



(12) 发明专利

(10) 授权公告号 CN 102193746 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201010259796. 4

US 2008/0089390 A1 , 2008. 04. 17,

(22) 申请日 2010. 08. 23

US 6516425 B1 , 2003. 02. 04,

(30) 优先权数据

审查员 杨牛

12/721, 712 2010. 03. 11 US

(73) 专利权人 LSI 公司

地址 美国加利福尼亚州米尔皮塔斯市

(72) 发明人 安德鲁·J·斯派瑞 凯文·肯德尼

(74) 专利代理机构 北京律盟知识产权代理有限公司
11287

代理人 王田

(51) Int. Cl.

G06F 3/06(2006. 01)

G06F 11/14(2006. 01)

(56) 对比文件

CN 101611376 A , 2009. 12. 23,

US 2006/0218422 A1 , 2006. 09. 28,

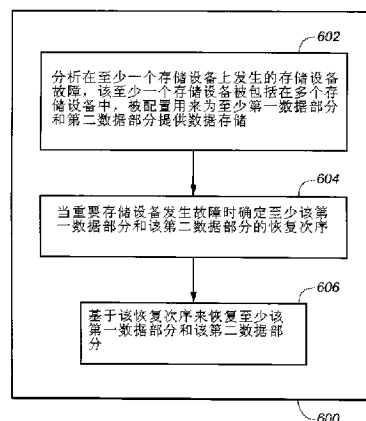
权利要求书2页 说明书7页 附图8页

(54) 发明名称

在分布式数据布局环境中优化冗余恢复的系统和方法

(57) 摘要

本发明涉及一种在分布式数据布局环境中优化冗余恢复的系统和方法。所述系统可包括多个被配置用以提供数据存储的存储设备。所述系统可包括优先级模块, 其被通信地耦合到所述多个存储设备上。当紧要数据发生故障时, 可配置所述优先级模块以确定至少第一数据部分和第二数据部分的恢复次序。所述系统可包括恢复模块, 其被通信地耦合到所述多个存储设备和所述优先级模块上, 所述恢复模块被配置用以基于所述恢复次序来恢复至少所述第一数据部分和所述第二数据部分。



1. 一种恢复存储在多个存储设备上数据的方法,包括:

映射一个或多个物理存储设备的地址到虚拟廉价磁盘冗余阵列 (RAID) 的一个或多个虚拟卷数据块;

创建一恢复列表,其包括一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块;

根据所述虚拟廉价磁盘冗余阵列 (RAID) 的一个或多个虚拟卷数据切片冗余等级,优化包括所述一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的所述恢复列表;和

根据所述恢复列表,将所述一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的数据重建到至少一个替代物理存储设备。

2. 如权利要求 1 所述的方法,所述虚拟 RAID 具有虚拟驱动器的数量少于所述物理存储设备的数量。

3. 如权利要求 1 所述的方法,其特征在于,所述创建包括一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的恢复列表包括:

从元数据处将一列标记拷贝到存储在所述故障物理存储设备上数据块。

4. 如权利要求 1 所述的方法,其特征在于,所述创建包括一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的恢复列表包括:

应用一映射算法以判定所述虚拟 RAID 的哪个数据块被存储在所述一个或多个故障物理存储设备上,所述映射算法与映射一个或多个物理存储设备的一或多个地址到所述虚拟 RAID 的一个或多个虚拟卷数据块相关联。

5. 一种恢复存储在多个存储设备上数据的方法,包括:

映射一个或多个物理存储设备的地址到虚拟廉价磁盘冗余阵列 (RAID) 的一个或多个虚拟卷数据块;

创建一恢复列表,其包括一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块;和

根据所述虚拟廉价磁盘冗余阵列 (RAID) 的一个或多个虚拟卷数据切片冗余等级,优化包括所述一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的所述恢复列表,包括:

比较与所述一个或多个故障物理存储设备相关联的第一虚拟切片的虚拟卷数据块的数量与第二虚拟切片的虚拟卷数据块的数量;

设定具有与所述一个或多个故障物理存储设备相关联的最大数量的虚拟卷数据块的至少一个虚拟卷块在所述恢复列表中为最高优先级;和

根据所述恢复列表,将所述一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的数据重建到至少一个替代物理存储设备。

6. 如权利要求 1 所述的方法,其特征在于:所述虚拟 RAID 的一个或多个虚拟卷数据形成镜像虚拟卷。

7. 一种用于恢复存储在多个存储设备上数据的硬件系统,包括:

用于映射一个或多个物理存储设备的地址到虚拟廉价磁盘冗余阵列 (RAID) 的一个或多个虚拟卷数据块的装置;

用于创建一恢复列表的装置,所述恢复列表包括一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块;

用于根据所述虚拟廉价磁盘冗余阵列(RAID)的一个或多个虚拟卷数据切片冗余等级优化包括所述一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的所述恢复列表的装置;和

用于根据所述恢复列表将所述一个或多个映射到一个或多个故障物理存储设备的虚拟卷数据块的数据重建到至少一个替代物理存储设备的装置。

在分布式数据布局环境中优化冗余恢复的系统和方法

技术领域

[0001] 本发明总体涉及分布式存储设备领域,特别是涉及一种在分布式数据布局环境中优化冗余恢复的系统和方法。

背景技术

[0002] 存储设备,如 RAID 镜像架构,允许数据被存储并防止潜在数据丢失。但是,如果多个存储设备在同一架构中发生故障,在所述存储架构有机会重建或恢复所述数据之前,可能潜在地丢失了所述数据。

发明内容

[0003] 相应的,本发明的一个实施例描述了一种恢复存储在多个存储设备上的数据的系统。所述系统可包括多个被配置用来提供数据存储的存储设备。所述系统可包括优先级模块,所述优先级模块被通信地耦合到所述多个存储设备上。当紧要数据发生故障时,可配置所述优先级模块以确定至少第一数据部分和第二数据部分的恢复次序。所述系统可包括耦合到所述多个存储设备和所述优先级模块的恢复模块,所述恢复模块可被配置以基于所述恢复次序来恢复至少所述第一数据部分和所述第二数据部分。

[0004] 本发明同样描述了一种恢复存储在多个存储设备上的数据的方法。所述方法可包括分析发生在至少一个存储设备上的存储设备故障,所述至少一个存储设备被包括在多个被配置用来为至少第一数据部分和第二数据部分提供数据存储的存储设备中。所述方法可包括当紧要数据发生故障时确定至少所述第一数据部分和所述第二数据部分的恢复次序。所述方法可包括基于所述恢复次序来恢复至少所述第一数据部分和所述第二数据部分。

