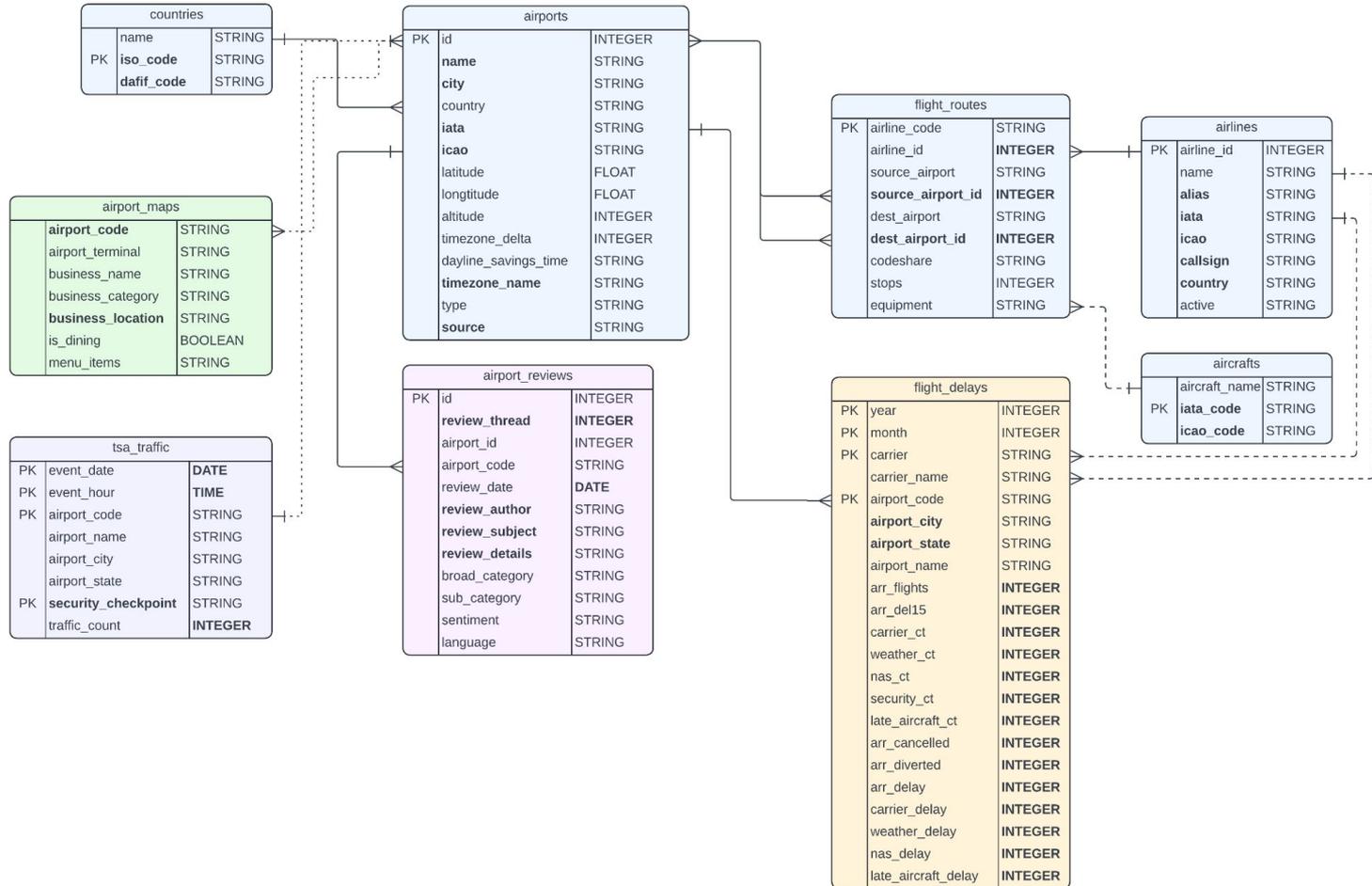


# Data Modeling

Elements of Data Integration (CS 333E)

Mar 6, 2026

# Air Travel Data Model - Staging Layer as of Project 5



## Data anomaly 8: example 1

The raw layer of the lakehouse includes a table containing nested lists within a single field. Unlike Criterion 7, these elements represent multiple values of a single attribute or feature.

Staging table: countries

Row	name	iso_code	dafif_code	_data_source
1	Afghanistan	AF	AF	openflights
2	Albania	AL	AL	openflights
3	India	IN	IN	openflights
4	India	IN	BS	openflights
5	Palestine	PS	GZ	openflights
6	Palestine	PS	WE	openflights
7	United States	US	US	openflights



Normalized table: Country

Row	iso_code	name	dafif_codes	_data_source
1	AF	Afghanistan	AF	openflights
2	AL	Albania	AL	openflights
3	IN	India	IN	openflights
			BS	
4	PS	Palestine	GZ	openflights
			WE	
5	US	United States	US	openflights

Convert the dafif code from a STRING type to ARRAY<STRING>. This makes sense because dafif codes is not a critical feature, it will only be used for direct data retrievals (i.e. it won't be involved in filters, joins, group bys, etc.).

## Data anomaly 8: example 2

The raw layer of the lakehouse includes a table containing nested lists within a single field. Unlike Criterion 7, these elements represent multiple values of a single attribute or feature.

