Study on a Data Warehousing for E-commerce Logistics

Gulzat Turken[1], Zhanerke Temirbekova[1,2], Lyazat Naizabayeva[2] and Manuel M. Barata[3]

[1]*Al-Farabi Kazakh National University, Almaty, Kazakhstan*
[2]*International Information Technology University, Manas St. 34/1, Almaty, 050040, Kazakhstan*
[3]*ISEL, Lizbon, Portugal*

Abstract

As a result of advancements in science and technology and the elevation of people's living standards, the growth of the e-commerce sector has been swift. This growth has been accompanied by a surge in demand for logistics services. The fundamental challenge faced by many enterprises is how to curtail logistics expenses and enhance logistics effectiveness in order to better support the development and operations of e-commerce businesses. This, in turn, improves the core competitiveness of these enterprises and enhances customer service. At the enterprise level, achieving these objectives represents a crucial aspect of cost reduction and operational efficiency enhancement. Numerous researchers and enterprise management decision-makers are actively exploring strategies to lower logistics costs and maximize the role of logistics in enterprise functioning. As e-commerce experiences swift global expansion, the need to enhance logistics efficiency has become a necessity for both businesses and consumers. The utilization and widespread adoption of big data technology in conjunction with its development and application have contributed to increased efficiency and reduced logistics expenses. The update and maintenance of the E-commerce logistics data warehouse system now only demand 7 hours, resulting in a remarkable 92% reduction in maintenance costs. Consequently, this leads to more proficient and informed professional management of logistics warehouses.

Keywords

Data warehouse, E-Commerce logistics, ETL (Extract-Transform- Load), DWHA (Data Warehouse Architecture).

1. Introduction

Data warehouse technology is also more used in business decision-making. When the enterprise is planning for increasing its productivity during financial investment, frequently encountered problems are how to use the limited resources to maximize the profit margin and decrease the loss [1]. In order to communicate data between various departments and facilitate the analysis of massive amounts of data, using the data warehouse technology, Spark technology, OLAP technology, etc. To design a business intelligence platform based on data warehouse, so as to achieve the purpose of improving the company's decision- making ability. The term" data warehouse " (Bill Inmon) was coined in 1991. It is a domain-oriented, integrated, relatively stable dataset that reflects historical trends and is often used for Decision Analysis [2]. In the ever-evolving landscape of e-commerce, efficient and data-driven logistics management is paramount for businesses seeking to thrive in the digital marketplace. The ability to store, process, and analyze vast volumes of data is crucial for making informed decisions, optimizing supply chains, and ultimately delivering an exceptional customer experience. This is where data warehousing comes into play as a pivotal tool that empowers e-commerce logistics to harness the power of data. In the context of increasing market competition, many top domestic and foreign corporations are using information management methods to improve their operational efficiency [3].

DTESI 2023: Proceedings of the 8th International Conference on Digital Technologies in Education, Science and Industry, December 06–07, 2023, Almaty, Kazakhstan
turken.gulzat@gmail.com (G. Turken); zh.temirbekova@iitu.edu.kz (Zh. Temirbekova); l.naizabayeva@iitu.edu.kz (L. Naizabayeva)
0000-0003-4981-514X (G. Turken); 0000-0003-3909-0210 (Zh. Temirbekova); 0000-0002-4860-7376 (L. Naizabayeva); 0000-0002-8335-4052 (M.M. Barata)

The data warehouse is based on the system data of various departments of the e-commerce company and uses some technologies and algorithms to display the results of the report according to the needs of the company's users' order data, business data, third-party data, and other related data. Based on the results of the report, decision-makers make relevant measures for e-commerce logistics [4]. A data warehouse, in the context of e-commerce logistics, is a centralized repository that consolidates and integrates data from various sources, such as sales transactions, inventory levels, shipment tracking, and customer interactions. By harmonizing and storing this wealth of information, e-commerce companies can gain valuable insights, make data-driven decisions, and enhance their operational efficiency.

To achieve the formulated goal, the following tasks were set:

Create data warehouse architecture

- Analyze the architecture of the main data warehouse most commonly used in large corporations.
- Summarize the shortcomings of the existing architecture.

Development of a logical model based on enterprise data warehouse architecture

- Creating data dimensions based on the design of the logical model architecture method.
- Classifying the identification and use of data.
- Developing a systematic and comprehensive data structure by determining the degree of data and selecting dimensions.

Implementation of methods

As for the implementation of enterprise data warehouse solutions, we design and implement the following steps:

- Implementation of the data model in Erwin modeling.
- Use Talend to implement the ETL process.
- Data warehouse implementation in Oracle SQL Developer.
- Use the open-source platform Grafana for monitoring the data in the Data warehouse.
- Display graphical reports using BI (Tableau) - to analyze sales data in the created data warehouse.

Realization of the enterprise data warehouse for the E-commerce logistics

- Realize the creation of an enterprise data warehouse according to the architecture model.
- Create appropriate tests for implemented functions and analyze, compare, and verify test results.

2. Review of literature

E-commerce has seen exponential growth in recent years, and logistics are at the core of its success. Efficient management of the supply chain, order fulfillment, inventory, and customer satisfaction are all dependent on data-driven decision-making. Data warehousing plays a pivotal role in this scenario, serving as the backbone for effective data analysis and decision support. In this literature review, we delve into the key themes, methodologies, and findings from various research studies related to e-commerce logistics data warehousing.

