Business Template

ADVENTURE MOTOCYCLES SHOPS

Logo / Image

CONTENTS

1 BUSINESS DESCRIPTION

1.1 BUSINESS BACKGROUND

Motorcycles is a specific products in a market.

Target market for this business is the type of person who wants the independence of their own transport, so that they can travel when they want. Possibly, a person on a small budget, unable or unwilling to pay for the running costs of a car. Different models of motorcycles are more popular in different regions, genders, perhaps seasons of the year, and for sure for different budgets. This kind of business is very specific and competitive, so if you want to be successful in this field you should very responsibly approach this case and learn a lot of factors which influence on people's choice of buying motorcycles. First of all it can be done by collecting product sales information and analyzing the one using special tools.

The business has its stores in different regions of the world. The main countries are USA, Canada, UK, France, Germany, and Australia. Products are distributed both online and offline.

1.2 PROBLEMS BECAUSE OF POOR DATA MANAGEMENT

Poor data management doesn't let to do successful business because of insufficient information about what should you do next. If you don't use instruments which can give you information for analysis and which can help you to come up with a business strategy you won't be competitive in this or that field.

Businesses are interested in questions about sales in the first place, in order to be able to make decisions about business development based on facts supported by numbers.

1.3 BENEFITS FROM IMPLEMENTING A DATA WAREHOUSE

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

- Which category of the product has the highest prices
- Which ones have the widest distribution of prices
- Which product or model is the most popular by region/season

Further processing data would also let you:

- What is model more popular among women/men, of different ages
- Define how look the average customer of motorcycles (gender, age)
- Determine the frequency of purchase of subcategories of goods (which products wear out the fastest and customers return for them most often)
- Define the most popular motorcycle's model by region/country
- Define trends for buying motorcycles and equipment through the internet or offline.

2 DIMENSIONS OF A BUSINESS

The main two sources of data – online and offline stores. Data from this sources does not match perfectly.

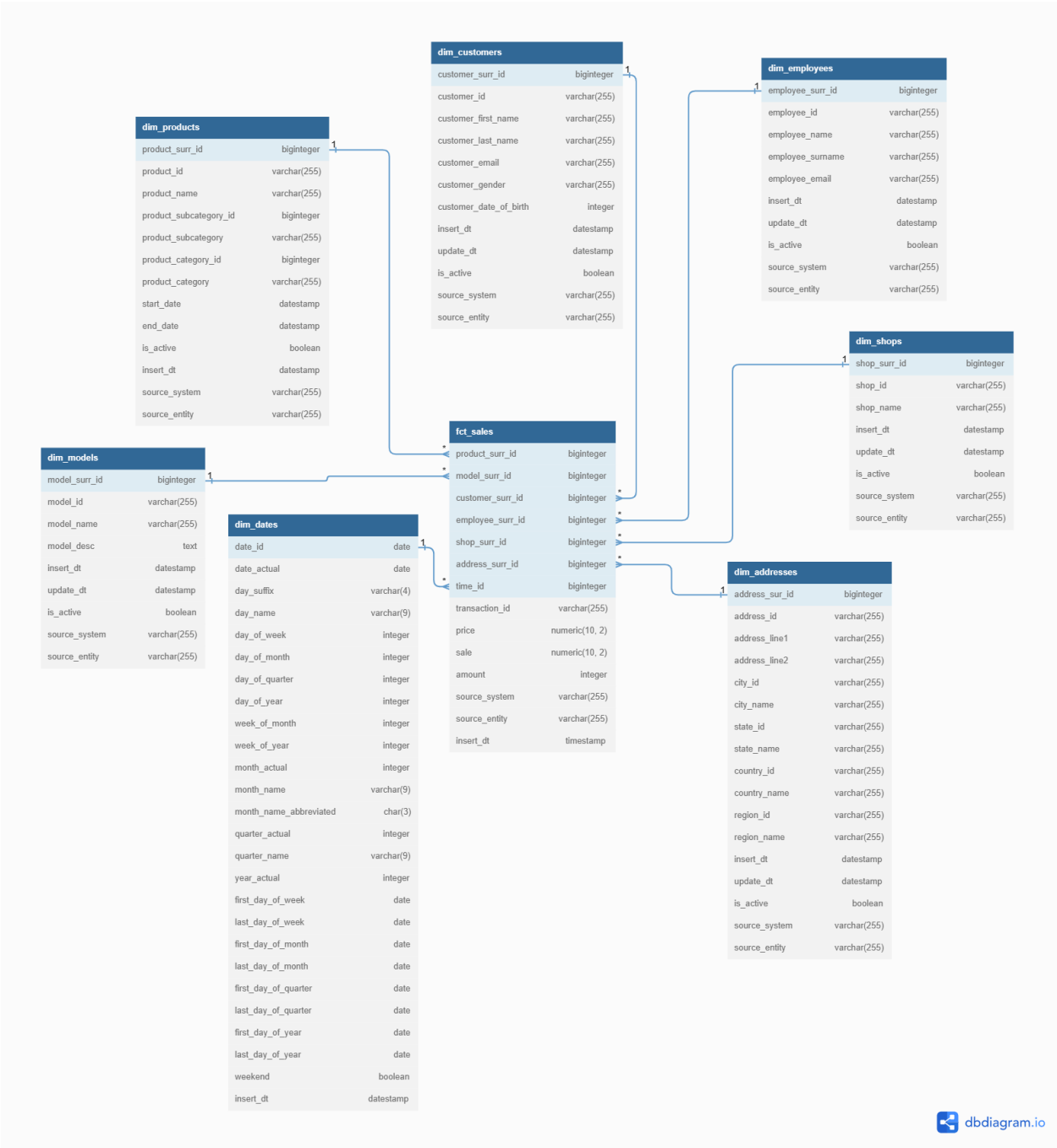
For online data, there is no information about employees and stores. For shops has address data. All transactions from both sources have geo data about city.

