

ПИТАННЯ ВИЗНАЧЕННЯ ДОСТУПНОСТІ СИСТЕМ НА БАЗІ ХМАРНИХ РІШЕНЬ ПІД ЧАС ЇХ РОЗРОБКИ

Вінницький національний технічний університет

Анотація

SLA або Service Level Agreement – це угода про рівень надання послуги. Одним з ключових елементів SLA є доступність послуги (Service Availability, SA) – час працездатності послуги у відсотках від часу надання послуги. Запропоновано метод визначення доступності SLA для систем побудованих на базі компонентів у хмарах AWS, GCP, Azure та інших, що враховує період активної розробки системи.

Ключові слова: Угода про рівень надання послуги, Доступність сервісу, Хмарні рішення

Abstract

SLA is Service Level Agreement. One of key elements of SLA is Service Availability (SA). Service availability is simply the measure of the service being available and accessible to the customers during the time you promised to keep the service available. It's usually calculated as a percentage. Proposed the method of service availability definition for systems with cloud components from AWS, GCP, Azure and others. Also, system active development lifecycle considered.

Keywords: Service Level Agreement, Service Availability, Cloud Solutions

Вступ

Доступність систем (Service availability) є ключовим параметром, який визначає, наскільки доступна і працездатна система в будь-який час [1, 2]. Недооцінка цього параметру може привести до побудови систем без достатнього запасу надійності, що може викликати серйозні проблеми в роботі системи і витрати на її вдосконалення. Доступність є однією з ключових нефункціональних вимог, яка повинна бути врахована при побудові архітектури системи. Її недооцінка може вплинути на ефективність роботи системи і додаткові витрати на її вдосконалення.

Метою даної роботи є методів визначення доступності SLA для систем, побудованих на базі компонентів у хмарах AWS, GCP, Azure та інших із врахуванням періоду активної розробки системи, оскільки цей параметр може суттєво вплинути на доступність системи в цілому. Результати дослідження можуть бути корисними для фахівців з інформаційних технологій, які займаються плануванням і розробкою систем в хмарному середовищі, а також для підприємств, які використовують хмарні технології для побудови своїх інформаційних систем.

Визначення доступності систем на базі хмарних рішень

Доступність систем на базі хмарних рішень визначається за допомогою різних метрик, які враховують різні аспекти функціонування системи. Однією з найбільш поширених метрик є доступність сервісу (Service availability) або "uptime" - час, протягом якого система була доступною для користувачів. У випадку з хмарними рішеннями, uptime зазвичай визначається на основі SLA (Service Level Agreement), який визначає мінімальний час доступності системи протягом певного періоду [2].

Для визначення доступності системи на базі хмарних рішень можуть використовуватися такі метрики, як MTBF (Mean Time Between Failures) - середній час між відмовами системи, MTTR (Mean Time To Repair) - середній час на відновлення системи після відмови, RTO (Recovery Time Objective) - максимальний допустимий час відновлення системи після відмови, RPO (Recovery Point Objective) - максимальна втрата даних, допустима під час відновлення системи після відмови та інші.

Оцінка доступності систем на базі хмарних рішень зазвичай здійснюється за допомогою спеціальних інструментів моніторингу та аналізу стану системи, які дозволяють вчасно виявляти проблеми та усувати їх, що забезпечує максимальну доступність системи для користувачів.

SLA виражають у відсотках, що відображає мінімальний час, протягом якого система повинна бути доступною для користувачів. Наприклад, якщо SLA становить 99%, то це означає, що система повинна бути доступною протягом не менше 99% часу від загального часу роботи. Необхідно розуміти, що це означає у реальному житті. Наприклад, якщо SLA становить 99,9%, то система може бути недоступною

протягом 8 годин 45 хвилин на рік. Якщо SLA становить 99,99%, то цей час складає 52 хвилини на рік. Тому, при визначенні SLA необхідно враховувати реальні потреби користувачів і ризики, пов'язані з можливими перервами у роботі системи. Також варто пам'ятати, що додатковий запас надійності та доступності може забезпечити більш високу якість обслуговування користувачів, але може призвести до додаткових витрат на розробку та підтримку системи.

У хмарному середовищі SLA системи визначається як композиція SLA її блоків, що поєднані послідовно або паралельно. Оскільки хмарні сервіси часто базуються на різних компонентах та сервісах, які надаються провайдером хмарних послуг, SLA кожного з цих компонентів також є важливим [3].

