https://doi.org/10.31891/csit-2024-1-13
UDC 004.9: 347.151

Olexandr HOLOTA, Vasyl KUT
Uzhhorod National University
Sergyi PASICHNYK, Oleh HRYBOVSKYI
Lviv Polytechnic National University

ANALYSIS OF DATA ORGANIZATION SYSTEM ARCHITECTURES OF THE SMART EUROPEAN REGION "CENTER OF EUROPE" BASED ON THE DATA LAKE TECHNOLOGICAL SOLUTION

In the contemporary landscape of digital transformation, the concept of a "smart region" emerges as a key strategy for regional development, utilizing data integration technologies to enhance the quality of life, optimize service delivery, and promote sustainable growth. The designation of the "Centre of Europe" as a smart region is justified by its strategic geographical location, encompassing Transcarpathia and adjacent areas of neighboring countries, coupled with a relatively low population density. This unique context creates compelling premises for rethinking traditional urban frameworks of smart cities on a broader scale, transforming them into a comprehensive smart region model tailored to the specific needs and opportunities of Central Europe.

The rationale for implementing the "Centre of Europe" as a smart region stems from its advantageous position at the heart of the continent, which generates unprecedented opportunities for cross-border collaboration, innovation, and connectivity. The low population density of the region further highlights the potential for the adoption of advanced technological solutions capable of addressing a wide array of challenges and needs. By achieving the status of a smart region, the "Centre of Europe" aspires to become a leading example of how digital and technological innovations can be utilized to create an inclusive, efficient, and progressive regional ecosystem.

Objective: The primary objective of this research is to conduct a comparative analysis of advanced data organization system architectures and their role in facilitating the transformation of the "Centre of Europe" into an intelligent European region. It also aims to explore the capabilities of utilizing data lakes for the effective management of large data sets and regional development, taking into account the unique geographical and demographic characteristics of the region.

The subject of the study encompasses the architectures of big data organization systems in the smart region of the "Centre of Europe". The object of the study involves the processes of storing and processing large data within the smart region.

Main characteristics: This article illuminates the main aspects of implementing big data processing technologies in the smart region of the "Centre of Europe", necessitating scalable and flexible solutions for large data storage, the use of efficient data management practices, and the integration of advanced technologies such as cloud computing, the Internet of Things (IoT), and artificial intelligence (AI) to enhance data processing and analytical capabilities. Furthermore, recommendations for regional authorities on adopting these technologies have been formulated, highlighting the importance of collaboration, support for investments in the region's digital infrastructure as key components for achieving the smart region vision.

The construction of modern, high-tech data storage facilities using the Data Lake architecture enables the simultaneous use and processing of large volumes of data from various fields and sources. The authors have proposed conceptual models for forming data complexes from different functional areas of the smart region.

Keywords: smart European region, data organization system architecture, "Center of Europe", big data, Data Lake

Олександр ГОЛОТА, Василь КУТ
Ужгородський Національний університет
Сергій ПАСІЧНИК, Олег ГРИБОВСЬКИЙ
Національний університет «Львівська політехніка»

АНАЛІЗ АРХІТЕКТУР СИСТЕМИ ОРГАНІЗАЦІЇ ДАНИХ РОЗУМНОГО ЄВРОПЕЙСЬКОГО РЕГІОНУ «ЦЕНТР ЄВРОПИ» НА ОСНОВІ ТЕХНОЛОГІЧНОГО РІШЕННЯ ОЗЕР ДАНИХ

У сучасному ландшафті цифрової трансформації концепція "розумного регіону" виступає як ключова стратегія регіонального розвитку, використовуючи технології інтеграції даних для підвищення якості життя, оптимізації надання послуг та сприяння сталому зростанню. Призначення проєкту розумного регіону "Центр Європи" обумовлене його стратегічним географічним розташуванням, що охоплює Закарпаття та прилеглі райони сусідніх країн, разом з відносно низькою густотою населення. Цей унікальний контекст створює переконливі передумови для переосмислення традиційних міських рамок розумного міста на більш широкому рівні, перетворюючи їх на комплексну модель розумного регіону, адаптовану до конкретних потреб і можливостей Центральної Європи.

Обґрунтування реалізації проєкту розумного регіону "Центр Європи" випливає з його вигідного положення в серці континенту, що генерує безпрецедентні можливості для транскордонної співпраці, інновацій та зв'язку. Низька густота населення регіону додатково підкреслює потенціал для впровадження передових технологічних рішень, які можуть вирішувати широкий спектр викликів і потреб. Набуттям статусу розумного регіону "Центр Європи" – регіон прагне стати провідним прикладом того, як цифрові та технологічні інновації можуть бути використані для створення інклюзивної, ефективної та прогресивної регіональної екосистеми. Мета: Основною метою цього дослідження є порівняльний аналіз передових архітектур систем організації даних та їх ролі у сприянні перетворенню "Центру Європи" на інтелектуальний європейський регіон, а також вивчення можливостей використання концепції озер даних для ефективного управління великими наборами даних та регіонального розвитку з врахуванням унікальних географічних та демографічних характеристик регіону. Предметом дослідження є архітектури систем організації великих даних розумного регіону "Центр Європи". Об'єктом дослідження є процеси збереження та опрацювання великих даних у розумному регіоні.

