

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/30 (2006.01)



[12] 发明专利说明书

专利号 ZL 02824989.5

[45] 授权公告日 2006 年 10 月 18 日

[11] 授权公告号 CN 1280760C

[22] 申请日 2002.12.13 [21] 申请号 02824989.5
[30] 优先权
[32] 2001.12.14 [33] FR [31] 01/16204
[86] 国际申请 PCT/FR2002/004351 2002.12.13
[87] 国际公布 WO2003/054735 法 2003.7.3
[85] 进入国家阶段日期 2004.6.14
[71] 专利权人 储存中介公司
地址 法国科洛米耶
[72] 发明人 A·姆佐希 C·吉特尼特
审查员 齐 霁

[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 程天正 陈景峻

权利要求书 4 页 说明书 13 页 附图 2 页

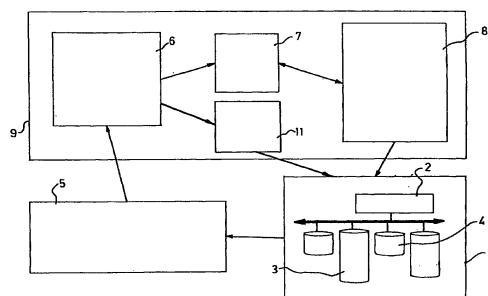
[54] 发明名称

在数据存储装置的异构组中分配对象的方法和
和设备

[57] 摘要

本发明涉及一种在一个数据存储装置的异构组中分配对象的方法,该方法包含将每个对象分解成多个块,按照一个分配准则在不同的存储装置中分配所述块,该分配准则是:在所述存储装置的每一个中分配均包括一个或多个块的对象的片。按照本发明,用于管理对象片的分配的参数是一个称作弹性系数 $CF(i)$ 的系数,代表各对象 (i) 的片的权重之间的差;定期地测量和计算代表每个对象引用率的可变性的值;按照一个向每个对象指派一个与其引用率的可变性成反比例的弹性系数的原则,在时刻 (t) 根据测量和计算的所述值,为每个对象计算一个要被指派给所述对象的期望弹性系数 $CF_v(i)$;在时刻 (t) 为每个对象 (i) 测量和计算所述对象的真实弹性系数 $CF_r(i)$,其代表所述对象各片的权重之间

的差;在存储对象片的各块的装置之间引起一次移动,以便为每个对象 (i) 获得一个与所述对象的期望弹性系数 $CF_v(i)$ 对应的真实弹性系数 $CF_r(i)$ 。



1. 一种在数据存储装置的异构组中分配对象的方法，该方法包含将每个对象分解成多个块，按照一个分配准则在不同的存储装置中分配所述块，该分配准则是：在每个所述存储装置中分配包括一个或多个块中的每一个块的对象片，所述方法的特征在于，它用一个称作弹性系数 $CF(i)$ 的系数，作为管理对象片的分配的参数，弹性系数代表各对象 (i) 的片的权重之间的差，在添加任何新对象时，通过规定一个给定的弹性系数值，来在所述存储装置中分配对象块，与此同时，对于被安置在存储装置中的各对象：
- 5 一定期地测量和计算代表每个对象的引用率的可变性的值；
- 在时刻 t 根据测量和计算的前述值为每个对象计算一个期望弹性系数 $CF_v(i)$ ，该系数按照一个原则而被指派给所述对象，该原则在于：向每个对象指派一个与其引用率的可变性成反比例的弹性系数；
- 在时刻 t 为每个对象 (i) 测量和计算所述对象的真实弹性系数
- 15 $CF_r(i)$ ，其代表所述对象的片的权重之间的差；和
- 在存储对象片的块的装置之间引起一次移动，以便为每个对象 (i) 获得一个与这个对象的期望弹性系数 $CF_v(i)$ 相对应的真实弹性系数 $CF_r(i)$ 。
2. 如权利要求 1 所要求的分配方法，其中，为每个对象 (i) 所计算的真实弹性系数 $CF_r(i)$ ，使得：
- 20 $CF_r(i) = P_{dev}(i) / P_{mean}(i)$
- 其中： $P_{mean}(i)$ 是对象 (i) 的片的权重的平均数，并且 $P_{dev}(i)$ 是关于对象 (i) 的片的权重的平均数的标准偏差。
3. 如权利要求 1 或 2 中任一个所要求的分配方法，其中：
- 25 —在一个预备步骤中，将引用率的可变性的各预定值指派给不同块，将所述块按引用率的可变性的降序来分类，并通过各被相邻地分类的块的关联创建对象；
- 在管理存储装置组的过程中，定期地将各块按引用率的可变性的降序重新分类，所述块作为关于所述块的引用率和可变性的测量信息
- 30 的函数。
4. 如权利要求 1 或 2 中任一个所要求的分配方法，其中，通过一种称为“决策梯形”的决策方法为每个对象 (i) 确定期望弹性系数

CFv(i), “决策梯形”包括一个具有第一底和第二底的梯形, 第一底由一个有刻度的轴组成, 其向量是所述对象的引用率的可变性, 第二底由一个有刻度的轴组成, 其向量是所述对象的期望弹性系数, 所述向量与前一个向量方向相反, 所述决策方法包括如下步骤:

- 5 规定每个变量—即引用率的可变性和期望弹性系数—的上限和下限值, 以便产生一个其两个边分别是由线段 [var 上限-CF 下限] 和 [var 下限-CF 上限] 组成的梯形,

—如果可变性大于或等于 var 上限, 则把这个可变性投影到弹性系数的 CF 下限值上;

- 10 —类似地, 如果可变性小于或等于 var 下限, 则把这个可变性投影到 CF 上限值上;

