

|  |
| --- |
| Business Template  **BOOK SALES** |
|  |

Contents

[1 Business Description 3](#_Toc68080597)

[1.1 Business background 3](#_Toc68080598)

[1.2 Problems because of poor data management 3](#_Toc68080599)

[1.3 Benefits from implementing a Data Warehouse 3](#_Toc68080600)

[2 Dimensions of a Business 3](#_Toc68080601)

[3 Logical Scheme 3](#_Toc68080602)

[4 Data Flow 3](#_Toc68080603)

# 

# Business Description

## Business background

Books are companion every person during in the life. In current times many people prefer to read electronic books. The Internet allows finding necessary literature don't go out from home. But paper books also demand in this hard competition. So that the business will succeed, you should very responsible approach to analyze different metrics influencing sales. To collect, transform and analyze data you can use special tools.

## Problems because of poor data management

Poor data management doesn’t let to correctly analyze the situation on the sales market. You can't do truth choice for growing the business. If you don’t use special instruments for common analysis business data, you won’t be competitive on market book sales.

## Benefits from implementing a Data Warehouse

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

* Which category of book has the highest sales?
* What distribution sales by region?
* Quantity of new customers by period.
* Profit by Customer.
* Category book ranking.

Further processing data would also let you:

* Analyze places for business extension.
* Analyze time period when increase sales.
* Customer preferences by book category.
* Price dynamics.
* And many other.

# Dimensions of a Business

Design the dimensional model use four steps.

**Step 1. Select business project**

The main target to analyze the business – sales books from different stores and shops.

**Step 2. Declare the Grain**

Analyzing the business process can define the grain entity «sale» as the atomic element for research.

**Step 3. Identify of the dimensions**

Define dimension table:

dim\_store – data about shops and warehouses

dim\_book\_scd – data about product of the company (slowly changing dimensions type 2 – storing history change row from the source system)

dim\_customers – retail and wholesale buyers

dim\_employee - store information about the staff business

dim\_type\_payments – kind of payments

dim\_address – geography

dim\_dates – date values

All dimension tables exclude “dim\_time” generate own unique surrogate primary key and

store link with source system use composite natural key(source\_id + table\_name + system\_name).

**Step 4. Identify the Facts**

Table fact «fct\_sales» store information:

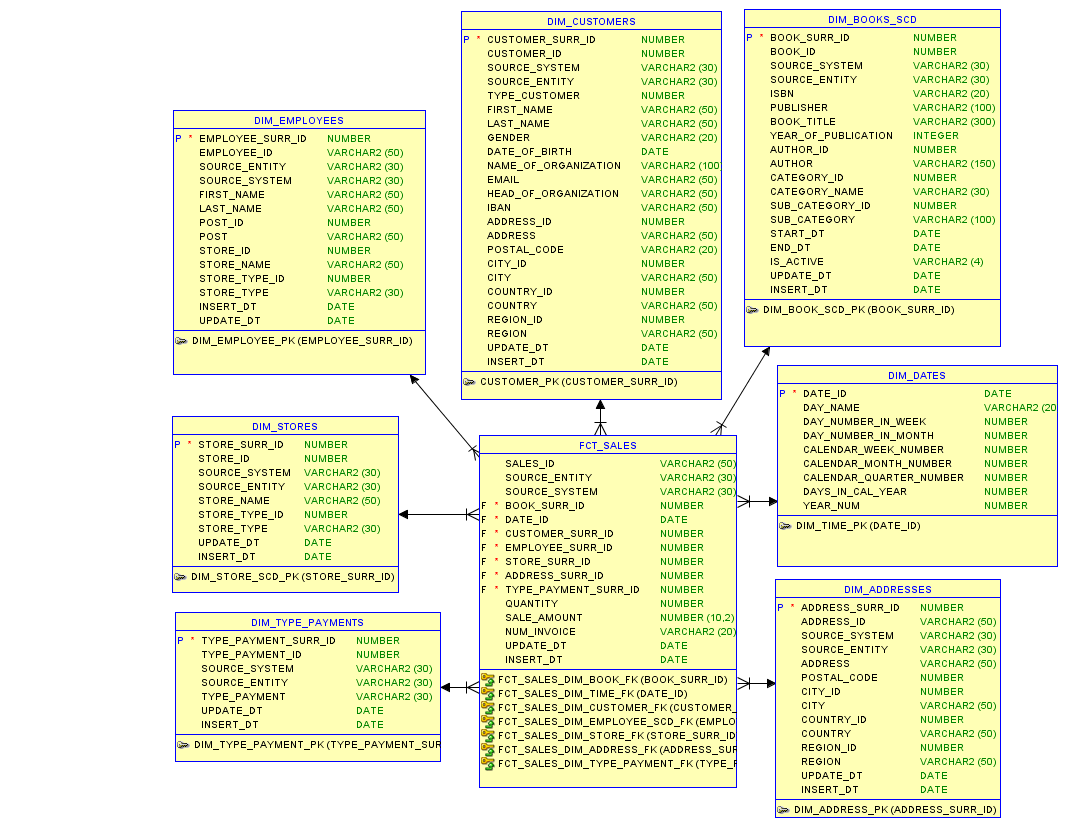
- composite natural key – sales\_id + table\_namу + system\_name

