

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Штучного інтелекту  
(повна назва)

**КВАЛІФІКАЦІЙНА РОБОТА**  
**Пояснювальна записка**

рівень вищої освіти другий (магістерський)

Дослідження та розробка методів оптимізації збереження та отримання  
доступу до великих даних для NoSQL сховищ  
(тема)

Виконав:  
студент 2 курсу, групи ДСМ-21-1  
Рогачов І.Д.  
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки  
(код і повна назва спеціальності)

Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)

Освітня програма Науки про дані  
(повна назва спеціалізації)

Керівник проф. Філатов В.О.  
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри \_\_\_\_\_  
(підпис)

В.О. Філатов  
(прізвище, ініціали)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)  
Кафедра Штучного інтелекту  
(повна назва)  
Рівень вищої освіти другий (магістерський)  
Спеціальність 122 Комп'ютерні науки  
(код і повна назва)  
Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)  
Освітня програма Науки про дані  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

«\_\_\_\_\_» \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Рогачову Ігорю Дмитровичу  
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження та розробка методів оптимізації збереження та отримання доступу до великих даних для NoSQL сховищ

затверджена наказом університету від 14 листопада 2022 р. № 1491Ст

2. Термін подання студентом роботи до екзаменаційної комісії 14 грудня 2022 р.

3. Вихідні дані до роботи Науково-технічні публікації, інформація з Інтернет-джерел та відомих наукових проектів щодо розробки та дослідження методів оптимізації збереження та отримання доступу до великих даних для NoSQL сховищ

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_

1) Аналіз предметної області

2) Постановка задачі

3) Опис принципів роботи NoSQL сховищ

4) Розробка та аналіз методів оптимізації

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) \_\_\_\_\_

---



---



---



---



---



---



---



---



---



---

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1 )

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання	21.11.2022	виконано
2	Аналіз завдання та літератури з теми роботи	22.11.2022	виконано
3	Вибір засобів для розробки технічних вимог до програми	23.11.2022	виконано
4	Структурне проектування	25.11.2022	виконано
5	Вибір середовища розробки програми	27.11.2022	виконано
6	Розробка програми	28.11.2022	виконано
7	Тестування програми	01.12.2022	виконано
8	Оформлення пояснювальної записки	10.12.2022	виконано
9	Оформлення презентації	11.12.2022	виконано
10	Попередній захист	12.12.2022	виконано
11	Захист перед ЕК	14.12.2022	виконано

Дата видачі завдання 21 листопада 20 22 р.



Студент \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_  
(підпис)

проф. Філатов В.О.  
(посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка: 50 с., 16 рис., 1 дод., 8 джерел.

ВЕЛИКІ ДАНІ, ЗАПИТИ, ІНДЕКСИ, ОПТИМІЗАЦІЯ, СХОВИЩА, JSON, NOSQL.

Об'єкт дослідження – збереження та отримання доступу до великих даних.

Предмет дослідження – використання конфігурацій NoSQL сховищ для збереження та отримання великих даних.

Мета роботи – пошук найбільш оптимальних за результативністю та продуктивністю конфігурацій NoSQL сховищ для оптимізації збереження та отримання великих даних за ціною та швидкістю.

Методи дослідження – аналіз технічної літератури з області NoSQL сховищ, експериментальний підбір конфігурацій, порівняльний аналіз експериментів.

Проведено теоретичний аналіз різних методів збереження та отримання великих даних, конфігурацій сховищ, методів оптимізації та інших параметрів. Проведення практичних дослідів та підбір оптимальних значень параметрів та опис результатів кожного дослідів. Практичні дослідів проводилися на різних вибірках даних за розміром та складністю структури. На основі отриманих результатів розроблено методи оптимізації, що дозволяють зменшити ціну збереження даних та підвищити швидкість отримання доступу до них.

## **ABSTRACT**

Explanatory note: 50 p., 16 fig., 1 ann., 8 sources.

**BIG DATA, JSON, INDEXES, NOSQL, OPTIMIZATION, QUERIES, STORAGE.**

The object of the research is to save and gain access to big data.

The subject of research is the use of NoSQL storage configurations for storing and retrieving big data.

The goal of the work is to find the most effective and efficient NoSQL storage configurations to optimize the storage and retrieval of large data at a price and speed.

Research methods – analysis of technical literature in the field of NoSQL storage, experimental selection of configurations, comparative analysis of experiments.

A theoretical analysis of various methods of storing and retrieving big data, storage configurations, optimization methods and other parameters was carried out. Conducting practical experiments and selecting optimal parameter values and describing the results of each experiment. Practical experiments were conducted on different samples of data according to the size and complexity of the structure. On the basis of the obtained results, optimization methods have been developed that allow to reduce the price of data storage and increase the speed of obtaining access to them.

## ЗМІСТ

Вступ.....	7
1 Аналіз предметної галузі .....	9
1.1 Опис предметної галузі .....	9
1.2 Огляд методів та технік аналізу, які застосовуються до великих даних .....	11
2 Постановка задач.....	13
2.1 Класифікація обраної задачі .....	13
2.2 Визначення вимог до результатів.....	13
2.3 Обрання вибірки даних.....	16
2.4 Програмні ресурси .....	19
3 Опис принципів роботи NoSQL сховищ.....	23
3.1 Загальний опис роботи сховища.....	23
3.2 Опис принципів роботи із конфігураціями .....	26
3.3 Опис принципів роботи збереження та розташування даних .....	27
3.4 Опис принципів роботи виконання та обробки запитів.....	28
4 Розробка та аналіз методів оптимізації.....	31
4.1 Розробка методів оптимізації конфігурацій сховища .....	31
4.2 Розробка методів оптимізації збереження та розташування даних ...	35
4.3 Розробка методів оптимізації запитів .....	39
Висновки .....	48
Перелік джерел посилання .....	49
Додаток А Відомість кваліфікаційної роботи магістра.....	50

## ВСТУП

Починаючи з 90-х років минулого століття зі збільшенням потужностей процесорів та збільшенням кількості пам'яті на засобах збереження людство почало створювати інформаційні системи для автоматизації та підвищення ефективності бізнес процесів в компаніях різних напрямів. Саме в цей час відбувся розквіт SQL баз даних, правил створення таблиць, об'єднання їх у структури та правил отримання доступу до цих даних. Із плином часу, в 00-х роках цього століття, класичні методи збереження даних почали поступово втрачати свою актуальність через розповсюдження інтернету та інтернет сервісів, які потребували більшої гнучкості до зміни схеми та зменшення вартості збереження. Саме в 10-х роках цього століття почали з'являтися та активно розповсюджуватися NoSQL сховища, та разом з ними постала нова проблема – великі дані.

Щоб зрозуміти проблеми великих даних можна дослідити розвиток інформаційних систем представників медичного сектору, як одних з головних користувачів великих даних в наш час. Із бумом комп'ютерного сектору саме вони стали одними із перших користувачів та першопроходців які вирішували найбільш актуальні проблеми свого часу стосовно збереження даних. В 90-х та 00-х роках найбільш поширеним методом збереження даних своїх клієнтів був виділення окремого ізольованого оточення із власним сховищем для кожного користувача, що полегшувало співпрацю із різними лікарнями та страховими компаніями, але з плином часу даних ставало все більше та більше і починаючи з 10-х років розпочалася міграція до NoSQL сховищ. Завдяки цьому вдалося зменшити витрати на збереження даних клієнтів, оскільки NoSQL дозволяє легко зберігати дані всіх клієнтів відразу в одному місці та відмовитися від обслуговування інфраструктури кожного клієнта індивідуально. Через те, що всі дані зберігаються в одному місці постає питання – як зробити збереження дешевше та пришвидшити отримання даних.

Це питання постає через те, що хоча і вдалося зменшити витрати на кожного клієнта, але більшість даних які зберігаються не використовуються більшість часу, але заважають отриманню корисних даних, збільшуючи час запитів та збільшуючи вартість на виконання запитів. Саме через це класичні методи отримання даних мають низьку ефективність для великих даних і постає питання оптимізації цього процесу.

У цій роботі буде розглянуто методи оптимізації ціни збереження та швидкості отримання даних для великих даних, що дозволить збільшити ефективність роботи з великими даними будь яких категорій – від даних пацієнтів медичних компаній до даних користувачів соціальних мереж.

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

### 1.1 Опис предметної галузі

Великі дані – позначення структурованих та неструктурованих даних величезних об'ємів та значного різноманіття, що ефективно обробляються горизонтально масштабованими програмними інструментами, що з'явилися наприкінці 2000-х років та альтернативних традиційним системам управління базами даних та рішенням класу Business Intelligence.

У широкому сенсі про «великі дані» говорять як про соціально-економічний феномен, пов'язаний з появою технологічних можливостей аналізувати величезні масиви даних, у деяких проблемних галузях – весь світовий обсяг даних, і з цього трансформаційних наслідків.

Як визначальних характеристик великих даних традиційно виділяють «три V»: обсяг (англ. volume, у сенсі величини фізичного обсягу), швидкість (velocity у сенсі як швидкості приросту, і необхідності високошвидкісної обробки та отримання результатів), різноманіття (variety, у сенсі можливості одночасної обробки різних типів структурованих та напівструктурованих даних); надалі виникли різні варіації та інтерпретації цієї ознаки.

З погляду інформаційних технологій, у сукупність підходів та інструментів спочатку включалися засоби масово-паралельної обробки невизначено структурованих даних, насамперед, системами управління базами даних категорії NoSQL, алгоритмами MapReduce та програмними каркасами та бібліотеками проекту Hadoop, що їх реалізують. Надалі до серії технологій великих даних стали відносити різноманітні інформаційно-технологічні рішення, у тому чи іншою мірою що забезпечують подібні за характеристиками можливості обробки надвеликих масивів даних.

Широке запровадження терміну «великі дані» пов'язують із Кліффордом Лінчем, редактором журналу Nature, який підготував до 3

вересня 2008 року спеціальний випуск із темою «Як можуть вплинути на майбутнє науки технології, що відкривають можливості роботи з великими обсягами даних?», у якому було зібрано матеріали про феномен вибухового зростання обсягів та різноманіття оброблюваних даних та технологічні перспективи в парадигмі ймовірного стрибка «від кількості до якості»; термін був запропонований за аналогією з розхожими в діловому англійському середовищі метафорами «велика нафта», «велика руда».

Незважаючи на те, що термін вводився в академічному середовищі і насамперед розбиралася проблема зростання та різноманіття наукових даних, починаючи з 2009 року термін широко поширився в діловій пресі, а до 2010 року відносять появу перших продуктів і рішень, що стосуються винятково і безпосередньо до проблеми обробки великих даних. До 2011 року більшість найбільших постачальників інформаційних технологій для організацій у своїх ділових стратегіях використовують поняття великих даних, зокрема IBM, Oracle, Microsoft, Hewlett-Packard, EMC, а основні аналітики ринку інформаційних технологій присвячують концепції виділені дослідження.

У 2011 році Gartner відзначив великі дані як тренд номер два в інформаційно-технологічній інфраструктурі (після віртуалізації і як суттєвіший, ніж енергозбереження та моніторинг). В той же час прогнозувалося, що впровадження технологій великих даних найбільше впливатиме на інформаційні технології у виробництві, охороні здоров'я, торгівлі, державному управлінні, а також у сферах та галузях, де реєструються індивідуальні переміщення ресурсів.

З 2013 року великі дані як академічний предмет вивчаються в вузівських програмах з науки про дані та обчислювальні науки і інженерії, що з'явилися.

У 2015 році Gartner виключив великі дані з циклу зрілості нових технологій і припинив випускати окремий цикл зрілості технологій великих даних, що виходив у 2011–2014 роках, мотивувавши це переходом від етапу

галасу до практичного застосування. Технології, що фігурували у виділеному циклі зрілості, здебільшого перейшли у спеціальні цикли з просунутої аналітики та науки про дані, з ВІ та аналізу даних, корпоративного управління інформацією, резидентних обчислень, інформаційної інфраструктури.

## 1.2 Огляд методів та технік аналізу, які застосовуються до великих даних

Методи та техніки аналізу, які застосовуються до великих даних, виділені у звіті McKinsey:

– методи класу Data Mining: навчання асоціативним правилам (англ. association rule learning), класифікація (методи категоризації нових даних на основі принципів, раніше застосованих до даних, що вже є), кластерний аналіз, регресійний аналіз;

– краудсорсинг – категоризація та збагачення даних силами широкого, невизначеного кола осіб, залучених на підставі публічної оферти, без вступу до трудових відносин;

– змішання та інтеграція даних (англ. data fusion and integration) – набір технік, що дозволяють інтегрувати різноманітні дані з різноманітних джерел для можливості глибинного аналізу, як приклади таких технік, що становлять цей клас методів, наводяться цифрова обробка сигналів і обробка природної мови (включаючи тональний аналіз);

– машинне навчання, включаючи навчання з учителем і без вчителя, а також Ensemble learning (англ.) – використання моделей, побудованих на базі статистичного аналізу або машинного навчання для отримання комплексних прогнозів на основі базових моделей (англ. constituent models, CP зі статистичним ансамблем у статистичній механіці);

– штучні нейронні мережі, мережевий аналіз, оптимізація, зокрема генетичні алгоритми;

- розпізнавання образів;
- прогнозна аналітика;
- імітаційне моделювання;
- просторовий аналіз (англ. Spatial analysis) – клас методів, що використовують топологічну, геометричну та географічну інформацію в даних;
- статистичний аналіз, як приклади методів наводяться A/B-тестування та аналіз часових рядів;
- візуалізація аналітичних даних – подання інформації у вигляді малюнків, діаграм, з використанням інтерактивних можливостей та анімації як для отримання результатів, так і для використання як вихідні дані для подальшого аналізу.

