

ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

(повне найменування вищого навчального закладу)

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ

(повне найменування інституту, назва факультету (відділення))

КАФЕДРА ПРОГРАМНИХ ЗАСОБІВ І ТЕХНОЛОГІЙ

(повна назва кафедри (предметної, циклової комісії))

Пояснювальна записка

до кваліфікаційної роботи

магістра

(освітній рівень)

на тему: «Дослідження методів масштабування розподілених
високонавантажених фінансових застосунків, що працюють в режимі
реального часу»

Виконав: студент 2 курсу, групи 6ПР
спеціальності

121 - «Інженерія програмного забезпечення»

(шифр і назва спеціальності)

Хмель Микита Олегович

(прізвище та ініціали)

Керівник к.т.н., доцент Ляшенко О. М.

(прізвище та ініціали)

Рецензент к.т.н., доцент Григорова А.А.

(прізвище та ініціали)

Херсонський національний технічний університет
(повне найменування вищого навчального закладу)

Факультет, відділення Інформаційних технологій та дизайну

Кафедра Програмних засобів і технологій

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення
(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Програмних засобів і технологій

к.т.н. доц. О.Є. Огнева

“ ___ ” _____ 2023 р.

З А В Д А Н Н Я

НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА

Хмель Микита Олегович

(прізвище, ім'я, по батькові)

Тема роботи «Дослідження методів масштабування розподілених високонавантажених фінансових застосунків, що працюють в режимі реального часу»

керівник роботи к.т.н. доцент Ляшенко. О. М.,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджена наказом вищого навчального закладу від 29.09.2023 р. № 509-С

2. Строк подання студентом роботи 01.12.2023

3. Вихідні дані до роботи літературні та періодичні джерела, матеріали науково-дослідної практики.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):

1) Аналіз предметної області. 2) Дослідження існуючих технологій та рішень описаної проблеми. 3) Збір даних та їх глибокий аналіз.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Огляд літературних джерел	17.09.2023	Виконано
2	Аналіз предметної області, детальне формування проблематики. Дослідження та вивчення існуючих технологій та рішень описаної проблеми	22.09.2023	Виконано
3	Накопичення та структуризація актуальних даних про пов'язані прецеденти у сфері бізнесу.	25.09.2023	Виконано
4	Проведення консультацій з авторами джерел та дотичними до кваліфікаційної роботи магістра працівників освіти.	28.09.2023	Виконано
5	Розроблення архітектури програмного застосунку	01.10.2023	Виконано
6	Розроблення макетів інтерфейсу користувача	30.10.2023	Виконано
7	Налагодження програмного застосунку	19.11.2023	Виконано
8	Складання програмної документації	11.12.2023	Виконано

Студент

(підпис)

Хмель М.О.

(прізвище та ініціали)

Керівник роботи

(підпис)

Ляшенко О.М.

(прізвище та ініціали)

АНОТАЦІЯ

У контексті постійного розвитку світу фінансових транзакцій та послуг у реальному часі, ця кваліфікаційна робота магістра досліджує важливі аспекти ефективності, надійності та швидкості обробки даних у фінтех-галузі. З урахуванням зростання попиту на фінансові послуги в реальному часі від нових учасників, таких як Facebook, Google та інших, вивчається роль хмарної інфраструктури в низькозатримкових транзакціях.

Основною спрямованістю роботи є проблеми високочастотної фінансової торгівлі та автоматизованої торгівлі, де мільйони доларів можуть бути втрачені через найменші затримки.

Обґрунтовується необхідність багаторегіонального розгортання та георозбиття даних у фінансових установах, які мають операції в різних регіонах. Висвітлюється роль технологій, таких як CockroachDB, Apache Spark та інших, які пропонують георозбиття для ефективного розгортання в кількох регіонах, забезпечуючи максимальну швидкість та конфіденційність даних.

Детально розглядаються виклики, пов'язані із стрімким зростанням обсягу фінансових даних, використанням передових технологій для оптимізації обчислень, а також проблеми аналітики великих даних. Зазначається, що затримка стає критичним елементом у вирішенні завдань, пов'язаних з великим обсягом фінансових даних у реальному часі.

В роботі висвітлюється також роль глибокого навчання у фінансовій аналітиці, висунуті вимоги до програмного/апаратного забезпечення, а також розглядаються методи масштабування для оптимізації ефективності обчислень і досягнення наднизьких затримок. Завдяки детальному аналізу наведених викликів із затримкою в обробці фінансових даних, робота надає важливий внесок у розуміння та розв'язання цих проблем у фінтех-галузі.

Ключові слова: фінтех, низькозатримкова обробка, георозбиття даних, великі дані, глибоке навчання, абсолютна затримка, IoT, ШІ.

