

Б. І. Мороз, доктор технічних наук,
декан факультету інформаційних
та транспортних систем і технологій
Університету митної справи та фінансів
Л. В. Кабак, кандидат технічних наук, доцент
кафедри інформаційних систем та технологій
Університету митної справи та фінансів
О. Н. Молотков, кандидат технічних наук,
доцент кафедри інформаційних
систем та технологій Університету митної
справи та фінансів

УПРОВАДЖЕННЯ GRID-ТЕХНОЛОГІЙ У СПЕЦІАЛІЗОВАНІ ІНФОРМАЦІЙНІ СИСТЕМИ

Проведено дослідження та розглянуто основні напрямки щодо впровадження Grid-технологій у спеціалізовані інформаційні системи на прикладі інформаційної системи Державної фіскальної служби України. Для переходу до Grid-технологій потрібно виконати консолідацію баз даних. Запропоновано модель консолідації та розроблено рекомендації щодо етапів і методів консолідації мережі, додатків, платформ та даних. Розроблено алгоритм проведення секціонування, який можна використовувати у секціонуванні великих таблиць. Розроблену модель та алгоритм секціонування доцільно використовувати під час перетворення розподілених баз даних до архітектури Grid.

Ключові слова: *Grid-системи; секціонування; консолідація даних; бази даних; спеціалізовані інформаційні системи.*

The main direction of implementation of Grid technologies in the specialized information system on the example of the information system of the State Fiscal Service of Ukraine are analyzed in the article. It is necessary to do database consolidation for transferring to the Grid technologies. The consolidation model is proposed in the article, appropriate recommendations with the stages and methods of network consolidation applications platforms and data are discussed. The algorithm the partitioning is developed, which can be used for partitioning large tables. This developed model and algorithm partitioning can be used for conversion of distributed database to the Grid architecture.

Key words: *Grid-system; partitioning; data consolidation; data bases; specialized information systems.*

Постановка проблеми. Нині в інформаційній системі департаменту організації митного контролю та оформлення Державної фіскальної служби України не використовують централізоване збереження інформації в одній базі даних. Важливі дані, до яких необхідний доступ усіх користувачів, звичайно розкидані серед безлічі баз даних, розташованих у різних фізичних місцях збереження. Щоб об'єднати можливості баз даних усіх підрозділів і дозволити всім користувачам звертатися до збереженої інфор-

© Б. І. Мороз, Л. В. Кабак, О. Н. Молотков, 2015

мації, використовуються засоби Oracle, які називаються розподіленою базою даних і тиражуванням даних. Кожна база даних керується власною локальною системою керування базою даних (СКБД). Усі сервери баз даних у розподіленій базі даних співпрацюють, щоб підтримувати узгодженість глобальної бази даних. Однак технологія розподілених баз даних досить застаріла і має деякі недоліки:

- витрачаються певні кошти на підтримку окремих серверів БД;
- швидкодія серверів не завжди відповідає потребам підрозділів фіскальної служби, спостерігається перенавантаження в пікові часи;
- не завжди коректне резервне копіювання та відновлення даних, досить часто виникає розбіжність даних після відновлення;
- можуть з'являтися проблеми у застосуванні системи кодування та шифрування даних під час з'єднання серверів Oracle та MS SQL-Server;
- дублювання даних.

Наразі в багатьох організаціях використовується технологія Oracle Grid. Вона надає завадостійкі та повні рішення для консолідації гетерогенних розподілених баз даних у єдину Grid-систему. Для зберігання інформації митного оформлення використовується СУБД Oracle 10G. Це не дуже нова версія, але й вона дає можливість створювати й виконувати додатки в середовищі Grid. Проте потрібно лише встановити необхідне програмне забезпечення і обладнання, що дозволить об'єднати комп'ютери в Grid. Йдеться про настання епохи Grid. Оскільки ж ідея Grid прозора і всім зрозуміла, а також дозволяє економити кошти й ефективніше використовувати наявні обчислювальні потужності, то, звісно, настання епохи Grid зупинити не можна.

