

Парфененко Ю.В.

<https://orcid.org/0000-0003-4377-5132>

Сумський державний університет

Кузнєцов Е.Г.

<https://orcid.org/0000-0003-3966-9724>

Сумський державний університет

АНАЛІЗ ПІДХОДІВ ТА ПРОЄКТУВАННЯ АРХІТЕКТУРИ ОБРОБКИ ВЕЛИКИХ ДАНИХ ДЛЯ ЦИФРОВИХ ДВІЙНИКІВ ЕНЕРГЕТИЧНИХ МІКРОМЕРЕЖ

У статті розглянуто підходи до зберігання та обробки великих наборів даних цифрових двійників енергетичних мікромереж. Метою роботи є аналітичний огляд підходів до застосування технологій великих даних у цифрових двійниках енергетичних мікромереж, а також розроблення хмарної архітектури обробки даних, яка забезпечує узгоджену інтеграцію процесів збору, оброблення та опрацювання даних. У дослідженні використано методологію аналізу та узагальнення підходів до застосування технологій великих даних у цифрових двійниках енергетичних мікромереж, а також методи архітектурного проєктування хмарної архітектури обробки даних. Проаналізовано та узагальнено сучасні технології й хмарні платформи для роботи з великими даними з позиції їх застосування у цифрових двійниках енергетичних мікромереж. За результатами дослідження обрано платформу AWS, що зумовлено її широкими можливостями масштабування, розвиненим набором сервісів для збирання, зберігання, обробки та аналітики даних.

Наукова новизна роботи полягає у розробленій архітектурі хмарної інфраструктури зберігання та обробки великих даних для цифрових двійників енергетичних мікромереж на платформі AWS.

Запропонована хмарна архітектура полягає в інтеграції потокового надходження даних від сенсорів цифрових двійників через AWS IoT Core із подальшою передачею та буферизацією за допомогою Apache Kafka і Amazon Kinesis Data Streams, їх збереженням у Amazon S3, Amazon Timestream та Amazon DynamoDB і подальшою обробкою. Обчислення та аналітика реалізуються з використанням Amazon EC2, Amazon EMR, AWS Lambda і AWS Glue, а моделювання – у Amazon SageMaker, з подальшою візуалізацією через Amazon QuickSight та моніторингом у Amazon CloudWatch.

Розроблена хмарна архітектура дозволяє дотримуватися ключових принципів роботи з великими даними, зокрема масштабованості, розподіленого зберігання та обчислень, а також забезпечення цілісності, безпеки й доступності даних.

Ключові слова: великі дані, аналіз, обробка даних, цифровий двійник, хмарні технології, хмарна платформа, розгортання, AWS.

Постановка проблеми. Застосування цифрових двійників як віртуальної копії фізичної системи, що інтегрує дані реального часу, симуляції та методи машинного навчання для створення динамічної, інтерактивної моделі реального середовища, на сьогодні набуває все більшого поширення. Для енергетичних мікромереж цифровий двійник дозволяє створити віртуальну модель енергосистеми, що відтворює її стан у реальному часі на основі даних сенсорів, забезпечуючи можливість

моніторингу, аналізу режимів роботи, прогнозування навантажень і генерації, а також підтримки прийняття рішень щодо оптимізації функціонування та підвищення надійності мікромережі. Цифрові двійники є особливо ефективними в оптимізації потоків енергії в мікромережах, моделюючи різні операційні сценарії, прогнозуючи потенційні збої та пропонуючи стратегії для підтримки стабільності системи за умов коливань навантажень та генерації з відновлюваних джерел енергії.



Водночас застосування цифрових двійників у контексті енергетичних мікромереж пов'язане з рядом специфічних викликів, зумовлених масштабом і складністю сучасних енергетичних систем. У таких системах функціонує велика кількість вимірювальних пристроїв, таких як, датчики, лічильники, а також контролерів, які підключені до мереж передачі даних через гетерогенні комунікаційні середовища Інтернету речей. Постійна взаємодія цих компонентів призводить до формування значних потоків різноманітних даних, що характеризують стан обладнання, параметри електричних мереж та режими їх функціонування. Ці дані використовуються при побудові методів штучного інтелекту для прогнозування енергоспоживання та генерації електроенергії, виявлення аномалій у роботі обладнання, оцінювання технічного стану енергетичних активів, а також оптимізації режимів роботи енергетичних мікромереж. Окрім цього, необхідно збирати дані безпосередньо про стан цифрового двійника, його параметри, конфігурацію моделей, результати симуляцій та обчислень, а також показники синхронізації з фізичною системою, що забезпечує актуальність і коректність відображення стану енергетичної мікромережі.