## 2 ПОСТАНОВКА ЗАДАЧ

### 2.1 Класифікація обраної задачі

Задача оптимізації збереження та отримання великих даних розподіляється на декілька низькорівневих задач:

- визначити необхідні конфігурації сховища, сервера та оточення для оптимізації використання фізичних ресурсів та зменшення часу простою процесору та пам'яті;
- визначити оптимальну структуру даних, відношення даних, посилення та індекси для зменшення об'єму зайвих даних та дешевшого фізичного розташування даних на диску;
- визначити найбільш ефективні конструкції запитів, програмного коду сховища для зменшення часу виконання запитів та використовуваної оперативної пам'яті.

### 2.2 Визначення вимог до результатів

Дана робота ставить на меті отримати методи оптимізації збереження та отримання доступу до великих даних для NoSQL сховищ.

Оптимізація може відбуватися за будь якими параметрами – від оптимізації швидкості запитів, до оптимізації завантаження ресурсів центрального процесору. Серед найбільш затребуваних видів оптимізації можна виділити три типи:

- оптимізація за вартістю. Тобто зменшення витрат на використання, обслуговування та підтримку робочого стану сховища з великими даними;
- оптимізація за швидкістю запису. Тобто оптимізація швидкості додавання нових даних до сховища для збільшення пропускну здатності системи загалом;

– оптимізація за швидкістю зчитування. Тобто оптимізація швидкості отримання даних для збільшення об'ємів даних, які можна отримати.

Оптимізація вартості зазвичай є найважливішим показником, адже великі об'єми даних та велика кількість навантаження на сервери потребує великої кількості ресурсів на обробку та збереження даних, тому оптимізація вартості надає значний результат відразу та його легко порахувати в грошовому еквіваленті. Через велику кількість навантаження та витрати в десятки, а інколи і в сотні тисяч доларів на день оптимізація навіть на кілька відсотків зменшує витрати на інфраструктуру на сотні тисяч та мільйони доларів на місяць.

Оптимізація швидкості запису нових даних до сховища зазвичай є найбільш важливим показником для систем безперервної обробки даних, адже це дозволяє підвищити швидкість всієї системи та обробляти більші потоки даних за теж самий час. Також ця оптимізація надає багато переваг в системах побудованих на основі подій, а саме надавати користувачам оброблені дані швидше.

Оптимізація швидкості зчитування є корисною майже в усіх випадках, адже зазвичай запитів на отримання даних більше, ніж запитів на додавання, також запити на отримання даних зазвичай потребують більше ресурсів та створюють більше навантаження, тому оптимізація зчитування часто йде поряд з оптимізацією витрат. Завдяки цій оптимізації користувачі та інші частини системи можуть отримувати дані швидше при цьому створюючи менше навантаження.

Для виконання завдання оптимізації витрат необхідно провести теоретичні дослідження, а саме:

– обрати актуальну конфігурацію сховища в залежності від наявних ресурсів таких як тип процесора, кількість процесорів, кількість оперативної пам'яті та обсяг диску на якому будуть зберігатися великі дані;

– обрати стратегію збереження даних на диску та їх розмежування в контейнерах;

– визначити стратегію індексування та валідації полів збережуваних моделей з даними.

Для виконання завдання оптимізації швидкості запису необхідно провести теоретичні дослідження, а саме:

– визначити стратегію побудови залежностей між різними сутностями для збереження;

– визначити стратегію індексування полів моделей та розподілу даних на фізичному диску;

– визначити стратегію написання програмного коду сховища та його використання при записі.

Для виконання завдання оптимізації швидкості зчитування необхідно провести теоретичні дослідження, а саме:

– обрати стратегію розподілення полів на фізичному диску та індексацію цих полів;

– обрати стратегію побудови залежностей між контейнерами та побудови вкладених сутностей;

– визначити конструкції для побудови запитів на зчитування даних з розподілених контейнерів.

Ці дослідження нададуть змогу отримати необхідні методи та практики оптимізації NoSQL сховищ за найбільш затребуваними напрямками підвищення ефективності, а саме:

– отримати методи та практики оптимізації сховищ за витратами для зменшення витрат на обслуговування та підтримку як самого сховища так і всієї інфраструктури;

– отримати методи та практики оптимізації сховищ за швидкістю запису нових даних до сховища для підвищення швидкості роботи всієї системи обробки та збереження даних;

– отримати методи та практики оптимізації сховищ за швидкістю зчитування даних із сховища, для підвищення швидкості отримання даних

користувачами та іншими частинами системи, що надасть змогу більш ефективно її використовувати.

## 2.3 Обрання вибірки даних

Для проведення дослідження необхідно мати велику кількість даних, тому для наповнення сховища було обрано рішення автоматично згенерувати 50 гігабайт даних взявши за основу датасет з лістингами на Airbnb, адже він має багато полів різних типів, таких як числа, текст, довгий текст, посилання, дати та інше. Приклад записів датасету можна побачити на рисунку 2.1.

```
{
  "id": "10006546", "listing_url": "https://www.airbnb.com/rooms/10006546", "name": "Ribeira Charming Duplex", "summary": "Fantastic duplex apartment with three bedrooms, located in the historic area of Porto, Ribeira (Cuba) - UNESCO World Heritage Site. Centenary building fully rehabilitated, without losing their original character.", "space": "Privileged views of the Douro River and Ribeira square, our apartment offers the perfect conditions to discover the history and the charm of Porto. Apartment comfortable, charming, romantic and cozy in the heart of Ribeira. Within walking distance of all the most emblematic places of the city of Porto. The apartment is fully equipped to host 8 people, with cooker, oven, washing machine, dishwasher, microwave, coffee machine (Nespresso) and kettle. The apartment is located in a very typical area of the city that allows to cross with the most picturesque population of the city, welcoming, genuine and happy people that fills the streets with his outspoken speech and contagious with your sincere generosity, wrapped in a only parochial spirit.", "description": "Fantastic duplex apartment with three bedrooms, located in the historic area of Porto, Ribeira (Cuba) - UNESCO World Heritage Site. Centenary building fully rehabilitated, without losing their original character. Privileged views of the Douro River and Ribeira square, our apartment offers the perfect conditions to discover the history and the charm of Porto. Apartment comfortable, charming, romantic and cozy in the heart of Ribeira. Within walking distance of all the most emblematic places of the city of Porto. The apartment is fully equipped to host 8 people, with cooker, oven, washing machine, dishwasher, microwave, coffee machine (Nespresso) and kettle. The apartment is located in a very typical area of the city that allows to cross with the most picturesque population of the city, welcoming, genuine and happy people that fills the streets with his outspoken speech and contagious with your sincere generosity, wrapped in a only parochial spirit. We are always available to help guests, without neighborhood overview." "In the neighborhood of the river, you can find several restaurants as varied flavors, but without forgetting the so traditional northern food. You can also find several bars and pubs to unwind after a day's visit to the magnificent Port. To enjoy the Douro River can board the boats that daily make the ride of six bridges. You can also embark towards Régua, Barca d'Alva, Pinhão, etc and enjoy the Douro wine Region, World Heritage of Humanity. The Infante's house is a few meters and no doubt it deserves a visit. They abound grocery stores, bakeries, etc. to make your meals. Souvenir shop, wine cellars, etc. to bring some souvenirs.", "notes": "Lose yourself in the narrow streets and staircases zone, have lunch in pubs and typical restaurants, and find the renovated cafes and shops in town. If you like exercise, rent a bicycle in the area and ride along the river to the sea, where it will enter beautiful beaches and terraces for everyone. The area is safe, find the bus stops 1min and metro line 5min. The bustling nightlife is a 10 min walk, where the streets are filled with people and entertainment for all. But Porto is much more than the historical center, here is modern museums, concert halls, clean and cared for beaches and surf all year round. Walk through the Ponte D. Luis and visit the different Caves of Port wine, where you will enjoy the famous port wine. Porto is a spoken city everywhere in the world as the best to be visited and savored by all... natural beauty, culture, tradition, river, sea, beach, single people, typical food, and we are among those who best receive tourists, confiram! Come visit us and feel at ho", "transit": "Transport: • Metro station and S. Bento railway 5min; • Bus stop a 50 meters; • Lift Guindais (Funicular) 50 meters; • Tuc Tuc to get around the city; • Buses tourist; • Cycling through the marginal drive; • Cable car in Gaia, overlooking the Port (just cross the bridge).", "access": "We are always available to help guests. The house is fully available to guests. We are always ready to assist guests, when possible we pick the guests at the airport. This service transfer have a cost per person. We will also have service 'meal at home' with a diverse menu and the taste of each. Enjoy the moment!" "Cot - 10 € / night Dog - € 7,5 / night", "house_rules": "Make the house your home...", "property_type": "House", "room_type": "Entire home/apt", "bed_type": "Real Bed", "minimum_nights": "2", "maximum_nights": "30", "cancellation_policy": "moderate", "last_scraped": {"$date": {"$numberLong": "1550293200000"}}, "calendar_last_scraped": {"$date": {"$numberLong": "1550293200000"}}, "first_review": {"$date": {"$numberLong": "1451797200000"}}, "last_review": {"$date": {"$numberLong": "1547960400000"}}, "accommodates": {"$numberInt": "8"}, "bedrooms": {"$numberInt": "3"}, "baths": {"$numberInt": "1"}, "number_of_reviews": {"$numberInt": "151"}, "bathrooms": {"$numberDecimal": "11.0"}, "amenities": [{"type": "Cable TV", "wifi", "Kitchen", "Paid parking off premises", "Smoking allowed", "Pets allowed", "Buzzer/wireless intercom", "Heating", "Family/kid friendly", "Washer", "First aid kit", "Fire extinguisher", "Essentials", "Hangers", "Hair dryer", "Iron", "Pack 'n Play/travel crib", "Room-darkening shades", "Hot water", "Bed linens", "Extra pillows and blankets", "Microwave", "Coffee maker", "Refrigerator", "Dishwasher", "Dishes and silverware", "Cooking basics", "Oven", "Stove", "Cleaning before checkout", "Waterfront"}, "price": {"$numberDecimal": "80.00"}, "security_deposit": {"$numberDecimal": "200.00"}, "cleaning_fee": {"$numberDecimal": "35.00"}, "extra_people": {"$numberDecimal": "15.00"}, "guests_included": {"$numberDecimal": "6"}, "images": [{"thumbnail_url": ""}, {"medium_url": ""}, {"picture_url": "https://a0.muscache.com/in/pictures/83e702f-ef49-40fb-8fa0-6512d7e26e9b.jpg?aki_policy=large"}, {"host": {"host_id": "51399391", "host_url": "https://www.airbnb.com/users/show/51399391", "host_name": "Ana&Goncalo", "host_location": "Porto, Porto District, Portugal", "host_about": "Gostamos de passear, de viajar, de conhecer pessoas e locais novos, gostamos de desporto e animais! Vivemos na cidade mais linda do mundo!!!", "host_response_time": "within an hour", "host_thumbnail_url": "https://a0.muscache.com/in/pictures/fab79f25-2e10-4f0f-9711-663cb69dc7d8.jpg?aki_policy=profile_small", "host_picture_url": "https://a0.muscache.com/in/pictures/fab79f25-2e10-4f0f-9711-663cb69dc7d8.jpg?aki_policy=profile_medium", "host_neighbourhood": "", "host_response_rate": {"$numberInt": "100"}, "host_is_superhost": false, "host_has_profile_pic": true, "host_identity_verified": true, "host_listings_count": {"$numberInt": "3"}, "host_total_listings_count": {"$numberInt": "3"}, "host_verifications": [{"email", "phone", "reviews", "jumio", "offline_government_id", "government_id"}], "address": {"street": "Porto, Porto, Portugal", "suburb": "", "government_area": "Cedofeita, Ildefonso, S6, Miragaia, Nicolau, V\u00edtor\u00eda", "market": "Porto", "country": "Portugal", "country_code": "PT", "location": {"type": "Point", "coordinates": [{"$numberDouble": "-8.61308"}, {"$numberDouble": "41.1413"}]}, "is_location_exact": false}, "availability": {"availability_30": {"$numberInt": "28"}, "availability_60": {"$numberInt": "47"}, "availability_90": {"$numberInt": "74"}, "availability_365": {"$numberInt": "239"}}, "review_scores_accuracy": {"$numberInt": "9"}, "review_scores_cleanliness": {"$numberInt": "9"}, "review_scores_checkin": {"$numberInt": "10"}, "review_scores_communication": {"$numberInt": "10"}, "review_scores_location": {"$numberInt": "10"}, "review_scores_value": {"$numberInt": "9"}, "review_scores_rating": {"$numberInt": "89"}}, "reviews": [{"id": "58663741", "date": {"$date": {"$numberLong": "1451797200000"}}, "listing_id": "10006546", "reviewer_id": "51483996", "reviewer_name": "G\u00e1tia", "comments": "A casa da Ana e do Goncalo foram o local escolhido para a passagem de ano com um grupo de amigos. Fomos super bem recebidos com uma grande simpatia e predisposi\u00e7\u00e3o a ajudar com qualquer coisa que fosse necess\u00e1rio.\r\n\r\nA casa era ainda melhor do que parecia nas fotos, totalmente equipada, com mantas, aquecedor e tudo o que ped\u00fassemos precisar.\r\n\r\nA localiza\u00e7\u00e3o n\u00e3o podia ser melhor! N\u00e3o h\u00e1 melhor do que acordar de manh\u00e3 e ao virar da esquina estar a ribeira do Porto.", "id": "62413197", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back next year. Porto is charming original.", "id": "69693942", "date": {"$date": {"$numberLong": "1455426000000"}}, "listing_id": "10006546", "reviewer_id": "40031996", "reviewer_name": "Theo", "comments": "We are french's students, we traveled some days in Porto, this space was good and we can cooking easily. It was raining so we eard every time the water fall to the ground in the street when we sleeping. But It was cool and or was well received by Ana et Goncalo", "id": "68310569", "date": {"$date": {"$numberLong": "1459742400000"}}, "listing_id": "10006546", "reviewer_id": "53859858", "reviewer_name": "Bart", "comments": "We had a splided time in the old centre of Porto.\r\n\r\nThe apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city.\r\n\r\nWe thank Ana & Goncalo, and we hope the apartment is free when we go back
```

- наявність полів різних типів;
- велика кількість даних;
- корисному навантаженню піддається лише мала частина датасету.

