

Databricks Data Eng: Delta Lake & Streaming

Learn to build data pipelines on Databricks using Delta Lake, Structured Streaming, Delta Live Tables and Databricks SQL.

4 Days

DURATION

11

MODULES

62

TOPICS

COURSE OVERVIEW

This course covers the Databricks Lakehouse Platform and the fundamentals of Delta Lake, including its architecture, ACID transactions, time travel, schema evolution, and table management. It proceeds through relational entities, ETL with Spark SQL, and incremental data processing using Structured Streaming and Auto Loader. It then addresses the Medallion Architecture, Delta Live Tables, and job orchestration, before covering Databricks SQL, permissions management with Unity Catalog, and productionizing dashboards and queries.

WHO THIS COURSE IS FOR

Data Engineers**Data Analysts****ETL Developers****BI Developers****Platform Administrators**

LEARNING OBJECTIVES

- Describe the Databricks Lakehouse Platform and workspace components
- Apply Delta Lake features such as ACID transactions and time travel
- Create and manage Delta tables, views, and catalogs
- Build ETL pipelines with Spark SQL and notebooks
- Implement streaming pipelines using Structured Streaming and Auto Loader
- Apply the Medallion Architecture across bronze, silver and gold layers
- Configure Delta Live Tables pipelines and job orchestration
- Implement access control and permissions with Unity Catalog

11 Modules, 62 Topics

01

Introduction to Databricks

TOPICS

- Overview of Databricks and the Lakehouse Platform
- Data Engineer Role in Databricks
- Key Components of Databricks Workspace

02

Delta Lake Fundamentals

TOPICS

Introduction to Delta Lake · Delta Lake Architecture and File Structure

KEY FEATURES OF DELTA LAKE

ACID Transactions · Time Travel and Versioning · Schema Evolution · Data Skipping and Caching

Creating and Managing Delta Tables · Updating, Deleting, and Merging Data · Delta Lake Optimization Techniques

03

Relational Entities on Databricks

TOPICS

Introduction to Relational Entities · Tables, Views, and Temporary Views · Managed and External Tables · Catalogs and Databases · Data Cataloging and Metadata Management

04

ETL with Spark SQL

TOPICS

- Introduction to Spark SQL
- Querying Data with Spark SQL
- ETL Process with Databricks Notebooks
- Extract data from a single file and from a directory of files
- Create a table from a JDBC connection and from an external CSV file
- Working with validations - duplicate values, unique values, missing values etc.
- Working with UDFs

05

Incremental Data Processing with Structured Streaming and Auto Loader

TOPICS

- Introduction to Structured Streaming
- Data Ingestion with Auto Loader
- Building Stream Processing Pipelines
- Watermarking and Late Data Handling
- Checkpointing and Fault Tolerance
- Streaming to Delta Lake
- ACID transactions in Delta Lake
- Working with Managed Tables

06

Medallion Architecture in the Data Lakehouse

TOPICS

- Overview of Medallion Architecture
- Bronze, Silver, and Gold Layers
- Best Practices for Medallion Architecture
- Implementing Medallion Architecture in Databricks

07

Delta Live Tables

TOPICS

- Introduction to Delta Live Tables
- Declarative Data Pipelines
- Creating and Managing Delta Live Tables Pipelines
- Monitoring and Troubleshooting Pipelines
- Incremental Data Processing with Delta Live Tables

08

Task Orchestration with Databricks Jobs

TOPICS

- Introduction to Databricks Jobs
- Creating and Scheduling Jobs
- Job Dependencies and Workflows
- Identify the benefits of using multiple tasks in Jobs
- Set up a retry policy in case of failure
- Create an alert in the case of a failed task
- Identify that an alert can be sent via email

09

Databricks SQL

TOPICS

Introduction to Databricks SQL · Building and Executing SQL Queries · Creating and Managing Dashboards · Visualization Techniques

10

Managing Permissions in the Lakehouse

TOPICS

- Overview of Permissions and Security
- Access Control with Unity Catalog
- Identify Unity Catalog securable
- Implementing Access Control
- Securing Delta Lake with Table Constraints
- Auditing and Compliance in Databricks

11

Productionizing Dashboards and Queries on Databricks SQL

TOPICS

Designing for Production Grade Dashboards · Performance Optimization Techniques · Scheduled Refresh and Alerts · Dashboard Sharing and Permissions

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