



[12] 发明专利说明书

专利号 ZL 200610101502.9

[45] 授权公告日 2009年9月9日

[11] 授权公告号 CN 100538661C

[22] 申请日 2006.7.18

[21] 申请号 200610101502.9

[30] 优先权

[32] 2005.9.29 [33] US [31] 11/239,597

[73] 专利权人 国际商业机器公司

地址 美国纽约

[72] 发明人 K·R·艾伦 K·C·沃森

W·A·布朗 R·K·柯克曼

[56] 参考文献

US6912625B2 2005.6.28

US6701421B1 2004.3.2

CN1085863C 2002.5.29

US6249802B1 2001.6.19

审查员 王燕_1

[74] 专利代理机构 北京市中咨律师事务所
代理人 于静 李峥

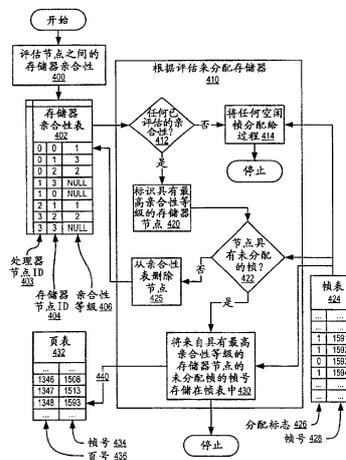
权利要求书 2 页 说明书 26 页 附图 9 页

[54] 发明名称

多节点计算机中存储器分配的方法和装置

[57] 摘要

本发明披露了多节点计算机中的存储器分配，包括评估节点之间的存储器亲合性并根据所述评估来分配存储器。评估存储器亲合性可以包括向节点指派存储器亲合性加权系数，其中每个加权系数都表示将节点的存储器分配给节点的处理器需求，并且分配存储器可以包括根据所述存储器亲合性加权系数来分配存储器。



1. 一种用于多节点计算机中的存储器分配的方法, 所述方法包括:
评估节点之间的存储器亲合性; 以及
根据所述评估来分配存储器;
其中, 评估存储器亲合性还包括向节点指派存储器亲合性加权系数, 每个加权系数都表示将节点的存储器分配给节点的处理器需求度; 以及
分配存储器还包括根据所述存储器亲合性加权系数来分配存储器。
2. 根据权利要求1的方法, 其中根据所述评估来分配存储器还包括按照要分配的存储器总量的比例从节点分配存储器。
3. 根据权利要求1的方法, 其中根据所述评估来分配存储器还包括按照存储器分配总数的比例从节点分配存储器。
4. 根据权利要求1的方法, 其中评估存储器亲合性还包括根据所述节点之间的存储器可用性来评估存储器亲合性。
5. 根据权利要求1的方法, 其中评估存储器亲合性还包括根据位于节点上的总系统存储器的比例来评估该节点的存储器亲合性。
6. 根据权利要求1的方法, 其中评估存储器亲合性还包括根据所述节点上的存储器的比例和所述节点上的处理器能力的比例来评估存储器亲合性。
7. 一种用于多节点计算机中的存储器分配的装置, 所述装置包括计算机处理器和操作地连接到所述计算机处理器的计算机存储器, 所述存储器分配的装置包括:
评估节点之间的存储器亲合性的第一装置; 以及
根据所述评估来分配存储器的第二装置;
其中, 所述第一装置进一步配置为向节点指派存储器亲合性加权系数, 每个加权系数都表示将节点的存储器分配给节点的处理器需求度; 以及
所述第二装置进一步配置为根据所述存储器亲合性加权系数来分配存储器。
8. 根据权利要求7的装置, 其中所述第二装置进一步配置为按照要

分配的存储器总量的比例从节点分配存储器。

9. 根据权利要求7的装置,其中所述第二装置进一步配置为按照存储器分配总数的比例从节点分配存储器。

多节点计算机中存储器分配的方法和装置

技术领域

本发明的领域是数据处理，或者更具体地说，是用于多节点计算机中存储器分配的方法、装置和产品。

背景技术

通常将1948年开发的EDVAC计算机系统作为计算机时代的开端。从那时起，计算机系统已经发展成非常复杂的设备。现今的计算机比早期的诸如EDVAC之类的系统更加完善。计算机系统通常包括硬件组件和软件组件、应用程序、操作系统、处理器、总线、存储器、输入/输出设备等的组合。由于半导体处理和计算机体系结构的进步使计算机的性能不断提高，更加完善的计算机软件已经发展为利用高性能的硬件，这导致今天的计算机系统比仅仅几年以前的计算机系统强大得多。

随着计算机系统变得更加复杂，它们的设计也日益模块化。通常使用多个模块化的节点来实现计算机系统，每个节点都包含一个或多个计算机处理器、一定数量的存储器，或同时包含处理器和存储器。复杂的计算机系统可能包括许多节点和用于在节点之间传输数据的复杂总线结构。

节点上的处理器访问节点上的存储器的访问时间随着哪个节点包含处理器以及哪个节点包含要访问的存储器的不同而不同。处理器对与其在同一节点上的存储器进行的存储器访问所用的时间短于处理器对在其他节点上的存储器进行的存储器访问所用的时间。访问同一节点上的存储器较快是因为访问远程节点上的存储器必须经过节点之间更多的计算机硬件、更多的总线、总线驱动器、存储器控制器等。

包含处理器和存储器的节点之间的计算机硬件分离的级别称为“存储器亲合性”或简称为“亲合性”。节点与其自身具有最大的存储器亲合性，

因为其处理器访问其存储器快于访问其他节点上的存储器。包含处理器的节点与其上安装了存储器的一个或多个节点之间的存储器亲合性随着硬件分离级别的增加而减小。

考虑具有下表中信息特征的计算机系统的实例:

节点	处理器能力的比例	存储器能力的比例
0	50%	50%
1	50%	5%
2	0%	45%

该表描述了具有三个节点（节点 0、1 和 2）的系统，其中处理器能力的比例表示相对于整个系统的每个节点上的处理器能力，并且存储器能力的比例表示相对于整个系统的安装在每个节点上的随机存取存储器的比例。操作系统可以实施亲合性，仅从与处理器在同一节点上的存储器将存储器分配给该处理器上的进程。在此实例中，节点 0 受益于实施亲合性，因为节点 0（具有系统中一半的存储器）很可能具有充足的存储器来满足在其处理器上运行的进程的需要。节点 0 还受益于实施存储器亲合性，因为访问与处理器在同一节点上的存储器比较快。

对于节点 1 就不是这种情况。节点 1（仅具有系统中百分之五的存储器）不太可能具有足够的存储器来满足在其处理器上运行的进程的需要。在实施亲合性中，每次执行进程或线程获得对节点 1 上的处理器的控制，所述进程或线程都可能遇到将 RAM 的内容交换到磁盘驱动器以清空存储器，并且从磁盘加载其存储器的内容，即称为‘交换’或‘系统失效’的效率非常低的操作。对处理器的本地节点上的存储器完全关闭亲合性实施可以减轻系统失效，但是在没有实施亲合性的情况下运行也会丧失在良好平衡的节点（如以上实例中的节点 0）上的处理器与存储器之间的亲合性实施的收益。

发明内容

本发明提供了一种用于多节点计算机中的存储器分配的方法，所述方法包括：评估节点之间的存储器亲合性；以及根据所述评估来分配存储器。其中，评估存储器亲合性还包括向节点指派存储器亲合性加权系数，每个加权系数都表示将节点的存储器分配给节点的处理器需求度；以及分配存储器还包括根据所述存储器亲合性加权系数来分配存储器。

本发明还提供了一种用于多节点计算机中的存储器分配的装置，所述装置包括计算机处理器和操作地连接到所述计算机处理器的计算机存储器，所述存储器分配的装置包括：评估节点之间的存储器亲合性的第一装置；以及根据所述评估来分配存储器的第二装置。其中，所述第一装置进一步配置为向节点指派存储器亲合性加权系数，每个加权系数都表示将节点的存储器分配给节点的处理器需求度；以及所述第二装置进一步配置为根据所述存储器亲合性加权系数来分配存储器。

本发明披露了通过评估节点之间的存储器亲合性并根据所述评估分配存储器来降低多节点计算机中存储器分配的系统失效的风险的方法、装置和产品。评估存储器亲合性可以包括：将存储器亲合性的加权系数指派给节点，其中每个加权系数都表示将节点的存储器分配给节点的处理器需求度，并且分配存储器可以包括根据所述存储器亲合性的加权系数来分配存储器。

如附图中示出的，从以下对本发明的示例性实施例的更具体的描述，本发明的上述和其他目标、特征和优点将变得显而易见，其中相同的标号通常代表本发明的示例性实施例的相同部件。

附图说明

图1是根据本发明的实施例的包括用于多节点计算机中的存储器分配的示例性计算机的自动计算机器的方块图；

图2是用于多节点计算机中的存储器分配的另一个示例性计算机的方块图；

图3是示出根据本发明的实施例的包括评估节点之间的存储器亲合性的用于多节点计算机中的存储器分配的示例性方法的流程图；

图 4 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图;

图 5 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图;

图 6 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图;

图 7 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图;

图 8 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图; 以及

图 9 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图。

具体实施方式

参考附图（开始于图 1）描述了根据本发明的实施例的用于多节点计算机中的存储器分配的示例性方法、装置和产品。通常使用计算机（即，自动计算机器）来实现根据本发明的多节点计算机中的存储器分配。因此，为了进一步说明，图 1 是根据本发明的实施例的包括用于多节点计算机中的存储器分配的示例性计算机（152）的自动计算机器的方块图。图 1 的计算机（152）包括至少一个节点（202）。节点是包含一个或多个计算机处理器、一定数量的存储器，或同时包含处理器和存储器的计算机硬件模块。在本说明书中，包含一个或多个处理器的节点有时称为‘处理器节点’，并且包含存储器的节点有时称为‘存储器节点’。同时包含一定数量的存储器和处理器的节点可以称为处理器节点和存储器节点两者。图 1 的节点（202）包括至少一个计算机处理器（156）或‘CPU’以及通过系统总线（160）连接到处理器（156）和计算机的其他组件的随机存取存储器（168）（‘RAM’）。实际上，根据本发明的实施例的用于多节点计算机中的存储器分配的系统通常包括多个节点、多个计算机处理器以及多个 RAM 电路。

存储在 RAM（168）中的是应用程序（153），即实现执行线程的用户级数据处理的计算机程序指令。还存储在 RAM（168）中的是操作系统（154）。根据本发明的实施例，计算机中使用的操作系统包括 UNIXTM、LinuxTM、Microsoft XPTM、AIXTM、IBM 的 i5/OSTM 和本领域的技术人员可想到的其他操作系统。操作系统（154）包含称为内核（157）的核心组件，该核心组件用于将诸如处理器和物理存储器之类的系统资源分配给应用程序（153）实例或操作系统（154）的其他组件。在图 1 的方法中，包括内核（157）的操作系统（154）显示在 RAM（168）中，但是此类软件的许多组件通常还存储在非易失性存储器（166）中。

图 1 的操作系统（154）包括加载器（158）。加载器（158）是从诸如盘驱动器、磁带或网络连接之类的加载源加载可执行程序以便例如由计算机处理器执行的计算机程序指令模块。加载器读取并解释可执行程序的元数据内容、分配程序所需的存储器、将程序的代码和数据段加载到存储器

中，以及向操作系统中的调度器登记程序以便执行（通常通过将新程序的标识符置于调度器的准备队列中）。在此实例中，加载器（158）是根据本发明的实施例改进的计算机程序指令模块，它通过评估节点之间的存储器亲合性并根据所述评估分配存储器来在多节点计算机中分配存储器。

