

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

Каланжова Анастасія Сергіївна

УДК 004.06

**Масштабування та оптимізація серверів баз даних в умовах підвищеного
навантаження**

Напрямок підготовки 6.050101 – «Комп'ютерні науки»

ДР.ПЗ-401.10790516

Автореферат
дипломної роботи на здобуття освітньої кваліфікації
«Бакалавр комп'ютерних наук»

Миколаїв – 2019

Дипломна робота є рукопис.

Робота виконана в Чорноморському національному університеті імені Петра Могили Міністерства освіти і науки України на кафедрі інтелектуальних інформаційних систем

Науковий керівник: Сіденко Євген Вікторович, кандидат технічних наук, доцент (б.в.з.) кафедри

Рецензент: Бойко Анжела Петрівна, декан факультету. К.т.н., доцент кафедри комп'ютерної інженерії.

Захист відбудеться «25» червня 2019 р. о 9³⁰ год. на засіданні екзаменаційної комісії (ауд. 2-403) у Чорноморському національному університеті імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

З дипломною роботою можна ознайомитися в бібліотеці Чорноморського національного університету імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

Автореферат представлений «20» червня 2019 р.

Секретар
екзаменаційної комісії,
ст.викл.

С.В.Дворецька

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

1. Актуальність теми.

У сучасному світі обмін даними займає важливе місце у житті суспільства. В той же час ці потоки даних настільки масштабні, що вибірка потрібної інформації може займати невиправдано багато часу та ресурсів.

Для кращого управління даними було введено спеціальний принцип обробки даних - масштабування баз даних. Зараз подібна техніка має багато варіантів роботи з даними. Для роботи з ним існує багато спеціалізованих засобів. Вони відрізняються функціоналом та направленістю (кожен тип даних, будь-то музика, відео, картинки, документи, або що має свій, часто не сумісний з іншими, формат даних), але кожна з них допомагає дізнатися ту чи іншу інформацію про наявні дані.

Масштабування даних забезпечує декілька переваг, головне з яких - зниження витрат на забезпечення узгодженого читання (яке для ряду низькорівневих операцій вимагає монополізації ресурсів сервера баз даних, вносячи обмеження на кількість одночасно оброблюваних запитів користувачів, незалежно від обчислювальної потужності використовуваного обладнання).

Робота з великими базами даних завжди потребувала більшої кількості ресурсів та наявності певних методів, які б могли працювати з динамічно зростаючим об'ємом інформації. З кожним роком все більш актуальнішою стає проблема пошуку оптимального методу обробки та масштабування баз даних.

2. Метою дипломної роботи є створення сучасного актуального способу обробки запиту великої бази даних та максимальне пришвидшення роботи застосунку. Це відбувається за допомогою оптимізації роботи вузлів серверів баз даних в умовах підвищеного навантаження.

Об'єкт досліджень – системи та сервери баз даних.

Предмет досліджень – методи масштабування та оптимізації роботи великих баз даних.

Для досягнення мети необхідно дотримуватись основних завдань, які протягом розробки ПЗ необхідно виконати:

- аналіз сучасних БД, визначення особливостей, переваг і недоліків;
- дослідження існуючих методів та алгоритмів масштабування;
- створення структури ПЗ (конструювання БД, користувацького інтерфейсу);
- розробка БД та наповнення її даними;
- розробка застосунку;
- тестування розробленого ПЗ.

3. Практичне значення отриманих результатів.

У сучасному світі бази даних та робота з ними займають дуже важливе місце. Обробка довгих запитів, виявлення помилок, аналіз даних серед мільйонів записів має бути максимально прискорене та оптимізоване. Саме тому, необхідно зробити більш продуктивною та ефективною взаємодію між серверами баз даних.

В даному дипломному проекті біло розглянуто систему масштабування бази даних вертикального секціонування. За приклад, була взята база даних користувачів мережі.

Практичне значення отриманих результатів полягає у тому, що використовуючи дану систему масштабізації баз даних можливе поліпшення обробки даних, пришвидшення та уникнення фатальних помилок під час роботи. Дана технологія може використовуватися майже в будь-якій галузі інформаційних технологій.