[0005] 本发明描述了一种具有计算机可执行指令的计算机可读介质,所述计算机可执行指令用于执行一种恢复存储在多个存储设备上的数据的方法。所述方法可包括至少映射第一虚拟数据块到第一存储设备和映射第二虚拟数据块到第二存储设备,所述第一存储设备和所述第二存储设备被包括在多个存储设备中,所述第一虚拟数据块与第一虚拟数据切片相关,以及所述第二虚拟数据块与第二虚拟数据切片相关。所述方法可包括检测所述第一存储设备的故障。所述方法可包括确定零驱动冗余事件是否发生在所述第一虚拟数据切片上。所述方法可包括当所述零冗余事件发生在所述第一虚拟数据切片上时,在所述第二虚拟数据块被恢复到第二替代存储设备之前,将所述第一虚拟数据块恢复到第一替代存储设备上。

[0006] 应所述理解的是,上述概括的描述和后面详细的描述都仅仅是示范性和说明性的,并且不仅限于当前发明要求的内容。包括并组成说明书一部分的附图揭示了本发明的主题。同时,说明书和附图用来解释本发明的原理。

附图说明

[0007] 通过参考附图,所述发明的众多优点可以被本领域的技术人员更好地理解:

- [0008] 图 1 是分布式 RAID(廉价磁盘冗余阵列) 数据布局的一个示范性实施例 ;
- [0009] 图 2A 是描述在分布式数据布局环境中优化冗余恢复的系统实施例的方框图 ;
- [0010] 图 2B 是描述在分布式数据布局环境中优化冗余恢复的系统的替代实施例的方框图 ;
- [0011] 图 3 是图 1 中描述两个存储设备故障的分布式 RAID(廉价磁盘冗余阵列) 数据布局的一个示范性实施例 ;
- [0012] 图 4 是分布式镜像数据布局的一个示范性实施例 ;
- [0013] 图 5 是图 4 中描述两个存储设备故障的分布式镜像数据布局的一个示范性实施例 ;
- [0014] 图 6 是描述恢复存储在多个存储设备上的数据的方法的流程图 ;
- [0015] 图 7 是描述驱动器故障特性方法的流程图 ;
- [0016] 图 8 是描述紧要恢复方法的流程图 ;
- [0017] 图 9 是描述背景恢复方法的流程图。

具体实施方案

[0018] 这里将参照相关附图中所描述的内容来详细揭示本发明的主题。

[0019] 总体上参考图 1-9, 本发明涉及一种在分布式布局环境中优化冗余恢复的系统 100。所述系统 100 可包括多个存储设备 102(图 1 中的 Drv0, Drv1, Drv2, Drv3, Drv4, Drv5, Drv6 和 Drv7), 用来存储数据。每一个存储设备 102 可以是物理上独立的存储设备。例如, 多个存储设备 102 可以都被包含在同一个存储系统中。相反地, 每一个存储设备 102 可以被定位在一个与其它存储设备 102 不同的位置。此外, 一些存储设备 102 可以在一个地理位置, 而其它存储设备 102 可以在其它地理位置。多个存储设备 102 可以是一个 RAID(廉价磁盘冗余阵列) 配置。每一个存储设备 102 可以被通信地耦合到其它存储设备 102 上。

[0020] 所述系统 100 可包括通过所述多个存储设备 102 所形成的虚拟卷 104(即, 存储虚拟化)。所述虚拟卷 104 可提供计算机设备访问所述虚拟卷 104 中数据。所述虚拟卷 104 可以是一个虚拟盘 RAID 组 104。所述虚拟卷 104 可包括多个虚拟驱动器 106(图 1 中的 VDrv0, VDrv1, VDrv2, VDrv3 和 VDrv4)。所述虚拟卷可以数据字或数据切片的形式将数据存储在所述虚拟驱动器 106 上。系统 100 可设置所述数据字大小或数据切片大小。例如, 单个数据切片可以被称作跨度行 0(如, 数据块 (DC) (0,0,0), DC(0,0,1), DC(0,0,2), DC(0,0,3), DC(0,0,4)) 的数据部分。每一数据部分(如, DC(0,0,0), DC(0,0,1), 等) 可被称为一数据块。可通过卷、列和行索引来识别所述数据块。所述数据块大小可等于存储在一个虚拟盘上的 RAID 条带(即, DC(0,0,0) 的大小)。以下, 术语数据可被称为数据切片, 以及术语数据部分可被称为数据块。

[0021] 如图 1 所示, 所述多个存储设备 102 在数量上超过所述虚拟卷 104 中虚拟驱动器 106 的个数。这会允许所述数据块被稀疏地分布在所述多个存储设备 102 上, 其可以允许优化的冗余恢复。

[0022] 图 2A, 描述了所述系统 100 的可能实施例。客户端系统 202 可以与所述虚拟卷 104 进行通信。所述虚拟卷 104 可以与网络虚拟化设备 204 进行通信。所述网络虚拟化设备可以驻留在存储区域网络 200 中。所述网络虚拟化设备 204 可以与所述多个设备 102 进行通

信。映射算法可以提供虚拟卷寻址方案,其可以允许所述数据块从所述虚拟卷 104 映射到所述多个存储设备 102,基于所述虚拟卷寻址方案。可将所述映射算法加入到一个映射程序或模块 206 中。所述映射算法可确定存储每一个数据块的存储设备 102。可以预料到,所述映射算法可以是任意算法,只要所述算法是可重复的并只从一个数据切片中映射至多个数据块到同一个存储设备 102。

[0023] 所述系统 100 可包括标记的恢复列表 208 或恢复队列 208,所述标记用于需要被重建以获取全冗余的数据块。所述恢复列表 208 可被通信地耦合到所述多个存储设备 102 和所述虚拟卷 104 上。所述标记可以包括指针或其类似。在一个存储设备 102 发生故障时,一个新恢复列表 208 会被创建,列出用于需要从所述故障存储设备 102 中恢复的每一个数据块的标记。特别地,所述恢复列表 208 可以通过从元数据处将一系列标记拷贝到存储在所述故障存储设备 102 上数据块的方式进行创建。可选地,所述恢复列表 208 可以通过在所述虚拟卷 104 上运行所述映射算法以确定哪些数据块被存储在所述故障存储设备 102 上以及通过基于上述确定而创建一系列标记的方式进行创建。所述恢复列表 208 可以驻留在所述网络虚拟化设备 204 上。

