

郵件歸檔技術大解析

劉立國

前言

隨著電腦對企業組織運作的日漸緊密與資料量的快速增長，資料歸檔作業（Data Archiving）已漸成爲 IT 領域一個重要的議題，但隨之而來的是一堆針對各類型資料作歸檔的商用歸檔軟體，像雨後春筍般地不斷產生。而如何才能在眾多的歸檔軟體中選擇一套最適切於自身 IT 環境需求的軟體也深深困擾著 IT 管理者。有鑑於此，特就目前業界中較常見之歸檔軟體特性、運作模式與優劣得失作一比較，藉以提供 IT 管理者選購此類產品時一項參考。

一、當談到資料歸檔作業，首先我們必須了解資料歸檔的目的爲何？由實際經驗歸納後得出下列數項常見的需求：

1.符合法規:

爲符合政府法令規定，IT 人員對資料必須提供長達數年的保存。常見的資料保存法規如美國沙賓法案、日本個人情報保護法、歐洲 BSI7799 法規等…

2.落實資料生命週期管理以降低成本:

依據各企業對資料存取頻率、重要性、資料大小與檔案建立/修改時間等條件來設定其歸檔策略，將使用率與重要性低的資料由一線儲存裝置移至二線儲存裝置（較便宜之 Disk Array 或磁帶）以降低儲存裝置成本。

3.法務稽核/興訟所需：

達成公司法務部門規定之資料保存政策，以及配合法院訴訟時，提供特定之郵件與相關資料檔案等數位證據。一般 IT 人員會被要求在短時間內還原某一段日期內，設定”特定查詢條件”下的所有檔案與郵件資料以供稽核。

4.還原誤刪資料：

針對一般使用者誤刪重要資料時，IT 部門提供一道較傳統備份還原機制更爲細緻與便於查詢的保護機制。

二、一般而言，資料歸檔技術會依據各企業與組織對各種不同類型的資料，執行歸檔作業的比例與該資料之商業價值雙重考量，進而衍生出相對應的商用歸檔軟體。目前業界的主流爲”電子郵件”以及”文件”的歸檔軟體，本篇文章主要會針對郵件歸檔軟體作一詳細說明與分析。以下便先就郵件歸檔時效性與歸檔儲存模式作一分析：

1. 三種主要郵件歸檔時效性：

1.1 手動歸檔(Manual Archiving)：

依據企業與組織之郵件歸檔政策制定線上郵件保存期限與一線儲存裝置高低水位，當線上郵件資料超過了上述的保存期限或超過了儲存裝置高水位時，IT 人員便啟動歸檔軟體手動執行郵件歸檔作業。此方式優點有二：

- a. 可利用非常廉價之歸檔軟體達成郵件歸檔之目的。
- b. 以人力判斷歸檔與否，故具有相當彈性。

相對的，此歸檔模式缺點主要有二：

- a. 無自動化排程作業，長期需要人力加以關注此作業，當遇長假或 IT 管理員疏忽時恐會造成一線儲存裝置空間使用殆盡。
- b. 缺乏完善的分類、監控與查詢機制，僅可以人為制定之規則粗略加以管理。

1.2. 排程歸檔(Schedule Archiving)：

在郵件歸檔軟體上設定周期性排程時間並依據企業與組織之郵件歸檔政策制定出歸檔工作，其中包含了設定郵件存在的時間多長、附檔資料大小超過多少、附檔資料的類型為何、儲存裝置剩餘空間等…歸檔條件。當排程時間到點時，歸檔軟體會自動將滿足歸檔條件的郵件資料進行歸檔作業。此作業方式優點有二：

- a. 自動化排程作業，IT 管理員僅需設定好歸檔條件後即可依排程進行歸檔作業，同時具備完整之監控、查詢與還原機制，大幅節省人力支出並減少錯誤產生。
- b. 可依 IT 環境需要設定適切之歸檔條件，避免因不當的歸檔規則導致 IT 人員需經常性進行還原作業。

而其缺點主要亦有二：

a. 郵件線上運作與歸檔作業緊密結合，往往會因歸檔伺服器的運作而使得線上 MS-Exchange Server 之效能大幅降低。

b. 依排程執行歸檔作業，會造成線上伺服器瞬間資源使用量飆高之現象。且當災難發生時，郵件資料僅可回復至上次歸檔時間點。

1.3 即時歸檔(Live Content Archiving)：

歸檔伺服器會先透過網路將郵件伺服器中所有的資料完整複製一份回來(Shadowing)，然後持續不斷地將郵件伺服器運作中新產生的資料複製一份回來並寫回先前完整複製的資料中(Log Replay)，以達成即時歸檔儲存之目的。此作業方式主要的優點有三：

