



## **Business Template**

# **" OFFICE GOODS " SALES**

---



# **Office Goods**

---

**Legal Notice:**

This document contains privileged and/or confidential information and may not be disclosed, distributed or reproduced without the prior written permission of EPAM®.

Confidential

CONTENTS

1 BUSINESS DESCRIPTION .....3

1.1 Business background .....3

1.2 Problems because of poor data management .....3

1.3 Benefits from implementing a Data Warehouse .....3

2 DIMENSIONS OF A BUSINESS.....3

3 LOGICAL SCHEME.....6

4 DATA FLOW .....6

5 FACT TABLE PARTITIONING STRATEGY ..... 10

# 1 BUSINESS DESCRIPTION

## 1.1 BUSINESS BACKGROUND

Nowadays a lot of people use different office equipment and furniture. Also during last 2 years many people switched their business office to home office, which led to additional demand on specific kind of goods. That is why it is very important to analyze data and update offer according to customers' needs.

## 1.2 PROBLEMS BECAUSE OF POOR DATA MANAGEMENT

Poor data management leads to mistakes in arranging the assortments, which should be adjusted taking into account nowadays tendencies and customer demand.

## 1.3 BENEFITS FROM IMPLEMENTING A DATA WAREHOUSE

Using of data warehouse can help to solve the problem with updating assortment according customers' needs.

Implementing a data warehouse can answer the following questions:

- What is the difference in demand between ordinary consumers and corporate and home office segments of clients?
- Which product category gives the highest profit?
- Which products are most popular in each segment?
- Which products have biggest increase in sales during last year?
- And many other.

# 2 DIMENSIONS OF A BUSINESS

4-step process:

- This time we have selected next **business process: sales**. Analyze of sales will help us to understand the demand of our clients, how should we update the assortment in order to widespread our customers' base.
- Taking into account our business model and selected business process, the **grain** should be **one sale**. It will represent a single fact table row, because 1 specific sale for one specific customer.
- For this business model and business process we have identified next **dimensions**: suppliers, customers, shipments, dates, markets, products, employees, loyal programs, orders. These dimensions will give us a good context, which will describe the facts.

Dim **products**: this dimension includes all information about our assortment: which categories of products we offer, these categories consist of which groups, which products are in each group.

This dimension includes hierarchy: product - group - category (for ex.:

Bevis Wood Table, Rectangular - tables - furniture).

Also there is information about product status (active/discontinued) which will show, whether this product is active for ordering or can't be ordered.

Dim **suppliers**: this dimension includes all information about our suppliers: name, geographical location, status (active/blocked) which will show, whether this supplier is active, and we can order goods from him.

Dim **customers**: this dimension includes all information about our customers, such as: name, age, gender, contact information, location.

Dim **employees**: this dimension includes all information about our employees: name, date of birth, contact information, position.

Dim **orders**: this dimension includes information about order priority (for ex. high, low, etc.).

**Dim shipments:** this dimension includes such info as shipment mode (for ex. standart class, first class), shipment type (car, plain, etc.)

**Dim markets:** this dimension includes info about our markets. Here we have hierarchy: address - city- country- region (for ex.: 7575 Main Street - Groningen - Netherlands - EU).

**Dim loyal programs:** this dimension includes info about possible discounts which we could offer to our loyal clients.

- For this business model and business process we have identified next **measurements:** sales amount, quantity, discount, shipping cost, profit.