[0024] 所述系统 100 可包括一个优先级模块 210,其被通信地耦合到所述多个存储设备 102 和所述虚拟卷 104 上。所述优先级模块 210 可基于紧要数据故障来确定所述数据块的恢复次序。紧要数据故障以下可被称为紧要切片故障。紧要切片故障可能发生在数据切片处于当另一个存储设备 102 发生故障时被丢失的危险中。紧要切片故障也可被称为零驱动冗余。

[0025] 紧要切片故障或零驱动冗余的一个例子在图 3 中进行了描述。在当前实施例中所述,数据切片在所述数据切片丢失前能承担两个数据块丢失。所述丢失的数据块可以通过多个 RAID 技术,例如在 RAID 6 配置中的奇偶校验和 / 或分条,来进行重组或重建。图 3 描述了当存储设备 Drv2 和 Drv3 已经发生故障的情况。被存储在存储设备 Drv2 和 Drv3 上的数据块已经被丢失。来自存储设备 Drv2 和 Drv3 和每一个数据块必须在全冗余已经恢复之前进行重组。恢复列表 208 可被创建,列出用于来自存储设备 Drv2 和 Drv3 的将被恢复数据块的标记。行 4 的数据切片可能处于被丢失的危险中。例如,如果另一个存储设备 102 发生故障,行 4 的数据切片会由于无法重组行 4 中包含的丢失数据块而被丢失或毁坏。这样类型的情况被称为紧要切片故障或零驱动冗余。所述优先级模块 210 可以修改所述恢复列表 208,通过将 DC(0,4,1) 的标记或 DC(0,4,4) 的标记之一,或 DC(0,4,1) 的标记和 DC(0,4,4) 的标记一起,放置在所述恢复列表 208 的顶部。在所述恢复列表 208 中,由于所述 DC(0,4,1) 的标记具有较早的列识别符,所述 DC(0,4,1) 的标记可以被放置在所述 DC(0,4,4) 的标记之前。本领域技术人员将认识到,仅仅 DC(0,4,1) 或 DC(0,4,4) 之一必须被重建以恢复单个驱动冗余。所述数据块,DC(0,4,1) 或 DC(0,4,4),会被重建到一个替换存储驱动器。

[0026] 一般地参考图 2A,所述系统 100 可包括恢复模块 212,其通信地耦合到所述多个存储设备 102 和所述优先级模块 210 上。所述恢复模块 212 可基于恢复次序将数据块和不同的数据块恢复给替换存储设备。所述替换存储设备可以是所述多个存储设备 102 内的一个存储设备。由于上述映射算法会获得所述丢失或故障存储设备 102,采用相同的映射算法,被恢复的数据块可以被映射到所述存储系统或存储池上。

[0027] 图 2B 中描述了系统 100 的一个可选实施例。客户端系统 202 可以与存储区域网

络 200 进行通信。所述存储区域网络 200 可以与存储虚拟化设备 204 进行通信。多个存储设备 102、虚拟卷 104、映射模块 206、恢复列表 208、优先级模块 210、以及恢复模块 212 可以驻留在所述存储虚拟化设备 204 中。

[0028] 所述存储区域网络 200 可以进一步与虚拟卷 104 进行通信。所述虚拟卷 104 可与映射算法进行通信。所述映射算法可与所述多个存储设备 102 进行通信。所述映射算法可提供一个虚拟卷寻址方案,其可以允许所述数据块从所述虚拟卷 104 被映射到所述多个存储设备 102,基于所述虚拟卷寻址方案。

[0029] 所述优先级模块 210 和 / 或所述恢复模块 212 可被实施为一组在计算机可读介质上可执行的计算机可读代码。所述优先级模块 210 和 / 或所述恢复模块 212 也可以被实施为固件。所述优先级模块 210 和 / 或所述恢复模块 212 也可以被实施为硬件。

[0030] 总体上参照图 3,所述恢复模块 212 可在任意其它被标记数据块之前重建 DC(0,4,1) 或 DC(0,4,4),因为 DC(0,4,1) 的标记或 DC(0,4,4) 的标记通过所述优先级模块 210 被放在所述恢复列表 208 的较高位置。一旦上述数据块中的一个被重建,所述恢复模块 212 会继续重建任意其它被标记在所述恢复列表 208 上的数据块。一旦一个数据块被重建,所述数据块的标记将被移出所述恢复列表 208。

[0031] 总体上参照图 4,描述了系统 100 的一个具体实施例。所述系统 100 可被实施为一个分布式镜像架构 400。被镜像的虚拟卷 404 可包括多个数据块(如,DC(0,0,0),DC(0,0,1),等)。所述被镜像虚拟卷 404 可以是一个虚拟卷 104。所述多个数据块可通过上述映射算法被映射到多个存储设备 102(图 4 中的 Drv0,Drv1,Drv2,Drv3,Drv4,Drv5,Drv6 和 Drv7)。所述映射算法可被并入映射模块 206 中。

[0032] 图 5 描述了存储设备 Drv2 和 Drv3 发生故障的场景。存储在存储设备 Drv2 和 Drv3 上的数据块会由于所述存储设备 102 发生故障而丢失。如果另一个存储设备 102 发生故障,与所述被镜像虚拟卷 404 中行 5 相关的数据切片处于被丢失的危险中。因此,所述优先级模块 210 会在所述恢复列表 208 的顶部修改或创建一个与行 5(DC(0,5,0) 和 DC(0,5,1)) 相关的数据块的标记。所述恢复模块 212 可以基于所述恢复列表 208 将所述数据块重建给所述多个存储设备 102。一旦完成,所述恢复模块 212 可继续恢复具有被包括在所述恢复列表 208 内标记的任意其它数据块。

[0033] 总体上参照图 6,描述了一种恢复存储在多个存储设备上数据的方法。所述方法 600 可包括分析在至少一个存储设备上发生的存储设备故障,所述至少一个存储设备被包括在多个存储设备(如,Drv0,Drv1,Drv2,Drv3,Drv4,Drv5,Drv6 和 Drv7)中,被配置用来为至少第一数据部分(即数据块)和第二数据部分(即数据块)提供数据存储 602。所述方法 600 可包括当重要切片发生故障时确定至少所述第一数据部分和所述第二数据部分的恢复次序 604。所述方法 600 可包括基于所述恢复次序来恢复至少所述第一数据部分和所述第二数据部分 606。