Таким чином, у результаті оперативного моніторингу в режимі реального часу генеруються значні обсяги даних, які необхідно накопичувати та зберігати для подальшого використання під час застосування моделей машинного навчання, зокрема для аналізу режимів роботи мікромережі, прогнозування навантажень і генерації електроенергії тощо.

Разом з тим значні обсяги даних, що надходять з різних джерел, характеризуються високою швидкістю генерації, різноманітністю форматів та нерідко містять неповні або зашумлені значення. Такі особливості ускладнюють їх безпосереднє використання в системах цифрових двійників та вимагають застосування ефективних методів попередньої обробки, інтеграції та аналізу даних. Для забезпечення коректного функціонування цифрових двійників необхідно реалізувати архітектури обробки великих даних, здатні забезпечувати збирання, зберігання, обробку та аналітику значних потоків інформації в режимі, близькому до реального часу.

Аналіз останніх досліджень і публікацій. Цифрові двійники є інструментом ефективного та безпечного керування великими масивами історичних даних та потоками даних у режимі реального часу, підтримуючи роботу мікромережі, та

забезпечуючи підтримку прийняття рішень при проектуванні, експлуатації та обслуговуванні енергетичних мікромереж [1].

У роботах Цюара Т. (Cioara T.) та ін., Лай Й. (Lai Y.) та ін. [2,3] зазначається, що джерелами даних для цифрових двійників енергетичних систем є різноманітні сенсори та пристрої моніторингу, включно з інтелектуальними лічильниками, датчиками напруги та струму, фазорними вимірювальними блоками (PMU), датчиками стану обладнання, а також пристроями Інтернету речей (IoT), з яких збирають дані в режимі реального часу. Технологія цифрових двійників використовує великі дані для моделювання та підтримки прийняття рішень щодо роботи й управління енергомережею в нормальних умовах функціонування та для виявлення відхилень від стаціонарного стану [4].

Підхід до управління енергетичною мікромережею, заснованих на обробці великих даних, представлено у роботі Лю Л. (Liu L.) та ін. [5]. Він забезпечує інтелектуальне управління розподіленою генерацією та локальним споживанням, а також оптимізує взаємодію всіх наявних джерел розподіленої енергії, систем зберігання та навантажень у межах мікромережі. Інтелектуальне управління складною мікромережею, що реалізується через центр управління на основі великих даних, дозволяє раціонально розподіляти енергію та підвищувати безпеку і надійність системи, сприяючи її інтелектуальному розвитку.

Вибір платформи обробки великих даних залежить від конкретних вимог до обробки та характеристик даних у застосунках цифрових двійників мікромереж. Apache Spark розглядається як найбільш відповідна технологія для застосунків інтелектуальних мереж, оскільки вона підтримує як пакетну, так і потокову обробку, що робить її застосування можливим обробки даних систем цифрових двійників [6]. Однак Spark вимагає зовнішніх систем зберігання даних, тому він найкраще працює, коли його встановлено поверх Hadoop, щоб використовувати розподілену файлову систему Hadoop (HDFS) для паралельного розподіленого зберігання [7]. Такий розподілений підхід гарантує, що потоки даних у режимі реального часу, необхідні для точного представлення цифрових двійників, можуть підтримуватися та оброблятися ефективно, підтримуючи двонаправлений обмін даними між фізичними мікромережами та їхніми віртуальними аналогами. Для цифрового двійника, що потребує потокової обробки даних у реальному часі, переважною є Apache Storm, тоді як Hadoop залишається опти-

мальним для пакетної офлайн-обробки великих наборів даних [7].