Для генерації додаткових даних створено окреме програмне забезпечення, яке буде використовуватися як для заповнення сховища даними, так і для проведення досліджень по оптимізації швидкості запису.

Завдяки використанню окремого програмного додатку для роботи з даними сховища можна масштабувати навантаження для проведення більш актуального дослідження.

Хоча цей датасет і є репрезентативним для більшості сховищ, але найбільшою проблемою NoSQL сховищ є розподілення даних між різними контейнерами та відсутність відношень між сутностями в різних контейнерах. Тому для дослідження та оптимізації роботи з даними потрібно розглянути сховище де наявні декілька пов'язаних між собою контейнерів.

Для цієї мети було обрано датасет з аналітикою транзакцій користувачів. Цей датасет має базові моделі, які пов'язані між собою та розподілені між різними контейнерами. Приклади моделей датасету можна побачити на рисунках 2.2, 2.3 та 2.4.

Для збільшення кількості даних було використано додаткове програмне забезпечення. З його допомогою створено додаткові дані, також воно буде використано для проведення дослідження по оптимізації швидкості запису та зчитування даних із сховища.

```

{"_id":{"$oid":"5ca4bbc7a2dd94ee5816238c"},"account_id":{"$numberInt":"371138"},"limit":{"$numberInt":"9000"},"products":["Derivatives","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816238d"},"account_id":{"$numberInt":"557378"},"limit":{"$numberInt":"10000"},"products":["InvestmentStock","Commodity","Brokerage","Currency"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816238e"},"account_id":{"$numberInt":"198100"},"limit":{"$numberInt":"10000"},"products":["Derivatives","CurrencyService","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816238f"},"account_id":{"$numberInt":"674364"},"limit":{"$numberInt":"10000"},"products":["InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162390"},"account_id":{"$numberInt":"278603"},"limit":{"$numberInt":"10000"},"products":["Commodity","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162391"},"account_id":{"$numberInt":"383777"},"limit":{"$numberInt":"10000"},"products":["CurrencyService","Derivatives","InvestmentFund","C"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162392"},"account_id":{"$numberInt":"794875"},"limit":{"$numberInt":"9000"},"products":["InvestmentFund","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162393"},"account_id":{"$numberInt":"328304"},"limit":{"$numberInt":"10000"},"products":["Derivatives","InvestmentStock","CurrencyService"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162394"},"account_id":{"$numberInt":"487188"},"limit":{"$numberInt":"10000"},"products":["Brokerage","CurrencyService","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162395"},"account_id":{"$numberInt":"910579"},"limit":{"$numberInt":"10000"},"products":["Brokerage","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162396"},"account_id":{"$numberInt":"260499"},"limit":{"$numberInt":"10000"},"products":["InvestmentFund","Derivatives","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162397"},"account_id":{"$numberInt":"668949"},"limit":{"$numberInt":"10000"},"products":["Brokerage","CurrencyService","InvestmentStock","Dei"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162398"},"account_id":{"$numberInt":"976027"},"limit":{"$numberInt":"10000"},"products":["Brokerage","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee58162399"},"account_id":{"$numberInt":"135185"},"limit":{"$numberInt":"10000"},"products":["CurrencyService","Brokerage","InvestmentStock","Cor"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816239a"},"account_id":{"$numberInt":"370583"},"limit":{"$numberInt":"10000"},"products":["Brokerage","Commodity","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816239b"},"account_id":{"$numberInt":"870466"},"limit":{"$numberInt":"10000"},"products":["Derivatives","Brokerage","Commodity","InvestmentSt"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816239c"},"account_id":{"$numberInt":"692278"},"limit":{"$numberInt":"10000"},"products":["Commodity","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816239d"},"account_id":{"$numberInt":"864905"},"limit":{"$numberInt":"10000"},"products":["Commodity","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816239e"},"account_id":{"$numberInt":"299072"},"limit":{"$numberInt":"10000"},"products":["InvestmentFund","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee5816239f"},"account_id":{"$numberInt":"137994"},"limit":{"$numberInt":"10000"},"products":["CurrencyService","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a0"},"account_id":{"$numberInt":"572981"},"limit":{"$numberInt":"10000"},"products":["InvestmentStock","CurrencyService"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a1"},"account_id":{"$numberInt":"977982"},"limit":{"$numberInt":"10000"},"products":["InvestmentFund","Derivatives","InvestmentStock","C"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a2"},"account_id":{"$numberInt":"616040"},"limit":{"$numberInt":"10000"},"products":["InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a3"},"account_id":{"$numberInt":"775273"},"limit":{"$numberInt":"10000"},"products":["Brokerage","Commodity","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a4"},"account_id":{"$numberInt":"212024"},"limit":{"$numberInt":"10000"},"products":["InvestmentFund","Brokerage","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a5"},"account_id":{"$numberInt":"433811"},"limit":{"$numberInt":"10000"},"products":["CurrencyService","InvestmentFund","Brokerage","Inv"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a6"},"account_id":{"$numberInt":"353465"},"limit":{"$numberInt":"10000"},"products":["CurrencyService","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a7"},"account_id":{"$numberInt":"464470"},"limit":{"$numberInt":"10000"},"products":["Brokerage","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623a9"},"account_id":{"$numberInt":"324287"},"limit":{"$numberInt":"10000"},"products":["Commodity","CurrencyService","Derivatives","Invest"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623aa"},"account_id":{"$numberInt":"403363"},"limit":{"$numberInt":"10000"},"products":["CurrencyService","InvestmentFund","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623ac"},"account_id":{"$numberInt":"276528"},"limit":{"$numberInt":"10000"},"products":["InvestmentFund","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623ad"},"account_id":{"$numberInt":"383701"},"limit":{"$numberInt":"10000"},"products":["InvestmentFund","InvestmentStock"]}
{"_id":{"$oid":"5ca4bbc7a2dd94ee581623ae"},"account_id":{"$numberInt":"423171"},"limit":{"$numberInt":"10000"},"products":["Brokerage","CurrencyService","Derivatives","Invest"]}

```

Рисунок 2.2 – Уривок датасету акаунтів

```

{"_id":{"$oid":"5ca4bbcea2dd94ee58162a68"},"username":"fmiller","name":"Elizabeth Ray","address":"9286 Bethany Glens\nVasqueztown, CO 22939","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a69"},"username":"valenciajennifer","name":"Lindsay Cowan","address":"Unit 1047 Box 4089\nNDPO AA 57348","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a6a"},"username":"hillrachel","name":"Katherine David","address":"55711 Janet Plaza Apt. 865\nChristinachester, CT 62716","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a6b"},"username":"serranobrian","name":"Leslie Martinez","address":"Unit 2676 Box 9352\nNDPO AA 38560","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a6c"},"username":"charleshudson","name":"Brad Cardenas","address":"2765 Powers Meadow\nHeatherfurt, CT 53165","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a6d"},"username":"gregoryharrison","name":"Natalie Ford","address":"17677 Mark Crest\nWalterberg, IA 39017","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a6e"},"username":"hmyers","name":"Dana Clarke","address":"50047 Smith Point Suite 162\nWilkinsstad, PA 04106","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a6f"},"username":"andrewhamilton","name":"Gary Nichols","address":"633 Miller Turnpike\nJonathanland, OR 62874","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a70"},"username":"matthewray","name":"John Parks","address":"38456 Rachael Causeway Apt. 735\nEvanfort, AR 33893","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a71"},"username":"glopez","name":"Jennifer Lawrence","address":"4140 Pamela Hollow Apt. 849\nEast Elizabeth, TN 29566","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a72"},"username":"wesley20","name":"James Sanchez","address":"8681 Karen Roads Apt. 096\nLowehaven, IA 19798","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a73"},"username":"thomasdavid","name":"Ashley Lopez","address":"18637 Jessica Ridge Apt. 157\nGrossberg, ME 84127","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a74"},"username":"patricia44","name":"Dr. Angela Brown","address":"2129 Joel Rapids\nLisahaven, NE 08609","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a75"},"username":"nelsonmaria","name":"John Vega","address":"86636 Maria Viaduct\nKennethhaven, SD 21876","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a76"},"username":"porterchristian","name":"Lauren Clark","address":"1579 Young Trail\nJessechester, OH 88328","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a77"},"username":"johnsonshelly","name":"Jacqueline Haynes","address":"USNS Howard\nFPO AP 30863","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a78"},"username":"hunterdaniel","name":"Brian Flores","address":"70092 Adams Prairie\nTurnerborough, TX 38603","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a79"},"username":"james75","name":"Christopher Gomez","address":"7322 Owens Inlet Apt. 688\nPort Leslie, OR 81893","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a7a"},"username":"eric10","name":"Robert Burns","address":"86176 Katherine Common\nWebbham, WA 51980","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a7b"},"username":"millerrenee","name":"Joshua Parker","address":"932 Jeremy Springs Suite 144\nJohnmouth, NM 02561","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a7c"},"username":"michael58","name":"Christine Douglas","address":"USNV Chavez\nFPO AP 78727","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a7d"},"username":"zsanders","name":"Derek Curtis","address":"565 Hodge Motorway Suite 101\nWendyberg, FL 57099","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a7e"},"username":"taylorbullock","name":"Shirley Rodriguez","address":"7637 Johnson Circles\nNew Laurahaven, KY 21914","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a7f"},"username":"archersarah","name":"Brittany Ellis","address":"81536 Underwood Freeway\nTurnerfort, OK 97967","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a80"},"username":"thomasboyd","name":"Clinton Shelton","address":"41852 Hodges Plaza Suite 567\nRachelhaven, MA 95137","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a81"},"username":"anthonyandrada","name":"Anna Perkins","address":"4399 Tony Turnpike Suite 439\nNorth Veronica, MD 57595","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a82"},"username":"andreawalker","name":"Kenneth Murphy","address":"24615 Black Mountain\nEast Carrieport, AL 41910","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a83"},"username":"rodriguezjeffrey","name":"Steven Campbell","address":"69809 Morris Extensions\nTaylorhaven, ME 37916","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a84"},"username":"kanelinda","name":"Ann Williams","address":"46556 Sutton Points\nNew Luis, CT 82525","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a85"},"username":"vmartinez","name":"Crystal Brown","address":"145 David Walks\nJenniferberg, TX 29872","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a86"},"username":"ethanarias","name":"Christopher Watson","address":"196 Baker Key\nNorth Billystad, AL 12800","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a87"},"username":"stephanie68","name":"Matthew Jones","address":"0297 Jacob Ranch Apt. 019\nNorth Judith, NV 27455","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a88"},"username":"paul182","name":"Joseph Dawson","address":"Unit 6696 Box 1901\nNDPO AE 45202","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a89"},"username":"jessica94","name":"Brandon Moore","address":"4523 Bright Roads Apt. 229\nHamptonfurt, ND 58176","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a8a"},"username":"anthonygarza","name":"Jeffrey Lester","address":"50132 Moore Meadows\nNew Stephanie, GA 22761","birthdate":{"$date":{"$numberLong":}}}
{"_id":{"$oid":"5ca4bbcea2dd94ee58162a8b"},"username":"thomasmendoza","name":"David Lawrence","address":"7193 Gibbs Key\nPort Lindsey, CO 60258","birthdate":{"$date":{"$numberLong":}}}