The data does not contain information about region but business would like to have such data to make analysis by regions.

The dimensions have been chosen:

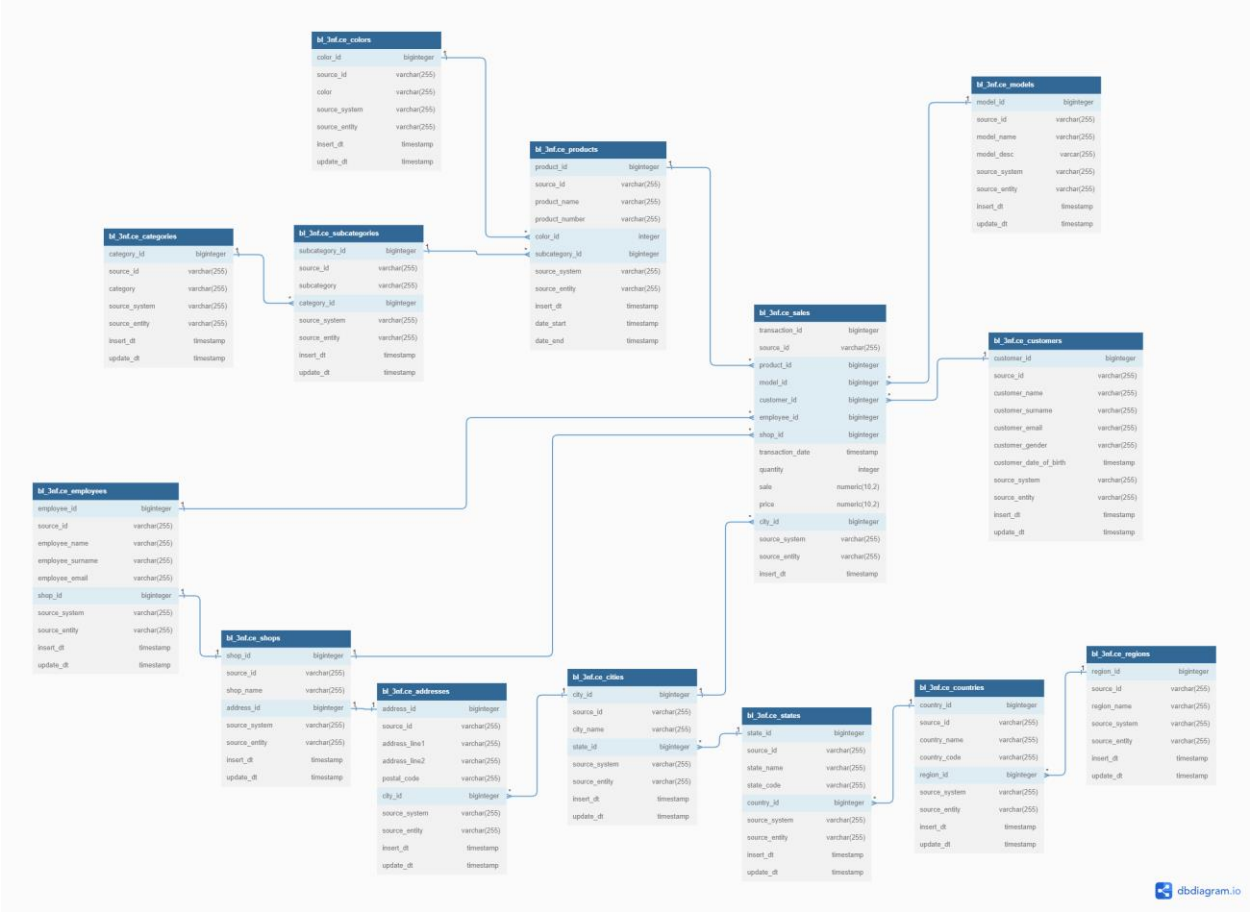
- Employees SCD-1
- Customers (for business is very important to know about customers to make decisions for developing their business) SCD-1