—如果可变性位于区间 [var 下限-var 上限], 则线性地把这个可变性投影到区间 [CF 下限-CF 上限] 上。

5. 如权利要求 1 或 2 中任一个所要求的分配方法, 其中, 为了为
15 每个对象 (i) 获得一个对应于期望弹性系数 CFv(i) 的真实弹性系数 CFr(i):

—为每个对象 (i) 计算一个等于绝对值 $|CFv(i) - CFr(i)|$ 的、被称作 CF 距离的参数;

- 为 $CFv(i) < CFr(i)$ 的对象 (i), 并且首先为 CF 距离最大的对象,
20 将属于最高权重的片的这些对象的块移向含有最低权重的片的存储装置;

—为 $CFv(i) > CFr(i)$ 的对象 (i), 并且首先为 CF 距离最大的对象, 将这些对象的块从最满的存储装置移向最不满的存储装置。

6. 如权利要求 1 或 2 中任一个所要求的分配方法, 其中:

- 25 —选择最多负荷的存储装置—称作热 SD 和选择最少负荷的存储装置—称作冷 SD;

—在最少负荷的存储装置中查找一个很少被访问的块—称作冷块, 并在最多负荷的存储装置中查找一个频繁被访问的块—称作热块, 致使这两个块能被交换, 以重新平衡负荷;

- 30 —按以下方式对热 SD 中的热块排序:

—将对象按平均引用率的降序分类,

—在这个分类后, 选择形成一个区域的 n 个最流行的对象, 所述

区域被称作高引用率区域，

—将 n 个对象分解为块，这些块被单独地处理，以便对它们排序并将它们分配在四个类中：

5 —类 1：对于其从热 SD 向冷 SD 的移动减少它们所属的对象的真实弹性系数的块；

 —类 2：向最满的存储装置以外的存储装置移动的块；

 —类 3：向最满的存储装置移动的块；

 —类 4：其移动增加 CF 距离的块，

10 —同样地，为了对冷 SD 中的冷块排序，在分类后选择 n 个最不流行的对象，这些对象构成一个被称作低引用率区域的区域，并且这 n 个最不流行的对象被划分成块，这些块被单独地处理，以便对它们排序并将它们分配在前述的四个类中：

 —然后，从热块和冷块组这两个已排序的块组开始：

15 —寻找热 SD 中要被转移的块，该块包括位于所述热 SD 中、其分类号最低且先前已被排序的块，

 —寻找冷 SD 中要被转移的块，该块包括位于所述冷 SD 中、其分类号最低且先前已被排序的块，

 —将所选择的、热 SD 中的块与所选择的、冷 SD 中的块交换。

20 7. 一种用于在存储装置 (3, 4) 的异构组 (1) 中按照一个分配准则分配被分解成多个块的对象 (i) 的装置，该分配准则是：在所述存储装置中分配均包括一个或多个块的对象的片，所述分配装置的特征在于它包含：

 一个称作分析模块 (5) 的模块，用于定期地测量和计算代表每个对象的引用率的可变性的值；

25 一个称作决策模块 (6) 的模块，用于在时刻 t 根据由分析模块 (5) 测量和计算的参数来为每个对象 (i) 计算一个称作真实弹性系数的、代表这个对象 (i) 的各片的权重之间差的系数 $CF_r(i)$ 和一个称作期望弹性系数的、要被指派给对象 (i) 的系数 $CF_v(i)$ ，该系数是按照向每个对象指派一个与其引用率的可变性成反比的弹性系数的原则计算的；

30 一个称作重新配置模块 (8) 的模块，用于在存储装置之间引起一次对象片的各个块的移动，以便为每个对象 (i) 获得一个对应于该对象的期望弹性系数 $CF_v(i)$ 的真实弹性系数 $CF_r(i)$ 。

8. 如权利要求7中所要求的分配装置,其中,它包含一个称作重新平衡模块(11)的模块,用于执行如下操作,这些操作包含对较少负荷的存储装置中很少被访问的块与高负荷的存储装置中频繁被访问的块进行交换。

在数据存储装置的异构组中分配对象的方法和设备

5 本发明涉及一种用于在数据存储装置的异构组中分配对象的方法和装置。

在数据存储装置领域中的当前趋势是在被称作异构存储系统的、日益复杂的存储系统中存储信息。之所以被称作异构存储系统，是因为它们包括存储容量(存储装置中能存储的信息量)不同、或者带宽(能从存储装置器读取或向存储装置写信息的速率)不同的存储装置。

10 按照惯例，在不同存储装置中存储信息的第一步，目前包含将每个对象划分成多个块，并按照某个分配准则在不同存储装置中分配所述块，包括在存储装置的每一个中分配对象片，其中每个对象片由一个或多个块组成。

15 为了清楚起见，在整个文本(说明书和权利要求书)中使用以下术语：

—存储装置：用于记录或获取数据的装置；

—客户机：访问在存储装置中存储的数据的装置；

—数据组的引用率(popularity)：与在此数据组被访问时存储装置与客户机之间每单位时间转移的数据量成正比例的值；

20 —引用率的可变性(variability)：表征数据组的引用率随时间变化的值；

—块：在存储装置中被连续存储的数据组；

—对象：具有可比较的、引用率的可变性的块组；

25 —对象片(piece of object)：所有属于同一个对象并被放置在同一个存储装置中的块组；

—存储装置的负荷：每单位时间在这个存储装置与客户机之间转移的数据组；

30 一片的权重：表示存储此片的存储装置中因为此片导致的工作负荷(workload)的值。该值可用多种方法估量，例如，以按存储装置的平均带宽划分的片尺寸来估量，以存储装置处理对这个片中的数据的读或写请求所用的平均时间来估量，等等。