The research [5] assesses the data loading speed into the information system. By comparing the throughput of optimized and unoptimized systems, an average throughput difference of 85% is observed. This suggests that the optimization of the ETL process and the data warehousing strategy leads to substantial improvements in both query performance and data loading speed, even as data volumes continue to increase. Praveen Kumar and Dr. Kavita (2015) [6] explored the diverse methodologies pertaining to the conception and administration of data warehouses. They delved into the construction of data stores, emphasizing that these can be fashioned through a bottom-up approach, a top-down approach, or a hybrid fusion of both. The study also delineated the overarching design procedure involved in this undertaking. In this manuscript [7], authors embrace a data-centric methodology and present Order Monitor, an innovative visual analytics platform engineered to support warehouse managers in their endeavors to assess and enhance real-time order processing efficiency. This system is grounded in streaming warehouse event data and plays a pivotal role in the warehouse operation landscape. This visualization is constructed around a pioneering pipeline design concept inspired by the sedimentation metaphor. It is specifically designed to facilitate real-time order monitoring while also proactively highlighting any orders that exhibit signs of potential irregularity or abnormality. In this study [8], researchers engineered a Data Warehouse system and developed a business intelligence dashboard by leveraging Superstore Europe's dataset. their primary objective was to establish a robust framework for tracking and enhancing sales performance and the efficiency of goods delivery. This was achieved through the utilization of a PostgreSQL database and the study meticulously followed the Kimball Lifecycle methodology. The following article [9], the comprehensive analysis presented here furnishes the research community with a panoramic view of the cutting-edge methodologies. Authors anticipate that this overview will serve as a catalyst, inspiring researchers and industry practitioners to elevate the standards of current practices and to innovate fresh approaches in this domain.

3. Research framework

The data warehouse (DW) comprises two primary components: the initial component is a cohesive decision support database, while the latter is software designed to gather, sanitize, convert, and archive data from diverse operational data origins and external data sources. These components are harmonized to fulfill the requisites of historical, analytical, and business intelligence (BI) purposes. Additionally, a data warehouse may encompass multiple associated data repositories, which constitute a subset of the data warehouse database. Broadly speaking, any system for data retrieval or storage aimed at delivering support for business analytics can be referred to as a data warehouse.

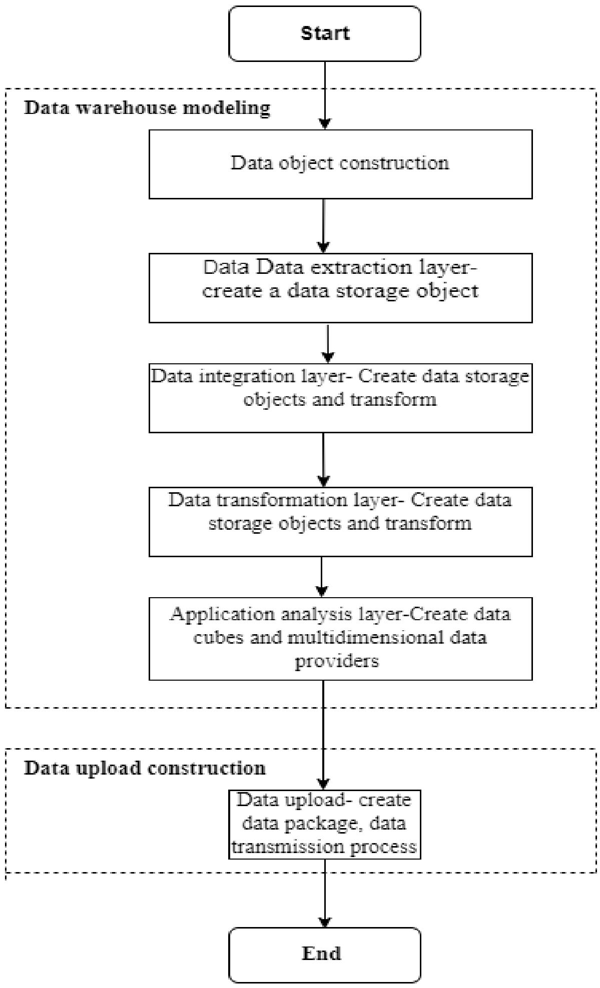

Figure 1: Enterprise SAP BW general realization process

The overall implementation process is mainly carried out in two steps, namely data warehouse modeling and data upload construction. The specific implementation process details are shown in Figure 1.

The main work in the data warehouse modeling part consists of the following five steps:

1. Creating the characteristics and key values of information objects, as the smallest unit of data storage in a data warehouse and the basis for forming data storage objects and information cubes, creating information objects is the most fundamental part of modeling;
2. At the data extraction layer. Create two-dimensional table-structured data storage objects for storing the finest granularity of business data.
3. In the data consolidation layer, a data storage object of two-dimensional tables is created and transformations are established to store the data from the data extraction layer in the data storage object after certain logical transformations and cleansing.
4. In the business transformation layer, data storage objects of two-dimensional tables are created and transformations are created according to the special needs of certain departments, and special logical transformations and aggregations of data in the data consolidation layer are performed.
5. Application analytics layer, building information cubes and multi-information providers with coarse data granularity. The information cube builds the physical data layer for the application analysis layer, and the multi-information provider builds the virtual data layer for the application analysis layer, which is used for the final business analysis and data reporting.