- Products (this dimension table contains all information about the product, its category, subcategory, color and other characteristics) SCD-2
- Models (this dimension table contains information about models and descriptions. We have a connection between products and models through the sales table. Connection between product and models many-to-many) SCD-1
- Shops (this data is available only for offline shops and contains information about shops and their addresses) SCD-1
- Addresses (This table contains data about shop's addresses and cities in case if it's online sales. Address information is available only for offline sales (for shops). Information about cities is available for all transactions) SCD-1
- Dates (this table is used for ease of reporting and speed up performance) SCD-1



3 3NF SCHEMA

After Data staging (consolidation data from different sources), we load the data into the bl_3nf DWH schema. For that, we define common entities and attributes from different sources. We select the main entities and their characteristics, determine what is the primary key and what attributes directly depend on it. Define the connection between tables.



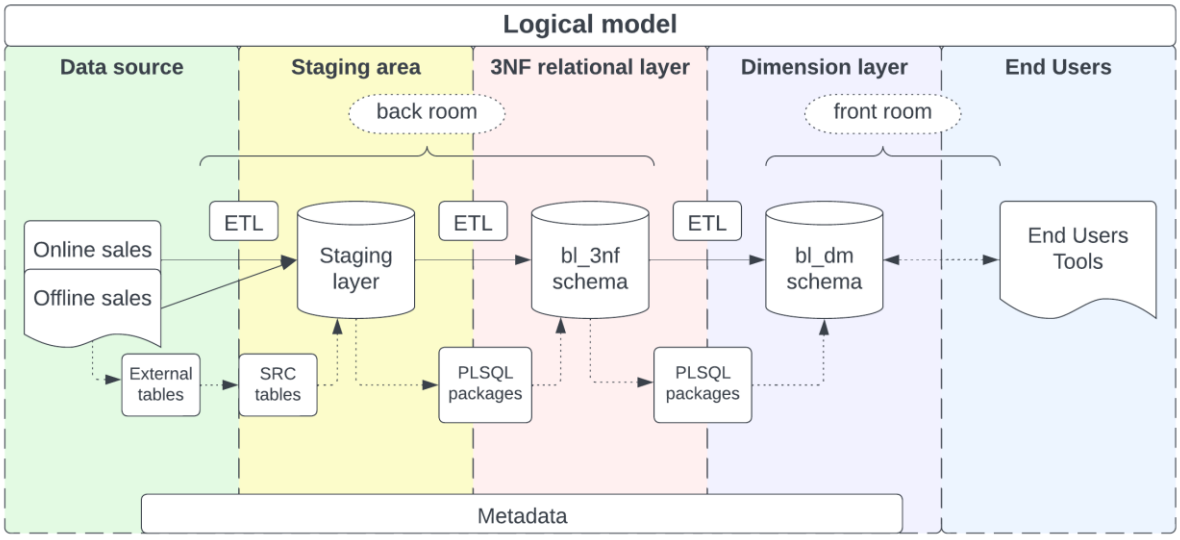
4 LOGICAL SCHEME

Business have 2 sources of data – online and offline. Data will be extract from source systems per day at night into flat files and then transported to the target platform using FTP mechanism. Agreed with business that incremental loading data in DWH run every night.

Transactions made during the day and transactions that came later (foreign transactions come with a delay of 2-3 days) and their characteristics will be extracted from the systems. Input data – 2 files online_sales.csv and offline_sales.csv, which contain only updated data.
The data does not contain information about region, this data will be taken from ISO3166 standard as confirmed.

We have duplicate data from 2 sources. In agreement with the business, we will make deduplication for countries and regions. Duplication is allowed for other data.

Country data is inconsistent between sources, so we will take data from the ISO3166 standard for countries as well. We will create a table to compare data from sources and the ISO standard.

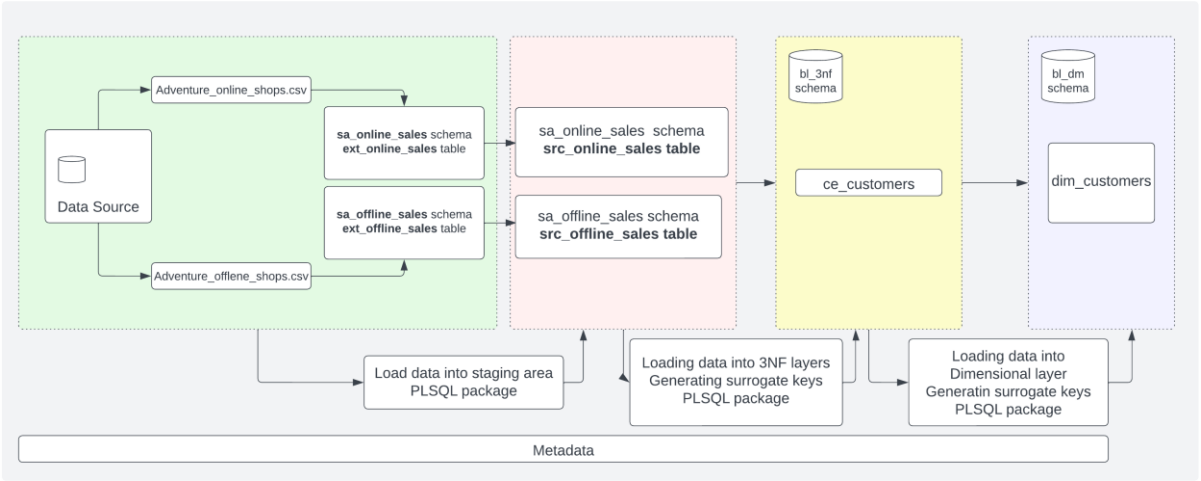


5 DATA FLOW

Customers

Duplication is allowed. It's SCD-1. Source data doesn't contain a whole list of customers. Transaction file contains just customers, whose were active for some period (day, month etc.). Therefore, we will only create and update clients, and not make them inactive.