迄今为止所有开发都是用了基于两个主要原则的分配方法。

第一个方法可称为平衡分配安置法，其根据的原则是，在每个存储装置中分配大小与所述存储装置的带宽成正比的对象片。这个方法因此倾向于充分使用存储装置的带宽。

然而，这个方法的缺点是，有些存储装置比其它存储装置更快变满，因此，一旦一个存储装置是满的，则如果不使安置不平衡，就不再可能在其它存储装置中增加片。因此，在大多数情况下，这个方法并不能使一个异构组的存储容量被充分利用。尽管这种存储容量的损失难以估计，但是由于它是配置的一个函数，所以已经表明存储容量损失一般大于存储容量的 25%，能多达约存储容量的 80%。

第二个方法可称为非平衡分配安置法，其根据的原则是，以充分利用存储装置组的存储容量的方式分配对象块，同时试图平衡这些存储装置上的工作负荷。然而，这个方法的主要缺点是，这种工作负荷的平衡不稳定，因为对象的引用率的变化往往干扰这种平衡。

因此，对这个方法所做的大部分工作在于提出重新平衡工作负荷的解决方案。然而，这种重新平衡消耗带宽，并且实际上导致存储装置能同时处理的访问数的减少。

总之，这第二个方法看起来使存储容量能被充分利用，但是却导致对带宽利用的显著减少（量级在 50%，依所开发的重新平衡方法而改变）。

迄今为止，对用于在数据存储装置的异构组中安置对象的方法所作的所有工作都旨在改善前述的安置方法的一个或另一个，即平衡分配安置法或非平衡分配安置法。

然而，由于与这些安置方法每一个的设计相关联的特定缺点，这种工作全部已经导致了存储容量得不到充分利用的解决方案，或者是在低水平下利用带宽的解决方案。

本发明建议减少这种缺点，其主要目的是提供一种在异构存储装置中分配对象的方法，致使既对存储容量的利用、又对所述存储装置的带宽利用的最优化。

为此，本发明提出一种在数据存储装置的异构组中分配对象的方法，该方法用一个称作弹性系数 $CF(i)$ 的系数，作为管理对象片的分配的参数，弹性系数代表对象 (i) 的各片的权重之间的差，在添加任何新的对象时，通过规定弹性系数的一个给定值，在所述存储装置中分配

对象块，与此同时，对于被安置在存储装置组中的各对象：

—定期地测量和计算代表每个对象的引用率的可变性的值；

—在时刻 t 根据测量和计算的前述值为每个对象计算一个期望的弹性系数 $CF_v(i)$ ，该系数按照一个原则而被指派给所述对象，该原则是：向每个对象指派一个与其引用率的可变性成反比例的弹性系数；

—在时刻 t 为每个对象 (i) 测量和计算所述对象的真实弹性系数 $CF_r(i)$ ，其代表所述对象各片的权重之间的差；

—在存储各对象片的块的装置之间引起一次移动，以便为每个对象 (i) 获得一个与这个对象的期望弹性系数 $CF_v(i)$ 相对应的真实弹性系数 $CF_r(i)$ 。

因此，本发明的基本原理是，用一个最好与引用率的可变性成反比、被称作弹性系数的系数作为管理这些移动的参数，定期地连续修改起初在不同存储装置中分配的对象片的权重。

按照这个原理，弹性系数一般在两个极值 $CF_{\min}$ （最小值）与 $CF_{\max}$ （最大值）之间变化，确定不同的安置方式：

—如果 $CF(i) = CF_{\min}$ ，对象 (i) 的所有片的权重相等，所使用的安置是平衡分配安置；

—如果 $CF(i) = CF_{\max}$ ，则对象 (i) 仅包括被安置在一个存储装置中的一个片，因此安置是纯粹的非平衡分配安置；

—在这两个极值之间，不同的 $CF(i)$ 值允许调整各片的权重，以便通过使 $CF(i)$ 靠近 $CF_{\min}$ 而获得负荷的更好平衡，或者以便通过使 $CF(i)$ 靠近 $CF_{\max}$ 而获得对存储容量的更好利用。

本发明因此在于一种混合安置方法，其主要结果是优先按照非平衡分配安置法（ $CF(i)$ 靠近 $CF_{\max}$ ）安置具有稳定的(stable)引用率的对象，优先地按照平衡分配安置法（ $CF(i)$ 靠近 $CF_{\min}$ ）安置具有不稳定引用率的对象。

实践中，对存储装置组进行的模拟已经显示，按照本发明的方法产生了大于 80% 的存储装置使用率和大于 85% 的带宽使用率。

在一个有益的实施例，按下式为每个对象 (i) 计算真实弹性系数 $CF_r(i)$ ：

$$CF_r(i) = P_{dev}(i) / P_{mean}(i)$$

其中 $P_{mean}(i)$ 是对象 (i) 的各片权重的平均数， $P_{dev}(i)$ 是关于对

象(i)的各片权重的平均数的标准偏差。

在另一个有益的实施例中：

- 在一个预备步骤中，将引用率的可变性的各预定值指派给不同块，将各所述块按引用率的可变性的降序分类，并通过被相邻分类的
- 5 块的关联来创建这些对象；

- 在管理存储装置组的过程中，定期地按引用率的可变性的降序重新分类各块，这些块作为关于所述块的引用率和可变性的测量信息的函数。

- 由于按照本发明的方法特别在于为每个对象(i) 计算理想弹性
- 10 系数（期望弹性系数），并因此在于对对象进行工作，这个有益实施例导致创建包括具有非常相似的、引用率的可变性的块的对象，该实施例使得对所述对象的每一个的工作能被有效进行。

