



УДК 004.67:004.424.7

НОВІТНІ СИСТЕМИ ЗБЕРЕЖЕННЯ КОНСОЛІДОВАНИХ ДАНИХ

Б. Мельник, М. Когут

*Львівський національний університет імені Івана Франка,
79008, м. Львів, проспект Свободи, 18,
E-mail: melb2@ukr.net*

*Розглянуто систему для збереження консолідованих даних. Визначено поняття та етапи консолідації даних. Сформульовано критерії оцінювання та основні характеристики систем збереження даних. Наведено приклад сучасної реалізації системи корпоративного рівня. Окреслено основні технологічні рішення, притаманні цій системі.
Ключові слова: інформаційна система, консолідація даних, сховище даних, система збереження даних, технологія.*

У сучасних великих підприємствах, екстериторіальних і транснаціональних корпораціях керівники верхнього і середнього рівня здійснюють свої управлінські функції, спираючись на великі обсяги інформації. Для підтримки власних управлінських рішень вони використовують відповідні автоматизовані корпоративні інформаційні системи, які оперують даними, отриманими зазвичай з чисельних і різнотипних джерел інформації. Щоб узгодити між собою усі масиви використовуваних даних необхідно провести їхню консолідацію.

Під консолідацією даних розуміють застосування комплексу методів і процедур, спрямованих на вилучення даних із різноманітних джерел, забезпечення необхідного рівня їхньої якості та інформативності, перетворення їх у єдиний формат, який дає змогу завантажити їх у сховище даних чи аналітичну систему [1].

Процес консолідації даних передбачає три основні етапи:

1. вилучення потрібної інформації з різноманітних джерел;
2. перетворення інформації у дані, які можуть бути у майбутньому використані у середовищі інформаційної системи;
3. завантаження даних у сховище даних для довготривалого зберігання.

Для реалізації третього етапу необхідно створити відповідне сховище даних, яке забезпечить ефективне використання консолідованих даних під час інтелектуального аналізу функціонування складних систем.

Для користувача консолідованих даних найважливішими характеристиками сховища даних є його продуктивність, доступність, гнучкість, масштабованість, консолідованість та безпека даних. Щоб забезпечити оптимальне співвідношення усіх цих характеристик необхідно застосувати системний підхід у процесі створення програмно-технічних засобів для збереження даних. Він дає змогу створити високоєфективні системи збереження даних (СЗД, *System Storage*).

Сьогодні ціла низка ІТ компаній розробляє СЗД для різного рівня управління. Багато з них використовують передові технології запропоновані такими ІТ гігантами як IBM, Dell, Fujitsu, NetApp.



Як приклад реалізації корпоративної СЗД варто розглянути поширену нині систему з архітектурою IBM XIV Storage System, створену на базі дисккових станцій серії DS 8000 [2]. Ціла низка рішень, реалізованих у рамках згаданої архітектури, зокрема останньої версії Generation 3 XIV, нині стає загальноприйнятим стандартом [3].

Система IBM XIV Storage System – це сукупність модулів, кожен з яких є самостійним комп'ютером з власною оперативною пам'яттю, кешем, дисками та іншими характерними компонентами. Усі модулі об'єднані у мережу.

Програмна платформа системи складається з модифікованого ядра Linux, а також низки програмних модулів з відкритим програмним кодом. Це дає змогу забезпечити взаємодію не лише з Linux, але й з операційними системами Windows, AIX, Solaris, HP-UX [4], а також операційними системами мобільних пристроїв.

Підвищена продуктивність, яка характерна системі, забезпечується цілою низкою чинників, зокрема, за рахунок використання високопродуктивних керованих контролерів запису даних, швидкісних каналів передавання даних та швидкісних носіїв інформації. Керовані контролери створені на основі багатопроцесорної системи (*Symmetric Multiprocessing, SMP*) багатоядерних (до 8 ядер) RISC процесорів класу IBM POWER6 і вище [3].

Застосування шини послідовної передачі даних PCI Express [5] дає змогу здійснювати обмін даними зі швидкістю до 8 Гбіт/с. Пропускна здатність системи при обміні даними між кеш-пам'яттю і диском може досягати 240 Гбіт/с [6]. Контролери 4- і 8-портових хост-адаптерів Fibre Channel підтримують одномодові і багатомодові кабелі і з пропускною здатністю 8 Гбіт/с [3].

У системі використовують швидкісні твердотільні диски (*solid-state drive, SSD*) і жорсткі диски з інтерфейсом SAS (*Serial Attached SCSI*) і NL-SAS (*Near Line SAS*). Застосування iSCSI-портів (*Internet Small Computer System Interface*) дає змогу створити відповідні мережеві з'єднання у швидкісних Gigabit Ethernet мережах [6].

Ще одним важливим чинником підвищення продуктивності є застосування новітніх алгоритмів кешування: Intelligent Write Caching, Adaptive Multistream Pre-fetching, Cooperative Caching.