Структура дипломної роботи. Пояснювальна записка до дипломної роботи складається із вступу, 3 розділів, висновків, додатків. Загальний обсяг роботи складає 76 сторінок, 9 рисунків, та 30 посилань на літературні джерела.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** подано загальну характеристику досліджуваної теми, обґрунтовано актуальність дипломної роботи, сформульовано мету, завдання досліджень, відзначено наукову новизну та практичну цінність отриманих результатів, подано інформацію про структуру роботи.

У **першому розділі** було проведено аналіз систем масштабування даних, їх різновидів та форм. Також, було розглянуто технічні форми контролю та сучасні алгоритми оптимізації роботи баз даних. Для подальшого розгляду було обрано тестову форму застосунку обробки таблиць баз даних. Проведено аналіз роботи баз даних та виявлено основні недоліки сучасних систем.

Загалом можна розглядати масштабування баз даних двома способами: посилити сам сервер, або додати йому помічників. Ці два способи називаються вертикальним та горизонтальним масштабуванням.

Вертикальним способом можна додавати пам'ять, процесори, ядра, але це більш затратно і обмежено швидкістю розвитку технологічного прогресу (рис.1).

У горизонтальному масштабуванні свободи більше. Воно здійснюється за допомогою додавання нових серверів, запустивши більше процесів, можна перенести частину навантаження на обчислення новим серверам (рис.2).

