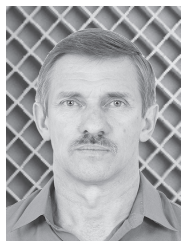




## ТЕХНОЛОГІЯ ПОБУДОВИ ОПТИМАЛЬНОЇ МОДЕЛІ СХОВИЩА ДАНИХ У ДЕРЖАВНИХ ОРГАНАХ



**М.Б. Вітер**, канд. фіз.-мат. наук

**Постановка проблеми.** Унаслідок підвищення рівня автоматизації документообігу спостерігається лавиноподібне зростання інформації, яка зберігається в електронному вигляді. Тенденції ринку останніх років свідчать про різке збільшення частки неструктурованих даних у загальному обсязі інформації, до яких можна віднести: цифрову графіку, відео-, аудіодані, результати комп'ютерного моделювання, довідкову інформацію тощо. За оцінками фахівців, їхній обсяг щорічно збільшується приблизно на 80%, а структурованих даних – на 20–30% [1].

Як свідчить досвід [2], 90% світового обсягу збережуваних даних – це дубльована інформація, причому до більш ніж 80% цих даних повторно ніхто не звертається. Слід зазначити, що основна частина цих даних зберігається на дорогих пристроях досить довго, що є економічно не вигідним, оскільки потребує додаткових затрат часу і коштів. Тому задача розробки ефективної технології збереження даних в електронних сховищах є вельми актуальною.

**Аналіз останніх досліджень і публікацій.** У сучасних дослідженнях [3–7] більше уваги приділяється загальним питанням організації корпоративного сховища електронних даних, підходам до способів архівації, пошуку

інформації тощо. При цьому розглядається переважно структурована інформація.

Важливою слід назвати і задачу оптимізації розміщення (як у просторі, так і в часі) великих обсягів (терабайти і більше) неструктурованої архівної інформації і забезпечення ефективного доступу до неї користувачів. Оптимальне вирішення її має передбачати не лише раціональне збереження даних, але й максимальну ефективність їхнього використання. При цьому вартість зберігання інформації не повинна бути вищою від вартості самої інформації.

**Мета статті** – аналіз сучасних технологій зберігання великих обсягів даних і формулювання рекомендацій щодо ефективного застосування їх у державних органах.

**Виклад основного матеріалу.** Інтенсивність роботи з різними документами неоднакова. Є документи, до яких часто звертаються (поточні матеріали), інші досить рідко використовуються для роботи (звіти, нормативні акти загального характеру тощо). Крім того, згідно з номенклатурою справ у державних органах для різних документів установлюються різні строки оперативного зберігання – від кількох місяців до постійного і тривалого (понад десять років). Після закінчення терміну зберігання документ знищується або відправляється в архів, звідки його можна отримати в разі потреби [8].

У таких державних органах, як Міністерство доходів і зборів України, Державна казначейська служба України, Державна служба статистики України й інших зберіга-

ються величезні обсяги звітної інформації, дані про платників, статистичні дані тощо. З часом ця інформація втрачає актуальність, використовується дедалі рідше. Тому з метою оптимізації витрат зберігання і забезпечення ефективності використання даних документи, з якими часто працюють, доцільно зберігати на швидких дорожчих носіях, а неактуальні – на повільних дешевших.

Одним із варіантів оптимізації збереження документів є розподіл їх між базою даних (оперативне сховище) і файловим сховищем. Така технологія використовується, наприклад, у модулі Storage Services системи електронного документообігу (СЕД) DIRECTUM (<http://www.directum.ru/system>). У цій системі документи, що знаходяться в оперативному обороті, можуть розташовуватися у сховищі SQL-сервера, а для рідко використовуваних документів і документів великого обсягу передбачена організація розподілених файлових сховищ. При цьому користувач може працювати з усіма документами з одного середовища. За рахунок винесення архівних даних зі сховища SQL-сервера збільшується швидкість їхнього відновлення і включення користувачів у роботу. Відновлення решти

частин системи може здійснюватись у фоновому режимі. На основі цієї технології можна побудувати декілька схем зберігання документів:

1. Усі документи зберігаються в оперативному сховищі. Це є доцільним на етапі впровадження СЕД до моменту накопичення критичного обсягу документів.

