ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

Харківський національний університет радіоелектроніки

КВАЛІФІКАЦІЙНА РОБОТА

«Методи обробки запитів в системах паралельних баз даних»

Виконав: студент групи СПм-20-1 Бессараб Є.В.

Керівник: доц. каф. ЕОМ Філімончук Т.В.

2

Аналіз предметної області

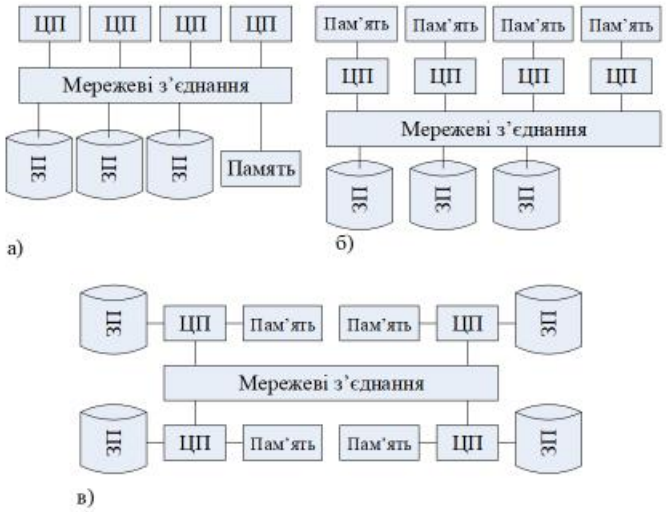
Метою кваліфікаційної роботи є дослідження методів паралельного виконання SQL - запитів у великомасштабних системах для підвищення продуктивності розподілених СКБД та зменшення часу виконання запиту.

Об'єкт досліження: методи паралельного виконання SQL-запитів.

Завдання:

- докладний аналіз ефективності існуючих методів розпаралелювання запитів у гетерогенних середовищах;
- дослідження алгоритмів та програмних засобів для паралельного виконання SQL – запитів, які забезпечують сумісність з існуючими СКБД;
- отримання апріорних оцінок часу виконання запитів;
- отримання експериментальної оцінки методів та алгоритмів, які були використані для розпаралелювання запитів у паралельних системах баз даних.

Архітектури паралельних систем 3



Класифікація Стоунбрекера для паралельних систем 4

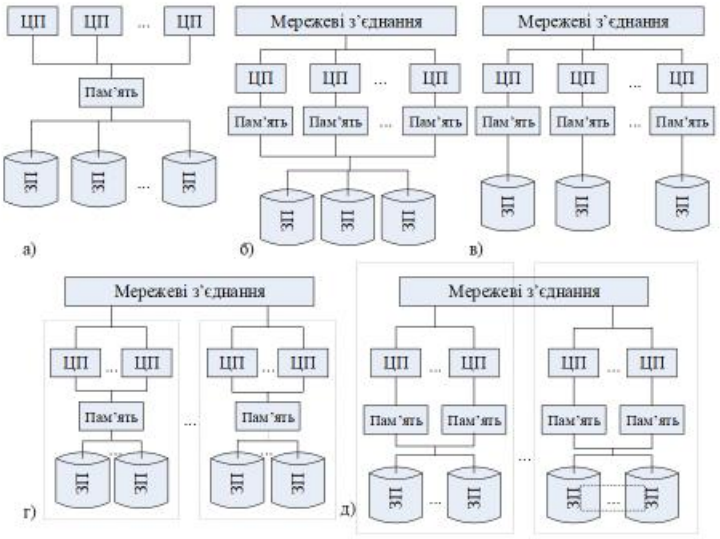
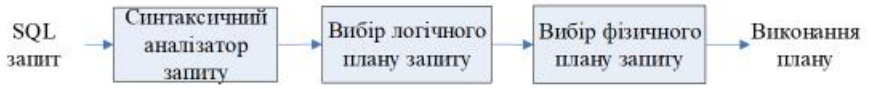


