Dimensional Data Model ERD

For building the dimensional model [Attached as ProjectDimensionalModel.png], I decided to use the snowflake schema with 7 dimension tables and 3 fact tables. Initially, I created a star schema and then decided to use a snowflake schema by keeping the Location as a separate dimension table. Keeping the location details of store, reseller and customer in a separate table reduces the space consumed by the dimension tables and avoids data redundancy as it stores a particular location only once even if a Store, Reseller or Customer belong to the same location as another.

The prefix Dim is used to indicate Dimension tables and Fact is used to indicate Fact tables. Next to each column, I have included the data type depending on the values it will be containing. For each dimension table, I have created Surrogate keys and they serve as a Primary key for the corresponding table. The source system keys (Natural Keys) for each table are marked as SourceID along with the column name to distinguish them easily.

The cardinality between dimension and fact tables is represented as One to Many as each dimension has many facts associated with it but every fact will be associated to only one row in a dimension. For example, the Sales fact table will have only one date associated with each transaction but for each date, there can be multiple sales. The fact tables are defined at a date granularity.

Dimension Tables	Fact Tables
1. Dim_Date 2. Dim_Channel 3. Dim_Product 4. Dim_Customer 5. Dim_Reseller 6. Dim_Store 7. Dim_Location	1. Fact_Sales 2. Fact_StoreSalesTarget 3. Fact_ProductSalesTarget

Dimension Tables

1. Dim_Date

Surrogate Key & Primary Key	Dim_Date_Key [Data type: INTEGER]
Natural Key	FullDate [Data type: DATE]
Grain	Individual date

This is the date dimension table that stores the date of each transaction. In addition to having the complete date (FullDate) the table stores the part of each date as a day of the week, day of the month, month, and year to make querying and retrieval easy. The table is connected to Fact_Sales, Fact_StoreSalesTarget, and Fact_ProductSalesTarget with a cardinality of One to Many as each date might have multiple transactions associated with it in each of the fact tables.

2. Dim_Channel

Surrogate Key & Primary Key	Dim_Channel_Key [Data type: INTEGER]
Natural Key	● ChannelID [Data type: INTEGER] ● ChannelCategoryID [Data type: INTEGER]
Grain	Individual Channel

This is the channel dimension table that stores the different channels and their categories. The table has two natural keys Channel ID and Channel category ID which exist in the provided source system. Each row in the table will represent an individual channel with the corresponding category to which it belongs. The table is connected to Fact_Sales and Fact_StoreSalesTarget with a cardinality of One to Many as channel detail needs to be considered while calculating the sales transactions for a day.

3. Dim_Product

Surrogate Key & Primary Key	Dim_Product_Key [Data type: INTEGER]
Natural Key	● ProductID [Data type: INTEGER] ● ProductTypeID [Data type: INTEGER]

	• ProductCategoryID [Data type: INTEGER]
Grain	Individual Product

This is the product dimension table that stores the different products along with their type and category. Each row in the table will have details of an individual product along with its type and category details. It also includes the cost, wholesale, and retail price of each product. The table has three natural keys Product ID, Product Type ID, and Product category ID which helps to connect the data from three of the data tables provided [Product, ProductType, ProductCategory]. The table is connected to Fact_Sales and Fact_ProductSalesTarget with a cardinality of One to Many as each product detail needs to be considered while calculating the sales transactions for a day and to see whether each product has met its expected daily target.

4. Dim_Customer

Surrogate Key & Primary Key	Dim_Customer_Key [Data type: INTEGER]
Natural Key	CustomerID [Data type: INTEGER]
Grain	Individual Customer

This is the customer dimension table where each row contains the details like first name, last name, gender, and phone number of an individual customer. Customer Id is the source system key. The table is connected to the Fact_Sales table with a cardinality of One to Many as each customer can be associated with multiple sales transactions for a day but each sale can have only one customer associated with it. The customer table is connected to a separate location dimension table with a cardinality of Many to One as each customer can have only one location but multiple customers can be residing in the same location.

