

(21)申請案號：102145155

(22)申請日：中華民國 102 (2013) 年 12 月 09 日

(51)Int. Cl. : G06F15/167 (2006.01)

G06F11/14 (2006.01)

G06F11/16 (2006.01)

(71)申請人：安然國際科技有限公司 (中華民國) (TW)

臺北市文山區木新路 2 段 113 號

(72)發明人：陳勳元 (TW)

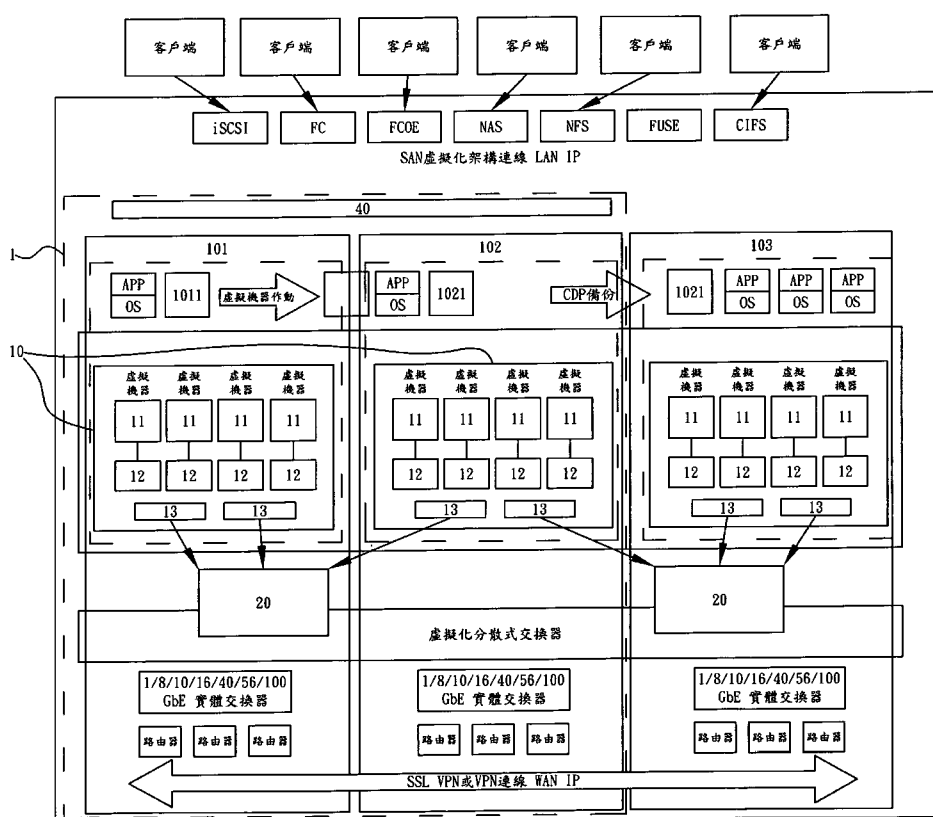
申請實體審查：無 申請專利範圍項數：28 項 圖式數：2 共 36 頁

(54)名稱

分散式記憶體磁碟群集儲存系統運作方法

(57)摘要

本發明係有關於一種分散式記憶體磁碟群集儲存系統運作方法，其採用分散式記憶體儲存系統，可以滿足四種需求的擴充，即網路頻寬傳輸的擴充、磁碟容量的擴充、IOPS 速度的擴充、記憶體 I/O 傳輸速度增加的擴充，同時可以做到滿足跨區域、資料中心、WAN 的使用，讓用戶需求可以在當地的群集記憶體磁碟集中提供服務、也可以漸進式的擴充群集記憶體磁碟容量，進一步提供跨區域或跨國的資料服務。



- 1：群集架構
- 10：電腦單元
- 11：區塊記憶體磁碟
- 12：硬碟儲存裝置
- 13：網路卡
- 20：連接埠群組現用鍵
- 40：分散式記憶體檔案系統
- 101：第一資料中心
- 102：第二資料中心
- 103：備份中心
- 1011：群集資料主控站
- 1021：群集資料備份站

第一圖

發明摘要

※ 申請案號：102145155

※ 申請日：102.12.09

※IPC 分類：G06F 15/167 2006.01

G06F 11/14 2006.01

G06F 14/16 2006.01

【發明名稱】(中文/英文)

分散式記憶體磁碟群集儲存系統運作方法

【中文】

本發明係有關於一種分散式記憶體磁碟群集儲存系統運作方法，其採用分散式記憶體儲存系統，可以滿足四種需求的擴充，即網路頻寬傳輸的擴充、磁碟容量的擴充、IOPS 速度的擴充、記憶體 I/O 傳輸速度增加的擴充，同時可以做到滿足跨區域、資料中心、WAN 的使用，讓用戶需求可以在當地的群集記憶體磁碟集中提供服務、也可以漸進式的擴充群集記憶體磁碟容量，進一步提供跨區域或跨國的資料服務。

【英文】

【代表圖】

【本案指定代表圖】：第（一）圖。

【本代表圖之符號簡單說明】：

群集架構(1)

電腦單元(10)

第一資料中心(101)

群集資料主控站(1011)

第二資料中心(102)

群集資料備份站(1021)

備份中心(103)

區塊記憶體磁碟(11)

硬碟儲存裝置(12)

網路卡(13)

連接埠群組現用鏈(20)

分散式記憶體檔案系統(40)

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

發明專利說明書

(本說明書格式、順序，請勿任意更動)

【發明名稱】(中文/英文)

分散式記憶體磁碟群集儲存系統運作方法

【技術領域】

【0001】本發明係關於一種分散式記憶體磁碟群集儲存系統運作方法，特指一種具備快速多對多傳輸、高擴充性及穩定性等多特徵之網路資訊交流儲存系統。

【先前技術】

【0002】在近年來中，隨著網路應用的快速普及與行動裝置的網路需求增長，讓企業訊息系統有了很大的改變。計算機資源的需求從未達到如此的高度，在面對同時間暴增的用戶量下，以目前的儲存設備是無法應付這樣龐大數量用戶的連線與頻寬需求。

【0003】首先，於習知的網路處理伺服器硬體系統中，電子資訊需透過網路橋接做為彼此之間相互對傳之途徑，然而於資訊量位於正常水平之時，網路傳輸量雖然足以應付，但在資訊量突增之時，因為網路橋接傳輸之傳輸率有其上限值，因此無論伺服器計算機本身硬體等級為何，亦無法增加龐大資訊之處理速度，而會使大眾使用網路系統時面對資訊延遲或者中斷等等窘境，此為一目前技術之缺失。

【0004】此外，於現有技術中，因為記憶體之技術發

展未至現今之高度，因此其儲存量受限之故，僅能做為暫存之用，故傳統之伺服器主機硬體規格皆是以硬碟設備做為主要之儲存以及主運作系統之配置空間。

【0005】而於硬體與硬體間之資訊傳輸而言，處理單元與記憶體之資訊對傳速度遠遠高於處理單元與硬碟設備間之傳輸速度，亦即，習知技術中以硬碟設備為主要儲存空間來做運算的支援，是一導致處理效能無法達到最佳之主因，尤其於處理資訊量大時此問題更是突顯放大，此外更遑論硬碟設備於使用壽命上更是遠遠不能與記憶體相比，透過硬碟設備做為主要的儲存方式對於整體系統而言，於應用上之保障有限，非為一最佳之模式。

【0006】綜上可知，於現有技術中網路資訊之處理設備處理龐大的資訊時，其效能無法發揮處理器之真實處理效率之問題可見一斑，故綜觀前所述，本發明之發明人思索並設計一種分散式記憶體磁碟群集儲存系統運作方法，以期針對現有技術之缺失加以改善，進而增進產業上之實施利用。

【發明內容】

【0007】有鑒於前述之現有技術的不足點，本發明係設計一種具備新穎性、進步性及產業利用性等專利要件之分散式記憶體磁碟群集儲存系統運作方法，以期克服現有技術之難點。

【0008】為達到上述目的，本發明所採用的技術手段為設計一種分散式記憶體磁碟群集儲存系統運作方法，其為：首先做分散記憶體式儲存設備之建置，其係包含將複數台的電腦單元組成群集架構而形成群集記憶體磁碟；而後在電腦單元安裝系統虛擬機器平台作業系統，藉此安裝令電腦單元形成多個虛擬機器，而該電腦單元設定利用作業系統或程式軟體來設定記憶體容量占用之方式，將該記憶體規劃成儲存裝置，做成複數個區塊記憶體磁碟；進而將每一個檔案拆成一到多份的資料，保持一到多份的複本複製平均分散在所有區塊記憶體磁碟上，並使用多通道的記憶體匯流排平行存取記憶體模組，而將記憶體模組之容量規劃成磁碟使用，而其中該記憶體模組之存取支援所有作業系統檔案格式，並且採用分散式存放架構，將資料可以複製複本一份以上；當虛擬機器之作業系統在記憶體模組中直接由 CPU 存取需要的檔案，處理資料後存回記憶體模組中，虛擬機器做計算使用的記憶體也在記憶體模組中，計算完依然直接存回在所存放虛擬機器作業系統的記憶體模組位置，讓這個儲存系統內資料可以在電腦單元中以平行運算的方式在區塊記憶體磁碟處理資料；其他沒有在電腦單元的資料係透過網路卡連接一連接埠群組現用鏈以存取其他電腦單元的區塊記憶體磁碟；而各電腦單元進一步依功能性區分為一第一資料中心、至少一個第二資料中心及一備份中心；其中該第一資料備份中心係以一虛擬群集資料主控站控管，其中各該第二資料中心係以

一 虛擬群集資料備份站控管，且其中該備份中心係以虛擬群集資料備份站控管，而其中該第一資料中心與該第二資料中心共同形成一分散式記憶體檔案系統；此外，更進一步是為形成堆疊架構擴展儲存容量架構，其使用網路層介面的存取方式將一台電腦單元中多個區塊記憶體磁碟規劃成一個群集概念的群集記憶體磁碟單元資源池，同步使用所有區塊記憶體磁碟取得資料；當群集架構產生後，每一個群集架構都可以獨立使用，並且可以將每一個群集架構當作分散記憶體式儲存設備單元，同時使用網路層介面堆疊，可以將每個分散記憶體式儲存設備單元模擬成大型的記憶體磁碟，並採用新的群集資料主控站與群集資料備份站來控制資料的處理量分散在所有大型的記憶體磁碟。

【0009】 其中，各該區塊記憶體磁碟係分別電性連接於至少一硬碟儲存裝置，該硬碟儲存裝置係可於一間隔時間將該區塊記憶體磁碟之資料做備份。

【0010】 其中，所有的電腦單元中的區塊記憶體磁碟，都會使用連續資料保護方式，無間斷的連續將資料備份到統一的大型物理硬碟群集磁碟陣列中備份。

【0011】 其中，該電腦單元係俱備了 CPU、記憶體、硬碟、網路卡、主機板、I/O 介面卡、連接線、外殼之計算機設備。

【0012】 其中，每份複本都可以採用 1-4096 位元 AES

與 RSA 混用加密資料存放在記憶體上，每當要存取資料時，資料在記憶體與 CPU 之間流動，虛擬機器係以檔案形式存在於記憶體模組中，虛擬記憶體所規劃使用的記憶體容量也在同一區塊。

【0013】其中，各該區塊記憶體磁碟透過一監控單元來做其運作情況之監控，該偵測單元利用 Splunk 或者是其他搜索引擎之軟體程式來監控之，當檢測到問題時更可以配置重新啟動應用程式服務，達到一回復之功能。

【0014】其中，虛擬機器平台作業系統採用 VMware vSphere ESXi 4.1 之後的版本、Microsoft Server 2012 R2 Hyper-V 之後的版本、Citrix XenServer Oracle VM、Oeacle VM、Red Hat KVM 或 Linux-Base KVM。其中虛擬機器作業系統採用 Linux、FreeBSD、OpenSolaris 或 MacOS X。