```

Рисунок 2.3 – Уривок датасету користувачів

```

{"_id":{"_id":"5ca4b1a2dd94e58161c1"}, "account_id":{"$numberInt":"443178"}, "transaction_count":{"$numberInt":"66"}, "bucket_start_date":{"$date":{"$numberLong":"-285984000000"}}, "bucket_end_date":{"$date":{"$numberLong":"1483401600000"}}, "transactions":{"$date":{"$date":{"$numberLong":"106306500000"}}, "amount":{"$numberInt":"7514"}, "transaction_code":"buy", "symbol":"adbe", "price":"19.107280265007418051936838102118968963623046875", "total":{"$numberInt":"143572.1039112657392422534031"}, {"date":{"$date":{"$numberLong":"1465862400000"}}, "amount":{"$numberInt":"9240"}, "transaction_code":"buy", "symbol":"team", "price":"24.152563238771480580340721644461154937744140625", "total":{"$numberInt":"223169.6843263008480562348268"}, {"date":{"$date":{"$numberLong":"103896000000"}}, "amount":{"$numberInt":"2824"}, "transaction_code":"buy", "symbol":"msft", "price":"21.046193953245431629284212249331176280975341796875", "total":{"$numberInt":"59434.45172396509892109861539"}, {"date":{"$date":{"$numberLong":"148529800000"}}, "amount":{"$numberInt":"7418"}, "transaction_code":"sell", "symbol":"sap", "price":"76.38514540536692720706923864781856536865234375", "total":{"$numberInt":"566625.0086170118660220396123"}, {"date":{"$date":{"$numberLong":"131976000000"}}, "amount":{"$numberInt":"5638"}, "transaction_code":"buy", "symbol":"adbe", "price":"28.36565760864117535788864188361912965745361328125", "total":{"$numberInt":"159925.5775975189466677761629"}, {"date":{"$date":{"$numberLong":"1292803200000"}}, "amount":{"$numberInt":"6774"}, "transaction_code":"sell", "price":"45.327971108758159743956639431416988372802734375", "total":{"$numberInt":"307023.4016290727774105562276"}, {"date":{"$date":{"$numberLong":"144590400000"}}, "amount":{"$numberInt":"7251"}, "transaction_code":"sell", "symbol":"nflx", "price":"102.936674830205050264150486327707767486572265625", "total":{"$numberInt":"746393.8329518248194653515764"}, {"date":{"$date":{"$numberLong":"132814800000"}}, "amount":{"$numberInt":"8828"}, "transaction_code":"buy", "symbol":"nflx", "price":"117.838862313630283296106426860205829143524169921875", "total":{"$numberInt":"157481.4765047281409380275363"}, {"date":{"$date":{"$numberLong":"145791360000"}}, "amount":{"$numberInt":"5585"}, "transaction_code":"sell", "symbol":"team", "price":"22.645934804740733170547173358500003814697265625", "total":{"$numberInt":"126477.545884476994757059632"}, {"date":{"$date":{"$numberLong":"133056000000"}}, "amount":{"$numberInt":"5789"}, "transaction_code":"buy", "symbol":"nflx", "price":"16.0804676963220600782733527012169361114501953125", "total":{"$numberInt":"93089.8274940084057931243879"}, {"date":{"$date":{"$numberLong":"978564000000"}}, "amount":{"$numberInt":"7856"}, "transaction_code":"buy", "symbol":"adbe", "price":"26.710984086786963245687156450033855438232421875", "total":{"$numberInt":"209841.4909857985645658118301"}, {"date":{"$date":{"$numberLong":"1458051200000"}}, "amount":{"$numberInt":"8539"}, "transaction_code":"buy", "symbol":"nflx", "price":"120.200206730275638734738974459469318389892578125", "total":{"$numberInt":"1026389.565269823679145133610"}, {"date":{"$date":{"$numberLong":"147562500000"}}, "amount":{"$numberInt":"4967"}, "transaction_code":"buy", "symbol":"sap", "price":"90.3506675309771652392111718654632568359375", "total":{"$numberInt":"448771.7656263663179743161891"}, {"date":{"$date":{"$numberLong":"1474848000000"}}, "amount":{"$numberInt":"2893"}, "transaction_code":"sell", "symbol":"team", "price":"30.227075585186970130280315061099827289581298828125", "total":{"$numberInt":"87446.92966794590458690095147"}, {"date":{"$date":{"$numberLong":"711331200000"}}, "amount":{"$numberInt":"9559"}, "transaction_code":"sell", "symbol":"ibm", "price":"19.83986843992497739463942707516252994537353515625", "total":{"$numberInt":"189649.3024172428589153828234"}, {"date":{"$date":{"$numberLong":"256435200000"}}, "amount":{"$numberInt":"8269"}, "transaction_code":"buy", "symbol":"ibm", "price":"13.2722697458426655003904670593287739753724153125", "total":{"$numberInt":"109748.398528373001022728721"}, {"date":{"$date":{"$numberLong":"1464739200000"}}, "amount":{"$numberInt":"8081"}, "transaction_code":"sell", "symbol":"ibm", "price":"143.758609271150135100469924509525299072265625", "total":{"$numberInt":"1161713.321520164241746897460"}, {"date":{"$date":{"$numberLong":"530409600000"}}, "amount":{"$numberInt":"6145"}, "transaction_code":"sell", "symbol":"adbe", "price":"0.2800000011920928955078125", "total":{"$numberInt":"1720.600007325410842895507812"}, {"date":{"$date":{"$numberLong":"1287705600000"}}, "amount":{"$numberInt":"6973"}, "transaction_code":"buy", "symbol":"msft", "price":"21.268423447276756377277066332194089895263671875", "total":{"$numberInt":"148304.716697860822187574484"}, {"date":{"$date":{"$numberInt":"3082"}, "transaction_code":"sell", "symbol":"sap", "price":"86.029571359374330086211557500064373016357421875", "total":{"$numberInt":"265143.138925916853257040202"}, {"date":{"$date":{"$numberLong":"595814400000"}}, "amount":{"$numberInt":"2978"}, "transaction_code":"buy", "symbol":"msft", "price":"0.243200014591217041015625", "total":{"$numberInt":"724.309164345264434814531250"}, {"date":{"$date":{"$numberLong":"1014163200000"}}, "amount":{"$numberInt":"9265"}, "transaction_code":"sell", "symbol":"ibm", "price":"81.091017566265037430639495141804218229236328125", "total":{"$numberInt":"751308.27775144557179487492425"}, {"date":{"$date":{"$numberLong":"148032000000"}}, "amount":{"$numberInt":"9192"}, "transaction_code":"buy", "symbol":"sap", "price":"74.557668636076840584792080335319042205810546875", "total":{"$numberInt":"685334.0501028183186554088024"}, {"date":{"$date":{"$numberLong":"59184000000"}}, "amount":{"$numberInt":"7408"}, "transaction_code":"sell", "symbol":"ibm", "price":"12.426717296569131931960328074637800450933837890625", "total":{"$numberInt":"92057.12173298412935196211038"}, {"date":{"$date":{"$numberLong":"1324598400000"}}, "amount":{"$numberInt":"3267"}, "transaction_code":"sell", "symbol":"nflx", "price":"10.5489864757904783232340117000043392181396484375", "total":{"$numberInt":"34463.53881640749268200352162"}, {"date":{"$date":{"$numberLong":"1337817600000"}}, "amount":{"$numberInt":"7582"}, "transaction_code":"buy", "symbol":"nflx", "price":"10.1614266582902761371087763189244406223297119140625", "total":{"$numberInt":"77043.93692315687367155874199"}, {"date":{"$date":{"$numberLong":"134542000000"}}, "amount":{"$numberInt":"643"}, "transaction_code":"sell", "symbol":"nflx", "price":"9.158372036088804790665529049055278301239013671875", "total":{"$numberInt":"5888.833219205101480397956948"}, {"date":{"$date":{"$numberLong":"1364342400000"}}, "amount":{"$numberInt":"3129"}, "transaction_code":"buy", "symbol":"sap", "price":"74.335708210354198399727465584874513132703125", "total":{"$numberInt":"232596.430990198286792742398"}, {"date":{"$date":{"$numberLong":"144542000000"}}, "amount":{"$numberInt":"4075"}, "transaction_code":"buy", "symbol":"msft", "price":"45.3716259275177193899253250711858272552490234375", "total":{"$numberInt":"184889.3756546347065139457300"}, {"date":{"$date":{"$numberLong":"146905200000"}}, "amount":{"$numberInt":"4386"}, "transaction_code":"sell", "symbol":"team", "price":"27.93690667695467376680611693680286407470703125", "total":{"$numberInt":"122531.2726445123199141562509"}, {"date":{"$date":{"$numberLong":"754790400000"}}, "amount":{"$numberInt":"3093"}, "transaction_code":"buy", "symbol":"adbe", "price":"2.9789810807763135170812347496394068002700805640625", "total":{"$numberInt":"9213.98848284521377083225908"}, {"date":{"$date":{"$numberLong":"1306886400000"}}, "amount":{"$numberInt":"5053"}, "transaction_code":"buy", "symbol":"adbe", "price":"34.76661562291676683013066548528373241424560546875", "total":{"$numberInt":"175675.7097702518422792650199"}

```

Рисунок 2.4 – Уривок датасету транзакцій

## 2.4 Програмні ресурси

Для реалізації дослідження було використано NoSQL сховище MongoDB. MongoDB – документоорієнтована СУБД, вона не вимагає опису схеми. Вважається класичним прикладом NoSQL-систем, використовує JSON-подібні документи та схему бази даних. Написана на мові C++, але дарма, що не слов'їною. Застосовується зазвичай у веб-розробці, зокрема, у рамках JavaScript стека MEAN, та використовується для проектів з великими даними.

Система підтримує ad-hoc запити. Вони можуть повертати конкретні поля документів та користувацькі JavaScript функції. Також є підтримка пошуку за регулярними виразами. Можна налаштувати запити на отримання випадкових документів.

Є підтримка індексів, що дозволяє підвищити швидкість.

Система може працювати з набором реплік, тобто може містити дві або більше копій даних на різних вузлах, які знаходяться в незалежних місцях. Кожен екземпляр набору реплік може будь-якої миті виступати в

ролі мастера або слейва. Усі операції запису та отримання доступу за замовчуванням здійснюються з майстер реплікою. Допоміжні репліки підтримують копію даних у актуальному стані. У разі коли основна репліка дає збій, набір реплік проводить вибір, яка з реплік має стати основною. Слейв репліки можуть додатково бути джерелом для отримання доступу до даних.

Система масштабується горизонтально, використовуючи техніку сегментування об'єктів баз даних – розподіл їх частин різними вузлами кластера. Адміністратор вибирає ключ сегментування, який визначає, за яким критерієм дані будуть рознесені вузлами (залежно від значень хеша ключа сегментування). Завдяки тому, що кожен вузол кластера може приймати запити забезпечується балансування навантаження.

Система може бути використана як файлове сховище з балансуванням навантаження та реплікацією даних (функція Grid File System; поставляється разом з драйверами MongoDB). Надаються програмні засоби для роботи з файлами та їх вмістом. GridFS використовується в плагінах для Nginx та lighttpd. GridFS поділяє файл на частини та зберігає кожен частину як окремий документ.

Може працювати відповідно до парадигми MapReduce. Для агрегації даних передбачено аналог SQL-виразу GROUP BY; оператори агрегації можуть бути пов'язані в ланцюзі подібно до Unix-конвєсів. Фреймворк також має оператор \$lookup для зв'язування документів при розвантаженні та статистичні операції, такі як середньоквадратичне відхилення.

Підтримується JavaScript у запитах, функціях агрегації (наприклад, MapReduce).

Підтримуються колекції із фіксованим розміром. Такі колекції зберігають порядок вставки і після досягнення заданого розміру поводяться як кільцевий буфер.

У червні 2018 року (у версії 4.0) додано підтримку транзакцій, які відповідають вимогам ACID.

Є офіційні драйвери для основних мов програмування. Існує також велика кількість неофіційних або підтримуваних спільнотою драйверів для інших мов програмування та фреймворків.

Основним інтерфейсом бази даних була командна оболонка mongo. З версії MongoDB 3.2 як графічна оболонка поставляється «MongoDB Compass». Існують продукти та сторонні проекти, які пропонують інструменти з графічним інтерфейсом для адміністрування та перегляду даних.

Для реалізації програмного додатку було використано мову C#.

C# (вимовляється сі шарп) – об'єктно-орієнтована мова програмування загального призначення. Розроблено в 1998-2001 роках групою інженерів компанії Microsoft під керівництвом Андерса Хейлсберга і Скотта Вільтаумота як мова розробки додатків для платформи Microsoft .NET Framework та .NET Core. Згодом був стандартизований як ECMA-334 та ISO/IEC 23270.

C# відноситься до сім'ї мов з C-подібним синтаксисом, їх синтаксис найбільш близький до C++ і Java. Мова має статичну типізацію, підтримує поліморфізм, перевантаження операторів (у тому числі операторів явного та неявного приведення типу), делегати, атрибути, події, змінні, властивості, узагальнені типи та методи, ітератори, анонімні функції з підтримкою замикань, LINQ, винятки, коментарі у форматі XML.

Перейнявши багато від своїх попередників – мов C++, Delphi, Модула, Smalltalk і, особливо, Java – C#, спираючись на практику їх використання, виключає деякі моделі, що зарекомендували себе проблематичні при розробці програмних систем, наприклад, C# на відміну від C++ не підтримує множинне успадкування класів (між тим допускається множинна реалізація інтерфейсів).

C# розроблявся як мова програмування прикладного рівня для CLR і, як такої, залежить, перш за все, від можливостей самої CLR. Це стосується, перш за все, системи типів C#, яка відображає BCL. Присутність або

відсутність тих чи інших виразних особливостей мови диктується тим, чи конкретна мовна особливість може бути трансльована у відповідні конструкції CLR. Так, з розвитком CLR від версії 1.1 до 2.0 значно збагатився сам C#; подібної взаємодії слід очікувати і надалі (проте, ця закономірність була порушена з виходом C# 3.0, що є розширення мови, що не спираються на розширення платформи .NET). CLR надає C#, як і всім іншим .NET-орієнтованим мовам, багато можливостей, яких позбавлені «класичні» мови програмування. Наприклад, складання сміття не реалізована в самому C#, а проводиться CLR для програм, написаних на C#, так само, як це робиться для програм на VB.NET, J# та ін.

Для збереження даних було використано формат JSON.

JSON (англ. JavaScript Object Notation, зазвичай вимовляється як /'dʒeɪsən/ JAY-sən) – текстовий формат обміну даними, заснований на JavaScript. Як і багато інших текстових форматів, JSON легко читається людьми. Формат JSON був розроблений Дугласом Крокфордом.

Незважаючи на походження від JavaScript (точніше, від підмножини мови стандарту ECMA-262 1999), формат вважається незалежним від мови і може використовуватися практично з будь-якою мовою програмування. Для багатьох мов існує готовий код для створення та обробки даних у форматі JSON.

## 3 ОПИС ПРИНЦИПІВ РОБОТИ NoSQL СХОВИЩ

### 3.1 Загальний опис роботи сховища

Для подальшої розробки методів оптимізації потрібно розуміти базові принципи роботи NoSQL сховищ. Більшість сховищ схожим шляхом зберігають, оброблюють та надають доступ до даних на диску. В ході роботи буде використовуватися сховище MongoDB, тому для опису буде використано саме опис роботи MongoDB.