- foreign keys to dimension tables (book\_surr\_id, time\_id, customer\_suur\_id, store\_surr\_id, employee\_surr\_id, address\_surr\_id, type\_payment\_surr\_id)

- additive facts: quantity, sale\_amount

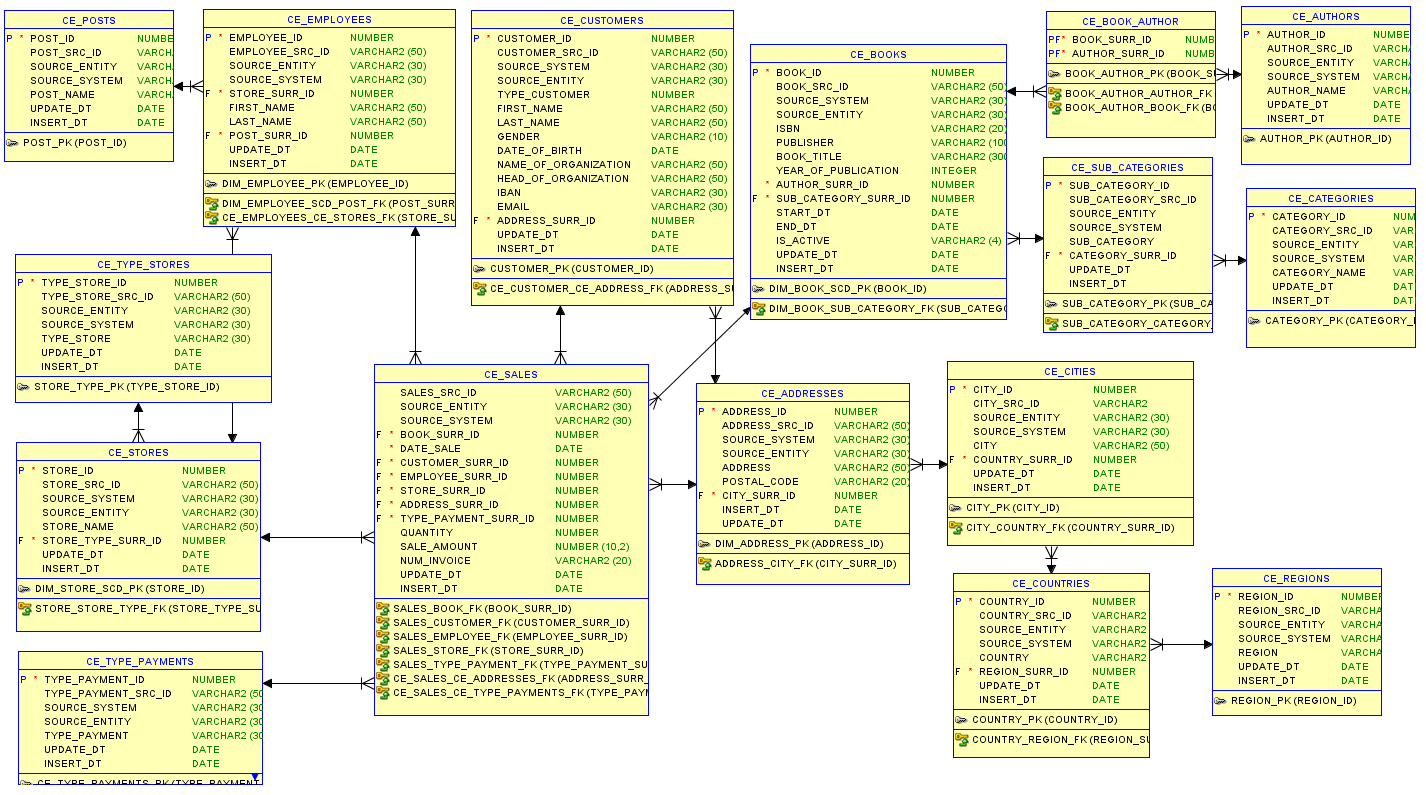
- degenerate field: num\_invoice

**The Star schema model «Book sales»**

****

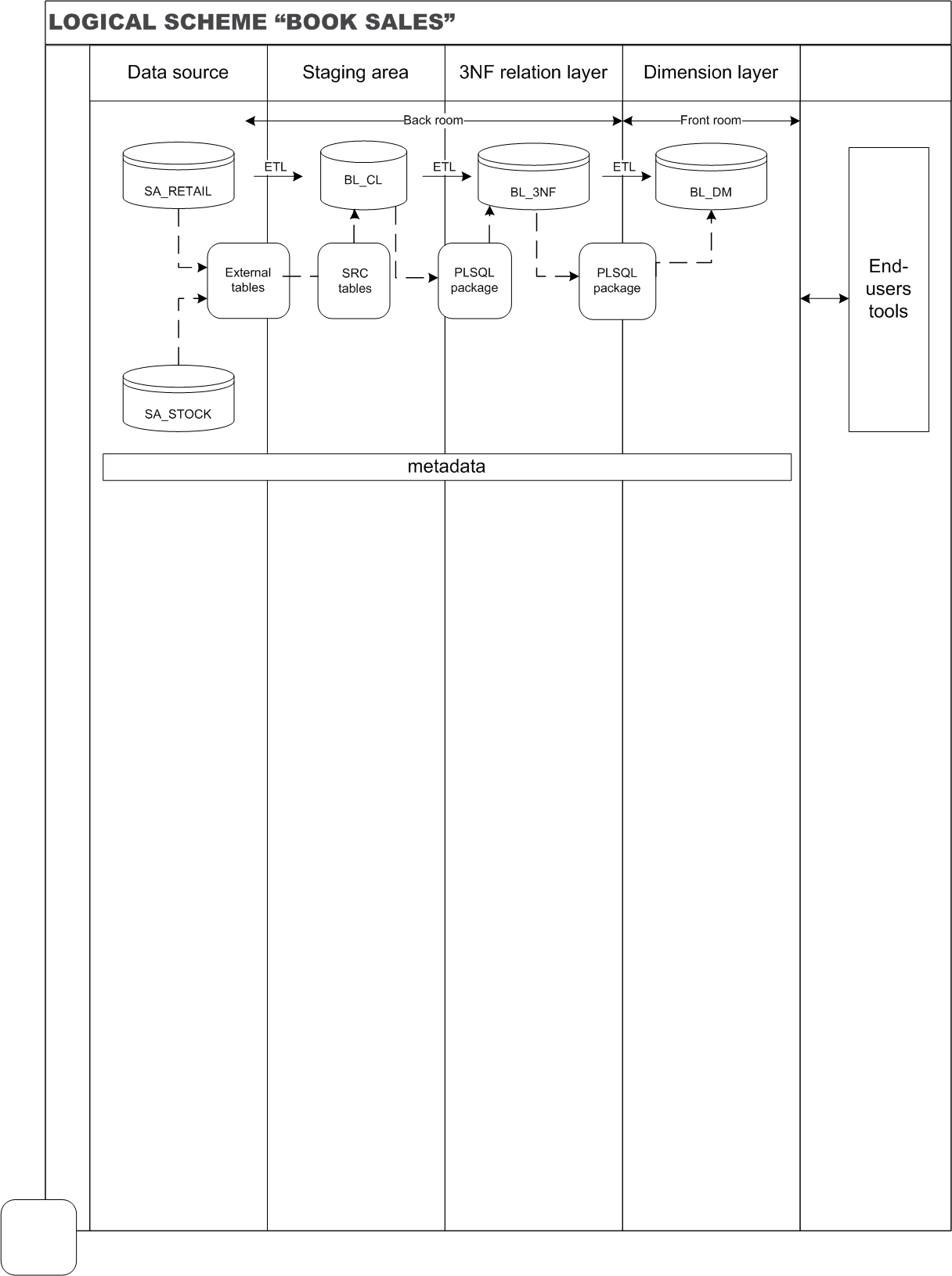
**The 3NF schema model «Book sales»**

For creating many-level architecture DWH also should create a high-level data model in 3NF. Model 3NF creating with description business process and yet created the star model. The dimension tables were divided into some different tables. For all tables exclude “ce\_book\_author” generate own unique surrogate primary key and store link with source systems use composite natural key (source\_id + source\_entity + source\_system).



