

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※ 申請案號：94111494

※ 申請日期：94.4.12

※IPC 分類：G06F 15/17 9/00

一、發明名稱：(中文/英文)

(2006.01)

利用動態開機影像和虛擬碟片進行電腦叢集虛擬化之系統和方法

SYSTEM AND METHOD FOR COMPUTER CLUSTER VIRTUALIZATION USING DYNAMIC BOOT
IMAGES AND VIRTUAL DISK

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

雷神公司/RAYTHEON COMPANY

代表人：(中文/英文)

杉德 約翰 J./SNYDER, JOHN J.

住居所或營業所地址：(中文/英文)

美國麻州瓦勒生·冬天街870號/870 Winter Street, Waltham, MA 02451-1449, USA

國 籍：(中文/英文)

美國/USA

三、發明人：(共 2 人)

姓 名：(中文/英文)

1. 大衛森 夏儂 V./DAVIDSON, SHANNON V.

2. 彼得森 羅伯特 J./PETERSON, ROBERT J.

國 籍：(中文/英文)

美國/USA

四、聲明事項：

主張專利法第二十二條第二項 第一款或 第二款規定之事實，其事實發生日期為： 年 月 日。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

1. 美國; 2004,04,15; 10/825,345

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

發明領域

概略言之，本揭示係有關資料處理領域，特別係有關
5 使用動態開機影像及虛擬化碟片存取之電腦叢集虛擬化方法。

【先前技術】

發明背景

典型地，企業應用程式係於專屬電腦資源上執行。經
10 常企業包括多種不同運算環境來用於不同應用實例，例如製造、測試與開發。此等多重運算環境典型具有與實際情況或製造情況相等的尺寸及容量。此外，非製造環境經常閑置一段長時間。如此通常導致大量電腦資源及勞力費用浪費於維修與行政管理此等各項環境。

15 目前企業可使用供應作業(provisioning)來試圖解決此等問題。一般而言，供應作業係經由將本地碟片由儲存庫拷貝之電腦資源，而例示(instantiating)電腦資源給企業應用程式之程序。然後該電腦資源以所供應作業之作業系統及軟體，經由通常耗時超過10分鐘之過程來開機。

20 【發明內容】

發明概要

本揭示提供一種電腦叢集虛擬化之系統及方法包括選擇一分散式應用程式。取還一種與該分散式應用程式相關聯之策略。動態選定多數節點之一。然後至少部分基於所

取還之策略而復置所選定之節點之開機影像，該開機影像係與該分散式應用程式可相容。然後，一虛擬碟片影像與該節點關聯。然後使用相關聯之虛擬碟片影像，於該復置節點執行至少部分該分散式應用程式。

- 5 本發明有數項重要技術優勢。例如本發明之一項可能優勢為其允許運算節點於飛行中再度做供應作業，來變成分散式應用程式之一虛擬叢集之一成員，藉此可將供應作業時間縮短至15秒或15秒以下。本揭示之另一項可能優勢係減少資訊技術(IT)硬體成本及維護成本達至少30%。此外，當一應用程式並非處於尖峰處理期時，該應用程式之閉置節點可被動態重新配置或重新供應給其它分散式應用程式。又另一項可能之優勢為，本揭示提供集中式容量規劃、效能監視、及簡化之行政管理。此外，本發明允許有較佳節點故障管理。各個本發明之具體例可不具有此等優勢或具有部分或全部此等優勢。其它本發明之技術優勢對熟諳技藝人士顯然自明。
- 10
- 15

圖式簡單說明

為求更完整了解本揭示及其優點，現在參照詳細說明部分連同附圖做說明，附圖者：

- 20 第1圖顯示根據本揭示之一具體例，提供動態開機之一實例分散式系統；以及

第2圖顯示根據本揭示之一具體例，動態重新開機一節點之實例方法。

【實施方式】

較佳實施例之詳細說明

第1圖為方塊圖，顯示使用動態開機影像131來執行軟體應用程式114及處理之一種分散式運算系統100。通常，系統100為企業或其它分散式應用程式用之可擴充式分散式運算環境。系統100提供可擴充式、容許錯誤之運算環境，其可基於運算需求動態成長，且同時可經由提供各個應用程式114有其本身之可擴充式虛擬叢集來同時提供多個運算資源給多數應用程式114。例如系統100包括伺服器102，伺服器102係經由網路116而連結至一或多個行政工作站或本地客端120。但系統100另外可為孤立運算環境，或為任何其它之適當環境。簡言之，系統100可為視應用程式114之需求、參數、及處理需要變化，而允許節點104於飛行中動態自動配置之任一種運算環境。「動態」用於此處通常係表示某些處理至少部分係基於一或多項變數來決定回合時間。「自動」一詞用於此處通常係表示適當處理實質上係藉至少部分系統100執行。須了解「自動」進一步涵蓋任何適當使用者或行政管理者與系統100互動，而未悖離本揭示之範圍。

伺服器102包含任何可操作來執行多數應用程式114跨一個節點或多個節點108之任一種本地電腦或分散式電腦。通常，伺服器102包含分散式電腦，諸如框架安裝伺服器、刀鋒伺服器、或其它分散式伺服器。節點108包含任一種電腦或處理裝置，諸如刀鋒伺服器、通用個人電腦(PC)、麥金塔電腦、工作站、基於優尼士(Unix)之電腦或任何其它

適當裝置。概略言之，第1圖僅提供可用於本揭示之電腦或刀鋒伺服器實例。舉例言之，雖然第1圖顯示可用於本揭示之刀鋒伺服器102，但伺服器102也可使用伺服器以外之電腦實作以及使用伺服器區實作。換言之，本揭示意圖涵蓋通用電腦以外之電腦以及不含習知作業系統之電腦。如本文件使用，「電腦」意圖涵蓋個人電腦、工作站、網路電腦或任何其它適當處理裝置。伺服器102或組成節點108可調整配合來執行任一種作業系統，包括Linux、UNIX、視窗伺服器(Windows Server)、或任何其它適當作業系統。根據一具體例，伺服器102也包括遠端網路伺服器或可通訊式耦接遠端網路伺服器。

