

# Data Models

Foundations of Data Warehousing (CS 378)  
Second Edition

September 26, 2025

# Where are we in our data journey?

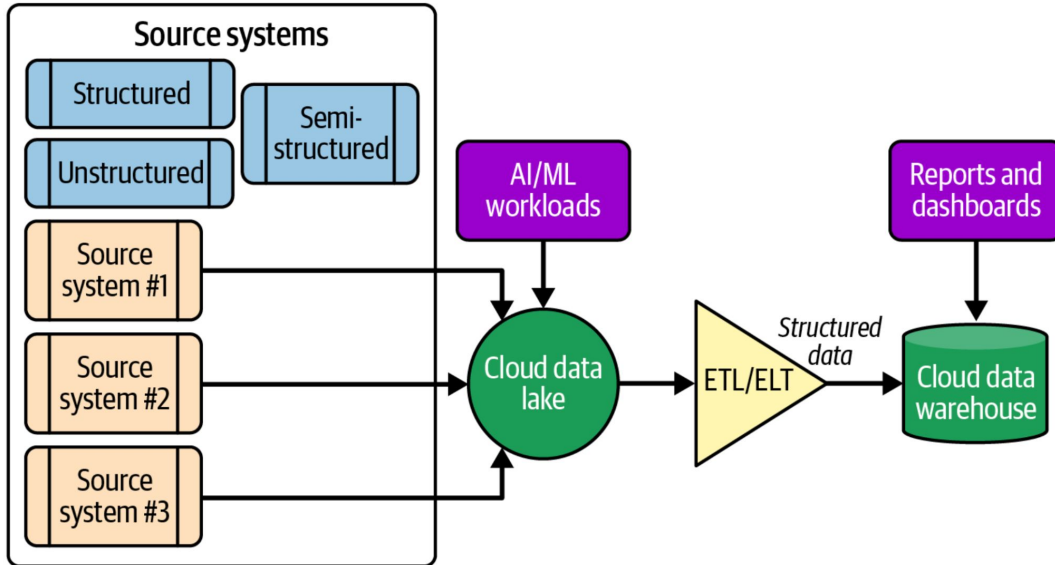


Figure 2-3. Data platform based on a combined architecture (cloud data lake + cloud data warehouse)

What we have done so far:

- Data extraction
- Data loads
- Data exploration
- Data augmentation

What we will do next:

- Data transformation

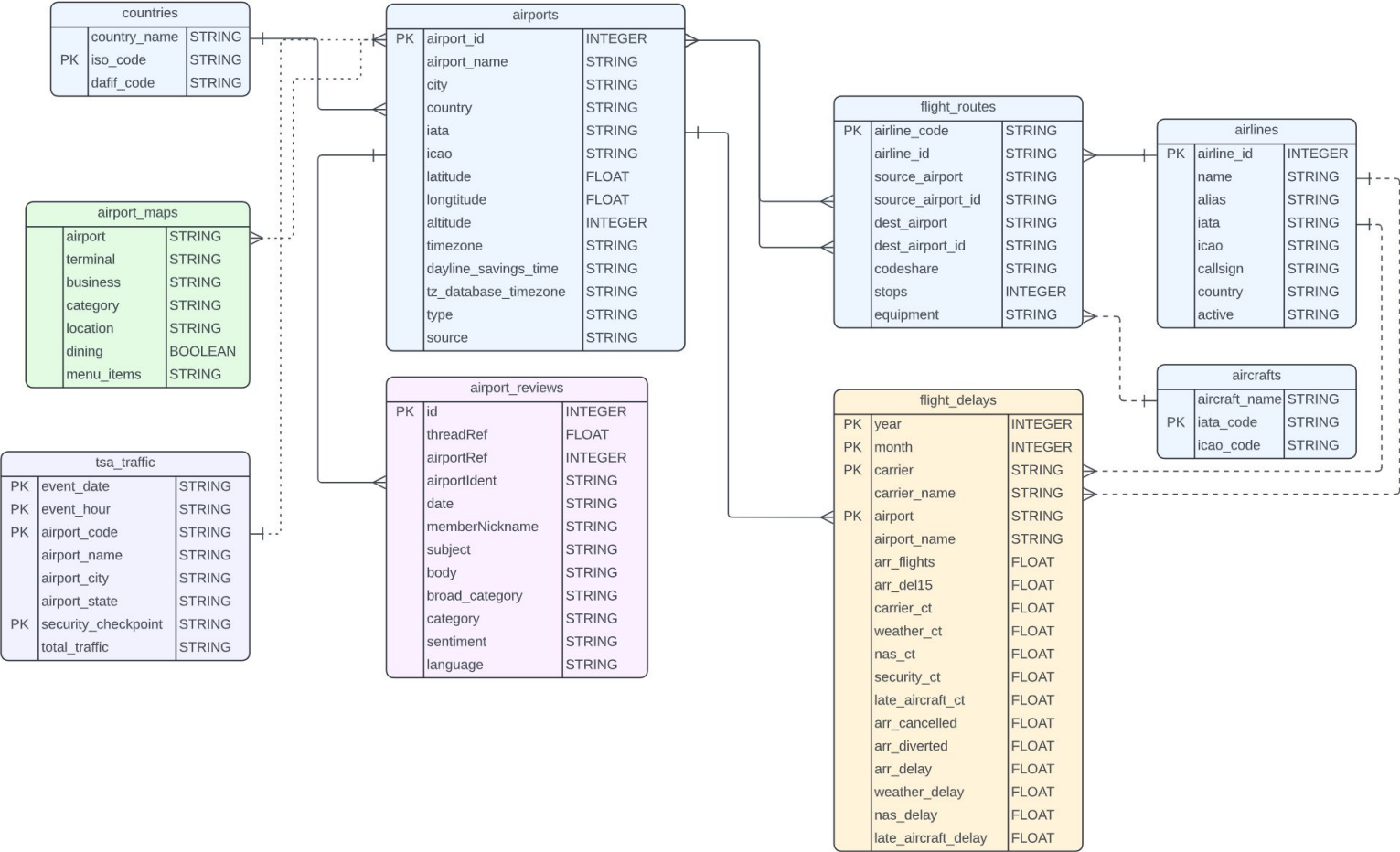
## **The Normalized Data Model**

Tables that join across multiple sources.

They contain fields that are relevant to different personas (data scientists, business analysts, etc.).

They enforce referential integrity.

Air Travel Data Model - Raw Layer as of Project 4



## The Staging Layer

- Relevant fields
- Semantic data types (anomaly 6)
- Descriptive field names
- Descriptive entity names
- Scalar values (anomaly 7)

## **The Intermediate Layer**

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

## **The Dimensional Data Model**

Tables are exposed in a simple schema, hiding the complexities and subtleties of sources.

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

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

## The Mart Layer

- Dimensions
- Facts
- History
- Snapshots
- Sampled

