



(12) 发明专利

(10) 授权公告号 CN 101075177 B

(45) 授权公告日 2010. 10. 13

(21) 申请号 200710005592. 6

(56) 对比文件

(22) 申请日 2007. 02. 13

CN 1661574 A, 2005. 08. 31, 全文.

CN 1662903 A, 2005. 08. 31, 全文.

(30) 优先权数据

2006-136869 2006. 05. 16 JP

审查员 刘娜

(73) 专利权人 株式会社日立制作所

地址 日本东京都

(72) 发明人 川村俊二 山本康友 江口贤哲

(74) 专利代理机构 北京银龙知识产权代理有限公司 11243

代理人 许静

(51) Int. Cl.

G06F 3/06 (2006. 01)

G06F 12/08 (2006. 01)

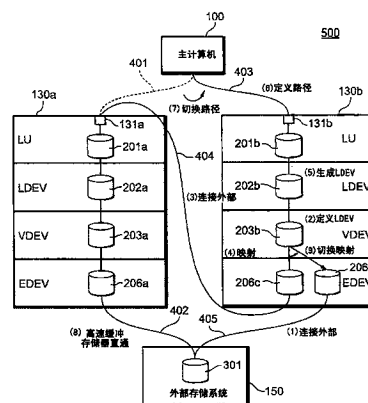
权利要求书 2 页 说明书 42 页 附图 47 页

(54) 发明名称

计算机系统

(57) 摘要

本发明涉及具有使用了存储虚拟化技术的存储系统的计算机系统。在使用存储虚拟化技术的计算机系统中,可使伴随计算机系统结构改变的以逻辑单元为单位的数据移动,并且可抑制存储系统的性能降低。移动目的地存储系统(130a),生成用于虚拟化移动源逻辑单元(201a)的扩展设备(206c)。主计算机(100),通过移动目的地逻辑单元(201b)→移动目的地存储系统(130b)→移动源存储系统(130a)→外部卷301的访问路径访问外部卷(301)。如果在移动源存储系统(130a)的盘高速缓冲存储器中积蓄的脏数据已经全部降级到外部卷(301),则把用于虚拟化外部卷(301)的扩展设备(206b)向移动目的地逻辑单元(201b)映射。



1. 一种计算机系统,其特征在于,包括:

主计算机;

具有外部卷的外部存储系统;

第一存储系统,具有虚拟化所述外部卷形成的第一扩展设备以及虚拟化所述第一扩展设备形成的、由所述主计算机可识别的第一逻辑单元;和

第二存储系统,具有虚拟化所述外部卷形成的第二扩展设备以及虚拟化所述第二扩展设备形成的、由所述主计算机可识别的第二逻辑单元。

2. 根据权利要求1所述的计算机系统,其中,

所述外部存储系统,对于构成所述外部卷的各存储区域,控制所述第一或者第二存储系统中仅任何一方的排他访问。

3. 根据权利要求2所述的计算机系统,其中,

所述主计算机,对于构成所述外部卷的各存储区域,为了连接具有访问权的存储系统进行路径切换。

4. 根据权利要求1所述的计算机系统,其中,

所述外部卷包含从所述第一以及第二存储系统双方访问的共用区域。

5. 一种计算机系统,其特征在于,包括:

主计算机;

具有外部卷的外部存储系统;

第一存储系统,具有虚拟化所述外部卷形成的第一扩展设备、以及从所述主计算机可识别那样经过中间存储层虚拟化所述第一扩展设备而形成的第一逻辑单元;和

可连接所述主计算机以及所述外部存储系统的第二存储系统,

所述第二存储系统,当接收到使所述第一逻辑单元从所述第一存储系统移动到所述第二存储系统的指示时,在所述第二存储系统内生成虚拟化所述外部卷的第二扩展设备,在所述第二存储系统内生成虚拟化所述第一逻辑单元的第三扩展设备,定义连接所述第一逻辑单元和所述第三扩展设备的路径,生成从所述主计算机可识别那样经过所述第二存储系统内的中间存储层虚拟权的存储系统进行路径切换。

18. 根据权利要求5所述的计算机系统,其中,

将所述第二存储系统对写入所述第三扩展设备的数据进行加密的功能设定为关,将所述第一存储系统对写入所述第一扩展设备的数据进行加密的功能设定为开。

19. 一种计算机系统,其特征在于,包括:

主计算机;

具有外部卷的外部存储系统;

多个第一存储系统,具有虚拟化所述外部卷而形成的第一扩展设备、以及从所述主计算机可识别那样经过中间存储层虚拟化所述第一扩展设备而形成的第一逻辑单元;

可连接所述主计算机以及所述外部存储系统的第二存储系统;和

与所述第一以及所述第二存储系统连接的管理服务器,

所述管理服务器,将所述多个第一存储系统中产生故障或者性能降低的存储系统设定为移动源存储系统,将所述移动源存储系统具有的所述第一逻辑单元设定为移动源逻辑单元,将所述第二存储系统设定为移动目的地存储系统,

所述移动目的地存储系统,在所述移动目的地存储系统内生成虚拟化所述外部卷的第二扩展设备,在所述移动目的地存储系统内生成虚拟化所述移动源逻辑单元的第三扩展设备,定义连接所述移动源逻辑单元和所述第三扩展设备的路径,生成从所述主计算机可识别那样经过所述移动目的地存储系统内的中间存储层虚拟化所述第三扩展设备而形成的移动目的地逻辑单元,将连接所述移动目的地逻辑单元和所述主计算机的路径设定为连接所述移动源逻辑单元和所述主计算机的路径的替换路径,

所述主计算机,经由所述替换路径,从所述移动目的地逻辑单元经过所述移动目的地存储系统内的中间存储层、连接所述移动目的地存储系统和所述移动源存储系统的路径、所述移动源逻辑单元、以及所述移动源存储系统内的中间存储层,来访问所述外部卷,

当从所述主计算机将已写入所述移动目的地逻辑单元的数据全部写入到所述外部卷时,所述第二存储系统,解除所述第三扩展设备和所述移动目的地逻辑单元的对应关系,将所述第二扩展设备与所述移动目的地逻辑单元关联对应。

20. 一种计算机系统,其特征在于,包括:

主计算机;

具有外部卷的外部存储系统;

第一存储系统,具有虚拟化所述外部卷而形成的第一扩展设备、以及从所述主计算机可识别那样通过中间存储层虚拟化所述第一扩展设备而形成的第一逻辑单元;和

第二存储系统,具有虚拟化所述外部卷而形成的第二扩展设备、以及从所述主计算机可识别那样经过中间存储层虚拟化所述第二扩展设备而形成的第二逻辑单元,

所述第一扩展设备是虚拟化所述第二逻辑单元而得到的逻辑单元,

将连接所述第二逻辑单元和所述主计算机的路径,设定为连接所述第一逻辑单元和所述主计算机的路径的替换路径,

所述主计算机,从所述第一逻辑单元经过所述第一存储系统内的中间存储层、连接所述第一和所述第二存储系统的路径、所述第二逻辑单元、以及所述第二存储系统内的中间存储层,来访问所述外部卷,当从所述主计算机将已写入所述第一逻辑单元的数据全部写入到所述外部卷时,所述主计算机经由所述替换路径访问所述第二逻辑单元。

计算机系统

技术领域

[0001] 本发明涉及具有使用了存储虚拟化技术的存储系统的计算机系统。

背景技术

[0002] 在因特网业务的扩展、手续的电子化等,信息系统迅速发展,计算机收发的数据量飞跃增加。在这样的数据量的飞跃增加以外,通过向盘装置的数据备份(Disk-to-Disk Backup)或者审核对应等引起的企业业务活动记录(交易信息或者邮件等)的长期保管等,在系统中存储的数据量也持续飞跃增加。与此相伴,在企业信息中,谋求各部门/各系统的存储的增强,另一方面谋求复杂的IT基础设施的管理的简单化和效率化。特别是对于简化存储的管理、根据数据的价值利用最佳的存储、谋求总成本最优的技术的期待日益增高。

[0003] 作为降低具有大规模存储的系统的管理成本的方法,可以举出在特开平2005-011277号公报中公开的存储虚拟化技术。在特开平2005-011277号公报中,公开了如下存储虚拟化技术(以下也称外部存储连接技术):把在主计算机上连接的存储系统(以下称第一存储系统)连接到一个以上的外部存储系统(以下称第二存储系统)上、将第二存储系统具有的设备(以下称外部设备)虚拟化如同第一存储系统内的逻辑设备那样并提供给主计算机。第一存储系统,当接收来自主计算机的向逻辑设备的输入输出请求时,判定访问目的地设备是否与第二存储系统内的外部设备对应,或者是否与第一存储系统搭载的盘装置等物理设备(内部设备)对应,根据判定结果,向适当的访问目的地分配输入输出请求。

[0004] 当使用在特开平2005-011277号公报中所示的使用外部存储连接技术的存储系统时,可以构建综合了性能、可靠性、价格等属性不同的多个存储系统的系统。例如,通过将使用外部存储连接技术的高成本、高功能、高可靠性的第一存储系统与低成本、低功能、低可靠性的第二存储系统相连,把存储区域在逻辑上分层,由此,能够实现对对应数据的更新日期时间或者数据的价值的最佳数据配置。通过使用具有分层的存储区域的存储系统,能够以审核对应等为目的,将在每日的业务中产生的交易信息或者邮件等大量的信息,根据各信息具有的价值以最合适的成本长期保存。

[0005] 但是,为了经历超过存储系统寿命的数十年的长时间保存大容量的数据,在数据的保存期间中需要进行更换构成计算机系统的设备的作业。在特开平2005-011277号公报中,公开了如下内容:为了把使用外部存储连接技术的第一存储系统(以下称移动源存储系统)更换为其他的第一存储系统(以下称移动目的地存储系统),把移动源存储系统具有的逻辑单元(以下称移动源逻辑单元)内的数据复制到移动目的地存储系统具有的逻辑单元(以下称移动目的地逻辑单元)来进行移动。

[0006] 【专利文献1】特开平2005-011277号公报

发明内容

[0007] 但是,担心由于从移动源逻辑单元向移动目的地逻辑单元的复制处理对移动源存

储系统进行的设备的输入输出处理造成吞吐量 (through-put) 降低等负面影响。并且, 因为需要把移动源逻辑单元内的数据全部复制到移动目的地逻辑单元, 所以产生从移动源存储系统向移动目的地存储系统的设备移交需要较长时间的问题。

[0008] 另外, 以存储层的复杂化为背景, 不仅可以设想移动源逻辑单元和外部设备一对一对应的情况, 而且也可以设想两者不是一对一对应的情况。特别即使在移动源逻辑单元和外部设备不是一对一对应的情况下, 也希望能够以作为主计算机识别的存储区域的单位的逻辑单元单位来进行数据移动。

[0009] 本发明的目的是在使用了存储虚拟化技术的计算机系统中, 能够抑制存储系统的性能降低, 并且进行伴随计算机系统的结构变更的、以逻辑单元为单位的数据移动。

[0010] 本发明的另一目的是在使用了存储虚拟化技术的计算机系统中, 由多个存储系统共享外部卷。

[0011] 为解决上述课题, 本发明的计算机系统具有: 主计算机; 具有外部卷的外部存储系统; 第一存储系统, 具有虚拟化外部卷形成的第一扩展设备和虚拟化第一扩展设备形成的、从所述主计算机可识别的第一逻辑单元; 以及第二存储系统, 具有虚拟化外部卷形成的第二扩展设备和虚拟化第二扩展设备形成的、从主计算机可识别的第二逻辑单元。

[0012] 这里, 例如外部存储系统, 对于构成外部卷的各存储区域, 控制第一或者第二存储系统中仅某一方的排他访问。

[0013] 主计算机, 例如为了连接对构成外部卷的各存储区域有访问权的存储系统而进行路径切换。

[0014] 外部卷也可以包含从第一以及第二存储系统的双方访问的共用区域。

[0015] 关于本发明其他方面的计算机系统, 具有: 主计算机; 具有外部卷的外部存储系统; 第一存储系统, 具有虚拟化外部卷形成的第一扩展设备、和使从主计算机可识别那样通过中间存储层虚拟化第一扩展设备形成的第一逻辑单元; 和可连接主计算机以及外部存储系统的第二存储系统。第二存储系统, 当接收到使第一逻辑单元从第一存储系统移动到第二存储系统的指示时, 在第二存储系统内生成虚拟化外部卷的第二扩展设备, 在第二存储系统内生成虚拟化第一逻辑单元的第三扩展设备, 定义连接第一逻辑单元和第三扩展设备的路径, 生成从主计算机可识别那样经过第二存储系统内的中间存储层虚拟化第三扩展设备而形成的第二逻辑单元, 把连接第二逻辑单元和主计算机的路径设定为连接第一逻辑单元和主计算机的路径的替换路径。主计算机, 经由替换路径, 从第二逻辑单元经过第二存储系统内的中间存储层、连接第一以及第二存储系统的路径、第一逻辑单元、以及第一存储系统内的中间存储层访问外部卷。当从主计算机写入第二逻辑单元的数据被全部写入到外部卷时, 第二存储系统解除第三扩展设备与第二逻辑单元的对应关系, 使第二扩展设备与第二逻辑单元关联对应。

[0016] 第一存储系统还具有对已写入第一逻辑卷的数据进行临时存储的盘高速缓冲存储器, 第一存储系统, 当接收到使第一逻辑单元从第一存储系统移动到第二存储系统的指示时, 把第一存储系统的动作模式设定为高速缓冲存储器直通模式。

[0017] 第一存储系统内的中间存储层, 包含用于虚拟化第一扩展设备的虚拟设备, 第一逻辑单元是虚拟化虚拟设备的一部分而得到的逻辑单元。

[0018] 第一存储系统, 在构成外部卷的存储区域中禁止对与虚拟设备的一部分对应的存

储区域的访问。

[0019] 第一存储系统内的中间存储层,包含用于虚拟化第一扩展设备的逻辑设备,第一逻辑单元是将多个逻辑设备虚拟化汇总为一个而得到的逻辑单元。

[0020] 第一存储系统内的中间存储层,通过在第一扩展设备的存储区域内使用 RAID 结构,包含虚拟化扩展设备的虚拟设备,第一逻辑单元是虚拟化虚拟设备的一部分而形成的逻辑单元。

[0021] 外部卷包含存储主计算机读写的用户数据的第一存储区域、和存储第一逻辑单元的管理信息的第二存储区域,第一存储系统内的中间存储层,包含用于虚拟化第一扩展设备的逻辑设备,逻辑设备包含虚拟化外部卷的第一存储区域的第一存储区域、和虚拟化外部卷的第二存储区域的第二存储区域,第一逻辑单元是虚拟化逻辑设备的第一存储区域而形成的逻辑单元。

[0022] 第二存储系统,将在外部卷的第二存储区域中存储的管理信息设定为第二逻辑单元的属性信息。

[0023] 本发明的计算机系统,也可以具有使多个第一逻辑单元从第一存储系统移动到第二存储系统的功能。

[0024] 本发明的计算机系统,也可以具有使第一逻辑单元的一部分从第一存储系统移动到第二存储系统的功能。

[0025] 在本发明的计算机系统中,第一逻辑单元存储有多个文件系统,逻辑单元的一部分是多个文件系统中任何一个以上的文件系统。

[0026] 在本发明的计算机系统中,外部存储系统,对于构成外部卷的存储区域中与第一逻辑单元的一部分对应的存储区域,排他地接受来自第二存储系统的访问,对于构成外部卷的存储区域中与除第一逻辑单元的一部分以外的部分对应的存储区域,排他地接受来自第一存储系统的访问。

[0027] 在本发明的计算机系统中,主计算机为了连接对于构成外部卷的各存储区域有访问权的存储系统而进行路径切换。

[0028] 在本发明的计算机系统中,将第二存储系统对写入第三扩展设备的数据进行加密的功能设定为关,将第一存储系统对写入第一扩展设备的数据进行加密的功能设定为开。

[0029] 本发明的计算机系统,具有:主计算机;具有外部卷的外部存储系统;多个第一存储系统,具有虚拟化外部卷形成的第一扩展设备、和使从主计算机可识别那样经过中间存储层虚拟化第一扩展设备而形成的第一逻辑单元;可连接主计算机以及外部存储系统的第二存储系统;和在第一以及第二存储系统上连接的管理服务器。管理服务器,把多个第一存储系统中产生故障或者性能降低的存储系统设定为移动源存储系统,把移动源存储系统具有的第一逻辑单元设定为移动源逻辑单元,把第二存储系统设定为移动目的地存储系统。移动目的地存储系统,在移动目的地存储系统内生成虚拟化外部卷的第二扩展设备,在移动目的地存储系统内生成虚拟化移动源逻辑单元的第三扩展设备,定义连接移动源逻辑单元和第三扩展设备的路径,生成使从主计算机可识别那样经过移动目的地存储系统内的中间存储层虚拟化第三扩展设备而形成的移动目的地逻辑单元,把连接移动目的地逻辑单元和主计算机的路径设定为连接移动源逻辑单元和主计算机的路径的替换路径。主计算机经由替换路径,从移动目的地逻辑单元经过移动目的地存储系统内的中间存储层、连接移动

目的地存储系统以及移动源存储系统的路径、移动源逻辑单元、以及移动源存储系统内的中间存储层访问外部卷。当从主计算机将已写入移动目的地逻辑单元的数据全部写入到外部卷时,第二存储系统解除第三扩展设备和移动目的地逻辑单元的对应关系,使第二扩展设备与移动目的地逻辑单元关联对应。

[0030] 本发明的计算机系统,具有:主计算机;具有外部卷的外部存储系统;第一存储系统,具有虚拟化外部卷形成的第一扩展设备、和从主计算机可识别那样经过中间存储层虚拟化第一扩展设备而形成的第一逻辑单元;和第二存储系统,具有虚拟化外部卷而形成的第二扩展设备、和从主计算机可识别那样经过中间存储层虚拟化第二扩展设备而形成的第二逻辑单元。第一扩展设备是虚拟化第二逻辑单元而得到的扩展设备。将连接第二逻辑单元和主计算机的路径,设定为连接第一逻辑单元和主计算机的路径的替换路径。主计算机,从第一逻辑单元经过第一存储系统内的中间存储层、连接第一以及第二存储系统的路径、第二逻辑单元、以及第二存储系统内的中间存储层访问外部卷。当从主计算机将已写入第一逻辑单元的数据全部写入到外部卷时,主计算机经由替换路径访问第二逻辑单元。

[0031] 根据本发明,在使用存储虚拟化技术的计算机系统中,可使伴随计算机系统结构变更的以逻辑单元为单位的数据移动,同时可抑制存储系统的性能降低。

[0032] 另外,根据本发明,在使用存储虚拟化技术的计算机系统中,可以由多个存储系统共享外部卷。

附图说明

[0033] 图 1 是关于本实施形态的逻辑单元移动处理的原理图。

[0034] 图 2 是关于实施例的计算机系统的网络结构图。

[0035] 图 3 是在存储系统内构建的存储层的说明图。

[0036] 图 4 是表示 CVS 结构的映射关系的说明图。

[0037] 图 5 是表示 LUSE 结构的映射关系的说明图。

[0038] 图 6 是表示外部卷 RAID 结构的映射关系的说明图。

[0039] 图 7 是附有 VMA 的卷的说明图。

[0040] 图 8 是大型机卷 (mainframe volume) 的说明图。

[0041] 图 9 是表示 VDEV 连接的映射关系的说明图。

[0042] 图 10 是表示 VDEV 离散的映射关系的说明图。

[0043] 图 11 是表示 VDEV 离散的映射关系的说明图。

[0044] 图 12 是表示 VDEV 离散的映射关系的说明图。

[0045] 图 13 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

[0046] 图 14 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

[0047] 图 15 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

[0048] 图 16 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

[0049] 图 17 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

[0050] 图 18 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

[0051] 图 19 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

[0052] 图 20 是表示 DEVGr-EDEV/PDEV 映射关系的说明图。

- [0053] 图 21 是表示 AOU 结构的存储层的说明图。
- [0054] 图 22 是表示具有快照 (snapshot) 功能的存储层的说明图。
- [0055] 图 23 是关于主机组 (host group) 的说明图。
- [0056] 图 24 是关于替换路径的说明图。
- [0057] 图 25 是关于替换路径的说明图。
- [0058] 图 26 是主机组管理信息的说明图。
- [0059] 图 27 是 LU 路径管理信息的说明图。
- [0060] 图 28 是逻辑设备管理信息的说明图。
- [0061] 图 29 是虚拟设备管理信息的说明图。
- [0062] 图 30 是设备组管理信息的说明图。
- [0063] 图 31 是 VDEV-DEVGr 映射信息的说明图。
- [0064] 图 32 是物理设备管理信息的说明图。
- [0065] 图 33 是扩展设备管理信息的说明图。
- [0066] 图 34 是计算机系统的软件结构图。
- [0067] 图 35 是表示使具有 CVS 结构的逻辑单元移动的处理的概要的说明图。
- [0068] 图 36 是表示使具有 LUSE 结构的逻辑单元移动的处理的概要的说明图。
- [0069] 图 37 是表示使具有外部卷 RAID 结构的逻辑单元移动的处理的概要的说明图。
- [0070] 图 38 是表示使具有附有 VMA 的卷结构的逻辑单元移动的处理的概要的说明图。
- [0071] 图 39 是表示使具有附有 VMA 的卷结构的逻辑单元移动的处理的概要的说明图。
- [0072] 图 40 是表示使具有大型机卷结构的逻辑单元移动的处理的概要的说明图。
- [0073] 图 41 是表示使多个逻辑单元移动的处理的概要的说明图。
- [0074] 图 42 是表示逻辑单元移动指示处理例行程序 (routine) 的说明图。
- [0075] 图 43 是表示逻辑单元移动处理子例行程序 (sub-routine) 的说明图。
- [0076] 图 44 是表示移动对象选择处理子例行程序的说明图。
- [0077] 图 45 是表示移动准备处理子例行程序的说明图。
- [0078] 图 46 是表示移动处理子例行程序的说明图。
- [0079] 图 47 是表示存储系统 I/O 处理例行程序的说明图。
- [0080] 图 48 是表示主机 (host) I/O 处理例行程序的说明图。
- [0081] 图 49 是表示存储负荷监视处理例行程序的说明图。
- [0082] 图 50 是表示逻辑单元移动处理的其他方法的说明图。
- [0083] 图 51 是表示逻辑单元移动处理的其他方法的说明图。
- [0084] 图 52 是表示逻辑单元移动处理的其他方法的说明图。
- [0085] 图 53 是表示逻辑单元移动处理的其他方法的说明图。
- [0086] 图 54 是表示逻辑单元移动处理的其他方法的说明图。
- [0087] 图 55 是表示使逻辑单元的一部分移动的处理的概要的说明图。
- [0088] 图 56 是表示使逻辑单元的一部分移动的处理的概要的说明图。
- [0089] 图 57 是高可用性计算机系统的系统结构图。
- [0090] 图 58 是高可用性计算机系统的系统结构图。
- [0091] 图 59 是高可用性计算机系统的系统结构图。

[0092] 符号说明

[0093] 100 主计算机, 120 光纤通道交换机, 130 存储系统, 140 管理终端, 150 外部存储系统, 500 计算机系统, 201 逻辑单元, 202 逻辑设备, 203 虚拟设备, 204 设备组, 205 物理设备, 206 扩展设备

具体实施方式

[0094] 以下参照各图说明本发明的实施形态。首先参照图 1 说明关于本实施形态的逻辑单元移动处理的原理。关于本实施形态的计算机系统 500 具有主计算机 (host computer) 100、多个存储系统 130a、130b、和外部存储系统 150。

[0095] 存储系统 130a 具有一个以上的逻辑单元 (LU) 201a。对该逻辑单元 201a 指定 (assign) LUN (Logical Unit Number (逻辑单元号))。逻辑单元 201a 作为逻辑存储区域被提供给主计算机 100。主计算机 100 能够向逻辑单元 201a 进行数据写入或者从逻辑单元 201a 进行数据读出。在逻辑单元 201a 的下层设置有多个中间存储层 (LDEV/VDEV/EDEV)。该中间存储层是用于将逻辑单元 201a 与外部卷 (volume) 301 之间关联对应起来的逻辑存储资源。来自主计算机 100 的访问, 从逻辑单元 201a 经过中间存储层被传送到外部卷 301。

[0096] 在存储系统 130a 内的中间存储层中可以包含例如逻辑设备 (LDEV) 202a、虚拟设备 (VDEV) 203a、以及扩展设备 (EDEV) 206a。这些设备不是必需的, 也可以省略一部分设备。另外, 在这些设备之外, 还可以包含其他设备 (例如后述的设备组 (DEVGr) 等) (参照图 3)。