所示伺服器102包括一管理節點104，管理節點104係通訊式耦接多個節點108，且可操作來執行動態開機引擎105。但須了解伺服器102及節點108可能並未包括全部所舉例說明之組成元件。管理節點104包含實質專用於管理伺服器102或輔助行政管理之至少一刀鋒裝置或運算裝置。舉例言之，管理節點104可包含二熱插拔刀鋒裝置，二刀鋒裝置或框架安裝伺服器之一為冗餘(諸如主動/被動組配結構)。

動態開機引擎105可包括任一種硬體、軟體、韌體或其組合，其可操作而使用虛擬節點叢集108(或應用程式環境)來動態配置節點與管理節點108，以及執行應用程式114。舉例言之，動態開機引擎105可以任一種適當電腦語言寫出或描述，電腦語言例如包括C、C⁺⁺、爪哇、虛擬基本語言、

組合語言、4GL之任何適當版本等，或其組合。須了解雖然動態開機引擎105於第1圖係以單一多重任務型模組舉例說明，但藉本動態開機引擎所執行之特色及功能也可藉多重模組執行，諸如實體層模組、虛擬層模組、工作排程器、及呈現引擎。此外，動態開機引擎105可為另一軟體模組之子模組或次模組，而未悖離本揭示之範圍。因此動態開機引擎105包含一或多個軟體模組，其可操作而基於策略132來智慧型管理節點108及應用程式114。

通常動態開機引擎105管理一或多項應用程式114之方式係經由於個別節點108開始及停止應用程式環境。舉例言之，動態開機引擎105可以來自開機影像檔案130之不同開機影像131來復置該特定節點108，該開機影像131為該預定之應用程式環境之特定影像或與該環境可相容。換言之，動態開機引擎105支援於任何經過控制之節點108動態開機任何適當應用程式環境。如此，動態開機引擎105也可支援於任何經過控制之電腦對公用IP介面動態設定IP位址或MAC位址。動態開機引擎105也可使用網路開機協定而直接由網路開機任何節點108，或由所連結之碟片儲存裝置開機來開機任何節點108。動態開機引擎105也可對任何經過控制之電腦利用高速網路存取來存取含虛擬本地碟片影像之作業系統、服務及應用程式。須了解動態開機引擎105可基於日曆日期或時間、或使用任何其它預定參數來啟動或關閉應用程式環境。

動態開機引擎105也支援各項錯誤容忍技術與復原技

術。例如動態開機引擎105可經由對故障之節點108自動更換且動態重新開機更換節點108，而由單一硬體組成元件故障中自動復原伺服器102。此外，由於動態開機引擎105可自動辨識新節點108且可從事任何所需組配結構、復置或開機，故動態開機引擎105可輔助安裝新節點108。

節點108包含任何可操作來執行至少部分應用程式114(諸如一項任務或一項程序)之任何電腦、刀鋒伺服器或伺服器。舉例說明之節點108於高階，包括記憶體109及處理器110。記憶體109可包括任一種記憶體或資料庫模組，記憶體109可呈依電性記憶體或非依電性記憶體形式包括(但非限制性)磁性媒體、光學媒體、隨機存取記憶體(RAM)、唯讀記憶體(ROM)、活動式媒體或任何其它適當本地或遠端記憶體組成元件。記憶體109可包括多項本地資訊之任一種。節點108也包括處理器110。處理器110執行指令且操控資料，來執行伺服器102之操作，處理器110例如為中央處理單元(CPU)或欄位可規劃陣列(FPGA)。雖然第1圖顯示於各個節點108有單一處理器110，但根據特定需求可使用多數處理器110，述及處理器110表示適用時包含多數處理器110。處理器110可包括開機影像之任一種指標器，例如可電子抹除之可規劃唯讀記憶體(EEPROM) 111。但須了解節點108可包含以任一種適當方式組配之任何數目之組成元件，而未悖離本揭示之範圍。節點108也包括一或多個本地硬體驅動器，用於提供本地暫時檔案空間及虛擬記憶體插拔空間之目的。

應用程式114可包含任何企業應用程式或分散式應用程式，諸如資料庫管理系統(DBMS)、財務軟體等。典型地，應用程式114包含以任一種適當語言寫出之軟體，且可操作來執行任一項資料處理。但非習知應用程式也屬於本揭示之範圍。應用程式114可於應用程式環境或於虛擬叢集執行，虛擬叢集以邏輯方式界定應用程式之執行環境。一具體例中，應用程式環境包含i)應用程式環境之名稱及說明；ii)最小/最大節點108數目；iii)軟體組配資訊，諸如作業系統軟體版本及應用程式114軟體版本；及iv)各個節點108之硬體組配結構，諸如開機影像、主機名稱及IP位址、節點108開機後施加之顧客組配狀態、虛擬本地碟片影像、本地檔案系統、欲安裝之檔案系統及網路組配結構。但須了解任一種適當參數、變數、或特徵皆可用來輔助動態開機引擎105界定、定位、以及處理該應用程式環境。舉例言之，應用程式環境也包括應用程式114之啟動、關閉、與健康情況監視等資訊。

伺服器102包括於賓主式環境或其它分散式環境透過網路116而與其它電腦系統諸如客端120通訊之介面115。若干具體例中，伺服器102由網路116接收開機影像131、虛擬本地碟片影像134、策略132或應用程式資料140，用以透過高速介面115來儲存或處理。通常，介面115包含以適當組合編碼於軟體及/或硬體之邏輯，且可操作來與網路116通訊。特別，介面115包含支援與通訊網路116相關聯之一或多項通訊協定之軟體，或介面115包含可操作來通訊實體信

