ВІСНИК ПРИАЗОВСЬКОГО ДЕРЖАВНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ
2024р. Серія: Технічні науки Вип. 48
p-ISSN: 2225-6733; e-ISSN: 2519-271X

122 КОМП'ЮТЕРНІ НАУКИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.75 doi: 10.31498/2225-6733.48.2024.310669

© Boyko N.I.[1], Chernenko A.V.[2]

MODERN APPROACHES TO DATA STORAGE: COMPARISON OF RELATIONAL AND CLOUD DATA WAREHOUSES USING ETL AND ELT METHODS

The paper analyses various aspects of the use of relational and cloud data warehouses as well as methods of integrating ETL and ELT data. A comparative analysis of these approaches, their advantages and disadvantages are provided. A central relational data warehouse is proposed that provides a single version of truth (SVOT), which allows standardising and structuring data, avoiding differences and providing the access to the same information for all users of an organisation. It is analysed the methodological approaches to implementing a data warehouse: top-down, bottom-up, and from middle. It is described cloud data warehouses that use cloud technologies to provide scalability, availability and fault tolerance, which is important for the companies with huge amounts of data. The advantages and disadvantages of ETL and ELT are analysed: ETL transforms data before it is loaded into the warehouse, which makes it easier to maintain data confidentiality. ELT performs transformation after loading, which allows for more flexible data processing directly in the warehouse. In the article, we deal with the approaches to implementing a data warehouse: top-down is suitable for strategic planning, bottom-up allows for faster results, and the middle approach combines both methods to achieve optimal efficiency. We considered cloud data storage: compared to relational storage, cloud storage is more flexible, scalable and efficient, providing speed and reducing infrastructure costs. It is described cloud storage architectures: massive parallel processing, hybrid architectures, lambda architectures, and multi-structured architectures. They provide high performance and flexibility in data processing. It is described data storage technologies: Data Lake, Polyglot Persistence, Apache Iceberg, Apache Parquet, and columnar databases that provide efficient storage and processing of large amounts of data.
Key words: *data warehouse, relational data warehouse, Data Lake, Polyglot Persistence, Apache Iceberg, Apache Parquet.*

Бойко Н.І., Черненко А.В. Сучасні підходи до зберігання даних: порівняння реляційних і хмарних сховищ з використанням ETL та ELT методів. *У дослідженні проаналізовано різні аспекти використання реляційних і хмарних сховищ даних, а також методи інтеграції даних ETL та ELT. Наведено порівняльний аналіз цих підходів, їх переваги та недоліки. Запропоновано центральне реляційне сховище даних, яке забезпечує єдину версію правди (Single Version of Truth), що дозволяє стандартизувати та структуризувати дані, усуваючи розбіжності і забезпечуючи доступ до однакової інформації для всіх користувачів організації. Проаналізовано методологічні підходи для реалізації сховища даних: зверху вниз, знизу вгору та із середини. Розглянуті хмарні сховища даних, які використовують хмарні технології для забезпечення масштабованості, доступності та відмовостійкості, що є важливим для компаній з великими обсягами даних. Проаналізовано переваги та недоліки ETL та*

[1] *Ph.D. in Economics, Associate Professor, Lviv Polytechnic National University, Lviv, ORCID: 0000-0002-6962-9363, Nataliya.i.boyko@gmail.com*
[2] *student, Lviv Polytechnic National University, Lviv, ORCID: 0009-0006-7910-2393, chernenko.anastasiia789@gmail.com*

ELT: ETL виконує трансформацію даних перед їх завантаженням у сховище, що спрощує дотримання конфіденційності даних. ELT виконує трансформацію після завантаження, що дозволяє більш гнучко обробляти дані безпосередньо в сховищі. В статті розглянуто підходи до реалізації сховища даних: зверху вниз підходить для стратегічного планування, знизу вгору дозволяє швидше отримувати результати, а підхід із середини комбінує обидва методи для досягнення оптимальної ефективності. Розглянуто хмарні сховища даних: порівняно з реляційними сховищами, хмарні сховища є більш гнучкими, масштабованими та ефективними, забезпечуючи швидкість і зниження витрат на інфраструктуру. Наведені архітектури хмарних сховищ: масова паралельна обробка, гібридні архітектури, лямбда-архітектури та багатоструктурні архітектури. Вони забезпечують високу продуктивність та гнучкість у обробці даних. Наведені технології зберігання даних: Data lake, Polyglot Persistence, Apache Iceberg, Apache Parquet та стовпчасті бази даних, які забезпечують ефективне зберігання та обробку великих обсягів даних.

Ключові слова: *база даних, реляційні бази даних, Озеро даних, Polyglot Persistence, Apache Iceberg, Apache Parquet.*

Description of the problem. Nowadays, processing huge amounts of information is a highly important task for government agencies, scientific research and business. Relational data warehouses, which have been around for a long time, play a key role in this, but they were not designed to cope with today's explosive data growth or to deal with the ever-changing needs of end users.

With cloud storage, there are no more limitations to physical data centres, and it is possible to dynamically expand or contract your data stores to quickly respond to changing business budgets and requirements. Similar to traditional data warehousing, cloud data warehousing stores information from a variety of data sources, such as IoT and financial systems [1].

The possibilities of relational and cloud data warehouses are determined by the need to choose the optimal technology for efficient management of large amounts of data, ensuring high performance and security, and optimising IT infrastructure costs. Analysing these technologies will not only help to measure their current state, but also predict trends in their development and use in the future.

