

| Business Template  **COFFEE SHOP SALES** |
| --- |
| **Logo / Image** |

**Contents**

[1](#_heading=h.1fob9te) Business Description 3

[1.1](#_heading=h.3znysh7) Business background 3

[1.2](#_heading=h.2et92p0) Problems because of poor data management 3

[1.3](#_heading=h.tyjcwt) Benefits from implementing a Data Warehouse 3

[1.4](#_heading=h.3dy6vkm) DATASETS DESCRIPTION 3

[1.5](#_heading=h.1t3h5sf) GRAIN / DIM / FACT 4

[2](#_heading=h.4d34og8) Business Layer 3NF 4

[3](#_heading=h.2s8eyo1) Business Layer Dimensional Model 4

[4](#_heading=h.17dp8vu) Logical Scheme 4

[5](#_heading=h.3rdcrjn) Data Flow 4

[6](#_heading=h.26in1rg) Fact Table Partitioning Strategy 4

# 

# Business Description

## Business background

Coffee is more than just a beverage; it's a cultural phenomenon that transcends borders and brings people together. In today's fast-paced world, the demand for quality coffee experiences is ever-growing.

Coffee shops have become an integral part of modern urban culture, offering a cozy ambiance and a variety of beverages and snacks to patrons seeking a respite from their busy lives. Our coffee shop stands out among the competition by providing a diverse selection of high-quality coffee, tea, bakery items, and Drinking Chocolate, catering to the varied preferences of our discerning customers.

## Problems because of poor data management

Inefficient data management poses significant challenges for our coffee shop's success. Without proper data collection and analysis tools, we lack crucial insights into customer preferences, sales trends, and inventory management, hindering our ability to make informed business decisions.

The absence of robust data management systems makes it difficult to identify patterns and trends in customer behavior, limiting our ability to tailor our offerings to meet their needs effectively.

Furthermore, without comprehensive data on our expenses and revenue streams, we struggle to develop sound financial strategies and forecast future growth accurately.

## Benefits from implementing a Data Warehouse

Implementing a Data Warehouse in our coffee shop can revolutionize the way we understand and manage our business operations, offering a plethora of benefits that directly address the challenges we face in today's competitive market. Implementing a data warehouse can answer you the following questions:

* How have sales at Maven Roasters coffee shop changed over time?
* Which coffee shop is more/less profitable and why?
* Which days of the week and times of the day are busiest and why?
* What products sell the most and least often? What generates the most revenue for the business?
* What is the customers' preferred method of payment?
* Which bank card payment is most frequently made with which bank?

## DATASETS DESCRIPTION

The first dataset contains the following information about Coffee Shop Sales Maven Roasters operating out of three NYC locations and **card payment method** with bank’s ids and title**.**

**Information about:**

transaction\_id : Unique identifier of the transaction

transaction\_date : Date of the transaction

transaction\_time : Time of the transaction

quantity\_sold : Quantity of items sold

discount\_amount : Total sale discount

sales : Total sales amount for each product

**Payment**:

payment\_id : Unique identifier of the payment

payment\_method : Payment by card

**Bank:**

bank\_id : Unique identifier of the store

bank\_title : Title of the cardholder bank

**Store:**

store\_id : Unique identifier of the store

store\_description : Description of the coffee shop where the transaction took place

store\_city : City of the store

store\_address : Address of the store

store\_manager\_id : Unique identifier of the store’s manager

store\_manager\_first\_name : Name of the store’s manager

store\_manager\_last\_name : Name of the store’s manager

store\_manager\_passport\_id : Passport number of the store’s manager

store\_manager\_phone\_number : Phone\_number of the store’s manager

store\_manager\_city : City of the store’s manager

store\_manager\_address : Address of the store’s manager

**Product:**

product\_id : Unique identifier of the product

product\_title : Title of the product

product\_category : Description of the product category

product\_description : Description of the product

product\_price : Cost of the product

product\_size : Cup size - Large, Regular, Small

**Employee:**

employee\_id : Unique identifier of the employee

employee\_first\_name : Name of the employee

employee\_last\_name : Name of the employee

employee\_passport\_id : Passport number of the employee

employee\_phone\_number : Phone\_number of the employee

employee\_city : City of the employee

employee\_address : Address of the employee

employee\_position : Job position held by an employee

**Customer:**

customer\_id : Unique identifier of the customer, who is logged in

customer\_first\_name : Name of the customer

customer\_last\_name : Name of the customer

customer\_email : email of the customer

customer\_phone\_number : Phone\_number of the customer

The second dataset contains the following information about Coffee Shop Sales Maven Roasters operating out of three NYC locations and **cash payment method**.