號之硬體。

網路116可輔助電腦伺服器102與任何其它電腦諸如客
端120間之無線通訊或有線通訊。確實雖然於駐在伺服器
102與客端120間舉例說明，但網路116也可駐在各個節點
5 108間，而未悖離本發明之範圍。換言之，網路116涵蓋任
何可輔助各個運算組成元件間之通訊之單一網路、多數網
路或次網路。網路116例如可介於各網址間通訊例如網際網
路協定(IP)封包、框架中繼框架、異步傳輸模型(ATM)單
元、語音、視訊、資料及其它適當資訊。網路116也可根據
10 任一種適當通訊協定如無線頻寬(IB)、十億位元乙太網路
(GE)、或光纖通道(FC)而處理資料封包與路由資料封包。
資料封包典型係用來傳輸應用程式資料140內部之資料。資
料封包可包括標頭，標頭有來源識別符及目的地識別符。
來源識別符例如為來源位址，來源識別符可識別該資訊之
15 發送器；而目的識別符例如為目的地位址則可識別該資訊
之接收端。網路116包括一或多區域網路(LAN)、無線存取
網路(RAN)、都會區域網路(MAN)、廣域網路(WAN)、全部
或部分稱作為網際網路之通用電腦網路及/或任何其它於
一或多個位置之通訊系統。

20 開機表130可為任何包括多數開機影像131之碟片場或
網路檔案系統。雖然係以遠端舉例說明，但開機影像131可
藉動態開機引擎105前置載入，來簡化初始化及安裝。開機
影像131為任一種形式、影像、指標器或述及至少部分開機
驅動一次作業系統區隔。開機影像131典型係呈二元形式。

開機影像型別包括來自檔案之核心影像、來自方塊裝置或軟碟之核心影像、或若干作業系統之開機扇區。例如Linux

開機影像呈現：

```

5      0x1B031336 ,
      0x4 ,
      0x90000000 ,
      0x90000200 ,

10     0x4 ,
      0x90200 ,
      0x800 ,
      0x800 ,

15     0x4 ,
      0x10000 ,
      0x80000 ,
      0x80000 ,

20     0x04000004 ,
      0x100000 ,
      0x80000 ,
      0x80000

```

須了解前述實例開機影像131僅供舉例說明之用，可未包括
 25 所示元件或包括部分或全部所示元件以及額外未顯示之其它
 元件。此外，開機影像131可為與前述實例不同之佈局或
 格式，而未悖離本揭示之範圍。

策略表132包括管理節點108及應用程式114之任何參
 數。例如策略132可對應用程式環境自動加減節點108。另
 30 外或此外，策略132可由伺服器102用來解決競爭應用程式
 114間之問題。通常，策略表132可包含以SQL陳述或SQL
 稿本描述之一或多個儲存於相關資料庫之表。另一具體例
 中，策略表132可儲存或定義多項資料結構為XML文件、虛

擬儲存存取方法(VSAM)檔案、平坦檔案、Btrieve檔案，或
隔開之CSV檔案。策略表132也包含多數表或檔案儲存於一
部電腦或儲存於多部電腦。此外，策略表132可為本地或遠
端，而未悖離本揭示之範圍，策略表132可儲存任何型別之
5 適當資料。舉例言之，策略表132可儲存個別虛擬叢集策
略，包括：i)指定給一應用程式環境之最小/最大節點108數
目；ii)指定給該應用程式之內設伺服器數目；iii)動態增加
節點108至應用程式環境之條件；iv)由應用程式環境動態移
除節點108之條件；v)移除節點108(例如關閉網路接取)，但
10 留下直到問題研究調查之條件；以及vi)因應用程式114主動
執行異動處理或程序，故節點108不可被移除之條件。

另一例中，策略表132可包括任何數目之虛擬叢集間之
策略，諸如優先順位策略、資源共享策略及先占策略。若
有策略衝突，則優先順位典型判定何者應用程式環境將取
15 得資源。舉例言之，若一特定應用程式環境之優先順位較
高，則可優先存取節點108。資源共享經常係基於應用程式
環境之經界定之資格。例如各個應用程式環境可被授與某
個百分比節點108之資格。資源共享也可基於於滑動時間窗
口之電腦之使用。先占策略允許較高優先順位環境優於來
20 自較低優先順位之應用程式環境之節點108。

虛擬本地碟片影像表133為包括複數個虛擬本地碟片
影像134之任何碟片場或網路檔案系統。雖然係以遠端舉例
說明，但管理節點104可預先載入作業系統及應用程式軟體
來簡化初始化與安裝。虛擬本地碟片影像134可為任何形

式、影像、指標器或述及用於各應用程式之各個虛擬節點之本地碟片儲存裝置。虛擬本地碟片影像134典型包括作業系統、組配之服務、以及各個應用程式之虛擬節點所安裝之應用程式。各個虛擬本地碟片影像134可含有多重檔案系統，該等檔案系統可為唯讀來於多個節點間共享；或為可修飾之檔案系統，其通常為一應用程式節點之特定檔案系統。虛擬本地碟片影像134可以階層式目錄儲存於傳統檔案系統內部，或可儲存於可復原資料庫，帶有網路檔案系統介面提供給應用程式節點。

10 通常，應用程式資料140為任何可用於儲存應用程式114之資料之記憶體、資料庫、儲存區域網路(SAN)或網路連結儲存裝置(NAS)。應用程式資料140可包含一或多表係儲存於以SQL陳述或稿本描述之相關資料庫。另一具體例
15 文件、VSAM檔案、平坦檔案、Btrieve檔案或CSV檔案。應用程式資料140也包含多個表或檔案儲存於一部電腦或儲存跨多部電腦。此外，應用程式資料140可為本地或遠端，而未悖離本揭示之範圍。

客端120可為任一種可操作而透過圖形使用者介面
20 (GUI) 122呈現行政管理畫面給使用者之任一種裝置。於高階時，所示客端120至少包括GUI 122，以及包含電子運算裝置，其可操作來接收、發送、處理與儲存與系統100相關聯之任何適當資料。須了解可有任何數目之客端120通訊式耦接至伺服器102。此外，「客端120」及「客端120之使用

者」可視情況互換使用，而未悖離本揭示之範圍。此外，為求方便說明，各個客端係以由一位使用者使用做說明。但本揭示涵蓋多位使用者使用一部電腦來通訊命令，或使用同一部GUI 122來觀視圖形呈現。

