Design of Heterogeneous Data Warehouse Architecture for Supply Chain Management System

Ruiqin Lin[1], Wenan Tan[2], Pan Liu[3] and Lu Zhang[1]

[1]*College of Resources and environmental engineering, Shanghai Polytechnic University, Jinhai Road, Shanghai, China*

[2]*College of Computer and Information Engineering, Shanghai Polytechnic University, Jinhai Road, Shanghai, China*

[3]*Information Technology Center, Shenzhen Easttop Supply Chain Management CO.,LTD., Zhouhai Road, Shanghai, China*

Keywords: Data Warehouse, Data Quality, Data Integration, Big Data Development.

Abstract: Multiple systems are covered by the supply chain management system portal website, which includes tasks such as basic data entry, bill of lading generation, customs declaration, transportation, expenditure settlement, statistical reports, and more. Each system is spread among several departments, and data is kept by numerous departments at various phases. The storage format and semantics are vastly different. Data is complicated and varied, and data quality is challenging to ensure during the data integration process. Data is frequently lost, and storage types are incompatible with one another. As a result, appropriate technical solutions are required to ensure the supply chain management system's data quality after data integration. This article focuses on the current state of the supply chain management system data and the issues that it faces. The business requirements for the unified and standardized storage of supply chain management system data are derived from the description of the challenges. Finally, it situates the primary issues discussed in this article within the framework of this company.

1 INTRODUCTION

1.1 Background in the Industry

The supply chain industry's informatization has been strengthened internally as a result of the rapid development of Internet technology, and industry data has shown explosive growth (HUANG, 2021). Massive data contains enormous value, and how to mine these values more effectively and quickly has steadily become the focus of data owners' attention. The information system's basic data is a valuable resource that has a significant impact on the enterprise's economic development and management, and serves as the foundation for scientific management and decision-making. Currently, most supply chain management systems spend a significant amount of money and time developing online transaction processing OLTP business systems and office automation systems to record various transaction processing related data. These data have a lot of commercial worth. Enterprises did not make the best use of their existing data resources, wasting more time and money while also missing out on the best opportunity to make critical business decisions. Most traditional data warehouses are still in use in the business, and the majority of existing supply chain management systems are built using old methods, such as acquiring pricey large-scale servers. Database fragmentation divides the data on this basis. The data is stored on a disk array, which makes system growth and upgrading more difficult and expensive, and the entire system is tightly coupled, making it impossible to meet the demands of high efficiency, dependability, and economy. As a result, figuring out how to turn data into information and knowledge via various technical means has become a major barrier in improving the company's fundamental competitiveness. ETL technology is the most important technological tool among them.

1.2 Scenario of a Business

The data from the supply chain management system is spread across several departments. Different departments are in charge of developing, managing, and maintaining various enterprises, and different departments keep track of various basic business data. Basic information is frequently kept and defined

Lin, R., Tan, W., Liu, P. and Zhang, L.
Design of Heterogeneous Data Warehouse Architecture for Supply Chain Management System.
DOI: 10.5220/0011183600003440
In *Proceedings of the International Conference on Big Data Economy and Digital Management (BDEDM 2022)*, pages 441-445
ISBN: 978-989-758-593-7

independently. Each department's basic data is like a data island that can't be linked to other departments' data. This type of data island manifests itself mostly in physical and logical features. Different business departments understand and define data at their own business level, and there is no linkage communication; data is stored and maintained independently in different departments, and there is no linkage communication; data is stored and maintained independently in different departments, and there is no linkage communication; data is stored and maintained independently in different departments, and there is no linkage communication. Different departments' data semantics differ, resulting in a disparity in data semantics that subtly increases the difficulty of data cooperation and data association between departments. Figure 1 shows the association approach, however these push data packets are frequently associated and related. The data push industry is complicated by this highly connected, low coherent, and inaccurate data representation. Department-based data, considerable autonomy, and a lack of cohesive management ideas characterize the current data push situation. There is a danger of data inconsistency because information on the present stage of the project cannot be obtained centrally and must be gathered from multiple departments of numerous firms.

