engineer
2. Azure Databricks Data Engineer
3. AI Enabled Data Engineer
4. ETL Developer with Gen AI
5. Azure ETL & DWH Developer
6. Generative AI Engineer
7. Databricks Data Engineer
8. Big Data Engineer

Azure Data Engineering with AI & Generative AI Course Content

This hands-on training program is divided into **FIVE comprehensive modules**:

Module 1: SQL Server (MSSQL) & TSQL Queries, Tuning
Module 2: Azure Funda, Data Factory & Synapse with Fabric
Module 3: Azure Databricks (Python, Spark, PySpark, Unity Catalog)
Module 4: Gen AI, Agentic AI & LLMs
Module 5: Realtime Projects, Certifications & Resume

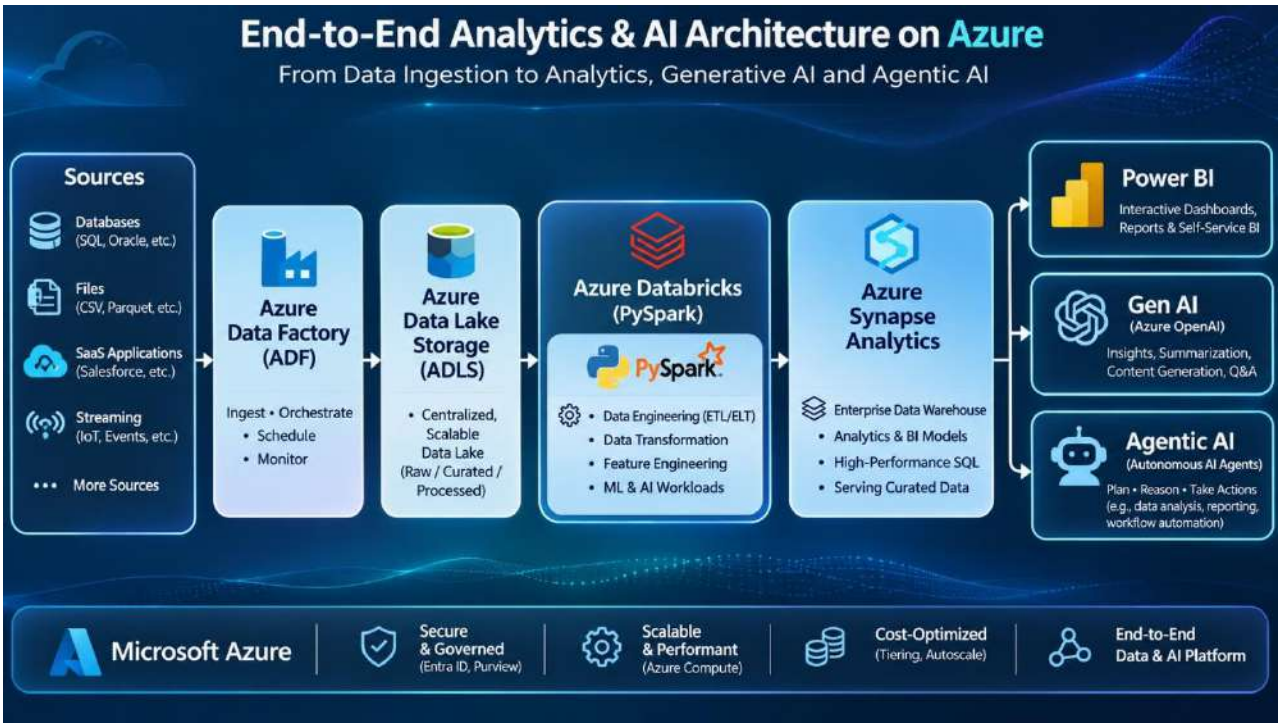
Training Modes: LIVE Online Instructor-Led Training / Self Paced Videos
Duration: 3.5 Months

Trainer: Mr. Sai Phanindra Tholeti (20+ Yrs of expertise)
Trainer Profile: https://www.linkedin.com/in/saiphanindra/

Who can join this course?

