

Experiment No. 1

Aim: Case study on building Data warehouse/Data Mart.

Problem Statement: Case study on Electronic sales data warehouse and designing star and snowflake schema for it.

Theory:

What is an electronic sales data warehouse?

=>A data warehouse, in this context, is a cloud-based system for gathering, organizing, and storing information about your customers. The name has a brick-and-mortar feel, but it's a modern concept. "Warehouse" is an apt term. A data warehouse creates a single digital place for you to review your information. You can then use that warehouse to run analytics, reports, and measure what's going on throughout the entire company.

Few key benefits of data warehouse:

Faster time to insights

Data analysis always requires data gathering first. If you already have a system in place to collect and store all relevant data, you can run analysis whenever you like.

Reducing the silo effect

You may have incoming data from places like Shopify, Google Analytics, and Klaviyo. Data may still be sitting on paid advertising networks like Facebook, Google, and Taboola. The problem? You can't get a sense of the bigger picture, because that data is stuck within those systems.

Whenever you try to measure your data across multiple channels, things get messy. Your siloed data might be useful, but if you can only see a fraction of the big picture at once that data will by definition only be partly useful to you. A data warehouse reduces the pains of the silo effect and helps you visualize big-picture trends.

Full ownership of data

When you silo your data, you lack a single source of truth for business insights. Even worse: unless you warehouse your own data, you're at the mercy of the data retention policy of every platform you're on. If they decide to ditch your data and you don't already have it, you're out of luck.

With data warehousing, you can migrate that data into your own reference source. If you ever want to refer to it for predictive models (like for personalized product recommendations), all the historical data you need is ready and waiting.

What can you do with an electronic sales data warehouse?

So far, so good. Data warehouse management sounds great. But what do you do with the data? What kinds of returns should you expect on your investment? Let's explore the possibilities.

Attribution modelling

An attribution model means you "tag" your incoming revenue with its appropriate source. You set the rules here. You assign partial or full credit of a sale to individual touchpoints in your sales pipeline.

As a result, you'll have a clearer measurement of internal Return of investment (ROI). Who's making the sales? Which channels are providing the best results? For mostly offline brick-and-mortar retail, it's nearly impossible to pull this off. But in an ecommerce data warehouse environment, these insights are invaluable.

Predictive analytics

In electronic sales, your predictive analytics aren't just for guessing at next-quarter sales. They help you build real, practical product and content recommendations for your customer segments.

A 2015 Forrester study found that predictive "lead scoring" was one of the top use cases here. With lead scoring, you can leverage data to predict which leads are most likely to convert into customers. This creates immediate leverage in marketing: you know who to market to, where to put your money, and what kind of ROI to expect.

Or, take another example: Netflix. When the streaming company created "House of Cards," it wasn't throwing darts at the wall. It used predictive analytics via historical data to determine the kind of show customers had already demonstrated they wanted. Netflix then simply went about creating that show.

Customer segmentation

It's a simple fact of economics, as defined by the Pareto principle: a small portion of your customers are likely to have the greatest impact on your bottom line. Customer segmentation is all about identifying that impact and using it to your advantage.

Typically, customer segmentation has focused on traditional variables, like customer demographics. But an ecommerce data warehouse opens all sorts of possibilities. You can identify and differentiate customers by products purchased, how likely they are to open your emails, and their behavior upon a previous visit. Some ecommerce outlets even offer weather-specific recommendations based on geolocation.

Optimizing paid ads and marketing spend on campaigns

Once you have a more accurate view of your customer segments, you'll have more precise targets for your paid ads. And with your pipelines transferring data into your data warehouse, every new advertising campaign doubles as a fresh learning experience.

For example, A/B split testing lets you target different variables in your campaigns. This includes ad channel selection, high-level messaging, audience targeting, and even the specific copy you use in the ads. Properly channelled into your data warehouse, you'll have the results of every campaign ready for comparison. What works and what doesn't work? Now, you'll know.