- 此外，并且有益地，通过一种称为“决策梯形”的决策方法为每个对象(i)确定期望弹性系数 $CF_v(i)$ ，“决策梯形”包括一个具有第
- 15 一底和第二底的梯形，第一底由一个有刻度的轴组成，其向量是所述对象的引用率的可变性，第二底由一个有刻度的轴组成，其向量是所述对象的期望弹性系数，所述向量与前一个向量方向相反，所述决策方法包括如下步骤：

- 规定每个变量—即引用率的可变性和期望弹性系数—的上限和下
- 20 限值，以便创建一个其两个边分别是由线段 $[var \text{ 上限}-CF \text{ 下限}]$ 和 $[var \text{ 下限}-CF \text{ 上限}]$ 组成的梯形，

—如果可变性大于或等于 $var \text{ 上限}$ ，则把这个可变性投影到到弹性系数的值 $CF \text{ 下限}$ 上；

- 类似地，如果可变性小于或等于 $var \text{ 下限}$ ，则把这个可变性投
- 25 影到到值 $CF \text{ 上限}$ 上；

—如果可变性位于区间 $[var \text{ 下限}-var \text{ 上限}]$ ，则线性地把这个可变性投影到区间 $[CF \text{ 下限}-CF \text{ 上限}]$ 上。

这个决策梯形构成一个使对象的期望弹性系数能根据该对象的引用率的可变性被立即确定的决策工具。

- 30 此外，这个决策梯形提供一种行为的灵活性，这是因为事实上能按照引用率的可变性和期望弹性系数的上限和下限值的選擇修改全局 (global) 行为。

在另一个有益实施例中, 为了为每个对象(i)获得一个对应于期望弹性系数 $CF_v(i)$ 的真实弹性系数 $CF_r(i)$:

—为每个对象(i)计算一个等于绝对值 $|CF_v(i) - CF_r(i)|$ 的、被称作 CF 距离的参数;

- 5 —为 $CF_v(i) < CF_r(i)$ 的对象(i), 并且首先为 CF 距离最大的对象, 将属于最高权重的片的这些对象的块移向含有最低权重片的存储装置;

—为 $CF_v(i) > CF_r(i)$ 的对象(i), 并且首先为 CF 距离最大的对象, 将这些对象的块从最满存储装置移向最不满(least full)存储装置;

- 10 这种旨在给予每个对象一个与期望弹性系数相等的真实弹性系数的实现方式, 对真实弹性系数正在减少的对象导致带宽上的增加并且对真实弹性系数正在增加的对象导致在存储容量上增加。

- 此外, 这种实现模式能够根据 $CF_v(i)$ 大于还是小于 $CF_r(i)$ 而被划分为两个相互独立的重新配置过程, 该实现模式能够使用小的存储装置
15 的带宽的百分数来获得期望结果 $CF_v(i) \cong CF_r(i)$ 。

而且, 前述这两种重新配置过程有不同的工作空间, 因此不会互相干扰。

在另一个旨在减少由于管理对象块的分配的不完善而引起的工作负荷不均衡的有益实施例中:

- 20 —选择最多负荷的存储装置-称作热 SD 和最少负荷的存储装置-称作冷 SD;

—在最少负荷的存储装置中查找一个很少被访问的块-称作冷块并在最多负荷的存储装置中查找一个频繁被访问的块-称作热块, 致使这两个块能被交换, 以重新平衡负荷;

- 25 —按以下方式对热 SD 中的热块排序:

—将对象按平均引用率的降序分类,

—在这个分类后, 选择形成一个区域的 n 个最流行的对象, 其中所述区域被称作高引用率区域,

- 将 n 个对象分解为块, 这些块被单独地处理, 以便将对它们排
30 序并将它们分在四个类(class)中:

—类 1: 对于其从热 SD 向冷 SD 的移动减少它们所属的对象的真实弹性系数的块;

-类 2: 向最满的存储装置以外的存储装置移动的块;

-类 3: 向最满的存储装置移动的块;

-类 4: 其移动增加 CF 距离的块,

5 一同样地, 为了对冷 SD 中的冷块排序, 在分类后选择 n 个最不流行的对象, 这些对象构成一个被称作低引用率区域的区域, 并且这 n 个最不流行的对象被划分成块, 这些块被单独地处理, 以便对它们排序并将它们分配在前述的四个类中:

一然后, 从热块和冷块组这两个已排序的块组开始:

10 一寻找热 SD 中要被转移的块, 该块包括位于所述热 SD 中、其分类号最低且先前被排序的块,

一寻找冷 SD 中要被转移的块, 该块包括位于所述冷 SD 中、其分类号最低且先前被排序的块,

一将所选择的、热 SD 中的块与所选择的、冷 SD 中的块交换。

15 这个“重新平衡”就是如此被设计来执行较少负荷的存储装置中很少被访问的块与高负荷的存储装置中频繁被访问的块的交换操作。

这个重新平衡过程于是使得在工作负荷上的不平衡能够被重新配置, 同时避免对重新配置过程的干扰问题或者至少将干扰问题最小化。

20 本发明包括一种用于应用按照本发明的方法的装置。本发明因此涉及一种按照分配准则在数据存储装置的异构组中分配被分解成多个块的对象的装置, 该分配准则在于: 在所述存储装置中分配均包括一个或多个块的对象的片, 所述分配装置包含:

一个称作分析模块的模块, 用于定期地测量和计算代表每个对象的引用率的可变性的值;

25 一个称作决策模块的模块, 用于在时刻 t 根据由分析模块测量和计算的参数来为每个对象 (i) 计算一个称作真实弹性系数的、代表这个对象 (i) 的各片的权重之间差的系数 $CF_r(i)$ 和一个称作期望弹性系数的、要被指派给对象 (i) 的系数 $CF_v(i)$, 该系数是按照向每个对象指派一个与其引用率的可变性成反比的弹性系数的原则计算的。

