# Cosmetics store

## CONTENTS

Business Description 3
Business background 3
Problems because of poor data management 3
Benefits from implementing a Data Warehouse 3
Dimensions of a Business 3
Logical Scheme 7
Data Flow 8

**Business Description**

**Business background**

Cosmetics are something that every person uses. There are many shops that offer different types of cosmetics for different purposes. This type of business has a large number of competitors, so in order to achieve success, it is necessary to approach this matter responsibly and study many factors that affect the choice of cosmetics. First of all, this can be done by collecting information about product sales and analyzing it with the help of special tools.

**Problems because of poor data management**

Poor data management can affect incorrect decision-making. To maintain competitiveness, it is necessary to use tools that can provide you with information for analyzing and developing a business management strategy

**Benefits from implementing a Data Warehouse**

Using a data warehouse can help you solve the problems described above. Implementing a data warehouse can answer the following questions for you:

Which producers have the highest prices?

Which of them have the widest price distribution?

Is there a typical price distribution (for example, normal) between brands or within specific brands? Is there a normal distribution by product prices?

Data processing will also allow you to:

correlate specific product characteristics with price changes.

identify brand preferences by country

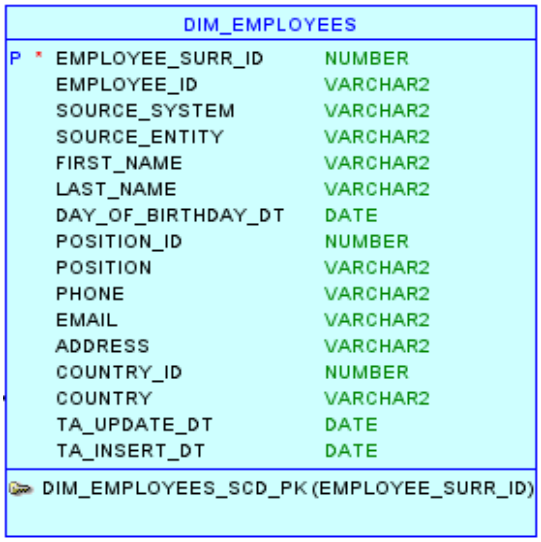
And much more.

**Dimensions of a Business**

Star schema of the model comprises 7 dimensional tables:

Dim_dates - includes dates for better querying of business reports.


Dim_employees - includes a list of employees of the cosmetics stores




3.Dim_customers - includes a list of customers of the cosmetics stores

| DIM_CUSTOMERS | | |
|---|---|---|
| P * | CUSTOMER_SURR_ID | NUMBER |
| | CUSTOMER_ID | VARCHAR2 |
| | SOURCE_SYSTEM | VARCHAR2 |
| | SOURCE_ENTITY | VARCHAR2 |
| | FIRST_NAME | VARCHAR2 |
| | LAST_NAME | VARCHAR2 |
| | COMPANY_NAME | VARCHAR2 |
| | PHONE | VARCHAR2 |
| | COMPANY_NUMBER | VARCHAR2 |
| | GENDER | VARCHAR2 |
| | YEAR_OF_BIRTH | NUMBER |
| | EMAIL | VARCHAR2 |
| | COUNTRY | VARCHAR2 |
| | COUNTRY_ID | NUMBER |
| | ADDRESS | VARCHAR2 |
| | TA_UPDATE_DT | DATE |
| | TA_INSERT_DT | DATE |
| DIM_CUSTOMERS_PK (CUSTOMER_SURR_ID) | | |