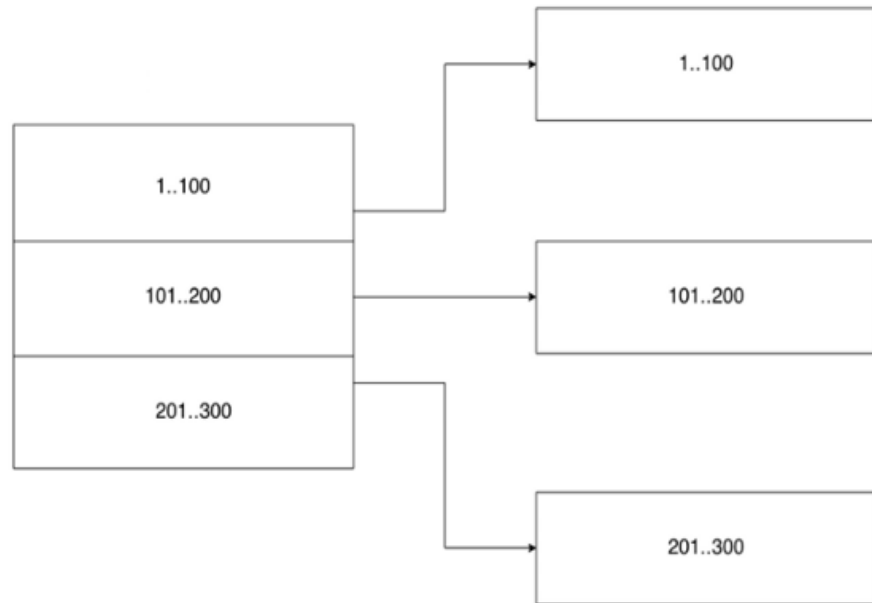


Рис.1. Схема вертикального масштабування даних

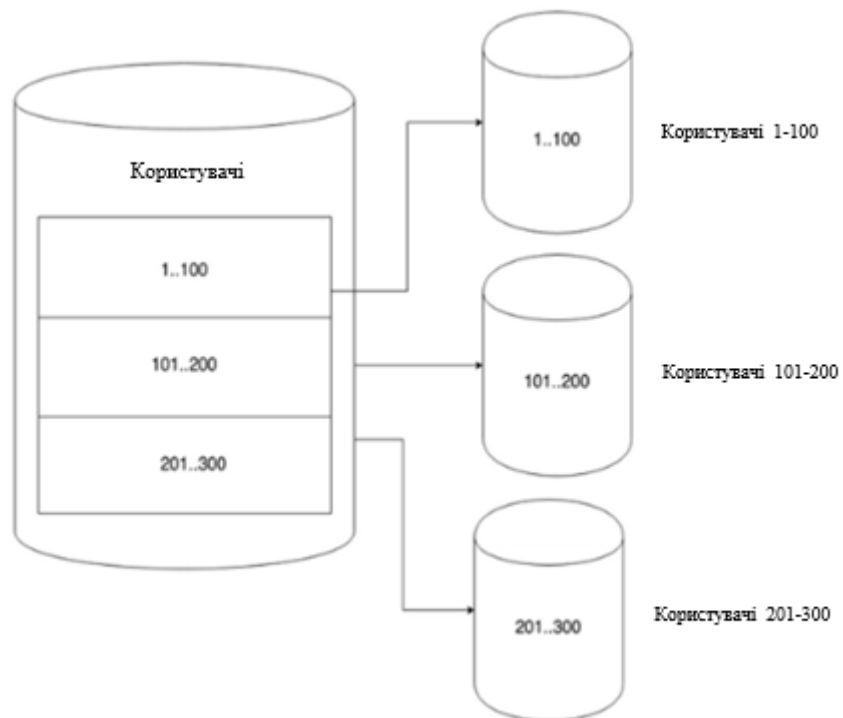


Рис.2 Схема горизонтального масштабування даних

Вертикальне масштабування зазвичай індивідуально, і вирішується швидше апаратно, ніж програмно. А ось для горизонтального існують загальновизнані програмні підходи, які отримали реалізацію в різних СУБД.

У другому розділі після аналізу предметної області, а саме форм масштабування баз даних їх видів та класифікацій, було обрано найбільш ефективний варіант для одного сервера, а саме вертикальне сегментування. Після постановки задачі, виникла необхідність у більш детальному аналізі. Так було проаналізовано сучасні системи масштабування та шардингу. Було досліджено сучасні методології використання масштабування великих баз даних. Розглядались як їх переваги так і недоліки. З огляду на отримані дані було виокремлено декілька правил, якими слід керуватися під час створення застосунку.

В даному розділі було проведено аналіз існуючих систем масштабування, а саме у MongoDB, SQL Azure Federations, Apache Cassandra, Redis, 2GIS, Grails Sharding Plugin. Було розглянуто основні підходи до створення систем БД та на основі цих даних висунуто вимоги до створюваної системи.

Були сформульовані вимоги до системи. Було виділено декілька основних функціональних характеристик для роботи застосунку:

- реєстрація користувачів в системі;
- авторизація користувачів;
- вибір мови інтерфейсу;
- створення та редагування записів;
- обробка результатів;
- відображення результатів;
- пошук користувача;
- пошук міста;
- надати можливість переглянути помилки при введенні;
- детальний перегляд інформації.

Повинно бути реалізовано поділ користувачів на наступні групи:

- модуль адміністратора;
- модуль користувача.

Застосунок повинен забезпечити для користувача інтуїтивно зрозумілий інтерфейс відповідно до вимог створення графічних інтерфейсів. Інтерфейс має бути однаковим для обох модулів, без зайвих елементів, що можуть відволікти від роботи.

Окрім створення нових записів, також необхідна реалізація їх редагування, а у випадку, якщо запис все був введений невірно, була можливість робити редагування, видалення.

У третьому розділі було проведено моделювання та проектування майбутнього застосунку. Змодельована ситуація проходження запиту у застосунку та роботи БД. На підставі цих даних біло спроектовано оптимальну обробку даних у таблицях, та розглянуто основні підходи до проектування інтерфейсу користувача.

Для виконання дипломної роботи було обрано об'єктно-орієнтовану мову програмування C# і платформу .Net. Ці засоби було обрано через відносно великий досвід роботи з ними. Також обрані технології мають переваги.

- 1) Об'єктно-орієнтований підхід мови програмування, що дозволяє максимально ефективно використовувати один раз написаний код для побудови нових компонентів.
- 2) Відсутність необхідності слідкувати за виділенням пам'яті.
- 3) Строга типізація, яка відфільтровує велику кількість помилок на етапі компіляції.
- 4) Переносимість на рівні скомпільованої збірки (зібраний на Windows застосунок може бути без змін запущений на Linux або будь-якій операційній системі, що підтримує будь-яку з реалізацій .Net).
- 5) Велика стандартна бібліотека.
- 6) Розробка засобу великою корпорацією (.Net — платформа Microsoft, яка активно розвивається).

Для створення застосунку були використані такі програмні технології:

- .Net Core;
- .Net Framework;
- My SQL;
- WPF;
- Visual Studio.

При написанні застосунку та створенні бази даних було досліджено ефективність використання методів вертикального сегментування.

Була створена база даних користувачів мережі. Переглянувши базу даних є можливість редагування записів (рис.3) та пошук за певними критеріями (рис.4).