MongoDB представляє найбільш популярну на даний момент документоорієнтовану систему управління базами даних. За різними оцінками входить до десятка найбільш використовуваних баз даних у світі.

На момент написання останньою версією платформи була версія 6.0, що побачила світ у липні 2022 року. Використання конкретної версії може дещо відрізнятись від інших версій платформи MongoDB. Але в цілому відмінності у побудові запитів у MongoDB між версіями мінімальні, має місце досить велика сумісність попередніх версій з наступними.

Якщо реляційні бази даних зберігають рядки, MongoDB зберігає документи. На відміну від рядків, документи можуть зберігати складну за структурою інформацію. Документ можна подати як сховище ключів та значень.

Ключ представляє просту мітку, з якою асоційовано певний шматок даних.

Однак за всіх відмінностей є одна особливість, яка зближує MongoDB та реляційні бази даних. У реляційних СУБД зустрічається таке поняття як первинний ключ. Це поняття описує стовпець, який має унікальні значення. У MongoDB для кожного документа є унікальний ідентифікатор, який називається `_id`. І якщо явно не вказати його значення, то MongoDB автоматично згенерує для нього значення.

Кожному ключу порівнюється певне значення. Але тут також треба враховувати одну особливість, а саме якщо в реляційних базах є чітко окреслена структура, де є поля, і якщо якийсь поле не має значення, то йому можна надати значення NULL. У MongoDB все інакше. Якщо якомусь ключу не зіставлено значення, цей ключ легко опускається у документі і використовується.

Якщо традиційному світі SQL є таблиці, то світі MongoDB є колекції. І якщо в реляційних БД таблиці зберігають однотипні жорстко структуровані об'єкти, то в колекції можуть містити різні об'єкти, що мають різну структуру і різний набір властивостей.

Вся система MongoDB може представляти не лише одну базу даних, яка розміщується на одному фізичному сервері. Функціональність MongoDB дозволяє розташувати декілька баз даних на кількох фізичних серверах, і ці бази даних зможуть легко обмінюватись даними та зберігати цілісність.

Система зберігання даних MongoDB представляє набір реплік. У цьому наборі є основний вузол і може бути набір вторинних вузлів. Усі вторинні вузли зберігають цілісність та автоматично оновлюються разом із оновленням головного вузла. І якщо основний вузол з якихось причин виходить з ладу, то один із вторинних вузлів стає головним.

Одним із популярних стандартів обміну даними та їх зберігання є JSON (JavaScript Object Notation). JSON ефективно описує складні структури даних. Спосіб зберігання даних MongoDB у цьому плані схожий на JSON, хоча формально JSON не використовується. Для зберігання MongoDB застосовується формат, який називається BSON або скорочення від binary JSON.

BSON дозволяє працювати з даними швидше, а саме швидше виконується пошук та обробка. Хоча треба зазначити, що BSON на відміну від зберігання даних у форматі JSON має невеликий недолік: загалом дані в

JSON-форматі займають менше місця, ніж у форматі BSON, з іншого боку, цей недолік з лишком окупається швидкістю.

MongoDB написана на C++, тому її легко портувати на різні платформи. MongoDB може бути розгорнуто на платформах Windows, Linux, MacOS, Solaris. Можна також завантажити вихідний код і самому скомпілювати MongoDB, але рекомендується використовувати бібліотеки з офсайту.

Відсутність жорсткої схеми бази даних та у зв'язку з цим потреби при найменшій зміні концепції зберігання даних перестворювати цю схему значно полегшують роботу з базами даних MongoDB та подальшим їх масштабуванням. Крім того, заощаджується час розробників. Їм більше не треба думати про перестворення бази даних та витратити час на побудову складних запитів.

Але, навіть з огляду на всі недоліки традиційних баз даних та переваги MongoDB, важливо розуміти, що завдання бувають різні та методи їх вирішення бувають різні. У якійсь ситуації MongoDB дійсно покращить продуктивність вашої програми, наприклад, якщо треба зберігати складні структури дані. В іншій ситуації краще буде використовувати традиційні реляційні бази даних. Крім того, можна використовувати змішаний підхід: зберігати один тип даних у MongoDB, а інший тип даних – у традиційних БД.

Однією із проблем під час роботи з будь-якими системами баз даних є збереження даних великого розміру. Можна зберігати дані у файлах, використовуючи різні мови програмування. Деякі СУБД пропонують спеціальні типи даних для зберігання бінарних даних БД (наприклад, BLOB в MySQL).

На відміну від реляційних СУБД MongoDB дозволяє зберігати різні документи з різним набором даних, проте розмір документа обмежується 16 мб. Але MongoDB пропонує рішення – спеціальну технологію GridFS, яка дозволяє зберігати дані за розміром більше 16 мб.

Система GridFS складається із двох колекцій. У першій колекції, яка називається `files`, зберігаються імена файлів, а також їх метадані, наприклад розмір. А в іншій колекції, яка називається `chunks`, у вигляді невеликих сегментів зберігаються дані файлів зазвичай сегментами по 256 кб.

Для тестування GridFS можна використовувати спеціальну утиліту `mongofiles`, яка йде у пакеті `mongodb`.

### 3.2 Опис принципів роботи із конфігураціями

Ми можемо конфігурувати роботу `mongod` за допомогою файлу конфігурації. `Mongod` екземпляр дивиться конфігураційний файл перед стартом, який у свою чергу містить налаштування для роботи бази. `MongoDB` для конфігураційного файлу використовує формат `YAML`.

Стандартний файл конфігурації лежить на шляху `etc/mongod.conf` ми можемо запустити `mongod` з використанням будь-якого конфігураційного файлу за допомогою прапора `--config`.

```
systemLog:
  destination: file
  path: "/var/log/mongodb/mongod.log"
  logAppend: true
storage:
  journal:
    enabled: true
processManagement:
  fork: true
net:
  bindIp: 127.0.0.1
  port: 27017
setParameter:
  enableLocalhostAuthBypass: false
```

Рисунок 3.1 – Приклад конфігураційного файлу

Конфігураційний файл розподіляється на частини, які відповідають за різні можливості серверу. До головних можна віднести:

- `systemLog` – відповідає за логування помилок, внутрішньої інформації виконання та розміщення цих логів на диску;
- `net` – відповідає за конфігурації мережі серверу;
- `security` – відповідає за конфігурації безпеки та доступу до серверу;
- `setParameter` – відповідає за додаткові параметри конфігурації сервера;
- `storage` – відповідає за конфігурації збереження даних сховища.

### 3.3 Опис принципів роботи збереження та розташування даних

WiredTiger використовує керування паралелізмом на рівні документа для операцій запису. У результаті кілька клієнтів можуть змінювати різні документи колекції одночасно.

Для більшості операцій читання та запису WiredTiger використовує оптимістичне керування паралелізмом. WiredTiger використовує лише блокування намірів на глобальному рівні, рівні бази даних і колекції. Коли система зберігання даних виявляє конфлікти між двома операціями, одна з них викличе конфлікт запису, що змусить MongoDB прозоро повторити цю операцію.

Деякі глобальні операції, як правило, короточасні операції, пов'язані з кількома базами даних, все ще потребують глобального блокування «всього екземпляра». Деякі інші операції, такі як `collMod`, все ще потребують ексклюзивного блокування бази даних.

WiredTiger використовує MultiVersion Concurrency Control (MVCC). На початку операції WiredTiger надає моментальний знімок даних для операції. Знімок представляє послідовне уявлення про дані в пам'яті.

Під час запису на диск WiredTiger записує всі дані зі знімка на диск узгоджено для всіх файлів даних. Тепер довговічні дані діють як контрольна

точка у файлах даних. Контрольна точка забезпечує узгодженість файлів даних до останньої контрольної точки включно; тобто контрольні точки можуть діяти як точки відновлення.

Починаючи з версії 3.6, MongoDB налаштовує WiredTiger для створення контрольних точок (тобто запису даних знімка на диск) з інтервалом у 60 секунд. У попередніх версіях MongoDB встановлює контрольні точки в WiredTiger для даних користувача з інтервалом у 60 секунд або після запису 2 ГБ даних журналу, залежно від того, що відбудеться раніше.

Під час запису нової контрольної точки попередня контрольна точка залишається дійсною. Таким чином, навіть якщо MongoDB завершує роботу або виявляє помилку під час запису нової контрольної точки, після перезапуску MongoDB може відновити роботу з останньої дійсної контрольної точки.

Нова контрольна точка стає доступною та постійною, коли таблиця метаданих WiredTiger атомарно оновлюється для посилання на нову контрольну точку. Коли нова контрольна точка стає доступною, WiredTiger звільняє сторінки зі старих контрольних точок.

Використання WiredTiger, навіть без ведення журналу, MongoDB може відновлюватися з останньої контрольної точки; однак, щоб відновити зміни, зроблені після останньої контрольної точки, запустить зведення журналу.

### 3.4 Опис принципів роботи виконання та обробки запитів

`db.collection.insertOne()` вставляє один документ у колекцію. На рисунку 3.2 новий документ вставляється в колекцію. Якщо в документі не вказано поле `_id`, MongoDB додає поле `_id` зі значенням `ObjectId` до нового документа.

```
1 db.test.insertOne(  
2   {  
3     item: "canvas",  
4     qty: 100,  
5     tags: ["cotton"],  
6     size:  
7     {  
8       h: 28,  
9       w: 35.5,  
10      uom: "cm"  
11    }  
12  }  
13 )
```

Рисунок 3.2 – Приклад додавання документу до сховища

Щоб вибрати всі документи в колекції, потрібно передати порожній документ як параметр фільтра запиту в метод `find`. Приклад можна побачити на рисунку 3.3.

```
db.inventory.find( {} )
```

Рисунок 3.3 – Приклад пошуку всіх документів

Якщо натомість потрібно знайти масив, який містить елементи із зазначених значень, незалежно від порядку чи інших елементів у масиві, потрібно використовувати оператор `$all`. Приклад використання можна побачити на рисунку 3.4.

```
db.inventory.find( { tags: { $all: ["red", "blank"] } } )
```

Рисунок 3.4 – Приклад пошуку документів з використанням `$all`

У наступному прикладі використовується метод `db.collection.updateOne()` для оновлення першого документа, де `item` дорівнює "paper". Приклад можна побачити на рисунку 3.5.

```
db.inventory.updateOne(  
  { item: "paper" },  
  {  
    $set: { "size.uom": "cm", status: "P" },  
    $currentDate: { lastModified: true }  
  }  
)
```

Рисунок 3.5 – Приклад оновлення документа

Щоб видалити один документ, який відповідає вказаному фільтру (навіть якщо кілька документів можуть відповідати зазначеному фільтру), використовується метод `db.collection.deleteOne()`. На рисунку 3.6 наведено приклад видалення першого документа із статусом "D".

```
db.inventory.deleteOne( { status: "D" } )
```

Рисунок 3.6 – Приклад видалення документа

## 4 РОЗРОБКА ТА АНАЛІЗ МЕТОДІВ ОПТИМІЗАЦІЇ

### 4.1 Розробка методів оптимізації конфігурацій сховища

Хоча MongoDB і дозволяє змінювати конфігурації сховища, але значення за замовчуванням зазвичай надають найбільшу продуктивність. Тому залишається не так багато простору для оптимізації, ніж це є для запитів.

По-перше потрібно вимкнути логування, архівування та аудіювання. Також потрібно вимкнути журналювання. Ці заходи дозволять прибрати додаткове навантаження на диск. Вплив кожної зміни на продуктивність можна побачити нижче. В якості заміру використовується запит на 527 документів із бази розміром 50 гігабайт на локальному сервері.

На рисунку 4.1 можна побачити вимкнення логування не надає значних переваг в продуктивності на дистанції, але робить роботу з диском більш стабільною за швидкістю. Це відбувається через відсутність потреби розподіляти диск між декількома процесами на комп'ютері. Чим менше процесів використовує диск, тим стабільніше працюють запити.

Порівняння продуктивності стандартної конфігурації та без логування

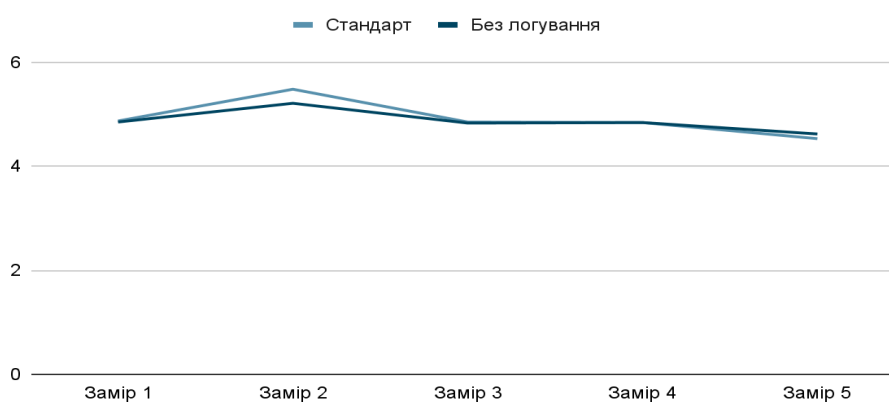


Рисунок 4.1 – Порівняння продуктивності стандартної конфігурації та без логування

Як можна побачити на рисунку 4.2 вимкнення архівування, як і вимкнення логування, не надало значного приросту продуктивності, але також надало стабільності виконанню запитів.

