

І. Ю. Петровська, Г. А. Кучук

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

РОЗПОДІЛ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ У ХМАРНИХ СИСТЕМАХ

Анотація. У статті розглянуто класифікацію існуючих технологій надання хмарних послуг. Для кожної технології визначено переваги та недоліки. Також визначено характерні особливості хмарних обчислень, які необхідно враховувати при розподілі ресурсів. Визначено, що у всіх технологіях враховуються тільки необхідні обсяги процесорного ресурсу, оперативної пам'яті та місця у сховищі даних. При цьому не враховується специфіка роботи застосунків. Також не враховується розділення ресурсів між різними додатками. Таким чином, не завжди має місце вибір оптимального ресурсу для розміщення заявок клієнтів. Це часто призводить до істотного зниження продуктивності додатків. Крім того, неефективно використовується хмарний ресурс. Існуючі підходи та методи розподілу ресурсів не враховують усіх особливостей хмарних обчислень. Це може призвести до неефективного використання хмарної інфраструктури. Тому **метою даної статті** є визначення характерних рис хмарних обчислень, які необхідно буде враховувати при розподілі ресурсів. Завдання полягає в забезпеченні рівномірного розподілу навантаження на всі хмарні сервери. У результаті проведеного аналізу доведено, що технологія IaaS є найбільш уразливою до якості розподілу ресурсів. Визначено необхідний набір показників для технології IaaS, які необхідно враховувати при розподілі ресурсів.

Ключові слова: хмарні технології, хмарні ресурси, оптимальний розподіл ресурсів.

Вступ

Розвиток інформаційного суспільства змінило уявлення користувачів про надання інформаційних сервісів. На перший план виходять не технічні характеристики послуг, а якісні показники. До них відносяться своєчасне задоволення потреб бізнесу, простота використання, швидкість виконання типових операцій та ін. У зв'язку з цим дедалі більшої популярності набувають хмарні обчислення (англ. cloud computing). Крім того, додатковим стимулом стали епідемія COVID-19 та агресія Росії у 2022 році. Хмарні обчислення – це можливість надання обчислювального ресурсу через Інтернет. При цьому обов'язкові такі характеристики [1]: **самообслуговування на вимогу**, споживач самостійно визначає та змінює обчислювальні потреби, такі як серверний час, швидкості доступу та обробки даних, обсяг даних, що зберігаються, без взаємодії з представником постачальника послуг; **універсальний доступ по мережі**, послуги доступні споживачам по мережі передачі даних незалежно від термінального пристрою, що використовується; **об'єднання ресурсів**, постачальник послуг об'єднує ресурси обслуговування великої кількості споживачів у єдиний пул для динамічного перерозподілу мощностей між споживачами за умов постійного зміни попиту потужності; при цьому споживачі контролюють лише основні параметри послуги (наприклад, обсяг даних, швидкість доступу); фактичний розподіл ресурсів, що надаються споживачеві, здійснює постачальник (у деяких випадках споживачі таки можуть керувати деякими фізичними параметрами перерозподілу, наприклад, вказувати бажаний центр обробки даних з міркувань географічної близькості); **еластичність**, послуги можуть бути надані, розширені, звужені будь-якої миті часу, без додаткових витрат; **облік споживання**, постачальник послуг автоматично обчислює спожиті ресурси на певному рівні абстракції (наприклад, обсяг даних, що зберігаються, пропускна спроможність, кількість користувачів, кількість транзакцій); на основі цих даних він оцінює обсяг наданих споживачам послуг.

Виконання цих вимог суттєво впливає на розподіл ресурсів у хмарній інфраструктурі. У науковій літературі розглядається багато методів розподілу ресурсів в інформаційних та обчислювальних структурах. У статті [2] автор пропонує алгоритм random swap clustering. Однак він суттєво втрачає ефективність при застосуванні для хмарних інфраструктур. При проведенні оптимального розподілу пропускних здібностей для самовідновлювального (англ. self-healing) сегменту комп'ютерної мережі в [3] специфіка хмарних складових системи не розглядається. Методи, запропоновані у [4], орієнтовані тільки на мобільні компоненти. ЇВ [5] більшою мірою враховуються особливості рухомих об'єктів, а не характеристики cloud computing. У статті [6] розглядаються методи розподілу ресурсів, що відсуюють на другий план часові показники, роблячи акцент на безпекових питаннях. Праці [7, 8] спрямовані тільки на специфіку Big Data. Стаття [9] орієнтована тільки на системи, що обробляють відеозображення. А в [10] акцент зроблено тільки на специфічні застосунки.

