202046119

(19)中華民國智慧財產局

(12)發明說明書公開本　(11)公開編號：TW 202046119 A

(43)公開日： 中華民國 109 (2020) 年 12 月 16 日

(21)申請案號：109115964　(22)申請日： 中華民國 109 (2020) 年 05 月 14 日

(51)Int. Cl.：　*G06F12/0871(2016.01)*

(30)優先權：2019/06/07　美國　16/435,238

(71)申請人：美商斯泰拉斯科技股份有限公司 (美國) STELLUS TECHNOLOGIES, INC.　(US)
美國

(72)發明人：雅庫拉　維傑亞 JAKKULA, VIJAYA (US)；拉米尼尼　西瓦 RAMINENI, SIVA (US)；葛拉普迪　文卡塔 巴努 普拉卡什 GOLLAPUDI, VENKATA BHANU PRAKASH (US)

(74)代理人：林孟閱；盧姵君；陳怡如

申請實體審查：無　申請專利範圍項數：20 項　圖式數：2　共 28 頁

(54)名稱

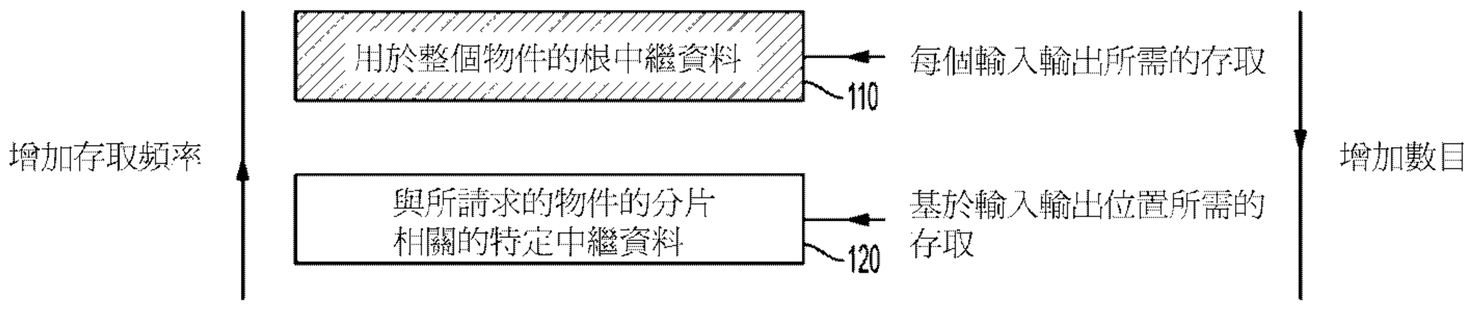
快取表項管理的方法、物件儲存器和電腦可讀媒體

(57)摘要

本揭露提供一種快取表項管理的方法，所述方法包括：由快取管理器確定多個快取表項中的每一者的重要程度；由快取管理器基於所確定的重要程度為快取表項中的每一者指配中繼資料類型；由快取管理器確定快取表項中的每一者的存取頻率；由快取管理器基於所確定的快取表項的存取頻率產生熱圖；以及由快取管理器基於快取表項各自的中繼資料類型及各自的存取頻率確定將快取表項中的至少兩者中的哪一者逐出。也提供一種分散式物件儲存器和非暫時性電腦可讀媒體。

Provided is a method of cache entry management, the method comprising determining, by a cache manager, a level of importance for each of a plurality of cache entries, assigning, by the cache manager, a metadata type for each of the cache entries based upon the determined level of importance, determining, by the cache manager, a frequency of access of each of the cache entries, generating, by the cache manager, a heat map based upon the determined frequency of access of the cache entries, and determining, by the cache manager, which one of at least two of the cache entries to evict based upon the respective metadata types and the respective frequencies of access. A distributed object store and a non-transitory computer readable medium are also provided.

指定代表圖：



符號簡單說明：

110:根中繼資料

120:特定中繼資料/分片中繼資料

【圖1】

202046119

【發明摘要】

【中文發明名稱】快取表項管理的方法、物件儲存器和電腦可讀媒體

【英文發明名稱】METHOD OF CACHE ENTRY MANAGEMENT, OBJECT STORE AND COMPUTER READABLE MEDIUM

【中文】本揭露提供一種快取表項管理的方法，所述方法包括：由快取管理器確定多個快取表項中的每一者的重要程度；由快取管理器基於所確定的重要程度為快取表項中的每一者指配中繼資料類型；由快取管理器確定快取表項中的每一者的存取頻率；由快取管理器基於所確定的快取表項的存取頻率產生熱圖；以及由快取管理器基於快取表項各自的中繼資料類型及各自的存取頻率確定將快取表項中的至少兩者中的哪一者逐出。也提供一種分散式物件儲存器和非暫時性電腦可讀媒體。

【英文】Provided is a method of cache entry management, the method comprising determining, by a cache manager, a level of importance for each of a plurality of cache entries, assigning, by the cache manager, a metadata type for each of the cache entries based upon the determined level of importance, determining, by the cache manager, a frequency of access of each of the cache entries, generating, by the cache manager, a heat map based upon the

determined frequency of access of the cache entries, and determining, by the cache manager, which one of at least two of the cache entries to evict based upon the respective metadata types and the respective frequencies of access. A distributed object store and a non-transitory computer readable medium are also provided.

【指定代表圖】圖1。

【代表圖之符號簡單說明】

110：根中繼資料

120：特定中繼資料/分片中繼資料

【特徵化學式】

無

【發明說明書】

【中文發明名稱】快取表項管理的方法、物件儲存器和電腦可讀媒體

【英文發明名稱】METHOD OF CACHE ENTRY MANAGEMENT, OBJECT STORE AND COMPUTER READABLE MEDIUM

【技術領域】

【0001】 本揭露的實施例的一個或多個方面一般來說涉及快取表項管理的方法，以及被配置成實行所述方法的一種分散式物件儲存器。

【先前技術】

【0002】 為了實現操作，包括記憶體驅動器的網路的“分散式物件儲存器（distributed object store）”一般來說使用複雜的中繼資料來管理所儲存的物件資料的不同方面。由中繼資料管理的幾個方面可包括物件的增長或收縮、資料位置、與物件對應的資料管理策略（例如，複製副本是否被授權）、與分散式物件儲存器的不可變性質對應的物件版本化（object-versioning）、資料儲存器上給定物件標識（identity，ID）的物件資料的位置跟蹤等。此外，物件可能不得不在分散式物件儲存器內四處移動（例如，在分散式系統中發生局部故障的情況下）。