【0015】其中，虛擬機器之記憶體間係使用存儲區域網路架構運行，採取軟體虛擬出來的網路層介面，將所有的區塊記憶體磁碟連接在一起協同運作。

【0016】其中，該網路層介面，係使用 SAN、SAN iSCSI、SAN FC、SAN FCoE、NFS、NAS、JBOD、CIFS 或 FUSE 介面進行伺服器 and 磁碟驅動器設備之間的通信，並且開發有 RAMSTORAGE™ API 作為應用程式支援用。其中該 RAMSTORAGE™ API 係採用 REST、Restful、C++、PHP、Python、Java、Perl 或 Javascript 程式開發

軟體所形成的 RAMSTORAGE™ API，存取分散記憶體式儲存設備單元 API 功能，其包含容錯、備份、移轉、快速佈署虛擬機器、管理磁碟大小、自動依照實際需求容量自動增加區塊記憶體磁碟、區塊間資料負載平衡移轉、備份回復的功能、連續資料保護、快照、監控資源等。

【0017】其中，未使用的 CPU、記憶體、實際物理硬碟的資源，係使用虛擬機器平台作業系統將所有資源規劃成統一的資源池，令每一個所需要電腦資源，可以調整移轉到其他資源較充裕的電腦單元。

【0018】其中，多個分散記憶體式儲存設備單元之連接方式依照實體的網路協定傳送，並且採用 SSL VPN、VPN 或加密演算方式傳送封包，每當網路連線不通時，每個區域都可以單獨運作，當連線回復時，資料將完整同步到每個分散記憶體式儲存設備單元中的每個區塊記憶體磁碟。

【0019】其中，該 CPU 係為 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 其中之一種處理器。

【0020】其中，該電腦單元之記憶體使用方式是直接由 CPU 的記憶體控制器直接以三通多或多通道以 800 MHz 到 1333 MHz 以上的速度直接存取記憶體資料。

【0021】其中，記憶體容量為 1MB 至 16 ZB。其中所採用的記憶體種類為動態異步記憶體 DRAM、動態同步

記憶體 DRAM、動態移動平台記憶體、動態繪圖處理記憶體、動態 Rambus 記憶體、靜態隨機存取存儲器、唯讀記憶體或快閃記憶體。

【0022】其中，動態異步記憶體 DRAM 為 FPM RAM 或 EDO RAM。其中動態同步記憶體 DRAM 為 SDRAM、DDR SDRAM、DDR2 SDRAM、DDR3 SDRAM、DDR4 SDRAM 或 DDR5 SDRAM。其中動態移動平台記憶體為 LPDDR、LPDDR2、LPDDR3 或 LPDDR4。其中動態繪圖處理記憶體為 VRAM、WRAM、MDRAM、SGRAM、SDRAM、GDDR、GDDR2、GDDR3、GDDR4、GDDR5、GDDR6 或 GDDR7。其中動態 Rambus 記憶體為 RDRAM、XDR DRAM 或 XDR2 DRAM。其中快閃記憶體為 NOR Flash、NAND Flash、3D NAND Flash、V-Flash、SLC、MLC、eMMC 或 TLC。

【0023】其中，硬碟儲存裝置為傳統磁頭、磁片的硬碟、固態硬碟、網路硬碟、SAS 介面硬碟、SATA 介面硬碟、mSATA 介面硬碟、mPCIe 介面硬碟、FC 介面硬碟、SCSI 介面硬碟、ATA 介面硬碟、NAND Flash 外接卡或 FCoE 介面硬碟。

【0024】其中，網路卡為乙太網路、快速乙太網路、千兆乙太網路、光纖、令牌環網、InfiniBand、FCoE 或無線網路。

【0025】其中，網路速度採用 2Mbit/s、10 Mbit/s、11Mbit/s、40Mbit/s、54Mbit/s、80Mbit/s、100 Mbit/s、150Mbit/s、300Mbit/s、433Mbit/s、1000

Mbit/s、1Gbit/s、8Gbit/s、10Gbit/s、16Gbit/s、32Gbit/s、40Gbit/s、56Gbit/s、100Gbit/s、160Gbit/s 或 1000Gbit/s 以上的網路速度。

【0026】其中，主機板相容 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 中任何一種處理器。

【0027】其中，作業系統檔案格式是 AdvFS、BFS、Btrfs、Coda、CrossDOS、DFS、Episode、EFS、exFAT、ext、FAT、GFS、HFS、HFS Plus、高效能檔案系統、IBM 通用並列檔案系統、JFS、Macintosh 檔案系統、MINIX、NetWare 檔案系統、NILFS、Novell 儲存服務、NTFS、QFS、QNX4FS、ReiserFS、SpadFS、UBIFS、Unix 檔案系統、Veritas VxFS、VFAT、WAFL、XFS、Xsan、ZFS、CHFS、FFS2、F2FS、JFFS、JFFS2、LogFS、NVFS、YAFFS、UBIFS、DCE/DFS、MFS、CXFS、GFS2、Google 檔案系統、OCFS、OCFS2、QFS、Xsan、AFS、OpenAFS、AFP、MS-DFS、GPFS、Lustre、NCP、NFS、POHMELFS、Hadoop、HAMMER、SMB (CIFS)、cramfs、FUSE、SquashFS、UMSDOS、UnionFS、configfs、devfs、procfs、specfs、sysfs、tmpfs、WinFS、EncFS、EFS、ZFS、RAW、ASM、LVM、SFS、MPFS 或 MGFS。

【0028】本發明之分散式記憶體磁碟群集儲存系統運作方法於設計上係透過本發明所提出之系統架構方法之下，我們採用分散式記憶體儲存系統，可以滿足四種需求的擴充，即網路頻寬傳輸的擴充、磁碟容量

的擴充、IOPS 速度的擴充、記憶體 I/O 傳輸速度增加的擴充。同時可以做到滿足跨區域、資料中心、WAN 的使用，讓用戶需求可以在當地的群集記憶體磁碟 (Memory Disk Cluster) 集中提供服務、也可以漸進式的擴充群集記憶體磁碟容量，進一步提供跨區域或跨國的資料服務。

【0029】在隨著儲存器的增加，每增加一台伺服器就累加一台機器的網路頻寬與磁碟容量，當形成一個資源池 (Resource Pool) 時，這分散記憶體式儲存設備單元就像一顆實體硬碟，其中一台實體主機故障時不會影響整體運作，複本中區塊記憶體磁碟會重新複製資料到新的區塊記憶體磁碟，維持一定資料備份基礎，同時也採取連續資料複製 (Continuous Data Protector, CDP) 為企業資料的備份和回復提供創新技術，以破除磁帶備份經常失敗，一天只能備份一次問題。

【0030】也讓複本產生資料可以由不同區塊記憶體磁碟送出達到多對一傳輸資料，當用戶數量增加，只要增加區塊記憶體磁碟數量，可以達到多對多的傳輸，讓 RAID 的多數硬碟崩潰造成整體資料遺失的缺點，儲存器網路介面數量與速度的限制造成用戶過多的資料擁擠造成的延遲，LUN 擴充數量的問題，資料中心無法跨區使用的問題，因為採用記憶體式磁碟方式，每個檔案或是每個虛擬機都可以用檔案形式存在於記憶體，直接使用記憶體匯流排最高的 I/O 高速度，資料在 CPU 與記憶體中搬遷移動，I/O 次數、距離、

速度都是最高的，綜觀上述優勢，是為習知技術所不能及者，故可見其增益性所在。

【0031】而為了讓上述目的、技術特徵以及實際實施後之增益性更為明顯易懂，於下文中將係以較佳之實施範例輔佐對應相關之圖式來進行更詳細之說明。

【圖式簡單說明】

【0032】

第一圖為本發明之分散式記憶體磁碟群集儲存系統運作方法之實施示意圖。

第二圖為本發明之分散式記憶體磁碟群集儲存系統運作方法之實施示意圖。

【實施方式】

【0033】為利貴審查員瞭解本發明之發明特徵、內容與優點及其所能達成之功效，茲將本發明配合附圖，並以實施例之表達形式詳細說明如下，而其中所使用之圖式，其主旨僅為示意及輔助說明書之用，未必為本發明實施後之真實比例與精準配置，故不應就所附之圖式的比例與配置關係解讀、侷限本發明於實際實施上的權利範圍。

【0034】本發明之優點、特徵以及達到之技術方法將參照例示性實施例及所附圖式進行更詳細地描述而更容易理解，且本發明或可以不同形式來實現，故不應被理解僅限於此處所陳述的實施例，相反地，對所屬技術領域具有通常知識者而言，所提供的實施例將

使本揭露更加透徹與全面且完整地傳達本發明的範疇，且本發明將僅為所附加的申請專利範圍所定義。

【0035】而除非另外定義，所有使用於後文的術語(包含科技及科學術語)與專有名詞，於實質上係與本發明所屬該領域的技術人士一般所理解之意思相同，而例如於一般所使用的字典所定義的那些術語應被理解為具有與相關領域的內容一致的意思，且除非明顯地定義於後文，將不以過度理想化或過度正式的意思理解，合先敘明。

【0036】請配合參看第一圖所示，本發明提出一種分散式記憶體磁碟群集儲存系統運作方法，其中該分散式記憶體磁碟群集儲存系統運作方法於一較佳之實施方式可為：

【0037】首先做分散記憶體式儲存設備之建置，其係包含將複數台的電腦單元(10)組成群集架構(1)而形成群集記憶體磁碟；而其中該電腦單元(10)係俱備了CPU、記憶體、硬碟、網路卡、主機板、I/O 介面卡、連接線、外殼之計算機設備。

【0038】主要在電腦單元(10)安裝系統虛擬機器平台作業系統，藉此安裝令電腦單元(10)形成多個虛擬機器，而該電腦單元(10)設定所須要使用機器記憶體資源容量，係利用作業系統來設定記憶體容量占用之方式，或者是利用程式軟體將該記憶體規劃成硬碟裝置，做成一個區塊記憶體磁碟(chunk memory disk)(11)，

這便如同硬碟的磁軌一般。

【0039】進而得以將一個檔案拆成一到多份的資料，檔案大小或可為 64MB 以上，保持一到多份的複本複製平均分散在所有區塊記憶體磁碟(11)上，故資料實際存放在記憶體模組裡，並使用多通道的記憶體匯流排平行存取記憶體模組，而將記憶體模組之容量規劃成磁碟使用，而其中該記憶體模組之存取支援所有作業系統檔案格式，並且採用分散式存放架構，將資料可以複製複本一份以上，藉此方式，即使機器損毀與資料中心損毀，資料中心依然可以運作。

【0040】每份複本資料都可以採用 1-4096 位元 AES 與 RSA 混用加密資料存放在記憶體上，每當要存取資料時，資料在記憶體與 CPU 之間流動，所耗費的 I/O 存取次數與距離是最少的，虛擬機器係以檔案形式存在於記憶體模組中，虛擬記憶體所規劃使用的記憶體容量也在同一區塊。

【0041】當虛擬機器之作業系統在記憶體模組中直接由 CPU 存取需要的檔案，處理資料後存回記憶體模組中，虛擬機器做計算使用的記憶體也在記憶體模組中，計算完依然直接存回在所存放虛擬機器作業系統的記憶體模組位置，相對依照存取路徑的減少與使用最快的 I/O 速度，讓這個儲存系統大部分資料可以在電腦單元(10)中以平行運算的方式在區塊記憶體磁碟(11)處理資料，其他沒有在電腦單元(10)的資料才透過網路卡(13)連接一連接埠群組現用鏈(20)以存取

其他電腦單元(10)的區塊記憶體磁碟(11)。

【0042】其中，虛擬機器平台作業系統，可以採用 VMware vSphere ESXi 4.1 之後的版本、Microsoft Server 2012 R2 Hyper-V 之後的版本、Citrix XenServer Oracle VM、Oeacle VM、Red Hat KVM、Linux-Base KVM；而虛擬機器作業系統，可以採用可以採用 Linux (Linux 2.6.14 and up have FUSE support included in the official kernel)、FreeBSD、OpenSolaris、MacOS X。