Платформи хмарних обчислень забезпечують масштабовану інфраструктуру, що може бути основою для розгортання цифрових двійників мікромереж. Хмарні платформи підтримують як структуровані, так і неструктуровані дані через розподілені архітектури, побудовані на Hadoop, Spark та інших проєктах великих даних [8]. О варіанти хмарних сховищ для аналізу великих даних, включаючи Amazon S3, Google BigQuery, Microsoft Azure та Google Drive, з практичними впровадженнями за допомогою таких інструментів, як Google Colab та PySpark, для прогнозування стабільності мережі [9]. Amazon Web Services (AWS) пропонує комплексні рішення, включаючи AWS IoT Greengrass, що використовується для периферійних обчислень та дозволяє локальну обробку й аналіз даних, знижуючи витрати на масштабне розгортання обчислювальної інфраструктури для енергетичних мікромереж [10]. Платформа Azure Digital Twins дозволяє цифрові моделі складних фізичних середовищ, зокрема, енергетичних мереж або інфраструктури, що інтегрують дані з IoT-пристроїв і роблять можливими аналітику, моніторинг і оптимізацію роботи систем у реальному часі [11]. Хоча використання хмарних технологій для обробки великих даних цифрових двійників стає все більш поширеним, питання вибору хмарної платформи та архітектури для зберігання та обробки великих даних цифрових двійників енергетичних мікромереж досліджене недостатньо. Існує потреба у розробленні хмарної архітектури, що забезпечуватиме ефективне зберігання, обробку й аналіз великих даних цифрових двійни-

ків для підтримки прийняття рішень в управлінні енергомережею.

Постановка завдання. Метою дослідження є аналіз підходів до використання технологій великих даних у цифрових двійниках енергетичних мікромереж та проєктування хмарної архітектури обробки великих даних, що забезпечує інтеграцію процесів збирання даних, їх обробку та подальшу аналітику.

Виклад основного матеріалу. Інтеграція технологій великих даних стає важливою, оскільки енергетичні мікромережі генерують величезні обсяги даних часових рядів з різних джерел, включаючи системи відновлюваної енергії, пристрої зберігання даних, навантаження та умови навколишнього середовища. Схема потоків даних при управлінні енергетичними мікромережами від збирання до зберігання великих обсягів даних показана на рис. 1.

Традиційні методи обробки даних не можуть впоратися з таким масштабом та складністю, що робить розподілені обчислювальні платформи, такі як Apache Spark та Hadoop, необхідними для аналітики в реальному часі та програм машинного навчання. Платформи хмарних обчислень, такі як AWS, забезпечують масштабовану інфраструктуру, необхідну для зберігання та аналітики наборів великих даних.

Перевагою використання хмарних технологій у системах обробки великих даних є еластичність ресурсів, можливість горизонтального масштабування обчислювальних потужностей, висока відмовостійкість та зменшення витрат на створення і підтримку власної інфраструктури. Крім того, хмарні сервіси інтегрують інструменти потокової