[0034] 总体上参照图 7,描述了一种驱动器(存储设备)故障特性的驱动器故障特性方法 700。所述方法 700 可使用标志来控制恢复方法(方法 800 和方法 900)以及用来控制所述恢复列表 208。可替换地,所述方法 700 可使用工作队列或信号来替代标志。所述驱动器故障特性方法 700 在一个驱动器(存储设备 102)发生故障时候开始 702。在另一个驱动器发生故障时确定一个驱动器故障是否潜在地将数据置于危险中。如果没有潜在的危险,进入

背景恢复方法 900。如果存在潜在的危险,在所有紧要恢复排成队列后,进入紧要恢复方法 800,接着是背景恢复方法 900。

[0035] 设置了驱动器故障标志 704 和紧要恢复标志 706。所述驱动器故障标志中止任意紧要恢复方法 800,从先前驱动器故障(先前存储设备 102 故障)812。所述紧要恢复标志中止任意背景恢复方法 900,其由先前驱动故障所引起 914。一旦表明了当前驱动故障的特征,紧要恢复方法 800 或背景恢复方法 900 被启动。用于从最新故障驱动器处恢复的数据块列表(恢复列表 208 或恢复队列)被创建 708。所述恢复列表可以从存储系统元数据处复制。所述恢复列表也可以通过运行基于被存储在所述故障驱动器上所有卷的映射算法而被创建。

[0036] 接着,通过检查上述恢复列表是否为空,来进行检查以确定是否存在进行中的先前恢复 710。如果所述恢复列表不为空,一个先前恢复已在进行。此时,所述驱动器故障表明特征为潜在紧要,来自最新故障驱动器的块列表被补充到现有恢复列表 712,以及紧要恢复方法 800 被启动 714。如果所述恢复列表为空,当所述驱动器发生故障时没有已进行的先前恢复。此时,所述故障存储盘的数据块列表变为所述恢复列表 716。但是,所述驱动器故障仍没有完全表明特征。

[0037] 检查在所述存储阵列(多个存储设备 102)上的所有容器(卷或对象)的最小冗余等级 718。如果最小冗余等级是 1,单个驱动故障必须被当作一个紧要故障;因此,所述紧要恢复方法 800 进入 714。当没有已进行的先前恢复以及所述最小冗余等级大于 1 时,仍旧保护所有数据,这样所述驱动故障能够使用背景恢复进行处理 720。

[0038] 在驱动器故障已经被所述驱动故障特性方法 700 归类为潜在紧要后,进入所述紧要恢复方法 800。搜索所述恢复列表以寻找故障盘组包含多个来自同一个卷和行的数据块的情况。如果在一个行的数据块数小于所述卷的最大冗余,将这些数据块的恢复留给所述背景恢复方法 900。如果在一个行的数据块数等于所述卷的最大冗余,立即对一个块启动紧要恢复。如果在一个行的数据块数大于所述卷的最大冗余,已产生了数据丢失且所述卷被离线以进一步恢复。

[0039] 总体上参照图 8,描述了背景恢复方法 800。所述驱动器故障标志被清空 802。以卷和行归类的所述恢复列表的备份被创建 804。所述被归类的恢复列表随后按行访问直到其末端 806。对于每一行,在所述行中故障驱动器上的数据块数目被计算 808。如果所述行的数据块总数超过所述卷的最大冗余等级 810,一个数据丢失事件被报告 826。在所述报告期间一个故障卷被离线,这样它的块被移出所述恢复列表 828。如果所述行的数据块总数小于所述卷的最大冗余等级 812,跳过对所述行所有数据块的恢复,通过在所述恢复列表中移过它们 822。否则,所述行的数据块总数等于所述卷的最大冗余等级,导致所述行中的数据处于危险中。

[0040] 对所述行中所有数据块进行检查以了解是否有当前等待恢复 814。如果上述数据块中的任意一个在阵列中等待恢复,在所述行中的所有数据块被跳过,通过在所述恢复列表中移过它们 822。否则,所述行的至少一个数据块必须被作为紧要恢复来进行恢复。所述行中的第一数据块在阵列中等待紧要恢复 818,以及所述数据块被标作等待恢复 816。最后,所有所述行中留下的数据块被跳过,通过在所述恢复列表中移过它们 822。

[0041] 下一步,检查以确定是否更多紧要恢复应被排入队列。检查所述驱动器故障标志

以确定是否以产生新的驱动器故障 820。注意一个附加驱动器故障并不意味着数据丢失,因为它会仅仅包括来自具有至少单独驱动冗余的行的数据块。如果设置了驱动器故障标志,当前紧要恢复方法 800 被中止,这样下一个恢复方法 800 能够带有来自最新故障驱动器所增加数据块进行启动 830。对所述恢复列表末端情况进行检查 806。如果已经达到所述列表末端,例如,具有紧要恢复的所有行已经定位,所述紧要恢复方法 800 结束,以及所述背景恢复方法 900 启动 824。

[0042] 在所有紧要恢复已被排入队列后,进入所述背景恢复方法 900。在背景恢复中,系统性能优先于恢复,因为数据不再处于危险中。在背景恢复中,按照驱动器发生故障的顺序来处理所述恢复列表。当最大数目的恢复被排入队列时,所述方法 900 暂停。在每一个恢复完成后,所述方法 900 被重启,这样附加的恢复可被排入队列直到所述恢复列表完全被处理。被排列恢复的限制能够在任意时刻进行改变以控制被分配给背景恢复的系统性能的总量。

[0043] 最初,背景恢复方法 900 开始于所述驱动器故障特性方法 720 之后,或在紧要恢复方法 824 之后。在上述情况之一,所述紧要恢复标志被清空 906 和所述驱动故障标志被清空 908。当先前排列的恢复完成时,所述背景恢复方法 900 被重启 902。此时,所述恢复数据块被移出所述恢复列表 904。

[0044] 如果在所述恢复列表中的当前位置未定义,将所述当前位置设置在所述列表头部 910。所述背景恢复方法 900 随后在所述排列恢复的恢复列表中循环,直到满足下列条件之一:(1) 如果所述恢复列表为空 912,所述方法 900 完成并会退出 926;或(2) 如果设置了所述紧要恢复标志 914,一个新驱动器已发生故障,则所述当前背景恢复方法 900 会退出 926。如果所述恢复队列超出所述性能阈值 916,排列添加的恢复将损害系统性能,因此,当前方法 900 被暂停或退出 926。如果一个恢复能够被排入队列,进行最终检查以确定当前数据块是否已经在等待恢复 918。如果所述当前数据块已经在等待恢复,将它跳过。如果所述数据块不是处于被恢复中,将其标明为等待恢复 920 以及排入恢复队列 922。在所述恢复列表中的当前位置随后被移过所述当前数据块,以及所述处理重复 924。

