UDC 004. 056.55:004.312.2 **doi: 10.26906/SUNZ.2025.3.208-212**

V. Rossikhin, Y. Tarapata, O. Iashchenko

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

METHODS OF DATA PROCESSING AND ANALYSIS IN A CORPORATE NETWORK

Abstract. Relevance. The relevance of studying methods of data processing and analysis in corporate networks is driven by the rapid development of information technologies, the growing volume of corporate data, and the need for prompt and accurate analysis to support effective decision-making. The importance of such methods is also linked to the challenges of ensuring cybersecurity in corporate networks, particularly the protection of information at all stages of its processing and transmission. The quality of the technologies and methods applied to data handling affects a company's performance, competitiveness, and ability to adapt to rapidly changing market conditions. Therefore, systematizing knowledge about the most effective methods of working with corporate data remains a relevant task for both modern science and practical applications. **The object of research** is data processing processes within corporate networks, including the collection, transmission, storage, filtering, analysis, and protection of information circulating within the information and communication infrastructure of a modern enterprise. These processes are considered in the context of their impact on the efficiency of the corporate information system, data security, support for managerial decision-making, and integration with analytical and cloud platforms. **Purpose of the article** is to investigate the main stages and methods of data processing and analysis in the corporate environment, including data collection, preprocessing, storage, analytical evaluation, and information security, as well as to analyze modern tools and platforms that ensure efficient and secure data handling at the scale of a large organization. **Research results.** During the study, a comprehensive analysis of the stages and methods of data processing in the corporate environment was conducted. A systemic model of the corporate data lifecycle was established, covering all key stages – collection, preprocessing, storage, analytics, and protection. Each of these stages requires the application of a specific class of technologies and presents its own implementation challenges. Tools for data collection and preprocessing were classified, with ETL processes, logging, aggregation technologies, sampling, and normalization playing an important role in ensuring data cleanliness and suitability for further analysis. Modern data storage platforms were analyzed, including cloud-based (Azure, AWS, GCP) and on-premises solutions. The effectiveness of using OLAP and machine learning in analytical processing was evaluated. The role of information security in corporate networks was addressed. The necessity of implementing cryptographic protection, access control, as well as anonymization and masking mechanisms was demonstrated as a response to the risks of data loss or compromise. **Conclusions.** Modern methods of data processing in corporate networks have been examined, covering the stages of data collection, preprocessing, storage, analytics, and information security. It has been established that effective data management is achievable only through a comprehensive approach that combines technical tools, software platforms, and security policies. Particular attention was given to the use of cloud technologies, ETL tools, OLAP analytics, and machine learning algorithms. The importance of cryptographic protection, anonymization, and access control was emphasized. The obtained results can be applied to improve the efficiency of enterprise information systems and to implement analytical solutions in business practice.

Keywords: corporate network, data processing, ETL, cloud technologies, data storage, OLAP, machine learning, information security, anonymization, cryptography, analytics, Big Data.

Introduction

In the current context of business digital transformation, corporate networks have become a key environment for the generation, transmission, storage, and processing of information. They serve not only as a communication tool between individual departments of an organization, but also as a foundation for building integrated information systems that support business process automation, decision-making, and increased competitiveness. The constant growth in the volume of data circulating within corporate networks generates the need for efficient, scalable, and secure methods of data processing. Modern corporate networks are characterized by complex architectures, heterogeneity of data sources, and the necessity to handle various types of data – from numerical tables and logs to unstructured text, images, and streaming sensor data. In such conditions, effective information flow management requires the implementation of specialized technologies at every stage of the data lifecycle: from collection and preliminary cleaning to analytical processing, visualization, and protection.

A critical component of this process is data preprocessing, which ensures the quality of information for further analysis. Methods such as normalization, aggregation, filtering, and compression enable the adaptation of data streams to the requirements of computing platforms. Additionally, at the data storage stage, the choice of appropriate infrastructure (on-premises, cloud, or hybrid) and the assurance of data availability and integrity are of fundamental importance.

Analytical processing tools play a crucial role in corporate information systems, including OLAP models [1], Data Mining technologies [2], and modern machine learning methods [3], which not only analyze large volumes of data but also identify patterns, generate forecasts, and support well-founded managerial decisions. Therefore, considerable attention is also given to platforms that support such processes, as well as to cloud-based analytics tools.

