

# ERD - Renuka Dixit - IMT 577B

---

## Schema and modeling decisions

My data model is designed as a Snowflake schema.

I initially wanted to build a Star schema because I had read that outrigger dimensions should be used sparingly and I didn't want to complicate my queries by having to add more joins. But the address and date data was repeated in multiple tables so I decided to make separate dimension tables for them (Dim\_Address and Dim\_Date) and linked them to other dimension tables.

In some scenarios though, I decided not to have outriggers. E.g. Initially, Dim\_Product was one dimension table and Dim\_Product\_Category was another dimension table that was connected to it. Since it only connected to one dimension table, I decided to merge them in a way that I added ProductCategoryID in the Dim\_Product dimension table. Also, I noticed PhoneNumber and EmailAddress in multiple tables and thought about creating a separate Dim\_Contact\_Info table but decided against it because it felt like I was creating too many of these outriggers and this one was less important than the other two.

## Business process

The Entity Relationship Diagram is organized around the Sales of the company.

## Dimension tables

I have 7 dimension tables:

### 1. Dim\_Store

- This table contains data from the Store spreadsheet
- It has 1 foreign key for store address
- It has 1 natural and 1 surrogate key for StoreID

### 2. Dim\_Product

- This table contains data from the Product spreadsheet
- It has 1 natural and 1 surrogate key for ProductID
- I decided to add ProductCategoryID to this table from the ProductCategory spreadsheet instead of creating a separate dimension table for ProductCategory

### 3. Dim\_Customer

- This table contains data from the Customer spreadsheet
- It has 1 foreign key for customer address

- It has 1 natural and 1 surrogate key for CustomerID

#### 4. Dim\_Reseller

- This table contains data from the Reseller spreadsheet
- It has 1 foreign key for reseller address
- It has 1 natural and 1 surrogate key for ResellerID

#### 5. Dim\_Channel

- This table contains data from the Channel spreadsheet
- It has 1 natural and 1 surrogate key for ChannelID
- I decided to add ChannelCategory to this table from the ChannelCategory spreadsheet instead of creating a separate dimension table for ChannelCategory

#### 6. Dim\_Date

- This table contains date information
- It's a separate dimension table because it's present in most of the other dimension tables and is related to all the fact tables
- I've broken the date down into DayName (e.g., Monday), DayNumber (e.g., 28), MonthName, MonthNumber and Year because we'll be looking at trends in several ways. E.g., I might want to see if more sales occur on a certain day of the week or a certain month of the year and this break-down will help me find patterns/trends.

#### 7. Dim\_Address

- This table contains address information
- It's a separate dimension table because it's present in multiple other dimension tables so it made sense to reduce the duplication of data
- It has 1 surrogate key for address

## Fact tables

I have 3 fact tables:

#### 1. Fact\_Sales\_Detail

- This table contains sales data (such as SalesQuantity and SalesAmount) from the Sales Detail spreadsheet.
- The grain is day (for daily sales)
- It has 7 foreign keys
- It has 2 natural keys of SalesHeaderID and SalesDetailID

#### 2. Fact\_Target\_Channel\_Reseller\_Store

- This table contains target sales amounts for different channels, stores and years from the Target Data Channel Reseller & Store spreadsheet
- The grain is year (for yearly target sales amount for specific stores and channels)
- It has 4 foreign keys

### 3. Fact\_Target\_Product

- This table contains target sales quantity for specific products and years from the Target Data Product spreadsheet
- The grain is year (yearly target sales quantity for specific products)
- It has 2 foreign keys