

DATA ENGINEERING

INTERMEDIATE

2 DAYS

Databricks Lakehouse Fundamentals

Learn Databricks Lakehouse fundamentals – Delta Lake tables, Unity Catalog governance, and ADLS Gen2-backed Bronze-to-Gold analytics patterns for teams new to the platform.

Why Databricks Lakehouse?

Lakehouse clarity

See how the lakehouse improves on pure data lakes and traditional warehouses.

Governed analytics

Use Unity Catalog for namespace, permissions, row-level security, and masking.

Delta Lake foundation

Create, query, update, and merge reliable tables with ACID guarantees.

Cloud storage ready

Connect Azure Data Lake Storage Gen2 as the storage layer under Databricks.

What you'll be able to do

- ✓ Understand Lakehouse architecture and how it differs from traditional data warehouses and data lakes
- ✓ Navigate the Databricks workspace: clusters, notebooks, DBFS, and the core interface
- ✓ Create, read, update, and merge Delta Lake tables using both SQL and Python
- ✓ Apply Delta Lake advanced features: time travel, table history, RESTORE, OPTIMIZE, and VACUUM
- ✓ Understand Unity Catalog: metastore hierarchy, catalogs, schemas, and tables
- ✓ Apply data governance controls using Unity Catalog: permissions, row-level security, and column masking
- ✓ Connect Databricks to Azure Data Lake Storage Gen2 as an external storage layer
- ✓ Execute an end-to-end analytics use case from raw ingestion through to a governed Delta table

Who this is for

- Data Engineers
- Data Analysts
- Analytics Engineers
- BI Professionals
- Cloud Engineers
- Data Platform Teams

Curriculum

01 Lakehouse Architecture vs Warehouse and Data Lake

- The evolution of data architectures: data warehouse, data lake, and lakehouse compared
- Why the lakehouse exists: lake limitations, ACID needs, and schema enforcement
- How open table formats and cloud object storage enable analytics and engineering together
- When to choose lakehouse patterns for migration from on-prem warehouses

02 Databricks Workspace, Clusters, Notebooks, and DBFS

- Workspace overview: navigation, notebooks, and compute clusters
- Interactive versus job-oriented cluster thinking for learning workloads
- Databricks File System (DBFS) concepts and common data access patterns
- Unity Catalog metastore awareness at the platform level
- Organising notebooks for Delta labs and end-to-end exercises

03 ADLS Gen2 as the Lakehouse Storage Layer

- How Azure Data Lake Storage Gen2 underpins Databricks lakehouse storage
- Landing raw files and referencing paths from Databricks
- External storage concepts used later with Unity Catalog external locations
- Practical checklist for secure storage access in training environments

04 Delta Lake Foundations: Tables, ACID, and Schema

- Delta Lake basics: Parquet files plus the `_delta_log` transaction log
- Schema enforcement and schema evolution behaviour
- Creating Delta tables with `CREATE TABLE USING DELTA` and `CTAS`
- Writing DataFrames to Delta and validating table metadata
- **Hands-on:** Build and query a Delta Lake table from raw ADLS Gen2 data

05 Reading, Writing, Updates, and MERGE on Delta

- Reading and querying Delta tables with SQL and PySpark
- `UPDATE`, `DELETE`, and `MERGE INTO` for upsert patterns
- Append versus overwrite write modes for curated tables
- Validating row counts and schemas after DML operations

06 Time Travel, History, RESTORE, OPTIMIZE, and VACUUM

- Time travel with `VERSION AS OF` and `TIMESTAMP AS OF`
- Table history and audit using `DESCRIBE HISTORY`
- Restoring tables to a previous state with `RESTORE`
- `OPTIMIZE` for compacting small files and improving scans
- `VACUUM` retention, time travel trade-offs, and safe purge practices

07 Z-Ordering and Delta Performance Basics

- Z-Ordering introduction: co-locating related data for faster filters
- Choosing Z-Order columns for common analytics predicates
- Linking OPTIMIZE and Z-Order to query performance improvements
- Simple performance checklist for new lakehouse tables

08 Unity Catalog: Namespace, Permissions, RLS, and Masking

- Unity Catalog three-level namespace: catalog, schema, and table
- Metastore setup and workspace assignment concepts on Azure
- GRANT, REVOKE, and SHOW GRANTS on Unity Catalog objects
- Row-level security with dynamic views and column masking for sensitive fields
- External locations: registering ADLS Gen2 paths for governed external tables

09 Capstone: Bronze–Silver–Gold to a Governed Gold Table

- Ingest raw ADLS Gen2 data into Bronze Delta tables
- Transform into Silver and business-ready Gold layers
- Register and govern the Gold table in Unity Catalog with least-privilege access
- **Hands-on:** Unity Catalog setup, table registration, access control, and end-to-end pipeline walkthrough