Equally important is the issue of data processing security – encryption, masking, anonymization, and access control has become integral components of corporate system architecture, as data loss or compromise may have critical consequences for an organization.

Thus, the article examines key approaches to data processing in corporate networks, systematizes the methods and tools used at all stages of the information

lifecycle, and explores ways to improve the efficiency and security of working with corporate data in large-scale information environments.

Analysis of Recent Research and Publications. Recent research highlights that effective data processing within corporate networks is critically important for modern organizations. Real-time data processing, the use of cloud technologies, process optimization, and the provision of reliable network communication are key aspects in achieving competitive advantages. At the same time, the implementation of these technologies is accompanied by challenges such as data security, the integration of heterogeneous sources, and resource management, which require further study and the development of effective strategies. Article [4] explores the impact of real-time data processing on the timeliness of business decision-making. It examines technologies such as Apache Kafka, Apache Flink, Google Cloud Dataflow, and Spark Streaming, which enable companies to respond quickly to market changes, optimize operations, and improve customer service. The study also analyzes implementation challenges, including data integration, scalability, data quality, and security. Article [5] focuses on factors affecting data processing efficiency, such as volume, variety, velocity, and veracity. It examines optimization techniques, including hardware improvements, software innovations, and architectural approaches such as distributed computing and cloud solutions. Special attention is given to the processing of unstructured data, integration of heterogeneous sources, and the assurance of energy efficiency and data confidentiality. Article [6] analyzes the use of cloud technologies for processing large volumes of industrial data. It discusses the advantages of cloud platforms in ensuring efficiency, security, and cost-effectiveness of data processing, as well as challenges related to data security, confidentiality protection, performance, and scalability. The study also considers the potential for integrating artificial intelligence and machine learning in this domain. Article [7] addresses the impact of modern technologies such as hyper automation, process mining, and predictive monitoring on business process management. It emphasizes the importance of integrating data from various sources to optimize processes and support well-informed decision-making. Article [8] presents a method for predictive resource allocation in networks supported by Multi-Access Edge Computing, aimed at ensuring the required quality of service. The study highlights the importance of effective resource management to guarantee reliable data processing in corporate networks.

The purpose of this work is to investigate data processing methods in corporate networks, with a focus on technological solutions that ensure efficient data collection, preprocessing, storage, analysis, and protection.

Main part

In the current conditions of global digitalization, the corporate network serves not only as a channel for information transmission, but also as a full-fledged information and communication environment forming the foundation of an organization's internal and external interactions. Corporate networks are the primary environment for collecting, routing, storing, and processing critical information related to production processes, financial activities, customer bases, marketing strategies, and more.

Corporate networks are typically built on a multi-level architecture that includes local, wide area, and virtual private networks. The main goal of such a structure is to provide stable access to information resources regardless of the user's physical location. The architecture of a corporate network may be centralized (with a single data center) or distributed (with multiple computing nodes). The data circulating within corporate networks is highly diverse in both structure and origin. It can be broadly categorized as structured, unstructured, or semi-structured. In addition, data may be historical, operational or streaming, which requires different processing approaches depending on the context of use.

Information flows in corporate networks are divided into internal (between departments, servers, internal users) and external (interactions with clients, suppliers, cloud services). Each type of flow has specific requirements in terms of transmission latency, reliability and fault tolerance, scalability, and security. Processing such flows requires the use of complex routing protocols, traffic monitoring tools, intelligent load balancing systems, and the application of access control and encryption policies. Given the growing number of devices and services, corporate networks are increasingly integrating cloud solutions, edge computing, and Internet of Things (IoT) components [9], which expand the possibilities for data collection and processing directly at the network edge. The initial stage of corporate data processing is crucial for subsequent analytics, as it is at this stage that the raw data sets are formed, determining the quality of analytical insights, the accuracy of forecasts, and the effectiveness of decision-making. Data collection and preprocessing approaches include a range of technical and logical processes aimed at transforming unstructured or raw data streams into structured, clean, and analysis-ready formats.