**Information about:**

transaction\_id : Unique identifier of the transaction

transaction\_date : Date of the transaction

transaction\_time : Time of the transaction

quantity\_sold : Quantity of items sold

discount\_amount : Total sale discount

sales : Total sales amount for each product

**Payment**:

payment\_id : Unique identifier of the payment

payment\_method : Payment by cash

**Store:**

store\_id : Unique identifier of the store

store\_description : Description of the coffee shop where the transaction took place

store\_city : City of the store

store\_address : Address of the store

store\_manager\_id : Unique identifier of the store’s manager

store\_manager\_first\_name : Name of the store’s manager

store\_manager\_last\_name : Name of the store’s manager

store\_manager\_passport\_id : Passport number of the store’s manager

store\_manager\_phone\_number : Phone\_number of the store’s manager

store\_manager\_city : City of the store’s manager

store\_manager\_address : Address of the store’s manager

**Product:**

product\_id : Unique identifier of the product

product\_title : Title of the product

product\_category : Description of the product category

product\_description : Description of the product

product\_price : Cost of the product

product\_size : Cup size - Large, Regular, Small

**Employee:**

employee\_id : Unique identifier of the employee

employee\_first\_name : Name of the employee

employee\_last\_name : Name of the employee

employee\_passport\_id : Passport number of the employee

employee\_phone\_number : Phone\_number of the employee

employee\_city : City of the employee

employee\_address : Address of the employee

employee\_position : Job position held by an employee

**Customer:**

customer\_id : Unique identifier of the customer, who is logged in

customer\_first\_name : Name of the customer

customer\_last\_name : Name of the customer

customer\_email : email of the customer

customer\_phone\_number : Phone\_number of the customer

The datasets provide a comprehensive overview of Coffee sales by bank's сards and cash, allowing for analysis and exploration of trends, sales performance, customer preferences, and more within the coffee industry.

## GRAIN / DIM / FACT

**Grain Description (4-steps)**

**Step 1.** Define the business process: tracking sales data

We're dealing with sales data, which involves transactions where products are sold.

**Step 2.** Declare the grain: single unique transaction

If we are tracking sales data, the grain could indeed be a unique transaction. This means that each record in our dataset will contain information about a specific product, and the grain will be a unique transaction identifier that will be used to combine these rows into a single transaction when building the data warehouse; capturing details such as the date and time of the transaction, the coffee shop where the transaction took place, location of the coffee shop and employee who made the sale, the products bought, the quantity sold, price, size, total sales with discount, customer if he is registered in the system and payment method.

**Step 3.** Identify the dimension tables: information about payments, banks, stores, products, employees and customers.

Dimension tables contain descriptive attributes that provide context to the facts (information about payments, stores, banks, products, employees, customers).

**Step 4.** Identify the facts: quantity\_sold - quantity of items sold, discount\_amount - total sale discount, sales - total sales amount for each product. These facts provide essential quantitative information about the sales transactions.

**Dim Description**

**Table “PAYMENT” Description**

The payment dimensional table contains descriptive information about the payment types used for processing payments in sales transactions.

| Column name | Description | Data Type |
| --- | --- | --- |
| payment\_id | unique identifier for the payment, PK | BIGINT |
| payment\_method | method of payment (cash or card) | VARCHAR(100) |

Example with filled data:

| payment\_id | payment\_method |
| --- | --- |
| 1 | cash |

**Table “BANK” Description**

The bank dimensional table contains descriptive information about the banks used for processing payments in sales transactions.

| Column name | Description | Data Type |
| --- | --- | --- |
| bank\_id | unique identifier for the bank, PK | BIGINT |
| bank\_title | official name or title of the bank entity | VARCHAR(100) |