The importance of this paper is based on the fact that in today's world, with the help of a large amount of properly structured and processed data, it is possible not only to view user data, but also to apply machine learning methods to it, such as prediction, classification or recommendation systems that help in various areas of our lives.

The practical value of this research includes the possibility of real improvement of data storage and processing data in organisations, which will increase productivity, reduce costs, improve data reliability and security, and support innovative business development.

The purpose of this publication is to conduct a comprehensive analysis of the capabilities of relational and cloud data warehouses, as well as to assess their efficiency and perspectives for storing and processing large amounts of data.

Analysis of the latest research and publications. Modern research is actively exploring new approaches to data storage and processing. In general, they often describe the perspectives of cloud technologies, their advantages and disadvantages, as well as the process of migration from relational warehouses.

Literature analysis is an important step in conducting scientific research, as it allows you to get acquainted with the existing theoretical and practical knowledge on the subject, identify the main concepts and trends, and detect gaps in knowledge that require further study. As part of this paper, a number of scientific articles, books and research papers on relational and cloud data warehouses were analysed.

In the article «A comparative analysis of traditional and cloud data warehouse» [1], the author compares traditional data warehouses with cloud data warehouses and notes that traditional warehouses might not be suitable for modern data analysis needs due to changes in the industry. Traditional warehouses are difficult to scale and cannot effectively cope with the growing number of users. The author describes the appearance of a new type of data warehouse – cloud warehouses, which have new aspects of design and implementation. Cloud data warehouses have evolved over time, which has an impact on applications and business domains, allowing for efficient management of huge amounts of data, easy scaling of resources, and serving an unlimited number of users. In conclusion, the author notes that the future belongs to cloud data warehouses.

In the article «Migrating a research data warehouse to a public cloud: challenges and opportunities» [2], the author shares his experience of moving from relational data warehouses to cloud-based ones. It is noted that the needs for computing resources and warehouses in research environments can quickly grow beyond the capabilities of local systems. New research data warehouses (RDWs) are moving to cloud platforms as they provide the scalability and flexibility needed to meet modern challenges.

The paper «Data-warehousing on Cloud Computing» [3] describes that cloud processing has become a new paradigm for storing data and providing data processing services over the Internet. Cloud computing is beneficial to business owners because it eliminates the need to plan for services in advance, allowing businesses to start small and increase resources only as demand for services grows.

There is also the article «Integration Methods and Advantages of Machine Learning with Cloud Data Warehouses» [4], which focuses on the integration of cloud data warehouses with machine learning and the importance of parallel integration methods. First, the article describes how the integration of cloud data warehouses and machine learning can support business innovation and productivity. It continues with a discussion of the challenges of managing machine learning models in production environments and the role of cloud data warehousing in solving these problems. Next, the authors present the Snowflake cloud computing integration in details, including the implementation stages and parallel integration processes. The authors analyse the results of the parallel integration method and believe that it has good prospects for application and development potential in cloud data warehouses.

The paper «A Data Warehouse Approach for Business Intelligence» [5] is devoted to the study of perception, understanding, querying and managing spatiotemporal data for online analytical processing (OLAP). The authors point out that in a cloud data warehouse, business users can access and query data from multiple sources and geographically distributed locations. Business analysts and decision makers rely on data warehouses, especially for data analysis and reporting. Temporal and spatial data are critical to decision-making and marketing strategies, and many applications require modelling and specialised processing of such data because it cannot be effectively handled in a conventional multidimensional database.

The article «Optimising Data Warehousing Performance Through Machine Learning Algorithms in the Cloud» [6] explores the integration of machine learning (ML) into cloud data warehouses, focusing on optimisation challenges, methodology, results, and future trends. The authors examine data warehouses, which are central to reporting and analysis, and are transformed by ML, solving problems such as high maintenance costs and failure rates. Integration improves performance through query optimisation, indexing, and automated data management. The results that they presented demonstrate the use of ML in predictive analytics for working load management, automated query optimisation, and adaptive resource management, which improves performance.

In the paper «From Classical DW to Cloud Data Warehouse» [7], the author researches Data Lake tools and notes that the technology has evolved significantly over the past decade. It responds to ever-growing data volumes with the support of distributed computing tactics. The study provides an example of applications of different types of structured, semi-structured and unstructured data. Data warehouse processing, namely data lakes, is moving from on-premises server rooms to cloud data centres. The data warehouse solution must also be able to meet the requirements of machine learning.

The authors of the article «Privacy and Availability in Cloud Data Warehouse» [8] write that with the development of cloud computing, data warehouses can benefit from this new technology in terms of cost reduction, computing power and response time. In line with technological advances, cloud computing has disadvantages, particularly in terms of security, which must be taken into account to take full advantage of this solution. To solve this problem, the authors proposed a new secure scheme for external data storage in the cloud. This new scheme is based on the (n, k) Shamir secret sharing scheme, IDA information dissemination algorithm, and MV-OPE multi-valued order preserving encryption. The scheme can provide high-confidentiality data, it also ensures data availability and minimises dependence on a single cloud provider. Experimental results in their paper show that the scheme allows sending queries to the data store in the cloud without further processing with high performance. The disadvantage of this solution is that if sensitive data and descriptive attributes are large in data volume and make up the majority of the data in the storage and are costly. Therefore, such solutions are more suitable for a repository that contains little confidential data [9].

A review of the literature (Table 1) shows that both technologies have their advantages and disadvantages, and the choice between them depends on the specific needs of the organisation. Relational databases are traditionally reliable and efficient for structured data and transactional operations, while cloud storage offers flexibility, scalability and cost-effectiveness for processing huge amounts of data. Further data processing experiments will allow us to draw more detailed conclusions and recommendations on the optimal technology.