- 5 如本揭示使用，客端120意圖涵蓋個人電腦、觸控式終端裝置、工作站、網路電腦、資訊服務站、無線資料埠、蜂巢式電話、個人資料助理器(PDA)、於此等或其它裝置內部之一或多處理器、或任何其它適當處理裝置。舉例言之，客端120可包含電腦，該電腦包括一輸入裝置，諸如數字小
- 10 鍵盤、觸控式螢幕、滑鼠或其它可接收資訊之裝置，以及可傳遞與伺服器102或客端120之操作以及可傳遞與伺服器102或客端120之操作相關聯之資訊(包括數位資料、視訊資訊或GUI 122)之輸出裝置。輸入裝置及輸出裝置二者可包括固定式儲存媒體或活動式儲存媒體，諸如電腦磁碟、
- 15 CD-ROM或其它適當媒體來經由行政管理顯示畫面或工作提交顯示畫面亦即GUI 122而接收來自客端120之使用者之輸入，以及提供輸出給客端120之使用者。

GUI 122包含一圖形使用者介面，該GUI可操作來允許系統(或網路)行政管理者與系統100介接，來監視應用程式

20 114或系統之效能、來修改虛擬叢集、或來做任何適當監督目的。通常，GUI 122對客端120使用者提供有效且使用者友善之由系統100所提供之資料呈現。GUI 122包含複數個客製訊框或視圖，其具有互動式欄位、下拉表單、及可由使用者操控之鈕。一具體例中，GUI 122呈現顯示畫面，該

顯示畫面提供應用程式環境之各圖形視圖或策略畫面；以及GUI 122透過輸入裝置之一而接收來自客端120之使用者之命令。此等圖形視圖包括i)應用程式環境、節點資源、以及所監視之負載之目前狀態之圖形呈現；ii)應用程式環境及節點負載與節點使用隨著時間經過之圖形呈現；iii)精靈；及iv)於各個應用程式環境及各個節點108執行之應用程式114之視圖。簡言之，GUI 122可呈現節點108之任何實體上與邏輯上之狀態及特性給系統行政管理者，以及GUI可接收來自行政管理者之各個命令。

10 一具體例中，GUI 122允許行政管理者形成、刪除、拷貝及修改應用程式環境。行政管理者也使用GUI 122來設定應用程式環境共享策略，來激發與鈍化應用程式環境，來監視應用程式環境及節點108之狀態及負載。此外，GUI 122也允許由主動應用程式環境加減節點108。GUI 122也可基於各項系統100之特性來警告行政管理者，該等系統特性諸如於節點108或應用程式環境達到可組配之負載程度；節點108變成無法利用；應用程式環境開始或停止；由應用程式環境加減節點108；伺服器102無法滿足最低應用程式環境需求，或超過服務需求程度(例如異動處理之回應時間)。

20 須了解，圖形使用者介面一詞可以單數形或多數形用來描述一或多圖形使用者介面，以及特定圖形使用者介面之個別顯示器。因此GUI 122涵蓋任何可於系統100處理資訊，且有效呈現處理資訊結果給使用者之任一種圖形使用者介面，諸如通稱之網路瀏覽器。GUI 122也可操作而透過

SSL-HTTPS通訊之安全性瀏覽器。伺服器102可透過網路瀏覽器(例如微軟公司網際網路探索者(Microsoft Internet Explorer)或網景公司遨翔者(Netscape Navigator))而接收來自客端120之資訊，以及使用網路116來送返適當XTML或XML回應。

於操作之一方面，動態開機引擎105選擇分散式應用程式114。基於一或多項相關策略132，動態開機引擎105可動態加減一或多個經選定之節點108至特殊應用程式環境或虛擬叢集。基於取還之策略132，動態開機引擎105為選定之節點108選擇適當開機影像132。舉例言之，若已經有四個節點108執行部分應用程式114，則動態開機引擎105可自動選擇與應用程式114相容之第五開機影像132(至少部分係基於節點108之硬體及其它特性、以及基於一或多策略132)。基於取還之策略132，動態開機引擎105也可為所選定之節點108選擇適當之虛擬本地碟片影像134。一旦已經選定適當開機影像132及/或虛擬本地碟片影像134，則動態開機引擎105對節點108使用指標器或其它參照來閃現所選定之開機影像132及虛擬本地碟片影像134，以及將節點108重新開機。一旦節點108已經初始化(通常係少於15秒)，動態開機引擎105(或若干其它工作排程器)於該選定之節點108執行適當任務、處理程序或其它應用程式104部分。

第2圖為流程圖，顯示根據本發明之一具體例，一節點108之動態重新開機之實例方法200。第2圖顯示方法200，其概略描述動態配置多數節點108之一給一虛擬叢集或一

應用程式環境。當然於本揭示之範圍內，任何數目之節點108皆可循序或同時復置、重新開機、或以其它方式配置。於高階，方法200包括選擇節點108用於配置給一應用程式114之環境；復置該選定之節點108之開機影像；以及重新開機該節點108。後文說明之重點集中於實施方法200時動態開機引擎105之操作。但系統100涵蓋使用邏輯元件之任一種適當組合及排列來實作部分或全部此處所述功能。

方法200始於步驟205，此處動態開機引擎105決定軟體應用程式114須被配置較多節點108。此項決定可使用適當技術達成。例如，行政管理者以手動方式將節點108加至應用程式114之應用程式環境。於另一例中，動態開機引擎105可動態決定基於策略132而可使用或應使用之節點108。其次於步驟210，動態開機引擎105判定是否有任何未經利用之運算節點108可資利用。若有較多節點108可資利用，則動態開機引擎105於步驟215，使用任一種適當技術來選擇第一個可用之運算節點108。例如，動態開機引擎105可經於實體位置、虛擬位置、應用程式114之相容性、處理器速度、或任何其它適當特性來選擇節點108。於決策步驟220，動態開機引擎105判定選定之節點是否與應用程式114可相容。若節點108並非與應用程式114可相容，則於步驟225，動態開機引擎105使用任一種適當技術取下該選定之節點。其次，於步驟230，動態開機引擎105基於軟體應用程式114而動態選擇策略132。例如動態開機引擎105決定三個節點108目前正在執行軟體應用程式114。基於此項判定，