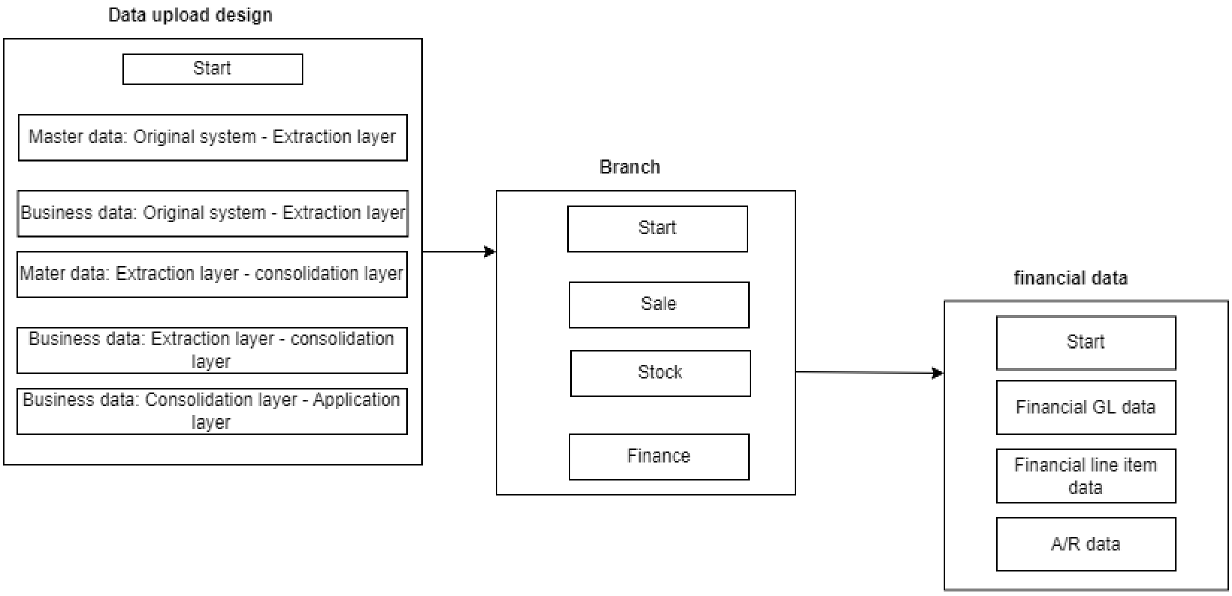

Figure 2: Enterprise SAP BW Data Upload Design Diagram

The data upload in the data warehouse is executed in a time-phased manner, and the design is split and executed according to the branches, and for each subsidiary, the processing design can be done in the order of the database hierarchy. As shown in Figure 2.

Our research work uses a hybrid Inmon and Kimball architecture [12], which grafts Kimball and Inmon architectures, stores the extracted data in a standardized data warehouse, and then extracts the data representation based on dimension modeling on this basis, and develops it to data analysts. Data Warehouse Architecture refers to the design and implementation of an enterprise-wide data warehouse solution that integrates and manages data from various sources to enable business intelligence and analytics. Figure 3 shows the process of building a data warehouse involves extracting data from multiple sources, transforming and consolidating the data, and storing it in a centralized location for easy access and analysis. The data sources could be external data in the form of CSV files, or relational databases. After loading, the data is analyzed using various services to create graphs, test models, and different analytical data. This analysis provides valuable insights into business performance and helps organizations make informed decisions. The insights generated from the analysis are then integrated with the operational systems to support decision-making processes, which is known as the integration of knowledge and action in the data center. Finally, the analyzed data is presented through various front-end applications, reporting files, web applications, and query tools. These tools enable users to access and analyze data quickly and easily. Overall, a well-designed Data Warehouse Architecture provides organizations with a scalable and reliable platform to manage data, enabling them to make informed business decisions and achieve their strategic goals.