Основні характеристики: У статті висвітлюється основні аспекти впровадження технологій опрацювання великих даних у розумному регіоні "Центрі Європи", які передбачають необхідність масштабованих та гнучких рішень для зберігання великих даних, використання ефективних практик управління даними та інтеграції передових технологій, таких як хмарні обчислення, Інтернет речей (IoT) та штучний інтелект (AI), для покращення опрацювання даних та аналітичних можливостей. Крім того, сформовано рекомендації для регіональних органів влади щодо впровадження цих технологій, підкреслюючи важливість співпраці, підтримки інвестицій у цифрову інфраструктуру регіону як ключових компонентів для досягнення візії розумного регіону.

Побудова сучасних високотехнологічних сховищ даних з використанням архітектури озер даних (Data Lake) створює можливість одночасного використання та опрацювання великих об'ємів даних з різних сфер та джерел. Авторами запропоновано концептуальні моделі формування комплексів даних з різних сфер функціонування розумного регіону.

Ключові слова: розумний європейський регіон, архітектура системи організації даних, «Центр Європи», великі дані, Data Lake.

Introduction

Motivation. The era of digitalization and technological progress leads to the global changes in all spheres of human activity. Data consumers, businesses, government structures, and the scientific community face the need for an optimal and efficient use of voluminous information flows. In the context of expanding digital integration geography, the concept of a smart region that is a cohesive system of cities and settlements based on the principles of intelligent data analysis, becomes particularly important. The smart European region "Center of Europe" possesses a unique potential to become a pilot project where big data becomes the key to harmoniously combining technologies, culture, economy, and social innovations of several regions of neighboring countries. However, despite the obvious advantages, research in this area remains limited, particularly due to the complexity of regional characteristics, requiring detailed understanding of data organization system architectures. Therefore, the study of big data use in the context of the smart region "Center of Europe" not only opens new horizons for the development of information technologies at the regional level but also proposes the latest approaches to management, development, and modernization of the region in the face of modern challenges.

Analysis of Sources. In the modern scientific context, data storage systems are considered as a key component of effective management and analysis of big data. Works by W.H. Inmon [1], R. Kimball [2] focus on the main principles and architectural solutions related to data storage methods, with particular attention to storage, integration, and data processing issues. Scientific literature often emphasizes specific types of systems. For example, M. Stonebraker, U. Cetintemel, etc. [3] explore the issues of creating and optimizing data lakes, highlighting their advantages in storing unstructured data.

The objective of this work is to analyze the architectural concept of the data lake organization system in the context of information technologies of the smart European region "Center of Europe". The main focus of the research is to identify the key features of this architectural concept and its impact on the implementation of the smart region system "Center of Europe".

The goal is also to systematize and analyze the aforementioned concepts to identify their advantages and disadvantages, as well as to determine areas of their application and potential interactions between them. Through in-depth analysis of literature sources and real use cases, the researcher aims to identify optimal conditions and scenarios for the application of each architectural concept within the smart information infrastructure of the European region "Center of Europe". The results of this study are intended to serve as a basis for developing recommendations and strategies for implementing architectural solutions in practical situations requiring the organization of large volumes of data. Considering the specifics of the European region "Center of Europe", the work proposes a correspondence between the chosen architecture and the needs for storage, processing, and analysis of data in modern information systems. This research plays a crucial role in the development of intelligent technologies in the European region "Center of Europe", promoting the improvement of strategies for smart integration and data analysis, which is a necessary condition for the further development of this geographical region.

Data Storage Issues and the Development of Data Storage Systems

In the modern digital landscape, where data plays a critical role in strategy selection and innovation achievement [4], digitization covers absolutely all spheres of human activity and there is a need for distributed processing and storage of information [5], the challenges of storage, processing, and management of data become paramount [6]. The increasing volumes of data and the diversity of information present the new challenges for information systems and technologies [7]. The global issue of data storage necessitates new approaches and architectural solutions [8]. One of the key challenges is the efficient storage of large volumes of data from various sources [9]. Ensuring the integrity, confidentiality, and accessibility of information is of great importance [10, 11]. A rapid access to data becomes crucial for making informed decisions and ensuring competitiveness [12]. The problem also lies in processing various data formats - from structured to unstructured ones [13]. The diversity of sources and types of information requires specialized approaches to analysis and processing. This article explores various architectural concepts aimed at solving the issues of data storage and processing. We consider both traditional and modern approaches, such as "data lakes" and others. Overall, the issue of data storage requires the selection of optimal architectural approaches for effective processing and access to information [14]. Understanding the different concepts of data storage methods and their impact on the operation of organizations helps professionals in developing and managing information systems to make informed decisions [15]. Modern data architecture refers to the latest approach to designing and implementing data systems that are capable of effectively addressing challenges and leveraging opportunities presented by the ever-increasing volume, diversity, and speed of data transfer.

Key characteristics and components of modern data architecture [16]:

Integration and Data Loading. Modern methods offer powerful data integration capabilities, collecting information from various sources, such as structured databases, unstructured data, and streaming transfer. This ensures rapid data loading in real time.

Data Storage and Management. Various data storage technologies are used, such as data lakes or databases, providing fast and efficient access to data [17].