30 一个称作重新配置模块的模块, 用于在存储装置之间引起一次对象片的各个块的移动, 以便为每个对象 (i) 获得一个对应于该对象的期望弹性系数 $CF_v(i)$ 的真实弹性系数 $CF_r(i)$ 。

此外,并且有益地,这个分配装置还包含一个称作重新均衡模块的模块,用于执行如下操作,这些操作对较少负荷的存储装置中很少被访问的块与高负荷的存储装置中频繁被访问的块进行交换。

5 本发明的其它特点、目的和优点,将通过以下结合代表本发明最佳实施例的各附图的详细说明而阐明,这种说明是示例性的而不是限制性的。

图 1 是按照本发明的装置的概要图;

图 2 表示一种确定每个对象(i)的期望弹性系数 $CF_v(i)$ 的方式的示意图;

10 图 3 表示按照本发明方法管理的一个存储装置的异构组所利用的存储容量和所利用的带宽的示意图。

图 1 中所示的按照本发明的装置,用于管理对象在异构存储系统 1 中的安置,该系统常规地包含一个用于管理该系统的控制器 2,其通过一个互连而与多个存储装置 3、4 相连。

15 此装置首先包含一个分析器 5,可用于根据对异构系统 1 所作的测量来计算负荷平衡的质量、对象的引用率的可变性、系统存储容量的占用、以及服务质量的维持。它将这些估计发送到对象弹性控制器 9 的决策部分 6。

20 决策器 6 以能对存储容量的过载或不良占用情况进行纠正的方式来监控在由分析器 5 提供的估计上的变化。例如,它被设计用来在发现负荷质量降低、因此将要导致不良服务质量时,决定例如减少在存储系统 1 中存储的所有对象的弹性。这些决定被发送到一个对象弹性声明 7 中。

25 弹性声明 7 为在存储系统 1 中存储的每个对象将其期望弹性系数 $CF_v(i)$ 和真实弹性系数 $CF_r(i)$ 存储在一起。按照决策器 6 的决定(导致期望弹性系数的修改)并且按照以下说明的重新配置单元 8 的动作(导致真实弹性系数的修改)来更新弹性声明 7。

30 此装置也包含上面提及的重新配置单元 8,该单元被用来(通过发出输入/输出请求)移动数据,使得对象的真实弹性系数靠近由决策器 6 设定的期望值。

然而应当注意,弹性控制器 9 的决定不可能是完善的,理由如下:
一分析器 5 不能进行完全精确且即时的测量。

—引用率是不断变化的：弹性控制器 9 在任何时刻 t 所作的工作都是为了解决在时刻 $t-1$ 所测量的情况问题。这个工作实际上将影响系统 1 在时刻 $t+1$ 的性能。如果引用率在时刻 $t-1$ 与 $t+1$ 之间已经变化，则灵活安置将没有最佳效果，因此有些负荷将失去平衡。

- 5 —没有引用率保持恒定的对象。因此，一旦一个安置不再是对所有对象来说是平衡地分配的，则发生不平衡。

这就是灵活安置需要一个动态重新平衡器 11 的支持的原因。但是，与在非平衡分配安置的情形中所发生的情况比起来，重新平衡器要做的工作甚少：它只需要弥补由于弹性控制器 9 的不完善所引起的不平衡。

灵活安置的质量因此受制于这四个机制的性能：决策器 6、重新配置单元 8、重新平衡器 11 和分析器 5。

为了达到这个安置质量，必须将这些机制设计为使得：

- 15 —它们向每个对象指派正确的弹性系数：这是因为，如果使用与所要求的弹性系数相反的弹性系数，结果可能比非平衡分配安置的结果更糟糕；

—它们把弹性控制器的反应时间限制到最小：分析器 5 以这样的方式检测变化，其中，系统 1 反应(周期 d_1)，然后决策器 6 指派期望弹性系数(周期 d_2)，然后重新配置单元 8 移动数据，以提供期望弹性系数与真实弹性系数之间的一个匹配(周期 d_3)。 d_2 非常短， d_1 适度地短(有些分析要求系统 1 监控特定的时间才能使分析是合适的)，如果需要许多的重新配置移动，则 d_3 可以长。随着 $d_1+d_2+d_3$ 的长度的增加，负荷变得更不平衡。

- 25 为了满足这些目的，下面详细说明混合安置机制的每一个的操作。

首先，决策器 6 用以下准则处理由分析器 5 测量和/或计算的参数：

- 存储容量的利用：自由的存储容量应当是适当分布的，换言之，它的排列应当把能被添加到系统 1 的数据量最大化（给定每个新存储对象的初始弹性系数）；

—引用率的可变性：对象的可变性越高，不平衡负荷的风险越大；

—满足服务质量的约束条件：如果某对象相关联的服务质量约束

条件不被满足（例如，如果不以足够的数据速率为显示屏服务，从而导致显示屏闪烁），则必须通过减少这个对象的弹性系数，使得它被平衡地分配并因此从更大的带宽得到好处，使服务质量得到稳定。如果一个对象组的服务质量是混乱的，则可以通过全局减少所有对象的弹性系数而使服务质量得到稳定。

一负荷平衡的质量：如果全局的平衡不良，则必须减少数据的全局弹性系数。

决策器 6 因此被设计为用来为每个对象 (i) 计算将是期望弹性系数 $CF_v(i)$ 的理想弹性系数。

10 首先，当决策器 6 对对象工作时，并且为了最优化决策机制，每个对象最好由具有相同量级引用率的可变性的块组成。

为此，在一个预备步骤中，将以随机方式或者作为从以前调查中获得的一个函数而确定的引用率的可变性的值，指派给各块，然后将所述块按引用率的可变性的降序分类。

15 然后通过被相邻分类的块的关联创建各对象。例如，第一个对象可以由头 n 个分类的块组成，第二个对象可以由接着的 n 个分类的块组成，如此等等。

当这个对象创建已经完成时，系统被启动，按照由分析器 5 获得的关于各块的引用率和可变性的信息，定期地按引用率的可变性的降序重新分类各块。

类似地，特别是在一个稳定化步骤中，决策器 6 对内容不断变化的各对象工作，所述对象包括含有属于不同文件的不同数据的块，但是具有非常相似的、引用率的可变性，使决策器 6 能有效地对每一个对象工作。

