

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ
ФАКУЛЬТЕТ КОМП'ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ І ПРОГРАМНОЇ
ІНЖЕНЕРІЇ

ЦАПКО АНДРІЙ ВОЛОДИМИРОВИЧ

УДК 004.62

**АЛГОРИТМІЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ МІГРАЦІЇ
СХОВИЩ ДАНИХ З РЕЛЯЦІЙНОЇ НА НЕРЕЛЯЦІЙНУ МОДЕЛЬ У
ВИСОКОНАВАНТАЖЕНИХ КОМП'ЮТЕРНИХ СИСТЕМАХ**

123 «Комп'ютерна інженерія»

Автореферат

дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль
2018

Роботу виконано на кафедрі комп'ютерних систем та мереж Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України

Керівник роботи: кандидат технічних наук, доцент кафедри комп'ютерних систем та мереж
Луцків Андрій Мирославович,
Тернопільський національний технічний університет
імені Івана Пулюя,

Рецензент: кандидат фізико-математичних наук, доцент кафедри фізики
Крамар Олександр Іванович,
Тернопільський національний технічний університет
імені Івана Пулюя,

Захист відбудеться 27 грудня 2018 р. о 9⁰⁰ годині на засіданні екзаменаційної комісії № 34 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд.1-603

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність роботи. Сучасне покоління програмних засобів, призначених для задоволення потреб великих груп користувачів або великих організацій, має справлятися з величезним зростанням обсягу даних, що підлягають обробці та зберіганню. Додатки, що стосуються даних соціальних мереж, ілюструють цей сценарій, в якому розвивається децентралізоване створення та обмін користувальним контентом та призводить до надзвичайно цінної кількості даних. Протягом багатьох десятиліть система управління реляційними базами даних (RDBMS), заснована на реляційній моделі, задовільно виконує вимоги самих різноманітних програмних застосувань. Популярність використання цієї моделі пояснюється простотою використання, а також надійністю, високою продуктивністю та сумісністю. Проте останнім часом традиційні RDBMS виявилися неспроможними впоратися з вимогами нового покоління програмних засобів, які зосереджені, зокрема, на неструктурованих даних та розподіленому зберіганню. З метою задоволення потреб в надійному зберіганні та швидкому доступі до великих обсягів даних була запропонована NoSQL бази. Вони забезпечують швидкість обробки даних, різноманіття типів збережених даних і гнучкість при роботі. На сьогодні, є можливим перехід від реляційної моделі до нереляційної, й відповідно з реляційної бази даних на нереляційну. Для цього, існує низка методологій, алгоритмічних та програмних засобів, використання базується на сфері використання сховища даних, технологіях доступу до даних та низці інших факторів. Узагальнення методології та формулювання рекомендацій по міграції з реляційної на нереляційну модель набуває все більшої популярності. Розв'язанням даної проблеми займалися: Е. Ф. Кодд, Е. Брюер, Дж. Грей, М. Фаулер та інші.

Мета і завдання дослідження. Метою роботи є дослідження алгоритмічного та програмного забезпечення для міграції сховищ даних з реляційної на нереляційну модель у високонавантажених комп'ютерних системах.

Для досягнення поставленої мети потрібно розв'язати такі задачі:

- проаналізувати сучасне алгоритмічне забезпечення для міграції сховищ даних з реляційної на нереляційну модель і обґрунтувати вибір оптимальної;
- проаналізувати і обґрунтувати вибір програмного забезпечення для міграції сховищ даних;
- проаналізувати наявні нереляційні бази даних у контексті міграції на них;
- застосувати програмне забезпечення для реалізації процесу міграції;
- перевірити продуктивність реляційного і нереляційного підходу, при великих наборах даних.

Об'єкт дослідження – процес міграції сховищ даних з реляційної на нереляційну модель у високонавантажених комп'ютерних системах.

Предметом дослідження є алгоритми, моделі і програмні засоби для реалізації процесу міграції з реляційної на нереляційну модель.