【0003】 分散式物件儲存器將通常會儲存對應於同一資料物件的多個中繼資料物件，以有助於管理資料物件的不同方面。此外，

分散式物件儲存器可採用中繼資料的快取來隨著分散式物件儲存器的大小的按比例縮放而保持高性能，因為由於存取快取一般來說比存取其中駐留有物件的媒體快得多的事實，因此通過簡單地對頻繁存取的中繼資料進行快取可改善性能。

【0004】　然而，當按比例縮放時，快取的有效性隨著工作集（例如，資料物件及中繼資料物件的數目及大小）的增加而降低。也就是說，當“熱”（例如，頻繁存取的）項重複地相互逐出時，快取可能變得不太有效，從而導致被稱為“快取抖動（cache thrashing）”的現象。隨著工作集增加，快取抖動可能會影響輸入/輸出（input/output，I/O）路徑的性能，且因此對於大型分散式物件儲存器（large scale distributed object store）來說，這可能是更為重要的問題。

【發明內容】

【0005】　本文闡述的實施例通過基於多種因素確定標記哪些快取表項用於逐出以提供對快取表項管理的改善。

【0006】　根據本揭露的一個實施例，提供一種快取表項管理的方法，所述方法包括：由快取管理器確定多個快取表項中的每一者的重要程度；由所述快取管理器基於所確定的所述重要程度為所述快取表項中的每一者指配中繼資料類型；由所述快取管理器確定所述快取表項中的每一者的存取頻率；由所述快取管理器基於所確定的所述快取表項的所述存取頻率產生熱圖；以及由所述快取管理器基於所述快取表項各自的所述中繼資料類型及各自的所述存取頻率確定將所述快取表項中的至少兩者中的哪一者逐出。

【0007】　所述方法可更包括：由所述快取管理器使所述熱圖的與所述快取表項中具有較低重要程度的快取表項對應的部分比所述熱圖的與所述快取表項中具有較高重要程度的快取表項對應的其他部分衰變得快。

【0008】　所述方法可更包括：一旦所述熱圖的與所述快取表項中的一者對應的部分達到零，便由所述快取管理器逐出所述快取表項中的所述一者。

【0009】　所述方法可更包括：由所述快取管理器使熱圖的與物件的分片對應的部分比所述熱圖的與所述物件的整體對應的部分衰變得快。

【0010】　所述方法可更包括：由所述快取管理器對所述快取進行移行，且由所述快取管理器將所述快取表項中與低於參考重要程度的重要程度對應的所有快取表項逐出。

【0011】　所述方法可更包括：一旦所使用的快取資源的量達到參考程度，便由所述快取管理器觸發所述快取的所述移行。

【0012】　所述方法可更包括：由所述快取管理器將最高重要程度指配給代表物件的整體的根中繼資料（root metadata），且由所述快取管理器將較低重要程度指配給與所述物件相關的用於分片中繼資料的其他中繼資料，所述分片中繼資料代表所述對象的分片。

【0013】　根據本揭露的另一實施例，提供一種被配置成允許在與工作集的大小一起按比例縮放的同時進行快取管理的分散式物件儲存器，所述分散式物件儲存器包括快取，所述快取被配置成由快取管理器進行管理，其中所述快取管理器被配置成：確定多個快取表項中的每一者的重要程度；基於所確定的所述重要程度為

所述快取表項中的每一者指配中繼資料類型；確定所述快取表項中的每一者的存取頻率；基於所確定的所述快取表項的所述存取頻率產生熱圖；以及基於所述快取表項各自的所述中繼資料類型及各自的所述存取頻率確定將所述快取表項中的至少兩者中的哪一者逐出。

【0014】 所述快取管理器可更被配置成使所述熱圖的與所述快取表項中具有較低重要程度的快取表項對應的部分比所述熱圖的與所述快取表項中具有較高重要程度的快取表項對應的其他部分衰變得快。

【0015】 所述快取管理器可更被配置成一旦所述熱圖的與所述快取表項中的一者對應的部分達到零，便逐出所述快取表項中的所述一者。

【0016】 所述快取管理器可更被配置成使熱圖的與物件的分片對應的部分比所述熱圖的與所述物件的整體對應的部分衰變得快。

【0017】 所述快取管理器可更被配置成對所述快取進行移行，且將所述快取表項中與低於參考重要程度的重要程度對應的所有快取表項逐出。

【0018】 所述快取管理器可更被配置成一旦所使用的快取資源的量達到參考程度，便觸發所述快取管理器對所述快取進行移行。

【0019】 所述快取管理器可更被配置成將最高重要程度指配給代表物件的整體的根中繼資料，且將較低重要程度指配給與所述物件相關的用於分片中繼資料的其他中繼資料，所述分片中繼資料代表所述對象的分片。

【0020】　根據本揭露的又一實施例，提供一種實施在用於顯示器件的顯示轉碼器上的非暫時性電腦可讀媒體，所述非暫時性電腦可讀媒體實施在分散式物件儲存器上，所述非暫時性電腦可讀媒體具有電腦代碼，所述電腦代碼在處理器上執行時通過所述分散式物件儲存器的快取的快取管理器實施一種快取管理方法，所述方法包括：由所述快取管理器確定多個快取表項中的每一者的重要程度；由所述快取管理器基於所確定的所述重要程度為所述快取表項中的每一者指配中繼資料類型；由所述快取管理器確定所述快取表項中的每一者的存取頻率；由所述快取管理器基於所確定的所述快取表項的所述存取頻率產生熱圖；以及由所述快取管理器基於所述快取表項各自的所述中繼資料類型及各自的所述存取頻率確定將所述快取表項中的至少兩者中的哪一者逐出。

【0021】　所述電腦代碼在由所述處理器執行時可更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：由所述快取管理器使所述熱圖的與所述快取表項中具有較低重要程度的快取表項對應的部分比所述熱圖的與所述快取表項中具有較高重要程度的快取表項對應的其他部分衰變得快。

【0022】　所述電腦代碼在由所述處理器執行時可更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：一旦所述熱圖的與所述快取表項中的一者對應的部分達到零，便由所述快取管理器逐出所述快取表項中的所述一者。

【0023】　所述電腦代碼在由所述處理器執行時可更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：由所述快取管理器使熱圖的與物件的分片對應的部分比所述熱圖

