

ПОКАЗНИКИ ОЦІНКИ МЕРЕЖЕВОЇ ДОСТУПНОСТІ ТА ЗВ'ЯЗНОСТІ ІНФОРМАЦІЙНИХ СИСТЕМ У ОБЧИСЛЮВАЛЬНИХ ХМАРАХ

Ілля Малініч, Ярослав Іванчук

***Анотація.** В публікації розглянуто різні показники доступності та зв'язності інформаційних систем, які розміщуються у обчислювальних хмарах. Приділено увагу інформаційним системам, які складаються з мікросервісів. Розглянуті показники є важливими для розміщення мікросервісів у хмарах, а також визначення оптимального їх розміщення: у приватних чи публічних хмарах. Визначення показників дозволить полегшити розробку нових методів для спрощення процесу розміщення мікросервісів.*

Ключові слова: комп'ютерні мережі, обчислювальні хмари, аптайм, затримка пакетів

I. ВСТУП

При проектуванні інформаційних систем у обчислювальних хмарах, зокрема на основі мікросервісів, перед архітекторами та DevOps-інженерами ставиться ціль щоб такі системи могли найбільш ефективно працювати, була швидко та постійно доступними з мережі Інтернет для персоналу компанії-користувача та її клієнтів, а також не створювали надлишкових витрат на їх утримання. При проектуванні інформаційних систем у гібридних хмарах завжди виникає вузьке місце (англ. bottleneck) між обома частинами такої хмари: приватної та публічної. Виникнення проблем на межі обох хмар часто призводить до деградації якості роботи сервісів, що працюють у гібридних хмарах. Для проведення оцінювання якості роботи різних розгорнутих у хмарі сервісів зазвичай потрібно проводити моніторинг різних метрик, які для кожного хмарного сервісу часто є індивідуальними.

II. ДАНІ ТА МЕТОДИ

Різні дослідники [1,2] та хмарні архітектори [3,4] виділяють різні показники мережевої доступності розміщених у хмарі сервісів як найбільш універсальні метрики для оцінки якості їх роботи зі сторони кінцевого користувача. Серед цих показників можна виділити найбільш універсальні, такі як аптайм (англ. uptime), час відгуку сервера (або TTFB), затримка пакетів (або затримка пінгу, англ. ping), кількість мережевих переходів (англ. traceroute hops). Ці чотири основних показника можна використати для того щоб оцінити мережеву доступність інформаційних систем у гібридних хмарах. Аптайм – це основний параметр, за допомогою якого можна оцінити наскільки система стабільно працює та зберігає зовнішню доступність з мережі Інтернет. Аптайм вимірюється у відсотках і відображає скільки часу в рік система була доступною. Протилежним поняттю аптайму є даунтайм (англ. downtime), який відображає скільки часу в рік система була недоступною. Тобто аптайм рівний різниці між сотнею відсотків та даунтаймом. Навіть якщо роботу обчислювальної хмари буде зупинено з інших причин, не пов'язаних з доступом до мережі Інтернет – наприклад знеструмлення, вихід з ладу обладнання та інших – це у будь-якому випадку відобразиться на значенні аптайму, що робить його доволі універсальним показником оцінки доступності. Для аналізу цих показників можна використовувати наступні методи: описативна статистика, кореляційний аналіз, регресійний аналіз, кластеризація, а також методи числових рядів. Завдяки застосуванню цих методів можна буде провести аналіз їх впливу на розміщення у гібридній хмарі (або навпаки).

III. РЕЗУЛЬТАТИ ТА АНАЛІЗ

Згідно загальноприйнятим стандартам хмарної та телеком-індустрій прийнятне значення SLA коливається між 99.9% та 99.999% [5]. Подібне значення забезпечується постачальником хмарних послуг згідно SLA (Угодою про рівень послуг, англ. Service Level Agreement). До прикладу при рівні SLA 99.9% для постачальника допускається наступний час недоступності послуг: щоденно: 1.44 хвилини, щотижнево: 10.08 хвилин, щомісячно: 43.8 хвилин, щорічно: 8.76 годин [5]. Відповідний рівень SLA закладений у договір між постачальником хмарних послуг та компанією-клієнтом, і він передбачає штраф для постачальника у разі недотримання рівня послуг, тобто певна сума сплачується клієнту назад, або вираховується з вартості послуг при виставленні рахунку у постоплаті. Зменшення аптайму у комерційних сервісах, розміщених у хмарі, призводить до втрат доходів компаній та організацій, оскільки кінцеві споживачі не зможуть скористатись їх онлайн-сервісами і відповідно не будуть платити за сервіс, який не працює. З іншої сторони, забезпечення високого рівня SLA у хмарах, наприклад 99.999% коштуватиме значно дорожче ніж відповідні хмарні послуги з SLA 99.9% [6]. В умовах використання гібридної хмари забезпечення високого рівня аптайму можна легко досягнути лише у публічній частині хмари завдяки SLA, у той час як у своїй власній приватній хмарі компанія-клієнт має сама дбати про рівень аптайму що не завжди досяжно і фінансово виправдано.