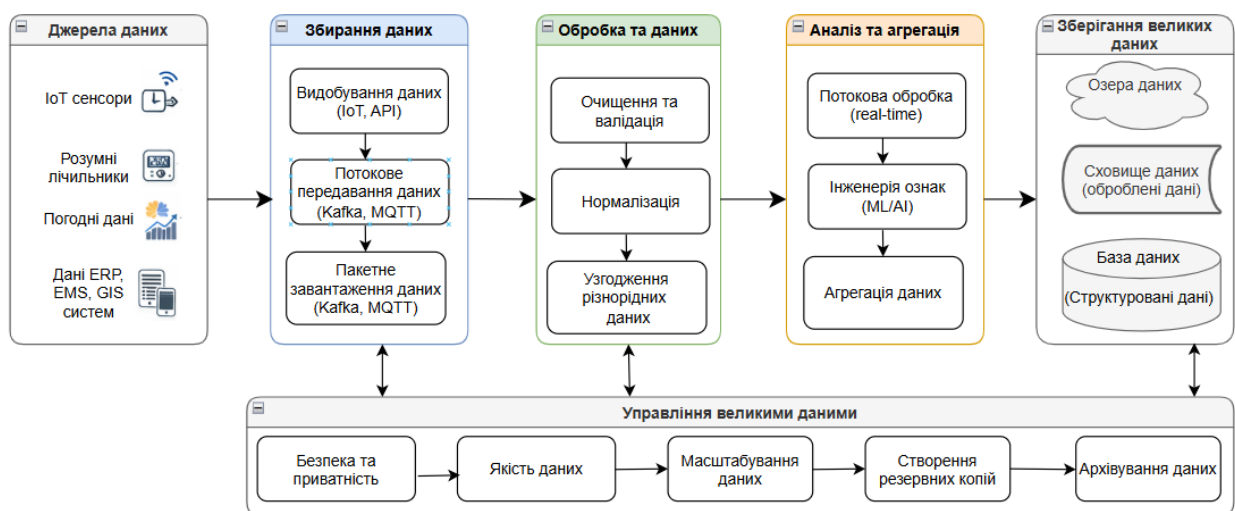


Рис. 1. Схема потоків даних при управлінні енергетичними мікромережами

обробки, аналітики великих даних та машинного навчання, що дозволяє швидко розгортати аналітичні платформи для підтримки функціонування цифрових двійників енергетичних систем [12].

Було проаналізовано та узагальнено технології та хмарні платформи для роботи з великими даними за напрямом використання у цифрових двійника енергетичних мікромереж [13, 14] (табл. 1).

Проведений аналіз застосування технологій великих даних у цифрових двійниках енергетичних мікромереж (табл. 1) показує, що для реалізації задач збору, обробки та аналітики даних використовуються різні хмарні платформи, серед яких найбільш поширеними є AWS, Microsoft Azure та Google Cloud Platform. У даному дослідженні для проєктування архітектури обробки даних цифрового двійника енергетичної мікромережі обрано платформу AWS, оскільки вона надає широкий набір спеціалізованих сервісів для роботи з великими даними, потоковими даними та IoT (зокрема AWS IoT, Amazon S3, Amazon Kinesis, AWS Lambda), забезпечує високу масштабованість, надійність і гнучкість розгортання аналітичних компонентів у хмарному середовищі.

Для ефективного проєктування архітектури обробки великих даних у цифрових двійниках енергетичних мікромереж доцільно розглядати три функціональні рівні: інтеграція даних, масштабована обробка та аналітика. На рівні інтеграції даних забезпечується збір і структуриза-

ція інформації з різнорідних джерел, включно з IoT-пристроями, SCADA-системами, сенсорами енергетичних активів та даними про навантаження [13, 14]. Таким чином буде забезпечено дотримання принципу BDIS–BDPS–BDAS (Big Data Integration System – Big Data Processing System – Big Data Analytics System), що передбачає три рівні управління та обробки даних.

У запропонованій архітектурі реалізується масштабована обробка даних цифрових двійників енергетичних мікромереж із використанням сервісів хмарної платформи AWS (рис. 2).

Джерела даних формуються сенсорами та IoT-пристроями, розташованими у розподілених мікромережах, які безперервно передають інформацію про енергоспоживання, напругу, струм та інші фізичні параметри системи.

Потокові дані надходять до AWS IoT Core, що забезпечує безпечний і надійний прийом повідомлень від сенсорів та їх маршрутизацію до обчислювальних і аналітичних сервісів. Для організації високопродуктивної передачі поточкових даних використовується Apache Kafka, яка виконує роль брокера повідомлень, забезпечує буферизацію та гарантовану доставку даних, а також дозволяє масштабувати обробку великих потоків інформації від численних сенсорних вузлів. Додатково, Amazon Kinesis Data Streams інтегрує потокові дані та передає їх до сховищ і аналітичних сервісів для подальшої обробки.

