

# Особливості архітектури Amazon EC2 для масштабування обчислювальних ресурсів

<https://doi.org/10.31713/MCIT.2024.040>

Богдан Красько

Національний університет водного господарства та природокористування  
м. Рівне, Україна  
[b.v.krasko@nuwm.edu.ua](mailto:b.v.krasko@nuwm.edu.ua)

Петро Грицюк

Національний університет водного господарства та природокористування  
м. Рівне, Україна  
[p.m.hrytsiuk@nuwm.edu.ua](mailto:p.m.hrytsiuk@nuwm.edu.ua)

**Анотація** – Особливості архітектури Amazon EC2 для масштабування обчислювальних ресурсів відображають ключові можливості та стратегії, які дозволяють динамічно керувати ресурсами для забезпечення продуктивності та ефективності хмарних сервісів. Amazon Elastic Compute Cloud дозволяє створювати масштабовані віртуальні сервери, які можна налаштувати відповідно до потреб користувачів за допомогою таких функцій, як автоматичне масштабування, підтримка різних типів екземплярів, поєднання з іншими службами AWS, S3, RDS, Lambda, а також використання контейнеризації та еластичного балансування навантаження. Як наслідок, потужності мінімізують витрати на інфраструктуру, забезпечують відновлення після збоїв та автоматично масштабують додатки залежно від навантаження.

**Ключові слова** – AWS; EC2; Autoscaling

## I. ВСТУП

Amazon EC2 є одним із ключових сервісів AWS, які надають масштабовані хмарні обчислювальні ресурси [1]. Архітектурні особливості Amazon EC2 роблять його чудовим рішенням для динамічних обчислювальних вимог, які можуть швидко змінюватися залежно від бізнес-вимог. Автоматично збільшуючи та зменшуючи кількість віртуальних серверів або екземплярів, EC2 забезпечує оптимальне використання ресурсів. Ключовими аспектами масштабованої архітектури EC2 є її еластичність, її: автоматизація у формі керування екземплярами EC2 за допомогою автоматичного масштабування, інтеграція з іншими службами AWS і широкий спектр типів екземплярів для різних обчислювальних потреб. Автоматичне масштабування може допомогти встановити правила для автоматичного додавання та видалення екземплярів на основі навантаження, що забезпечує покращену продуктивність і щільність ресурсів [4]. Крім того, EC2 дає можливість використовувати різні патерни, такі як розподілені системи, кластеризація та безсерверні рішення через AWS Lambda, що робить його дуже гнучким для імплементації малих і великих масштабованих варіацій [3]. Amazon EC2 також підтримує високий рівень доступності через можливість розгортання екземплярів у кількох зонах доступності та регіонах.

Це дозволяє забезпечити стійкість до збоїв і високу надійність хмарної інфраструктури для критично важливих додатків [2].

## II. ВИКЛАДЕННЯ ОСНОВНОГО МАТЕРІАЛУ

Amazon Elastic Compute Cloud (Amazon EC2) — це хмарна служба, яка надає масштабовані обчислювальні потужності в хмарі AWS [5]. Основною перевагою цієї платформи є гнучкість і можливість динамічного масштабування обчислювальних ресурсів в залежності від потреб користувачів [6].

Amazon EC2 (Elastic Compute Cloud) надає можливість автоматично збільшувати або зменшувати кількість обчислювальних ресурсів у відповідь на зміну навантаження. Це досягається за допомогою механізму Auto Scaling, який дозволяє додавати нові екземпляри (інстанси) або вимикати їх залежно від налаштувань і політик.

Auto Scaling групи дозволяють визначити мінімальну, максимальну та бажану кількість інстансів. Наприклад, встановивши мінімум у 2 інстанси, максимум у 10, і бажану кількість у 5. Auto Scaling автоматично забезпечить, щоб кількість інстансів відповідала цим параметрам [4].

### Основні функції Auto Scaling включають

- Автоматичний моніторинг стану інстансів і заміна несправних для підтримки бажаної кількості [6].
- Розподіл інстансів між різними зонами доступності для забезпечення високої доступності та стійкості.
- Використання різних типів інстансів і варіантів покупки (Spot, On-Demand) для зниження витрат [5].

Amazon EC2 (Elastic Compute Cloud) пропонує широкий вибір типів інстансів, кожен з яких оптимізований для певних завдань. Це дозволяє користувачам обирати найбільш підходящий набір ресурсів для своїх потреб, що значно підвищує ефективність та гнучкість масштабування [1].

### Основні типи інстансів

- Загального призначення - інстанси забезпечують баланс між обчислювальними потужностями, пам'яттю та мережею. Вони підходять для різноманітних навантажень, таких як веб-сервери та сховища коду.
- Оптимізовані для обчислень - інстанси мають високу обчислювальну потужність і підходять для завдань, що потребують інтенсивних обчислень, таких як наукові моделювання та обробка великих даних.
- Оптимізовані для пам'яті - інстанси мають велику кількість оперативної пам'яті і підходять для баз даних та інших додатків, що потребують багато пам'яті.
- Оптимізовані для зберігання - інстанси забезпечують високу пропускну здатність для зберігання даних і підходять для завдань, що потребують інтенсивного доступу до дискових ресурсів, таких як обробка великих даних та сховища [3].
- Прискорені обчислення - інстанси використовують графічні процесори (GPU) або інші апаратні прискорювачі для завдань, що потребують високої обчислювальної потужності, таких як машинне навчання та обробка відео [6].