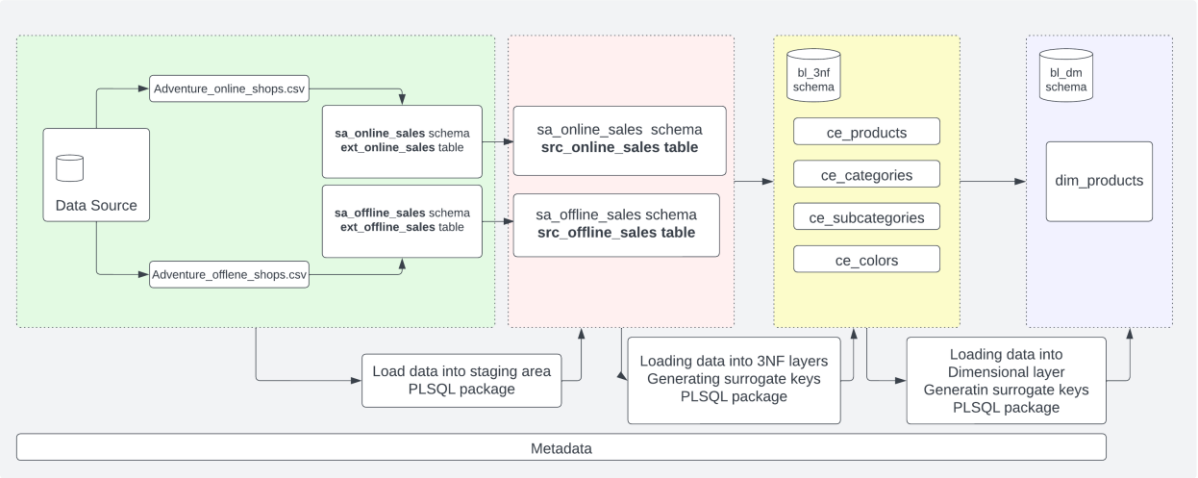
In the future, if the client has the opportunity to send a complete list of active clients, we can deactivate the rest of the clients <TBD>.



Products

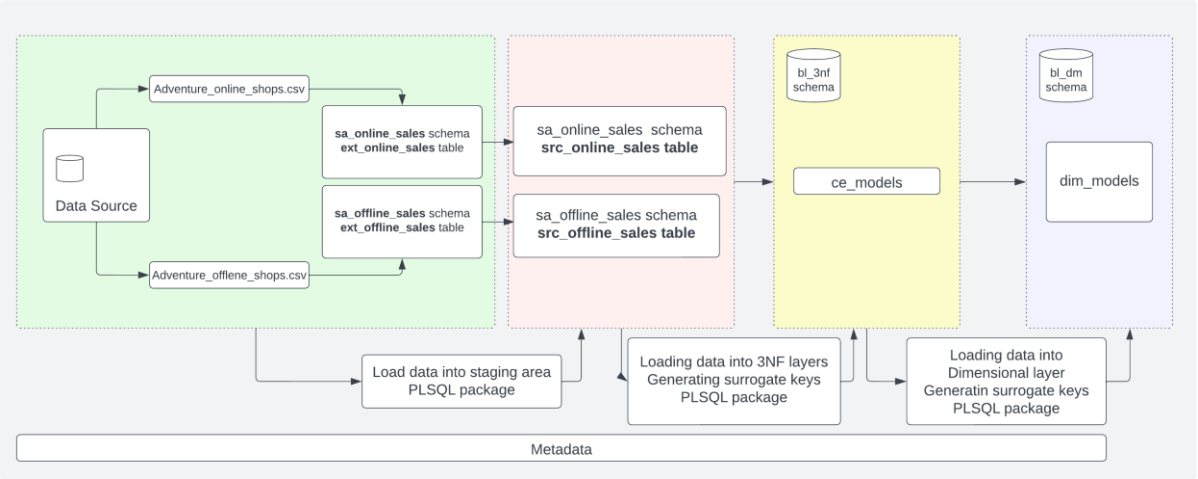
Duplication is allowed. It's SCD-2. Source data doesn't contain a whole list of products. Transaction file contains just products, where was sailed for period of transaction data updates (one day). Therefore, we will only create and update products, and not delete them.

In the future, if the client has the opportunity to send a complete list of active products, categories, subcategories, colors, we can deactivate the rest of them <TBD>.