[0097] 此外, 如图 3 所示, 在存储系统 130 具有物理设备 (PDEV) 205 的情况下, 物理设备 205 和扩展设备 206 共同属于同一中间存储层。

[0098] 为便于说明, 有时把逻辑单元属于的存储层称为“LDEV 层”。同样, 有时把逻辑设备属于的存储层称为“LDEV 层”, 把虚拟设备属于的存储层称为“VDEV 层”, 把物理设备和/或扩展设备属于的存储层称为“PDEV/EDEV 层”。

[0099] 存储系统 130a 具有用于通过路径 401 连接主计算机 100 的一个以上的端口 131a、分配给端口 131a 的逻辑单元 201a、向逻辑单元 201a 映射 (mapping) 的逻辑设备 202a、向逻辑设备 202a 映射的虚拟设备 203a、向虚拟设备 203a 映射的扩展设备 206a。这里, 所谓“映射”是指将某设备的地址空间与其他设备的地址空间关联对应。在本说明书中, 将某设备向其他设备映射, 和将某设备与其他设备关联对应 (或者分配) 同义。

[0100] 存储系统 130a 通过路径 402 与外部存储系统 150 连接。外部存储系统 150 具有在盘驱动器等有实际存储区域的物理设备上定义的外部卷 301。

[0101] 这里, 扩展设备 206a 是虚拟化作为在存储系统 130a 的外部存在的存储资源的外部卷 301 而形成的扩展设备。亦即, 扩展设备 206a 是虚拟化外部资源而形成的存储层。

[0102] 虚拟设备 203a 是连接上位存储层 (例如 LU/LDEV) 与下位存储层 (例如 DEVGr/EDEV) 的存储层。在下位存储层是物理设备的情况下, 虚拟设备 203a 例如是将多个物理设备各自提供的存储区域进行 RAID 构成形成的存储层。另一方面, 在下位存储层是扩展设备 206a 的情况下, 虚拟设备 203a 例如是集合多个扩展设备 206a 各自的存储区域的全部或一部分而形成的存储区域、或者是选出扩展设备 206a 的存储区域的一部分而形成的存储区域。

[0103] 逻辑设备 202a 例如是集合多个虚拟设备 203a 各自的存储区域的全部或者一部分

而形成的存储区域、或者是选出虚拟设备 203a 的存储区域的一部分而形成的存储区域。在主计算机 100 是开放类型系统的情况下,主计算机 100 把逻辑设备 202a 识别为一个物理设备,通过指定 LUN 或者逻辑块地址,访问所希望的逻辑设备 202a。在主计算机 100 是大型机(mainframe) 类型系统的情况下,主计算机 100 直接识别逻辑设备 202a。

[0104] 逻辑单元 201a 是虚拟化外部卷 301 而得到的单元,而且是主计算机 100 识别的逻辑存储区域。例如,在主计算机 100 是 UNIX(注册商标) 类型系统的情况下,逻辑单元 201a 与设备文件(Device File) 对应关联。另一方面,在主计算机 100 是 Windows(注册商标) 类型系统的情况下,逻辑单元 201a 与驱动器号(drive letter 驱动器名) 对应关联。对逻辑单元 201a 指定固有的 LUN(Logical Unit Number)。

[0105] 为便于说明逻辑单元移动处理,设逻辑单元 201a 和逻辑设备 202a 的对应关系是一对一、逻辑设备 202a 和虚拟设备 203a 的对应关系是一对一、虚拟设备 203a 和扩展设备 206a 的对应关系是一对一、扩展设备 206a 和外部卷 301 的对应关系是一对一。在这样的对应关系下,主计算机 100 识别的逻辑单元 201a 的存储容量等于外部卷 301 的存储容量。

[0106] 另一方面,存储系统 130b 有与存储系统 130a 同样分层的多个存储层。存储系统 130b 具有:用于通过路径 403 与主计算机 100 连接的一个以上的端口 131b、分配给端口 131b 的逻辑单元 201b、向逻辑单元 201b 映射的逻辑设备 202b、向逻辑设备 202b 映射的虚拟设备 203b、向虚拟设备 203b 映射的扩展设备 206c、206b。

[0107] 为便于说明逻辑单元移动处理,设逻辑单元 201b 和逻辑设备 202b 的对应关系是一对一、逻辑设备 202b 和虚拟设备 203b 的对应关系是一对一、虚拟设备 203b 和扩展设备 206c 的对应关系是一对一、虚拟设备 203b 和扩展设备 206b 的对应关系是一对一、扩展设备 206c 和逻辑单元 201a 的对应关系是一对一、扩展设备 206b 和外部卷 301 的对应关系是一对一。另外,假设扩展设备 206c 和扩展设备 206b 具有相同的存储容量,选择扩展设备 206c 和扩展设备 206b 中任何一方向虚拟设备 203b 上进行映射。

[0108] 下面概述把位于存储系统 130a(移动源存储系统) 内的逻辑单元 201a(移动源逻辑单元) 中的数据向存储系统 130b(移动目的地存储系统) 内的逻辑单元 201b(移动目的地逻辑单元) 移动的处理。逻辑单元移动处理包含下述(1)~(9) 的处理过程。

[0109] (1) 存储系统 130b,定义用于外部连接外部卷 301 和存储系统 130b 之间的路径 405,在存储系统 130b 内生成用于虚拟化外部卷 301 的扩展设备 206b。

[0110] (2) 存储系统 130b,在存储系统 130b 内生成具有与外部卷 301 的存储容量相同的存储容量的虚拟设备 203b。

[0111] (3) 存储系统 130b,定义用于外部连接逻辑单元 201a 和存储系统 130b 之间的路径 404,在存储系统 130b 内生成用于虚拟化逻辑单元 201a 的扩展设备 206c。

[0112] (4) 存储系统 130b,将扩展设备 206c 向虚拟设备 203b 映射(mapping)。

[0113] (5) 存储系统 130b,在存储系统 130b 内生成具有与逻辑设备 202a 的逻辑结构相同的逻辑结构的逻辑设备 202b,把虚拟设备 203b 向逻辑设备 202b 映射。

[0114] (6) 存储系统 130b,在存储系统 130b 内生成具有与逻辑单元 201a 的逻辑结构相同的逻辑结构的逻辑单元 201b,定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。把逻辑设备 202b 向逻辑单元 201b 映射。

[0115] (7) 主计算机 100,从路径 401 切换到路径 403,作为用于访问外部卷 301 的路径。

此时,进行删除路径 401 等操作,禁止从主计算机 100 向逻辑单元 201a 的访问。此外,从路径 401 到路径 403 的路径切换,也可以由存储系统 130a、130b 或者管理服务器 110 等进行。另外,也不一定需要删除路径 401,例如可以把路径 401 设定为路径 403 的替换路径(例如用于在系统故障时使用的路径),使路径 401 的状态成为无效不变,保留路径 401。

[0116] 通过上述处理,因为逻辑单元 201b 不仅虚拟化逻辑单元 201a,而且也虚拟化外部卷 301,所以在逻辑单元移动处理中的、从主计算机 100 向逻辑单元 201a 的数据输入输出路径成为:路径 403 → 存储系统 130b 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 404 → 存储系统 130a 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 402 → 外部卷 301。

[0117] 此外,在参与逻辑单元移动处理的存储系统 130a、130b 上搭载加密功能时,当把两个存储系统 130a、130b 的加密功能设定为开时,有时不能解密在外部卷 301 中写入的数据。在这种情况下,把存储系统 130b 的加密功能设定为关,把存储系统 130a 的加密功能设定为开。即,不通过存储系统 130b 加密在扩展设备 206c 中写入的数据,而通过存储系统 130a 加密在扩展设备 206a 中写入的数据。通过这样设定加密功能,能够使对写入外部卷 301 中的数据进行加密的次数仅为一次,所以能够解密在外部卷 301 中写入的数据。

[0118] (8) 存储系统 130a,为把在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据(dirty data)据降级(destage)到外部卷 301,将存储系统 130a 的动作模式设定为高速缓冲存储器直通模式(cache through mode)(写直通模式 write through mode)。所谓脏数据指未从盘高速缓冲存储器降级到外部卷 301 的未反映数据。所谓高速缓冲存储器直通模式,指接收来自主计算机 100 的数据写入请求、在盘高速缓冲存储器中存储数据,进而在设备(内部设备或者外部设备)中也写入该数据后,向主计算机 100 报告数据更新结束的动作。此外,存储系统 130a 可以自发地把动作模式设定为高速缓冲存储器直通模式,或者也可以根据来自外部(例如,存储系统 130b、主计算机 100、或者后述的管理服务器 110 等)的指示,把动作模式设定为高速缓冲存储器直通模式。

[0119] (9) 如果在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷 301,则存储系统 130b 解除扩展设备 206c 和虚拟设备 203b 之间的映射关系,把扩展设备 206b 向虚拟设备 203b 映射。并且,存储系统 130b 删除路径 404。

[0120] 根据这样的方法,可以维持继续从主计算机 100 向逻辑单元 201a 的数据输入输出处理并使逻辑单元 201a 移动。由此,根据来自主计算机 100 的 I/O 请求所引起的多个存储系统 130 之间的负荷变动的发生,把向外部卷的 I/O 处理责任在存储系统 130a、130b 之间切换,由此可以实现计算机系统 500 的负荷分散。而且,通过把存储系统 130a、130b 中移动源存储系统的动作模式设定为高速缓冲存储器直通模式,把移动目的地存储系统的动作模式设定为后写模式(write after mode),能够进行使用盘高速缓冲存储器的非同步写处理,所以能够抑制伴随逻辑单元移动处理的存储系统 130a 的性能降低。

[0121] 此外,把移动目的地存储系统 130b 的动作模式设定为高速缓冲存储器直通模式也能够使逻辑单元 201a 移动,但是只要不将移动源存储系统 130a 内积蓄的全部脏数据降级到外部卷 301 中,则逻辑单元 201a 的移动处理不结束。因此,为缩短逻辑单元 201a 的移动处理时间,优选把移动源存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。

[0122] 另外,在图 1 所示的系统结构中,各个存储系统 130a、130b 具有虚拟化外部卷 301 而形成的扩展设备 206,但是没有内部设备(物理设备)。涉及本发明的逻辑单元移动处

理,不限于应用与图 1 所示那样的系统结构,例如,也可以应用于使通过没有外部卷虚拟化功能的存储系统 130a、130b 内置的盘装置等物理设备所提供的逻辑单元移动的处理。

[0123] 在上述的逻辑单元移动处理的说明中,为便于说明,举例表示了使各中间存储层相互之间的映射关系为一对一的情况,但是各中间存储层相互之间的映射关系不一定限定为一对一的关系。例如,作为各中间存储层相互之间的映射关系,可以考虑 CVS (Custom Volume Size)、LUSE (LU SizeExpansion)、外部卷 RAID、附有 VMA (Volume Management Area) 的卷、大型机卷 (mainframe volume)、VDEV 连接、VDEV 离散、AOU (Allocation onUse)、快照 (snapshot) 等各种变体 (variation)。关于其详情,在后述的实施例中说明 (参照图 4 到图 22)。

[0124] 图 2 表示关于本实施例的计算机系统 500 的网络结构。

[0125] 计算机系统 500 具有:多个主计算机 100a、100b;一个以上的管理服务器 110;一个以上的光纤通道交换机 (fiber channal swtich) 120;多个存储系统 130a、130b;多个管理终端 140a、140b 以及多个外部存储系统 150a、150b。在以下的说明中,在不区别主计算机 100a、100b 时,总称为主计算机 100。在不区别存储系统 130a、130b 时,总称为存储系统 130。在不区别管理终端 140a、140b 时,总称为管理终端 140。在不区别外部存储系统 150a、150b 时,总称为外部存储系统 150。

[0126] 主计算机 100、存储系统 130 以及外部存储系统 150,分别通过端口 107、131、151 和光纤通道交换机 120 的端口 121 连接。另外,主计算机 100、存储系统 130、外部存储系统 150 以及光纤通道交换机 120,分别从接口控制部 106、138、157、123 通过 IP 网络 175 与管理服务器 110 相连。管理服务器 110 管理这些网络节点 (主计算机 100、存储系统 130、外部存储系统 150 以及光纤通道交换机 120)。

[0127] 此外,在本实施例中,存储系统 130 具有通过管理终端 140 连接管理服务器 110 的网络结构,但也可以是存储系统 130 直接与 IP 网络 175 相连的网络结构。

[0128] 主计算机 100 具有 CPU101、存储器 102、存储装置 103、输入装置 104、显示器 105、接口控制部 106 以及端口 107。存储装置 103 例如是盘装置、光磁盘装置等。主计算机 100 把在存储装置 103 中存储的操作系统或者应用程序等软件读入存储器 102,通过由 CPU101 读出该软件后执行,来执行信息处理。输入装置 104 是键盘或者鼠标等。主计算机 100,通过输入输出装置 104 接收来自主机管理者等的输入,在显示器 105 上显示信息处理结果等。接口控制部 106 是用于与 IP 网络 175 相连的 LAN 适配器等。端口 107 是用于与存储系统 130 相连的主机总线适配器 (host bus adapter) 等。

[0129] 管理服务器 110 具有 CPU111、存储器 112、存储装置 113、输入装置 114、显示器 115 以及接口控制部 116。存储装置 113 例如是盘装置、光磁盘装置等。管理服务器 110 把在存储装置 113 中存储的存储管理软件等读入存储器 112,通过由 CPU111 读出该存储管理软件后执行,来维护管理计算机系统 500 整体。当通过 CPU111 执行存储管理软件时,管理服务器 110 从接口控制部 116 通过 IP 网络 175 从计算机系统 500 内的各网络节点收集结构信息、资源利用率、性能监视信息、故障日志 (log) 等。然后,管理服务器 110 把收集到的这些信息向显示器 115 等输出装置输出,提示存储系统管理者。另外,管理服务器 110,通过键盘或者鼠标等输入装置 114 从存储系统管理者接收关于维护管理的指示,将该指示向网络节点发送。

[0130] 存储系统 130, 具有用于与存储网络相连的一个以上的端口 131、用于控制向盘装置 137 的数据输入输出处理的一个以上的控制处理器 132、作为各个控制处理器 132 的工作区而工作的一个以上的存储器 133、用于临时存储向盘装置 137 输入输出的数据的一个以上的盘高速缓冲存储器 134、用于存储存储系统 130 的结构信息等的一个以上的控制存储器 135、作为用于与盘装置 137 相连的接口而工作的一个以上的端口 136、用于存储数据的盘装置 137。

[0131] 控制处理器 132, 根据在通过端口 131 从主计算机 100 接收到的输入输出请求中包含的信息 (端口 ID 以及 LUN), 确定访问目的地的逻辑单元。如果访问目的地的逻辑单元与盘装置 137 (内部设备) 对应, 则控制处理器 132 控制向盘装置 137 的数据输入输出。如果访问目的地的逻辑单元与外部存储系统 150 (外部设备) 对应, 则控制处理器 132 控制向外部存储系统 150 的数据输入输出。

[0132] 此外, 在本实施例中, 作为端口 131, 设为与把 SCSI (Small Computer System Interface) 作为上位协议的光纤通道接口对应的端口, 但是也可以是与把 SCSI 作为上位协议的 IP 网络接口相对应的端口。

[0133] 存储系统 130 具有如下存储层。通过多个盘装置 137 构成盘阵列 (diskarray), 该盘阵列作为物理设备由控制处理器 132 管理。进而, 控制处理器 132 给在存储系统 130 内搭载的物理设备分配逻辑设备 (即, 控制处理器 132 将物理设备和逻辑设备关联对应)。逻辑设备, 在存储系统 130 内被管理, 其 LDEV 号码由每一存储系统 130 独立管理。

[0134] 逻辑设备与分配给各端口 131 的 LUN 关联对应, 作为存储系统 130 内的设备被提供给主计算机 100。即, 主计算机 100 识别的是存储系统 130 内的逻辑设备, 主计算机 100 使用用于识别分配给端口 131 的逻辑设备的 LUN, 访问在存储系统 130 内存储的数据。

[0135] 此外, 在本实施例中, 控制处理器 132 把外部存储系统 150 内的逻辑设备作为外部设备来管理, 作为存储系统 130 内的设备 (内部设备) 进行虚拟化。即, 由存储系统 130 管理的外部设备是虚拟设备。使用外部存储连接技术, 取入到存储系统 130 内的外部设备与存储系统 130 内的物理设备同样地作为存储系统 130 内的逻辑设备被进行管理。

[0136] 为实现以上那样的存储层, 控制处理器 132 管理存储系统 130 内的逻辑设备、物理设备以及盘装置 137、和外部存储系统 150 内的逻辑设备、以及盘装置 156 的对应关系。并且, 控制处理器 132, 把来自主计算机 100 的 I/O 请求变换为对存储系统 130 内的盘装置 137 的访问请求、或者对外部存储系统 150 内的盘装置 156 的访问请求。

[0137] 此外, 如上所述, 本实施例中的存储系统 130, 汇总多个盘装置 137 的存储区域定义一个以上的物理设备 (亦即, 汇总多个盘装置 137 的存储区域, 与一个以上的物理设备关联对应), 向一个物理设备分配一个逻辑设备, 将其存储区域提供给主计算机 100。也可以使各个盘装置 137 分别与一个物理设备关联对应。

[0138] 控制处理器 132 不仅控制对设备的数据输入输出处理, 而且控制卷之间的数据复制或者数据再配置等。控制处理器 132 也可以通过接口控制部 138 连接管理终端 140, 接收由存储系统管理者向管理终端 140 输入的结构变更指示, 进行存储系统 130 的结构变更。

[0139] 盘高速缓冲存储器 134, 临时存储主计算机 100 和盘装置 137 之间输入输出的数据。把存储系统 130 从主计算机 100 接收到的数据存储于盘高速缓冲存储器 134 中, 在向主计算机 100 返回写结束报告后, 把向盘装置 137 降级数据的处理称为后写处理。

[0140] 在存储系统 130 进行后写处理那样构成的情况下,为防止在盘高速缓冲存储器 134 中存储的数据在被写入盘装置 137 前消失,理想的是给盘高速缓冲存储器 134 备份电源,或者将盘高速缓冲存储器 134 二重化等,来提高抗故障性。

[0141] 在控制存储器 135 中存储有存储系统 130 的结构信息(例如用于管理设备之间的对应关系的映射信息、高速缓冲存储器直通模式的设定信息等)。当在控制存储器 135 中存储的结构信息消失时,由于控制处理器 132 不能进行对盘装置 137 的访问,所以理想的是给控制存储器 135 备份电源,或者将控制存储器 135 二重化等,来提高抗故障性。

[0142] 管理终端 140 具有接口控制部 141、CPU142、存储器 143、存储装置 144、输入装置 145、显示器 146 以及接口控制部 147。接口控制部 141 是用于与存储系统 130 相连的通信接口。存储装置 144 例如是盘装置、光磁盘装置等。输入装置 145 是键盘或者鼠标等。显示器 146 提供存储管理用的用户界面环境,例如显示存储系统 130 的结构信息或者管理信息等。CPU142,通过把存储装置 144 中存储的存储管理程序读出到存储器 143 中并执行,进行结构信息的参照、结构信息的变更、特定功能的动作指示等。

[0143] 此外,在本实施例中也可以采用如下网络结构:不通过管理终端 140 而把存储系统 130 与管理服务器 110 直接连接,通过在管理服务器 110 上运行的管理软件来管理存储系统 130。

[0144] 外部存储系统 150 具有:用于通过光纤通道交换机 120 与存储系统 130 的端口 131 连接的一个以上的端口 151、用于控制对盘装置 156 的数据输入输出处理的一个以上的控制处理器 152、作为各控制处理器 152 的工作区域而工作的一个以上的存储器 153、用于临时存储向盘装置 156 输入输出的数据的一个以上的盘高速缓冲存储器 154、作为用于控制处理器 152 连接盘装置 156 的接口而工作的一个以上的端口 155、用于存储数据的一个以上的盘驱动器 156、作为用于与 IP 网络相连的接口而工作的接口控制部 157。

[0145] 此外,在本实施例中,因为采用通过光纤通道交换机 120 将存储系统 130 上装备的端口 131、和外部存储系统 150 上装备的端口 151 相连的网络结构,所以为了制止从主计算机 100 向外部存储系统 150 的直接访问,理想的是在光纤通道交换机 120 上设定分区(zoning)。也可以不一定在存储网络上设置光纤通道交换机 120、而采用例如将存储系统上装备的端口 131、和外部存储系统 150 上装备的端口 151 直接相连的网络结构。

[0146] 图 3 表示关于本实施例的存储系统 130 内构建的存储层。与图 1 到图 2 所示符号相同的符号,表示相同的逻辑资源或者物理资源,省略其详细的说明。

[0147] 主计算机 100 具有链路管理器 108 以及 RAID 管理器 109。链路管理器 108 是用于切换向存储系统 130 的访问路径的软件程序。RAID 管理器 109 是用于管理存储系统 130 的存储资源的存储系统管理软件,向存储系统 130 给出各种指示(例如逻辑单元移动指示)。

[0148] 在存储系统 130 上装备的某端口 131,通过存储网络 510 与主计算机 100 相连,另一方面,另一端口 131 通过存储网络 510 与外部存储系统 150 相连。存储网络 510 是包含一个以上的光纤通道交换机 120 的 FC-SAN(FiberChannel Storage Area Network) 等的网络。

[0149] 存储系统 130 具有由逻辑单元 201、逻辑设备 202、虚拟设备 203、设备组 204、物理设备 205 以及扩展设备 206 组成的存储层。设备组 204,例如是集合多个物理设备 205 或者扩展设备 206 中任何一方或两者各自的存储区域的全部或者一部分而形成的存储区域、或

者选出物理设备 205 或扩展设备 206 中任何一方或者两者的存储区域的一部分而形成的存储区域。

[0150] 此外,物理设备 205 是虚拟化存储系统 130 所具有的盘装置 137 的存储区域(内部设备),与之相对,扩展设备 206 是虚拟化外部存储系统 150 具有的盘装置 156 的存储区域(外部设备),虽然两者这点不同,但是两者都虚拟化实际设备这点共同,因此属于同一层(最下层)。

[0151] 存储系统 130 具有分配给端口 131 的 CMD(Command Device)207。CMD207 是用于在主计算机 100 和存储系统 130 之间交接命令或者状态(status)的逻辑单元。将从主计算机 100 向存储系统 130 发送的命令写入 CMD207。存储系统 130 执行与写入 CMD207 的命令对应的处理,将其执行结果作为状态写入 CMD207。主计算机 100 读出写入 CMD207 的状态并确认,在 CMD207 中写入下一要执行的命令。这样,主计算机 100 可以向存储系统 130 发出各种指示。

[0152] 作为主计算机 100 向存储系统 130 的指示内容,例如有以逻辑单元单位的设备移动指示(逻辑单元移动指示)或者以逻辑设备单位的设备移动指示(逻辑设备移动指示)。关于逻辑单元移动处理的概要如上述。在移动源逻辑单元 201 有 LUSE 结构的情况下,主计算机 100 能够向存储系统 130 发出逻辑设备移动指示。在具有 LUSE 结构的逻辑单元 201 中,因为把多个逻辑设备 202 映射到一个逻辑单元 201(参照图 5),所以通过指定多个逻辑设备 202 中一个逻辑设备 202(例如分配给逻辑单元 201 的开始地址的逻辑设备 202),也可以把映射到该逻辑单元 201 的其他全部逻辑设备 202 指定为移动对象。

[0153] 此外,存储系统 130,也可以把满足规定条件(例如在存储系统 130 内发生故障的情况,或者存储系统 130 的性能降低的情况等)作为契机,自发地(不等待来自主计算机 100 的指示)执行逻辑单元移动处理。关于把存储系统 130 内发生故障的情况或者存储系统 130 的性能降低的情况作为契机来移动逻辑单元的处理的概要,在后面进行描述。

[0154] 另外,管理服务器 110 也可以经由 IP 网络 175 以及管理终端 140 对存储系统 130 给予逻辑单元移动指示。亦即可以从主计算机 100 通过 in-band 途径(存储网络途径)对存储系统 130 给予逻辑单元移动指示,或者可以从管理服务器 110 通过 out-band 途径(IP 网络途径)对存储系统 130 给予逻辑单元移动指示。

[0155] 下面参照图 4 到图 22,对各存储层相互之间的映射关系变化加以说明。

[0156] (1)CVS

[0157] 如图 4 所示,把一个虚拟设备 203 分割为多个 VDEV 区域 203-1、203-2、203-3,给各 VDEV 区域 203-1、203-2、203-3 分配一个逻辑设备 202 的映射关系称为 CVS。各 VDEV 区域 203-1、203-2、203-3 的大小不一定相同。

