

Title

The Delivery Challenge

Persona

You are Lemlem, a data team member building a client's data warehouse.

Executive Summary

Lemlem discovered **non-trivial reconciliation gaps** between Finance and Delivery during proactive validation. Because the program is on a fast-track schedule with no validation buffer, the immediate goal is to secure **client agreement for a time-boxed investigation** to isolate whether gaps are caused by data quality, timing/reporting windows, or definition mismatches-before these issues become embedded in the warehouse as "truth."

This report includes a **client-ready update** that balances urgency with diplomacy, proposes a concrete and limited scope, and requests explicit approval to allocate time to root-cause the discrepancies.

Scenario

Lemlem is a recently appointed mid-level Data Engineer tasked with spearheading the development of a centralized data warehouse for a primary client. As a new addition to the team, Lemlem faces the "Liability of Newness"-a sociological phenomenon where a lack of established trust and institutional knowledge complicates the navigation of internal politics. This role serves as Lemlem's debut in high-stakes client delivery, requiring a delicate balance between technical rigor and executive diplomacy.

The project environment is characterized by a significant internal schism regarding digital transformation. While senior leadership views the centralized warehouse as a catalyst for organizational scaling, the operational staff exhibits varying degrees of "information hoarding." This defensive posture suggests a perceived threat to departmental autonomy or a fear of exposure regarding legacy process inefficiencies. Consequently, data discovery is hindered by stakeholders who are protective of their datasets and resistant to cross-functional inquiry.

The project is currently operating under a compressed "fast-track" schedule with zero allocated buffer for iterative validation. The discovery of discrepancies between Finance and Delivery datasets introduces a critical project management dilemma. Because the project plan is rigid, the effort required for a comprehensive root-cause analysis represents a zero-sum trade-off. Lemlem must navigate the "Project Management Triangle," where the pursuit of data integrity (Quality) directly threatens either the project timeline (Time) or the team's human capital (Scope/Burnout), necessitating a strategic decision on resource prioritization.

During validation that you undertook on your own, you discover some curious discrepancies between finance data and delivery data. Totals and counts do not fully reconcile.

The root cause is unclear. Possible explanations include:

- * Data quality issues
- * Timing/reporting differences
- * Definition mismatches
- * Team misunderstanding

- * Something nefarious
- * System inconsistencies

There is no evidence of wrongdoing, yet.

Your manager asks you to draft a client update to ask for agreement to allocate some time to look into the discrepancies.

What We Observed (High Level)

During initial reconciliation checks (using sample extracts as a stand-in for the broader legacy landscape), the types of mismatches that commonly appear include:

- * **Data quality issues**: missing/blank join keys, duplicate transaction rows, inconsistent identifiers across systems
- * **Timing/reporting differences**: payments recorded in one period while delivery events land in another; late-arriving events
- * **Definition mismatches**: "paid" vs "delivered" vs "completed" are not necessarily equivalent business events
- * **Process/system inconsistencies**: retries, partial captures/refunds, re-shipments, returns, and operational overrides

These are normal in legacy environments with many contributors, but they must be resolved (or explicitly modeled) to avoid downstream reporting disputes.

Client Update Draft (Request for Time-Boxed Reconciliation Investigation)

Subject: Request to time-box reconciliation of Finance vs Delivery metrics

Hi [Client Name/Team],

As part of our ongoing validation while building the centralized data warehouse, we ran an initial reconciliation between the Finance payments dataset and the Delivery status dataset. We're seeing **small but material differences** in counts and totals that prevent a clean tie-out between "paid" and "delivered" outcomes.

At this point, the **root cause is not yet clear**. Common explanations in systems like this include:

- * Duplicate or retried transactions (e.g., multiple payment attempts for the same order)
- * Missing or inconsistent identifiers used to link Finance and Delivery records
- * Timing differences between when a payment is recorded vs when a delivery event is recorded (month-end and late-arriving updates)
- * Definition differences (e.g., paid vs delivered vs returned/refunded vs failed)

Importantly, there is **no evidence of wrongdoing**-this appears consistent with typical legacy-system behavior and reporting logic drift over time. However, if we proceed without alignment, the warehouse could hard-code assumptions that later cause executive reporting disagreements.

Request: We'd like your agreement to allocate a **time-boxed 1-2 business days** (or equivalent effort) to isolate and document the discrepancy drivers and agree on the "source-of-truth" definitions and join rules. This will allow us to move forward with confidence and avoid rework.

Proposed approach (time-boxed):

1. **Define reconciliation rules** with Finance + Delivery stakeholders (30-60 minutes)
 - What constitutes a unique payment? How to treat retries/partials/refunds?
 - What constitutes a successful delivery? How to treat returns/failures/re-shipments?
2. **Run targeted checks** across the legacy extracts (same day)
 - Duplicates, missing keys, late events, status transitions
3. **Decision memo** (same or next day)
 - Final definitions, mapping rules, known exceptions, and any gaps to track

Decision needed from you: Please confirm whether you prefer:

- * **Option A (recommended):** Approve the 1-2 day time-box now to resolve and document reconciliation rules, or
- * **Option B:** Proceed on schedule with provisional rules (we can do this, but it increases the risk of later rework and metric disputes).

If you approve Option A, we will share a short memo and proposed implementation rules immediately after the time-box completes.

Thanks,
Lemlem
Data Engineering

Materials Provided

You receive:

- * Client brief (above)
- * Sample data (extracts)
- * Guidance on desired next actions
- * (Assume materials may be imperfect or incomplete - like real projects.)