

МЕТРИКИ ЕФЕКТИВНОСТІ ВИКОРИСТАННЯ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ У ХМАРНИХ СЕРЕДОВИЩАХ ДЛЯ МІКРОСЕРВІСНИХ АРХІТЕКТУР

Систематизовано та проаналізовано багатовимірні метрики ефективності мікросервісів у хмарних середовищах з точки зору технічних, економічних та екологічних аспектів, їх взаємозв'язків та компромісів. Запропоновано концептуальну багаторівневу модель оцінки ефективності та підходи до багатоцільової оптимізації ресурсів. Особлива увага приділяється моделюванню системи прийняття рішень з урахуванням суперечливих цілей оптимізації на різних рівнях архітектури.

Ключові слова: мікросервіси, хмарні обчислення, метрики ефективності, багатоцільова оптимізація, енергоефективність, економічна ефективність.

Сучасна розробка програмного забезпечення переживає значний зсув парадигми від традиційних монолітних архітектур на користь більш гнучких та масштабованих мікросервісних підходів, особливо в контексті хмарно-орієнтованих систем [1]. Галузеві дослідження демонструють стрімке прийняття цієї архітектури в останні роки, що свідчить про її практичну цінність для вирішення сучасних задач розробки.

Хмарні обчислення стали каталізатором цього переходу, надаючи інфраструктуру, що дозволяє повною мірою реалізувати переваги мікросервісів. Технології контейнеризації, такі як Docker, разом із платформами оркестрації, зокрема Kubernetes, сформували екосистему інструментів, що дозволяє ефективно розгортати, масштабувати та керувати мікросервісними додатками [2].

Однак, попри очевидні переваги, мікросервісна архітектура породжує низку викликів – від підвищеної складності управління розподіленими системами до проблем з мережевою затримкою та ускладненого моніторингу. З поширенням мікросервісних систем у корпоративному середовищі зростає потреба у комплексній оцінці їхньої ефективності, що виходить за межі суто технічних аспектів та охоплює економічну доцільність і екологічну стійкість [3].

Аналіз ефективності мікросервісних систем вимагає інтегрованого підходу, що враховує технічні, економічні та екологічні аспекти. Ці три виміри формують цілісну картину продуктивності та стійкості системи у сучасному хмарному середовищі.

У технічному вимірі ключове значення мають показники продуктивності та використання ресурсів. Затримка та час відгуку стають критично важливими через часті мережеві взаємодії між сервісами, що є невід'ємною характеристикою мікросервісної архітектури [4]. Пропускна здатність системи визначає її можливість обробляти високі навантаження. Не менш важливими є показники використання обчислювальних ресурсів – рівень завантаження процесора та об'єм спожитої оперативної пам'яті, що безпосередньо впливають на масштабованість та еластичність системи [5].

Економічний вимір ефективності охоплює як прямі витрати на споживання хмарних ресурсів [6], так і непрямі витрати, що формують загальну вартість володіння (TCO). Вибір оптимальної моделі ціноутворення – від оплати за фактичним використанням (PAYG) до зарезервованих або спотових екземплярів – стає стратегічним рішенням, що суттєво впливає на фінансову ефективність [7]. При цьому економічна ефективність тісно пов'язана з технічними показниками: недостатнє використання ресурсів призводить до невиправданих витрат, тоді як надмірне використання може негативно впливати на якість обслуговування користувачів [8].

Екологічний вимір, який раніше часто залишався поза увагою, набуває все більшої значущості з огляду на зростаючий вплив IT-індустрії на довкілля. Споживання енергії центрами обробки даних, що вимірюється в кіловат-годинах [9], та пов'язаний з цим вуглецевий слід стають важливими показниками стійкості системи. Особливої уваги заслуговує розрізнення операційного вуглецю (від споживання енергії) та втіленого вуглецю (від виробництва обладнання) [10], а також показники ефективності використання енергії, зокрема PUE (Power Usage Effectiveness) [11].