In a corporate environment, data may originate from a wide variety of sources: software systems, internal registers, event logs, API requests, IoT device sensors, network traffic, or external information systems. To integrate such heterogeneous streams, the following technologies are used: event and transaction logging, ETL processes, integration platforms, and streaming technologies. Data collection must meet criteria for reliability, accuracy, and completeness, while also accounting for the format and semantics of the source.

After the collection stage, the data often contain duplicates, missing values, logical inconsistencies, or input errors. Preprocessing involves data cleaning, normalization, standardization, type conversion, and filtering. These procedures are implemented using SQL queries, data processing languages, and specialized environments such as KNIME, RapidMiner, and SAS.

To ensure the scalability of corporate data storage and reduce storage costs, a variety of methods are widely used: compression algorithms to minimize the volume of stored data without losing integrity; aggregation for

analytical purposes; sampling to reduce computational load; and window aggregation for analyzing metric changes over time intervals. Compression and aggregation help reduce the load on storage and processing systems and accelerate query execution for databases or analytical modules.

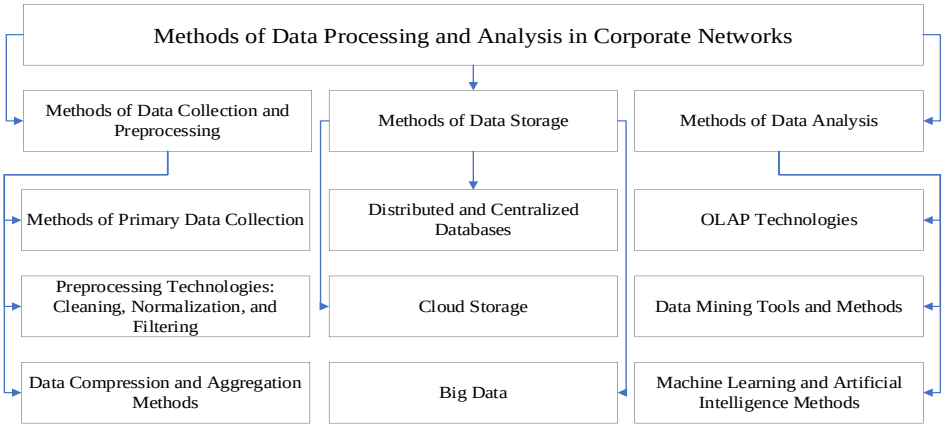

Fig. 1. Methods of Data Processing and Analysis in Corporate Networks

Data storage is one of the key stages in the information lifecycle within corporate networks. It is in storage systems where data is accumulated, organized, and structured for subsequent access, analysis, and use in business processes. In the context of modern companies, data storage encompasses not only the physical placement of information on servers but also concepts of logical organization, dynamic scalability, availability, and security. There are two primary approaches to database architecture in corporate environments: centralized and distributed. Centralized databases involve storing all data in a single data center or server repository. This simplifies administration and backup, but it scales poorly and is vulnerable to hardware failures. Distributed databases allow information to be stored across multiple nodes within the corporate network. They provide high fault tolerance, load balancing, and fast access to information in geographically distributed offices, but require more complex mechanisms for consistency (consensus) and replication. The choice between these models depends on data volume, access time requirements, risk levels, and existing IT infrastructure. Cloud storage has become an integral part of corporate infrastructure due to its high flexibility, scalability, and reduced hardware costs. Leading platforms offer automatic scaling based on data volume, high availability, integration with analytical and machine learning services, and support for access policies, encryption, version control, and logging. Hybrid models allow companies to store confidential or critical data locally, while hosting less sensitive information in the cloud –optimizing costs while maintaining control over data. Corporate systems increasingly operate with Big Data arrays. Traditional database management systems are insufficient for their efficient storage and processing, prompting the use of file systems for distributed environments, NoSQL solutions, and unified repositories for structured and unstructured data with flexible processing capabilities. A distinctive feature of Big Data storage is support for parallel processing, decentralized access management, protection against data loss through replication, and fast horizontal scalability. Once corporate data is collected, preprocessed, and stored, the next key stage is analytical processing, which involves identifying patterns, building decision-making models, and supporting strategic planning. In the modern corporate environment, analytics goes far beyond traditional reporting. It has evolved into a comprehensive data intelligence system combining classical statistical methods with machine learning and artificial intelligence algorithms.

