

# Data Modeling

Elements of Data Integration (CS 333E)

Feb 13, 2026

# Where we are in our data infrastructure journey

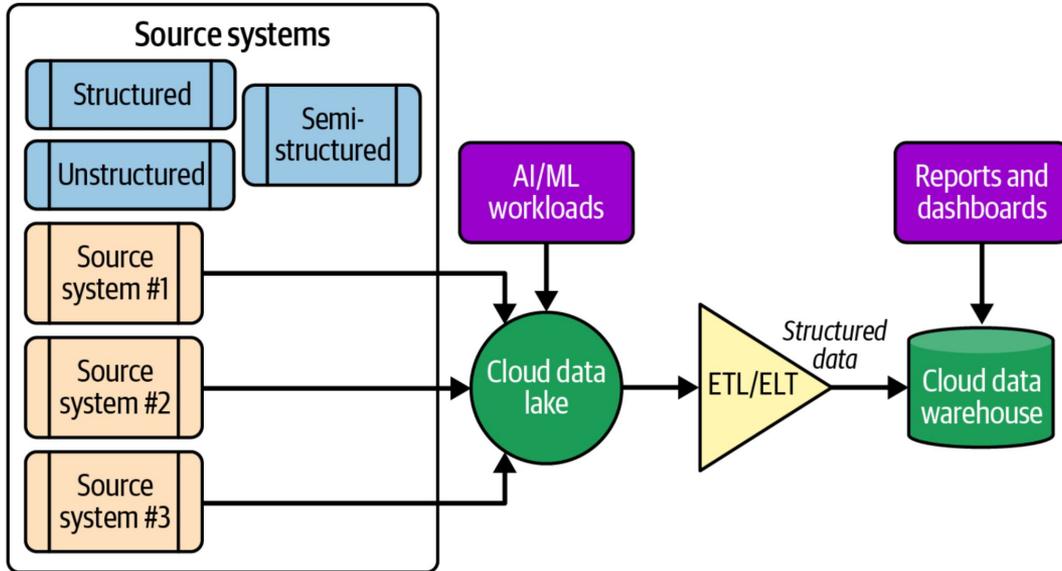


Figure 2-3. Data platform based on a combined architecture (cloud data lake + cloud data warehouse)

What we have done:

- Extraction
- Ingestion
- Enrichment
- Augmentation

What we will do next:

- Modeling
- Transformation

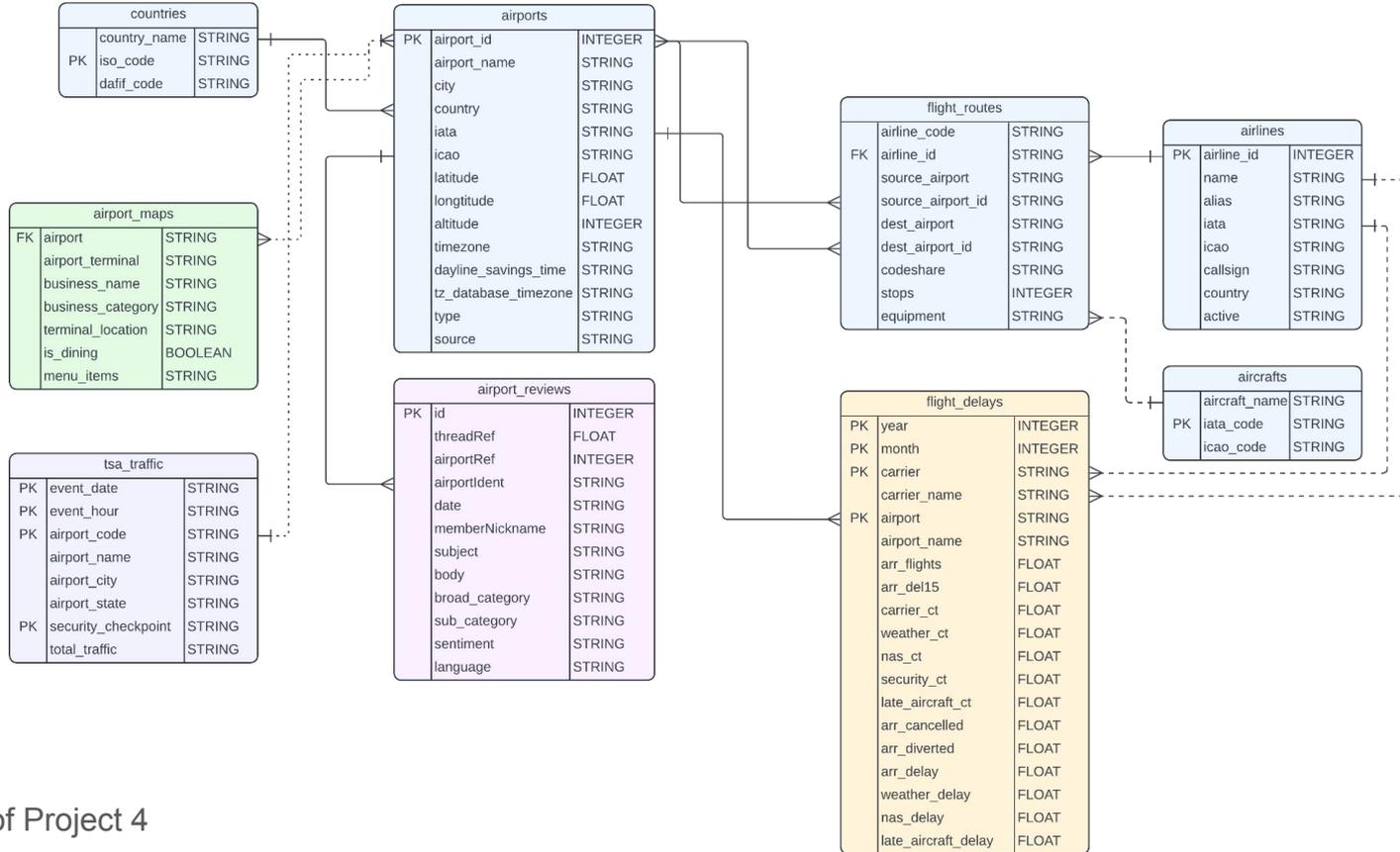
## **Future State: Normalized Data Model for our Lakehouse**

Goal 1: Tables that connect across sources.

Goal 2: Fields which are of interest to multiple stakeholder groups and personas.

Goal 3: Foreign keys that enforce referential integrity.

# Current State: Air Travel Raw Layer





## The Staging Layer

- Only relevant fields
- Semantic data types (anomaly 6)
- Descriptive field names
- Descriptive entity names
- Only scalar values (anomaly 7)

## The Normalized Layer

- Harmonized identifiers (anomaly 9)
- Referential integrity (anomalies 8 and 10)

## The Mart Layer

Marts act as simple interface to underlying model, hiding the complexities and subtleties of ingestion.

They are tailored to a specific use case and persona.

They are backed by a normalized schema, its canonical source.

They can come in multiple flavors: snapshots, real-time, sampled.  
These flavors are consistent with each other.

# Mart Concepts

- Dimensions
- Facts
- History
- Snapshots
- Sampled