Example with filled data:

| bank\_id | bank\_title |
| --- | --- |
| 4 | HSBC USA Inc. |

**Table “STORE” Description**

The store dimension table contains descriptive information about the coffee shops where sales transactions take place.

| Column name | Description | Data Type |
| --- | --- | --- |
| store\_id | unique identifier for the store, PK | BIGINT |
| store\_description | coffee shop where the transaction took place | VARCHAR(250) |
| store\_city | city of the store | VARCHAR(100) |
| store\_address | address of the store | VARCHAR(150) |
| store\_manager\_id | unique identifier of the store’s manager | BIGINT |
| store\_manager\_first\_name | name of the store’s manager | VARCHAR(100) |
| store\_manager\_last\_name | name of the store’s manager | VARCHAR(100) |
| store\_manager\_passport\_id | passport number of the store’s manager | VARCHAR(100) |
| store\_manager\_phone\_number | phone\_number of the store’s manager | VARCHAR(50) |
| store\_manager\_city | city of the store’s manager | VARCHAR(100) |
| store\_manager\_address | address of the store’s manager | VARCHAR(150) |

Example with filled data:

| store\_id | store\_description | store\_city | store\_address |
| --- | --- | --- | --- |
| 3 | Coffee Shop Maven Roasters №3, a charming corner of the city welcomes early risers and late lingerers alike. | New York | Astoria 7/1 |

| store\_manager\_id | store\_manager\_first\_name | store\_manager\_last\_name | store\_manager\_passport\_id |
| --- | --- | --- | --- |
| 12 | Tariq | Whitaker | 2642125269 |

| store\_manager\_phone\_number | store\_manager\_city | store\_manager\_address |
| --- | --- | --- |
| +16464861486 | New York | Crosby Street 59 |

**Table “PRODUCT” Description**

The product dimensional table contains descriptive information about the products sold in sales transactions.

| Column name | Description | Data Type |
| --- | --- | --- |
| product\_id | unique identifier for the product, PK | BIGINT |
| product\_title | title if the product | VARCHAR(255) |
| product\_category | categorization of the product | VARCHAR(255) |
| product\_description | description of the product | VARCHAR(255) |
| product\_price | cost of the product | VARCHAR(255) |
| product\_size | cup size - Large, Regular, Small | VARCHAR(255) |

Example with filled data:

| product\_id | product\_title | product\_category | product\_description | product\_price | product\_size |
| --- | --- | --- | --- | --- | --- |
| 1 | Brewed Chai tea | Tea | Spicy Eye Opener Chai | 3 | Small |

**Table “EMPLOYEE” Description**

The employee dimensional table contains descriptive information about the employees involved in sales transactions.

| Column name | Description | Data Type |
| --- | --- | --- |
| employee\_id | unique identifier for the employee, PK | BIGINT |
| employee\_first\_name | first name of the employee | VARCHAR(100) |
| employee\_last\_name | last name of the employee | VARCHAR(100) |
| employee\_passport\_id | passport number of the employee | VARCHAR(100) |
| employee\_phone\_number | phone\_number of the employee | VARCHAR(50) |
| employee\_city | city of the employee | VARCHAR(100) |
| employee\_address | address of the employee | VARCHAR(150) |
| employee\_position | job position held by an employee | VARCHAR(100) |

Example with filled data:

| employee\_id | employee\_first\_name | employee\_last\_name | employee\_passport\_id | employee\_phone\_number |
| --- | --- | --- | --- | --- |
| 9 | Wojciech | John | 4642253237 | +16469363585 |

| employee\_city | employee\_address | employee\_position |
| --- | --- | --- |
| New York | Wojciech | Astoria 47 |

**Table “CUSTOMER” Description**

The customer dimensional table contains descriptive information about the customers who made the purchase (if they are registered in the system).

| Column name | Description | Data Type |
| --- | --- | --- |
| customer\_id | unique identifier for the customer, PK | BIGINT |
| customer\_first\_name | first name of the customer | VARCHAR(100) |
| customer\_last\_name | last name of the customer | VARCHAR(100) |
| customer\_email | email of the customer | VARCHAR(100) |
| customer\_phone\_number | phone\_number of the customer | VARCHAR(50) |