Під час упровадження Grid виникають питання щодо консолідації та розміщення даних у Grid-системах. При консолідації даних у Grid-системах порушується проблема, пов'язана з тим, що об'єкти бази даних, а саме таблиці, які були у розподіленій базі даних на різних вузлах завдяки горизонтальній фрагментації, об'єднуються в одну велику таблицю. Для прискорення доступу та обробки даних великих об'єктів використовуються різні типи секціонування. Отже, для розв'язання проблеми швидкодії Grid потрібні розробити метод вибору типу секціонування та дослідити швидкодію Grid-системи.

Аналіз останніх досліджень і публікацій. Нині все частіше в організації обчислень застосовуються поділ робіт, даних і процесорних потужностей, а також інші режими взаємодії, що передбачають використання розподілених ресурсів. Це змушує звертати особливу увагу на забезпечення зв'язку систем між організаціями та всередині їх. Крім того, підприємства можуть значно економити кошти за рахунок передачі деяких елементів своєї IT-інфраструктури на аутсорсинг. Для організації таких Grid-систем проводять консолідацію додатків та даних.

Багато авторів досліджують питання консолідації даних у Grid-системи. Так, у праці [1] досліджено формалізовані характеристики якості консолідованих даних у просторах даних. Уведено поняття корисності даних, отриманих з джерел даних. Розроблено архітектуру системи оцінювання якості різнотипних даних. Уведено метамову для формування запитів користувачів до різнотипних джерел. Розроблено структури даних для опису інформаційних продуктів та схеми метаданих. Досліджено якість даних у просторі даних туризму та експериментально визначено значення коефіцієнтів характеристик якості даних. У дослідженні [2] розглянуто Grid-технології як важливий, перспективний для України напрямок в організації IT-інфраструктур для корпоративного співробітництва і спільного використання розподілених обчислювальних та інформаційних ресурсів в урядуванні, бізнесі, науці й освіті. Виконано порівняння та аналіз

можливостей розподілених технологій, типів Grid-систем як засобу спільного використання ресурсів; напрямків подальших досліджень, використання, розвитку та впровадження Grid-технологій. Підкреслено необхідність координації праць вітчизняних учених і фахівців у даному напрямку, який розвивається дуже динамічно. Досліджено проект та перспективи створення національної Grid-інфраструктури для розвитку українського суспільства [3]. Розглянуто технологію та інфраструктуру Grid, яка підтримує спільне і скоординоване використання різномірних ресурсів у динамічних, розподілених віртуальних організаціях, дозволяючи з географічно розосереджених компонентів, що застосовуються в різних організаціях із різними правилами роботи, створювати віртуальні обчислювальні системи, здатні спільно підтримувати необхідний рівень обслуговування [4]. Зокрема, вільно поширюваний інструментарій Globus Toolkit став фактичним стандартом конструювання Grid-систем. У праці [5] розглянуто методи та моделі підвищення швидкодії серверу Oracle.

Мета статті – запропонувати метод переходу від архітектури розподілених гетерогенних баз даних до Grid-архітектури у спеціалізованих інформаційних системах.

У розглянутих працях проводились дослідження методів і технологій консолідації даних, але питання швидкодії та надійності Grid під час консолідації залишається відкритим. У дослідженні [5] наведено методи підвищення швидкодії серверу, коли він працює у режимі Single instance.

Виклад основного матеріалу. Сучасна модель інформаційної системи департаменту організації митного контролю та оформлення Державної фіскальної служби України для користувача – це єдина логічна база даних (рис. 1), але насправді вона становить сукупність баз даних, розташованих на вузлах розподіленої бази даних (тобто на комп'ютерах або серверах, на яких встановлено програмне забезпечення СКБД і зберігаються дані).