[0045] 应所述注意的是,根据本发明的所述实施例可以使用传统的通用数字计算机容易地实现,其中所述计算机根据本说明书所教导的内容被编程,说明书中的内容对计算机领域的技术人员来说是显而易见的。基于当前公开的教导,适合的软件代码可已被熟练的编程员准备好,所述实现对软件领域的技术人员来说也是显而易见的。

[0046] 应所述理解本发明可以以软件包的形式来方便地实现。这样一个软件包可以是利用计算机可读媒介的计算机程序产品,所述计算机可读媒介包括储存的用于编程计算机以实现本发明已公开的功能和过程的计算机代码。所述计算机可读媒介/计算机可读存储媒介可以包括但不限于:任何类型的传统软盘、光盘、CD-ROM、磁盘、硬盘驱动器、磁光盘、ROM、RAM、EPROM、EEPROM、磁或光卡、或者其它任何合适用来存储电子指令的媒介。

[0047] 应当相信,通过前述将理解本发明和许多其伴随的优点,显而易见,其组成的形式、架构和布置可以进行不同的变化,而不脱离本发明的主题或不牺牲所有其物质优点。上述的形式只是说明性的,下列权利要求的目的是包含和包括此类变化。

[0048] 应当相信,通过前述将理解本发明和许多其伴随的优点。同样应当相信,显而易见,在这里,其组成的形式、架构和布置可以进行不同的变化,而不脱离本发明的范围和精