[0158] (2)LUSE

[0159] 如图 5 所示,把多个逻辑设备 202 的存储区域集合,分配给一个逻辑单元 201 的映射关系称为 LUSE。

[0160] (3)外部卷 RAID

[0161] 如图 6 所示,把进行过 RAID 构成的扩展设备 206 分配给设备组 204 的映射关系称为外部卷 RAID。例如,在通过 RAID1 进行 RAID 构成的多个扩展设备 206 中一方的扩展设备 206 与数据盘(data disk)对应,另一方的扩展设备 206 与镜像盘(mirror disk)对应。

[0162] (4) 附有 VMA 的卷

[0163] 图 7 表示附有 VMA 的卷。附有 VMA 的卷是在开放系统类型的系统中使用的逻辑设备 202。该逻辑设备 202 包含用于存储用户数据的用户区域 202-1、和用于存储管理信息的管理区域 202-2。在管理信息中,例如包含附有 VMA 的卷的访问属性。访问属性包含可读/写、只读、不可读/写、读出容量零 (read capacity zero)、抑制查询 (Inquiry)、禁止辅助卷 (S-VOL disable) 等信息。用户区域 202-1 为主计算机 100 可以识别、而且被分配给逻辑单元 201,与之相对,管理区域 202-2 为主计算机 100 不能识别、而且也不分配给逻辑单元 201。

[0164] (5) 大型机卷

[0165] 图 8 表示大型机卷 (mainframe volume)。大型机卷是大型机类型的系统中使用的逻辑设备 202。该逻辑设备 202 包含用于存储用户数据的用户区域 202-1、和用于存储管理信息的管理区域 202-3。在管理信息中,例如包含大型机卷的访问属性等。用户区域 202-1 为主计算机 100 可以识别、而且被分配给逻辑单元 201,与之相对,管理区域 202-3 主计算机 100 不能识别、而且也不被分配给逻辑单元 201。

[0166] (6) VDEV 连接

[0167] 如图 9 所示,把多个虚拟设备 203 的存储区域集合,分配给一个逻辑设备 202 的映射关系称为 VDEV 连接。

[0168] (7) VDEV 离散 (基本形态)

[0169] 如图 10 所示,把多个设备组 204 的存储区域集合,分配给一个虚拟设备 203 的映射关系称为 VDEV 离散。在 VDEV 离散的映射关系中,包含后述的变形例 1 ~ 2。

[0170] (8) VDEV 离散 (变形例 1)

[0171] 作为 VDEV 离散的一个变形例,例如图 11 所示,也可以把一个设备组 204 分割为多个存储区域,对分割所得的各个存储区域分配一个虚拟设备 203。

[0172] (9) VDEV 离散 (变形例 2)

[0173] 作为 VDEV 离散的另一个变形例,例如图 12 所示,也可以把多个设备组 204 的各个分割为多个存储区域,将分割所得的各个存储区域分配给多个虚拟设备 203。

[0174] (10) DEVGr/EDEV

[0175] 关于设备组 204 与扩展设备 206 之间的映射关系,例如图 13 所示,也可以将多个扩展设备 206 的各个的存储区域的全部或者一部分集合,分配给一个设备组 204。在扩展设备 206 中,也可以存在包含未分配给设备组 204 的存储区域的设备。还例如图 14 所示,也可以将多个扩展设备 206 的各个的存储区域的全部或者一部分集合,分配给多个设备组 204。在扩展设备 206 中,也可以存在包含未分配给任何设备组 204 的存储区域的设备。还例如图 15 所示,也可以存在分配给多个设备组 204 的扩展设备 206、或者仅分配给一部分设备组 204 的扩展设备 206。

[0176] (11) 不同种类介质并存的 RAID

[0177] 如图 16 所示,也可以在一个设备组 204 的存储区域内并存不同种类设备 (例如物理设备 205 以及扩展设备 206) 的形态下进行分配。还例如图 17 所示,也可以在一个设备组 204 的存储区域内并存不同种类盘驱动器 (例如 FC 盘驱动器 520 以及 SATA 盘驱动器 530) 的形态下进行分配。还例如图 18 到图 20 所示,也可以在一个设备组 204 的存储区域内并

存闪速存储器 (flashmemory) 540 以及盘驱动器 550 的形态下进行分配。

[0178] 在图 18 中,通过多个闪速存储器 540 和多个盘驱动器 550 构成 RAID0/1 结构的设备组 204。由多个闪速存储器 540 构成数据卷,由多个盘驱动器 550 构成镜像卷。在这样的驱动器结构中,在随机读时,从闪速存储器 540 读出数据。在顺序读时分别从闪速存储器 540 和盘驱动器 550 读出数据。在写时,分别在闪速存储器 540 和盘驱动器 550 中写入数据。通过这样的驱动器结构,发挥闪速存储器 540 的高速性,并且比仅有闪速存储器 540 的 RAID 成本低。

[0179] 在图 19 中,由三个闪速存储器 540 和一个盘驱动器 550 构成 RAID4 结构的设备组 204。闪速存储器 540 存储数据,盘驱动器 550 存储奇偶校验数据 (parity data)。在这样的驱动器结构中,通过不在访问数多的奇偶校验驱动器中使用有写入寿命的闪速存储器 540,能够延长闪速存储器 540 的使用时间。

[0180] 在图 20 中,由一个闪速存储器 540 和三个盘驱动器 550 构成 RAID4 结构的设备组 204。闪速存储器 540 存储数据,盘驱动器 550 存储奇偶校验数据。在这样的驱动器结构中,作为访问数多的奇偶校验驱动器,通过使用闪速存储器 540,可以提高写入性能。

[0181] (12) AOU

[0182] 图 21 表示存储系统 130 内的、称为 AOU 的存储层相互间的映射关系。逻辑单元 201 属于 LU 层。AOU 卷 202c 和多个池卷 (pool volume) 202d 属于 LDEV 层。多个池卷 203c 属于 VDEV 层。多个扩展设备 206e 属于 PDEV/EDEV 层。给主计算机 100 识别的逻辑单元 201 分配 AOU 卷 202c。给 AOU 卷 202c 分配多个池卷 203c 的各个的存储区域的一部分。可以给 AOU 卷 202c 动态分配池卷 203c 的存储区域,能够自由地增减 AOU 卷 202c 的容量。给多个池卷 203c 各个的存储区域分配多个扩展设备 206e 的各个的存储区域。此外,在 PDEV/EDEV 层中,也可以代替扩展设备 206e 配置物理设备。

[0183] (13) 快照

[0184] 图 22 表示在具有快照功能的存储系统 130 内的存储层之间的映射关系。逻辑单元 210 属于 LU 层。主卷 (primary volume) 202e、辅助卷 (secondary volume) 202f、多个池卷 202g 属于 LDEV 层。虚拟设备 203d、多个池卷 203e 属于 VDEV 层。扩展设备 206f、206g 属于 PDEV/EDEV 层。给逻辑单元 210 分配逻辑设备 202e。给逻辑设备 202e 分配虚拟设备 203d。给虚拟设备 203d 分配扩展设备 206f。辅助卷 202f 是:用于根据某时刻在主卷 202e 中存储的数据和某时刻以后从主卷 202e 保存到池卷 203e 的数据,来恢复某时刻的主卷 202e 的数据图像的虚拟逻辑卷。

[0185] 下面参照图 23 对主机组 (host group) 加以说明。属于主机组号码 #0 的主机组的多个主计算机 100a、100c 中的任何一个,通过指定端口号码 #0 作为端口 131 的号码、以及指定 LUN#0 作为逻辑单元 201 的号码,都能够访问逻辑设备号码 #0 的逻辑设备 202h。另一方面,属于主机组号码 #1 的主机组的主计算机 100b,通过指定端口号码 #0 作为端口 131 的号码、以及指定 LUN#0 作为逻辑单元 201 的号码,能够访问逻辑设备号码 #1 的逻辑设备 202i。这样,即使某主计算机 100 指定相同的端口号码以及相同的逻辑单元号码,根据该主计算机 100 所属的主机组号码,可以访问的逻辑设备 202 不同。

[0186] 在关于本实施形态的逻辑单元移动处理中,在使属于某主机组的一个主计算机 100 可利用的逻辑单元 201 向其他存储系统 130 移动的情况下,对于属于该主机组的其他主

计算机 100 也执行同样的逻辑单元移动处理。

[0187] 下面参照图 24 到图 25 对替换路径加以说明。如图 24 所示,当存在用于从主计算机 100 连接存储系统 130 内的某逻辑设备 202j 的多个路径 406、407 的情况下,当把该多个路径 406、407 中某一个路径 406 作为常用类型路径时,另一路径 407 成为替换路径。替换路径有时也称为待机类型路径。主计算机 100,通过指定端口号码 #0 作为端口 131 的号码、指定 LUN#0 作为逻辑单元 201c 的号码,可以通过路径 406 访问逻辑设备 202j。另外,主计算机 100,通过指定端口号码 #1 作为端口 131 的号码、以及指定 LUN#0 作为逻辑单元 201d 的号码,可以通过路径 407 访问逻辑设备 202j。链路管理器 108 执行如下处理:在路径 406、407 中选择某一方来作为用于访问逻辑设备 202j 的路径的处理。逻辑单元 201c、201d 都是虚拟化同一逻辑设备 202j 的单元,所以在使一个逻辑单元(例如逻辑单元 201c)向另一存储系统 130 移动的情况下,也需要使另一个逻辑单元(例如逻辑单元 201d)向另一存储系统 130 移动。

[0188] 如图 25 所示,当存在用于从存储系统 130 连接外部存储系统 150 内的某逻辑设备 202k 的多个路径 408、409 的情况下,当把该多个路径 408、409 中某一个路径 408 作为常用类型路径时,另一路径 409 成为替换路径。存储系统 130,通过指定端口号码 #0 作为端口 151 的号码、指定 LUN#0 作为逻辑单元 201e 的号码,可以通过路径 408 访问逻辑设备 202k。另外,存储系统 130,通过指定端口号码 #1 作为端口 151 的号码、以及指定 LUN#0 作为逻辑单元 201f 的号码,可以通过路径 409 访问逻辑设备 202k。逻辑单元 201e、201f 都是虚拟化同一逻辑设备 202k 的单元,所以在使一个逻辑单元(例如逻辑单元 201e)向另一存储系统 130 移动的情况下,也需要使另一个逻辑单元(例如逻辑单元 201f)向另一存储系统 130 移动。

[0189] 下面参照图 26 到图 33 对在逻辑单元移动处理中使用的各种信息加以说明。

[0190] 图 26 表示主机组管理信息 601。主机组管理信息 601 保存关于主机组的的信息。主机组管理信息 601 保存从条目(entry)701 到条目 705 的信息。

[0191] 在条目 701 中,作为计算机系统 500 内用于识别主机组的识别信息,存储主机组号码。

[0192] 在条目 702 中,存储属于通过主机组号码确定的主机组的主计算机 100 的台数。

[0193] 在条目 703 中,存储属于通过主机组号码确定的主机组的全部主计算机 100 的主机名(host name)的列表(list)。作为主机名,只要是能够唯一识别主计算机 100 的信息,就不特别限定,例如,可以使用对端口 107 指定的 WWN(World Wide Name)。

[0194] 在条目 704 中,存储可从主计算机 100 访问的逻辑单元的个数,该主计算机 100 属于通过主机组号码而确定的主机组。

[0195] 在条目 705 中,以列表形式存储后述的 LU 路径管理信息 602。

[0196] 图 27 表示 LU 路径管理信息 602。LU 路径管理信息 602,关于存储系统 130 内的各端口 131,保存对各端口 130 定义的有效的 LUN 的个数数量的信息。LU 路径管理信息 602 保存从条目 801 到条目 805 的信息。

[0197] 在条目 801 中,存储用于唯一识别存储系统 130 内的端口 131 的端口号码。

[0198] 在条目 802 中,存储用于识别逻辑单元 201 的信息(例如目标 ID 或者 LUN)。目标 ID 以及 LUN 是用于识别逻辑单元 201 的识别符。例如使用在从主计算机 100 通过 SCSI 协

议访问设备时所使用的 SCSI-ID 和 LUN, 作为用于识别逻辑单元 201 的识别符。

[0199] 在条目 803 中, 存储分配有在条目 802 中存储的 LUN 的逻辑设备的号码 (LDEV 号码)。存储在逻辑单元 201 具有 LUSE 结构的情况下的、分配给逻辑单元 201 的多个逻辑设备 202 中代表性的逻辑设备 202 的号码。

[0200] 在条目 804 中, 存储可对通过在条目 803 中存储的 LDEV 号码确定的逻辑设备进行访问的主计算机 100 的台数。

[0201] 在条目 805 中, 存储可对通过在条目 803 中存储的 LDEV 号码确定的逻辑设备进行访问的全部主计算机 100 的主机名的列表。

[0202] 图 28 表示逻辑设备管理信息 603。逻辑设备管理信息 603 关于各逻辑设备 202 保存从条目 901 到 918 的信息。

[0203] 在条目 901 中存储用于识别逻辑设备 202 的 LDEV 号码。

[0204] 在条目 902 中存储通过 LDEV 号码确定的逻辑设备 202 的容量 (大小)。

[0205] 在条目 903 中存储通过 LDEV 号码确定的逻辑设备 202 的 LDEV 类型。在存储系统 130 中, 可以定义盘高速缓冲存储器上的数据管理单位或者设备管理信息的存储形态 (有无向盘空间存储管理信息或者存储形态等) 不同的、多个设备类型的逻辑设备 202。LDEV 类型表示各逻辑设备 202 是哪种设备类型。

[0206] 在条目 904 中存储通过 LDEV 号码确定的逻辑设备 202 的 LDEV 状态。LDEV 状态表示逻辑设备 202 的设备状态。在 LDEV 状态中有“联机 (on line)”、“脱机 (offline)”、“未安装”、“封锁 (blocked)”。“联机”表示逻辑设备 202 正常运行、在一个以上的端口 131 上定义 LU 路径、从主计算机 100 可以访问的状态。“脱机”表示逻辑设备 202 被定义、正常运行、但是由于未定义 LU 路径等原因处于不能从主计算机 100 访问的状态。“未安装”表示由于未给逻辑设备 202 分配物理设备或者外部设备等原因处于不能从主计算机 100 访问的状态。“封锁”表示在逻辑设备 202 上发生故障不能从主计算机 100 访问的状态。条目 603d 的初始值是“未安装”, 但是通过逻辑设备定义处理更新为“脱机”, 进而通过 LU 路径定义处理更新为“联机”。

[0207] 在条目 905 中存储通过 LDEV 号码而确定的逻辑设备 202 的 LUSE 状态。LUSE 状态, 表示逻辑单元 201 是否具有 LUSE 结构, 并且表示在逻辑单元 210 具有 LUSE 结构时分配给逻辑单元 201 的多个逻辑设备 202 中代表性的逻辑设备的 LDEV 号码。

[0208] 在条目 906 中存储连接 LDEV 数。连接 LDEV 数表示在逻辑单元 210 具有 LUSE 结构的情况下分配给逻辑单元 201 的逻辑设备 202 的个数。

[0209] 在条目 907 中存储连接 LDEV 号码列表。连接 LDEV 号码列表表示在逻辑单元 210 具有 LUSE 结构的情况下分配给逻辑单元 201 的逻辑设备 202 的 LDEV 号码的列表。

[0210] 在条目 908 中存储 LU 路径数。所谓 LU 路径, 是用于主计算机 100 与逻辑单元 201 连接的路径。LU 路径数表示 LU 路径的条数。

[0211] 在条目 909 中, 以列表形式存储: 各个主机组的号码、用于识别可从属于主机组的主计算机 100 访问的端口 131 的端口号码、用于识别分配给端口 131 的逻辑单元 201 的信息 (目标 ID 或者 LUN)。

[0212] 在条目 910 中存储属于各个主机组的主计算机 100 的台数。

[0213] 在条目 911 中以列表形式存储在许可多个主计算机 100 对一个逻辑单元 201 访问

时的、主计算机 100 的主机名。

[0214] 在条目 912 中,存储在逻辑设备 202 具有 VDEV 连接结构时的、分配给逻辑设备 202 的虚拟设备 203 的个数。

[0215] 在条目 913 中,以列表形式存储在逻辑设备 202 和虚拟设备 203 不是一对一对应的情况下的虚拟设备 203 的号码 (VDEV 号码)。

[0216] 在条目 914 中存储表示逻辑设备 202 与虚拟设备 203 之间的映射关系的信息。例如在条目 914 中存储:在逻辑设备 202 具有 CVS 结构时的、虚拟设备 203 上的逻辑设备 202 的偏移 (offset) 位置,或者在逻辑设备 202 具有 VDEV 连接结构时的、逻辑设备 202 上的虚拟设备 203 偏移位置等。

[0217] 在条目 915 中存储在逻辑设备 202 处于同步复制中或者移动中的情况下的、复制目的地或者移动目的地逻辑设备 202 的号码。在逻辑设备 202 既不处于同步复制中也不处于移动中的情况下,在条目 915 中存储无效值。

[0218] 在条目 916 中,存储在逻辑设备 202 处于数据移动中的情况下的、表示数据移动结束后的最终地址的信息 (以下称为数据移动进展指针 (point))。在如下情况下使用数据移动进展指针:存储系统 130 从主计算机 100 接收到向逻辑设备 202 的 I/O 请求时判定访问目的地存储区域已完成移动还是未完成移动、并进行选择处理。

[0219] 在条目 917 中,存储表示逻辑设备 202 处于同步复制中、或处于数据移动中、或不对应于两种状态中任意一种状态的信息 (以下称同步复制 / 数据移动中标志)。

[0220] 在条目 918 中存储加密标志。所谓加密标志,是表示是否加密在物理设备 205 或者扩展设备 206 中写入的数据的标志信息。当把加密标志设定为开 (ON) 时,存储系统 130 加密在物理设备 205 或者扩展设备 206 中写入的数据。

[0221] 图 29 表示虚拟设备管理信息 604。虚拟设备管理信息 604 关于各虚拟设备 203 保存从条目 1001 到 1010 的信息。

[0222] 在条目 1001 中存储用于识别虚拟设备 203 的 VDEV 号码。

[0223] 在条目 1002 中存储通过 VDEV 号码确定的虚拟设备 203 的容量 (大小)。

[0224] 在条目 1003 中存储通过 VDEV 号码确定的虚拟设备 203 的 VDEV 状态。VDEV 状态表示虚拟设备 203 的设备状态。在 VDEV 状态中有“联机”、“脱机”、“未安装”、“封锁”。“联机”表示虚拟设备 203 正常运行、而且定义了从主计算机 100 到虚拟设备 203 的路径、从主计算机 100 可以访问的状态。“脱机”表示虚拟设备 203 被定义、正常运行、但是由于未定义向虚拟设备 203 的路径等原因处于不能从主计算机 100 访问的状态。“未安装”表示由于未给虚拟设备 203 分配物理设备或者外部设备等原因处于不能从主计算机 100 访问的状态。“封锁”表示在虚拟设备 203 上发生故障,不能从主计算机 100 访问的状态。条目 1003 的初始值是“未安装”,但是通过虚拟设备定义处理,更新为“脱机”,进而通过路径定义处理,更新为“联机”。

[0225] 在条目 1004 中存储通过 VDEV 号码确定的虚拟设备 203 的 RAID 结构。在 RAID 结构中,例如有 RAID 级 (RAID level)、数据奇偶校验数、驱动器结构、条带大小 (stripe size) 等。

[0226] 在条目 1005 中,存储在逻辑设备 202 具有 CVS 结构时的、从一个虚拟设备 203 分配的逻辑设备 202 的个数。

[0227] 在条目 1006 中,以列表形式存储有分配通过 VDEV 号码确定的虚拟设备 203 的逻辑设备 202 的号码。

[0228] 在条目 1007 中,以列表形式存储表示逻辑设备 202 和虚拟设备 203 之间的映射关系的信息。例如,在条目 1007 中存储在逻辑设备 202 具有 CVS 结构时的、虚拟设备 203 上的逻辑设备 202 的偏移位置、或者在逻辑设备 202 具有 VDEV 连接结构时的、逻辑设备 202 上的虚拟设备 203 的偏移位置等。

[0229] 在条目 1008 中存储分配给虚拟设备 203 的设备组 204 的个数。例如,在条目 1008 中存储在虚拟设备 203 具有 VDEV 离散的结构时的、分配给虚拟设备 203 的设备组 204 的个数。还例如条目 1008 用于在逻辑单元移动处理中对一个虚拟设备 203 分配多个设备组 204。

[0230] 在条目 1009 中以列表形式存储分配给虚拟设备 203 的设备组 204 的号码。

[0231] 在条目 1010 中存储表示虚拟设备 203 与设备组 204 之间的映射关系的信息。

[0232] 图 30 表示设备组管理信息 605。设备组管理信息 605 关于各设备组 204 保存从条目 1101 到 1110 的信息。

[0233] 在条目 1101 中存储用于识别设备组 204 的设备组号码。

[0234] 在条目 1102 中存储通过设备组号码确定的设备组 204 的容量(大小)。

[0235] 在条目 1103 中存储通过设备组号码确定的设备组 204 的 DEVGr 状态。DEVGr 状态表示设备组 204 的设备状态。在 DEVGr 状态中有“联机”、“脱机”、“未安装”、“封锁”。“联机”表示设备组 204 正常运行、而且定义了从主计算机 100 到设备组 204 的路径、从主计算机 100 可以访问的状态。“脱机”表示设备组 204 被定义、正常运行、但是由于未定义向设备组 204 的路径等原因处于不能从主计算机 100 访问的状态。“未安装”表示由于未给设备组 204 分配物理设备或者外部设备等原因处于不能从主计算机 100 访问的状态。“封锁”表示在设备组 204 上发生故障,不能从主计算机 100 访问的状态。条目 1103 的初始值是“未安装”,但是通过设备组定义处理,更新为“脱机”,进而通过路径定义处理,更新为“联机”。

[0236] 在条目 1104 中存储通过设备组号码确定的设备组 204 的 RAID 结构。在 RAID 结构中,例如有 RAID 级、数据奇偶校验数、驱动器结构、条带大小等。但是,如在逻辑单元移动处理中的情况下,在给一个设备组 204 分配一个扩展设备 206 时,对于设备组 204 不应用 RAID 结构。

[0237] 在条目 1105 中,存储分配给设备组 204 的虚拟设备 203 的个数。在使具有 LUSE 结构的逻辑单元 201 移动时,因为有时给一个设备组 204 分配多个虚拟设备 203,所以在这样的情况下更新条目 1105 的值。

[0238] 在条目 1106 中,以列表形式存储分配给通过设备组号码确定的设备组 204 的虚拟设备 203 的号码。

[0239] 在条目 1107 中,以列表形式存储表示设备组 204 与虚拟设备 203 之间的映射关系的信息。

[0240] 在条目 1108 中存储分配给设备组 204 的设备(物理设备 205 或者扩展设备 206)的个数。

[0241] 在条目 1109 中以列表形式存储分配给设备组 204 的设备(物理设备 205 或者扩展设备 206)的号码。

[0242] 在条目 1110 中,以列表形式存储表示设备组 204 与设备(物理设备 205 或者扩展设备 206)之间的映射关系的信息。例如在条目 1110 中,如图 13 到图 15 所示,存储在给设备组 204 仅分配扩展设备 206 的一部分那样的情况下的、从扩展设备 206 的先头起的地址和容量等。

[0243] 图 31 表示 VDEV-DEVGr 映射信息 606。VDEV-DEVGr 映射信息 606 保存从条目 1201 到 1207 的信息。

[0244] 在条目 1201 中存储用于识别 VDEV 区域的号码。所谓 VDEV 区域指虚拟设备 203 的一部分的区域。

[0245] 在条目 1202 中存储与 VDEV 区域对应关联的设备组 204 的个数。

[0246] 在条目 1203 中以列表形式存储与 VDEV 区域对应关联的设备组 204 的号码。

[0247] 在条目 1204 中以列表形式存储有效标志。所谓有效标志,是表示是否在分配给逻辑单元 201 的虚拟设备 203 中执行镜像或者奇偶校验生成等处理的标志信息。在逻辑单元移动处理中,把有效标志设定为关(OFF),不实施针对分配给移动源的逻辑单元 201 的虚拟设备 203 的镜像或者奇偶校验生成等处理。

[0248] 在条目 1205 中,以列表形式存储与设备组 204 对应关联的 VDEV 区域的偏移地址。

[0249] 在条目 1206 中,关于各 VDEV 区域,以列表形式存储表示是否以高速缓冲存储器直通读的动作模式动作的信息。

[0250] 在条目 1207 中,关于各 VDEV 区域,以列表形式存储表示是否以高速缓冲存储器直通写的动作模式动作的信息。

[0251] 图 32 表示物理设备管理信息 607。物理设备管理信息 607 关于在存储系统 130 内定义的各物理设备 205,保存从条目 1301 到 1306 的信息。