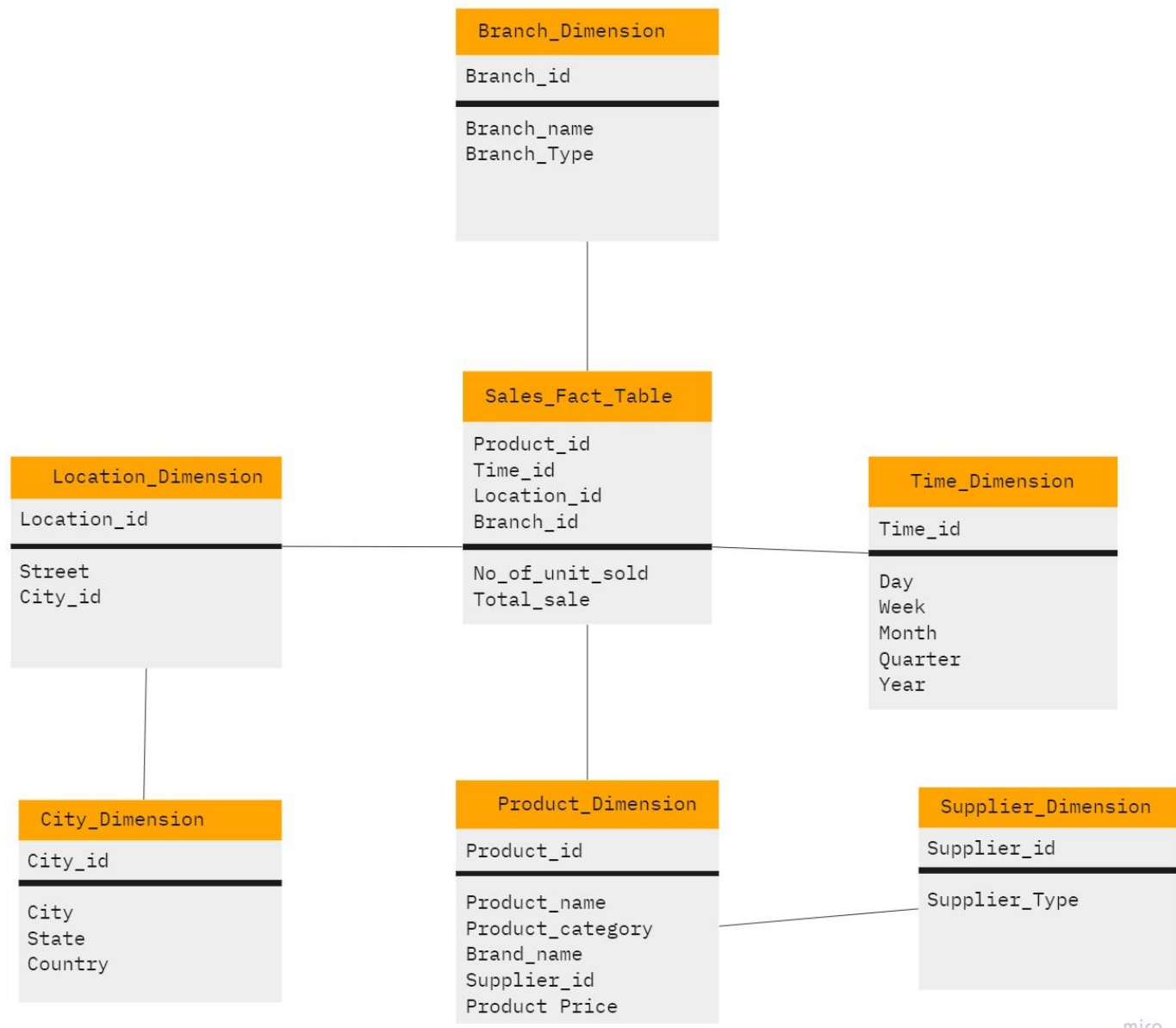
Design Dimensional Modelling:

Star Schema: Star Schema in data warehouse, in which the centre of the star can have one fact table and a number of associated dimension tables. It is known as star schema as its structure resembles a star. The Star Schema data model is the simplest type of Data Warehouse schema. It is also known as Star Join Schema and is optimized for querying large data sets.



Star Schema of Electronic Sales

Snow Flake Schema: Snowflake Schema in data warehouse is a logical arrangement of tables in a multidimensional database such that the ER diagram resembles a snowflake shape. A Snowflake Schema is an extension of a Star Schema, and it adds additional dimensions. The dimension tables are normalized which splits data into additional tables.



miro

Snow Flake Schema of Electronic Sales

Conclusion: We have successfully done case study on Electronic sales data warehouse and designed star and snow flake schema for it.