

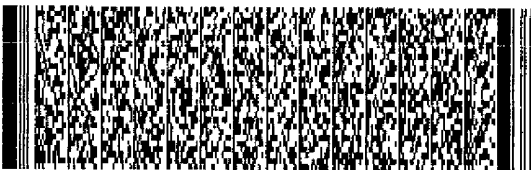
# 公告本

申請日期： 88. 3. 8	案號： 88103519
類別： G06F 11/14	

(以上各欄由本局填註)

## 發明專利說明書 452636

一、 發明名稱	中文	高度可用之檔案伺服器
	英文	HIGHLY AVAILABLE FILE SERVERS
二、 發明人	姓名 (中文)	1. 史帝芬 羅勃特 克萊門
	姓名 (英文)	1. STEVEN ROBERT KLEIMAN
	國籍	1. 美國
	住、居所	1. 美國加州洛杉磯市艾爾曼地路157號
三、 申請人	姓名 (名稱) (中文)	1. 美商奈特渥克家電用品公司
	姓名 (名稱) (英文)	1. NETWORK APPLIANCE, INC.
	國籍	1. 美國
	住、居所 (事務所)	1. 美國加州聖塔克萊拉市聖湯瑪士路2770號
	代表人 姓名 (中文)	1. 丹尼爾 J. 溫門侯芬
	代表人 姓名 (英文)	1. DANIEL J. WARMENHOVEN



本案已向

國(地區)申請專利

申請日期

案號

主張優先權

美國 US

1998/03/10 09/037,652

有

有關微生物已寄存於

寄存日期

寄存號碼

無



## 五、發明說明 (1)

發明背景

## 1. 發明領域

本發明係關於儲存系統。

## 2. 相關技術

電腦儲存系統是用以記錄及擷取資料。在一些電腦系統中，儲存系統與一組從屬裝置進行通訊，且提供服務以記錄及擷取資料至該等從屬裝置。因為對於許多應用而言資料儲存非常重要，所以儲存系統所提供之服務及資料最好能儘量可供用以服務。因此最好提供一種即使儲存系統之元件發生故障仍能夠維持可供用以服務之儲存系統。

一種用以提供維持可用以服務之儲存系統之已知技術是提供多個冗餘之儲存組件，且該等儲存組件具有下列特性：當第一儲存組件故障時，第二儲存組件可用以提供原來由第一儲存組件所提供之服務及資料。轉移提供服務之功能自第一儲存組件至第二儲存組件稱為故障服務轉移 (failover)。第二儲存組件保存第一儲存組件所保存之資料之拷貝，以致故障服務轉移可繼續進行而無重大之中斷。

用以達成故障服務轉移之第一已知技術是導致第二儲存組件拷貝第一儲存組件之所有運作。因此，第二儲存組件也完成第一儲存組件所完成之每一儲存運作。此種第一已知技術具有下列缺點：(1) 其使用第二儲存組件之極多處理容量以複製第一儲存組件之努力，而該等複製努力之大部份皆是浪費。(2) 其使得第一儲存組件在確認運作完成



## 五、發明說明 (2)

上變慢，因為第一儲存組件必須等候至第二儲存組件也完成相同之運作。

用以達成故障服務轉移之第二已知技術是辨識一系列之檢查點，而在該等檢查點第一儲存組件是處於一致及已知之狀態。一旦發生故障服務轉移，則第二儲存組件可自最近之檢查點開始繼續運作。例如，NFS(網路檔案系統)通訊協定要求所有寫入運作在他們獲得確認之前必須儲存於磁碟，以致寫入運作之確認顯示穩定之檔案系統組態。此第二已知技術具有下列缺點：(1)其使得第一儲存組件在執行寫入運作上變慢，因為第一儲存組件必須等候至寫入運作完全儲存於磁碟。(2)其使得故障服務轉移之回復變慢，因為第二儲存組件解決第一儲存組件之故障在所辨識之檢查點間所留下之任何不一致。

因此，最好提供一種儲存系統，及一種用以操作儲存系統之方法，且該種方法可高效率使用所有儲存系統組件，快速完成及確認運作，且自任何儲存組件之故障轉移快速回復。本發明之一實例達成此種優點，其中儲存系統建構頻繁且快速之檢查點，且其中儲存系統在檢查點之間快速分佈該等運作之複製指令於其之儲存組件。

發明摘要

本發明提供一種儲存系統，與一種用以操作該儲存系統之方法，且即使該儲存系統之元件發生故障該儲存系統仍維持高度可用。第一及第二檔案伺服器皆包含一檔案伺服器請求記錄器以儲存輸入之檔案伺服器請求。第一及第二



## 五、發明說明 (4)

檔案伺服器皆可存取一組共用之大量儲存組件。第一輸入之檔案伺服器請求皆拷貝至第一及第二檔案伺服器；第一檔案伺服器處理檔案伺服器請求，而第二檔案伺服器保存一拷貝於其之檔案伺服器請求記錄器。每一檔案伺服器藉由使用一檔案系統來運作，且該檔案系統在每一檔案伺服器請求以後保存一致狀態。一旦出現故障服務轉移，則第二檔案伺服器可自最近之一致狀態開始執行位於其之檔案伺服器請求記錄器之該等檔案伺服器請求。

在本發明之第二方面，一檔案伺服器系統提供一或更多之大量儲存組件之鏡射。每一輸入之檔案伺服器請求皆拷貝至第一檔案伺服器及第二檔案伺服器。第一檔案伺服器執行檔案伺服器請求以修改一組之主要大量儲存組件，且也執行該等相同之檔案伺服器請求以修改一組之鏡射大量儲存組件。該等鏡射大量儲存組件是配置成為實體上與主要大量儲存組件實體分離，例如位於另一位置，且萬一整組之主要大量儲存組件需要獲得回復則提供一資源。

附圖簡短說明

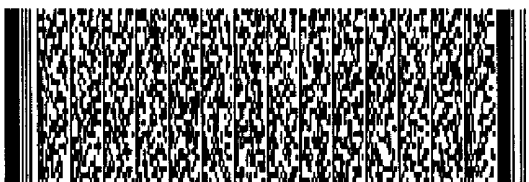
圖1展示一高度可用之檔案伺服器系統之方塊圖。

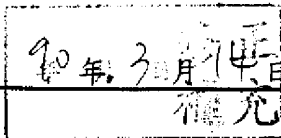
圖2展示該檔案伺服器系統之一檔案伺服器之方塊圖。

圖3展示該檔案伺服器系統之運作之處理流程圖。

元件符號說明

100	檔案伺服器系統	110	檔案伺服器
111	處理器	112	程式及資料記憶體
114	網路介面組件	115	磁碟介面組件





## 五、發明說明 (5)

117	匯流排	120	大量儲存裝置
121	檔案系統	130	中間匯流排
140	網路介面	150	網路
151	檔案伺服器請求	160	伺服器請求記憶體
171	記憶體匯流排	173	輸入／輸出控制器

較佳實例詳細說明

在下列說明中，本發明之一較佳實例之較佳處理步驟及資料結構獲得說明。但是，熟悉本技術領域者應可認知，在熟讀本專利申請之後，本發明之實例可使用在程式控制



## 五、發明說明 (4)

之下運作之一或更多通用處理器(或針對特定處理步驟及資料結構來調整之專用處理器)來加以建構，且使用此種設備之本文所述之較佳處理步驟及資料結構的建構無需不適當之實驗或進一步之發明。

檔案伺服器對與故障服務轉移運作

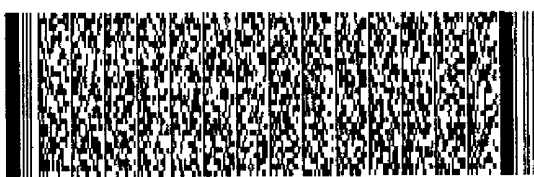
圖1展示一高度可用之檔案伺服器系統之方塊圖。

檔案伺服器系統100包含一對檔案伺服器110，且該對檔案伺服器皆耦接至一組共用之大量儲存裝置120。該二檔案伺服器110之第一檔案伺服器耦接至第一輸入/輸出匯流排130，以控制選定之大量儲存裝置120之第一子集。同樣地，該二檔案伺服器110之第二檔案伺服器耦接至第二輸入/輸出匯流排130，以控制選定之大量儲存裝置120之第二子集。

雖然該二檔案伺服器110皆耦接至所有之共用儲存裝置120，在任何指定之時間只有一檔案伺服器110運作以控制任一大量儲存裝置120。因此，縱使大量儲存裝置120在任一時間皆只可由一檔案伺服器110來控制，即使其之二相關檔案伺服器110之一發生故障，該等大量儲存裝置120之任一裝置仍維持可用。

在一較佳實例中，檔案伺服器系統100包含一對此種檔案伺服器110；但是，在其他實例中，多於二此種檔案伺服器110可包含於單一檔案伺服器系統100。

在一較佳實例中，第一輸入/輸出匯流排130及第二輸入/輸出匯流排130皆包含一中間匯流排，例如PCI匯流排架



## 五、發明說明 (5)

構。

在一較佳實例中，大量儲存裝置120包含磁碟機，光碟機，或磁光碟機。但是，在其他實例中，可使用其他儲存系統，例如磁泡記憶體，快閃記憶體，或使用其他儲存技術之系統。大量儲存裝置120之元件稱為“磁碟”(disk)，即使該等元件可包含其他型態或形狀。

每一大量儲存裝置120皆可包含單一磁碟或多個磁碟。在一較佳實例中，每一大量儲存裝置120包含多個磁碟且是配置及操作成為一RAID儲存系統。

在一較佳實例中，第一檔案伺服器110使用一共用連結來耦接至第二檔案伺服器110。該共用連結提供遠端記憶體存取能力給每一檔案伺服器110，以致資料可自遠端位置儲存至每一檔案伺服器110。在一較佳實例中，共用連結包含一Tandem "ServerNet" 連結。該共用連結使用一裝置控制器來耦接至每一檔案伺服器110，且該裝置控制器耦接至每一檔案伺服器110之輸入/輸出匯流排。

第一檔案伺服器110耦接至第一網路介面140，而第一網路介面140是配置成為接收來自網路150之檔案伺服器請求151。同樣地，第二檔案伺服器110耦接至第二網路介面140，而第二網路介面140也是配置成為接收來自網路150之檔案伺服器請求151。

第一檔案伺服器110包含第一伺服器請求記憶體160，而第一伺服器請求記憶體160接收檔案伺服器請求151並記錄該等請求。萬一第一檔案伺服器110自電源故障或其他服