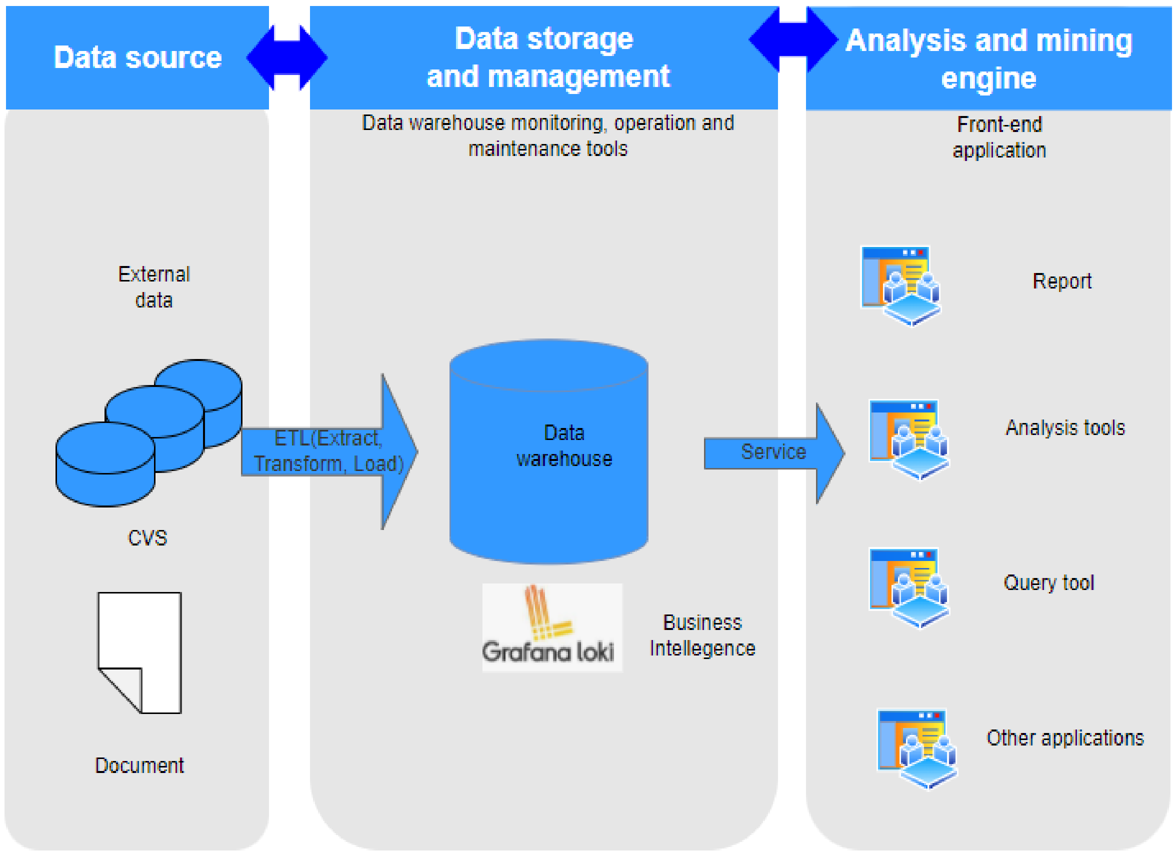

Figure 3: Data warehouse architecture

In most instances, it becomes evident that the data warehouse best lends itself to adopting the star schema model for constructing fundamental data tables. This approach is notable for its ability to significantly enhance query efficiency while introducing a degree of redundancy. The star schema, in particular, is particularly well-suited for bolstering the OLAP (Online Analytical Processing) analytics engine, a characteristic that has garnered significant recognition in research endeavors. This schema is ubiquitously employed in relational databases, such as MySQL and Oracle, especially within the realm of e-commerce database tables. The use of star schema models in data warehousing is a commonly observed practice. In this context, we used a part of the real-time data (Excel, CVS) from Iron International Logistic Group which is size of nearly 6000 GB, a logical data warehouse model was conceived, with a primary focus on establishing a data warehouse for the exhaustive analysis of data from an e-commerce establishment, specifically geared toward assessing goods sales levels. The visual representation of this model is depicted in Figure 4, and it is loaded into the data warehouse. The conceptual underpinning of this model is rooted in dimensional modeling, a concept initially introduced by Kimball. In essence, this approach involves creating a data warehouse that organizes data around a fact table, complemented by dimension tables. The colloquial term often used to describe this method is the "Star schema".

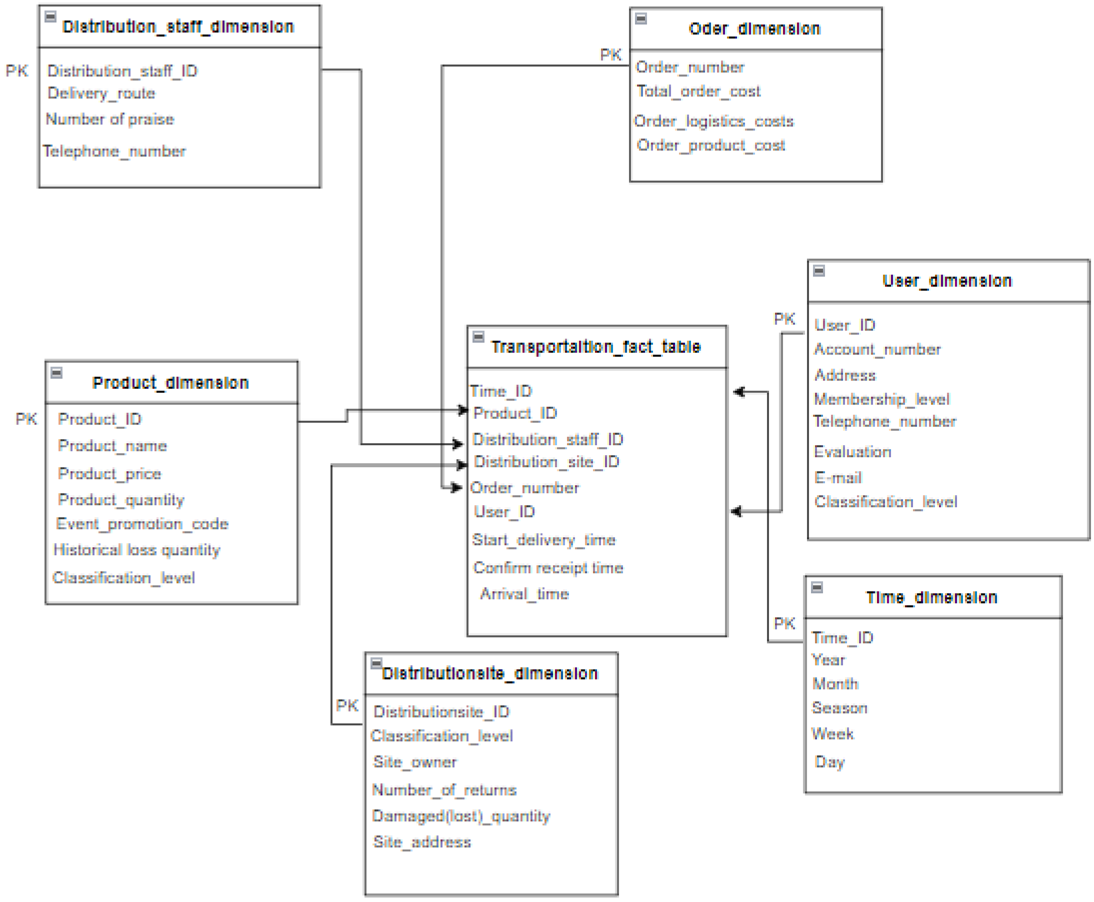

Figure 4: Star schema

We have thoroughly examined the key technologies employed in the establishment of a data warehouse centered on e-commerce logistics. Our primary focus has been dedicated to advancing the development of this e-commerce logistics-based data repository. This involves utilizing the Erwin data modeling tool for structuring the data, leveraging the Talend open studio ETL tool to ready data for the data warehouse, implementing the data warehouse with Oracle SQL Developer, and utilizing the Tableau analytical platform for data analysis and visualization.

As shown in Figure 5, the implementation of dimensional modeling using the Erwin data modeling tool at the very first stage of the methods and technologies used in this article for data Modeling, below describes the implementation of dimensional modeling.