- a. 持續不斷地進行歸檔作業，當郵件伺服器發生故障時，損失的資料相較於其它歸檔作業模式為最少。

- b. 所有新增的資料都會進行歸檔作業，不需考量歸檔條件亦不需任何人力負擔，卻能提供最完整的資料回復功能。
- c. 分散定期歸檔作業一次資料量過大的問題，以避免郵件及歸檔伺服器效能負擔爆增的狀況。

此歸檔方式較無明顯的缺點，僅在二線儲存空間的需求量上較大(因所有資料均作歸檔作業)，故在選擇此類型歸檔軟體時可留意是否具備重覆資料刪除功能(De-Duplication)以降低二線儲存空間使用量。

2. 二種主要的郵件歸檔儲存模式:

2.1.複製模式(Copy)：

歸檔伺服器以複製方式對郵件伺服器進行歸檔作業，郵件伺服器中有的資料，在歸檔伺服器上也會有相對應的一份。歸檔伺服器的存在與否不會影響郵件伺服器原本的運作，但相對所需的二線儲存空間量也必定較大。

2.2.歸檔模式(Archive)

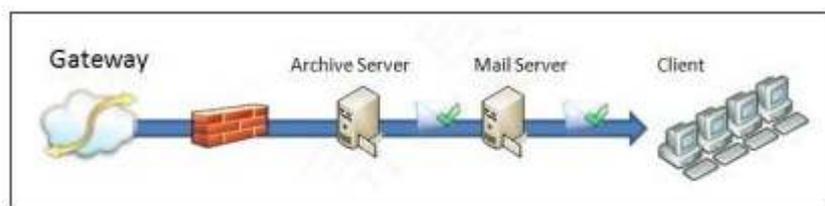
歸檔伺服器以搬移方式對郵件伺服器進行歸檔作業，原先在郵件伺服器中的資料經歸檔後僅存一個連結圖示(Stub)或完全消失，可有效節省一線儲存空間。當前端使用者使用已歸檔的資料時，會由歸檔伺服器先將資料還原至郵件伺服器後，使用者方可使用。因此歸檔伺服器發生故障時郵件伺服器同時也會運作失敗。

三、Mail Archiving

常見的郵件系統包括了 MS-Exchange、Lotus Notes 以及 Unix Base 的 Sendmail、Postfix 等等…依照其市佔率與商業市場價值考量筆者僅針對在 MS-Exchange 上，常見的歸檔技術作一分析與比較：

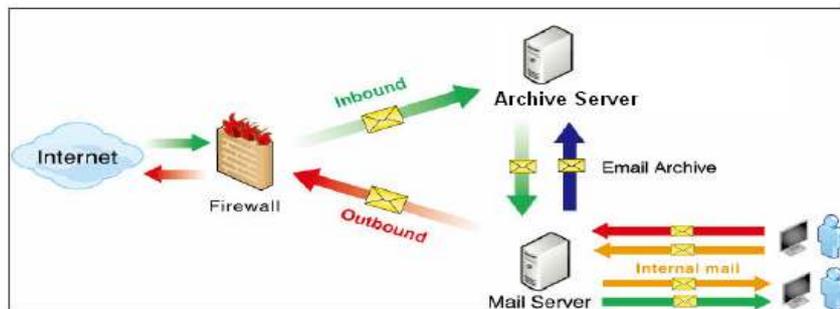
1.業界常見 MS-Exchange 歸檔架構及軟體：

a. Gateway Mode:



在 MS-Exchange 伺服器對外連結端建置一台歸檔伺服器，所有 Inbound 及 Outbound 的電子郵件在經過此歸檔伺服器時會被歸檔儲存下來。在此架構下歸檔作業將不會影響到 MS-Exchange 伺服器的效能，但其缺點在於公司內部人員

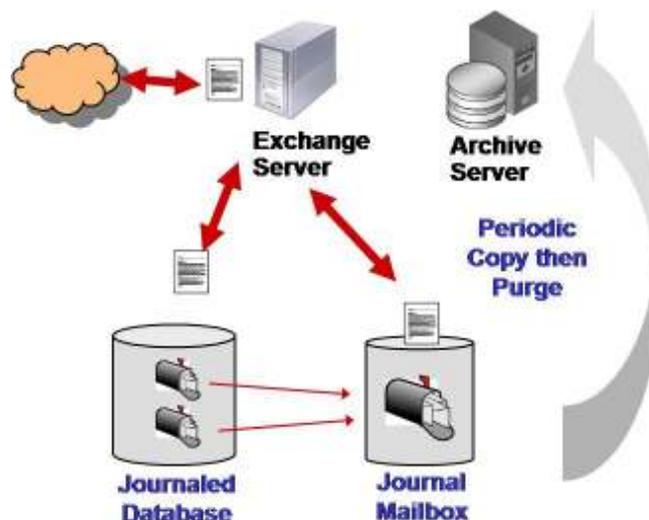
彼此收發的信件，因不會經過歸檔伺服器故無法進行歸檔保存。此外當歸檔伺服器發生故障時，可採 By Pass 模式使信件直接進入郵件伺服器，但在此時進出的所有信件將無法被歸檔保存。