Однак розглянуті підходи та методи розподілу ресурсів не враховують усіх особливостей хмарних обчислень. Це може призвести до неефективного використання хмарної інфраструктури. Тому метою даної статті є визначення характерних рис хмарних обчислень, які необхідно буде враховувати при розподілі ресурсів.

Також необхідно виділити технології надання хмарних послуг, які є найбільш вразливими при неоптимальному розподілі або перерозподілі ресурсів.

1. Класифікація технологій надання хмарних послуг

На сьогодні технології надання хмарних послуг можна поділити на такі типи, які визначаються відповідними моделями хмарних послуг: IaaS, інфраструктура як послуга; PaaS, платформа як послуга; SaaS, програмне забезпечення як послуга; CaaS, розрахункові обчислення як сервіс).

За технологією на базі моделі обслуговування IaaS споживач отримує інформаційно-технологічні

ресурси – віртуальні сервери з певною обчислювальною потужністю та обсягами пам'яті, тобто повне Software та віртуальне Hardware. Провайдер забезпечує фізичне Hardware (обладнання). Він встановлює на нього Software для створення віртуальних машин, але не займається установкою і підтримкою Software (програмне забезпечення) користувача. Провайдер контролює тільки фізичну та віртуальну інфраструктуру, яку надає клієнтові у віртуальному вигляді за допомогою хмарних серверів. Зазвичай, клієнтами IaaS є системні адміністратори компаній.

У випадку використання технології на базі моделі обслуговування PaaS хмарний провайдер надає доступ до операційних систем, засобів розробки і тестування, систем управління базами даних, тобто тільки до системного та прикладного Software. Провайдер контролює не тільки сервери, системи зберігання даних і обчислювальні потужності, але також пропонує користувачеві на вибір певні платформи і засоби управління ними. Найбільш часто дану технологію використовують розробники Software.

Хмарна модель SaaS є найпоширенішою серед клієнтів хмарних провайдерів. У відповідній технології програми та сервіси розробляє і обслуговує провайдер, розміщує їх у хмарі і пропонує кінцевому користувачеві через браузер або застосунок на його комп'ютері. Основні клієнти технології SaaS – звичайні користувачі, наприклад, користувачі налагоджених бізнес-застосунків через Інтернет.

Деякі хмарні провайдери надають послуги з використанням технології SaaS. При цьому клієнт отримує обчислювальні ресурси для виконання у хмарі власних застосунків. Зазвичай цю послугу використовують у випадку, якщо власне Hardware неспроможне виконати завдання з якихось причин.

Відповідно, для кожної технології є свої особливості щодо контролю та прогнозування завантаження фізичної та віртуальної інфраструктури хмарного середовища. Це суттєво залежить від рівня доступу користувача до хмарних послуг.

2. Переваги та недоліки технологій надання хмарних послуг

IaaS – це найбільш гнучка модель хмарних послуг із простим процесом розгортання обладнання. IaaS дозволяє підприємствам нарощувати обчислювальні ресурси при необхідності, замість того, щоб купувати дороге обладнання для власної інфраструктури. Вартість IaaS варіюється і в основному залежить від потреб клієнта в CPU та RAM. IaaS — це також економічна модель, у тому числі через високу масштабованість та автоматизацію хмарних послуг. Послуги IaaS актуальні як для стартапів та невеликих компаній, так і для великого бізнесу. Отже, перехід на IaaS заощаджує час та гроші клієнта.

Але слід зазначити, що побудувати готове рішення на базі IaaS за пару годин практично неможливо. Крім того, при наявності власного Hardware немає сенсу орендувати хмарні віртуальні пристрої. В цих випадках переважає платформа PaaS. Подібно до інших хмарних сервісів, PaaS дозволяє клієнтам користуватися сучасними потужними інструментами

розробки, підтримку яких бере на себе провайдер. Платформа як послуга хороша тим, на відміну IaaS, що відразу готова до роботи. За допомогою PaaS підвищується швидкість розробки, тестування та доставки додатків. Отже, технологія PaaS допомагає компаніям різного розміру оптимізувати процес розробки. Наприклад, PaaS може спростити роботу великої команди розробників, які займаються одним і тим самим проектом. Ця технологія є кращою для компаній з існуючою IT-інфраструктурою.

