

## **АРХІТЕКТУРИ КОМП'ЮТЕРНИХ СИСТЕМ ОПРАЦЮВАННЯ ВЕЛИКИХ ДАНИХ**

При створенні систем опрацювання великих даних ключовим компонентом виступає програмне забезпечення. Оскільки, саме воно визначає вимоги до апаратних засобів, які мають забезпечувати належний рівень гнучкості й адаптивності. На сьогодні домінуючою є кластерна архітектура на основі доступних за ціною та уніфікованих типових обчислювальних систем (x86\_64, 10G Ethernet, InfiniBand, HDD, SSD). Використання таких засобів дає змогу:

- відносно просто будувати системи з можливістю горизонтального (збільшення кількості обчислювальних вузлів), а іноді й вертикального (нарощення параметрів окремих вузлів) масштабування;

- просто мігрувати обчислювальним задачам між різними датацентрами та хмарними сервісами;

- швидко розгортати обчислювальні системи та припиняти їх роботу за відсутності обчислювальних потреб, що в результаті дає змогу економити кошти.

Узагальнені архітектури обчислювальних систем базуються здебільшого на архітектурах програмних систем й орієнтовані на конкретні обчислювальні задачі.

Обчислювальні задачі визначаються характером вхідних даних та специфікою отриманих результатів. Вхідні дані можуть надходити пакетами (певними порціями через певні інтервали часу) або потоково (дискретний неперервний потік). До отриманих результатів обчислень можуть висуватись різні вимоги:

- актуальність (отримання результатів у режимі реального часу, або через певні часові інтервали);

- точність (результати можуть бути з певною, наперед заданою точністю);

- формат представлення.

З урахуванням наведеного, умовно прийнято виділяти лямбда- та каппа-архітектури. Лямбда архітектура забезпечує надійне та тривале зберігання даних, передбачає елементи пакетного опрацювання, опрацювання даних відбувається із задовільною швидкістю, вихідні дані в сховищі є незмінними (immutable). Водночас дані з такого сховища важко мігрувати та складно їх реструктурувати. Системи на основі лямбда-архітектури мають велику кількість компонент, що зумовлює деяку їх складність (багатошаровість). Водночас, такі системи добре підходять для опрацювання даних на основі історичних даних. Використовуються вони для різноманітних задач аналізу даних у соціальних мережах, моніторингу хмарних сервісів та опрацюванні даних, які отримані від InternetOfThings.

Каппа-архітектура може частково розглядатись як альтернативна. Надає можливості опрацювання даних у режимі реального часу, може працювати на дешевшому обладнанні, але не характеризується достатньою стійкістю до помилок та відмов. Доволі часто каппа-архітекта використовується при побудові систем для розв'язання задач машинного навчання.

### **Література**

1. Samizadeh I. A brief introduction to two data processing architectures—Lambda and Kappa for Big Data / Iman Samizadeh// Medium. Towards Data Science. [Електронний ресурс] Режим доступу: URL: <https://towardsdatascience.com/a-brief-introduction-to-two-data-processing-architectures-lambda-and-kappa-for-big-data-4f35c28005bb>