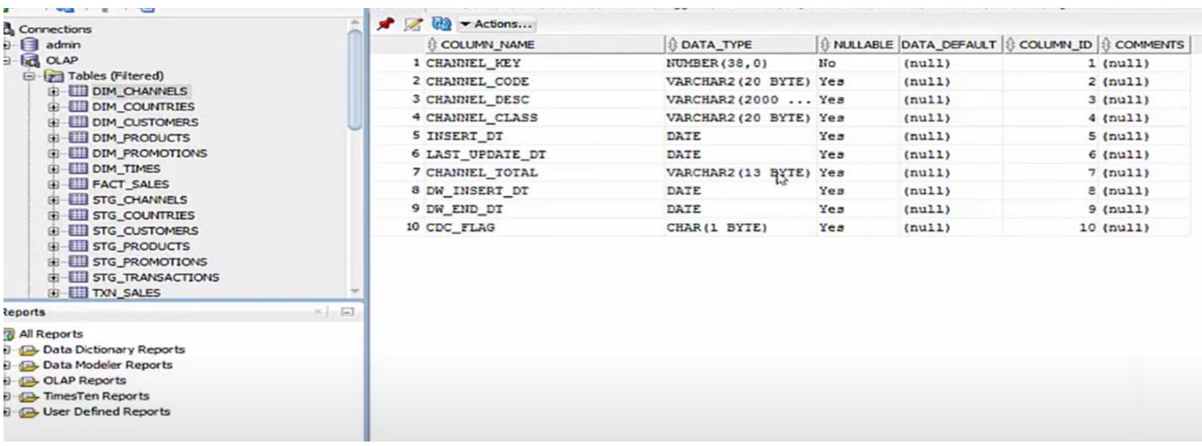

Figure 5: Dimension tables with Erwin

Extract, Transform, Load or ETL is the main data warehouse key [10]. The second stage: the ETL process Talend open studio was also implemented. In loading intermediate tables, dimension tables, and fact tables, we first load all intermediate tables using the same task, with all temporary tables loaded one after the other. We also upload the transaction data coming from the OLTP system to the intermediate tables. so, after loading all the intermediate tables, we gradually carry out the loading of the measurement table one after the other. Figure 6 and Figure 7 describe in detail the progress of the implementation of the ETL process in Talend Open Studio.