Однак багато дрібних і середніх компаній не мають можливості купувати on-premise-рішення. Крім того, короткострокові проекти, що потребують швидких, простих і доступних рішень, доцільніше реалізовувати самій компанії. У таких випадках найвигідніша технологія SaaS. Віддалена настройка та обслуговування Software провайдером надають компанії-замовнику більше часу для вирішення інших важливих питань та завдань. SaaS-рішення керуються централізовано і розміщуються на віддаленому сервері. Виробник, а не користувач, несе відповідальність за налаштування необхідного обладнання та програмного забезпечення. Найчастіше для роботи SaaS не потрібно завантаження та встановлення Software на пристрій.

У випадку, коли потрібна тимчасова хмарна послуга, найбільш прийнятною є технологія SaaS. Її застосування дозволяє користувачеві доповнити на деякий період власні ресурси за рахунок придбання потрібних віртуальних ресурсів.

Отже, кожна хмарна модель пропонує певні функції та можливості. Коли бізнес має набір конкретних завдань і розуміння переваг різних типів хмарних сервісів, простіше вибрати відповідний. Технологія IaaS дає практично повний контроль за готовою інфраструктурою.

Це дозволяє організації створити стек технологій, повністю адаптований до потреб бізнесу. Підприємства, які вже мають деякі ресурси та IT-відділ, можуть вибрати технологію PaaS. Готова платформа допоможе компаніям розробляти індивідуальні рішення, які легко інтегрувати з існуючими робочими процесами. Технологія SaaS дозволяє підприємствам заощаджувати гроші.

Клієнтам не потрібно самостійно займатися розробкою та підтримкою програмного забезпечення. А за браком ресурсів незамінна технологія SaaS.

Хоча, на перший погляд, ці моделі схожі, SaaS, PaaS та IaaS надають різні рівні послуг, а оптимальне використання ресурсів залежить від використовуваних методів розподілу та перерозподілу ресурсів.

3. Методи розподілу ресурсів у хмарних інфраструктурах

У хмарних інфраструктурах, як і серед віртуалізації, проблема планування та оптимізації використання ресурсів вирішується двома способами: перерозподіл ресурсів адміністратором середовища віртуалізації вручну; автоматичне планування ресурсів за допомогою методів розподіленого виділення ресурсів DRS (Distributed Resource Scheduling) та розподіленого управління енергоспоживанням DPM.

У різних середовищах методи DRS та DPM працюють по-різному, проте логіка роботи зберігається.

Алгоритм роботи DRS у загальному випадку складається з набору простих кроків:

- 1) вибирається найменш навантажений хост серед тих, які мають достатній ресурс;
- 2) під час роботи хостів збирається статистика завантаження CPU на них;
- 3) якщо CPU на одному з хостів протягом встановленого інтервалу навантажено більше за певний поріг, то приймається рішення про перерозподіл ресурсів;
- 4) якщо приймається рішення про перерозподіл ресурсів, то метод DRS або перенесе завдання на більш продуктивний або менш навантажений хост, або перенесе інші завдання інші сервери.

Внаслідок роботи методу DRS вдається оптимально розподілити навантаження між серверами хмарного середовища. Це підвищує як пікову продуктивність хостів, і ефективність використання ресурсів. Однак метод оцінює лише навантаження на CPU і не прогнозує зміну навантаження за іншими параметрами. З цієї причини метод нечасто застосовування в середовищах cloud computing, де адміністратор хмарного обчислювального середовища не може передбачити всіх небажаних наслідків від перерозподілу ресурсів у такий спосіб.

Аналогічно, метод DPM вирішує задачу зниження енергоспоживання, проте не вирішує задачу максимізації ефективного використання наявних ресурсів. Отже існуючі методи не дозволяють оптимально розподіляти та перерозподіляти хмарні ресурси. При цьому якість розподілу суттєво залежить від обраної технології. Результати порівняльного аналізу можливості оптимізації за деякими суттєвими параметрами наведено в табл. 1.

Таблиця 1 – Порівняльний аналіз хмарних технологій

Технологія Параметр	IaaS	PaaS	SaaS	SaaS
Вартість	←	←	←	←
Витрати клієнта на обслуговування	←	←	←	←
Витрати провайдера на обслуговування	←	←	←	←
Можливість перерозподілу ресурсів	–	←	←	←
Прогнозування використання ресурсів	–	←	←	+
Балансування навантаження	–	←	+	+
Можливість оцінки надійності інфраструктури	–	+	+	+

За результатами порівняльного аналізу хмарних технологій можна дійти невтішного висновку, що методи, застосовувані в хмарних обчислювальних середовищах, які завжди дозволяють оптимально розподілити хмарні ресурси. Найбільш ефективно завдання розподілу ресурсів у хмарному середовищі вирішується під час використання технології SaaS.