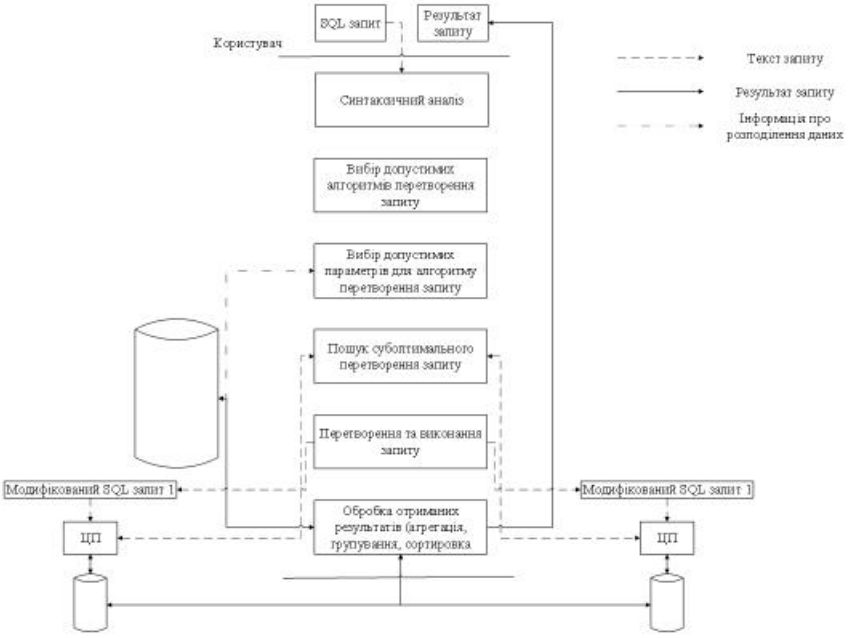
Схема компіляції SQL-запиту. Методи вибору субоптимального запиту

5



Архітектура системи для паралельного використання запитів

6



Агрегація атрибутів проміжних співвідношень

7

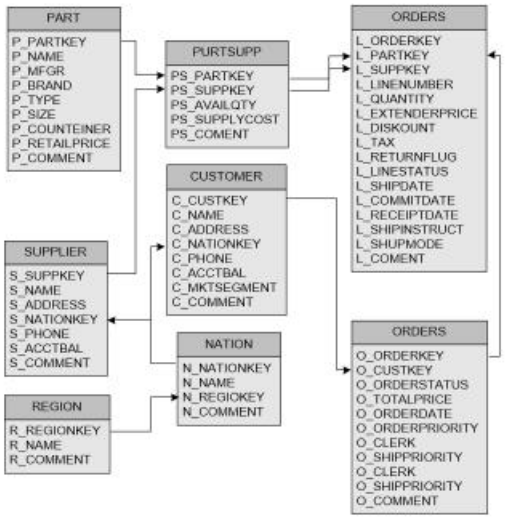
8

Структура інформаційної системи підприємства

Назва атрибуту	Тип	Коментар
SysID	int	Унікальний ключ опису індексу
Tablename	varchar(255)	Ім'я відношення, якому належить атрибут
FieldName	varchar(255)	Назва атрибуту в таблиці
IndexName	varchar(255)	Назва індексу
IndexType	int	Тип индексу: 0 = B-Tree, 1= Cluster
FieldType	varchar(255)	Тип атрибуту
MinValue	varchar(255)	Мінімальне значення атрибуту на всьому співвідношеніні
MaxValue	varchar(255)	Максимальне значення атрибуту на всьому співвідношеніні

9

Схема тестової бази даних



Результати виконання Q1

10

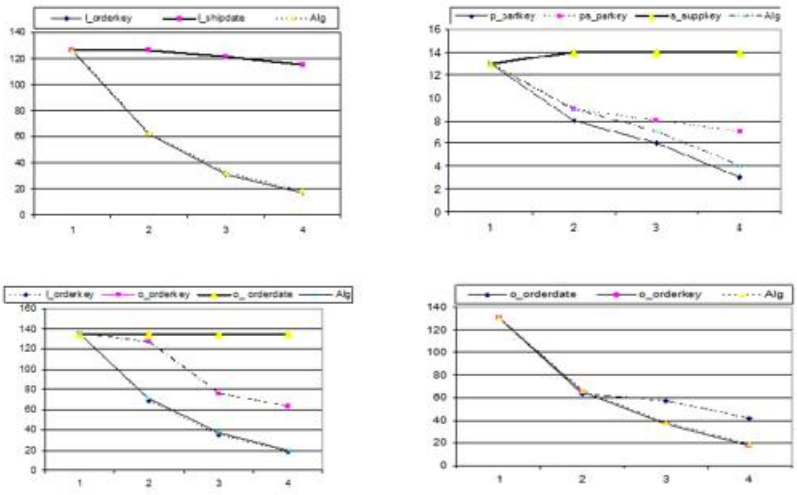
Кількість вузлів	Рядків у виборці	Проміжна вибірка	Додаткові плани запиту	Загальний час	K_y
1	4	0	0	126	1
2	4	8	1	62	1.02
4	4	16	1	31	1.02
8	4	32	1	17	0.93

