

Monday's number changed on Wednesday.

D8A.ACADEMY/RESOURCES/LATE-ARRIVING-DATA

THE PROBLEM

late events	logged Monday, loaded Wednesday
job at 00:05	← Monday is not all there yet

THE ANSWER

lookback window	recompute the last N days, every run
upsert those days	← insert again = duplicates
watermark	the day before which nothing moves

REMEMBER

pick N	from the lag you measured, not a guess
two timestamps	keep event_time and loaded_at
tell the readers	the last 3 days may still move

WORTH MEMORIZING

01 The daily job, with a lookback

Recompute the last three days on every run and write them in place. Yesterday is always included, and so are the two days that may still receive events.

```
INSERT INTO daily_orders (day, orders)
SELECT event_time::date, COUNT(*)
FROM events
WHERE event_time >= CURRENT_DATE - 3
GROUP BY 1
ON CONFLICT (day) DO UPDATE
SET orders = EXCLUDED.orders;
```

→ Three days of rows are rewritten each night. A rerun gives the same table (sheet O79). Readers see Monday move until Wednesday, then it settles: that is the watermark.

02 Trap: the lookback that is too short

Three ways the daily table stays wrong after the fix.

YOU DID	WHAT HAPPENS	DO INSTEAD
insert once at 00:05	Monday frozen, late events never counted	lookback + upsert
N = 1, guessed	the batch that lands 2 days late never shows	measure loaded_at - event_time
bucket by loaded_at	Monday's events counted as Wednesday's	bucket by event_time

→ The lag is a distribution. Plot loaded_at minus event_time once, pick N where 99% of the rows have arrived, and write that N next to the table for its readers.