

I'm **Mr. Sai Phanindra**, your trainer and mentor for this course. With 21 years of experience in Database Technologies, Data Engineering, Business Intelligence, and Data Warehousing, I bring extensive real-world expertise into every training session. **Let's connect:** linkedin.com/in/saiphanindra/

Learning Outcomes

By the end of this course, you will be able to:

- Design enterprise-grade Lakehouse architectures
- Build scalable ETL pipelines using Databricks
- Implement Bronze, Silver, and Gold data layers
- Perform advanced data transformations using PySpark
- Secure data using Unity Catalog with Gen AI
- Deliver business-ready datasets for analytics
- Build executive dashboards in Power BI
- Gain hands-on experience through production-style Data Engineering projects
- Implement Genie AI, Gen AI, Agentic AI & LLMs
- Automated End to End Workflows with GIT

By end of this course, you can apply for these Trending Job Roles:

1. Databricks Developer
2. ETL Developer (Databricks)
3. Azure Databricks Engineer
4. AI Enabled Databricks Engineer

Course pre-requisites:

None. We shall start from the very basics and proceed with the real-time job activities with scenarios, step by step.

This course includes total of FOUR Modules:

Module 1: MSSQL & TSQL, Query Tuning
Module 2: Databricks (Azure, Python, Spark, PySpark, Genie)
Module 3: Gen AI, Agentic AI & LLMs
Module 4: Realtime Projects, Resume

Total Duration: 11 to 12 Weeks

🔥 Detailed Course Content 🔥

Module 1: MSSQL & TSQL Queries, Tuning

Ch 1: SQL Database Job Roles

- ✓ Database Intro. OLTP, DWH, OLAP
- ✓ DBMS Basics & Technologies. SQL Database Job Roles

Ch 2: SQL Server Configurations

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

Ch 3: SQL Basics V1 (Commands)

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

Ch 4: SQL Basics V2 (Commands, Operators)

- ✓ Create, Alter, Drop, Add
- ✓ Insert, Update, Delete, Truncate
- ✓ Select Options. Basic Data Types

Ch 5: Schemas & Data Imports

- ✓ Data Imports with Excel, CSV. Data Types Detection
- ✓ SQL Operators. Schemas, Table Grouping
- ✓ SQL Batches & 4 Part Naming

Ch 6: Data Types & Variables

- ✓ Integer, Character, MAX Data Types. Decimal & Boolean Data Types
- ✓ Date, Time, Other Data Types. SQL_Variant Types, Variables

Ch 7: Constraints, Keys & RDBMS

- ✓ Null, Not Null Constraints. Unique Key & Primary Key Constraint
- ✓ Foreign Keys, Default, Check. DB Diagrams & ER Models

Ch 8: Normal Forms & RDBMS

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

Ch 9: Joins & TSQL Queries

- ✓ Joins: Table Comparisons. Inner Joins & Matching Data

- ✓ Outer Joins: LEFT, RIGHT. Full Outer Joins & Aliases
- ✓ Self Joins & Aliases

Ch 10: Sub Queries

- ✓ Basic Sub Queries. Aggregations
- ✓ Combining TSQL Queries, Correlated Sub Queries
- ✓ Union, Union All

Ch 11: Group By Queries

- ✓ Group By, Distinct, Having. Cube( ), Rollup( ), Grouping()
- ✓ IsNull(), Coalesce() in TSQL

Ch 12: Joins with Group By, Sub Queries

- ✓ 3 Table, 4 Table Joins. Join & Group By, Sub Queries
- ✓ IIF(), CASE Statement. Exists
- ✓ Query Execution Order

Ch 13: Views & RLS

- ✓ Views: Realtime Usage. DML & SELECT
- ✓ WITH CHECK OPTION, Row Level Security (RLS)
- ✓ Important System Views

Ch 14: Functions & Queries

- ✓ User Defined Functions, Scalar, Table Value Functions
- ✓ Date-Time & String Functions
- ✓ Aggregated & Data Conversion Functions

Ch 15: Advanced SQL & Window Functions

- ✓ Window Functions (Rank), Row_Number, DenseRank
- ✓ Partition By & Order By. Lag & Lead Functions
- ✓ Running Totals. Moving Average

Ch 16: Stored Procedures - 1

- ✓ Stored Procedures: Realtime Use
- ✓ Parameters, SPs with Select, DML
- ✓ System Stored Procedures

Ch 17: Stored Procedures - 2

- ✓ Merge Statement (Upsert),Matched and Not Matched
- ✓ Merge & SP Recompilations

Ch 18: Triggers & Automations

- ✓ Need for Triggers in Real-world
- ✓ DDL & DML Triggers. For / After Triggers
- ✓ Instead Of Triggers. Disabling DMLs Triggers

Ch 19: Transactions & ACID

- ✓ Auto Commit Transaction , Explicit, Implicit Transactions

- ✓ Commit, Rollback, Query Blocking, Checkpoint, ReadPast & NoLock

Ch 20: SQL Query Performance & Indexing
- ✓ Clustered vs Non-clustered Indexes
- ✓ Index Seek vs Scan. Tuning Queries
- ✓ Avoiding SELECT *, Query Optimization

Ch 21: CTEs & TSQL Tuning
- ✓ Common Table Expression, CTEs for Data Retrieval, DML
- ✓ Data Cleansing in TSQL, Duplicate Detection/Removal