Три виміри ефективності мікросервісних архітектур не є ізольованими, а утворюють складну систему взаємозв'язків та компромісів, що потребує комплексного аналізу та балансування. Оптимізація одного аспекту системи часто має наслідки для інших вимірів, створюючи потребу в пошуку оптимальних компромісів (Рис. 1).

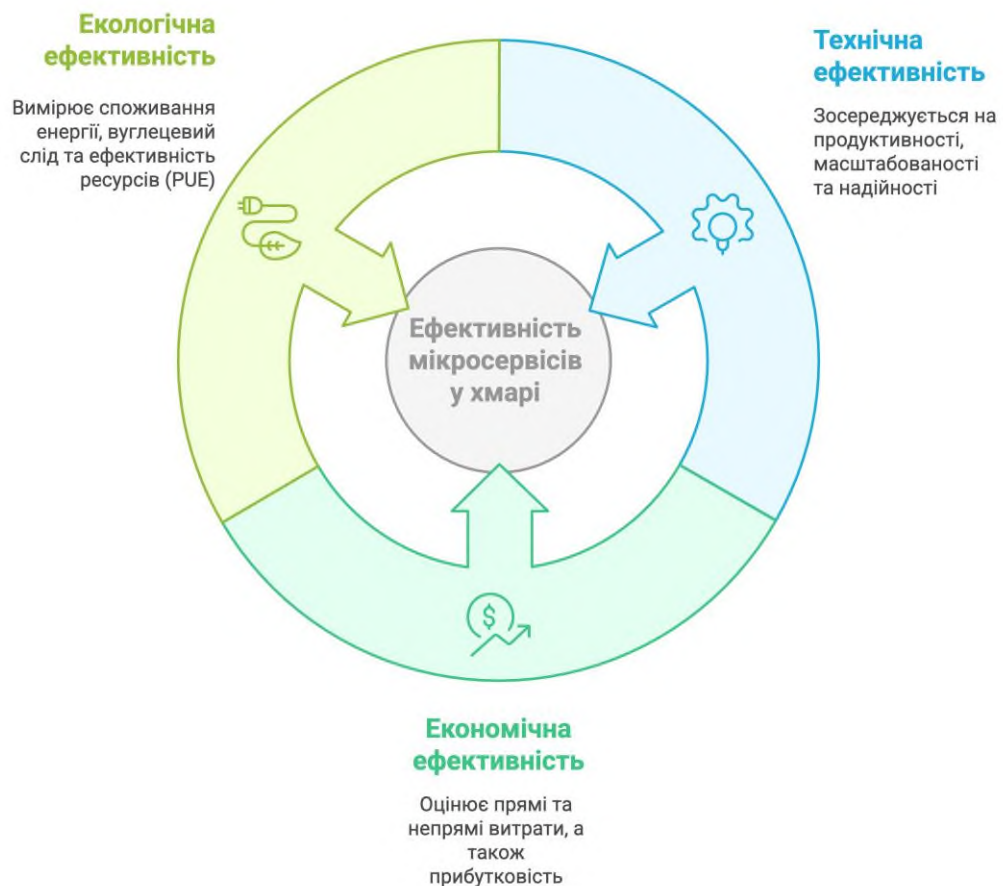


Рис. 1. Взаємозв'язки та компроміси між вимірами ефективності мікросервісів у хмарі

Зокрема, підвищення технічної продуктивності шляхом виділення додаткових обчислювальних ресурсів зазвичай призводить до збільшення фінансових витрат [6]. При цьому інтенсивніше використання ресурсів для досягнення кращої продуктивності також збільшує споживання енергії, що має як економічні (вища вартість електроенергії), так і екологічні (більший вуглецевий слід) наслідки [12].