Data Processing and Analytics. Different methods of data processing and analytics are supported, including batch processing, stream processing, and real-time analytics.

Cloud and Hybrid Systems: Modern methods actively utilize cloud computing and support hybrid environments.

Data Management and Security. Priority is given to data management and security practices.

Data Access and Analytics for Users. Tools are provided for self-service access and data analysis.

Scalability and Flexibility: Solutions are designed to scale according to the needs.

DataOps and DevOps Principles. Collaboration and flexibility in data-related projects are supported.

The advantages of this approach include improved data access, accelerated time to results, enhanced decision-making, and the ability to use advanced analytical methods [18]. However, it should be noted that modern data architecture requires a personalized approach. Organizations must carefully assess their needs and capabilities and build infrastructure accordingly [19].

Data Lake

A data lake is a centralized data storage service that stores large volumes of structured, semi-structured, and unstructured data in their raw or original format. Unlike a data warehouse or data mart that adheres to a predefined schema, a data lake allows the storage of data in its original form, without the need for prior transformation [20]. In a data lake, data is stored in a flat architecture, typically using a distributed file system such as Apache Hadoop, or cloud storage systems like Amazon S3 or Azure Data Lake Storage [21]. This repository can accommodate various types of data such as text files, documents, sensor data, log files, social media streams, etc. [22].

Key characteristics and advantages of data lakes:

Diverse Data Storage. A data lake can store structured, semi-structured, and unstructured data from various sources without imposing a specific schema [23]. This flexibility allows organizations to store vast amounts of data in their primary format, maintaining data integrity for future analysis.

Scalability. Data lakes are designed to handle massive volumes of data, making them highly scalable. They can scale horizontally by adding more storage nodes or utilizing cloud storage resources.

Schema-on-Read. In a data lake, the structure and schema of data are applied at the time of data access or query, not during their acquisition. This approach provides greater flexibility for data exploration and analysis, as users can interpret and transform data according to their specific needs during an analysis.

Data Exploration and Discovery. Data lakes facilitate data exploration, offering a wide range of sources and formats. Analysts and researchers can explore various datasets, combine, and correlate data from different sources, discovering hidden patterns or insights.

Integration with Analytical Tools. Data lakes integrate with various analytical and processing tools, such as Apache Spark and Apache Hive, allowing users to perform complex analyses and derive valuable insights from data stored in the lake.

Data Management and Security. Data lakes require effective management and security methods to ensure data confidentiality, compliance, and access control [24]. Data management practices, such as data cataloging, metadata management, and data lineage tracking, help maintain data quality, track data origin, and establish data usage policies.

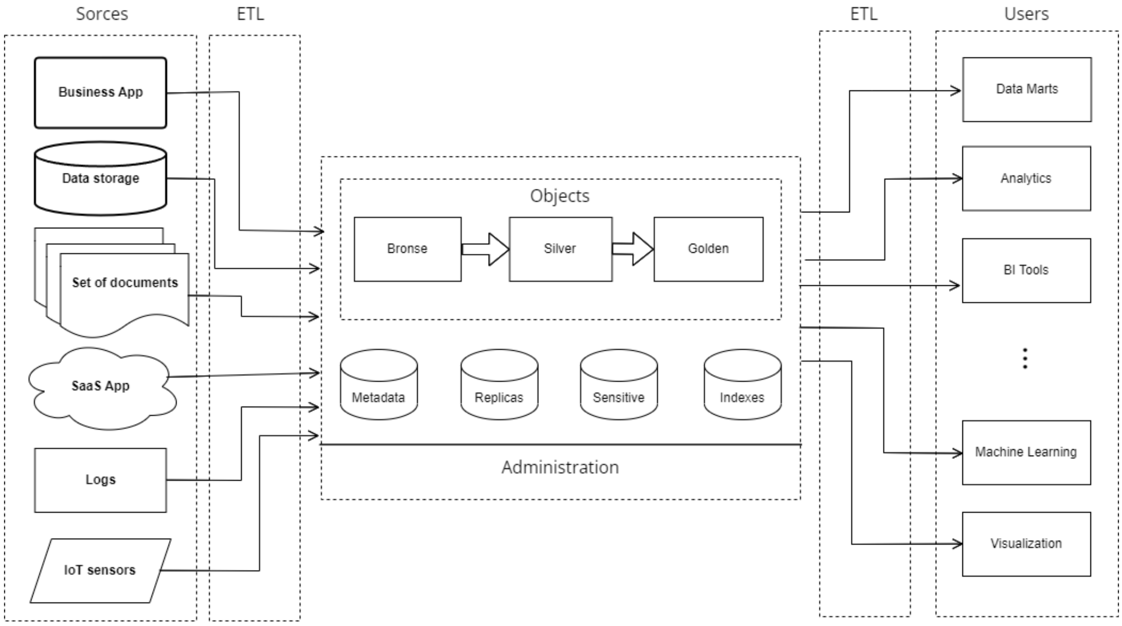

Fig. 1. Data Lake structure visualization

Data Democratization. Data lakes promote data democratization, providing self-service data access for different user groups within an organization [25]. Business users, analysts, and researchers can access the lake, discover relevant datasets, and derive insights without significant reliance on IT or data engineering teams. The input data for the developed "Methodology for predicting the success of the implementation of scientific IT projects based on the analysis of their characteristics" is a set of indicators on the basis of which requirements are formed, as well as the category of project success is determined [26]. Data democratization refers to the process of granting users an access to data and analytical tools, typically without direct intervention or support from the IT department. At the same time, it is possible to use a multi-agent system for taking into account the information of the subject field, which: performs parsing of documents based on semantic similarity and contributes to the formation of the ontological presentation of documents [27]. The goal of data democratization is to enable self-service access to information for informed decision-making across various organizational levels. Key aspects of data democratization include:

Self-Service Data Access. Users can independently access databases or other information sources instead of relying on IT specialists.