## 五、發明說明 (6)

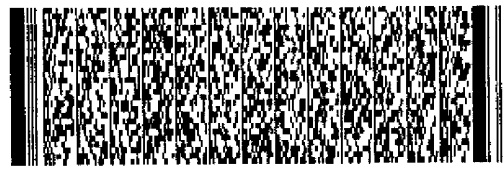
務中斷回復，則第一伺服器請求記憶體160之未決檔案伺服器請求151受到重新執行，以使他們進入第一檔案伺服器110所保存之檔案系統之下一一致狀態。

同樣地，第二檔案伺服器110包含第二伺服器請求記憶體160，而第二伺服器請求記憶體160接收檔案伺服器請求151並記錄該等請求。萬一第二檔案伺服器110自電源故障或其他服務中斷回復，則第二伺服器請求記憶體160之未決檔案伺服器請求151受到重新執行，以使他們進入第二檔案伺服器110所保存之檔案系統之下一一致狀態。

當第一檔案伺服器110收到來自網路150之檔案伺服器請求151時，該檔案伺服器請求151拷貝至第一伺服器請求記憶體160。檔案伺服器請求151也藉由使用遠端記憶體存取經由共用連結來拷貝至第二伺服器記憶體160。同樣地，當第二檔案伺服器110收到來自網路150之檔案伺服器請求151時，該檔案伺服器請求151拷貝至第二伺服器請求記憶體160。檔案伺服器請求151也藉由使用遠端記憶體存取經由共用連結來拷貝至第一伺服器記憶體160。使用遠端記憶體存取相當快速且其之通訊架空少於使用網路通訊協定。

萬一檔案伺服器110之任一發生故障，則另一檔案伺服器110可藉由使用儲存於其本身之伺服器請求記憶體160之檔案伺服器請求151來繼續運作。

在一較佳實例中，每一伺服器請求記憶體160皆包含一非揮發性記憶體，所以儲存於伺服器請求記憶體160之任



## 五、發明說明 (7)

一之該等檔案伺服器請求不會由於電源故障或其他服務中斷而喪失。

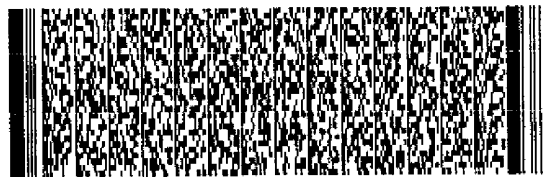
回應檔案伺服器110處理檔案伺服器請求151，且可能修改儲存於大量儲存裝置120之一之檔案。非回應檔案伺服器110保存儲存於其之伺服器請求記憶體160之檔案伺服器請求151，以為回應檔案伺服器110可能發生故障做準備，而非回應檔案伺服器110是回應檔案伺服器110之夥伴。萬一回應檔案伺服器110發生故障，則非回應檔案伺服器110處理檔案伺服器請求151以做為故障服務轉移技術之一部份。

在一較佳實例中，每一檔案伺服器110控制其之相關大量儲存裝置120，以利用下列專利申請所述之發明來構成冗餘陣列，例如RAID儲存系統：

- 專利申請號碼08/471,218，於1995年6月5日歸檔，發明者為David Hitz等人，名稱為"A Method for Providing Parity in a Raid Sub-System Using Non-Volatile Memory"，流水號NET-004；

- 專利申請號碼08/454,921，於1995年5月31日歸檔，發明者為David Hitz等人，名稱為"Write Anywhere File-System Layout"，流水號NET-005；

- 專利申請號碼08/464,591，於1995年5月31日歸檔，發明者為David Hitz等人，名稱為"Method for Allocating Files in a File System Integrated with a Raid Disk Sub-System"，流水號NET-006；

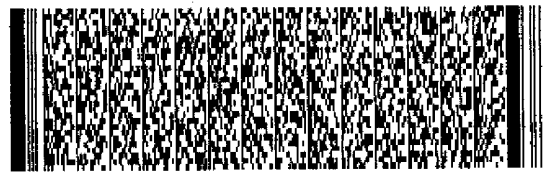


## 五、發明說明 (8)

在此提及每一該等專利申請以供參考。該等專利申請統稱為“WAFL發明。”

做為WAFL發明所展示之技術之一部份，每一檔案伺服器110控制其之相關大量儲存裝置120，以藉由一原子方式來回應檔案伺服器請求151。任何檔案伺服器請求151之最終動作是融入最近之一致狀態於檔案系統121。因此，檔案系統121在完成每一檔案伺服器請求151之後是處於內部一致狀態。因此，定義於大量儲存裝置120之檔案系統121將以內部一致狀態出現，無論那一檔案伺服器110控制該等大量儲存裝置120。內部一致狀態之例外只包含一些最近檔案伺服器請求151，且該等請求將儲存於該二檔案伺服器110之伺服器請求記憶體160。該等最近檔案伺服器請求151可藉由針對一最近一致狀態來執行該等請求以融入該一致狀態。

對於任何檔案伺服器請求151而言，萬一通常回應該檔案伺服器請求151之檔案伺服器110發生故障，則另一檔案伺服器110將認知該故障，且執行一故障服務轉移方法以控制先前指配給故障之檔案伺服器110之大量儲存裝置120。故障服務轉移檔案伺服器110將發現該等大量儲存裝置120及他們之檔案系統121處於內部一致狀態，但是一些最近檔案伺服器請求151仍未獲執行。故障服務轉移檔案伺服器110將包含該等最近檔案伺服器請求151之拷貝於其之伺服器請求記憶體160，且將執行該等檔案伺服器請求151以回應該等拷貝。



## 五、發明說明 (9)

## 檔案伺服器節點

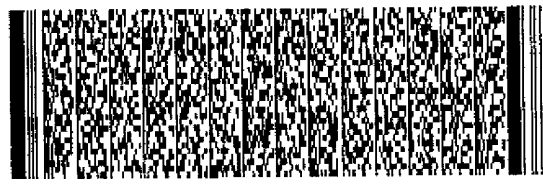
圖2展示該檔案伺服器系統之一檔案伺服器之方塊圖。

每一檔案伺服器110皆包含至少一處理器111，一程式及資料記憶體112，伺服器請求記憶體160（包含一非揮發性RAM），一網路介面組件114，與一磁碟介面組件115。該等組件是藉由匯流排117或其他已知系統架構來互連以進行處理器，記憶體，及週邊設備間之通訊。

在一較佳實例中，網路介面組件114包含一已知網路介面來與網路150一起運作。例如，網路介面組件114可包含一介面以與FDDI標準介面或100BaseT標準介面來一起運作。

在故障服務轉移之後，檔案伺服器110回應針對其本身或其之（故障）夥伴檔案伺服器110之檔案伺服器請求。因此一旦發生故障服務轉移，則每一檔案伺服器110皆能夠具有一額外之網路身份，一身份是其本身而另一身份是其之故障夥伴檔案伺服器110。在一較佳實例中，每一檔案伺服器110之網路介面組件114包含一網路轉接器，且該網路轉接器能夠依照檔案伺服器110之指令來回應二分別之位址。在另一實例中，每一檔案伺服器110可具有二此種網路轉接器。

在一較佳實例中，磁碟介面組件115包含一已知磁碟介面，以與磁碟，光碟，或磁光碟一起運作，且該介面具有二獨立埠，且每一埠耦接至一分別之檔案伺服器110，例如FC-AI介面。這有助於防止一檔案伺服器110之故障影響



## 五、發明說明 (10)

另一檔案伺服器110之低階運作。

在一較佳實例中，匯流排117包含至少一記憶體匯流排171與中間匯流排130。記憶體匯流排171耦接處理器111與程式及資料記憶體112。中間匯流排130耦接網路介面組件114及磁碟介面組件115。記憶體匯流排171使用輸入/輸出控制器173或其他已知之匯流排轉接器技術來耦接至中間匯流排130。

在一較佳實例中，大量儲存裝置120之每一磁碟是靜態指配給第一檔案伺服器110或第二檔案伺服器110，回應於是否磁碟是連接成為主要由第一檔案伺服器110或第二檔案伺服器110來控制。每一磁碟具有二控制埠A及B；連接至埠A之檔案伺服器110具有該磁碟之主要控制，而另一檔案伺服器110只當另一檔案伺服器110發生故障時方具有該磁碟之控制。

## 運作處理流程

圖3展示該檔案伺服器系統之運作之處理流程圖。

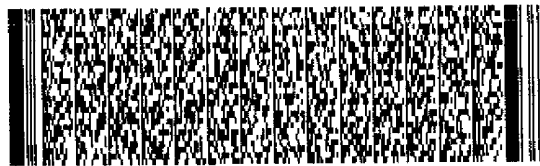
檔案伺服器110之元件執行方法300，且方法300包含一組流程點及處理步驟，如本文所述。

在流程點310，耦接至網路150之一裝置想要提出檔案系統請求151。

在步驟311，該裝置傳送檔案系統請求151至網路150。

在步驟312，網路150傳送檔案伺服器請求151至檔案伺服器110。

在步驟313，檔案伺服器系統100之第一檔案伺服器110



## 五、發明說明 (11)

接收檔案伺服器請求151。第一檔案伺服器110拷貝檔案伺服器請求151至第一伺服器請求記憶體160，且也使用共用連結來拷貝檔案伺服器請求151至第二伺服器請求記憶體160。第二伺服器請求記憶體160之拷貝運作之目標是為此目的所預留之區域。拷貝運作無需第二檔案伺服器110之進一步處理，且第二檔案伺服器110通常不處理或回應檔案伺服器請求151。

在步驟314，第一檔案伺服器110回應檔案伺服器請求151。

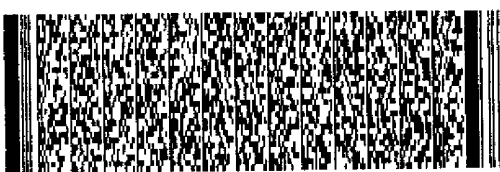
在流程點320，檔案伺服器請求已成功受到處理。

在本發明之第二方面，第一檔案伺服器110提供其之大量儲存裝置120之一或更多之鏡射。

如同本發明之第一方面，每一輸入之檔案伺服器請求拷貝至第一檔案伺服器110及第二檔案伺服器110。第一檔案伺服器110執行檔案伺服器請求以修改在其控制之下之一或更多之主要大量儲存裝置120。第一檔案伺服器110也執行檔案伺服器請求以修改在其控制之下之一組鏡射大量儲存裝置120，但該等鏡射大量儲存裝置120遠離主要大量儲存裝置120。因此，鏡射大量儲存裝置120將是主要大量儲存裝置120之重要拷貝。