25 决策方法使得能够为每个对象确定期望弹性系数，作为分析器 5 所监测的引用率的可变性的一个函数，该决策方法接着以图 2 中所示的“决策梯形”的使用为基础。

这个梯形的其中一个底由一个有刻度的轴组成，其向量是引用率的可变性；第二个底由一个有刻度的轴组成，其向量是期望弹性系数，所述向量与前一个向量方向相反。

30 这个决策方法的第一个步骤在于，规定每个变量—即引用率的可变性和期望弹性系数—的上限和下限值，以便创建图 2 中所示的其两

个边分别是由线段 [var 上限-CF 下限]和[var 下限-CF 上限]组成的梯形。定义了这个决策梯形后，该决策方法由下列步骤组成：

- 如果可变性大于或等于 var 上限，则把这个可变性投影到弹性系数的值 CF 下限上，换言之，将一个其可变性大于或等于 var 上限的对象的期望弹性系数设定在值 CF 下限；

—类似地，如果可变性小于或等于 var 下限，则把这个可变性投影到值 CF 上限上；

—如果可变性位于区间[var 下限-var 上限]，则线性地把这个可变性投影到区间[CF 下限-CF 上限]上。

- 这个决策梯形因此构成一个使对象的期望弹性系数能根据该对象的引用率的可变性被立即确定的决策工具。

此外，还能按照引用率的可变性和期望弹性系数的上限和下限值的選擇修改决策器 6 的行为。

因此，例如：

- CF 下限和 CF 上限值的增加(在图 2 中向右移动)，导致由决策器 6 生成的期望弹性系数的全局增加，因此导致对存储容量的更好利用，同时在带宽方面有不利；

- 相反，CF 下限和 CF 上限的减少导致存储系统的带宽的利用的最优化，同时在存储容量方面有不利；对 var 上限和 var 下限值也能做出类似的结论；

- CF 下限的减少和 CF 上限的增加(这两个限度之间间距的增加)将导致对象处理中的更大差异：稳定的对象将更易被“弹性化”(通过它们真实弹性系数的增加)，而易变的对象将更易被“固定化”(通过它们真实弹性系数的减少)。因此，弹性控制器 9 的能力得到增加。相反地，通过将弹性系数区间的两个限度向一起更靠近，也会减少弹性控制器 9 的这个能力。

决策梯形因此使弹性系数能被指派给要每个被逐个确定的对象，同时也使对所有对象的全局策略能按照介入存储系统 1 的操作中的事件而被修改。

- 因此，例如，如果存储系统 1 的存储装置 3 或 4 出故障，则一部分带宽就丧失，因此就不再能维持服务质量。在这种情况下，可以修改决策梯形，以便以存储容量方面的不利为代价，增加带宽，以便在

替换出故障存储装置的同时，暂时重新建立可接受的服务质量。

第二，重新配置单元 8 适于这样进行操作，使得每个对象(i)的真实弹性系数 $CF_r(i)$ 等于由决策器 6 确定的期望弹性系数 $CF_v(i)$ ，以便为真实弹性系数减少的对象增加带宽，并为真实弹性系数增加的对象增加存储容量。为此，首先，这个重新配置单元 8 的操作方法使用一个等于绝对值 $|CF_v(i) - CF_r(i)|$ 的、被称作“CF 距离”的参数。

这个操作方法可以被划分成两个过程：

—一个被称作 BW 重新配置的重新配置过程，被设计为用来增加带宽，并适用于其期望 CF 小于真实 CF、因此需要被固定化的对象；为此，这个过程包括对于要被固定化的对象而言，将一个属于最高权重的片的块向含有最低权重的片的存储装置移动；这个过程也被设计为首先处理 CF 距离最大的对象；

—一个被称作 SC 重新配置的重新配置过程，被设计为用来增加存储容量，并适用于其期望 CF 大于真实 CF、因此需要被弹性化的对象；为此，这个过程包含选择具有最大 CF 距离的对象，将这些对象从最满存储装置向最不满存储装置移动。

两个不同过程的这种选择，使得有可能在使用存储系统的小百分率带宽的同时，很好地利用决策器 6 的决策能力。还应当注意的是，BW 重新配置过程作用于需要被固定化的对象，而 SC 重新配置过程则作用于需要被弹性化的对象。因此，这两个过程有不同的工作空间，因此不能互相干扰。

最后，如上文所提及的那样，重新平衡器 11 的目的是减少因弹性控制器 1 的不完善而产生的工作负荷不平衡。

为此，重新平衡方法被设计为用来提供包含交换位于较少负荷的存储装置(冷 SD)中很少被访问的块(冷块)与位于高负荷的存储装置(热 SD)中频繁被访问的块(热块)的操作。

这个与重新配置完全分开的重新平衡方法，要被设计来避免或者至少最小化任何对重新配置过程—特别是 BW 重新配置过程—的干扰问题，因此以一种特殊方式操作。

这个重新平衡方法需要一种按照由分析器 5 传送的信息而变化的特殊分类方法，这在下文作解释。

首先，选择最高负荷的存储装置(热 SD)和最少负荷的存储装置

(冷 SD)。

为了把一个热块从热 SD 向冷 SD 移动, 将对象按平均引用率的降序分类, 在这个分类后, 选择 n 个最流行的对象, 其中 n 是预定的, 例如等于 25, 这构成一个引用率的“高区域”。