Table 1

Analysis of the main publications on the researched topic

Title of the paper (author)	Approaches to data analysis and storage	Advantages of applying the approaches	Disadvantages of applying the approaches
Rehman, K. U. ur, Ahmad, U., & Mahmood, S. (2018).	Comparison of traditional data warehouses with cloud	The processes that take place in cloud data warehouse have an impact on applications and business domains. This has given rise to the need to control huge data. It is possible to increase or decrease the storage itself at any time. There are no limitations on the number of users. The advantage of the traditional approach is the simple implementation of processes.	Traditional data warehousing is unsuitable for data analytics needs due to the problem of scaling data, and the difficulty with the growing number of users. Cloud warehouses can be hard to integrate.
Michael G. Kahn, Joyce Y. Mui, Michael J. Ames, Anoop K. Yamsani, Nikita Pozdeyev, Nicholas Rafaels, Ian M. Brooks (2022)	Migration from relational to cloud warehouses	Moving from a traditional data warehouse is not easy. Migration of tools and processes and access to resources is performed on an as-needed basis, which slows down the implementation of cloud warehouses.	Relational data warehouses are being forced to move to cloud platforms for the scalability and flexibility needed to cope with new challenges and increasing data volumes
Hemlata Verma (2013)	Comparison of traditional data warehouses with cloud	Cloud computing data storage is elastic, scalable, takes less time to deploy, is reliable and has lower costs.	Data storage in the cloud is largely hypothetical, and cloud comparisons can lead to new insights into the possibilities and impossibilities of deploying storage in a modern cloud environment.

Summary of the main material. A relational data warehouse acts as a central repository for many subject areas and provides a Single Version of Truth (SVOT). This means that all data in the repository is standardised and structured. This allows to eliminate differences and provide all users with access to the same information, which improves decision-making, collaboration and efficiency throughout the organisation [10, 11].

ETL (Extract, Transform, Load) is a process used in data warehouses to extract data from a variety of sources, transform it into a suitable format, and load it into the data warehouse. ETL applies a set of business rules to cleanse and organise data from raw data, preparing it for storage, analysis, and machine learning (ML). The ETL process consists of three main steps (Fig. 1) and is iterative and repeated as new data is added to the warehouse. It ensures that the data in the data warehouse is accurate, complete and up-to-date, and in the right format for analysis and reporting [12, 13].

ELT (Extract, Load, Transform) is an alternative data integration process. This method moves raw data from the source system to the destination data warehouse. The data integration process combines data from multiple sources into one consistent warehouse. Traditional ETL tools were created to build data warehouses that support business intelligence (BI) and artificial intelligence (AI) applications [14, 15].

The main difference between ELT and ETL is the order in which the steps are performed: ELT first loads and then transforms the data, which means that the second and third steps of the ETL process are reversed (Fig. 1). ELT copies or exports data from the source locations, but instead of intermediate storage for transformation, it loads the raw data directly into the target data warehouse where it can be transformed as needed. Thus, ELT does not transform the data during transfer.

Another difference between ETL and ELT is that ETL transforms data before it is moved to a central repository, which can make it easier or more systematic to maintain data privacy than ELT (Fig. 1).

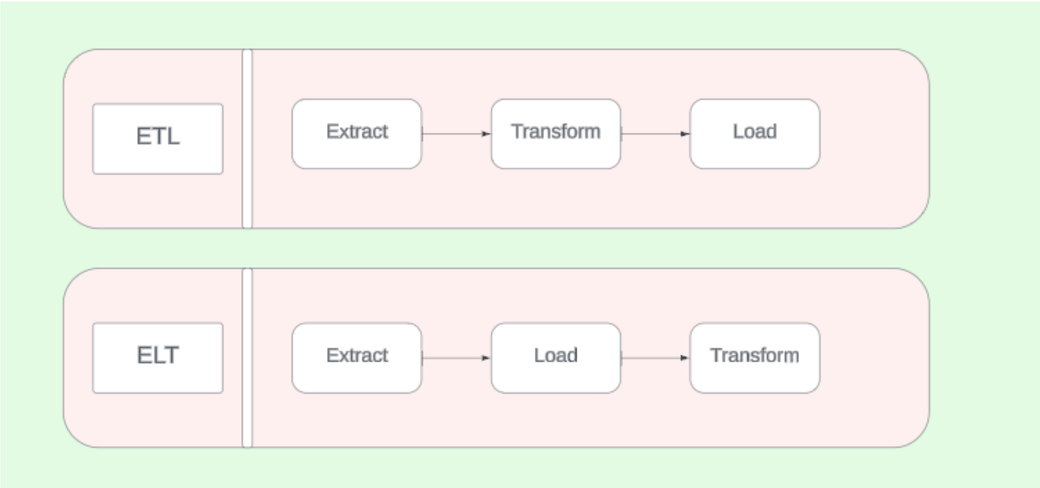

Fig. 1 – Steps of the ETL and ELT processes

Methodological approaches for implementing a data warehouse [16]:

1. Top-down design
2. Bottom-up design
3. Middle of design

The top-down approach works well for historical type reports, to determine what happened (descriptive analytics) and why it happened (diagnostic analytics) (Fig. 2). In a top-down approach, the general planning, design, and architecture of the data warehouse is established first, and then specific components are developed. This method highlights the importance of defining the overall enterprise vision and understanding the organisation's strategic goals and information requirements before starting to develop the data warehouse.