動態開機引擎105於策略132定位第四邏輯節點108。基於所選用之策略132，於步驟235，動態開機引擎105以指標器閃現選定之節點給新開機影像131；以及於步驟237，關聯虛擬本地碟片影像134。如前文說明，動態開機引擎105也可

5 閃現EEPROM 111或任何其它適當組成元件。其次，於步驟240，動態開機引擎105使用新的開機影像131來開機所選定之節點108。一旦節點108已經被重新開機(或若該節點已經與應用程式114可相容)，則於步驟245，動態開機引擎105於選定之節點108執行應用程式114，以及方法200結束。

10 回頭參照決定步驟210，若無任何運算節點108可資利用，則於步驟250，動態開機引擎105對應用程式114選擇最佳利用之節點108。此項最佳節點108之選擇可以任一種方式進行，例如決定最少被利用之節點108，選擇可相容之節點108，或判定某些其它「最佳匹配」狀況。於步驟255，

15 動態開機引擎105消除於選定之節點108之目前處理程序。動態開機引擎105可使用任一種適當技術來結束處理，該等技術諸如執行特殊應用命令、使用作用系統來消除一處理程序等。於決策步驟260，動態開機引擎105判定所選定之節點108是否與應用程式114可相容。若節點108與應用程式

20 114不可相容，則於步驟265，動態開機引擎105使用任一種適當技術來取代所選定之節點。其次，於步驟270，動態開機引擎105基於軟體應用程式114而動態選擇策略132。例如動態開機引擎105決定三個節點108目前正在執行應用程式114。基於此項決定，動態開機引擎105定位於策略132之第

四邏輯節點108。基於所選用之策略132，於步驟275，動態開機引擎105以指標器閃現選定之節點給新開機影像131；以及於步驟277，關聯虛擬本地碟片影像134。如前文說明，動態開機引擎105也可閃現EEPROM 111或任何其它適當組成元件。其次，於步驟280，動態開機引擎105使用新的開機影像131及虛擬本地碟片影像134來開機所選定之節點108。一旦節點108已經被重新開機(或若該節點已經與應用程式114可相容)，則於步驟285，動態開機引擎105於選定之節點108執行應用程式114，以及方法200結束。

10 先前流程圖及伴隨之說明僅舉例說明實例方法200。系統100涵蓋使用任一種適當技術來執行此等及其它任務。如此，本流程圖之多個步驟可同時進行及/或以所示順序以外之順序進行。此外，系統100可使用有額外步驟、較少步驟、及/或不同步驟之方法，只要該等方法為適當方法即可。

15 雖然本揭示已經就若干具體例及概略相關聯之方法做說明，但熟諳技藝人士顯然易知此等具體例及方法之變更與變異。如此，前文具體實施例之說明絕非界定或約束本揭示。可未悖離本揭示之精髓及範圍做出其它變化、取代及變更。

20 【圖式簡單說明】

第1圖顯示根據本揭示之一具體例，提供動態開機之一實例分散式系統；以及

第2圖顯示根據本揭示之一具體例，動態重新開機一節點之實例方法。

【主要元件符號說明】

100...分散式運算系統	120...客端，行政管理工作站或本地客端
102...伺服器	
104...管理節點	122...圖形使用者介面，GUI
105...動態開機引擎	130...開機影像檔案
108...節點	131...開機影像
109...記憶體	132...策略，策略表
110...處理器	133...虛擬本地碟片影像表
111...可電子抹除可規劃唯讀記憶體(EEPROM)	134...虛擬本地碟片影像
114、114a-b...應用程式	140...應用程式資料
115...介面	200...方法
116...網路	205-285...步驟