# Logical Scheme

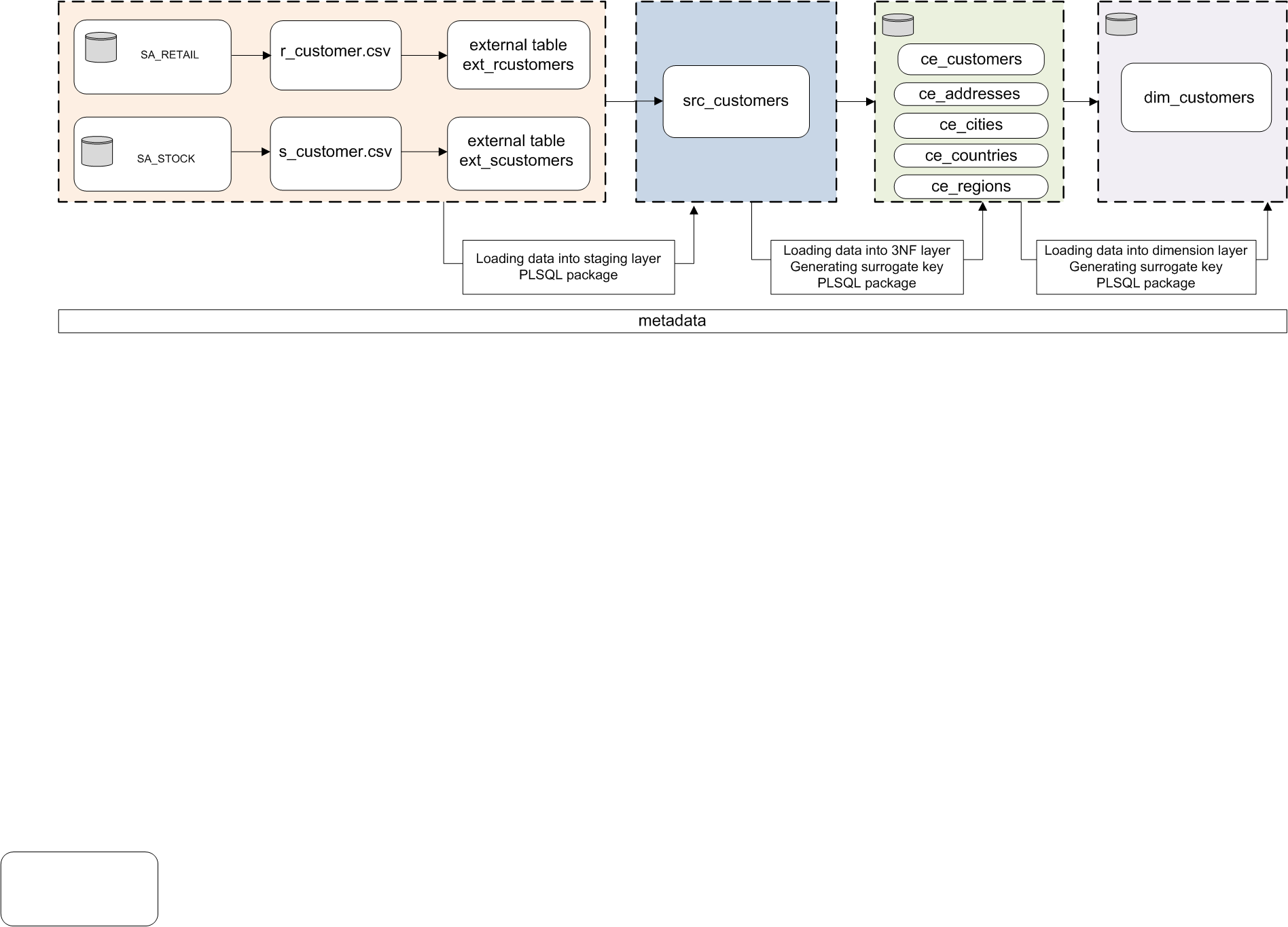
Create logical model of DWH load

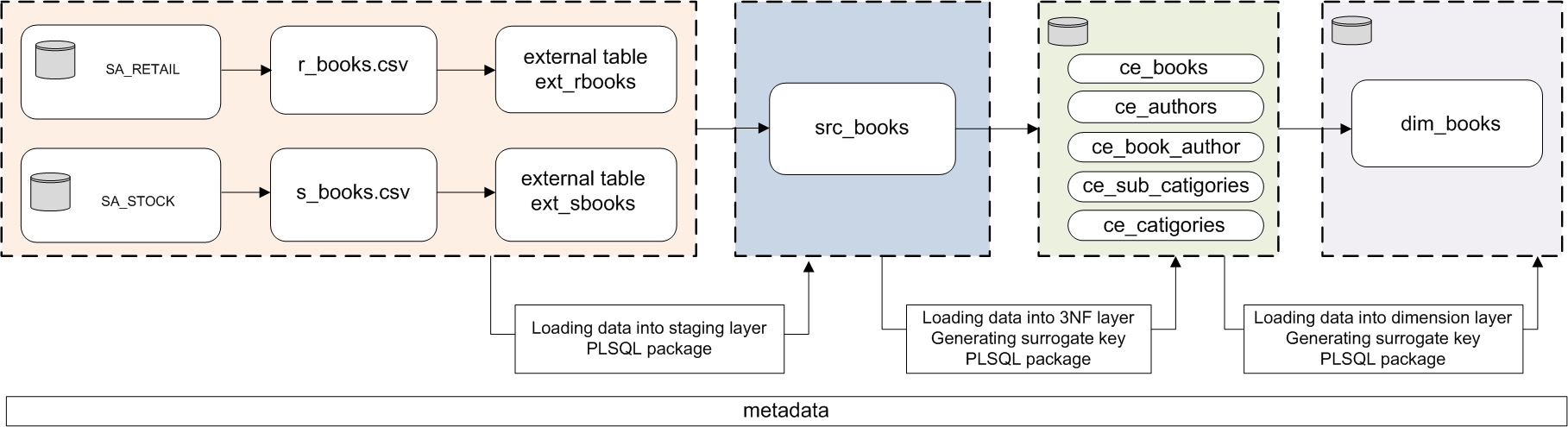


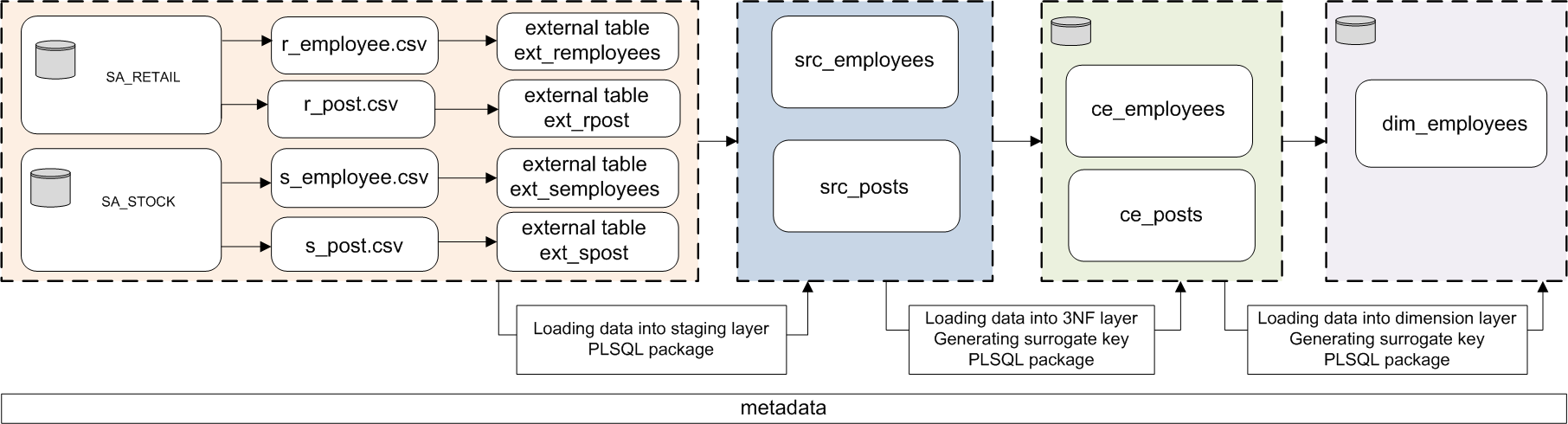