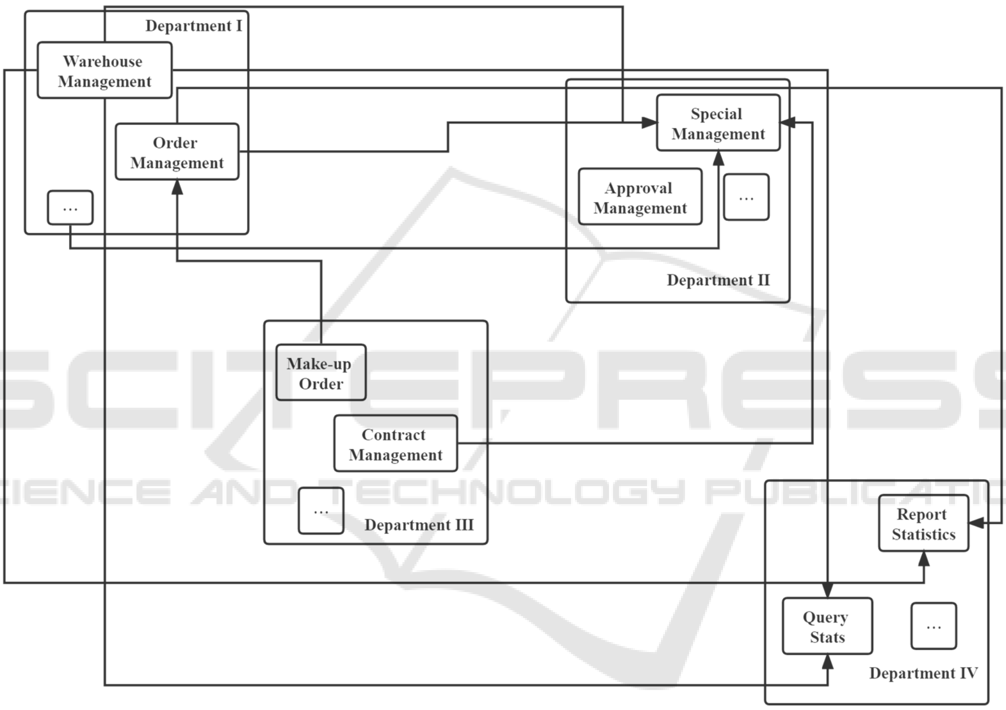

Figure 1: The original way of business data association.

1.3 Scholar of Research

The creation of a new type of hierarchical data warehouse has become a current research hotspot with the onset of the big data era (Di Tria, et al., 2014). To collect, process, and store unstructured data, Wu Wentai (Wu, et al., 2018) and others employed Hadoop technology. The challenge of enterprise data processing is handled by combining big data fusion technology with a data warehouse. (Zhang, 2017) evaluated Hive technology's logistics data warehouse and offered a specific logistics data warehouse deployment plan. Simultaneously, as corporate complexity has increased, big data development has evolved from just performing computations on data to being more standardized and process-oriented. (ZHANG, 2017, Dongya, 2017, Jinghua, 2017)

1.4 Solution

As a result, it is required to upgrade the current state of data in the supply chain management system in order to make data easier to gather, store, interpret, and value. Data storage can be standardized, and data representation standards can be specified consistently for logical data islands. Physical data islands are being phased out in favor of a unified supply chain management system data warehouse and the application of ETL technologies to the supply chain management system data integration company to complete the supply chain management system. Data integration is used to store data information centrally and offer initialization data for each system. The coupling of data between various departments may be decreased, the interdependence of multiple departments can be reduced, and the data push process can be simplified by establishing a data warehouse. Figure 2 depicts the new and better business data association condition.

