Інформаційні технології

УДК 004.7

Ніконова Анна Вадимівна
студентка
Харківського національного університету радіоелектроніки

Лєсна Наталя Совєтівна
кандидат технічних наук,
професор кафедри програмної інженерії
Харківський національний університет радіоелектроніки

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДІВ І МОДЕЛЕЙ РЕЗЕРВНОГО КОПІЮВАННЯ ТА ПІСЛЯАВАРІЙНОГО ВІДНОВЛЕННЯ ДАНИХ У SAP HYBRIS COMMERCE

Анотація. *У роботі розглянуті сучасні методи створення бекапів із подальшим післяаварійним відновленням у популярній платформі Hybris. Особлива увага дослідження приділена проблемі пошуку оптимального балансу між необхідною частотою створення резервних копій даних та продуктивністю сервера в контексті електронної комерції.*

Ключові слова: *електронна комерція, післяаварійне відновлення як послуга, ефективність, DRaaS, RPO, RTO, Hybris, Java.*

Проблема і актуальність дослідження. 2020 рік став періодом винятково швидкого зростання обсягів сховищ даних - як активних, так і резервних, що використовуються на вимогу у випадку пікових навантажень на сервері [1]. Це глобальне явище пов'язане переважно із потребами бізнесу, обумовленими розвитком електронної комерції. Водночас, 71% американських IT-компаній визнають, що аспектом їхньої поточної інфраструктури зберігання, який у першу чергу потребує

вдосконалення чи розширення, є саме здатність до післяаварійного відновлення [2].

Із корпоративної точки зору, втрата або порушення цілісності інформації може коштувати сотні тисяч доларів у грошовому, часовому й навіть репутаційному еквівалентах [3].

Огляд поточного стану об'єкта дослідження. Найважливішим критерієм для бізнесу є надійний стек технологій обраної платформи. Найпопулярнішим програмним рішенням для електронної комерції є SAP Hybris Commerce, на якій реалізовані веб-магазини таких великих корпорацій, як Porsche, Volkswagen, Nikon, Philips, Indesit, Deutsche Post, Paco Rabbane та інші.

Hybris надає багато готових і потужних рішень для управління продуктами, автентифікації і контролю доступу для користувачів та груп, процесу створення замовлень від додавання продукту в кошик до оплати й доставки. Платформа дозволяє створення користувацької бізнес-логіки. Існує також багато додаткових модулів для інтеграції із хмарними провайдерами (Amazon S3, Microsoft Azure), медіа-ресурсами (CloudFlare), сервісами оплати (PayPal, LiqPay) тощо. Також Hybris забезпечує швидке перемикання між різними типами СУБД (MySQL, Oracle SQL, SAP HANA, Postgre SQL, HSQL) без необхідності зміни коду.

У той же час на даний момент у платформі не існує єдиного стандартизованого методу створення резервних копій та подальшого відновлення даних із бекапів. Така задача є нетривіальною, і майже на кожному новому проєкті розробники витрачають понад 40-80 годин на написання логіки користувацьких сервісів для експорту та імпорту даних. Окрім того, поки що не існує програмних засобів для оцінки ефективності тих чи інших методів створення резервних копій у середовищі Hybris. Тому процес планування післяаварійного відновлення є досить проблематичним і неструктурованим.

Пошук оптимальної технології для експорту та імпорту даних із метою резервного копіювання є об'єктом дослідження IT-компаній, які співпрацюють із SAP Hybris Commerce. Так, українська компанія Aimprosoft створила модуль Commerce Cloud Backup and Migration (Commerce CBM), який автоматизує створення бекапів і їх використання на рівні сайтів чи версій каталогів [4].

Німецька компанія SAP, яка є власником сімейства продуктів Hybris, працює над покращенням показників технологій імпорту, представлених "із коробки". До них належать методи із шару сервісів та традиційний менеджер імпорту-експорту, представлений на рівні Jalo. Існують дослідження білоруської компанії EPAM, присвячені підвищенню швидкості імпорту даних за рахунок відмови від стандартного протоколу ImpEx та використання альтернативної технології, яка однак потребує додаткової Java-конфігурації і не є гнучкою.

Як бачимо, всі існуючі методи резервного копіювання в Hybris потребують значних ресурсів сервера. Водночас цінність втрачених даних може бути дуже високою. Саме тому необхідним є пошук оптимального баласу між частотою створення бекапів, яка визначається об'єктивною точкою відновлення RPO, та швидкодією системи, яка описується комбінацією різних метрик (рівень завантаженості CPU, кількість звернень до бази даних тощо). Очікуваним результатом є така комбінація, при якій досягається найнижчий рівень об'єктивного часу RTO, тобто проміжку часу від аварії до повного відновлення стану системи.

