

A promotional banner for a Databricks training course. The background is dark blue with glowing orange and blue geometric shapes. On the left, the Databricks logo is at the top, followed by the word 'MASTER' in a red banner, and 'DATABRICKS' in large white letters. Below this is 'WITH REAL-TIME PROJECTS' and a blue banner with the text 'BUILD. ANALYZE. AUTOMATE. INNOVATE.'. Further down, it says 'The Unified Data Platform for Data Engineering, Analytics & AI'. Below this are four icons: a database cylinder for 'DATA ENGINEERING', a bar chart for 'DATA ANALYTICS', a brain for 'MACHINE LEARNING', and a network diagram for 'AI SOLUTIONS'. On the right, there is a portrait of Mr. Sai Phanindra Tholeti, a man with a mustache wearing a grey blazer. Above his portrait is a blue box with 'LEARN BUILD DEPLOY' and a red box with 'BECOME JOB READY!'. Below his portrait is a dark blue box with his name, title 'Chief Mentor & DataStack Expert', and '20+ Years of Experience'. At the bottom, there is a red banner with 'SQL SCHOOL Premium Quality Training', a logo for '120+ MNC Clients', an 'ISO 9001:2015 Certified' logo, an 'MSME Registered' logo, and a red box with a phone icon, 'Call Today for FREE DEMO +91 99514 40801', and the website 'www.sqlschool.com'.

I'm **Mr. Sai Phanindra**, your trainer and mentor for this course. With 20+ 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/](https://www.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
- Deliver business-ready datasets for analytics
- Build executive dashboards in Power BI
- Gain hands-on experience through production-style Data Engineering projects

## Course Outline:

Module 1: MSSQL & TSQL Queries, Tuning

Module 2: Python

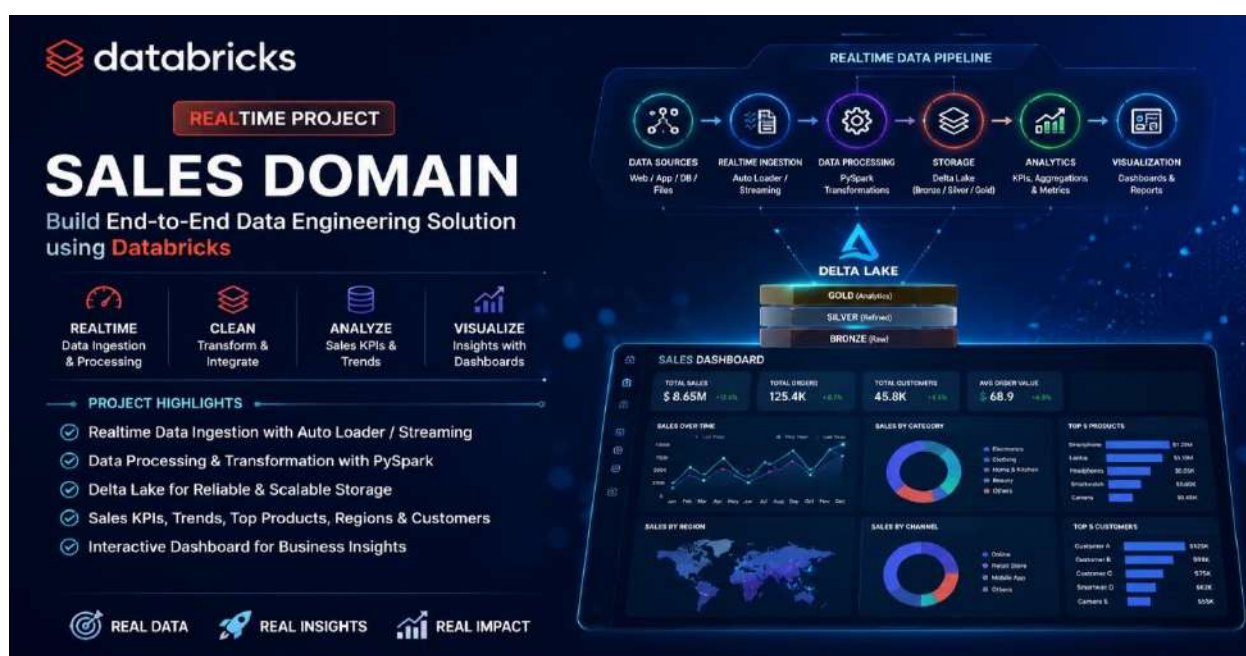
Module 3: Databricks (Spark, PySpark, Genie AI)

