

ETL Validation SQL Query Pack

techfitflow.com | Generic object names; map to your environment | Run in non-production first | Report keys and counts, never PHI

1. Eleven checks

#	Check	What it catches	Pass rule
1	Load log reconciliation	Internal row leaks	read = written + rejected
2	Independent source-to-target count	Selection errors	Gap = 0 for same window
3	Control totals (sum, min, max)	Lost or altered measures	Diff = 0
4	Row-level hash, FULL OUTER JOIN	Missing, extra, changed rows	Zero rows returned
5	Transformation rule CASE test	Mapping errors	Zero mismatches
6	Uniqueness at grain	Duplicate keys	Zero rows
7	Unknown members and orphans	Broken dimension keys	Orphans = 0; unknown rate under threshold
8	SCD2 overlaps, multiple current rows	Fact duplication on join	Zero rows
9	Late-arriving anti-join	Watermark loss	Zero rows
10	Truncation and precision	Silent cuts, rounding	Zero rows; types match
11	Rerun fingerprint	Append instead of upsert	Fingerprint unchanged

2. Failing check decision tree

- [] 1. Load finished? Check etl.load_log status. No: rerun, retest.
- [] 2. Source changed after extract? Compare update timestamps. Yes: widen window or freeze snapshot.
- [] 3. Test wrong? Trace one key by hand. Yes: fix the query.
- [] 4. One key explains it? Trace through staging to the stage that dropped it.
- [] 5. Systematic? Group by date, department, status. Look for a rule or mapping gap.

3. Validation pipeline checkpoints

Checkpoint	Question	Owner
V1 Extract counts	Source count equals extract?	ETL developer
V2 Staging fidelity	Staging equals extract?	ETL developer
V3 Transformation rules	Rules produce expected values?	QA
V4 Load integrity and SCD	Unique, linked, no overlaps?	QA
V5 Reject review	Rejects explained and owned?	Data steward
V6 Report reconciliation	Report equals source totals?	QA and BA

4. Core SQL

Hash comparison (Check 4)

```
SELECT COALESCE(s.k, t.k) AS key_value,
       CASE WHEN s.k IS NULL THEN 'extra in target'
            WHEN t.k IS NULL THEN 'missing in target'
            ELSE 'value mismatch' END AS problem
FROM s FULL OUTER JOIN t ON t.k = s.k
WHERE s.k IS NULL OR t.k IS NULL OR s.h <> t.h;
-- h = HASHBYTES('SHA2_256', CONCAT_WS('|', ISNULL(col, '~'), ...))
-- Convert datetimes with style 121 on both sides
```

SCD2 overlap (Check 8)

```
FROM dw.dim_department a JOIN dw.dim_department b
     ON a.department_id = b.department_id AND a.department_key < b.department_key
AND a.effective_from < ISNULL(b.effective_to, '9999-12-31')
AND b.effective_from < ISNULL(a.effective_to, '9999-12-31');
```

Late-arriving rows (Check 9)

```
WHERE s.UPDATE_DATE >= DATEADD(day, -3, @run_date)
AND NOT EXISTS (SELECT 1 FROM dw.fact_encounter f
                WHERE f.encounter_csn = s.PAT_ENC_CSN_ID AND f.source_update_date >= s.UPDATE_DATE)
```

5. Source-to-target mapping template

Target column	Source	Rule	Null handling	Test	Check

6. Load sign-off checklist

Item	Evidence	Owner	Done
Counts reconcile to independent source query	Check 2 output	QA	
Control totals match	Check 3 output	QA	
Row hashes match for the window	Check 4 output	QA	
No duplicates, overlaps, or orphans	Checks 6-8	QA	
Late-update replay passed	Check 9 output	QA	
Rerun fingerprint unchanged	Check 11 output	ETL developer	
Thresholds signed by data steward	Mapping document	Data steward	

Fail any check that examined zero rows. Sources: Microsoft Learn HASHBYTES (T-SQL); Kimball Group dimensional modeling techniques.