УДК 004.852

Журавльов Павло Володимирович
студент
Національний технічний університет України «Київський політехнічний інститут»
Медведський Андрій Миколайович
студент
Національний технічний університет України «Київський політехнічний інститут»
Журавлев Павел Владимирович
студент
Национальный технический университет Украины «Киевский политехнический институт»
Медведский Андрей Николаевич
студент
Национальный технический университет Украины «Киевский политехнический институт»
Zhurzvlyov P.
student
National Technical University of Ukraine «Kyiv Polytechnic Institute»
Medvedskyi A.
student
National Technical University of Ukraine «Kyiv Polytechnic Institute»

ПРОБЛЕМА ВИБОРУ ТИПУ СХОВИЩА ДАНИХ ПІДПИЄМСТВА

ПРОБЛЕМА ВЫБОРА ТИПА ХРАНИЛИЩА ДАННЫХ ПРЕДПРИЯТИЯ

CHOOSING THE RIGHT TYPE OF ENTERPRISE STORAGE ARCHITECTURE

Анотація. *Досліджені різні типи файлових сховищ, їх проблеми та різниця між ними. Також сформована відповідність між вимогами до сховища та сховищами, що їх задовільняють.*

Ключові слова: *сховище даних, DAS, NAS, SAN, об'єктне сховище.*

Аннотация: *Исследованы разные типы файловых хранилищ, их проблемы и разница между ними. Так же сформировано соответствие между требованиями к хранилищам и хранилищами, которые их удовлетворяют.*

Ключевые слова: *хранилище данных, DAS, SAN, SAN, объектное хранилище.*

Summary. *Different types of file storage, their problems and the difference between them studied. Also storage requirements and storage types use cases studied.*

Key Words: *data storage, object storage, DAS, SAN, SAN.*

Закон Мура свідчить, що «за всю історію обчислювальної техніки, кількість транзисторів у щільній інтегральній схемі подвоювалась приблизно кожні два роки». В той час як ця думка, головним чином, стосується обчислювальної потужності комп'ютерів, цю ж закономірність можна побачити у світлі росту та зберігання даних. Об'єм даних, що генеруються сьогодні величезний. До 2020 р. він, як очікується, досягне 40 000 екзабайт (EB), що більше, ніж 5,2 ТБ даних на одну людину на планеті.

Діяльність сучасних компаній заснована на роботі з даними. Користувацькі дані збираються, використовуються та зберігаються у великих кількостях. При такій великій кількості даних стає складно підтримувати їх на стандартних засобах зберігання. Більшість даних існує в публічних та внутрішніх «хмарах», віддалених сховищах, організованих з використанням технологій віртуалізації.

Оскільки сховище даних повинно справлятися зі зростом кількості різноманітних даних, виникає задача порівняння та вибору серед різних архітектур сховищ з урахуванням їх переваг та потреб підприємства. Існує чотири основних типи сховищ:

- DAS (англ. Direct-attached storage) — система зберігання даних з прямим підключенням;

- NAS (англ. Network Attached Storage) — мережеве зберігання даних;
- SAN (англ. Storage Area Network) — мережа зберігання даних;
- Об'єктне сховище.

Direct Attached Storage

Direct Attached Storage — цифрове сховище, яке безпосередньо підключено до сервера, який отримує доступ до пам'яті, на противагу віддаленого доступу через мережу. Найбільш поширені типи пристроїв DAS — внутрішні / зовнішні жорсткі диски і оптичні приводи дисків (наприклад, CD / DVD-ROM). Сховище DAS — це фізичний диск, який може бути безпосередньо підключеним до сервера всередині або зовні, так само, як зовнішній диск для ПК.

До основних переваг DAS-систем можна віднести їх низьку вартість (у порівнянні з іншими рішеннями СЗД), простоту розгортання і адміністрування, а також високу швидкість обміну даними між системою зберігання і сервером. Власне, саме з цієї причини вони стали дуже популярні в сегменті малих офісів і невеликих корпоративних мереж. У той же час DAS-системи мають і свої недоліки — в першу чергу це висока вартість зберігання і управління даними внаслідок їх розкиданості по організації, а також вимушений простій мережі в момент додавання нових дисків і необхідність нарощування пам'яті або процесорної потужності сервера при перевищенні певного розміру дискового простору. Перевантаженість мережевого трафіку з додаванням нових серверів ускладнює проблему захисту даних, перешкоджає ефективному використанню ресурсів і т.д.