Під час створення нового запису надається інформація про головні дані користувача, а саме ім'я, прізвище, вулиця проживання, місто, дата народження (рис. 5).

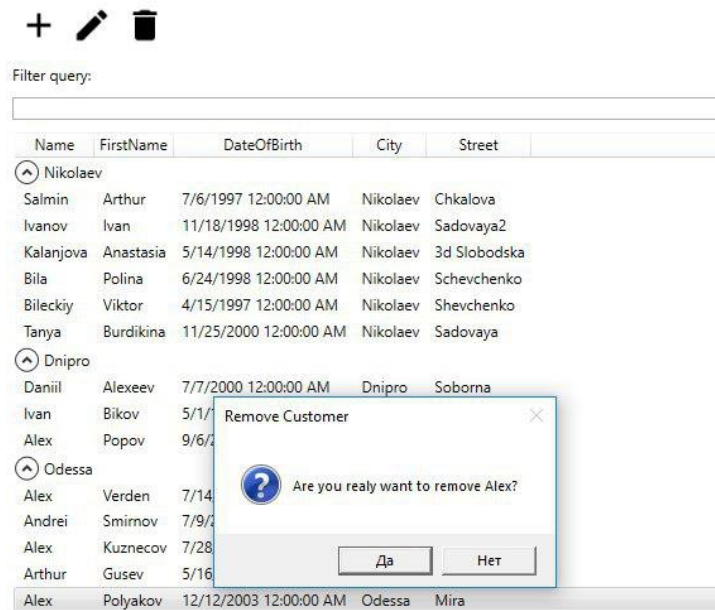
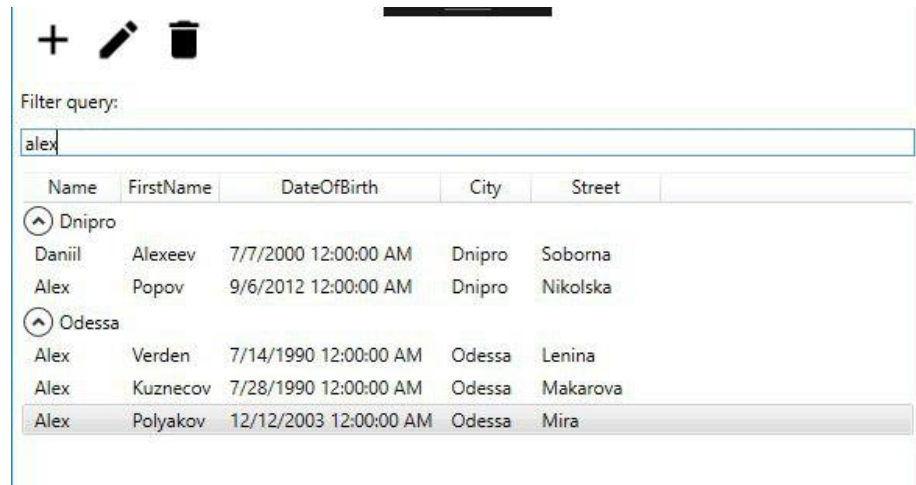