- 5 这些被选择的对象然后被分解成块, 这些块通过被排序并被分配在以下四个类中而被单独地处理:

-类 1: 对于其从热 SD 向冷 SD 的移动减少它们所属的对象的真实弹性系数、因此导致这些对象的固定化的块。因此, 这个类 1 是指不产生与 BW 重新配置有冲突风险的块。

- 10 -类 2: 向最满存储装置以外的存储装置移动的块(换言之, 当冷 SD 不同于最满的存储装置时); 因此, 这个类 2 是指不引起与 SC 重新配置有冲突风险的块。

-类 3: 向最满存储装置移动的块(与类 2 相反的分类), 因此, 该类能引起与 SC 重新配置有冲突的风险。

- 15 -类 4: 其移动增加 CF 距离的块(与类 1 相反的分类), 该类可能引起与 BW 重新配置有冲突的风险。

同样地, 为了处理一个冷块从冷 SD 向热 SD 的移动, 对构成一个“低引用率”区域的 n 个最不流行对象的块进行相同的分类, 将它们分在与相同的四个类中。

- 20 在这个分类后, 平衡方法的步骤如下:

- 1) 寻找热 SD 中要被转移的块, 该块属于一个被分类在高引用率区域的对象。为此, 选择热 SD 中的分类号且有最低分类号(采用优先级的分类 1)、并且其对象具有最小 CF 距离的块。应当注意的是, 对最小 CF 距离的这种选择对于位于类 3 中的块而言导致-特别是导致减少与 SC 重新配置之间有冲突的风险, 以及对于位于类 4 中的块而言导致与 BW 重新配置之间有冲突的风险, 因为这些重新配置导致实际上仅仅移动具有大的 CF 距离的对象的数据。

2) 寻找冷 SD 中要被转移的块, 这个块属于一个被分类在低引用率区域的对象(该寻找过程与上面的步骤 1 的寻找过程类似)。

- 30 3) 将在步骤 1) 中选择的热 SD 中的块与在步骤 2) 中所选择的冷 SD 中的块交换。

导致“灵活的安置”的弹性控制器 9 的设计与这个重新平衡过程

的组合，导致对存储系统的存储容量的利用和带宽的利用的同时最优化，如以下结合图 3 所述的使用例子中所示的那样。

所使用的存储系统 1 包含两个具有 18.6GB 的存储容量和 34.2MB/S 的带宽的硬盘、以及两个具有 43GB 的存储容量和 28.8MB/S 的带宽的硬盘。

在表示结果的图 3 中，按时间计的操作时间在横坐标上表示，而带宽的和存储容量的利用率则在纵坐标上表示。

在操作的开始，以随机方式填充存储系统 1。如图 3 中的曲线所示，在初始的稳定化阶段，弹性控制器 9 趋向于释放存储容量，与此同时，尽可能地稳定带宽。随后，由分析器 5 提供的信息变得逐渐地越来越精确并且越来越代表各对象的特征，使决策器 6 能计算越来越合适的期望弹性系数。

对于每个对象来说，这些期望弹性系数也逐渐接近真实弹性系数的值，作为重新配置单元 7 的动作的结果。

在这个稳定化阶段的结尾，弹性控制器 9 提供一段巡航状态 (cruising regime)，在此期间实现 77% 的存储容量使用率和 89% 的带宽使用率。

相比之下，“平衡分配”安置在相同的条件下提供 100% 的带宽使用率、但是仅提供 35% 的带宽使用率。

因此，与已知的各安置方法相比，按照本发明的“弹性交织” (flexible interfacing) 的带宽和存储容量的使用率二者都高，但并不导致“牺牲”这两个使用率的一个来提高另一个。