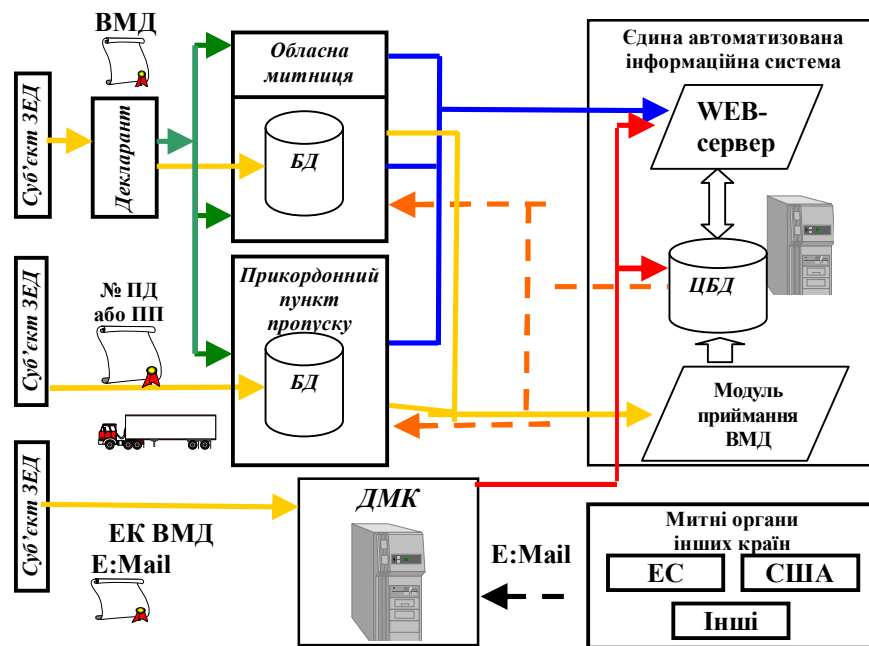


Рис. 1. Архітектура EAIC департаменту митного контролю та оформлення

До даних на всіх вузлах можна одночасно звертатися і модифікувати їх, використовуючи мережу. Кожна база даних керується власною локальною СКБД. Усі сервери баз даних у розподіленій базі даних співпрацюють, щоб підтримувати погодженість глобальної бази даних. За узгодженість даних відповідає схема фрагментації РБД. На всіх серверах РБД департаменту організації митного контролю та оформлення використовується СКБД Oracle 10G. Усі сервери обласних митниць пов'язані між собою високошвидкісною мережею Fibre Channel, швидкість передачі даних 3 Gbt/c. Це дозволяє об'єднати Центральний сервер та сервери обласних митниць у Grid-систему й застосувати модель консолідації даних (рис. 2).



Рис. 2. Модель консолідації даних Grid-систем

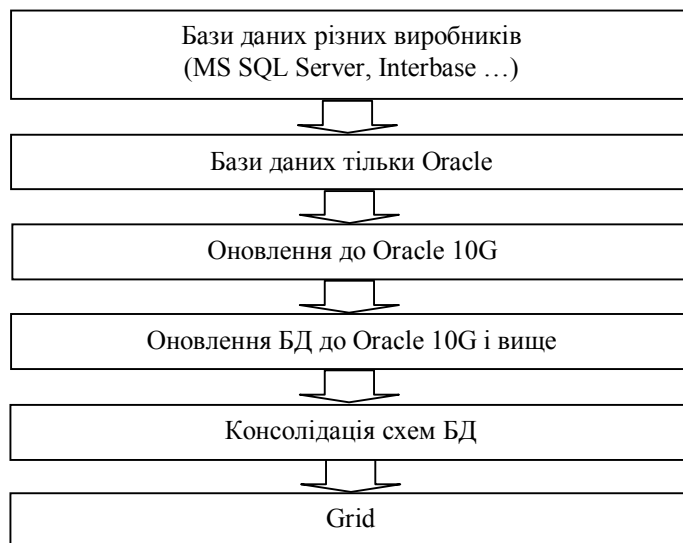


Рис. 3. Етапи консолідації для переходу до Grid

Основні переваги Grid:

1. Координоване використання безлічі малих серверів як єдиного загального ресурсу, які в сукупності працюють як один великий сервер.
2. Динамічне виділення ресурсів.
3. Швидкодія за вимогами.
4. Спільне використання та розподілення інформації.
5. Автоматичне ребалансування навантаження.
6. Висока доступність.

Під час використання методу консолідації дані збираються з декількох первинних систем та інтегруються в одне постійне місце зберігання. Воно може бути використано для підготовки звітності й проведення аналізу, як у випадку переходу до Grid, або у випадку операційного складу даних. Процес заповнення Сховища даних складається з трьох етапів:

1. Консолідація платформ додатків.
2. Консолідація самих додатків.
3. Консолідація даних.