【0043】再者，各該區塊記憶體磁碟(11)更可能透過一監控單元來做其運作情況之監控，該偵測單元或可利用 Splunk 或者是其他搜索引擎之軟體程式來監控之，當檢測到問題時更可以配置重新啟動應用程式服務，達到一回復之功能，而這類型之程式為習知技術，故不在此部份進一步的多加贅述。

【0044】而各電腦單元(10)進一步依功能性又可區分為一第一資料中心(101)、至少一個第二資料中心(102)及一備份中心(103)；其中該第一資料中心(101)係以一虛擬群集資料主控站(1011)控管，其中各該第二資料中心(102)係以一虛擬群集資料備份站(1021)控管，且其中該備份中心(103)係以虛擬群集資料備份站(1021)控管，而其中該第一資料中心(101)與該第二資料中心(102)共同形成一分散式記憶體檔案系統(40)。

【0045】而請進一步配合參看第二圖所示，其中更進一步是為形成堆疊架構擴展儲存容量架構，其使用網路層介面的存取方式將一台電腦單元(10)中多個區塊記憶體磁碟(11)規劃成一個群集概念的群集記憶體磁碟單元資源池，其運作原理如同電腦的匯流排使用方式。

【0046】當採用多台電腦單元(10)中 64 位元之區塊記憶體磁碟(11)，就如同 64 位元的 CPU 匯流排形式使用，同步使用所有區塊記憶體磁碟(11)取得資料，而當向上擴展區塊記憶體磁碟(11)數量時，如同將 64 位元的 CPU 匯流排提升至 128 位元或 256 位元，其存取的速度以累加的方式提升，故其中記憶體磁碟容量也會因為區塊記憶體磁碟(11)增加，磁碟容量上限也可以增加，並且提升資料存取速度與資料的可靠度，使用者可以依照需求漸進式增加。

【0047】當群集架構(1)產生，每一個群集架構都可以獨立使用，並且可以將每一個群集架構(1)當作分散記憶體式儲存設備單元(DMDCS Distributed Memory Disk Cluster Storage)(1A)，同時使用網路層介面堆疊，可以將每個分散記憶體式儲存設備單元(1A)模擬成大型的記憶體磁碟(chunk memory disk)，採用新的群集資料主控站(1011)與群集資料備份站(1021)來控制資料的處理量分散在所有大型的記憶體磁碟(chunk memory disk)。

【0048】這便如同平行運算使用每一台主機的資源，

將資料分成區塊送到每一台機器運算，最後再組合成一個結果。

【0049】其中一個區塊記憶體磁碟(11)或是其中一個分散記憶體式儲存設備單元(1A)失效，都不會影響整個磁碟的使用或使整個磁碟崩潰。

【0050】此外，當一個一台電腦單元(10)中的記憶體單元故障，該伺服器電腦單元會將損壞的記憶體標示故障，並且不再使用該 DIMM 記憶體中的區塊記憶體 IC，當更換記憶體後，才會恢復該資源之使用。

【0051】而其中虛擬機器之記憶體間係使用存儲區域網路 (Storage Area Network, SAN) 架構運行，採取軟體虛擬出來的網路層介面，將所有的區塊記憶體磁碟連接在一起協同運作；而該網路層介面，係使用 SAN、SAN iSCSI、SAN FC、SAN FCoE、NFS、NAS、JBOD、CIFS、FUSE 介面進行伺服器和磁碟驅動器設備之間的通信，並且開發有 RAMSTORAGE™ API 作為應用程式支援用。

【0052】又其中，該 RAMSTORAGE™ API 係採用 REST、Restful、C++、PHP、Python、Java、Perl、Javascript 與其他程式開發軟體所形成的 RAMSTORAGE™ API，存取分散記憶體式儲存設備單元 (DMDCS Distributed Memory Disk Cluster Storage)(1A) API 功能，其包含容錯、備份、移轉、快速佈署虛擬機器、管理磁碟大小、自動依照實際需求容量自動增加區塊記憶體磁碟

(11)、區塊間資料負載平衡移轉、備份回復的功能、連續資料保護(CDP)、快照、監控資源等。

【0053】此外各該區塊記憶體磁碟(11)係分別電性連接於至少一硬碟儲存裝置(12)，該硬碟儲存裝置(12)係可於一間隔時間將該區塊記憶體磁碟(11)之資料做備份，以防止非預期之故障產生，例如每一個區塊記憶體磁碟(11)每一分鐘內都會將所改變過的資料，將差異的部分實際複製一份到實際物理硬碟儲存裝置(12)備份資料。

【0054】當每一個區塊記憶體磁碟(11)重新開機後，都會將最後備份在實際物理硬碟儲存裝置(12)上資料，完整的還原到區塊記憶體磁碟(11)上，並且通知群集資料主控站(1011)加入群集環境運作。

【0055】所有的電腦單元(10)中所有的區塊記憶體磁碟(11)，都會使用連續資料保護方式(Continuous Data Protector, CDP)，無間斷的連續將資料備份到統一的大型物理硬碟群集磁碟陣列中備份，每當部分伺服器電腦單元或部分區塊記憶體磁碟(11)因為環境或其他因素失效，都可以依照需要的時間找出快照備份或是時間點還原虛擬機器，而其中所述之大型物理硬碟為前述之備份中心(103)，當然群集磁碟陣列也可以用傳統的磁帶方式作第三份完整備份。

【0056】每個區塊記憶體磁碟(11)都可以使用自動部屬的方式規劃所需要的容量與CPU資源，網路層介面

也可以自動設定所需要的網路 IP 與 MAC address，並且一種實際需求設定 AP 在虛擬機器上，並可以自動分派所需設定條件。

【0057】另外，若未使用的 CPU、記憶體、實際物理硬碟的資源，可以使用虛擬機器平台作業系統將所有資源規劃成統一的資源池，每一個所需要電腦資源，可以自動調整移轉到其他資源較充裕的電腦單元(10)。

【0058】多個分散記憶體式儲存設備單元(1A)之連接方式則可以依照實體的網路協定傳送，並且採用 SSL VPN、VPN 或加密演算方式傳送封包，並可以達到跨區域、國家、WAN IP 的方式運作，每當網路連線不通時，每個區域都可以單獨運作。當連線回復時，資料將完整同步到每個分散記憶體式儲存設備單元(1A)中的每個區塊記憶體磁碟(11)。

【0059】而其中於一些硬體規格上之實施態樣選擇中，該 CPU 係為 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 其中之一種處理器。

【0060】該電腦單元(10)之記憶體使用方式是直接由 CPU 的記憶體控制器直接以三通多或多通道以 800 MHz 到 1333 MHz 以上的速度直接存取記憶體資料。

【0061】其中使用的單一記憶體容量為 1MB(megabyte) 至 16 ZB(zettabyte)，所採用的記憶體種類可以為

動態異步記憶體 DRAM 列舉形式為:FPM RAM 、 EDO RAM；或動態同步記憶體 DRAM 列舉形式為:SDRAM、DDR SDRAM、DDR2 SDRAM、DDR3 SDRAM、DDR4 SDRAM、DDR5 SDRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或動態移動平台記憶體列舉形式為：LPDDR、LPDDR2、LPDDR3、LPDDR4 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或動態繪圖處理記憶體列舉形式為 VRAM、WRAM、MDRAM、SGRAM、SDRAM、GDDR、GDDR2、GDDR3、GDDR4、GDDR5、GDDR6、GDDR7 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或動態 Rambus 記憶體列舉形式為：RDRAM、XDR DRAM、XDR2 DRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或靜態隨機存取存儲器（Static Random-Access Memory, SRAM）、或唯讀記憶體（Read-Only Memory, ROM）、或快閃記憶體（Flash Memory）列舉形式為:NOR Flash、NAND Flash、3D NAND Flash、V-Flash、SLC、MLC、eMMC、TLC 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號。

【0062】再者，硬碟儲存裝置(12)可使用包含傳統磁頭與磁片的硬碟、固態硬碟、網路硬碟、SAS 介面硬碟、SATA 介面硬碟、mSATA 介面硬碟、mPCIe 介面硬碟、

FC 介面硬碟、SCSI 介面硬碟、ATA 介面硬碟、NAND Flash 外接卡、FCoE 介面硬碟，以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態硬碟。

【0063】又其中，網路卡可以選擇乙太網路、快速乙太網路、千兆乙太網路、光纖、令牌環網、InfiniBand、FCoE(Fibre Channel over Ethernet)或無線網路，而其網路速度可依照網路協定的不同採用 2Mbit/s、10 Mbit/s、11Mbit/s、40Mbit/s、54Mbit/s、80Mbit/s、100 Mbit/s、150Mbit/s、300Mbit/s、433Mbit/s、1000 Mbit/s、1Gbit/s、8Gbit/s、10Gbit/s、16Gbit/s、32Gbit/s、40Gbit/s、56Gbit/s、100Gbit/s、160Gbit/s、1000Gbit/s 以上的網路速度與任何新型網路通訊協定的網路卡。

【0064】其中，主機板可相容 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 任何一種處理器相容的任何架構與品牌電腦製造商所產生的包括 BeagleBoneBlack、Raspberry Pi 架構主機板。

【0065】另外值得注意的是，作業系統檔案格式，可以是 AdvFS、Be 檔案系統 (BFS)、Btrfs、Coda、CrossDOS、光碟存檔系統 (DFS)、Episode、EFS、exFAT、ext、FAT、全域檔案系統 (GFS)、分層檔案系統 (HFS)、HFS Plus、高效能檔案系統、IBM 通用並列檔案系統、JFS、Macintosh 檔案系統、MINIX、NetWare 檔案系

統、NILFS、Novell 儲存服務、NTFS、QFS、QNX4FS、ReiserFS (Reiser4)、SpadFS、UBIFS、Unix 檔案系統、Veritas 檔案系統 (VxFS)、VFAT、任意位置寫入檔案布局 (WAFL)、XFS、Xsan、ZFS、CHFS、FFS2、F2FS、JFFS、JFFS2、LogFS、NVFS、YAFFS、UBIFS、DCE/DFS、MFS、CXFS、GFS2、Google 檔案系統、OCFS、OCFS2、QFS、Xsan、AFS、OpenAFS、AFP、MS-DFS、GPFS、Lustre、NCP、NFS、POHMEFS、Hadoop、HAMMER、SMB (CIFS)、cramfs、FUSE、SquashFS、UMSDOS、UnionFS、configfs、devfs、procfs、specfs、sysfs、tmpfs、WinFS、EncFS、EFS、ZFS、RAW、ASM、LVM、SFS、MPFS、MGFS 其中之一。

【0066】而實體的網路協定傳送，以選擇乙太網路、快速乙太網路、千兆乙太網路、光纖、令牌環網、SS7、GSM、GPRS、EDGE、HSPA、HSPA+、CDMA、WCDMA、TD-WCDMA、LTE、GSM、cdmaOne、CDMA2000、UMTS、WCDMA、TD-SCDMA、WiMAX、3G 播報網、CDMA2000 1X、Wi-Fi、SuperWiFi、Wi-Fi GO、以及向上相容 IEEE 所制定出新的網路傳輸協定。

【0067】透過本發明所提出之系統架構方法之下，我們採用分散式記憶體儲存系統，可以滿足四種需求的擴充，即網路頻寬傳輸的擴充、磁碟容量的擴充、IOPS 速度的擴充、記憶體 I/O 傳輸速度增加的擴充。同時可以做到滿足跨區域、資料中心、WAN 的使用，讓用戶需求可以在當地的群集記憶體磁碟 (Memory