的與所述物件的整體對應的部分衰變得快。

【0024】 所述電腦代碼在由所述處理器執行時可更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：由所述快取管理器對所述快取進行移行，且由所述快取管理器將所述快取表項中與低於參考重要程度的重要程度對應的所有快取表項逐出。

【0025】 所述電腦代碼在由所述處理器執行時可更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：一旦所使用的快取資源的量達到參考程度，便由所述快取管理器觸發所述快取的所述移行。

【0026】 因此，本揭露實施例的自調整快取（self-tuning cache）允許儲存裝置隨著儲存裝置大小的增加而在性能方面按比例縮放。

【圖式簡單說明】

【0027】 結合附圖，根據以下對實施例的說明，上述和/或其他方面將變得顯而易見且更容易理解，在附圖中：

圖 1 是繪示根據本揭露的一個或多個實施例的分散式物件儲存器的中繼資料的一般性質的方塊圖。

圖 2A 是繪示根據本揭露的一個或多個實施例的用於對相應快取表項進行快取的快取的兩個不同分區的方塊圖。

圖 2B 是繪示根據本揭露的一個或多個實施例，在從分區中的一個分區逐出快取表項中的一些快取表項之後，圖 2A 的快取的所述兩個不同分區的方塊圖。

【實施方式】

【0028】　通過參照實施例的詳細說明及附圖，可更容易理解本發明概念的特徵及其實現方法。在下文中，將參照附圖更詳細地闡述實施例。然而，所闡述的實施例可被實施成各種不同的形式，而不應被視為僅限於本文中所例示的實施例。確切來說，提供這些實施例是作為實例來使本揭露將透徹及完整，並將向所屬領域中的技術人員充分傳達本發明概念的各個方面及特徵。因此，可不再闡述對於所屬領域中的普通技術人員完整地理解本發明概念的各個方面及特徵而言並非必需的製程、元件及技術。除非另外注明，否則在所有附圖及本書面說明通篇中相同的參考編號表示相同的元件，且因此，可不再對其予以重複說明。此外，可能未示出與實施例的說明無關的部件，以使說明清楚。在圖式中，為了清晰起見，可放大元件、層及區的相對大小。

【0029】　在本文中參照剖視圖闡述各種實施例，所述剖視圖為實施例和/或中間結構的示意性例示。因此，預期會因例如製造技術和/或容差而導致相對於例示形狀的變化。此外，本文所揭露的具體結構說明或功能說明僅是例示性的，目的在於闡述根據本揭露概念的實施例。因此，本文所揭露的實施例不應被視為僅限於各個區的特定例示形狀，而是應包含由例如製造引起的形狀偏差。舉例來說，被例示為矩形的植入區通常應具有圓形特徵或曲線特徵和/或在其邊緣處存在植入濃度的梯度而非從植入區到非植入區為二元改變。同樣地，通過植入而形成的掩埋區可在所述掩埋區與在進行植入時所經過的表面之間的區中引起一些植入。因此，

圖式中所例示的區本質上為示意性的且其形狀並非旨在例示器件的區的實際形狀且並非旨在進行限制。另外，如所屬領域中的技術人員將認識到的，所闡述的實施例可以各種不同的方式修改，所有這些都不背離本揭露的精神或範圍。

【0030】　如上所述，隨著工作集增加，由重複相互逐出的熱資料項目（hot data item）導致的快取抖動會對快取的性能產生負面影響。

【0031】　傳統上，可按比例縮放的分散式物件儲存器（scalable distributed object store）的性能取決於分散式物件儲存器在存取對應於物件的中繼資料時的性能有多高。有效地存取中繼資料通常需要對物件的頻繁存取的中繼資料進行快取，以便可相對快速地存取中繼資料。然而，用於在可按比例縮放的分散式對象儲存器中對更頻繁存取的中繼資料進行快取的快取通常面臨挑戰。

【0032】　舉例來說，可按比例縮放的分散式對象儲存器可期望地針對數十億個對象進行按比例縮放，且更可期望地允許成千上萬個使用不同應用的併發用戶端，各種用戶端中的每一者潛在地針對用戶端的不同相應工作負載存取不同的資料集。然而，由於從具有不同要求的併發存取中產生的大的工作集，因此對應於存取要求的可按比例縮放性（scalability）需求可能會使快取的效率變低。

【0033】　此外，簡單地增加快取的大小僅僅會局部地緩解由於可被存取的大量物件以及由於物件中的每一者潛在地具有可被快取的多個相關聯的中繼資料物件而導致的問題。舉例來說，I/O 可能無法成功地實行查找操作以在快取中找到中繼資料，且由於快取

的大的大小，因此必須對記憶體儲存裝置進行存取來找到中繼資料。

【0034】 因此，儘管傳統的快取可能有助於解決分散式物件儲存器的 I/O 路徑的性能，但是隨著分散式物件儲存器按比例縮放，從對更頻繁存取的中繼資料進行快取實現的好處可能會由於“快取抖動”被抵消，“快取抖動”可能在併發存取對快取資源相互進行競爭時發生。本揭露的實施例通過減少快取抖動的不希望的影響來解決這個挑戰。

【0035】 本揭露的實施例提供一種自學習及自調整快取，所述自學習及自調整快取能夠做出關於決定何時快取中繼資料、快取哪個中繼資料、逐出哪個中繼資料以及已被快取的中繼資料應該保留多長時間的智慧選擇。

【0036】 本揭露的實施例更提供一種快取，所述快取知道快取表項的類型，且能夠基於快取表項的類型來判斷重要程度。

【0037】 因此，通過向相應的快取表項指配不同的重要性價值，快取能夠朝隨著工作集增加而指配有高重要程度的高價值快取表項偏置或使所述高價值快取表項優先。

【0038】 圖 1 是繪示根據本揭露的一個或多個實施例的分散式物件儲存器的中繼資料的一般性質的方塊圖。

【0039】 參照圖 1，各種類型的中繼資料可作為中繼資料物件儲存在分散式物件儲存器的快取中。一種類型的中繼資料是可代表整個資料物件的根中繼資料 110。根中繼資料 110 一般來說將被試圖存取物件的某個方面的任何 I/O 執行緒存取。舉例來說，即使 I/O 執行緒試圖僅存取物件的分片，但 I/O 執行緒仍將存取根中繼