Таблиця 1

Напрямок використання	Приклад великих даних	Технології великих даних	Платформи	Інструменти
Моніторинг та збір даних	IoT-сенсори, SCADA, смарт-лічильники	Потокова обробка даних у реальному часі	Azure IoT Hub, AWS IoT Core, Google Cloud IoT	Apache Kafka, Spark Streaming, Apache Flink
Прогнозування споживання	Історичні профілі навантаження, погодні дані	Машинне навчання та глибинне навчання	Azure Machine Learning, AWS SageMaker, Google Vertex AI	TensorFlow, PyTorch, Scikit-learn
Інтеграція відновлюваних джерел енергії	Дані про сонячну радіацію, вітер, метеопрогнози	Аналітика великих даних та сховища даних (Data Lake)	Google Cloud Platform, AWS Data Analytics, Cloudera Data Platform	BigQuery, Apache Hadoop, Spark SQL
Оптимізація роботи	Тарифи, графіки навантаження, стан обладнання	Хмарні обчислення та оптимізаційна аналітика	AWS Lambda, Azure Functions, Google Cloud Functions	Gurobi, SciPy Optimize, Pyomo
Надійність та безпека	Журнали систем, кіберінциденти, аномалії	Виявлення аномалій та пояснюваний штучний інтелект (XAI)	Microsoft Sentinel, IBM Watson OpenScale, Google Chronicle	Splunk, Elastic Stack (ELK), Apache Metron
Підтримка рішень	Дані енергоринку, регуляторні вимоги	Бізнес-аналітика та візуалізація даних	Power BI, Tableau, Qlik Sense	Grafana, Apache Superset
Технічне обслуговування	Вібрації, температури, історія ремонтів	Предиктивне технічне обслуговування та цифрові двійники	Azure Digital Twins, AWS IoT TwinMaker, Siemens MindSphere	IBM Maximo, ThingWorx Analytics

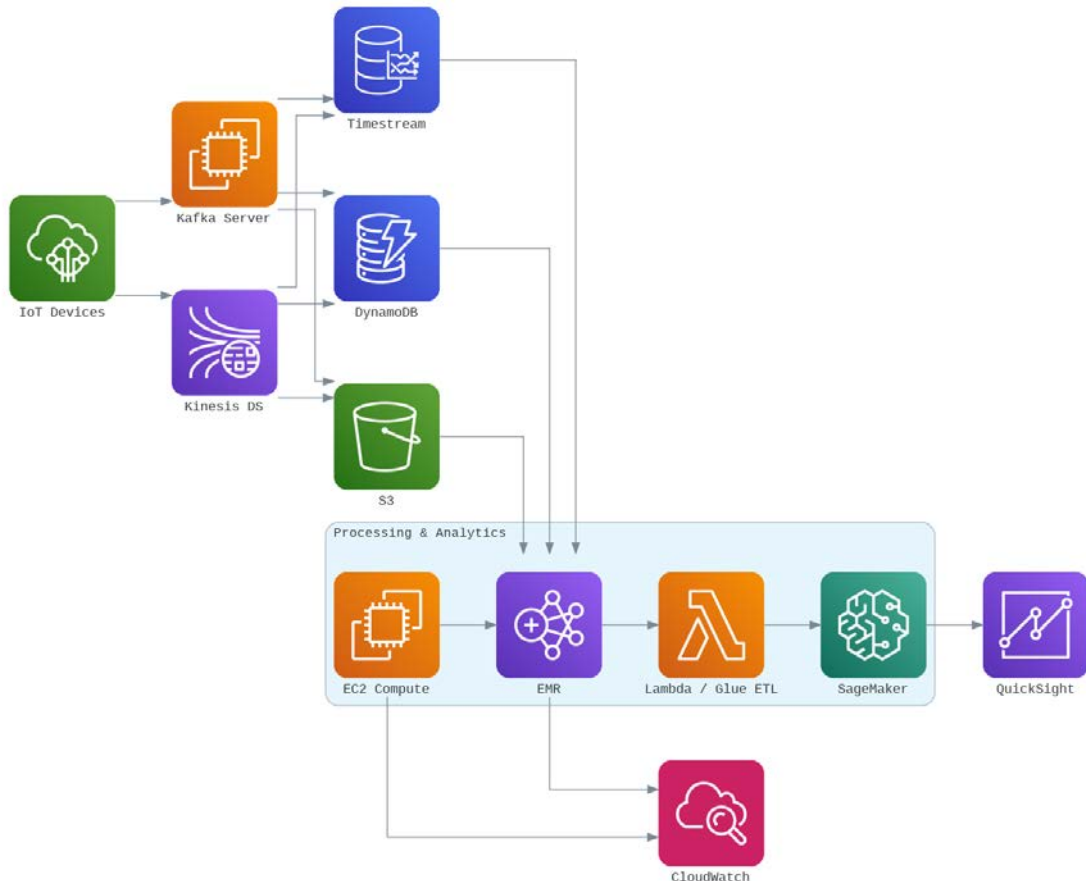


Рис. 2. Хмарна архітектура зберігання та обробки великих даних на платформі AWS

Збереження даних здійснюється через Amazon S3 для необроблених та історичних даних, а також Amazon Timestream та DynamoDB для структурованих даних часових рядів цифрових двійників. Така організація сховищ забезпечує одночасно ефективний доступ до історичних даних та високошвидкісну обробку поточних показників системи.