Use of Intuitive Tools. The market offers a lot of data visualization and analysis tools designed to be user-friendly, even for non-technical users.

Supporting Decision-Making. Access to data enables professionals from various fields (e.g., marketers, financiers, managers) to make decisions based on factual information, not just intuition.

Data Security and Management. While democratization aims to provide broader data access, it's important to ensure proper control and data security, potentially including role-based access models, user action audits, and other control mechanisms.

Maintaining Data Quality. Data access is only useful if the data is accurate and up-to-date. Therefore, some measures must be taken to ensure high data quality given the rapid development of data analysis technologies, data democratization is becoming a key component of many modern organizations, fostering a culture of data-based decision-making and enhancing market competitiveness.

It's important to note that while a data lake provides a flexible and scalable storage solution, it requires proper data management and governance practices to ensure data quality, data integration, and appropriate access control [25]. Without proper management, a data lake can become a data swamp, where data becomes unorganized and challenging to find, making it nearly impossible to derive valuable insights. In summary, a data lake is a centralized repository that stores a variety of raw data, allowing organizations to capture, store, and analyze large volumes of information in a flexible and scalable manner, facilitating data exploration and discovery for various analytical and decision-making purposes. Regarding the potential use of a data lake in developing the smart region "Center of Europe," the following areas can be highlighted:

Initial Information Accumulation. A data lake is an ideal platform for collecting and storing large volumes of data from various information sources in a smart region. This may include data from sensors, geodata, citizen information, government registries, and more. Centralized storage facilitates an easy and quick access to required data.

*Data Transformation and Optimization***.** Before data can be used for analytics and decision-making, it often requires processing and normalization. Data lakes allow for such transformations "on the fly," using the "schema-on-read" principle, providing flexibility in data utilization.

Geodata Aggregation. In smart regions like "Center of Europe," the focus is on integration and interaction between cities and communities. Data lakes can aggregate geodata from all types of sensors, creating centralized cartographic images for planning and monitoring.

Integration with Data Warehouses. While data lakes focus on storing raw data, data warehouses help organize, aggregate, and optimize this data for specific analytical needs. Integration between the lake and the warehouse can provide a harmonious approach to data management and analysis at the regional level.

Facilitating Data Democratization. As smart regions involve active citizen participation, data democratization through data lakes can be key to collaboration and interaction between government bodies, the private sector, and the public.

Ensuring Security and Compliance. An important part of any data infrastructure is information protection. Data lakes can implement stringent security mechanisms and ensure compliance with regional and international standards for data processing and storage.

In conclusion, the approach to using a data lake in the smart region "Center of Europe" can serve as an essential component of the successful implementation of the smart region concept, providing integration, analysis, and access to key information at all management levels.

Possible Use Cases.

The Data Lake concept, thanks to its flexibility and ease of use, can be applied within a smart region to solve a wide range of tasks. The most interesting aspect of using Data Lake storage is the ability to create common information domains (both structured and unstructured), areas similar in origin, or data usage, to which statistical analysis and artificial intelligence analysis approaches can be applied. It is possible to find common patterns and data connections within one domain and, most importantly, at the intersection of several data domains. Potential domains that can be created:

- *Social Domain***.** Data related to education, social protection systems, healthcare, libraries, support for national minorities.
- *Natural Resources Domain*. The state of natural resources, seismological status, weather, etc.
- *Economic Domain*. Information on the economic activity of the population, tax burden, etc.
- *Infrastructure Domain***.** Information about transport flows, road conditions, etc.
- *International Cooperation Domain*. Considering that the smart region "Center of Europe" is located at the intersection of many countries, information on integration and communication of different legislation and restrictions should also be stored in a separate domain.

Other Domains. The necessity of which will become more evident during real integrations and analysis.

Examples of creating services based on the intersection of data from different information domains:

Health and Environment Monitoring. Combining data from the Social and Natural Resources Domains to monitor public health trends in relation to environmental changes, such as air quality or weather patterns.

Economic Forecasting. Using data from the Economic Domain in conjunction with information from the Infrastructure Domain to predict economic trends based on transportation and infrastructure development.

Disaster Response and Planning. Integrating data from the Natural Resources Domain with Social and Infrastructure Domains to improve disaster response strategies and plan for future infrastructure development in areas prone to natural disasters.

Education and Labor Market Analysis. Combining data from the Social Domain with Economic Domain insights to align educational programs with current and future labor market demands.