資料 110。

【0040】　相反，其他特定中繼資料（例如，分片中繼資料）120 僅特定地與所請求的物件的分片相關，所述物件的整體由根中繼資料 110 代表。也就是說，如果 I/O 執行緒不試圖存取整個資料物件，而是僅試圖存取物件的分片，則 I/O 執行緒可存取對應的分片中繼資料 120，以檢索僅對應於所尋找的物件分片的資訊。因此，分片中繼資料 120 代表物件的各個分片，且可用於分別管理特定於對應的物件分片的屬性。

【0041】　I/O 執行緒對於物件的存取利用了對與作為整體的物件對應的根中繼資料 110 的存取，接著是根據所請求的 I/O 存取的位置及性質進行的對與物件分片相關聯的分片中繼資料 120 的存取。因此，對應於分片中繼資料 120 的中繼資料物件通常會相對多，因為一般來說將會存在與由分片中繼資料 120 管理的單個資料物件對應的許多相應方面。相反，僅根中繼資料 110 的一個例子被用於管理作為整體的資料物件。

【0042】　然而，分片中繼資料 120 的每一單獨例子一般來說遠不如根中繼資料 110 被存取得頻繁，因為對資料物件的某個方面關注的各種 I/O 執行緒一般來說將與物件的一些數目的單獨分片無關。也就是說，由於針對資料物件實行查找操作的所有 I/O 執行緒一般來說將存取根中繼資料 110，因此分片中繼資料 120 的更多例子將不如根中繼資料 110 被單獨存取得頻繁。

【0043】　本揭露的實施例利用物件的中繼資料的固有性質來對快取進行自調整以確保即使當工作集隨著用戶端連接及併發存取的增加而按比例縮放時，快取也保持有效。為了實現這一點，快

取可使用快取管理器來利用中繼資料的類型進行自調整。也就是說，本揭露的實施例可通過利用由物件的各個方面產生的固有中繼資料特性來實現大型分散式物件儲存器中的有效快取。

【0044】 圖 2A 是繪示根據本揭露的一個或多個實施例的用於對相應的快取表項進行快取的快取的兩個不同分區的方塊圖，且圖 2B 是繪示根據本揭露的一個或多個實施例，在從分區中的一個分區逐出快取表項中的一些快取表項之後，圖 2A 的快取的所述兩個不同分區的方塊圖。

【0045】 參照圖 2A，基於與要儲存在其中的快取表項的中繼資料對應的中繼資料類型來對快取 200 進行分區。應注意，圖 2A 及圖 2B 中例示的兩個分區僅僅是用於例示本揭露的方面的實例。在其他實施例中，可存在許多這樣的分區（即，多於兩個）。在本揭露的實施方式中，可存在幾個分區用於採用所揭露的實施例的方面來實現有效的快取。在本實施例中，快取 200 包括：根物件中繼資料分區 210，用於處理指配有與根中繼資料 110 對應的中繼資料類型的快取表項；以及物件分片中繼資料分區 220，用於處理指配有與僅對應於物件的分片的分片中繼資料 120 對應的中繼資料類型的快取表項。

【0046】 與物件分片對應的分片中繼資料 120 的存取模式使得只有當 I/O 存取的位置落於相應的分片內時才需要進行存取。因此，隨著 I/O 活動在關於物件的範圍內從一個分片移動到另一分片，I/O 存取不再需要前一個物件分片的分片中繼資料 120。然而，由於對物件進行的每個 I/O 存取都需要獨立於物件中的 I/O 位置來存取根中繼資料 110，因此代表整個物件的根中繼資料 110 的存取模

式是不同的。因此，如前所述，根中繼資料 110 一般來說將比分片中繼資料 120 被存取得頻繁。

【0047】 根據本實施例，作為實例，快取管理器將兩種不同的“中繼資料類型”中的一者指配給快取表項。然而，應注意，在本揭露的其他實施例中，可使用更多數目的相應類型的中繼資料（例如，3 個或更多個）。因此，在本實例中，對應於根中繼資料 110 的快取表項指配有第一中繼資料類型（例如，“類型 0”），且對應於分片中繼資料 120 的快取表項指配有第二中繼資料類型（例如，“類型 1”）。

【0048】 通過將不同的中繼資料類型指配給不同種類的中繼資料，能夠形成層次（hierarchy）。因此，快取 200 可進行自調整以通過分配額外的資源來實現與指配有高重要程度的快取表項對應的高價值資料的快取，從而在工作集增加時保持有效。由於用於自調整的計算是基於恒定大小輸入（例如，基於中繼資料的類型），而不是基於可能變得非常大的工作集的大小，因此自調整機制隨著物件的工作集中物件的數目而按比例縮放。

【0049】 此外，如以下進一步論述，可將對於每一分區 210、220 的逐出策略調整成與分區相關聯的特定中繼資料類型固有的或一般來說對應的預期熱度。這些及其他逐出操作可通過以低優先順序運行的專用後臺執行緒來實現，以確保隨著工作集增加，快取保持有效。

【0050】 參照圖 2B，隨著快取 200 的資源被消耗，快取 200 可確定應該逐出的快取表項中的一些快取表項。快取 200 可接著開始對各種快取表項進行移行，以便它可確定要將哪些快取表項逐

出。

【0051】 舉例來說，快取 200 可使用參考容量程度（例如，90%滿）來確定何時開始將其中快取的快取表項中的一些快取表項逐出。然而，應注意，根據本揭露的實施例，可使用不同的容量程度及不同的逐出觸發條件（例如，快取可在管理員的指導下開始逐出）。

【0052】 因此，快取 200 可基於指配給快取表項的中繼資料類型來確定將哪些快取表項逐出來釋放額外的資源。舉例來說，快取 200 可決定逐出特定中繼資料類型、或者特定中繼資料類型或“較低”（例如，特定重要程度或較低）的快取表項中的每一者。另外，快取 200 可決定將特定中繼資料類型中的每一快取表項逐出，直到發生觸發（例如，直到達到快取 200 的給定容量程度，例如 80%滿，或者直到給定百分比的容量被逐出，例如 10%）。

【0053】 在本實例中，快取 200 可確定指配有“類型 0”中繼資料類型的快取表項一般來說比指配有“類型 1”中繼資料類型的快取表項重要。即使分別指配有“類型 0”中繼資料類型及“類型 1”中繼資料類型的兩個特定快取表項以近似相等的頻率被存取，這也可能是真的。因此，快取 200 可有效地將其自身朝更高價值中繼資料偏置。