图1的操作系统（154）包括存储器分配模块（159）。图1的存储器分配模块（159）是提供应用编程接口（‘API’）的计算机程序指令模块，应用程序和操作系统的其他组件可以通过该接口来动态地分配存储器、重新分配存储器或释放先前分配的存储器。对存储器分配模块（159）的API的函数调用（例如，‘malloc()’、‘realloc()’和‘free()’）满足了程序执行期间的动态存储器分配要求。在此实例中，存储器分配模块（159）是根据本发明的实施例改进的计算机程序指令模块，它通过评估节点之间的存储器亲合性并根据所述评估分配存储器来在多节点计算机中分配存储器。

还存储在RAM（168）中的是页表（432），页表（432）将计算机系统的虚拟存储器地址空间与图1的系统中的物理存储器地址空间之间的映射表示为数据结构。虚拟存储器地址空间被分成称为‘页’的固定大小的块，而物理存储器地址空间被分成称为‘帧’的相同大小的块。虚拟存储器地址空间为程序提供的用于在其中执行的存储器块要远大于计算机系统中安装的实际物理存储器的数量。虽然程序在似乎连续的虚拟存储器空间块中执行，但是包含该程序的实际物理存储器可以分散在整个计算机系统中。当在程序执行期间引用虚拟存储器的页时，操作系统（154）在与做出引用的程序关联的页表（432）中查找物理存储器的相应的帧。因此，页表（432）允许程序在虚拟地址空间中执行而不考虑其在物理存储器中的位置。在将图1的页表（432）与程序关联中，某些操作系统为每个执行程序维护页表（432），而其他操作系统可能将为整个系统维护的大型页表（432）的一部分指派给每个程序。

在创建、扩展或修改用于程序的页表（432）时，操作系统（154）将物理存储器的帧分配给页表（432）中的页。操作系统（154）通过帧表（424）定位未分配的帧以指派给页表（432）。帧表（424）存储在RAM（168）中

并表示与图 1 的系统中的物理存储器的帧有关的信息。在将图 1 的帧表 (424) 与节点上的帧关联中, 某些操作系统可能为每个节点维护包含该节点上未分配帧的列表的帧表 (424), 而其他操作系统可能为整个系统维护包含与所有节点中的所有帧有关的信息的大型帧表 (424)。帧表 (424) 指示帧是否被映射到虚拟存储器空间中的页。未映射到页的帧是未分配的并且因此可用于存储代码和数据。

还存储在 RAM (168) 中的是表示处理器节点与存储器节点之间的存储器亲合性的评估的存储器亲合性表 (402)。非常接近的处理器节点与存储器节点之间存在较高的存储器亲合性评估, 因为向与处理器节点具有高存储器亲合性的节点写入数据 (或从其读取数据) 时, 在向 (或者从) 此类高亲合性存储器节点的传输中, 将通过较少的计算机硬件、存储器控制器以及总线驱动器。此外, 对于具有相对较多部分的可用存储器的存储器节点, 存储器亲合性评估也会很高。例如, 比其他存储器节点 (具有相似的对处理器节点的物理接近度) 包含更多未分配帧的存储器节点就该处理器节点而言可以具有较高的存储器亲合性评估。可以使用存储器亲合性等级或存储器亲合性的加权系数在存储器亲合性表 (402) 中表示存储器亲合性的评估。存储器亲合性等级可以例如是指示从其将帧分配给执行程序的处理器节点的存储器节点顺序的序数。存储器亲合性的加权系数可以例如指示将做出的从存储器节点到处理器节点的帧分配比例。在将图 1 的存储器亲合性表 (402) 与处理器节点关联中, 某些操作系统为每个处理器节点维护存储器亲合性表 (402), 而其他操作系统可能将为整个系统维护的大型存储器亲合性表 (402) 的一部分指派给每个处理器节点 (156)。

图 1 的计算机 (152) 包括通过系统总线 (160) 连接到处理器 (156) 和计算机 (152) 的其他组件的非易失性计算机存储器 (166)。非易失性计算机存储器 (166) 可以被实现为硬盘驱动器 (170)、光盘驱动器 (172)、电可擦除可编程只读存储器空间 (所谓的“EEPROM”或“闪速”存储器) (174)、RAM 驱动器 (未示出) 或本领域的技术人员可想到的任何其他类型的计算机存储器。在图 1 的方法中, 页表 (432)、帧表 (424)、存储器亲合性表 (402) 和应用程序 (153) 在 RAM (168) 中示出, 但是此类软

件的许多组件通常存储在非易失性存储器(166)中。

图1的实例计算机包括一个或多个输入/输出接口适配器(178)。计算机中的输入/输出接口适配器通过例如软件驱动程序和计算机硬件来实现面向用户的输入/输出,以便控制到诸如计算机显示屏幕之类的显示设备(180)的输出以及来自诸如键盘和鼠标之类的输入设备(181)的用户输入。

图1的示例性计算机(152)包括用于实现与其他计算机(182)的数据通信(184)的通信适配器(167)。可以通过串行RS-232连接、外部总线(如USB)、数据通信网络(如IP网络)和本领域的技术人员可想到的其他方式来执行此类数据通信。通信适配器实现硬件级别的数据通信,通过所述适配器,一台计算机直接地或通过网络将数据通信发送到另一台计算机。根据本发明的实施例,用于确定目的地的可用性的通信适配器的实例包括用于有线拨号通信的调制解调器、用于有线网络通信的以太网(IEEE 802.3)适配器和用于无线网络通信的802.11b适配器。

为了进一步说明,图2是用于多节点计算机中的存储器分配的另一个示例性计算机(152)的方块图。图2的系统包括实现为存储器集成电路(称为‘存储器芯片’(205))的随机存取存储器,所述芯片包括在安装在背板(206)上的节点(202)中,每个背板通过系统总线(160)连接到计算机(152)的其他组件。节点(202)还可以包括计算机处理器(204),它也以集成电路的形式安装在节点上。连接背板上的节点以便通过背板总线(212)进行数据通信,并且连接节点上的处理器芯片和存储器芯片以便通过节点总线进行数据通信,所述节点总线在节点(222)上的标号(210)处示出,其扩展了节点(221)的图形表示。

节点可以例如被实现为多芯片模块(‘MCM’)。MCM是具有两个或更多组装在衬底上的裸集成电路(裸片)或‘芯片尺寸的封装’的电子系统或子系统。在图2的方法中,MCM中的芯片是计算机处理器和计算机存储器。例如,衬底可以是印刷电路板或具有互连图形的厚或薄的陶瓷膜或硅膜。衬底可以是MCM封装的整体部分,也可以安装在MCM封装内。MCM在计算机硬件体系结构中非常有用,因为它们代表了专用集成电路(‘ASIC’)

与印刷电路板之间的封装级别。

图 2 的节点示出了硬件存储器分离或存储器亲合性的级别。节点(222)上的处理器(214)可以访问在以下存储器芯片中的物理存储器:

- 在与访问存储器芯片的处理器(214)位于同一节点上的存储器芯片(216)中,
- 在同一背板(208)的另一个节点上的存储器芯片(218)中, 或者
- 在另一个背板(206)的另一个节点上的存储器芯片(220)中。

就处理器(214)而言, 存储器芯片(216)被称为‘本地’, 因为存储器芯片(216)与处理器(214)位于同一节点上。但是, 就处理器(214)而言, 存储器芯片(218和220)被称为‘远程’, 因为存储器芯片(218和220)位于与处理器(214)所在节点不同的节点上。访问同一背板上的远程存储器所用的时间比访问本地存储器要长, 因为由处理器写入远程存储器或从远程存储器读取的数据在向(或者从)远程存储器的传输中, 将经过更多的计算机硬件、存储器控制器和总线驱动器。出于相同的原因, 远程地访问其他背板上的存储器将花费更长的时间。处理器节点的最高存储器亲合性是其自身的亲合性; 本地存储器提供了最快的可用存储器访问。与处理器节点位于同一背板上的存储器节点与该处理器节点具有的存储器亲合性评估要高于位于其他背板上的存储器节点。如此描述的计算机体系结构仅用于说明, 而不是限制计算机存储器。例如, 可以将若干节点安装在印刷电路板上, 并且将该印刷电路板插入背板, 由此创建图 2 中未示出的其他存储器亲合性级别。本领域的技术人员可想到的计算机体系结构的其他方面都可能影响处理器-存储器亲合性, 并且所有这些方面都在根据本发明的实施例的在多节点计算机中分配存储器的范围中。

为了进一步说明, 图 3 是示出根据本发明的实施例的包括评估(400)节点之间的存储器亲合性的用于多节点计算机中的存储器分配的示例性方法的流程图。在图 3 的方法中, 通过根据系统参数来计算可用于处理器节点的每个存储器节点的存储器亲合性等级(406), 可以完成评估(400)

节点之间的存储器亲合性。在图 3 的方法中，可以由指示其中操作系统将存储器从存储器节点分配到处理器节点的顺序的序数来表示存储器亲合性等级（406）。在计算存储器亲合性等级（406）中使用的系统参数可以是静态的并由系统管理员在安装计算机系统时存储到非易失性存储器中，所述参数例如是处理器节点数、节点上安装的存储器量，或节点的物理位置（MCM、背板等）。但是，所述系统参数可以随计算机系统的运行而动态改变，例如，在通过释放、分配或重新分配来动态改变每个节点中未分配帧的数量时。此外，可以在系统上电或初始程序加载（‘引导’）期间计算系统参数并将其存储到 RAM 或非易失性存储器中。

图 3 的存储器亲合性表（402）存储了节点之间的存储器亲合性评估。表（402）中的每个记录都指定了存储器节点（404）到处理器节点（403）的存储器亲合性评估（406）。在图 3 的方法中，存储器亲合性评估（406）是由指示其中操作系统将存储器从存储器节点（404）分配给处理器节点（403）的顺序的序数存储器亲合性等级（406）表示的存储器亲合性值。较低的序数表示较高的存储器亲合性等级（406）-序数 1 是高于序数 2 的存储器亲合性等级，序数 2 是高于序数 3 的存储器亲合性等级，依此类推，并且最低的序数对应于与处理器节点具有最高存储器亲合性评估的存储器节点，最高的序数对应于与处理器节点具有最低存储器亲合性评估的存储器节点。

图 3 的方法还包括根据所述评估来分配（410）存储器。按照图 3 的方法的根据所述评估来分配（410）存储器包括判定（412）系统中是否存在任何与处理器节点（即，要为其分配存储器的处理器节点）具有已评估的亲合性的存储器节点。在图 3 的实例中，通过判定表中是否存在用于要向其分配存储器的特定处理器节点的已评估亲合性，可以完成判定系统中是否存在任何与处理器节点具有已评估的亲合性的存储器节点。在此实例中，缺少已评估的存储器亲合性由表中的空表项来表示。

如果系统中没有与处理器节点具有已评估的亲合性的存储器节点，则图 3 的方法包括在不考虑存储器亲合性的情况下分配（414）在系统的任意位置可用的任何空闲存储器帧。例如，存储器亲合性表（402）中的处理器

节点 1 没有与存储器节点的已评估亲合性（由列（406）中的空值指示），所以可以从系统存储器中的任意位置的任何空闲帧向处理器节点 1 分配存储器而不考虑其位置。

如果系统中存在与处理器节点具有已评估亲合性的存储器节点，则图 3 的方法继之以标识（420）具有最高存储器亲合性等级（406）的存储器节点，并且如果该节点具有未分配的帧，则通过将来自该存储器节点的存储器帧的帧号（428）存储（430）在页表（432）中来从该节点分配存储器。页表（432）的每个记录都将页号（436）与帧号（434）相关联。根据图 3 的方法，如箭头（440）所示，帧号‘1593’表示已将来自具有最高存储器亲合性等级（406）的存储器节点的帧分配给页表（432）中的页号‘1348’。