Online Analytical Processing (OLAP) enables interactive exploration of large volumes of data across multiple dimensions. Its main advantages include quick access to aggregated indicators, category-based analytics, and operations that allow users to drill down or roll up information. OLAP systems are widely used in management accounting, financial planning, and marketing analysis. They require preprocessed, structured data and typically support regular reporting. For more complex analytical tasks, machine learning and artificial intelligence algorithms are increasingly applied. These techniques model non-linear dependencies, handle unstructured data, enable predictive analytics, and even support automated decision-making. Typical use cases for machine learning in corporate environments include demand forecasting and supply chain optimization; risk assessment in banking and insurance; personalization of marketing campaigns; sentiment analysis based on customer reviews or social media; automatic document classification; and recommendation systems. In corporate settings, effective data processing requires not only appropriate methods but also powerful tools that can scale, integrate with various sources, automate processes, and ensure the reliability of results. Depending on their objectives, companies use both universal platforms and specialized solutions for ETL, analytics, storage, or visualization. ETL is a core mechanism for preparing data for analysis. Its main function is to extract data from various sources, transform it into the appropriate format, and load it into a data warehouse. Among the most popular ETL tools are Apache NiFi, Talend, Microsoft SQL Server Integration Services, Pentaho Data Integration, and Airflow. These tools automate data collection and transformation, reduce the workload on IT departments, and ensure data consistency across storage systems.

Analytical platforms enable the creation of dynamic reports, dashboards, visualizations, and automated analytical models. ERP and CRM platforms also include built-in analytical functions, allowing the integration of planning, accounting, customer interaction, and analytics processes within a unified environment.

Table 1 presents the key capabilities of three cloud platforms: Microsoft Azure, AWS, and Google Cloud.

Table 1 – **Comparative Analysis of Models**

Characteristic	Microsoft Azure	Amazon Web Services	Google Cloud Platform
Core Services	Azure Synapse, Data Factory, ML Studio	Redshift, Glue, SageMaker	BigQuery, Dataflow, Vertex AI)
Big Data Analytics	Integration with Power BI, support Data Lake	Powerful toolset for Data Lakehouse	Vertex AI, AutoML Tables, TensorFlow
Machine Learning	Platform AutoML, Azure ML Designer	SageMaker SageMaker with advanced customization	Vertex AI, AutoML Tables, TensorFlow
Programming Language Support	Python, R, SQL, .NET	Python, Scala, Java, SQL	Python, SQL, TensorFlow
Security	Azure Active Directory, Encryption, RBAC	IAM, KMS, DLP	Identity and Access Management, Data Loss Prevention

Each of the platforms has its own specialization: Azure offers close integration with office products, AWS provides flexibility and scalability, and GCP is optimized for analytics and artificial intelligence.

It is also important to mention the issue of ensuring data security in corporate networks, as the constant growth in the volume of information circulating within corporate systems makes information security a matter of primary importance. Data is a strategic asset of an enterprise, and its loss, modification, or unauthorized access can lead to significant financial, reputational, and legal consequences. Therefore, at all stages of the data lifecycle – from collection to archiving – reliable protection measures must be implemented in accordance with the internal policies of the organization and international standards. A comprehensive approach to security requires implementing protection at every stage of data processing: at the collection stage, this includes the use of secure transmission protocols, authentication of data sources, and client-side encryption; at the processing stage, this involves isolation of processing environments, access control to computing resources, and logging of data operations; at the storage stage, encryption at the disk level, access control to file systems, backup, and data retention policies are applied.

The integration of cryptographic mechanisms into software and network protocols allows information to be protected even in the event of physical access to storage media or traffic interception. In addition to encryption, methods for restricting access to personal or confidential information are increasingly applied, including data masking, anonymization, pseudonymization, and differential privacy. These methods make it possible to comply with legislation on the protection of personal data while maintaining the analytical value of the data.

Conclusions

As a result of the conducted research, it has been established that effective data processing in corporate networks is a critically important prerequisite for the stable functioning of information systems, strategic enterprise management, and ensuring competitiveness under conditions of digital transformation. The examined structure of data processing methods covers the entire data lifecycle – from the moment of collection to the final stages of analytical use and protection.

