EasyChair Preprint

№ 11695

Elastic Data Warehousing: Adapting to Fluctuating Workloads with Cloud-Native Technologies

Sina Ahmadi

January 6, 2024

ISSN: 2959-6386 (Online), Vol. 2, Issue 3, December 2023

Journal of Knowledge Learning and Science Technology

journal homepage: https://jklst.org/index.php/home

Elastic Data Warehousing: Adapting To Fluctuating Workloads With Cloud-Native Technologies

Sina Ahmadi [1]
[1] Independent Researcher, United States

Abstract

This research focuses on the development of elastic data warehousing while adapting to changing workloads with the help of cloud-based technologies. The traditional methods of data warehousing need innovative and creative strategies in order to improve their efficiency. Thus, this research focuses on analyzing innovative methods which can improve the future of data warehousing, such as machine learning, encryption, artificial intelligence, etc. Moreover, the study also focuses on specific industries that require customized solutions to data warehousing. These include the manufacturing, finance, and healthcare industries. The study uses qualitative data gained through a comprehensive review of literature. The findings reveal a great level of significance of modern data warehousing techniques that assist in improving the overall efficiency of traditional methods.

Keywords: Data Warehousing, Artificial Intelligence, Data Integration, Automation, Cloud Native..

Article Information:
Article history: *Received:* 28/08/2023 *Accepted:* 25/09/2023 *Online:* 20/10/2023 *Published*: 12/12/2023
DOI: https://doi.org/10.60087/jklst.vol2.n3.p301
[i]**Correspondence author: Sina Ahmadi** **Email:** sina0@acm.org,

Introduction

Elastic data warehousing can be defined as a process that enables a database to adapt to the growth in data rate, data volume and concurrency without affecting the overall performance. Such a data warehouse plays a vital role in providing dedicated resources for undergoing different business operations within the organization and managing all the tasks [1]. In this digital world, data warehousing has been advanced with the help of advanced and modern technologies like clout-native technologies that are helpful in managing workloads effectively and efficiently. These technologies play a vital role in providing flexibility and scalability in the organization. The purpose of this research study is to explore the role of cloud-native technologies in enhancing elastic data warehousing.

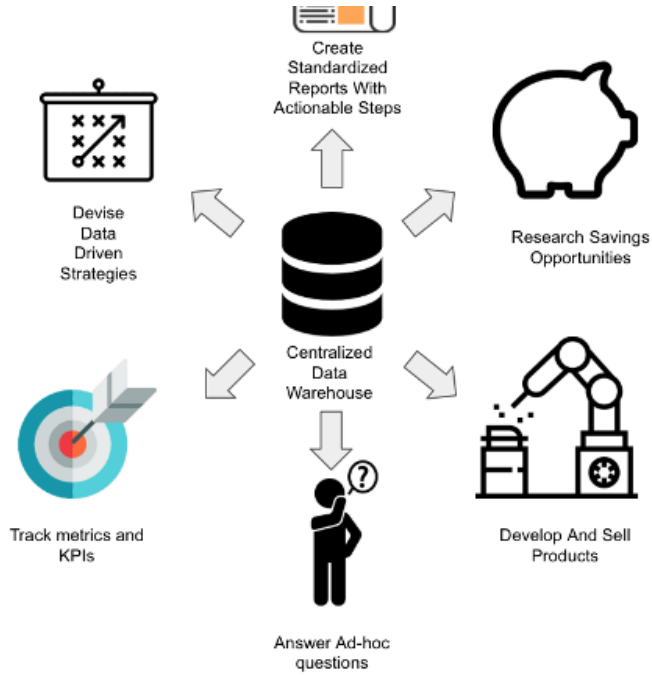

Figure 1: Data Warehouse Using Cloud Computing [2]

In this digital world, organizations have to go through impulsive workloads that necessitates the use of data warehousing strategies for the purpose of workload management. Elasticity is the key of cloud-based technologies that allows the organizations to make their internal and external structure strong and meet the changing demands of the organizations. This research aims to go through cloud-based technologies such as containerization that play an important role in shaping the business environments in terms of managing data. With the help of such tools and technologies, organizations can ensure enhanced performance, efficiency in business operations and cost effectiveness to meet the requirements of data collection, management and maintenance.

1. Literature Survey

1.1. Challenges in Traditional Data Warehousing for Fluctuating Workloads

Data warehousing encounters many challenges like scalability and speed limits, unreliable results, high costs, inflexibility, issues with data management, etc. Research conducted by [3] in this regard, focused on the challenges and issues with data warehousing, mainly in terms of big data. According to the research, big data is characterized by the aspects of value, veracity, volume, variety, and velocity. Traditional methods of data warehousing are usually based on multidimensional types of data. However, the use of new data types is different from traditional methods of data warehousing. This creates difficulty in managing the workload. Thus, every company needs big data analytics to overcome this issue.

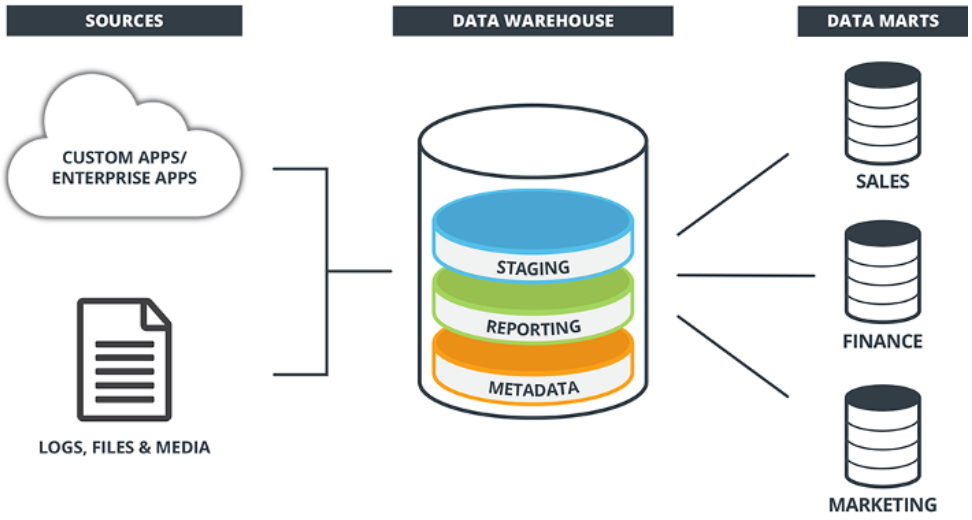

Figure 2: Architecture of a Traditional Data Warehouse [4]