Перевагою консолідації даних є те, що цей підхід дозволяє здійснювати трансформацію значних обсягів даних (реструктуризацію, узгодження, очистку та/або агрегування) у процесі їх передачі від первинних систем до кінцевих місць зберігання. Деякі складнощі, пов'язані з даним підходом, – це значні обчислювальні ресурси, потрібні для підтримки процесу консолідації даних, а також істотні ресурси пам'яті, необхідні для підтримки кінцевого місця зберігання. Та з урахуванням постійного вдосконалення апаратних і програмних засобів це не проблема.

На рис. 3 зображено етапи, які потрібно пройти для перетворення різних апаратних і програмних засобів, що існують у системі, в Grid.

Під час об'єднання БД в Grid великі таблиці, якими незручно керувати, та на SQL-запити витрачається дуже багато часу для пошуку необхідної інформації. Механізм Automatic Storage Management (ASM) співіснує з технологією секціонування, а саме Automatic Storage Management дозволяє розкидати дані по дисках і автоматично виконує їх перерозподіл. Ці дві технології не виключають, а доповнюють одна одну: секціонування працює на “верхньому поверсі” схеми розподілу даних, ASM – на “нижньому”. Тільки за допомогою секціонування можна домогтися *рівного* розміру секцій та ідентичного при цьому секціонування таблиць.

Відомі такі типи секціонування:

- діапазонне;
- за індексами;
- за інтервалом;
- хеш-секціонування;
- за списком;
- композитне;
- еквісекціонування і посеційне з'єднання;
- посиальне;
- секціонування за віртуальним стовпцем;
- системне.

Алгоритм секціонування таблиць (рис. 4).

Цей алгоритм базується на необхідності вибору типу й методу секціонування залежно від обсягу та типу інформації, яка зберігається в таблиці, та аналізу SQL-запитів, які

найчастіше використовуються в цих системах. Не кожна таблиця потребує секціонування. Незважаючи на те, що секціонування прозоре щодо програми, воно ускладнює адміністрування та реалізацію заданих об'єктів. У той час як секціонування може запропонувати значні переваги, використовувати його для маленьких таблиць недоцільно.