Алгоритм Intelligent Write Caching передбачає використання різних механізмів кешування операцій читання і запису на диск. Інтелектуальний вибір необхідного механізму залежить від потреб поточної задачі [6].

Алгоритм Adaptive Multistream Pre-fetching (AMP) є адаптивним та асинхронним. Його адаптивність полягає у зміні величини попередньої вибірки даних залежно від робочого навантаження на систему. Це зменшує "забруднення" кешу зайвими даними. Асинхронність передбачає застосування попередньої вибірки не лише у випадку кеш-промаху, але й тоді, коли відбулося кеш-попадання [7].

Застосування алгоритму Cooperative Caching дає змогу створювати спільний кеш-масив для різних серверів. Це усуває дублювання спільно використовуваних різними додатками даних [8].

За рахунок повного апаратного резервування на основі архітектури серверів IBM Power System СЗД забезпечує найвищий рівень доступності до даних (більше "п'яти дев'яток") та безвідмовності. Високу гнучкість системи, зокрема, забезпечують за рахунок прискореного паралельного оновлення мікрокоду системного програмного забезпечення і "гарячої" заміни елементів обладнання [9].

Ще одним чинником підвищення гнучкості системи, а також її надійності і продуктивності, стала технологія автоматичного переміщення даних Easy Tier. Вона



дає змогу динамічно переносити в автоматичному режимі найбільш активно використовувані дані на швидкісні твердотільні диски [10].

Шляхом зміни ємності дисків і оперативної пам'яті, збільшення кількості адаптерів і процесорних комплексів СЗД може бути масштабована за обсягами даних зберігання вище 2 петабайт. Полегшення масштабованості досягається шляхом віртуалізації за допомогою технології PowerVM [9].

СЗД дає змогу консолідувати дані, які знаходяться у вже існуючих системах, шляхом створення багаторівневої структури збереження даних. Отже, є можливість для подальшого розвитку та інтеграції інформаційних систем, які тривалий час експлуатують відокремлено.

Розробники СЗД звернули значну увагу на забезпечення безпеки даних. У системі використовують резервування дисків типу RAID-5, RAID-6, RAID-10.

Кожен з жорстких дисків забезпечений захистом за технологією Seagate Secure, яка відповідає найвищим вимогам забезпечення конфіденційності інформації, регламентованих американським національним стандартом NSTISS #11 (*National Security Telecommunications and Information Systems Security Policy*) [11]. Ця технологія передбачає безперервне апаратне шифрування даних за алгоритмом блочного шифрування Advanced Encryption Standard [12]. Шифрування відбувається на швидкості роботи інтерфейсу.

Технологія Seagate Secure орієнтована на розв'язання проблем, які виникають в корпоративних інформаційних системах. Вона дає змогу модифікувати систему захисту відповідно до змін законодавства щодо захисту інформації, проводити багатofакторну ідентифікацію користувача з можливим блокуванням доступу до даних, за допомогою програмної оболонки IBM Security Key Lifecycle Manager [13] здійснювати централізоване управління ключами шифрування. Криптографічне стирання даних з дисків унеможливує витікання корпоративної інформації під час виведення дисків з експлуатації чи їх утилізації. У випадку перепрофілювання комп'ютера чи диска без небезпеки для збереження конфіденційності усі дані попереднього користувача, які залишаються на диску, стають недоступними після вилучення ключа шифрування [11].

Для діагностики і контролю справності дисків застосовують інтелектуальну діагностичну панель Light Path Diagnostics [14], а також механізм ведення журналу, де фіксують зміни системних налаштувань [2].

СЗД IBM Storage System DS 8000 Series розрахована на обслуговування великих та розподілених компаній, у яких багато користувачів системи на локальному рівні, а також віддалені підрозділи. Тому дуже важливо забезпечити надійний і швидкий доступ до даних у багатокористувацькому і віддаленому режимах. Для цього у системі реалізовано механізми локального і віддаленого копіювання даних.

Для створення локальних копій використовують технологію IBM FlashCopy, яка дає змогу здійснювати асинхронне копіювання у будь-який момент часу без додаткового навантаження на сервер. Під час копіювання у фоновому режимі реалізована можливість збереження копій лише змінених даних, що зменшує часові затрати та забезпечує економію дискового простору. Програмне забезпечення IBM Tivoli Storage FlashCopy Manager допомагає забезпечити найвищий рівень захисту для одночасно працюючих додатків, які обслуговують різні системи керування базами даних, наприклад, IBM DB2, SAP, Oracle чи Microsoft SQL Server [6], а також підтримує сховища даних VSS (*Visual SourceSafe*) корпорації Microsoft, EMC



Symmetrix, SVC (*SAN Volume Controller*) фірми IBM, файлові системи NetApp на платформах AIX, HP-UX, Linux і Solaris, програмне середовище VMware [8].

