

行政院國家科學委員會專題研究計畫 成果報告

Linux-based 網路多磁碟儲存系統

計畫類別：個別型計畫

計畫編號：NSC91-2213-E-034-004-

執行期間：91年08月01日至92年07月31日

執行單位：中國文化大學資訊科學系

計畫主持人：戴文彬

計畫參與人員：石旭本, 張志豪

報告類型：精簡報告

處理方式：本計畫涉及專利或其他智慧財產權，2年後可公開查詢

中 華 民 國 92 年 10 月 1 日

行政院國家科學委員會專題研究計畫成果報告

Linux-based 網路多磁碟儲存系統

計畫編號：NSC 91-2213-E-034-004

執行期限：91 年 8 月 1 日至 92 年 7 月 31 日

主持人：戴文彬 中國文化大學資訊科學系助理教授

一、中文摘要

本計畫主要針對資料使用的儲存特性，以 Linux 系統作為開發環境，研究資料分享、儲存效能、大量儲存、資料安全、網路通訊等資料的儲存特性，提出一個網路磁碟儲存管理系統(NDSMS)的新架構。網路磁碟儲存系統是一連結於網路上的儲存伺服器系統，是利用網路共通協定，將網路上之儲存空間作集合式的管理，以達到高效率、高度分享的一種儲存型態。在這個大量資料的儲存管理系統中，資料正確性、使用效能分析、以及資料備份復原等功能必須包含其中，才能使儲存系統具有實用性。另外，網路磁碟儲存系統必須具備完善的管理系統，以提供多磁碟環境的使用管理以及即時服務機制。最後所實作出的網路磁碟儲存管理系統，將能運用於網路儲存，提供多使用者環境的各式資料儲存格式(如 Ext2、FAT、NTFS 等)。

關鍵詞：網路磁碟儲存、網路附加儲存、儲存管理、Linux

Abstract

In this project, In this paper, we propose the Linux-based Network Disk Storage Management Systems (NDSMS). In this storage architecture, the characteristics on

data sharing, storage efficiency, storage capacity, and data safety are discussed. Different from the other storage systems, the network disk systems are devised to store user data in various file system formats. The functions of NDS management systems focus on the practical usage types of the users. The Linux-based NDSMS been implemented to provide a new efficient storage mechanism for the data storage in multi-user network environment.

Keywords: Network Disk Storage, Network Attached Storage, Storage Management, Linux.

二、緣由與目的

在目前資訊通訊頻繁的數位時代下，由於使用者對於數位資料的使用層面逐漸地普遍廣泛並且精緻分工，使得電腦儲存的資料量急遽膨脹，設備容量的擴充成為必然的趨勢。而另一方面，隨著電腦技術不斷的進步，容量更大、速度更快的儲存設備也隨之不斷地開發。然而，如果僅是增加原有設備的容量，而忽視使用資料的儲存管理、以及儲存維護的話，並不是可以長期依賴的有效作法，必須採用系統管理方式加以整合管理。不過，這其中涉及網路資料使用的儲存特性不同，管理系統的功能設計常需要有特殊化的考量，而且

其中又牽涉到多種不同的檔案系統，及之間的各項關係，是非常複雜而耗時間精力的工作。

本計畫主要針對資料使用的儲存特性，以 Linux 系統作為開發環境[1][2]，研究資料分享、儲存效能、大量儲存、資料安全、網路通訊等資料的儲存特性[3][4][5]，提出一個網路磁碟儲存管理系統(Network Disk Storage Management System, NDSMS)的新架構。網路磁碟儲存管理系統是一連結於網路上的儲存伺服器系統，是利用網路共通協定，將網路上之儲存空間作集合式的管理，以達到高效率、高度分享的一種儲存型態。其中的管理系統，將能運用於網路儲存，提供一個符合多使用者環境的各式資料儲存格式(如 ext2、minix、proc、fat、ntfs、msdos 等)，以及即時服務機制。