如果具有最高存储器亲合性等级（406）的存储器节点没有未分配的帧，则图 3 的方法继之以从存储器亲合性表（402）移除（425）该节点的表项，并且循环以再次判定（412）系统中是否存在与处理器节点具有已评估亲合性的存储器节点，标识（420）具有最高存储器亲合性等级（406）的存储器节点等。

通过使用诸如在图 3 中的标号（424）处示出的帧表之类的帧表，可以判定（422）具有最高存储器亲合性等级（406）的节点是否具有未分配的帧。帧表（424）中的每个记录都表示由帧号（428）标识的存储器帧并由分配标志（426）指定该帧是否被分配。已分配帧的关联分配标志被设置为‘1’，并且空闲帧的分配标志被重置为‘0’。从此类帧表（424）分配帧包括将帧的分配标志设置为‘1’。在图 3 的帧表（424）中，分配了帧号‘1591’、‘1592’和‘1594’。但是帧号‘1593’仍未分配。

帧表的替代形式可以被实现为仅包含可分配帧的帧号的‘空闲帧表’。从空闲帧表分配帧包括从该空闲帧表删除已分配帧的帧号。本领域的技术人员可以想到其他形式的帧表，指示空闲帧和已分配帧的方法，并且所有这些形式都在本发明的范围之内。

为了进一步说明，图 4 示出了根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图，该方法包括评估（400）节点之间的存储器亲合性并根据所述评估来分配（410）存储器。在图 4

的方法中，评估（400）节点之间的存储器亲合性包括向节点指派（500）存储器亲合性的加权系数（502），其中每个加权系数（502）表示将节点存储器分配给节点的处理器需求度。通过根据系统参数来计算每个处理器节点和与处理器节点具有已评估存储器亲合性的存储器节点的存储器亲合性加权系数（502），并且将存储器亲合性加权系数（502）存储在如标号（402）处所示的存储器亲合性表中，可以完成指派（500）存储器亲合性的加权系数（502）。存储器亲合性表（402）的每个记录都指定了存储器节点（404）与处理器节点（403）的存储器亲合性的加权系数（502）。如图所示，处理器节点0具有与存储器节点0的存储器亲合性系数0.80，即，处理器节点0与自身的存储器亲合性系数为0.80。处理器节点0与存储器节点1的存储器亲合性系数为0.55。依此类推。在计算存储器亲合性加权系数（502）中使用的系统参数可以包括例如系统中的处理器节点数、节点的物理位置（MCM、背板等）、每个存储器节点上的存储器量、每个存储器节点中未分配的帧数，以及本领域的技术人员可以想到的与存储器亲合性评估有关的其他系统参数。

存储器亲合性表（402）中的存储器亲合性（502）的评估是存储器亲合性加权系数（502）。较高的存储器亲合性加权系数（502）表示存储器亲合性的评估较高。加权系数0.65表示的存储器亲合性评估高于加权系数0.35表示的存储器亲合性评估；加权系数1.25表示的存储器亲合性评估高于加权系数0.65表示的存储器亲合性评估；依此类推，并且最高的存储器亲合性加权系数对应于与处理器节点具有最高存储器亲合性评估的存储器节点，最低的存储器亲合性加权系数对应于与处理器节点具有最低存储器亲合性评估的存储器节点

图4的方法还包括根据所述评估来分配（410）存储器。按照图4的方法根据所述评估来分配（410）存储器包括根据存储器亲合性的加权系数来分配（510）存储器。在图4的方法中，根据存储器亲合性的加权系数来分配（510）存储器包括判定（410）系统中是否存在任何与处理器节点（即，要为其分配存储器的处理器节点）具有已评估亲合性的存储器节点。在图4的实例中，通过判定表中是否存在用于要向其分配存储器的特定处理器

节点的已评估亲合性，可以完成判定系统中是否存在任何与处理器节点具有已评估的亲合性的存储器节点。在此实例中，缺少已评估的存储器亲合性由表中的空表项来表示。

如果系统中没有与处理器节点具有已评估的亲合性的存储器节点，则图 4 的方法包括在不考虑存储器亲合性的情况下分配 (414) 在系统的任意位置可用的任何空闲存储器帧。例如，存储器亲合性表 (402) 中的处理器节点 1 没有与存储器节点的已评估亲合性 (由列 (502) 中的空值指示)，所以可以从系统存储器中的任意位置的任何空闲帧向处理器节点 1 分配存储器而不考虑其位置。

如果系统中存在与处理器节点具有已评估亲合性的存储器节点，则图 4 的方法继之以标识 (520) 具有最高存储器亲合性加权系数 (502) 的存储器节点，并且如果该节点具有未分配的帧，则通过将来自该存储器节点的存储器帧的帧号 (428) 存储 (430) 在页表 (432) 中来从该节点分配存储器。如果具有最高存储器亲合性加权系数 (502) 的存储器节点没有未分配的帧，则图 4 的方法继之以从存储器亲合性表 (402) 移除 (525) 该节点的表项，并且循环以再次判定 (412) 系统中是否存在与处理器节点具有已评估亲合性的存储器节点，标识 (520) 具有最高存储器亲合性加权系数 (502) 的存储器节点等。

可以从具有最高存储器亲合性加权系数 (502) 的节点的帧表 (424) 来判定 (422) 该节点是否具有未分配的帧。图 4 的帧表 (424) 和图 4 的页表 (432) 类似于图 3 的帧表和页表。在图 4 中，帧表 (424) 被表示为将分配标志 (426) 与存储器节点中的帧的帧号 (428) 相关联的数据结构。图 4 的页表 (432) 被表示为将存储器节点中的帧的帧号 (434) 与虚拟存储器空间中的页号 (436) 相关联的数据结构。根据图 4 的方法，如箭头 (440) 所示，帧号 '1593' 表示已将来自具有最高存储器亲合性加权系数 (502) 的存储器节点的帧分配给页表 (432) 中的页号 '1348'。

为了进一步说明，图 5 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图，所述方法包括评估 (400) 节点之间的存储器亲合性并根据所述评估来分配 (410) 存储器。通过根据

系统参数来计算每个处理器节点和与处理器节点具有已评估存储器亲合性的存储器节点的存储器亲合性加权系数(502)，并且将存储器亲合性加权系数(502)存储在存储器亲合性表(402)中，可以完成根据图5的方法的评估(400)节点之间的存储器亲合性。每个记录都指定了存储器节点(404)与处理器节点(403)的存储器亲合性的评估(502)。存储器亲合性表(402)中的存储器亲合性评估(502)是指示要分配的存储器总量的比例的存储器亲合性加权系数。

图5的方法还包括根据存储器亲合性评估，即根据存储器亲合性的加权系数(502)来分配(410)存储器。按照图5的方法根据所述评估来分配(410)存储器包括按照要分配的存储器总量的比例来从节点分配(610)存储器。通过按照要分配给处理器节点的存储器总量的比例来从节点分配存储器，可以完成按照要分配的存储器总量的比例来从节点分配(610)存储器。可以将要分配的存储器总量标识为要分配的存储器的预定量，例如，下一个要分配的5兆字节。

根据图5的方法的按照要分配的存储器总量的比例来从节点分配(610)存储器包括从节点的存储器亲合性加权系数(502)来计算(612)要分配的存储器总量的比例(624)。存储器节点要从与处理器具有已评估亲合性的存储器节点分配给处理器节点的存储器总量的比例(610)可以按以下方法来计算：要分配的存储器总量乘以存储器节点的存储器亲合性的加权系数(502)的值与相对于处理器节点具有已评估亲合性的存储器节点的所有存储器亲合性加权系数(502)的总值的比率。

对于表(402)中的处理器节点0，与处理器节点0具有已评估亲合性的存储器节点(即，存储器节点0、1和2)的所有存储器亲合性加权系数的总和为1.5。使用图5的实例中的5兆字节的要分配的存储器总量，可以如下分别计算要从与存储器节点0、1和2关联的节点的存储器分配的存储器总量的比例(624)：

- 节点0: $(0.75 \text{ 节点0的已评估存储器亲合性}) \div (1.5 \text{ 总的已评估存储器亲合性}) \times 5\text{MB} = 2.5\text{MB}$

- 节点 1: $(0.60 \text{ 节点 1 的已评估存储器亲合性}) \div (1.5 \text{ 总的已评估存储器亲合性}) \times 5\text{MB} = 2.0\text{MB}$
- 节点 2: $(0.15 \text{ 节点 0 的已评估存储器亲合性}) \div (1.5 \text{ 总的已评估存储器亲合性}) \times 5\text{MB} = 0.5\text{MB}$

在此实例中, 根据图 5 的方法的按照要分配的 5MB 的存储器总量的比例从节点分配 (610) 存储器可以通过将下一个 5MB 分配给节点 0 (通过从节点 0 分配 5MB 分配量的最初 2.5MB, 从节点 1 分配下一个 2.0MB, 以及从节点 2 分配 5MB 分配量的最后 0.5MB) 来完成。所有这些分配都根据存储器节点中的帧的可用性。具体地说, 在图 5 的实例中, 按照要分配的存储器总量的比例从节点分配 (610) 存储器还包括根据帧的可用性从节点上的存储器分配 (630) 要分配的存储器总量的已计算的比例 (624)。可以通过使用帧表 (424) 来判定存储器节点上是否存在未分配的帧。帧表 (424) 将存储器节点中的帧的帧号 (428) 与指示是否分配了存储器帧的分配标志 (426) 相关联。

根据图 5 的方法的分配 (630) 存储器总量的已计算的比例 (624) 可以包括计算分配要分配的存储器总量的已计算的比例 (624) 所需的帧数。计算所需的帧数可以通过将帧的大小分为要分配的存储器总量的比例 (624) 来完成。继续以上实例的计算, 其中与处理器节点 0 具有已评估亲合性的存储器节点的所有存储器亲合性加权系数的总和为 1.5, 要分配的存储器总量为 5 兆字节, 要从节点 0、1 和 2 分配的存储器总量的比例分别为 2.5MB、2.0MB 和 0.5MB, 并且帧大小为 2KB, 则按照下面的方法计算要从节点 0、1 和 2 分配的帧数:

- 节点 0: $2.5\text{MB} \div 2\text{KB}/\text{帧} = 1280 \text{ 帧}$
- 节点 1: $2.0\text{MB} \div 2\text{KB}/\text{帧} = 1024 \text{ 帧}$
- 节点 2: $0.5\text{MB} \div 2\text{KB}/\text{帧} = 256 \text{ 帧}$

根据图 5 的方法的分配 (630) 存储器总量的已计算的比例 (624) 还

可以通过以下步骤来完成：对于在处理器节点上执行的程序，将来自存储器节点的所有未分配帧的帧号（428）（最高到从存储器节点分配要分配的存储器总量的已计算的比例（624）所需的帧数并且包括该帧数）存储在页表（432）中。图5的页表（432）的每个记录都将存储器节点上的帧的帧号（434）与处理器节点上执行的程序所使用的虚拟存储器空间中的页号（436）相关联。因此，在图5的实例中，如箭头（440）所示，帧号‘1593’表示已将来自具有最高存储器亲合性加权系数（502）的存储器节点的帧分配给页表（432）中的页号‘1348’。