萬一整組之主要大量儲存裝置120皆要受到回復，例如如果災難降臨該組主要大量儲存裝置120，則該組之鏡射大量儲存裝置120提供一資源。

在流程點330，檔案伺服器系統100之第一檔案伺服器



## 五、發明說明 (12)

110 發生故障。

在步驟331，檔案伺服器系統100之第二檔案伺服器110  
認知第一檔案伺服器110之故障。

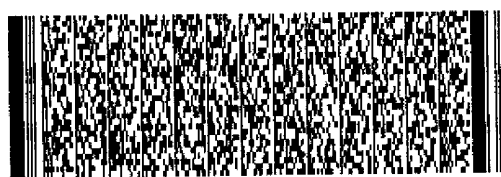
在一較佳實例中，第二檔案伺服器110以下列方式執行  
步驟331：

- 每一檔案伺服器110保存其之大量儲存裝置120之二磁碟  
(因此對於二檔案伺服器110而言總共存在四此種磁碟)，  
以記錄關於檔案伺服器110之狀態資訊。存在二此種磁碟  
(稱為“信箱磁碟”)以致一磁碟可供做為主要儲存而另一  
磁碟可供做為備份儲存。如果二信箱磁碟之一發生故障，  
則使用該信箱磁碟之檔案伺服器110指定另一磁碟來做為  
其之二信箱磁碟之一信箱磁碟。

- 每一檔案伺服器110保存每一信箱磁碟之至少一磁區，  
且檔案伺服器110週期性寫入狀態資訊於該等磁區。每一  
檔案伺服器110也使用連結來傳送其之狀態資訊至另一檔  
案伺服器110，而該連結使用遠端記憶體存取。每一檔案  
伺服器110寫入信箱磁碟之狀態資訊會隨著每一更新而改  
變。

- 每一檔案伺服器110自另一檔案伺服器110之該等信箱磁  
碟之至少一信箱磁碟週期性讀取狀態資訊。每一檔案伺服  
器110也使用連結來接收來自另一檔案伺服器110之狀態資  
訊，且該連結使用遠端記憶體存取。

- 每一檔案伺服器110藉由注意另一檔案伺服器110之信箱  
磁碟是否無狀態資訊之更新，來判斷是否另一檔案伺服器



## 五、發明說明 (13)

110 發生故障。

在一較佳實例中，第二檔案伺服器110決定是否第一檔案伺服器110之故障是硬體錯誤或軟體錯誤，且只認知第一檔案伺服器110由於硬體錯誤所造成之故障。在其他實例中，第二檔案伺服器110也可認知第一檔案伺服器110由於軟體錯誤所造成之故障。

在步驟332，第二檔案伺服器110取得先前指配給第一檔案伺服器110之所有大量儲存裝置120之控制。由於WAFI發明所展示之技術之本質，定義於該等大量儲存裝置120之檔案系統121將處於內部一致狀態。標示成為完成之所有該等檔案伺服器請求151將受到處理，且結果將包含於大量儲存裝置120之儲存區塊。

在正常運作中，檔案伺服器110之任一皆未在大量儲存裝置120之任一裝置進行預留。在步驟332（只有當發生故障服務轉移時），第二檔案伺服器110取得先前由第一檔案伺服器110所控制之大量儲存裝置120之控制，且維持該等大量儲存裝置120之控制直到其認知第一檔案伺服器110回復為止。

當第一檔案伺服器110回復時，其傳送回復訊息至第二檔案伺服器110。在一較佳實例中，第二檔案伺服器110藉由操作員指令來放棄所取得之大量儲存裝置120之控制。但是，在其他實例中，第二檔案伺服器110可認知來自第一檔案伺服器110之回復訊息，並放棄所取得之大量儲存裝置120之控制以做為回應。



## 五、發明說明 (14)

在步驟333，第二檔案伺服器110注意位於其之伺服器請求記憶體160之區域之所有檔案伺服器請求151，且該等請求151是由第一檔案伺服器110拷貝至該區域。結果已包含於儲存裝置120之儲存區塊之該等檔案伺服器請求151會受到捨棄。

在步驟334，當第二檔案伺服器110遇到其之每一檔案伺服器請求151之拷貝時，第二檔案伺服器110以正常方式來處理該檔案伺服器請求151。

在流程點340，自第一檔案伺服器110至第二檔案伺服器110之故障服務轉移已成功受到處理。

## 其他實例

雖然本文已說明較佳實例，存在許多屬於本發明之概念，範疇及精神之變型，且在熟讀本專利申請之後熟悉本技術領域者應可明瞭該等變型。

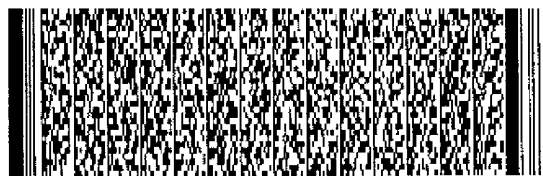


## 四、中文發明摘要 (發明之名稱：高度可用之檔案伺服器)

本發明提供一種儲存系統，及一種用以操作該儲存系統之方法，且即使該儲存系統之元件發生故障，該儲存系統仍是高度可用。第一及第二檔案伺服器皆包含一檔案伺服器請求記錄器以儲存輸入之檔案伺服器請求。第一及第二檔案伺服器皆可存取一組共用之大量儲存組件。每一輸入之檔案伺服器請求皆拷貝至第一及第二檔案伺服器；第一檔案伺服器處理檔案伺服器請求，而第二檔案伺服器保存一拷貝於其之檔案伺服器請求記錄器。每一檔案伺服器藉由使用一檔案系統來運作，且該檔案系統在每一檔案伺服器請求以後保存一致狀態。一旦出現故障服務轉移，第二檔案伺服器可自最近之一致狀態開始執行位於其之檔案伺服器請求記錄器之該等檔案伺服器請求。沒有任何單點之

## 英文發明摘要 (發明之名稱：HIGHLY AVAILABLE FILE SERVERS)

The invention provides a storage system that is highly available even in the face of component failures in the storage system, and a method for operating that storage system. A first and a second file server each includes a file server request log for storing incoming file server requests. Both the first and second file servers have access to a common set of mass storage elements. Each incoming file server request is copied to both the first and



四、中文發明摘要 (發明之名稱：高度可用之檔案伺服器)

故障可阻止對於任何個別大量儲存組件之存取。

英文發明摘要 (發明之名稱：HIGHLY AVAILABLE FILE SERVERS)

second file servers; the first file server processes the file server request while the second file server maintains a copy in its file server request log. Each file server operates using a file system that maintains consistent state after each file server request. On failover, the second file server can perform those file server requests in its file server request log since the most recent consistent state. There is no single point of failure



四、中文發明摘要 (發明之名稱：高度可用之檔案伺服器)

英文發明摘要 (發明之名稱：HIGHLY AVAILABLE FILE SERVERS)

that prevents access to any individual mass storage element.

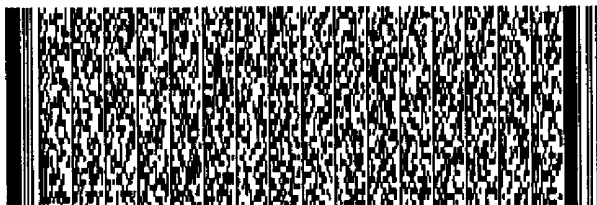


90年3月14日

六、申請專利範圍

1. 一種檔案伺服器系統，且該檔案伺服器系統包含
  - 包含一檔案伺服器變更記憶體之第一檔案伺服器；
  - 包含一檔案伺服器變更記憶體之第二檔案伺服器；
  - 一大量儲存組件；
  - 該第一檔案伺服器及該第二檔案伺服器耦接至該大量儲存組件；
  - 用以拷貝一檔案系統變更之說明碼至該第一及第二檔案伺服器之裝置，其中該第一檔案伺服器處理該檔案系統變更，而該第二檔案伺服器保存其之該說明碼之拷貝於其之檔案伺服器變更記憶體；及
  - 用以使該第二檔案伺服器執行其檔案伺服器變更記憶體之檔案系統變更之裝置，以回應該第一檔案伺服器之服務中斷。
2. 如申請專利範圍第1項之系統，且該系統包含針對每一該檔案伺服器之至少一該大量儲存組件。
3. 如申請專利範圍第1項之系統，其中一第一該檔案伺服器係配置以微細處理該等檔案系統變更，其中一旦發生故障服務轉移，則第二該檔案伺服器可處理一些檔案系統變更，且該等檔案系統變更尚未受到該第一檔案伺服器之處理。
4. 如申請專利範圍第1項之系統，其中第一該檔案伺服器受到配置以同樣回應其本身及第二該檔案伺服器之服務中斷。
5. 如申請專利範圍第1項之系統，其中至少一該檔案伺

煩請未實明未90年3月14日所撰之修正本有無變更實質內容是否准予修正。



## 六、申請專利範圍

伺服器是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案系統變更之回應。

6. 如申請專利範圍第1項之系統，其中至少一該檔案伺服器回應一檔案系統變更，在提交該檔案系統變更之結果至大量儲存組件之前。

7. 如申請專利範圍第1項之系統，其中

該等檔案伺服器之每一檔案伺服器使用本地記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份；且

該等檔案伺服器之每一檔案伺服器使用遠端記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份。

8. 如申請專利範圍第1項之系統，其中該說明碼包含一檔案伺服器請求。

9. 如申請專利範圍第1項之系統，其中該檔案伺服器變更記憶體包含一磁碟區塊。

10. 如申請專利範圍第1項之系統，其中該檔案伺服器變更記憶體包含一檔案伺服器請求。

11. 如申請專利範圍第1項之系統，其中該檔案伺服器變更記憶體是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

12. 如申請專利範圍第1項之系統，其中

該大量儲存組件包含一檔案儲存系統；

每一該檔案伺服器是配置成為在處理檔案系統變更之後使得該檔案儲存系統處於內部一致狀態；

該內部一致狀態相關於一組完成之檔案系統變



## 六、申請專利範圍

更；

該組完成之檔案系統變更可由每一該檔案伺服器來加以辨識。

13. 如申請專利範圍第1項之系統，其中該大量儲存組件包含一檔案儲存系統，且每一該檔案伺服器是配置成為在處理每一該檔案系統變更之後使得該檔案儲存系統處於內部一致狀態。