**Sales amount:** the sum, which client has paid by exact order;

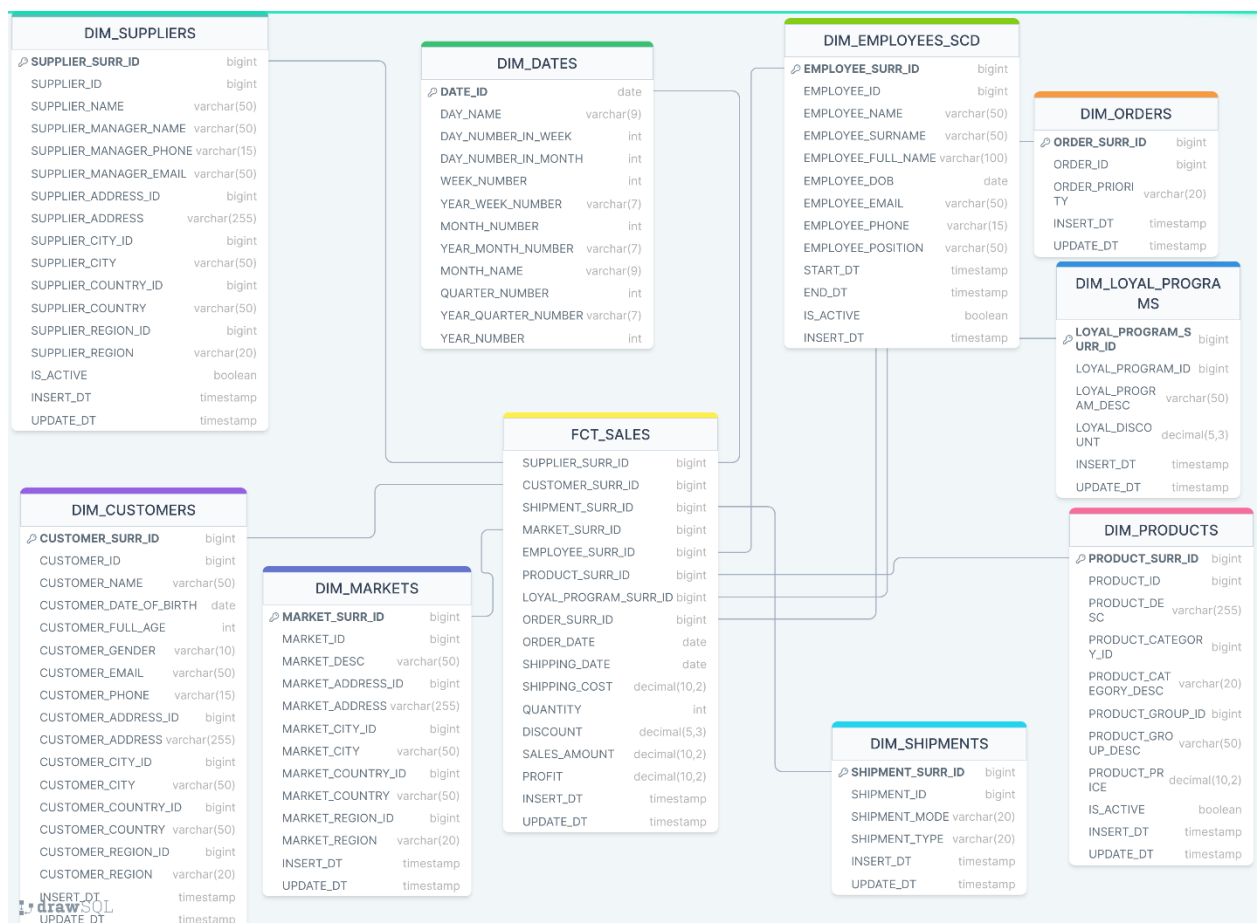
**Quantity** - how many pcs of product have been sold in exact order;

**Discount** - size of discount, which has been granted for exact order;

**Shipping cost** - expenses for delivery of exact order;

**Profit** - amount, which has been earned after sale is finished.

### STAR SCHEMA:



On the basis of star schema created 3NF schema, using normalization rules.

Created separate entity **CE\_addresses**, where will be stored info about addresses from customers (entity **ce\_customers**), suppliers (entity **ce\_suppliers**) and markets (entity **ce\_markets**). Here we have geographical hierarchy: address - city - country - region.