Another research conducted by [5] focused on the challenges encountered in big data warehousing. The research showed that due to the dynamic and complex nature of supply chains, today's firms need proper data solutions that can help integrate their big data sets and manage big data analytics. This helps them ensure that proper protective efforts are made rather than reactive ones. The researchers also developed a proof-of-concept BDW (Big Data Warehouse) that can help companies manage data requirements in the area of data warehousing. Overall, the use of data-driven, user-focused, and goal-driven strategies is necessary to manage the issues encountered in data warehousing.

1.2. Elasticity in Data Warehousing: Conceptual Framework

The elasticity aspect in the area of data warehousing is very important since it helps in provisioning server resources when required. Then, it reallocates all the resources in order to manage the workload. Overall, this elasticity helps ensure the high scalability and performance of data warehousing systems while simplifying the aspects of capacity tweaking and planning. The same aspect was explored by [6], who focused on managing and maintaining elastic data warehousing. According to the researchers, traditional data warehousing systems could be more flexible to manage resource control. This restraints both the users and the cloud providers to manage the desired quality of service and optimize the total cost.

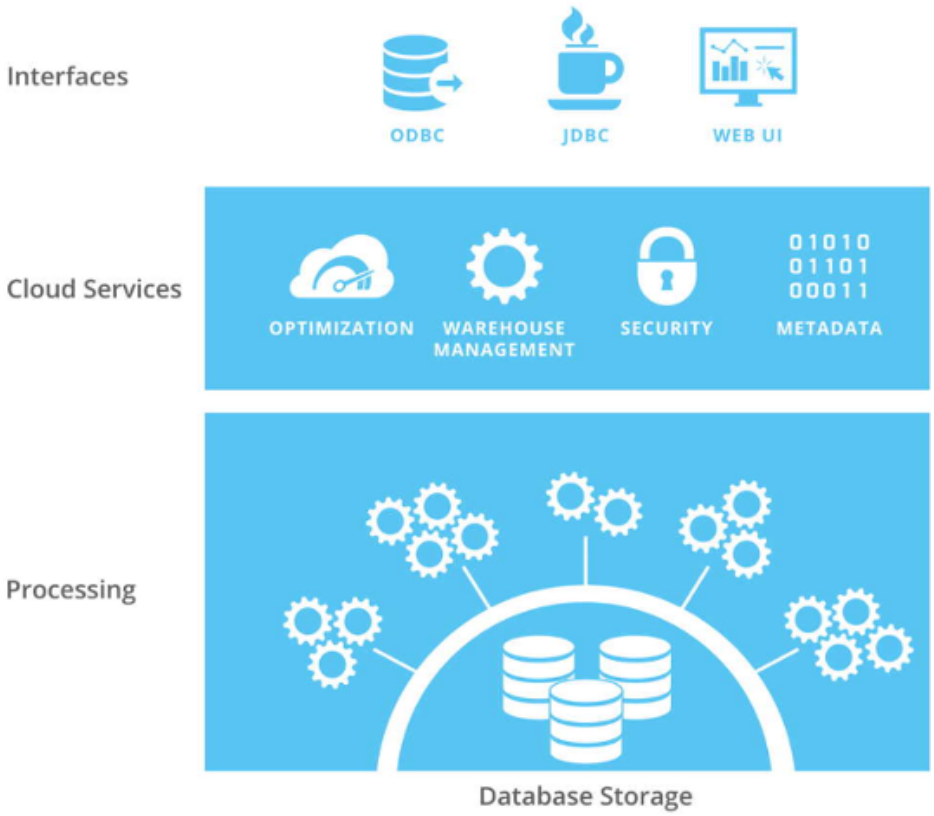

Figure 3: Elastic Data Warehouse [7]

Another study was conducted by [8], which focused on the elasticity of data warehousing. According to the research, an elastic data warehouse has the ability to quickly adapt to growth in concurrency, data volume, query intensity, and data rate. Moreover, it can also provide the required resources for diverse workloads and utilize additional resources whenever needed. Moreover, their elasticity also helps them easily adapt to fluctuations in content and data structure without needing any new data storage structures, additional data migration, and schema redesign.

1.3. Key Cloud-Based Technologies for Elastic Data Warehousing

Different cloud-based technologies can be used in elastic data warehousing. Cloud data warehouses can help by using powerful computing methods that support streaming data. In this way, they help make queries in real-time. As a result, users can use and access data much faster than with traditional methods of data warehousing. This allows the users to gain more accurate and correct insights faster and make better business decisions. Research by [9] was also conducted in this regard. According to the research, cloud-based technologies in data warehousing can provide services in an open network [6].

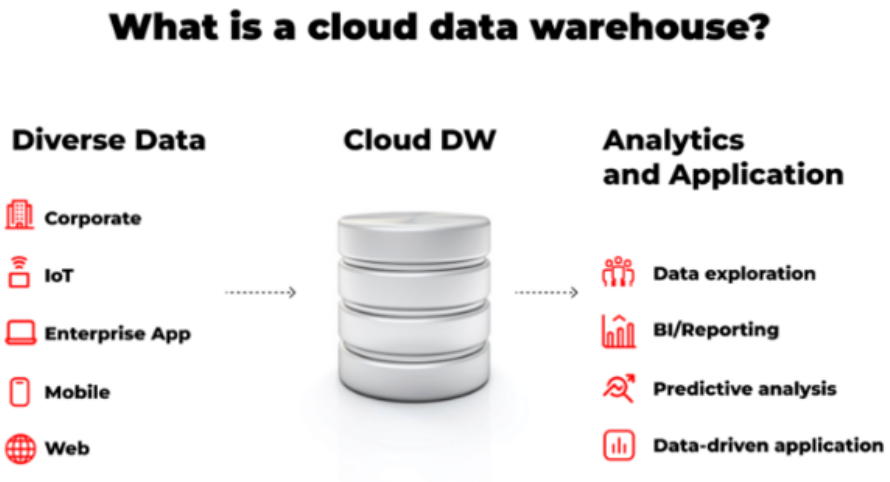

Figure 4: Cloud-based Data Warehouse [10]

Today, many cloud-based technologies have been developed in the area of data warehousing. These include Amazon Redshift, Azure Synapse Analytics, Google BigQuery, Snowflake, etc. All such data warehouses ensure high levels of scalability, security, and cost-effectiveness. The use of such cloud-based systems can help in improving the overall efficiency of data warehousing. Arora and Gosain (2021) also focused on this aspect. According to the researchers, an enhanced encryption model can be used in this regard to secure data warehouses [8].