[0252] 在条目 1301 中存储用于识别物理设备 205 的物理设备号码。

[0253] 在条目 1302 中存储通过物理设备号码确定的物理设备 205 的容量(大小)。

[0254] 在条目 1303 中存储通过物理设备号码确定的物理设备 205 的设备状态。作为设备状态,有“联机”、“脱机”、“未安装”、“封锁”。“联机”表示物理设备 205 正常运行、而且该物理设备 205 被分配给逻辑设备 202 的状态。“脱机”表示物理设备 205 被定义、正常运行、但是该物理设备 205 未被分配给逻辑设备 202 的状态。“未安装”表示物理设备 205 未在盘装置 37 上定义的状态。“封锁”表示在物理设备 205 上发生故障不能访问的状态。条目 1303 的初始值设定为“未安装”,通过物理设备定义处理,更新为“脱机”,以在逻辑设备 202 上定义物理设备 205 为契机更新为“联机”。

[0255] 在条目 1304 中,在把一个物理设备 205 的存储区域部分地分配给多个设备组 204 那样的情况下(参照图 15),存储分配给一个物理设备 205 的设备组 204 的个数。

[0256] 在条目 1305 中,以列表形式存储分配给物理设备 205 的设备组 204 的号码。

[0257] 在条目 1306 中存储表示物理设备 205 与设备组 204 之间的映射关系的信息。

[0258] 图 33 表示扩展设备管理信息 608。扩展设备管理信息 608,是用于把外部存储系统 150 具有的外部卷 301 虚拟为存储系统 130 内的扩展设备 206(内部卷)的信息。扩展设备管理信息 608,关于各外部卷 301 保存从条目 1401 到条目 1411 的信息。

[0259] 在条目 1401 中存储用于识别扩展设备 206 的 EDEV 号码。EDEV 号码是用于在存储系统 130 内唯一识别外部卷 301 的识别号码。

[0260] 在条目 1402 中存储通过 EDEV 号码确定的扩展设备 206 的容量（大小）。

[0261] 在条目 1403 中存储通过 EDEV 号码确定的扩展设备 206 的设备状态。作为设备状态，有“联机”、“脱机”、“未安装”、“封锁”。“联机”表示扩展设备 206 正常运行、而且该扩展设备 206 被分配给逻辑设备 202 的状态。“脱机”表示扩展设备 206 被定义、正常运行、但是该扩展设备 206 未被分配给逻辑设备 202 的状态。“未安装”表示未定义扩展设备 206 的状态。“封锁”表示在扩展设备 206 上发生故障不能访问的状态。条目 1403 的初始值设定为“未安装”，通过扩展设备定义处理更新为“脱机”，以在逻辑设备 202 上定义扩展设备 206 为契机更新为“联机”。

[0262] 在条目 1404 中，在把一个扩展设备 206 的存储区域部分地分配给多个设备组 204 那样的情况下（参照图 15），存储分配给一个扩展设备 206 的设备组 204 的个数。

[0263] 在条目 1405 中，以列表形式存储分配给扩展设备 206 的设备组 204 的号码。

[0264] 在条目 1406 中存储表示扩展设备 206 与设备组 204 之间的映射关系的信息。

[0265] 在条目 1407 中存储用于识别虚拟化扩展设备 206 的外部存储系统 150 的存储系统识别信息。作为存储系统识别信息，例如可以考虑用于识别制造销售外部存储系统 150 的供货商的供货商识别信息和各供货商唯一分配的序列号的组合等。

[0266] 在条目 1408 中存储外部存储系统 150 管理的外部卷 301 的 LDEV 号码。

[0267] 在条目 1409 中存储端口号码，用于识别在存储系统 130 内的多个端口 131 中的、与外部存储系统 150 连接的启动器端口（initiator port）。

[0268] 在条目 1410 中，以列表形式存储在外部存储系统 150 内的一个以上的端口上定义外部卷 301 的情况下的、用于识别这些端口 151 的端口 ID、以及用于识别外部卷 301 的目标 ID 以及 LUN。

[0269] 在条目 1411 中存储访问控制信息。访问控制信息，是在多个存储系统 130 共享外部卷 301 的情况下，用于管理对于外部卷 301 的各区域哪个存储系统 130 具有访问权限的信息。

[0270] 图 34 表示计算机系统 500 的软件结构。在存储系统 130 内的控制存储器 135 中，存储各种结构管理信息（主机组管理信息 601、LU 路径管理信息 602、逻辑设备管理信息 603、虚拟设备管理信息 604、设备组管理信息 605、VDEV-DEVGr 映射信息 606、物理设备管理信息 607、扩展设备管理信息 608、以及设备功能管理信息 609）。

[0271] 设备功能管理信息 609 保存表示在各逻辑设备 202 中设定的各种属性的信息。作为逻辑设备 202 的属性信息的例子，可以举出：仅把对逻辑设备 202 的访问限制于特定的主计算机 100 上的访问限制信息、禁止对逻辑设备 202 的读访问或者写访问的访问属性信息、对于在逻辑设备 202 中存储的数据有无使用加密或者在数据的加密或者解密中使用的密钥信息等加密信息。

[0272] 在存储系统 130 内的存储器 133 中，在存储上述各种结构管理信息之外，存储存储系统 I/O 处理程序 221、以及逻辑单元移动处理程序 222。关于存储系统 I/O 处理程序 221 执行的存储系统 I/O 处理的细节、以及逻辑单元移动处理程序 222 执行的逻辑单元移动处理的细节，将在后面进行描述。

[0273] 在管理服务器 110 内的存储器 112 中，在存储上述各种结构管理信息之外，存储逻辑单元移动指示处理程序 241、以及存储系统负荷监视程序 242。关于逻辑单元移动指示处

理程序 241 执行的 LU 移动指示处理的细节、以及存储系统负荷监视程序 242 执行的处理的细节,将在后面进行描述。

[0274] 在主计算机 100 内的存储器 102 中,存储设备路径管理信息 251、以及主机 (host) I/O 处理程序 261。主计算机 100,识别存储系统 130 或者外部存储系统 150 提供的逻辑设备,使逻辑设备与设备文件关联对应进行管理。在设备路径管理信息 251 中,包含表示逻辑设备与设备文件的对应关系的信息。例如,在给存储系统 130 具有的多个端口 131 分配同一逻辑设备的情况下,从同一逻辑设备生成多个设备文件。主计算机 100 对与同一逻辑设备对应关联的多个设备文件进行一元管理,分散向逻辑设备的输入输出负荷、或控制网络故障时的路径切换。

[0275] 此外,在从管理服务器 110 或者管理终端 140 接收来自存储系统管理软件或者存储系统管理者的指示、变更存储系统 130 的结构的情况下,或者在存储系统 130 内的各部位由于故障或者自动交替而变更结构的情况下,位于存储系统 130 内的一个控制处理器 132 更新控制存储器 135 内的结构管理信息。进而,控制处理器 132 将由于存储系统 130 的结构变更引起结构管理信息被更新的情况通知其他控制处理器 132、管理终端 140、以及管理服务器 110,使其他控制存储器 135 中存储的全部结构管理信息更新为最新的信息。

[0276] 下面参照图 35 说明具有 CVS 结构的逻辑单位的移动处理的概要。

[0277] 存储系统 130a 具有逻辑单元 201a-1、201a-2、逻辑设备 202a-1、202a-2、虚拟设备 203a 以及扩展设备 206a。逻辑单元 201a-1 通过路径 401-1 连接主计算机 100。逻辑单元 201a-2 通过路径 401-2 连接主计算机 100。虚拟设备 203a 分割为多个存储区域 203a-1、203a-2。一个存储区域 203a-1 通过逻辑设备 202a-1 映射到逻辑单元 202a-1,另一个存储区域 203a-1 通过逻辑设备 202a-1 映射到逻辑单元 202a-1。各个逻辑单元 201a-1、201a-2 为虚拟化虚拟设备 203a 的一部分的单元,具有 CVS 结构。

[0278] 扩展设备 206a 是虚拟化外部卷 301 而得到的设备,被分割为多个存储区域 206a-1、206a-2。一个存储区域 206a-1 映射到存储区域 203a-1,另一个存储区域 206a-2 映射到存储区域 203a-2。

[0279] 对于将位于存储系统 130a(移动源存储系统)内的逻辑单元 201a-2(移动源逻辑单元)中的数据移动到存储系统 130b(移动目的地存储系统)内的逻辑单元 201b(移动目的地逻辑单元)的处理进行概括描述。逻辑单元移动处理包含下述的(1)~(9)的处理过程。

[0280] (1) 存储系统 130b,定义用于外部连接外部卷 301 和存储系统 130b 之间的路径 405,在存储系统 130b 内生成虚拟化外部卷 301 的扩展设备 206b。扩展设备 206b,被分割为多个存储区域 206b-1、206b-2。扩展设备 206b 的容量和外部卷 301 的存储容量相等。存储区域 206b-2 的存储容量等于逻辑单元 201a-2 的存储容量。

[0281] (2) 存储系统 130b,在存储系统 130b 内生成具有与外部卷 301 的存储容量相同的存储容量的虚拟设备 203b。虚拟设备 203b 被分割为多个存储区域 203b-1、203b-2。存储区域 203b-2 的存储容量等于逻辑单元 201a-2 的存储容量。

[0282] (3) 存储系统 130b,定义用于外部连接逻辑单元 201a2 和存储系统 130b 之间的路径 404,在存储系统 130b 内生成用于虚拟化逻辑单元 201a-2 的扩展设备 206c。

[0283] (4) 存储系统 130b,把扩展设备 206c 向虚拟设备 203b(更详细说,存储区域

203b-2) 进行映射。

[0284] (5) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑设备 202a-2 的逻辑结构相同的逻辑结构的逻辑设备 202b。虚拟设备 203b(更详细说, 存储区域 203b-2) 向逻辑设备 202b 映射。

[0285] (6) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑单元 201a-2 的逻辑结构相同的逻辑结构的逻辑单元 201b, 定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。虚拟设备 202b 向逻辑单元 201b 映射。

[0286] (7) 主计算机 100 从路径 401-2 切换到路径 403, 作为用于访问外部卷 301 的路径。此时, 删除路径 401-2 等, 禁止从主计算机 100 向逻辑单元 201a-2 的访问。此外, 从路径 401-2 到路径 403 的路径切换, 也可以由存储系统 130a、130b、或者管理服务器 110 等进行。另外, 不一定需要删除路径 401-2, 例如也可以把路径 401-2 设定为路径 403 的替换路径(例如用于在系统故障时使用的路径), 使路径 401-2 的状态成为无效不变, 保留路径 401-2。

[0287] 通过上述处理, 逻辑单元 201b 因为不仅是虚拟化逻辑单元 201a-2 而且也是虚拟化外部卷 301 的一部分, 所以在逻辑单元移动处理中的、从主计算机 100 向逻辑单元 201a-2 的数据输入输出路径成为: 路径 403 → 存储系统 130b 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 404 → 存储系统 130a 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 402 → 外部卷 301。

[0288] (8) 存储系统 130a, 为把在存储系统 130a 的盘高速缓冲存储器 (diskcache) 中积蓄的全部脏数据降级到外部卷 301, 把存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。此外, 存储系统 130a 可以自发地把动作模式设定为高速缓冲存储器直通模式, 或者也可以根据来自外部(例如, 存储系统 130b, 主计算机 100, 或者管理服务器 110 等)的指示, 把动作模式设定为高速缓冲存储器直通模式。

[0289] (9) 如果在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷 301, 则存储系统 130b 解除扩展设备 206c 和虚拟设备 203b 之间的映射关系, 把扩展设备 206b 向虚拟设备 203b 映射。并且, 存储系统 130b 删除路径 404。

[0290] 此外, 在具有 CVS 结构的逻辑单元的移动处理中, 因为通过多个存储系统 130a、130b 共享外部卷 301 的一部分, 所以需要针对外部卷 301 的排他控制。针对外部卷 301 的排他控制, 可以由外部存储系统 150 执行, 或者也可以由多个存储系统 130a、130b 联合执行。

[0291] 下面参照图 36 说明具有 LUSE 结构的逻辑单元的移动处理的概要。

[0292] 存储系统 130a, 具有逻辑单元 201a、逻辑设备 202a-1、202a-2、虚拟设备 203a-1、203a-2、以及扩展设备 206a-1、206a-2。逻辑单元 201a 通过路径 401 连接主计算机 100。逻辑单元 201a 是把多个逻辑设备 202a-1、202a-2 集成为一个的存储区域, 具有 LUSE 结构。扩展设备 206a-1 是虚拟化外部卷 301-1 的设备, 向虚拟设备 203a-1 映射。虚拟设备 203a-1 向逻辑设备 202a-1 映射。扩展设备 206a-2 是虚拟外部卷 301-2 的设备, 向虚拟设备 203a-2 映射。虚拟设备 203a-2 向逻辑设备 202a-2 映射。

[0293] 对于将位于存储系统 130a(移动源存储系统)内的逻辑单元 201a(移动源逻辑单元)中的数据移动到存储系统 130b(移动目的地存储系统)内的逻辑单元 201b(移动目的地逻辑单元)的处理进行概括描述。逻辑单元移动处理包含下述的(1)~(9)的处理过程。

[0294] (1) 存储系统 130b, 定义用于外部连接外部卷 301-1、301-2 和存储系统 130b 之间的路径 405-1、405-2, 在存储系统 130b 内生成虚拟化外部卷 301-1、301-2 的扩展设备 206b-1、206b-2。

[0295] (2) 存储系统 130b, 在存储系统 130b 内生成具有与外部卷 301-1 的存储容量相同的存储容量的虚拟设备 203b-1、以及具有与外部卷 301-2 的存储容量相同的存储容量的虚拟设备 203b-2。

[0296] (3) 存储系统 130b, 定义用于外部连接逻辑单元 201a 和存储系统 130b 之间的路径 404, 在存储系统 130b 内生成用于虚拟化逻辑单元 201a 的扩展设备 206c。扩展设备 206c 被分割成多个存储区域 206c-1、206c-2。

[0297] (4) 存储系统 130b, 把扩展设备 206c 向虚拟设备 203b-1、203b-2 映射 (更详细说, 把存储区域 206c-1 向虚拟设备 203b-1 映射, 把存储区域 206c-2 向虚拟设备 203b-2 映射)。

[0298] (5) 存储系统 130b, 在存储系统 130b 内分别生成具有与逻辑设备 202a-1 的逻辑结构相同的逻辑结构的逻辑设备 202b-1、以及具有与逻辑设备 202a-2 的逻辑结构相同的逻辑结构的逻辑设备 202b-2。虚拟设备 203b-1 向逻辑设备 202b-1 映射, 虚拟设备 203b-2 向逻辑设备 202b-2 映射。

[0299] (6) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑单元 201a 的逻辑结构相同的逻辑结构的逻辑单元 201b, 定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。逻辑设备 202b-1、202b-2 向逻辑单元 201b 映射。

[0300] (7) 主计算机 100, 作为用于访问外部卷 301-1、301-2 的路径, 从路径 401 切换到路径 403。此时, 删除路径 401 等, 禁止从主计算机 100 向逻辑单元 201a 的访问。此外, 从路径 401 到路径 403 的路径切换, 也可以由存储系统 130a、130b、或者管理服务器 110 等进行。另外, 不一定需要删除路径 401, 例如也可以把路径 401 设定为路径 403 的替换路径 (例如用于在系统故障时使用的路径), 使路径 401 的状态成为无效不变, 保留路径 401。

[0301] 通过上述处理, 逻辑单元 201b 因为不仅是虚拟化逻辑单元 201a 而且也是虚拟化外部卷 301-1、301-2, 所以在逻辑单元移动处理中的、从主计算机 100 向逻辑单元 201a 的数据输入输出路径成为: 路径 403 → 存储系统 130b 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 404 → 存储系统 130a 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 402-1、402-2 → 外部卷 301-1、301-2。

[0302] (8) 存储系统 130a, 为了把在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据降级到外部卷 301-1、301-2, 把存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。此外, 存储系统 130a 可以自发地把动作模式设定为高速缓冲存储器直通模式, 或者也可以根据来自外部 (例如, 存储系统 130b, 主计算机 100, 或者管理服务器 110 等) 的指示, 把动作模式设定为高速缓冲存储器直通模式。

[0303] (9) 如果在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷 301-1、301-2, 则存储系统 130b 解除扩展设备 206c 和虚拟设备 203b-1、203b-2 之间的映射关系, 把扩展设备 206b-1 向虚拟设备 203b-1 映射, 把扩展设备 206b-2 向虚拟设备 203b-2 映射。进而, 存储系统 130b 删除路径 404。

[0304] 下面参照图 37 说明具有外部卷 RAID 结构的逻辑单元的移动处理的概要。

[0305] 存储系统 130a, 具有逻辑单元 201a、逻辑设备 202a、虚拟设备 203a、以及扩展设备 206a-1、206a-2。逻辑单元 201a 经过路径 401 与主计算机 100 连接。逻辑单元 201a 是通过在多个外部卷 301-1、301-2 中应用 RAID 结构, 把这些外部卷 301-1、301-2 虚拟化为一个存储区域。作为在多个外部卷 301-1、301-2 中使用的 RAID 结构, 在本实施例中, 举例表示 RAID1 (镜像 (mirroring))。在 RAID1 结构中, 相对于外部卷 301-1 是主卷, 外部卷 301-2 是辅助卷 (镜像卷 (mirror volume))。作为在外部卷 301-1、301-2 中使用的 RAID 结构, 不限于 RAID1, 也可以使用 RAID0、RAID5、RAID6 等。扩展设备 206a-1 是虚拟化外部卷 301-1 的设备。扩展设备 206a-2 是虚拟化外部卷 301-2 的设备。虚拟设备 203a 是通过在多个扩展设备 206a-1、206a-2 中使用 RAID 结构由此虚拟为一个存储区域的设备。在虚拟设备 203a 中与主卷对应的存储区域映射到逻辑设备 202a。逻辑设备 202a 映射到逻辑单元 201a。

[0306] 对于将位于存储系统 130a (移动源存储系统) 内的逻辑单元 201a (移动源逻辑单元) 中的数据移动到存储系统 130b (移动目的地存储系统) 内的逻辑单元 201b (移动目的地逻辑单元) 的处理进行概括描述。逻辑单元移动处理包含下述的 (1) ~ (9) 的处理过程。

[0307] (1) 存储系统 130b, 定义用于外部连接外部卷 301-1、301-2 和存储系统 130b 之间的路径 405-1、405-2, 在存储系统 130b 内生成虚拟化外部卷 301-1、301-2 的扩展设备 206b-1、206b-2。

[0308] (2) 存储系统 130b, 在存储系统 130b 内生成用于虚拟化多个外部卷 301-1、301-2 的虚拟设备 203b。作为用于虚拟化多个外部卷 301-1、301-2 的 RAID 结构, 在使用镜像时, 虚拟设备 203b 的存储容量等于多个外部卷 301-1、301-2 的各自的存储容量的总和。

[0309] (3) 存储系统 130b, 定义用于外部连接逻辑单元 201a 和存储系统 130b 之间的路径 404, 在存储系统 130b 内生成用于虚拟化逻辑单元 201a 的扩展设备 206c。

[0310] (4) 存储系统 130b, 把扩展设备 206c 向虚拟设备 203b (更详细说, 虚拟设备 203b 中与主卷对应的存储区域) 映射。

[0311] (5) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑设备 202a 的逻辑结构相同的逻辑结构的逻辑设备 202b。虚拟设备 203b 中与主卷对应的存储区域向逻辑设备 202b 映射。

[0312] (6) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑单元 201a 的逻辑结构相同的逻辑结构的逻辑单元 201b, 定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。逻辑设备 202b 向逻辑单元 201b 映射。

[0313] (7) 主计算机 100, 作为用于访问外部卷 301-1、301-2 的路径, 从路径 401 切换到路径 403。此时, 删除路径 401 等, 禁止从主计算机 100 向逻辑单元 201a 的访问。此外, 从路径 401 到路径 403 的路径切换, 也可以由存储系统 130a、130b、或者管理服务器 110 等进行。另外, 不一定需要删除路径 401, 例如也可以把路径 401 设定为路径 403 的替换路径 (例如用于在系统故障时使用的路径), 使路径 401 的状态成为无效不变, 保留路径 401。

[0314] 通过上述处理, 逻辑单元 201b, 因为不仅是虚拟化逻辑单元 201a、而且也是虚拟化外部卷 301-1、301-2, 所以在逻辑单元移动处理中的、从主计算机 100 向逻辑单元 201a 的数据输入输出路径成为: 路径 403 → 存储系统 130b 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 404 → 存储系统 130a 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 402-1、402-2 → 外部卷 301-1、301-2。

[0315] (8) 存储系统 130a, 为了把在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据降级到外部卷 301-1、301-2, 把存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。此外, 存储系统 130a 可以自发地把动作模式设定为高速缓冲存储器直通模式, 或者也可以根据来自外部 (例如, 存储系统 130b, 主计算机 100, 或者管理服务器 110 等) 的指示, 把动作模式设定为高速缓冲存储器直通模式。

[0316] (9) 如果在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷 301-1、301-2, 则存储系统 130b 解除扩展设备 206c 和虚拟设备 203b 之间的映射关系, 把扩展设备 206b-1、206b-2 向虚拟设备 203b 映射。进而, 存储系统 130b 删除路径 404。

[0317] 下面参照图 38 说明附有 VMA 的卷的移动处理的概要。

[0318] 存储系统 130a, 具有逻辑单元 201a、逻辑设备 202a、虚拟设备 203a、以及扩展设备 206a。逻辑单元 201a 通过路径 401 连接主计算机 100。逻辑设备 202a 是附有 VMA 的卷, 包含可从主计算机 100 访问的用户区域 202a-1、和用于存储管理信息的管理区域 202a-2。在管理信息中, 例如包含附有 VMA 的卷的访问属性。用户区域 202a-1 由于被分配给逻辑单元 201a, 所以可为主计算机 100 识别。管理区域 202a-2 没有被分配给逻辑单元 201a 所以主计算机 100 不能识别。虚拟设备 203a 包含可从主计算机 100 访问的用户区域 203a-1、和用于存储管理信息的管理区域 203a-2。扩展设备 206a 包含可从主计算机 100 访问的用户区域 206a-1、和用于存储管理信息的管理区域 206a-2。扩展设备 206a 内的用户区域 206a-1 是虚拟化外部卷 301 内的用户区域 301-1 而得到的区域。扩展设备 206a 内的管理区域 206a-2 是虚拟化外部卷 301 内的管理区域 301-2 而得到的区域。用户区域 206a-1 向用户区域 203a-1 映射。管理区域 206a-2 向管理区域 203a-2 映射。用户区域 203a-1 向用户区域 202a-1 映射。管理区域 203a-2 向管理区域 202a-2 映射。

[0319] 此外, 移动源的存储系统 130a, 在不在外部卷 301 内存储管理信息、而在存储系统 130a 内的盘高速缓冲存储器或者其他存储器内存储管理信息时, 作为执行逻辑单元 201a 的移动处理的前准备, 需要将管理信息降级到外部卷 301。

[0320] 对将位于存储系统 130a (移动源存储系统) 内的逻辑单元 201a (移动源逻辑单元) 中的数据移动到存储系统 130b (移动目的地存储系统) 内的逻辑单元 201b (移动目的地逻辑单元) 的处理进行概括描述。逻辑单元移动处理包含下述的 (1) ~ (9) 的处理过程。

[0321] (1) 存储系统 130b, 定义用于外部连接外部卷 301 和存储系统 130b 之间的路径 405, 在存储系统 130b 内生成虚拟化外部卷 301 的扩展设备 206b。扩展设备 206b 包含可从主计算机 100 访问的用户区域 206b-1、和用于存储管理信息的管理区域 206b-2。存储系统 130b 将管理区域 301-2 向管理区域 206b-2 映射。

[0322] (2) 存储系统 130b, 在存储系统 130b 内生成用于虚拟化外部卷 301 的虚拟设备 203b。虚拟设备 203b 包含可从主计算机 100 访问的用户区域 203b-1、和用于存储管理信息的管理区域 203b-2。存储系统 130b 把管理区域 206b-2 向管理区域 203b-2 映射。

[0323] (3) 存储系统 130b, 定义用于外部连接逻辑单元 201a 和存储系统 130b 之间的路径 404, 在存储系统 130b 内生成用于虚拟化逻辑单元 201a 的扩展设备 206c。

[0324] (4) 存储系统 130b, 把扩展设备 206c 向虚拟设备 203b 内的用户区域 203b-1 映射。

[0325] (5) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑设备 202a 的逻辑结构相同的逻辑结构的逻辑设备 202b。逻辑设备 202b 包含可从主计算机 100 访问的用户区域