Cross-Border Cooperation. Using the International Cooperation Domain to facilitate legal and logistical aspects of cross-border projects is to incorporate data from other domains to ensure compliance and efficiency.

Community Health Initiatives. Leveraging data from the Social and Economic Domains is to identify areas in need of healthcare resources and plan for community health initiatives.

Transportation and Urban Planning. Utilizing data from the Infrastructure Domain in combination with Social and Economic Domains is to optimize urban planning and public transportation routes according to population needs and economic activities.

Cultural and Minority Support Programs. Integrating data from the Social Domain, particularly focusing on national minorities is to develop targeted cultural and support programs.

These examples demonstrate how the intersection of data from various domains can lead to innovative solutions and services in a smart region, enhancing the quality of life, economic development, and environmental sustainability.

Traffic Management

Utilizing the Infrastructure Domain, International Cooperation Domain, and Natural Resources Domain. One example of Data Lake application within a smart region is an automated traffic management system. Such a system uses data from surveillance cameras and analyzes it with artificial intelligence (AI) systems to obtain road traffic load statistics. In all areas of the smart region, there is already a developed system of various quality road traffic surveillance cameras. Given that the Data Lake is optimized for storing large amounts of unstructured data, this information can be easily analyzed by artificial intelligence to create road traffic status information and, for example, the condition of road surfaces. Surveillance cameras, placed at key road junctions, collect real-time traffic data. This data is then uploaded to the Data Lake for further analysis. AI systems analyze the video data to determine the quantitative and qualitative characteristics of traffic (e.g., the number of vehicles, speed of movement, etc.). Based on the obtained data, the system automatically adjusts the operation mode of traffic lights, increasing or decreasing passage time to improve road traffic capacity. The system also provides reports and analytics to local authorities to assist in planning infrastructure improvements. Under certain conditions, the road restrictions can be automatically created in case of road surface problems (earthquake, landslide, heavy rain, river flooding, etc.), or border crossing needs, etc. By integrating with municipal and social and medical services (ambulance, fire brigade, police, etc.), green corridors can be created for emergencies.

Predicting Weather Disasters

Utilizing the Natural Resources Domain and the Social Domain. We propose using Data Lake for predicting weather disasters, providing critically important information for prevention and response to emergencies. For this project, Data Lake will collect weather data such as:

Satellite Images. These can be used to detect anomalies in weather conditions, such as changes in temperature, humidity, or atmospheric pressure.

Data from Meteorological Stations. This can be used to monitor current weather conditions.

Historical Weather Data. This can be used to identify trends that may indicate the possibility of a disaster.

Information from IoT Sensors. This can be used for real-time weather condition monitoring.

This will provide a comprehensive overview of weather conditions and assist in accurately predicting weather disasters. Artificial Intelligence (AI) systems can analyze the integrated data to identify patterns that indicate the possibility of weather disasters. AI uses machine learning algorithms to analyze data, enabling predictions of hurricanes, floods, heavy rains, and other extreme weather conditions. For instance, AI systems can be used for:

- Predicting Hurricane Trajectories
- Assessing Flood Strength
- Identifying Potential Dam Break Points
- Predicting Soil Erosion

This predictive capability is critically important for timely response to potential dangers. Integrating and analyzing weather data in Data Lake can significantly improve the efficiency of responses to weather disasters. This not only allows smart cities to better prepare and respond to extreme weather conditions but also enhances the overall safety and quality of life of residents. There are several challenges in using Data Lake for predicting weather disasters. One of the main challenges is ensuring accurate data integration. Weather data may come from various sources, and it's important that they are correctly integrated and cleansed before analysis. Another challenge is the effective analysis of data. Analyzing weather data can be a complex task, as many different factors need to be considered. AI systems can be used to automate this process, but it's important to ensure they are trained on a sufficient data set to provide accurate predictions. Despite these challenges, using Data Lake for predicting weather disasters has the potential to significantly improve the effectiveness of responding to weather disasters. Population resettlement information can be used for early warnings to the population, organizing evacuations, signaling social services for residents who need extra checking (health problems and restrictions, remote living, etc.).

Earthquake Prediction

Utilizing the Natural Resources Domain, Social Domain, Economic Domain, and Infrastructure Domain. Earthquakes are among the most dangerous natural phenomena, capable of leading to significant human casualties and material damages. The use of Data Lake for earthquake prediction is a promising direction that has the potential to significantly reduce the risks associated with these disasters. Data Lake allows the integration and analysis of vast amounts of data from various sources, enhancing the accuracy and timeliness of predictions. This data can include:

- Data from Seismographs. Seismic data is used to record earthquakes.
- Data on Geological Structures. There are such as faults and fold belts, potential sources of earthquakes.
- Weather Data. Meteorological data that can influence seismic activity.
- Historical Earthquake Data. It can help identify patterns and trends indicating the possibility of future earthquakes.
- Information from IoT Sensors. It can be used for real-time weather condition monitoring.