【0054】 因此，通過比較圖 2A 與圖 2B 可看出，快取 200 可從對象分片中繼資料分區 220 逐出包括分片中繼資料 120 的“類型 1”快取表項中的一些“類型 1”快取表項，從而釋放可用於在根中繼資料分區 210 中產生包括根中繼資料 110 的附加快取表項的資源。

【0055】 此外，如上所述，包括快取 200 的分散式物件儲存器可

通過在適當時對低價值快取表項進行快取來有效地處理 I/O 熱點，而不管指配給快取表項的中繼資料類型如何。這可通過考慮附加參數來實現。

【0056】 舉例來說，快取 200 更可採用快取表項的熱圖來保持跟蹤對快取表項的存取頻率。這使得本揭露的實施例能夠學習如何基於中繼資料物件被存取的頻率來在相同中繼資料類型的不同中繼資料物件中進行區分。

【0057】 快取管理器可維持被快取的每一中繼資料物件的熱圖。快取管理器可接著針對對於對應的中繼資料物件的每次存取增加熱圖的一部分。然而，熱圖的衰變速率可基於指配給中繼資料的中繼資料類型來決定。

【0058】 舉例來說，熱圖的與指配有低價值的分片中繼資料對應的部分可能衰變得更快，而熱圖的與更重要的根中繼資料對應的部分可能衰變得更慢。也就是說，儘管可能發生熱圖衰變，但是可基於快取表項的中繼資料類型來調節這種衰變。舉例來說，根中繼資料可能更熱，因為對物件的任何分片進行存取的所有 I/O 執行緒都將存取根中繼資料。如熱圖上所呈現，一旦對應於中繼資料物件的物件溫度達到零，便可從快取逐出中繼資料物件。

【0059】 應注意，低價值的中繼資料（例如，指配有與較低重要程度相關聯的中繼資料類型的中繼資料）可被視為具有高價值。低價值的中繼資料被視為具有高價值的實例是突然被頻繁存取的新聞故事。即使當與新聞故事相關聯的中繼資料與低價值相關聯時，其存取頻率也會使快取管理器將中繼資料視為具有更高的價值。

【0060】　舉例來說，通過考慮熱圖，如果中繼資料物件非常熱，則儘管與低價值相關聯，但快取管理器可決定在中繼資料否則將被逐出時將中繼資料保持在快取中。這確保了當工作集的大小增加時，快取會自調整以區分更有價值的中繼資料與較低價值的中繼資料，從而通過僅快取高價值的中繼資料來保持有效性。同時，快取對 I/O 熱點保持有效，因為它將對用於作為此類熱點的目標的物件分片的較低價值的中繼資料進行快取。

【0061】　因此，本揭露的實施例既使用中繼資料的中繼資料類型以具有中繼資料類別的層次，且也使用熱圖以保持對本地熱點的敏感。

【0062】　如上所述，本揭露的實施例為快取提供可按比例縮放的自調整機制，提供隨著工作集的大小而按比例縮放的可按比例縮放的快取性能，且通過隨著工作集的大小的增加而管理 I/O 熱點的性能來保持有效。因此，本揭露的實施例通過確保 I/O 路徑性能不會隨著工作集的大小按比例縮放而劣化，且通過允許 I/O 路徑性能即使在工作集的大小增加的情況下也能處理 I/O 熱點，從而形成快取的高效自調整機制，來提供對分散式物件儲存器技術的改善。

【0063】　在說明中，出於闡釋目的，闡述各種具體細節來提供對各種實施例的透徹理解。然而，顯而易見的是，可不使用這些具體細節或者可使用一種或多種等效配置來實踐各種實施例。在其他情況下，以方塊圖形式示出眾所周知的結構及器件以避免不必要地使各種實施例模糊不清。

【0064】　本文所用術語僅是出於闡述特定實施例的目的而並非

旨在對本發明進行限制。除非上下文清楚地另外指明，否則本文所用單數形式“一（a 及 an）”旨在也包括複數形式。更應理解，當在本說明書中使用用語“包括（comprises、comprising）”、“具有（have、having）”及“包含（includes、including）”時，是指明所陳述特徵、整數、步驟、操作、元件和/或元件的存在，但不排除一個或多個其他特徵、整數、步驟、操作、元件、元件和/或其群組的存在或添加。本文所用用語“和/或”包括相關聯的列出項中的一個或多個項的任意及所有組合。

【0065】　當某一實施例可被以不同方式實施時，特定製程次序可與所闡述的次序不同地實行。舉例來說，兩個連續闡述的製程可實質上同時實行或以與所闡述的次序相反的次序實行。

【0066】　根據本文所述本揭露的實施例的電子器件或電氣器件和/或任何其他相關器件或元件可利用任何適當的硬體、韌體（例如，專用積體電路（application-specific integrated circuit））、軟體、或軟體、韌體、及硬體的組合來實施。舉例來說，可將這些器件的各種元件形成在一個積體電路（integrated circuit，IC）晶片上或單獨的積體電路晶片上。此外，可將這些器件的各種元件實施在柔性印刷電路膜（flexible printed circuit film）、載帶封裝（tape carrier package，TCP）、印刷電路板（printed circuit board，PCB）上、或形成在一個基板上。此外，這些器件的各種元件可為在一個或多個計算器件中由一個或多個處理器運行、執行電腦程式指令並與用於實行本文所述各種功能性的其他系統元件進行交互的過程或執行緒。電腦程式指令儲存在可在使用例如（舉例來說）隨機存取記憶體（random access memory，RAM）等標準記憶體器

件的計算器件中實施的記憶體中。電腦程式指令也可儲存在例如（舉例來說）壓縮磁碟唯讀記憶體（compact disc read only memory，CD-ROM）、快閃記憶體驅動器（flash drive）或類似元件等其他非暫時性電腦可讀媒體中。另外，所屬領域中的技術人員應知，在不背離本揭露實施例的精神及範圍的條件下，可將各種計算器件的功能性組合或整合成單一的計算器件，或者可使一特定計算器件的功能性跨越一個或多個其他計算器件分佈。

【0067】　除非另外定義，否則本文所用所有用語（包括技術及科學用語）的含義均與本發明概念所屬領域中的普通技術人員所通常理解的含義相同。更應理解，用語（例如在常用詞典中所定義的用語）應被解釋為具有與其在相關技術的上下文和/或本說明書中的含義一致的含義，且除非在本文中明確定義，否則不應將其解釋為具有理想化或過於正式的意義。