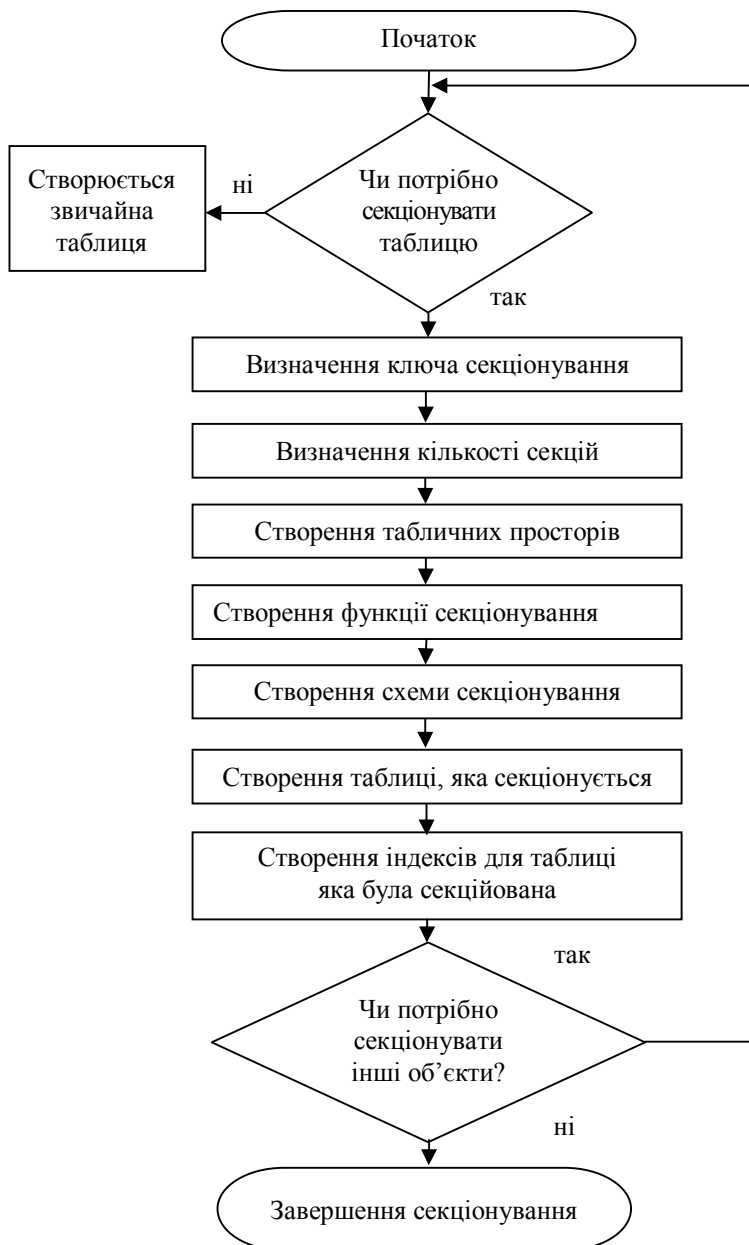


Рис. 4. Алгоритм секціонування таблиць

Дослідження типів секціонування

Для тестування типів секціонування використовуватимемо дані, що зберігаються у демо-БД, які стандартно поставляються для тестування з сервером Oracle: таблицю Sales, в якій зберігається інформація щодо продажу товарів схеми користувача SH, в якій зберігаються дані торговельної фірми. Вона містить 1 002 12 записів, за замовчанням ця таблиця не секціонована. Створюємо декілька таблиць, які були секціоновані за різними типами, до яких перенесемо інформацію з таблиці Sales (табл. 1)

Таблиця 1

Типи секціонування таблиць

Назва таблиці	Тип секціонування
S1	діапазонне секціонування
S2	секціонування за індексами
S3	секціонування за інтервалом
S4	хеш-секціонування
S5	секціонування за списком
S6	компонентне секціонування
S7	еквісекціонування і посеційне з'єднання
S8	посилальне секціонування
S9	секціонування за віртуальним стовпцем
S10	системне секціонування

На основі Алгоритму створення секціонованої таблиці створюємо таблиці з різними типами секціонування. Для тестування обрано 4 звичайних комп'ютери, що об'єднані у Grid (Процесор Intel 5i, ПЗУ 4 Gb, 2 HDD 750Gb), на яких встановлено демо-версія Oracle 11G, але і в такій конфігурації отримано необхідні результати.

Тестування швидкодії запитів визначалось за таким анонімним блоком.

```
DECLARE
  start_time CHAR(5);
  finish_time CHAR(5);
  elapsed_time NUMBER(5);
BEGIN
  /* Отримаємо системний час у секундах після опівночі */
  SELECT TO_CHAR(SYSDATE, 'SSSSS')
  INTO start_time FROM sys.dual;
  -----
  -- Виконуємо потрібний SQL запит
  -----
  /* Знову отримаємо системний час у секундах після опівночі */
  SELECT TO_CHAR(SYSDATE, 'SSSSS')
```

```

        INTO finish_time FROM sys.dual;
        /* Вираховуємо час, що був витрачений на запит */
        elapsed_time := finish_time - start_time;
        /* Записуємо результати у відповідну таблицю */
        INSERT INTO results VALUES (... , elapsed_time);
    END;

```

У табл. 2 наведено результати тестових запитів у секундах. Як бачимо з цієї таблиці, секціонування значно підвищує швидкість запитів, у яких використовуються агрегатні функції. На строкові функції це майже ніяк не впливає. Значно пришвидшує DML-операції. Тому вибір типу секціонування доцільно здійснювати, виходячи з того, які SQL-запити найчастіше генеруватимуться в інформаційній системі.

Таблиця 2

Результати тестів

SQL \ Табл. Sales	S1 мс	S2 мс	S3 мс	S4 мс	S5 мс	S6 мс	S7 мс	S8 мс	S9 мс	S10 мс
Агрегатні функції (min, avg, max)	6	5	4	4	3	5	4	3	5	6
Строкові функції (init cap, lower, upper)	9	10	11	13	12	12	13	14	15	12
Insert (100 записів)	3	1	1	1	1	1	1	1	1	1
Update (за датами)	5	3	2	2	2	2	2	2	2	2