在已分配要从存储器节点分配要分配的存储器总量的比例（624）所需的帧数之后，或从存储器节点分配所有未分配的帧之后（无论哪一个在前），图5的方法继续（632）以循环到存储器亲合性表（402）中与存储器节点关联的下一个表项，并且再次从节点的存储器亲合性加权系数（502）来计算（612）要分配的存储器总量的比例，根据帧的可用性来分配（630）要从节点上的存储器分配的存储器总量的已计算的比例（624）等，直到根据帧的可用性，为与处理器节点（将为其分配一定数量的存储器）具有已评估存储器亲合性（502）的每个存储器节点分配了要分配的存储器总量的比例（624）发生为止。在根据帧的可用性，按照图5的方法为与处理器节点（将为其分配一定数量的存储器）具有已评估存储器亲合性（502）的每个存储器节点分配要分配的存储器总量的比例（624）时，分配总量中任何未分配的部分都可以来自系统上任何位置的存储器而不考虑存储器亲合性。

为了进一步说明，图6是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图，该方法包括评估（400）节点之间的存储器亲合性并根据所述评估来分配（410）存储器。通过根据系统参数为每个处理器节点计算每个存储器节点的存储器亲合性加权系数（502），并且将存储器亲合性加权系数（502）存储在存储器亲合性表（402）中，可以完成根据图6的方法的评估（400）节点之间的存储器亲合性。存储器亲合性表（402）的每个记录都指定了存储器节点（404）与处理器节点（403）的存储器亲合性的评估（502）。存储器亲合性表（402）中的存储器亲合性评估（502）是指示要从存储器节点分配给处理器节点的存储器

分配总数的比例的存储器亲合性加权系数 (502)。

图 6 的方法还包括根据存储器亲合性评估, 即根据存储器亲合性的加权系数 (502) 来分配 (410) 存储器。按照图 6 的方法根据所述评估来分配 (410) 存储器包括按照存储器分配总数的比例来从节点分配 (710) 存储器。通过按照到处理器节点的存储器分配总数的比例来从节点分配存储器, 可以完成按照存储器分配总数的比例来从节点分配 (710) 存储器。在图 6 中, 可以将存储器分配总数标识为预定的存储器分配数, 例如, 到处理器节点的下一个 500 次存储器分配。

根据图 6 的方法的按照存储器分配总数的比例从节点分配 (710) 存储器包括从节点的存储器亲合性加权系数 (502) 来计算 (712) 存储器分配总数的比例 (724)。存储器节点要从与处理器具有已评估亲合性的存储器节点分配给处理器节点的存储器分配总数的比例 (724) 可以按照以下方法来计算: 将存储器分配的总数乘以存储器节点的存储器亲合性加权系数 (502) 的值与相对于处理器节点具有已评估亲合性的存储器节点的所有存储器亲合性加权系数 (502) 的总值的比率。

对于表 (402) 中的处理器节点 0, 与处理器节点 0 具有已评估亲合性的存储器节点 (即, 存储器节点 0、1 和 2) 的所有亲合性加权系数的总和为 1.5。使用图 6 的实例中的 500 次分配的存储器分配总数, 可以如下分别计算从存储器节点 0、1 和 2 到处理器节点的存储器分配总数的比例 (724):

- 节点 0: $(0.75 \text{ 节点 0 的已评估存储器亲合性}) \div (1.5 \text{ 总的已评估存储器亲合性}) \times 500 \text{ 次分配} = 250 \text{ 次分配}$
- 节点 1: $(0.60 \text{ 节点 1 的已评估存储器亲合性}) \div (1.5 \text{ 总的已评估存储器亲合性}) \times 500 \text{ 次分配} = 200 \text{ 次分配}$
- 节点 2: $(0.15 \text{ 节点 0 的已评估存储器亲合性}) \div (1.5 \text{ 总的已评估存储器亲合性}) \times 500 \text{ 次分配} = 50 \text{ 次分配}$

在此实例中, 根据图 6 的方法的按照 500 次存储器分配总数的比例从

节点分配(710)存储器可以通过将下一个500次分配分配给节点0(通过从节点0分配500次分配的最初250次,从节点1分配下一个200次,以及从节点2分配500次的最后50次)来完成。所有这些分配都根据存储器节点中的帧的可用性,并且所有这些分配都在不考虑已分配的存储器量的情况下实现。具体地说,在图6的实例中,按照存储器分配总数的比例从节点分配(710)存储器还包括根据帧的可用性从节点上的存储器分配(730)存储器分配总数的已计算的比例(724)。可以通过使用帧表(424)来判定存储器节点上是否存在未分配的帧。帧表(424)将存储器节点中的帧的帧号(428)与指示是否分配了存储器帧的分配标志(426)相关联。

根据图6的方法的分配(730)存储器分配总数的已计算的比例(724)还可以通过以下步骤来完成:对于在处理器节点上执行的程序,将来自存储器节点的所有未分配帧的帧号(428)(最高到并且包括存储器节点的存储器分配总数的已计算的比例(724))存储在页表(432)中。图6的页表(432)的每个记录都将存储器节点上的帧的帧号(434)与处理器节点上执行的程序所使用的虚拟存储器空间中的页号(436)相关联。因此,在图6的实例中,如箭头(440)所示,帧号‘1593’表示已将来自具有到处理器节点的已评估存储器亲合性(此处为加权的存储器亲合性)的存储器节点的帧分配给页表(432)中的页号‘1348’。

在从存储器节点分配存储器分配总数的已计算的比例(724)后,或从存储器节点分配所有未分配的帧之后(无论哪一个在前),图6的方法继续(732)以循环到存储器亲合性表(402)中与存储器节点关联的下一个表项,并且再次从节点的存储器亲合性加权系数(502)来计算(712)存储器分配总数的比例(724),根据帧的可用性来从节点上的存储器分配(730)存储器分配总数的已计算的比例(724)等,直到根据帧的可用性,为与处理器节点(将为其分配存储器)具有已评估存储器亲合性(502)的每个存储器节点分配了存储器分配总数的已计算的比例(724)发生为止。在根据帧的可用性,按照图6的方法为与处理器节点(将为其分配存储器)具有已评估存储器亲合性(502)的每个存储器节点分配存储器分配总数的已计算的比例(724)时,分配总数中任何未分配的部分都可以来自系统上

任何位置的存储器而不考虑存储器亲合性。

为了进一步说明，图 7 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图，该方法包括评估 (400) 节点之间的存储器亲合性并根据所述评估来分配 (410) 存储器。根据图 7 的方法的评估 (400) 节点之间的存储器亲合性包括根据节点之间的存储器可用性来评估 (800) 存储器亲合性。

在图 7 的方法中，根据节点之间的存储器可用性来评估 (800) 存储器亲合性包括确定 (804) 每个存储器节点的未分配的帧数。可以从帧表 (424) 来确定每个存储器节点的未分配的帧数。在图 7 的方法中，帧表 (424) 被表示为将存储器节点中的帧的帧号 (428) 与指示是否分配了存储器帧的分配标志 (426) 相关联的数据结构。根据图 7 的方法的确定 (804) 每个存储器节点的未分配的帧数可以通过以下步骤来完成：对位于每个存储器节点中的未分配帧的数量进行计数并将每个存储器节点的未分配帧的总数存储在未分配帧总数表 (806) 中。在某些实施例中，操作系统可以以空闲帧列表的形式为每个存储器节点维护帧表 (424)。在这些实施例中，确定 (804) 每个存储器节点的未分配帧的数量可以通过以下步骤来完成：对每个存储器节点的空闲帧列表中的表项数进行计数并将每个存储器节点的未分配帧的总数存储在未分配帧总数表 (如标号 (806) 处示出的表) 中。

图 7 的未分配帧总数表 (806) 存储了系统的每个节点上安装的存储器中的未分配帧的数量。未分配帧总数表 (806) 的每个记录都将存储器节点 (404) 与未分配帧总数 (808) 相关联。

按照图 7 的方法的根据节点之间的存储器可用性来评估 (800) 存储器亲合性还包括根据以下公式 1 来计算 (810) 处理器节点与存储器节点之间的存储器亲合性的加权系数 (502)：

$$\text{公式 1: } A_i = \frac{F_i}{\sum_{n=0}^{N-1} F_n}$$

其中 A_i 是处理器节点与第 i 个存储器节点的存储器亲合性加权系数 (502)， F_i 是第 i 个存储器节点上的未分配帧的数量， N 是系统中的存储

器节点数，公式 1 的分母是所有存储器节点上的所有未分配帧的总数。例如，对于存储器亲合性表 (402) 中的处理器节点 0 和存储器节点 0，可以根据公式 1 来计算存储器亲合性加权系数 A_i ，其中从表 (806) 获得的第 i 个存储器节点上的未分配帧数 F_i 为 100，存储器节点数 N 为 3，从表 (806) 的列 (808) 相加的所有存储器节点上的所有未分配帧的总数为 200，计算出的 A_i 为 $0.5=100 \div 200$ 。

在图 7 的方法中，存储器亲合性的评估 (502) 是存储器亲合性加权系数 (502)，但是这些存储器亲合性加权系数 (502) 仅用于示例性目的。实际上，图 7 的存储器亲合性的评估 (502) 还可以被表示为指示其中操作系统将存储器从存储器节点分配给处理器节点的顺序和本领域的技术人员所想到的其他方式的存储器亲合性等级。

在图 7 的方法中，计算 (810) 存储器亲合性加权系数 (502) 可以包括将每个存储器节点的存储器亲合性加权系数 (502) 存储在存储器亲合性表 (402) 中。存储器亲合性表 (402) 的每个记录都将存储器节点 (404) 的存储器亲合性评估 (502) 与处理器节点 (403) 相关联。

图 7 的方法还包括根据存储器亲合性的评估来分配 (410) 存储器。如以上在本说明书中详细描述，根据所述评估来分配 (410) 存储器可以通过以下步骤来完成：判定系统中是否存在与处理器节点具有已评估亲合性的任何存储器节点，标识具有最高存储器亲合性等级的存储器节点，以及判定具有最高存储器亲合性等级的节点是否具有未分配的帧等。

为了进一步说明，图 8 是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图，该方法包括评估 (400) 节点之间的存储器亲合性并根据所述评估来分配 (410) 存储器。根据图 8 的方法的评估 (400) 节点之间的存储器亲合性包括根据位于节点上的总系统存储器的比例来评估 (900) 该节点的存储器亲合性。总系统存储器表示系统的存储器节点上安装的随机存取存储器的总量。

在图 8 的方法中，根据位于节点上的总系统存储器的比例来评估 (900) 该节点的存储器亲合性包括确定 (902) 每个存储器节点上安装的存储器量。根据图 8 的方法的确定 (902) 每个存储器节点上的存储器量可以通过读取

安装存储器节点时系统管理员输入的包含存储器节点上的存储器量 (912) 的每个存储器节点的系统参数来完成。在其他实施例中, 确定 (902) 每个存储器节点上的存储器量可以通过在初始启动系统期间 (即, 当系统 ‘引导’ 时) 对存储器进行计数来完成。

在图 8 的方法中, 确定 (902) 每个存储器节点上的存储器量可以包括将每个存储器节点的存储器量 (912) 存储在总存储器表 (904) 中。图 8 的总存储器表 (904) 的每个记录都将存储器节点 (404) 与表 (904) 中标识的每个存储器节点的存储器量 (912) 相关联。

