

УДК 004.65

RESEARCH OF REPLICATION LAGS IN NON-RELATIONAL DATABASES

ДОСЛІДЖЕННЯ ЗАТРИМОК РЕПЛІКАЦІЇ У НЕРЕЛЯЦІЙНИХ БАЗ ДАНИХ

Arkhyrova V.V. / Архипова В.В.*s.t.s., as.prof. / к.т.н., доц.*

ORCID: 0000-0002-9722-5622

*Dnipro State Agrarian and Economic University, Dnipro, Efremova, 25, 49009**Дніпровський державний аграрно-економічний університет, Дніпро, Єфремова, 25, 49009**Ukrainian State University of Chemical Technology, Dnipro, Lazariana, 2, 49010**Український державний університет науки і технологій, Дніпро, Лазаряна, 2, 49010***Reznichenko O.V. / Резніченко О.В.**

ORCID: 0009-0001-0194-5079

*Ukrainian State University of Chemical Technology, Dnipro, Lazariana, 2, 49010**Український державний університет науки і технологій, Дніпро, Лазаряна, 2, 49010***Liashenko O.A. / Ляшенко О.А.***s.t.s., as.prof. / к.т.н., доц.*

ORCID: 0000-0002-99838-5504

*Ukrainian State University of Chemical Technology, Dnipro, Lazariana, 2, 49010**Український державний університет науки і технологій, Дніпро, Лазаряна, 2, 49010*

Анотація. В роботі досліджено та проаналізовано затримки реплікації в нереляційній базі даних MongoDB. За допомогою Python-коду зібрані дані, визначено, що розподіл затримок є асиметричним, з мультимодальною структурою. Визначені локальні максимуми, що відповідають затримкам в 5 та 10 секунд, та найбільші величини затримок, що можуть впливати на ефективність роботи системи.

Ключові слова: MongoDB, кластер, аналіз.

Abstract. The paper investigates and analyzes replication lags in the non-relational MongoDB database. Using Python code, data was collected and it was determined that the distribution of lags is asymmetric, with a multimodal structure. Local maximums corresponding to latencies of 5 and 10 seconds were identified, as well as the largest latency values that can affect system performance.

Key words: MongoDB, cluster, analysis.

Вступ.

Нереляційні бази даних (NoSQL) отримали значне поширення у теперішній час [1]. Це пов'язано, по-перше, з їх здатністю легко масштабуватися, підлаштовуючись під сучасні величезні об'єми інформації, по-друге, завдяки можливості збереження інформації різних типів і відсутності жорсткої структури [2].

Для забезпечення надійності збереження інформації, вона повинна знаходитися на декількох серверах. Тоді дані спочатку записується на первинну (primary) репліку, а потім на вторинну (secondary), і оскільки це відбувається не

миттєво, можливе виникнення розбіжностей між їх вмістом. Нереляційні бази даних підтримують декілька видів узгодженості, найбільш відомі з яких є наступні [3]:

- сильна узгодженість (strong consistency), для якої операція запису інформації вважається завершеною тільки при синхронізації всіх реплік;
- підсумкова узгодженість, або слабка (eventual consistency, weak consistency), при якій через деякий час інформація на всіх репліках буде однаковою.

Також існують інші проміжні види, які дозволяють обирати поведінку системи.

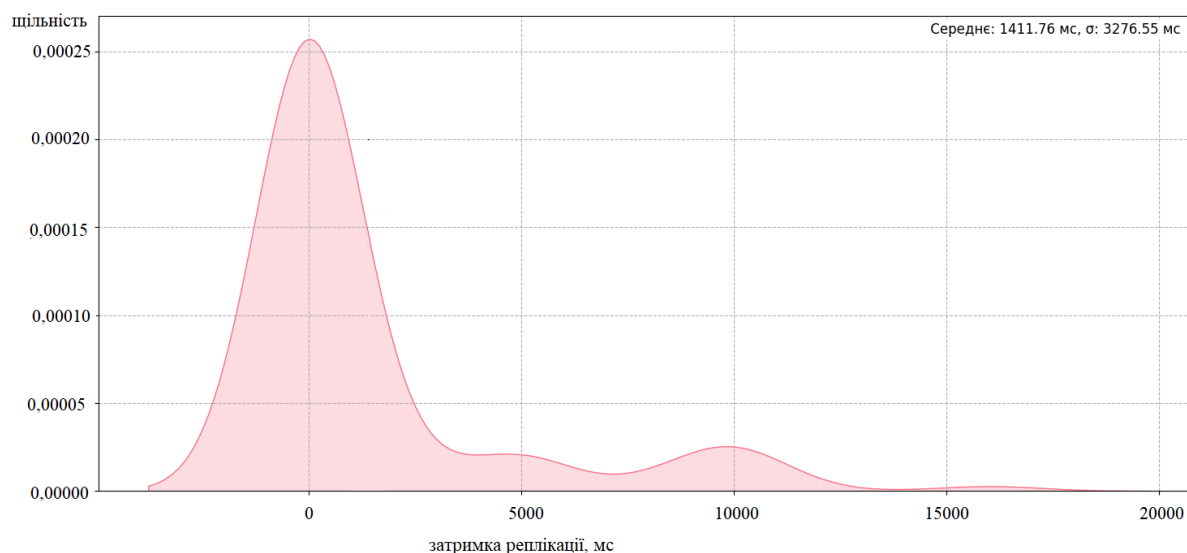
Показником ефективності роботи баз даних є затримка реплікації, що характеризує відмінності між первинними та вторинними репліками, тому аналіз цього параметра є важливим для оцінки роботи баз даних. Таким чином, метою дослідження було визначення розміру та характеру затримок реплікації нереляційної бази даних.

Основний текст.

В даній роботі було досліджено роботу бази даних MongoDB, яка була змодельована на комп'ютері. Було створено один кластер, що містив одну первинну та дві вторинні репліки. За допомогою Python-коду зібрано дані щодо затримок реплікації.

На основі отриманої інформації побудовано діаграму розподілу затримок реплікації (рис. 1).

Аналіз отриманої інформації свідчить про те, що розподіл є асиметричним, тобто не є нормальним, і може бути описано за допомогою суміші розподілів. Більша частина затримок (0,7019) розташована близько нуля, що свідчить про високу ефективність реплікації в досліджуваній системі. Але, незважаючи на це, розподіл містить декілька локальних максимумів, які відповідають значенням затримок 5000 мс (5 с) та 10000 мс (10 с). І, хоча їх кількість зовсім незначна, наявність затримок більше 15000 мс, свідчить про те, що в деякий момент часу система відчуває значного зниження продуктивності роботи.



Авторська розробка

Рисунок 1 – Розподіл затримок реплікації

Висновки. У роботі проаналізовано розподіл затримок реплікації у нереляційних базах даних на прикладі MongoDB, визначено форму розподілу та розташування локальних максимумів.

Література:

1. Nirnaya T. (2025). NoSQL database education: A review of models, tools and teaching methods. Journal of Systems and Software. 226. 112391. <https://doi.org/10.1016/j.jss.2025.112391>.
2. Bansari H. K., Hetal J. (2017). A Survey of Non -Relational Databases with Big Data. International Journal on Recent and Innovation Trends in Computing and Communication. 5. 11. 143-148. https://www.academia.edu/36852552/A_Survey_of_Non_Relational_Databases_with_Big_Data
3. Diogo M., Cabral B., & Bernardino J. (2019). Consistency Models of NoSQL Databases. Future Internet. 11. 43. DOI: 10.3390/fi11020043.