Team,

After finishing the transformation and cleaning of the staging tables for customers, products, and sales, I’ve moved forward with designing the structure of our warehouse. I’ve decided to build it using a star schema, and I want to talk a little bit about what that means.

The star schema works well for generating efficient reports. Reports pull data from one central fact table (sales) that connects to dimension tables (like products, customers, and dates). Think of a central hub where data is aggregated and several spokes that feed raw data into it.

When deciding how to organize the data we have into tables, here is my process:

* If the data describes something (like a product name, region, or category), it belongs in a dimension table.
* If it’s something we measure (like quantity sold or sale amount), it goes in the fact table.

For example:

* DimCustomer includes names and regions.
* DimProduct has names, categories, and unit prices.
* DimDate covers months, quarters, and years.
* FactSales includes the quantity, total sale amount, and foreign keys pointing to the other tables.

With this setup, our BI tools (Power BI, Excel, etc.) can now easily show things like sales by month, customer trends, top-performing products, and more — all in a way that’s fast and accurate.

Next steps:

We will continue to monitor reports to ensure accuracy and performance

I’m sure you all have noticed the 0’s showing up in Sales reports... As we discussed last week, there is still a non-insignificant amount of missing data regarding figures and dates. I look forward to working with the auditing team to patch in missing data and come to a resolution on this shortly – thank you for your patience.

Let me know if you want a walkthrough or if you have any questions.

Best,

Connor Gaskill