2. Наявність у системі оперативного і файлового сховищ. Це забезпечує високу швидкість роботи з документами, що знаходяться в оперативному обороті, при низькій вартості зберігання і практично необмеженому обсязі архіву документів.

3. Аналогічно до сховищ (пункти 1 і 2) можуть створюватись додатково оперативні і файлові сховища для документів великого обсягу: відео, графіка, карти тощо (рис. 1).

Серед сучасних технологій оптимізації збереження великих обсягів інформації також варто виділити технологію HSM (Hierarchical storage management – менеджмент ієрархічного зберігання). Ця технологія передбачає розміщення актуальної інформації на нижчому рівні на швидких носіях (жорсткі диски), а менш використовуваної інформації – на вищому рівні на дешевших, менш продук-



Рис.1. Варіант схеми зберігання документів

тивних носіях (магнітооптичні диски, диски WORM, магнітні стрічки). На нижчому рівні дані знаходяться строго регламентований час. Якщо вони за цей час не використовувались, система переміщує їх на наступний рівень зберігання. Черговий запит до файла запускає процес переміщення його назад на жорсткий диск нижчого рівня ієрархії. Системи HSM, як правило, включають у себе два або три рівні зберігання інформації (рис. 2.) Процес міграції здійснюється автоматично при виконанні відповідних критеріїв: інтервал часу з моменту створення/оновлення файлів або досягнення межі заповнення сховища поточного рівня. Налаштування розкладу архівування за годинами, тижнями, місяцями згідно з документообігом державного органу надає можливість досягти максимальної продуктивності.

Технологія HSM має ряд недоліків, зокрема:

- досить низькою є оперативність доступу до даних. Так, чим більше часу минуло з моменту створення, оновлення або останнього доступу до файла, тим більше часу потрібно для його завантаження. Найбільш гостро ця проблема постає в разі одночасного звернення декількох користувачів до старих файлів.

- системи HSM досить вимогливі стосовно апаратного забезпечення.

У [9] описано приклад системи електронного архівування з ієрархічним зберіганням даних, яка реалізована на основі обладнання IBM – серверів IBM pSeries 550 під управлінням ОС AIX 5L і двох типів носіїв даних: з оперативним доступом – масивів IBM TotalStorage DS4300 і з доступом на вимогу – стрічковою бібліотекою IBM TotalStorage 3584. Сервери, система зберігання і стрічкова бібліотека об'єднані в єдину мережу зберігання даних (SAN) за допомогою двох FC-комутаторів IBM TotalStorage SAN-switch.