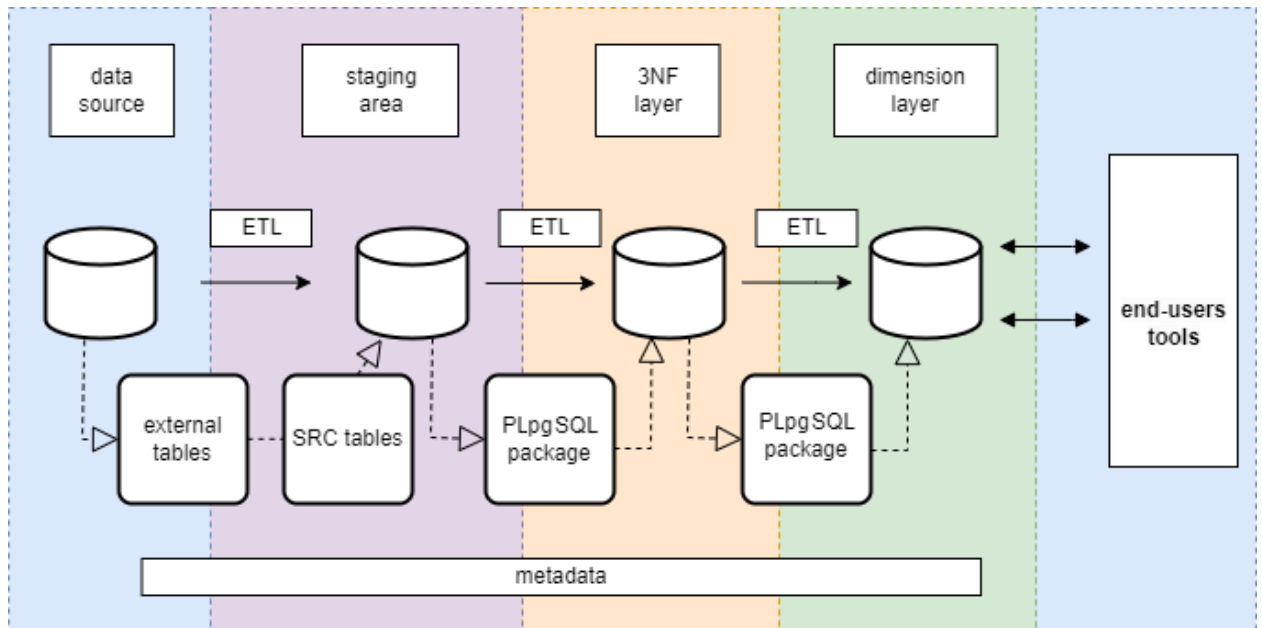
From entity **ce\_suppliers** created separate entity **ce\_supplier\_managers** with info about managers, who work there. Supplier can have several managers, but 1 manager works on 1 supplier.

Entity **ce\_products** was separated into 3 entities (hierarchy): products - groups - categories.



### 3 LOGICAL SCHEME

There is a logical schema, which represents, how DWH is loaded:



Firstly, data exists in data source. Then we create external tables, where all data from source is copied, this is like a link, not materialized table.

After that data from external table is saved in src\_table (materialization 1:1). This is done on staging area. After this step data is safe, even if data source will be deleted, we won't lose data.

Then data is loaded into BL\_3NF layer, where it is normalized (3NF schema).

After that data is loaded into BL\_DM dimension layer, which is denormalized (star schema).

End users have access to this data via end-users tools.

### 4 DATA FLOW

From the beginning till the stage, where data is loaded into 3NF, data flow is the same for all tables.