202b-1、和用于存储管理信息的管理区域 202b-2。使用户区域 203b-1 向用户区域 202b-1 映射。使管理区域 203b-2 向管理区域 202b-2 映射。

[0326] (6) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑单元 201a 的逻辑结构相同的逻辑结构的逻辑单元 201b, 定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。将用户区域 202b-1 向逻辑单元 201b 映射。

[0327] (7) 主计算机 100, 作为用于访问外部卷 301 的路径, 从路径 401 切换到路径 403。此时, 删除路径 401 等, 禁止从主计算机 100 向逻辑单元 201a 的访问。此外, 从路径 401 到路径 403 的路径切换, 也可以由存储系统 130a、130b、或者管理服务器 110 等进行。另外, 不一定需要删除路径 401, 例如也可以把路径 401 设定为路径 403 的替换路径 (例如为在系统故障时使用的路径), 使路径 401 的状态成为无效不变, 保留路径 401。

[0328] 通过上述处理, 逻辑单元 201b 因为不仅是虚拟化逻辑单元 201a、而且也是虚拟化外部卷 301 内的用户区域 301-1, 所以在逻辑单元移动处理中的、从主计算机 100 向逻辑单元 201a 的数据输入输出路径成为: 路径 403 → 存储系统 130b 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 404 → 存储系统 130a 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 402 → 外部卷 301。

[0329] 此外, 存储系统 130b, 因为通过虚拟设备 203b 虚拟化外部卷 130 内的管理区域 301-2, 所以可以通过路径 405 参照管理信息。在逻辑单元 201a 的移动处理中, 为了存储系统 130a 通过路径 402 参照外部卷内 130 内的管理信息, 需要把从虚拟设备 203b 内的管理区域 203b-2 向外部卷 301 内的管理区域 301-2 的写入设定为高速缓冲存储器直通模式, 始终更新管理区域 301-2 内的管理信息。通过移动目的地的存储系统 130b 进行在外部卷 301 内的管理区域 301-2 中写入的管理信息的更新。

[0330] (8) 存储系统 130a, 为了把在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据降级到外部卷 301, 把存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。此外, 存储系统 130a 可以自发地把动作模式设定为高速缓冲存储器直通模式, 或者也可以根据来自外部 (例如, 存储系统 130b, 主计算机 100, 或者管理服务器 110 等) 的指示, 把动作模式设定为高速缓冲存储器直通模式。

[0331] (9) 如果在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷 301, 则存储系统 130b 解除扩展设备 206c 和虚拟设备 203b 之间的映射关系, 把扩展设备 206b 内的用户区域 206b-1 向虚拟设备 203b 内的用户区域 203b-1 映射。进而, 存储系统 130b 删除路径 404。

[0332] 此外, 图 38 中表示的方法, 不仅在外部卷 301 是附有 VMA 的卷时、而且在是大型机卷时, 经过和上述的处理过程 (1) ~ (9) 同样的过程, 也可以使逻辑单元 201a 移动。

[0333] 下面参照图 39 说明附有 VMA 的卷的移动处理的另一种方法。

[0334] 存储系统 130a, 具有逻辑单元 201a-1、202a-1、逻辑设备 202a、虚拟设备 203a、以及扩展设备 206a。逻辑单元 201a-1 通过路径 401 连接主计算机 100, 可由主计算机 100 识别。相反地, 逻辑单元 201a-2 不连接主计算机 100, 则不能由主计算机 100 识别。仅在从存储系统 130a 向存储系统 130b 移动逻辑单元 201a-1 时临时生成逻辑单元 201a-2。逻辑设备 202a 是附有 VMA 的卷, 包含可从主计算机 100 访问的用户区域 202a-1、和用于存储管理信息的管理区域 202a-2。在管理信息中, 例如包含附有 VMA 的卷的访问属性。用户区域

202a-1 通过被分配给逻辑单元 201a-1, 可为主计算机 100 识别。尽管管理区域 202a-2 被分配给逻辑单元 201a-2, 但主计算机 100 不能识别。虚拟设备 203a 包含可从主计算机 100 访问的用户区域 203a-1、和用于存储管理信息的管理区域 203a-2。扩展设备 206a 包含可从主计算机 100 访问的用户区域 206a-1、和用于存储管理信息的管理区域 206a-2。扩展设备 206a 内的用户区域 206a-1 是虚拟化外部卷 301 内的用户区域 301-1 的区域。扩展设备 206a 内的管理区域 206a-2 是虚拟化外部卷 301 内的管理区域 301-2 的区域。用户区域 206a-1 向用户区域 203a-1 映射。管理区域 206a-2 向管理区域 203a-2 映射。用户区域 203a-1 向用户区域 202a-1 映射。管理区域 203a-2 向管理区域 202a-2 映射。

[0335] 对把位于存储系统 130a(移动源存储系统)内的逻辑单元 201a(移动源逻辑单元)中的数据移动到存储系统 130b(移动目的地存储系统)内的逻辑单元 201b(移动目的地逻辑单元)的处理进行概括描述。逻辑单元移动处理包含下述的(1)~(9)的处理过程。

[0336] (1) 存储系统 130b, 定义用于外部连接外部卷 301 和存储系统 130b 之间的路径 405, 在存储系统 130b 内生成虚拟化外部卷 301 的扩展设备 206b。扩展设备 206b 包含可从主计算机 100 访问的用户区域 206b-1、和用于存储管理信息的管理区域 206b-2。

[0337] (2) 存储系统 130b, 在存储系统 130b 内生成用于虚拟化外部卷 301 的虚拟设备 203b。虚拟设备 203b 包含可从主计算机 100 访问的用户区域 203b-1、和用于存储管理信息的管理区域 203b-2。

[0338] (3) 存储系统 130b, 分别定义用于外部连接逻辑单元 201a-1 和存储系统 130b 之间的路径 404-1、用于外部连接逻辑单元 201a-2 和存储系统 130b 之间的路径 404-2, 在存储系统 130b 内分别生成用于虚拟化逻辑单元 201a-1 的扩展设备 206c-1、用于虚拟化逻辑单元 201a-2 的扩展设备 206c-2。

[0339] (4) 存储系统 130b, 把扩展设备 206c-1 向虚拟设备 203b 内的用户区域 203b-1 映射, 把扩展设备 206c-2 向虚拟设备 203b 内的管理区域 203b-2 映射。

[0340] (5) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑设备 202a 的逻辑结构相同的逻辑结构的逻辑设备 202b。逻辑设备 202b 包含可从主计算机 100 访问的用户区域 202b-1、和用于存储管理信息的管理区域 202b-2。将用户区域 203b-1 向用户区域 202b-1 映射。将管理区域 203b-2 向管理区域 202b-2 映射。

[0341] (6) 存储系统 130b, 在存储系统 130b 内生成具有与逻辑单元 201a-1 的逻辑结构相同的逻辑结构的逻辑单元 201b, 定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。将用户区域 202b-1 向逻辑单元 201b 映射。

[0342] (7) 主计算机 100, 作为用于访问外部卷 301 的路径, 从路径 401 切换到路径 403。此时, 删除路径 401 等, 禁止从主计算机 100 向逻辑单元 201a-1 的访问。此外, 从路径 401 到路径 403 的路径切换, 也可以由存储系统 130a、130b、或者管理服务器 110 等进行。另外, 不一定需要删除路径 401, 例如也可以把路径 401 设定为路径 403 的替换路径(例如为在系统故障时使用的路径), 使路径 401 的状态成为无效不变, 保留路径 401。

[0343] 通过上述处理, 逻辑单元 201b 因为不仅是虚拟化逻辑单元 201a-1、而且也是虚拟化外部卷 301 内的用户区域 301-1, 所以在逻辑单元移动处理中的、从主计算机 100 向逻辑单元 201a 的数据输入输出路径成为: 路径 403 → 存储系统 130b 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 404-1 → 存储系统 130a 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 402 →

外部卷 301。

[0344] 此外,存储系统 130a,因为通过虚拟设备 203a-1 虚拟化外部卷 130 内的管理区域 301-1,所以可以通过路径 402 参照管理信息。在逻辑单元 201a 的移动处理中,为存储系统 130b 参照外部卷 130 内的管理信息,也可以通过路径 404-2 以及逻辑单元 201a-2 访问外部卷 130 内的管理区域 301-2。在外部卷 301 内的管理区域 301-2 中写入的管理信息的更新,通过移动源的存储系统 130a 进行。

[0345] (8) 存储系统 130a,为把在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据降级到外部卷 301,把存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。此外,存储系统 130a 可以自发地把动作模式设定为高速缓冲存储器直通模式,或者也可以根据来自外部(例如,存储系统 130b,主计算机 100,或者管理服务器 110 等)的指示,把动作模式设定为高速缓冲存储器直通模式。

[0346] (9) 如果在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷 301,则存储系统 130b 解除扩展设备 206c-1 和用户区域 203b-1 之间的映射关系,并且解除扩展设备 206c-2 和管理区域 203b-2 之间的映射关系。进而,存储系统 130b 把扩展设备 206b 内的用户区域 206b-1 向用户区域 203b-1 映射,并且把管理区域 206b-2 向管理区域 203b-2 映射。进而,存储系统 130b 删除路径 404。

[0347] 此外,图 39 中表示的方法,不仅在外卷 301 是附有 VMA 的卷时,而且是在大型机卷时,经过和上述的处理过程(1)~(9)同样的过程,也可以使逻辑单元 201a-1 移动。

[0348] 图 38 所示方法和图 39 所示方法之间的主要不同点如下。在图 39 所示的方法中,移动目的地的存储系统 130b,通过移动源的存储系统 130a 访问管理区域 301-2,但是在图 38 所示的方法中,移动目的地的存储系统 130b,直接访问管理区域 301-2。在图 39 所示的方法中,管理区域 301-2 内的管理信息的更新,由移动源的存储系统 130a 执行,但是在图 38 所示的方法中,管理区域 301-2 内的管理信息的更新,由移动目的地的存储系统 130b 执行。

[0349] 下面参照图 40 说明附有 VMA 的卷的移动处理的另一种方法。

[0350] 存储系统 130a 具有逻辑单元 201a、201a'、逻辑设备 202a、虚拟设备 203a、以及扩展设备 206a。逻辑单元 201a 通过路径 401 连接主计算机 100,可由主计算机 100 识别。相反地,逻辑单元 201a' 不连接主计算机 100,不能由主计算机 100 识别。仅在从存储系统 130a 向存储系统 130b 移动逻辑单元 201a 期间临时生成逻辑单元 201a'。逻辑单元 201a' 包含用户区域 201a-1'、管理区域 201a-2'。逻辑设备 202a 是附有 VMA 的卷,包含可从主计算机 100 访问的用户区域 202a-1、和用于存储管理信息的管理区域 202a-2。在管理信息中,例如包含附有 VMA 的卷的访问属性。用户区域 202a-1 通过被分配给逻辑单元 201a,可为主计算机 100 识别。用户区域 202a-1 也被分配给逻辑单元 201a' 内的用户区域 201a-1'。管理区域 202a-2 不被分配给逻辑单元 201a' 内的管理区域 201a-2'。虚拟设备 203a 包含可从主计算机 100 访问的用户区域 203a-1、和用于存储管理信息的管理区域 203a-2。扩展设备 206a 包含可从主计算机 100 访问的用户区域 206a-1、和用于存储管理信息的管理区域 206a-2。扩展设备 206a 内的用户区域 206a-1 是虚拟化外部卷 301 内的用户区域 301-1 的区域。扩展设备 206a 内的管理区域 206a-2 是虚拟化外部卷 301 内的管理区域 301-2 的区域。用户区域 206a-1 向用户区域 203a-1 映射。管理区域 206a-2 向管理区域 203a-2 映射。用户区域 203a-1 向用户区域 202a-1 映射。管理区域 203a-2 向管理区域 202a-2 映射。

[0351] 对把位于存储系统 130a(移动源存储系统)内的逻辑单元 201a(移动源逻辑单元)中的数据移动到存储系统 130b(移动目的地存储系统)内的逻辑单元 201b(移动目的地逻辑单元)的处理进行概括描述。逻辑单元移动处理包含下述的(1)~(9)的处理过程。

[0352] (1) 存储系统 130b,定义用于外部连接外部卷 301 和存储系统 130b 之间的路径 405,在存储系统 130b 内生成虚拟化外部卷 301 的扩展设备 206b。扩展设备 206b 包含可从主计算机 100 访问的用户区域 206b-1、和用于存储管理信息的管理区域 206b-2。

[0353] (2) 存储系统 130b,在存储系统 130b 内生成用于虚拟化外部卷 301 的虚拟设备 203b。虚拟设备 203b 包含可从主计算机 100 访问的用户区域 203b-1、和用于存储管理信息的管理区域 203b-2。

[0354] (3) 存储系统 130b,定义用于外部连接逻辑单元 201a' 和存储系统 130b 之间的路径 404,在存储系统 130b 内生成用于虚拟化逻辑单元 201a' 的扩展设备 206c。扩展设备 206c 包含可从主计算机 100 访问的用户区域 206c-1、和用于存储管理信息的管理区域 206c-2。

[0355] (4) 存储系统 130b,把扩展设备 206c 内的用户区域 206c-1 向虚拟设备 203b 内的用户区域 203b-1 映射,把扩展设备 206c 内的管理区域 206c-2 向虚拟设备 203b 内的管理区域 203b-2 映射。

[0356] (5) 存储系统 130b,在存储系统 130b 内生成具有与逻辑设备 202a 的逻辑结构相同的逻辑结构的逻辑设备 202b。逻辑设备 202b 包含可从主计算机 100 访问的用户区域 202b-1、和用于存储管理信息的管理区域 202b-2。将用户区域 203b-1 向用户区域 202b-1 映射。将管理区域 203b-2 向管理区域 202b-2 映射。

[0357] (6) 存储系统 130b,在存储系统 130b 内生成具有与逻辑单元 201a 的逻辑结构相同的逻辑结构的逻辑单元 201b,定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。将用户区域 202b-1 向逻辑单元 201b 映射。

[0358] (7) 主计算机 100,作为用于访问外部卷 301 的路径,从路径 401 切换到路径 403。此时,删除路径 401 等,禁止从主计算机 100 向逻辑单元 201a-1 的访问。此外,从路径 401 到路径 403 的路径切换,也可以由存储系统 130a、130b、或者管理服务器 110 等进行。另外,不一定需要删除路径 401,例如也可以把路径 401 设定为路径 403 的替换路径(例如用于在系统故障时使用的路径),使路径 401 的状态成为无效不变,保留路径 401。

[0359] 通过上述处理,逻辑单元 201b,因为不仅是虚拟化逻辑单元 201a、而且也是虚拟化外部卷 301 内的用户区域 301-1,所以在逻辑单元移动处理中的、从主计算机 100 向逻辑单元 201a 的数据输入输出路径成为:路径 403 → 存储系统 130b 内的存储层(LU/LDEV/VDEV/EDEV) → 路径 404 → 存储系统 130a 内的存储层(LU/LDEV/VDEV/EDEV) → 路径 402 → 外部卷 301。

[0360] 此外,存储系统 130a,因为通过虚拟设备 203a-1 虚拟化外部卷 130 内的管理区域 301-1,所以可以通过路径 402 参照管理信息。在逻辑单元 201a 的移动处理中,为了存储系统 130b 参照外部卷 130 内的管理信息,也可以通过路径 404 以及逻辑单元 201a' 访问外部卷 130 内的管理区域 301-2。通过移动源的存储系统 130a 进行在外部卷 301 内的管理区域 301-2 中写入的管理信息的更新。

[0361] (8) 存储系统 130a,为了把在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏

数据降级到外部卷 301,把存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。此外,存储系统 130a 可以自发地把动作模式设定为高速缓冲存储器直通模式,或者也可以根据来自外部(例如,存储系统 130b,主计算机 100,或者管理服务器 110 等)的指示,把动作模式设定为高速缓冲存储器直通模式。

[0362] (9) 如果在存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷 301,则存储系统 130b 解除扩展设备 206c 和虚拟设备 203b 之间的映射关系,并且将扩展设备 206b 向虚拟设备 203b 映射。更详细说,存储系统 130b,把扩展设备 206b 内的用户区域 206b-1 向虚拟设备 203b 内的用户区域 203b-1 映射,并且把扩展设备 206b 内的管理区域 206b-2 向虚拟设备 203b 内的管理区域 203b-2 映射。进而,存储系统 130b 删除路径 404。

[0363] 下面参照图 41,说明移动多个逻辑单元的处理的概要。

[0364] 存储系统 130a 具有:多个逻辑单元 201a-1、201a-2、多个逻辑设备 202a-1、202a-2、多个虚拟设备 203a-1、203a-2、以及多个扩展设备 206a-1、206a-2。扩展设备 206a-1 是虚拟化外部卷 301-1 的扩展设备,通过路径 402-1 连接外部存储系统 150。扩展设备 206a-1 向虚拟设备 203a-1 映射。虚拟设备 203a-1 向逻辑设备 202a-1 映射。逻辑设备 202a-1 向逻辑单元 201a-1 映射。逻辑单元 201a-1 是虚拟化外部卷 301-1 的逻辑单元,通过路径 401-1 连接主计算机 100。

[0365] 另一方面,扩展设备 206a-2 是虚拟化外部卷 301-2 的扩展设备,通过路径 402-2 连接外部存储系统 150。扩展设备 206a-2 向虚拟设备 203a-2 映射。虚拟设备 203a-2 向逻辑设备 202a-2 映射。逻辑设备 202a-2 向逻辑单元 201a-2 映射。逻辑单元 201a-2 是虚拟化外部卷 301-2 的逻辑单元,通过路径 401-2 连接主计算机 100。

[0366] 在上述结构中,为把多个逻辑单元 201a-1、201a-2 从存储系统 130a 向存储系统 130b 移动,和上述的处理过程同样,存储系统 130b,生成成为移动目的地的逻辑单元 201b-1、201b-2。逻辑单元 201b-1 通过路径 403-1 连接主计算机 100。逻辑单元 201b-2 通过路径 403-2 连接主计算机 100。进而,存储系统 130b 为了虚拟化多个逻辑单元 201a-1、201a-2 而分别生成多个扩展设备 206c-1、206c-2、多个虚拟设备 203b-1、203b-2、多个逻辑设备 202b-1、202b-2。扩展设备 206c-1,在逻辑单元 201a-1 的移动处理中,虚拟化逻辑单元 201a-1,另一方面,在逻辑单元 201a-1 的移动处理后(更详细说,在对于写入虚拟设备 203a-1 的脏数据的降级结束后),通过路径 405-1 连接外部卷 301-1,虚拟化外部卷 301-1。扩展设备 206c-2,在逻辑单元 201a-2 的移动处理中,虚拟化逻辑单元 201a-2,另一方面,在逻辑单元 201a-2 的移动处理后(更详细说,在写入虚拟设备 203a-2 的脏数据的降级结束后),通过路径 405-2 连接外部卷 301-2,虚拟化外部卷 301-2。

[0367] 此外,为了移动多个逻辑单元的指示,可以从管理服务器 110(参照图 2)给予各存储系统 130a、130b,或者也可以从主计算机 100 给予各存储系统 130a、130b。

[0368] 下面参照图 42 到图 46 对于逻辑单元移动处理的细节加以说明。

[0369] 图 42 是表示逻辑单元移动指示处理例行程序(routine)的说明图。通过管理服务器 110 内的逻辑单元移动指示处理例行程序 241 执行逻辑单元移动指示处理例行程序。

[0370] 管理服务器 110 对于存储系统 130a(移动源存储系统)、以及存储系统 130b(移动目的地存储系统)给予使移动一个以上的逻辑单元的指示(步骤 4201)。此时,可以通过

操作管理服务器 110 的存储器管理者对各存储系统 130a、130b 给予逻辑单元移动指示,或者也可以由管理服务器 110 内的管理软件对于各存储系统 130a、130b 给予逻辑单元移动指示。此外,逻辑单元的移动指示也可以从主计算机 100 对于各存储系统 130a、130b 给予。

[0371] 各存储系统 130a、130b,当接收逻辑单元移动指示时,执行为使指定的逻辑单元移动的处理(步骤 4202)。

[0372] 在指定的全部逻辑单元中有一部分逻辑单元的移动尚未结束时(步骤 4203:NO),各存储系统 130a、130b 执行逻辑单元移动处理(步骤 4202)。

[0373] 当指定的全部逻辑单元的移动结束时(步骤 4203:YES),各存储系统 130a、130b 向管理服务器 110 报告移动结束(步骤 4204)。此外,在从主计算机 100 对各存储系统 130a、130b 给予逻辑单元移动指示时,各存储系统 130a、130b 向主计算机 100 报告移动结束。

[0374] 图 43 表示逻辑单元移动处理子例行程序(sub-routine)。通过存储系统 130 内的逻辑单元移动处理程序 222 执行逻辑单元移动处理子例行程序。

[0375] 当逻辑单元移动处理的子例行程序起动时,各存储系统 130a、130b 执行移动体对象选择处理(步骤 4301)、移动准备处理(步骤 4302)、以及移动处理(步骤 4303)。

[0376] 图 44 表示移动对象选择处理子例行程序。当移动对象选择处理子例行程序起动时,存储系统 130a 选择与成为移动对象的逻辑单元对应的逻辑设备(成为移动对象的逻辑设备)(步骤 4401)。例如,在 LUSE 结构中,对于一个逻辑单元选择多个逻辑设备。

[0377] 接着,存储系统 130a,选择使用成为移动对象的逻辑设备的全部主计算机 100(步骤 4402)。例如,不仅属于同一主机组的主计算机 100,也选择属于其他主机组的主计算机 100。

[0378] 接着,存储系统 130a,选择与成为移动对象的逻辑设备对应的 VDEV 区域(成为移动对象的 VDEV 区域)(步骤 4403)。VDEV 区域,有时与一个虚拟设备的一部分对应,有时也与多个虚拟设备对应。例如,在 CVS 结构中,作为与成为移动对象的逻辑设备对应的 VDEV 区域,选择虚拟设备的一部分。另一方面,在 VDEV 连接结构中,作为与成为移动对象的逻辑设备对应的 VDEV 区域,选择多个虚拟设备。

[0379] 接着,存储系统 130a,选择与成为移动对象的 VDEV 区域对应的设备组(成为移动对象的设备组)(步骤 4404)。设备组有时与一个 VDEV 区域的一部分对应,有时也与多个 VDEV 区域对应。例如,在 VDEV 离散结构中,作为与成为移动对象的 VDEV 区域对应的设备组,选择多个设备组。

[0380] 接着,存储系统 130a,选择与成为移动对象的设备组对应的扩展设备(成为移动对象的扩展设备)(步骤 4405)。扩展设备有时与一个设备组的一部分对应,有时也与多个设备组对应。例如,把一个扩展设备的一部分的存储区域给某设备组分配、把其他存储区域给其他设备组分配时,作为与成为移动对象的设备组对应的扩展设备,选择扩展设备的一部分。另一方面,在外部卷 RAID 结构中,作为与成为移动对象的设备组对应的扩展设备,选择多个扩展设备。

[0381] 接着,存储系统 130a,选择与成为移动对象的扩展设备对应的外部卷(步骤 4406)。

[0382] 通过上述处理,能够分别选择与成为移动对象的逻辑单元对应的逻辑设备、与该逻辑设备对应的虚拟设备、与该虚拟设备对应的设备组、与该设备组对应的扩展设备、与该

扩展设备对应的外部卷。

[0383] 图 45 表示移动准备处理子例行程序。当移动准备处理子例行程序启动时,存储系统 130a,设定从存储系统 130b 向外部卷的访问许可(步骤 4501)。

[0384] 存储系统 130b,定义用于从存储系统 130b 连接外部卷的路径,在存储系统 130b 内生成用于虚拟化外部卷的第一扩展设备(与图 1 的扩展设备 206b 相当)(步骤 4502)。

[0385] 存储系统 130b,使存储系统 130b 内的第一扩展设备与存储系统 130b 内的设备组关联对应(步骤 4503)。例如,在外部卷 RAID 结构中,使多个第一扩展设备与一个设备组关联对应。

[0386] 存储系统 130b,使与存储系统 130b 内的第一扩展设备关联对应的设备组、与存储系统 130b 内的虚拟设备关联对应(步骤 4504)。

[0387] 存储系统 130a,设定从存储系统 130b 向存储系统 130a 内的移动源逻辑单元的访问许可(步骤 4505)。

[0388] 存储系统 130b,使存储系统 130b 内的第二扩展设备(与图 1 的扩展设备 206c 相当)与存储系统 130a 内成为移动对象的逻辑单元关联对应(步骤 4506)。

[0389] 例如,在成为移动对象的逻辑单元具有外部卷 RAID 结构时,使存储系统 130b 内的第二扩展设备与存储系统 130a 内的虚拟设备的一部分(除存储奇偶校验数据以及镜像数据的存储区域外。)关联对应。