14. 如申請專利範圍第1項之系統，其中

該大量儲存組件包含一主要大量儲存組件及一鏡射大量儲存組件；且

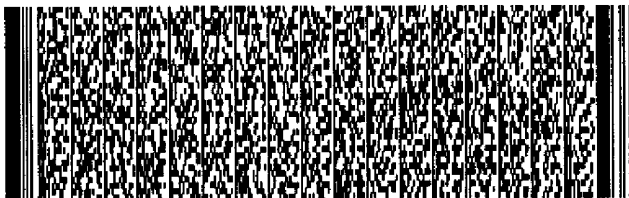
該第一檔案伺服器處理該主要大量儲存組件及該鏡射大量儲存組件之該檔案系統變更。

15. 如申請專利範圍第1項之系統，其中用以拷貝之該裝置包含使用一NUMA網路來存取該第一及第二檔案伺服器變更記憶體之至少一記憶體。

16. 如申請專利範圍第1項之系統，其中用以拷貝之該裝置包含對於該第一及第二檔案伺服器變更記憶體之至少一記憶體之遠端記憶體存取。

17. 如申請專利範圍第1項之系統，其中該第二檔案伺服器用以執行其之檔案伺服器變更記憶體之檔案伺服器請求之該裝置也運作以回應該第二檔案伺服器之服務中斷。

18. 一種檔案伺服器系統，且該種檔案伺服器系統包含  
耦接至第一組大量儲存裝置之第一檔案伺服器；  
耦接至第二組大量儲存裝置之第二檔案伺服器；



## 六、申請專利範圍

一 伺服器變更記憶體；

該第一檔案伺服器是配置成為接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至該伺服器變更記憶體以做為回應；及

該第一檔案伺服器是配置成為處理該第一組大量儲存裝置及該第二組之至少一該大量儲存裝置之該檔案系統變更。

19. 如申請專利範圍第18項之系統，其中

該第二檔案伺服器是配置成為接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至該伺服器變更記憶體以做為回應；及

該第二檔案伺服器是配置成為處理該第二組大量儲存裝置及該第一組之至少一該大量儲存裝置之該檔案系統變更。

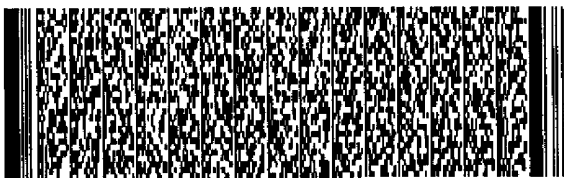
20. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體包含一磁碟區塊。

21. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體包含一檔案伺服器請求。

22. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份。

23. 如申請專利範圍第18項之系統，其中

該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且



## 六、申請專利範圍

該第一檔案伺服器是配置成為拷貝該說明碼至該第一部份及該第二部份。

24. 如申請專利範圍第18項之系統，其中

該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且

該第一檔案伺服器及該第二檔案伺服器皆配置成為拷貝該說明碼至該第一部份及該第二部份。

25. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

26. 一種檔案伺服器系統，且該種檔案伺服器系統包含多個檔案伺服器，且該等多個檔案伺服器耦接至一大量儲存組件及至少一檔案伺服器變更記憶體；

每一該檔案伺服器皆配置成為接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至該檔案伺服器變更記憶體以做為回應；及

每一該檔案伺服器皆配置成為回應執行該檔案伺服器變更記憶體之檔案系統變更所造成之服務中斷。

27. 如申請專利範圍第26項之系統，且該系統包含針對每一該檔案伺服器之至少一該大量儲存組件。

28. 如申請專利範圍第26項之系統，且該系統包含針對每一該檔案伺服器之至少一該伺服器變更記憶體。

29. 如申請專利範圍第26項之系統，其中一第一該檔案伺服器係配置以微細處理該等檔案系統變更，其中一旦發



## 六、申請專利範圍

生故障服務轉移，則第二該檔案伺服器可處理一些檔案系統變更，且該等檔案系統變更尚未受到該第一檔案伺服器之處理。

30. 如申請專利範圍第26項之系統，其中第一該檔案伺服器受到配置以同樣回應其本身及第二該檔案伺服器之服務中斷。

31. 如申請專利範圍第26項之系統，其中至少一該檔案伺服器延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

32. 如申請專利範圍第26項之系統，其中至少一該檔案伺服器回應一檔案系統變更，在提交該檔案系統變更之結果至大量儲存組件之前。

33. 如申請專利範圍第26項之系統，其中

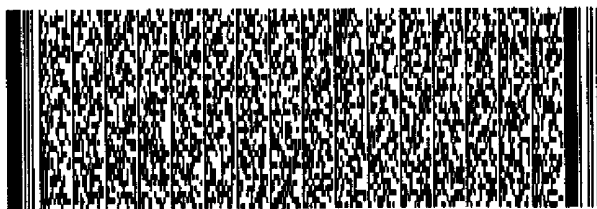
該等檔案伺服器之每一檔案伺服器使用本地記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份；且

該等檔案伺服器之每一檔案伺服器使用遠端記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份。

34. 如申請專利範圍第26項之系統，其中每一該檔案伺服器皆配置使用一NUMA網路來拷貝該等說明碼。

35. 如申請專利範圍第26項之系統，其中每一該檔案伺服器皆配置使用遠端記憶體存取來拷貝該等說明碼。

36. 如申請專利範圍第26項之系統，其中該檔案伺服器變更記憶體包含一磁碟區塊。



## 六、申請專利範圍

37. 如申請專利範圍第26項之系統，其中該檔案伺服器變更記憶體包含一檔案伺服器請求。

38. 如申請專利範圍第26項之系統，其中該檔案伺服器變更記憶體配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

39. 如申請專利範圍第26項之系統，其中該大量儲存組件包含一檔案儲存系統，且每一該檔案伺服器皆配置成為在處理每一該檔案系統變更之後使得該檔案儲存系統處於內部一致狀態。

40. 如申請專利範圍第26項之系統，其中

該大量儲存組件包含一檔案儲存系統；

每一該檔案伺服器皆配置成為在處理檔案系統變更之後使得該檔案儲存系統處於內部一致狀態；

該內部一致狀態相關於一組完成之檔案系統變更；

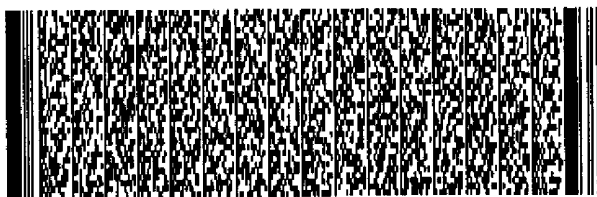
該組完成之檔案系統變更可由每一該檔案伺服器來加以辨識。

41. 如申請專利範圍第26項之系統，其中

該大量儲存組件包含一主要大量儲存組件及一鏡射大量儲存組件；且

該第一檔案伺服器處理該主要大量儲存組件及該鏡射大量儲存組件之該檔案系統變更。

42. 一種用以操作檔案伺服器系統之方法，且該種方法包含下列步驟



## 六、申請專利範圍

藉由拷貝檔案系統變更之說明碼至第一檔案伺服器及第二檔案伺服器來回應於一輸入之檔案伺服器請求；

在該第一檔案伺服器處理該檔案系統變更，且保存該說明碼拷貝於該第二檔案伺服器；及

在該第二檔案伺服器執行一檔案系統變更，以回應一拷貝之說明碼及該第一檔案伺服器之服務中斷。

43. 如申請專利範圍第42項之方法，且該方法包含用以相關第一檔案伺服器及第二檔案伺服器與一大量儲存組件之步驟。

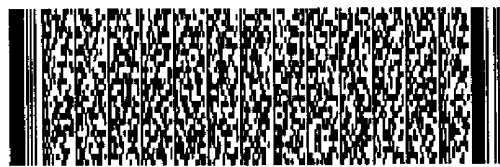
44. 如申請專利範圍第42項之方法，且該方法包含延遲通往該大量儲存系統之至少一該檔案伺服器之輸出，而不延遲對於檔案系統變更之回應之步驟。

45. 如申請專利範圍第42項之方法，其中一第一該檔案伺服器係配置以微細處理該等檔案系統變更，其中一旦發生故障服務轉移，則第二該檔案伺服器可處理一些檔案系統變更，且該等檔案系統變更尚未受到該第一檔案伺服器之處理。

46. 如申請專利範圍第42項之方法，其中第一該檔案伺服器受到配置以同樣回應其本身及第二該檔案伺服器之服務中斷。

47. 如申請專利範圍第42項之方法，其中至少一該檔案伺服器回應一檔案系統變更，在提交該檔案系統變更之結果至大量儲存組件之前。

48. 如申請專利範圍第42項之方法，其中



## 六、申請專利範圍

每一該檔案伺服器皆包含一檔案伺服器變更記憶體；

該等檔案伺服器之每一檔案伺服器使用本地記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份；且

該等檔案伺服器之每一檔案伺服器使用遠端記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份。

49. 如申請專利範圍第42項之方法，其中該檔案伺服器變更記憶體包含一磁碟區塊。

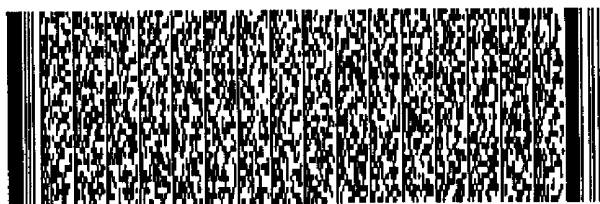
50. 如申請專利範圍第42項之方法，其中該檔案伺服器變更記憶體包含一檔案伺服器請求。

51. 如申請專利範圍第42項之方法，其中該檔案伺服器變更記憶體配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

52. 如申請專利範圍第42項之方法，其中該大量儲存組件包含一檔案儲存系統，且每一該檔案伺服器皆配置成為在處理每一該檔案系統變更之後使得該檔案儲存系統處於內部一致狀態。

53. 如申請專利範圍第42項之方法，其中用以執行檔案系統變更以回應拷貝之說明碼之該步驟也運作以回應該第二檔案伺服器之服務中斷。

54. 如申請專利範圍第42項之方法，其中用以處理之該步驟包含在一主要大量儲存組件及一鏡射大量儲存組件處理該檔案系統變更之步驟。



## 六、申請專利範圍

55. 一種用以操作檔案伺服器系統之方法，且該種方法包含下列步驟

在該等多個檔案伺服器之一檔案伺服器接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至伺服器變更記憶體做為回應；