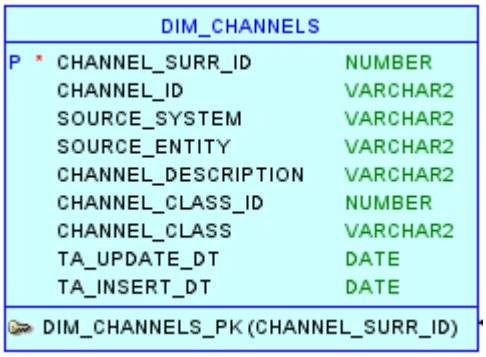
4. Dim_stores - includes a list of the cosmetics stores

| DIM_STORES | | |
|---|---|---|
| P * | STORE_SURR | NUMBER |
| | STORE_ID | VARCHAR2 |
| | SOURCE_SYSTEM | VARCHAR2 |
| | SOURCE_ENTITY | VARCHAR2 |
| | PHONE | VARCHAR2 |
| | STREET_ADDRESS | VARCHAR2 |
| | POSTAL_CODE | VARCHAR2 |
| | CITY_ID | NUMBER |
| | CITY | VARCHAR2 |
| | STATE_ID | NUMBER |
| | STATE | VARCHAR2 |
| | COUNTRY_ID | NUMBER |
| | COUNTRY | VARCHAR2 |
| | TA_UPDATE_DT | DATE |
| | TA_INSERT_DT | DATE |
| DIM_STORES_PK (STORE_SURR) | | |

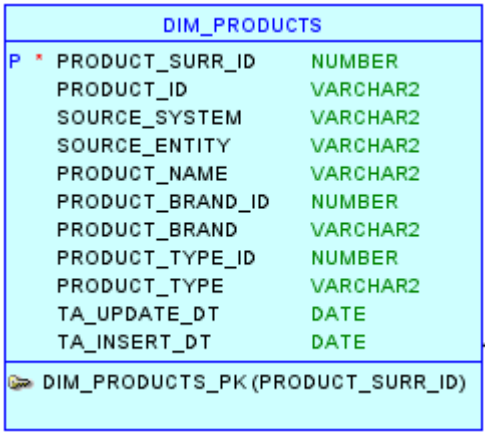
5. Dim_promotions_scd - includes a list of the promotions

| DIM_PROMOTIONS_SCD | | |
|---|---|---|
| P * | PROMOTION_SURR_ID | NUMBER |
| | PROMOTION_ID | VARCHAR2 |
| | SOURCE_SYSTEM | VARCHAR2 |
| | SOURCE_ENTITY | VARCHAR2 |
| | PROMOTION_NAME | VARCHAR2 |
| | PROMOTION_CATEGORY_ID | NUMBER |
| | PROMOTION_CATEGOTY | VARCHAR2 |
| | BEGIN_SOP_DT | DATE |
| | END_EOP_DT | DATE |
| | IS_ACTIVE | VARCHAR2 |
| | COST | NUMBER (15,2) |
| | TA_UPDATE_DT | DATE |
| | TA_INSERT_DT | DATE |
| DIM_PROMOTIONS_SCD_PK (PROMOTION_SURR_ID) | | |

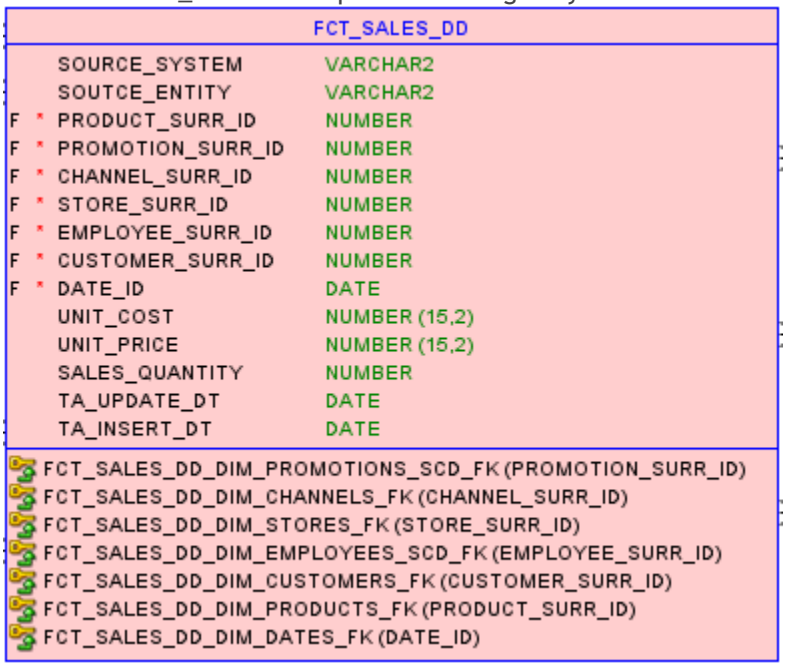
6. Dim_channels - includes a list of sales channels




7. Dim_products - includes a list of products of cosmetics stores




Fact table FCT_SALES - comprises of foreign keys from dimensional tables and additional measurements.