1.4. Scalability and Resource Management in Elastic Data Warehousing

The aspect of scalability mainly relates to the ability of a data warehouse to manage large volumes of unstructured, structured, or partially structured data. On the other hand, resource management is a kind of managerial activity that uses different data management tools in order to manage the data resources of a firm in order to meet its organizational objectives. One such research was held by Agiwal et al. (2021) who focused on developing a scalable data warehouse to improve the performance of Google. The researchers mainly developed Napa, which is a kind of analytical data management system that enhances the scalability of warehouses at Google [9]. It also facilitates robust query performance and offers a great level of flexibility.

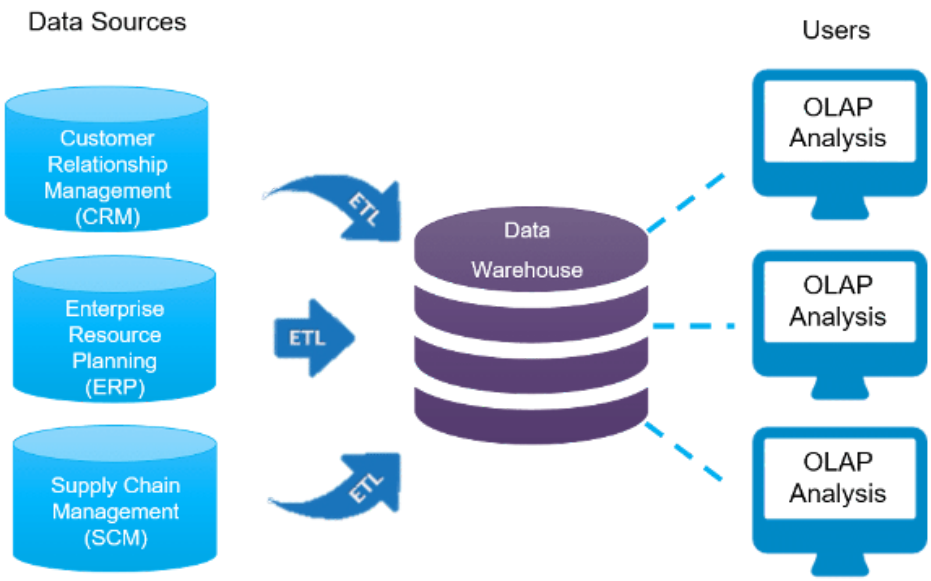

Figure 5:Resource Management in Data Warehousing [10]

Resource management is also a very important activity in data warehousing. This is because the adequate management of resources ensures that the system runs smoothly without any operational difficulties. It also helps in enhancing the visibility of the data assets of a firm, which makes it easier for individuals to confidently and quickly find accurate data for analysis. This data visibility in turn helps a firm in becoming more productive and organized, which assists the workers in finding the data they require to perform their jobs in the best way. According to Bogojevic (2020), the aspect of resource management in the arena of data warehousing can help in managing costs, size, and time constraints [11]. Overall, the use of resource management techniques can assist in meeting the requirements of data warehousing.

1.5. Best Practices for Designing and Managing Elastic Data Warehousing Systems

In order to manage flexible data warehouses, it is important to integrate different efficient practices. For instance, a firm should always align its objectives within each department in order to ensure smooth data warehousing. Moreover, they should also develop some disaster recovery protocols in case of data breaches. The first and foremost technique is to choose the most accurate warehouse architecture. The next step is to focus on the aspect of data cleaning. According to Huang et al. (2020), the use of blockchain-based data warehousing systems can also prove to be highly beneficial in this regard [12]. Such systems can help a firm in improving its decision-making procedure while cutting extra costs. Therefore, such systems should be implemented by firms operating complex data warehouses.

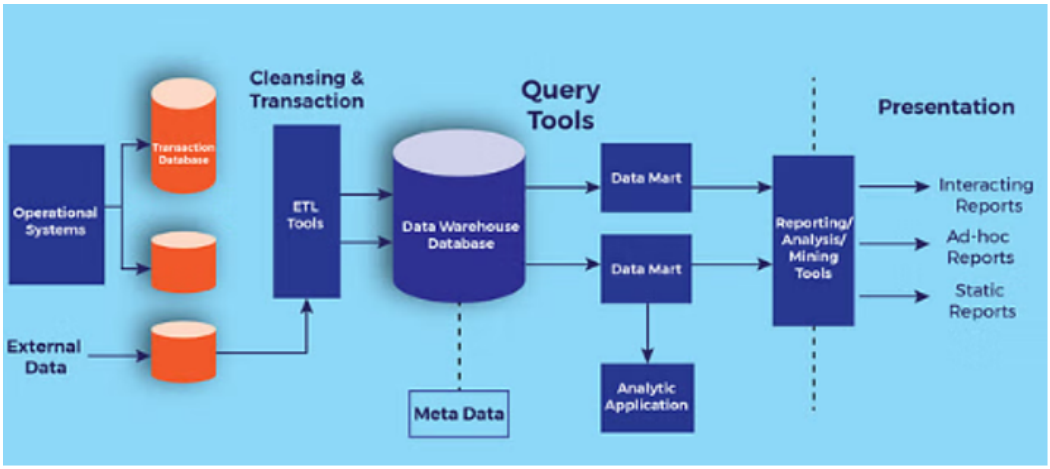

Figure 6: Designing and Managing Data Warehousing [13]

Another important way is to manage the process flow in data warehousing. This procedure is mainly based on four steps. The first step focuses on extracting and loading the required data. The second step is to clean and transform the data. This is very important as it ensures the smooth transformation of data into the data warehouse. The next step focuses on creating a backup for the data and archiving it. The last step is to manage the queries and direct them to the specified data sources. Garani et al. (2019) also focused on this smooth procedure of managing the process flow. The researchers emphasized the importance of ensuring the smooth transformation of data into the warehouses [14].