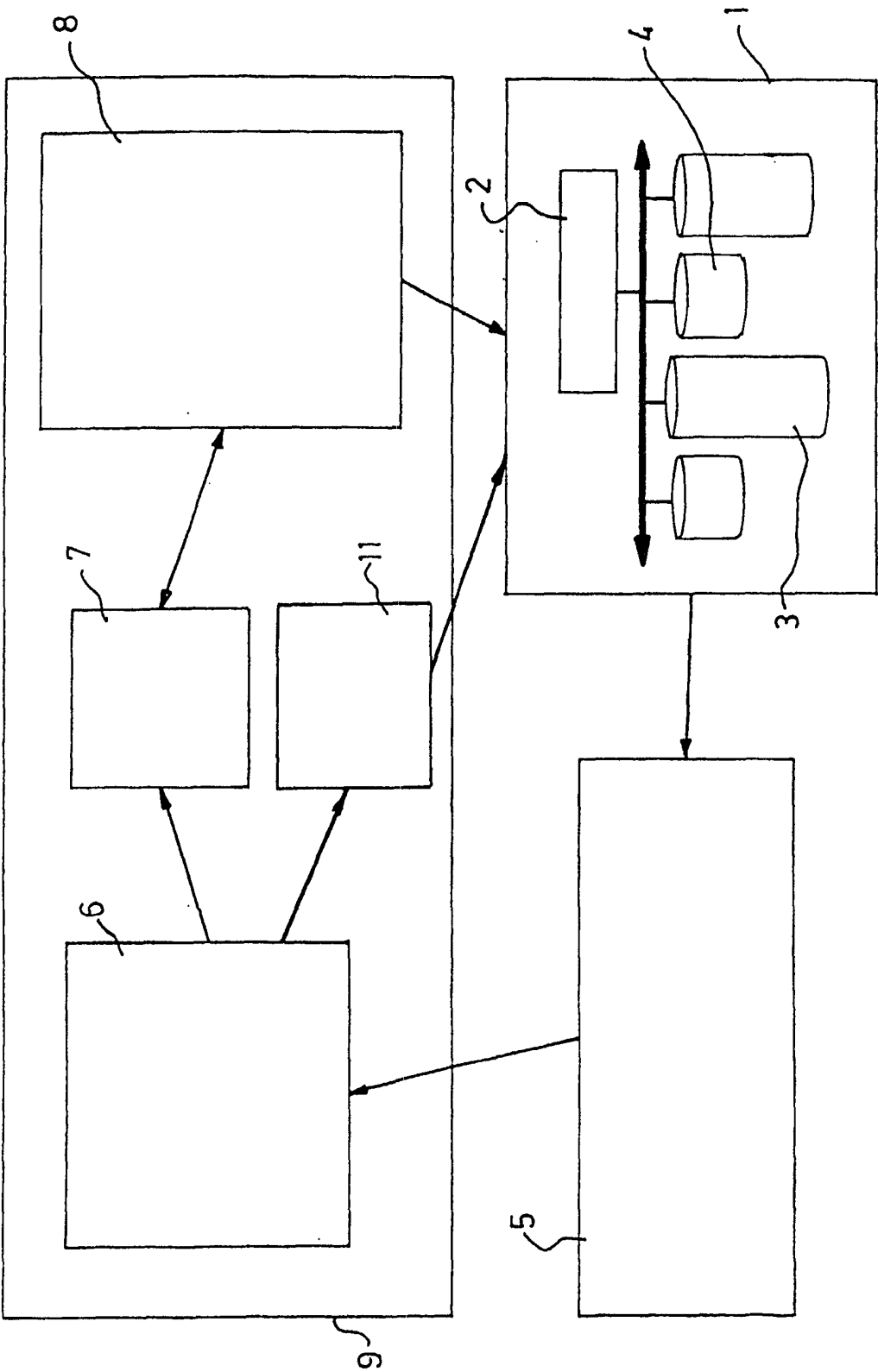


图 1

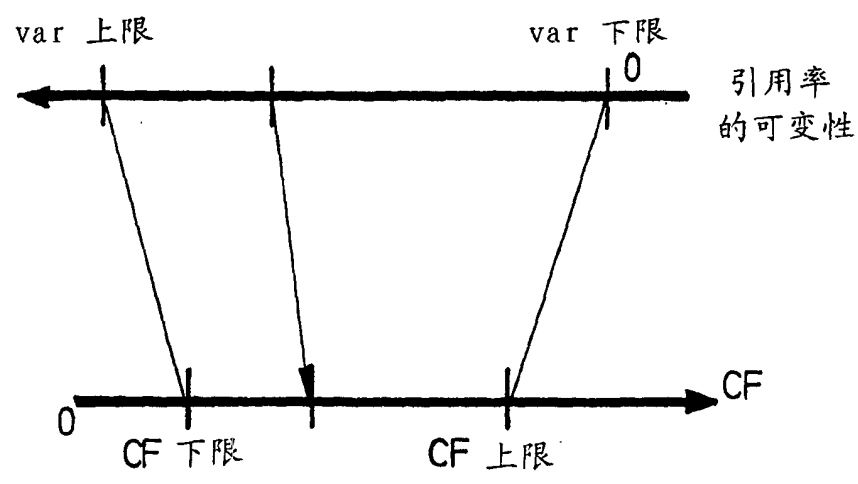


图 2

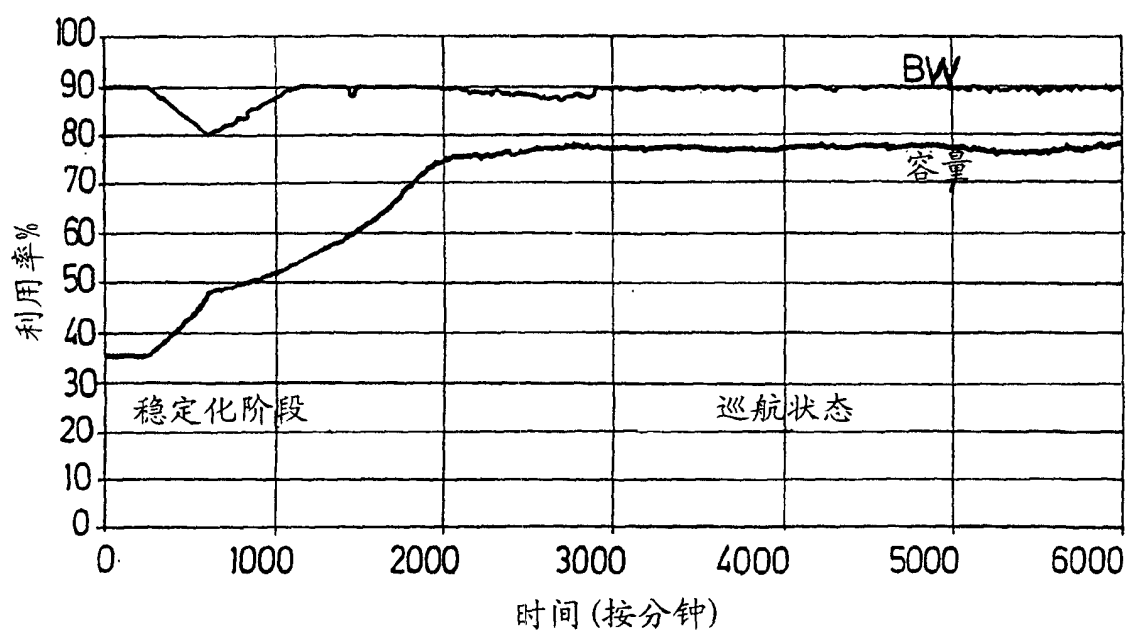


图 3