Суттєвих проблем не виникає і при використанні складніших технологій SaaS та PaaS, хоча є низка особливостей, які перешкоджають використанню методів DRS та DPM у певних сценаріях.

Однак у хмарних обчислювальних середовищах на основі технології IaaS виникають проблеми оптимального планування ресурсів як на етапі їх надання, так і під час використання. Ці проблеми неможливо вирішити вручну, як у простих технологіях.

По-перше, адміністратору хмарної інфраструктури невідомо, які потреби додатків для ефективного їх використання. По-друге, заважає наявність «статичності» платформи – відсутність реакцію зміна потреб у часі.

Отже, у технології IaaS не вирішується завдання оптимального перерозподілу обчислювальних ресурсів у середовищі хмар. Так, наприклад, одне з найбільш популярних в даний час рішень з управління хмарною інфраструктурою OpenStack, підтримує всього 3 методи виділення ресурсів: випадковий, випадковий у межах зони доступності та простий (виділення ресурсів відбувається по черзі). Що стосується перерозподілу навантаження за її нерівномірності, таких підходів у середовищах хмарних обчислень поки що не існує.

Простий метод розподілу виділяє ресурси кожного з хостів, які задовольняють запити клієнта, по черзі, доки вони не закінчаться. При випадковому методі розподілу ресурсів хост для запуску екземпляра вибирається випадково серед хостів, на яких є необхідна кількість ресурсів. Метод розподілу «випадковий у межах зони доступності» працює за тим самим принципом, що й випадковий, проте задіє лише ті хости, які заздалегідь визначені адміністратором хмарного обчислювального середовища як зона доступності для даного екземпляра.

Таким чином, порівняльний аналіз показав, що всі три описані методи розподілу ресурсів мають два істотні недоліки. Перший у тому, що під час ухвалення рішення про використання тієї чи іншої ресурсу немає оцінка оптимальності прийнятого рішення. Другий недолік у тому, що це методи працюють лише з етапі початкового виділення ресурсів примірнику. Крім того, жоден із розглянутих методів розподілу обчислювальних ресурсів не враховує таких важливих складових, як дискова підсистема та мережа. Через це програми з підвищеними вимогами до продуктивності диска (бази даних, аналітичні системи) та підвищеними вимогами до продуктивності мережі (відеоконференції, інтернет-шлюзи) випадають із логіки розподілу ресурсів.

Тому у зв'язку зі специфікою хмарних обчислювальних середовищ необхідно в технології IAAS враховувати при розподілі ресурсів розширений набір показників, у тому числі змінюваних у часі:

- продуктивність (кількість процесорів; число ядер процесорів; тактова частота процесорів; швидкість роботи мережі);
- навантаженість (зайняте RAM, завантаження CPU; навантаження на дискову підсистему; відгук дискової підсистеми завантаження мережі; кількість екземплярів, що виконуються);

- надійність (середній час між збоями; фізичний стан обладнання; якість складання обладнання; рівень компетенцій персоналу для обслуговування обладнання).

Висновки

У статті розглянуто класифікацію існуючих технологій надання хмарних послуг, для кожної з яких визначено переваги та недоліки, визначено характерні особливості хмарних обчислень, які необхідно враховувати при розподілі ресурсів.

Визначено, що недоліком кожної з перерахованих технологій є те, що потреби додатків, що працюють усередині примірників, враховуються лише у контексті необхідних обсягів процесорного ресурсу, оперативної пам'яті та місця у сховищі даних. При цьому не враховується специфіка роботи додатків, а також те, як із цим додатком розділятимуть ресурси вже розгорнуті на тих самих хостах інші додатки. Таким чином, не завжди має місце вибір оптимального

ресурсу для розміщення заявок клієнтів, що часто призводить до істотного зниження продуктивності додатків та ефективності використання хмарного ресурсу. Завдання полягає в тому, щоб забезпечити рівномірний розподіл навантаження на всі сервери хмарного обчислювального середовища різної продуктивності, надавши при цьому найкращі умови для функціонування додатків і одночасно максимального ефективного використання наявних обчислювальних ресурсів. Доведено, що технологія IAAS найбільш уразлива при неоптимальному розподілі чи перерозподілі ресурсів. Визначено розширений набір показників для технології IAAS, у тому числі змінюваних у часі, які необхідно враховувати під час розподілу ресурсів.