Цікавим є взаємозв'язок між різними часовими горизонтами оптимізації. Наприклад, подовження терміну служби серверного обладнання зменшує втілений вуглець, пов'язаний з виробництвом нового обладнання, але старіше обладнання зазвичай споживає більше енергії, збільшуючи операційний вуглецевий слід [10]. Це створює необхідність врахування повного життєвого циклу системи при прийнятті рішень щодо оновлення інфраструктури.

Особливо складними є компроміси при горизонтальному масштабуванні мікросервісів – додавання нових екземплярів сервісів покращує продуктивність та доступність, але збільшує витрати та споживання енергії. Ефективні механізми еластичності, що забезпечують адаптивне масштабування відповідно до поточного навантаження, стають ключовим фактором оптимізації всіх трьох вимірів ефективності [5].

Зважаючи на складність взаємозв'язків між різними аспектами ефективності, традиційні методи оптимізації, орієнтовані на єдиний критерій, виявляються недостатніми. Натомість необхідним стає застосування методів багатоцільової оптимізації (БЦО), що дозволяють знаходити компромісні рішення між суперечливими цілями.

Концепція Парето-оптимальності лежить в основі сучасних підходів до БЦО, дозволяючи ідентифікувати набір рішень, де неможливо покращити один показник без погіршення іншого. Серед алгоритмів, що успішно застосовуються для БЦО в контексті мікросервісних архітектур, особливе місце займає NSGA-II (Non-dominated Sorting Genetic Algorithm II) – еволюційний алгоритм, що ефективно знаходить різноманітні рішення вздовж фронту Парето [13]. Також широко використовуються генетичні алгоритми та оптимізація роєм частинок (PSO) для розв'язання складних задач розподілу ресурсів у хмарних середовищах [8].

Стратегії управління ресурсами, що застосовуються для оптимізації мікросервісних систем, охоплюють різні аспекти – від розподілу ресурсів (призначення контейнерів конкретним фізичним машинам) до

динамічного масштабування відповідно до змін навантаження. Особливо перспективним напрямком є консолідація робочих навантажень – переміщення контейнерів між серверами для підвищення рівня використання ресурсів та зниження енергоспоживання [12].

З розвитком технологій штучного інтелекту та машинного навчання з'являються нові можливості для автоматизованого управління ресурсами мікросервісних систем. Методи прогнозування навантаження дозволяють реалізувати проактивне масштабування, а алгоритми навчання з підкріпленням здатні автоматично адаптувати стратегії управління ресурсами до змінних умов функціонування системи [14].

На основі проведеного аналізу пропонується концептуальна багаторівнева модель оцінки ефективності мікросервісних архітектур, що інтегрує технічні, економічні та екологічні аспекти на різних рівнях абстракції системи (Рис. 2).

	Технічна ефективність	Економічна ефективність	Екологічна ефективність
Рівень мікросервісу	Технічні метрики: - Час відгуку - Використання ЦП/пам'яті	Економічні метрики: Вартість контейнера	Екологічні метрики: Енергоспоживання контейнера
Рівень додатку	Технічні метрики: - Наскрізна затримка - Загальна пропускна здатність - Надійність додатку	Економічні метрики: - Загальна вартість експлуатації - ТСО	Екологічні метрики: Загальне енергоспоживання додатку
Рівень інфраструктури	Технічні метрики: - Використання фізичних ресурсів - Продуктивність інфраструктури	Економічні метрики: Вартість інфраструктури	Екологічні метрики: - PUE - Вуглецевий слід - Частка відновлюваної енергії

Інтенсивність кольору ≈ Ступінь впливу на загальну ефективність

Рис. 2. Концептуальна багаторівнева модель оцінки ефективності мікросервісів

Модель структурована за трьома ключовими рівнями мікросервісних систем. На рівні мікросервісу оцінюються локальні показники – час відгуку та пропускна здатність, споживання ресурсів контейнером, вартість функціонування та енергоспоживання на рівні контейнера.