Integrating this data into Data Lake allows researchers to detect complex patterns that are not identifiable using traditional methods, leading to more accurate earthquake forecasts. Using machine learning and other analytical tools, researchers analyze Data Lake data to identify potential earthquakes, including predicting the location, time, and potential strength of future seismic events. Predictions obtained from Data Lake are used by local authorities for civil defense planning. These forecasts can help rescue services prepare for possible earthquakes and save lives. Among the challenges of earthquake prediction is ensuring data accuracy and adapting to rapid technological changes. Data from seismographs may be incomplete or distorted, leading to false predictions. Additionally, machine learning technologies are constantly evolving, and researchers must continually refine their algorithms to ensure prediction accuracy. The future development prospects of earthquake prediction include improving machine learning algorithms and expanding data volume for more accurate forecasts. The researchers are also working on developing early warning systems that can provide warnings about imminent earthquakes in a few minutes or even hours in advance. Using Data Lake in seismology plays a crucial role in predicting earthquakes, reducing risks, and preparing society for possible natural disasters. Further development of these technologies opens new possibilities for ensuring safety and stability in vulnerable regions of the world. There are some additional explanations regarding the inclusion of historical earthquake information in Data Lake:

- Historical Earthquake Information. It can help identify patterns and trends that may be useful for predicting future earthquakes. Researchers may find that earthquakes with a certain location or strength tend to recur at a certain frequency.
- Historical Information. It can also help assess the potential strength of future earthquakes. If researchers know that an area has experienced a 7-magnitude earthquake, they can predict that future earthquakes in that area might also be quite strong.

Including historical earthquake information in Data Lake it can significantly increase the accuracy of earthquake predictions. In addition to the recommendations of the system for special care for certain individuals, as mentioned in the previous section, the system can predict potentially risky areas and advise against laying roads (or additionally reinforcing them), or building large-scale productions in particularly dangerous places, etc.

Impact of Schools on Life Indicators in Residential Areas

Domains: Infrastructure, Social, Economic. Schools are not just educational institutions but also important catalysts for social and economic development in residential areas. Investment and support for schools are critically important to ensure their positive impact on the life indicators of the areas where they are located. Schools contribute to social cohesion and community development. They become centers for local events such as meetings, cultural activities, and sports competitions, strengthening interpersonal connections and supporting community engagement. Schools also contribute to raising the level of education and training of the population, which can lead to increased employment rates and incomes. The quality of schools directly affects property values in the area, making it more attractive to investors. Schools also contribute to the local economy through job creation and support for small businesses, especially in the service sector. Experts estimate that the quality of schools can increase property values in the area by 5-10%. Schools also contribute to the creation of new jobs for teachers, school administration staff, and personnel. Additionally, schools can promote the development of small businesses in the areas where they are located, such as through the creation of food establishments, trade, and other services that are in demand among parents and students. We propose that all these indicators are collected in a database, then statistically processed, and as a result, we obtain a ranking of residential areas. This will help young families and highly qualified educators in choosing a place of work or residence. To ensure the positive impact of schools on the life indicators of the areas where they are located, it is necessary to provide them with sufficient funding and support. In some areas, schools are underfunded, which can lead to a decrease in the quality of education and negatively affect the social and economic development of the area. Additionally, there are areas where schools cannot provide proper conditions for learning and student development, for example, in areas with high crime rates or poverty. Addressing these challenges requires joint efforts by government authorities, local communities, and educators. For example, government bodies can provide additional funding to schools located in low-income areas. Local communities can support schools through volunteer activities or participation in parent committees. Educators can develop innovative programs and teaching methods that meet the needs of students from different social groups. Using schools as catalysts for the social and economic development of residential areas has the potential to significantly improve the quality of life of residents.

Epidemic Forecasting

Domains: Social, International, Infrastructure, Economic, etc. Predicting epidemics plays a key role in preventing large-scale disease outbreaks and reducing their negative impact on society. Early detection of potential epidemics allows health authorities and governments to distribute medical and logistical resources more effectively and plan response strategies. For example, predicting a flu epidemic allowed US health authorities to quickly initiate population vaccination, helping to prevent serious outbreaks. Predicting a polio epidemic in Africa allowed health authorities to coordinate vaccination efforts for children, leading to the eradication of polio in that region. Various data sources are used for epidemic forecasting, including:

- Medical Records. Data on patients who have been infected with a disease can help health authorities track the spread of the disease and assess its severity.
- Social Media Information. Data on social media posts can help identify trends in the spread of the disease, such as an increase in posts about symptoms.

- Meteorological Data. Weather data can help predict conditions that facilitate the spread of disease, such as temperature and humidity.
- Population Mobility Data. Data on people's movements can help identify areas where the disease is spreading most rapidly.

Artificial intelligence (AI) and machine learning play a crucial role in analyzing large volumes of data to detect patterns that may indicate the possibility of an epidemic outbreak. These technologies help increase the accuracy and timeliness of predictions, allowing for a quick response to threats. Using AI and machine learning for epidemic forecasting poses several challenges. One of the main challenges is that these technologies can be imperfect. For example, they can be susceptible to noise in the data or may not consider all the factors influencing the spread of the disease. Another challenge is the need for confidence in the prediction results. If the predictions are not accurate, it can lead to incorrect decisions that can worsen the situation. Overcoming these challenges requires further research and development. However, thanks to continuous improvements in artificial intelligence and machine learning, epidemic forecasting is becoming more accurate and effective.