Disk Cluster)集中提供服務、也可以漸進式的擴充群集記憶體磁碟容量，進一步提供跨區域或跨國的資料服務。

【0068】在隨著儲存器的增加，每增加一台伺服器就累加一台機器的網路頻寬與磁碟容量，當形成一個資源池(Resource Pool)時，這分散記憶體式儲存設備單元(1A)就像一顆實體硬碟，其中一台實體主機故障時不會影響整體運作，複本中區塊記憶體磁碟(11)會重新複製資料到新的區塊記憶體磁碟(11)，維持一定資料備份基礎，同時也採取連續資料複製(Continuous Data Protector, CDP)為企業資料的備份和回復提供創新技術，以破除磁帶備份經常失敗，一天只能備份一次問題。

【0069】也讓複本產生資料可以由不同區塊記憶體磁碟(11)送出達到多對一傳輸資料，當用戶數量增加，只要增加區塊記憶體磁碟(11)數量，可以達到多對多的傳輸，讓 RAID 的多數硬碟崩潰造成整體資料遺失的缺點，儲存器網路介面數量與速度的限制造成用戶過多的資料擁擠造成的延遲，LUN 擴充數量的問題，資料中心無法跨區使用的問題，因為採用記憶體式磁碟方式，每個檔案或是每個虛擬機都可以用檔案形式存在於記憶體，直接使用記憶體匯流排最高的 I/O 高速度，資料在 CPU 與記憶體中搬遷移動，I/O 次數、距離、速度都是最高的，綜觀上述優勢，是為習知技術所不能及者，故可見其增益性所在。

【0070】 以上所述之實施例僅係為說明本發明之技術思想及特點，其目的在使熟習此項技藝之人士能夠瞭解本發明之內容並據以實施，當不能以之限定本發明之專利範圍，即大凡依本發明所揭示之精神所作之均等變化或修飾，仍應涵蓋在本發明之專利範圍內。

【0071】 綜觀上述，可見本發明在突破先前之技術下，確實已達到所欲增進之功效，且也非熟悉該項技藝者所易於思及，其所具之進步性、實用性，顯已符合專利之申請要件，爰依法提出專利申請，懇請 貴局核准本件發明專利申請案，以勵發明，至感德便。

【符號說明】

【0072】

群集架構(1)

分散記憶體式儲存設備單元(1A)

電腦單元(10)

第一資料中心(101)

群集資料主控站(1011)

第二資料中心(102)

群集資料備份站(1021)

備份中心(103)

區塊記憶體磁碟(11)

硬碟儲存裝置(12)

網路卡(13)

連接埠群組現用鏈(20)

分散式記憶體檔案系統(40)

【生物材料寄存】

國內寄存資訊【請依寄存機構、日期、號碼順序註記】

無

國外寄存資訊【請依寄存國家、機構、日期、號碼順序註記】

無

【序列表】 (請換頁單獨記載)

無

申請專利範圍

1. 一種分散式記憶體磁碟群集儲存系統運作方法，其為：

首先做分散記憶體式儲存設備之建置，其係包含將複數台的電腦單元組成群集架構而形成群集記憶體磁碟；

而後在電腦單元安裝系統虛擬機器平台作業系統，藉此安裝令電腦單元形成多個虛擬機器，而該電腦單元設定利用作業系統或程式軟體來設定記憶體容量占用之方式，將該記憶體規劃成儲存裝置，做成複數個區塊記憶體磁碟；進而將每一個檔案拆成一到多份的資料，保持一到多份的複本複製平均分散在所有區塊記憶體磁碟上，並使用多通道的記憶體匯流排平行存取記憶體模組，而將記憶體模組之容量規劃成磁碟使用，而其中該記憶體模組之存取支援所有作業系統檔案格式，並且採用分散式存放架構，將資料可以複製複本一份以上；

當虛擬機器之作業系統在記憶體模組中直接由CPU存取需要的檔案，處理資料後存回記憶體模組中，虛擬機器做計算使用的記憶體也在記憶體模組中，計算完依然直接存回在所存放虛擬機器作業系統的記憶體模組位置，讓這個儲存系統內資料可以在電腦單元中以平行運算的方式在區塊記憶體磁碟處理資料；其他沒有在電腦單元的資料係透過網路卡連接一連接埠群組現用鏈以存取

其他電腦單元的區塊記憶體磁碟；

而各電腦單元進一步依功能性區分為一第一資料中心、至少一個第二資料中心及一備份中心；其中該第一資料備份中心係以一虛擬群集資料主控站控管，其中各該第二資料中心係以一虛擬群集資料備份站控管，且其中該備份中心係以虛擬群集資料備份站控管，而其中該第一資料中心與該第二資料中心共同形成一分散式記憶體檔案系統；此外，更進一步是為形成堆疊架構擴展儲存容量架構，其使用網路層介面的存取方式將一台電腦單元中多個區塊記憶體磁碟規劃成一個群集概念的群集記憶體磁碟單元資源池，同步使用所有區塊記憶體磁碟取得資料；

當群集架構產生後，每一個群集架構都可以獨立使用，並且可以將每一個群集架構當作分散記憶體式儲存設備單元，同時使用網路層介面堆疊，可以將每個分散記憶體式儲存設備單元模擬成大型的記憶體磁碟，並採用新的群集資料主控站與群集資料備份站來控制資料的處理量分散在所有大型的記憶體磁碟。

2. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中各該區塊記憶體磁碟係分別電性連接於至少一硬碟儲存裝置，該硬碟儲存裝置係可於一間隔時間將該區塊記憶體磁碟之資料做備份。

3. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中所有的電腦單元中的區塊記憶體磁碟，都會使用連續資料保護方式，無間斷的連續將資料備份到統一的大型物理硬碟群集磁碟陣列中備份。
4. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該電腦單元係俱備了 CPU、記憶體、硬碟、網路卡、主機板、I/O 介面卡、連接線、外殼之計算機設備。
5. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中每份複本都可以採用 1-4096 位元 AES 與 RSA 混用加密資料存放在記憶體上，每當要存取資料時，資料在記憶體與 CPU 之間流動，虛擬機器係以檔案形式存在於記憶體模組中，虛擬記憶體所規劃使用的記憶體容量也在同一區塊。
6. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中各該區塊記憶體磁碟透過一監控單元來做其運作情況之監控，該偵測單元利用 Splunk 或者是其他搜索引擎之軟體程式來監控之，當檢測到問題時更可以配置重新啟動應用程式服務，達到一回復之功能。
7. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中虛擬機器平台作業系統採用

VMware vSphere ESXi 4.1 之後的版本、Microsoft Server 2012 R2 Hyper-V 之後的版本、Citrix XenServer Oracle VM、Oeacle VM、Red Hat KVM 或 Linux-Base KVM。

8. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中虛擬機器作業系統採用 Linux、FreeBSD、OpenSolaris 或 MacOS X。
9. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中虛擬機器之記憶體間係使用存儲區域網路架構運行，採取軟體虛擬出來的網路層介面，將所有的區塊記憶體磁碟連接在一起協同運作。
10. 如申請專利範圍第 9 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該網路層介面，係使用 SAN、SAN iSCSI、SAN FC、SAN FCoE、NFS、NAS、JBOD、CIFS 或 FUSE 介面進行伺服器 and 磁碟驅動器設備之間的通信，並且開發有 RAMSTORAGE™ API 作為應用程式支援用。
11. 如申請專利範圍第 10 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該 RAMSTORAGE™ API 係採用 REST、Restful、C++、PHP、Python、Java、Perl 或 Javascript 程式開發軟體所形成的 RAMSTORAGE™ API，存取分散記憶體式儲存設備單元 API 功能，其包含容錯、備份、移轉、快速佈署虛擬機器、管理磁碟大小、自動依照實際需求容量自動增加區塊記憶體磁碟、區塊間資料負載平衡移轉、備份回復的功能、連續資料保護、快照、監控資源等。
12. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群

集儲存系統運作方法，其中未使用的 CPU、記憶體、實際物理硬碟的資源，係使用虛擬機器平台作業系統將所有資源規劃成統一的資源池，令每一個所需要電腦資源，可以調整移轉到其他資源較充裕的電腦單元。

13. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中多個分散記憶體式儲存設備單元之連接方式依照實體的網路協定傳送，並且採用 SSL VPN、VPN 或加密演算方式傳送封包，每當網路連線不通時，每個區域都可以單獨運作，當連線回復時，資料將完整同步到每個分散記憶體式儲存設備單元中的每個區塊記憶體磁碟。
14. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該 CPU 係為 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 其中之一種處理器。
15. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該電腦單元之記憶體使用方式是直接由 CPU 的記憶體控制器直接以三通多或多通道以 800 MHz 到 1333 MHz 以上的速度直接存取記憶體資料。
16. 如申請專利範圍第 15 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中記憶體容量為 1MB 至 16 ZB。
17. 如申請專利範圍第 15 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中所採用的記憶體種類為動態異步記憶體 DRAM、動態同步記憶體 DRAM、動態移動平台記

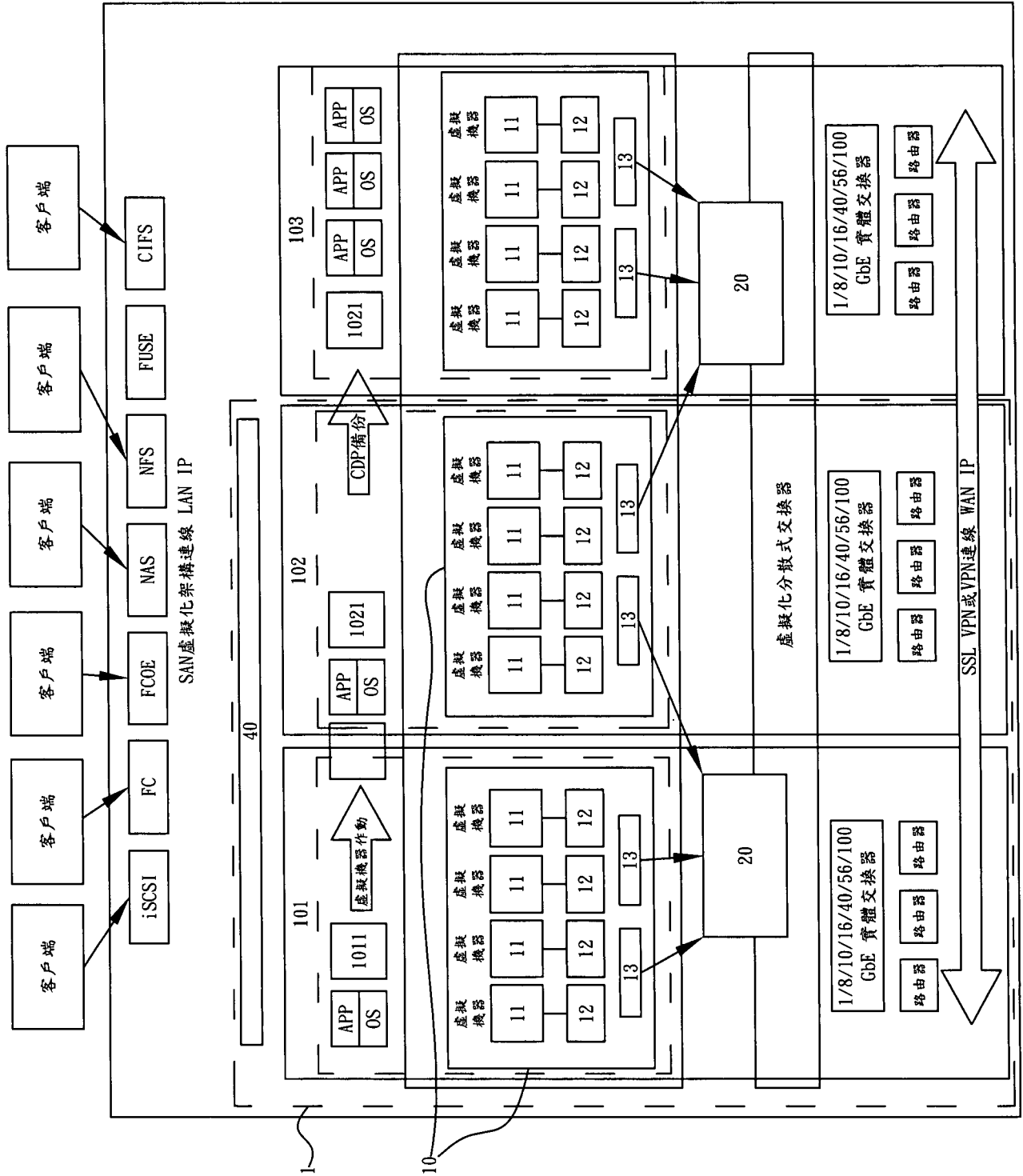
憶體、動態繪圖處理記憶體、動態 Rambus 記憶體、靜態隨機存取存儲器、唯讀記憶體或快閃記憶體。

18. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態異步記憶體 DRAM 為 FPM RAM 或 EDO RAM。
19. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態同步記憶體 DRA 為 SDRAM、DDR SDRAM、DDR2 SDRAM、DDR3 SDRAM、DDR4 SDRAM 或 DDR5 SDRAM。
20. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態移動平台記憶體為 LPDDR、LPDDR2、LPDDR3 或 LPDDR4。
21. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態繪圖處理記憶體為 VRAM、WRAM、MDRAM、SGRAM、SDRAM、GDDR、GDDR2、GDDR3、GDDR4、GDDR5、GDDR6 或 GDDR7。
22. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態 Rambus 記憶體為 RDRAM、XDR DRAM 或 XDR2 DRAM。
23. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中快閃記憶體為 NOR Flash、NAND Flash、3D NAND Flash、V-Flash、SLC、MLC、eMMC 或 TLC。

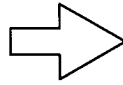
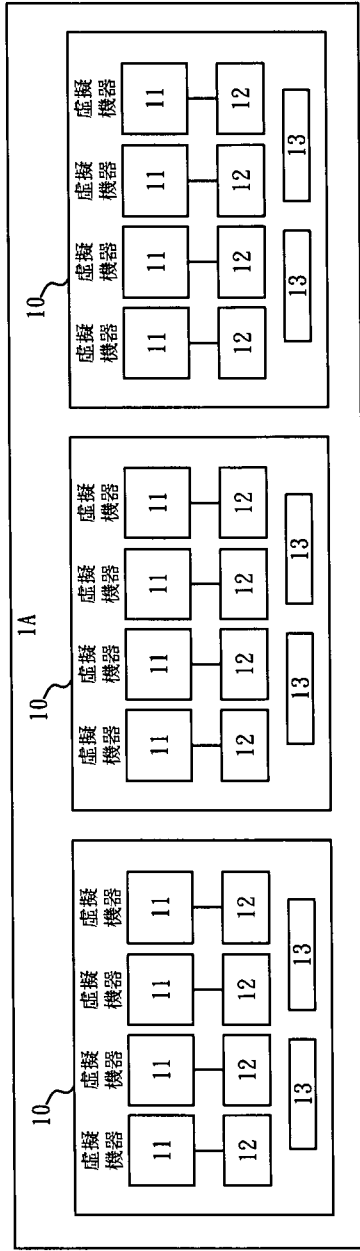
24. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中硬碟儲存裝置為傳統磁頭、磁片的硬碟、固態硬碟、網路硬碟、SAS 介面硬碟、SATA 介面硬碟、mSATA 介面硬碟、mPCIE 介面硬碟、FC 介面硬碟、SCSI 介面硬碟、ATA 介面硬碟、NAND Flash 外接卡或 FCoE 介面硬碟。
25. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中網路卡為乙太網路、快速乙太網路、千兆乙太網路、光纖、令牌環網、InfiniBand、FCoE 或無線網路。
26. 如申請專利範圍第 25 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中網路速度採用 2Mbit/s、10 Mbit/s、11Mbit/s、40Mbit/s、54Mbit/s、80Mbit/s、100 Mbit/s、150Mbit/s、300Mbit/s、433Mbit/s、1000 Mbit/s、1Gbit/s、8Gbit/s、10Gbit/s、16Gbit/s、32Gbit/s、40Gbit/s、56Gbit/s、100Gbit/s、160Gbit/s 或 1000Gbit/s 以上的網路速度。
27. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中，主機板相容 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 中任何一種處理器。
28. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中作業系統檔案格式是 AdvFS、BFS、Btrfs、Coda、CrossDOS、DFS、Episode、EFS、exFAT、ext、FAT、GFS、HFS、HFS Plus、高效能檔案系統、IBM 通用

並列檔案系統、JFS、Macintosh 檔案系統、MINIX、NetWare 檔案系統、NILFS、Novell 儲存服務、NTFS、QFS、QNX4FS、ReiserFS、SpadFS、UBIFS、Unix 檔案系統、Veritas VxFS、VFAT、WAFL、XFS、Xsan、ZFS、CHFS、FFS2、F2FS、JFFS、JFFS2、LogFS、NVFS、YAFFS、UBIFS、DCE/DFS、MFS、CXFS、GFS2、Google 檔案系統、OCFS、OCFS2、QFS、Xsan、AFS、OpenAFS、AFP、MS-DFS、GPFS、Lustre、NCP、NFS、POHMELFS、Hadoop、HAMMER、SMB (CIFS)、cramfs、FUSE、SquashFS、UMSDOS、UnionFS、configfs、devfs、procfs、specfs、sysfs、tmpfs、WinFS、EncFS、EFS、ZFS、RAW、ASM、LVM、SFS、MPFS 或 MGFS。

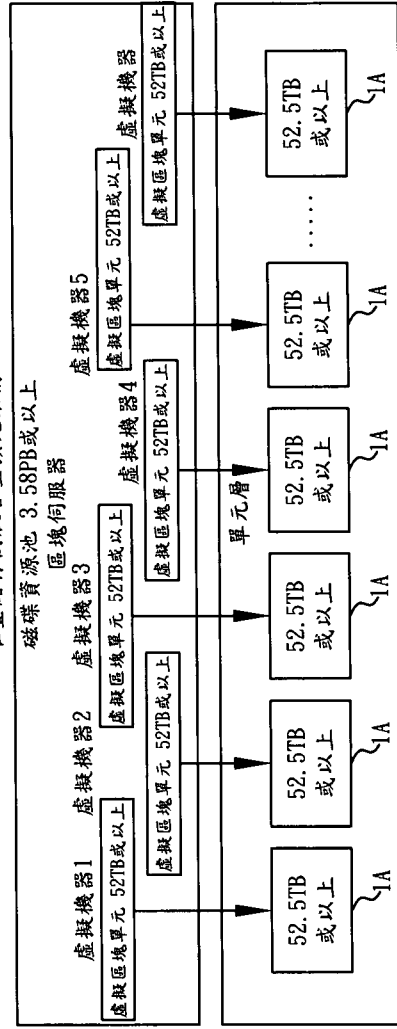
圖式



第一圖



堆疊儲存樹狀容量擴充架構



第二圖

與 RSA 混用加密資料存放在記憶體上，每當要存取資料時，資料在記憶體與 CPU 之間流動，虛擬機器係以檔案形式存在於記憶體模組中，虛擬記憶體所規劃使用的記憶體容量也在同一區塊。

【0013】其中，各該區塊記憶體磁碟透過一監控單元來做其運作情況之監控，該偵測單元利用 Splunk 或者是其他搜索引擎之軟體程式來監控之，當檢測到問題時更可以配置重新啟動應用程式服務，達到一回復之功能。

【0014】其中，虛擬機器平台作業系統採用 VMware vSphere ESXi 4.1 之後的版本、Microsoft Server 2012 R2 Hyper-V 之後的版本、Citrix XenServer Oracle VM、Oeacle VM、Red Hat KVM、Red Hat Control groups(cgroups)、Red Hat Linux Containers(LXC)、KVM、Eucalyptus、OpenStack、User Mode Linux、LXC、OpenVZ、OpenNebula、Enomaly's Elastic Computing、OpenFlow 或 Linux-Base KVM。其中虛擬機器作業系統採用 Linux、FreeBSD、OpenSolaris 或 MacOS X。

【0015】其中，虛擬機器之記憶體間係使用存儲區域網路架構運行，採取軟體虛擬出來的網路層介面，將所有的區塊記憶體磁碟連接在一起協同運作。

【0016】其中，該網路層介面，係使用 SAN、SAN iSCSI、SAN FC、SAN FCoE、NFS、NAS、JBOD、CIFS 或 FUSE 介面進行伺服器 and 磁碟驅動器設備之間的通信，並且

開發有 RAMSTORAGE™ API 作為應用程式支援用。其中該 RAMSTORAGE™ API 係採用 REST、Restful、C++、PHP、Python、Java、Perl 或 Javascript 程式開發軟體所形成的 RAMSTORAGE™ API，存取分散記憶體式儲存設備單元 API 功能，其包含容錯、備份、移轉、快速佈署虛擬機器、管理磁碟大小、自動依照實際需求容量自動增加區塊記憶體磁碟、區塊間資料負載平衡移轉、備份回復的功能、連續資料保護、快照、監控資源等。

【0017】其中，未使用的 CPU、記憶體、實際物理硬碟的資源，係使用虛擬機器平台作業系統將所有資源規劃成統一的資源池，令每一個所需要電腦資源，可以調整移轉到其他資源較充裕的電腦單元。

【0018】其中，多個分散記憶體式儲存設備單元之連接方式依照實體的網路協定傳送，並且採用 SSL VPN、VPN 或加密演算方式傳送封包，每當網路連線不通時，每個區域都可以單獨運作，當連線回復時，資料將完整同步到每個分散記憶體式儲存設備單元中的每個區塊記憶體磁碟。

【0019】其中，該 CPU 係為 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 其中之一種處理器。

【0020】其中，該電腦單元之記憶體使用方式是直接由 CPU 的記憶體控制器直接以三通多或多通道以 800

MHz 到 1333 MHz 以上的速度直接存取記憶體資料。

【0021】其中，記憶體容量為 1MB 至 16 ZB。其中所採用的記憶體種類為動態異步記憶體 DRAM、動態同步記憶體 DRAM、動態移動平台記憶體、動態繪圖處理記憶體、動態 Rambus 記憶體、靜態隨機存取存儲器、唯讀記憶體、磁阻式隨機存取記憶體或快閃記憶體。

【0022】其中，動態異步記憶體 DRAM 為 FPM RAM 或 EDO RAM。其中動態同步記憶體 DRA 為 SDRAM、DDR SDRAM、DDR2 SDRAM、DDR3 SDRAM、DDR4 SDRAM 或 DDR5 SDRAM。其中磁阻式隨機存取記憶體為 MRAM，或鐵電式記憶體，為 FeRAM；或相變化記憶體為 PCRAM；或可變電阻式記憶體為 ReRAM。其中動態移動平台記憶體為 LPDDR、LPDDR2、LPDDR3 或 LPDDR4。其中動態繪圖處理記憶體為 VRAM、WRAM、MDRAM、SGRAM、SDRAM、GDDR、GDDR2、GDDR3、GDDR4、GDDR5、GDDR6 或 GDDR7。其中動態 Rambus 記憶體為 RDRAM、XDR DRAM 或 XDR2 DRAM。其中快閃記憶體為 NOR Flash、NAND Flash、3D NAND Flash、V-Flash、SLC、MLC、eMMC 或 TLC。

【0023】其中，硬碟儲存裝置為傳統磁頭、磁片的硬碟、固態硬碟、網路硬碟、SAS 介面硬碟、SATA 介面硬碟、mSATA 介面硬碟、mPCIe 介面硬碟、FC 介面硬碟、SCSI 介面硬碟、ATA 介面硬碟、NAND Flash 外接卡或 FCoE 介面硬碟。

