МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Факультет програмної інженерії та бізнесу

Кафедра інженерії програмного забезпечення

Пояснювальна записка до дипломної роботи

магістра
(освітній ступінь)

на тему: «Експериментальне дослідження методів балансування навантаження високонавантажених комп'ютерних систем»

ХАІ.603.667П1.22О.121.9580854 ПЗ

Виконала: студентка 6 курсу групи 667П1
Галузь знань 12 «Інформаційні технології»
Спеціальність 121 «Інженерія програмного забезпечення»
(код та найменування)
Освітня програма Інженерія програмного забезпечення
(найменування)
Кир'ян О. П.
(прізвище й ініціали студента)
Керівник: Феоктистова О. І.
(прізвище й ініціали)
Рецензент: Ляшенко О. С.
(прізвище й ініціали)

Харків – 2022

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Факультет програмної інженерії та бізнесу
(повне найменування)
Кафедра інженерії програмного забезпечення
(повне найменування)
Рівень вищої освіти другий (магістерський)
Галузь знань 12 «Інформаційні технології»
Спеціальність 121 «Інженерія програмного забезпечення»
(код та найменування)
Освітня програма «Інженерія програмного забезпечення»
(найменування)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Ігор ТУРКІН
(підпис) (ім'я та прізвище)
«___» _____ 2022 р.

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТЦІ

Кир'ян Олені Петрівні
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи «Експериментальне дослідження методів балансування навантаження високонавантажених комп'ютерних систем»
керівник кваліфікаційної роботи Феоктистова Олена Ігорівна, к.т.н, доцент
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом Університету № _____ від "____" _____ 2022 року
2. Термін подання студенткою кваліфікаційної роботи _____
3. Вихідні дані до кваліфікаційної роботи: результати огляду та аналізу проблем високонавантажених комп'ютерних систем
4. Зміст пояснювальної записки (перелік питань, які потрібно розробити):

1) огляд та аналіз проблем високонавантажених комп'ютерних систем;
2) планування експерименту для досліджень розвантаження високонавантажених систем;
3) аналіз результатів експерименту для досліджень розвантаження високонавантажених систем та формування практичних рекомендацій.

5. Перелік графічного матеріалу: пояснювальна записка – 78 сторінок, 23 рисунка, 5 таблиць, 37 джерел, 20 слайдів презентації

6. Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Феоктистова О. І., доцент каф. 603		
2	Феоктистова О. І., доцент каф. 603		
3	Феоктистова О. І., доцент каф. 603		

Нормоконтроль _____ Вероніка ПОСТЕРНАКОВА «11» грудня 2022 р.
(підпис) (ім'я та прізвище)

7. Дата видачі завдання 09.09.2021 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	Отримання і затвердження теми роботи	09.09.21	
2	Огляд та аналіз проблем високонавантажених комп'ютерних систем	10.09.21 - 20.12.21	
3	Планування експерименту для досліджень розвантаження високонавантажених систем	15.02.22 - 26.05.22	
4	Аналіз результатів експерименту для досліджень розвантаження високонавантажених систем та формування практичних рекомендацій	02.09.22 - 23.10.22	
5	Наповнення контентом розділів пояснювальної записки	24.10.22 - 10.11.22	
6	Форматування пояснювальної записки до дипломної роботи відповідно до правил та вимог оформлення письмової роботи Національного аерокосмічного університету ім. М. Є. Жуковського «ХАІ»	11.11.22 - 24.11.22	
7	Передзахист дипломної роботи	09.12.22	
8	Захист дипломної роботи	14.12.22	

Здобувач _____ Олена КИР'ЯН
(підпис) (ім'я та прізвище)

Керівник кваліфікаційної роботи
_____ Олена ФЕОКТИСТОВА
(підпис) (ім'я та прізвище)

РЕФЕРАТ

Дипломна робота на тему «Експериментальне дослідження методів балансування навантаження високонавантажених комп’ютерних систем»: 78 сторінок, 23 рисунка, 5 таблиць, 37 джерел.

Об'єкт дослідження – процеси, що протікають у високонавантажених системах типу Google, Facebook, YouTube.

Предмет дослідження – використання методів статистичного аналізу, а саме методів планування експерименту для порівняльного аналізу алгоритмів балансування навантаження Round Robin та Least Connection.