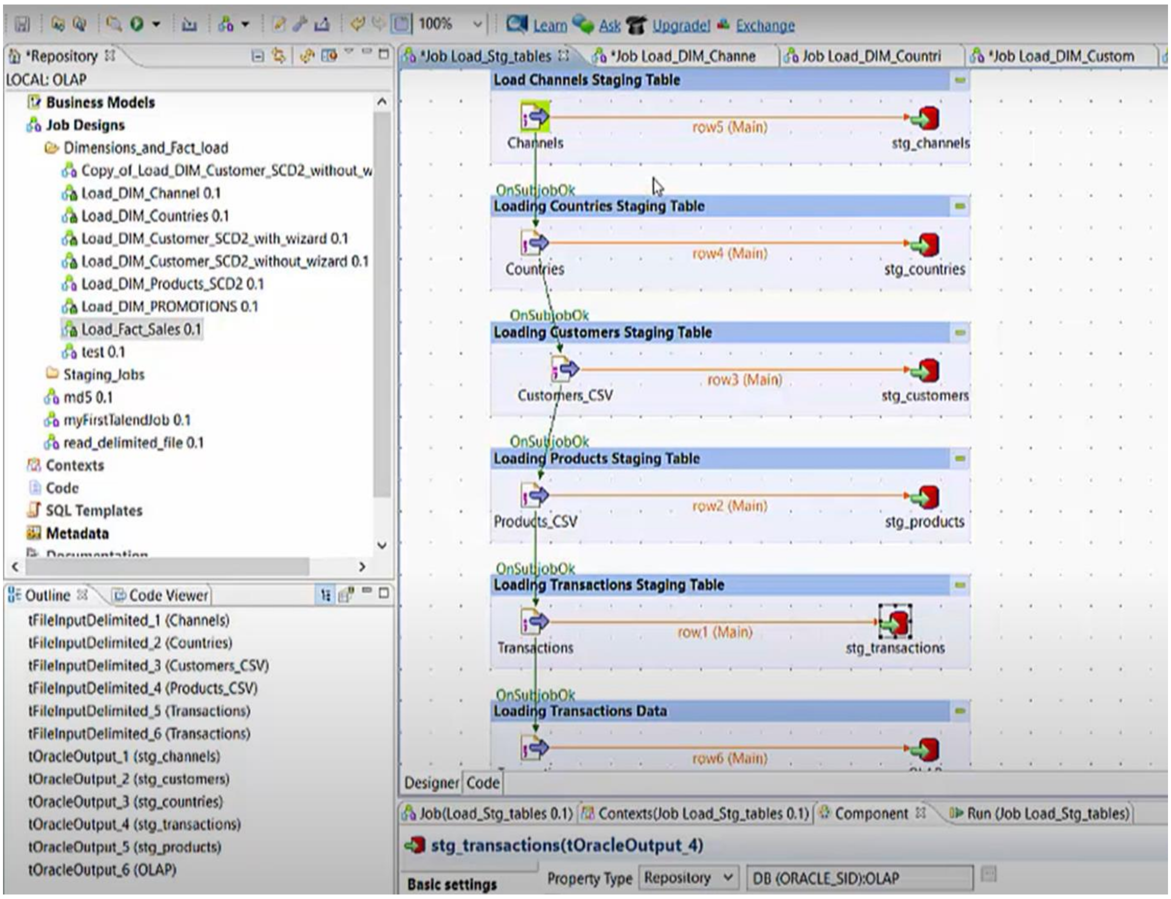

Figure 6: Upload the tables to Talend studio

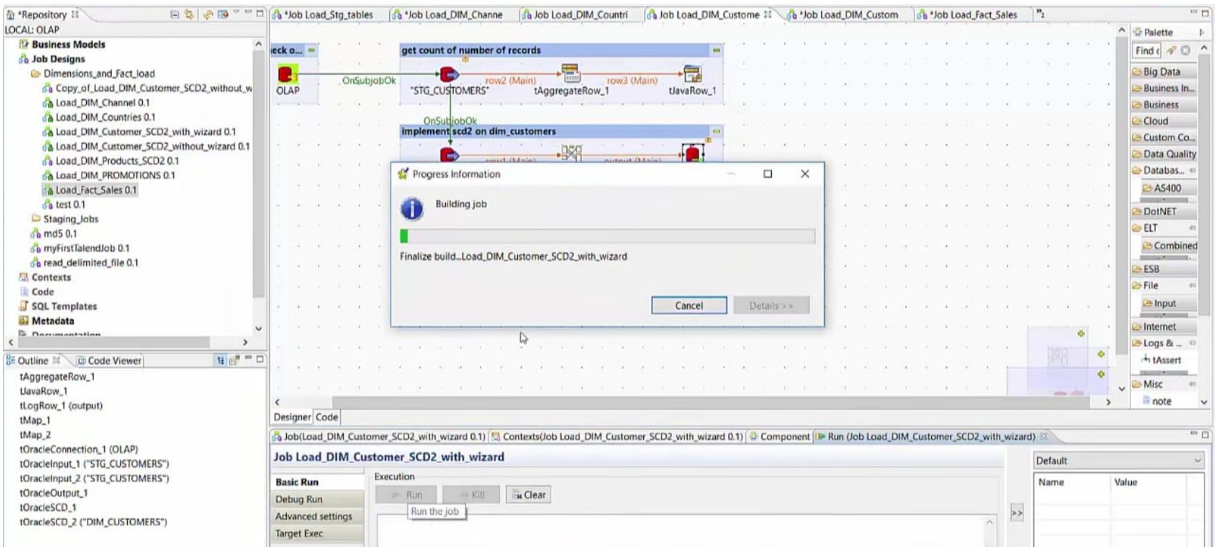

Figure 7: ETL process in Talend Open studio

After the implementation of the ETL process in Talend Open Studio is completed, a connection is made between Talend Open Studio and Oracle SQL Developer. That is, the data from which the integration is made is fully automatically entered into the Oracle SQL Developer. So, the creation of the data warehouse Oracle SQL Developer was also carried out. Figure 8 below describes this created data warehouse.

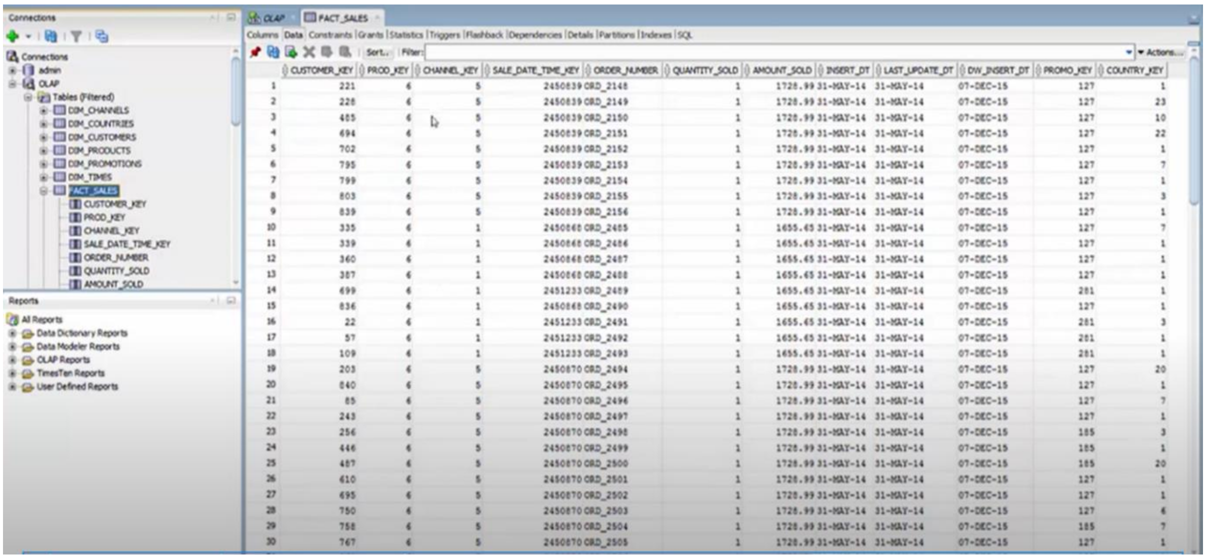

Figure 8: Created data warehouse

We used the Grafana Dashboard to provide a platform for monitoring, visualization, and observability of data from various sources for creating interactive and customizable dashboards that display real-time data metrics, trends, and analytics. In Figure 9, you can see clearly that on the Grafana Dashboard, we made the query for the price average of products. This platform not only makes queries but also can manage all the data that is able to improve the data quality.