[0390] 还例如,在成为移动对象的逻辑单元具有 CVS 结构时,使存储系统 130b 内的第二扩展设备与存储系统 130a 内的虚拟设备的一部分关联对应。

[0391] 还例如,在成为移动对象的逻辑单元具有 LUSE 结构时,使存储系统 130b 内的第二扩展设备与存储系统 130a 内的多个虚拟设备关联对应。

[0392] 还例如,在成为移动对象的逻辑单元是附有 VMA 的卷时,使存储系统 130b 内的第二扩展设备与存储系统 130a 内的虚拟设备的一部分(用户区域)关联对应。

[0393] 存储系统 130b,使存储系统 130b 内的第二扩展设备与存储系统 130b 内的设备组关联对应(步骤 4507)。

[0394] 存储系统 130b,使与存储系统 130b 内的第二扩展设备关联对应的设备组与存储系统 130b 内的虚拟设备关联对应(步骤 4508)。

[0395] 例如,在成为移动对象的逻辑单元具有外部卷 RAID 结构时,使存储系统 130b 内的第二扩展设备与存储系统 130b 内的虚拟设备的一部分(除存储奇偶校验数据以及镜像数据的存储区域外。)关联对应。

[0396] 还例如,在成为移动对象的逻辑单元具有 CVS 结构时,使存储系统 130b 内的第二扩展设备与存储系统 130b 内的虚拟设备的一部分关联对应。

[0397] 还例如,在成为移动对象的逻辑单元具有 LUSE 结构时,使存储系统 130b 内的第二扩展设备与存储系统 130b 内的多个虚拟设备关联对应。

[0398] 还例如,在成为移动对象的逻辑单元是附有 VMA 的卷时,使存储系统 130b 内的第二扩展设备与存储系统 130b 内的虚拟设备的一部分(用户区域)关联对应。

[0399] 存储系统 130b,使与第一扩展设备关联对应的存储系统 130b 内的设备组和存储系统 130b 内的虚拟设备之间的映射关系无效,使与第二扩展设备关联对应的存储系统 130b 内的设备组和存储系统 130b 内的虚拟设备之间的映射关系有效(步骤 4509)。

[0400] 存储系统 130b,把存储系统 130b 的动作模式设定为高速缓冲存储器直通模式,把存储系统 130a 的动作模式设定为后写模式(步骤 4510)。

[0401] 存储系统 130b,在存储系统 130b 内生成移动目的地的逻辑单元(步骤 4511)。

[0402] 存储系统 130b,给移动目的地的逻辑单元设定访问属性(步骤 4512)。作为访问属性,例如有关于主机组的设定或者关于替换路径的设定等。

[0403] 存储系统 130b,解除加密从存储系统 130b 写入外部卷的数据的功能(步骤 4513)。在逻辑单元的移动处理中,写入外部卷的数据,通过移动源的存储系统 130a 加密。

[0404] 存储系统 130b,定义用于连接移动目的地逻辑单元和主计算机 100 的路径(步骤 4514)。在存在多个移动目的地逻辑单元时,定义用于连接各个移动目的地的逻辑单元和主计算机 100 的路径。

[0405] 存储系统 130b,使主计算机 100 识别移动目的地的逻辑单元(步骤 4515)。在存在多个使用成为移动对象的逻辑单元的主计算机 100 时,使各个主计算机 100 识别移动目的地的逻辑单元。

[0406] 存储系统 130b,把连接移动目的地的逻辑单元和主计算机 100 的路径作为连接移动源的逻辑单元和主计算机 100 的路径的替换路径设定(步骤 4516)。

[0407] 图 46 表示移动处理子例行程序。当起动移动处理子例行程序时,主计算机 100 停止向移动源的逻辑单元的 I/O 请求(步骤 4601)。也可以由存储系统 130a 禁止来自主计算机 100 的 I/O 请求。在存在多个使用移动源的逻辑单元的主计算机 100 时,停止或者禁止从各个主计算机 100 向移动源的逻辑单元的 I/O 请求。

[0408] 主计算机 100,从连接移动源的逻辑单元的路径切换到连接移动目的地的逻辑单元的路径,并且删除连接移动源的逻辑单元的路径(步骤 4602)。在存在多个使用移动源的逻辑单元的主计算机 100 时,各主计算机 100 同时执行路径切换。

[0409] 移动源的存储系统 130a,禁止来自主计算机 100 的访问(步骤 4603)。

[0410] 移动目的地的存储系统 130b,作为存储系统 130b 的动作模式,解除高速缓冲存储器直通模式,设定为后写模式(步骤 4604)。

[0411] 移动源的存储系统 130a,作为存储系统 130a 的动作模式,解除后写模式,设定为高速缓冲存储器直通模式(步骤 4605)。

[0412] 如果把在移动源的存储系统 130a 的盘高速缓冲存储器中积蓄的全部脏数据已经降级到外部卷中(步骤 4606:YES),则移动目的地的存储系统 130b,使与第二扩展设备关联对应的存储系统 130b 内的设备组和存储系统 130b 内的虚拟设备之间的映射关系无效、使与第一扩展设备关联对应的存储系统 130b 内的设备组和存储系统 130b 内的虚拟设备之间的映射关系有效(步骤 4607)。

[0413] 存储系统 130b,进行许可从存储系统 130b 对于外部卷的访问、另一方面禁止从存储系统 130a 对于外部卷的访问的设定(排他设定访问)(步骤 4608)。

[0414] 存储系统 130b,把加密写入外部卷的数据的功能设定为活动(步骤 4609)。

[0415] 存储系统 130b,删除连接存储系统 130b 内的第二扩展设备和存储系统 130a 内的移动源的逻辑单元的路径(步骤 4610)。

[0416] 在由移动源的存储系统 130a、和移动目的地的存储系统 130b 共有外部卷时(例如移动源的逻辑单元具有 CVS 结构时)(步骤 4611:YES),移动源的存储系统 130a,使移动处

理后的逻辑单元和存储系统 130a 内的虚拟设备之间的映射关系成为无效（步骤 4612）。

[0417] 在未由移动源的存储系统 130a、和移动目的地的存储系统 130b 共有外部卷的情况下（步骤 4611 :NO），存储系统 130a，删除连接存储系统 130a 和外部卷的路径（步骤 4613）。

[0418] 移动源的存储系统 130a，将为虚拟化外部卷而在存储系统 130a 内生成的扩展设备、与该扩展设备关联对应的设备组、与该设备组关联对应的虚拟设备、与该虚拟设备关联对应的逻辑设备、与该逻辑设备关联对应的逻辑单元、用于连接该逻辑单元和主计算机 100 的路径全部删除（步骤 4614）。

[0419] 外部存储系统 150 禁止从存储系统 130a 对于外部卷的访问（步骤 4615）。

[0420] 图 47 表示存储器 I/O 处理例行程序。存储器 I/O 处理例行程序，通过存储器 I/O 处理程序 211 执行。

[0421] 接收来自主计算机 100 的 I/O 请求的存储系统 130，参照逻辑设备管理信息 603，确认主计算机 100 的访问权（步骤 4701）。

[0422] 在主计算机 100 没有访问权的情况下（步骤 4701 :NO），存储系统 130 拒绝 I/O 请求（步骤 4708）。

[0423] 在主计算机 100 有访问权的情况下（步骤 4701 :YES），存储系统 130，参照逻辑设备管理信息 603，判断是否设定为高速缓冲存储器直通模式（步骤 4702）。

[0424] 在存储系统 130 的动作模式是高速缓冲存储器直通模式（步骤 4702 :YES）、而且来自主计算机 100 的 I/O 请求是写请求时（步骤 4703 :YES），存储系统 130 在盘高速缓冲存储器 134 中写入从主计算机 100 接收的数据（步骤 4704）。

[0425] 其后，存储系统 130，把数据写入外部卷（步骤 4705），把高速缓冲存储器数据的管理信息更新为降级完毕（步骤 4706），向主计算机 100 报告写结束（步骤 4707）。

[0426] 在存储系统 130 的动作模式不是高速缓冲存储器直通模式（步骤 4702 :NO）、存储系统 130 判断访问目的地的逻辑单元是否在移动中（步骤 4710）。

[0427] 在访问目的地的逻辑单元不在移动中时（步骤 4710 :NO），存储系统 130，对于移动源的逻辑单元执行 I/O 处理（步骤 4709），向主计算机 100 报告写结束（步骤 4707）。

[0428] 另一方面，在访问目的地的逻辑单元在移动中时（步骤 4710 :YES），存储系统 130，判断与访问地址对应的存储区域的移动是否结束（步骤 4711）。

[0429] 在与访问地址对应的存储区域的移动未结束时（步骤 4711 :NO），存储系统 130，对于移动源的逻辑单元执行 I/O 处理（步骤 4709），向主计算机 100 报告写结束（步骤 4707）。

[0430] 在与访问地址对应的存储区域的移动已结束时（步骤 4711 :YES），存储系统 130，对于移动目的地的逻辑单元执行 I/O 处理（步骤 4712），向主计算机 100 报告写结束（步骤 4707）。

[0431] 图 48 表示主机 I/O 处理例行程序。主机 I/O 处理例行程序，由主计算机 100 内的主机 I/O 处理程序 261 执行。

[0432] 主计算机 100，接收从在主计算机 100 上执行的应用程序对于与存储系统 130 的逻辑单元对应的设备组的 I/O 请求（步骤 4801）。

[0433] 主计算机 100，参照设备路径管理信息 251，判定在访问目的地的设备文件中是否设定有多条路径（多路径）（步骤 4802）。

[0434] 在未设定多条路径时(步骤 4802:NO),主计算机 100,把指定设备文件作为访问对象,前进到步骤 4804。

[0435] 在设定多条路径时(步骤 4802:YES),主计算机 100,根据各路径的状态和分配的优先顺序,决定成为访问对象设备文件(步骤 4803)。

[0436] 接着,主计算机 100,变换作为访问对象决定的设备文件,计算存储系统 130 以及端口 131 的识别符和 LUN(步骤 4804)。

[0437] 主计算机 100 向访问目的地的存储系统 130 发送 I/O 请求(步骤 4805)。

[0438] 当 I/O 处理正常结束时(步骤 4806:YES),主计算机 100 对于应用程序报告 I/O 处理结束(步骤 4809)。

[0439] 在 I/O 处理未正常结束时(步骤 4806:NO),主计算机 100 检查有无替换路径(步骤 4807)。

[0440] 如果存在替换路径(步骤 4807:YES),则主计算机 100 经由替换路径再次向存储系统 130 发行 I/O 处理请求(步骤 4808),对于应用程序报告 I/O 处理结束(步骤 4809)。

[0441] 如果不存在替换路径(步骤 4807:NO),则主计算机 100 对于应用程序报告 I/O 处理结束(步骤 4809)。

[0442] 图 49 表示存储系统负荷监视处理例行程序。执行存储系统负荷监视处理例行程序的目的是,检查各存储系统 130 的负荷,把处于高负荷状态或者性能降低状态的存储系统 130 内的逻辑单元移动到其他的存储系统 130。存储系统负荷监视处理例行程序通过管理服务器 110 内的存储系统负荷监视程序 242 执行。

[0443] 管理服务器 110,经由 IP 网络 175,周期地取得存储系统 130 的结构信息(步骤 4901),检查在各存储系统 130 内是否发生结构异常(步骤 4902)。

[0444] 如果由于盘高速缓冲存储器 134 或者控制处理器 132 等的故障,某存储系统 130 的 I/O 处理能力降低(步骤 4902),则管理服务器 110 把出现性能降低的存储系统 130 决定为移动源存储系统 130a(步骤 4908)。

[0445] 在不能检测到出现性能降低的存储系统 130 时(步骤 4902:NO),管理服务器 110,周期地取得从各存储系统 130 对于各逻辑设备的 I/O 频度、各控制处理器 132 或者盘高速缓冲存储器 134 的利用率等的运行信息(步骤 4903)。

[0446] 在各存储系统 130 之间的 I/O 负荷的差在规定水平以上时(步骤 4904:YES),管理服务器 110 把处于高负荷状态下的存储系统 130 决定为移动源存储系统 130a(步骤 4905)。

[0447] 管理服务器 110,对于通过步骤 4905 或者步骤 4908 决定的移动源存储系统 130a 选定成为移动源的逻辑单元,进而选定移动目的地的存储器 130b 同时选定移动目的地的逻辑单元(步骤 4906),向存储系统 130a、130b 指示移动逻辑单元(步骤 4907)。

[0448] 构成计算机系统 500 的各存储系统 130 的负荷通过存储系统负荷监视处理例行程序监视,在由于盘高速缓冲存储器 134 或者控制处理器 132 的故障等某存储系统 130 中发生故障时,或者由于某种原因存储系统 130 的 I/O 处理性能降低时,通过把外部卷的控制责任切换到处于正常状态或者低负荷状态的存储系统 130,能够最大限度发挥计算机 500 整体的 I/O 处理性能。

[0449] 下面参照图 50 到 54 关于逻辑单元移动处理的其他方法加以说明。

[0450] 图 50 表示计算机系统 501 的系统结构。计算机系统 501,具有主计算机 100、存储

系统 130a、130b、以及外部存储系统 150。与图 1 所示符号相同符号的系统,表示相同的系统,省略其详细的说明。

[0451] 存储系统 130a 的动作模式,设定为高速缓冲存储器直通模式,在外部卷 301 中存储的数据始终处于最新的状态。主计算机 100,仅通过把访问路径从路径 401 切换为路径 403,不进行设备彼此的映射切换等,可以使逻辑单元 201a 移动到逻辑单元 201b。

[0452] 此外,在把存储系统 130a 的动作模式设定为后写模式时,临时停止从主计算机 100 对于存储系统 130a 的 I/O 请求,把在存储系统 130a 内的盘高速缓冲存储器上积蓄的脏数据降级到外部卷 301 上。其后,主计算机 100,仅通过把访问路径从路径 401 切换为路径 403,不进行设备彼此的映射切换等,可以使逻辑单元 201a 移动到逻辑单元 201b。

[0453] 图 51 表示计算机系统 502 的系统结构。计算机系统 502 具有主计算机 100、以及存储系统 130a、130b。存储系统 130a 具有与具有盘驱动器等实存储区域的存储资源关联对应的物理设备 205a,通过各个存储层 (PDEV/VEDV/LDEV/LU) 把物理设备 205a 虚拟化为逻辑单元 201a。另一方面,存储系统 130b 具有与具有盘驱动器等实存储区域的存储资源关联对应的物理设备 205b,通过各个存储层 (PDEV/VEDV/LDEV/LU) 把物理设备 205b 虚拟化为逻辑单元 201b。逻辑单元 201b 被预定为逻辑单元 201a 的移动目的地。

[0454] 在计算机系统 502 中,设定为在成为移动源的逻辑单元 201a 和成为移动目的地的逻辑单元 201b 之间进行同步复制的复制对 (copy pair)。通过这样设定复制对,从主计算机 100 向逻辑单元 201a (实际上高速缓冲存储器) 写入的数据,经由存储网络上的路径 (未图示),在写入逻辑单元 201b (实际上高速缓冲存储器) 后,向主计算机 100 报告写结束。

[0455] 当在存储系统 130a 中发生故障时,主计算机 100,仅通过把访问路径从路径 401 切换到路径 403,不进行设备彼此的映射变更等,能够使逻辑单元 201a 移动到逻辑单元 201b。

[0456] 图 52 表示计算机系统 503 的系统结构。计算机系统 503 具有主计算机 100、以及存储系统 130a、130b。和图 51 所示符号相同符号的系统等表示相同的系统等,省略其详细的说明。

[0457] 概述把位于存储系统 130a (移动源存储系统) 内的逻辑单元 201a (移动源逻辑单元) 中的数据移动到存储系统 130b (移动目的地存储系统) 内的逻辑单元 201b (移动目的地逻辑单元) 的处理。逻辑单元移动处理包含下述 (1) ~ (9) 的处理过程。

[0458] (1) 存储系统 130b,定义用于外部连接移动源的逻辑单元 201a 和移动目的地的存储系统 130b 之间的路径 404,在存储系统 130b 内生成用于虚拟化逻辑单元 201a 的扩展设备 206b。

[0459] (2) 存储系统 130b,在存储系统 130b 内生成具有与虚拟设备 203a 相同逻辑结构的虚拟设备 203b-1、203b-2。

[0460] (3) 存储系统 130b,在存储系统 130b 内生成具有与逻辑设备 202a 的逻辑结构相同的逻辑结构的逻辑设备 202b-1、202b-2。

[0461] (4) 存储系统 130b,把扩展设备 206b 向虚拟设备 203b-1 映射,并且把物理设备 205b 向虚拟设备 203b-2 映射。

[0462] (5) 存储系统 130b,把虚拟设备 203b-1 向逻辑设备 202b-1 映射,并且把虚拟设备 203b-2 向逻辑设备 202b-2 映射。

[0463] (6) 存储系统 130b,在存储系统 130b 内生成具有与逻辑单元 201a 的逻辑结构相

同的逻辑结构的逻辑单元 201b, 定义用于连接主计算机 100 和逻辑单元 201b 的路径 403。使逻辑设备 202b-1 向逻辑单元 201b 映射。

[0464] (7) 主计算机 100, 把访问路径从路径 401 切换到路径 403。此时, 删除路径 401 等, 禁止从主计算机 100 向逻辑单元 201a 的访问。此外, 从路径 401 到路径 403 的路径切换, 也可以由存储系统 130a、130b、或者管理服务器 110 等进行。另外, 不一定需要删除路径 401, 例如也可以把路径 401 设定为路径 403 的替换路径 (例如为在系统故障时使用的路径), 使路径 401 的状态成为无效不变, 保留路径 401。

[0465] (8) 存储系统 130b, 向逻辑设备 202b-2 复制逻辑设备 202b-1。通过这样的处理, 能够使物理设备 205a 和物理设备 205b 同步。

[0466] 此外, 在逻辑单元移动处理中的、从主计算机 100 向存储系统 130b 的数据输入输出路径成为: 路径 403 → 存储系统 130b 内的存储层 (LU/LDEV/VDEV/EDEV) → 路径 404 → 存储系统 130a 内的存储层 (LU/LDEV/VDEV/EDEV)。

[0467] (9) 存储系统 130b, 如果能够使物理设备 205a 和物理设备 205b 同步化, 则解除虚拟设备 203b-2 和逻辑设备 202b-2 之间的映射关系, 把虚拟设备 203b-2 向逻辑设备 202b-1 映射。并且, 存储系统 130b, 解除虚拟设备 203b-1 和逻辑设备 202b-1 之间的映射关系, 把虚拟设备 203b-1 向逻辑设备 202b-2 映射。进而, 存储系统 130b 删除路径 404。

[0468] 图 53 表示计算机系统 504 的系统结构。计算机系统 504 具有主计算机 100、存储系统 130a、130b、以及外部存储系统 150。和图 52 所示符号相同符号的系统等表示相同的系统等, 省略其详细的说明。

[0469] 计算机系统 504, 和上述计算机系统 503 的主要不同点如下。即, 存储系统 130a 具有扩展设备 206a 代替物理设备 205a。扩展设备 206a 是用于虚拟化外部卷 301 的虚拟的存储区域。计算机系统 504 执行的逻辑单元移动处理的过程, 和计算机系统 503 执行的逻辑单元移动处理的过程相同。

[0470] 图 54 表示计算机系统 505 的系统结构。计算机系统 505 具有主计算机 100、存储系统 130a、130b、以及外部存储系统 150。和图 52 所示符号相同符号的系统等表示相同的系统等, 省略其详细的说明。

[0471] 计算机系统 505, 和上述计算机系统 503 的主要不同点如下。即, 存储系统 130b 具有扩展设备 206c 代替物理设备 205b。扩展设备 206c 是用于虚拟化外部卷 301 的虚拟的存储区域。计算机系统 505 执行的逻辑单元移动处理的过程, 和计算机系统 503 执行的逻辑单元移动处理的过程相同。

[0472] 下面, 参照图 55 到图 56, 关于具有把逻辑单元的一部分从移动源存储系统移动到移动目的地存储系统的功能的计算机系统 506、507 加以说明。

[0473] 图 55 表示计算机系统 506 的系统结构。计算机系统 506 具有主计算机 100、存储系统 130a、130b、以及外部存储系统 150。和图 1 所示符号相同符号的系统表示相同的系统, 省略其详细的说明。

[0474] 外部卷 301 分割为多个存储区域 301-1、301-2。各个存储区域 301-1、301-2 是分割外部卷 301 的分区。在各个存储区域 301-1、301-2 中可以构建独立的文件系统。例如, 存储区域 301-1 与 C 驱动器对应, 存储区域 301-2 与 D 驱动器对应。

[0475] 存储系统 130a 通过路径 401 向主计算机 100 提供的逻辑单元 201a 虚拟化外部卷

301,分割为多个存储区域 201a-1、201a-2。存储区域 201a-1 虚拟化存储区域 301-1。存储区域 201a-2 虚拟化存储区域 301-2。

[0476] 为把逻辑单元 201a 的一部分(例如与 D 驱动器对应的存储区域 201a-2)从存储系统 130a 移动到存储系统 130b,经过和上述处理过程(1)~(9)同样的过程,使逻辑单元 201a 从存储系统 130a 移动到存储系统 130b。在存储系统 130b 内生成的逻辑单元 201b 的逻辑结构,和逻辑单元 201a 的逻辑结构相同。亦即,存储系统 130b 通过路径 403 向主计算机 100 提供的逻辑单元 201b 是虚拟化外部卷 301 的逻辑单元,并被分割成多个存储区域 201b-1、201b-2。存储区域 201b-1 是虚拟化存储区域 301-1 的存储区域。存储区域 201b-2 是虚拟化存储区域 301-2 的存储区域。

[0477] 这里,如果禁止从主计算机对于存储区域 201a-2 的访问并且禁止从主计算机对于存储区域 201b-1 的访问那样设定计算机系统 506 的系统结构,则能够从存储系统 130a 到存储系统 130b 移动逻辑单元 201a 的一部分。

[0478] 此外,需要在禁止从存储系统 130a 对于存储区域 301-2 的访问,并且禁止从存储系统 130b 对于存储区域 301-1 的访问。

[0479] 主计算机 100 内的链路管理器 108,对应访问目的地切换路径。例如,在主计算机 100 访问 C 驱动器时,选择路径 401,在主计算机 100 访问 D 驱动器时,选择路径 403。

[0480] 图 56 表示计算机系统 507 的系统结构。计算机系统 507 具有主计算机 100、存储系统 130a、130b、以及外部存储系统 150。和图 55 所示符号相同符号的系统表示相同的系统,省略其详细的说明。

[0481] 计算机系统 507 的基本的系统结构,因为和计算机系统 506 的系统结构大体相同,所以以两者的不同点为中心说明。外部卷 301 分割为多个存储区域 301-1、301-2、301-3。存储区域 301-3,是通过各个逻辑单元 201a、201b 被虚拟化的共用区域。逻辑单元 201a 的存储区域 201a-3 是虚拟化外部卷 301 的存储区域 301-3 的存储区域。并且,逻辑单元 201b 的存储区域 201b-3 是虚拟化外部卷 301 的存储区域 301-3 的存储区域。

[0482] 主计算机 100,作为对于外部卷 301 的共用区域的存储区域 301-3 的访问路径,具有通过逻辑单元 201a 访问的路径和通过逻辑单元 201b 的访问的路径这两条路径。在存储区域 301-3 中,例如,在主计算机 100 执行业务外,存储重要的数据。主计算机 100,因为具有多条用于访问存储区域 301-3 的路径,所以即使在某一条路径上发生故障,也可以经由其他路径进行访问。

[0483] 下面,作为用于通过多个逻辑单元 201a、201b 虚拟化作为外部卷 301 的共用区域的存储区域 301-3 的方式,说明三种方式。

[0484] (1) 第一方式

[0485] 把各存储系统 130a、130b 的动作模式设定为高速缓冲存储器直通模式。当主计算机 100 通过逻辑单元 201a 或者逻辑单元 201b 在作为外部卷 301 的共用区域的存储区域 301-3 中写入数据时,在存储区域 301-3 上的数据被更新后向主计算机 100 报告写结束。在存储区域 301-3 上存储的数据,因为始终处于最新的状态,所以主计算机 100 通过逻辑单元 201a 访问外部卷 301 的共用区域,或者主计算机 100 通过逻辑单元 201b 访问外部卷 301 的共用区域,在共用区域上的数据中都不会产生矛盾。