На рівні додатку аналізуються агреговані показники системи мікросервісів – наскрізна затримка при проходженні запиту через ланцюжок сервісів, загальна пропускна здатність додатку, сукупна вартість експлуатації та загальне енергоспоживання. На цьому рівні проявляються ефекти взаємодії між сервісами, що суттєво впливають на загальну ефективність системи.

Рівень інфраструктури охоплює показники базової хмарної платформи – використання фізичних ресурсів серверів, вартість інфраструктури, енергоефективність центру обробки даних (через PUE) та вуглецевий слід.

Ключова перевага моделі – здатність відображати взаємозв'язки між рівнями та вимірами ефективності. Зміни в архітектурі мікросервісу впливають на його локальну продуктивність, загальну ефективність додатку та використання ресурсів. Така перспектива дозволяє приймати обґрунтовані рішення щодо проєктування та оптимізації мікросервісних систем.

Перспективними напрямками досліджень є розробка інструментів для вимірювання вуглецевого сліду, вдосконалення архітектурних тактик та розвиток методів оптимізації, що враховують динамічний характер хмарних середовищ. Особливо перспективним є застосування штучного інтелекту для автоматизованого управління ресурсами з урахуванням множинних цілей оптимізації.

Проведене дослідження демонструє, що ефективність мікросервісних архітектур у хмарному середовищі є багатогранним поняттям, що потребує комплексного підходу. Інтеграція технічних, економічних та екологічних метрик дозволяє сформулювати цілісне уявлення про систему. Ключовим результатом є визнання

неминує компромісів між аспектами ефективності, що робить багатоцільову оптимізацію необхідною для пошуку збалансованих рішень. Запропонована модель створює структуровану основу для аналізу цих компромісів на різних рівнях архітектури.

Таким чином, поєднання багаторівневої моделі оцінки ефективності з методами багатоцільової оптимізації створює фундамент для розробки гібридних технологій оптимізації ресурсів мікросервісів, що забезпечують баланс між продуктивністю, вартістю та екологічною стійкістю хмарних застосунків.