Crop Yield Forecasting

Domains: Natural, Social, Economic, etc. The use of Data Lake in the agricultural sector for predicting crop yields opens the new opportunities to enhance productivity and efficiency in farming. Integrating large volumes of data allows for deeper analysis of factors affecting yield and making more accurate forecasts. Data Lake facilitates the collection and analysis of data from various sources, such as:

- Weather Data. Factors like temperature, precipitation, and humidity can significantly impact crop yields.
- Soil Data. Information such as soil type, acidity, and nutrient content also affects crop yields.
- Previous Years' Yield Data. It can help identify patterns and trends useful for forecasting future yields.

This data integration reveals trends and patterns critical for yield forecasting. For example, researchers might discover that the yield of a certain crop depends on specific weather conditions or soil characteristics. Data Lake collects and integrates data from weather stations, satellite images, IoT sensors in fields, providing a comprehensive picture of agricultural conditions. This ensures a thorough analysis necessary for accurate forecasting. Applying machine learning algorithms to Data Lake data helps identify key factors affecting yields. Models assist in predicting potential yield, considering various scenarios and conditions. The forecasts help farmers and agricultural companies in planning agronomic measures, efficient resource allocation, and risk management. Integrating this data with decision support systems enhances the overall productivity of agribusiness. Using Data Lake for predicting crop yields is an important step in modernizing agriculture. It enables the identification and utilization of key data to improve yield, planning, and management of agricultural resources, contributing significantly to food security and stability in the agricultural sector. There are some additional explanations regarding the inclusion of weather conditions and soil characteristics data in Data Lake:

- Weather Conditions. It is one of the most critical factors affecting yield. Weather data, such as temperature, precipitation, and humidity, can help researchers identify patterns and trends useful for forecasting yields. For example, researchers might find that the yield of a particular crop depends on soil temperature.
- Soil Characteristics. It also impacts yields. Data on soil, such as a soil type, acidity, and nutrient content, can help researchers assess the yield potential of a specific field. For instance, researchers might discover that a field with a certain soil type has the potential for high yields of a particular crop.

Including these data in Data Lake allows researchers to make more accurate crop yield predictions.

Conclusions

In the evolving tapestry of the European digital landscape, the "Centre of Europe" region is positioned uniquely, both as a beneficiary and a driver of the relentless integration of data and emerging technologies. The contemporary push towards a more digitally coherent and intelligent European region is not just a trend but an imperative, and the tools and techniques to manage data form the cornerstone of this transformative journey.

The diverse technological concepts explored in this research, ranging from traditional databases and data warehouses to more novel constructions like data lakes and data meshes, illuminate the myriad ways in which data can be stored, organized, and retrieved. Each technology presents its unique attributes, advantages, and potential pitfalls, and their relevance and applicability can vary based on specific regional requirements and objectives.

However, the true essence of this study transcends the mere understanding of these technologies. It emphasizes the strategic synthesis of these systems to harness the full potential of big data. When judiciously selected and integrated, these technologies can unlock unprecedented insights, optimize operational efficiencies, and catalyze innovations, thereby enhancing the quality of life and services in the "Centre of Europe".

Furthermore, the potential applications of these data technologies stretch across various sectors. From optimizing urban infrastructure and public services to fostering innovations in healthcare, education, and commerce, the possibilities are expansive. But with great potential the responsibility also comes to ensure ethical data practices, maintain data security, and prioritize user privacy.

In wrapping up, it's evident that the road ahead for the "Centre of Europe" is both promising and challenging. The foundation has been laid with a profound understanding of modern data storage and organizational techniques. The next steps involve strategic implementation, continuous adaptation, and the vision to merge technological prowess with regional aspirations. The "Centre of Europe", with its intelligent approach to harnessing data, is poised to become a beacon of digital excellence in the European context.

