# Homework / Lab 2

# Collecting Business Requirements

MGS 657LEC F1S: Cloud Data Wrhsing & Data Engn (19774 Fall 24)

**Submitted by**

**Anjali Pandey**

# Contents

# Data Warehouse Design Report for The Guac Stop (TGS)

## Design

This report outlines the design of a data warehouse for The Guac Stop (TGS), integrating sales, inventory, and promotional data, based on requirements from the VP of Sales, with a high-level dimensional model for improved analytics and decision-making, along with methodology and recommendations.

## Methodology

The first step is to build design process to better understand the business needs and expectation of data warehouse followed by the star-schema design to address the current shortcomings in existing report and helping the business make informed decisions, enabling us to explore data from different perspectives and gain valuable insights.

- The interview questions primarily focus on KPIs, reporting frequency, promotional metrics to ensure data warehouse support accurate and timely analysis.
- A TGS's system review identified key data warehouse, sales transactions, customer service data, employee management, promotions, inventory and order fulfilment.
- High-level dimensional model includes key dimensions for time, product, order, store, promotions with sales fact table to store retail transaction data.

## Activities

The data warehouse design and implementation will involve the following activities:

- **Stakeholder Interviews & Business Requirement –** Detailed interviews of key stakeholders such as VP, IT managers, and regional managers. This will help to understand the requirement for the business, understanding existing data challenges, and capture the needs of the business, necessary metrics to measure vendor performance and sales trends. And then document the new metrics that the new warehouse should track.
- **Data Source Evaluation –** In this review all available source system, and identify the gaps in the existing systems, missing supplier data, and building the requirement and process on how to improve the vendor's business. The lack of data is a big challenge and ETL process is required to generate this data, creating surrogate keys to collaborate with IT to ensure future data collection with these missing details.
- **Dimensional Modelling –** A high-level dimensional modelling is developed to include key dimensions such as Date, Product, Supplier, Store and Customer to capture the structured data and necessary analytics.
- **Data Flow Mapping –** The data will be mapped out from various source and systems such as CRM, retail transactions and inventory management. This will help us to identify the data, and how the data is moved from operations to reporting including ETL processes.
- **Entity-Relationship Diagram (ERD) -** Design an ERD result that had the relationships showing between main the data warehouse entities which are the (orders, customers, suppliers, products).
- **Environment Setup and ETL Pipeline Configuration –** This step includes creating the necessary pipeline for the data. The pipeline is established to automate the process of

gathering the data from different sources, transforming into usable format and loading of cleaned data into warehouse.

- **Reporting and Analysis –** Once the data is operational the reporting tool is set-up, and it will allow the user to query the data, create reports and data analysis is done to derive actionable insights. Analyse data and generate reports to provide to stakeholders that supports decision making.
- **End User Validation (UAT) –** In this step, the system will validate with real users and confirms the expected outcomes and requirements are met or not. UAT is a platform that ensures that all functions, reports and data queries work correctly, and system is still user friendly.
- **Performance Optimization –** After initial testing, the system goes for improving the efficiency of ETL process and query performance. This will help in providing quick and accurate access to data for reporting and analysis.
- **Initial Pilot Deployment –** To test the system in a controlled environment before rolling out the full plan and gather feedback and make final adjustments before full deployment.

## Limitations

The data warehouse will include all these limitations as follows:

- **Data Availability –** Some of key data points are missing, and it becomes difficult to aggregate and analyse data across order, suppliers and products. This means that reports and analytics are incomplete or inaccurate due to absence of essential data.
- **Data Integration Complexity –** The store consists of various systems to manage its operation, databases and cloud-based platforms. This makes various systems to store various data in various other formats making it challenging to integrate all data into a unified warehouse. ETL is required to extract the necessary and structured data from various format, transform it into consistent structure and load it into warehouse.
- **Scalability –** For the store expansion into international market or nationally, the data warehouse must be scalable to handle the large volume of data and maintain its complexities. This means systems should be capable enough to grow in terms of storage, processing power, number of data source it can handle without affecting the system performance or its accuracy.