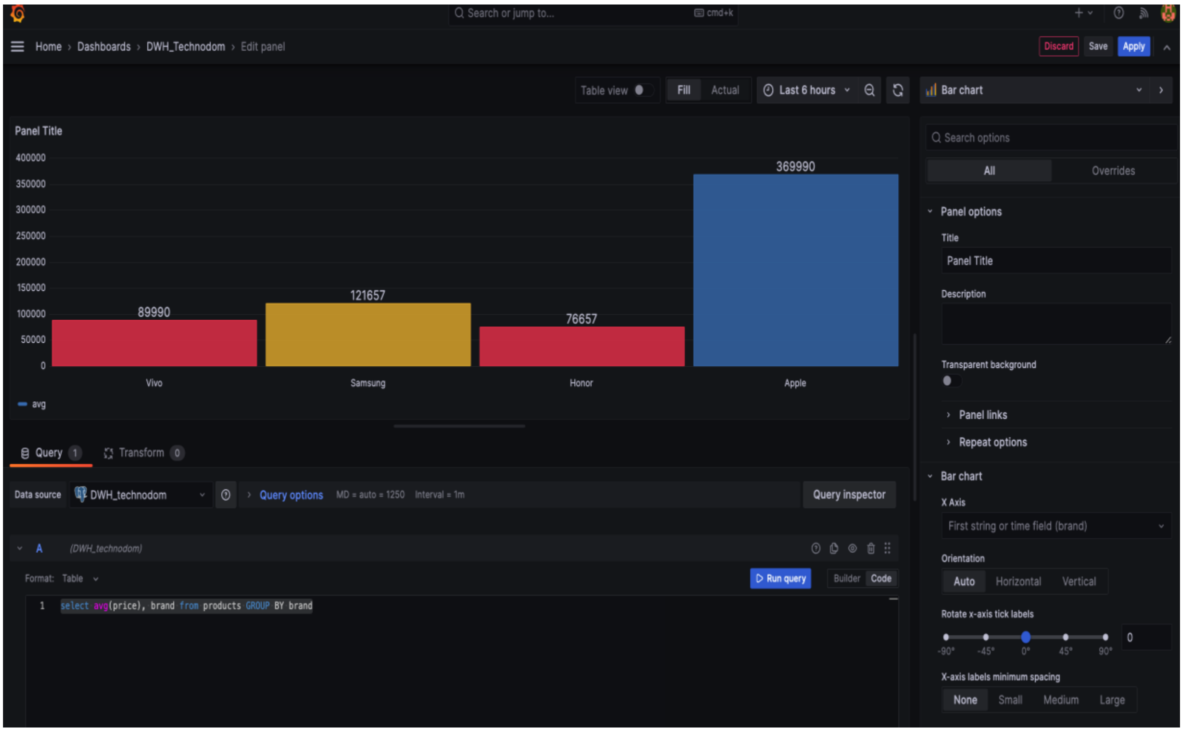

Figure 9: Grafana Dashboard

The most recent stage of the methods and technologies used in my research work: data analysis using Tableau. As we described in the above chapters the relationship between a data warehouse and Business Analytics, it is not enough just to create a data warehouse, business analytics methods are considered as an integral part of the data warehouse, and data analysis in the created data warehouse is also considered an important part of the research work. In the last sections of my thesis, I performed a detailed analysis of the data using the Tableau business analytics method and showed the result of visualization. Figures 10 and 11 below show a visualization of the results of the analyses carried out in Tableau Business Analytics.