此為 Gateway 模式的另一種型態，不同於上述典型的 Gateway 模式，所有 Outbound 的信件不論發送對象是內部或外部人員都會直接發出，在此同時 MS-Exchange 伺服器會將 Outbound 的信件再轉寄一份至歸檔伺服器進行歸檔作業。此架構雖然解決了原先 Gateway 模式內部對內部郵件無法歸檔的問題，但在 MS-Exchange 伺服器上必須安裝歸檔軟體之代理程式負責信件的歸檔轉寄作業，這會對 MS-Exchange 伺服器造成很大的效能負擔。當歸檔伺服器無法正常運作時，龐大的歸檔轉寄郵件會累積至 MS-Exchange 伺服器當中進而造成空間與效能不足的現象。

目前市面上用此二種 Gateway 模式的歸檔軟體有：中華數位 MAE (Gateway mode)、Cellopoint、Nopam、桓基 Mail Sherlock 等...

b. Journal Mode :



此類型架構首先需將 Exchange Server 設定成 Journal Mode，所有進出往來的信件會轉寄一份至單一的 Journal Mailbox 當中，同時會將 Archive 作業的資訊寫入 Journal Database 內。接下來需在 Exchange Server 上安裝歸檔

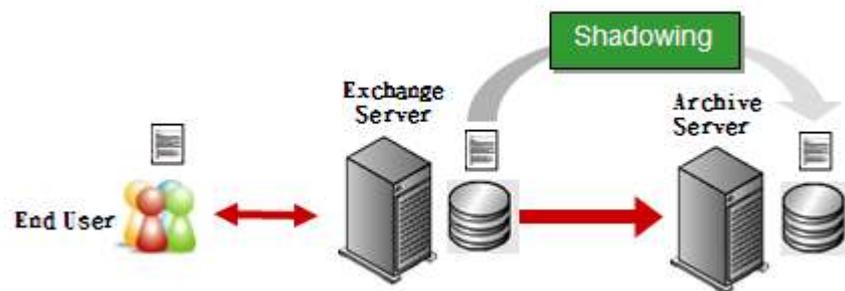
軟體郵件主機端代理程式，負責接受歸檔伺服器的排程指令將 Journal mailbox 中所有郵件資料定期地移至歸檔儲存裝置後端儲存。

由於啟動 Journal Mode 會使用到 MS-Exchange Server 上非常多的資源，同時所有歸檔的信件會集中於一個 Journal mailbox 當中，當進行歸檔作業時，該 mailbox 所在的伺服器與磁碟陣列，資源使用率會飆至極高，嚴重影響導致 MS-Exchange 原先正常運作。另外此歸檔作業是以信件為單元進行歸檔，故無法完整保存密件附本郵件、行事曆、通訊錄、郵件樹狀結構、刪除的信箱與共享資料匣等資訊...

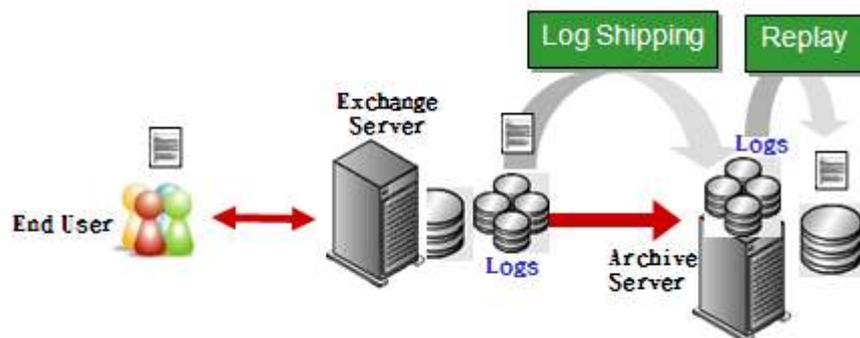
目前市面採此架構的歸檔軟體有：中華數位 MAE(Standalone Mode)、Symantec Enterprise Vault、Openfind MailBase、IBM CommonStore for Exchange、EMC EmailXtender 等...