## Outcomes

- **Improved Analytics –** The current TGS store is facing many challenges in reporting because TGS lacks the ability to cumulate sales data across multiple dimensions. The whole new warehouse will be able to gain insights into how granular view of sales performance, tracking how individual products and vendors performances across various regions and stores. This will help in more detail analysis and comparison, trend analysis and performance tracking providing a clear view of business operations.
- **Actionable Insights –** This data warehouse will help in actionable insights, particularly regarding vendor's performance and international expansion decisions. By this centralized data warehouse it will help in tracking vendor contributions, it can help in which suppliers are delivering values and which may need to be replaced or negotiated.

# Appendix – I: Entity Relationship Diagram (ERD)

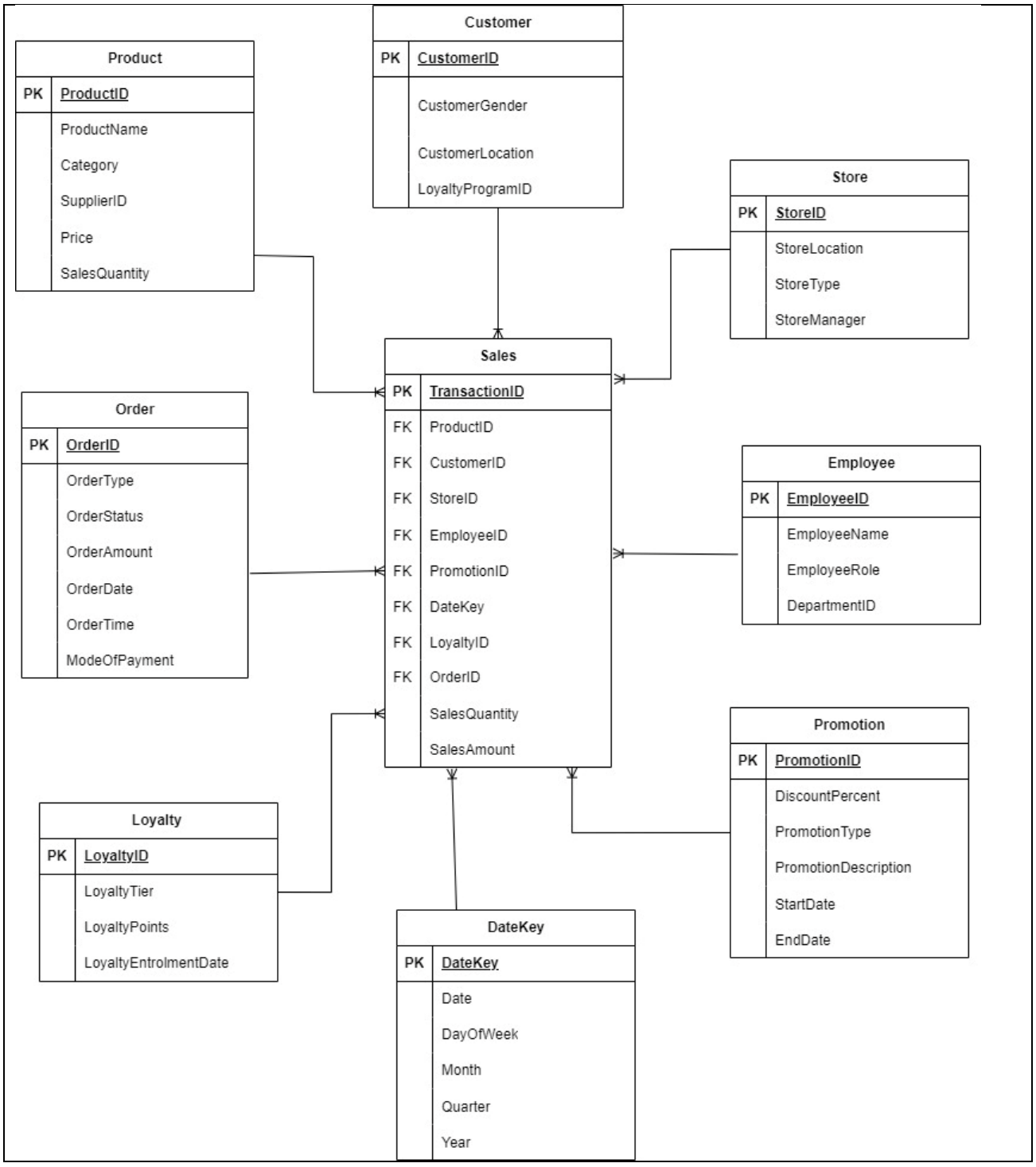


**Fig.**1 – Entity Relationship Diagram (ERD)

## Appendix – II: Enterprise Bus Matrix

The Bus Matrix gives a good way of presentation of information of TGS's basic business processes–Sales Transactions, Order Management, Loyalty Programs, Promotions and Employee Performance–in connection with fundamental conditions like Product, Customer, Store and Date. It permits multi-dimensional analysis for monitoring of sales as well as promotions and employee's work. This makes sure flawless reports and crucial reveals into customer behaviour, sales trends and performance, turning scalable and cohesive data storage into result.

| Business Process | Date | Product | Customer | Store | Supplier | Promotion | Employee | Order | Loyalty |
|---|---|---|---|---|---|---|---|---|---|
| **Sales** | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |
| **Inventory Management** | ✓ | ✓ | | ✓ | ✓ | | | | |
| **Order Management** | ✓ | ✓ | ✓ | ✓ | | | | ✓ | |
| **Promotions and Discounts** | ✓ | ✓ | | ✓ | | ✓ | | | |
| **Employee Performance** | ✓ | ✓ | | | | | ✓ | | |
| **Loyalty Program** | ✓ | | ✓ | | | | | | ✓ |

**Fig.2 –** Bus Matrix

- **Sales**: This is responsible for recording sales transactions and linking the transactions to the dimensions Date, Product, Customer, Store, Promotion, and Employee.
- **Inventory Management**: It is designed to check product availability and stock levels in stores and suppliers and it contains dimensions Date, Product, Store, and Supplier.
- **Order Fulfillment**: In addition to customer and product details, data warehousing involves linkages with Date, Product, Customer, and Store dimensions to monitor orders.