Any one. Course starts with fundamentals. Progresses to advanced Data Engineer concepts!

Book Your FREE Demo Today. Call us on +91 9951 440801, +91 9666 640801



DETAILED COURSE CONTENT

Module 1: MSSQL & TSQL Queries, Tuning

Ch 1: SQL Database Job Roles

- ✓ Database Intro
- ✓ OLTP, DWH, OLAP
- ✓ DBMS Basics
- ✓ Data Engineer Job Roles

Ch 2: Database Intro & Installations

- ✓ SQL Server Installations
- ✓ Instance & Collations
- ✓ SSMS Tool Installation
- ✓ Connections, Authentications

Ch 3: SQL Basics V1 (Commands)

- ✓ SQL Basics (DDL, DML, etc..)
- ✓ Creating Databases, Tables
- ✓ Data Inserts (GUI, SQL)
- ✓ Basic SELECT Queries

Ch 4: SQL Basics V2 (Commands, Operators)

- ✓ DDL: Create, Alter, Drop
- ✓ DML: Insert, Update, Delete
- ✓ DQL: Select, Fetch
- ✓ SQL Operators

Ch 5: Data Types & Variables

- ✓ Integer, Character Data Types
- ✓ Decimal & Boolean Data Types
- ✓ Date and Time Data Types
- ✓ SQL_Variant Type

Ch 6: Data Imports

- ✓ Data Imports with Excel
- ✓ Data Types Detection
- ✓ OLE-DB Connections
- ✓ ORDER BY, TOP, OFFSET

Ch 7: Schemas & Batches

- ✓ Schemas, Table Grouping
- ✓ Real-world Banking Database
- ✓ 2 Part, 3 Part, 4 Part Naming
- ✓ Batch & “Go” Command

Ch 8: Constraints, Keys & RDBMS

- ✓ Null, Not Null Constraints

- ✓ Unique Key & Check
- ✓ Primary Key & Default
- ✓ Foreign Keys, DB Diagrams

Ch 9: Normal Forms & RDBMS

- ✓ Normal Forms: 1 NF, 2 NF
- ✓ 3 NF, BCNF and 4 NF
- ✓ 1:1, 1:M, M:1 Cardinality
- ✓ Self Referencing Keys, ER Models

Ch 10: Joins & Queries

- ✓ Joins & TSQL Queries
- ✓ Inner Joins, Matching Data
- ✓ Outer Joins: Left, Right, Full
- ✓ Self Joins, Conditions

Ch 11: Sub Queries

- ✓ Sub Queries
- ✓ Nested Queries
- ✓ Correlated Sub Queries
- ✓ Union, Union All

Ch 12: Group By Queries

- ✓ Group By, Having
- ✓ Cube( ) and Rollup( )
- ✓ Grouping( )
- ✓ IsNull(), Coalesce()

Ch 13: Joins with Group By, Sub Queries

- ✓ Join, Group By, Sub Queries
- ✓ IIF(), CASE Statement
- ✓ EXISTS, NOT EXISTS
- ✓ Query Execution Order

Ch 14: Views & RLS

- ✓ Views: Realtime Usage
- ✓ DML, SELECT with Views
- ✓ WITH CHECK OPTION
- ✓ RLS, Important System Views

Ch 15: Functions & Queries

- ✓ Scalar, Table Value Functions
- ✓ Variables & Parameters
- ✓ Date Time, Conversion Functions
- ✓ String, Aggregated Functions

Ch 16: Advanced SQL, Window Functions

- ✓ Window Functions (Rank)

- ✓ Row_Number, DenseRank
- ✓ Partition By, Lead & Lag
- ✓ Pivot, UnPivot, Running Totals

Ch 17: Stored Procedures - 1

- ✓ Stored Procedures & Parameters
- ✓ SELECT & Conditional DML
- ✓ IF Statement, SP Recompilation
- ✓ System Stored Procedures