Наукова новизна отриманих результатів:

- узагальнено та сформульовано рекомендації, які дають змогу спростити процес міграції сховищ даних;
- проведено обчислювальний експеримент в ході якого проведено порівняння продуктивності реляційної і нереляційної БД після міграції, при великій кількості даних.

Практичне значення отриманих результатів. Розроблені рекомендації та практики дають змогу спростити та автоматизувати процес міграції даних.

Апробація. Результати роботи апробовано на VII Міжнародній науково-технічній конференції молодих учених та студентів «Актуальні задачі сучасних технологій» м. Тернопіль 28-29 листопада 2018 року та VI науково-технічній конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі, системи та технології». м. Тернопіль 12-13 грудня 2018 року.

Структура роботи. Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 6 частин, висновків, переліку посилань та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 145 арк. формату А4, графічна частина – 10 аркушів формату А1.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність дослідження, мету роботи, задачі, об'єкт, предмет, наукову новизну, практичне значення та публікації дипломних досліджень.

У першому розділі роботи «Аналіз предметної області сховищ зберігання даних» проаналізовано реляційні і нереляційні бази даних, їх концепції побудови. Здійснено порівняльний аналіз SQL та NoSQL. Розглянуто основні моделі нереляційних баз, а також проаналізовано праці інших дослідників в даній предметній області.

У другому розділі «Аналіз процесу міграції з реляційної на нереляційну модель» проаналізовано сучасне алгоритмічне забезпечення та методи міграції сховищ даних з реляційної на нереляційну модель, до яких належить: ETL, ELT, алгоритм Map Reduce, міграція на основі зв'язаних списків та на основі графів. Розглянуто моделі міграції на основі зв'язаних списків та на основі графів. Приведено рекомендації, щодо реалізації ETL процесу

У третьому розділі «Реалізація переходу з реляційної на нереляційну модель» проаналізовано дві БД, які є найпопулярнішими представниками реляційної (MySQL) та нереляційної (MongoDB) моделей. Також розглянуто дві платформи для міграції даних: Cosmos DB та Apache Sqoop; застосовано програмне забезпечення на основі ETL моделі для реалізації процесу міграції.

Проведено обчислювальний експеримент в ході якого проведено порівняння продуктивності реляційної і нереляційної БД після міграції на прикладі MySQL та MongoDB, при великій кількості даних.

Проведений експеримент підтвердив твердження, що нереляційна модель є більш продуктивною, ніж реляційна, при великих обсягах даних.

У четвертому розділі «Обґрунтування економічної ефективності» зроблено обчислення показників економічної ефективності від застосування методу вибору оптимального рішення при реалізації програмних проектів, що дало можливість зробити висновок про доцільність проведення НДР.

П'ятий розділ роботи «Охорона праці та безпека в надзвичайних ситуаціях» проведено аналіз вимог з охорони праці і техніки безпеки при використанні комп'ютерної техніки, розглянуто небезпечні і шкідливі фактори при роботі з комп'ютерними системами, а також наслідки їх дії і описано заходи, щодо запобігання цих факторів; охарактеризовано заходи по підвищенню стійкості об'єктів в НС. Проаналізовано методи захисту персоналу об'єктів та членів їх сімей.

Шостий розділ роботи «Екологія» містить дослідження питання про збір екологічної інформації на базі комп'ютерних технологій, а також характеристику рівнів та видів моніторингу навколишнього середовища.

ВИСНОВКИ

У даній магістерській роботі проаналізовано алгоритмічне та програмне забезпечення міграції сховищ даних з реляційної на нереляційну модель у високонавантажених розподілених комп'ютерних системах. Основні результати та висновки проведених теоретичних та експериментальних досліджень такі:

- проаналізовано сучасне алгоритмічне забезпечення для міграції сховищ даних з реляційної на нереляційну модель і обґрунтовано вибір оптимальної моделі;
- проаналізовано програмне забезпечення для міграції сховищ даних;
- проаналізовано різні методики та алгоритмічне забезпечення міграції з точки зору ефективності та відповідності наявним структурам даних;
- на прикладі документо-орієнтованої MongoDB проаналізовано наявні нереляційні бази даних у контексті міграції на них;
- застосовано програмне забезпечення для реалізації процесу міграції, що дало змогу провести верифікацію запропонованих підходів;
- на основі аналізу методологій та практик, узагальнено та сформульовано рекомендації, які дають змогу спростити процес міграції сховищ даних;
- проведено обчислювальний експеримент в ході якого проведено порівняння продуктивності реляційної і нереляційної БД після міграції, при великій кількості даних.