目前相關技術以 SAN (Storage Area Network) 與 NAS (Network Attached Storage)技術，可以作為較能接受的解決方案。這些技術所著眼的是藉由儲存設備集中存放，來處理資料儲存與檔案分享的問題。主要目的是要使資訊系統更方便於管理與維護，而使得企業投資在管理資訊系統的人事、時間、軟硬體上的成本減少而卻能得到更好的效能[6][7]。然而，僅僅單純的集中式服務管理技術，如果未能針對資料的儲存特性作有效管理，將只能解決資訊系統的資料儲存、分享與管理上的部分的問題而已。

事實上，有效與合理的資料儲存管理應該以用戶端的磁碟使用(disk usage)為單位，作為管理考量。不過，用戶端的磁碟使用卻是涉及複雜的各種情況與問題，例如：不同的電腦設備、不同時期建構的資訊系統、以及不同世代的應用軟體、不同的檔案系統上等。加上作業系統及應用軟體不因應使用者更廣泛的使用需求，不斷地膨脹，使得每一系統的維修、軟體的更新、維護形成管理的一大負擔。這些問題都有可能使得儲存管理無法有效運用。

本研究主要目的正是著眼於此，針對用戶端的磁碟空間的集中管理服務，藉由引進新的核心架構技術及分享觀念，來解決管理多台電腦的網路使用環境的問題。

在這個資料儲存管理系統中，資料正確性、使用效能分析、以及資料備份復原等功能必須包含在其中，才能使儲存系統具有實用性。使得多台電腦使用因集中化的服務管理，變得更容易維護可用性也更強。

三、結果與討論

3.1 系統設計

用戶端的磁碟使用的管理問題，主要是在於網路連結的多使用者環境中，其儲存的情況存在著各式不同的檔案系統(file systems)，例如：umsdos、msdos、vfat、ntfs、ext、ext2、xia、minix、proc、smb、ncp、iso9660、sysv、hpfs、affs、ufs 等等。經常，這些不同的檔案系統都是同時共存、並且以分散方式一起來使用。只要有一個有效的網路磁碟儲存管理系統(NDSMS)存在，便可以來管理不同的檔案系統，讓它們同時集中共存，使得不同用戶端的磁碟使用有效管理，以提供網路多使用者的使用環境。

網路磁碟儲存管理系統是一連結於網路上的 disks 儲存服務管理系統，它利用網路共通協定，將網路上之儲存空間作集合式的管理，以達到高效率、高度分享的一種儲存型態。因此，這個 NDSMS 新架構具備有以下的優點：

- 讓不同檔案系統磁碟同時集中共存。
- 使儲存資料達到高度共享使用。
- 減少使用者間相同資料的重複儲存。
- 提高大量資料的儲存效能。
- 做到有效的資料備份、復原等管理。
- 管理使用端資料將變得非常容易。
- 做到網路資料流量的使用分析。

在這些優點中，資料分享、減少相同資料的重複儲存、提高資料儲存效能，是讓不同用戶端的磁碟使用同時集中共存的最大理由。而容易維護與管理，更是使用一個新的儲存架構時的一個重要考量。在 NDSMS 設計中，都具有這些特性，因此適合做為網路上大量資料、高度分享、高效率的一種儲存型態。

本研究就針對用戶端的磁碟使用的儲存特性，以 Linux 系統作為開發環境，設計 NDSMS 架構，投入相關技術之研究，最後實作出 NDSMS，作為網路資料儲存的

有效解決方案。

3.2 架構

NDSMS 著眼於用戶端磁碟與檔案系統之間實際使用關係。NDSMS 可以管理多台電腦的硬式磁碟機，並能提供服務多台電腦透過網路連線(TCP/IP protocol)來分享和存取儲存空間。使用者可以透過不同的電腦，從 NDSMS 獲得包含有屬於自己的電腦使用環境的邏輯磁碟區域(logical volumes)。而 NDSMS 支援對使用端磁碟區域的管理操作，例如：製作用戶端磁碟區域的影像檔、對用戶端磁碟區域提供日誌的功能、及快速復原用戶端磁碟區域…等操作。