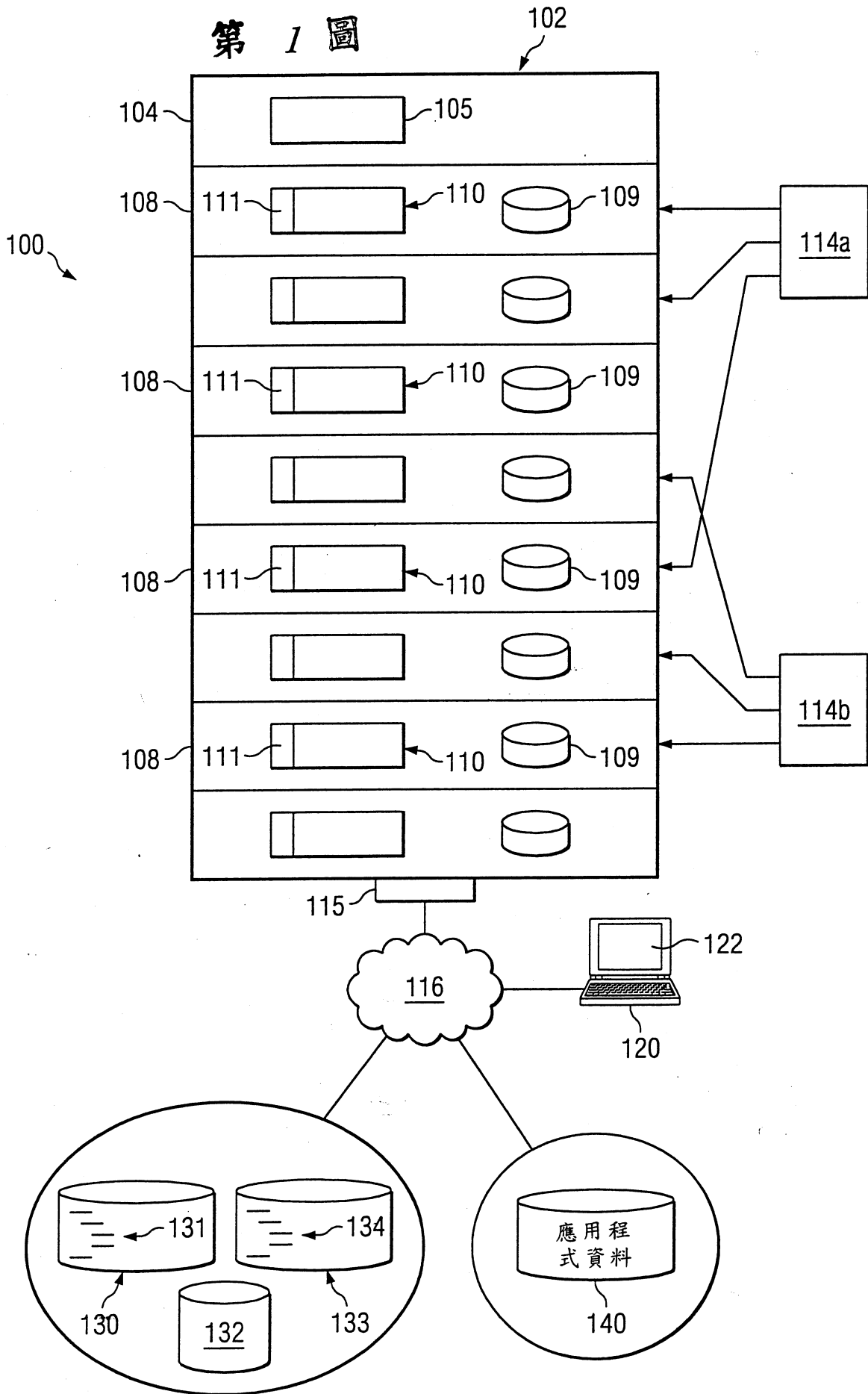
五、中文發明摘要：

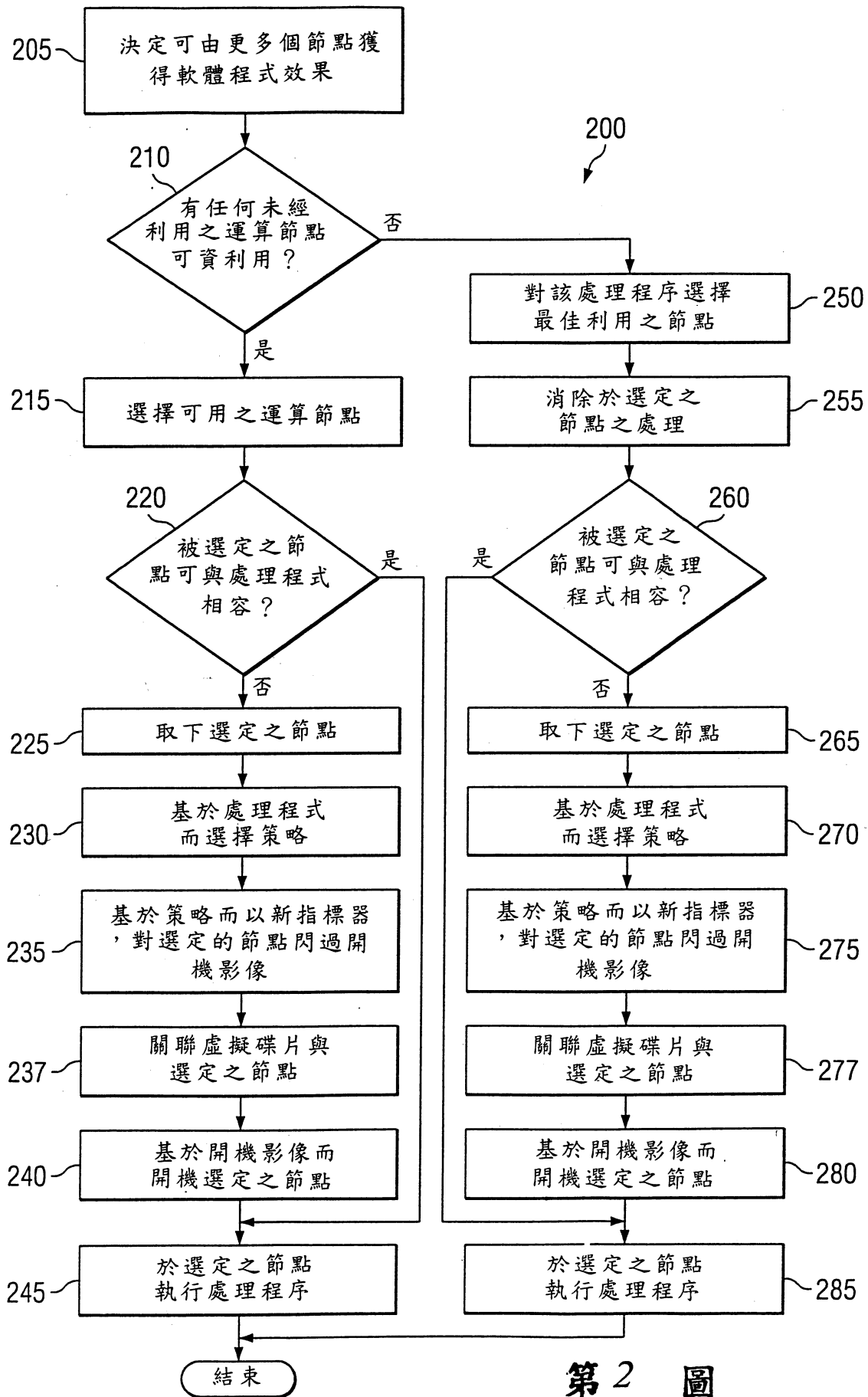
一種電腦叢集虛擬化之方法包括選擇一分散式應用程式。取還一種與該分散式應用程式相關聯之策略。動態選定多數節點之一。然後至少部分基於所取還之策略而復置所選定之節點之開機影像，該開機影像係與該分散式應用程式可相容。然後，一虛擬碟片影像與該節點關聯。然後使用相關聯之虛擬碟片影像，於該復置節點執行至少部分該分散式應用程式。

六、英文發明摘要：

A method for computer cluster virtualization includes selecting a distributed application. A policy associated with the distributed application is retrieved. One of a plurality of nodes is dynamically selected. Then, a boot image of the selected node is reset based, at least in part, on the retrieved policy, with the boot image being compatible with the distributed application. Then, a virtual disk image is associated with the node. At least a portion of the distributed application is then executed on the reset node using the associated virtual disk image.