Мета – аналіз ефективності високонавантажених систем на прикладі використання методів балансування навантаження.

Мета роботи полягає в використанні методів: статистичного аналізу, планування експерименту для аналізу ефективності високонавантажених систем.

На етапі аналізу проблеми проаналізовані високонавантажені системи: розглянуті відмінні риси «high load» систем, проаналізовано проблеми високонавантажених систем, розглянуті основні концепції «high load» систем, розглянуті основні технологічні рішення таких високонавантажених систем, як Google, Facebook і YouTube.

Актуальність даної роботи полягає в тому, що зростає значимість інформаційних технологій в сучасному світі, відповідно зростає чисельність відвідувачів веб-додатків і питання забезпечення відмово стійкості при високих навантаженнях встає на передній план. Таким чином, на сьогоднішній день стоїть завдання розробки оптимального програмного коду, організації баз даних, удосконалення апаратних ресурсів для підвищення швидкодії систем.

ВИСОКОНАВАНТАЖЕНІ СИСТЕМИ, КОНЦЕПЦІЇ HIGHLOAD, ПЛАТФОРМИ GOOGLE, FACEBOOK, YOUTUBE

ABSTRACT

Course thesis on the topic “Study of software features of high-load systems”: 78 pages, 23 figures, 5 tables, 37 sources.

The object of the research is high-load systems such as Google, Facebook and YouTube.

The study subject is technologies for developing high-load systems using examples of Google, Facebook and YouTube.

The goal is to analyze the efficiency of high-load systems using examples of Google, Facebook and YouTube.

The purpose of the work is based on the use of methods: statistical analysis, experimental planning methods for analyzing the effectiveness of high-load systems.

At the problem analysis stage, high-load systems were analyzed: the distinctive features of high-load systems were considered, the problems of high-load systems were analyzed, the basic concepts of high-load systems were considered, the main technological solutions of high-load systems such as Google, Facebook and YouTube were considered.

The relevance of this work is that the importance of information technologies in the modern world increases, the number of visitors of web applications is increasing accordingly, and the issue of ensuring resiliency under high loads comes to the fore. So, today there is still and issue with of developing an optimal software, organizing databases, improving hardware resources to improve the speed of systems.

HIGH-LOAD SYSTEMS, HIGH-LOAD CONCEPTS, GOOGLE PLATFORM, FACEBOOK, YOUTUBE

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ 8
ВСТУП 9
1 ОГЛЯД ТА АНАЛІЗ ПРОБЛЕМ ВИСОКОНАВАНТАЖЕНИХ КОМП'ЮТЕРНИХ СИСТЕМ 11
1.1 Відмінні риси високонавантажених систем «High load» 11
1.2 Проблеми високонавантажених інформаційних систем 13
1.3 Основні концепції «high load» систем 16
1.4 Огляд використовуваних технологій в високо навантажених системах 21
1.4.1 Технології Google 21
1.4.2 Технології Facebook 27
1.4.3 Технології YouTube 30
1.5 Висновки з розділу 1 33
2 ПЛАНУВАННЯ ЕКСПЕРЕМЕНТУ ДЛЯ ДОСЛІДЖЕНЬ РОЗВАНТАЖЕННЯ ВИСОКОНАВАНТАЖЕНИХ СИСТЕМ 34
2.1 Постановка цілей і завдань експериментального дослідження 34
2.2 Аналіз факторів та відгуків експерименту 34
2.2.1 Основні положення теорії планування експерименту 34
2.2.2 Визначення факторів експерименту 35
2.2.3 Визначення відгуків експерименту 37
2.3 Методика обробки результатів експерименту 37
2.3.1 Статистична обробка 37
2.3.2 Попередня обробка даних 38
2.3.3 Гістограми 38
2.3.4 Кореляція 39
2.3.5 Аналіз використання методу Weighted Round Robin для розвантаження трафіку на сервери 41
2.3.6 Аналіз використання методу LeastConn для розвантаження трафіку на серверу 47
2.4 JMeter 52
2.5 Azure services 53
2.6 Алгоритм проведення експерименту 55
2.7 Висновки з розділу 2 55