【0068】　本文中已揭露了實施例，且儘管採用了特定用語，但是它們僅用於一般意義及闡述性意義或將以一般意義及闡述性意義解釋，而並非用於限制目的。在一些情況下，除非例如另外指明，否則對於所屬領域中的普通技術人員來說，在提交本申請時，結合特定實施例闡述的特徵、特性和/或元件可單獨使用，或者與結合其他實施例闡述的特徵、特性和/或元件結合使用。因此，所屬領域中的技術人員應理解，在不背離如以上權利要求書中所提出的本揭露的精神及範圍的條件下可作出形式及細節上的各種改變，其中權利要求的功能等效形式包含在本文中。

【符號說明】

【0069】

110：根中繼資料

120：特定中繼資料/分片中繼資料

200：快取

210：根對象中繼資料分區/分區

220：對象分片中繼資料分區/分區

【發明申請專利範圍】

【請求項1】 一種快取表項管理的方法，所述方法包括：

由快取管理器確定多個快取表項中的每一者的重要程度；

由所述快取管理器基於所確定的所述重要程度為所述多個快取表項中的每一者指配中繼資料類型；

由所述快取管理器確定所述多個快取表項中的每一者的存取頻率；

由所述快取管理器基於所確定的所述多個快取表項的所述存取頻率產生熱圖；以及

由所述快取管理器基於所述多個快取表項各自的所述中繼資料類型及各自的所述存取頻率確定將所述多個快取表項中的至少兩者中的哪一者逐出。

【請求項2】 如請求項1所述的方法，更包括：由所述快取管理器使所述熱圖的與具有較低重要程度的快取表項對應的部分比所述熱圖的與具有較高重要程度的快取表項對應的其他部分衰變得快。

【請求項3】 如請求項1所述的方法，更包括：一旦所述熱圖的與所述多個快取表項中的一者對應的部分達到零，便由所述快取管理器逐出所述多個快取表項中的所述一者。

【請求項4】 如請求項1所述的方法，更包括：由所述快取管理器使熱圖的與物件的分片對應的部分比所述熱圖的與所述物件的整體對應的部分衰變得快。

【請求項5】 如請求項1所述的方法，更包括：由所述快取管理器對快取進行移行，且由所述快取管理器將與低於參考重要程度的重要程度對應的所有快取表項逐出。

【請求項6】 如請求項5所述的方法，更包括：一旦所使用的快取資源的量達到參考程度，便由所述快取管理器觸發所述快取的移行。

【請求項7】 如請求項1所述的方法，更包括：由所述快取管理器將最高重要程度指配給代表物件的整體的根中繼資料，且由所述快取管理器將較低重要程度指配給與所述物件相關的用於分片中繼資料的其他中繼資料，所述分片中繼資料代表所述對象的分片。

【請求項8】 一種被配置成允許在與工作集的大小按比例縮放的同時進行快取管理的分散式物件儲存器，所述分散式物件儲存器包括快取，所述快取被配置成由快取管理器進行管理，其中所述快取管理器被配置成：

確定多個快取表項中的每一者的重要程度；

基於所確定的所述重要程度為所述多個快取表項中的每一者指配中繼資料類型；

確定所述多個快取表項中的每一者的存取頻率；

基於所確定的所述多個快取表項的所述存取頻率產生熱圖；以及

基於所述多個快取表項各自的所述中繼資料類型及各自的所述存取頻率確定將所述多個快取表項中的至少兩者中的哪一者逐

出。

【請求項9】 如請求項8所述的分散式物件儲存器，其中所述快取管理器更被配置成使所述熱圖的與具有較低重要程度的快取表項對應的部分比所述熱圖的與具有較高重要程度的快取表項對應的其他部分衰變得快。

【請求項10】 如請求項8所述的分散式物件儲存器，其中所述快取管理器更被配置成一旦所述熱圖的與所述多個快取表項中的一者對應的部分達到零，便逐出所述多個快取表項中的所述一者。

【請求項11】 如請求項8所述的分散式物件儲存器，其中所述快取管理器更被配置成使熱圖的與物件的分片對應的部分比所述熱圖的與所述物件的整體對應的部分衰變得快。

【請求項12】 如請求項8所述的分散式物件儲存器，其中所述快取管理器更被配置成對所述快取進行移行，且將與低於參考重要程度的重要程度對應的所有快取表項逐出。

【請求項13】 如請求項12所述的分散式物件儲存器，其中所述快取管理器更被配置成一旦所使用的快取資源的量達到參考程度，便觸發所述快取管理器對所述快取進行移行。

【請求項14】 如請求項8所述的分散式物件儲存器，其中所述快取管理器更被配置成將最高重要程度指配給代表物件的整體的根中繼資料，且將較低重要程度指配給與所述物件相關的用於分片中繼資料的其他中繼資料，所述分片中繼資料代表所述對象的分片。

【請求項15】　一種實施在分散式物件儲存器上的非暫時性電腦可讀媒體，所述非暫時性電腦可讀媒體具有電腦代碼，所述電腦代碼在處理器上執行時通過所述分散式物件儲存器的快取的快取管理器實施一種快取管理方法，所述方法包括：

由所述快取管理器確定多個快取表項中的每一者的重要程度；

由所述快取管理器基於所確定的所述重要程度為所述多個快取表項中的每一者指配中繼資料類型；

由所述快取管理器確定所述多個快取表項中的每一者的存取頻率；

由所述快取管理器基於所確定的所述多個快取表項的所述存取頻率產生熱圖；以及

由所述快取管理器基於所述多個快取表項各自的所述中繼資料類型及各自的所述存取頻率確定將所述多個快取表項中的至少兩者中的哪一者逐出。

【請求項16】　如請求項15所述的非暫時性電腦可讀媒體，其中所述電腦代碼在由所述處理器執行時更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：由所述快取管理器使所述熱圖的與具有較低重要程度的快取表項對應的部分比所述熱圖的與具有較高重要程度的快取表項對應的其他部分衰變得快。