【0024】其中，網路卡為乙太網路、快速乙太網路、

千兆乙太網路、光纖、令牌環網、InfiniBand、FCoE 或無線網路。

【0025】其中，網路速度採用 2Mbit/s、10 Mbit/s、11Mbit/s、40Mbit/s、54Mbit/s、80Mbit/s、100 Mbit/s、150Mbit/s、300Mbit/s、433Mbit/s、1000 Mbit/s、1Gbit/s、8Gbit/s、10Gbit/s、16Gbit/s、32Gbit/s、40Gbit/s、56Gbit/s、100Gbit/s、160Gbit/s 或 1000Gbit/s 以上的網路速度。

【0026】其中，主機板相容 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 中任何一種處理器。

【0027】其中，作業系統檔案格式是 VMFS3、VMFS5 與格式不同所產生新的向上相容型態型號、VHD 與格式不同所產生新的向上相容型態型號、VHDX 與格式不同所產生新的向上相容型態型號、VMDK 與格式不同所產生新的向上相容型態型號、HDFS 與格式不同所產生新的向上相容型態型號、Isilon OneFS 與格式不同所產生新的向上相容型態型號、記憶體式的分頁檔案所產生的任何格式與格式不同所產生新的向上相容型態型號、VEs 與格式不同所產生新的向上相容型態型號、VPSs 與格式不同所產生新的向上相容型態型號、Ceph、GlusterFS、SphereFS、Taobao File System、ZFS、SDFS、MooseFS、AdvFS、BFS、Btrfs、Coda、CrossDOS、DFS、Episode、EFS、exFAT、ext、FAT、GFS、HFS、HFS Plus、高效能檔案系統、IBM 通用並列檔案系統、JFS、Macintosh 檔案系統、MINIX、NetWare 檔案系統、NILFS、Novell 儲存服務、NTFS、QFS、QNX4FS、

ReiserFS、SpadFS、UBIFS、Unix 檔案系統、Veritas VxFS、VFAT、WAFL、XFS、Xsan、ZFS、CHFS、FFS2、F2FS、JFFS、JFFS2、LogFS、NVFS、YAFFS、UBIFS、DCE/DFS、MFS、CXFS、GFS2、Google 檔案系統、OCFS、OCFS2、QFS、Xsan、AFS、OpenAFS、AFP、MS-DFS、GPFS、Lustre、NCP、NFS、POHMELFS、Hadoop、HAMMER、SMB (CIFS)、cramfs、FUSE、SquashFS、UMSDOS、UnionFS、configfs、devfs、procfs、specfs、sysfs、tmpfs、WinFS、EncFS、EFS、ZFS、RAW、ASM、LVM、SFS、MPFS 或 MGFS。

【0028】本發明之分散式記憶體磁碟群集儲存系統運作方法於設計上係透過本發明所提出之系統架構方法之下，我們採用分散式記憶體儲存系統，可以滿足四種需求的擴充，即網路頻寬傳輸的擴充、磁碟容量的擴充、IOPS 速度的擴充、記憶體 I/O 傳輸速度增加的擴充。同時可以做到滿足跨區域、資料中心、WAN 的使用，讓用戶需求可以在當地的群集記憶體磁碟 (Memory Disk Cluster) 集中提供服務、也可以漸進式的擴充群集記憶體磁碟容量，進一步提供跨區域或跨國的資料服務。

【0029】在隨著儲存器的增加，每增加一台伺服器就累加一台機器的網路頻寬與磁碟容量，當形成一個資源池 (Resource Pool) 時，這分散記憶體式儲存設備單元就像一顆實體硬碟，其中一台實體主機故障時不會影響整體運作，複本中區塊記憶體磁碟會重新複製資

料到新的區塊記憶體磁碟，維持一定資料備份基礎，同時也採取連續資料複製(Continuous Data Protector, CDP)為企業資料的備份和回復提供創新技術，以破除磁帶備份經常失敗，一天只能備份一次問題。

【0030】也讓複本產生資料可以由不同區塊記憶體磁碟送出達到多對一傳輸資料，當用戶數量增加，只要增加區塊記憶體磁碟數量，可以達到多對多的傳輸，讓 RAID 的多數硬碟崩潰造成整體資料遺失的缺點，儲存器網路介面數量與速度的限制造成用戶過多的資料擁擠造成的延遲，LUN 擴充數量的問題，資料中心無法跨區使用的問題，因為採用記憶體式磁碟方式，每個檔案或是每個虛擬機都可以用檔案形式存在於記憶體，直接使用記憶體匯流排最高的 I/O 高速度，資料在 CPU 與記憶體中搬遷移動，I/O 次數、距離、速度都是最高的，綜觀上述優勢，是為習知技術所不能及者，故可見其增益性所在。

【0031】而為了讓上述目的、技術特徵以及實際實施後之增益性更為明顯易懂，於下文中將係以較佳之實施範例輔佐對應相關之圖式來進行更詳細之說明。

【圖式簡單說明】

【0032】

第一圖為本發明之分散式記憶體磁碟群集儲存系統運作方法之實施示意圖。

第二圖為本發明之分散式記憶體磁碟群集儲存系統運作方法之實施示意圖。

【實施方式】

【0033】為利貴審查員瞭解本發明之發明特徵、內容與優點及其所能達成之功效，茲將本發明配合附圖，並以實施例之表達形式詳細說明如下，而其中所使用之圖式，其主旨僅為示意及輔助說明書之用，未必為本發明實施後之真實比例與精準配置，故不應就所附之圖式的比例與配置關係解讀、侷限本發明於實際實施上的權利範圍。

【0034】本發明之優點、特徵以及達到之技術方法將參照例示性實施例及所附圖式進行更詳細地描述而更容易理解，且本發明或可以不同形式來實現，故不應被理解僅限於此處所陳述的實施例，相反地，對所屬技術領域具有通常知識者而言，所提供的實施例將使本揭露更加透徹與全面且完整地傳達本發明的範疇，且本發明將僅為所附加的申請專利範圍所定義。

【0035】而除非另外定義，所有使用於後文的術語（包含科技及科學術語）與專有名詞，於實質上係與本發明所屬該領域的技術人士一般所理解之意思相同，而例如於一般所使用的字典所定義的那些術語應被理解為具有與相關領域的內容一致的意思，且除非明顯地定義於後文，將不以過度理想化或過度正式的意思理解，合先敘明。

【0036】請配合參看第一圖所示，本發明提出一種分散式記憶體磁碟群集儲存系統運作方法，其中該分散

式記憶體磁碟群集儲存系統運作方法於一較佳之實施方式可為：

【0037】首先做分散記憶體式儲存設備之建置，其係包含將複數台的電腦單元(10)組成群集架構(1)而形成群集記憶體磁碟；而其中該電腦單元(10)係俱備了CPU、記憶體、硬碟、網路卡、主機板、I/O 介面卡、連接線、外殼之計算機設備。

【0038】主要在電腦單元(10)安裝系統虛擬機器平台作業系統，藉此安裝令電腦單元(10)形成多個虛擬機器，而該電腦單元(10)設定所須要使用機器記憶體資源容量，係利用作業系統來設定記憶體容量占用之方式，或者是利用程式軟體將該記憶體規劃成硬碟裝置，做成一個區塊記憶體磁碟(chunk memory disk)(11)，這便如同硬碟的磁軌一般。

【0039】進而得以將一個檔案拆成一到多份的資料，檔案大小或可為 64MB 以上，保持一到多份的複本複製平均分散在所有區塊記憶體磁碟(11)上，故資料實際存放在記憶體模組裡，並使用多通道的記憶體匯流排平行存取記憶體模組，而將記憶體模組之容量規劃成磁碟使用，而其中該記憶體模組之存取支援所有作業系統檔案格式，並且採用分散式存放架構，將資料可以複製複本一份以上，藉此方式，即使機器損毀與資料中心損毀，資料中心依然可以運作。

【0040】每份複本資料都可以採用 1-4096 位元 AES 與

RSA 混用加密資料存放在記憶體上，每當要存取資料時，資料在記憶體與 CPU 之間流動，所耗費的 I/O 存取次數與距離是最少的，虛擬機器係以檔案形式存在於記憶體模組中，虛擬記憶體所規劃使用的記憶體容量也在同一區塊。

【0041】當虛擬機器之作業系統在記憶體模組中直接由 CPU 存取需要的檔案，處理資料後存回記憶體模組中，虛擬機器做計算使用的記憶體也在記憶體模組中，計算完依然直接存回在所存放虛擬機器作業系統的記憶體模組位置，相對依照存取路徑的減少與使用最快的 I/O 速度，讓這個儲存系統大部分資料可以在電腦單元(10)中以平行運算的方式在區塊記憶體磁碟(11)處理資料，其他沒有在電腦單元(10)的資料才透過網路卡(13)連接一連接埠群組現用鏈(20)以存取其他電腦單元(10)的區塊記憶體磁碟(11)。

【0042】其中，虛擬機器平台作業系統，可以採用 VMware vSphere ESXi 4.1 之後的版本、Microsoft Server 2012 R2 Hyper-V 之後的版本、Citrix XenServer Oracle VM、Oeacle VM、Red Hat KVM、Red Hat Control groups(cgroups)、Red Hat Linux Containers(LXC)、KVM、Eucalyptus、OpenStack、User Mode Linux、LXC、OpenVZ、OpenNebula、Enomaly's Elastic Computing、OpenFlow、Linux-Base KVM；而虛擬機器作業系統，可以採用 Linux (Linux 2.6.14 and up have FUSE support included in the official kernel)、

FreeBSD、OpenSolaris、MacOS X。

【0043】再者，各該區塊記憶體磁碟(11)更可能透過一監控單元來做其運作情況之監控，該偵測單元或可利用 Splunk 或者是其他搜索引擎之軟體程式來監控之，當檢測到問題時更可以配置重新啟動應用程式服務，達到一回復之功能，而這類型之程式為習知技術，故不在此部份進一步的多加贅述。

【0044】而各電腦單元(10)進一步依功能性又可區分為一第一資料中心(101)、至少一個第二資料中心(102)及一備份中心(103)；其中該第一資料中心(101)係以一虛擬群集資料主控站(1011)控管，其中各該第二資料中心(102)係以一虛擬群集資料備份站(1021)控管，且其中該備份中心(103)係以虛擬群集資料備份站(1021)控管，而其中該第一資料中心(101)與該第二資料中心(102)共同形成一分散式記憶體檔案系統(40)。

【0045】而請進一步配合參看第二圖所示，其中更進一步是為形成堆疊架構擴展儲存容量架構，其使用網路層介面的存取方式將一台電腦單元(10)中多個區塊記憶體磁碟(11)規劃成一個群集概念的群集記憶體磁碟單元資源池，其運作原理如同電腦的匯流排使用方式。

【0046】當採用多台電腦單元(10)中 64 位元之區塊記憶體磁碟(11)，就如同 64 位元的 CPU 匯流排形式使

用，同步使用所有區塊記憶體磁碟(11)取得資料，而當向上擴展區塊記憶體磁碟(11)數量時，如同將 64 位元的 CPU 匯流排提升至 128 位元或 256 位元，其存取的速度以累加的方式提升，故其中記憶體磁碟容量也會因為區塊記憶體磁碟(11)增加，磁碟容量上限也可以增加，並且提升資料存取速度與資料的可靠度，使用者可以依照需求漸進式增加。

【0047】當群集架構(1)產生，每一個群集架構都可以獨立使用，並且可以將每一個群集架構(1)當作分散記憶體式儲存設備單元(DMDCS Distributed Memory Disk Cluster Storage)(1A)，同時使用網路層介面堆疊，可以將每個分散記憶體式儲存設備單元(1A)模擬成大型的記憶體磁碟(chunk memory disk)，採用新的群集資料主控站(1011)與群集資料備份站(1021)來控制資料的處理量分散在所有大型的記憶體磁碟(chunk memory disk)。