第 1 圖





第 2 圖

七、指定代表圖：

(一)本案指定代表圖為：第(1)圖。

(二)本代表圖之元件符號簡單說明：

100...分散式運算系統	115...介面
102...伺服器	116...網路
104...管理節點	120...客端，行政管理工作站或本地客端
105...動態開機引擎	122...圖形使用者介面，GUI
108...節點	130...開機影像檔案
109...記憶體	131...開機影像
110...處理器	132...策略，策略表
111...可電子抹除可規劃唯讀記憶體 (EEPROM)	133...虛擬本地碟片影像表
114、114a-b...應用程式	134...虛擬本地碟片影像
	140...應用程式資料

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

十、申請專利範圍：

第94111494號申請案申請專利範圍修正本 96.01.31.

1. 一種電腦叢集虛擬化之方法，該方法包含下列步驟：
 - 選擇一分散式應用程式；
 - 5 取還與該分散式應用程式相關聯之一策略；
 - 動態選定多數節點之一；
 - 至少部分基於所取還之策略，而復置該經選定之節點之一開機影像，該開機影像係與該分散式應用程式可相容；以及
 - 10 至少部分基於該取還之策略，而關聯一虛擬碟片影像與該選定之節點；以及
 - 使用該相關聯之虛擬碟片影像，來與該經復置之節點執行至少部分分散式應用程式。
2. 如申請專利範圍第1項之方法，該應用程式係於多數節點之一子集執行，以及該方法進一步包含：
 - 15 將該節點子集與該取還之策略做比較；以及
 - 基於該項比較而選出多數可相容之開機影像之一。
3. 如申請專利範圍第2項之方法，其中比較該節點子集與該經取還之策略包含：
 - 20 決定於該子集之節點數目；以及
 - 基於於該策略之一鏈路以及該節點之數目來選擇開機影像。
4. 如申請專利範圍第2項之方法，該節點子集各自係與多數可相容之開機影像之一相關聯。

5. 如申請專利範圍第1項之方法，其中動態選擇多數節點之一包含：
- 判定多數節點中之一或多者是否未被一第二分散式應用程式所利用；以及
- 5 響應於該等節點中之至少一者未被利用，選定該等未被利用之節點之一。
6. 如申請專利範圍第5項之方法，響應於並無任何節點未被利用，進一步包含基於下列之一或多項，來選出由該第二分散式應用程式所利用之該等節點之一：
- 10 取還之策略；
- 選定之節點之利用率低；
- 選定之分散式應用程式之優先順位；以及
- 選定之節點與該選定之分散式應用程式之可相容性。
- 15 7. 如申請專利範圍第6項之方法，其中復置該選定之節點之開機影像包含：
- 自動關閉該經選定之節點；
- 復置該選定之節點之開機影像；以及
- 使用復置後之開機影像來重新啟動該經選定之節
- 20 點。
8. 如申請專利範圍第7項之方法，進一步包含於關閉該節點前，結束與該第二分散式應用程式相關聯之任何處理程序。
9. 如申請專利範圍第1項之方法，該策略包含鏈結至開機

影像之多數鏈路，各鏈路係與可與該分散式應用程式相容之多個節點之一相關聯。

10. 如申請專利範圍第1項之方法，該策略包含決定該節點之選擇時序之一或多項參數。
- 5 11. 如申請專利範圍第1項之方法，該開機影像包含儲存於儲存區網路(SAN)之一遠端開機影像。
12. 如申請專利範圍第1項之方法，該節點於復置前係與一第一開機影像相關聯，以及於復置後係與一第二開機影像相關聯，該第一開機影像與該第二開機影像之差異在於下列特徵中之至少一者：
 - 作業系統；
 - 系統組配結構；以及
 - 分散式應用程式參數。
13. 如申請專利範圍第1項之方法，進一步包含：
 - 15 判定多數節點之一故障，該故障節點執行所選定之分散式應用程式之至少一部分；以及
 - 其中選出多數節點之一包含響應於該故障而選出其餘節點之一。
14. 如申請專利範圍第1項之方法，該等多數節點各自包含相同之處理器架構。
- 20 15. 如申請專利範圍第1項之方法，其中選擇該等多數節點之一包含於預定時間選擇該等多數節點之一。
16. 一種儲存有用於電腦叢集虛擬化之軟體的電腦可讀媒體，該軟體可操作來進行下列動作：

選擇一分散式應用程式；

取還與該分散式應用程式相關聯之一策略；

動態選定多數節點之一；

5 至少部分基於所取還之策略，而復置該經選定之節點之一開機影像，該開機影像係與該分散式應用程式可相容；以及

至少部分基於該取還之策略，而關聯一虛擬碟片影像與該選定之節點；以及

10 使用該相關聯之虛擬碟片影像，來與該經復置之節點執行至少部分分散式應用程式。

17. 如申請專利範圍第16項之電腦可讀媒體，其中該應用程式係於多數節點之一子集執行，以及該軟體進一步可操作來：

將該節點子集與該取還之策略做比較；以及

15 基於該項比較而選出多數可相容之開機影像之一。

18. 如申請專利範圍第17項之電腦可讀媒體，其中該可操作來比較該節點子集與該所取還之策略之軟體包含可操作來進行下列動作之軟體：

決定於該子集之節點數目；以及

20 基於於該策略之一鏈路以及該節點之數目來選擇開機影像。

19. 如申請專利範圍第17項之電腦可讀媒體，其中該節點子集各自係與多數可相容之開機影像之一相關聯。

20. 如申請專利範圍第16項之電腦可讀媒體，其中該可操作

來動態選定多數節點之一之軟體包含可操作來進行下列動作之軟體：

判定多數節點中之一或多者是否未被一第二分散式應用程式所利用；以及

5 響應於該等節點中之至少一者未被利用，選定該等未被利用之節點之一。

21. 如申請專利範圍第20項之電腦可讀媒體，其中該軟體響應於並無任何節點未被利用，進一步包含基於下列之一或多項，來選出由該第二分散式應用程式所利用之該等節點之一：