Example with filled data:

| customer\_id | customer\_first\_name | customer\_last\_name | customer\_email | customer\_phone\_number |
| --- | --- | --- | --- | --- |
| 722 | Ronald | Pineda | pineda\_r@gmail.com | +16465813713 |

**Table “TRANSACTION” Description**

A table named transaction would contain information about each sales transaction.

| Column name | Description | Data Type |
| --- | --- | --- |
| transaction\_id | unique identifier of the transaction | BIGINT |
| transaction\_date | date of the transaction | DATE |
| transaction\_time | time of the transaction | TIME |

Example with filled data:

| transaction\_id | transaction\_date | transaction\_time |
| --- | --- | --- |
| 21 | 2022-09-01 | 07:34:53 |

**Fact Description**

**Table “PRODUCT\_TRANSACTION” Description**

A fact table named product\_transaction would contain detailed information about each sales transaction.

| Column name | Description | Data Type |
| --- | --- | --- |
| payment\_id | foreign key referencing the payment, FK | BIGINT |
| transaction\_id | foreign key referencing the transaction, FK | BIGINT |
| product\_id | foreign key referencing the product, FK | BIGINT |
| store\_id | foreign key referencing the store, FK | BIGINT |
| employee\_id | foreign key referencing the employee, FK | BIGINT |
| customer\_id | foreign key referencing the customer, FK | BIGINT |
| quantity\_sold | amount of items sold | VARCHAR(10) |
| discount\_amount | total sale discount | VARCHAR(10) |
| sales | total sales amount for each product | VARCHAR(10) |

Example with filled data:

| payment\_id | transaction\_id | product\_id | store\_id | employee\_id | customer\_id |
| --- | --- | --- | --- | --- | --- |
| 1 | 25867 | 54 | 3 | 8 | 567 |

| quantity\_sold | discount\_amount | sales |
| --- | --- | --- |
| 2 | 0.95 | 6.7 |

# Business Layer 3NF



To design a data model in third normal form (3NF) based on the given entities and attributes, we need to ensure that there are no transitive dependencies and that each non-prime attribute is fully functionally dependent on the primary key.

`CE\_Transactions` entity contains details of each transaction, including the date and time of the transaction.

`CE\_Products\_SCD` entity contains details of each product sold, price, size and product\_categories\_id referencing the corresponding product\_categories.

`CE\_Product\_Transactions` - bridge table between `CE\_Products` entity and `CE\_Transactions` entity. Also contains essential quantitative information about the sales transactions: the quantity sold, total sales with discount and information about: the store where it occurred (store\_id), the employee who made the sale (employee\_id), the customer if he is registered in the system (customer\_id) and information about payment (payment\_id).

`CE\_Product\_categories` entity contains details about the category of products.

`CE\_Payments` entity stores payment information, referencing the corresponding bank.

`CE\_Banks` entity holds information about banks involved in payment.

`CE\_Customers` table contains details about the customer.

`CE\_Employees` entity stores information about employees working at different stores, referencing the store they work in, position and city.

`CE\_Positions` entity holds information about the position of the employee.

`CE\_Stores` entity describes each store, including its location details (city, address) and store\_manager\_id referencing employee\_id (one to one relation).

`CE\_Cities` entity holds information about the city.

This design ensures that: There are no repeating groups within the tables. Each attribute is dependent only on the primary key. There are no transitive dependencies between non-prime attributes.

