INTRODUCTION

The Guac Stop (TGS) is a multinational retail organization, founded in 2016. It started as a retail shop for avocado and avocado-themed merchandise, but it soon grew to become a grocery chain with a physical and online presence. With the growth of business, TGS has hired Bullish Consulting, LLP to design a warehouse for their organization to leverage their data for business decisions and streamline the processes.

METHODOLOGY

Based on the discussions with the VP for Sales at TGS and the process outline received by the IT department at TGS, the methodology for building the data warehouse will be the Kimball method. This will be well suited to TGS's needs as it allows us to work with the stakeholders to understand the business requirements and integrate the individual business requirements using fact tables and dimensional tables.[1]

The Kimball methodology leverages a bottom-up approach and will focus on designing the data warehouse using a dimensional modelling approach i.e. a star schema. The star schema makes it easier to create reports due to the fact that it facilitates fast data retrieval. It also gives way to a flexible data warehouse that will cater to the future needs of TGS.[2]

The Kimball method is also a favorable choice economically as its start-up costs are low and do not need highly specialized skill sets. It will also facilitate quicker delivery of the first data warehouse.[3]

ACTIVITIES

The activities involved in building the data warehouse for TGS were,

Requirement gathering interview: The Kimball approach starts with understanding the business requirements and determining how best to add value to the organization. This is the ideal starting point to gather and prioritize enterprise-wide high-level business requirements for information. As a result of these interviews, we were able to obtain a list of business processes that generate data. It also helped us determine the analytical opportunities supported by the data.

Source System Analysis & Data Source Integration: Once we had the business process list, we then analyzed the different processes and systems in detail to identify the relevant data sources. This gave us a clear idea of where the data is stored, its structure and the relationships between the data points. This step was crucial to ensure that all sources were mapped correctly and that the data integration was seamless.

In this step, we were able to figure out that TGS uses different systems such as relational databases (sales transactions), spreadsheets (promotion and store data) and cloud-based (employee management) platforms to store data allowing us to understand how to consolidate the data from these sources.

Dimensional Model Design: Once the sources were identified and the integration plan was complete, we started working on the design of the star schema for the warehouse. Based on our experience in the retail domain, we decided that our fact table has been designed around sales data. The fact table will store metrics such as sales totals, profits, taxes etc. With the sales table at the

---

[1] (Naeem, 2024)
[2] (Naeem, 2024)
[3] (*Kimball Vs. Inmon in Data Warehouse Architecture*, 2021)

center of our star schema, we designed the stores, orders, employees, products, vendors, partners, customers, and promotions dimension tables (*see Appendix Figure 2.1*).

Enterprise Bus Matrix Creation: Moving further in the design phase, we created the enterprise bus matrix for the data warehouse (*see Appendix Table 3.1*). An enterprise bus matrix provides a roadmap for how the facts and dimensions align with the business process such as sales and promotions. It plays an important role in avoiding data storage redundancy and ensuring consistency in reporting across different processes. Moreover, it plays a crucial role in scaling the warehouse as the business needs grow since it allows the addition of new processes (dimensions) in a structured manner.

These activities will be followed by implementation, testing, deployment and maintenance activities as we move forward with building the data warehouse.

LIMITATIONS

Data Inconsistency: The data is being stored in different ways across different departments at TGS. The sales transactions are stored in relational databases, promotion and store data is tracked using spreadsheets and employee management is done using cloud-based systems. The different systems could lead to inconsistent data and make the integration of the data while ensuring consistency a challenge.

Data Complexity: With the prospects for expansion to different countries in the future. The expansion can introduce different operational practices and tax systems. Complexity can also occur when handling different currencies and conversion rates. The business models in the different countries could also vary if TGS chooses to operate only physical stores or online stores. There can be multiple different scenarios that could contribute to the data complexity.

Lack of Historical Data: It is a possibility that the current systems at TGS might not have enough data to support trend analysis. The lack of past data will limit the ability to provide seasonal trends, consumer behavior patterns, vendor performance and yearly performance comparisons. We acknowledge that the data currently stored might be recent with limited retention of older data.