Завдяки такому різноманіттю типів інстансів, користувачі можуть обирати найбільш оптимальні ресурси для своїх конкретних завдань, що дозволяє ефективно масштабувати свої додатки та сервіси.

Amazon EC2 (Elastic Compute Cloud) є потужним інструментом для розгортання та управління обчислювальними ресурсами в хмарі. Він інтегрований з іншими сервісами AWS, такими як Amazon Elastic Block Store (EBS) та Amazon S3, що дозволяє масштабувати як обчислювальні ресурси, так і системи зберігання даних без переривання роботи додатків.

Amazon EBS забезпечує стійке блокове зберігання для використання з EC2, що дозволяє зберігати дані на довготривалій основі та забезпечує високу продуктивність для критично важливих робочих навантажень. Це особливо корисно для баз даних, файлових систем та інших додатків, які потребують постійного доступу до даних.

Amazon S3 (Simple Storage Service) надає масштабоване об'єктне зберігання, яке ідеально підходить для зберігання та захисту великих обсягів даних. Завдяки інтеграції з EC2, користувачі можуть легко зберігати резервні копії, архіви та інші важливі дані, забезпечуючи їх доступність та безпеку [1].

Автоматизоване розподілення трафіку між декількома інстансами EC2 за допомогою Elastic Load Balancer (ELB) забезпечує рівномірне завантаження ресурсів, підвищуючи ефективність використання серверів. Це дозволяє уникнути перевантажень окремих інстансів, що може

привести до зниження продуктивності або навіть до відмови додатків.

ELB автоматично розподіляє вхідний трафік між доступними інстансами, враховуючи їх поточне завантаження та стан. Це забезпечує високу доступність додатків, оскільки у разі виходу з ладу одного з інстансів, трафік буде перенаправлено на інші, що залишилися в робочому стані. Крім того, ELB підтримує масштабування додатків, дозволяючи додавати або видаляти інстанси в залежності від поточних потреб, що забезпечує гнучкість і адаптивність системи.

Користувачі можуть масштабувати ресурси, зберігаючи високий рівень безпеки за допомогою різноманітних механізмів контролю доступу. Одним із таких механізмів є Virtual Private Cloud (VPC), який дозволяє створювати ізольовані мережеві середовища в хмарі. Security Groups забезпечують контроль доступу на рівні екземплярів, дозволяючи визначати правила вхідного та вихідного трафіку. Network ACLs (Access Control Lists) додають додатковий рівень безпеки, контролюючи трафік на рівні підмереж. Крім того, шифрування даних гарантує, що інформація залишається захищеною як під час передачі, так і при зберіганні. Завдяки цим інструментам користувачі можуть ефективно керувати своїми ресурсами, не жертвуючи безпекою.

Amazon EC2 (Elastic Compute Cloud) працює в багатьох регіонах по всьому світу, кожен з яких складається з кількох зон доступності (Availability Zones, AZ). Це дозволяє масштабувати ресурси глобально, забезпечуючи високу відмовостійкість та низьку затримку для користувачів у будь-якому регіоні.

### III. ВИСНОВКИ

Таким чином, архітектура Amazon EC2 забезпечує ефективне масштабування обчислювальних ресурсів завдяки високій еластичності, різноманітності інстансів, інтеграції з іншими сервісами AWS, балансуванню навантаження та глобальній доступності.

### СПИСОК ЛІТЕРАТУРИ

- [1] Amazon Web Services (AWS). Amazon Elastic Compute Cloud (EC2) Documentation. <https://docs.aws.amazon.com/ec2/>.
- [2] Jassy, A. (2017). *The Cloud Computing Revolution: The AWS Story*. New York: AWS Publishing.
- [3] Barroso, L. A., & Hölzle, U. (2009). *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. *Synthesis Lectures on Computer Architecture*, 4(1), 1-108.
- [4] Kratzke, N., & Quint, P.-C. (2017). *Understanding Cloud-native Applications after 10 Years of Cloud Computing – A Systematic Mapping Study*. *Journal of Systems and Software*, 126, 1-16.
- [5] Duan, J., Faker, P., Fesak, A., & Stuart, T. (2012). *Benefits and Drawbacks of Cloud-Based vs Traditional ERP Systems*. *Proceedings of the 2012-13 Course on Advanced Cloud Computing*, 1-10.
- [6] Villamizar, M., et al. (2015). *Evaluating the Impact of Auto-scaling Mechanisms on the Performance of Web Applications in the Cloud*. *International Journal of Cloud Computing*, 7(2), 116-124.