Ch 22: Temp Tables, Cursors & Queries
- ✓ Local & Global Temp Tables
- ✓ Cursor Basics – When to Use & Avoid

Ch 23: Bonus / Advanced Module: SQL Server Internals
- ✓ Database Engine Components, Parser, Compiler & Optimizer
- ✓ Parsing and Compilation, Memory Manager & IO Managers

Module 2: Azure Databricks (Python, 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 Tables
- ✓ 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, DDL & DML, DCL
- ✓ Spark Joins, Aggregations, Data Analytics

Ch 7: Spark SQL: Functions
- ✓ Aggregate Functions, Math, Sort Functions

- ✓ String, DateTime Functions, Window & Null Functions

Ch 8: Spark SQL: Time Travel

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

Ch 9: Python Introduction, Print

- ✓ Python Intro & Versions, Python Implementations
- ✓ Python Print(), display(), Single, Multiline Statements

Ch 10: Python Variables

- ✓ Variable Declarations, Variable Values
- ✓ Multi Variable Values, Type Casting, Collections

Ch 11: Python Operators

- ✓ Arithmetic, String Operators, Logical, Assignment Operators
- ✓ Comparison, Format Operators, Indexing, Bitwise Operators

Ch 12: Python Data Types

- ✓ Integer, Float Data Types, String, List Data Types
- ✓ Dictionary Data Type, Tuple Data Type

Ch 13: Python Control Statements

- ✓ If ... Else Statement, ELIF & ELSE IF Statements
- ✓ OR, AND Concepts, FOR and WHILE Loops

Ch 14: Python Modules & DataFrames

- ✓ Pandas, NumPy, DataFrame Concepts
- ✓ Handling Nulls, Data Cleansing Concepts

Ch 15: Python DataFrames Transformations

- ✓ Concat(), Indexing, Merge()
- ✓ Join KIND, ON Clauses, Merging 3 Dataframes

Ch 16: PySpark Concepts

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

Ch 17: Medallion Architecture - 1

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

Ch 18: Medallion Architecture - 2

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

Ch 19: Databricks Delta Lake

- ✓ Schema Evolution, Delta Table API

- ✓ MERGE / Upsert Operations, Delta Transaction Log

Ch 20: PySpark: Widgets

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

Ch 21: Lake Flow Jobs

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

Ch 22: PySpark: Auto Loader

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

Ch 23: LakeFlow Declarative Pipelines

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

Ch 24: Databricks Optimizations

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

Ch 25: Databricks Security, AI

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

Ch 26: Databricks with Power BI

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

Ch 27: Genie AI

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

Ch 28: GitHub & Asset Bundle

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

Ch 29: Apache Airflow

- ✓ Apache Airflow Concepts. Python Scripts, Airflow Jobs
- ✓ DatabricksRunNowOperator. DatabricksSubmitRunOperator

Ch 30: Azure Fundamentals

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

Ch 31: Azure Databricks

- ✓ Azure Databricks Deployments. Azure Databricks Regions
- ✓ Classic Deployments. Access connector for Databricks

Ch 32: Azure Storage & ADLS

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

Ch 33: Azure Databricks with ADLS

- ✓ Creating IAM Users. ADLS Security for Databricks
- ✓ Creating Service Principals. Enterprise Applications
- ✓ Client ID, Secret ID. External Tables (Delta)

Module 3: Databricks AI (GenAI, LLM, Agents, RAG)

Ch 1: AI Concepts

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

Ch 2: Large Language Models (LLM) & Prompt Engineering

- ✓ Introduction to LLMs and foundation models
- ✓ Tokens, context windows and inference
- ✓ Prompt Engineering Techniques
- ✓ Structured outputs and JSON responses

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

- ✓ 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 With LLMs, Customer Reviews & Feedbacks

Ch 6: Unstructured Data Engineering

- ✓ Working with PDF, JSON and text files
- ✓ Ingesting unstructured data, Document Parsing
- ✓ 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, 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. Realtime Implementations.**

Module 4: Real-Time Projects (E-Commerce Domain)

Project 1: HealthCare Domain

Project Requirement:
Solve 20+ Real-World Healthcare Business Requirements.

Project Workflow:
Database Design → SQL Development → 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…

Project 2: 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
- Track sales performance across regions
- Analyse customer purchasing behaviour
- Monitor inventory levels
- 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

Technologies Covered

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

Resume Highlights

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

- PySpark & Spark SQL
- Delta Lake
- Medallion Architecture
- ETL Pipeline Development
- Incremental Data Processing
- Data Quality & Validation
- Unity Catalog, RAG, LLM
- Workflow Automation with Agentic AI
- Performance Optimization

Why Choose This Training?

🚀 Industry-Oriented Curriculum
💼 End-to-End Real-Time Projects & Solutions
🏗️ Medallion Architecture Implementation
📊 100% Hands-on Practical Sessions
⚡ Production-Ready ETL Pipelines
🔥 PySpark & Spark SQL from Basics to Advanced
🎯 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.



Your Learning Journey Starts Here!

Our Address (For Classroom Trainings / In-person Vist)
SQL SCHOOL – Training Institute
201, Sai Anu Avenue, Street No. 3,
Patrika Nagar, Hitech City, Hyderabad, Telangana – 500081.

Call / WhatsApp: +91 99514 40801
Website: www.sqlschool.com
Google Maps: https://share.google/bmMW3OxuAWDHCxSyD