Ch 18: Stored Procedures - 2

- ✓ Merge Statement (Upsert)
- ✓ Merge with OLTP & DWH
- ✓ Matched and Not Matched
- ✓ Merge & SP Recompilations

Ch 19: Triggers & Automations

- ✓ DDL & DML Triggers
- ✓ For / After Triggers
- ✓ Instead Of Triggers
- ✓ Disabling DMLs Triggers

Ch 20: Transactions & ACID

- ✓ Auto Commit & Explicit
- ✓ Commit, Rollback, ReadPast
- ✓ Checkpoint, Query Blocking
- ✓ TRY...CATCH, Error Handling

Ch 21: SQL Query Performance & Indexing

- ✓ Clustered, NonClustered Indexes
- ✓ Execution Plans, Tuning Queries
- ✓ SARGable Queries
- ✓ Query Optimization Examples

Ch 22: CTEs & Tuning

- ✓ Common Table Expression
- ✓ Data Cleansing Techniques
- ✓ Duplicate Detection/Removal
- ✓ Cursor Basics – When to Use & Avoid

Module 2: Azure Data Factory, Synapse [ADLS, ARM, GIT, Fabric]

Ch 1: Azure Fundamentals

- ✓ Cloud Intro, Implementations
- ✓ Azure Account, Subscription
- ✓ Azure ETL & DWH Resources
- ✓ Azure Storage, IoT Resources

Ch 2: Azure Storage & ADLS

- ✓ Azure Storage Account
- ✓ Azure Data Lake Storage
- ✓ BLOB Containers
- ✓ Folders & Files

Ch 3: Azure Security Concepts

- ✓ Microsoft Entra, RBAC
- ✓ Users & Groups. Storage Security, IAM, ACL
- ✓ Access Keys, SAS Keys

Ch 4: Azure Deployments, Azure SQL

- ✓ Azure SQL Server Deployment
- ✓ Azure SQL DB Deployment
- ✓ Azure SQL Database (OLTP)
- ✓ Firewall Rules, SSMS Connections

Ch 5: Azure SQL DB Migrations

- ✓ On-Premise SQL DB, bacpac
- ✓ Azure SQL DB Migration
- ✓ SQL Server Imports from BLOB
- ✓ Migration Verifications

Ch 6: Azure Synapse Analytics (DWH)

- ✓ Synapse Pool Architecture
- ✓ Control Node, Compute Node
- ✓ DMS (Data Movement Service)
- ✓ Partitions, Distributions & MAXDOP

Ch 7: Azure Data Factory – Pipelines

- ✓ ADF Intro & Pipeline Artifacts
- ✓ IR, Linked Services, Datasets
- ✓ Copy Data Activity
- ✓ Azure SQL DB → Synapse Loads

Ch 8: ADF: Azure Blob Data Loads

- ✓ ADF Author & Pipeline Runs
- ✓ Azure BLOB Storage Access
- ✓ Pipeline Staging & Logging
- ✓ Validations & Schedules

Ch 9: Table Incremental Loads

- ✓ SCD & Incremental Loads
- ✓ Business Keys & Data Delta
- ✓ Copy Method: Upsert
- ✓ Initial & Incremental Loads

Ch 10: On-Premises Loads, Optimizations

- ✓ Self Hosted IR Setup
- ✓ IR Keys, Security
- ✓ On-Premise Data Loads
- ✓ ADF Optimizations

Ch 11: ADF Data Flow - 1

- ✓ Data Flow Activity
- ✓ Debug Cluster, Data Preview
- ✓ Data Cleansing, Conditional Split
- ✓ Select, Sort, Union, Loops

Ch 12: ADF Data Flow - 2

- ✓ Join Transform, Broadcast
- ✓ Surrogate Keys, Derived Cols
- ✓ ETL Loads Dates, Sink Options
- ✓ Aggregated Data Loads

Ch 13: ADF Data Flow - 3

- ✓ Pivot, Data Duplication
- ✓ Group By & Pivot Keys
- ✓ Column Name Pattern
- ✓ Tuning Data Flow, Spark