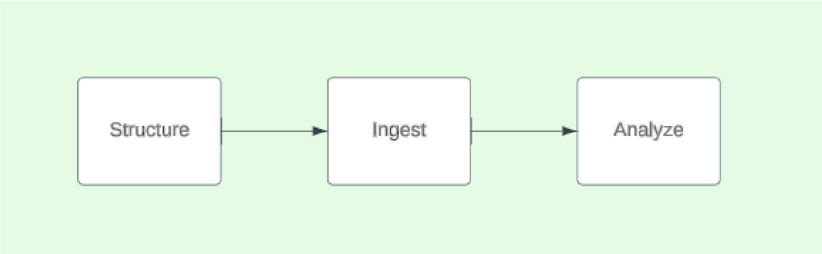

Fig. 2 – Top-down approach

Without knowing what questions to ask about the data, a *bottom-up approach* can quickly explore the data to identify relevant ones when data is initially collected, before generating any theories or hypotheses [17]. The bottom-up approach starts with planning and designing departmental data warehouses without first developing the organisation's global computing infrastructure. This approach is preferable in many cases because it leads to final results faster (Fig. 3).

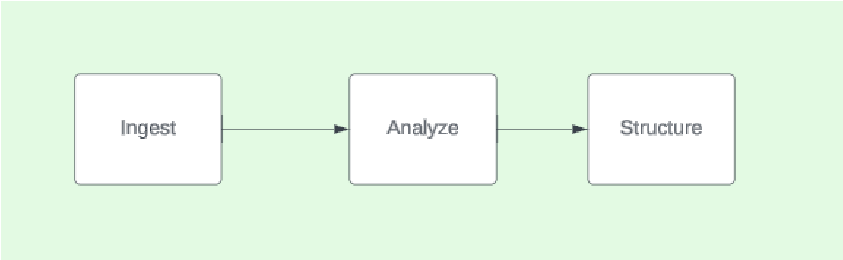

Fig. 3 – Bottom-up approach

Bottom-up and top-down approaches can be combined depending on the goals of the data warehouse project manager. The *design from the middle* approach is a combination of the above approaches applied in a spiral process. First, the core of the system is created (the top-down approach), and then it is gradually expanded by adding new or additional functionality (the bottom-up approach). Thus, each of the two approaches mentioned above can be used at each turn of the spiral [15].

Relational databases (RDBs) are one of the most common ways to organise and store structured data. The main features of RDBs are shown in Table 2.

Table 2

Main features of relational data warehouses

Features of relational data warehouses	Possibilities of using the method
Data model	- The relational model is based on tables containing rows and columns, where each row is a separate record and columns define the attributes of the record. - Popular management systems: MySQL, PostgreSQL, Oracle Database, Microsoft SQL Server
Query language	- To work with relational databases, you can use SQL (Structured Query Language), which allows performing various operations with data: selecting, inserting, updating, deleting.
Data integrity	- RDB ensure high data integrity through the use of transactions and various constraints (keys, indexes, uniqueness, etc.).
Scalability	- RDB are usually scaled vertically, which means adding resources to a single server.

A cloud data warehouse is a data warehouse that exists in a cloud environment and is capable of combining exabytes of data from multiple sources. Cloud data warehouses are designed to handle complex queries and are optimised for business intelligence. The «cloud» part means that instead of managing physical servers and infrastructure, everything happens online - third-party servers do the heavy lifting, and you can access your data and analytics tools over the Internet without having to download or configure any software or applications [18].

Cloud data warehousing is critical for quick data-driven decision making, offering improved computing power and simplified data management, enabling valuable insights to be derived from updated, accurate and enriched data as needed. The data in a cloud data warehouse is well-structured and unified, ready to support a wide range of specific business intelligence use cases.

In the modern world, cloud data warehouses play a critical role in collecting, storing and processing data for decision-making. There are certain types of cloud storage architectures:

1. Massively parallel processing architecture distributes and processes complex data queries across multiple nodes, delivering high performance and scalability for huge datasets [19].

2. Hybrid architectures use a combination of on-premises and cloud infrastructure, typically with on-premises resources used only to extend the cloud where needed. They are designed to leverage the strengths and overcome the limitations of each architecture to achieve optimal efficiency and performance [20].

3. Lambda architectures process huge amounts of data using a combination of layers (batch, speed and serving). The data is delivered simultaneously to the batch and speed layers, with the batch layer supporting the processing of raw data and the speed layer supporting low-latency data that has not yet been delivered to the batch layer. At the same time, the service layer supports real-time queries [21, 22].

Multi-structured data architectures can acquire, store, and process data in a variety of formats for more complete and flexible data analysis.

A data lake is a central location that contains huge amounts of data in a raw format. Compared to a hierarchical data warehouse that stores data in files or folders, a data lake uses a flat architecture and an object store to contain data. The object store saves data with metadata tags and a unique identifier, which makes it easier to find and retrieve data across regions and improve performance [23].

Polyglot Persistence is the use of multiple storage technologies to store different types of data in a single application or system depending on how the data is used. Polyglot Persistence is about choosing the right tool for the right use case, using multiple data stores in an organisation or enterprise. Each data warehouse is optimised for a specific type of data or use case. This approach allows organisations to use different data warehouses for different projects or business units, rather than a single approach for the entire organisation [24].

Apache Iceberg is a table format that solves the problems of traditional directories and is quickly becoming the industry standard for managing data in a data lake. Iceberg introduces capabilities that allow multiple applications to work together on the same data in a consistent transaction order and define additional information about the state of datasets as they evolve and change over time [13].