為了有效管理多台電腦的硬式磁碟機，及提供分享和其他服務。NDSMS 被建構成一主從式架構，包含 NMS (Network Management Server)與 NDS (Network Disk Server)。NMS 負責管理使用者資訊和使用環境資訊，實體儲存裝置已被集中由 NDS 管理，多台沒有配備磁碟的電腦藉由 VAA(Volume Attached Adaptor)提供介面與 NMS 和 NDS 連結，做為管理多台電腦使用環境和存取邏輯磁碟的橋樑。

NMS 在 NDSMS 架構中，扮演管理者的角色。NMS 負責協調 clients 與 NDS 之間的互動，記錄了使用者的登入資訊以及使用者在 NDS 中所擁有的邏輯磁碟的進入資訊和一些管理上所需的資訊。一旦使用者透過 VAA 向 DMS 登入後，使用者即可取得所擁有的邏輯磁碟的進入資訊，進一步地透過邏輯磁碟的進入資訊對位在於 NDS 中的邏輯磁碟進行操作或存取。而系統管理者可以透過 console 來管理 NMS 的運作。

而 NDS 在 NDSMS 架構中，單純地扮演資料管理者的角色。NDS 集中管理儲存裝置，並能提供使用者要求所需的儲存空間以供使用。NDS 把儲存裝置看做是一個 volume 裝置，而每一個 volume 裝置皆被格式化為上層管理系統(SMS, Superior Management System)的資料格式，並且使用 SMS operator 來管理。當使用者向 NDS 要求建立所需要的儲存空間時，SMS operator 會從被集中管理的 volume 裝置中，配置一

個符合使用者要求的大小的邏輯磁碟區域，並把這個邏輯磁碟區域的進入點(entry)的資訊通知 NMS。爾後，使用者就可通過邏輯磁碟區域的 entry 來存取或操作屬於使用者的邏輯磁碟區域。

藉由儲存設備集中存放管理及核心技術與資源分享，NDSMS 能夠輕易地管理多台電腦的使用環境，整合可利用的軟體資源，使得軟體資源更有效率地被使用。管理及維護多台電腦不再是系統維修人員的惡夢，大量降低企業在軟硬體上的投資，增進整個資訊系統的效能與管理效率。

3.2 核心機制

在此我們設計 NDSMS 的核心機制：分享式讀寫 SRW (Sharing Read/Write) 用以提供支援邏輯磁碟區域的操作，使得將儲存裝置集中所獲得到的利益不僅僅只是分享磁碟區域和管理方便而已。此以核心機制改變以往對磁碟區域的使用方式，使得儲存於磁碟區域內的資料能夠比以前更有效率的分享及利用。

NDSMS 的 SRW 機制其實就是把儲存空間看成是記憶體一般，機制可以使不同的邏輯磁碟區域分享共同的資料，而當邏輯磁碟區域被寫入資料時，SRW 機制才為寫入的資料配置新的空間。這和虛擬記憶體的管理機制有異曲同工之處，唯一的不同是因為儲存空間是非揮發性的，所以 NDSMS 的 SRW 機制是能夠保存住各邏輯磁碟區域之間的差異性及共通性。

由於 NDSMS 的 SRW 機制能夠保存住各邏輯磁碟區域之間的差異性及共通性，所以運用 SRW 機制來對邏輯磁碟區域進行某種操作就變得可能。以下分別是目前利用 SRW 機制在邏輯磁碟區域上最主要的幾項操作：

- 複製操作：目的是將一邏輯磁碟區域當時的資料狀況保存下來。一邏輯磁碟區域經複製後，仍能正常讀取到以前的資料，後來寫入的資料則會儲存於別處，而不會影響到原來資料的完整性。

- 回復操作：回復操作的目的是將一指定的邏輯磁碟區域的資料狀況恢復到上一次複製操作後的狀況。一邏輯磁碟區域經回復操作後會釋放自上一次複製操作後

到現在為止所寫入的資料。

- 清除操作：清除操作的目的是將一指定的邏輯磁碟區域的數個指定的複製操作的版本，清除各個版本間資料狀況的差異性，保留最後的共通性而成為一新的複製操作的版本。

- 邏輯磁碟區域分享：一邏輯磁碟區域經複製操作後，其資料狀況就獲得了完整的保留，因而分享給其他的邏輯磁碟區域。這樣的分享資料方式，能夠使得其他的邏輯磁碟區域快速地擁有數量可觀的資料，而不必個別的建置。