Ch 14: ADF Data Flow - 4

- ✓ Get Metadata Activity
- ✓ IF Activity Expressions
- ✓ Lookup & Loops in ADF
- ✓ Data Flow Cluster

Ch 15: Azure AD Security (Entra), Parameters

- ✓ Creating AD Users, Groups
- ✓ RBAC & IAM with ADF
- ✓ Azure SQL Pool Logins, Users
- ✓ ADF Connection Parameters

Ch 16: SCD & ETL with Control Tables

- ✓ ADF Templates, Data Delta
- ✓ Control Tables, Watermarks
- ✓ Lookup, SP Activity For SCD
- ✓ Pipeline Parameters, SPs

Ch 17: ADF Metrics, Alerts

- ✓ ADF Insights & Metrics
- ✓ DTU, CPU & Memory Metrics
- ✓ Alerts and Notifications
- ✓ Action Groups, Tuning Options

Ch 18: ADF with GIT, ARM Templates

- ✓ ADF Git Integration For CI-CD
- ✓ ADF ARM Templates Export & Import
- ✓ Dev / Test / Prod Environments
- ✓ Parameterized Deployments
- ✓ Production Release Workflow

Ch 19: Azure Stream Analytics

- ✓ Azure IoT Hubs
- ✓ Azure Stream Analytics
- ✓ LIVE Data Ingestions
- ✓ SAQL Queries, JSON

Ch 20: Azure Key Vaults

- ✓ Azure Key Vaults
- ✓ Azure Keys & Secrets
- ✓ Access Policies
- ✓ Secret Management

Module 3: Azure Databricks (Spark, PySpark, Unity Catalog)

Ch 1: Databricks Introduction

- ✓ Cloud ETL, DWH
- ✓ Databricks Concepts
- ✓ ETL, Lakehouse & Big Data
- ✓ Databricks Workspace UI

Ch 2: Databricks Architecture

- ✓ Two Plane Architecture
- ✓ Databricks Runtime (DBR)
- ✓ Spark Compute Concepts
- ✓ Databricks APIs

Ch 3: Unity Catalog

- ✓ Unity Catalog Concepts
- ✓ Managed & External
- ✓ Lineage, Data Governance
- ✓ Genie AI with Unity Catalog

Ch 4: Databricks Operations (GUI)

- ✓ Catalog & Schema Creations
- ✓ Volumes, Files & Tables
- ✓ Catalog Explorer. Need for ETL

Ch 5: Spark SQL

- ✓ Spark SQL Notebooks
- ✓ Catalog, Schemas
- ✓ Tables, DDL Operations
- ✓ Spark Data Types & Clones

Ch 6: Spark Queries

- ✓ Data Inserts, Sequences
- ✓ Spark Joins
- ✓ Aggregations
- ✓ Data Analytics

Ch 7: Spark SQL: Functions

- ✓ Aggregate Functions
- ✓ Math, Sort Functions
- ✓ String, Date Time Functions
- ✓ Window & Null Functions

Ch 8: Spark SQL: Time Travel

- ✓ Time Travel Concepts
- ✓ HISTORY, EXTENDED
- ✓ Versions & Time Stamp
- ✓ Time Travel Settings

Ch 9: Python Concepts in Databricks

- ✓ Python Implementations
- ✓ Print(), display()
- ✓ Variables, Operators
- ✓ Dataframes

Ch 10: PySpark Concepts

- ✓ Python Vs PySpark
- ✓ PySpark DataFrames
- ✓ select(), filter()
- ✓ withColumn()

Ch 11: Medallion Architecture - 1

- ✓ File Data Sources
- ✓ Bronze, Silver and Gold
- ✓ Temp Views, Aggregations
- ✓ Spark Tables (Parquet)

Ch 12: Medallion Architecture - 2

- ✓ Azure SQL DB Connections
- ✓ Joins & Aggregations
- ✓ Gold Data Consumption
- ✓ Bronze → Silver → Gold → BI/Analytics