Було здійснено економічні розрахунки, спрямовані на визначення економічної ефективності та вартості дослідження алгоритмічного та програмного забезпечення міграції сховищ даних з реляційної на нереляційну модель у високонавантажених комп'ютерних системах.

Розглянуто та описано вимоги з охорони праці та техніки безпеки відповідно до нормативних документів щодо: організації робочого місця, електробезпеки, шуму та вібрації, освітленості, мікроклімату та пожежної безпеки.

Розглянуто поставлені питання екології які стосуються магістерської роботи.

СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ

1. Цапко А.В. Концепції сховищ даних у контексті міграції з SQL на NoSQL/ А. В. Цапко // Матеріали VII Міжнародної науково-технічної конференції молодих учених та студентів. «Актуальні задачі сучасних технологій» - Тернопіль, 28-29 листопада 2018 р. – с. 107
2. Цапко А.В. Методи та засоби міграції з реляційної моделі даних до нереляційної/ А. В. Цапко // Матеріали VI науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі, системи та технології» - Тернопіль, 12-13 грудня 2018 р. – с.35

АНОТАЦІЯ

Цапко А.В. Алгоритмічне та програмне забезпечення міграції сховищ даних з реляційної на нереляційну модель у високонавантажених комп'ютерних системах.

Дипломна робота на здобуття освітнього ступеня магістра 123 – Комп'ютерні системи та мережі. – Тернопільський національний технічний університет імені Івана Пулюя 2018.

Мета роботи полягає у дослідженні алгоритмічного і програмного забезпечення міграції сховищ даних з реляційної на нереляційну модель у високонавантажених комп'ютерних системах.

У дипломній роботі проаналізовано сучасне алгоритмічне та програмне забезпечення для міграції сховищ даних з реляційної на нереляційну модель і обґрунтовано вибір оптимальної моделі; проаналізовано різні методики та алгоритмічне забезпечення міграції з точки зору ефективності та відповідності наявним структурам даних.

На прикладі документо-орієнтованої MongoDB проаналізовано наявні нереляційні бази даних у контексті міграції на них. Застосовано програмне забезпечення для реалізації процесу міграції, що дало змогу провести верифікацію запропонованих підходів, а також проведено обчислювальний експеримент в ході якого проведено порівняння продуктивності реляційної і нереляційної БД після міграції, при великій кількості даних.

Ключові слова: РЕЛЯЦІЙНА МОДЕЛЬ, НЕРЕЛЯЦІЙНА МОДЕЛЬ, МІГРАЦІЯ ДАНИХ, СХОВИЩА ДАНИХ.

ANNOTATION

Tsapko A.V. Algorithms and software of data storage migration from relational into non-relational model in highly loaded computer systems.

The diploma paper for obtaining the Master's degree 123 – Computer systems and network – Ternopil Ivan Puluj National Technical University 2018.

The purpose of the work is to investigate algorithmic and software migration of data warehouses from relational to non-relational model in high-capacity computer systems.

In the thesis the modern algorithmic and software for migration of data warehouses from relational to non-relational model is analyzed and the choice of optimal model is substantiated; Different methods and algorithmic support of migration in terms of efficiency and compliance with existing data structures are analyzed.

An example of a document-oriented MongoDB analyzes existing non-relational databases in the context of migration to them. Applied software for the implementation of the migration process, which made it possible to verify the proposed approaches, as well as a computational experiment was conducted during which a comparison of the performance of the relational and non-relational database after migration was performed, with a large amount of data.

Keywords: RELATIONAL MODEL, NON-RELATIONAL MODEL, DATA MIGRATION, DATA STORAGE.