Іншим показником який пропонується використовувати для оцінки інформаційних систем у гібридних хмарах є час відгуку сервера (або TTFB). Він найбільш чітко відображає наявність внутрішніх вузьких місць у розгорнутій у хмарі інформаційній системі. Зокрема зростання затримки TTFB у гібридній хмарі може бути обумовлене розривом критичних залежностей між спорідненими мікросервісами, які опинились у різних частинах хмари. У такому випадку на збільшення затримки TTFB може впливати велика дистанція між приватною та публічною частинами хмари, що характеризується ростом затримки пакетів та збільшенням кількості мережових переходів. Однак ці показники далеко не завжди корелюються з часом затримки TTFB, оскільки на цей параметр можуть впливати також помилки конфігурації сервісів у хмарі, а також помилки допущені при проектуванні та розробці мікросервісів. Тому ці два показники варто використовувати при оцінці мережевої доступності інформаційних систем окремо.

Ще одним мережевим параметром який впливає на якість роботи онлайн-сервісів є затримка передачі пакетів. Її можна виміряти відносно певного хоста (наприклад серверу) за допомогою використання мережевої утиліти `ping`. Подібна затримка зазвичай є критичною при використанні онлайн-сервісів, що працюють в режимі реального часу, наприклад VoIP-телефонія, конференц-зв'язок, онлайн-радіо, відеостріми та онлайн-ігри. При збільшенні затримок пакетів до неприйнятного рівня якість цих сервісів починає падати і ці затримки відображаються у вигляді відчутних кінцевому користувачеві затримок зв'язку. Зазвичай для веб-сервісів, де не використовуються комунікація в режимі реального часу це лише призводить до збільшення TTFB та зазвичай не дуже сильно погіршує роботу веб-додатків чи веб-сайтів. Однак цей параметр може грати важливу роль при виявленні проблем зі швидкодією міжхмарних зв'язків у гібридній хмарі. Зростання цього параметру може відобразитись на швидкодії певних сервісів, доступ до яких відбувається з іншої частини гібридної хмари. Також на даний параметр у деякій мірі впливає також наявність шифрування [7].

Зовнішня затримка пакетів поза межами приватної та публічної частинами гібридної хмари, які йдуть через Інтернет, не залежить від внутрішньої конфігурації хмари, а натомість залежить від топології взаємоз'єднання магістральних Інтернет-провайдерів, точок обміну трафіку та їх пропускних здатностей. Подібна кон'юнктура їх взаємоз'єднань не є постійною і може постійно змінюватись в залежності від аварійних ситуацій, DDoS-атак [8]. В Інтернеті затримка пакетів виміряна за допомогою ICMP Ping не завжди є показником якості зв'язку,

оскільки у деяких мережах протокол ICMP може мати нижчий пріоритет QoS ніж інші протоколи. Натомість для аналізу оптимальності розподілу мікросервісів всередині хмари даний параметр дозволяє визначати наявність вузьких місць у мережевій конфігурації та уникати їх виникнення у подальшому.

Разом із першими трьома показниками оцінки мережевої доступності є також четвертий – кількість мережевих переходів. Їх можна виміряти відносно певного хоста (наприклад серверу) за допомогою використання мережевої утиліти *traceroute*. Цей показник є також корисним для аналізу оптимальності розподілу мікросервісів всередині хмари, однак як і затримка пакетів може бути корисним лише для оцінки мережі всередині хмари. Хоча існує думка, що існує чітка кореляція між затримкою передачі пакетів та кількістю хопів, але вона спостерігається по більшій мірі лише у ідеальних умовах, наприклад в навчальних мережевих стендах, в яких немає справжнього трафіку. Річ у тім, що при проходженні трафіку через мережу Інтернет, пакети проходять через гетерогенні мережеві середовища з різними пропускними здатностями, а також є і інші чинники, такі як DDoS-атаки, які можуть обмежувати цю пропускну здатність. Також сповільнення швидкості може відбуватись через високу завантаженість мережевих екранів [9]. Однак для хмар, в яких розміщуються мікросервіси існують також і інші вагомні параметри, наприклад зв'язність і когезія, на які необхідно зважати при розподілі мікросервісів у різні частини гібридної хмари. У мікросервісній архітектурі є досить важливим застосування принципу "loose coupling" (слабкої зв'язності), зокрема тому що мікросервіси розділені між собою мережею, передача даних по якій є менш продуктивною ніж при взаємодії різних модулів монолітного додатку. Модулі монолітного додатку можуть обмінюватись даними в межах одного процесу, де між ними є спільний доступ до оперативної пам'яті. У мікросервісних середовищах також існує можливість створення кешу пам'яті, яка може бути розділена між різними мікросервісами, але її швидкодія буде значно меншою. Цей і інші способи комунікації та обміну інформації між реалізують різні принципи зв'язності, серед яких виділяють наступні [10]: доменна зв'язність (англ. Domain Coupling), темпоральна зв'язність (англ. Temporal Coupling), прохідна зв'язність (англ. Pass-Through Coupling), Контентна зв'язність (англ. Content Coupling) та загальна зв'язність (англ. Common Coupling). При розробці хмарної архітектури інженери багато уваги приділяють доменній зв'язності, яка стосується ступеня взаємозалежності та взаємозв'язку між різними доменами або модулями в програмному забезпеченні. У мікросервісній архітектурі у ролі домену виступає власно мікросервіс. Якщо доменна зв'язність висока – то це означає, що переінакшення в одному домені можуть вплинути на інші домени, ускладнюючи їх незалежний розвиток та тестування. Низька зв'язність, дозволяє доменам функціонувати більш автономно, зменшуючи ризик потенційних помилок взаємоінтеграції. Ідеальною вважається ситуація, коли доменна зв'язність є мінімальною, що сприяє більшій гнучкості та легкості підтримки програмного продукту. Інший тип зв'язності, темпоральний, означає що один програмний модуль залежатиме від іншого, і якщо модуль-провайдер не працюватиме, то і залежний від нього модуль теж буде не в стані продовжувати нормальну роботу.