4-step 3D design process:

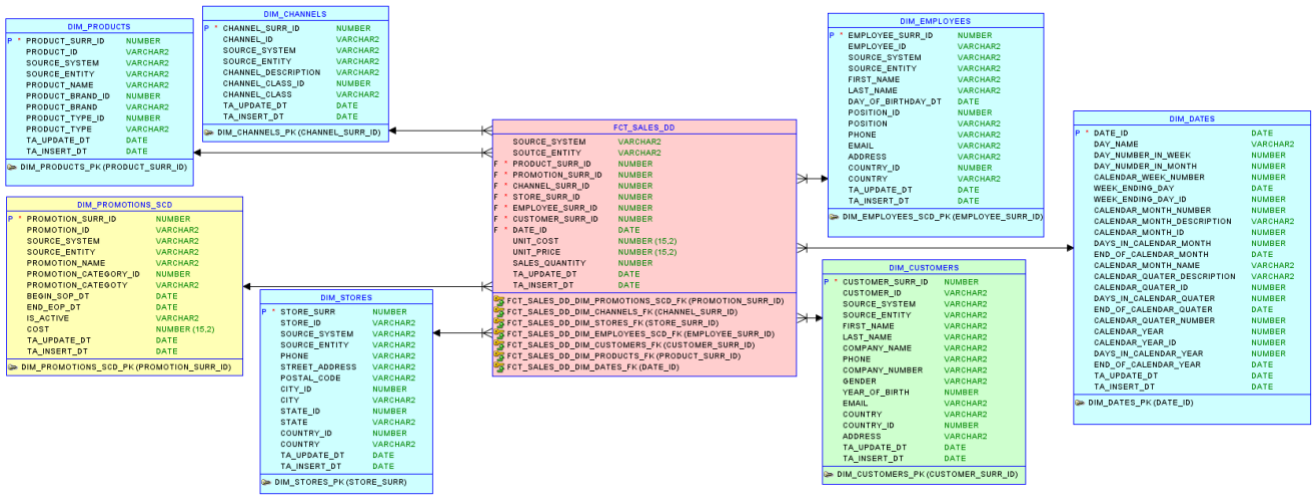
Select the business process. The process of business model is retail sales. Business process - sales products

Declaration the grain. In this scheme, the grain can be a sale of one product.