根据图 8 的方法, 根据位于节点上的总系统存储器的比例来评估 (900) 该节点的存储器亲合性还包括根据下面的公式 2 来计算 (906) 系统上安装的处理器节点与存储器节点之间的存储器亲合性加权系数 (502) :

$$\text{公式 2: } A_i = \frac{M_i}{\sum_{n=0}^{N-1} M_n}$$

其中 A_i 为处理器节点与第 i 个存储器节点的存储器亲合性加权系数 (502), M_i 为第 i 个存储器节点上的存储器量, N 为系统中的存储器节点数, 公式 2 的分母为所有存储器节点上的存储器总量。例如, 对于存储器亲合性表 (402) 中的处理器节点 0 和存储器节点 0, 可以根据公式 2 来计算存储器亲合性加权系数 A_i , 其中从表 (904) 获得的第 i 个存储器节点上的存储器量 M_i 为 500 MB, 存储器节点数 N 为 3, 从表 (904) 的列 (912) 相加的所有存储器节点上的存储器总量为 1000 MB, 计算出 A_i 为 $0.50=500 \div 1000$ 。

在图 8 的方法中, 计算 (906) 存储器亲合性加权系数 (502) 可以例如在系统上电期间或在早期引导阶段完成, 并且可以包括将每个存储器节点的存储器亲合性加权系数 (502) 存储在存储器亲合性表 (如在图 8 的标号 (402) 处所示的表) 中。存储器亲合性表 (402) 的每个记录都将存储器节点 (404) 的存储器亲合性评估 (502) 与处理器节点 (403) 相关联。

图 8 的方法还包括根据存储器亲合性的评估来分配 (410) 存储器。如以上在本说明书中详细描述, 根据所述评估来分配 (410) 存储器可以通

过以下步骤来完成：判定系统中是否存在与处理器节点具有已评估亲合性的任何存储器节点，标识具有最高存储器亲合性等级的存储器节点，以及判定具有最高存储器亲合性等级的节点是否具有未分配的帧等。

为了进一步说明，图9是示出根据本发明的实施例的用于多节点计算机中的存储器分配的另一个示例性方法的流程图，该方法包括评估(400)节点之间的存储器亲合性并根据所述评估来分配(410)存储器。根据图9的方法的评估(400)节点之间的存储器亲合性包括根据节点上的存储器的比例(1006)和节点上的处理器能力的比例(1008)来评估(1000)存储器亲合性。可以通过存储器节点上安装的存储器量与系统存储器总量的比来表示每个节点的存储器的比例(1006)。可以通过处理器节点上的处理器能力与系统中所有处理器节点的处理器总能力的总量的比来表示每个节点上的处理器能力的比例(1008)。在图9中，每个节点的存储器的比例(1006)和每个节点的处理器能力的比例(1008)可以从安装系统时由系统管理员输入的系统参数来获得。

图9的实例中的节点处理器-存储器配置(1002)是将存储器的比例(1006)和处理器能力的比例(1008)与节点标识符(1004)相关联的数据结构(在此实例中是表)。在此实例中，节点0包含50%的总系统存储器和50%的系统处理器能力，节点1包含5%的总系统存储器和45%的系统处理器能力，节点2包含45%的总系统存储器但是没有安装在节点上的处理器，节点3没有安装在其上的存储器并且包含5%的系统处理器能力。

在图9的方法中，根据节点上的存储器的比例(1006)和节点上的处理器能力的比例(1008)来评估(1000)存储器亲合性包括计算(1010)节点的处理器-存储器比。根据图9的方法的计算(1010)节点的处理器-存储器比可以通过以下步骤来完成：将节点上的处理器能力的比例(1008)除以节点上安装的存储器的比例(1006)，并将结果(1016)存储在处理器-存储器比率表(1012)中。

图9的处理器-存储器比率表(1012)将节点标识符(1004)与处理器-存储器比率(1016)相关联。在图9中，为‘1’的处理器-存储器比率(1016)指示节点相对于整个系统包含相等比例的处理器能力和存储器。处理器-

存储器比率 (1016) 大于 '1' 指示节点相对于整个系统包含的处理器能力的比例大于存储器的比例, 而处理器-存储器比率 (1016) 小于 '1' 指示节点相对于整个系统包含的处理器能力的比例小于存储器的比例。在图 9 中, 处理器-存储器比率 (1016) 为 '0' 指示节点上没有安装处理器, 而处理器-存储器比率 (1016) 为 'NULL' 指示节点上没有安装存储器。例如, 对于其上未安装存储器的节点 3, 将节点上的处理器能力的比例 (1008) 除以节点上安装的存储器的比例 (1006) 将会除零, 这由表 (1012) 中的节点 3 的 NULL 表项来指示。NULL 表项是适当的; 在处理器节点与其上没有存储器的其他节点之间, 没有用于存储器分配目的的有用的存储器亲合性。

根据图 9 的方法, 根据节点上的存储器的比例 (1006) 和节点上的处理器能力的比例 (1008) 来评估 (1000) 存储器亲合性还包括使用存储器-处理器比率来确定 (1020) 每个处理器节点对每个存储器节点的存储器亲合性等级。使用存储器-处理器比率来确定 (1020) 每个处理器节点对每个存储器节点的存储器亲合性等级可以包括将处理器节点对存储器节点的存储器亲合性等级存储在存储器亲合性表 (402) 中。每个记录都将存储器节点 (404) 的存储器亲合性评估 (406) 与处理器节点 (403) 相关联。存储器亲合性表 (402) 中的存储器亲合性评估是序数存储器亲合性等级 (406), 其指示了操作系统将存储器从表中标识的存储器节点 (404) 分配给处理器节点 (403) 的顺序。

存储器亲合性是在存储器节点与处理器节点之间, 而不是在存储器节点与其他存储器节点之间。节点具有为 0 的处理器-存储器比率 (1016) 意味着节点不包含处理器, 仅包含存储器并且因此在该节点与任何其他包含存储器的节点之间, 没有用于存储器分配目的的有用的存储器亲合性。为了良好的顺序以及完整性, 表 (402) 仍然将每个此类处理器的表项包含在其 '处理器节点' 列 (403) 中, 尽管此类节点实质上已不是 '处理器节点'。因此, 在图 9 的方法中, 对于节点 2 (具有为 '0' 的处理器-存储器比率 (1060) 的处理器节点), 确定 (1020) 该节点与其他存储器节点之间的存储器亲合性等级可以通过将 'NULL' 存储为此类节点的存储器亲合性等

级(406)来完成。例如,在图9中,将NULL存储在处理器节点2(不包含处理器的‘处理器节点’)的所有存储器亲合性等级(406)中。

节点具有的处理器-存储器比率等于或小于1指示该节点的资源通常是相当平衡的。可以合理地期望具有系统的一半处理能力和一半存储器的节点能够使用来自同一节点的存储器来满足所有其存储器要求。因此,在图9的方法中,对于节点0(处理器-存储器比率(1060)小于或等于‘1’的处理器节点),使用存储器-处理器比率来确定(1020)存储器亲合性还可以通过为表示同一节点的存储器节点(404)将‘1’存储在此类处理器节点的存储器亲合性等级(406)中以及将‘NULL’存储在与所述处理器节点关联的其他存储器亲合性等级(406)中来完成。在这种情况下,存储器亲合性等级为‘1’指示最高的存储器亲合性,‘2’指示较小的存储器亲合性,‘3’指示更小的存储器亲合性等。例如,在图9中,节点0具有处理器-存储器比率‘1’,并且为处理器节点0与存储器节点0(同一节点)指定为‘1’的存储器亲合性等级,而将‘NULL’存储为处理器节点0的所有其他存储器节点的存储器亲合性等级(406)。

处理器节点具有大于1的处理器-存储器比率意味着该节点相对于存储器具有较多的处理能力;此类节点很可能需要从其他节点分配的存储器。此类节点的最初存储器分配可以在节点具有可用存储器时来自其自身,并且当存储器必须来自其他节点时,从其他节点分配存储器可以首选来自具有小于1的处理器-存储器比率的节点(即,具有相对较多存储器的节点)的存储器。因此,在图9的方法中,对于节点1(具有大于‘1’的处理器-存储器比率(1016)的处理器节点),使用存储器-处理器比率来确定(1020)存储器亲合性等级可以通过以下步骤来完成:为表示同一节点的存储器节点(404)将值‘1’存储为此类处理器节点的存储器亲合性等级(406),将不断增大的序数存储为具有小于‘1’的处理器-存储器比率(1016)的其他存储器节点的存储器亲合性等级(406),以及为与处理器节点具有已评估亲合性的其他存储器节点将‘NULL’存储为存储器亲合性等级(406)。

在此实例中,低存储器亲合性等级值表示高存储器亲合性。存储器亲合性等级值为1表示最高的存储器亲合性,存储器亲合性等级值为2是较

低的存储器亲合性，3是更低的存储器亲合性等。按照具有最低处理器-存储器比率（1016）的存储器节点被评级为‘2’，具有第二低的处理器-存储器比率（1016）的存储器节点被评级为‘3’等对大于1的非空存储器亲合性等级值进行排序。例如，在图9的表（402）中，为存储器节点1将‘1’存储为处理器节点1的存储器亲合性等级。为存储器节点2将‘2’存储为处理器节点1的存储器亲合性等级。将NULL存储为处理器节点1的所有其他存储器亲合性等级。

处理器节点具有为空的处理器-存储器比率意味着该节点其上没有安装存储器；此类节点需要从其他节点分配的存储器。可以根据系统中存储器节点的处理器-存储器比率来完成为没有存储器的节点评估存储器亲合性。例如，就是说，为没有存储器的节点评估存储器亲合性可以通过将相对较高的存储器亲合性指派给具有小于1的处理器-存储器比率的存储器节点（即，具有相对较多存储器的节点）来完成。

因此，在图9的方法中，对于节点3（处理器-存储器比率（1016）为NULL的处理器节点），使用存储器-处理器比率来确定（1020）存储器亲合性等级可以通过将不断增大的序数存储为具有小于‘1’的处理器-存储器比率（1016）的存储器节点的存储器亲合性等级（406），并且为与处理器节点具有已评估亲合性的其他存储器节点将‘NULL’存储为存储器亲合性等级（406）来完成。在此实例中，低存储器亲合性等级值表示高存储器亲合性。存储器亲合性等级值为1表示最高的存储器亲合性，存储器亲合性等级为2是较低的存储器亲合性，存储器亲合性等级为3是更低的存储器亲合性等。按照具有最低处理器-存储器比率（1016）的存储器节点被评级为‘1’，具有第二低的处理器-存储器比率（1016）的存储器节点被评级为‘2’等对非空的存储器亲合性等级值进行排序。例如，在图9的表（402）中，将‘1’存储在处理器节点3和存储器节点2的存储器亲合性等级中。将NULL存储在处理器节点3的所有其他存储器亲合性等级中。

图9的方法还包括根据存储器亲合性的评估来分配（410）存储器。如以上在本说明书中详细描述，根据所述评估来分配（410）存储器可以通过以下步骤来完成：判定系统中是否存在与处理器节点具有已评估亲合性

的任何存储器节点，标识具有最高存储器亲合性等级的存储器节点，以及判定具有最高存储器亲合性等级的节点是否具有未分配的帧等。

很大程度上在用于多节点计算机中的存储器分配的完整功能计算机系统的上下文中描述了本发明的示例性实施例。但是，本领域的技术人员将认识到，本发明还可以包含在布置在用于与任何适合的数据处理系统一起使用的信号承载介质上的计算机程序产品中。此类信号承载介质可以是用于机器可读信息的传输介质或可记录介质，包括磁介质、光介质或其他适合的介质。可记录介质的实例包括硬盘驱动器中的磁盘或软盘、用于光学驱动器的光盘、磁带以及本领域的技术人员可想到的其他介质。传输介质的实例包括用于语音通信的电话网络、诸如以太网™之类的数字数据通信网络、使用网际协议通信的网络以及万维网。本领域的技术人员将立即认识到，任何具有适合的编程装置的计算机系统都将能够执行包含在程序产品中的本发明的方法的步骤。本领域的技术人员将立即认识到，虽然在说明书中描述的某些示例性实施例面向在计算机硬件上安装和执行的软件，但是作为固件或硬件实现的替代实施例也在本发明的范围之内。