1.6. Future Trends and Innovations in Elastic Data Warehousing

The area of data warehousing will experience many innovations and creative trends. For instance, artificial intelligence is expected to play a very crucial role in improving data warehousing. The incorporation of natural language processing, machine learning, and many other techniques will make the current data warehouses more effective and intelligent at interpreting large volumes of data. Pandian (2019) also focused on proposing the use of AI in a smart warehousing environment. The researcher stated that it is important to modify the current operations in data warehousing with the help of innovative techniques like AI [15]. It can help in enhancing the potential of functioning of data warehouses in the areas of coordination, management, and logistics.

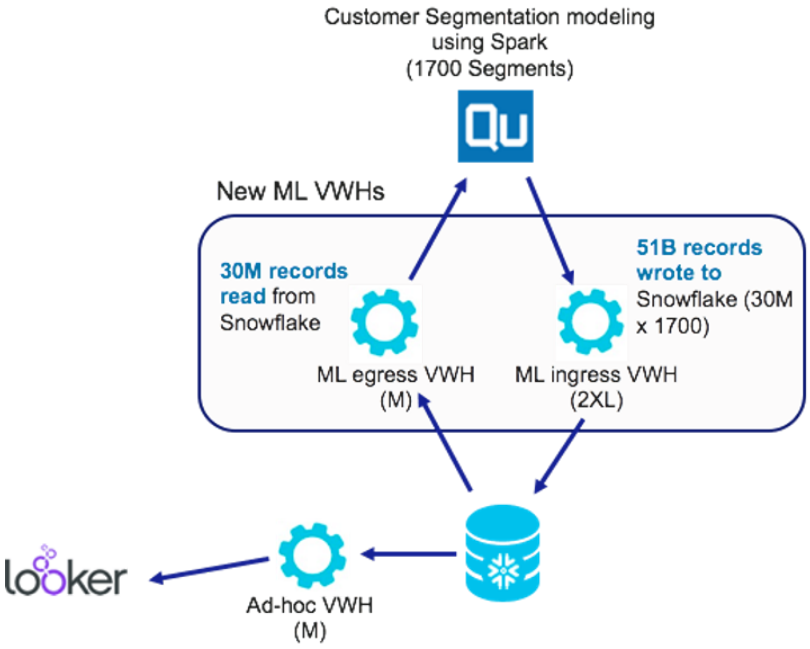

Figure 7: Machine Learning in Data Warehousing [16]

Another important trend that is likely to emerge in data warehousing is zero-copy data sharing. This trend will help in reducing the costs and risks associated with traditional methods. One platform trying to offer this feature is Snowflake. This method helps firms in sharing read-only database items with other firms without transferring the real data. The major advantage in this regard is the separation of compute and storage. According to Bell et al. (2021), the Snowflake data cloud can help in bringing huge innovation in the arena of data warehousing [17].

2. Problem Definition

2.1. Challenges in Traditional Data Warehousing

Modern organizations need a high level of scalability and agility but they both lack in traditional data warehousing especially when it comes to the security industry [18]. It is obvious that it is difficult to adjust according to the evolving changes and implementing them is also crucial just because of the rigid architecture. Thus, high speed and scalability are necessary to perform error-free operations within an organization. Another challenge regarding traditional data warehousing is the inefficiencies with non-technical users. There is a need for data teams that could manage and fulfill all the requests related to the data. Thus, it results in inefficiency and delays in tasks.

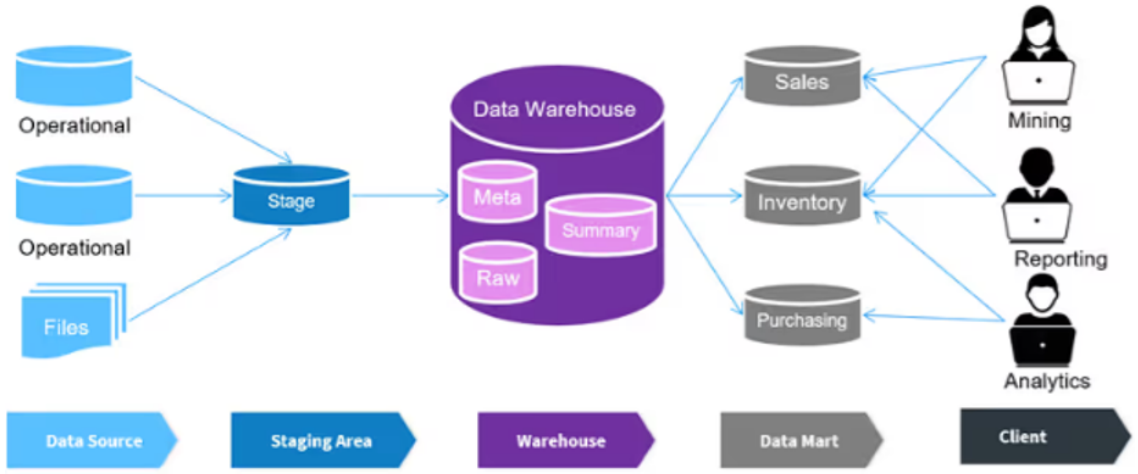

Figure 8: Challenges in Traditional Data Warehousing [19]

2.2. Lack of Agility in Adapting to Evolving Business Needs

The agility of a business network is the most important thing to consider when it comes to meeting the market needs and demands. In this modern world, industries are becoming data-centric where there is a need to pay a great attention to data management and maintenance for managing business workloads effectively. For this purpose, cloud-based technologies must be implemented to grow a business and make it competitive because traditional data warehouses lack agility [20].