The effectiveness of data collection and preprocessing within corporate networks determines the subsequent quality of all information analytics. Proper implementation of these stages allows for increased accuracy of models, minimized risk of erroneous conclusions, and preservation of data integrity within the corporate environment. Efficient data storage is also a vital condition for ensuring the stable operation of the information infrastructure. Modern approaches to storage organization enable flexible adaptation of corporate storage systems to business needs and provide a reliable foundation for analytical data processing.

Proper implementation allows not only for assessing the current situation in a company but also for projecting its future development, adapting strategies to market changes based on evidence-driven data. Modern tools and platforms for data processing offer enterprises a broad range of technologies for tasks of any scale. Effective use of these tools significantly improves the quality of management, reduces the time required for analytical operations, and enables rapid responses to market dynamics.

At the same time, ensuring the security of data processing requires a systematic approach that includes technical, organizational, and procedural measures. A well-considered combination of classical protection methods – encryption, access control, and authentication – with modern practices such as anonymization, machine-based incident monitoring, and continuous auditing allows organizations to minimize the risks of data leaks, manipulations, and privacy violations within a complex and dynamic corporate environment.

REFERENCES

1. U. Sirin, A. Ailamaki. Micro-architectural Analysis of OLAP: Limitations and Opportunities. Cornell University, 2019. 13 p. https://doi.org/10.48550/arXiv.1908.04718
2. Wu, W.-T., Li, Y.-J., Feng, A.-Z., et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. Military Medical Research, 8(1), 2021. 44 p. https://doi.org/10.1186/s40779-021-00338-z
3. Flach P. A. Machine Learning: The Art and Science of Algorithms that Makes Sense of Data. Cambridge: Cambridge University Press, 2012. 291 p. https://doi.org/10.1017/CBO9780511973000
4. M, Achanta.The Impact of Real - Time Data Processing on Business Decision - making." International Journal of Science and Research (IJSR), vol. 13, no. 7, 2024, pp. 400-404, https://www.doi.org/10.21275/SR24708033511

5. R.Abu-Zaid, A.Hammad. Streamlining Data Processing Efficiency in Large-Scale Applications: Proven Strategies for Optimizing Performance, Scalability, and Resource Utilization in Distributed Architectures. International Journal of Machine Intelligence for Smart Applications, 14(8), 2024. P. 31-49. https://dljournals.com/index.php/IJMISA/article/view/27 .
6. Ziyan Yao. Application of cloud computing platform in industrial big data processing. Penn State University, 2024. 8 p. https://doi.org/10.48550/arXiv.2407.09491
7. L. Ackermann, M. Käppel, L. Marcus, L. Moder, et al. Recent Advances in Data-Driven Business Process Management. Cornell University, 2024. 34 p. https://doi.org/10.48550/arXiv.2406.01786
8. W. Symbor, L. Falas. Ensuring Reliable Network Communication and Data Processing in Internet of Things Systems with Prediction-Based Resource Allocation, MDPI Sensors. Vol. 25, iss. 1, 2025. 33 p. https://doi.org/10.3390/s25010247
9. O. Aouedi, T. Vu, A. Sacco, D. Nguyen, K. Piamrat, G. Marchetto, Q. Pham. A Survey on Intelligent Internet of Things: Applications, Security, Privacy, and Future Directions. IEEE Communications Surveys & Tutorials, 2024. 56 p. https://doi.org/10.1109/COMST.2024.3430368

Received (Надійшла) 28.05.2025
Accepted for publication (Прийнята до друку) 20.08.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Россіхін Василь Васильович – доктор юридичних наук, проректор з інноваційно-корпоративної роботи та адміністрування, професор кафедри філософії, Харківський національний університет радіоелектроніки, Харків, Україна;
Vasyl Rossikhin – Doctor of Low, Vice-Rector for Innovative Corporate Work and Administration, Professor of Philosophy Department, Kharkiv National University of Radio Electronics Kharkiv, Ukraine;
e-mail: vasyl.rossikhin@nure.ua; ORCID Author ID: http://orcid.org/0000-0003-3423-8896;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57207769187.

