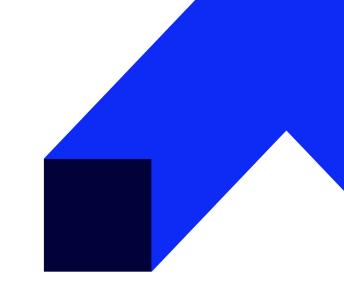


training code: DBX-LKH / ENG DL 1d / EN

# Databricks Lakehouse Architecture

The Databricks Lakehouse training is the third step in the structured training path Fundamental → Explorer → Lakehouse → Transformation. Participants will learn how to design and implement Lakehouse architecture using Delta Lake, Auto Loader, and Structured Streaming. The training also covers cost management, optimization, and governance.





## Training recipients

The training is intended for data engineers and DataOps teams who want to learn how to build and maintain Lakehouse architecture and data processing workflows in Databricks.



#### **Benefits**

- understand the Lakehouse concept and Medallion Architecture
- can create, load, and update data in Delta Lake
- are familiar with batch and streaming data ingestion techniques
- can optimize and monitor Delta tables
- understand governance and lineage in Unity Catalog
- know how to combine transformations, optimization, and quality control in practical workflows
- are prepared for the next stage of the training path Databricks Transformation



### Training program

1.Introduction to Lakehouse architecture



- Lakehouse concept combining Data Lake and Data Warehouse
- Medallion Architecture structure (Bronze, Silver, Gold)
- The role of Delta Lake and Unity Catalog in data management
- Designing data flow logic between layers

#### 2.Delta Lake in practice

- ACID operations and schema enforcement
- MERGE, UPDATE, DELETE, and INSERT modifying Delta tables
- Time travel and change history (DESCRIBE HISTORY)
- Creating managed and external tables in Unity Catalog

#### 3.Data ingestion - batch and stream

- · COPY INTO as a batch data loading method
- Auto Loader (cloudFiles) incremental ingest and schema evolution
- Monitoring streams in the new Streaming UI

#### 4. Optimization and data management

- OPTIMIZE, ZORDER, and VACUUM Delta Lake optimization mechanisms
- Partitioning and query plan analysis
- Liquid Clustering automatic data clustering
- Delta Sharing sharing data across teams and environments

#### 5.Cost management (practical)

- Batch vs streaming costs (Auto Loader, Structured Streaming)
- Impact of OPTIMIZE, ZORDER, and VACUUM on costs
- Resource planning for large tables and pipelines
- Cost architecture in the Bronze-Silver-Gold model

#### 6.Observability & Monitoring (light)

- · Monitoring in Streaming UI and Metrics UI
- · Alerts in Workflows and SQL dashboards
- Best practices for observability in Lakehouse

#### 7. Security fundamentals (light)

- Row-level security and column masking basics
- Token passthrough awareness and scenarios
- Unity Catalog as a governance layer

#### 8. Final project

• Design and implement a mini-Lakehouse with batch and stream data loading, Delta optimization, quality control, monitoring, and cost management



## Expected preparation of the participant

- Completion of Databricks Explorer or equivalent knowledge
- Experience with SQL and basic PySpark



- Basic understanding of cloud and data architecture concepts



# **Training Includes**

• access to Altkom Akademia student

Training method:

The training is conducted in the Databricks cloud environment. Each participant receives their own workspace with access to Unity Catalog, SQL Editor, Notebooks, and a catalog with test data.



# Duration

1 days / 7 hours

## Language

Training: English

• Materials: English