### Порівняння продуктивності стандартної конфігурації та без архівування

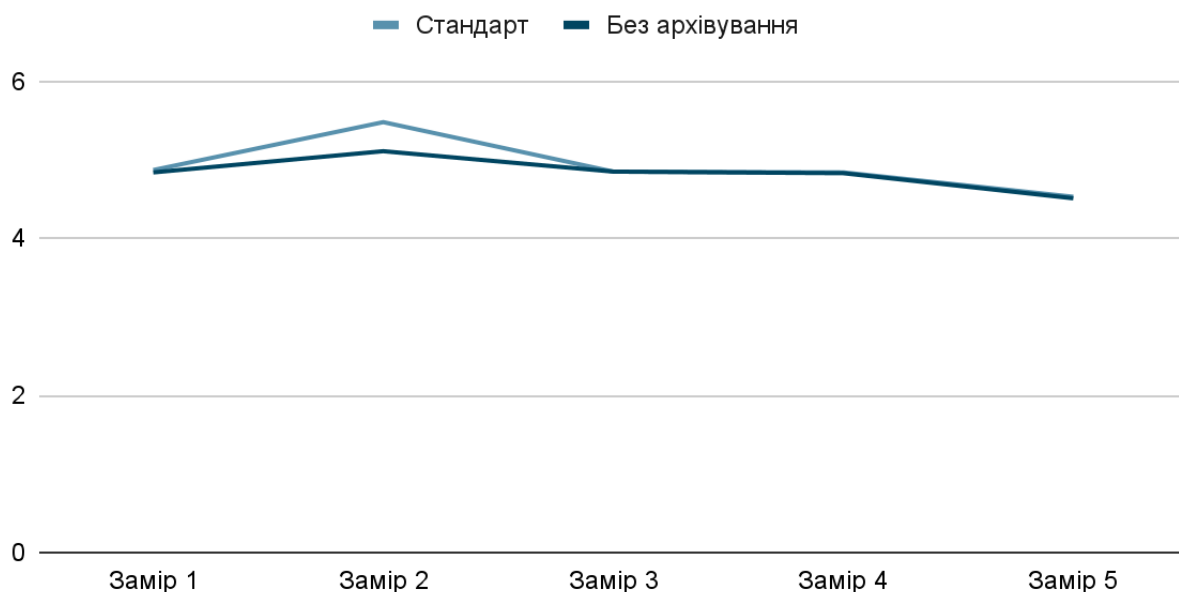


Рисунок 4.2 – Порівняння продуктивності стандартної конфігурації та без архівування

Як можна побачити на рисунку 4.3 вимкнення аудіювання додає певної продуктивності запитам. Це відбувається через відсутність потреби виконувати додаткові процеси, які завантажують диск. Також процес аудитування потребує додаткового навантаження на центральний процесор та використовує додаткову оперативну пам'ять для відокремлення змінених полів від всієї моделі та збереження у додатковій колекції.

### Порівняння продуктивності стандартної конфігурації та без аудіювання

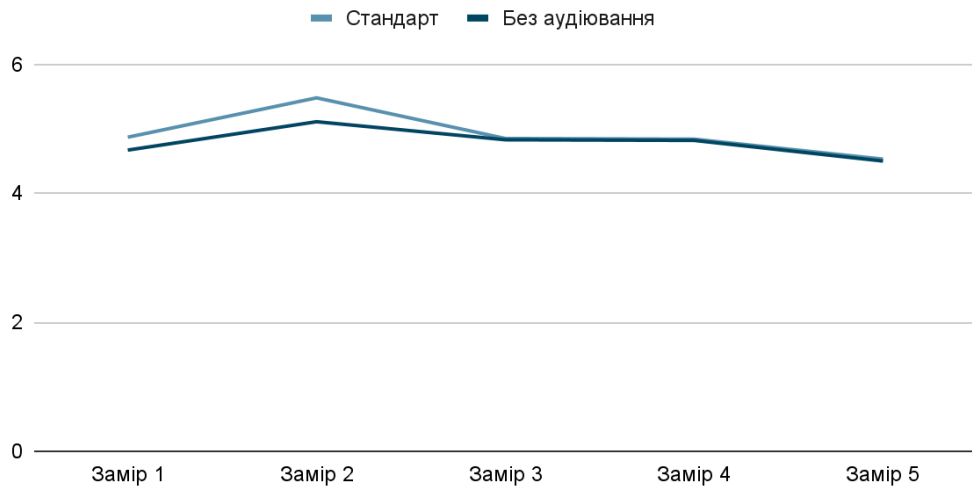


Рисунок 4.3 – Порівняння продуктивності стандартної конфігурації та без аудіювання

Як можна побачити на рисунку 4.4 вимкнення журналювання надає приріст продуктивності, але не значний.

### Порівняння продуктивності стандартної конфігурації та без журналювання

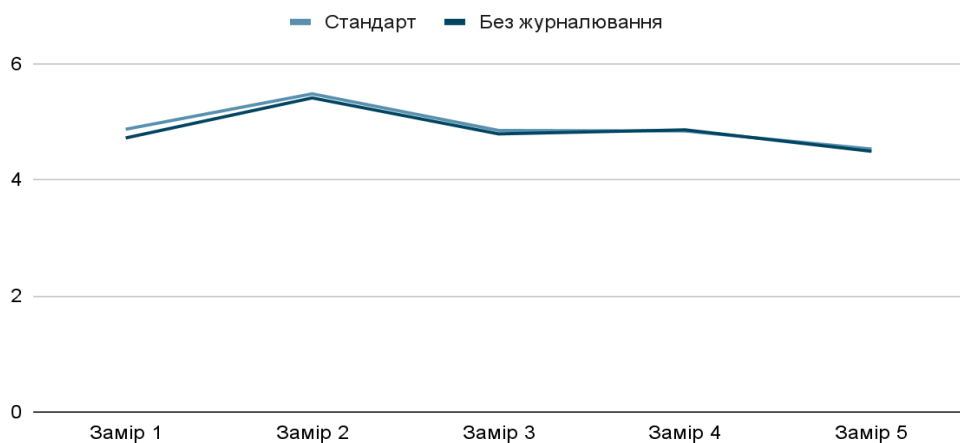


Рисунок 4.4 – Порівняння продуктивності стандартної конфігурації та без журналювання

Як можна побачити на рисунку 4.5 оновлення конфігурацій надає значний приріст продуктивності на перших запитах, та малий при подальших запитах. Підвищення продуктивності на перших запитах пов'язано із відсутністю кешу для запитів, тому відсутність додаткового навантаження на диск надає додаткову продуктивність, а з часом, коли з'являється кеш, він починає використовуватися і навантаження на сервер та диск значно спадає.

### Порівняння продуктивності стандартної конфігурації та оновленої конфігурації

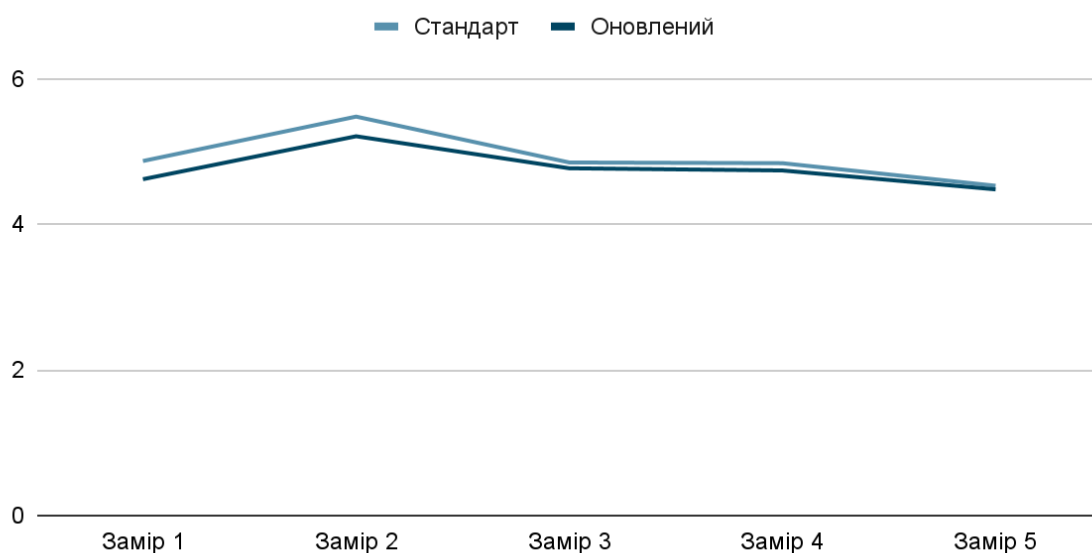


Рисунок 4.5 – Порівняння продуктивності стандартної конфігурації та оновленої конфігурації

Тобто оновлена конфігурація має містити наступні поля та їх значення:

- `storage.journal.enabled: false;`
- `auditLog.destination: null;`
- `systemLog.path: null;`
- `storage.wiredTiger.indexConfig.prefixCompression: false.`

## 4.2 Розробка методів оптимізації збереження та розташування даних

Оскільки MongoDB документоорієнтована СУБД, то для неї неможливо впроваджувати більшість практик, які використовувались до реляційних СУБД. Існує два основних підходи в створенні схеми документів – вбудування та посилання.

На рисунку 4.6 можна побачити приклад вбудування одного документу в інший, а саме документу який зберігає інформацію про автомобілі в документ який зберігає інформацію про людину.

```
1 {
2   "first_name": "Paul",
3   "surname": "Miller",
4   "cell": "447557505611",
5   "city": "London",
6   "location": [45.123, 47.232],
7   "profession": ["banking", "finance", "trader"],
8   "cars": [
9     {
10      "model": "Bentley",
11      "year": 1973
12     },
13    {
14      "model": "Rolls Royce",
15      "year": 1965
16     }
17  ]
18 }
```

Рисунок 4.6 – Приклад вбудування одного документу в інший

Перевагами вбудування є в першу чергу отримати всю інформацію одним запитом, що дозволяє не використовувати \$lookup, який є аналогом join, що значно підвищує продуктивність. Також це дозволяє оновлювати інформацію неподільними атомарними запитом. Хоча цей підхід і дозволяє використовувати транзакції, їх краще позбигати при написанні запитів, бо

для NoSQL використання транзакцій є антипатерном, бо викликає більше проблем, ніж вирішує.

Не дивлячись на переваги, головним недоліком є обмеження на розмір документу, а саме 16 мегабайт. Також при великій кількості великих документів страждає продуктивність, бо запити потребують витягати більше даних, використовувати більше пам'яті, більше ресурсів центрального процесору, та кеш біде зайнятий меншою кількістю документів, що також зменшує продуктивність з часом.

Підхід з посиланнями використовую посилання на інші документи, для цього використовуються унікальні поля, або Object ID. Перевагами цього підходу є можливість розподілювати документи на більшу кількість менших документів, що допомагає не перевищувати ліміт у 16 мегабайт, та підвищує швидкість одного запиту. Також це дозволяє не забирати зайву інформацію, на відміну від підходу із вбудованістю. Також перевагою є надання можливості зменшити кількість дублікатів документів.

Однак головним недоліком є потреба використовувати оператор \$lookup в запитах, що значно зменшує швидкість запитів.

На рисунку 4.7 можна побачити приклад створення документа з відношенням один-до-одного, де користувачу відповідає один аккаунт в соціальних мережах.

```
1 {
2   "_id": "ObjectId('AAA')",
3   "name": "Joe Karlsson",
4   "company": "MongoDB",
5   "twitter": "@JoeKarlsson1",
6   "twitch": "joe_karlsson",
7   "tiktok": "joekarlsson",
8   "website": "joekarlsson.com"
9 }
```

Рисунок 4.7 – Приклад відношення один-до-одного

Як можна побачити на рисунку 4.8, для створення відношення використовується метод вбудування. Саме метод вбудування надає можливість зробити запити більш швидкими, тому зазвичай потрібно використовувати саме цей метод, якщо це можливо.

```
1 {
2   "_id": "ObjectId('AAA')",
3   "name": "Joe Karlsson",
4   "company": "MongoDB",
5   "twitter": "@JoeKarlsson1",
6   "twitch": "joe_karlsson",
7   "tiktok": "joekarlsson",
8   "website": "joekarlsson.com",
9   "addresses": [
10    { "street": "123 Sesame St",
11      "city": "Anytown", "cc": "USA" },
12    { "street": "123 Avenue Q",
13      "city": "New York", "cc": "USA" }
14  ]
15 }
```

Рисунок 4.8 – Приклад документа з відношенням один-до-декількох

На рисунку 4.9 можна побачити приклад відношення один-до-багатьох. Це відношення має багато спільного з один-до-декількох, але головною відмінністю є використання методу посилання. Масив `parts` зберігає айді інших документів, які пов'язані із ним. Хоча досить часто трапляються ситуації, коли потрібно використовувати посилання, однак при можливості такого підходу варто уникати.

Як можна побачити на рисунку 4.10 відношення багато-до-багатьох відбувається через створення відношення. У цьому випадку використання відношень вирішує необхідну задачу, в той час як вбудування не завжди здатне створити та підтримувати багато-до-багатьох через обмеження на розмір документа в 16 мегабайт.

```

1  {
2      "name": "left-handed smoke shifter",
3      "manufacturer": "Acme Corp",
4      "catalog_number": "1234",
5      "parts": [
6          "ObjectID('AAAA')",
7          "ObjectID('BBBB')",
8          "ObjectID('CCCC')",
9      ]
10 }

```

Рисунок 4.9 – Приклад відношення один-до-багатьох

```

1  {
2      "_id": ObjectID("AAF1"),
3      "name": "Name",
4      "tasks": [ObjectID("ADF9"),
5              ObjectID("AE02"), ObjectID("AE73")]
6  }
7
8  {
9      "_id": ObjectID("ADF9"),
10     "description": "MongoDB",
11     "due_date": ISODate("2014-04-01"),
12     "owners": [ObjectID("AAF1"), ObjectID("BB3G")]
13 }

```

Рисунок 4.10 – Приклад відношення багато-до-багатьох

Незважаючи на те, що ці рекомендації та підходи потрібно використовувати, але найбільш оптимальна схема може бути створена тільки при створенні для конкретного програмного додатку.