Тарапата Ярослав Ігорович – студент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Yaroslav Tarapata – student, Department of Electronic Computers, Kharkiv National University of Radio Electronics Ukraine;
e-mail: yaroslav.tarapata@nure.ua; ORCID Author ID: https://orcid.org/0009-0008-4691-260X.

Ященко Олександр Миколайович – студент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Oleksandr Iashchenko – student, Department of Electronic Computers, Kharkiv National University of Radio Electronics Kharkiv, Ukraine;
e-mail: oleksandr.iashchenko1@nure.ua; ORCID Author ID: https://orcid.org/0009-0007-1033-365X.

Методи обробки та аналізу даних в корпоративній мережі

В. В. Россіхін, Я. І. Тарапата, О. М. Ященко

Анотація. Актуальність. Актуальність теми дослідження методів обробки та аналізу даних у корпоративних мережах зумовлена швидким розвитком інформаційних технологій, зростанням обсягів корпоративної інформації та необхідністю оперативного й точного аналізу для ефективного прийняття рішень. Важливість дослідження таких методів також пов'язана з викликами щодо забезпечення кібербезпеки корпоративних мереж, зокрема захисту інформації на всіх етапах її обробки та передачі. Від якості застосованих технологій та мето-дів роботи з даними залежить продуктивність компанії, її конкурентоспроможність та здатність адаптуватися до швидко змінюваних умов ринку. Тому систематизація знань про найбільш ефективні методи роботи з корпоративними даними є актуальним завданням сучасної науки та практики. **Об'єкт дослідження**: процеси обробки даних у корпоративних мережах, що включають збір, передавання, зберігання, фільтрацію, аналіз та захист інформації, яка циркулює в межах інформаційно-комунікаційної інфраструктури сучасного підприємства. Ці процеси розглядаються в контексті їх впливу на ефективність функціонування корпоративної інформаційної системи, забезпечення безпеки даних, підтримку управлінських рішень та інтеграцію з аналітичними й хмарними платформами. **Мета статті**: дослідження основних етапів та методів обробки та аналізу даних у корпоративному середовищі, які включають збір, попередню обробку, зберігання, аналітичний аналіз та захист інформації, а також аналіз сучасних інструментів та платформ, що забезпечують ефективну й безпечну роботу з даними в масштабах великої організації. **Результати дослідження**. У ході дослідження було здійснено комплексний аналіз етапів і методів обробки даних у корпоративному середовищі. Встановлено системну модель життєвого циклу корпоративних даних, яка охоплює всі ключові стадії – збір, попередню обробку, зберігання, аналітику та захист. Кожен із цих етапів вимагає застосування окремого класу технологій і має власні виклики у реалізації. Класифіковано інструменти збору й попередньої підготовки даних, серед яких важливу роль відіграють ETL-процеси, логування, технології агрегації, семплінгу та нормалізації, що забезпечують чистоту й придатність даних до подальшого аналізу. Проаналізовано сучасні платформи зберігання даних, включно з хмарними (Azure, AWS, GCP) та локальними рішеннями. Оцінено ефективність використання OLAP та машинного навчання в аналітичній обробці. Розкрито роль інформаційної безпеки в корпоративних мережах. Доведено необхідність впровадження криптографічного захисту, контролю доступу, а також механізмів анонімізації та маскування як відповідь на ризики втрати або компрометації даних. **Висновки.** розглянуто сучасні методи обробки даних у корпоративних мережах, що охоплюють етапи збору, попередньої обробки, зберігання, аналітики та захисту інформації. Встановлено, що ефективне управління даними можливе лише за умови комплексного підходу, який поєднує технічні засоби, програмні платформи та політики безпеки. Особливу увагу приділено використанню хмарних технологій, інструментів ETL, OLAP-аналітики та алгоритмів машинного навчання. Підкреслено важливість криптографічного захисту, анонімізації та контролю доступу. Отримані результати можуть бути використані для підвищення ефективності роботи інформаційних систем підприємств і впровадження аналітичних рішень у бізнес-практику.

Ключові слова корпоративна мережа, обробка даних, ETL, хмарні технології, зберігання інформації, OLAP, машинне навчання, інформаційна безпека, анонімізація, криптографія, аналітика, Big Data.