- **Promotions and Discounts**: This is a module that checks promotions' efficiency by store and by product, where dimensions Date, Product, Store, and Promotion are taken into account.
- **Employee Performance**: This is a way of gauging how much employees are contributing to sales and product performance relative to Date, Product, and Employee.

## Appendix – III: Interview Questions

*Question 1: How do you presently assess the effectiveness of the promotion, and what kinds of reports are you interested in about the promotion effectiveness?*

*Reason/Specification*: This question will decide how the Promotion Dimension is designed and how it should be connected with sales transactions. The response will let us know if TGS should need clear reports on promotion effectiveness, hence, an increase in the analytics on the promotional effect on produce and stores.

*Question 2: How are you monitoring and controlling customer loyalty schemes at the moment, and what kind of loyalty-based input do you want to have?*

*Reason/Specification*: This response will shape the creation of the Loyalty Dimension and its integration with the Sales Fact Table. Having a clear idea of the loyalty program that customers are anticipating will give way to a system that not only tracks customer loyalty tiers and points but also can be used in the future for analysis of other data.

*Question 3: How often v. How much do you require in the Hall you need the tech category if you have a program error and then need a part-order? (daily, weekly, monthly)*

*Reason/Specification*: This is the way to show the frequency of the data warehouse. Time dimension will help to get the time information of the data. In addition, it confirms if the system is in real-time or near real time reporting that is based on TGS's operational needs.

## Conclusion

The selected data warehouse framework for TGS will supply a common basis for the combination of data from several internal systems. TGS can make better use of a dimensional model with a Pos sales man the sales, customers, and promotions they make to their clients through their terms of sales will be one of the achievements of TGS. However, the company is requested to take into account the availability of the data sources together with the need for real-time reporting as the company expands.

## References-


1. https://www.integrate.io/blog/the-ultimate-guide-to-data-warehouse-design/
2. https://chatgpt.com/