Apache Parquet is an open source, column-oriented data file format designed for efficient data storage and retrieval. It provides efficient data compression and performance-enhancing coding schemes for mass processing of complex data. Apache Parquet is designed as a common exchange format for both batch and interactive workloads [20].

Columnar databases are particularly beneficial for analytical queries that often scan or aggregate large data sets but require only a few columns. Since only the necessary columns are read from the storage, import/export costs and time are minimised, providing an advantage over traditional row-based databases in analytics scenarios. For these use cases, columnar databases make better use of system resources and provide faster information [17].

In summary, cloud data storage is a new approach to data storage and processing that takes advantage of cloud technologies. The main features of cloud storage are shown in Table 3.

Table 3

Main features of cloud data warehouses

Features of cloud data warehouses	Possibilities of using the method
Data model	- Cloud data warehouses can support both relational and non-relational data models, providing flexibility in choosing a data storage approach. - Popular management systems: Amazon RDS, Google Cloud SQL, Microsoft Azure SQL Database
Payment model	- Cloud services usually operate on a pay-as-you-go model, which allows you to optimise costs
Availability and reliability	- Cloud providers deliver high availability and reliability of data through distributed architecture and backups.
Scalability	- Cloud storage is usually scaled horizontally, which allows you to efficiently process huge amounts of data by distributing the load across many servers.

Based on the literature analysis, the following key aspects can be identified for comparing relational and cloud data warehouses (Table 4).

Table 4

Comparing relational and cloud data warehouses

Main characteristics of data warehouses	Relational data warehouses	Cloud data warehouses
Productivity	- Relational databases demonstrate high performance when working with transactional data and complex queries.	- Cloud storage provides flexibility and scalability, which can be more efficient when processing huge amounts of data.
Scalability	- Relational databases are typically scaled vertically, which can limit their performance as data volumes grow.	- Cloud storage is typically scaled horizontally, which allows you to distribute the load across multiple servers and provides better performance and resilience to data volume increases.
Price	- The cost of relational databases can be high because of the need to maintain and upgrade the infrastructure.	- Cloud storage offers a pay-as-you-go model, which can be more cost-effective for organisations with intermittent workloads.
Reliability and security	- Relational databases provide high reliability due to the transactional model and integrity constraints.	- Cloud storage provides reliability and security through a distributed architecture, backup, and security tools provided by the providers.

To experiment with the efficiency of relational and cloud data warehouses we chose the dataset «Amazon Books Reviews» [10], which contains 3000000 reviews for 212404 books.

We chose the PostgreSQL database to work with a relational data warehouse. To offer with a preper analysis of the data, we divided it between three tables and added identifiers for each of them:

- Table of books «books» with columns: id- book identifier, title, price. It contains 221998 values
- Table of users «users» with columns: id - user identifier, profileName. It contains 1008972 values
- Table of reviews «reviews» with columns: uuid - review identifier, book_id, user_id, score, time, summary. It contains 2966688 values.

To load the data into PostgreSQL, we used Python and its libraries such as sqlalchemy to communicate with PostgreSQL and its server; and the pandas library to read CSV files in which the data for each of the tables were structured. Queries to the data warehouse were also implemented using sqlalchemy and the time library to calculate the query time.

Queries with COUNT, SUM, GROUP BY operations are most often used in data analytics because they aggregate and summarise large amounts of data, so the focus was on them.

The requests were implemented:

1. Search for book reviews by the book identifier:

Time of request execution	0.8293237686

2. Number of reviews for the book by its identifier:

Time of request execution	0.913532257

3. The number of reviews of a book by its identifier in 2000:

Time of request execution	0.7880356312

4. Average rating of the book among all users:

Time of request execution	0.9420266152

To work with cloud storage, Amazon S3 was chosen to upload data to the cloud and DynamoDB as a database.

Firstly, the CSV file data was converted to JSON format using Python to upload the data to DynamoDB. Then it was uploaded to an S3 bucket. And after that, the data was successfully transferred to the database. Queries to the data warehouse were implemented using boto3 to connect to the server on the cloud and the time library to calculate the query time.

The same queries were performed as in the relational storage, namely:

1. Search for book reviews by the book identifier:

Time of request execution	0.6458830833

2. Number of reviews for the book by its identifier:

Time of request execution	0.553743124

3. The number of reviews of a book by its identifier in 2000:

Time of request execution	0.667219638

4. Average rating of the book among all users:

Time of request execution	0.707470178

For a better testing analysis, a general table comparing the time of each query relative to the storages is displayed (Table 5).

Table 5

Relational and cloud storage query times

Relational data warehouses	Cloud data warehouses
0.8293237686	0.6458830833
0.913532257	0.553743124
0.7880356312	0.667219638
0.9420266152	0.707470178

Analysing the table, it can be said that in all cases, the relational data warehouse processed the query and displayed the result slower than the cloud one. To better see the difference, a diagram can be displayed (Fig. 4).