从以上描述可以理解，可以在本发明的各个实施例中做出修改和更改而不偏离本发明的真正精神。说明书中的描述只是出于示例目的并且不应被理解为进行限制。本发明的范围仅由以下权利要求的语言来限制。

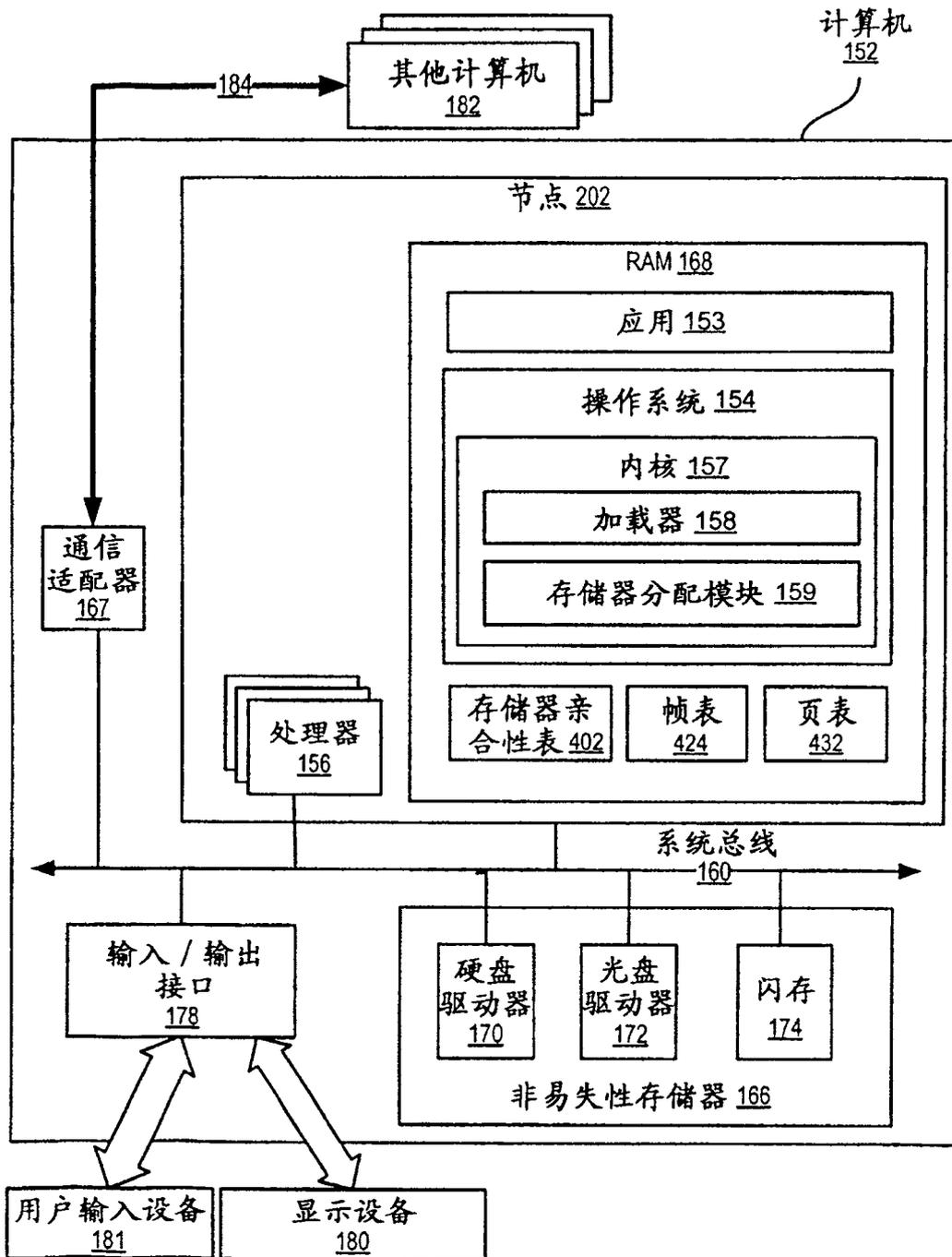


图 1

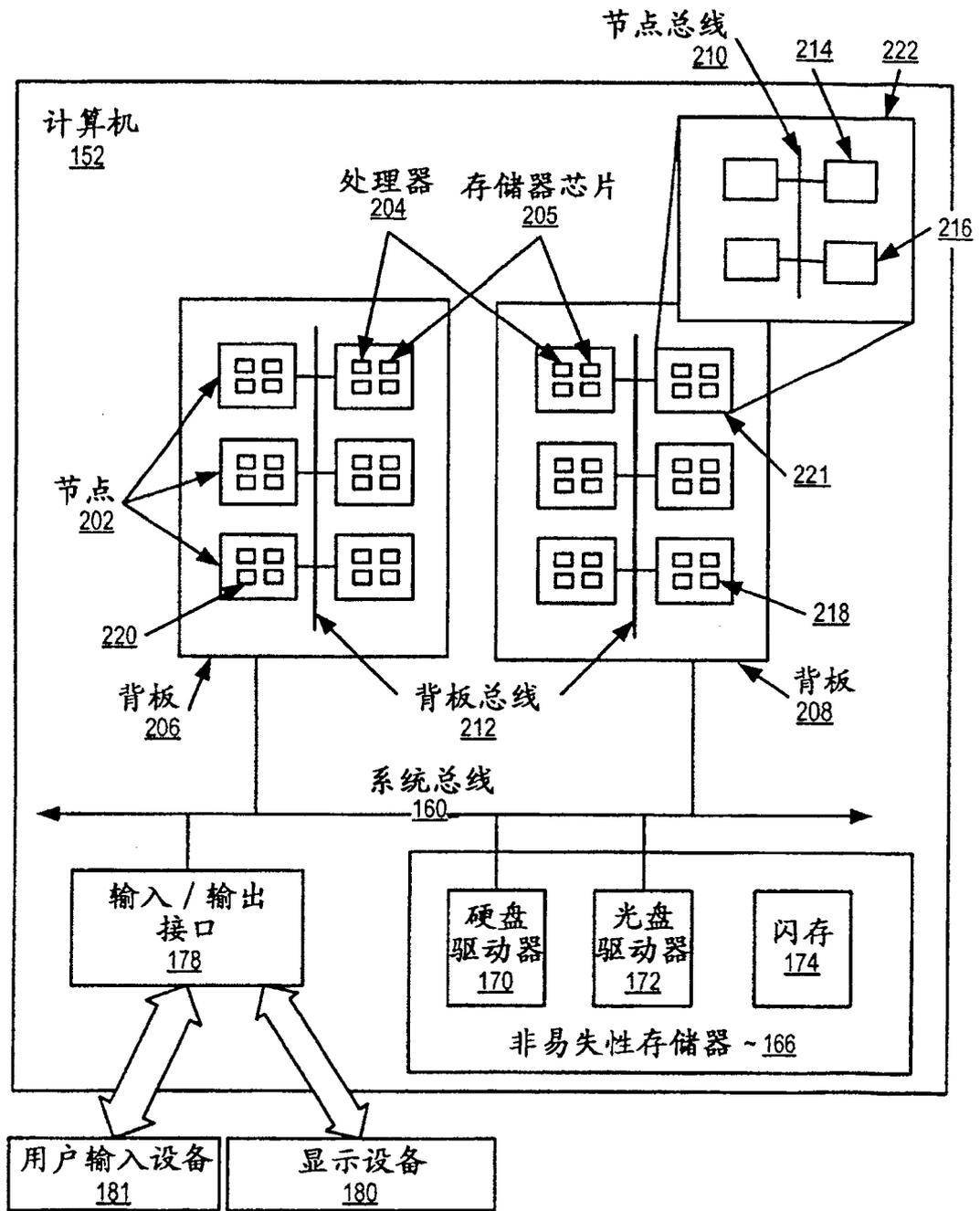