Віддалені дзеркальні копії даних в системі створюють за допомогою таких технологій IBM System Storage як Metro Mirror чи Global Mirror. Технологія Metro Mirror забезпечує синхронне копіювання даних на відстані до 300 км. Синхронність означає, що доступ до даних на основному диску може бути дозволений лише після завершення копіювання на диску-копії. Це є причиною часових затримок у використанні даних [15]. Копіювання за технологією Global Mirror [14] передбачає затримку (3-5 с) у створенні віддаленої копії без обмеження відстані між окремими сховищами даних.

СЗД IBM Storage System DS 8000 Series є потужним засобом консолідації даних в інформаційних системах компаній різного масштабу. Застосування її, або її аналогів, дає змогу здійснювати багаторівневе управління різноманітними системами і процесами.

1. Орешков В.И. Консолидация данных – ключевые понятия [Електронний ресурс] / Орешков В.И., Палкин Н.Б. // Глава из книги "Бизнес-аналитика: от данных к знаниям – <http://www.cfin.ru/itm/olap/cons.shtml>
2. Space IT. Системы хранения данных компании IBM [Електронний ресурс] – http://www.space-it.com.ua/solutions/shd_server/shd/sds-ibm/
3. IBM. IBM System storage DS8000 series [Електронний ресурс] – <http://www.03.ibm.com/systems/ru/storage/disk/ds8000/overview.html>
4. IBM. Tivoli storage FlashCopy Manager [Електронний ресурс] – <http://www-03.ibm.com/software/products/ru/tivoli-storage-fladhcory-manager>
5. IP-телефония. Дисковые системы хранения данных IBM Storage System DS 8000 [Електронний ресурс] – <http://www.i-p.kiev.ua/ibm/ds8000>
6. MUK. Система хранения данных IBM XIV [Електронний ресурс] – http://www.muk.ua.catalog/syst_savedata/stor_ibm/stor_ibm_5
7. Sitchedfabric. IBM DS8000 write cache -all 6GB of it [Електронний ресурс] – <http://www.switchedfabric.co.za/blog/2012/01/30/ibm-ds8000-write-cache-all-6-gb-of-it>
8. Sarkar P. Efficient Cooperative caching using hints [Електронний ресурс] / Sarkar P., Hartman J. – <http://www.cs.arizona.edu/projects/swarm/papers/ccache/>
9. ИЦ. Телеком-сервис. IBM POWER7. Power Systems – новый уровень производительности [Електронний ресурс] – http://www.tls-group.ru/server/-IBM/ibm_p_ser.html
10. Karma. Дисковая система среднего класса IBM Storwize V7000 [Електронний ресурс] – <http://www.karma-group.ru/Sites/karma/Uploads/V7000.1851.....pdf>
11. SecureIT. The latest IT secure news [Електронний ресурс] – http://www.secureit.online.com/news/read_article.php?i=1204
12. Wikipedia. Advanced Encryption Standard [Електронний ресурс] – http://www.uk.wikipedia.org/wiki/advanced_encryption_standard
13. IBM. IBM Security Key Lifecycle Manager [Електронний ресурс] – <http://www-03.ibm.com/software/products/ru/key-lifecycle-manager>
14. IBM. Light path diagnostics [Електронний ресурс] – <http://www.ibm.com/us/en>
15. Intellimagic. IBM Metro Mirror (PPRC) [Електронний ресурс] – <http://www.intellimagic.com/resources/copy-services-described/pprc>



16. Wikipedia. IBM Global Mirror [Електронний ресурс] – http://www.en.wikipedia.org/wiki/IBM-Global_Mirror

MODERN STORAGE SYSTEMS CONSOLIDATED DATA

B. Melnyk, M. Cogut

*Ivan Franko National University of Lviv,
Svobody av., 18. UA-79008 Lviv, Ukraine*

In this article we consider the system to save the consolidated data. The notion of stages and data consolidation. Formulated criteria for evaluation and the main characteristics of data storage systems. An example of a modern system for the implementation of corporate level. The basic technological solutions specific to this system.

Keywords: information system, data consolidation, data warehouse, storage system, technology.

НОВЕЙШИЕ СИСТЕМЫ СОХРАНЕНИЯ КОНСОЛИДИРОВАННЫХ ДАННЫХ

Б. Мельник, М. Когут

*Львовський національний університет імені Івана Франка,
79008, г. Львов, проспект Свободи, 18*

В статье рассмотрено систему для сохранения консолидированных данных. Определено понятие и этапы консолидации данных. Сформулировано критерии оценки и основные характеристики систем сохранения данных. Приведен пример современной реализации системы для корпоративного уровня. Описаны основные технологические решения, присущих этой системе.

Ключевые слова: информационная система, консолидация данных, хранилище данных, система сохранения данных, технология.