Тим не менш можна виділити 5 основних правил створення схеми даних:

- потрібно використовувати вбудовування, якщо немає вагомих причин використовувати посилання;

- якщо необхідно мати доступ до самостійного документа, то це може бути причиною створити посилання;
- потрібно уникати використання \$lookup;
- потрібно використовувати посилання, коли масив з додатковими документами може перевищувати сотню документів;
- при побудові схеми потрібно базуватися на вимогах програмного додатку.

### 4.3 Розробка методів оптимізації запитів

Хоча налаштування конфігурацій та схеми даних на надають значних переваг в продуктивності, вони необхідні для подальшої оптимізації запитів, саме оптимізація запитів і надає найбільший приріст продуктивності.

Найбільш часто використовуваними запитами є запити на отримання даних та запити на додавання нових даних, тому саме їх оптимізації і будуть досліджені.

В першу чергу слід розглянути оптимізацію саме запитів на отримання даних, адже зазвичай саме запитів на отримання даних виконується більше всього.

Найголовнішим з чого треба почати це індекси. Саме індекси, як і в реляційні СУБД, зазвичай надають значний приріст продуктивності. Індокси в NoSQL СУБД мало чим відрізняються від індоксів в реляційних СУБД, тому для вибору полів для індоксування можна використовувати реляційні правила, а саме:

- якщо передбачається пошук по полю;
- якщо передбачається сортування по полю;
- якщо поле має багато різних значень. При великій кількості однакових значень індоксування може надати незначний, або зовсім не надати, прирості продуктивності;

– якщо поле має тип короткого тексту, числа, бульовий тип, або тип дати-часу;

– якщо поле має тип довгого тексту, то індекси слід використовувати лише у разі необхідності виконання повнотекстного пошуку, за відсутності змоги підключення третього стороннього програмного забезпечення.

На рисунку 4.11 можна побачити результати порівняння запиту на отримання даних по фільтру { 'name': "Ocean View Waikiki Marina w/prkg" } з використанням індексації для цього поля та без індексації на 50 гігабайтах даних, запит отримує 527 документів, запит відбувається на локальному сервері. Після підключення індексації продуктивність в середньому збільшилася на 21%.

Порівняння запитів пошуку з індексом та без індекса

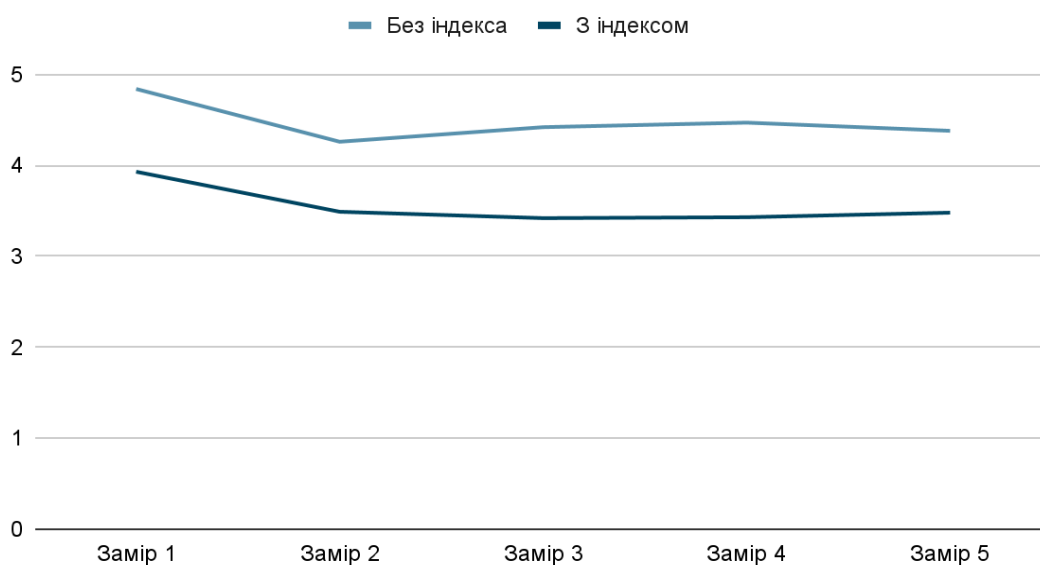


Рисунок 4.11 – Порівняння запитів пошуку з індексом та без індексу

Хоча індекси і надають додаткові переваги в швидкості виконання запитів, вони сприяють збільшенню розміру бази даних, адже збереження індексу потребує додаткового місця на диску, індекс займає стільки ж місця, скільки займає і оригінальне поле. Про це слід пам'ятати особливо

при створенні індексів на текстові поля. Зазвичай найбільшого приросту в продуктивності у пошуку по текстовим полям надають третьосторонні програмні засоби, такі як, наприклад, Elastic Search.

Незважаючи на те, що саме запити найбільш завантажують диск та центральний процесор, також важливим ресурсом залишається мережа. Саме мережа, її якість та швидкість передачі даних дуже часто безпосередньо впливає не стільки на швидкість виконання запиту, як на швидкість отримання даних. Тим самим, при низькій швидкості мережі, сервер із СУБД потребує додаткового часу на утримання даних для відправки на своїй стороні, що в свою чергу зменшує корисне навантаження на процесор та пам'ять та зменшує швидкість інших запитів, які виконуються на сервері в цей час.

Іноколи можливість зробити мережу із високою пропускнуою здатністю відсутня, в таких випадках, для підвищення продуктивності роботи сервера потрібно обмежувати кількість полів документів, які отримуються, або кількість документів які отримуються.

Для обмеження кількості полів в документі потрібно передати JSON об'єкт із переліком необхідних полів та вказати 1 для полів, які потрібно отримати, та 0 для полів які не треба отримувати. Наприклад запит `{"name":1,"_id":0}` поверне тільки поле name, та прибере поле \_id.

Для обмеження кількості документів в запиті потрібно використовувати `limit(N)`, де N це кількість документів які потрібно отримати. Ця операція обере перші N документів та поверне їх у запиті.

Обмежувати кількість полів та кількість документів у запитах краще у всіх можливих випадках, адже це дозволяє менше навантажувати процесор при розрахунках та використовувати менше даних пам'яті при виконанні пошуку. Але також варто пам'ятати, що якщо існує багато запитів з унікальним складом полів, то це заважає кешуванню, що робить запити значно повільнішими, а також впливає на швидкість інших запитів, бо займає простор кешу, який більше не буде використовуватися та серверу

знадобіться додатковий час на те, щоб перевірити кеш, виконати запит та записати його в кеш, який більше не знадобіться.

Як можна побачити на рисунку 4.12 використання обмежених документів надає в середньому близько 3% продуктивності. В пошуку було додано фільтр на отримання тільки параметра name. Продуктивність оптимізації може змінюватися залежно від фізичної відстані до сервера та якості мережі. Ця оптимізація надає незначні переваги в виконанні одиничних запитів, але здатна зменшити навантаження на сервер, що дозволяє підвищувати швидкість окремих важких запитів, які виконуються в цей час на сервері, через зменшення ресурсів процесу і пам'яті на виконання цього запиту.

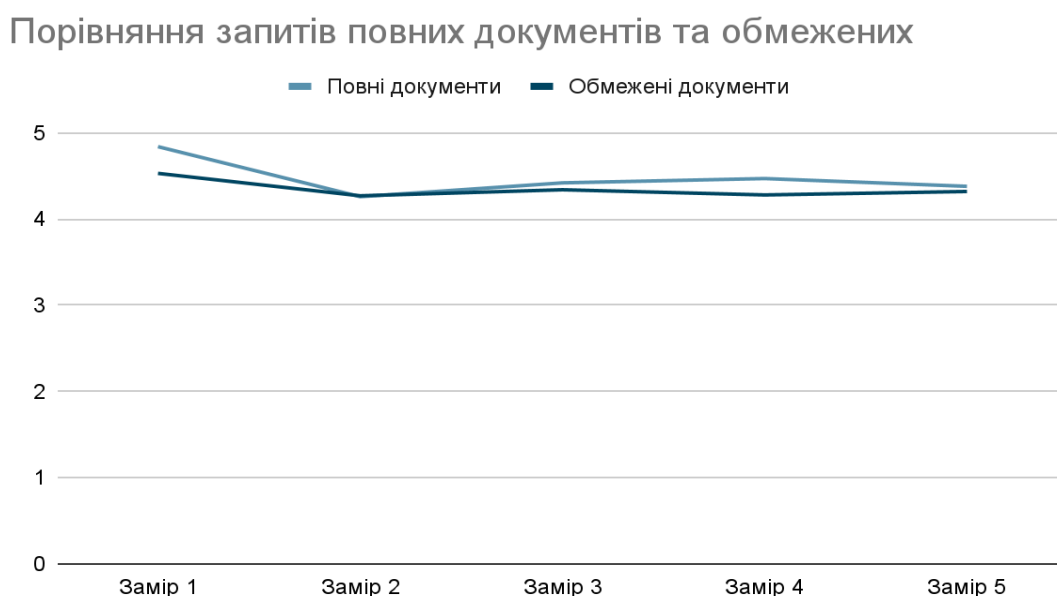


Рисунок 4.12 – Порівняння запитів на отримання повних документів та обмежених

Для виконання запитів на оновлення чисельних даних, замість отримання даних із сервера бази, оновлення в програмному додатку та ще однієї загрузки, можна використовувати оператор \$inc. Він виконується під

час запиту на отримання даних та оновлює чисельні дані в документах, які підпадають під фільтр. Це дозволяє не тільки виконувати менше запитів на сервер, що значно зменшує навантаження, але і оновлювати дані найбільш ефективним методом, адже `$inc` є вбудованою функцією та має власні процеси в пайплайні виконання запитів, що дозволяє оновлювати дані не обмежуючи інші запити, які виконуються в той самий час.

Однією з головних особливостей MongoDB є огорнуті запити. Цей механізм дозволяє отримувати доступ до даних, не виконуючи запит до сховища, замість цього використовується безпосередній доступ до значень індексу. Тобто для виконання огорнутого запиту потрібно щоб всі поля, які будуть отримані мали індекси.

На рисунку 4.13 можна побачити порівняння звичайного запиту та огорнутого запиту. Використання огорнутого запиту надало 71% зростання продуктивності.

#### Порівняння звичайного і огорнутого запиту

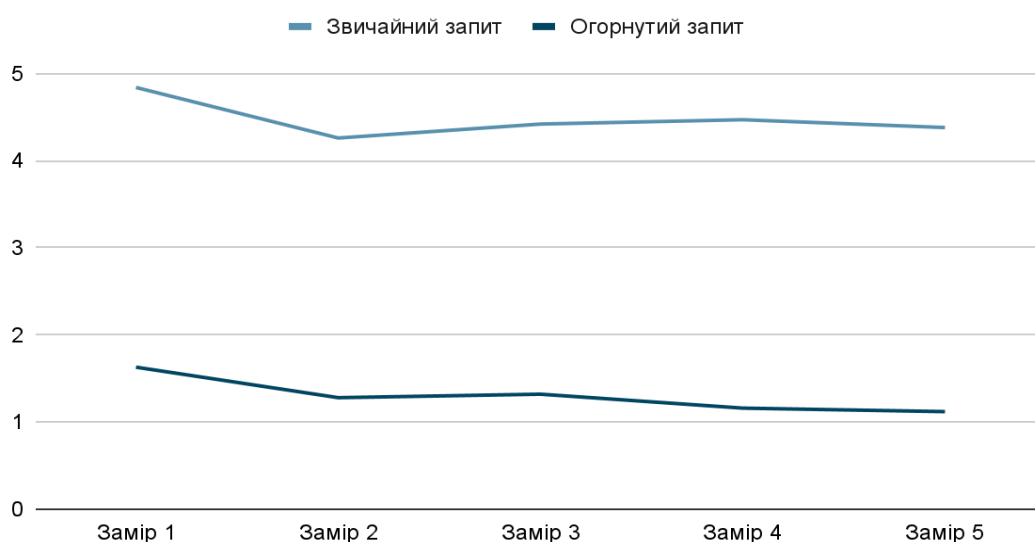


Рисунок 4.8 – Порівняння звичайного і огорнутого запиту

Отже, хоча інтерфейс MongoDB і має обмежений функціонал на виконання запитів, ця суюд надає можливість майже миттєво виконувати запити на отримання даних при правильному розподіленні індексів.

Збільшуючи кількість індексів значно зростає швидкість виконання запитів на отримання даних, але з іншого боку і значно падає швидкість додавання нових даних, оновлення існуючих, та видалення старих. Це відбувається через потребу створювати, змінювати, або видаляти індекси та перебудовувати дерево індексів.

На рисунку 4.14 можна побачити порівняння швидкості створення 5555 нових документів загальною вагою 97 мегабайт. В першому випадку було заіндексовано лише поле `_id`, яке є обов'язковим та генерується автоматично, в другому випадку були заіндексовані майже всі поля, а саме 42 поля. Загалом середнє зменшення продуктивності склало 32%.