處理第一組大量儲存裝置及第二組大量儲存裝置之至少一該大量儲存裝置之該檔案系統變更，且該第一組大量儲存裝置耦接至第一該檔案伺服器，而該第二組大量儲存裝置耦接至第二該檔案伺服器。

56. 如申請專利範圍第55項之方法，其中該說明碼包含一檔案伺服器請求。

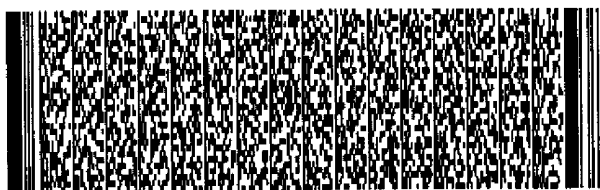
57. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含一磁碟區塊。

58. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含一檔案伺服器請求。

59. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份。

60. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且其中用以拷貝之該等步驟包含用以拷貝該說明碼至該第一部份及該第二部份之步驟。

61. 如申請專利範圍第55項之方法，其中該伺服器變更



## 六、申請專利範圍

記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且用以拷貝之該等步驟包含該第一檔案伺服器或該第二檔案伺服器拷貝該說明碼至該第一部份及該第二部份之步驟。

62. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

63. 如申請專利範圍第55項之方法，其中

用以接收之該等步驟包含在該第一檔案伺服器或該第二檔案伺服器接收一檔案伺服器請求，且用以拷貝該說明碼之該等步驟包含由該第一檔案伺服器或該第二檔案伺服器進行拷貝；且包含步驟以

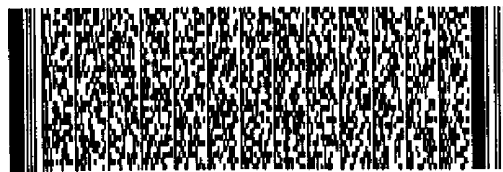
處理該第二組大量儲存裝置及該第一組大量儲存裝置之至少一該大量儲存裝置之該檔案系統變更。

64. 一種用以操作檔案伺服器系統之方法，且該種方法包含下列步驟

在多個檔案伺服器之一檔案伺服器接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至檔案伺服器變更記憶體做為回應；及

藉由執行一檔案系統變更以回應該檔案伺服器變更記憶體之一說明碼來回應一服務中斷。

65. 如申請專利範圍第64項之方法，且該方法包含用以相關多個檔案伺服器及至少一大量儲存組件及至少一檔案伺服器變更記憶體之步驟。



## 六、申請專利範圍

66. 如申請專利範圍第64項之方法，且該方法包含延遲通往一大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應之步驟。

67. 如申請專利範圍第64項之方法，且該方法包含在處理每一該檔案系統變更之後使得該大量儲存組件之一檔案儲存系統處於內部一致狀態。

68. 如申請專利範圍第64項之方法，且該方法包含下列步驟以

在處理檔案系統變更之後使得該大量儲存組件之一檔案儲存系統處於內部一致狀態；

相關該內部一致狀態與一組完成之檔案系統變更；及

藉由至少一該檔案伺服器來辨識該組完成之檔案系統變更。

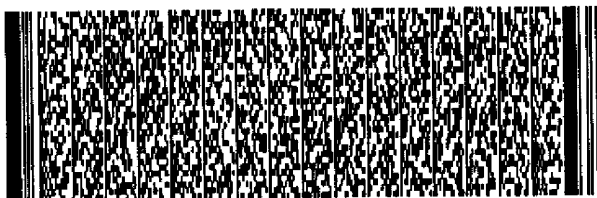
69. 如申請專利範圍第64項之方法，且該方法包含用以在一主要大量儲存組件及一鏡射大量儲存組件執行該接收之檔案伺服器請求之步驟。

70. 如申請專利範圍第64項之方法，且該方法包含下列步驟以

在一第一該檔案伺服器微細處理該檔案系統變更；及

一旦發生故障服務轉移，則處理該第一檔案伺服器尚未處理之該等檔案系統變更。

71. 如申請專利範圍第64項之方法，且該方法包含在第



## 六、申請專利範圍

一該檔案伺服器同樣回應其本身及第二該檔案伺服器之服務中斷之步驟。

72. 如申請專利範圍第64項之方法，其中該檔案伺服器變更記憶體包含一磁碟區塊。

73. 如申請專利範圍第64項之方法，其中該檔案伺服器變更記憶體包含一檔案伺服器請求。

74. 如申請專利範圍第64項之方法，其中該檔案伺服器變更記憶體是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

