

LAB 03: Delta DML & Time Travel

Duration: ~35 min

Day: 2

After module: M03: Delta Lake Fundamentals

Difficulty: Intermediate

Scenario

“RetailHub received an updated customer list – some are new customers, some have changed addresses. You need to merge this update into the existing customers table. Later, someone accidentally deletes critical records. Use Time Travel to recover them.”

Objectives

After completing this lab you will be able to: - Use `MERGE INTO` to handle insert/update/delete in a single operation - Perform `UPDATE` and `DELETE` on Delta tables - Query historical table versions using Time Travel - Use `RESTORE` to recover from accidental data loss - Inspect table history with `DESCRIBE HISTORY`

Part 1: MERGE INTO (~15 min)

Task 1: Examine the Update File

Load `customers_new.csv` and compare with the existing Bronze customers table. How many are new vs. existing?

Task 2: MERGE (Upsert)

Use `MERGE INTO` to: - **UPDATE** existing customers (match on `customer_id`) with new values - **INSERT** new customers that don't exist yet

Exam Tip: `MERGE` supports `WHEN MATCHED`, `WHEN NOT MATCHED`, and `WHEN NOT MATCHED BY SOURCE` clauses. Know all three.

Task 3: Conditional MERGE

Modify the `MERGE` to only update if the email has actually changed (add a condition to `WHEN MATCHED AND ...`).

Part 2: UPDATE & DELETE (~10 min)

Task 4: UPDATE Records

Update all customers from a specific city – change their `state` value.

Task 5: DELETE Records (Accidental!)

Simulate an accident: `DELETE FROM customers WHERE country = 'US'`.

This will delete a significant portion of data. (We'll fix it next!)

Part 3: Time Travel & Recovery (~10 min)

Task 6: DESCRIBE HISTORY

Run `DESCRIBE HISTORY` on the customers table. Observe the versions and operations.

<screen = DESCRIBE HISTORY output showing multiple versions: WRITE, MERGE, UPDATE, DELETE operations with timestamps>

Task 7: Query a Previous Version

Use `VERSION AS OF` to read the table before the accidental delete.

Task 8: RESTORE the Table

Use `RESTORE TABLE ... TO VERSION AS OF X` to undo the accidental delete.

Exam Tip: `RESTORE` creates a new version in the history. It doesn't delete history – it adds a new entry. The old versions remain accessible for further Time Travel.

Summary

In this lab you: - Used `MERGE INTO` for upsert operations - Performed `UPDATE` and `DELETE` on Delta tables - Queried historical versions with Time Travel - Restored a table after accidental deletion

What's next: LAB 04 - Optimize your Delta tables for query performance.