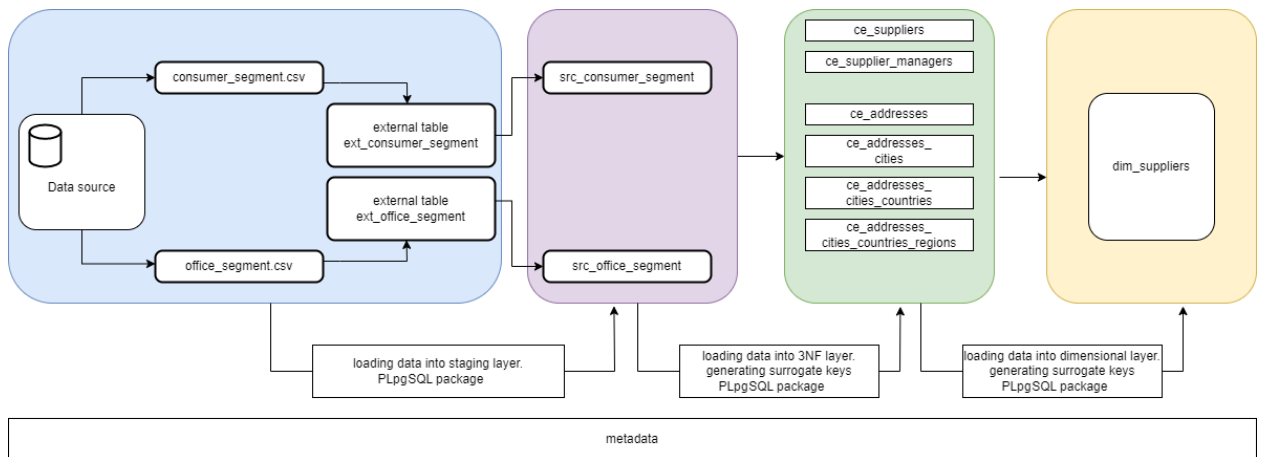
Firstly, from data source it is saved into 2 csv files: consumer\_segment.csv and office\_segment.csv. Then we create external tables ext\_consumer\_segment and ext\_office\_segment with data from csv files, these tables are not materialized.

After that from external tables we create tables src\_consumer\_segment and src\_office\_segment (staging area). From these table we load data into 3NF layer.

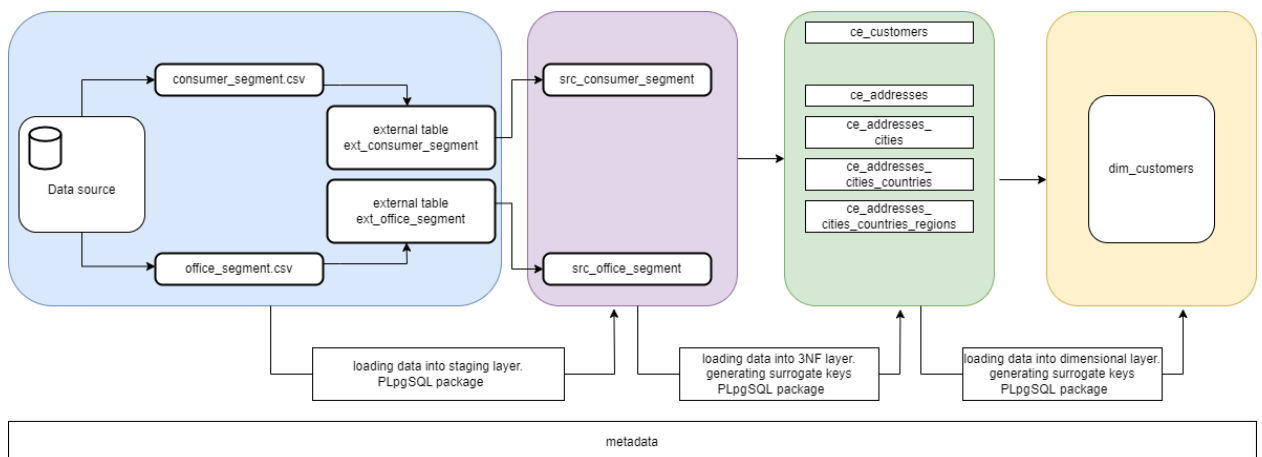
Here we already have differences by dimensions.