- 日誌功能：一邏輯磁碟區域可以使用複製操作、回復操作來為資料作簡單的日誌，複製操作可以保存特定的時間點的資料狀況，而回復操作可以回復資料狀況至某一特定的時間點。

- 快速的建置與恢復邏輯磁碟區域：有了複製操作、回復操作、清除操作這些基本的操作，再加上邏輯磁碟區域的影像的分享和簡單的日誌功能。使得建置一新的邏輯磁碟區域的資料內容變得相當容易且快速，而如果資料因人為疏失而流失時，亦可於相當短的時間內恢復資料，而不必經過複雜且耗時的從備份資料中把資料拷貝回來的過程。而當所使用的空間超過系統限制時，還可以利用清除操作來清除一些老舊而沒有用的資料，使得儲存空間的利用更有彈性。

利用 SRW 處理機制，使得資料的分享、儲存、備份變得更有效率，把儲存空間更徹底的使用。如此使得建置或管理電腦使用環境變得容易而且有效率，更不必只是單純的增加儲存容量來解決問題。

四、計畫成果自評

NDSMS 就是透過磁碟區域的分享、上層管理系統 SMS、及分享式讀寫 SRW 機制來達到簡易管理多台電腦使用環境的目的。不論軟體或硬體都獲得了更徹底的應用，而且透過集中式的管理與服務，使得資料分享更快速可使用的資源也更多元。藉由 SAN、NAS 及 NDSMS 之研究，未來會有更多的應用來解決日益膨脹的資訊問題，資料分享問題將得以完全的解決。

本計劃完成之工作項目包括：

- 提出一個網路儲存管理系統的新架構。
- 完成高效率、高度分享的一種儲存機制：
 1. 不同磁碟同時集中共存。
 2. 使儲存資料達到高度共享使用。
 3. 減少使用者間相同資料重複儲存。
 4. 提高大量資料的儲存效能。
 5. 做到有效的資料備份、復原等管理。
 6. 管理使用端資料將變得非常容易。
 7. 做到網路資料流量的使用分析。
- 實際在 Linux 環境下建置完成系統。

本計劃之成果對學術研究之貢獻如下：提出一個網路儲存管理系統的新架構解決網路資料儲存問題，並建置實作出相關系統，結果顯示其效能得到驗證。計劃的結果具有實際價值，可提供相關學術研究參考。

有關後續研究方面，這個新架構能擴展至其他的問題的解決，有進一步探討的必要。特別是透過 WAN 方式的資料傳輸、同步、維護、備援等問題，將作為本計劃日後的延續研究。

總體而言，本計劃完成一定的研究成果，相關成果已在研討會或期刊上發表 [8]，並完成相關的系統實作。

五、參考文獻

- [1] Card, R., Dumas, E., and Mevel, F., The Linux Kernel Book, John Wiley & Sons, 1997.
- [2] Beck, M., et al., The Linux Kernel Intervals, Addison-Wesley, 1996.
- [3] Kleiman, S., "Vnodes: An Architecture for Multiple File Types in Sun Unix", Proc. Summer USENIX Conf., 1986, pp.260-69.
- [4] Knowlton, K. C., "Fast Storage Allocator", Communications of the ACM, 8(10), 1965, pp.623-625.
- [5] McKusick, M., et al., "A Fast File System for Unix", ACM Transactions on Computer Systems, 2(3), pp.181-197, 1984.
- [6] Silberschatz, A. and Calvin, P., Operating System Concepts, 4th Ed., Addison-Wesley, 1994.
- [7] Tanenbaum, A., Operating Systems: Design and Implementation, Prentice Hall, 1987.
- [8] 戴文彬,石旭本,張志豪,"Linux-based 網路磁碟儲存管理系統",電子化企業經營管理理論暨實務研討會,E33,彰化,2003.