# Data Flow

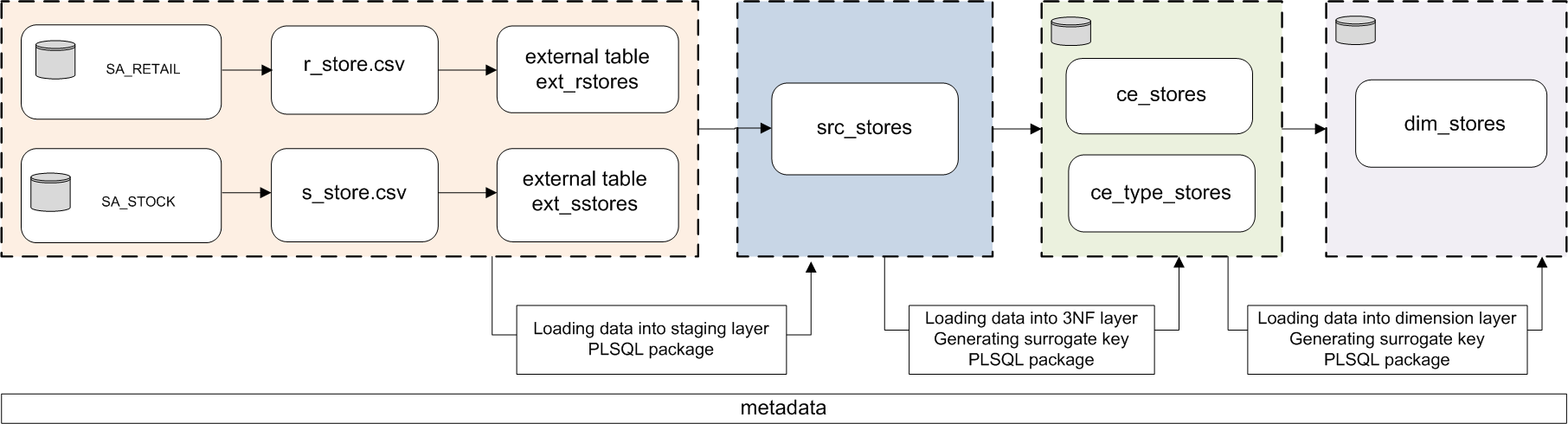
Design data flow diagrams for DWH load

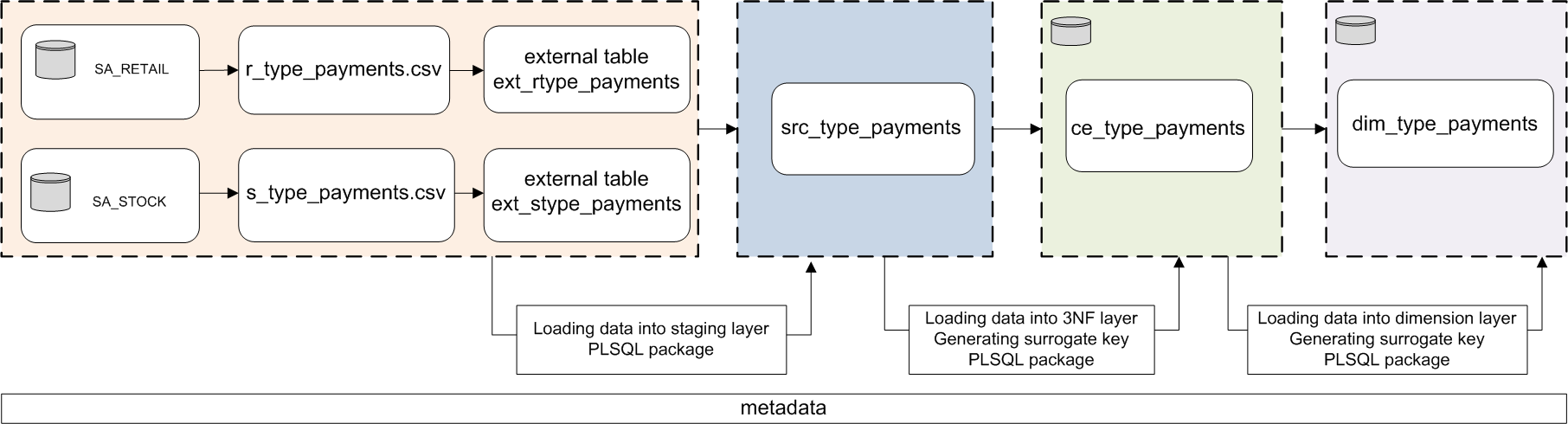
**DIMENSION TABLES**

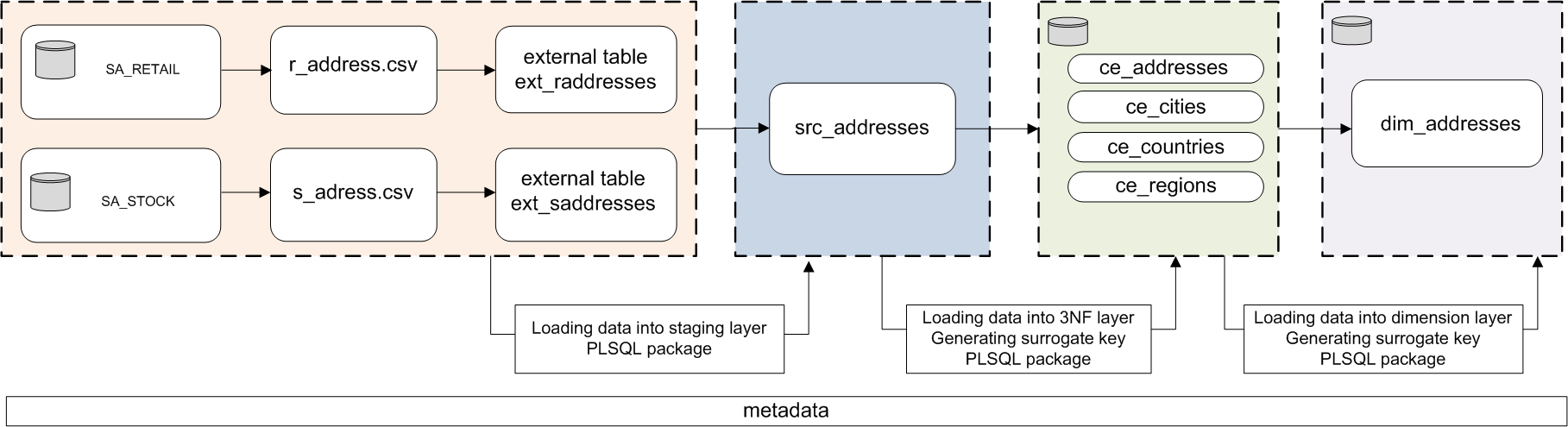