75. 如申請專利範圍第64項之方法，且該方法包含用以在提交該檔案系統變更之結果至一該檔案伺服器之大量儲存組件之前，回應一檔案系統變更之步驟。



圖式

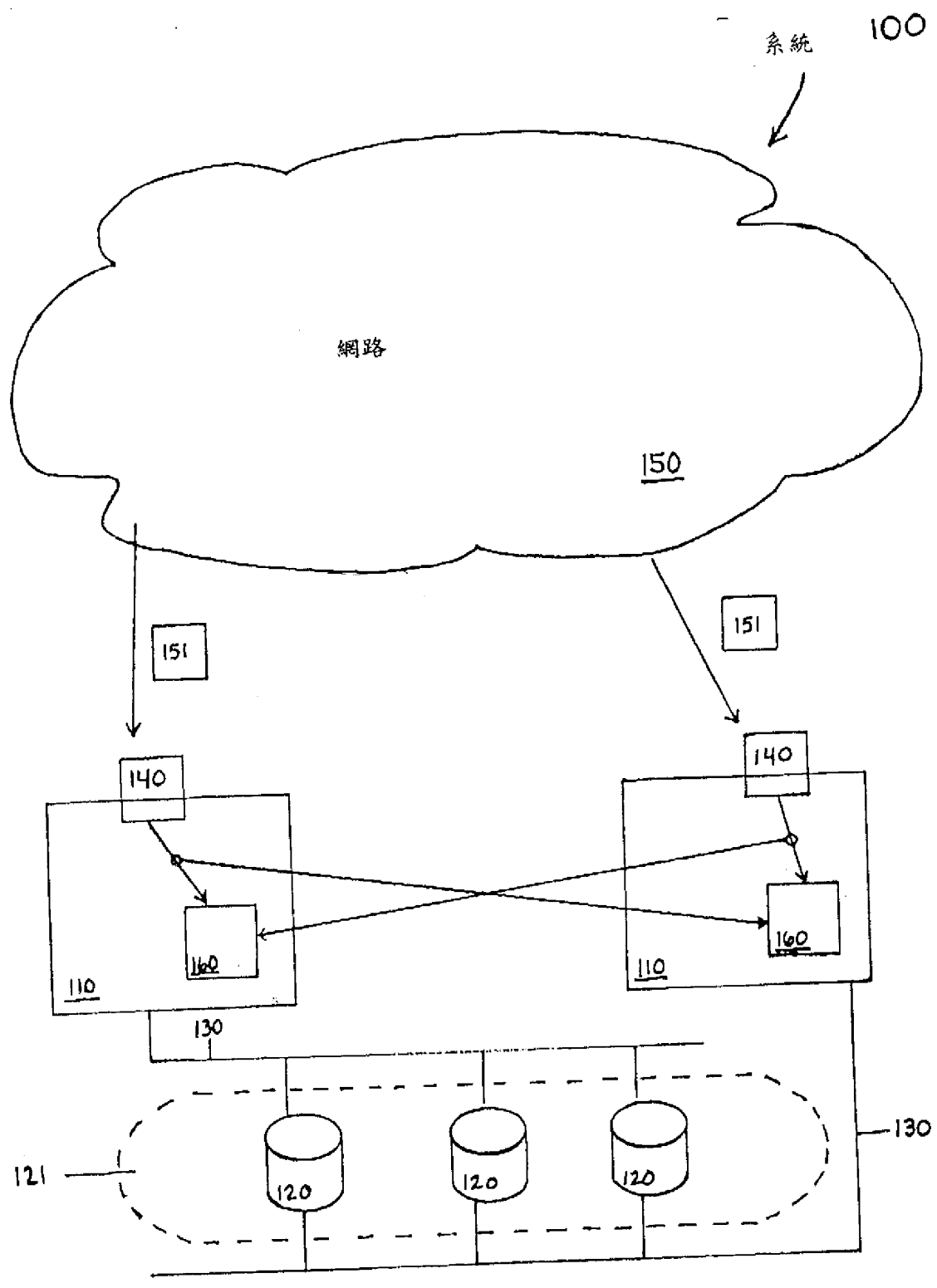


圖 1

圖式

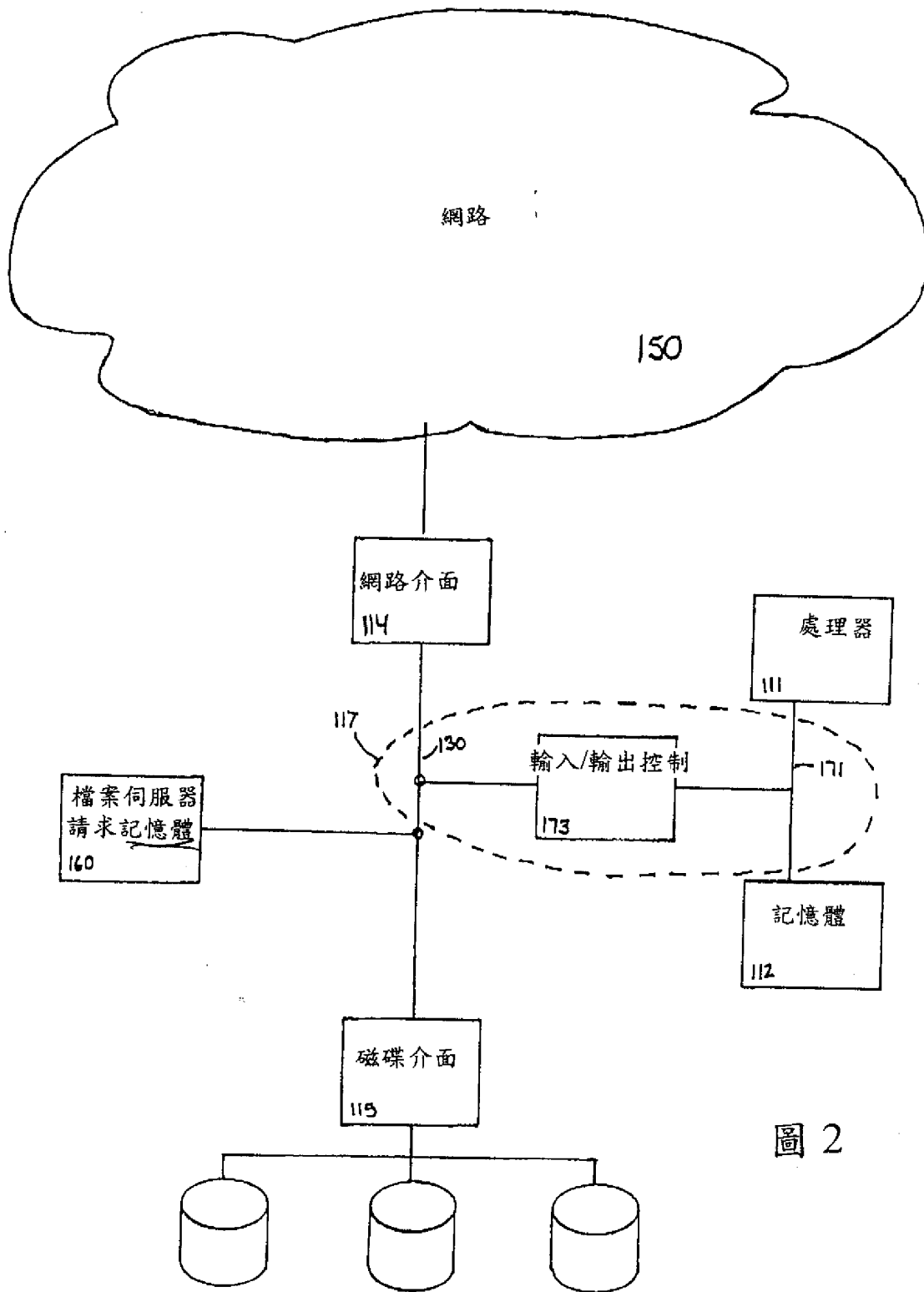


圖 2

圖式

300 方法

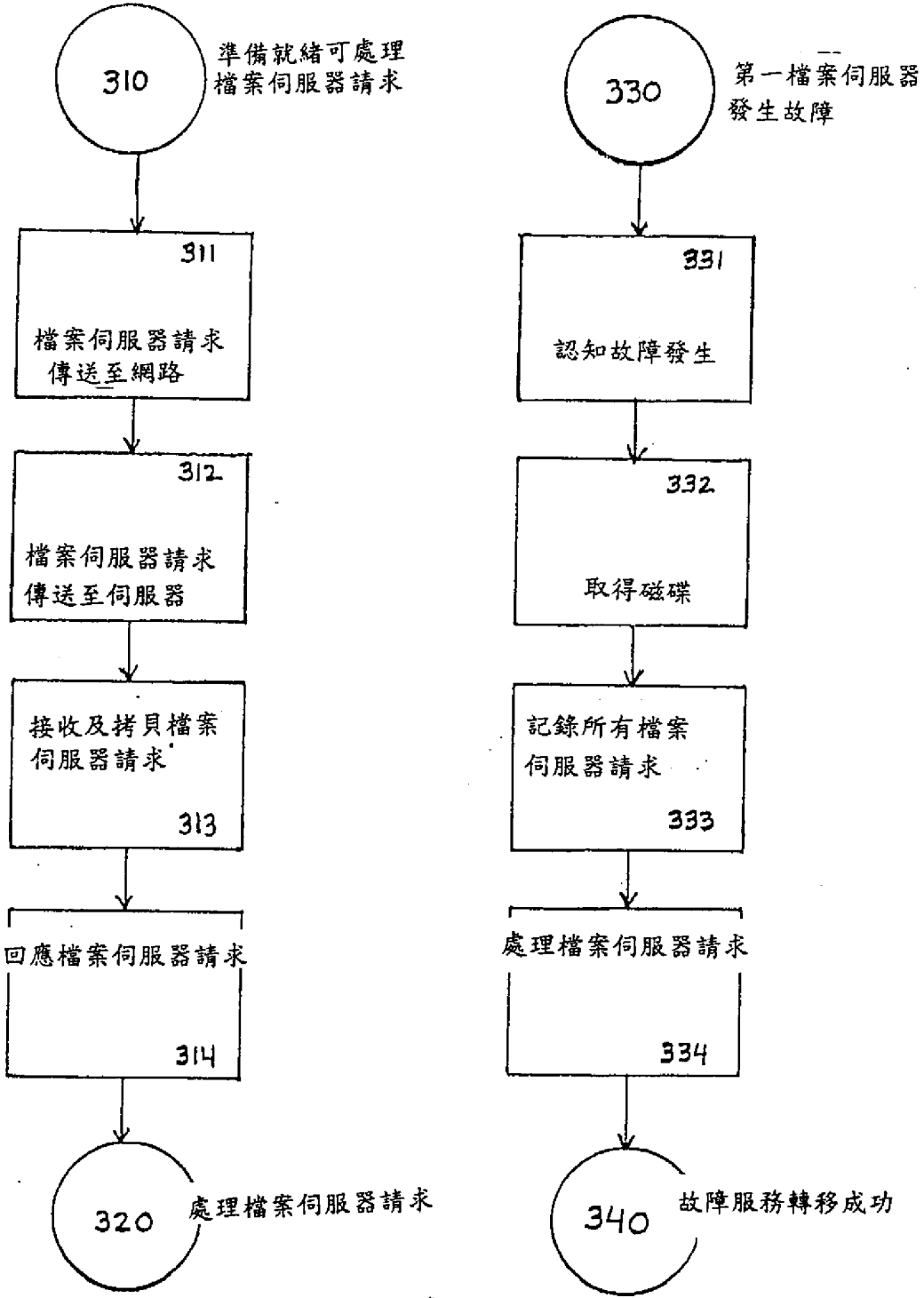


圖 3

## 五、發明說明 (4)

檔案伺服器皆可存取一組共用之大量儲存組件。第一輸入之檔案伺服器請求皆拷貝至第一及第二檔案伺服器；第一檔案伺服器處理檔案伺服器請求，而第二檔案伺服器保存一拷貝於其之檔案伺服器請求記錄器。每一檔案伺服器藉由使用一檔案系統來運作，且該檔案系統在每一檔案伺服器請求以後保存一致狀態。一旦出現故障服務轉移，則第二檔案伺服器可自最近之一致狀態開始執行位於其之檔案伺服器請求記錄器之該等檔案伺服器請求。

在本發明之第二方面，一檔案伺服器系統提供一或更多之大量儲存組件之鏡射。每一輸入之檔案伺服器請求皆拷貝至第一檔案伺服器及第二檔案伺服器。第一檔案伺服器執行檔案伺服器請求以修改一組之主要大量儲存組件，且也執行該等相同之檔案伺服器請求以修改一組之鏡射大量儲存組件。該等鏡射大量儲存組件是配置成為實體上與主要大量儲存組件實體分離，例如位於另一位置，且萬一整組之主要大量儲存組件需要獲得回復則提供一資源。

附圖簡短說明

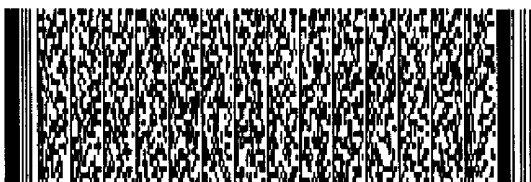
圖1展示一高度可用之檔案伺服器系統之方塊圖。

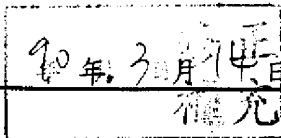
圖2展示該檔案伺服器系統之一檔案伺服器之方塊圖。

圖3展示該檔案伺服器系統之運作之處理流程圖。

元件符號說明

100	檔案伺服器系統	110	檔案伺服器
111	處理器	112	程式及資料記憶體
114	網路介面組件	115	磁碟介面組件





## 五、發明說明 (5)

117	匯流排	120	大量儲存裝置
121	檔案系統	130	中間匯流排
140	網路介面	150	網路
151	檔案伺服器請求	160	伺服器請求記憶體
171	記憶體匯流排	173	輸入／輸出控制器

較佳實例詳細說明

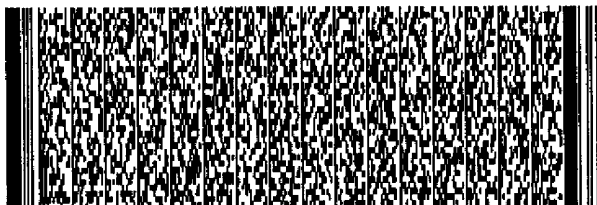
在下列說明中，本發明之一較佳實例之較佳處理步驟及資料結構獲得說明。但是，熟悉本技術領域者應可認知，在熟讀本專利申請之後，本發明之實例可使用在程式控制



六、申請專利範圍

1. 一種檔案伺服器系統，且該檔案伺服器系統包含
  - 包含一檔案伺服器變更記憶體之第一檔案伺服器；
  - 包含一檔案伺服器變更記憶體之第二檔案伺服器；
  - 一大量儲存組件；
  - 該第一檔案伺服器及該第二檔案伺服器耦接至該大量儲存組件；
  - 用以拷貝一檔案系統變更之說明碼至該第一及第二檔案伺服器之裝置，其中該第一檔案伺服器處理該檔案系統變更，而該第二檔案伺服器保存其之該說明碼之拷貝於其之檔案伺服器變更記憶體；及
  - 用以使該第二檔案伺服器執行其檔案伺服器變更記憶體之檔案系統變更之裝置，以回應該第一檔案伺服器之服務中斷。
2. 如申請專利範圍第1項之系統，且該系統包含針對每一該檔案伺服器之至少一該大量儲存組件。
3. 如申請專利範圍第1項之系統，其中一第一該檔案伺服器係配置以微細處理該等檔案系統變更，其中一旦發生故障服務轉移，則第二該檔案伺服器可處理一些檔案系統變更，且該等檔案系統變更尚未受到該第一檔案伺服器之處理。
4. 如申請專利範圍第1項之系統，其中第一該檔案伺服器受到配置以同樣回應其本身及第二該檔案伺服器之服務中斷。
5. 如申請專利範圍第1項之系統，其中至少一該檔案伺