Вирішення поставленого завдання дозволить підвищити ефективність використання ресурсів хмарного обчислювального середовища в цілому, підвищити продуктивність додатків та знизити витрати на інфраструктуру та її супровід.

СПИСОК ЛІТЕРАТУРИ

1. Zhen Xiao, Weija Song, and Qu Chen, "Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment", IEEE transaction on parallel and distributed systems, Vol/ 24, Is. 6, pp. 1107 – 1117, June 2013, doi: 10.1109/TPDS.2012.283.v/
2. P. Franti, "Efficiency of random swap clustering", Journal of Big Data, 2018, vol. 5, No. 13, pp. 1-29. doi: 10.1186/s40537-018-0122-y.
3. N. Kuchuk, O. Shefer, G. Cherneva, and F. A. Alnaeri, "Determining the capacity of the self-healing network segment", Advanced Information Systems, vol. 5, no. 2, pp. 114–119, Jun. 2021, doi: 10.20998/2522-9052.2021.2.16.
4. Ye. Qiang, and W. Zhuang, "Distributed and adaptive medium access control for internet-of-things-enabled mobile networks", IEEE Internet of Things Journal, 2017, vol. 4, no. 2, pp. 446-460, doi: 10.1109/IJOT.2016.2566659.
5. H. Khudov, K. Tahyan, V. Chepurnyi, I. Khizhnyak, K. Romanenko, A. Nevodnichii, and O. Yakovenko, "Optimization of joint search and detection of objects in technical surveillance systems", Advanced Information Systems, 2020, Vol. 4, No. 2, pp. 156-162, doi: 10.20998/2522-9052.2020.2.23.
6. S. Semenov, and Cao Weilin, "Testing process for penetration into computer systems mathematical model modification", Advanced Information Systems, Vol. 4, No. 3, pp. 133–138. 2020, doi: 10.20998/2522-9052.2020.3.19.
7. G. Kuchuk, A. Kovalenko, I.E. Komari, A. Svyrydov, and V. Kharchenko, "Improving big data centers energy efficiency: Traffic based model and method", Studies in Systems, Decision and Control, vol. 171, Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (Eds.), Springer Nature Switzerland AG, pp. 161-183, 2019, doi: 10.1007/978-3-030-00253-4_8.
8. A. Nechausov, I. Mamusuć, and N. Kuchuk, "Synthesis of the air pollution level control system on the basis of hyperconvergent infrastructures", Advanced Information Systems, vol. 1, no. 2, 2017, pp. 21–26. DOI: 10.20998/2522-9052.2017.2.04.
9. H. Kuchuk, A. Kovalenko, B.F. Ibrahim, and I. Ruban, "Adaptive compression method for video information", International Journal of Advanced Trends in Computer Science and Engineering, 8(1), pp. 66–69, 2019, doi: http://dx.doi.org/10.30534/ijatcse/2019/1181.22019.
10. S. Bulba, "Composite application distribution methods", Advanced Information Systems", vol. 2, no. 3, pp. 128–131, 2018, doi: 10.20998/2522-9052.2018.3.22.

Received (Надійшла) 22.03.2022

Accepted for publication (Прийнята до друку) 25.05.2022

Features of the distribution of computing resources in cloud systems

Inna Petrovska, Heorhii Kuchuk

Abstract. The article considers the classification of existing technologies for providing cloud services. Advantages and disadvantages are determined for each technology. Characteristic features of cloud computing, which must be taken into account when allocating resources, are also defined. It was determined that all technologies take into account only the necessary amounts of processing resources, RAM and storage space. At the same time, the specifics of the applications are not taken into account. It also does not take into account the sharing of resources between different applications. Thus, it is not always possible to choose the optimal resource for placing customer applications. This often leads to a significant decrease in application performance. In addition, the cloud resource is used inefficiently. Existing approaches and methods of resource allocation do not take into account all the features of cloud computing. This can lead to inefficient use of cloud infrastructure. Therefore, the purpose of this article is to determine the characteristic features of cloud computing, which will need to be taken into account when allocating resources. The task is to ensure an even distribution of the load on all cloud servers. As a result of the analysis, it was proved that the IAAS technology is the most vulnerable to the quality of resource allocation. The necessary set of indicators for IAAS technology, which must be taken into account when allocating resources, is determined.

Keywords: cloud technologies, cloud resources, optimal allocation of resources.