【0048】這便如同平行運算使用每一台主機的資源，將資料分成區塊送到每一台機器運算，最後再組合成一個結果。

【0049】其中一個區塊記憶體磁碟(11)或是其中一個分散記憶體式儲存設備單元(1A)失效，都不會影響整個磁碟的使用或使整個磁碟崩潰。

【0050】此外，當一個一台電腦單元(10)中的記憶體單元故障，該伺服器電腦單元會將損壞的記憶體標示故

障，並且不再使用該 DIMM 記憶體中的區塊記憶體 IC，當更換記憶體後，才會恢復該資源之使用。

【0051】而其中虛擬機器之記憶體間係使用存儲區域網路 (Storage Area Network, SAN) 架構運行，採取軟體虛擬出來的網路層介面，將所有的區塊記憶體磁碟連接在一起協同運作；而該網路層介面，係使用 SAN、SAN iSCSI、SAN FC、SAN FCoE、NFS、NAS、JBOD、CIFS、FUSE 介面進行伺服器 and 磁碟驅動器設備之間的通信，並且開發有 RAMSTORAGE™ API 作為應用程式支援用。

【0052】又其中，該 RAMSTORAGE™ API 係採用 REST、Restful、C++、PHP、Python、Java、Perl、Javascript 與其他程式開發軟體所形成的 RAMSTORAGE™ API，存取分散記憶體式儲存設備單元 (DMDCS Distributed Memory Disk Cluster Storage)(1A) API 功能，其包含容錯、備份、移轉、快速佈署虛擬機器、管理磁碟大小、自動依照實際需求容量自動增加區塊記憶體磁碟 (11)、區塊間資料負載平衡移轉、備份回復的功能、連續資料保護 (CDP)、快照、監控資源等。

【0053】此外各該區塊記憶體磁碟 (11) 係分別電性連接於至少一硬碟儲存裝置 (12)，該硬碟儲存裝置 (12) 係可於一間隔時間將該區塊記憶體磁碟 (11) 之資料做備份，以防止非預期之故障產生，例如每一個區塊記憶體磁碟 (11) 每一分鐘內都會將所改變過的資料，將差異的部分實際複製一份到實際物理硬碟儲存裝

置(12)備份資料。

【0054】當每一個區塊記憶體磁碟(11)重新開機後，都會將最後備份在實際物理硬碟儲存裝置(12)上資料，完整的還原到區塊記憶體磁碟(11)上，並且通知群集資料主控站(1011)加入群集環境運作。

【0055】所有的電腦單元(10)中所有的區塊記憶體磁碟(11)，都會使用連續資料保護方式(Continuous Data Protector, CDP)，無間斷的連續將資料備份到統一的大型物理硬碟群集磁碟陣列中備份，每當部分伺服器電腦單元或部分區塊記憶體磁碟(11)因為環境或其他因素失效，都可以依照需要的時間找出快照備份或是時間點還原虛擬機器，而其中所述之大型物理硬碟為前述之備份中心(103)，當然群集磁碟陣列也可以用傳統的磁帶方式作第三份完整備份。

【0056】每個區塊記憶體磁碟(11)都可以使用自動部屬的方式規劃所需要的容量與CPU資源，網路層介面也可以自動設定所需要的網路IP與MAC address，並且一種實際需求設定AP在虛擬機器上，並可以自動分派所需設定條件。

【0057】另外，若未使用的CPU、記憶體、實際物理硬碟的資源，可以使用虛擬機器平台作業系統將所有資源規劃成統一的資源池，每一個所需要電腦資源，可以自動調整移轉到其他資源較充裕的電腦單元(10)。

【0058】 多個分散記憶體式儲存設備單元(1A)之連接方式則可以依照實體的網路協定傳送，並且採用 SSL VPN、VPN 或加密演算方式傳送封包，並可以達到跨區域、國家、WAN IP 的方式運作，每當網路連線不通時，每個區域都可以單獨運作。當連線回復時，資料將完整同步到每個分散記憶體式儲存設備單元(1A)中的每個區塊記憶體磁碟(11)。

【0059】 而其中於一些硬體規格上之實施態樣選擇中，該 CPU 係為 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 其中之一種處理器。

【0060】 該電腦單元(10)之記憶體使用方式是直接由 CPU 的記憶體控制器直接以三通多或多通道以 800 MHz 到 1333 MHz 以上的速度直接存取記憶體資料。

【0061】 其中使用的單一記憶體容量為 1MB(megabyte) 至 16 ZB(zettabyte)，所採用的記憶體種類可以為動態異步記憶體 DRAM 列舉形式為：FPM RAM、EDO RAM；或動態同步記憶體 DRAM 列舉形式為：SDRAM、DDR SDRAM、DDR2 SDRAM、DDR3 SDRAM、DDR4 SDRAM、DDR5 SDRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或動態移動平台記憶體列舉形式為：LPDDR、LPDDR2、LPDDR3、LPDDR4 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或動態繪圖處理記憶體列舉形式為 VRAM、WRAM、

MDRAM、SGRAM、SDRAM、GDDR、GDDR2、GDDR3、GDDR4、GDDR5、GDDR6、GDDR7 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或磁阻式隨機存取記憶體，列舉形式為MRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或鐵電式記憶體，列舉形式為FeRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或相變化記憶體，列舉形式為PCRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或可變電阻式記憶體，列舉形式為ReRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或動態 Rambus 記憶體列舉形式為：RDRAM、XDR DRAM、XDR2 DRAM 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號；或靜態隨機存取存儲器（Static Random-Access Memory, SRAM）、或唯讀記憶體（Read-Only Memory, ROM）、或快閃記憶體（Flash Memory）列舉形式為：NOR Flash、NAND Flash、3D NAND Flash、V-Flash、SLC、MLC、eMMC、TLC 以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態型號。

【0062】再者，硬碟儲存裝置(12)可使用包含傳統磁頭與磁片的硬碟、固態硬碟、網路硬碟、SAS 介面硬碟、SATA 介面硬碟、mSATA 介面硬碟、mPCIe 介面硬碟、FC 介面硬碟、SCSI 介面硬碟、ATA 介面硬碟、NAND

Flash 外接卡、FCoE 介面硬碟，以及所謂列舉出型號為依照存取速度增加與存取方式不同所產生新的向上相容型態硬碟。

【0063】又其中，網路卡可以選擇乙太網路、快速乙太網路、千兆乙太網路、光纖、令牌環網、InfiniBand、FCoE(Fibre Channel over Ethernet)或無線網路，而其網路速度可依照網路協定的不同採用 2Mbit/s、10 Mbit/s、11Mbit/s、40Mbit/s、54Mbit/s、80Mbit/s、100 Mbit/s、150Mbit/s、300Mbit/s、433Mbit/s、1000 Mbit/s、1Gbit/s、8Gbit/s、10Gbit/s、16Gbit/s、32Gbit/s、40Gbit/s、56Gbit/s、100Gbit/s、160Gbit/s、1000Gbit/s 以上的網路速度與任何新型網路通訊協定的網路卡。

【0064】其中，主機板可相容 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 任何一種處理器相容的任何架構與品牌電腦製造商所產生的包括 BeagleBoneBlack、Raspberry Pi 架構主機板。

【0065】另外值得注意的是，作業系統檔案格式，可以是 VMFS3、VMFS5 與格式不同所產生新的向上相容型態型號、VHD 與格式不同所產生新的向上相容型態型號、VHDX 與格式不同所產生新的向上相容型態型號、VMDK 與格式不同所產生新的向上相容型態型號、HDFS 與格式不同所產生新的向上相容型態型號、Isilon OneFS 與格式不同所產生新的向上相容型態型號、記憶體式的分頁檔案所產生的任何格式與格式不同所產生新的向上相容

型態型號、VEs 與格式不同所產生新的向上相容型態型號、VPSs 與格式不同所產生新的向上相容型態型號、Ceph、GlusterFS、SphereFS、Taobao File System、ZFS、SDFS、MooseFS、AdvFS、Be 檔案系統 (BFS)、Btrfs、Coda、CrossDOS、光碟存檔系統 (DFS)、Episode、EFS、exFAT、ext、FAT、全域檔案系統 (GFS)、分層檔案系統 (HFS)、HFS Plus、高效能檔案系統、IBM 通用並列檔案系統、JFS、Macintosh 檔案系統、MINIX、NetWare 檔案系統、NILFS、Novell 儲存服務、NTFS、QFS、QNX4FS、ReiserFS (Reiser4)、SpadFS、UBIFS、Unix 檔案系統、Veritas 檔案系統 (VxFS)、VFAT、任意位置寫入檔案布局 (WAFL)、XFS、Xsan、ZFS、CHFS、FFS2、F2FS、JFFS、JFFS2、LogFS、NVFS、YAFFS、UBIFS、DCE/DFS、MFS、CXFS、GFS2、Google 檔案系統、OCFS、OCFS2、QFS、Xsan、AFS、OpenAFS、AFP、MS-DFS、GPFS、Lustre、NCP、NFS、POHMEFS、Hadoop、HAMMER、SMB (CIFS)、cramfs、FUSE、SquashFS、UMSDOS、UnionFS、configfs、devfs、procfs、specfs、sysfs、tmpfs、WinFS、EncFS、EFS、ZFS、RAW、ASM、LVM、SFS、MPFS、MGFS 其中之一。

【0066】而實體的網路協定傳送，以選擇乙太網路、快速乙太網路、千兆乙太網路、光纖、令牌環網、SS7、GSM、GPRS、EDGE、HSPA、HSPA+、CDMA、WCDMA、TD-WCDMA、LTE、GSM、cdmaOne、CDMA2000、UMTS、WCDMA、TD-SCDMA、WiMAX、3G 播報網、CDMA2000 1X、Wi-Fi、

SuperWiFi、Wi-Fi GO、以及向上相容 IEEE 所制定出新的網路傳輸協定。

【0067】 透過本發明所提出之系統架構方法之下，我們採用分散式記憶體儲存系統，可以滿足四種需求的擴充，即網路頻寬傳輸的擴充、磁碟容量的擴充、IOPS速度的擴充、記憶體 I/O 傳輸速度增加的擴充。同時可以做到滿足跨區域、資料中心、WAN 的使用，讓用戶需求可以在當地的群集記憶體磁碟 (Memory Disk Cluster) 集中提供服務、也可以漸進式的擴充群集記憶體磁碟容量，進一步提供跨區域或跨國的資料服務。

【0068】 在隨著儲存器的增加，每增加一台伺服器就累加一台機器的網路頻寬與磁碟容量，當形成一個資源池 (Resource Pool) 時，這分散記憶體式儲存設備單元 (1A) 就像一顆實體硬碟，其中一台實體主機故障時不會影響整體運作，複本中區塊記憶體磁碟 (11) 會重新複製資料到新的區塊記憶體磁碟 (11)，維持一定資料備份基礎，同時也採取連續資料複製 (Continuous Data Protector, CDP) 為企業資料的備份和回復提供創新技術，以破除磁帶備份經常失敗，一天只能備份一次問題。

【0069】 也讓複本產生資料可以由不同區塊記憶體磁碟 (11) 送出達到多對一傳輸資料，當用戶數量增加，只要增加區塊記憶體磁碟 (11) 數量，可以達到多對多的傳輸，讓 RAID 的多數硬碟崩潰造成整體資料遺失

的缺點，儲存器網路介面數量與速度的限制造成用戶過多的資料擁擠造成的延遲，LUN 擴充數量的問題，資料中心無法跨區使用的問題，因為採用記憶體式磁碟方式，每個檔案或是每個虛擬機都可以用檔案形式存在於記憶體，直接使用記憶體匯流排最高的 I/O 高速度，資料在 CPU 與記憶體中搬遷移動，I/O 次數、距離、速度都是最高的，綜觀上述優勢，是為習知技術所不能及者，故可見其增益性所在。