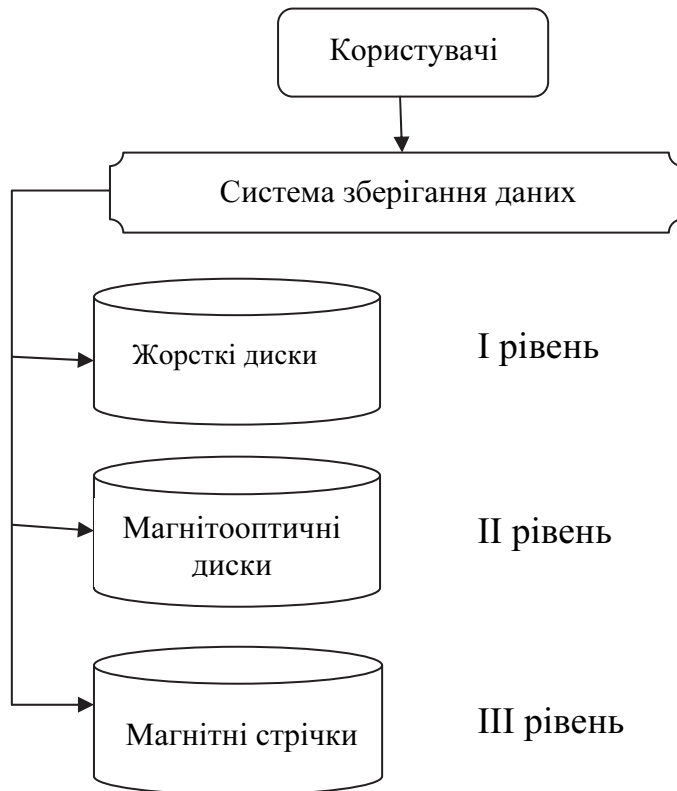


Рис. 2. Рівні зберігання даних

Технологія ILM (Information Lifecycle Management – управління життєвим циклом інформації), на відміну від HSM, враховує не лише частоту звертання до інформації, а й її якісний склад. Це спричинено появою нових актуальних типів даних, які потребують відповідної стратегії управління зберіганням: відеодані, web-дані тощо. У разі здійснення електронного листування часто необхідно зберігати всю його історію з безліччю вкладень.

Сучасні розробки в галузі технології зберігання сприяють появі нових видів ієрархій запам'ятовуючих пристроїв. Завдяки цьому багато сучасних інфраструктур зберігання тепер мають три і більше рівнів [2].

Так, п'ятирівнева структура зберігання архівних даних може мати такі рівні:

- I – відмовостійкий високопродуктивний дисковий накопичувач для найбільш важливих даних;
- II – недороге сховище з використанням оптичних каналів Fibre Channel для некритичних даних;
- III – дешева повільна буферна пам'ять для рідко використовуваних файлів;
- IV – стрічкові накопичувачі високої ємності для ведення архіву;
- V – виділені (відокремлені) стрічкові сховища для довготривалого зберігання.

У сфері програмного забезпечення технології ILM одним із лідерів є компанія EMC Documentum. Якщо організація вже використовує систему зберігання EMC, єдиний віртуальний дисковий простір і програмне середовище Documentum, то впровадження ILM здійснюється шляхом відповідного налаштування програмного забезпечення. В інших випадках необхідно здійснювати певну послідовність дій, що включає в себе як організаційні, так і інфраструктурні зміни [10].

Більшість систем управління документами (Document Management System, DMS) і систем управління корпоративним контентом (Enterprise Content Management, ECM) підтримують функції ILM.

## Висновки

На підставі аналізу сучасних технологій і засобів зберігання архівних даних можна сформулювати загальну методологію оптимізації системи зберігання архівних даних у державних органах.

Для формування системи для архівування електронних даних необхідно максимально враховувати вимоги сучасного документообігу в державних структурах [8], забезпечуючи: мінімальну вартість інфраструктури зберігання; оперативність доступу; особливості зберігання різних типів електронних даних; високий ступінь захисту від втрати інформації; можливість управління термінами гарантованого зберігання даних тощо. При цьому рішення для архівування даних повинні мати гнучку масштабовану архітектуру.

Для зберігання великих обсягів інформації найкращим рішенням є багаторівнева структура архівного зберігання електронних даних, яка передбачає наявність оперативних і архівних сховищ. Взаємна відповідність таких сховищ задається під час їхнього створення і може бути змінена.

Сучасне розв'язання поставленої задачі, як правило, включає в себе спеціалізовану систему зберігання архівних даних, програмне забезпечення, що реалізує механізми взаємодії з клієнтом і керування зберіганням, а також компоненти, що забезпечують взаємодію з суміжними системами (наприклад, система резервного копіювання і відновлення даних). При цьому обов'язково має формуватись гнучка схема міграції документів між сховищами. Новостворений документ розміщується у сховищі, що відповідає його виду документів, формуються відповідні параметри (час знаходження в оперативному доступі, умови і шляхи переміщення тощо), які задають умови зберігання і міграції. Протягом свого життєвого циклу документ також може бути переміщений вручну в інше сховище (оперативне чи в архівне). Використання workflow допомагає автоматизувати цей про-

цес. Інтелектуальні технології міграції даних (типу HSM і ILM) підвищують ефективність системи зберігання електронних даних.