References

1. Inmon, W.H. Building the Data Warehouse. Hoboken, New Jersey: John Wiley & Sons, 2002. 576 p.
2. Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. The Data Warehouse Lifecycle Toolkit. Hoboken, New Jersey: John Wiley & Sons, 2010
3. Stonebraker, M., & Cetintemel, U. "One size fits all": an idea whose time has come and gone. Proceedings of the 21st International Conference on Data Engineering. 2005. pp. 2-11. doi: 10.1109/ICDE.2005.1
4. Cloudera. Turn Your Data Lake into an Enterprise Data Hub. URL: https://vision.cloudera.com/turn-your-data-lake-into-an-enterprise-data-hub/, 2014.
5. Nikitin Valerii, Krylov Evgen. Primary-based spectral bloom filter for the ensuring consistency in distributed document-based nosql databases using active anti-entropy mechanism Computer Systems and Information Technologies,2023, №3, pp.75–80. https://doi.org/10.31891/csit-2023-3-9
6. Druzdzel M. J. and Flynn R. R. Decision support systems, 2000. doi: 10.1081/E-ELIS4-121
7. Fang H. Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem. International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). 2015. Pp. 820--824. doi: 10.1109/CYBER.2015.7288049
8. Gartner. Gartner Says Beware of the Data Lake Fallacy. URL: http://www.gartner.com/newsroom/id/2809117. doi: 10.2139/ssrn.2262249
9. Hortonworks. A Modern Data Architecture whith Apache Hadoop. 2014. URL: http://info.hortonworks.com/rs/h2source/images/Hadoop-Data-Lake-white-paper.pdf. doi: 10.1109/ICGCIoT.2015.7380530
10. IBM. Governing and Managing Big Data for Analytics and Decision Makers. 2014. URL: http://www.redbooks.ibm.com/abstracts/redp5120.html?Open, doi: 10.1515/97831106640651.
11. Hovorushchenko Tetiana, Zabelina Iryna, Rei Kostyantyn Hovorushchenko Olha. Method of creating an information system for monitoring infectious patients. Computer Systems and Information Technologies, 2023, №3, pp.59–64. https://doi.org/10.31891/csit-2023-3-7
12. Inmon W. H. Building the Data Warehouse. QED Information Sciences, MA: Wellesley, 1992.
13. Kimball R. and Caserta J. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data. Hoboken, New Jersey: John Wiley & Sons, 2004.
14. McKinney, W. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc, 2012.
15. White T. Hadoop: The Definitive Guide. O'Reilly Media, Inc., 2012. doi: 10.12691/ajss-5-1-2
16. Stonebraker M., & Cetintemel, U. Proceedings. 21st International Conference «One size fits all: an idea whose time has come and gone. Data Engineering», 2005. Pp. 2-11. doi: 10.1109/ICDE.2005.1
17. Redman T. C. Data quality for the information age. Artech House, Inc., 1996.
18. Zikopoulos, P., & Eaton, C. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media, 2011.
19. Marz, N., Warren, J. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications Co, 2015. doi: 10.12691/ajss-5-1-2
20. Buttle, F. Customer relationship management: concepts and technologies. Routledge, 2019. doi 10.4324/9781351016551
21. Pyts Markiyan, Dronyuk Ivanna. Theoretical background for creating real world data lake architecture. Computer Systems and Information Technologies,2023, №2, pp. 69–74. https://doi.org/10.31891/csit-2023-2-9
22. Hamoud, A.K.; Adday, H.; Obaid, T.A.S., Hameed, R.A. Design and implementing cancer data warehouse to support clinical decisions. International Journal of Scientific and Engineering Research. 2016. Vol.7(2). Pp.1271-1285.
23. Hamoud, A.K.; and Obaid, T. Building data warehouse for diseases registry: first step for clinical data warehouse. International Journal of Scientific and Engineering Research. 2013. Vol.4(7). Pp. 636-640. doi 10.2139/ssrn.3061599
24. Hamoud, A.K.; and Obaid, T.A.S. Using OLAP with diseases registry warehouse for clinical decision support. International Journal of Computer Science and Mobile Computing. 2014. Vol.3(4). Pp. 39-49. doi 10.2139/ssrn.3061597
25. Hamoud, A.K.; Hashim, A.S.; and Awadh, W.A. Clinical data warehouse: a review. Iraqi Journal for Computers and Informatics. 2018. Vol.44(2). Pp.1-11.
26. Govorushchenko T. O. The method of predicting the success of the implementation of scientific IT projects based on the analysis of their characteristics / T. O. Govorushchenko, K. Yu. Gavrylyuk // Computer systems and information technologies. - 2020. - No. 1. - P. 91-95. - Access mode: http://nbuv.gov.ua/UJRN/csit_2020_1_14.
27. Govorushchenko T. O. Multi-agent system of taking into account the information of the subject field at all stages of software development / T. O. Govorushchenko, I. Yu. Lopatto, M. V. Kapustyan // Scientific notes of V. I. Vernadskyi Tavri National University. Series: Technical sciences. - 2022. - Vol. 33(72), No. 2. - P. 74-79. - Access mode: http://nbuv.gov.ua/UJRN/sntuts_2022_33(72)_2_14.

Olexandr HOLOTA **Олександр ГОЛОТА**	Postgraduate student of the Department of Informatics and Physical and Mathematical Disciplines. "Uzhgorod National University" https://orcid.org/0009-0008-9717-3165 e-mail: oleksandr.holota@uzhnu.edu.ua	Аспірант кафедра інформатики та фізико-математичних дисциплін. ДВНЗ «Ужгородський національний університет»
Vasyl KUT **Василь КУТ**	PhD, associate professor, head of the department of informatics and physical and mathematical disciplines. "Uzhgorod National University" https://orcid.org/0000-0003-3007-2462 e-mail: vasilij.kut@uzhnu.edu.ua	Канд. техн. наук, доцент, завідувач кафедра інформатики та фізико-математичних дисциплін. ДВНЗ «Ужгородський національний університет»

Sergyi PASICHNYK **Сергій ПАСІЧНИК**	Postgraduate student of the Department of Information Systems and Networks of Lviv Polytechnic National University https://orcid.org/0009-0006-8011-5618 e-mail: sergiypasichnyk@gmail.com	Аспірант кафедри інформаційних систем та мереж Національного університету «Львівська політехніка»
Oleh HRYBOVSKYI **Олег ГРИБОВСЬКИЙ**	Master of the Department of Information Systems and Networks of Lviv Polytechnic National University https://orcid.org/0009-0005-6318-3611 e-mail: oleh.hrybovskyi.mitup.2022@lpnu.ua	Магістр кафедри інформаційних систем та мереж Національного університету «Львівська політехніка»