LAB 03: Delta Lake Fundamentals

Duration: ~40 min | **Day:** 1 | **After module:** M03: Delta Lake Fundamentals | **Difficulty:** Intermediate

---

Scenario

> *"New customer data has arrived! Use Delta Lake's MERGE to upsert records without duplicates. Then practice UPDATE, DELETE, and explore time travel to recover from mistakes."*

Objectives

After completing this lab you will be able to: - Read and inspect an update file - Use `MERGE INTO` for upsert operations - Perform `UPDATE` and `DELETE` operations - Use `DESCRIBE HISTORY` to inspect the transaction log - Query previous table versions with time travel - Use `RESTORE` to revert to an earlier version - Understand the impact of `VACUUM` on time travel

---

Prerequisites

- Cluster running and attached to notebook
- LAB 02 completed (customers table exists in Bronze)

Tasks Overview

Open **`LAB_03_code.ipynb`** and complete the `# TODO` cells.

Task	What to do	Key concept
Task 1	Examine the Update File	Read `customers_new.csv` and inspect its content
Task 2	MERGE INTO	Upsert — update existing + insert new records
Task 3	UPDATE	`UPDATE table SET col = value WHERE condition`
Task 4	DELETE	`DELETE FROM table WHERE condition`
Task 5	DESCRIBE HISTORY	View all operations in the transaction log
Task 6	Time Travel	`SELECT * FROM table VERSION AS OF n`
Task 7	RESTORE	`RESTORE TABLE table TO VERSION AS OF n`
Task 8	VACUUM Impact	Understand how VACUUM removes old file versions

---

Detailed Hints

Task 2: MERGE INTO

```
MERGE INTO target USING source
ON target.id = source.id
WHEN MATCHED THEN UPDATE SET *
WHEN NOT MATCHED THEN INSERT *
```

Task 5: DESCRIBE HISTORY

- `DESCRIBE HISTORY table_name` shows all versions
- Each operation creates a new version

Task 6: Time Travel

- By version: `SELECT * FROM table VERSION AS OF 2`
- By timestamp: `SELECT * FROM table TIMESTAMP AS OF '2024-01-01'`

Task 7: RESTORE

- `RESTORE TABLE table_name TO VERSION AS OF n`
- Creates a NEW version (does not delete history)

Task 8: VACUUM

- Default retention: 7 days
- After VACUUM, time travel to versions older than retention fails

---

Summary

In this lab you: - Performed MERGE to upsert customer data - Used UPDATE and DELETE for DML operations - Inspected the transaction log with DESCRIBE HISTORY - Queried historical data using time travel - Restored a table to a previous version - Understood VACUUM's impact on time travel

> **Exam Tip:** MERGE is the key pattern for CDC/upsert. RESTORE creates a new version (non-destructive). VACUUM removes files older than retention period — after VACUUM, time travel to those versions fails. Default retention is 7 days.

> **What's next:** Day 2 starts with LAB 04 — optimizing Delta tables with OPTIMIZE, Z-ORDER, VACUUM, and Liquid Clustering.