Хоча деякі DAS системи можуть бути розширені, вони мають більш низьку межу масштабованості, ніж інші типи пристроїв.

Network Attached Storage

Network Attached Storage — Зовнішній пристрій, який фізично з'єднаний з серверним оточенням по LAN або WAN мережі. На відміну від DAS пристроїв, які підключаються безпосередньо до сервера, NAS пристрої підключаються через існуючу мережу. Типові пристрої NAS є повністю автономними, і поставляються з власною операційною системоюі інтерфейсом адміністрування. Крім оптимізованої ОС, звільненої від всіх функцій, не пов'язаних з обслуговуванням файлової системи і реалізацією введення-виведення даних, NAS-системи мають оптимізовану по швидкості доступу файлову систему. NAS-системи проектуються таким чином, що вся їх обчислювальна потужність фокусується виключно на операціях обслуговування і зберігання файлів. Сама операційна система розташовується у флеш-пам'яті та встановлюється фірмою-виробником. У порівнянні з традиційними файловими серверами, NAS-пристрої є більш продуктивними і менш дорогими. В даний час практично всі NAS-пристрої орієнтовані на використання в мережах Ethernet (Fast Ethernet, Gigabit Ethernet) на основі протоколів TCP / IP. Доступ до пристроїв NAS здійснюється за допомогою спеціальних протоколів доступу до файлів. Найбільш поширеними протоколами файлового доступу є протоколи CIFS та NFS.

Так як пристрої NAS під'єднуються через мережу, вони не залежать від працездатності серверів, як DAS сховища.

Пристрої NAS ідеально пасують для людей чи компаній, що прагнуть обмінюватися файлами через мережу. Операції передачі даних можуть оброблюватися декількома серверами одночасно з великою кількістю з'єднань.

Простіший та зручніший метод доступу та керування, як наслідок, наявність своєї операційної системи та графічного інтерфейсу користувача.

Серед недоліків можна вказати відсутність додаткових функцій серверів NAS та залежність продуктивності пристроїв NAS від пропускної здатності мережі.

Storage Area Network

SAN — це спеціалізована мережева інфраструктура для зберігання даних (мережа зберігання даних). Ці мережі інтегруються у вигляді окремих спеціалізованих підмереж до складу локальної (LAN) або глобальною (WAN) мережі. По суті, SAN-мережі пов'язують один або кілька серверів (SAN-серверів) з одним або декількома пристроями зберігання даних. SAN-мережі дозволяють будь-якому SAN-сервера отримувати доступ до будь-якого пристрою зберігання даних, не завантажуючи при цьому ні інші сервери, ні локальну мережу. Крім того, можливий обмін даними між пристроями зберігання даних без участі серверів. SAN-мережі дозволяють дуже великому числу користувачів зберігати інформацію в одному місці (з швидким централізованим доступом) і спільно використовувати її. Як пристрої зберігання даних можуть застосовуватися RAID-масиви, різні бібліотеки (стрічкові, магнітооптичні і ін.), А також JBOD-системи (масиви дисків, не об'єднані в RAID). Для побудови мереж SAN використовується або стандарт Fibre Channel (FC), або стандарт ISCSI.

Основними перевагами SAN систем є висока продуктивність, додаткові набори функцій, таких як реплікація, дедуплікація, висока доступність, висока маштабованість, відсутність обмежень на кількість дисків, які можна додати.

Також забезпечується безперебійна робота, оскільки системи SAN не потребують перезавантаження для додавання нових дисків, для заміни дисків чи для налаштування RAID-групи.

Недоліком даного типу сховищ є висока собівартість зберігання даних.

Об'єктне сховище

Об'єктне сховище зберігає дані у вигляді об'єктів. Кожен «об'єкт» включає в себе власне дані, деяку кількість метаданих, а також глобально унікальний ідентифікатор. Об'єктне сховище є недорогим і зручним у масштабуванні способом для зберігання великих обсягів неструктурованих даних.