Staging table: airport\_maps

Row	business_name	is_dining	menu_items
1	Texas Monthly News/Starbucks	true	Coffee, Espresso drinks, Pastries, Sandwiches (Starbucks); Newspapers, Magazines (Texas Monthly News)
2	Mountain House Diner	true	Diner-style breakfast, lunch, and dinner options (e.g., burgers, sandwiches, eggs)
3	Tap and Pour	true	Impossible burger, chimichurri salmon, classic ribeye, breakfast, lunch, dinner
4	The Springs	true	American cuisine, breakfast, lunch, and dinner entrees, grab-and-go



Normalized table: Food\_Beverage

Row	business_name	menu_item
1	Mountain House Diner	Diner-style breakfast
2	Mountain House Diner	lunch
3	Mountain House Diner	and dinner options (e.g.
4	Mountain House Diner	burgers
5	Mountain House Diner	sandwiches
6	Mountain House Diner	eggs)

In this case, we want to be able to filter, aggregate, group on menu items. For example, what are all the businesses serving lunch? So, rather than convert it to an ARRAY<STRING> type, we flatten the items into their own records. Note that we still have some cleaning to do to the normalized table (e.g. remove stop words, parenthesis, etc.)

# Data anomaly 9: example 1

There exists two tables in the raw layer of the lakehouse which originated from different sources and which have similar data. These tables use two different identifier systems to refer to the same entity.

Staging table: airports

Row	id	name	country	iata	icao
1	11744	Alert Bay Airport	Canada	YAL	CYAL
2	11823	Burley Municipal Airport	United States	BYI	KBYI
3	13469	Ghazni Airport	Afghanistan	GZI	OAGN
4	11780	Pikwitonei Airport	Canada	PIW	CZMN
5	11030	Port Keats Airport	Australia	PKT	YPKT
6	11789	RAF Kinloss	United Kingdom	FSS	EGQK
7	12043	Warren Airport	Australia	QRR	YWRN
8	11669	Wilpena Pound	Australia	HWK	<i>null</i>



Normalized table: Airport

Row	icao	iata	name	country_code
1	CBB6	YAL	Alert Bay Airport	CA
2	KBYI	BYI	Burley Municipal Airport	US
3	OAOG	GZI	Ghazni Airport	AF
4	CZTP	PIW	Pikwitonei Airport	CA
5	YPKE	PKT	Port Keats Airport	AU
6	EGQK	FSS	RAF Kinloss	GB
7	YWRN	QRR	Warren Airport	AU
8	YWPL	HWK	Wilpena Pound	AU

Promoted the icao field in the Airport table to primary key by correcting errors and filling in nulls. The schemas of Airport's dependent tables (e.g. Airport\_Establishment, Airport\_Review, etc.) were re-created with the icao code as their foreign key.

# Data anomaly 10: example 1

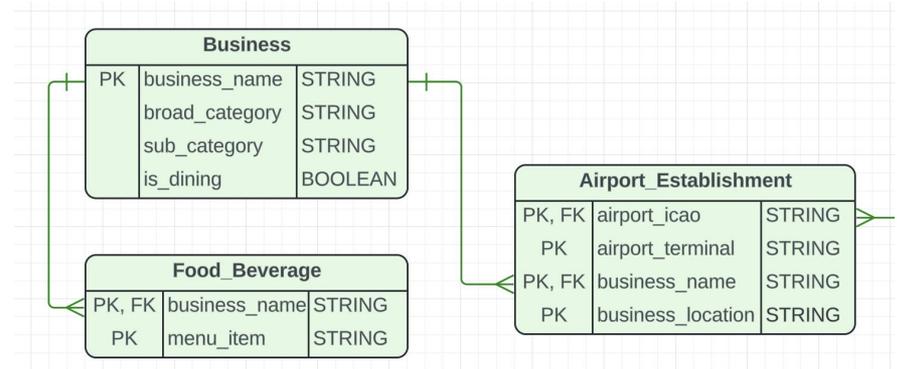
A table in the raw layer of the lakehouse represents multiple logical entities. This structure results in data redundancy and the unnecessary repetition of values.

Staging table: airport\_maps



Row	airport_code	business_name	business_category	business_location
1	hdn	jetBlue	Airline Services	Upper right concourse, near Vending Machines (6)
2	lax	iStore Boutique	Shopping	Near Gate 131
3	bdl	iStore	Shopping	Central concourse, near Black Bear Saloon
4	hou	iStore	Retail	Post-Security, Central Concourse
5	sfo	iStore	Shopping	Level 2 (Map item 5)
6	las	iCandy	Candy Store / Sweets	Gate C
7	syr	eSavvy	Retail/Electronics	Level 2, Item 10. Near Item 9 Stop The Bleed, Item 11 Military Courtesy Room, Item 12 Pay & Parking Kiosk.
8	sfo	Zoom Travel St...	Shopping	Level 2 (Map item 39)

Normalized tables: Airport\_Establishment, Business, and Food\_Beverage



Observe that a business can operate in multiple locations. We therefore separate its category information from its specific airport records. This decoupling minimizes redundancy by ensuring general business details aren't repeated for every site.

Air Travel Data Model - Normalized Layer as of Project 6