Мета дослідження. Дана робота містить дослідження існуючих методів резервного копіювання та післяаврійного відновлення даних у Hybris. Метою дослідження є виявлення оптимального балансу між показниклм RTO та швидкодією комп'ютера, а також формування рекомендацій щодо розроблення плану післяаварійного відновлення для SAP Hybris Commerce шляхом створення тестового середовища, імітації

порушення цілісності даних та заміру показників ефективності використаних методів при різних конфігураціях.

Дослідження методів резервного копіювання та післяаварійного відновлення з використанням технології імпорту-експорту ImpEx та розширення Commerce CBM здійснюється на основі реалізації періодичних задач генерації тестових даних замовлень в ізольованій версії каталогу. Генерація відбувається із нормальним розподілом зі змінними значеннями RPO, обсягу замовлень на сайті за годину, можливістю синхронного й асинхронного виконання. Прототипом аварії на сервері є моделювання порушення цілісності даних - повного або часткового - в тестовому середовищі. Заміри ефективності роботи сервісів та утиліт виконані за допомогою аспектно-орієнтованого програмування, яке підтримується в Hybris засобами AspectJ.

Для обчисленя відсотку втрати даних при післяаварійному імпорті реалізовані сервіс та API з функцією експорту тестових даних. HTTP-запит для експорту здійснюється до та після "аварії". Отримана відповідь у форматі JSON десеріалізується засобами JAXB, таким чином порівнюється стан об'єктів. Також засоби Java дозволяють оцінити відсоток завантаженості CPU при використанні того чи іншого методу.

Для серії експериментів використані задачі періодичної генерації сутності OrderModel - замовлень на сайті - зі значеннями RTO: 1, 2, 4, 8, 12, 24 години, та фіксованим щогодинним коефіцієнтом нових значень у базі даних. За час RTO прийнятий час, необхідний для здійснення імпорту із бекап-носія, створеного даною альтернативою під час останнього експорту.

Для отримання достовірних результатів для кожної комбінації RPO та альтернативи, що аналізується, виконано 10 запусків задачі. Найпріорітетнішим завданням є визначення залежності об'єктивного часу

відновлення RTO від об'єктивної точки відновлення RPO. Усереднені значення серії експериментів показані на рисунку 1.

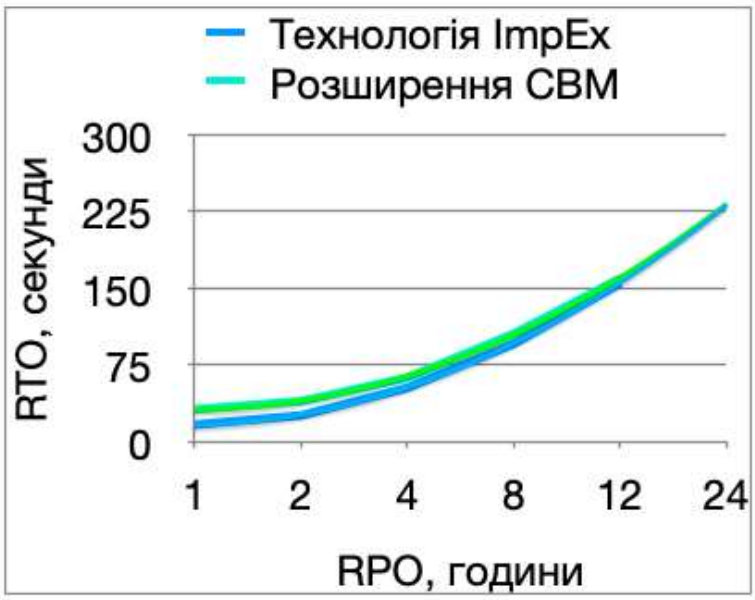

Рис. 1. Залежність RTO від RPO

Також важливим є встановлення залежності рівня навантаження на процесор від обраної RPO. Результати серії експериментів показані на рисунку 2.

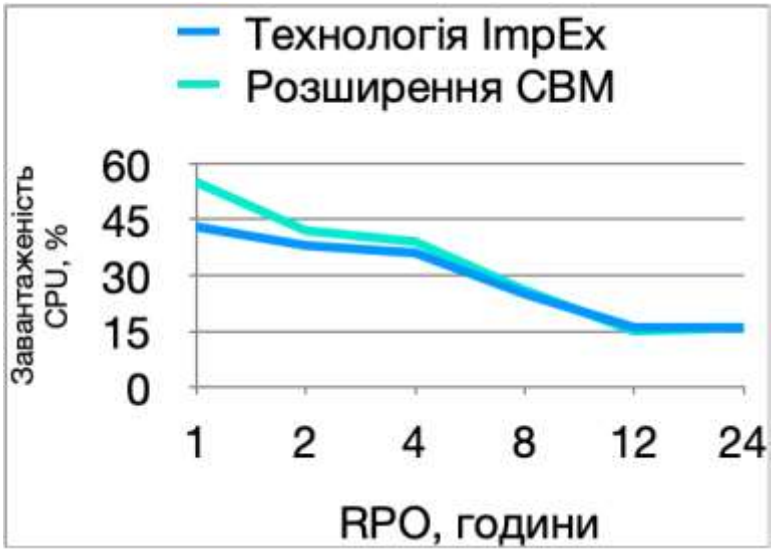

Рис. 2. Залежність завантаженості CPU від RPO

Для визначення оптимальної точки між RTO, збільшення якої еквівалентне збільшенню вартості трати даних, та RPO, ріст значень якої означає зниження вартості бекапів (тобто навантаження на сервер), слід визначити точку перетину графіків залежності RTO від RPO та RTO від навантаження на систему, яке у контексті даної моделі формується як відсоток завантаженості процесора. Така оптимальна точка позначена на рисунку 3 як A.