10

取還之策略；

選定之節點之利用率低；

選定之分散式應用程式之優先順位；以及

選定之節點與該選定之分散式應用程式之可相容

15

性。

22. 如申請專利範圍第21項之電腦可讀媒體，其中該可操作來復置所選定之節點之開機影像之軟體包含可操作來進行下列動作之軟體：

自動關閉該經選定之節點；

20

復置該選定之節點之開機影像；以及

使用復置後之開機影像來重新啟動該經選定之節點。

23. 如申請專利範圍第22項之電腦可讀媒體，其中該軟體進一步可操作來於關閉該節點前，結束與該第二分散式應

用程式相關聯之任何處理程序。

24. 如申請專利範圍第16項之電腦可讀媒體，其中該策略包含鏈結至開機影像之多數鏈路，各鏈路係與可與該分散式應用程式相容之多個節點之一相關聯。
- 5 25. 如申請專利範圍第16項之電腦可讀媒體，其中該策略包含決定該節點之選擇時序之一或多項參數。
26. 如申請專利範圍第16項之電腦可讀媒體，其中該開機影像包含儲存於儲存區網路(SAN)之一遠端開機影像。
- 10 27. 如申請專利範圍第16項之電腦可讀媒體，其中該節點於復置前係與一第一開機影像相關聯，以及於復置後係與一第二開機影像相關聯，該第一開機影像與該第二開機影像之差異在於下列特徵中之至少一者：
- 作業系統；
- 系統組配結構；以及
- 15 分散式應用程式參數。
28. 如申請專利範圍第16項之電腦可讀媒體，其中該軟體進一步可操作來：
- 判定多數節點之一故障，該故障節點執行所選定之分散式應用程式之至少一部分；以及
- 20 其中該可操作來選擇多數節點之一之軟體包含可操作來響應於該故障而選定其餘各節點之一的軟體。
29. 如申請專利範圍第16項之電腦可讀媒體，其中該等多數節點各自包含相同之處理器架構。
30. 如申請專利範圍第16項之電腦可讀媒體，其中該可操作

來選定多數節點之一之軟體包含可操作而於預定時間
選定多數節點之一的軟體。

31. 一種電腦叢集虛擬化之系統，包含：

多數運算節點；以及

5 可操作來進行下列動作之一管理節點：

選擇一分散式應用程式；

取還與該分散式應用程式相關聯之一策略；

動態選定多數節點之一；

10 至少部分基於所取還之策略，而復置該經選定之
節點之一開機影像，該開機影像係與該分散式應用程
式可相容；以及

至少部分基於該取還之策略，而關聯一虛擬碟片
影像與該選定之節點；以及

15 使用該相關聯之虛擬碟片影像，來與該經復置之
節點執行至少部分分散式應用程式。

32. 如申請專利範圍第31項之系統，其中該應用程式係於多
數節點之一子集執行，以及該管理節點進一步可操作
來：

將該節點子集與該取還之策略做比較；以及

20 基於該項比較而選出多數可相容之開機影像之一。

33. 如申請專利範圍第32項之系統，其中該可操作來比較該
節點子集與該所取還之策略之管理節點包含可操作來
進行下列動作之管理節點：

決定於該子集之節點數目；以及

基於於該策略之一鏈路以及該節點之數目來選擇開機影像。

34. 如申請專利範圍第32項之系統，其中該節點子集各自係與多數可相容之開機影像之一相關聯。

5 35. 如申請專利範圍第31項之系統，其中該可操作來動態選定多數節點之一之管理節點包含可操作來進行下列動作之管理節點：

判定多數節點中之一或多者是否未被一第二分散式應用程式所利用；以及

10 響應於該等節點中之至少一者未被利用，選定該等未被利用之節點之一。

36. 如申請專利範圍第35項之系統，其中響應於並無任何節點未被利用，進一步包含基於下列之一或多項，來選出由該第二分散式應用程式所利用之該等節點之一：

15 取還之策略；

選定之節點之利用率低；

選定之分散式應用程式之優先順位；以及

選定之節點與該選定之分散式應用程式之可相容性。

20 37. 如申請專利範圍第36項之系統，其中該可操作來復置所選定之節點之開機影像之管理節點包含可操作來進行下列動作之管理節點：

自動關閉該經選定之節點；

復置該選定之節點之開機影像；以及

使用復置後之開機影像來重新啟動該經選定之節點。

5 38. 如申請專利範圍第37項之系統，其中該管理節點進一步可操作來於關閉該節點前，結束與該第二分散式應用程式相關聯之任何處理程序。

39. 如申請專利範圍第31項之系統，其中該策略包含鏈結至開機影像之多數鏈路，各鏈路係與可與該分散式應用程式相容之多個節點之一相關聯。

10 40. 如申請專利範圍第31項之系統，其中該策略包含決定該節點之選擇時序之一或多項參數。

41. 如申請專利範圍第31項之系統，其中該開機影像包含儲存於儲存區網路(SAN)之一遠端開機影像。

15 42. 如申請專利範圍第31項之系統，其中該節點於復置前係與一第一開機影像相關聯，以及於復置後係與一第二開機影像相關聯，該第一開機影像與該第二開機影像之差異在於下列特徵中之至少一者：

作業系統；

系統組配結構；以及

分散式應用程式參數。

20 43. 如申請專利範圍第31項之系統，其中該管理節點進一步可操作來：

判定多數節點之一故障，該故障節點執行所選定之分散式應用程式之至少一部分；以及

其中該可操作來選擇多數節點之一之管理節點包

含可操作來響應於該故障而選定其餘各節點之一的管理節點。

44. 如申請專利範圍第31項之系統，其中該等多數節點各自包含相同之處理器架構。

- 5 45. 如申請專利範圍第31項之系統，其中該可操作來選定多數節點之一之管理節點包含可操作而於預定時間選定多數節點之一的管理節點。