神或不牺牲所有其物质优点。这里以前所述的形式只是示例说明,由此,下列权利要求的目的是包含和包括此类变化。

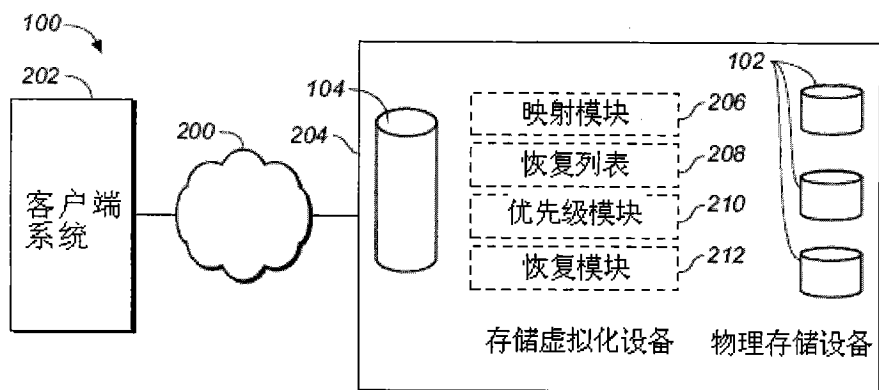


图 2B

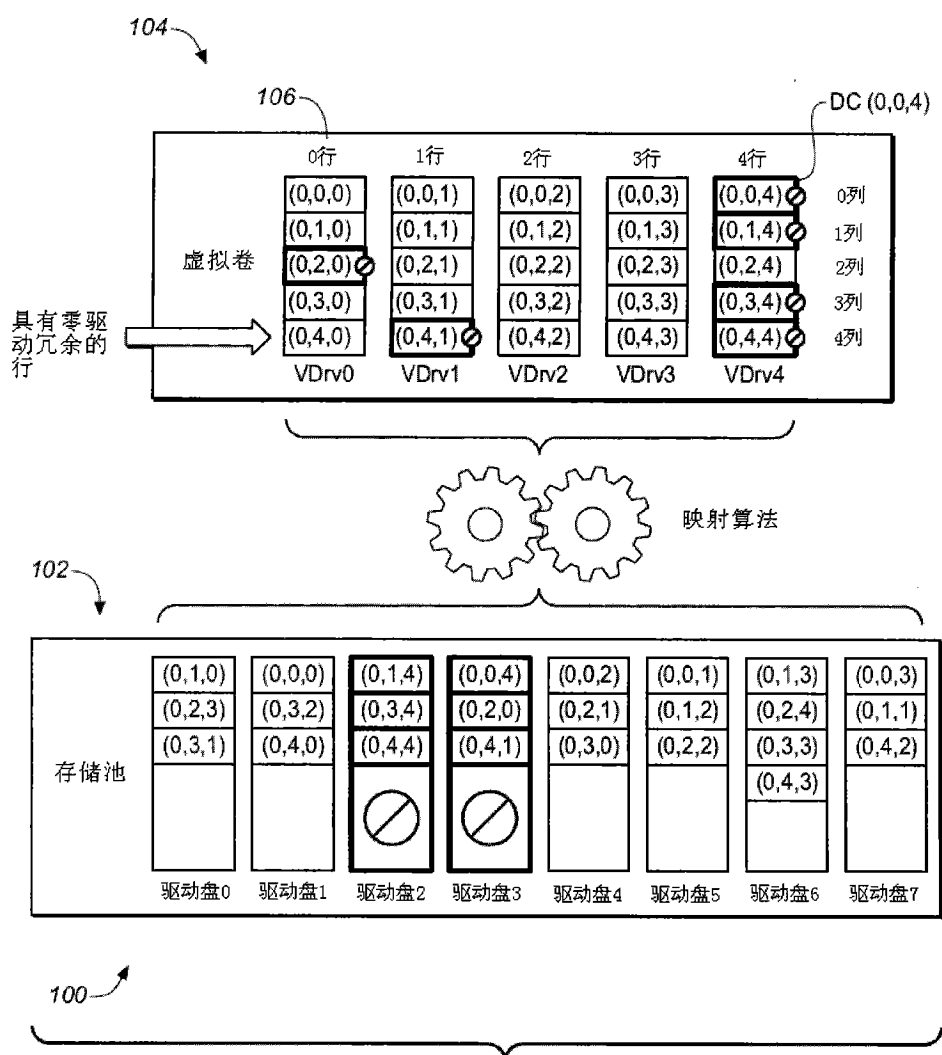


