

Your pipeline reloads two years to add one day.

D8A.ACADEMY/RESOURCES/INCREMENTAL-LOAD

FULL LOAD

TRUNCATE + INSERT	simple, correct, slow
every night	← reads everything to add one day

INCREMENTAL

WHERE updated_at > :mark	only what changed
watermark	store the max you loaded
upsert	updates too, not only inserts

REMEMBER

late data	← lands after the watermark
overlap	reload the last 2 days, dedupe
full load	still right for small tables

WORTH MEMORIZING

01 The watermark pattern

Read the max you loaded last time, pull everything above it, write it with an upsert, store the new max.

```
-- 1. where did we stop?
SELECT MAX(updated_at) FROM dw.orders; -- :mark

-- 2. only what changed since, with a 2-day overlap
INSERT INTO dw.orders
SELECT * FROM src.orders
WHERE updated_at > :mark - INTERVAL '2 days'
ON CONFLICT (order_id) DO UPDATE
SET status = EXCLUDED.status,
    updated_at = EXCLUDED.updated_at;
```

→ The overlap costs two days of rows and buys you the late ones. The upsert makes the overlap harmless: a row loaded twice is updated, not duplicated.

02 Trap: what an incremental load misses

Three cases a full load handles for free and an incremental one has to plan for.

deletes	A deleted source row never shows up above the watermark. Soft-delete flags, or a periodic full compare.
late data	A row updated at 23:59 that lands at 00:10, after the watermark moved. Hence the overlap.
no updated_at	No reliable change column, no incremental load. Fix the source, or use CDC.

→ Rule of thumb: if the table fits a full load in minutes, do the full load. Incremental is for when it does not.