8-Week Databricks Data Engineering Program

Week 1: Python for Data Engineering

Day 1: Python Basics - Variables, Data Types, and Control Flow

Day 2: Functions, Modules, and Error Handling

Day 3: File I/O and Working with JSON

Day 4: Introduction to Pandas for Data Manipulation

Day 5: Python with PySpark - DataFrames and RDDs

Day 6: Functional Programming - map, filter, reduce

Day 7: Mini Project - Log Parser with Python

Week 2: SQL for Data Engineering

Day 8: SQL Basics - SELECT, WHERE, and JOINS

Day 9: Aggregations, GROUP BY, and HAVING

Day 10: Subqueries and CTEs

Day 11: Window Functions - RANK, ROW_NUMBER

Day 12: CASE/WHEN, PIVOT, and Advanced Filters

Day 13: Nested JSON in SQL + Dot Notation

Day 14: SQL Lab - Data Exploration Project

Week 3: Apache Spark and DataFrames

Day 15: Spark Architecture and Execution Model

Day 16: Creating and Transforming DataFrames

Day 17: Actions, Lazy Evaluation, and Caching

Day 18: Joins and Aggregations in PySpark

Day 19: Window Functions in Spark

Day 20: Using UDFs in Spark SQL

Day 21: Spark Lab - ETL from Raw Logs

Week 4: Delta Lake & Lakehouse Fundamentals

Day 22: Databricks Lakehouse Overview

Day 23: Creating Managed and External Delta Tables

Day 24: Delta Lake Time Travel and History

Day 25: Schema Enforcement and Evolution

Day 26: Partitioning and Z-Ordering

Day 27: OPTIMIZE and VACUUM Commands

Day 28: Lakehouse Lab - Data Lifecycle Demo

Week 5: Advanced Delta Lake Operations

Day 29: MERGE INTO and Upserts

Day 30: COPY INTO for Bulk Loads

Day 31: Insert Overwrite vs Create or Replace

Day 32: Using Generated Columns and Table Comments

Day 33: Delta Table Lineage and Version Rollback

Day 34: Delta Table Constraints and Validation

Day 35: Delta Lab - Change Data Capture Pipeline

Week 6: Real-Time Data with Spark Structured Streaming

Day 36: Introduction to Structured Streaming

Day 37: Auto Loader for File Ingestion

Day 38: Streaming Joins and Watermarks

Day 39: Writing Streams to Delta Lake

Day 40: Triggered vs Continuous Streaming

Day 41: Using APPLY CHANGES INTO with DLT

Day 42: Real-Time Pipeline Lab - IoT Use Case

Week 7: AI Querying, Optimization & Governance

Day 43: Photon and Performance Optimization

Day 44: Unity Catalog - Data Governance and Access Control

Day 45: Audit Logging, Lineage, and Monitoring

Day 46: Generative AI for SQL and Pipeline Insights

Day 47: Auto-Documentation and AI-Assisted Debugging

Week 8: Capstone Project & Certification Prep

Day 48: Capstone Kickoff - Project Design and Planning

Day 49: Capstone Implementation - Bronze to Gold Pipeline

Day 50: Capstone Finalization - Governance and Automation

Day 51: Mock Exam and Certification Strategy