图 3

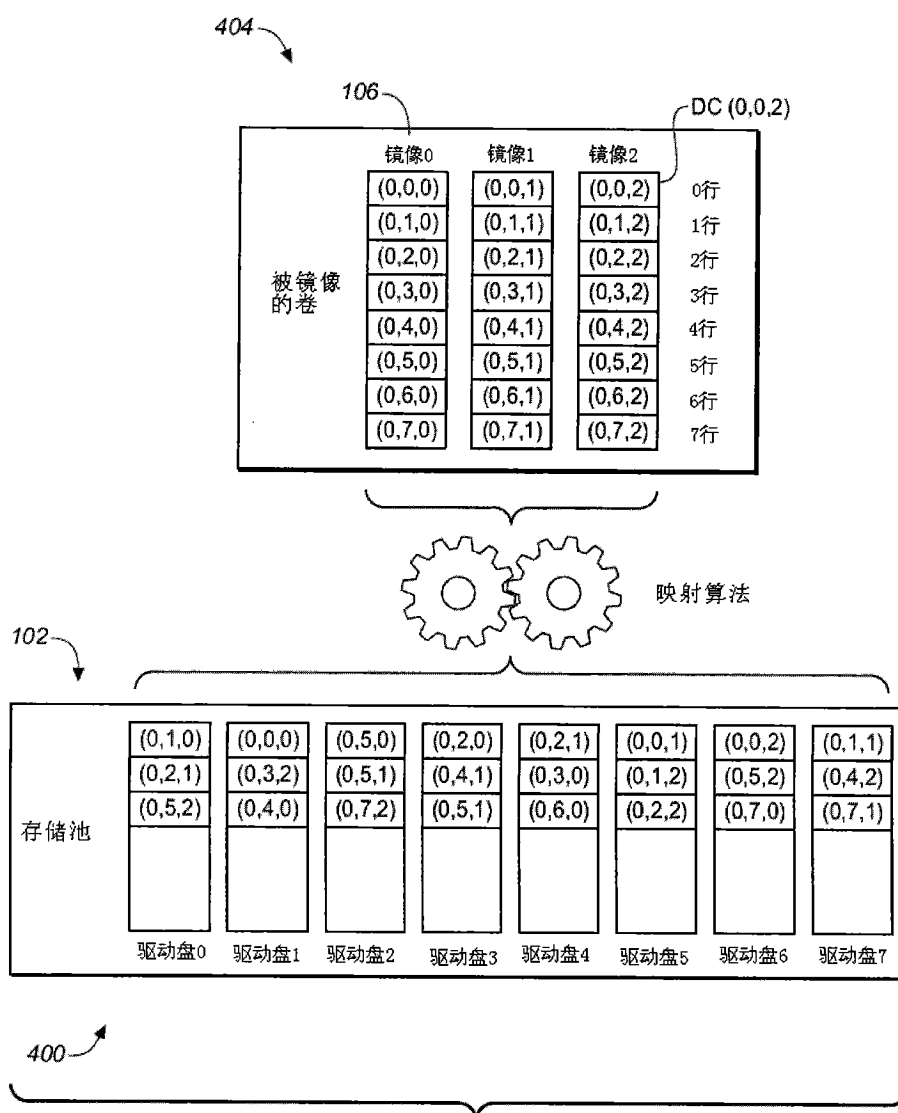


图 4

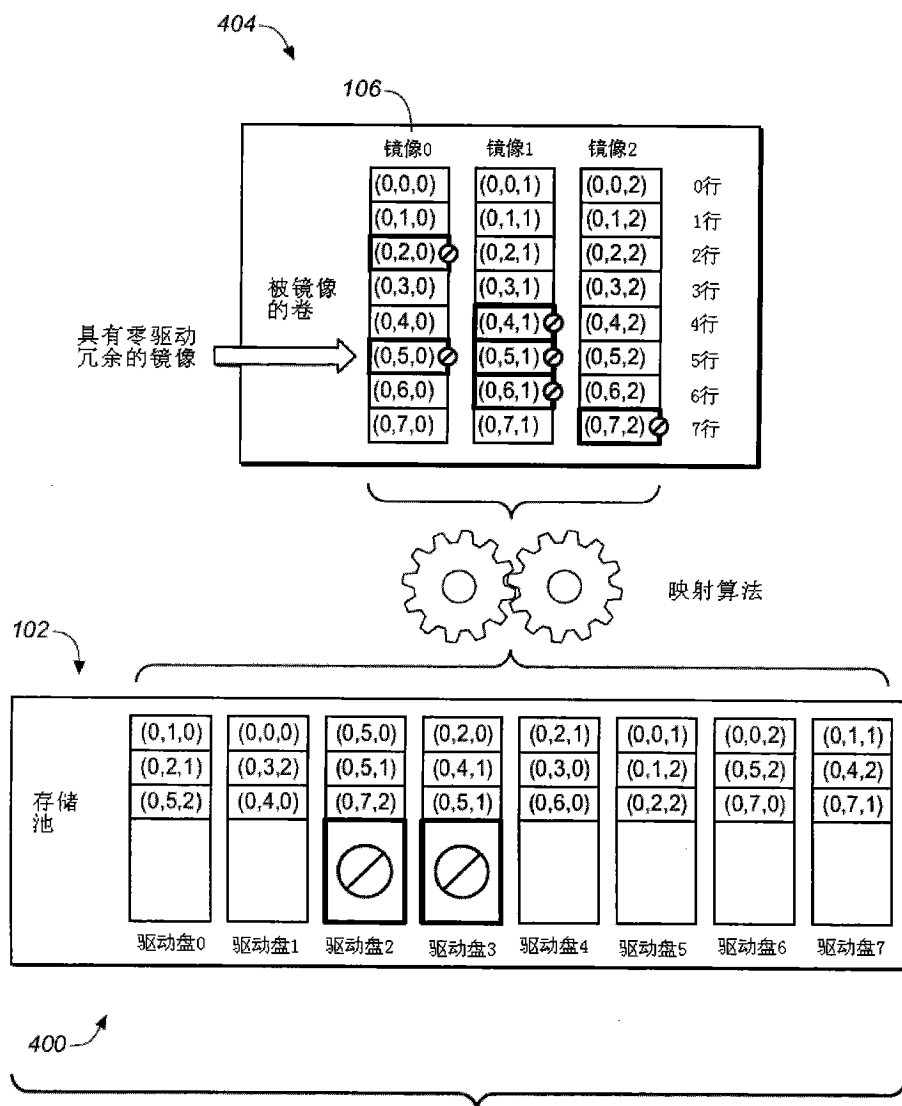


图 5

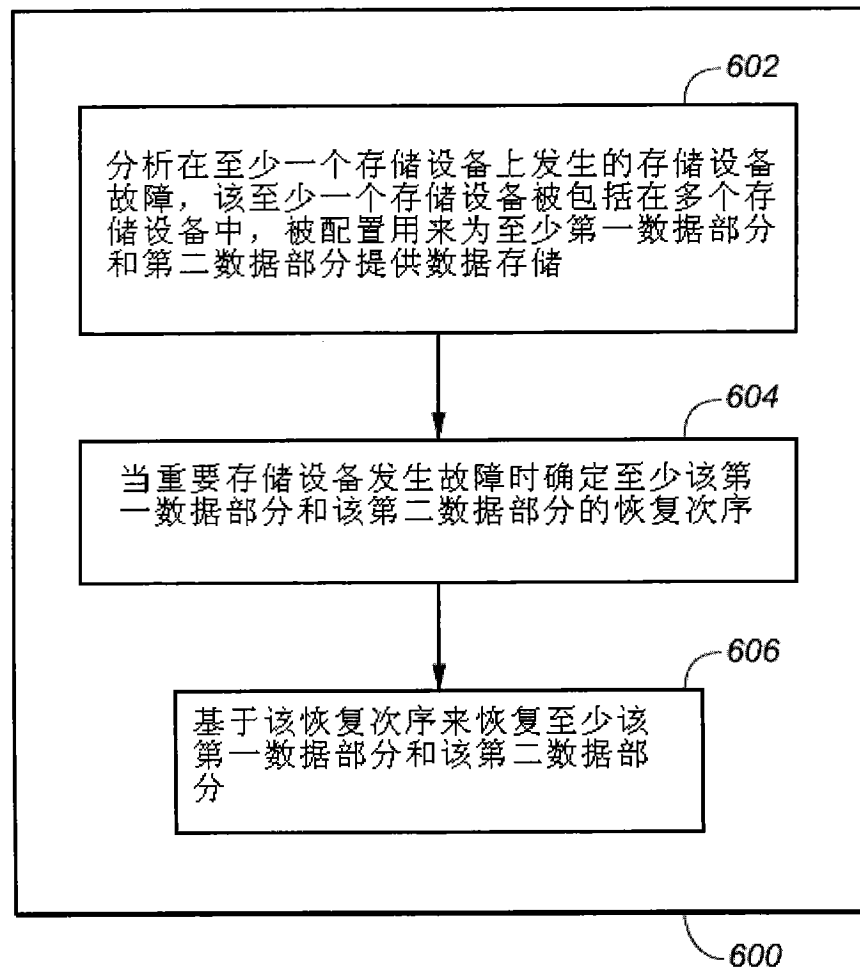