2.3. Inflexibility of Traditional Data Warehouses

It is stated that modern organizations must be flexible enough to adjust according to the changing market trends and demands. There is inflexibility in traditional data warehouses because of their rigid structures which makes them expensive as well as time-consuming [21]. As a result, they fail to meet the requirements of real-time data collection, management and maintenance. To manage such data warehouses, Oracle, Teradata or SQL Server is required but their licensing costs and maintenance costs are high. Thus, the implementation of cloud-based technologies is necessary to reduce these expenses and enhance flexibility in businesses.

2.4. Increased Data Quality Concerns

An organization uses different sources to collect the data, which is available in different structures. It is difficult for traditional data warehouses to maintain the quality of the data because errors and discrepancies occur when the data is collected from different sources. This issue leads to inaccurate data results which in turn affect the decision-making processes of an organization. Another issue that occurs is data duplication, which is caused when the data is not

updated timely.

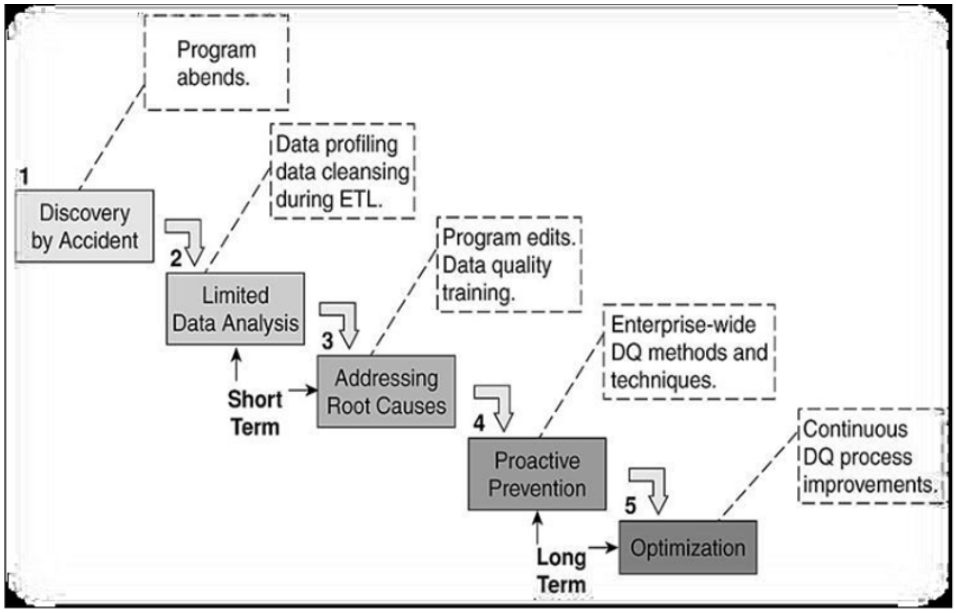

Figure 9: Data Quality Concerns in Data Warehouses [22]

3. Methodology/Approach

3.1. Research Design

This research utilizes a systematic literature review-based research design. This is because it helps in analyzing the changing attitudes of the target study. Besides, it also helps in gaining familiarity with the current research area. It helps in understanding the theories supporting the topic, allowing us to place the current research question into context. The approach of a systematic literature review is based on an organized and methodical search for relevant information while ensuring the analysis of a diverse range of opinions. Overall, it helps in the exploration of major gaps, trends, and key concepts in the current literature, which assists in developing the groundwork for a detailed acknowledgement of flexible data warehousing in the case of changing workloads.

3.2. Data Collection Methods

For the purpose of collecting data regarding the alignment of flexible data warehousing with changing workloads with the help of cloud-based techniques, a systematic approach has been developed. The secondary data collection method was used in this research to gain relevant information from existing literature. This section is based on a detailed depiction of the data collection procedure, showing the method used to select and assess the relevant journal articles.

3.3. Development of a Systematic Search Strategy

Before collecting data, a search strategy was developed to select relevant articles from different databases. The use of diverse databases was ensured in order to ensure the use of diverse opinions and perspectives regarding the research topic. Moreover, the literature search covered conference proceedings, journal articles, and peer-reviewed articles.

3.4. Inclusion Criteria

An inclusion criterion was also developed while collecting data. Only those articles were selected for the research that included the specific keywords of this research, i.e., elastic data warehousing, cloud-native technologies, and fluctuating workloads. Moreover, the research only included articles from 2019 to 2023. This helped in ensuring that only up-to-date data was gathered for the research study.

3.5. Keyword and Term-Based Search

The search procedure for the research was based on the use of specific keywords from the title of the research. These include: elastic data warehousing, cloud-native technologies, and fluctuating workloads. The use of specific term-based search helped in gaining the most relevant data for the research, which adds to its credibility and reliability.

3.6. Source Selection and Examination

The sources used for this research were analyzed in detail in order to gain useful insights into the aspects of flexible data warehousing. The analysis procedure was based on a critical examination of the selected articles while analyzing their credibility, reliability, and relevance to the research.

3.7. Data Analysis Technique

The information gathered from different articles was assessed with the help of thematic analysis. This technique provides an organized approach to synthesizing and organizing various sources, which makes it easy to analyze the patterns, gaps, and trends in the existing literature. Through this method, the key themes of the research have been identified, which provides a detailed analysis of particular aspects of the research questions.

3.8. Thematic Analysis

Thematic analysis is a very common but crucial way of analyzing research data. It is based on the analysis of the major themes or patterns in a research area. It helps in dividing the collected data into different themes, which makes it easier to analyze and assess them in detail. The use of

this technique has been very helpful in the current research since it organized the whole dataset into a particular form.