OUTCOMES

Centralized Repository: The data warehouse will consolidate the data from different systems into one single place. This will give way to better and more accurate reporting.

Real-time Reporting: One of the needs highlighted in the requirement gathering interview was being able to view reports in real-time. Real time data warehousing lets businesses process data in real time giving them access to quick insights using the most up-to-date information.[4]

Performance Tracking: The warehouse will allow TGS to monitor store, vendor, product, promotion and employee performances and give detailed insights that will lead to informed decisions.

Support for Future Growth: TGS will be able to scale the data warehouse to address the complexity arising as operations expand. This will allow them to integrate new processes and data sources seamlessly.

---

[4] (Richman, 2024)

REFERENCES

SQLServerCentral. (2012b, April 8). *The Kimball Approach – SQLServerCentral*. https://www.sqlservercentral.com/blogs/the-kimball-approach

Naeem, T. (2024, September 3). Data Warehouse Concepts: Kimball vs. Inmon Approach | Aomstera. *Astera*. https://www.astera.com/type/blog/data-warehouse-concepts/

Kimball vs. Inmon in Data Warehouse Architecture. (2021, June 5). Zentut. https://www.zentut.com/data-warehouse/kimball-and-inmon-data-warehouse-architectures/

Ross, M. (2016b, January 26). *Design Tip #115 Kimball Lifecycle in a Nutshell - Kimball Group*. Kimball Group. https://www.kimballgroup.com/2009/08/design-tip-115-kimball-lifecycle-in-a-nutshell/

Kimball DW/BI Lifecycle Methodology - Kimball Group. (n.d.). Kimball Group. https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/

Melanie. (2024, March 10). *Kimball Method: What is it? How do I use it?* Data Science Courses | DataScientest. https://datascientest.com/en/kimball-method-what-is-it-how-do-i-use-it

Richman, J. (2024, September 27). What Is a Real Time Data Warehouse? Benefits & Best Practices. *Estuary*. https://estuary.dev/what-is-real-time-data-warehouse

Chatgpt. (n.d.). https://chatgpt.com/

Note: Chat GPT & Grammarly have been used everywhere to enhance and improve text.

APPENDIX

1. **Requirement Gathering Interview Questions [5]**

    1.1. What is TGS's business model in terms of setting up stores, getting the product and fulfilling orders?

    This was crucial in understanding how the business works and determining how the data warehouse should be structured to support the different business processes such as order fulfillment, sales tracking, and inventory management.

    1.2. What are the top KPIs you want to see standardized and made available in the data warehouse?

    Understanding the KPIs that are important to the management helped align the data warehouse with the company's performance evaluation goals.

    1.3. What are the different sales mediums that TGS have and does this data get stored in the system somewhere?

    This question was important to understand if TGS is limited to physical grocery stores or online stores or both and if that data can be used to structure the sales from separate mediums differently.

    1.4. How frequently do you need access to real-time reports, and what specific types of reports should be prioritized for automation?

---

[5] Based on the notes from the interview, we asked Chat GPT to prepare the questionnaire.

This response to this question helped us get an idea about the reports that are crucial for the daily operations of TGS, which further helps define which data needs to be automated in the warehouse.

1.5. Do you see TGS's business model changing majorly in the future with its plans to expand to different countries?

Understanding the expectations in terms of changing business needs and models helped in deciding the methodology for the development of the data warehouse.

1.6. Does TGS only serve the B2C sector or also engage with the B2B sector as well? Does the leadership need this segregation in the reports?