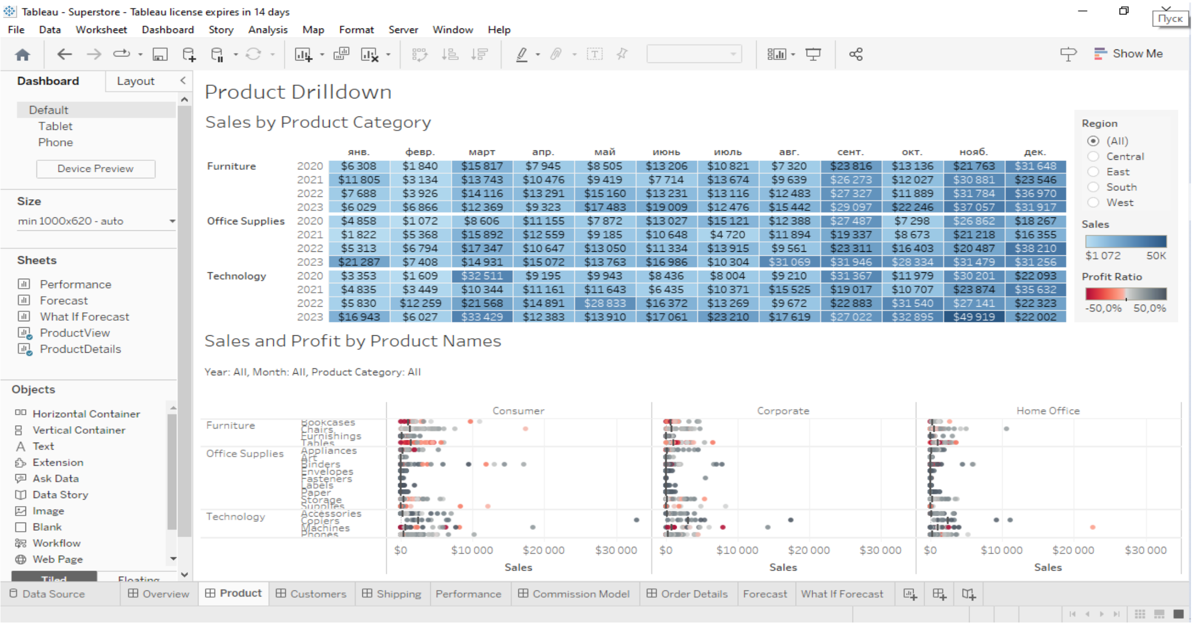

Figure 10: Visualization for sale by product category

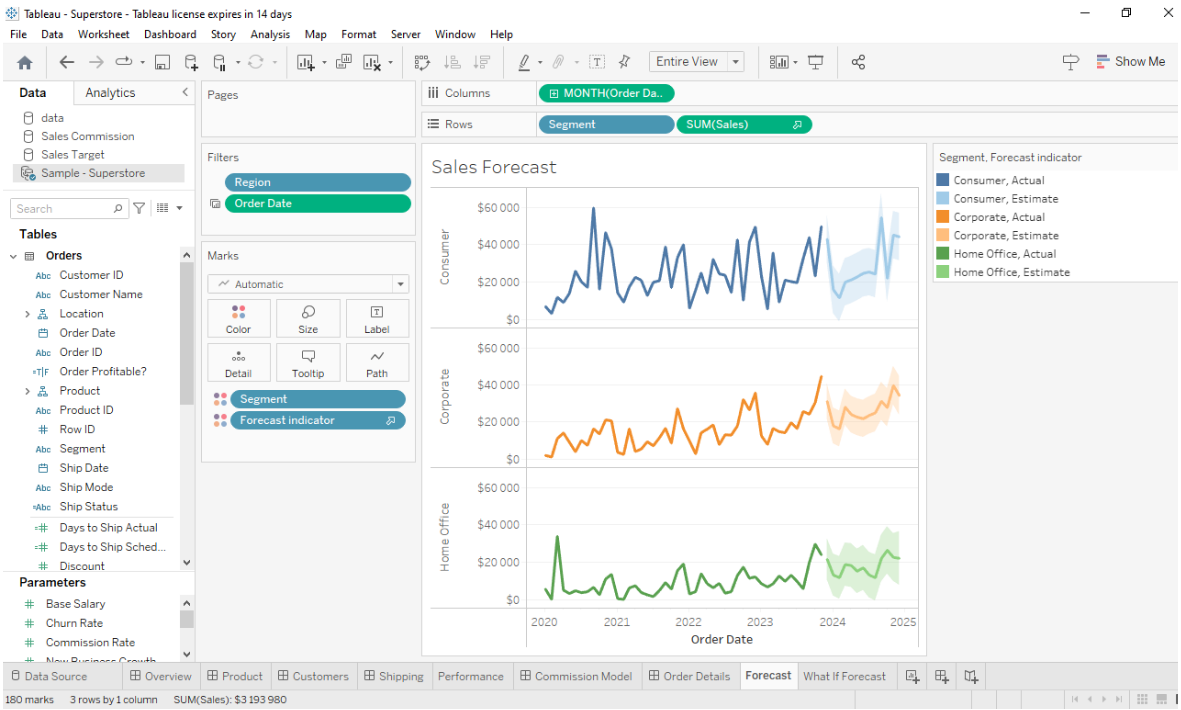

Figure 11: Prediction model for sales

We considered in comparison with the result of the data warehouse architecture classification model studied in the work [11]. The new data warehouse architecture classification model is proposed for better identification, analysis, and comparison in terms of characteristics and structural perspectives. In Table 1, we can see that the architecture of our data warehouse showed a better classification result than others.

Table 1
DWHA Classification Model

Data warehouse classification model		Data warehouse architecture		
		One Layer	Two Layer	Three Layer
Data type	Structured	✓	✓	
	Unstructured			✓
Schema	Single	✓	✓	
	Multiple			✓
ETL	Extract	✓		
	Extract & Load		✓	
	Extract & Transform & Load			✓
Storage	Distributed			
	Aggregated	✓	✓	✓

4. Conclusion

The establishment of a data warehouse within the e-commerce sector serves a fundamental purpose: to empower businesses and organizations with enhanced capabilities for the efficient management of their historical data. This reservoir of information becomes an invaluable resource, driving improvements in decision-making efficacy and overall operational efficiency. Data warehouses are versatile in their support for various types of data analysis, spanning descriptive analysis, predictive analysis, and optimization, all of which are vital components in the decision-making process.

The essence of a data warehouse lies in its capacity to offer a consolidated and dependable repository for the storage and management of data. This unified data hub equips businesses and organizations with the tools to efficiently dissect and interpret data, facilitating the creation of insightful reports.

Within the e-commerce realm, the ever-expanding volume of logistics data is a constant challenge. To address the issues tied to the accumulation and governance of substantial datasets, a data warehouse is conceived. This strategic data repository aims to streamline data control within enterprises, ultimately converting these datasets into valuable information assets. In turn, this transformation fuels the decision-making processes for production and operations, enhancing logistics efficiency and reducing costs. As a consequence, this elevation in operational efficiency and cost-effectiveness concurrently bolsters the competitive edge of the enterprise.

The upgrade of the previous system required a minimum of 68 hours in labor costs for completion. Following the establishment of the data warehouse, not only did the expense of implementing updates decrease, but it also eradicated errors in the routine maintenance process. The update and maintenance of the E-commerce logistics data warehouse system now only demand 7 hours, resulting in a remarkable 92% reduction in maintenance costs.

5. References

[1] T. Gulzat, N. Lyazat,V. Siladi, S. Gulbakyt, S. Maxatbek. (2020). Research on predictive model based on classification with parameters of optimization, Neural Network World 5, pp.295–308, DOI: 10.14311/NNW.2020.30.020.

[2] Bill Inmon Francesco Puppini. (2020). The Unified Star Schema: An Agile and Resilient Approach to DataWarehouse and Analytics Design, Technics Publications, Order 26952 by Kara Joyce on November 19, pp.6-7.
[3] Turken, G., Naizabayeva, L., Satymbekov, M., Abdiakhmetova, Z. (2023). Research and Development of Enterprise Data Warehouse Based on SAP BW. Modeling SIST 2023 - 2023 IEEE International Conference on Smart Information Systems and Technologies, Proceedings, pp.5–9.
[4] Turken, G., Pey, V., Abdiakhmetova, Z., Temirbekova, Z. (2023). Research on Creating a Data Warehouse Based on E-Commerce, SIST 2023 - 2023 IEEE International Conference on Smart Information Systems and Technologies, Proceedings, pp. 16–20.
[5] Suriansyah, Amil Ahmad Ilham, Wahyudi Paundu. (2023). Optimization of Data Warehouse Architecture to Improve Information System Performance, 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), DOI: 10.1109/ICCoSITE57641.2023.10127721.
[6] Praveen Kumar, 2Dr. Kavita, Data Warehouse Concept and Its Usage, Praveen Kumar, 2Dr. Kavita, ISSN: 2581 - 3730.
[7] Junxiu Tang; Yuhua Zhou; Tan Tang; Di Weng; Boyang Xie; Lingyun Yu. (2022). A Visualization Approach for Monitoring Order Processing in E-Commerce Warehouse, IEEE Transactions on Visualization and Computer Graphics. 28(1), DOI: 10.1109/TVCG.2021.3114878.
[8] Paquita Putri Ramadhani; Setiawan Hadi; Rudi Rosadi, Implementation of Data Warehouse in Making Business Intelligence Dashboard Development Using PostgreSQL Database and Kimball Lifecycle Method, 2021 International Conference on Artificial Intelligence and Big Data Analytics, DOI: 10.1109/ICAIBDA53487.2021.9689697.
[9] Jakub Belter, Marek Hering, Paweł Weichbroth. (2023). Motion Trajectory Prediction in Warehouse Management Systems: A Systematic Literature Review, Journal of "Human-Computer Interaction: Challenges, Opportunities and Emerging Developments, MDPI, *13*(17), 9780; https://doi.org/10.3390/app13179780.
[10] Suriansyah B, Suriansyah B, Suriansyah B. (2023). Optimization of Data Warehouse Architecture to Improve Information System Performance, 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), DOI: 10.1109/ICCoSITE57641.2023.1012772.
[11] Qishan Yang1, Mouzhi Ge and Markus Helfert. (2019). Analysis of Data Warehouse Architectures: Modeling and Classification, In Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019), pp.604-611, DOI: 10.5220/0007728006040611.
[12] Qishan Yang, A thesis: Benchmarking Data Warehouse Architectures A Feature-based Modelling and Evaluation Methodology, September 2019.