ЗМІСТ

ВСТУП	7
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	10
1.1 Обґрунтування актуальності дослідження	10
1.2 Аналіз проблем і викликів	11
1.2.1 Визначення прогалин або можливостей для дослідження	13
1.3. Аналіз останніх досліджень і публікацій	16
ВИСНОВКИ ДО РОЗДІЛУ 1	20
2 ПІДХОДИ ДЛЯ ОПТИМІЗАЦІЇ ЕФЕКТИВНОСТІ ОБЧИСЛЕНЬ ДЛЯ ДОСЯГНЕННЯ НАДНИЗЬКИХ ЗАТРИМОК	22
2.1 Огляд багаторівневого архітектурного підходу аналізу критично залежних від затримок фінансових даних	22
2.2 Центри обробки великих фінансових даних	23
2.3 Використання віртуальних машин для підтримки фінтех застосунків	25
2.4 Новаторські підходи до обробки великих даних з використанням суперкомп'ютерів	30
ВИСНОВКИ ДО РОЗДІЛУ 2	33
3 ВПРОВАДЖЕННЯ ТА РЕАЛІЗАЦІЯ ОПТИМІЗАЦІЙНИХ ПІДХОДІВ	35
3.1 Порівняльний аналіз з існуючих технологій	35
3.1.1 Великі дані та Apache Spark	37
3.1.2 Apache Hadoop та інші пов'язані технології	43
3.2 Дослідження етапів та технічних аспектів впровадження	47
3.2.1 Використання фреймворку MapReduce	51
3.2.2 Огляд стійкої та георозподіленої бази даних CockroachDB	55
3.3 Впровадження шляхом моделювання розповсюджених явищ у фінтех додатках розподіленого типу	59
3.3.1 Передача великого дата-сету фінансових даних для моделювання явищ	60

3.3.2 Розгортання та налаштування екземпляру Spark у хмарному середовищі	62
3.3.3 Результати обробки фінансових даних в Spark	63
3.4 Висновки щодо ефективності та впроваджуваності	66
ВИСНОВКИ ДО РОЗДІЛУ 3	68
ВИСНОВКИ	70
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	72
ДОДАТОК А	76
ДОДАТОК Б	81
ДОДАТОК В	87

ВСТУП

Сучасний світ фінансових транзакцій та фінансових послуг постійно розвивається, а разом із цим зростають вимоги до ефективності, надійності та швидкості обробки даних. Клієнти потребують багатофункціональні фінтех-додатки, які працюють в режимі реального часу та забезпечують стійку обробку транзакцій, незалежно від часових обмежень. З появою нових гравців у галузі фінтеху, таких як Facebook, Google, Betterment та Robinhood, з'являється зростаючий попит на хмарну інфраструктуру, яка має низьку затримку та може витримувати великі обсяги транзакцій. У контексті фінансових послуг, де кожна мілісекунда має величезне значення, низька затримка стає найважливішою складовою. Втрати прибутку через затримки або простої можуть вирівнювати мільйони доларів. Ця проблема особливо актуальна в сфері високочастотної фінансової торгівлі та автоматизованої торгівлі, де кожної мілісекунди робиться мільйони рішень, і швидкість отримання даних є критичною.

Не менш важливою є потреба в мінімізації затримки в інших сегментах фінансових послуг, таких як криптовалюта, периферійні обчислення, штучний інтелект та машинне навчання. Зниження затримки дозволяє фахівцям з обробки даних приймати обґрунтовані рішення в режимі реального часу, а банкам виявляти шахрайство негайно. Для розв'язання цієї проблеми важливо розглядати багаторегіональне розгортання та георозбиття даних. Багато фінансових установ мають операції в різних регіонах та країнах, і вони повинні забезпечувати доступ до даних і послуг з мінімальною затримкою.

Розгортання в кількох регіонах стає важливим рішенням для досягнення високої продуктивності та забезпечення надійності. Однак ця стратегія може бути складною та вимагати значних ресурсів для керування. Щоб вирішити ці складні завдання, розробники і фахівці з обробки даних повинні використовувати передові технології, такі як CockroachDB, яка пропонує георозбиття для розгортання в кількох регіонах. Це дозволяє організаціям

зберігати дані найближче до користувачів, забезпечуючи максимальну швидкість і виконуючи найвищі стандарти захисту даних і конфіденційності. Проблема зниження затримки в фінансових послугах стає все більш актуальною, і розглядані рішення допомагають забезпечити успішну роботу та конкурентоспроможність у цьому секторі.

У даній кваліфікаційній роботі магістра досліджені та структуровані актуальні на 2023 рік методи і підходи масштабування розподілених фінтех додатків, георозділення їх даних та зниження затримок при високих навантаженнях, а також висунуті обґрунтовані пропозиції щодо покращення їх окремих частин.