【請求項17】　如請求項15所述的非暫時性電腦可讀媒體，其中所述電腦代碼在由所述處理器執行時更通過以下方式實施所述分

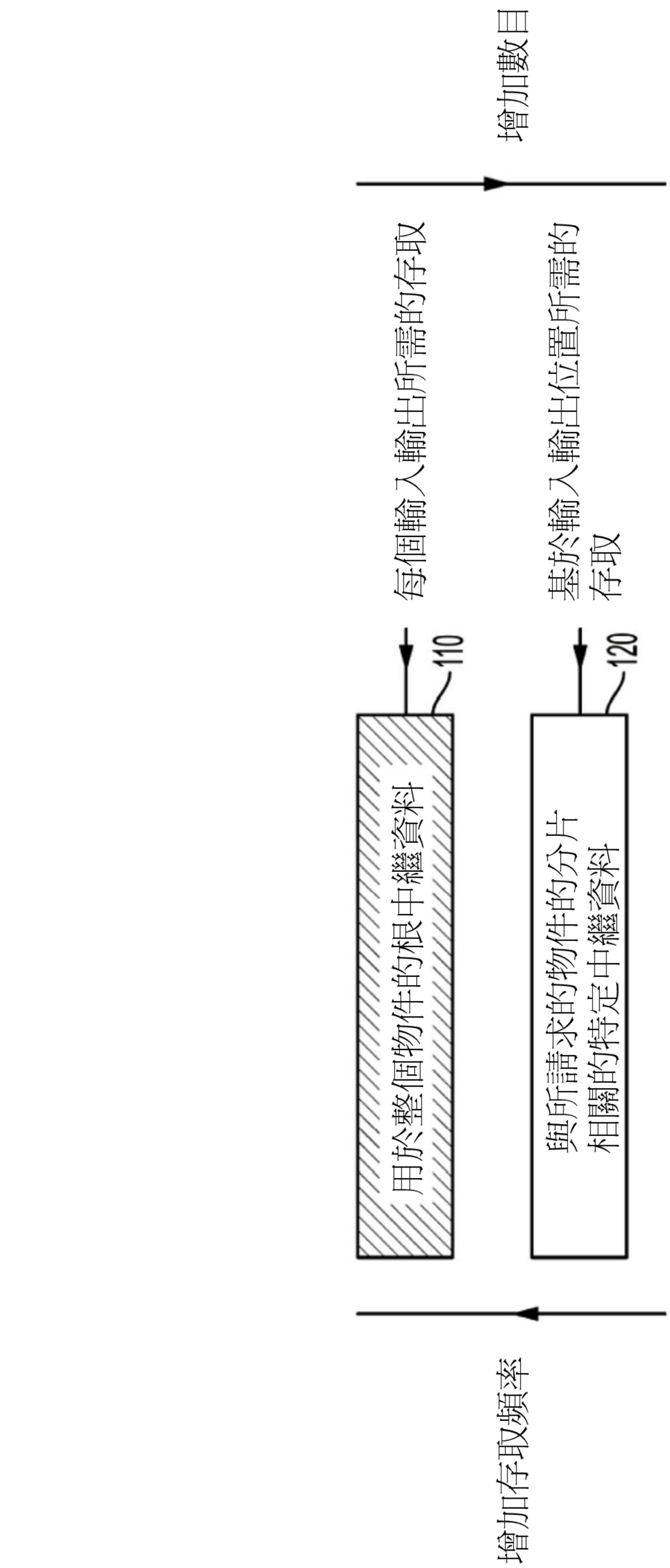
散式物件儲存器的所述快取的所述快取管理方法：一旦所述熱圖的與所述多個快取表項中的一者對應的部分達到零，便由所述快取管理器逐出所述多個快取表項中的所述一者。

【請求項18】　如請求項15所述的非暫時性電腦可讀媒體，其中所述電腦代碼在由所述處理器執行時更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：由所述快取管理器使熱圖的與物件的分片對應的部分比所述熱圖的與所述物件的整體對應的部分衰變得快。

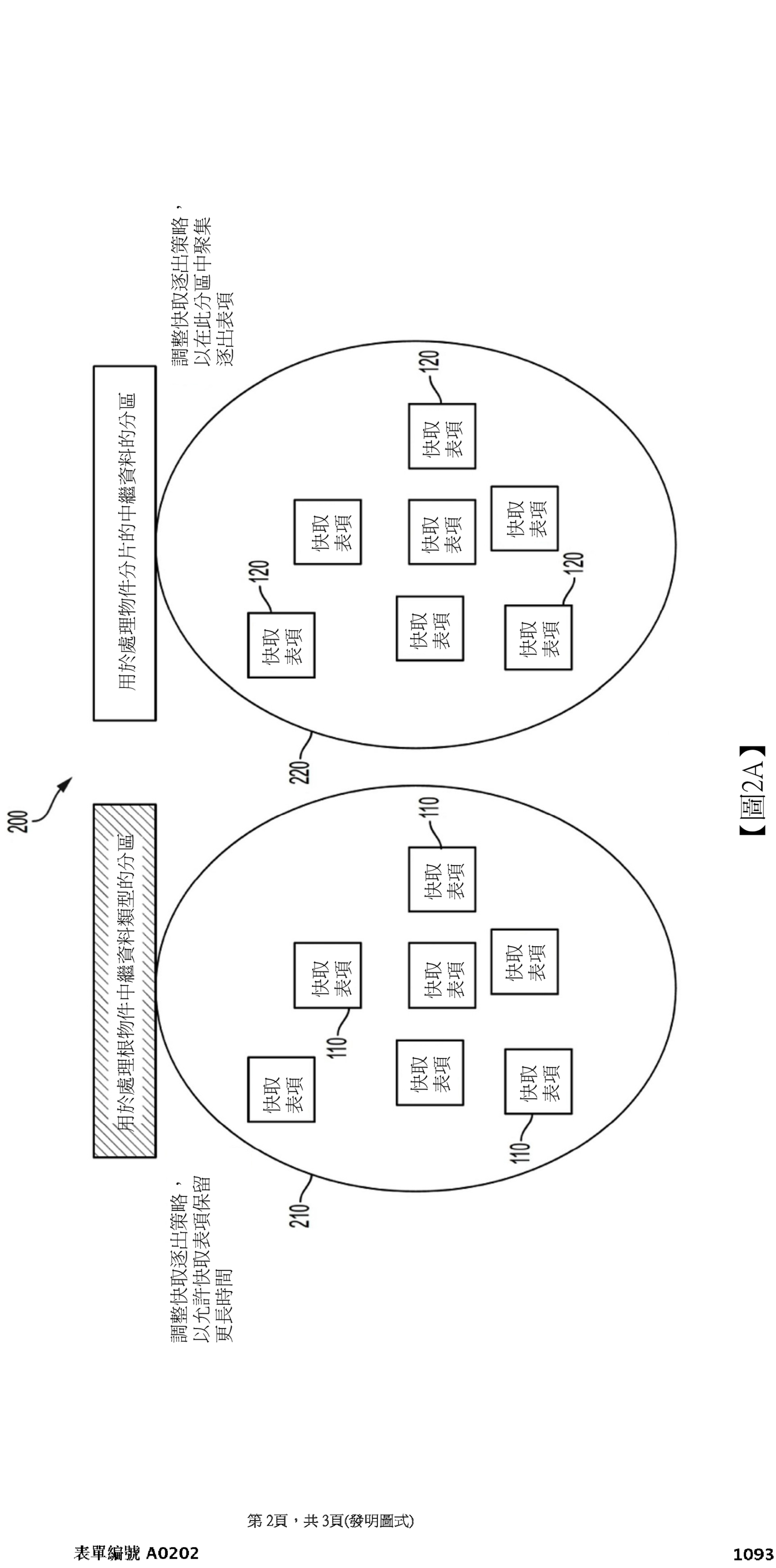
【請求項19】　如請求項15所述的非暫時性電腦可讀媒體，其中所述電腦代碼在由所述處理器執行時更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：由所述快取管理器對所述快取進行移行，且由所述快取管理器將與低於參考重要程度的重要程度對應的所有快取表項逐出。