图 2

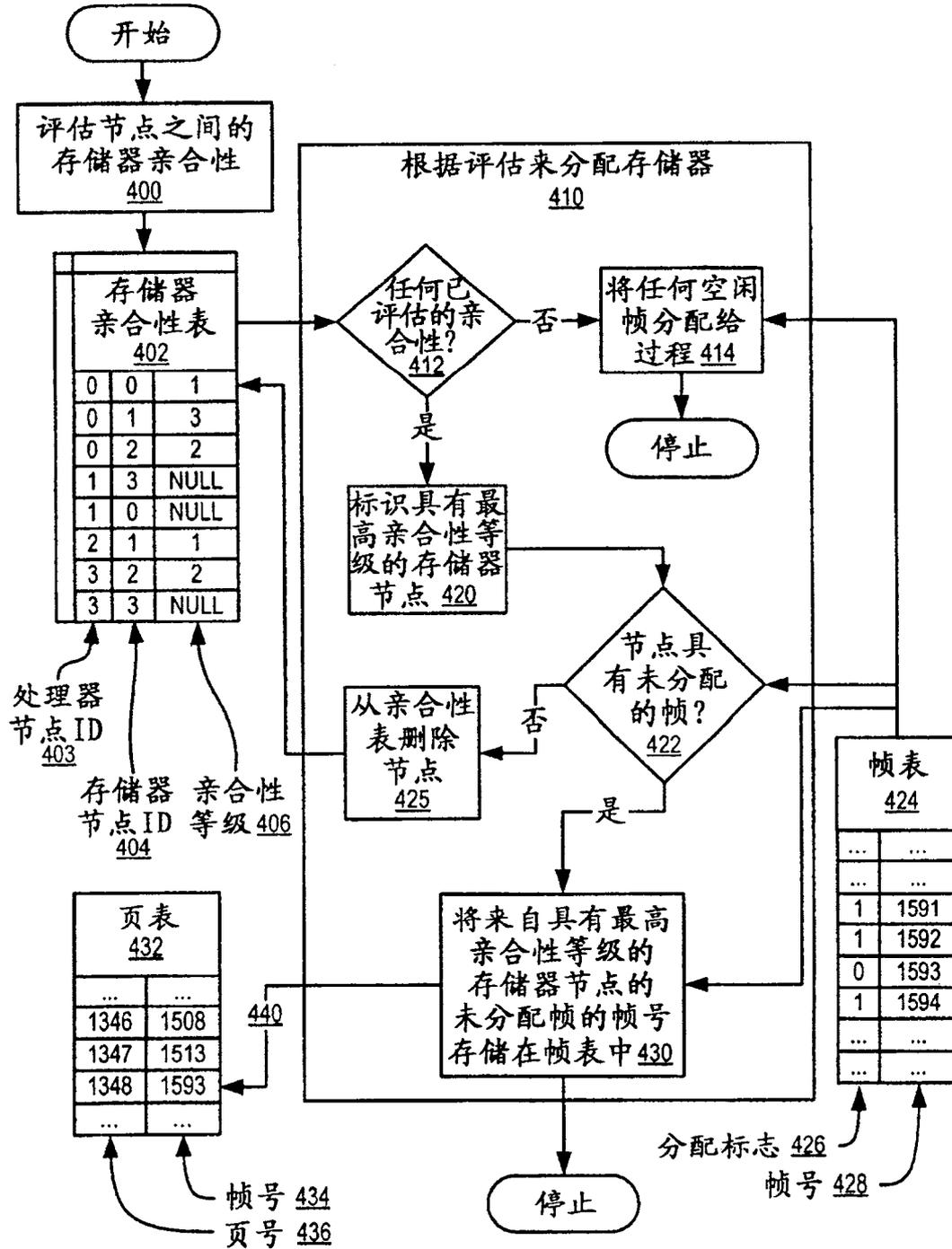


图 3

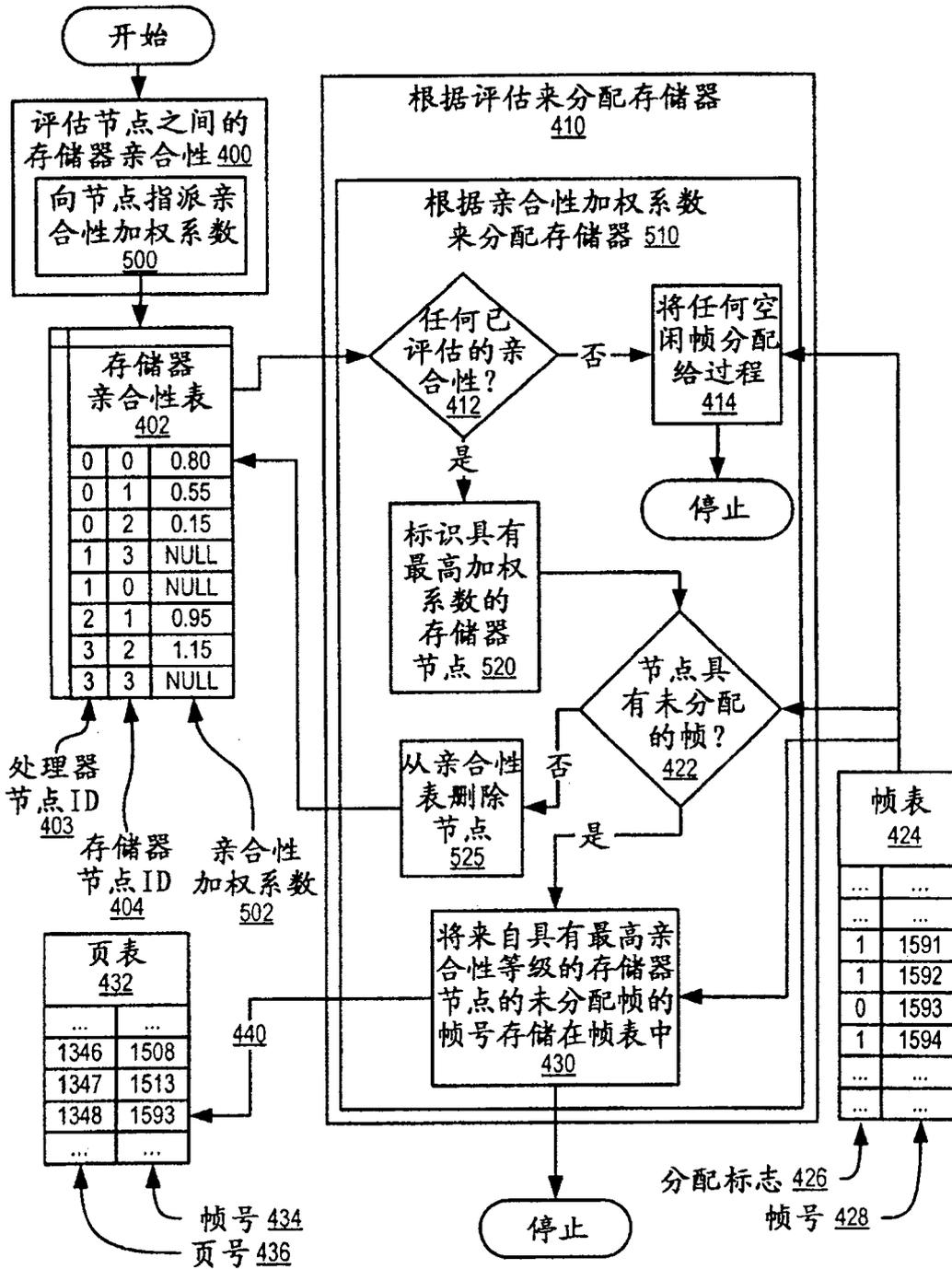


图 4

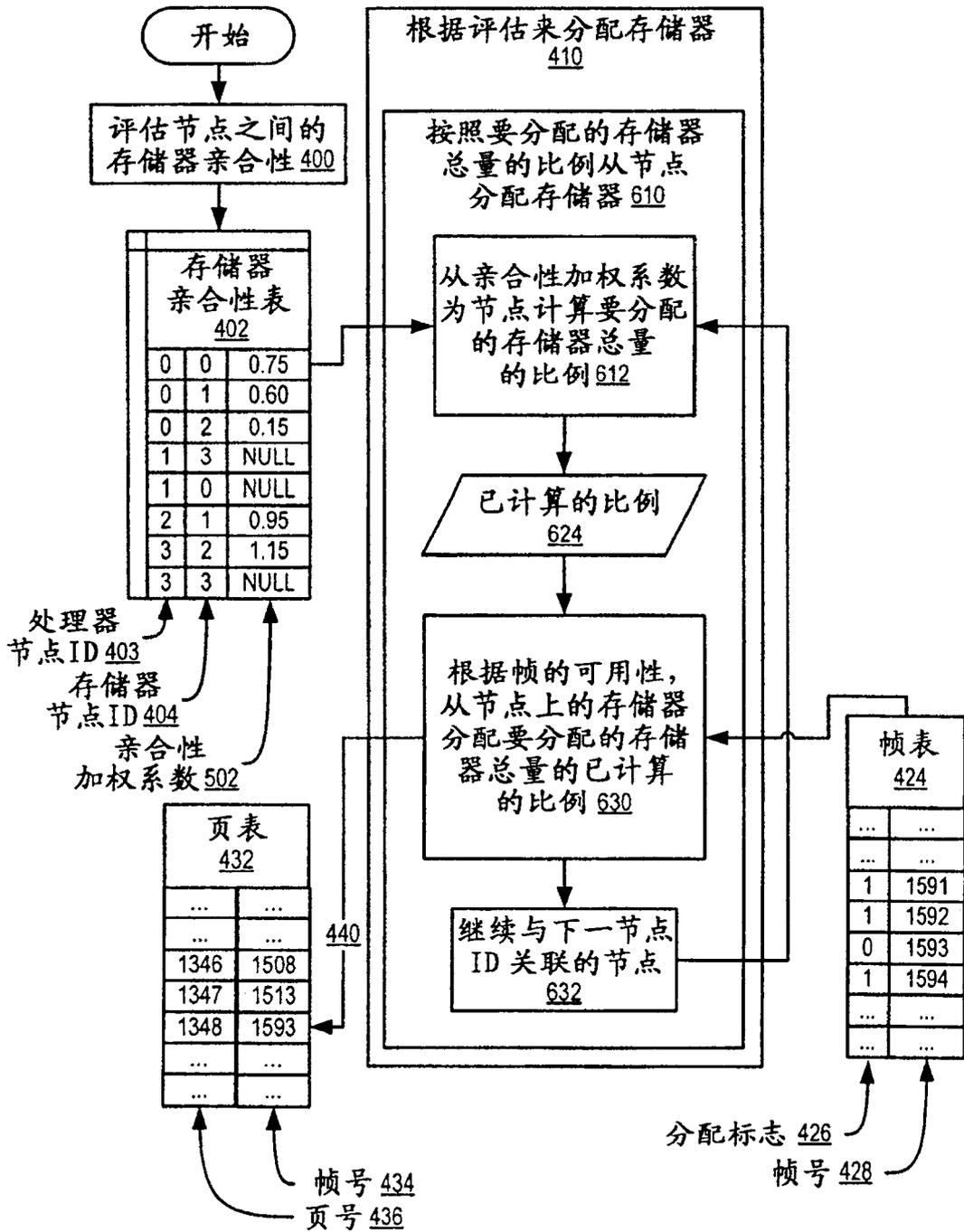


图 5