【0070】 以上所述之實施例僅係為說明本發明之技術思想及特點，其目的在使熟習此項技藝之人士能夠瞭解本發明之內容並據以實施，當不能以之限定本發明之專利範圍，即大凡依本發明所揭示之精神所作之均等變化或修飾，仍應涵蓋在本發明之專利範圍內。

【0071】 綜觀上述，可見本發明在突破先前之技術下，確實已達到所欲增進之功效，且也非熟悉該項技藝者所易於思及，其所具之進步性、實用性，顯已符合專利之申請要件，爰依法提出專利申請，懇請 貴局核准本件發明專利申請案，以勵發明，至感德便。

【符號說明】

【0072】

群集架構(1)

分散記憶體式儲存設備單元(1A)

電腦單元(10)

第一資料中心(101)

群集資料主控站(1011)

第二資料中心(102)

群集資料備份站(1021)

備份中心(103)

區塊記憶體磁碟(11)

硬碟儲存裝置(12)

網路卡(13)

連接埠群組現用鏈(20)

分散式記憶體檔案系統(40)

【生物材料寄存】

國內寄存資訊【請依寄存機構、日期、號碼順序註記】

無

國外寄存資訊【請依寄存國家、機構、日期、號碼順序註記】

無

【序列表】(請換頁單獨記載)

無

申請專利範圍

1. 一種分散式記憶體磁碟群集儲存系統運作方法，其為：
首先做分散記憶體式儲存設備之建置，其係包含將複數台的電腦單元組成群集架構而形成群集記憶體磁碟；
而後在電腦單元安裝系統虛擬機器平台作業系統，藉此安裝令電腦單元形成多個虛擬機器，而該電腦單元設定利用作業系統或程式軟體來設定記憶體容量占用之方式，將該記憶體規劃成儲存裝置，做成複數個區塊記憶體磁碟；進而將每一個檔案拆成一到多份的資料，保持一到多份的複本複製平均分散在所有區塊記憶體磁碟上，並使用多通道的記憶體匯流排平行存取記憶體模組，而將記憶體模組之容量規劃成磁碟使用，而其中該記憶體模組之存取支援所有作業系統檔案格式，並且採用分散式存放架構，將資料可以複製複本一份以上；
當虛擬機器之作業系統在記憶體模組中直接由CPU存取需要的檔案，處理資料後存回記憶體模組中，虛擬機器做計算使用的記憶體也在記憶體模組中，計算完依然直接存回在所存放虛擬機器作業系統的記憶體模組位置，讓這個儲存系統內資料可以在電腦單元中以平行運算的方式在區塊記憶體磁碟處理資料；其他沒有在電腦單元的資料係透過網路卡連接一連接埠群組現用鏈以存取

其他電腦單元的區塊記憶體磁碟；

而各電腦單元進一步依功能性區分為一第一資料中心、至少一個第二資料中心及一備份中心；其中該第一資料備份中心係以一虛擬群集資料主控站控管，其中各該第二資料中心係以一虛擬群集資料備份站控管，且其中該備份中心係以虛擬群集資料備份站控管，而其中該第一資料中心與該第二資料中心共同形成一分散式記憶體檔案系統；此外，更進一步是為形成堆疊架構擴展儲存容量架構，其使用網路層介面的存取方式將一台電腦單元中多個區塊記憶體磁碟規劃成一個群集概念的群集記憶體磁碟單元資源池，同步使用所有區塊記憶體磁碟取得資料；

當群集架構產生後，每一個群集架構都可以獨立使用，並且可以將每一個群集架構當作分散記憶體式儲存設備單元，同時使用網路層介面堆疊，可以將每個分散記憶體式儲存設備單元模擬成大型的記憶體磁碟，並採用新的群集資料主控站與群集資料備份站來控制資料的處理量分散在所有大型的記憶體磁碟。

2. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中各該區塊記憶體磁碟係分別電性連接於至少一硬碟儲存裝置，該硬碟儲存裝置係可於一間隔時間將該區塊記憶體磁碟之資料做備份。

3. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中所有的電腦單元中的區塊記憶體磁碟，都會使用連續資料保護方式，無間斷的連續將資料備份到統一的大型物理硬碟群集磁碟陣列中備份。
4. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該電腦單元係俱備了 CPU、記憶體、硬碟、網路卡、主機板、I/O 介面卡、連接線、外殼之計算機設備。
5. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中每份複本都可以採用 1-4096 位元 AES 與 RSA 混用加密資料存放在記憶體上，每當要存取資料時，資料在記憶體與 CPU 之間流動，虛擬機器係以檔案形式存在於記憶體模組中，虛擬記憶體所規劃使用的記憶體容量也在同一區塊。
6. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中各該區塊記憶體磁碟透過一監控單元來做其運作情況之監控，該偵測單元利用 Splunk 或者是其他搜索引擎之軟體程式來監控之，當檢測到問題時更可以配置重新啟動應用程式服務，達到一回復之功能。
7. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中虛擬機器平台作業系統採用

VMware vSphere ESXi 4.1 之後的版本、Microsoft Server 2012 R2 Hyper-V 之後的版本、Citrix XenServer Oracle VM、Oracle VM、Red Hat KVM、Red Hat Control groups(cgroups)、Red Hat Linux Containers(LXC)、KVM、Eucalyptus、OpenStack、User Mode Linux、LXC、OpenVZ、OpenNebula、Enomaly's Elastic Computing、OpenFlow 或 Linux-Base KVM。

8. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中虛擬機器作業系統採用 Linux、FreeBSD、OpenSolaris 或 MacOS X。
9. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中虛擬機器之記憶體間係使用存儲區域網路架構運行，採取軟體虛擬出來的網路層介面，將所有的區塊記憶體磁碟連接在一起協同運作。
10. 如申請專利範圍第 9 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該網路層介面，係使用 SAN、SAN iSCSI、SAN FC、SAN FCoE、NFS、NAS、JBOD、CIFS 或 FUSE 介面進行伺服器 and 磁碟驅動器設備之間的通信，並且開發有 RAMSTORAGE™ API 作為應用程式支援用。
11. 如申請專利範圍第 10 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該 RAMSTORAGE™ API 係採用 REST、Restful、C++、PHP、Python、Java、Perl 或 Javascript 程式開發軟體所形成的 RAMSTORAGE™ API，存取分散記憶體式儲存設備單元 API 功能，其包含容錯、備份、移轉、快

速佈署虛擬機器、管理磁碟大小、自動依照實際需求容量自動增加區塊記憶體磁碟、區塊間資料負載平衡移轉、備份回復的功能、連續資料保護、快照、監控資源等。

12. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中未使用的 CPU、記憶體、實際物理硬碟的資源，係使用虛擬機器平台作業系統將所有資源規劃成統一的資源池，令每一個所需要電腦資源，可以調整移轉到其他資源較充裕的電腦單元。
13. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中多個分散記憶體式儲存設備單元之連接方式依照實體的網路協定傳送，並且採用 SSL VPN、VPN 或加密演算方式傳送封包，每當網路連線不通時，每個區域都可以單獨運作，當連線回復時，資料將完整同步到每個分散記憶體式儲存設備單元中的每個區塊記憶體磁碟。
14. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該 CPU 係為 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 其中之一種處理器。
15. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該電腦單元之記憶體使用方式是直接由 CPU 的記憶體控制器直接以三通多或多通道以 800 MHz 到 1333 MHz 以上的速度直接存取記憶體資料。
16. 如申請專利範圍第 15 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中記憶體容量為 1MB 至 16

ZB。

17. 如申請專利範圍第 15 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中所採用的記憶體種類為動態異步記憶體 DRAM、動態同步記憶體 DRAM、動態移動平台記憶體、動態繪圖處理記憶體、動態 Rambus 記憶體、靜態隨機存取存儲器、唯讀記憶體、快閃記憶體或磁阻式隨機存取記憶體。
18. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態異步記憶體 DRAM 為 FPM RAM 或 EDO RAM。
19. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態同步記憶體 DRA 為 SDRAM、DDR SDRAM、DDR2 SDRAM、DDR3 SDRAM、DDR4 SDRAM 或 DDR5 SDRAM。
20. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態移動平台記憶體為 LPDDR、LPDDR2、LPDDR3 或 LPDDR4。
21. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態繪圖處理記憶體為 VRAM、WRAM、MDRAM、SGRAM、SDRAM、GDDR、GDDR2、GDDR3、GDDR4、GDDR5、GDDR6 或 GDDR7。
22. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中動態 Rambus 記憶體為 RDRAM、XDR DRAM 或 XDR2 DRAM。

23. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中快閃記憶體為 NOR Flash、NAND Flash、3D NAND Flash、V-Flash、SLC、MLC、eMMC 或 TLC。
24. 如申請專利範圍第 17 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中該磁阻式隨機存取記憶體為 MRAM。
25. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中硬碟儲存裝置為傳統磁頭、磁片的硬碟、固態硬碟、網路硬碟、SAS 介面硬碟、SATA 介面硬碟、mSATA 介面硬碟、mPCIe 介面硬碟、FC 介面硬碟、SCSI 介面硬碟、ATA 介面硬碟、NAND Flash 外接卡或 FCoE 介面硬碟。
26. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中網路卡為乙太網路、快速乙太網路、千兆乙太網路、光纖、令牌環網、InfiniBand、FCoE 或無線網路。
27. 如申請專利範圍第 26 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中網路速度採用 2Mbit/s、10 Mbit/s、11Mbit/s、40Mbit/s、54Mbit/s、80Mbit/s、100 Mbit/s、150Mbit/s、300Mbit/s、433Mbit/s、1000 Mbit/s、1Gbit/s、8Gbit/s、10Gbit/s、16Gbit/s、32Gbit/s、40Gbit/s、56Gbit/s、100Gbit/s、160Gbit/s 或 1000Gbit/s 以上的網路速度。
28. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群

集儲存系統運作方法，其中，主機板相容 x86、x86-64、IA-64、Alpha、ARM、SPARC 32 and 64、PowerPC、MIPS、Tilera 中任何一種處理器。

29. 如申請專利範圍第 1 項所述之分散式記憶體磁碟群集儲存系統運作方法，其中作業系統檔案格式是 VMFS3、VMFS5 與格式不同所產生新的向上相容型態型號、VHD 與格式不同所產生新的向上相容型態型號、VHDX 與格式不同所產生新的向上相容型態型號、VMDK 與格式不同所產生新的向上相容型態型號、HDFS 與格式不同所產生新的向上相容型態型號、Isilon OneFS 與格式不同所產生新的向上相容型態型號、記憶體式的分頁檔案所產生的任何格式與格式不同所產生新的向上相容型態型號、VEs 與格式不同所產生新的向上相容型態型號、VPSs 與格式不同所產生新的向上相容型態型號、Ceph、GlusterFS、SphereFS、Taobao File System、ZFS、SDFS、MooseFS、AdvFS、BFS、Btrfs、Coda、CrossDOS、DFS、Episode、EFS、exFAT、ext、FAT、GFS、HFS、HFS Plus、高效能檔案系統、IBM 通用並列檔案系統、JFS、Macintosh 檔案系統、MINIX、NetWare 檔案系統、NILFS、Novell 儲存服務、NTFS、QFS、QNX4FS、ReiserFS、SpadFS、UBIFS、Unix 檔案系統、Veritas VxFS、VFAT、WAFL、XFS、Xsan、ZFS、CHFS、FFS2、F2FS、JFFS、JFFS2、LogFS、NVFS、YAFFS、UBIFS、DCE/DFS、MFS、CXFS、GFS2、Google 檔案系統、OCFS、OCFS2、QFS、Xsan、AFS、OpenAFS、AFP、MS-DFS、GPFS、Lustre、NCP、NFS、POHMELFS、Hadoop、HAMMER、SMB (CIFs)、cramfs、FUSE、SquashFS、UMSDOS、UnionFS、configfs、devfs、procfs、specfs、sysfs、tmpfs、WinFS、

EncFS、EFS、ZFS、RAW、ASM、LVM、SFS、MPFS 或 MGFS。