2 Business Layer Dimensional Model



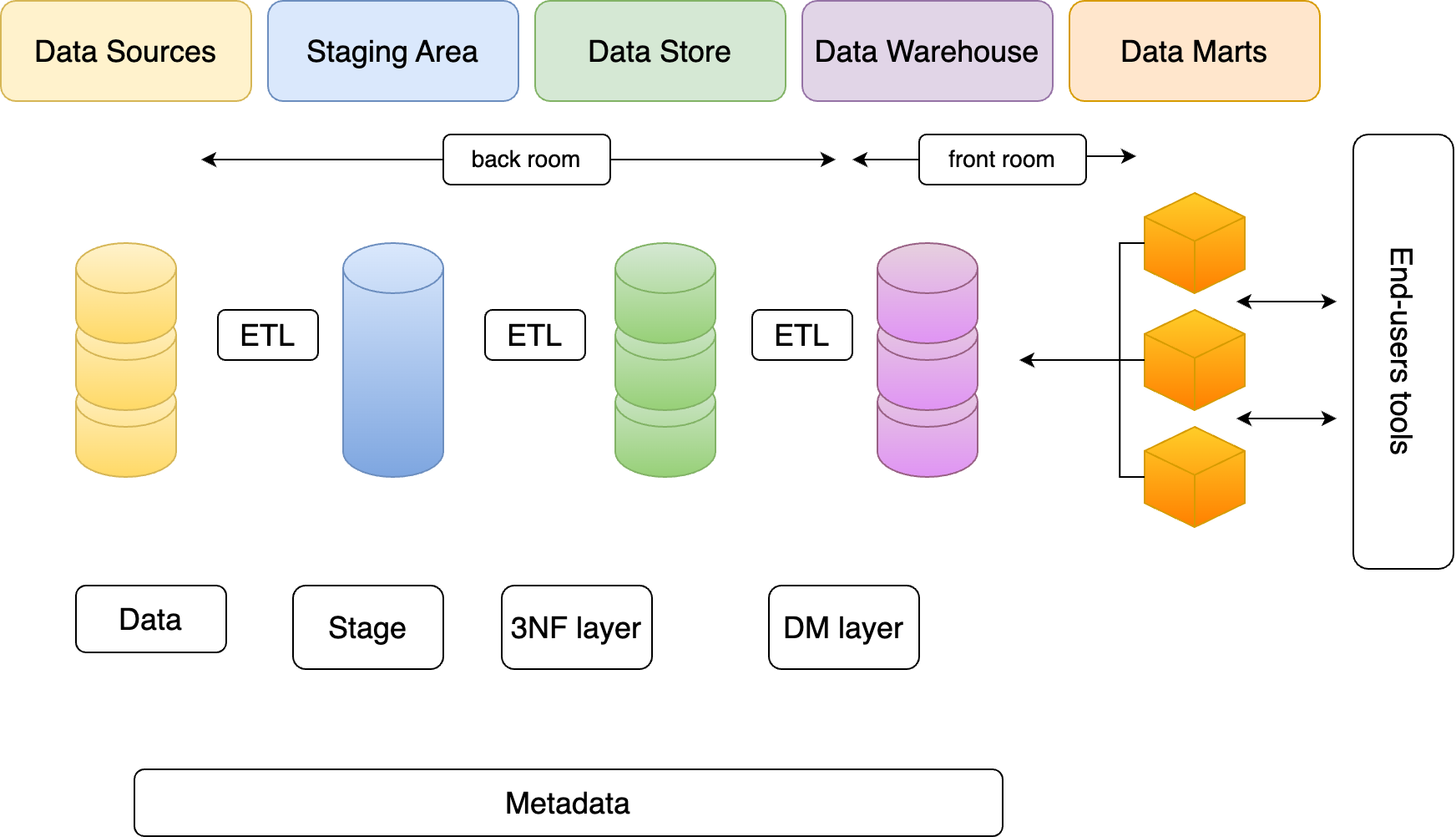
**Metrics**

Metric in the fact table (**FCT\_Transactions\_DD)** representing:

**Revenue per Quantity Sold:** Represents the revenue generated per unit sold (considering the fact that we have a discount on the item if the customer is registered in the system).