Ch 13: Databricks Delta Lake

- ✓ Schema Evolution
- ✓ Delta Table API
- ✓ MERGE / Upsert Operations
- ✓ Delta Transaction Log

Ch 14: PySpark: Widgets

- ✓ Parameters, Text Widgets
- ✓ dbutils() & get() Functions
- ✓ Widget Prompts, Edits
- ✓ UI & JSON for Widgets

Ch 15: Lake Flow Jobs

- ✓ Workflows & CRON
- ✓ Job Compute, Running Tasks
- ✓ Python Script Tasks, Dependencies
- ✓ Concurrency, Branch Control

Ch 16: PySpark: Auto Loader

- ✓ Cloud Files & Checkpoints
- ✓ Data Stream, Schemas
- ✓ Schema Evaluation Modes
- ✓ File Ingestions, Rescued Column

Ch 17: LakeFlow Declarative Pipelines

- ✓ SDP: Spark Declarative Pipelines
- ✓ Delta LIVE Tables, Data Streams
- ✓ Bronze, Silver, Gold Data
- ✓ Data Quality, Materialized Views

Ch 18: Databricks Optimizations

- ✓ Lazy Evaluation, Data Skew
- ✓ Data Shuffling, Partitions
- ✓ Broadcast Joins, Vacuum
- ✓ Liquid Clusters, Optimize, Z-Order

Ch 19: Databricks Security, AI

- ✓ ACL: Access Control List
- ✓ Workspace & Catalog Security
- ✓ Schema, Volume Security
- ✓ Table & Column Security

Ch 20: Databricks with Power BI

- ✓ Access Keys (Tokens)
- ✓ Server Host, HTTP Path
- ✓ Spark Connectors
- ✓ Databricks Connectors

Ch 21: Genie AI

- ✓ Unity Catalog & Genie AI
- ✓ AI Components in Databricks
- ✓ Debugging Controls
- ✓ Using AI for Notebook Design

Ch 22: GitHub & Asset Bundle

- ✓ GitHub & GIT
- ✓ Branches, Asset Bundle
- ✓ Integrating GIT
- ✓ Process Integration

Ch 23: Azure Databricks

- ✓ Azure Databricks Deployments
- ✓ Azure Databricks Regions
- ✓ Classic Deployments, Connectors
- ✓ ADF Versus ADB

Ch 24: Azure Databricks with ADLS

- ✓ ADLS Security for Databricks
- ✓ Creating Service Principals
- ✓ Enterprise Apps, Client IDs
- ✓ External Tables (Delta)

Module 4: Databricks AI (GenAI, LLM, Agents, RAG)

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, Foundation Models
- ✓ Tokens, context windows & 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. 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, 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, 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**

Module 5: Realtime Projects for Data Engineering

Real-Time Project 1: HealthCare Domain

Project Requirement:

Solve 20+ Real-World Healthcare Business Requirements.

Project Workflow:

Database Design → SQL → Data Analysis → Performance Optimization

Project Operational Flow:

Healthcare DB → Patients → Doctors → Appointments → Treatments → Billing → Insurance → SQL Analysis → Stored Procedures → Performance Tuning → Business Reports

Business Requirements:

Monthly Hospital Revenue, Top Doctors By Patients Treated, Repeat Patients, Department Performance, Average Treatment Cost, Unpaid Bills, Insurance Claim Analysis, Patient Trends, Month-Over-Month Revenue And Top Procedures…

Real-Time Project 2: Online Travel Data Engineering Platform

Project Objective:

Build an end-to-end Azure Data Engineering solution to process, transform, and analyze Online Travel business data from multiple sources.

Skills Gained:

- ✓ Data Ingestion & ETL Development
- ✓ Azure Data Factory Pipelines
- ✓ Data Orchestration (End to End)
- ✓ Data Lake Architecture, RBAC
- ✓ Real-Time Industry Experience
- ✓ Azure IoT, Stream Analytics
- ✓ Azure Databricks with Data Factory
- ✓ Azure Data Factory with Fabric Migrations

Components For Project

- ✓ ADF Pipelines
- ✓ Databricks Notebooks
- ✓ Monitoring & Alerts
- ✓ CI/CD Deployment
- ✓ End to End Integrations
- ✓ IAM & Managed Identity

Real-Time Project 3: 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 ETL practices, data quality validation, incremental processing and orchestration.