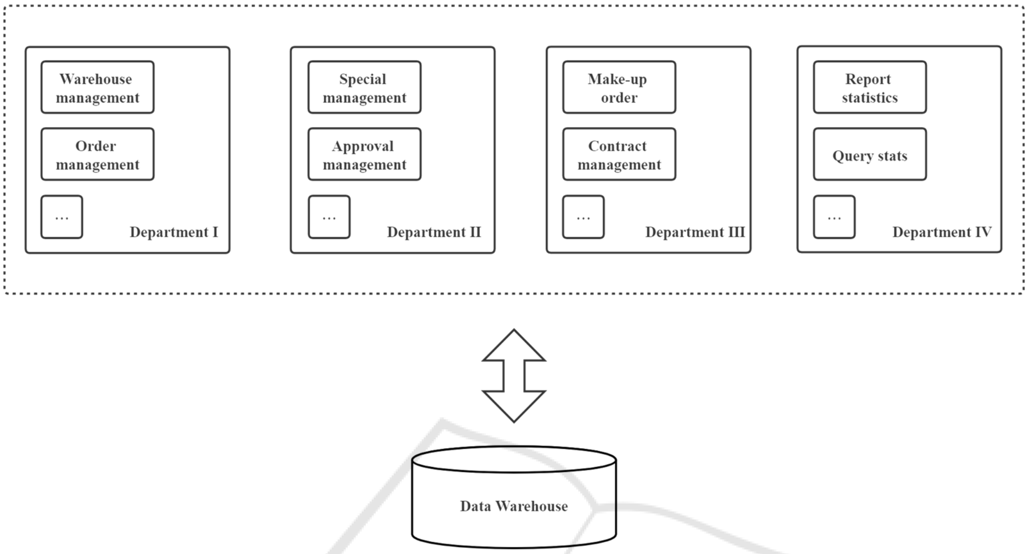

Figure 2: The new and improved business data association situation.

2 MANUSCRIPT PREPARATION

Because of the characteristics of big data, such as the rapid increase in data volume, various and complex data types, extremely low data value density, and fast processing speed, traditional data warehouse architecture has become increasingly unable to meet actual needs in the context of big data applications. This work builds a set of data warehouse architecture against the backdrop of big data on the basis of traditional data warehouse architecture, as shown in Figure 3, to address the unified storage and administration requirements of supply chain management system data.

In-depth research of business demands, in-depth understanding of the company's business, and the establishment of an effective model are among the issues to be resolved. Data warehouse modeling ensures the rationality of the data warehouse storage structure, efficient data query efficiency, and storage space savings by consulting relevant literature, analyzing cutting-edge hotspots, and implementing technical routes; data warehouse modeling, the storage structure of each information system is different, data warehouse modeling ensures the rationality of the data warehouse storage structure, efficient data query efficiency, and storage space savings by consulting relevant literature, analyzing cutting-edge hotspots, and implementing technical routes; Reduce data loss; ETL tools are used in the process of implementing ETL, the ETL process is designed, and the data is loaded into the data warehouse; ETL tools are used in the process of implementing ETL, the ETL process is designed, and the data is loaded into the data warehouse.

The data acquisition layer, data processing layer, data storage layer, and data application layer are the four tiers of the architecture. The data warehouse has subject-oriented and integrative properties. To face a choice, data integration is used to solve a specific subject query and visualize the analytical results. As a result, the foundation of a data warehouse is high-quality data. Through layered management, the data warehouse architecture realizes the step-by-step completion of work, and the processing logic of each layer becomes easy.

2.1 Data Acquisition Layer

To handle company data and offer raw data for data warehouses, employ relational databases. This is the step in which the source data is acquired. Other types of data, such as database script data, text data, and so on, are collected using open source tools, Sqoop, or program codes. The Oracle data source stores all data information; this layer is the foundation of the overall architecture.

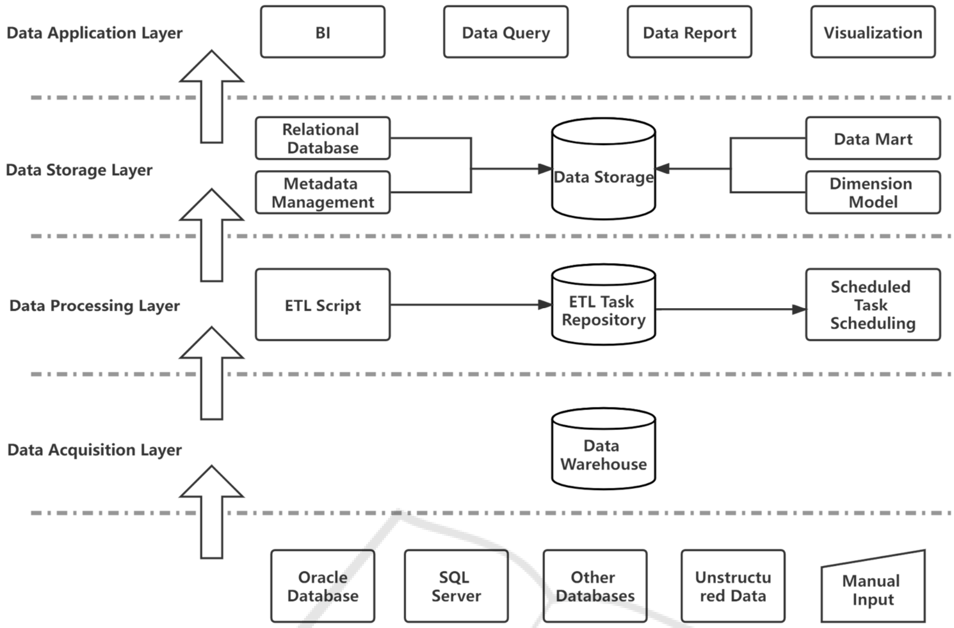

Figure 3: Design of a heterogeneous data warehouse architecture.