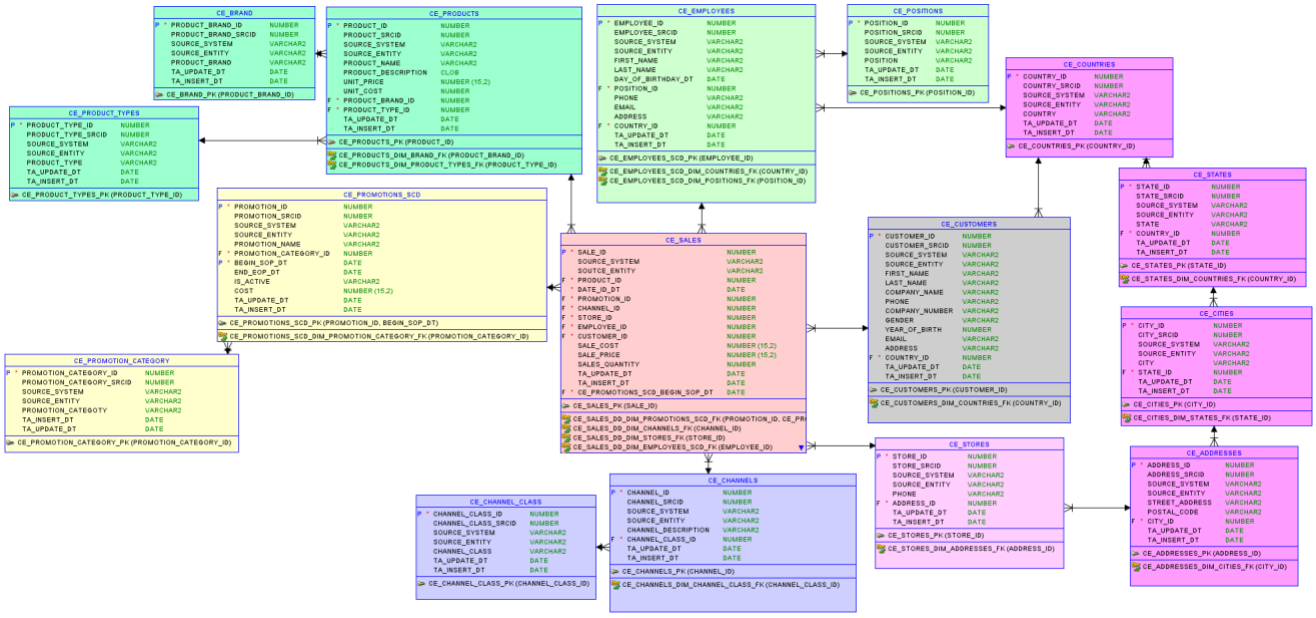
Identify the dimensions. The dimensions are product, customer, employee, store, dates, promotions,channels.

Identify the facts. The sale fact can be determined by cost, price and quantity attributes.

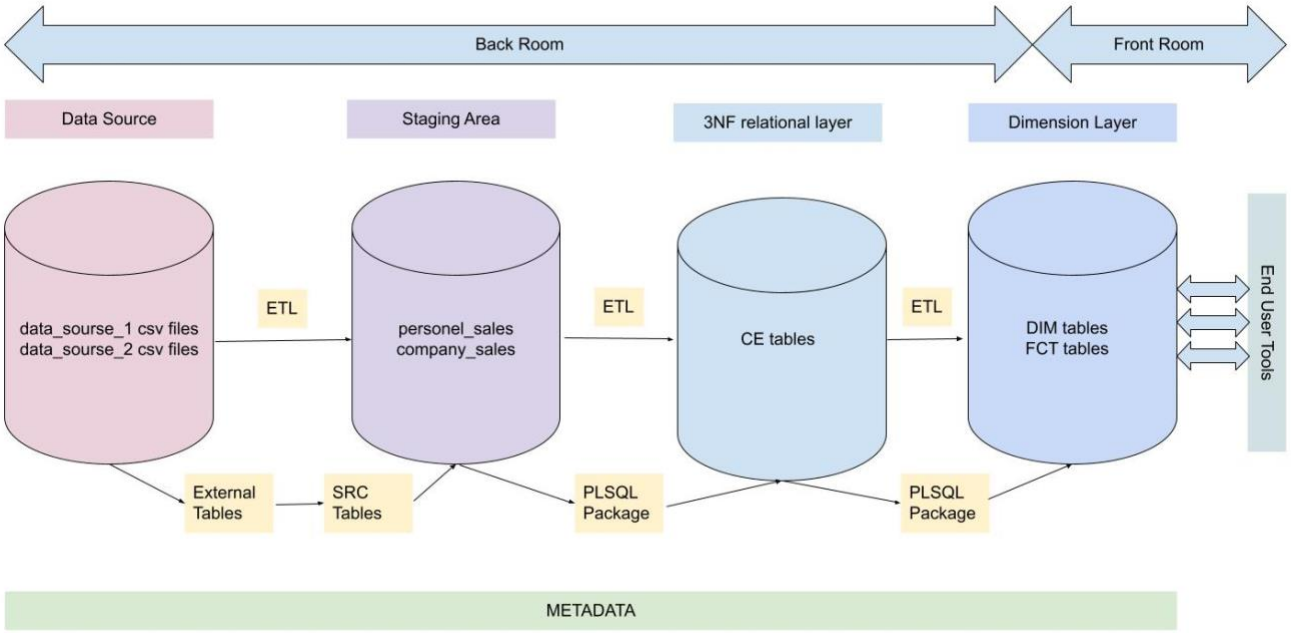
**Star Schema Model:**




**3NF Model:**



**Logical Scheme**




The logical model have five layers:

1. **Data source** - stores data that we periodically receive from our sources. In our warehouse, we access these tables using external tables.