c. Log Shipping Mode :

Step1



Step2



此架構在第一次歸檔作業時，會將 MS-Exchange Server 上所有的資料複製(Shadowing)一份至歸檔伺服器中。接下來當 User 收送郵件至 MS-Exchange Server 時或增、刪、修改行事曆、通訊錄、公用資料匣等...都會產生一份 Log，而歸檔伺服器偵測到有新的 Log 產生後便會自 MS-Exchange Server 上拉回(Log Shipping)並寫入歸檔磁碟中(Log Replay)。由於是持續分散地進行複製作業，故瞬間耗損 MS-Exchange Server 的效能甚低。

此架構作法與 MS-Exchange 2007 中的 CCR(Cluster continuous replication)運作模式相同，並不需要安裝任何代理程式，同時它會記錄 MS-Exchange 上所有的變動，故可儲存完整郵件樹狀結構、行事曆、通訊錄、密件附本郵件，此外連刪除的信箱與共享資料匣與等資訊亦可保留。這種架構對 Mail 的歸檔保存可說是非常完善，但在二線歸檔磁碟的使用量上會較其它的架構來的多，所以適切的調整歸檔策略與搭配 De-dup 功能來進行歸檔作業會是此架構較理想的作業方式。

目前市面採此架構的歸檔軟體僅有 Mimosa Nearpoint

2.MS-Exchange 各類型歸檔軟體之功能比較如下：

	GateWay Mode	Journal Mode	Log Shipping Mode
是否不消耗 MS-Exchange 伺服器效能		X	
MS-Exchange 是否不需安裝 Agent		X	
是否可保存所有內外部進出郵件			
是否能保存郵件架構、通訊錄、行事曆等資訊...	X	X	
Archiving Server 損毀時是否不會影響正常郵件收送	X	X	

 :是  : 部份是 X : 否

3. 附屬功能考量：

除上述功能外，下列幾項功能也是選擇 MS-Exchange 歸檔軟體時，重要的參考指標：

a. 歸檔郵件查詢介面完善與否：

當郵件回復需求發生時必定為使用者所關注之重要事件，而歸檔郵件查詢介面的好壞往往決定著郵件回復的快慢。要如何從龐大郵件資料中找尋自己所需的信件，首先擁有多種查詢方式為基本的要求，如關鍵字檢索、全文檢索、甚至與 Outlook 樹狀結構相同的瀏覽檢索均為目前常見的檢索方式。其次考量就是還原作業是否可由使用者自行作業或是需由管理者進行，這會與回復時效與管理者的人力負擔有關。

b. 未來擴充性簡易與否：

隨著郵件資料量的增長，MS-Exchange Server 的擴充需求與日俱增，當然歸檔伺服器擴充的簡易與否也必須納入購買郵件歸檔軟體時的一項考量。是否可不中斷 MS-Exchange 與郵件歸檔服務即可完成擴充作業，以及擴充時軟體設定之繁瑣程度為此一項目之觀察重點。

c. 具備 e-Discovery 功能與否：

e-Discovery 在國外 IT 業界為一項熱門的郵件歸檔功能，其功能內容為追蹤信件自發出後所有的寄轉送歷程，包含轉寄、密件副本等所有收發人員均會被記錄與有效搜尋。此項作業主要用於公司內部各單位文件傳送流程查詢與公司間的商業糾紛與興訟，擁有此項功能可使法務稽核人員快速的搜尋所需的資料。

d. 是否具備有效的災難復原機制：

當郵件伺服器發生損毀時，伺服器的重建與還原需仰賴備份軟體或叢集架構加以完成。選擇具備災難復原機制的郵件歸檔軟體可簡化作業手續同時減少郵件資料的損失。

e. 具備重覆資料刪除技術與否：

當郵件歸檔軟體具備重覆資料刪除技術時，可將各信件中相同的附檔資料、公用資料匣檔案等…加以剔除，進而大幅降低郵件歸檔儲存成本。

總結：

根據 Garner Report 對 E-Mail Archiving 市場的統計，由 2005 年全球市場 U.S 197Million 到 2007 年的 U.S 376Million，更預估至 2012 年市場總值將達 U.S 1.7Billion。這個數字說明了資料歸檔已成為未來 IT 環境中必備的一環，因此如

何能夠選擇一項最合適的郵件歸檔軟體將考驗著 IT 決策人員的智慧。希望由業界實務中所得到的經驗心得能提供一個選擇解決方案的思考方向，使 Mail Archiving 成爲一項輕鬆事。

(作者現任職於麟瑞科技)