Порівняння створення нових документів з індексами та без

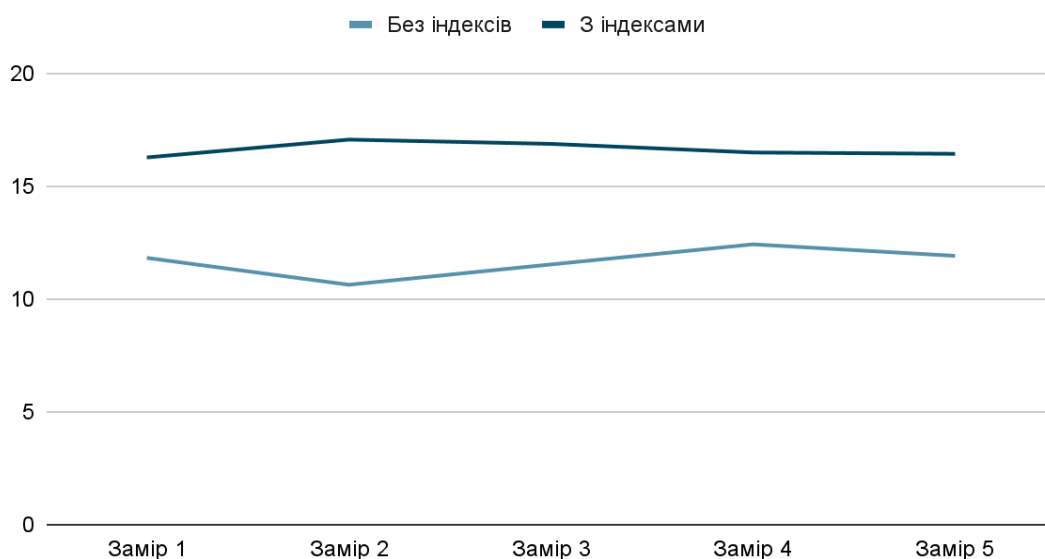


Рисунок 4.14 – Порівняння створення нових документів з індексами та без

Окрім зменшення кількості індексів, для підвищення продуктивності запису, існує не багато методів. Один з них це використання компресії великих документів при використанні посилань.

Оскільки \$lookup потребує багато ресурсів для свого виконання, але часом без нього не обійтись може виникати потреба. В таких випадках для підвищення продуктивності є сенс розмістити повні документи в масив замість посилань та зробити компресію для зменшення об'єму даних для подальшої декомпресії в програмному додатку. Для підвищення швидкості декомпресії потрібно додати додаткове поле з розміром скомпресованої частини даних.

При використанні алгоритму zstd можна досягти високого рівня компресії, не витрачаючи на це багато часу, тому це є найбільш ефективним вибором (рисунок 4.15).

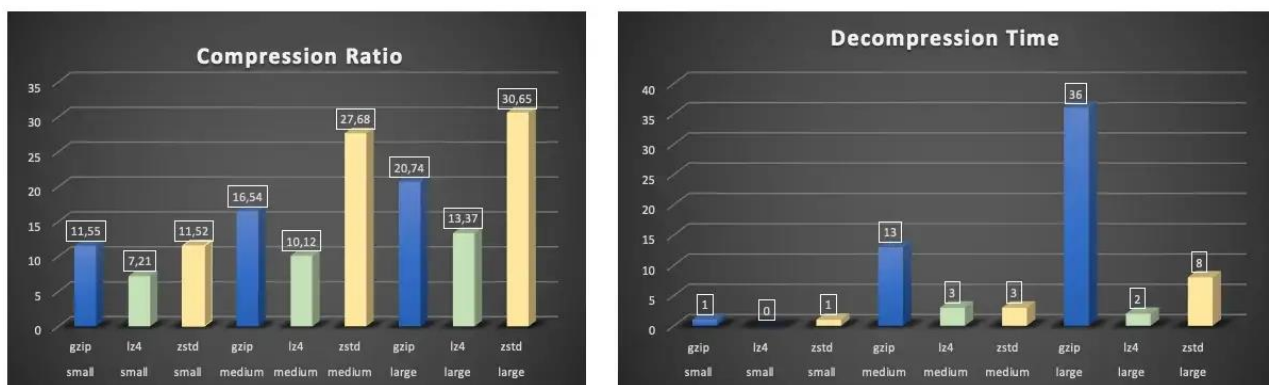


Рисунок 4.15 – Порівняння алгоритмів компресії за рівнем компресії (ліво) та часом компресії (право)

На рисунку 4.16 зображено приклад порівняння виконання запитів на отримання даних з використанням посилань та \$lookup та приклад запитів з використанням компресії замість посилань. Виконання запитів на отримання даних з компресією швидше на 52%. Для виконання запитів було створено 50 гігабайт даних, де кожен документ посилається на 100 інших.

### Порівняння отримання даних з посиланнями та компресією

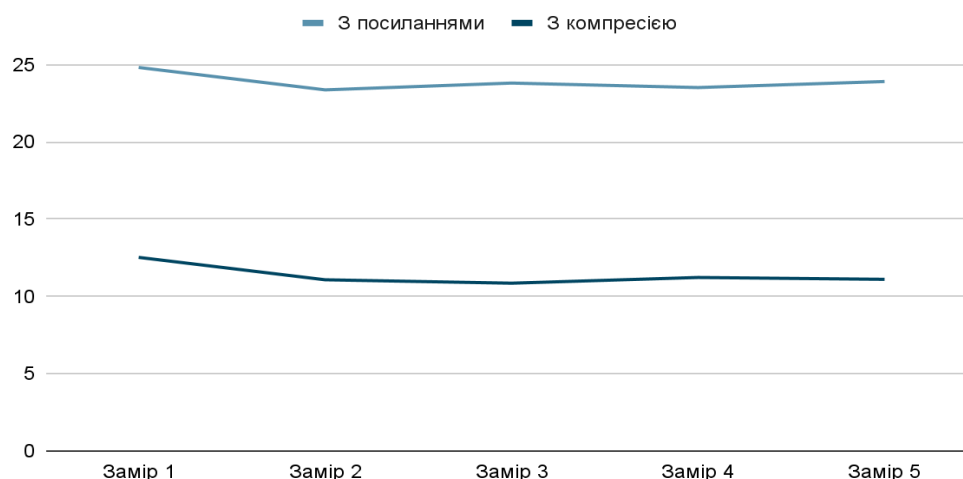


Рисунок 4.16 – Порівняння отримання даних з посиланнями та компресією

Окрім підвищення швидкості запитів на отримання даних використання компресії підвищує швидкість запису, бо позбавляє потреби робити додаткові запити в базу.

Також ефективним засобом підвищення ефективності запису є використання батч операцій. Вони дозволяють одним запитом додавати велику кількість записів до бази, тим самим задіюючи внутрішні пайплайни сервера з максимальною ефективністю, позбавляючи від потреби перерозподіляти багато ресурсів на виконання цієї операції з виконання запитів на зчитування даних.

На рисунку 4.17 зображено порівняння звичайного запиту на запис 5555 документів, та батч запиту. Батч запит в середньому на 35% швидше. Для реалізації батч запиту було застосовано офіційну бібліотеку для C#.

Також інколи трапляються випадки, коли потрібно видалити документи, для зменшення навантаження на сервер цей процес можна винести в окремий програмний додаток та застосувати стратегію помітки-підмітання.

## Порівняння звичайних та батч запитів

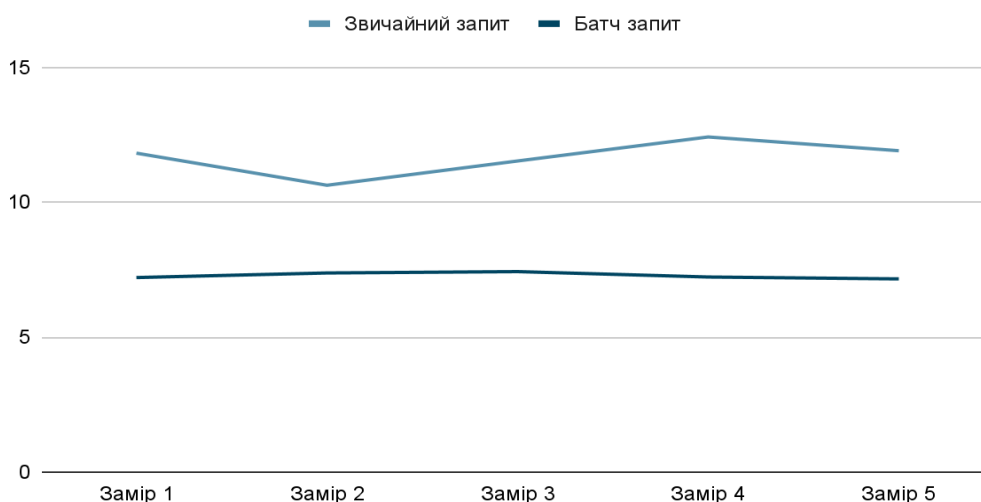


Рисунок 4.17 – Порівняння звичайних та батч запитів

Ця стратегія полягає в не миттєвому видаленні документів, а лише в створенні в них нового поля, яке буде відповідати за те, що цей документ готовий до видалення. Потім, в запланований час запускається окремий процес, який за допомогою батч методів видалляє всі відмічені документи одним запитом.

Завдяки цьому, як і у випадку із батч записом, вдається зменшити навантаження на сервер та задіяти вбудовані пайплайни, що дозволить ефективніше виконувати інші запити.

Отже, для підвищення швидкості запитів на отримання даних основним шляхом є створення індексів, але в той самий час це ускладнює додання нових даних та зміну вже існуючих. Для знаходження оптимального балансу в індексах потрібно в першу чергу дивитися на потреби, які буде вирішувати база.

## ВИСНОВКИ

У результаті виконання роботи було досліджено та розроблено методи оптимізації отримання доступу до великих даних. Це дозволило оволодіти професійним досвідом та покращили навички володіння базами даних та поглибило знання у сфері великих даних.

Було проведене ознайомлення з основними напрямками використання великих даних, їх потребами, перспективами та потенційними проблемами.

Були набуті практичні знання дослідження та реалізації методів оптимізації методів отримання доступу до великих даних.

Був проведений збір матеріалів, які стосуються теми роботи – «Дослідження та розробка методів оптимізації збереження та отримання доступу до великих даних для NoSQL сховищ».

Були вивчені сучасні потреби сектору великих даних.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

- 1) Офіційний сайт MongoDB. URL: <https://www.mongodb.com/>(дата звернення 23.11.2022).
- 2) Офіційний сайт C# URL: <https://learn.microsoft.com/en-us/dotnet/csharp/>(дата звернення 23.11.2022).
- 3) Офіційний сайт JSON URL: <https://www.json.org/json-en.html>(дата звернення 24.11.2022).
- 4) Сторінка вікіпедії про MongoDB URL: <https://uk.wikipedia.org/wiki/MongoDB>(дата звернення 24.11.2022).
- 5) Сторінка вікіпедії про C# URL: [https://en.wikipedia.org/wiki/C\\_Sharp\\_\(programming\\_language\)](https://en.wikipedia.org/wiki/C_Sharp_(programming_language))(дата звернення 25.11.2022).
- 6) Сторінка вікіпедії про JSON URL: <https://en.wikipedia.org/wiki/JSON>(дата звернення 25.11.2022).
- 7) Зразки датасетів URL: <https://github.com/neelabalan/mongodb-sample-dataset>(дата звернення 26.11.2022).
- 8) Сторінка вікіпедії про великі дані URL: <https://uk.wikipedia.org/wiki/%D0%92%D0%B5%D0%BB%D0%B8%D0%BA%D1%96%D0%B4%D0%B0%D0%BD%D1%96>(дата звернення 26.11.2022).
- 9) Філатов В. О. Реінжинірінг інформаційної системи на підставі пошуку і аналізу функціонально-залежних атрибутів реляційної бази даних / В. О. Філатов, О. В. Золотухін // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: тези доповідей XXVII міжн. наук.-практ. конф. MicroCAD-2019, 15-17 травня 2019р. – Харків: НТУ «ХПІ». – 2019. – С. 201.
- 10) Filatov, V. Model of semantic integration of information systems properties in relay database reengineering problems / O. Avrunin, O. Vlasov, V. Filatov. // Innovative Technologies and Scientific Solutions

for Industries, 2020 - (4 (14)), 5-12. doi:10.30837/itssi.2020.14.005

11) Filatov, V., Semenets, V. and Zolotukhin, O. (2020) “Data mining in relational systems”, Innovative Technologies and Scientific Solutions for Industries, (3 (13), pp. 65-76. doi: 10.30837/ITSSI.2020.13.065

12) Филатов, В.А. Методы и средства проектирования информационных систем и распределенных баз данных / В.А. Филатов, Р.В. Семенец. // Вестник Херсонского национального технического университета № 4(27) – 2007. – С. 203-207.

13) Filatov V., Voloshchuk O., Spivak N. Implementation and support fuzzy systems by means the relational data model, «European Cooperation», Vol 4, No 11 (2016). – P. 49-61.

14) Doskalenko S. N. On the Approach to Searching for Functional Dependences of Data in Relational Systems / S. N. Doskalenko, V. A. Filatov. // Innovative technologies and scientific solutions for industries. Kharkiv. 2018. No. 3 (1). P. 54–58. doi:10.30837/2522-9818.2018.3.054

15) Filatov V. Fuzzy models presentation and realization by means of relational systems // Econtechmod : an international quarterly journal on economics in technology, new technologies and modelling processes. – Lublin ; Rzeszow, 2014. – Vol.(3), № 3. – P. 99-102.

16) Филатов В.А. Модель мультиагентной системы автономного администрирования информационных систем и распределенных баз данных / В.А. Филатов, Е.Е. Цыбульник, Л.Э. Чалая // Новости искусственного интеллекта. – 2002. – № 4. – С. 620-627.