2. **Staging area** - stores materialized data from sources.

3. **3NF relational layer** - stores data in the Third Normal Form.

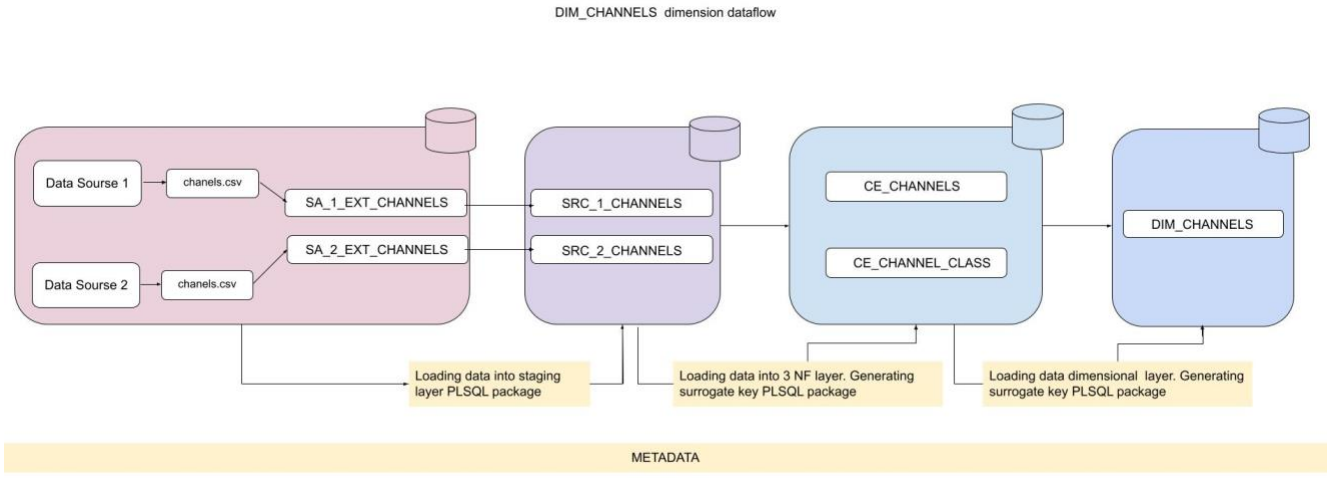
4. **Dimension layer** - stores denormalized data in the star schema.

5. **Data presentation layer** - the layer which our business users and business departments can use to retrieve information.

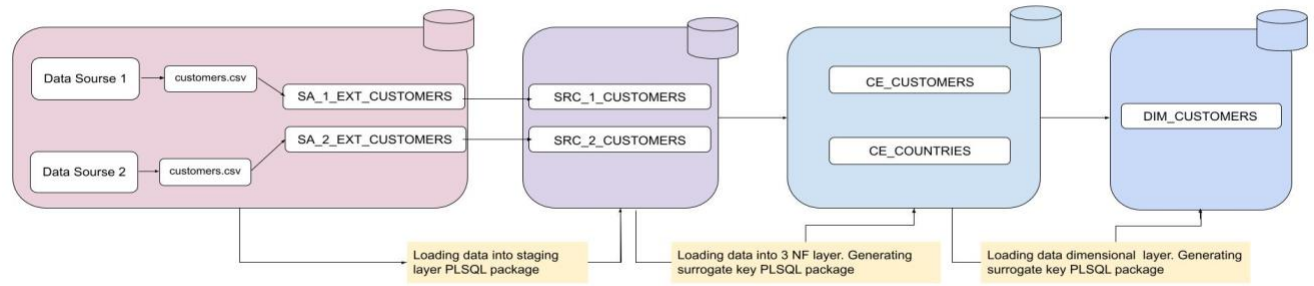
**Data Flow**

Source datasets contain information about COSMETICS sales. The first source has information about personal sales, the second - about sales to companies.