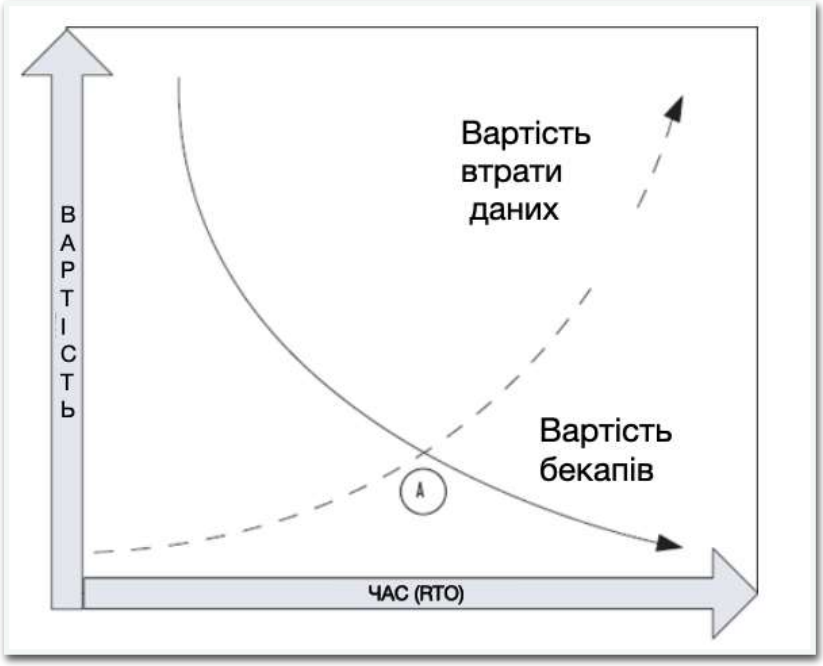

Рис. 3. Оптимальна точка балансу між RTO та RPO

Після нормалізації отриманих графіків залежностей було виявлено, що оптимальна точка балансу знаходиться приблизно на рівні 8-годинного RPO. Тобто із заданими коефіцієнтами щогодинної кількості нових даних (замовлень) на сервері та технічними характеристиками системи найкращою (із точки зору балансу між безпекою даних та швидкодією сервера) опцією є створення бекапів тричі на добу, тобто раз на 8 годин.

Висновки. У ході дослідження була розглянута проблема визначення ефективності методів резервного копіювання та післяаварійного відновлення даних на сервері SAP Hybris Commerce.

Створена програмна система для генерації записів у тестовому середовищі, запуску задач експорту, імітації порушення цілісності даних, а також імпорту. Реалізовано визначення швидкодії системи (рівня завантаженості CPU), часу роботи методів імпорту (післяаварійного відновлення з бекапносія).

Завдяки серії експериментів визначено рівень залежності RTO від RPO, а також відстоток завантаженоссті CPU від RPO для OOTB-технології ImpEx та розширення Aimprosoft Commerce CBM. Встановлено, що для низьких (до 4 годин) значень RPO бажаним є використання OOTB-методів. Водночас на великих RPO оптимальність розширення CBM практично наближається до значень "коробочних" технологій. Враховуючи факт автоматичної генерації скриптів та можливості прямого запису резервних копій у хмару, CBM Commerce рекомендується для використання у системах з RPO, більшим за 4 години.

Також експериментально виявлено точку оптимального балансу між вартістю втрати даних та вартістю створення резервних копій. При заданих характеристиках системи такою точкою є RPO, близьке до 8 годин, тобто здійснення процесу бекапу тричі на добу.

Література

1. Gai S. Building a Future-Proof Cloud Infrastructure: A Unified Architecture for Network, Security, and Storage Services. Boston, MA: Addison-Wesley, 2020. 118 p.
2. Harvard Business Review Press. HBR Guide to Data Analytics Basics for Managers. Boston, MA: Harvard Business Review Press, 2018. 256 p.
3. Stevens B. DRaaS basics: Veeam special edition. Baar: Veeam Press, 2015. 110 p.
4. Machnev A. Functional Specification Document: Hybris Commerce CBM v.1.0. Kharkiv: Aimprosoft, 2020. 4 p.