Для продовження дослідження необхідно зібрати різні сценарії open source систем для розгортання гібридних хмар, а також запустити хмарний стенд на основі Kubernetes. Іншими необхідними компонентами є системи стрес-тестування та збору статистики.

IV. ОБГОВОРЕННЯ ТА ВИСНОВКИ

Розглянуто різні показники оцінки мережевої доступності та зв'язності інформаційних систем у обчислювальних хмарах. Встановлено що показники оцінки мережевої доступності є метриками на основі яких можна оцінити вдалість вибору оптимального розташування мікросервісів у гібридній хмарі, однак у прийнятті рішень вони не можуть бути застосовані

у неоднакових конфігураціях хмари. Іншими показниками які необхідно враховувати є зв'язність та когезія між мікросервісами.

ДЖЕРЕЛА

1. Shi, J., Fu, K., Wang, J., Chen, Q., Zeng, D. & Guo, M. (2024) "Adaptive QoS-aware microservice deployment with excessive loads via intra-and inter-datacenter scheduling", IEEE Transactions on Parallel and Distributed Systems.
2. Newell, A., Skarlatos, D., Fan, J., Kumar, P., Khutornenko, M., Pundir, M., ... & Tang, C. (2021) "RAS: continuously optimized region-wide datacenter resource allocation", Proceedings of the ACM SIGOPS 28th Symposium on Operating Systems Principles, October, pp. 505–520.
3. Magda, D. (2022) "Run Java microservices across multiple cloud regions with Spring Cloud", DZone. Available at: <https://dzone.com/articles/run-java-microservices-across-multiple-cloud-regio> [Accessed 14 October 2024].
4. Saleem, S. (2024) "18 essential microservice best practices", Dev Skill Builder. Available at: <https://www.devskillbuilder.com/18-essential-microservice-best-practices-655fd4d20ee6> [Accessed 14 October 2024].
5. UptimeRobot. (2023) "How to calculate uptime? And 5 tips for achieving 99.999%", UptimeRobot Blog. Available at: <https://uptimerobot.com/blog/how-to-calculate-uptime/> [Accessed 14 October 2024].
6. Swain, A.K. & Garza, V.R. (2023) "Key factors in achieving Service Level Agreements (SLA) for Information Technology (IT) incident resolution", Information Systems Frontiers, vol. 25, no. 2, pp. 819–834.
7. Іванчук, Я.В., Горобець, Ю.В. та Коваль, К.О. (2022) 'Метод підвищення степеню захисту шифрування повідомлень на основі алгоритму за сталою складовою по часу', Системні технології, no. 139 (2), pp. 3–13. Dnipro. DOI: 10.34185/1562-9945-2-139-2022-01.
8. Arnold, T., He, J., Jiang, W., Calder, M., Cunha, I., Giotsas, V. & Katz-Bassett, E. (2020) "Cloud provider connectivity in the flat internet", Proceedings of the ACM Internet Measurement Conference, October, pp. 230–246.
9. Малініч, І.П., Месюра, В.І. та Арсенюк, І.Р. (2021). Аналіз використання трафіку при скануванні комп'ютерних мереж різними версіями NMAP. Вісник Вінницького політехнічного інституту, (2), pp. 92-97.



ІЛІЯ МАЛІНІЧ

асистент кафедри Комп'ютерних наук, ВНТУ, Вінниця, Україна

ORCID: 0000-0002-5862-3732

E-mail: malinich@vntu.edu.ua

член оргкомітету хакатону "Vinnytsia Linuxation" та журі конкурсу "Адміністрування Linux" (IT-Universe). Інтереси: веб-програмування, хмарні технології, стартапи.



ЯРОСЛАВ ІВАНЧУК

професор кафедри Комп'ютерних наук, ВНТУ, Вінниця, Україна

ORCID: 0000-0002-4775-6505

E-mail: ivanchuk@vntu.edu.ua

член вченої ради ВНТУ та вченої ради ФІТА. Науковий ступінь: доктор технічних наук. Галузь наукових інтересів: математичне моделювання складних систем та обчислювальні методи, стартапи.