revenue\_per\_quantity\_sold = fct\_sales\_unit / fct\_quantity\_sold\_unit

# Logical Scheme



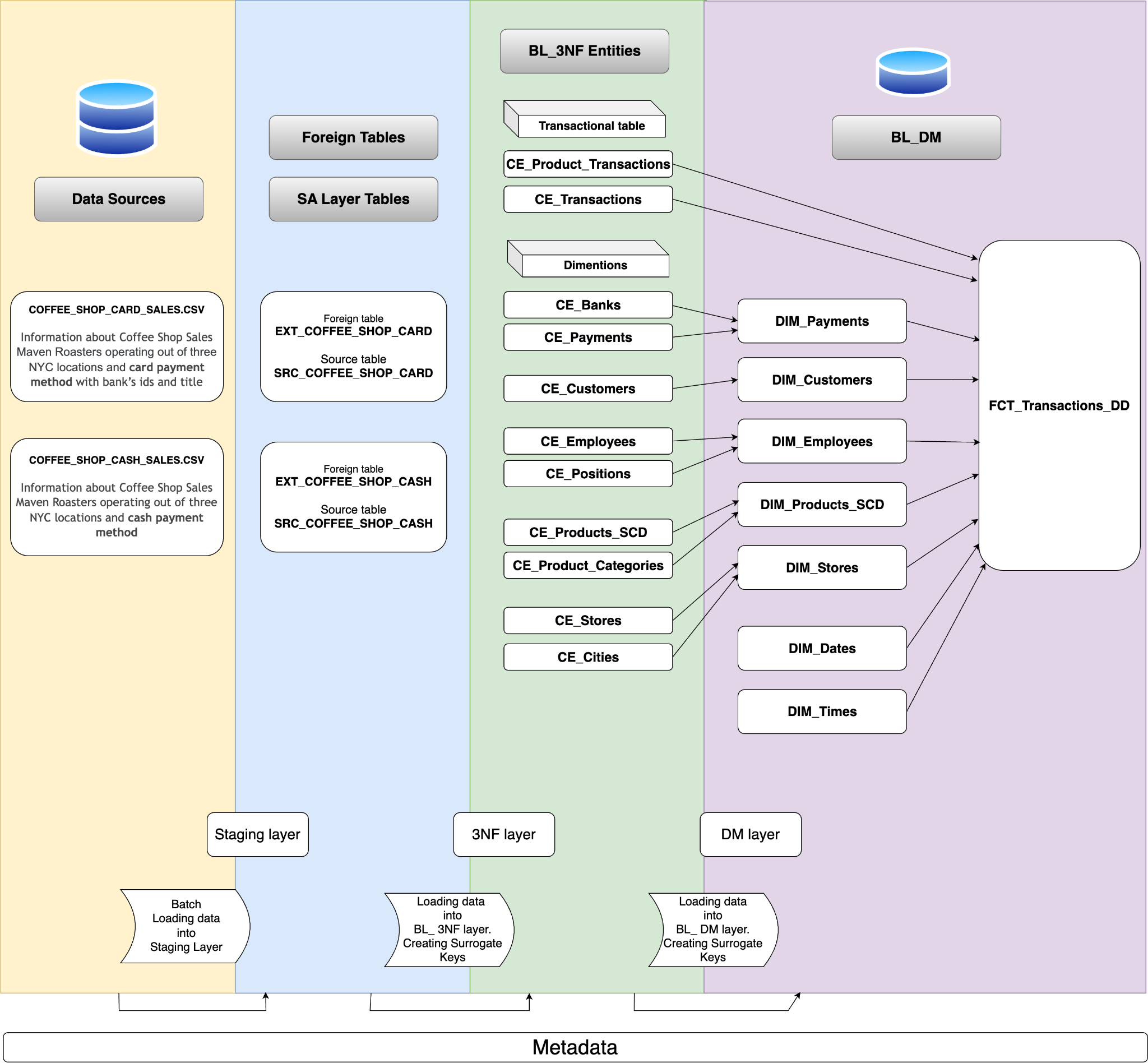
This logical model provides a structured approach to organizing and loading data into the data warehouse, enabling efficient querying and analysis of the data for business intelligence purposes.

Data flows from the data sources to the staging area, where it is temporarily stored.

From the staging area, data is transformed and loaded into the 3NF relational layer, which consists of multiple normalized tables.

Finally, the dimension layer provides descriptive context to the data stored in the relational layer, allowing for easier analysis and reporting.

# Data Flow



At the highest level, the data flow diagram (DFD) will depict the overall flow of data from various sources to the data warehouse.

Then to deep into the ETL processes involved in loading data into the data warehouse. Steps involved in extracting, transforming, and loading data.

Next the processes involved in loading data into the dimension and fact tables within the data warehouse.

# Fact Table Partitioning Strategy

During the creation of the fact table, necessary divided it by event\_dt. After that to create partitions with a rolling window period (3 months).

There is a procedure designed to manage partitions in the bl\_dm.bl\_dm\_fct\_sales\_dd table. It involves detaching an old partition, dropping it if it exists, calculating new partition dates based on the most recent transaction date, and then creating a new partition.  
  
This procedure is execute before inserting data into the fact table.