According to our star schema, we have 1 fact table (FCT\_SALES) and 8 dimension tables (besides DIM\_DATES, which won't be discussed here):

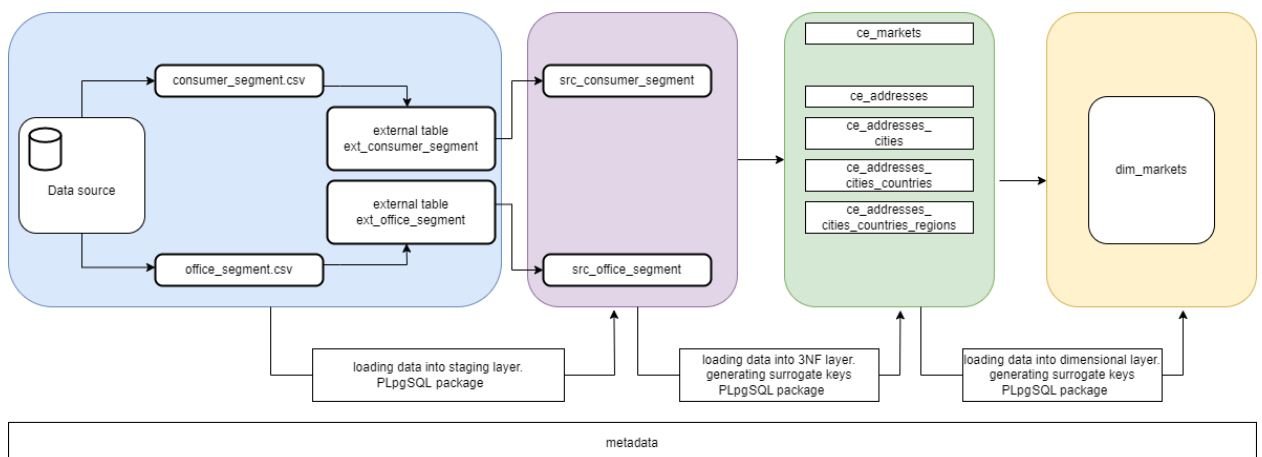
1. DIM\_SUPPLIERS. In 3NF layer info is stored in next entities: ce\_suppliers, ce\_supplier\_managers, ce\_addresses, ce\_addresses\_cities, ce\_addresses\_cities\_countries, ce\_addresses\_cities\_countries\_regions. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_SUPPLIERS.



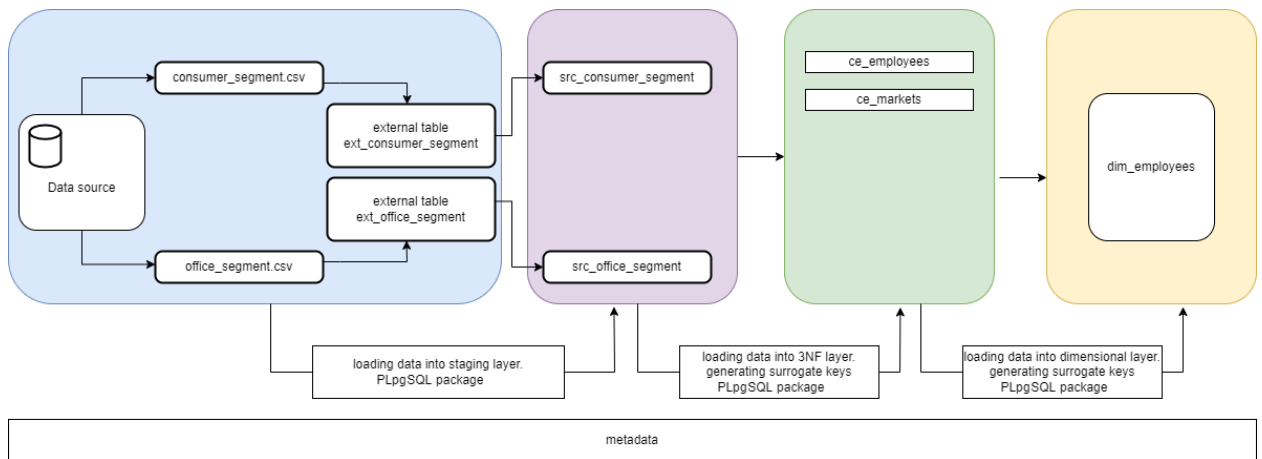
2. DIM\_CUSTOMERS. In 3NF layer info is stored in next entities: ce\_customers, ce\_addresses, ce\_addresses\_cities, ce\_addresses\_cities\_countries, ce\_addresses\_cities\_countries\_regions. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_CUSTOMERS.