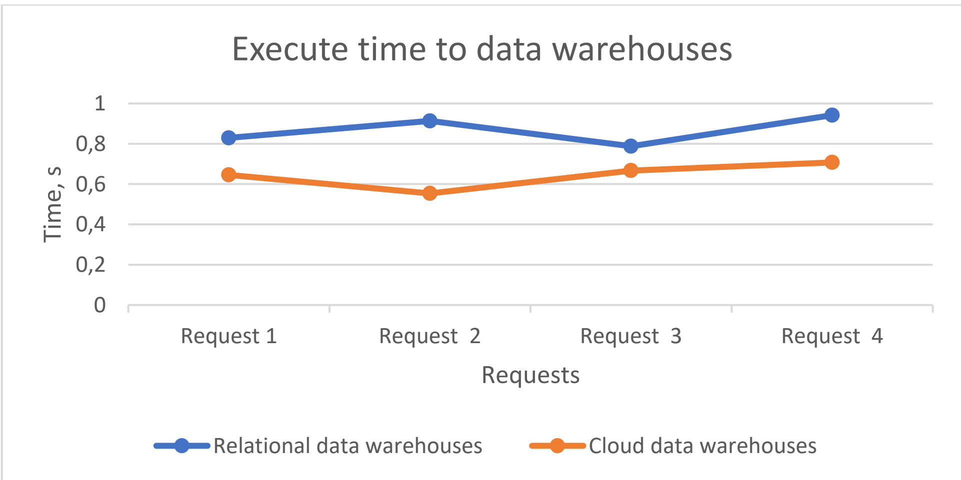

Fig. 4 – Comparison of the time spent on each request by relational and cloud warehouses

Based on the results of the experiments, a conclusion can be made that cloud data warehouses process and analyse tables and data faster than relational warehouses. The tests were conducted on 3 million records, and in large companies, these volumes reach terabytes and even exabytes. Storing such data on physical servers is extremely difficult, as opposed to cloud storage, which is easier to scale. An additional advantage is that data can be accessed from anywhere in the world, which saves on hardware and maintenance.

In addition, cloud storage has the advantage of scalability. They can be easily adapted to the growth of data, which is important for companies that process massive amounts of information. In fact, cloud storage provides high availability and fault tolerance, which ensures continuous access to data and minimises the risk of loss.

To summarise, cloud data warehouses offer a more flexible, scalable and efficient solution for processing huge amounts of data compared to relational warehouses. They are suitable for a wide range of applications, including business intelligence, machine learning, and analysing data from various sources.

Conclusions

Many companies still use relational data warehouses because they are familiar and well-known, or because they fear that data migration will be difficult. However, cloud computing is already well established and has shown to be a reliable and efficient business solution.

The main problem for Ukraine is the unstable supply of electricity and the Internet, which are critical for data processing. Generators also have a limited lifespan. Cloud technologies can significantly improve business operations, as when using physical servers, the lack of access to them stops the company's work. Storing data in the cloud allows you to work with it from anywhere in the world, even if the local infrastructure is down.

Our lives are changing and progressing rapidly. It could be remembered how data was stored on paper, where you had to rewrite the entire document to correct an error. Later, data was stored on physical devices, edited in notebooks or on physical storage. The introduction of relational data warehouses was a significant step forward, reducing the time required to store and process data. Today, cloud computing is the next step in this development, enabling faster and more optimised data processing using a wider range of tools.

From this, we can conclude that cloud computing is a more powerful solution for working with data, and the future of business belongs to it. However, it is possible that even more powerful and relevant technologies will appear in the future.

Comparison has shown that cloud data storage offers a more flexible, scalable, and efficient solution for processing large volumes of data compared to relational storage. Cloud technologies provide significant advantages for businesses, especially in the context of modern challenges such as unstable power and internet supply in Ukraine.

Relational storage, while remaining popular due to its familiarity and stability, lags behind cloud solutions in many aspects. Cloud storage allows savings on physical equipment and ensures data access at any time and from any location, which is a key factor for the successful operation of modern businesses. Cloud technologies are the next logical step in the evolution of data storage and processing, and their use will continue to grow.

Thus, cloud data storage is a promising direction for business development, allowing adaptation to rapidly changing conditions and optimizing the handling of large volumes of information.

Acknowledgement

The study was created within the framework of the project financed by the National Research Fund of Ukraine, registered No. 30/0103 from 01.05.2023, «Methods and means of researching markers of ageing and their influence on post-ageing effects for prolonging the working period», which is carried out at the Department of Artificial Intelligence Systems of the Institute of Computer Sciences and Information Technologies of the Lviv Polytechnic National University.

References:

1. K.U. Rehman, U. Ahmad, and S. Mahmood, «A Comparative Analysis of Traditional and Cloud Data Warehouse», *VAWKUM Transactions on Computer Sciences*, vol. 6(1), pp. 34-40, 2018. **doi: 10.21015/vtcs.v15i1.487**.
2. M.G. Kahn et al., «Migrating a research data warehouse to a public cloud: challenges and opportunities», *Journal of the American Medical Informatics Association*, vol. 29 (4), pp. 592-600, 2022. **doi: 10.1093/jamia/ocab278**.