3 АНАЛІЗ РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТУ ДЛЯ ДОСЛІДЖЕНЬ РОЗВАНТАЖЕННЯ ВИСОКОНАВАНТАЖЕНИХ СИСТЕМ ТА ФОРМУВАННЯ ПРАКТИЧНИХ РЕКОМЕНДАЦІЙ 57
3.1 Конфігурування серверів за допомогою Azure Services 57
3.2 Тестування навантаженням за допомогою програми JMeter 62
3.3 Аналіз результатів експерименту 70
3.4 Висновки з розділу 3 71
ВИСНОВКИ 73
ПЕРЕЛІК ПОСИЛАНЬ 76

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

API – інтерфейс створення додатку [1];

IMDG – InMemory Data Grid – це розподілене сховище об'єктів, схоже по інтерфейсу зі звичайною багато потокової хеш-таблицею [2];

GFS – Google File System – розподілена файлова система [3];

HighLoad – високонавантажених система;

HTML – Hyper Text Markup Language – мова гіпертекстової розмітки;

Nginx – веб-сервер і поштовий проксі-сервер, що працює на Unix-подібних операційних системах [4];

NoSQL – not only SQL – ряд підходів, спрямованих на реалізацію сховищ баз даних [5];

БД – база даних;

ІС – інформаційна система;

СУБД – система управління базами даних.

ВСТУП

Щороку кількість інтернет-користувачів неухильно зростає. Високонавантажені ресурси в першу чергу розраховані на багато користувачів додатків, більшість з яких розподілені системи, які працюють на більш ніж одному сервері. Ця конфігурація необхідна для ввімкнення обробки великих обсягів даних, створених під час пікових навантажень, а також його реплікації. Друге найважливіше завдання, описане в конфігурації застосунку ресурсів, полягає в забезпеченні стійкості системи .

В даний час немає добре налагодженого визначення дуже завантаженої системи, але деякі функції, які присутні в таких системах, можуть бути ідентифіковані з існуючої літератури. Як правило, високонавантажені системи реактивні це означає, що вони не просто повинні робити розрахунки, але і , якщо в такій системі щось виходить зі строю, користувачі не повинні цього відчувати. Навантаження в таких системах зазвичай починаються з 1000 RPS / QPS вони повинні швидко реагувати на прохання [7].

Високонавантажені ресурси в першу чергу розраховані на багато користувачів додатків, більшість з яких розподілені системи, які працюють на більш ніж одному сервері. Таким чином, дуже завантажені системи є безупинними системами доступу, тобто тими структурами, чиї дані запити дозволяють отримувати інформацію без тривалого переривання роботи при постійній роботі.

Актуальність цієї роботи полягає в тому, що важливість інформаційної технології в сучасному світі зростає, кількість відвідувачів веб-додатків зростає, і питання забезпечення стійкості при високих навантажень виходить на перший план. Таким чином, сьогодні завдання полягає у розробці оптимального коду, організації баз даних, поліпшенню апаратних ресурсів для підвищення швидкості роботи систем.

Мета дипломної роботи – аналіз ефективності високонавантажених систем на прикладі використання методів балансування навантаження.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

1. Проаналізувати особливості та проблеми при розробці високонавантажених систем на прикладах Google, Facebook и YouTube.

2. Спланувати експериментальне дослідження для оцінки ефективності роботи високонавантажених систем.

3. Провести експериментальне дослідження і проаналізувати результати експерименту.

4. Розробити практичні рекомендації для вирішення задачі вибору методу балансувальника навантаження.

Об'єкт дослідження – процеси, що протікають у високонавантажених системах типу Google, Facebook, YouTube.

Предмет дослідження – технології для оптимізації високонавантажених систем.

Методи дослідження – використання методів статистичного аналізу, а саме методів планування експерименту для порівняльного аналізу алгоритмів балансування навантаження Round Robin та Least Connection.

Наукова новизна дослідження. Дістали подальший розвиток методи оптимізації навантаження на сервер за рахунок використання балансувальника навантаження.

Практичне значення отриманих результатів полягає у тому, що рекомендації, отримані в результаті проведеного експериментального дослідження, можуть бути використані розробниками високонавантажених систем для зменшення навантаження на веб-сервер, що дозволить підвищити доступність серверу.