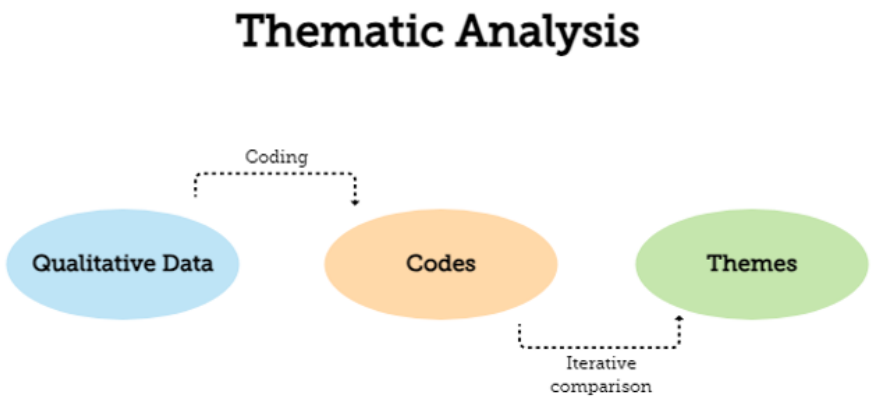

Figure 10: Illustration of Thematic Analysis [23]

3.9. Identifying Key Themes

The most important aspect of this research was to develop different themes based on the research topic. For this purpose, different sources were analyzed to identify the most important topics in this area. The major themes that were developed in this research are based on the challenges in traditional data warehousing, elasticity in warehousing, scalability and resource management, best practices for managing data warehouses, and future trends in this area.

3.10. Facilitating a Deeper Understanding of the Findings

The analysis procedure did not only involve the categorization of the major themes but was also based on the development of useful insights from the literature. The use of thematic analysis helped in the exploration of major aspects within each of them, which helped in gaining useful acknowledgement of the research topic.

3.11. Comparative Discussions

After developing the major themes for this research, a comparative discussion was also conducted in order to compare the major aspects in these themes. In this way, the differences and similarities among all the themes were identified, which helped in presenting the findings in a more organized manner.

4. Results and Discussion

4.1. Elastic Data Warehousing Frameworks

4.1.1. Amazon Redshift

There are a lot of elastic data warehousing frameworks among which Amazon Redshift is the one [24]. It can be defined as a framework that is responsible for maintaining scalable data within the Amazon Web Services ecosystem. It is efficient in handling and managing large sets of data as it uses a columnar storage approach where the data is stored in columns. Massively Parallel Processing (MPP) architecture has been employed in Amazon Redshift that plays a vital role in maintaining a large computing network and maintaining all the tasks among the connected users with the purpose of efficient and rapid query execution. This framework allows the organizations to adjust the sizes of clusters on the basis of organizational workload. This feature makes it cost-effective and efficient to get flexibility in data management.

4.1.2. Snowflake

Snowflake is one of the best data warehousing frameworks that is helpful for organizations to manage, handle and analyze their data to enhance their overall operations [25]. Its storage methods and computing resources make it distinctive among other data warehousing frameworks. These features support scalable data which in turn offers flexibility at higher levels. The multi-cluster feature of this framework provides independent access to all the nodes connected with the network which results in enhancing the overall performance of the network as well as the data. Moreover, automatic scaling is also offered by Snowflake that helps in meeting the requirements of workload management along with maintaining cost-effectiveness and efficiency of the operations. Security and privacy is also the main concern of this platform which secures the sensitive data.

4.1.3. Google BigQuery

BigQuery is a modern data warehousing platform that is highly scalable and supports real-time analytics and operations in cloud computing [26]. This platform comes with an easy-to-use interface that makes it useful and no technical knowledge is required to use it. Standard SQL queries can be used for storing the data no matter how big the data is. This framework is based on the serverless model which makes it the best for the organizations in this modern world to fulfill all the needs regarding data warehousing. Its capabilities are enhanced with the help of cloud computing services which play a vital role in the processing and analysis of big data. The data in different structures and collected from multiple sources can be managed with the help of federated queries.

4.2. Scalability Dynamics in Cloud-Native Solutions

4.2.1. Overview of Scalability in Cloud-Native Environments

Scalability can be defined as one of the major benefits offered by cloud-native development [27].

Organizations can scale their businesses with the help of containerization and microservices. These services help in allocating resources when necessary which results in optimizing costs as well as enhancing the overall performance during traffic spikes. The IT resources can be increased or decreased with a scalable cloud environment when the reason is to meet the changing needs and requirements of the market. Today, organizations are seeking scalable cloud-native environments that support scalable computing and storage operations. This, in turn, contributes to the overall success of the organization. Features like container orchestration are also helpful in scaling up cloud computing services to meet workload demands and enhance customer satisfaction

4.2.2. Mechanisms for Dynamic Scaling

Dynamic scaling plays a vital role in enhanced cloud-native environments because it involves those mechanisms that are helpful in automating system resources for the purpose of managing workloads [28]. In this concern, auto-scaling is applied, which includes predefined guidelines and policies for the removal and addition of cloud computing services in real time. This process makes sure that the resources are allocated efficiently when the activity is increased and demand is decreased. Dynamic scaling is usually supported by container orchestration platforms as they play an important role in deploying and managing the applications that support scaling of integrated applications and meet the customer demands at the same time.

4.2.3. Benefits of Scalability in Cloud-based Network

It is known that scalability offers a number of benefits to organizations when it comes to its involvement in cloud-native networks [29]. Enhanced performance is one of the major advantages as it lets the organizations handle the workloads effectively and efficiently without any delay. Another benefit is cost effectiveness as cloud-native solutions provide the organizations with the opportunity to allocate their resources without investing a lot of money especially when it is a low activity period. The cost of infrastructure is reduced when a scalable cloud-based environment is implemented. Thus, the organizations can spend their money on innovations and advanced technologies to enhance the business operations and performance.

4.3. Assessing Adaptation Efficiency through Performance Metrics and Benchmarks

4.3.1. Performance Metrics Overview

The performance metrics play a vital role when it comes to measure the efficiency of a cloud-based network within an organization. These metrics ensure that this system is reliable enough to manage workloads without compromising the quality of the data [30]. These performance metrics may include utilization of resources, throughput, response time etc. The response time can be defined as the period from the initialization of the query till the response of the system to that query. While, throughput includes the calculation of the number of queries made in a specified time period. When it comes to resource utilization, it explains the employment of resources when dealing with workloads.

