

ДОСЛІДЖЕННЯ ВПЛИВУ КОМПРЕСІЇ ДАНИХ НА ПРОДУКТИВНІСТЬ СИСТЕМИ ЗБЕРЕЖЕННЯ ДАНИХ

Поява алгоритмів компресії зумовлена постійним ростом кількості даних у світі. Щодня генеруються петабайти структурованих даних і одна із першочергових задач бізнес користувача полягає у здатності до їх зберігання

В якості апаратного рішення, яке здатне керувати такою кількістю даних використовуються системи зберігання (далі СЗД). На відміну від сервера, першочерговою задачею якого є обчислення і обробка даних, СЗД використовуються для їх зберігання, вони здатні вмещувати сотні дисків, які можна віртуалізувати за допомогою вбудованого програмного забезпечення.

Однією із важливих характеристик СЗД є можливість компресії даних, оскільки вона дозволяє помістити більшу кількість інформації на накопичувач. Тим самим бізнес користувач може заощадити свої кошти.

Існує декілька видів компресії:

- апаратна компресія;
- програмна компресія.

Як правило використання програмної компресії приводить до значного зниження продуктивності системи, це пов'язано з тим, що СЗД виділяє частину своїх обчислювальних ресурсів на роботу з алгоритмами скорочення даних.

Існують системи зберігання в яких компресія реалізована, як додатковий апаратний прискорювач. В такому випадку зниження продуктивності СЗД буде не значним, але варто відзначити, що такий функціонал, як правило необхідно оплачувати додатково.

У цій роботі було проведено дослідження впливу компресії даних на продуктивність системи збереження даних. На рис.1 наведені графіки залежності продуктивності вводу виводу (далі IOPS) від часу обслуговування однієї команди для SSD накопичувачів з використанням апаратної та програмної компресії

даних. Дана інформація була отримана внаслідок моделювання роботи СЗД IBM Storwize v5030 та Storwize 5100

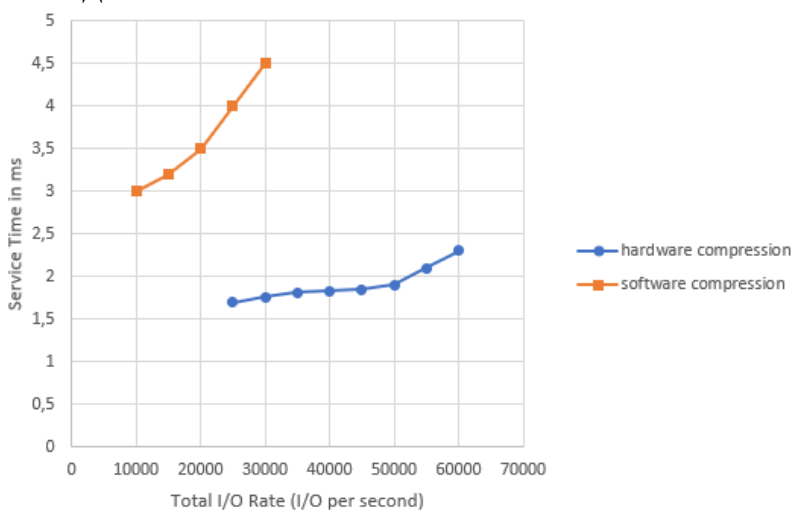


Рис.1. Графік залежності кількості IOPS від часу обслуговування для SSD накопичувачів з використанням апаратної та програмної компресії даних

Порівнюючи графіки між собою, можна підсумувати, що використовувати програмну компресію варто лише для задач, які не потребують великої продуктивності IOPS, адже час обслуговування одного запиту на 10000 операцій вводу виводу займає близько трьох мілісекунд, водночас для бази даних Oracle розробники рекомендують, щоб час обслуговування був менше однієї мілісекунди. Апаратна компресія є доцільним рішенням навіть при роботі з базами даних. Однак необхідно розраховувати на системи вищого рівня, наприклад IBM FlashSystem 7200. Дана система має два 8 ядерних процесори на контролер, це означає, що пікове завантаження процесора при включеній компресії відбудеться приблизно на позначці в 200 тисяч IOPS.

ВИКОРИСТАНІ ДЖЕРЕЛА

1. *IOPS [Електронний ресурс] – Режим доступу до ресурсу: <https://ru.wikipedia.org/wiki/IOPS>.*