For customers and channels loading we use both datasets. For products,promotions,stores,employees loading we use only personnel_sales dataset as default because these dimensions have the same data in both datasets.

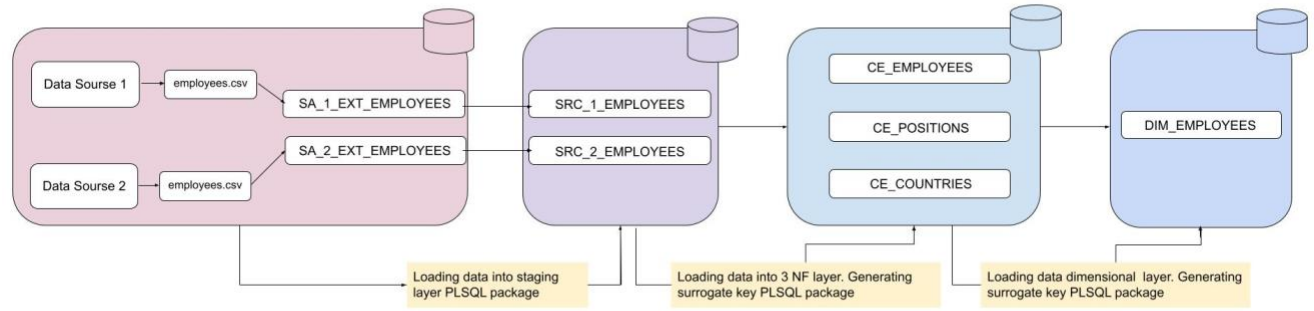


DIM_CUSTOMERS dimension dataflow



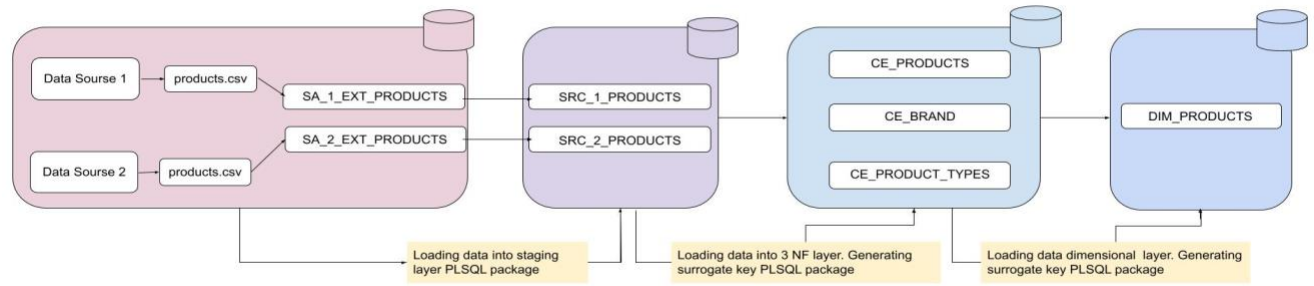

METADATA

DIM_EMPLOYEES dimension dataflow



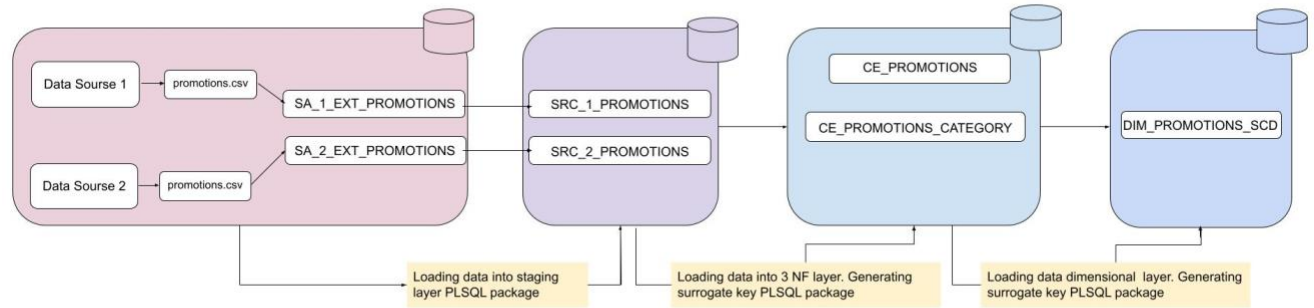

METADATA

DIM_PRODUCTS dimension dataflow



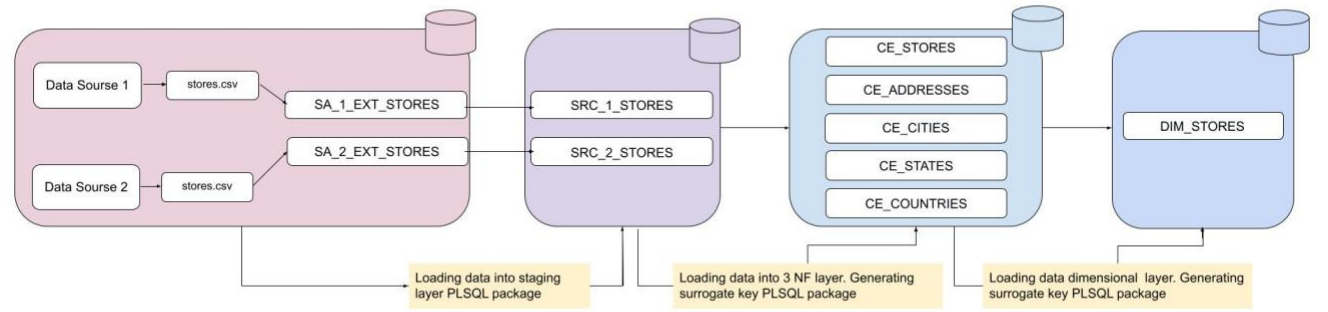

METADATA

DIM_PROMOTIONS_SCD dimension dataflow



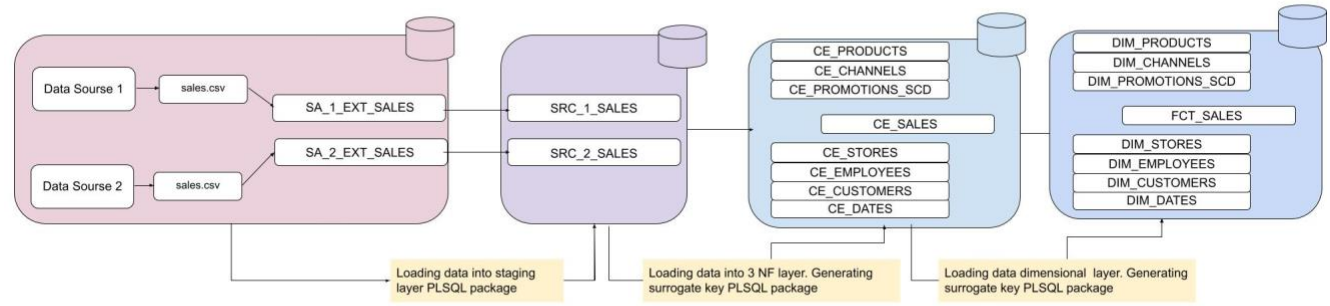

METADATA

DIM_STORES dimension dataflow




METADATA

FCT_SALES fact dataflow




METADATA

## FACT TABLE PARTITIONING STRATEGY

### Range Partition by time

In this project, I use a time range (by months) divided into equal segments. In this partitioning strategy, the fact table is partitioned based on a time period. Here, each time period represents a significant period of retention in business. Span partitioning provides an easy way to automatically create span partitions as data arrives.

I need to use partitioning strategy for the following reasons:

1. For Easy Management.
2. To Assist Backup/Recovery.
3. To Enhance Performance