Module 4: Real-Time Projects

## Databricks Data Engineer Training – 100% Practical & Project-Based

| Training Plan | Modules                              | Total Duration | Pre-Requisites |
|---------------|--------------------------------------|----------------|----------------|
| Plan A        | 1. Databricks                        | 4 Weeks        | SQL, Python    |
| Plan B        | 1. Python<br>2. Databricks           | 7 Weeks        | SQL            |
| Plan C        | 1. SQL<br>2. Python<br>3. Databricks | 10 Weeks       | None           |

**Training Modes:** LIVE Online Instructor-Led Training || Self-Paced Videos



## 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: Sub Queries**

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

### **Ch 11: Group By Queries**

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

### **Ch 12: Joins & Queries**

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

### **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: SQL Query Performance & Indexing

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

### Ch 20: CTEs & Tuning

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

## Module 2: Python For Data Engineers

### Ch 1: Python Introduction

- ✓ Python Introduction
- ✓ Python Versions
- ✓ Python for Data Engineers
- ✓ Python Architecture

### Ch 2: Python Installations

- ✓ Python Common Line
- ✓ Anaconda Web
- ✓ Anaconda Installations
- ✓ Jupyter Notebooks, IDE

### Ch 3: Python Basics

- ✓ print()
- ✓ display()
- ✓ Multi Line display
- ✓ Debug with AI

### Ch 4: Python Variables

- ✓ Variable Declarations
- ✓ Variable Rules
- ✓ Multi Value Variables
- ✓ Common Value Variables

### Ch 5: Python Operators

- ✓ Arithmetic, String Operators
- ✓ Bit, Comparison Operators
- ✓ Binary Operators
- ✓ Indexing Operators

### Ch 6: Python Data Types

- ✓ Integer, Float

- ✓ String Data Types
- ✓ Type Casting
- ✓ Type Identification

#### **Ch 7: Python Lists**

- ✓ Python Lists
- ✓ List Items, Slices
- ✓ Length & Type
- ✓ Empty Lists, Append

#### **Ch 8: Python Dictionaries**

- ✓ Indexing Dictionaries
- ✓ Edit Key Values
- ✓ Lists inside Dictionaries
- ✓ Delete & Clear

#### **Ch 9: Python Tuples**

- ✓ Python Tuples
- ✓ Defining, Indexing
- ✓ Length(), Type()
- ✓ Mixed Values in Tuples

#### **Ch 10: Python Control Statements**

- ✓ If..Else conditions
- ✓ if..elif..else, range
- ✓ Shorthand if
- ✓ in & negation operators

#### **Ch 11: Python Loops**

- ✓ For Loop, Range
- ✓ While Loop
- ✓ Exit Conditions
- ✓ Break, Continue, Pass

#### **Ch 12: Python SQL DB Access**

- ✓ SQL DB Access with Python
- ✓ pyodbc module, sql functions
- ✓ SQL DB Cursor Connections
- ✓ Filters, Aggregations with SQL

#### **Ch 13: Dataframe Transformations - 1**

- ✓ Concat & Append
- ✓ Merge Dataframes
- ✓ ON and KIND Options
- ✓ Loops with Dataframes

#### **Ch 14: Dataframe Transformations - 2**

- ✓ Pandas - Cleaning Data
- ✓ Replace, Transform Columns
- ✓ Identify & Remove Duplicates
- ✓ dropna(), fillna() Functions

### Ch 15: Dataframe Transformations - 3

- ✓ date, time Functions
- ✓ Now() Function in Python
- ✓ date, time Functions
- ✓ Calendar Data Generation

## Module 3: Databricks, PySpark, Spark, 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, 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 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

## **Module 4: 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

#### **Components For Project**

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

## **Certifications & Career**

### **Databricks Data Engineer Associate**

- ✓ 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

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



- 🔧 Ready to Upgrade Your Skills & Become Job-Ready?
- 📅 Join Our Upcoming Batch – Book Your FREE Demo Today!

☐ WhatsApp: +91 99514 40801 | +91 96666 40801

🌐 [www.sqlschool.com](http://www.sqlschool.com)

🔧 Learn Practical Skills. Build Confidence. Accelerate Your Career!