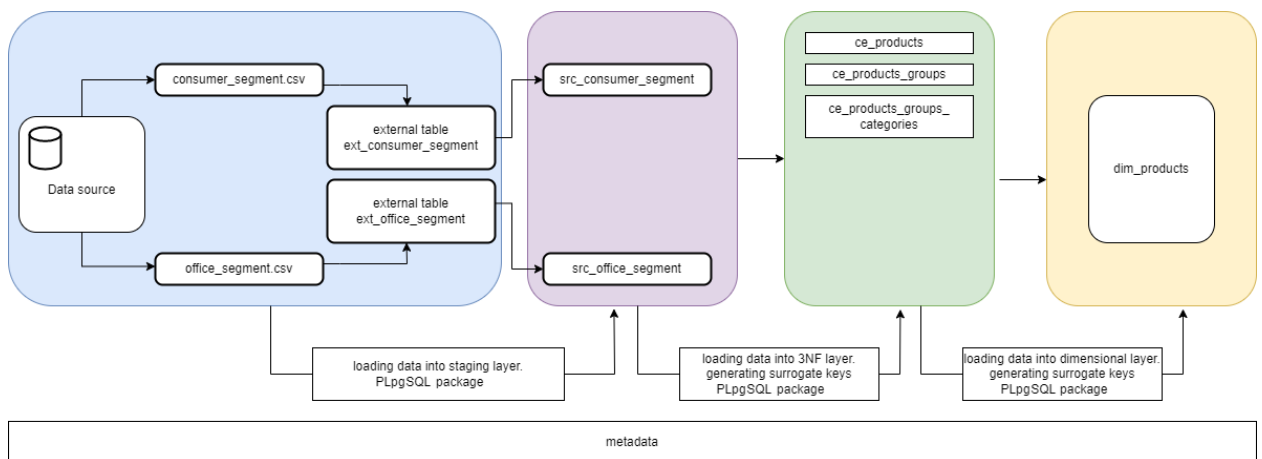
3. DIM\_MARKETS. In 3NF layer info is stored in next entities: ce\_markets, ce\_addresses, ce\_addresses\_cities, ce\_addresses\_cities\_countries, ce\_addresses\_cities\_countries\_regions. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_MARKETS.



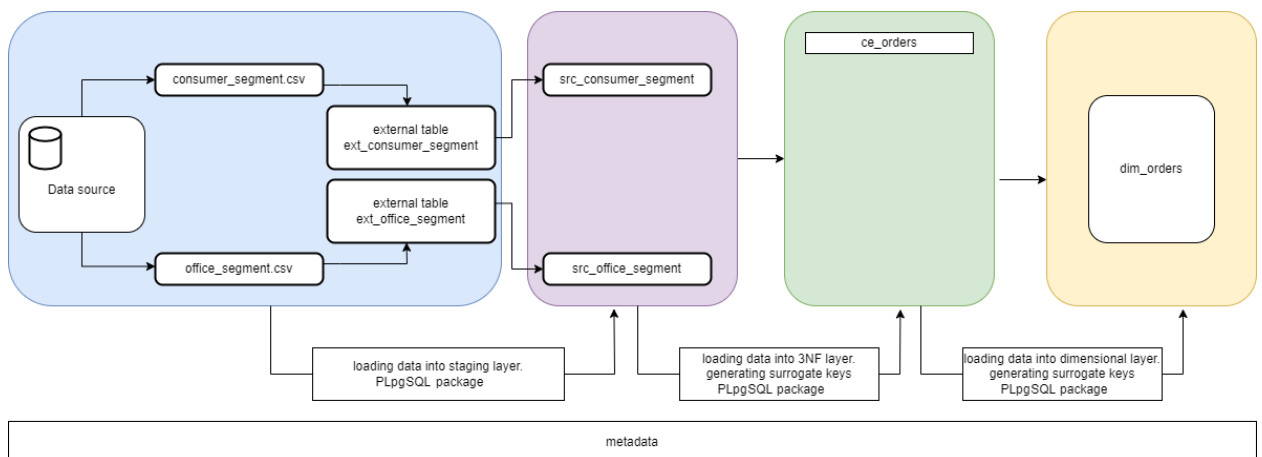
4. DIM\_EMPLOYEES. In 3NF layer info is stored in next entities: ce\_employees, ce\_markets. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_EMPLOYEES.