[0486] (2) 第二方式

[0487] 在逻辑单元 201a 的存储区域 201a-3、和逻辑单元 201b 的存储区域 201b-3 之间在双方向上设定用于进行同步复制的复制对。通过这样的设定,为了主计算机 100 更新作为外部卷 301 的共用区域的存储区域 301-3 的数据,当更新逻辑单元 201a 的存储区域 201a-3 上的数据时,逻辑单元 201b 的存储区域 201b-3 上的数据也同样被更新。另外,当主计算机 100 更新逻辑单元 201b 的存储区域 201b-3 上的数据时,逻辑单元 201a 的存储区域 201a-3 上的数据也同样被更新。

[0488] (3) 第三方式

[0489] 在用于虚拟化外部卷 301 的扩展设备 206a 的存储区域中使与共用区域对应的存储区域 206a-3 与逻辑单元 201a 的存储区域 201a-3 关联对应,并且也与逻辑单元 201b 的存储区域 201b-3 关联对应。同样,在用于虚拟化外部卷 301 的扩展设备 206b 的存储区域中使与共用区域对应的存储区域 206b-3 与逻辑单元 201b 的存储区域 201b-3 关联对应,并且也与逻辑单元 201a 的存储区域 201a-3 关联对应。这样,通过在双方向上应用外部连接技术,当主计算机 100 更新逻辑单元 201a 的存储区域 201a-3 上的数据时,逻辑单元 201b 的存储区域 201b-3 上的数据也同样被更新。另外,当主计算机 100 更新逻辑单元 201b 的存储区域 201b-3 上的数据时,逻辑单元 201a 的存储区域 201a-3 上的数据也同样被更新。

[0490] 此外,在第三方式中,为了避免无限环(loop)处理,需要区别主机访问和经由外部连接的访问。

[0491] 这样,在通过多个逻辑单元虚拟化外部卷 301 时,可以使用如下结构:对于构成外部卷 301 的某存储区域,可通过任何一个逻辑单元进行访问;对于其他的存储区域(共用区域),可通过多个逻辑单元进行访问。对于在多个逻辑单元的各个存储区域中与共用区域对应的存储区域,可以使用同步复制,或者也可以使用非同步复制。

[0492] 下面参照图 57 到图 58,从高可用性的观点对在上述的计算机系统 500 的系统结构上增加若干变更的计算机系统 508、509 加以说明。

[0493] 图 57 表示计算机系统 508 的系统结构。计算机系统 508 具有主计算机 100、存储系统 130a、130b、以及外部存储系统 150。主计算机 100、存储系统 130a、130b、以及外部存储系统 150 通过 FC-SAN 等存储网络(未图示)相互连接。存储系统 130a,作为常用类型系统(运行类型系统)发挥作用,存储系统 130b,作为替换类型系统(待机类型系统)发挥作用。存储系统 130a、130b 共有外部存储系统 150 内的外部卷 301。

[0494] 存储系统 130a 具有用于虚拟化外部卷 301 的多个存储层(逻辑单元 201a、逻辑设备 202a、虚拟设备 203a、以及扩展设备 206a)。扩展设备 206a 是虚拟化外部卷 301 而成的扩展设备,并通过路径 402 连接外部存储系统 150。扩展设备 206a 向虚拟设备 203a 映射。虚拟设备 203a 向逻辑设备 202a 映射。逻辑设备 202a 向逻辑单元 201a 映射。逻辑单元 201a,通过路径 401 连接主计算机 100。存储系统 130a 的动作模式设定为高速缓冲存储器直通模式。在外部卷 301 中始终存储伴随主机访问的最新的最新数据。

[0495] 另一方面,存储系统 130b 具有用于虚拟化外部卷 301 的多个存储层(逻辑单元 201b、逻辑设备 202b、虚拟设备 203b、以及扩展设备 206b)。扩展设备 206b 是虚拟化外部卷 301 而成的扩展设备,并通过路径 405 连接外部存储系统 150。扩展设备 206b 向虚拟设备 203b 映射。虚拟设备 203b 向逻辑设备 202b 映射。逻辑设备 202b 向逻辑单元 201b 映射。逻辑单元 201b,通过路径 403 连接主计算机 100。路径 403 相对于路径 401 有替换路径的

关系。存储系统 130b 的动作模式可设定为高速缓冲存储器直通模式,也可以设定为后写模式。

[0496] 在上述的系统结构中,当在常用类型的存储系统 130a 内发生故障时,主计算机 100 从路径 401 切换到路径 403,向替换类型的存储系统 130b 发行 I/O 请求。如上述,存储系统 130a 的动作模式,因为被设定为高速缓冲存储器直通模式,所以在存储系统 130a 内的盘高速缓冲存储器中不存在未降级到外部卷 301 的数据。因此,主计算机 100,在检测到在常用类型的存储系统 130a 内发生故障时,能够立即把 I/O 请求的发行目的地切换到替换类型的存储系统 130b。

[0497] 图 58 表示计算机系统 509 的系统结构。计算机系统 509 具有主计算机 100、存储系统 130a、130b、以及外部存储系统 150。和图 57 所示符号相同符号的系统表示相同的系统,省略其详细的说明。

[0498] 存储系统 130a 具有用于虚拟化外部卷 301 的多个存储层(逻辑单元 201a、逻辑设备 202a、虚拟设备 203a、以及扩展设备 206a、206d)。扩展设备 206a 是虚拟化外部卷 301 而成的扩展设备,并通过路径 402 连接外部存储系统 150。扩展设备 206d 是虚拟化逻辑单元 201b 而形成的扩展设备,并通过路径 410 连接存储系统 130b。各扩展设备 206a、206d 向虚拟设备 203a 映射。虚拟设备 203a 向逻辑设备 202a 映射。逻辑设备 202a 向逻辑单元 201a 映射。逻辑单元 201a 通过路径 401 连接主计算机 100。

[0499] 另一方面,存储系统 130b 具有用于虚拟化外部卷 301 的多个存储层(逻辑单元 201b、逻辑设备 202b、虚拟设备 203b、以及扩展设备 206b)。扩展设备 206b 是虚拟化外部卷 301 而成的扩展设备,并通过路径 405 连接外部存储系统 150。扩展设备 206b 向虚拟设备 203b 映射。虚拟设备 203b 向逻辑设备 202b 映射。逻辑设备 202b 向逻辑单元 201b 映射。逻辑单元 201b,通过路径 403 连接主计算机 100。路径 403 相对于路径 401 有替换路径的关系。

[0500] 下面,作为为提高计算机系统 509 的可用性的方式,说明三种方式。

[0501] (1) 第一方式

[0502] 把常用类型的存储系统 130a 以及替换类型的存储系统 130b 各自的动作模式设定为高速缓冲存储器直通模式。于是,从主计算机 100 到逻辑单元 201a 的数据输入输出路径成为:路径 401 → 存储系统 130a 内的存储层(LU/LDEV/VDEV/EDEV) → 路径 410 → 存储系统 130b 内的存储层(LU/LDEV/VDEV/EDEV) → 路径 405 → 外部卷 301。但是,作为用于从主计算机 100 向逻辑单元 201a 读访问的路径,例如也可以使用如下路径:路径 401 → 存储系统 130a 内的存储层(LU/LDEV/VDEV/EDEV) → 路径 402 → 外部卷 301 的路径。

[0503] 根据第一方式,因为在外卷 301 中始终存储伴随主机访问的最新的最新数据,所以即使在常用类型的存储系统 130a 中发生故障,主计算机 100,通过从路径 401 切换到路径 403,能够立即使用替换类型的存储系统 130b。

[0504] (2) 第二方式

[0505] 把位于常用类型的存储系统 130a 内的逻辑单元 201a 和位于替换类型的存储系统 130b 内的逻辑单元 201b 之间的数据输入输出动作设定为高速缓冲存储器直通模式。亦即从主计算机 100 向逻辑单元 201a(实际上盘高速缓冲存储器)写入的数据,从存储系统 130a 内的存储层(LU/LDEV/VDEV/EDEV)经由路径 410 写入逻辑单元 201b(实际上盘高速缓

冲存储器)后,向主计算机 100 报告写结束。另一方面,把位于常用类型的存储系统 130a 内的逻辑单元 201a 和外部卷 301 之间的数据输入输出动作设定为后写模式。亦即,从主计算机 100 向逻辑单元 201a(实际上盘高速缓冲存储器)写入的数据,在向主计算机 100 报告写结束后,被降级到外部卷 301。

[0506] 在常用类型的存储系统 130a 正常运行期间,仅存储系统 130a 具有对外部卷 301 的访问权,替换类型的存储系统 130b 没有对外部卷 301 的访问权。当存储系统 130b 从存储系统 130a 接收存储系统 130a 已向外部卷 301 写入数据的通知时,存储系统 130b 删除由存储系统 130a 写入到逻辑单元 201b(实际上盘高速缓冲存储器)的数据。

[0507] 根据第二方式,不仅常用类型的存储系统 130a,而且替换类型的存储系统 130b,也可以保存未降级到外部卷 301 中的脏数据。因此,在常用类型的存储系统 130a 中发生故障时,替换类型的存储系统 130b 取得对于外部卷 301 的访问权,通过把脏数据降级到外部卷 301,可以把在外部卷 301 中存储的数据更新为最新的状态。如果能够把在外部卷 301 中存储的数据更新为最新的状态,则主计算机 100,通过从路径 401 切换到路径 403,能够对替换类型的存储系统 130b 发行 I/O 请求。

[0508] (3) 第三方式

[0509] 在位于常用类型的存储系统 130a 内的逻辑单元 201a 和位于替换类型的存储系统 130b 内的逻辑单元 201b 之间形成用于同步复制复制对。通过这样设定复制对,从主计算机 100 写入到逻辑单元 201a(实际上盘高速缓冲存储器)的数据,在经由存储网络上的路径(未图示)写入逻辑单元 201b(实际上盘高速缓冲存储器)后,向主计算机 100 报告写结束。另一方面,把位于常用类型的存储系统 130a 内的逻辑单元 201a 和外部卷 301 之间的数据输入输出动作设定为后写模式。亦即从主计算机 100 写入到逻辑单元 201a(实际上盘高速缓冲存储器)的数据,在向主计算机报告写结束后,降级到外部卷 301。

[0510] 在常用类型的存储系统 130a 正常运行期间,仅存储系统 130a 具有对于外部卷 301 的访问权,替换类型的存储系统 130b 没有对于外部卷 301 的访问权。当存储系统 130b 从存储系统 130a 接收存储系统 130a 已向外部卷 301 写入数据的通知时,存储系统 130b 删除由存储系统 130a 写入到逻辑单元 201b(实际上盘高速缓冲存储器)的数据。这点和第二方式相同。

[0511] 根据第三方式,不仅常用类型的存储系统 130a,而且替换类型的存储系统 130b,也可以保存未降级到外部卷 301 中的脏数据。因此,在常用类型的存储系统 130a 中发生故障时,替换类型的存储系统 130b 取得对于外部卷 301 的访问权,通过把脏数据降级到外部卷 301,可以把在外部卷 301 中存储的数据更新为最新的状态。如果能够把在外部卷 301 中存储的数据更新为最新的状态,则主计算机 100,通过从路径 401 切换到路径 403,能够对替换类型的存储系统 130b 发行 I/O 请求。

[0512] 图 59 表示作为作为本实施例的变形例的计算机系统 580 的系统结构。计算机系统 580,在计算机系统的系统结构(参照图 2)中,具有把存储系统 130a、130b 置换成虚拟化装置 180a、180b、180c 而形成的系统结构。为说明方便,把虚拟化装置 180a、180b、180c 总称为虚拟化装置 180。虚拟化装置 180 具有:用于通过存储网络连接光纤通道交换机 120 的端口 181、控制虚拟化外部卷 301 的处理等的控制处理器 182、存储控制处理器 182 执行的控制程序等的存储器 183、用于通过 IP 网络 175 连接管理服务器 110 的端口 184。虚拟化装置

180 例如是具有虚拟化存储资源的功能的专用的虚拟化装置或者虚拟化转换器 (switch)。虚拟化装置 180 具有向主计算机 100 提供虚拟化外部卷 301 而形成的逻辑单元的功能。

[0513] 计算机系统 580, 例如能够使虚拟化装置 180a 向主计算机 100 提供的逻辑单元向虚拟化装置 180b 移动。从虚拟化装置 180a 向虚拟化装置 180b 的逻辑单元移动处理的概要和上述的处理内容大体相同。但是虚拟化装置 180, 因为不具有用于临时存储向外部卷 301 读写的数据的盘高速缓冲存储器, 所以不仅不需要把虚拟化装置 180 的动作模式设定为高速缓冲存储器直通模式, 而且也不需要等待从虚拟化装置 180 对于外部卷 301 的降级结束。仅通过将移动源的虚拟化装置 180a 向主计算机 100 提供的逻辑单元中设定的属性移交给移动目的地的虚拟化装置 180b 向主计算机 100 提供的逻辑单元, 由此能够使移动处理完成。