5. Dim_Reseller

Surrogate Key & Primary Key	Dim_Reseller_Key [Data type: INTEGER]
Natural Key	ResellerID [Data type: INTEGER]

Grain	Individual Reseller

This is the reseller dimension table where each row contains the details like name, contact and phone number of an individual reseller. Reseller Id is the natural key. The table is connected to Fact_Sales and Fact_StoreSalesTarget table with a cardinality of One to Many as each reseller can be associated with multiple sales transactions for a day but each sale can have only one reseller associated with it. The reseller table is connected to a separate location dimension table with a cardinality of Many to One as each reseller can have only one location, but multiple resellers can be residing in the same location.

6. Dim_Store

Surrogate Key & Primary Key	Dim_Store_Key [Data type: INTEGER]
Natural Key	StoreID [Data type: INTEGER]
Grain	Individual store

This is the store dimension table where each row contains the store number, store manager name and phone number of an individual store. Store Id is the natural key which uniquely identifies a store in the source system. The table is connected to Fact_Sales and Fact_StoreSalesTarget table with a cardinality of One to Many as each store can be associated with multiple sales transactions for a day but each sale can have only one store associated with it. The store table is connected to a separate location dimension table with a cardinality of Many to One as each store can have only one location, but multiple stores can have the same location.

7. Dim_Location

Surrogate Key & Primary Key	Dim_Location_Key [Data type: INTEGER]
Natural Key	PostalCode [Data type: INTEGER]
Grain	Postal Code

This is the location dimension table that includes the location details from Reseller, Channel, and Customer tables. Keeping the location details in a separate dimension table helps to reduce data redundancy and save space in each dimension table. PostalCode is the natural key that uniquely identifies each row in the source system. The table is connected to Fact_Sales table with a cardinality of One to Many as each location can be associated with multiple sales transactions for a day (depending on a store, reseller or customer) but each sale can have only one location associated with it. The location table is connected to store, reseller and customer table with a cardinality of One to Many as each location can have multiple stores, resellers or customers associated to it, but each store, reseller or customer can belong to only one location.

Fact Tables

1. Fact_Sales

Foreign Keys [All Foreign keys are Integers]	● Dim_Date_Key ● Dim_Store_Key ● Dim_Channel_Key ● Dim_Product_Key ● Dim_Location_Key ● Dim_Reseller_Key ● Dim_Customer_Key
Natural Key [All Natural keys are Integers]	● SalesHeaderID ● SalesDetailID
Grain	Sale for each date
Metric	SalesProfit [Data type: Float]

This fact table includes sales details like amount, quantity and profit for each date at an individual store, channel, product, reseller or customer level. As fact tables are used to capture measurements or metrics, I included a calculated field SalesProfit with a float data type. This will help to store the profit obtained for each sales transaction and avoids computing it each time during querying. The SalesProfit will be an additive field as it will have value corresponding to each day and can be summed to determine yearly profit.

SalesProfit will be calculated as (quantity * price) − (quantity × cost), where the required values for the computation will be obtained from the corresponding dimension tables.

The foreign keys are used to connect the fact table to the required dimension tables based on the details needed for representing sale for each day. The fact table has a cardinality of Many to One to the associated dimension tables as each sale transaction in the fact table can be associated to only a single row in a dimension table but each row in a dimension table can be associated with multiple sales transactions in the fact table.

2. Fact_StoreSalesTarget

Foreign Keys [All Foreign keys are Integers]	● Dim_Date_Key ● Dim_Store_Key ● Dim_Channel_Key ● Dim_Reseller_Key
Grain	Sales target for each date

This fact table includes the expected sales target for each date. The provided data included only the yearly target. To convert it into daily target values the given yearly target needs to be divided by 365. The foreign keys connect the target details to corresponding stores for each date. The fact table has a cardinality of Many to One to the associated dimension tables as each daily target in the fact table can be associated to only a single store for each date, but each store can be associated with different sales target for different dates.

3. Fact_ProductSalesTarget

Foreign Keys [All Foreign keys are Integers]	● Dim_Date_Key ● Dim_Product_Key
Grain	Daily individual product sales quantity target

This fact table stores the expected target sales quantity for each product for a date. The foreign keys connect the fact table to the date and product dimension tables that correspond to the target value. As the provided target values were yearly, they need to

be divided by 365 to calculate the daily product sales quantity target. The fact table has a cardinality of Many to One to the associated dimension tables as each daily target in the fact table can be associated to only a single product for each date, but each product can be associated with different sales target for different dates.