4.3.2. Benchmarks as Evaluation Tools

Benchmarks are essential tools that are helpful in evaluating the overall performance of a system according to some predefined industry standards to regulations. They measure the computing resources performance and have also been applied to cloud resources in the network. There are a number of benchmarks in cloud-based data warehousing, for example, OLTP-Bench is a common benchmark that plays a vital role in controlling workload skew, mixture and transaction rate during the execution of a task. Moreover, the TPC-H benchmark is used for enhancing the performance of the database as a whole. Benchmarks also come with some challenges because of the diversity in workloads, that's why it is important for the organizations to be selective when it comes to choosing benchmarks. They must select the ones that are in accordance with the organizations' specific use cases.

4.4. Security and Compliance in the Cloud

It is important to consider the factors of security and compliance when it comes to cloud computing because it is necessary to maintain confidentiality for the sensitive data of the business as well as the users connected with the network. Cloud-based technologies offer a lot of benefits but they also come with a number of challenges that are needed to be addressed for an efficient network.

4.4.1. Challenges in Cloud Security

One of the major challenges in cloud security is related to the protection and privacy of the data. The confidentiality of the data becomes complex and intricate because it travels from a lot of sources i.e. cloud services and services, that's why it is difficult to maintain efficiency of such data. The major challenges that occur in this concern include data breaches and unauthorized access which may result in tangible hazards. It is important for the cloud service providers to deal with such cyber threats and implement the shared responsibility model in the cloud-based environment where both customers and cloud services providers are responsible for securing data.

4.4.2. Strategies for Security in the Cloud

When the security concerns in the cloud are increased, it is important for the organizations to focus on implementing advanced and latest cloud-based technologies. Access to users must be limited so that the data could be protected from unauthorized access. Identity and Access Management systems must be implemented by the organizations to limit the controls over data handling and management so that only authorized individuals could have access to the sensitive information. Encryption techniques must be applied in this concern for the protection of existing and newly-added data. When the data is transferred from one user to another, the data is encrypted and can only be decrypted with the help of specified keys. Moreover, the organizations must monitor and audit their data regularly so that the discrepancies in data could be resolved timely.

4.4.3. Compliance in the Cloud

The cloud-based networks come with a number of compliances as well so it is important for the organizations to focus on these compliances and stay aligned with geographical as well as industrial rules and regulations. Some common compliances include the General Data Protection Regulation (GDPR), the Payment Card Industry Data Security Standard (PCI DSS) and the Health Insurance Portability and Accountability Act (HIPAA), where they all are related to the protection of data within an organization. In cloud computing, compliance risks may lead to legal penalties as well as financial loss if industry standards are not being followed. It is necessary for the policy-makers to focus on implementing robust security controls so that industry standards could be met.

4.5. Challenges and Opportunities in Elastic Data Warehousing

4.5.1. Challenges in Cloud Computing

- **Security Concerns**

Cloud computing comes with security challenges when they are implemented in small or large organizations. It is important to protect this data whether it is at rest or is getting transferred from one node to another. For this purpose, the shared responsibility model is the best to implement where both the customers and cloud services providers are responsible for the maintenance and security of the data so that it could be protected from cyber threats.

- **Cost Management**

Cloud computing comes with a lot of benefits but it is expensive to manage data in it. It is important for the organizations to optimize the process of resource allocation by choosing the best pricing model and focus on cost-tracking techniques so that expenses could be managed effectively. It is stated that cost efficiency of a cloud model is helpful in enhancing the overall performance of a business.

- **Interoperability and Vendor Lock-In**

Avoiding vendor lock-in and the interoperability between different cloud-based platforms are also included in the challenges associated with cloud computing. A cloud computing network must be able to move data and applications to other cloud platforms, ensuring flexibility and mitigating the reliance on just one provider. Thus, organizations must implement a multi-cloud approach in this concern.

4.5.2. Opportunities in Cloud Computing

- **Innovation Acceleration**

Cloud computing not only comes with challenges but also opportunities like innovation acceleration [31]. Organizations can introduce new prototypes and focus on developing innovative solutions by using advanced cloud computing services and tools. SaaS, PaaS and IaaS are helpful for the organizations to innovate solutions and make them competitive in the market.

- **Global Accessibility and Collaboration**

The organizations that use cloud computing services can get connected with other cloud networks globally which means they can share their data and resources worldwide [32]. Team members from different corners of the world can get connected and work seamlessly. Thus, cloud computing provides the opportunity to collaborate in real time, enhance communication, and share resources for the purpose of increasing the overall efficiency and productivity.

5. Conclusion

To conclude, the use of modern data warehousing techniques including cloud-based technologies can help in improving the efficiency of overall systems. The research has shown that the utilization of different methods like data virtualization and gradual migration can help in developing efficient data warehouses. While many issues are present in this regard, there are many opportunities for technology firms to dig deeper into the aspects of cost optimization, scalability, and innovation by using modern innovative techniques like machine learning and AI. It is also seen that it is important to enhance the operational efficiency of data warehouses by using modern techniques. It can also help in improving the working operations within the digital world by integrating new and novel approaches of data warehousing.

6. Future Scope

Future researchers should focus on exploring the integration of modern innovative techniques in data warehousing. They should try to develop such systems that overcome the challenges encountered in traditional data warehouses. More specifically, they should delve deeper into the implementation mechanisms of the aspects of machine learning, automation, and AI in data warehouses. Moreover, the implementation of these mechanisms in specific industries should also be ensured. The development of customized systems for each industry will help in ensuring that the unique requirements of each industry are being met by upgraded data warehousing methods.

References

[1] M. A. Georgiou, "Enabling workload scalability, strong consistency and elasticity with transactional database replication," Doctoral dissertation, 2020.

[2] Data Sleek, "What are the Advantages of Building a Cloud Data Warehouse?," 16 September 2022. [Online]. Available: https://data-sleek.com/blog/what-are-the-advantages-of-building-a-data-warehouse-in-the-cloud/.

[3] Z. Liu, H. Wu, T. Bai, Y. Wang and C. Xu, "Towards Elastic Data Warehousing by Decoupling Data Management and Computation," Proceedings of the 2020 4th International Conference on Cloud and Big Data Computing, pp. 52-57, 2020.