图 6

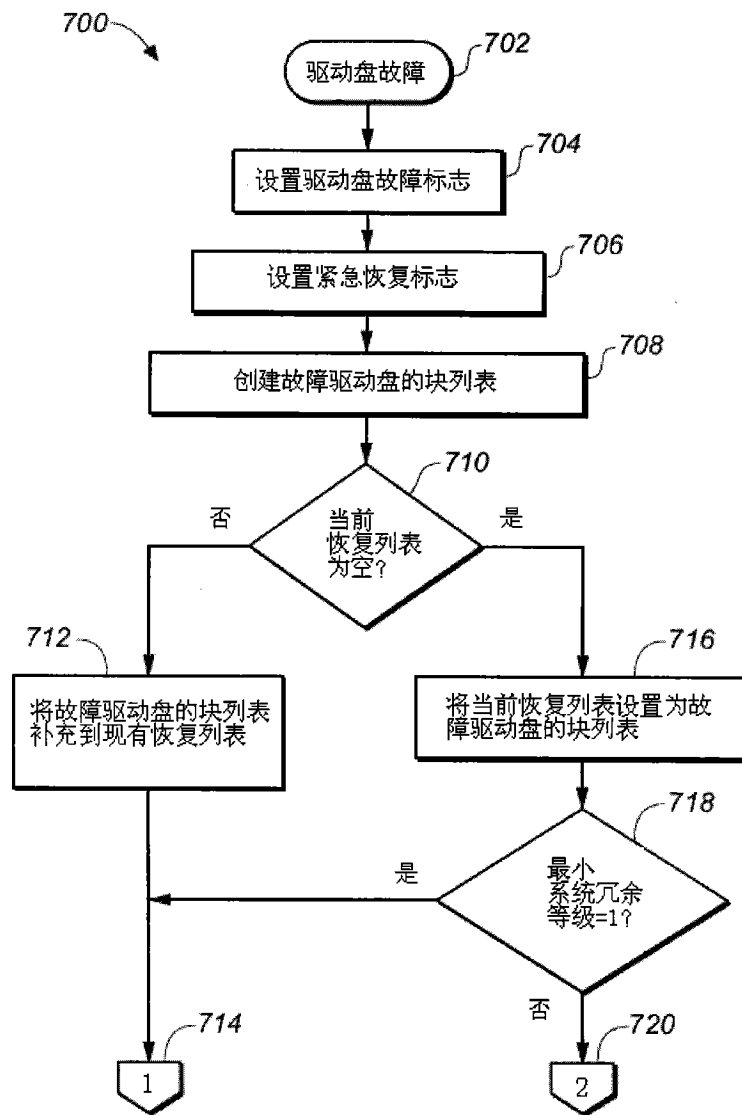


图 7

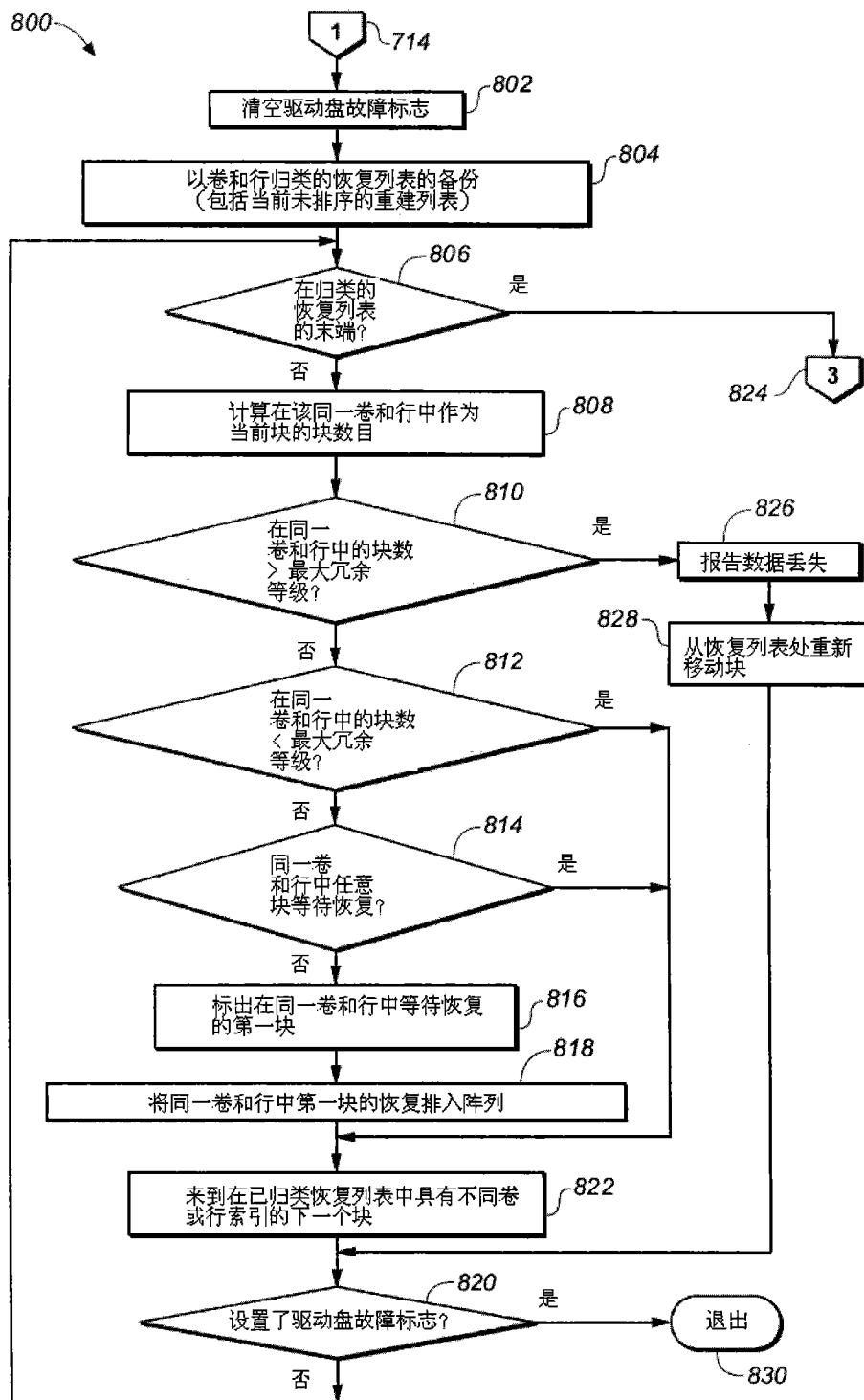


图 8

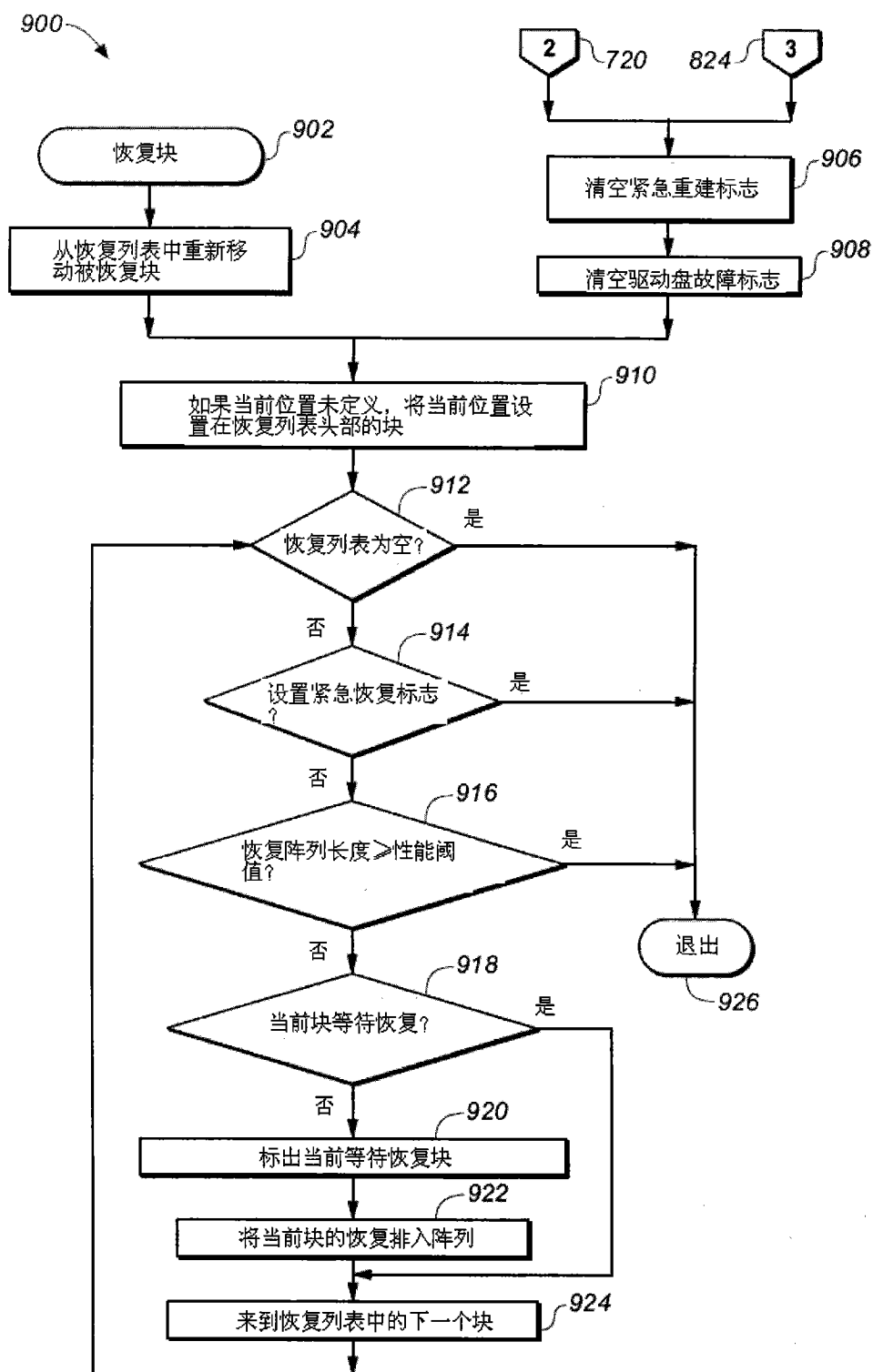


图 9