p-ISSN: 2225-6733; e-ISSN: 2519-271X

3. H. Verma, «Data-warehousing on cloud computing», *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 2(2), pp. 411-416, 2013.
4. H. Li, X. Wang, Y. Feng, Y. Qi, and J. Tian, «Integration Methods and Advantages of Machine Learning with Cloud Data Warehouses», *International Journal of Computer Science and Information Technology*, vol. 2(1), pp. 348-358, 2024. **doi: 10.62051/ijcsit.v2n1.36**.
5. G. Garani, A. Chernov, I. Savvas, and M. Butakova, «A Data Warehouse Approach for Business Intelligence», in Proceedings of IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Napoli, Italy, 2019, pp. 70-75. **doi: 10.1109/WETICE.2019.00022**.
6. A. Sina, «Optimizing Data Warehousing Performance Through Machine Learning Algorithms in the Cloud», *International Journal of Science and Research (IJSR)*, vol. 12(12), pp. 1859-1867, 2023. **doi: 10.21275/SR231224074241**.
7. J. Heinonen, «From Classical DW to Cloud Data Warehouse», M.S. thesis, Helsinki, Finland, 2020.
8. K. Kawthar, A. Nabli, and F. Gargouri, «Privacy and availability in cloud data warehouse», in Proceedings of the 10th International Conference on Education Technology and Computers, 2018, pp. 388-391. **doi: 10.1145/3290511.3290580**.
9. Cloud Data Warehouse. [Online]. Available: https://www.qlik.com/us/cloud-data-migration/cloud-data-warehouse. Accessed on: June 10, 2023.
10. Data Set: Amazon Books Reviews. [Online]. Available: https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews. Accessed on: 12.06.2023.
11. Deep Dive into AWS DynamoDB: A NoSQL Database for High-Performance Applications. [Online]. Available: https://medium.com/@christopheradamson253/deep-dive-into-aws-dynamodb-a-nosql-database-for-high-performance-applications-4c80d1410533. Accessed on: August 01, 2023.
12. PostgreSQL. [Online]. Available: https://kinsta.com/knowledgebase/what-is-postgresql/. Accessed on: September 12, 2023.
13. What is extract, load, transform (ELT). [Online]. Available: https://www.ibm.com/topics/elt. Accessed on: December 10, 2023.
14. Column databases. [Online]. Available: https://www.tinybird.co/blog-posts/what-is-a-columnar-database. Accessed on: January 08, 2024.
15. What is an MPP Database. [Online]. Available: https://airbyte.com/data-engineering-resources/mpp-database. Accessed on: November 13, 2023.
16. Exploring the Basics of Amazon Simple Storage Service (S3). [Online]. Available: https://medium.com/@dbrandonbawe/exploring-the-basics-of-amazon-simple-storage-service-s3-f8ad2af0a6f9. Accessed on: November 10, 2023.
17. What is a Cloud Data Warehouse. [Online]. Available: https://www.astera.com/type/blog/cloud-data-warehouse/. Accessed on: November 10, 2023.
18. The Lambda Architecture: A Hybrid Approach to Data Processing. [Online]. Available: https://www.linkedin.com/pulse/lambda-architecture-hybrid-approach-data-processing-midhun-pottammal. Accessed on: January 10, 2024.
19. Introduction to Data Lakes. [Online]. Available: https://www.databricks.com/discover/data-lakes. Accessed on: January 11, 2024.
20. What is Polyglot Persistence. [Online]. Available: https://www.harperdb.io/post/what-is-polyglot-persistence-and-why-is-it-awful. Accessed on: February 01, 2024).
21. Exploring the Benefits of Hybrid Architecture in Data Warehousing. [Online]. Available: https://reconfigured.io/blog/exploring-benefits-of-hybrid-architecture-in-data-warehousing. Accessed on: November 30, 2023.
22. The Apache Iceberg Open Table Format. [Online]. Available: https://www.dremio.com/resources/guides/apache-iceberg/. Accessed on: November 28, 2023.
23. Parquet. [Online]. Available: https://www.databricks.com/glossary/what-is-parquet. Accessed on: September 02, 2023.
24. ETL Process in Data Warehouse. [Online]. Available: https://www.geeksforgeeks.org/etl-process-in-data-warehouse/. Accessed on: October 10, 2023.

Перелік використаних джерел:

1. Rehman K. U., Ahmad U., Mahmood S. A comparative analysis of traditional and cloud data warehouse. *VAWKUM Transactions on Computer Sciences*. 2018. Vol. 6(1). Pp. 34-40. DOI: https://doi.org/10.21015/vtcs.v15i1.487.
2. Migrating a research data warehouse to a public cloud: challenges and opportunities / M. G. Kahn et al. *Journal of the American Medical Informatics Association*. 2022. Vol. 29(4). Pp. 592-600. DOI: https://doi.org/10.1093/jamia/ocab278.
3. Verma H. Data-warehousing on cloud computing. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. 2013. Vol. 2. Iss. 2. Pp. 411-416.
4. Integration methods and advantages of machine learning with cloud data warehouses / H. Li et al. *International Journal of Computer Science and Information Technology*. 2024. Vol. 2(1). Pp. 348-358. DOI: https://doi.org/10.62051/ijcsit.v2n1.36.
5. A Data Warehouse Approach for Business Intelligence / G. Garani, A. Chernov, I. Savvas, M. Butakova. *Infrastructure for Collaborative Enterprises (WETICE)*: IEEE 28th International Conference on Enabling Technologies, Napoli, Italy, 12-14 June, 2019. Pp. 70-75. DOI: https://doi.org/10.1109/WETICE.2019.00022.
6. Sina A. Optimizing data warehousing performance through machine learning algorithms in the cloud. *International Journal of Science and Research (IJSR)*. 2023. Vol. 12(12). Pp. 1859-1867. DOI: https://dx.doi.org/10.21275/SR231224074241.
7. Heinonen J. From classical DW to cloud data warehouse : Masters Thesis. Helsinki, 2020. 79 p.
8. Kawthar K., Nabli A., Gargouri F. Privacy and availability in cloud data warehouse. *Proceedings of the 10th International Conference on Education Technology and Computers*. 2018. Pp. 388-391. DOI: https://doi.org/10.1145/3290511.3290580.
9. Cloud Data Warehouse. URL: https://www.qlik.com/us/cloud-data-migration/cloud-data-warehouse (дата звернення: 10.06.2023).
10. Data Set: Amazon Books Reviews. URL: https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews (дата звернення: 12.06.2023).
11. Deep Dive into AWS DynamoDB: A NoSQL Database for High-Performance Applications. URL: https://medium.com/@christopheradamson253/deep-dive-into-aws-dynamodb-a-nosql-database-for-high-performance-applications-4c80d1410533 (дата звернення: 01.08.2023).
12. PostgreSQL. URL: https://kinsta.com/knowledgebase/what-is-postgresql/ (дата звернення: 12.09.2023).
13. What is extract, load, transform (ELT). URL: https://www.ibm.com/topics/elt (дата звернення: 10.12.2023).
14. Column databases. URL: https://www.tinybird.co/blog-posts/what-is-a-columnar-database (дата звернення: 08.01.2024).
15. What is an MPP Database. URL: https://airbyte.com/data-engineering-resources/mpp-database (дата звернення: 13.11.2023).
16. Exploring the Basics of Amazon Simple Storage Service (S3). URL: https://medium.com/@dbrandonbawe/exploring-the-basics-of-amazon-simple-storage-service-s3-f8ad2af0a6f9 (дата звернення: 10.11.2023).
17. What is a Cloud Data Warehouse. URL: https://www.astera.com/type/blog/cloud-data-warehouse/ (дата звернення: 10.11.2023).
18. The Lambda Architecture: A Hybrid Approach to Data Processing. URL: https://www.linkedin.com/pulse/lambda-architecture-hybrid-approach-data-processing-midhun-pottammal (дата звернення: 10.01.2024).
19. Introduction to Data Lakes. URL: https://www.databricks.com/discover/data-lakes (дата звернення: 11.01.2024).
20. What is Polyglot Persistence. URL: https://www.harperdb.io/post/what-is-polyglot-persistence-and-why-is-it-awful (дата звернення: 01.02.2024).
21. Exploring the Benefits of Hybrid Architecture in Data Warehousing. URL: https://reconfigured.io/blog/exploring-benefits-of-hybrid-architecture-in-data-warehousing (дата звернення: 30.11.2023).