2.2 Data Processing Layer

Pre-processing, pulling distant mapping data using scripts, and then querying and analyzing the data according to the system's business requirements are all required before doing data analysis on the original data stored in the data warehouse. The code execution time consumption will be recorded as a log log during the full job procedure. The automatic program will evaluate this log, and the start and end times of various programs will be emailed to the management and maintainer. At the same time, the query analysis and processing result data will be exported to the Mysql database. These are the data to which we must pay attention in order to assess the quality of later data integration. This procedure involves performing ETL operations on the obtained source data, which includes extraction, conversion, and loading. The ETL process can be described as an ETL task script, which is stored in the ETL task resource library in the form of metadata, and at the same time Combine Hadoop, Spark, Zookeeper, and other big data processing technologies to improve task scheduling and monitoring; the ETL process can be described as an ETL task script, which is stored in the ETL task resource library in the form of metadata, and at the same time Combine Hadoop, Spark, Zookeeper, and other big data processing technologies to improve task scheduling and monitoring. The data processing layer is in charge of the data warehouse architecture's key tasks. Not only must the data quality be guaranteed, but the efficiency of ETL task scheduling and resource usage must also be increased as much as feasible during this process.

2.3 Data Storage Layer

Designing dimensional models and data marts can help to support and improve data storage. The data storage model divides ETL task levels while providing data storage. It can also handle distributed data storage and relational database storage, such as Hbase and Hive, as well as integrated data migration.

2.4 Data Application Layer

Front-end data analysis and interface display, such as multi-dimensional analysis and report visualization, are supported. The application layer performs all of the activities of the Web display module, follows the B/S structure, creates J2EE projects, selects the MVC mode, and uses the Web terminal to display the data in the form of charts. Before and after, the data is transmitted using ajax asynchronous transmission, and the data is in json format.

3 RESULTS & DISCUSSION

To satisfy the current requirements for big data integration and migration storage, the data warehouse integrates associated big data processing architectural technologies. This article starts with the construction of a supply chain management data warehouse and divides it into two sections: data integration under the data processing layer and ETL task scheduling and monitoring, which includes data quality assurance technology after data integration and the study of ETL task script scheduling execution strategies.

4 CONCLUSIONS

This chapter primarily introduces the current specific business background of supply chain management data, as well as the various problems encountered, and then constructs a supply chain management data warehouse architecture against this backdrop, before describing and analyzing the architecture's hierarchy.

ACKNOWLEDGEMENTS

This work was financially supported by the National Natural Science Foundation of China (61672022, U1904186), Shanghai Polytechnic University Key Discipline Electronic Information Special Master Program Project (XXKZD1604). At the same time, I'd want to express my gratitude to my teachers, Wenan Tan and Pan Liu, at this time. They provided me with numerous scholarly and helpful comments and assisted me in revising my work during the writing process. He also gave me the opportunity to practice teaching in addition to these things.

REFERENCES

Di Tria, F., Lefons, E., & Tangorra, F. (2014). Design process for big data warehouses. In 2014 International Conference on Data Science and Advanced Analytics (DSAA) (pp. 512-518). IEEE.

HUANG B. (2021). Research on Intelligent Transformation of Logistics Industry in the Era of Big Data. J. Journal of Technical Economics & Management. 12, 118-121.

Wu, W., Lin, W., Hsu, C. H., & He, L. (2018). Energy-efficient hadoop for big data analytics and computing: A systematic review and research insights. Future Generation Computer Systems, 86, 1351-1367.

ZHANG R. (2017). Based on the Hive of data warehouse logistics research and design of the big data platform. J. Electronic Design Engineering. 25(09), 31-35.

ZHANG, Q., Dongya, W. U., & Jinghua, Z. H. A. O. (2017). Big data standards system. Big Data Research, 3(4), 11.