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
- ✓ Track sales performance across regions
- ✓ Analyse customer purchasing behaviour
- ✓ Generate business KPIs for decision-makers

Technologies Covered

- Apache Spark
- PySpark, Spark SQL
- Delta Lake, Unity Catalog
- Databricks Workflows
- Git Integration

Project Workflow

Sources → ADLS → Bronze → Silver → Gold → Databricks Workflows → Fabric, Power BI

Module 6: Certifications & Career

Exam 1: Databricks Data Engineer Associate

- ✓ Exam Details
- ✓ Exam Dumps
- ✓ Scenario-Based Questions
- ✓ Practice Assessments

Exam 2: Microsoft Databricks (DP-750)

- ✓ Exam Details
- ✓ Exam Dumps
- ✓ Scenario-Based Questions
- ✓ Practice Assessments

Resume Highlights

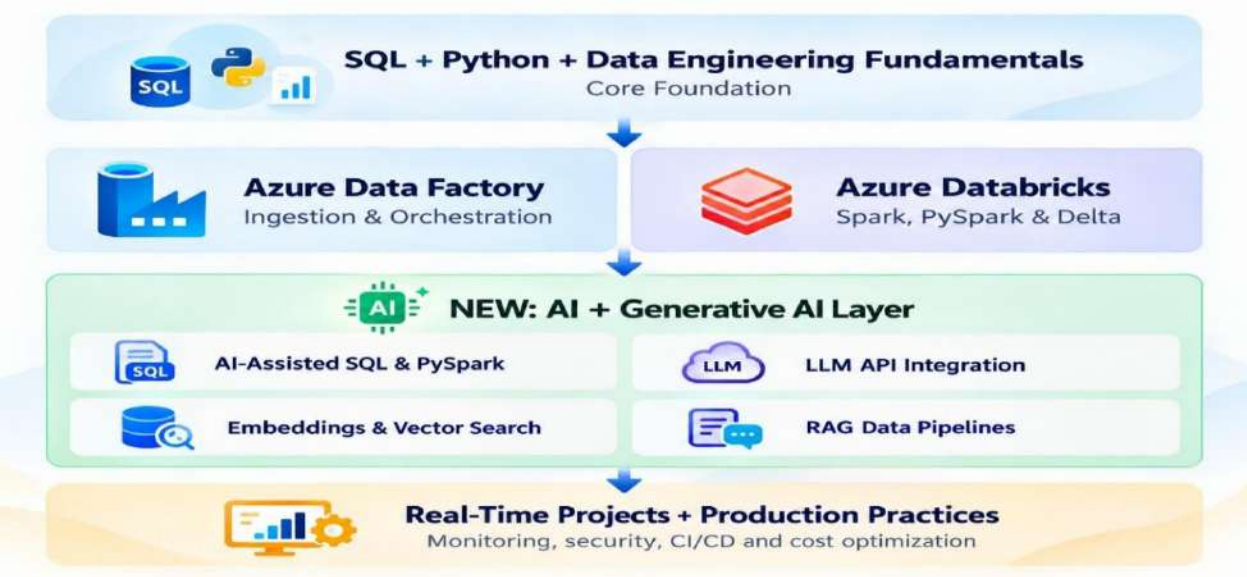
After completing this module, you can confidently showcase experience in:

- ✓ PySpark & Spark SQL
- ✓ ETL Pipeline Development
- ✓ Data Quality & Validation
- ✓ Automation, Optimizations

Interview & Career Preparation

- End-to-End Real-Time Projects & Solutions
- 100% Hands-on Practical Sessions
- 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 a Data Engineer with AI Expert?
Join Our Next Batch. Reach us today!
Call/WhatsApp: +91 9951440801
www.sqlschool.com
Wish you all the best!