References

1. A. Bogner et al., "On Microservice Analysis and Architecture Evolution: A Systematic Mapping Study," *Applied Sciences*, vol. 11, no. 17, 2021. DOI: 10.3390/app11177856. Retrieved from: <https://www.mdpi.com/2076-3417/11/17/7856>
2. K. Toczé et al., "SoK: Microservice Architectures from a Dependability Perspective," arXiv, 2025. DOI: 10.48550/arXiv.2503.03392. Retrieved from: <https://arxiv.org/abs/2503.03392>
3. A. Sorgalla et al., "Architectural Tactics to Improve the Environmental Sustainability of Microservices: A Rapid Review," arXiv, 2024. DOI: 10.48550/arXiv.2407.16706. Retrieved from: <https://arxiv.org/abs/2407.16706v1>
4. D. K. Pandiya, "Performance Analysis of Microservices Architecture in Cloud Environments," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 12, 2022. Retrieved from: <https://ijritcc.org/index.php/ijritcc/article/view/10745>
5. M. H. Fourati et al., "Cloud Elasticity of Microservices-based Applications: A Survey," February 2024. DOI: 10.21203/rs.3.rs-3925329/v1. Retrieved from: https://www.researchgate.net/publication/378053513_Cloud_Elasticity_of_Microservices-based_Applications_A_Survey
6. I. Papakonstantinou et al., "A Techno-Economic Assessment of Microservices," *Conference on Network and Service Management (CNSM)*, 2020. DOI: 10.23919/CNSM50824.2020.9269114. Retrieved from: <https://www.semanticscholar.org/paper/A-Techno-Economic-Assessment-of-Microservices-Papakonstantinou-Kalafatidis/39a533d5a8e3e91b3fb580d3d91df02e02a1743d>
7. S. Arshad et al., "A Survey of Cloud Computing Variable Pricing Models," *10th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2015)*, Barcelona, Spain, pp. 27-32, 2015. DOI: 10.5220/0005429900270032. Retrieved from: https://www.researchgate.net/publication/277341823_A_Survey_of_Cloud_Computing_Variable_Pricing_Models
8. J. Chen et al., "A multi-objective optimization for resource allocation of emergent demands in cloud computing," *Journal of Cloud Computing*, vol. 10, no. 1, 2021. DOI: 10.1186/s13677-021-00237-7. Retrieved from: https://www.researchgate.net/publication/349707086_A_multi-objective_optimization_for_resource_allocation_of_emergent_demands_in_cloud_computing
9. J. Wang et al., "Peeling Back the Carbon Curtain: Carbon Optimization Challenges in Cloud Computing," *HotCarbon '23: Proceedings of the 2nd Workshop on Sustainable Computer Systems*, Article No. 8, pp. 1-7, 2023. DOI: 10.1145/3604930.3605718. Retrieved from: <https://dl.acm.org/doi/abs/10.1145/3604930.3605718>
10. D. Patterson et al., "Carbon Emissions and Large Neural Network Training," arXiv, 2021. DOI: 10.48550/arXiv.2104.10350. Retrieved from: <https://arxiv.org/abs/2104.10350>
11. "Green Computing Reduces Data Centers' Carbon Footprint," *Sustainability*, January 2023. Retrieved from: <https://greenexdc.com/green-computing-reduces-data-centers-carbon-footprint/>
12. S. Bharany et al., "A Systematic Survey on Energy-Efficient Techniques in Sustainable Cloud Computing," *Sustainability*, vol. 14, no. 10, 2022. DOI: 10.3390/su14106256. Retrieved from: <https://www.mdpi.com/2071-1050/14/10/6256>
13. B. Tan et al., "A NSGA-II-based Approach for Multi-objective Micro-service Allocation in Container-based Clouds," *IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, 2020. DOI: 10.1109/CCGrid49817.2020.00-65. Retrieved from: https://www.researchgate.net/publication/342929909_A_NSGA-II-based_Approach_for_Multi-objective_Micro-service_Allocation_in_Container-based_Clouds
14. H. X. Nguyen et al., "A Survey on Graph Neural Networks for Microservice-Based Cloud Applications," *Sensors*, vol. 22, no. 23, 2022. DOI: 10.3390/s22239492. Retrieved from: <https://www.mdpi.com/1424-8220/22/23/9492>

DOI 10.34132/mspc2025.01.06.23

Dmytro Murzha

**EFFICIENCY METRICS FOR COMPUTATIONAL RESOURCES IN CLOUD ENVIRONMENTS FOR
MICROSERVICE ARCHITECTURES**

The thesis presents a systematization and analysis of multidimensional efficiency metrics for microservices in cloud environments. Technical, economic, and environmental aspects of efficiency, their interrelationships, and trade-offs are examined. A conceptual multi-level model for efficiency assessment is proposed, along with approaches to multi-objective resource optimization. Special attention is paid to modeling a decision-making system that considers conflicting optimization goals at different architecture levels – from an individual microservice to data center infrastructure.

Keywords: *microservices, cloud computing, efficiency metrics, multi-objective optimization, energy efficiency, economic efficiency.*

Відомості про автора / Information about the Author

Дмитро Муржа, аспірант кафедри кібербезпеки та інформаційних технологій, Харківський національний економічний університет імені Семена Кузнеця, м. Харків, Україна. E-mail: dmytro.murzha@hneu.net, orcid: <https://orcid.org/0009-0002-4214-0848>

Dmytro Murzha, PhD student of the Department of Cybersecurity and Information Technologies, Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine. E-mail: dmytro.murzha@hneu.net, orcid: <https://orcid.org/0009-0002-4214-0848>

© Д. Муржа
20.05.2025.

Стаття отримана редакцією