Вибір технології залежить від завдань, які розв'язує в організації система електронного документообігу. Так, система HSM може використовуватись в органах державної влади, для документообігу яких характерним є: стандартизація документів (як правило, текстових, плани, звіти); чітко визначені часові параметри життєвого циклу документів тощо. Технологію ILM доцільно використовувати органам, які працюють з різними типами документів (мультимедійні, веб-документи), мають розгалужену територіально розподілену структуру, великий обсяг електронної пошти.

При запровадженні системи ILM необхідно:

- класифікувати інформацію відповідно до її цінності, місця розташування, вартості обслуговування, ресурсомісткості тощо;
- визначити кількість рівнів зберігання інформації;
- розробити політику міграції даних і правила доступу до них;
- здійснювати впровадження поступово, спочатку на рівні додатків (електронна пошта), а потім – повномасштабне.

Розвиток сучасних ІТ-технологій надає можливість ефективно здійснювати багаторівневе архівне зберігання електронних даних у різних формах: на жорсткому диску, стрічковому носії, оптичному носії, у CAS-системі, «хмарній» системі тощо. При цьому забезпечується нормальна робота під керуванням основних відомих операційних систем: Windows, UNIX, Linux або Mac OS X.

Раціональна політика організації моделі електронного сховища документів надає

значні переваги як щодо вартості зберігання даних, так і щодо продуктивності роботи системи, сприяє збалансуванню навантаження на сервери, мережу, збільшенню максимального обсягу збережених даних, підвищенню відмовостійкості системи в цілому і зниженню вартості зберігання.

#### ЛІТЕРАТУРА

1. *Кристоф Ланге*. Решения для архивирования и управления данными [Електронний ресурс] / Кристоф Ланге. – Режим доступу: <http://www.osp.ru/lan/2012/09/13017517>
2. *Элиот Кинг*. ILM все разложит по местам [Електронний ресурс] / Элиот Кинг. – Режим доступу: <http://www.osp.ru/win2000/2005/02/177481>
3. *Шаховська Н.Б.* Організація просторів даних для забезпечення якості консолідованих даних / Н.Б. Шаховська // Реєстрація, зберігання і оброб. даних. – 2011. – Т. 13. – №2. – С. 87–96.
4. *Асеев Г.Г.* Архітектура корпоративного сховища даних / Г.Г. Асеев // Вісн. Кн. палати. – 2010. – № 10. – С. 20–28.
5. *Носевич В.Л.* Проблеми довгострокового зберігання електронних архівних матеріалів у роботі Білоруського науково-дослідного центру електронної документації // Арх. України. – 2001. – № 3. – С. 65–74.
6. *Марков А.* Концепция построения электронного архива / А. Марков // Открытые системы. – 1997. – № 1. – С. 54–58.
7. *Юмин И.Ф.* Электронные документы и архивы: теория и практика/ И.Ф. Юмин // Отеч. арх.– 1999. – №3. – С.109–112.
8. Постанова Кабінету Міністрів України від 30.11.2011 № 1242 «Про затвердження Типової інструкції з діловодства у центральних органах виконавчої влади, Раді Міністрів Автономної Республіки Крим, місцевих органах виконавчої влади» [Електронний ресурс]. – Режим доступу: <http://zakon1.rada.gov.ua/laws/main/>.
9. *Егоров Андрей*. HSM – прошлое или будущее систем хранения данных? [Електронний ресурс]. – Режим доступу: <http://www.cnews.ru/reviews/free/infrastructure2007/articles/hsm.shtml>
10. *Андрей Шуклин*. ILM: еще одна дорогая игрушка? [Електронний ресурс]. – Режим доступу: <http://citforum.ru/nets/storage/ilm/>