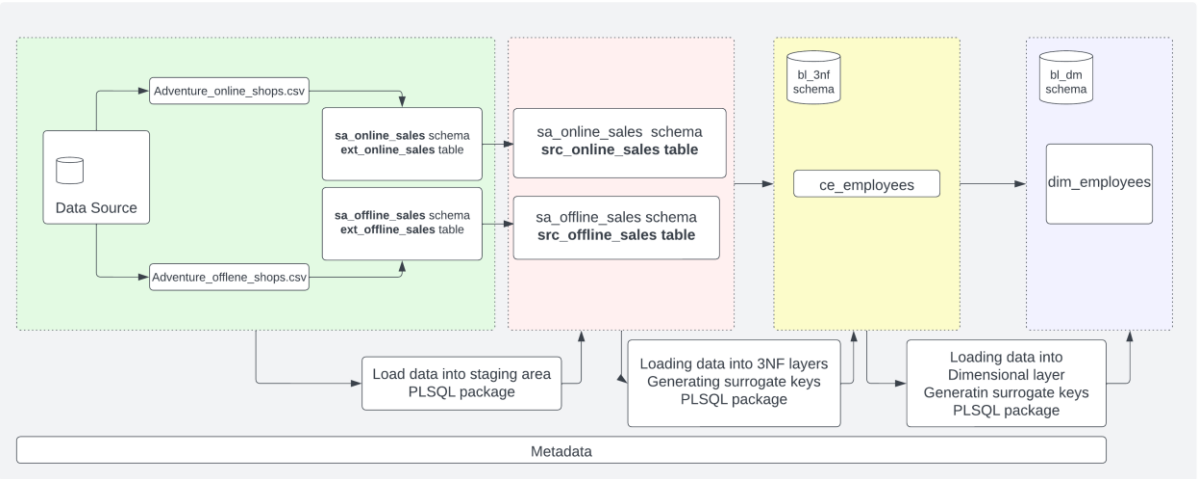
Models

Same behavior as for customers.