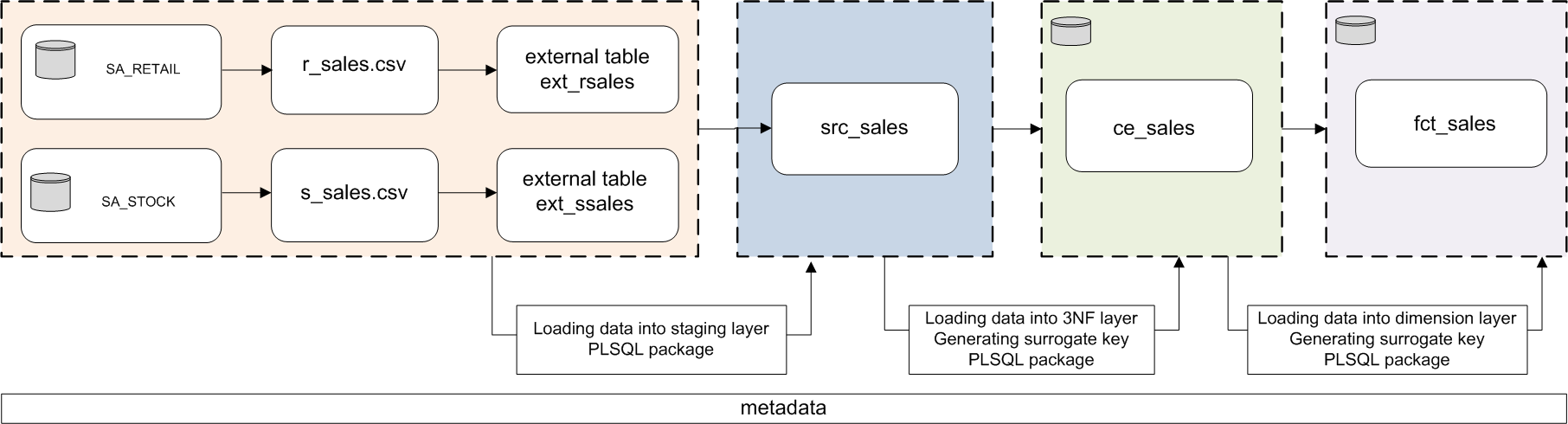








**FACT TABLE**



# Partitioning strategy

For optimization to work with the Datawarehouse will create partitions on the layer BL\_3NF for table CE\_SALES and on the dimension layer for table FCT\_SALE. It improves load data to DWH and query performance. These tables were chosen for partition because contain historical data and the size table will greater than 2 GB.

Since the analysis of sales data is planned to execute by period, the partitioning strategy by range is selected. Selected the period - month as the unit of the partition.