Understanding this segregation helped ensure that the data warehouse structures B2B and B2C sales and provides customer insights for both.
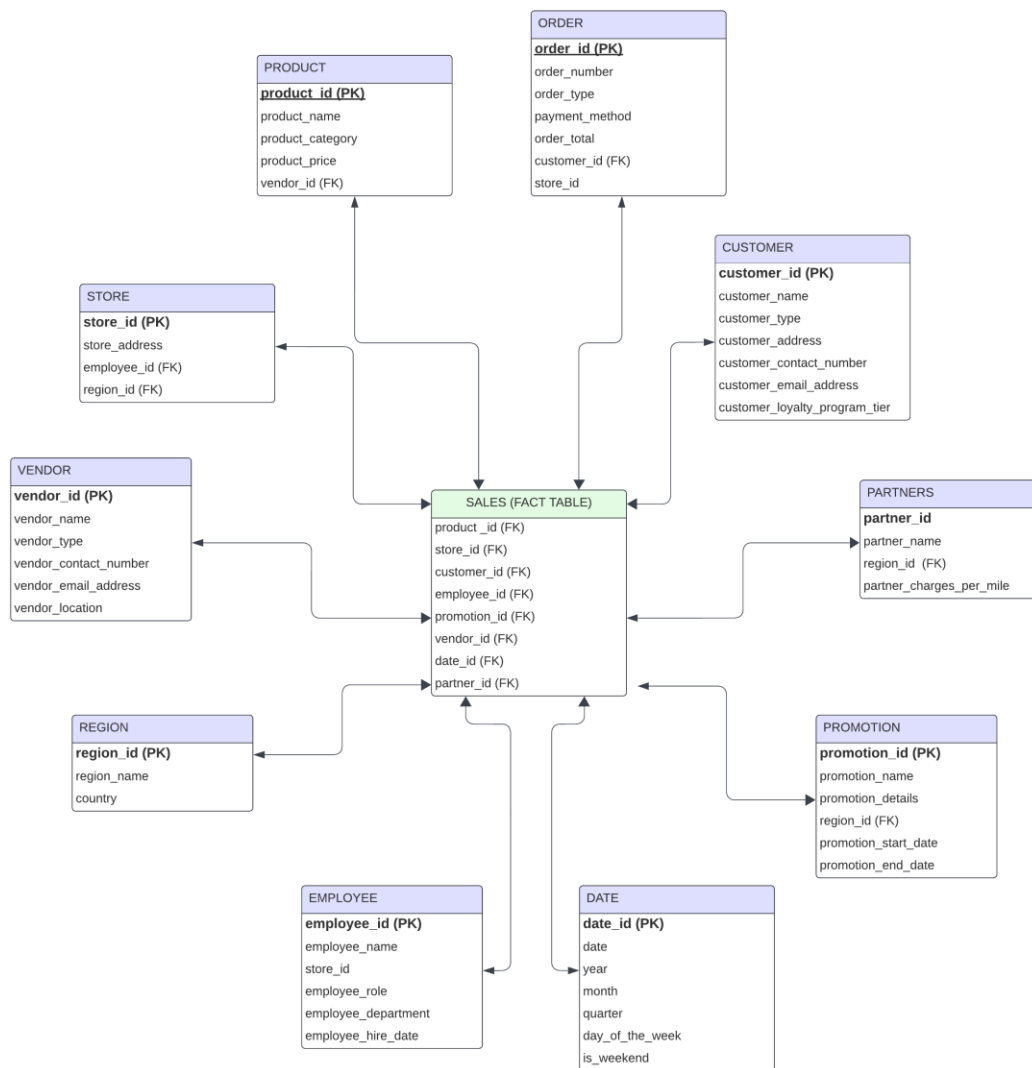
2. Entity Relationship Diagram

Figure 2.1: Entity Relationship Diagram

3. Enterprise Bus Matrix

Business Process/ Dimensions	Store	Order	Employee	Product	Vendor	Partner	Customer	Promotion
Sales	✓	✓	✓	✓	✓	✓	✓	✓
Customer Loyalty Program	✓	✓						
Promotion Tracking	✓			✓			✓	✓
Employee Performance	✓		✓					
Store Performance	✓		✓					✓
Vendor Performance	✓			✓	✓			

Table 3.1: Enterprise Bus Matrix