**Іванкевич Олексій Вікторович. Методи і засоби організації паралельних процесів у системах великих баз даних: дис... канд. техн. наук: 05.13.13 / НАН України; Інститут проблем моделювання в енергетиці ім. Г.Є.Пухова. - К., 2004.**

|  |  |
| --- | --- |
| |  | | --- | | ***Іванкевич О.В. Методи і засоби організації паралельних процесів у системах великих баз даних. -****Рукопис.*  *Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.13****"Обчислювальні машини, системи і мережі".****- Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, Київ, 2003.*  Дисертація присвячена розробці методів і засобів підвищення ефективності обробки великих масивів статистичних, економічних, радіолокаційних даних, обробка яких вимагає роботи в реальному режимі часу чи потребує мінімального часу відгуку. Розглянуто і проаналізовано інформаційні технології, апаратні та програмні методи і засоби, орієнтовані на побудову обчислювальних систем, призначених для вирішення задач оптимізації і прискорення обробки великих баз даних (БД). Здійснено створення на основі існуючих і модифікованих методів та засобів апаратної платформи, алгоритмів та програмних засобів, орієнтованих на розпаралелювання операцій і призначених для обробки великих масивів даних.  Під час виконання роботи були створені як спеціалізовані апаратні засоби, архітектура яких орієнтована на переробку великих обсягів даних, так і програмні засоби, що дозволяють прискорити обробку даних на комп'ютерах з новою архітектурою.  При розробці апаратної частини була створена модифікована трирівнева ахітектура системи бази даних, основою якої стали багатопроцесорні модулі, побудовані на *DSP*-процесорах. Програмна частина заснована на постреляційному *SQL*-сервері, що за допомогою спеціально розробленого програмного забезпечення може робити обчислення на створеній багатопроцесорній системі. Розроблена оптимізована організація міжпроцесорного обміну, запропоновані алгоритми, які використовуються для прискорення обчислень на багатопроцесорних системах. При цьому основна увага приділена трудомістким операціям (вибірки, з'єднання таблиць БД та ін.). Виконано оцінку ефективності паралельних алгоритмів і пристроїв, що входять до складу систем БД, за тимчасовими витратами, числом процесорів і значенням прискорення. | |
| |  | | --- | | У дисертаційній роботі досліджений і розроблений комплекс методів і засобів прискорення й оптимізації обробки великих масивів даних. Вирішено актуальну задачу побудови апаратної і програмної частин системи, призначеної для вирішення задач обробки великих БД і яка використовує принципи розпаралелювання обробки інформації. На основі проведених досліджень отримані наступні основні результати:   1. Проведено дослідження особливостей побудови сучасних високопродуктивних ОС із використанням багатопроцесорності, розпаралелювання в накопичувачах даних, кластерних рішень, спеціального програмного забезпечення доступу до БД і мовних засобів, що допускають програмування обчислювальних процесів у системах обробки великих БД. 2. Досліджено методи підвищення продуктивності СБД - апаратна і програмна реалізації операцій реляційної алгебри, вертикальний і горизонтальний паралелізм. Для збільшення продуктивності СБД запропоновано використовувати існуючі високопродуктивні компоненти, а також спеціальний багатопроцесорний акселератор, побудований на недорогих високопродуктивних процесорах. Запропонована трирівнева ієрархія побудови СБД, сформульовані вимоги до акселераторів СБД. На першому рівні ієрархії БОС поєднуються між собою високошвидкісною обчислювальною мережею, на другому будуються системи, що складаються з процесорних модулів, на третьому розглядаються процесорні модулі, що у свою чергу можуть складатися з декількох процесорів. 3. Виконано дослідження сучасних *DSP*-мікропроцесорів, заснованих на *Risc*-архітектурі, що можуть бути використані для побудови акселераторів, досліджені продуктивність, пропускна здатність портів, сумісність із сучасними серверними системами, вартість. Вперше запропоновано базові структури СБД: акселератора СБД мінімальної конфігурації на основі одного *DSP*; конфігурації АСБД з двома, чотирма і вісьмома процесорами; конфігурації акселератора, призначеного для конвеєрної обробки даних і одноплатні пристрої, що розташовуються усередині корпуса сервера і під'єднуються до *PCI-*шини. 4. Запропоновано різні архітектури багатопроцесорних *DSP*-систем (лінійна, матрична, лінійна з розподіленим ЗП, матричні з розподіленим і загальним ЗП) для різних видів обробки БД і реалізації функцій СКБД. Показано, що існує можливість нарощування кількості процесорів від 3 (у випадку використання однопроцесорних модулів) до 48 (при використанні модулів з 8 *DSP*). 5. Запропоновано способи розподілу функцій між *Host*-комп'ютером і співпроцесором на базі *DSP*-процесорів у СБД при виконанні запитів до даних, розглянуті як послідовні стадії виконання запиту на СБД. Розроблено систему, що розподіляє навантаження в СБД, у залежності від якої обробка невеликих запитів відбувається на *Host*-комп'ютері, а запитів, пов'язаних з великим обсягом обчислень (сортування, перетинання й ін.) - у співпроцесорі СБД. 6. На базі стандартного сервера розроблена СБД з оптимізованим процесом обробки даних. Сформульовано вимоги до апаратного і програмного забезпечення *Host*-комп'ютера в складі СБД. Найбільш значущим критерієм є ємність і швидкодія використовуваних ЗП, а також швидкодія центрального процесора особливо для варіанта з великою кількістю *DSP-*процесорів, при якому обмін даними в АСБД із ЗП найбільш інтенсивний. 7. Як основу програмної частини СБД запропоновано використовувати *SQL-* сервер, у якому реалізована базова *SQL* мова, реляційна структура збереження даних, система індексування і реплікації даних, система безпеки і т.д. На базі *SQL-* мови базової СКБД запропоновані семантичні і синтаксичні структури вхідної мови СБД, що задовольняють вимогам: реляційної орієнтації за структурами даних; операторам запитів і маніпулювання; багатоцільової орієнтації операторів мови; можливості вибору типу обробки даних СКБД. 8. Розроблено функції, що інтегруються в базову СКБД, і взаємодіючі з ними програми керування процесом обчислення в співпроцесорі БД. Розроблена оптимізована організація міжпроцесорного обміну, запропоновані алгоритми, що використовуються для прискорення обчислень на багатопроцесорних системах. При цьому основна увага приділена трудомістким операціям (вибірки, з'єднання таблиць БД та ін.). Виконано оцінку ефективності паралельних алгоритмів і пристроїв, що входять до складу СБД, за часовими витратами, числом процесорів і значенням прискорення. 9. Вперше створено два комплекси програм, призначених для обробки в режимі реального часу і з мінімальним часом чекання радіолокаційних даних, а також для аналізу і зміни великих масивів бухгалтерської й економічної інформації й оптимізовані для роботи з розробленою СБД, що базується на АСБД. | |