Оскільки Redis досить швидко виконує операції читання та запису, я намагався реалізувати кешування для зберігання необхідної інформації, яку повертаю. Коли користувач запитує скорочення посилання, спочатку відбувається пошук за цим посиланням у Redis, а вже тоді, коли там нічого не знайшли, перевіряємо у БД. Якщо знаходимо, то записуємо на певний проміжок часу у Redis, щоб при наступному запиті змогли швидше повернути відповідь. Те ж саме реалізовано і для переходу за скороченим посиланням. Також, можна збільшити проміжок часу, який буде зберігатись посилання у Redis (зараз встановлено 4 години для зберігання повного url по токену для redirect та 30 хвилин на випадок, якщо користувач швидко втратить посилання і знову введе ту саму адресу, щоб повернути йому скорочене посилання, що було створено раніше). Тим самим, ми зменшуємо навантаження на БД і збільшуємо швидкість відповіді.

Також можна індексувати певні поля бази даних, щоб пришвидшити пошук по ній.

Також, можна згадати про горизонтальне масштабування. Підняти декілька серверів і балансувальник навантаження, тим самим розподіляючи рівномірно навантаження. Якщо я правильно пам’ятаю, AWS дозволяє автоматично збільшувати кількість серверів при перевищенні певного відсотка їх завантаженості і так само автоматично зменшувати їх кількість, коли навантаження падає нижче певного рівня.