Для досягнення високого рівня SLA у хмарному середовищі, проєктувальники зазвичай вибирають високонадійні компоненти та сервіси, що мають відомі та гарантовані SLA. Вони також застосовують підхід з надлишковістю, що дозволяє забезпечити неперервну роботу системи в разі відмови одного з компонентів. Наприклад, якщо система складається з трьох компонентів, кожен з яких має SLA 99,9%, то загальна доступність системи буде $99,9\% * 99,9\% * 99,9\% = 99,6\%$ [3, 4]. Таким чином, при побудові системи важливо розглядати кожен компонент окремо та його внесок у загальну доступність системи. Важливо пам'ятати, що високий SLA може бути досягнутий, але це пов'язано з додатковими витратами на підтримку та розробку системи. Досвід показує, що кожна, що покращення на кілька сотих відсотка, щоб досягнути 99,99% може потребувати значних фінансових затрат і збільшити вартість системи в кілька разів.

Вплив активної розробки системи на її доступність

Під час активної розробки системи, зазвичай проводяться різноманітні зміни та додавання нового функціоналу, що може призводити до недоступності системи на певний час. Цей час включає у себе період розгортання нових змін та функціоналу, перевірки працездатності, тестування та інші процеси, пов'язані зі змінами в системі.

Таким чином, час недоступності системи може збільшуватись під час активної розробки, що потрібно враховувати при визначенні SLA. При цьому, для зменшення часу недоступності, можуть використовуватись різні методи, такі як розгортання нового функціоналу поетапно, проведення тестування в окремих середовищах та інші методи. Наприклад, як показано в табл. 1, якщо система недоступна при розгортанні нової версії 30 хвилин і нові версію впроваджуються щомісяця, то ми маємо значний вплив на SLA.

Таблиця 1 – Вплив недоступності під час розгортання на доступність системи

Базовий SLA системи за рік, %	Час недоступності за рік	SLA системи при щомісячних розгортаннях з 30хв недоступністю, %	SLA системи при щомісячних розгортаннях з 3хв недоступністю, %
99,9%	8 год. 45 хв. 36с.	99,83%	99,89%
99,95%	4 год. 22 хв. 48с.	99,88%	99,94%
99,99%	52 хв. 34 с.	99,82%	99,98%

Такий вплив можна компенсувати лише зменшенням часу недоступності під час розгортання, застосовуючи різноманітні методи розгортання: поетапне розгортання (rolling deployment), заміна системи на нову версію (blue/green deployment), часткове розгортання нової версії (canary deployment) та інших. Таким чином річна доступність системи, що перебуває в активній розробці можна визначити як

$$SA = \frac{t_{downtime.base} + \sum t_{downtime.deployment\ i} + t_{downtime.failure}}{t}, \quad (1)$$

де $t_{downtime.base}$ – базовий час недоступності системи, $t_{downtime.deployment\ i}$ – час недоступності кожного розгортання, $t_{downtime.failure}$ – інший час недоступності системи, який накопичується через відмови, t – період часу відносно якого проводимо розрахунок.

Висновки

Отже, при розрахунку SLA системи необхідно враховувати не лише базове значення доступності, отримане для компонентів системи, але і вплив, який має на систему її активна розробка та можливі відмови. Під час розробки системи можуть відбуватись зміни та додавання нового функціоналу, що може призводити до збільшення часу, коли система недоступна.

Тому, при визначенні SLA для систем, побудованих на базі хмарних рішень, необхідно враховувати композицію SLA її блоків, що поєднані послідовно або паралельно, а також проводити аналіз та оцінку ризиків, пов'язаних з активною розробкою та можливими відмовами. При цьому, важливо використовувати методи зменшення часу недоступності, такі як розгортання нового функціоналу поетапно та проведення тестування в окремих середовищах, для забезпечення високої доступності системи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Joe H. Service Availability: Calculations and Metrics, Five 9s, and Best Practices. [Електронний ресурс] – Режим доступу: <https://www.bmc.com/blogs/service-availability-calculation-metrics/>

2. Site Reliability Engineering / B. Beyer, C. Jones, J. Petoff, N. Murphy – O'Reilly Media, Incorporated, 2016. – 524 pages. – ISBN 978-149-192-912-4

3. Sastry S. How to Calculate and Design IT Service Availability. [Електронний ресурс] – Режим доступу: https://www.linkedin.com/pulse/how-calculate-design-service-availability-shashi-sastry/?trk=pulse-article_more-articles_related-content-card

4. Теорія ймовірностей та математична статистика / О. І. Кушлик-Дивульська, Н. В. Поліщук, Б. П. Орел, П. І. Штабальок – К: НТУУ «КПІ», 2014. – 212 с. – ISBN 978-966-622-654-2

Решетник Олександр Олександрович – студент групи ІПІ-22М, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: degratnik@gmail.com

Reshetnik Oleksandr O. – student of ІPI-22M group, Department of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: degratnik@gmail.com