Для обчислень та аналітики у хмарі використовується EC2 як центральний обчислювальний вузол, що координує роботу блоку “Processing & Analytics” і взаємодіє з потоковими брокерами та сховищами даних. Amazon EMR із фреймворками Apache Spark та Hadoop застосовується для масових аналітичних обчислень та підготовки даних, забезпечуючи високу продуктивність при роботі з великими обсягами інформації. Додатково, AWS Lambda та AWS Glue реалізують ETL-процеси для перетворення та очищення даних, підготовки їх до моделювання цифрових двійників. Навчання моделей та прогнозування поведінки мікромереж виконується у Amazon SageMaker, що дозволяє створювати аналітичні моделі з використанням машинного навчання та автоматизувати процес оцінки їх продуктивності.

Візуалізація аналітики та ключових показників цифрових двійників здійснюється через Amazon

QuickSight, що надає інтерактивні панелі моніторингу та інструменти для детального аналізу даних. Amazon CloudWatch забезпечує безперервний моніторинг стану всієї системи, збір логів і оповіщення про потенційні відхилення або помилки.

Безпека та контроль доступу забезпечуються через IAM Roles, шифрування даних за допомогою AWS KMS, а також розгортання компонентів у ізольованих середовищах VPC, що гарантує захищену обробку та зберігання великих обсягів даних цифрових двійників.

Висновки. У результаті проведеного дослідження узагальнено підходи до зберігання та обробки великих даних у цифрових двійниках енергетичних мікромереж. Обґрунтовано доцільність використання платформи AWS для побудови інфраструктури обробки великих даних завдяки її гнучкості, широкому набору сервісів та можливостям горизонтального масштабування. Розроблена хмарна архітектура зберігання та обробки великих даних, що забезпечує узгоджену взаємодію компонентів збору, передавання, зберігання, обробки та аналізу даних, і таким чином дозволяє реалізувати повний цикл роботи з даними цифрового двійника.

Практичне значення отриманих результатів полягає у можливості використання запропо-

нованої архітектури як основи для створення та впровадження цифрових двійників енергетичних мікромереж. Перспективи подальших досліджень пов'язані з оптимізацією витрат на обробку даних, підвищенням ефективності моделей машинного навчання та розширенням функціональних мож-

ливостей систем підтримки прийняття рішень на основі цифрових двійників.

Дослідження виконувалося в рамках держбюджетної НДР «Методи та технології штучного інтелекту для стійкого управління енергетичними системами із застосуванням цифрових двійників» (№ 0126U000866).

Список літератури:

1. Bazmohammadi N., Madary A., Vasquez J. C., Bazmohammadi H., Khan B., Wu Y. Et al. Microgrid digital twins: concepts, applications, and future trends. *IEEE Access*. 2021. Vol. 10. DOI: 10.1109/ACCESS.2021.3138990.
2. Cioara T. et al. An overview of digital twins application domains in smart energy grid. *arXiv*. 2021. arXiv:2104.07904. DOI: 10.48550/arXiv.2104.07904.
3. Lai Y. et al. Construction of a digital twin model for incremental aggregation of multi-type load information in hybrid microgrids under integrity constraints. *Energy Informatics*. 2024. Vol. 7. Article 112. DOI: 10.1186/s42162-024-00404-5.
4. He X. et al. Preliminary exploration on digital twin for power systems: challenges, framework, and applications. *arXiv*. 2019. arXiv:1909.06977. DOI: 10.48550/arXiv.1909.06977.
5. Liu L. et al. Intelligent control of micro grid: a big data-based control center. *IOP Conference Series: Earth and Environmental Science*. 2018. Vol. 108, № 5. P. 052037. DOI: 10.1088/1755-1315/108/5/052037.
6. Dihan M. S. et al. Digital twin: data exploration, architecture, implementation and future. *Heliyon*. 2024. Vol. 10, № 5. P. e26503. DOI: 10.1016/j.heliyon.2024.e26503.
7. Zainab A., Ghayeb A., Syed D. et al. Big data management in smart grids: technologies and challenges. *IEEE Access*. 2021. Vol. 9. P. 73046–73059.
8. Guo K. et al. Fault diagnosis of intelligent production line based on digital twin and improved random forest. *Applied Sciences*. 2021. Vol. 11, № 16. P. 7733. DOI: 10.3390/app11167733.
9. Xu J. et al. Research on integrated intelligent energy management system based on big data analysis and machine learning. *arXiv*. 2025. arXiv:2508.05583. DOI: 10.48550/arXiv.2508.05583.
10. Borra P., Mullapudi M., Nerella H., Prakashchand L. Analyzing AWS edge computing solutions to enhance IoT deployments. *International Journal of Engineering and Advanced Technology (IJEAT)*. 2024. Vol. 13, № 6. P. 8–12. DOI: 10.35940/ijeat.F4519.13060824.
11. Калайда А. В., Пан М. П. Створення цифрових двійників розумних будівель на платформі Azure Digital Twins. *Комунальне господарство міст*. 2024. № 6 (187). С. 2–7. DOI: 10.33042/2522-1809-2024-6-187-2-7.
12. Aslam S. et al. Machine learning applications in energy systems: current trends, challenges, and research directions. *Energy Informatics*. 2025. Vol. 8. Article 62. DOI: 10.1186/s42162-025-00524-6.
13. Hashem I. A. T. et al. The rise of “big data” on cloud computing: review and open research issues. *Information Systems*. 2015. Vol. 47. P. 98–115. DOI: 10.1016/j.is.2014.07.006.
14. Chen M., Mao S., Liu Y. Big data: a survey. *Mobile Networks and Applications*. 2014. Vol. 19, № 2. P. 171–209. DOI: 10.1007/s11036-013-0489-0.

Parfenenko Yu.V., Kuznetsov E.G. ANALYSIS OF APPROACHES AND DESIGN OF BIG DATA PROCESSING ARCHITECTURE FOR DIGITAL TWINS OF ENERGY MICROGRIDS

The article considers approaches to storage and processing big data sets of digital twins of energy microgrids. The purpose of the work is to provide an analytical overview of approaches to the application of big data technologies in digital twins of energy microgrids, as well as the development of a cloud data processing architecture that provides a consistent integration of data collection, processing and processing processes. The study uses a methodology for analyzing and summarizing approaches to the application of big data technologies in digital twins of energy microgrids, as well as methods for architectural design of a cloud data processing architecture. Modern technologies and cloud platforms for big data processing in digital twins of energy microgrids are analyzed and summarized. According to the results of the study, the AWS platform was chosen, which is due to its wide scalability capabilities and a developed set of services for data collection, storage, processing and analytics.

The scientific novelty of the work lies in the developed architecture of the cloud infrastructure for storing and processing big data for digital twins of energy microgrids on the AWS platform.

The proposed cloud architecture consists of integrating streaming data from digital twin sensors via AWS IoT Core with subsequent transmission and buffering using Apache Kafka and Amazon Kinesis Data Streams, their storage in Amazon S3, Amazon Timestream and Amazon DynamoDB and further processing. Computing and analytics are implemented using Amazon EC2, Amazon EMR, AWS Lambda and AWS Glue, and modeling is implemented in Amazon SageMaker, with subsequent visualization via Amazon QuickSight and monitoring in Amazon CloudWatch.

The proposed cloud architecture allows to adhere to the key principles of working with big data, in particular scalability, distributed storage and computing, as well as ensuring data integrity, security and availability.

Keywords: big data, analysis, data processing, digital twin, cloud technologies, cloud platform, deployment, AWS.

Дата першого надходження статті до видання: 18.03.2026

Дата прийняття статті до друку після рецензування: 14.04.2026

Дата публікації (оприлюднення) статті: 19.05.2026