煩請未實明示90年3月14日所提之修正本有無變更實質內容是否准予修正。



## 六、申請專利範圍

伺服器是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案系統變更之回應。

6. 如申請專利範圍第1項之系統，其中至少一該檔案伺服器回應一檔案系統變更，在提交該檔案系統變更之結果至大量儲存組件之前。

7. 如申請專利範圍第1項之系統，其中

該等檔案伺服器之每一檔案伺服器使用本地記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份；且

該等檔案伺服器之每一檔案伺服器使用遠端記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份。

8. 如申請專利範圍第1項之系統，其中該說明碼包含一檔案伺服器請求。

9. 如申請專利範圍第1項之系統，其中該檔案伺服器變更記憶體包含一磁碟區塊。

10. 如申請專利範圍第1項之系統，其中該檔案伺服器變更記憶體包含一檔案伺服器請求。

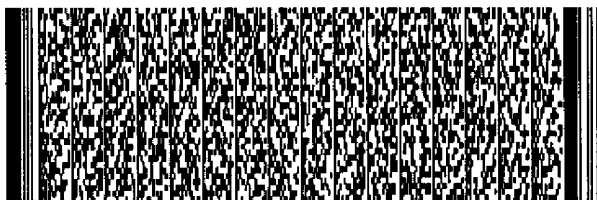
11. 如申請專利範圍第1項之系統，其中該檔案伺服器變更記憶體是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

12. 如申請專利範圍第1項之系統，其中

該大量儲存組件包含一檔案儲存系統；

每一該檔案伺服器是配置成為在處理檔案系統變更之後使得該檔案儲存系統處於內部一致狀態；

該內部一致狀態相關於一組完成之檔案系統變



## 六、申請專利範圍

更；

該組完成之檔案系統變更可由每一該檔案伺服器來加以辨識。

13. 如申請專利範圍第1項之系統，其中該大量儲存組件包含一檔案儲存系統，且每一該檔案伺服器是配置成為在處理每一該檔案系統變更之後使得該檔案儲存系統處於內部一致狀態。

14. 如申請專利範圍第1項之系統，其中

該大量儲存組件包含一主要大量儲存組件及一鏡射大量儲存組件；且

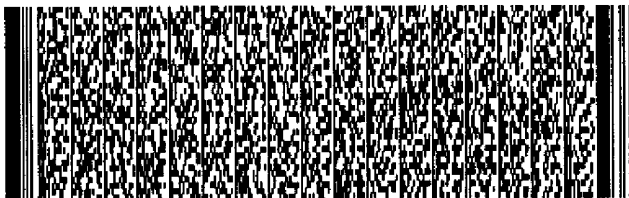
該第一檔案伺服器處理該主要大量儲存組件及該鏡射大量儲存組件之該檔案系統變更。

15. 如申請專利範圍第1項之系統，其中用以拷貝之該裝置包含使用一NUMA網路來存取該第一及第二檔案伺服器變更記憶體之至少一記憶體。

16. 如申請專利範圍第1項之系統，其中用以拷貝之該裝置包含對於該第一及第二檔案伺服器變更記憶體之至少一記憶體之遠端記憶體存取。

17. 如申請專利範圍第1項之系統，其中該第二檔案伺服器用以執行其之檔案伺服器變更記憶體之檔案伺服器請求之該裝置也運作以回應該第二檔案伺服器之服務中斷。

18. 一種檔案伺服器系統，且該種檔案伺服器系統包含  
耦接至第一組大量儲存裝置之第一檔案伺服器；  
耦接至第二組大量儲存裝置之第二檔案伺服器；



## 六、申請專利範圍

一 伺服器變更記憶體；

該第一檔案伺服器是配置成為接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至該伺服器變更記憶體以做為回應；及

該第一檔案伺服器是配置成為處理該第一組大量儲存裝置及該第二組之至少一該大量儲存裝置之該檔案系統變更。

19. 如申請專利範圍第18項之系統，其中

該第二檔案伺服器是配置成為接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至該伺服器變更記憶體以做為回應；及

該第二檔案伺服器是配置成為處理該第二組大量儲存裝置及該第一組之至少一該大量儲存裝置之該檔案系統變更。

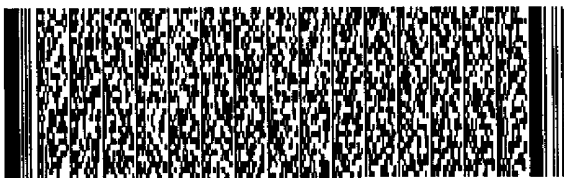
20. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體包含一磁碟區塊。

21. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體包含一檔案伺服器請求。

22. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份。

23. 如申請專利範圍第18項之系統，其中

該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且



## 六、申請專利範圍

該第一檔案伺服器是配置成為拷貝該說明碼至該第一部份及該第二部份。

24. 如申請專利範圍第18項之系統，其中

該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且

該第一檔案伺服器及該第二檔案伺服器皆配置成為拷貝該說明碼至該第一部份及該第二部份。

25. 如申請專利範圍第18項之系統，其中該伺服器變更記憶體是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

26. 一種檔案伺服器系統，且該種檔案伺服器系統包含多個檔案伺服器，且該等多個檔案伺服器耦接至一大量儲存組件及至少一檔案伺服器變更記憶體；

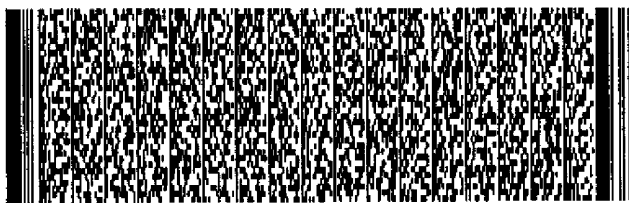
每一該檔案伺服器皆配置成為接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至該檔案伺服器變更記憶體以做為回應；及

每一該檔案伺服器皆配置成為回應執行該檔案伺服器變更記憶體之檔案系統變更所造成之服務中斷。

27. 如申請專利範圍第26項之系統，且該系統包含針對每一該檔案伺服器之至少一該大量儲存組件。

28. 如申請專利範圍第26項之系統，且該系統包含針對每一該檔案伺服器之至少一該伺服器變更記憶體。

29. 如申請專利範圍第26項之系統，其中一第一該檔案伺服器係配置以微細處理該等檔案系統變更，其中一旦發



## 六、申請專利範圍

生故障服務轉移，則第二該檔案伺服器可處理一些檔案系統變更，且該等檔案系統變更尚未受到該第一檔案伺服器之處理。

30. 如申請專利範圍第26項之系統，其中第一該檔案伺服器受到配置以同樣回應其本身及第二該檔案伺服器之服務中斷。

31. 如申請專利範圍第26項之系統，其中至少一該檔案伺服器延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

32. 如申請專利範圍第26項之系統，其中至少一該檔案伺服器回應一檔案系統變更，在提交該檔案系統變更之結果至大量儲存組件之前。

33. 如申請專利範圍第26項之系統，其中

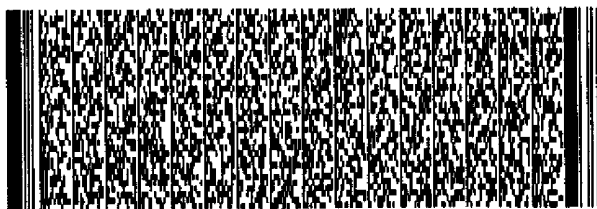
該等檔案伺服器之每一檔案伺服器使用本地記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份；且

該等檔案伺服器之每一檔案伺服器使用遠端記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份。

34. 如申請專利範圍第26項之系統，其中每一該檔案伺服器皆配置使用一NUMA網路來拷貝該等說明碼。

35. 如申請專利範圍第26項之系統，其中每一該檔案伺服器皆配置使用遠端記憶體存取來拷貝該等說明碼。

36. 如申請專利範圍第26項之系統，其中該檔案伺服器變更記憶體包含一磁碟區塊。



## 六、申請專利範圍

37. 如申請專利範圍第26項之系統，其中該檔案伺服器變更記憶體包含一檔案伺服器請求。

38. 如申請專利範圍第26項之系統，其中該檔案伺服器變更記憶體配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

39. 如申請專利範圍第26項之系統，其中該大量儲存組件包含一檔案儲存系統，且每一該檔案伺服器皆配置成為在處理每一該檔案系統變更之後使得該檔案儲存系統處於內部一致狀態。

40. 如申請專利範圍第26項之系統，其中

該大量儲存組件包含一檔案儲存系統；

每一該檔案伺服器皆配置成為在處理檔案系統變更之後使得該檔案儲存系統處於內部一致狀態；

該內部一致狀態相關於一組完成之檔案系統變更；

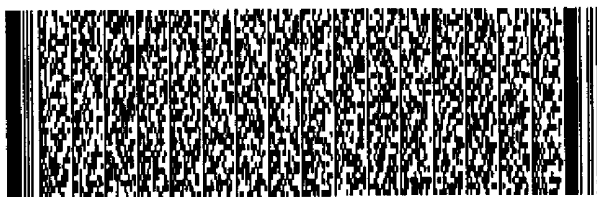
該組完成之檔案系統變更可由每一該檔案伺服器來加以辨識。

41. 如申請專利範圍第26項之系統，其中

該大量儲存組件包含一主要大量儲存組件及一鏡射大量儲存組件；且

該第一檔案伺服器處理該主要大量儲存組件及該鏡射大量儲存組件之該檔案系統變更。

42. 一種用以操作檔案伺服器系統之方法，且該種方法包含下列步驟



## 六、申請專利範圍

藉由拷貝檔案系統變更之說明碼至第一檔案伺服器及第二檔案伺服器來回應於一輸入之檔案伺服器請求；

在該第一檔案伺服器處理該檔案系統變更，且保存該說明碼拷貝於該第二檔案伺服器；及

在該第二檔案伺服器執行一檔案系統變更，以回應一拷貝之說明碼及該第一檔案伺服器之服務中斷。

43. 如申請專利範圍第42項之方法，且該方法包含用以相關第一檔案伺服器及第二檔案伺服器與一大量儲存組件之步驟。

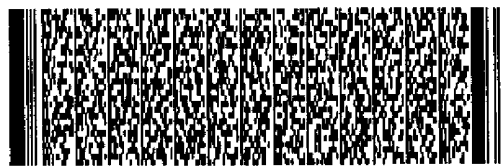
44. 如申請專利範圍第42項之方法，且該方法包含延遲通往該大量儲存系統之至少一該檔案伺服器之輸出，而不延遲對於檔案系統變更之回應之步驟。

45. 如申請專利範圍第42項之方法，其中一第一該檔案伺服器係配置以微細處理該等檔案系統變更，其中一旦發生故障服務轉移，則第二該檔案伺服器可處理一些檔案系統變更，且該等檔案系統變更尚未受到該第一檔案伺服器之處理。