5. DIM\_PRODUCTS. In 3NF layer info is stored in next entities: ce\_products, ce\_products\_groups, ce\_products\_groups\_categories. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_PRODUCTS.

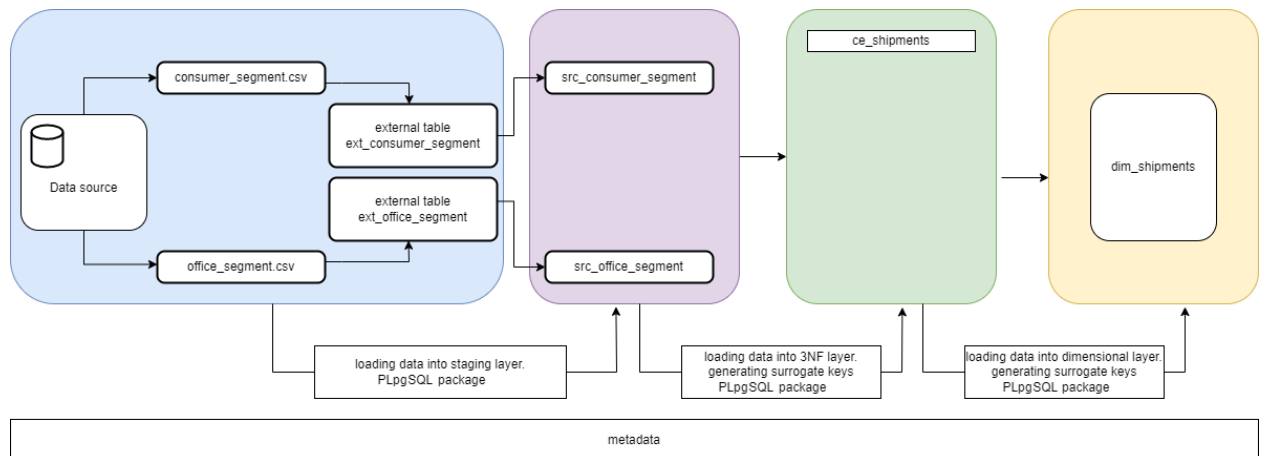


6. DIM\_ORDERS. In 3NF layer info is stored in next entities: ce\_orders. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_ORDERS.