22. The Apache Iceberg Open Table Format. URL: https://www.dremio.com/resources/guides/apache-iceberg/ (дата звернення: 28.11.2023).
23. Parquet. URL: https://www.databricks.com/glossary/what-is-parquet (дата звернення: 02.09.2023).
24. ETL Process in Data Warehouse. URL: https://www.geeksforgeeks.org/etl-process-in-data-warehouse/ (дата звернення: 10.10.2023).

Стаття надійшла 12.04.2024
Стаття прийнята 01.05.2024

УДК 004.9:658.818 **doi: 10.31498/2225-6733.48.2024.310670**

© Кіріакіді С.К.[1], Кривенко О.В.[2]

ДОСЛІДЖЕННЯ МЕТОДОЛОГІЙ РОЗРОБКИ ВЕБДОДАТКУ ДЛЯ МЕДИЧНОЇ КЛІНІКИ З ВИКОРИСТАННЯМ REACT

У статті детально досліджуються методології розробки вебдодатків для медичних клінік з використанням популярної бібліотеки React. Основною метою цього дослідження є аналіз сучасних підходів до створення вебдодатків, які відповідають специфічним вимогам медичних установ. Особлива увага приділяється функціональним та нефункціональним характеристикам, що є критично важливими для таких систем, зокрема безпеці даних, користувацькому досвіду, інтеграції з існуючими системами та відповідності нормативним стандартам. Перш за все, проведено всебічний аналіз вимог медичних клінік до вебдодатків. Безпека даних є одним з найважливіших аспектів, оскільки медичні установи працюють з конфіденційною інформацією. У дослідженні розглянуто методи забезпечення конфіденційності та цілісності медичних даних, зокрема шляхом дотримання стандартів HIPAA та GDPR. React дозволяє легко інтегрувати сучасні засоби безпеки, такі як JWT (JSON Web Tokens) для аутентифікації та авторизації, а також використовувати захищені HTTP-з'єднання для передачі даних. Крім того, особлива увага приділена користувацькому досвіду (UX/UI), який має забезпечувати зручність та ефективність використання вебдодатку як медичним персоналом, так і пацієнтами. React дозволяє створювати динамічні та інтерактивні інтерфейси, що можуть адаптуватися до різних пристроїв та екранів, забезпечуючи таким чином високу якість користувацького досвіду. Це особливо важливо для медичних додатків, де точність і швидкість взаємодії можуть впливати на якість надання медичних послуг. Наведено детальний опис найкращих практик та підходів до розробки вебдодатку з використанням React. Розглянуто архітектуру компонентів, яка забезпечує модульність та повторне використання коду, що є ключовим фактором у створенні масштабованих та легко підтримуваних додатків. Досліджено різні методи керування станом у додатку, такі як використання Context API та Redux, що дозволяє ефективно керувати даними та забезпечувати стабільну роботу додатку. Описано рекомендації щодо налаштування маршрутизації за допомогою бібліотеки React Router, що дозволяє створювати динамічні та багатосторінкові додатки, забезпечуючи при цьому плавну навігацію та хороший користувацький досвід. Крім того, розглянуто способи забезпечення безпеки додатку, включаючи захист від XSS-атак, CSRF та інших типів вразливостей, що є критично важливим у медичних додатках.

[1] *бакалавр, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, kiriakidi.sofiya60@gmail.com*
[2] *канд. техн. наук, доцент, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: 0009-0006-2860-6575, krivenkoov@gmail.com*