46. 如申請專利範圍第42項之方法，其中第一該檔案伺服器受到配置以同樣回應其本身及第二該檔案伺服器之服務中斷。

47. 如申請專利範圍第42項之方法，其中至少一該檔案伺服器回應一檔案系統變更，在提交該檔案系統變更之結果至大量儲存組件之前。

48. 如申請專利範圍第42項之方法，其中



## 六、申請專利範圍

每一該檔案伺服器皆包含一檔案伺服器變更記憶體；

該等檔案伺服器之每一檔案伺服器使用本地記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份；且

該等檔案伺服器之每一檔案伺服器使用遠端記憶體存取來耦接至該檔案伺服器變更記憶體之至少一部份。

49. 如申請專利範圍第42項之方法，其中該檔案伺服器變更記憶體包含一磁碟區塊。

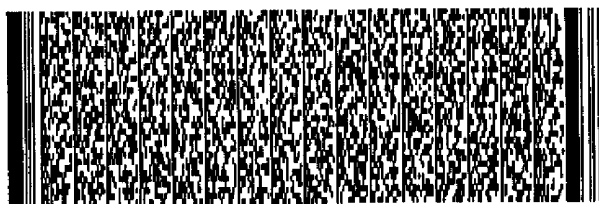
50. 如申請專利範圍第42項之方法，其中該檔案伺服器變更記憶體包含一檔案伺服器請求。

51. 如申請專利範圍第42項之方法，其中該檔案伺服器變更記憶體配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

52. 如申請專利範圍第42項之方法，其中該大量儲存組件包含一檔案儲存系統，且每一該檔案伺服器皆配置成為在處理每一該檔案系統變更之後使得該檔案儲存系統處於內部一致狀態。

53. 如申請專利範圍第42項之方法，其中用以執行檔案系統變更以回應拷貝之說明碼之該步驟也運作以回應該第二檔案伺服器之服務中斷。

54. 如申請專利範圍第42項之方法，其中用以處理之該步驟包含在一主要大量儲存組件及一鏡射大量儲存組件處理該檔案系統變更之步驟。



## 六、申請專利範圍

55. 一種用以操作檔案伺服器系統之方法，且該種方法包含下列步驟

在該等多個檔案伺服器之一檔案伺服器接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至伺服器變更記憶體做為回應；

處理第一組大量儲存裝置及第二組大量儲存裝置之至少一該大量儲存裝置之該檔案系統變更，且該第一組大量儲存裝置耦接至第一該檔案伺服器，而該第二組大量儲存裝置耦接至第二該檔案伺服器。

56. 如申請專利範圍第55項之方法，其中該說明碼包含一檔案伺服器請求。

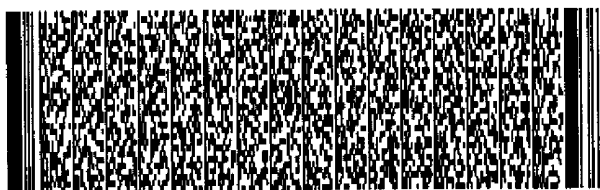
57. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含一磁碟區塊。

58. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含一檔案伺服器請求。

59. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份。

60. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且其中用以拷貝之該等步驟包含用以拷貝該說明碼至該第一部份及該第二部份之步驟。

61. 如申請專利範圍第55項之方法，其中該伺服器變更



## 六、申請專利範圍

記憶體包含配置於該第一檔案伺服器之第一部份及配置於該第二檔案伺服器之第二部份；且用以拷貝之該等步驟包含該第一檔案伺服器或該第二檔案伺服器拷貝該說明碼至該第一部份及該第二部份之步驟。

62. 如申請專利範圍第55項之方法，其中該伺服器變更記憶體是配置成為延遲通往該大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應。

63. 如申請專利範圍第55項之方法，其中

用以接收之該等步驟包含在該第一檔案伺服器或該第二檔案伺服器接收一檔案伺服器請求，且用以拷貝該說明碼之該等步驟包含由該第一檔案伺服器或該第二檔案伺服器進行拷貝；且包含步驟以

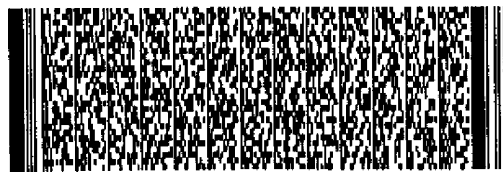
處理該第二組大量儲存裝置及該第一組大量儲存裝置之至少一該大量儲存裝置之該檔案系統變更。

64. 一種用以操作檔案伺服器系統之方法，且該種方法包含下列步驟

在多個檔案伺服器之一檔案伺服器接收檔案伺服器請求，且拷貝檔案系統變更之說明碼至檔案伺服器變更記憶體做為回應；及

藉由執行一檔案系統變更以回應該檔案伺服器變更記憶體之一說明碼來回應一服務中斷。

65. 如申請專利範圍第64項之方法，且該方法包含用以相關多個檔案伺服器及至少一大量儲存組件及至少一檔案伺服器變更記憶體之步驟。



## 六、申請專利範圍

66. 如申請專利範圍第64項之方法，且該方法包含延遲通往一大量儲存組件之輸出，而不延遲對於檔案伺服器請求之回應之步驟。

67. 如申請專利範圍第64項之方法，且該方法包含在處理每一該檔案系統變更之後使得該大量儲存組件之一檔案儲存系統處於內部一致狀態。

68. 如申請專利範圍第64項之方法，且該方法包含下列步驟以

在處理檔案系統變更之後使得該大量儲存組件之一檔案儲存系統處於內部一致狀態；

相關該內部一致狀態與一組完成之檔案系統變更；及

藉由至少一該檔案伺服器來辨識該組完成之檔案系統變更。

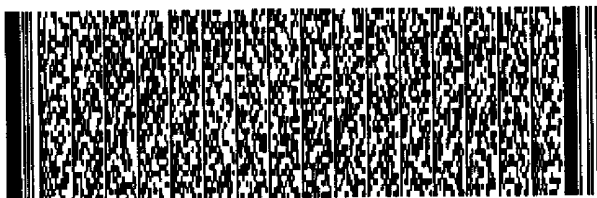
69. 如申請專利範圍第64項之方法，且該方法包含用以在一主要大量儲存組件及一鏡射大量儲存組件執行該接收之檔案伺服器請求之步驟。

70. 如申請專利範圍第64項之方法，且該方法包含下列步驟以

在一第一該檔案伺服器微細處理該檔案系統變更；及

一旦發生故障服務轉移，則處理該第一檔案伺服器尚未處理之該等檔案系統變更。

71. 如申請專利範圍第64項之方法，且該方法包含在第



圖式

300 方法

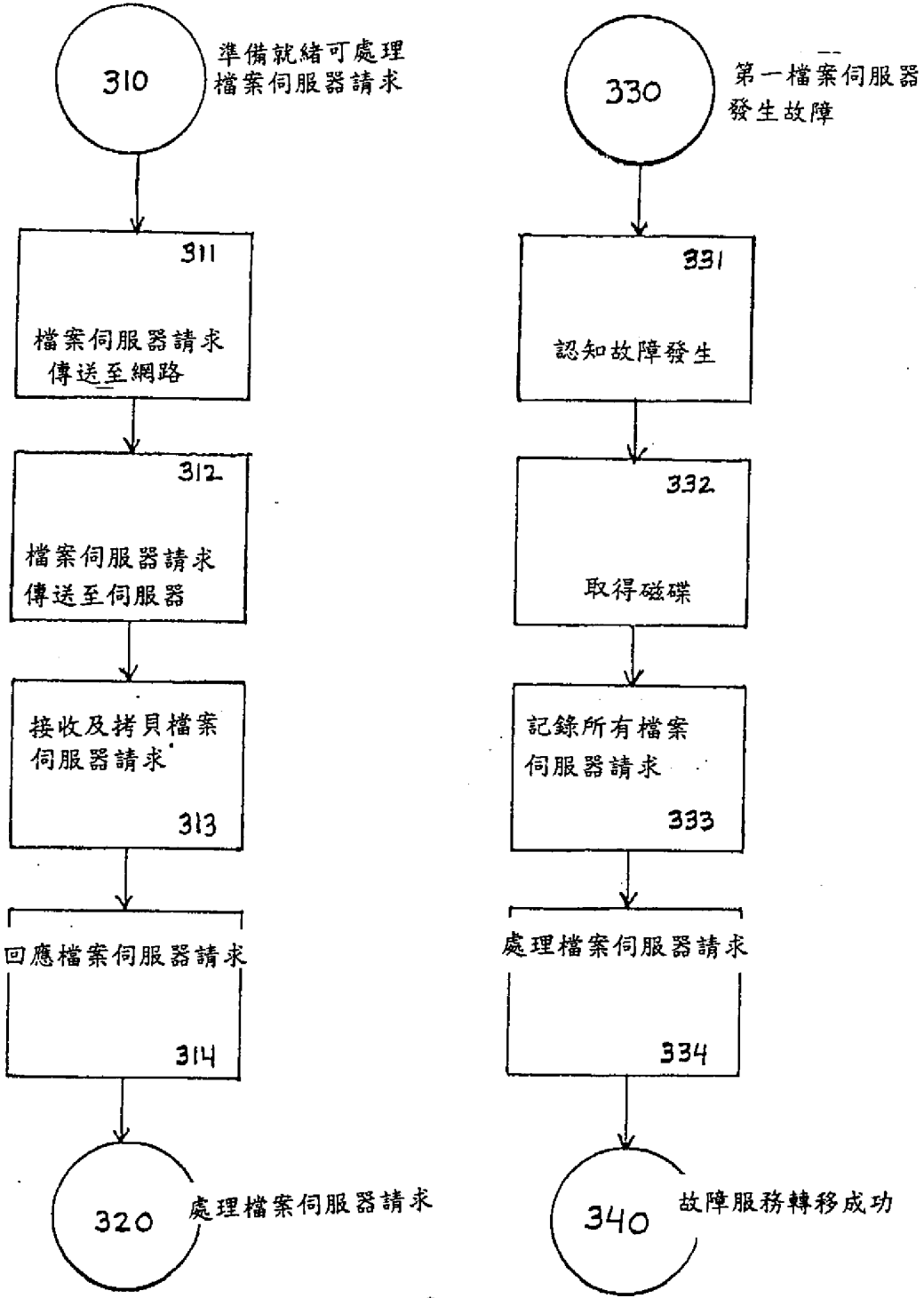


圖 3