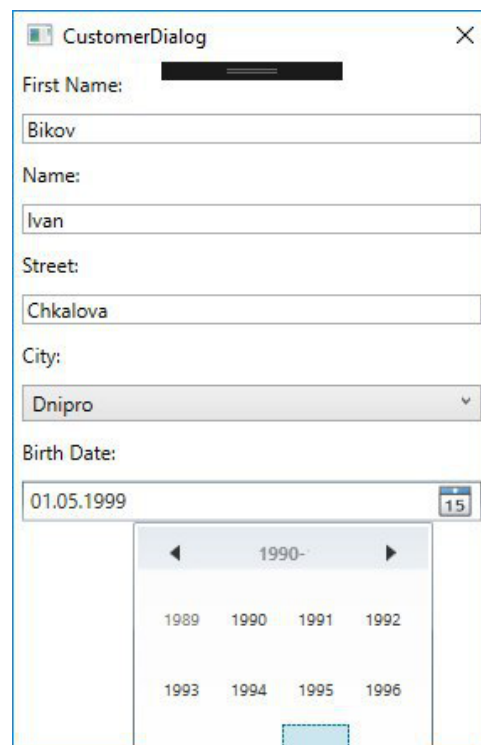
Рис.3. Видалення запису



Filter query: alex

Name	FirstName	DateOfBirth	City	Street
Dnipro				
Daniil	Alexeev	7/7/2000 12:00:00 AM	Dnipro	Soborna
Alex	Popov	9/6/2012 12:00:00 AM	Dnipro	Nikolska
Odessa				
Alex	Verden	7/14/1990 12:00:00 AM	Odessa	Lenina
Alex	Kuznecov	7/28/1990 12:00:00 AM	Odessa	Makarova
Alex	Polyakov	12/12/2003 12:00:00 AM	Odessa	Mira

Рис.4. Пошук за ім'ям користувача



CustomerDialog

First Name: [Redacted]

Name: Bikov

Name: Ivan

Street: Chkalova

City: Dnipro

Birth Date: 01.05.1999

1990

1989 1990 1991 1992

1993 1994 1995 1996

Рис.5 Створення нового запису

У процесі створення бази даних вона була заповнена певною кількістю даних, протестована та досліджена на наявність помилок.

Аналізуючи функціональну модель бази даних можна побачити 3 взаємопов'язані між собою таблиці, кожна з яких має ключ значення та ідентифікатори (рис. 6).

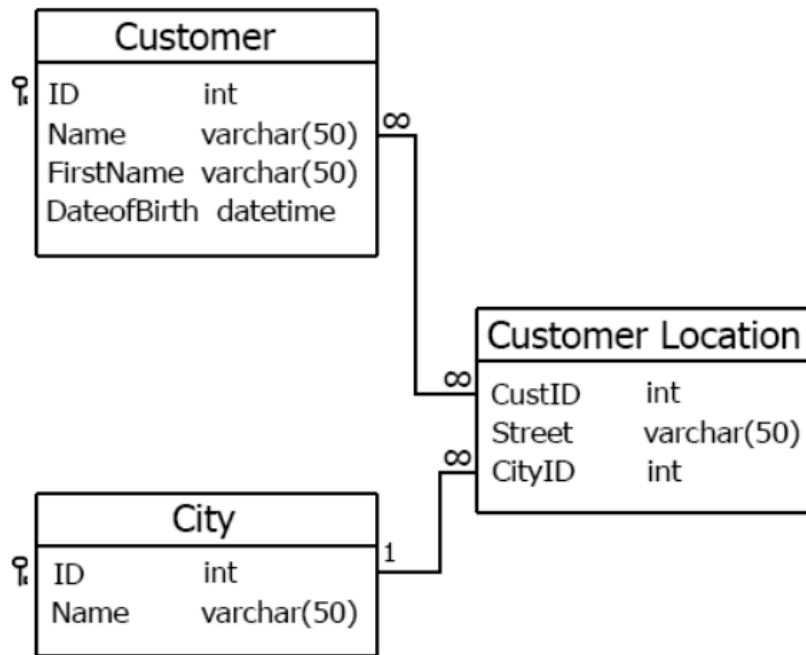


Рис.6 Фізична модель даних

Таблиця Customer містить у собі дані користувачів, їх повні імена у рядках, дату народження та ідентифікаційний номер у базі даних. При виконанні запиту за база перевіряє результат за повним ім'ям користувача, або за ідентифікаційним номером.

Таблиця Cust Location містить дані користувачів про місце проживання, а саме вулиця та ідентифікатор міста. Таблиця City містить дані про міста, їх назву та ідентифікатор.

ЗАГАЛЬНІ ВИСНОВКИ

В даному дипломному проєкті біло розглянуто систему масштабування бази даних вертикального секціонування. За приклад, була взята база даних користувачів мережі.

Після проведеного аналізу предметної області було зроблено ряд висновків:

- існуючі системи оптимізації роботи під навантаженням здебільшого мають мету прискорити обробку даних за рахунок збільшення кількості серверів;

– сучасні системи масштабування даних надають функціональні можливості лише частково;

– навіть найбільш популярні системи не дають гарантію на правильність обробки отриманих результатів.