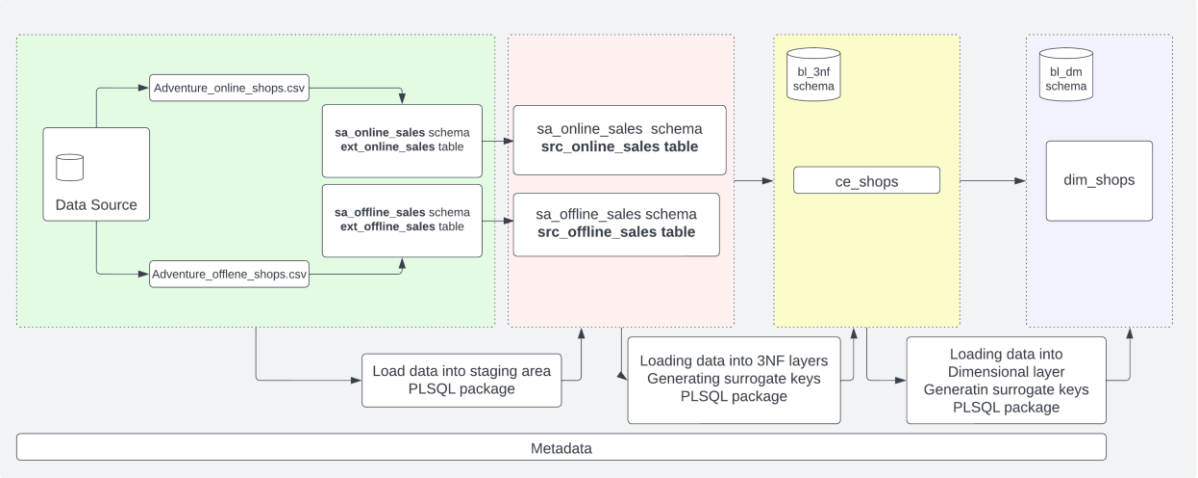
Employees

Same behavior as for customers



Shops

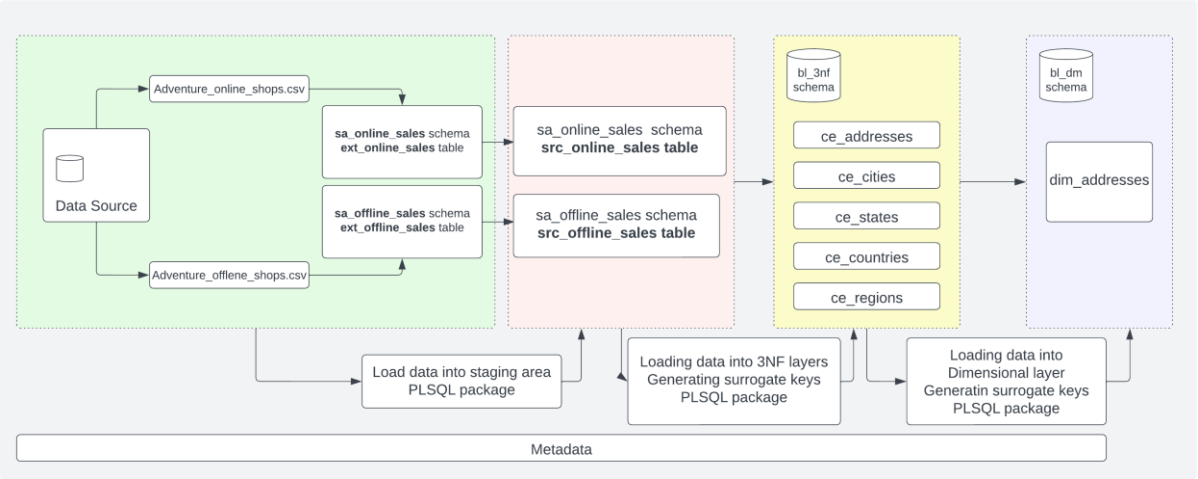
Same behavior as for customers



Addresses

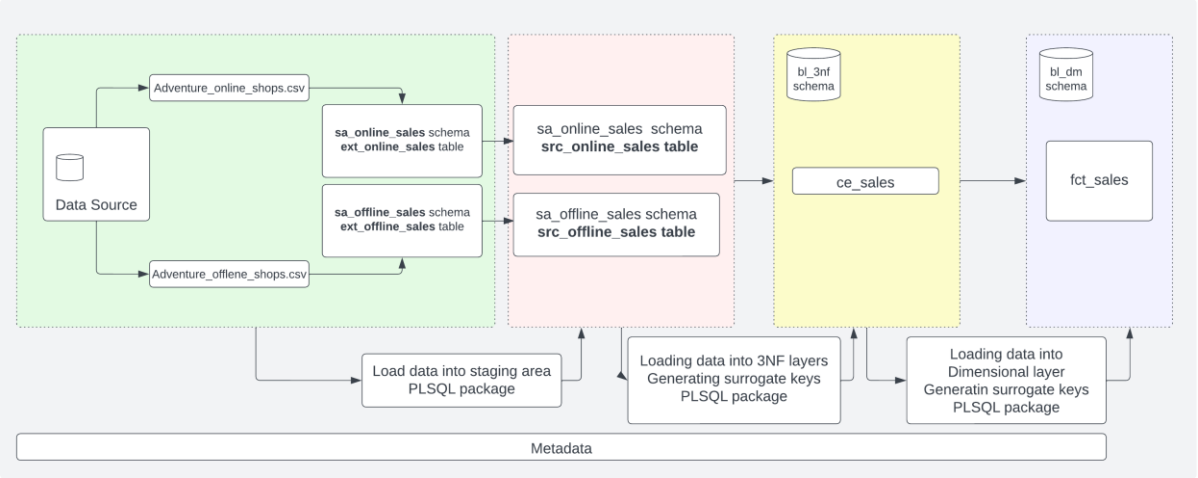
Same behavior as for customers

This dimension has additional logic. Offline data contains information about accurate address, online data contains information only about cities. In that case for transactions that will not have an accurate address, this data will be fill as -1, and this record will have data about cities, states, countries and regions.



Fact table – Sales

Fact table has the sale transaction grain.



6 FACT TABLE PARTITIONING STRATEGY