Особливістю об'єктного сховища є те, що кожен файл має ще й унікальний глобальний ідентифікатор крім користувацького. Якщо фізичне розташування об'єкта змінюється, зміни обробляються внутрішнім механізмом об'єктного сховища, а ідентифікатор, призначений користувачем або додатком, залишається незмінним, що дозволяє легко отримати доступ до нього в будь-який інший час.

Об'єктне сховище може зберігати велику кількість даних та доступ до них є більш простим, оскільки воно має декілька незалежних вузлів, які зберігають дані і центральний вузол не має знати місцезнаходження об'єкта для того, щоб отримати його.

Ідентифікатори можна використовувати для того, щоб легко порівняти два файла без необхідності завантаження. Об'єктне сховище також дає можливість використання великої кількості, простого, більш дешевого апаратного забезпечення різних типів, пов'язуючи все разом в єдину систему.

Кожен об'єкт, як правило, має кілька реплік, можливо в географічно розподілених кластерах. Кожен об'єкт також містить контрольну суму, яка дозволяє легко виявити пошкодження даних, в цьому випадку можлива генерація нової копії об'єкта для заміни пошкодженої.

Іншою характерною особливістю зберігання об'єкта є використання метаданих, або даних про дані. Кожен об'єкт при зберіганні маркований не тільки своїм унікальним ідентифікатором, але також інформацією про дані в об'єкті. Наприклад, оцифрована пісня поміщена в об'єктне сховище може зберігатися з додатковими метаданими про назву пісні, виконавця, альбомом, тривалістю, роком і так далі. Важливо відзначити, що метадані кожного об'єкта фіксуються і зберігаються з ним. Немає необхідності тримати їх в окремій централізованій базі даних, яка може некеровано збільшитись. Проте недоліком даного типу сховищ є те, що пошук по метаданим є складним в реалізації.

Критерії вибору сховищ

Перед тим як вибрати, яке рішення для зберігання найкраще працює для ваших потреб, є кілька факторів, які необхідно враховувати.

1. Чи необхідно захищати дані, підвищуючи надлишковість даних?

Найкращим варіантом за таких умов є об'єктне сховище, адже саме воно надає функцію зберігання кількох реплік одного об'єкта в в декількох місцях, що забезпечує максимальний час безперебійної роботи. Однак серед SAN пристроїв також є типи, що можуть виконувати цю функцію.

2. Наскільки важлива продуктивність операцій читання / запису?

При важливості такого критерію найкращим вибором є SAN сховища які забезпечують найвищу продуктивність.

3. Скільки людей повинні отримати доступ до даних і як часто?

При порівнянні сховищ за цим критерієм, найкращий показник забезпечення доступності мають об'єктні і деякі SAN сховища. Щодо DAS та NAS сховищ, то найнижчий рівень забезпечують DAS, а NAS, хоч і забезпечують низький рівень, однак він кращий, ніж у DAS.

4. Скільки коштів і ресурсів планується витратити на сховище?

У разі обмеженості бюджету відповідними варіантами будуть DAS або NAS сховища.

Висновки

В кінцевому рахунку немає одного типу сховищ, яке буде відповідати всім вимогам зберігання даних, і тому бажано визначити, які вимоги є найбільш важливими у тому чи іншому випадку. Це може стати частиною загальної стратегії організації зберігання даних, що має вирішальне значення для забезпечення ефективного виконання бізнес процесів. Правильний вибір типа сховища даних може допомогти підприємствам впоратись із зростаючими об'ємами даних.

В результаті дослідження було розглянуто чотири основні типи сховищ даних. Кожен з них пропонує величезні переваги в певних обставинах. Не можна однозначно визначити «правильне» або «неправильне» рішення — є тільки те, що найкраще підходить до потреб підприємства. Було зазначено основні критерії до вибору найбільш прийнятного рішення та відповідність розглянутих типів сховищ даним критеріям.

Література

1. Gnanasundaram, S., Shrivastava, A., & Services, E. E. (2012). Information storage and management: Storing, managing, and protecting digital information in classic, virtualized, and cloud environments (2nd ed.). Indianapolis, IN: Wiley, John & Sons. — pp. 64–69.

2. Poulton, Nigel. Data Storage Networking. N.p.: Sybex, 2014. — pp. 145–156.

3. Toigo, J.W. The holy Grail of enterprise data storage. / Prentice-Hall, 1999 — pp. 45–47.