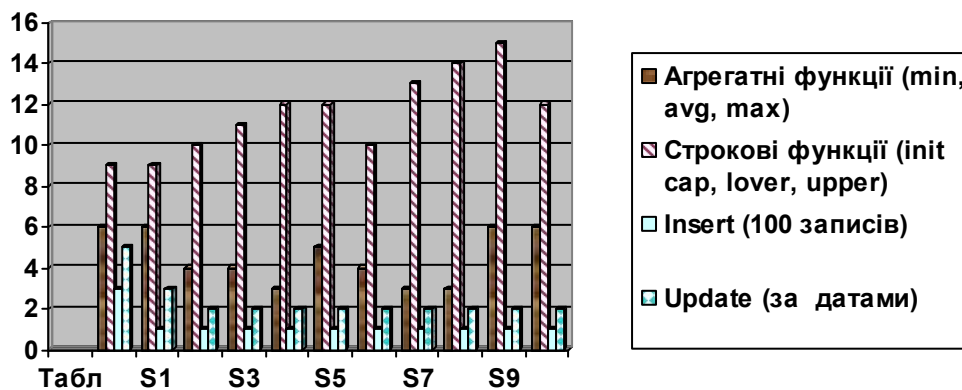


Рис. 5. Результати тестів секціонування таблиць

З отриманих результатів (рис. 5) видно, що секціонування значно підвищує швидкість виконання агрегатних функцій, особливо, коли виконуються групові функції і групування відбувається за діапазонами секціонування, ніяк не впливає на швидкість виконання строкових функцій (можливо тому, що всі секції розміщувались на одному диску), трішки прискорює операцію вставки й значно підвищує операцію Update, якщо умова операції збігається зі стовпцем, за яким відбувалось секціонування.

Висновки з даного дослідження та перспективи подальших розвідок у даному напрямку. В результаті проведеного дослідження набув подальшого розвитку метод керування консолідацією баз даних під час переходу від архітектури розподілених баз даних до Grid-систем. За допомогою цього методу розроблено рекомендації, щодо етапів і методів консолідації мережі, додатків, платформ і даних. При об'єднанні таблиць із різних серверів розподіленої бази даних формується таблиця, яка має дуже великий розмір. Для прискорення обробки даних, що зберігаються в цій таблиці, використовують різні типи секціонування. Для проведення секціонування був сформульований алгоритм, котрий можна використовувати для секціонування великих таблиць. Таким чином, у реальних Grid-системах виникає задача визначення оптимального типу секціонування залежно від розміру таблиці з інформацією, яка зберігається, та SQL-запитів, що найчастіше генеруються до цих таблиць. Сфера застосування методів обмежена системами керування базами даних, які використовують Oracle, та іншими СУБД, що підтримують Grid-технологію.

До напрямків подальших розвідок з даного напрямку зараховують більш нові версії Oracle, а саме версію 12c, яка дозволяє на базі Grid створювати, організовувати та впроваджувати хмарні технології.

Список використаних джерел:

1. Шаховська Н. Б. Проблеми якості консолідованих даних у просторах даних / Н. Б. Шаховська, О. Ю. Пшеничний, І. М. Чорней // Системи обробки інформації. – 2011. – № 3. – С. 80–84.
2. Матов О. Я. Перспективні інформаційні технології та розвиток Grid-систем у високопродуктивних глобально-розподілених обчислювальних інфраструктурах корпоративної співпраці / О. Я. Матов, І. О. Храмова // Реєстрація, зберігання і обробка даних. – 2004. – Т. 6. – № 1. – С. 85–98.
3. Згуровський М. З. Створення національної Grid-інфраструктури для забезпечення наукових досліджень / М. З. Згуровський, А. І. Петренко, Г. Д. Кисельов // Інформаційні технології в освіті. – 2009. – Вип. 4. – Т. 4. – С. 12–17.
4. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration / I. Foster, C. Kesselman, J. Nick, S. Tuecke // Open Grid Service Infrastructure WG, Global Grid Forum. – 2012. – June 22.
5. Методи та модель системи підвищення продуктивності серверу ORACLE // Б. І. Мороз, Л. В. Кабак, О. Н. Молотков, О. В. Трофімов // Вісник Академії митної служби України. Серія: “Технічні науки”. – 2013. – № 2. – С. 33–42.