Задачі, які були виконані в процесі дослідження:

– проведено аналіз існуючих систем масштабування даних;
– проаналізовано методи та алгоритми оптимізації запитів в умовах підвищеного навантаження;

– розроблено застосунок для обробки пошука бази даних;

– проведено тестування розробленої системи.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ

1. «Вертикальні алгоритми розбиття на розробка баз даних» Шаманта Навате, Стефано Сері, Гіо Відерхольда і Цзінлі Доу, Стенфордський університет 1984
2. «Високонавантажені застосунки. Програмування, масштабування, підтримка» Мартин Клеппман 2019
3. «NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence» Прамодкумар Дж. Садаладж, Мартин Фаулер 2017
4. «MongoDB в действии» ДМК Пресс, Кайл Бэнкер 2014
5. «Data Modeling for MongoDB» Стів Хоберман 2014
6. «Cassandra. Полное руководство» Карпентер Дж., Хевит Э. 2016
7. «MySQL: использование и администрирование» В. Васвани 2011
8. «PHP 6 and MySQL 6 Bible» Стів Суерінг, Тім Конверс, Джойс Парк 2010
9. «Beginning MySQL» Роберт Шелдон, Джоффри Мойе 2007
10. «MySQL на примерах» Кузнецов М., Сімдянов І. 2008.

АНОТАЦІЯ

Каланжова Анастасія Сергіївна. Масштабування та оптимізація серверів баз даних в умовах підвищеного навантаження. – На правах рукопису.

Дипломна робота на здобуття освітньої кваліфікації «Бакалавр комп'ютерних наук». – Чорноморський національний університет імені Петра Могили, Миколаїв, 2019.

В даному дипломному проекті розглянутий процес розробки застосунку для оптимізації роботи серверу в умовах підвищеного навантаження, шляхом масштабування баз даних.

Метою даної роботи є створення сучасного актуального способу обробки запиту великої бази даних та максимальне пришвидшення роботи застосунку.

Об'єкт дослідження – системи та сервери баз даних.

Предметом дослідження є методи масштабування та оптимізації роботи великих баз даних.

Загальна частина включає в себе наступні розділи: основні поняття, форми та методи масштабування баз даних; методи і моделі оптимізації роботи баз даних та моделювання та проектування програмної структури застосунку.

В дипломній роботі було розглянуто вертикальний метод масштабування даних. За приклад, була взята база даних користувачів мережі.

Після проведеного аналізу предметної області було зроблено ряд висновків:

– існуючі системи оптимізації роботи під навантаженням здебільшого мають мету прискорити обробку даних за рахунок збільшення кількості серверів;

– сучасні системи масштабування даних надають функціональні можливості лише частково;

– навіть найбільш популярні системи не дають гарантію на правильність обробки отриманих результатів.

Також в даній частині було проведено моделювання та проектування майбутнього застосунку. Результатом виконання загальної частини дипломного проекту є розроблене програмне забезпечення для обробки запитів бази даних користувачів мережі.

В спеціальній частині «Охорона праці» розглянуто умови праці на робочих місцях спеціалістів комп'ютерної фірми та користувачів, а також заходи по збереженню життя людини.

В цілому робота складається зі 75 сторінок, 9 рисунків, 2 додатків.

Ключові слова: масштабування даних, сервери баз даних, оптимізація.

ABSTRACT

In this graduation project was regarded the process of developing an application for optimizing the server in conditions of increased load, by scaling databases is considered.

The purpose of this work is to create a modern current method of processing a large database query and maximize the speed of the application.

Object of research - systems and servers of databases.

The subject of the study is the methods of scaling and optimizing the work of large databases.

The general part includes the following sections: basic concepts, forms and methods of database scaling; methods and models of database optimization and modeling and designing of the software structure of the application.

In the thesis the vertical method of data scaling was considered. For example, a database of network users were taken.

After the analysis of the subject area, a number of conclusions were made:

- Existing systems for optimizing work under load are for the most part aimed at speeding up data processing by increasing the number of servers;
- modern data scaling systems provide only partial functionality;
- Even the most popular systems do not guarantee the correctness of the processing of the results.

Also, in this part, the simulation and design of the future application were carried out. The result of the implementation of the general part of the diploma project is the developed software for processing the requests of the database users of the network.

In the special section "Labor protection" the conditions of work at the workstations of specialists of the computer firm and users are considered, as well as measures for the preservation of human life.

In total, the work is split with 75 pages, 9 figures, 2 attachments.

Key words: data scaling, database servers, optimization.