[4] StorageNewsletter, "Elastic Data Warehouse Through AWS Marketplace by Snowflake Computing," 9 December 2016 . [Online]. Available: https://www.storagenewsletter.com/2016/12/09/elastic-data-warehouse-through-aws-marketplace-by-snowflake-computing/.

[5] N. Kunnathuvalappil Hariharan, "Trends in Data Warehousing Techniques," Trends in Data Warehousing Techniques. International Journal of Innovations in Engineering Research and Technology, pp. 7-14, 2019.

[6] J. Ning, X. Huang, W. Susilo, K. Liang, X. Liu and Y. Zhang, "Dual access control for cloud-based data storage and sharing," IEEE Transactions on Dependable and Secure Computing, pp. 1036-1048, 2020.

[7] Symphony Solutions, "Benefits of Cloud Data Warehouse for Your Business," 2020. [Online]. Available: https://symphony-solutions.com/insights/cloud-data-warehouse-benefits-for-business.

[8] A. Arora and A. Gosain, "Mechanism for securing cloud based data warehouse schema," International Journal of Information Technology, pp. 171-184, 2021.

[9] A. Agiwal, K. Lai, G. N. B. Manoharan, I. Roy, J. Sankaranarayanan, H. Zhang and H. Hacıgümüş, "Napa: Powering scalable data warehousing with robust query performance at Google," Proceedings of the VLDB Endowment, 2021.

[10] K. Roy, "A Complete Guide To Data Warehousing — What Is Data Warehousing, Its Architecture, Characteristics & More!," 4 August 2021. [Online]. Available: https://medium.com/datatobiz/a-complete-guide-to-data-warehousing-what-is-data-warehousing-its-architecture-characteristics-863220d605d6.

[11] P. Bogojevic, "Project management in data warehouse implementations: a literature review," IEEE Access, pp. 225902-225934, 2020.

[12] S. Huang, G. Wang, Y. Yan and X. Fang, "Blockchain-based data management for digital twin of product," Journal of Manufacturing Systems, pp. 361-371, 2020.

[13] Pragna, "How to Design a Data Warehouse-Best Practices and Examples," 12 October 2023 . [Online]. Available: https://www.projectpro.io/article/how-to-design-a-data-warehouse/825.

[14] G. Garani, A. Chernov, I. Savvas and M. Butakova, "A data warehouse approach for business intelligence," IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 70-75, 2019.

[15] D. A. P. Pandian, "Artificial intelligence application in smart warehousing environment for automated logistics," Journal of Artificial Intelligence and Capsule Networks, pp. 63-72, 2019.

[16] S. Aggarwal, "Machine Learning Juggernaut," 2020. [Online]. Available: https://www.qubole.com/blog/qubole-brings-machine-learning-to-the-data-warehouse.

[17] F. Bell, R. Chirumamilla, B. B. Joshi, B. Lindstrom, R. Soni and S. Videkar, "The snowflake data cloud," Snowflake Essentials: Getting Started with Big Data in the Cloud , pp. 1-10, 2021.

[18] A. Nambiar and D. Mundra, "An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management," Big Data and Cognitive Computing, p. 132, 2022.

[19] J. Richman, "What Is A Traditional Data Warehouse? Examples & Challenges," 5 September 2023. [Online]. Available: https://estuary.dev/traditional-data-warehouse/.

[20] J. Gikandi and A. Njuguna, "PROMOTING COMPETITIVE ENTREPRENEURSHIP: USING CLOUD COMPUTING FOR OPTIMAL BUSINESS INTELLIGENCE IN SMALL AND MEDIUM ENTERPRISES," International Journal of Business Management and Processes, pp. 42-60, 2021.

[21] S. Roy, S. Raj, T. Chakraborty, A. Chakrabarty, A. Cortesi and S. Sen, "Efficient OLAP query processing across cuboids in distributed data warehousing environment," Expert Systems with Applications, p. 122481, 2024.

[22] R. K. Pandey, "Data Quality in Data warehouse: problems and solution," 2014. [Online]. Available: https://www.semanticscholar.org/paper/Data-Quality-in-Data-warehouse%3A-problems-and-Pandey/7f0f6b58ae156b7207c15462a4c17c253faedf83.

[23] A. Salomão, "The Art of Interpretation: A Journey through Thematic Analysis," 20 September 2023. [Online]. Available: https://mindthegraph.com/blog/thematic-analysis/.

[24] N. Armenatzoglou, S. Basu, N. Bhanoori, M. Cai, N. Chainani, K. Chinta and D. Terry, "Amazon Redshift re-invented," Proceedings of the 2022 International Conference on Management of Data, pp. 2205-2217, 2022.

[25] A. Gupta and A. Sahayadhas, "A comprehensive survey to design efficient data warehouse for betterment of decision support systems for management and business corporates," International Journal of Management, pp. 463-471, 2020.

[26] M. H. Ali, M. S. Hosain and M. A. Hossain, "Big Data analysis using BigQuery on cloud computing platform," Australian JofEng Inno Tech, pp. 1-9, 2021.

[27] S. Henning and W. Hasselbring, "A configurable method for benchmarking scalability of cloud-native applications," Empirical Software Engineering, p. 143, 2022.

[28] P. Singh, P. Gupta, K. Jyoti and A. Nayyar, "Research on auto-scaling of web applications in cloud: survey, trends and future directions," Scalable Computing: Practice and Experience, pp. 399-432, 2019.

[29] S. Henning and W. Hasselbring, "A configurable method for benchmarking scalability of cloud-native applications," Empirical Software Engineering, p. 143, 2022.

[30] M. S. Aslanpour, S. S. Gill and A. N. Toosi, "Performance evaluation metrics for cloud, fog and edge computing: A review, taxonomy, benchmarks and standards for future research," Internet of Things, p. 100273, 2020.

[31] S. Koehler, H. Desamsetti, V. K. R. Ballamudi and S. Dekkati, "Real World Applications of Cloud Computing: Architecture, Reasons for Using, and Challenges," Asia Pacific Journal of Energy and Environment, pp. 93-102, 2020.

[32] A. Sunyaev and A. Sunyaev, "Cloud computing," Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies, pp. 195-236, 2020.