

Azure Databricks: Delta Lake & PySpark ELT

Learn to build, optimise and deploy scalable ELT pipelines in Azure Databricks using Delta Lake, PySpark and Unity Catalog.

5 Days

40 HOURS

17

MODULES

61

TOPICS

5

HANDS-ON ACTIVITIES

COURSE OVERVIEW

This course covers the Databricks Lakehouse architecture, the Databricks workspace, and Delta Lake fundamentals including ACID operations, versioning and schema evolution, alongside Databricks SQL essentials. It then moves into building ELT pipelines with Spark DataFrames and Spark SQL, covering table management, custom schemas, data cleaning, wrangling and transformation. Coverage continues with structured streaming, Auto Loader, medallion architecture and Delta Live Tables, followed by performance optimisation, workflow orchestration, monitoring and logging, and cost tuning. The final modules address business intelligence dashboards, Unity Catalog access control, and deployment with Databricks Asset Bundles.

WHO THIS COURSE IS FOR

Data Engineers

ETL Developers

BI Developers

Data Analysts

Platform Administrators

Data Architects

LEARNING OBJECTIVES

By the end of this 5-day program, participants will be able to:

- Build robust, scalable ELT pipelines in Databricks using PySpark and Spark SQL
- Leverage Delta Lake and Delta Live Tables for both batch and streaming data
- Optimize pipeline performance and cost for real-world workloads
- Monitor and orchestrate data pipelines effectively
- Implement secure deployments using Databricks Asset Bundles and Unity Catalog

17 Modules, 61 Topics

DAY 1 Databricks and Lakehouse Foundations

01

Introduction to the Lakehouse Architecture

TOPICS

- What is a Lakehouse and how it unifies lakes and warehouses
- Lakehouse vs. Data Lake vs. Data Warehouse
- Databricks control plane vs. data plane architecture

02

Navigating the Databricks Environment

TOPICS

- **Databricks UI:** Notebooks, Repos, Jobs, Data, Compute
- All-purpose vs. Job clusters
- Multi-language support (SQL, Python, Scala)
- Git integration for notebook version control

03

Introduction to Delta Lake

TOPICS

- What is Delta Lake and how it enhances the data lake
- Creating Delta tables (managed vs. external)
- **ACID operations:** insert, update, delete, merge
- Table versioning and schema evolution

04

Databricks SQL Essentials

Case Study**TOPICS**

Executing SQL queries in the SQL workspace · Views, temporary views, CTEs · Visualizations and dashboards (intro level)

05

Working with Spark DataFrames and Spark SQL

TOPICS

Spark DataFrames and SQL · Load Dataset from Data Catalog · Load Data into DataFrames Using PySpark · Load Dataset Using Spark SQL and Create Table

06

Table Management and Schema Handling

TOPICS

- Create Temporary and Permanent Tables Using a Notebook
- Demonstrate Custom Schema
- Select and Index/Slice Columns in PySpark DataFrames

07

Case Study

Developing ELT Pipelines

TOPICS

- ELT Pipeline Development with DataFrames and Spark SQL
- Demonstrate Data Cleaning and Preprocessing
- Demonstrate Data Wrangling and Transformation
- Demonstrate Data Wrangling Using Filters and Sorting
- Writing transformed outputs to Delta

08

Structured Streaming and Incremental Processing

TOPICS

- Structured Streaming concepts
- Real-time ingestion with Auto Loader
- Deduplication using MERGE
- Watermarking, late-arriving data, checkpointing

09

Implementing Medallion Architecture

TOPICS

Bronze, Silver, and Gold layer design · Building modular and scalable pipelines

10**Case Study**

Delta Live Tables (DLT)

TOPICS

DLT concepts and pipeline creation · Triggered vs. continuous pipelines · APPLY CHANGES INTO for CDC · Debugging and monitoring DLT pipelines

DAY 4 Optimization, and Monitoring

11

Optimizing Delta Lake Performance

TOPICS

Z-ordering, file compaction, and VACUUM · Schema evolution and constraints · ON VIOLATION clause and conflict handling · Best practices for table size and layout

12

Workflow Orchestration and Monitoring

TOPICS

Creating multi-task workflows using Jobs · Setting dependencies, retries, scheduling · Logs, alerts, execution history

13

Monitoring and Logging Capabilities

TOPICS

- Monitoring pipeline health
- Integration with workspace logs and cluster metrics
- Logging strategies in notebooks and jobs

14**Case Study**

Cost Optimization and Performance Tuning

TOPICS

- Cluster tuning and scaling
- Job execution planning
- Storage optimization (cache, Auto Loader checkpointing)
- **Demo:** Schedule Automation and Optimization

15

Databricks SQL for Business Intelligence

TOPICS

- Writing analytical queries
- Building interactive dashboards and reports
- Sharing dashboards securely and scheduling refreshes

16

Unity Catalog and Access Control

TOPICS

- Catalogs, schemas, and tables in Unity Catalog
- Role-based access control (RBAC)
- Granting/revoking permissions (USAGE, SELECT)
- Cluster modes and data isolation with service principals

17

Case Study

Deployments in Azure Databricks Using Asset Bundles

TOPICS

- Deployments with Databricks Asset Bundles
- Introduction to Asset Bundles for CI/CD
- Packaging notebooks, pipelines, and jobs for deployment
- Automating deployment using workspace Git

HANDS-ON ACTIVITIES

Practise What You Learn

#	MODULE	ACTIVITY
04	Databricks SQL Essentials	Analyze operational data, implement version control, and explore ACID behavior using Delta tables
07	Developing ELT Pipelines	Create an ELT pipeline to clean and transform customer and order data using Spark SQL and PySpark DataFrames
10	Delta Live Tables (DLT)	Stream sales logs into a Medallion architecture and create a Gold layer analytics view with deduplication and DLT

#	MODULE	ACTIVITY
14	Cost Optimization and Performance Tuning	Build, optimize, and monitor an e-commerce analytics pipeline with performance tuning and alerting
17	Deployments in Azure Databricks Using Asset Bundles	Deploy a full-featured data pipeline to QA and Production using asset bundles, access controls, and automated scheduling

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training

Live Online Classes (ILO)

Fly-Me-A-Trainer (FMAT)

Flexi (Self-Paced)

1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com

✉ info@koenig-solutions.com