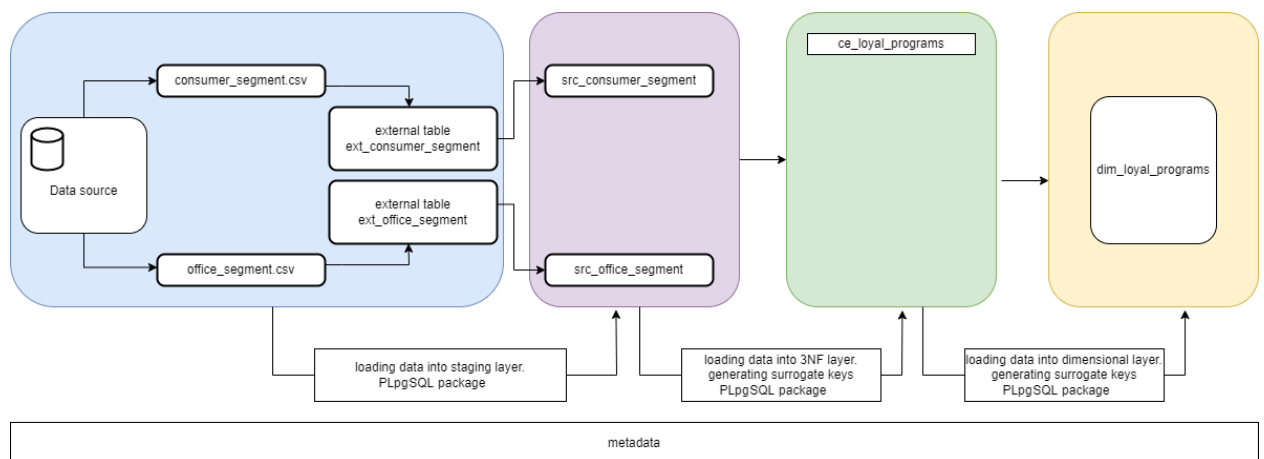




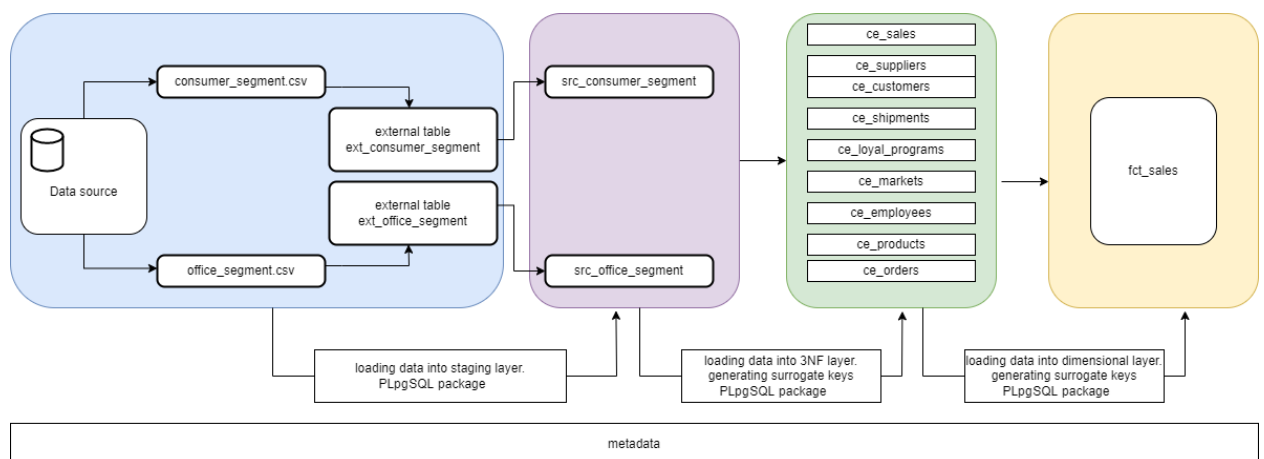
7. DIM\_SHIPMENTS. In 3NF layer info is stored in next entities: ce\_shipments. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_SHIPMENTS.



8. DIM\_LOYAL\_PROGRAMS. In 3NF layer info is stored in next entities: ce\_loyal\_programs. In the next stage we load data from 3NF into dimensional layer, where have 1 dimensional table DIM\_LOYAL\_PROGRAMS.



9. FCT\_SALES. In 3NF layer info is stored in next entities: ce\_sales, ce\_suppliers, ce\_customers, ce\_shipments, ce\_loyal\_programs, ce\_markets, ce\_employees, ce\_products, ce\_orders. In the next stage we load data from 3NF into dimensional layer, where have 1 fact table FCT\_SALES.



## 5 FACT TABLE PARTITIONING STRATEGY

By partitioning the fact table into sets of data (partitions), the query procedures can be enhanced, because query scan only particular partitions, not whole set of data.

In this case we have partition by range (month). The fact table is being refreshed by procedure `BL_CL.MASTER_PROC_LOAD_INC()`, which refresh the current and previous months. According to the business, refresh is planned on the last day of month.