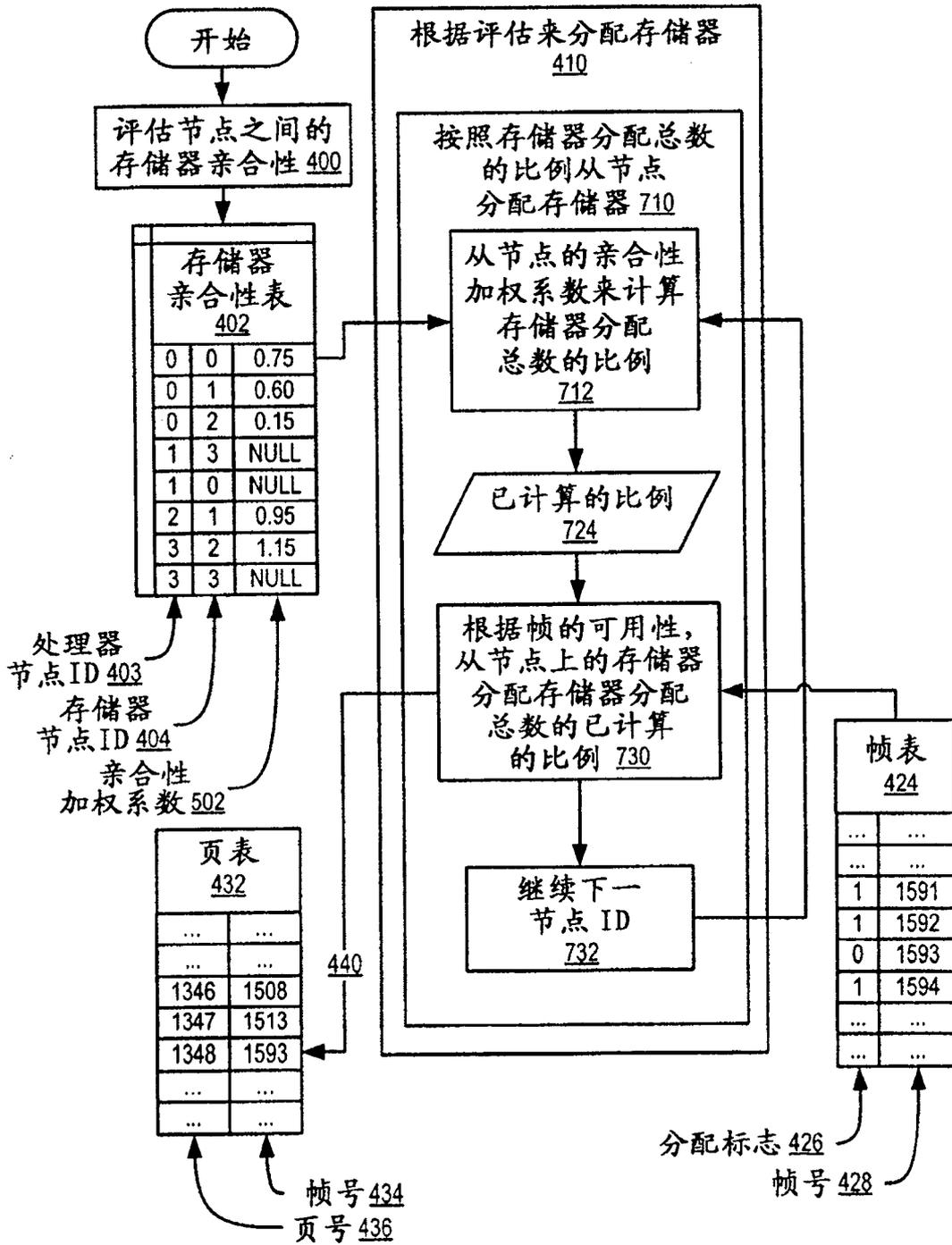


图 6

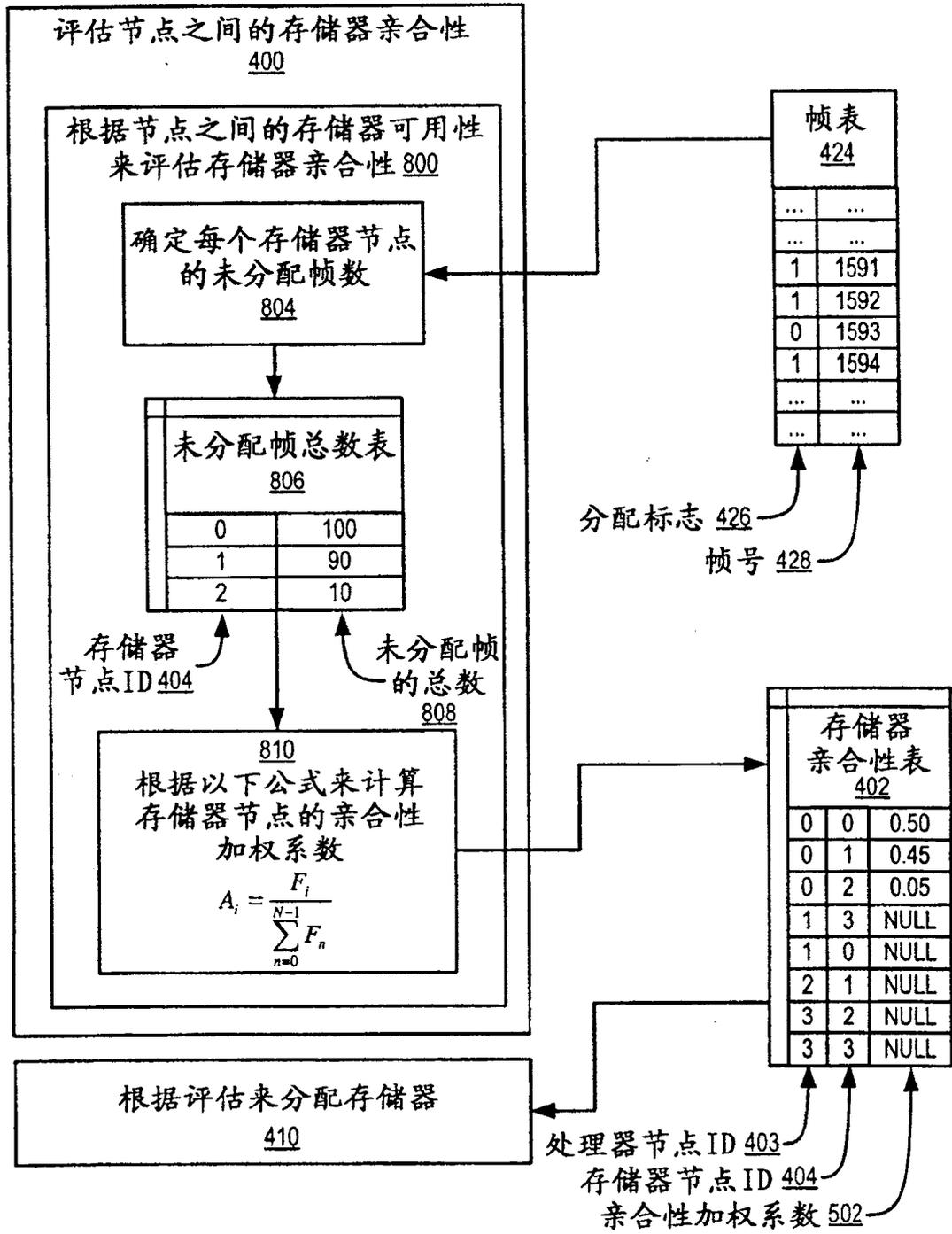


图 7

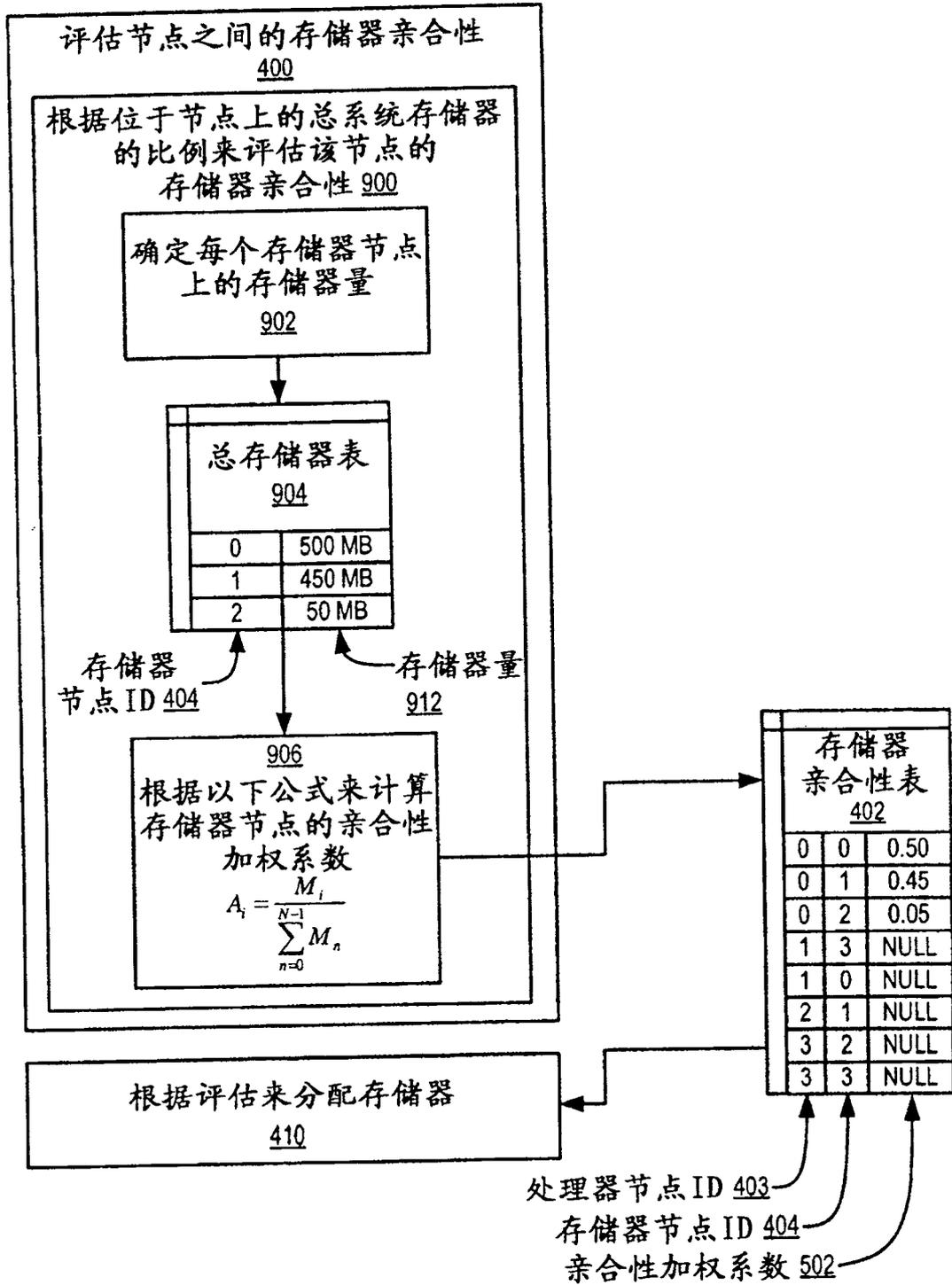


图 8

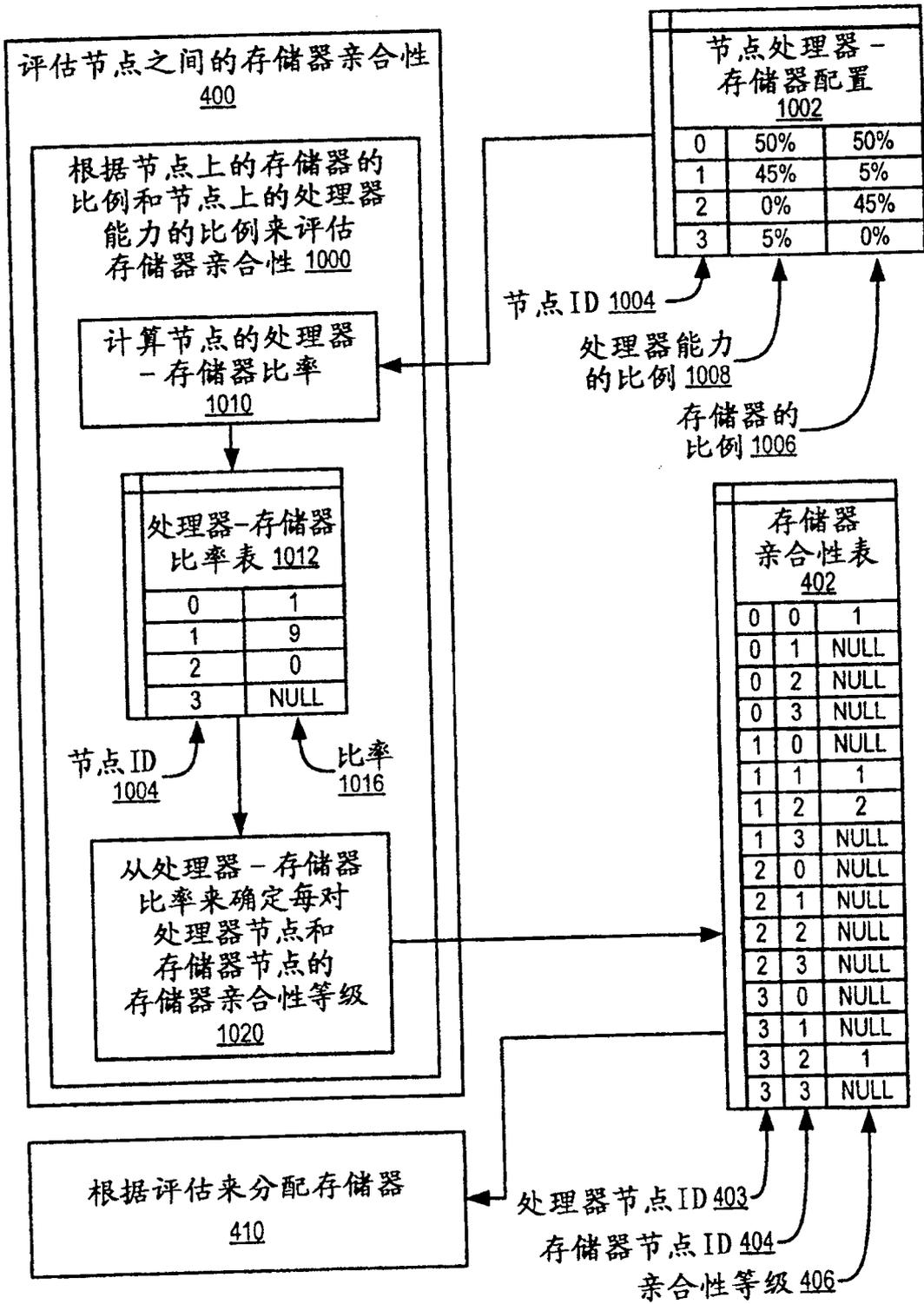


图 9