Кількість вузлів	Рядків у виборці	Проміжна вибірка	Додаткові плани запиту	Загальний час	K_y
1	4	0	0	126	1
2	4	8	1	126	0.5
4	4	16	1	121	0.26
8	4	32	1	115	0.14

Кількість вузлів	Рядків у виборці	Проміжна вибірка	Додаткові плани запиту	Загальний час	K_y
1	4	0	0	126	1
2	4	8	2	63	1
3	4	16	2	32	0.98
8	4	32	2	18	0.88

Результати

11



Висновки

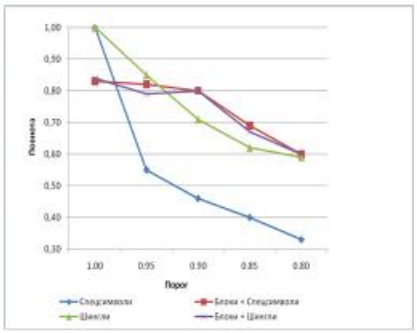
12

З використанням різних архітектур систем для паралельного виконання запитів з використанням надлишкової інформації було розглянуто та досліджено методи розпаралелювання SQL-запитів у СУБД на основі еквівалентного перетворення запитів до форми, що допускає паралельне виконання для застосування в гетерогенних середовищах. Внаслідок чого були отримані апріорні оцінки загального часу паралельного виконання перетворених SQL-запитів, що відрізняються урахуванням особливостей виконання запитів у гетерогенних системах з реплікацією даних, і дозволяють швидко оцінити доцільність досліджуваних методів розпаралелювання запитів на етапі конфігурування системи. Найкращі загальні показники продуктивності системи можна досягти, розробляючи систему на основі MPP архітектури, допускаючи використання SMP машин як вузлів системи, що фактично перетворює її на гібридну.

Програмний інтерфейс та результати

13

Ім'я класу	Описання
FileList	Містить множину екземплярів класу File, що описує файли колекції
File	містить ім'я, шлях і вміст файлу, а також множину екземплярів класу FileBlock
FileBlock	являє собою блок файлу, що включає текстовий вміст і множину відбитків
ShinglesFactory	статичний клас, що виробляє відбитки на основі заданого тексту і набору параметрів
ShinglesCollector	сховище відбитків і інтерфейс для зручної роботи з ними
PageProcessor	надає інтерфейс для роботи із конкретною сторінкою в частині «Виділення блоків»
Log	надає інтерфейс для роботи із конкретною сторінкою в частині «Виділення блоків»
GeneticAlg	клас для роботи з реалізацією генетичного алгоритму бібліотеки JGAP
FitnessFunc	клас, що представляє фітнес-функцію в відповідної зі специфікацією бібліотеки JGAP
StringExtractor	статичний клас, що надає можливість витягнення контенту з web-сторінки на базі бібліотеки HTMLParcer
VisitorRemover	клас, що надає можливість видалення вузлів і гілок DOM-дерева шляхом використання бібліотеки HTMLParcer
VisitorCalculator	реалізує простановку коефіцієнтів вузлів дерева, дозволяючи працювати ваговій функції алгоритму витягнення блоків, працює з використанням бібліотеки HTMLParcer
BlockListTableModel, FileListTableModel, utilLineCellRenderer, MyHTMLTreeCellRenderer, NodeTableModel, ShinglesListTableModel	розширення стандартних Java-класів візуальних компонентів, що мають додаткові можливості в частині візуалізації та інтерфейсу



Висновки

14

Запропоновано модель, що дозволяє оцінювати схожість документів, використовуючи інформацію про схожість їх блоків. Відповідно до моделі розроблено метод оцінки схожості документів на основі складових їх структурно-семантичних блоків, що дозволяє поліпшити якість розпізнавання дублікатів за рахунок збільшення середнього значення показника повноти. Це дозволяє отримувати більш повну вибірку схожих документів на одних і тих же наборах даних. Проведена експериментальна перевірка запропонованих алгоритмів і методів, у розробленому програмному забезпеченні.