【請求項20】　如請求項19所述的非暫時性電腦可讀媒體，其中所述電腦代碼在由所述處理器執行時更通過以下方式實施所述分散式物件儲存器的所述快取的所述快取管理方法：一旦所使用的快取資源的量達到參考程度，便由所述快取管理器觸發所述快取的移行。

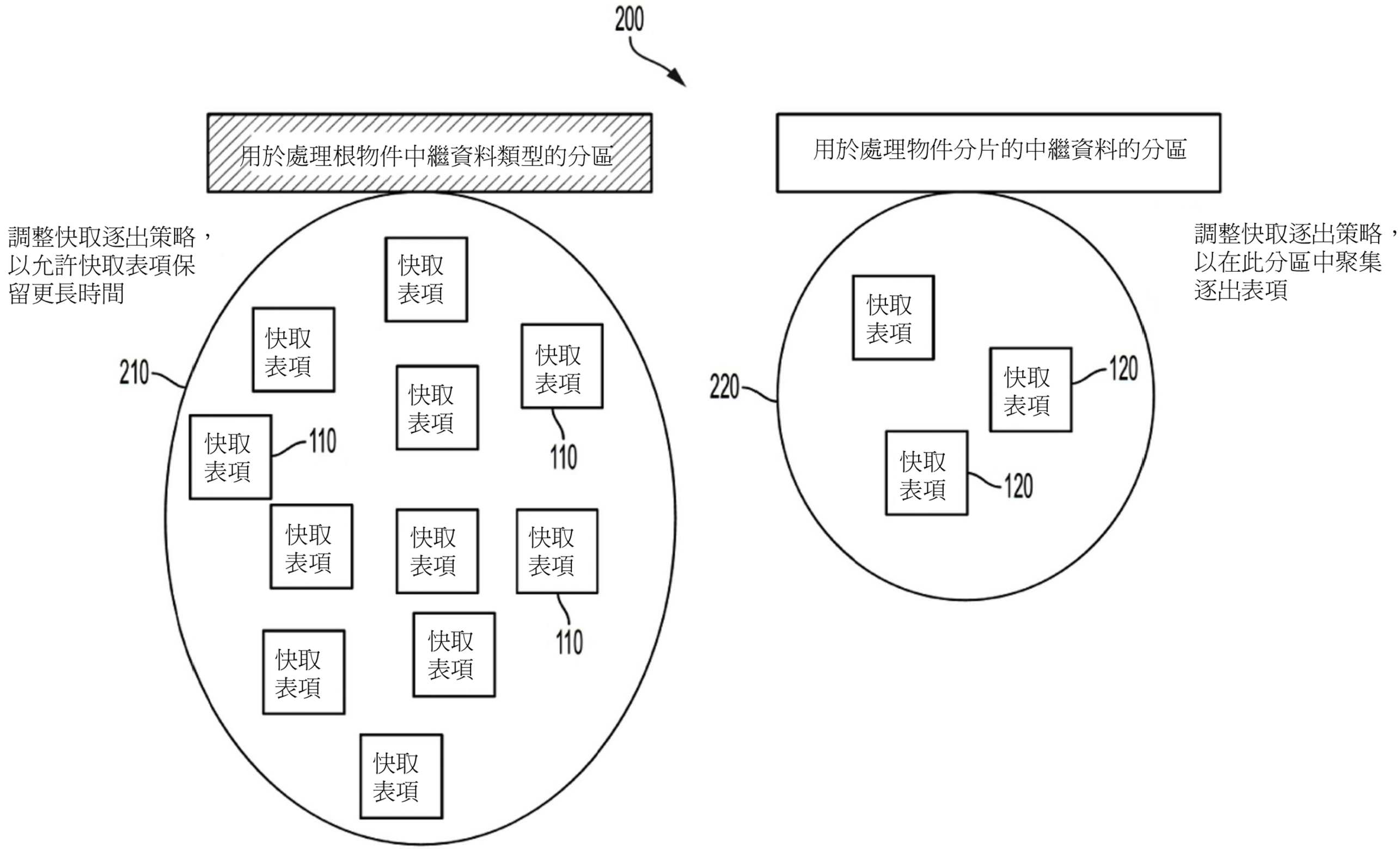
【發明圖式】


用於整個物件的根中繼資料
110
與所請求的物件的分片相關的特定中繼資料
120
每個輸入輸出所需的存取
基於輸入輸出位置所需的存取
增加數目
增加存取頻率

【圖1】


200
用於處理根物件中繼資料類型的分區
調整快取逐出策略，
以允許快取表項
更長時間
210
220
用於處理物件分片的中繼資料的分區
調整快取逐出策略，
以在此分區中聚集
逐出表項
快取表項
110
120

【圖2A】


200
用於處理根物件中繼資料類型的分區
調整快取逐出策略，以允許快取表項保留更長時間
210
快取表項
110
用於處理物件分片的中繼資料的分區
調整快取逐出策略，以在此分區中聚集逐出表項
220
120

【圖2B】