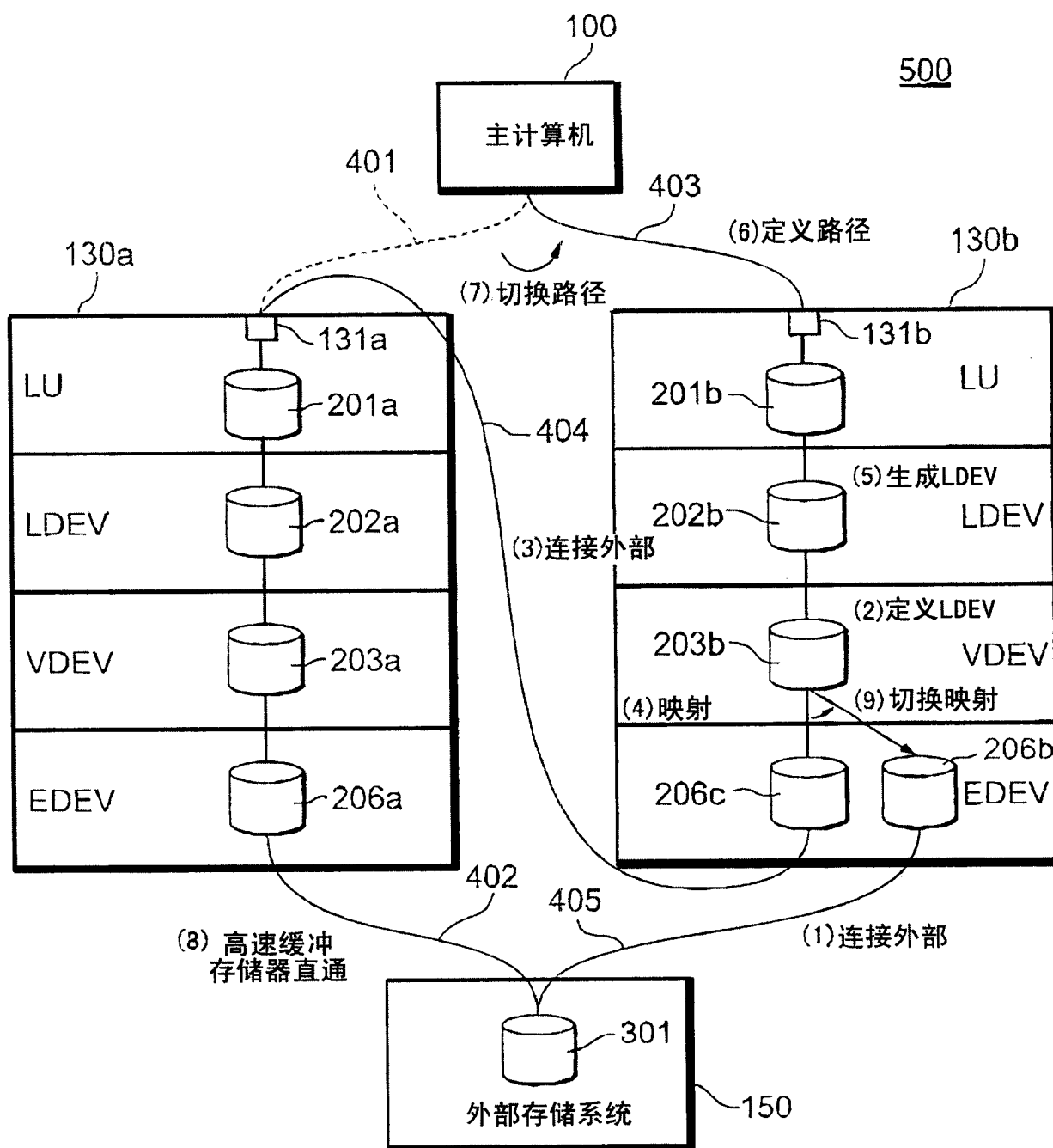


图 1

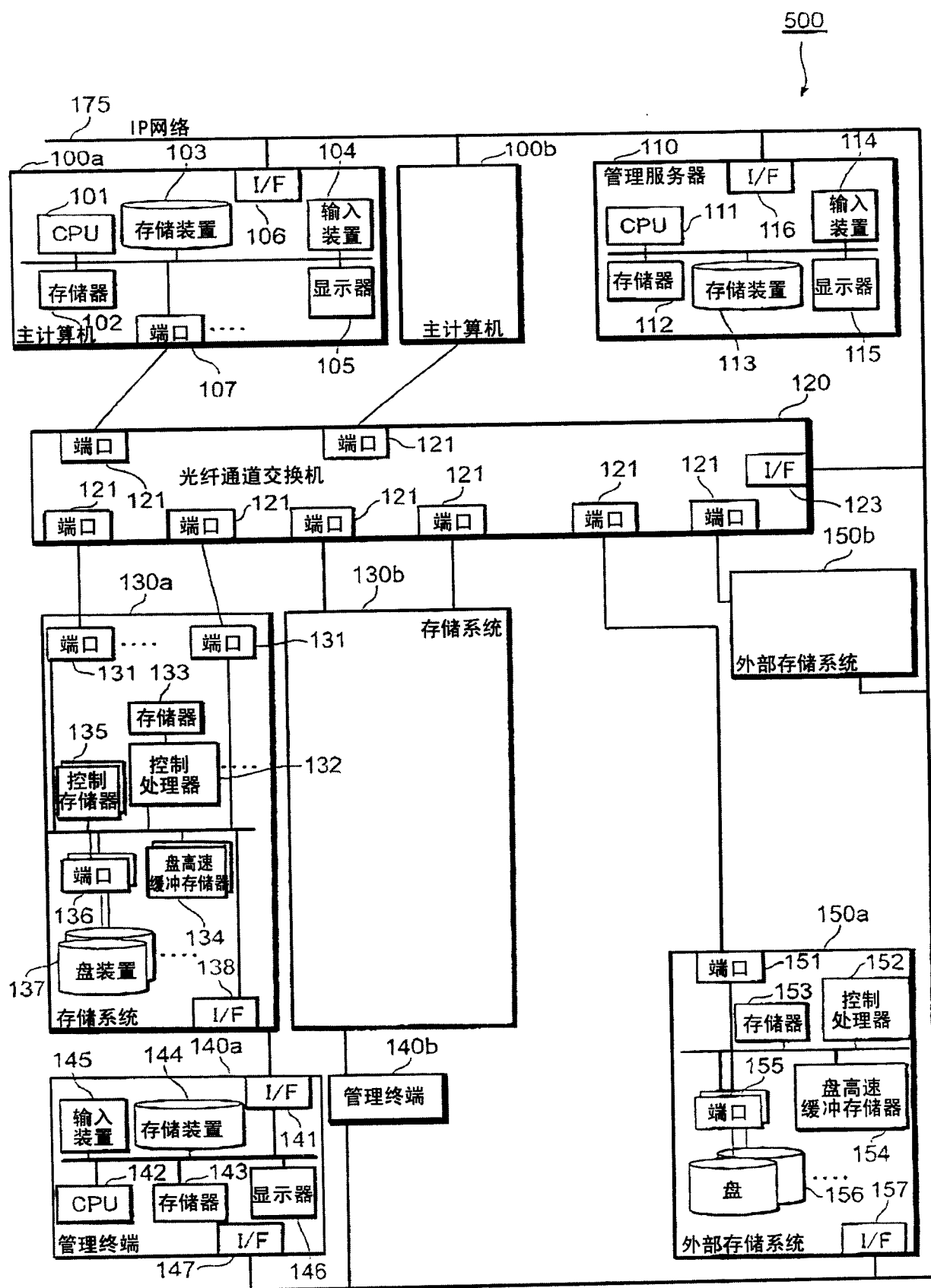


图 2

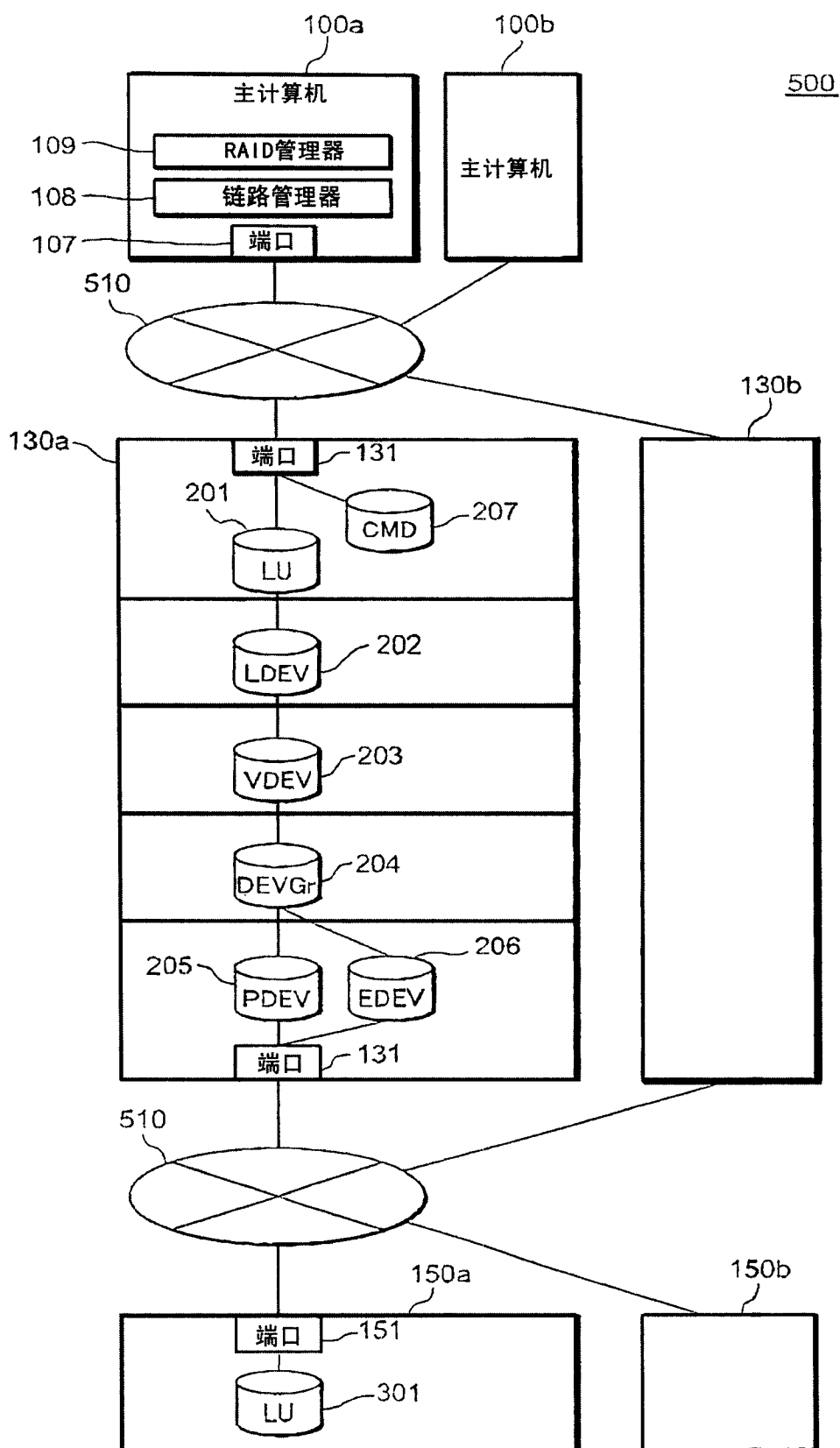


图 3

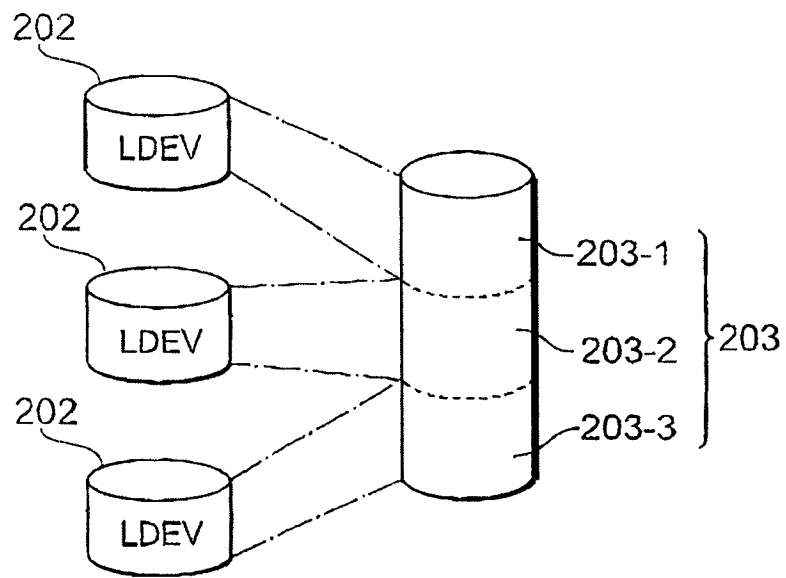


图 4

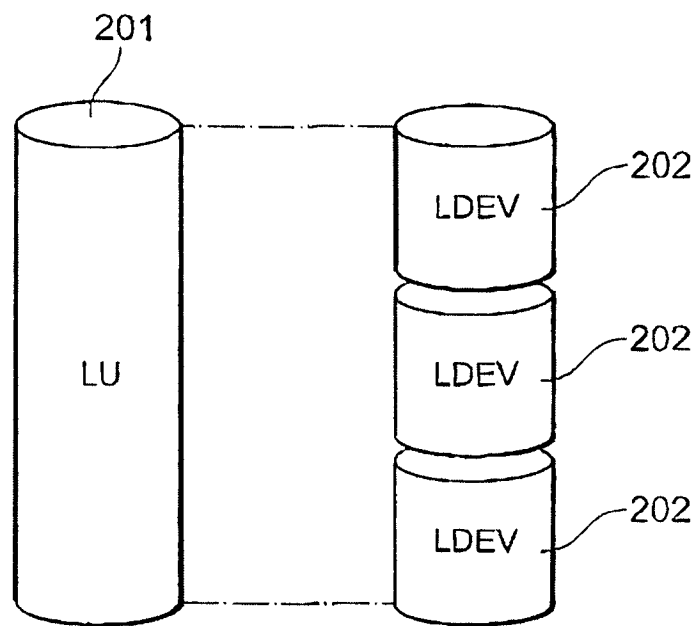


图 5

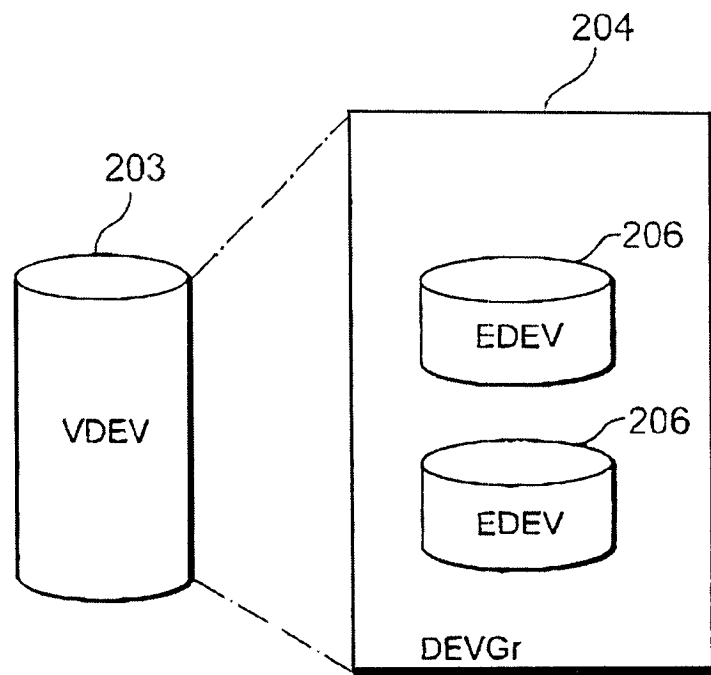


图 6

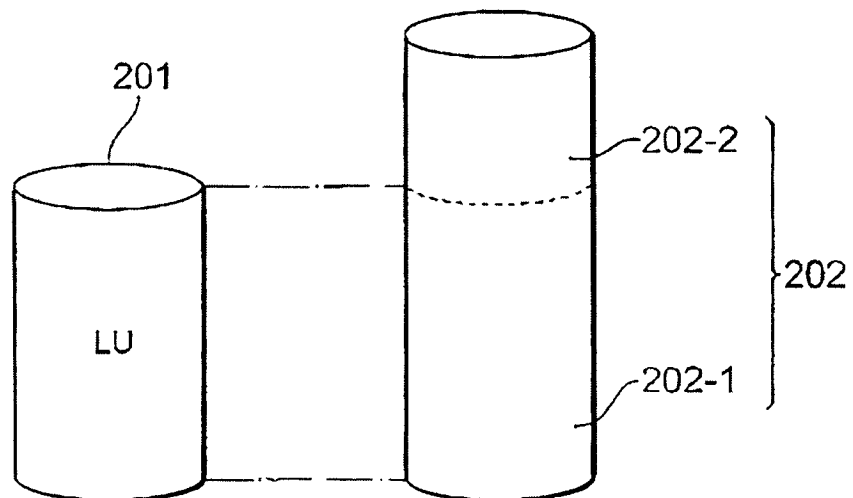


图 7

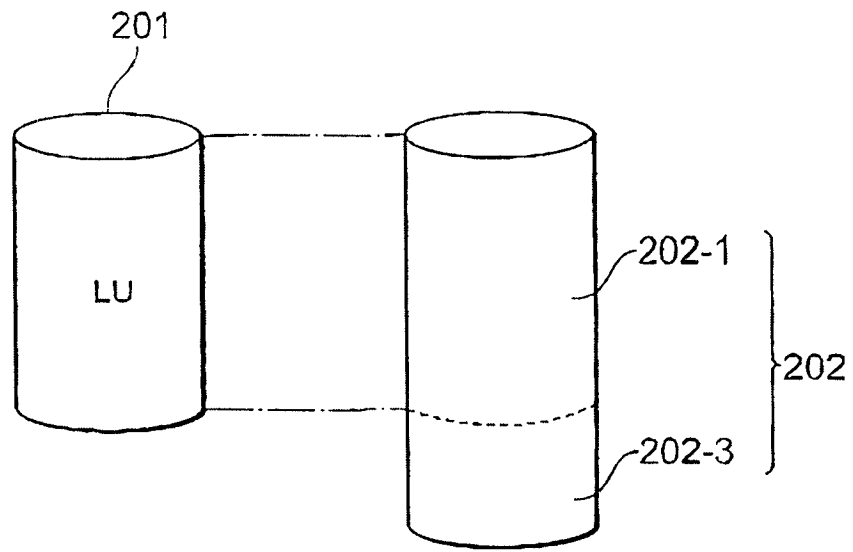


图 8

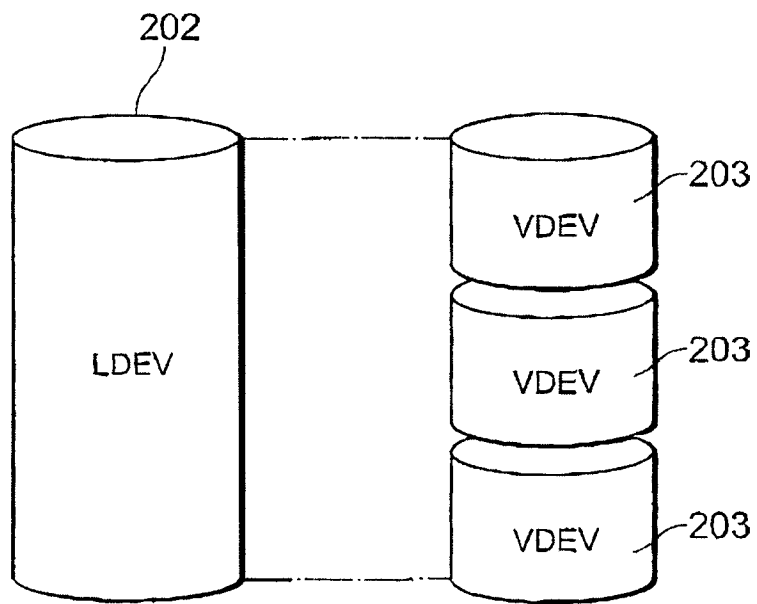


图 9

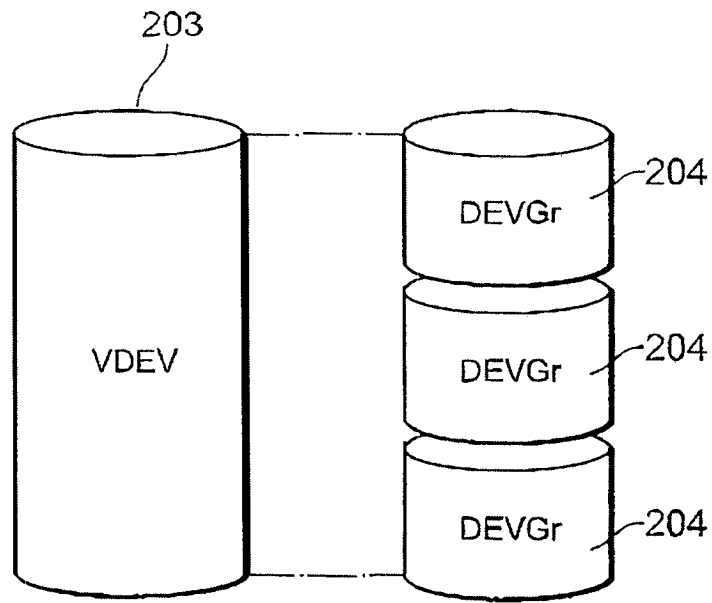


图 10

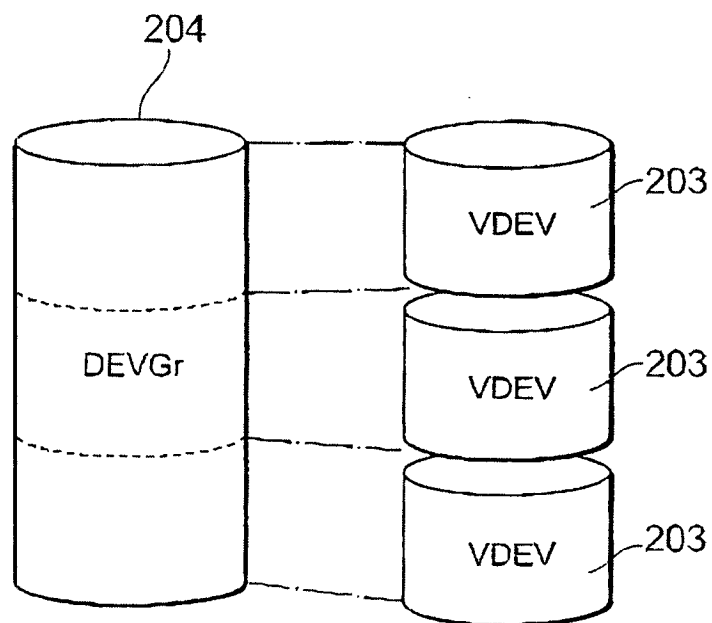


图 11

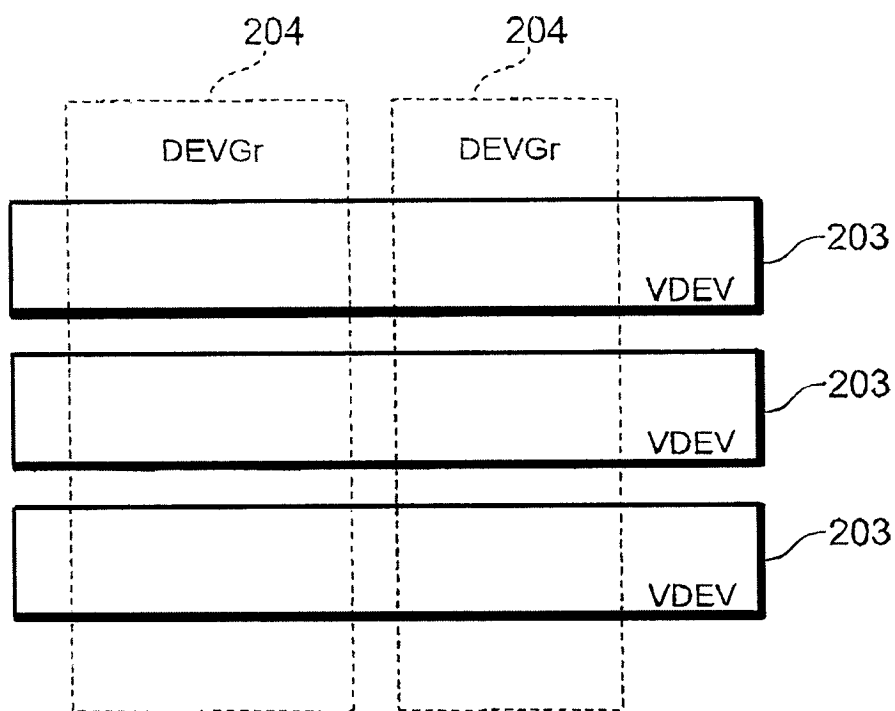


图 12

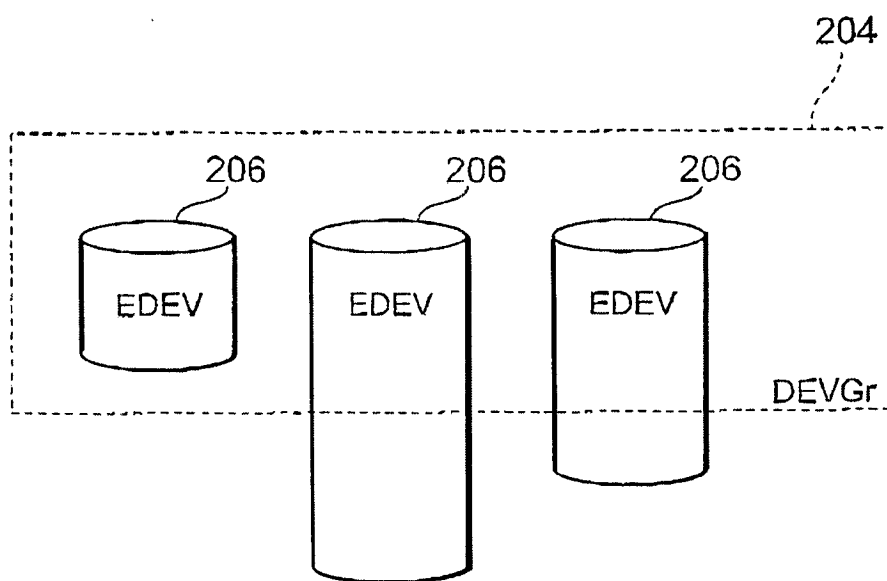


图 13

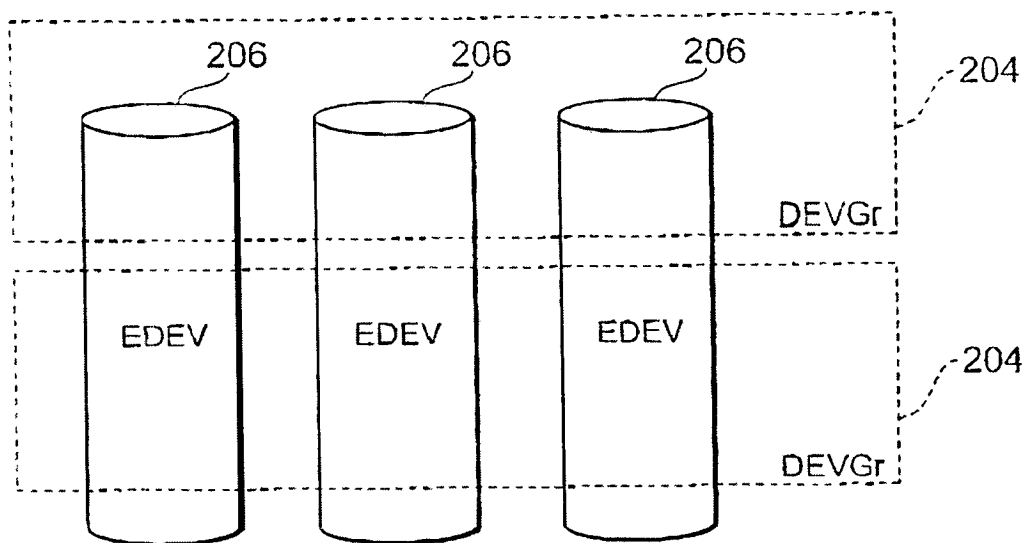


图 14

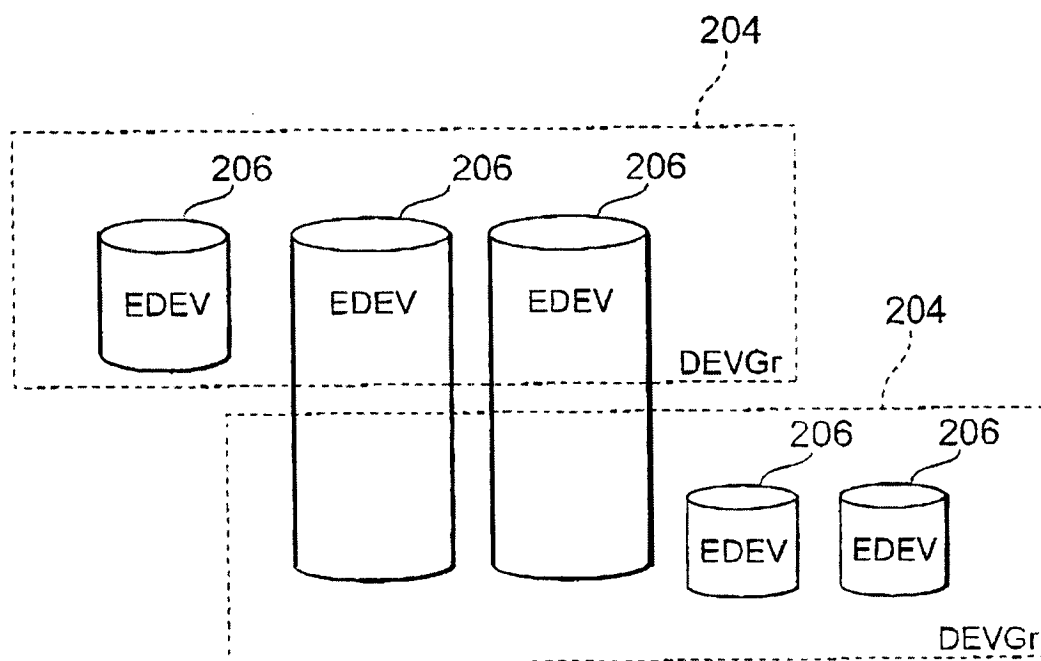


图 15

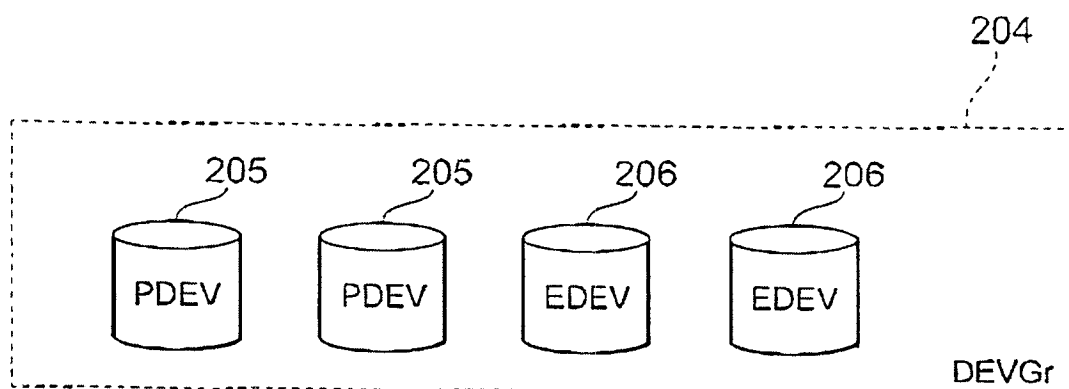


图 16

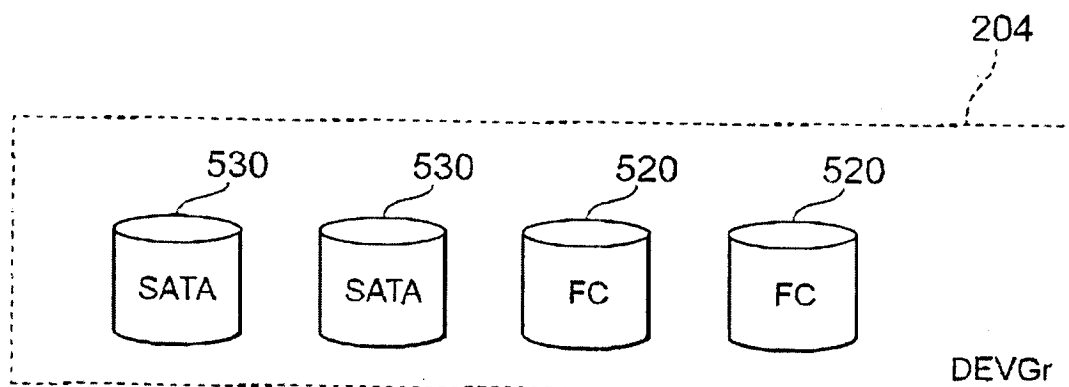


图 17

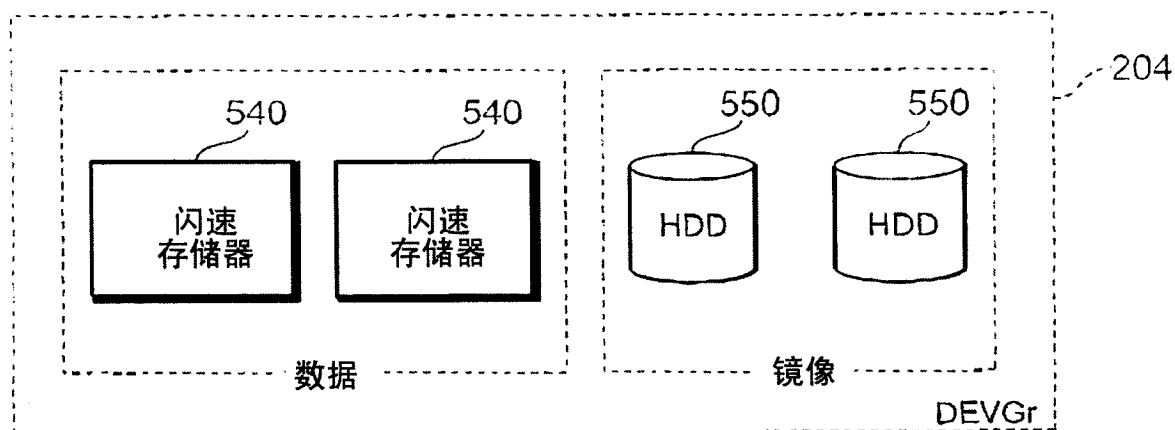


图 18

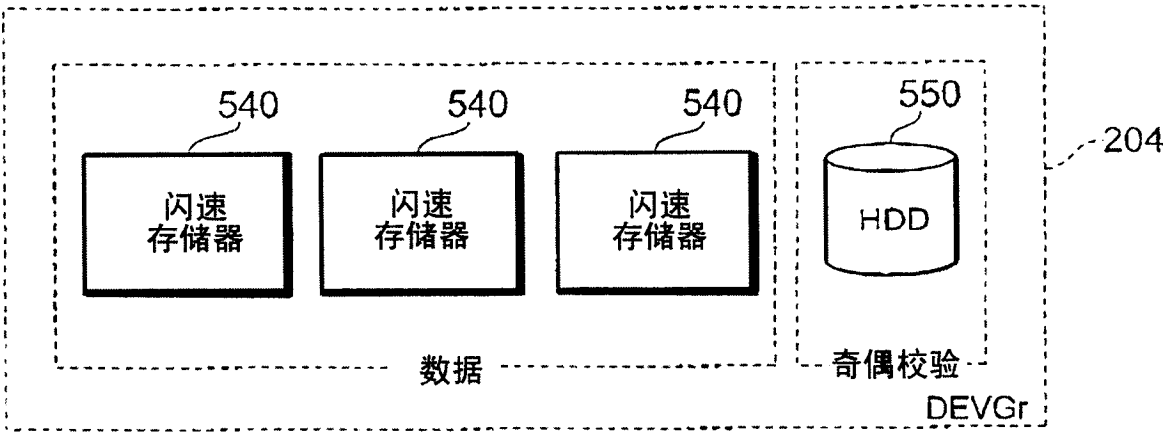


图 19

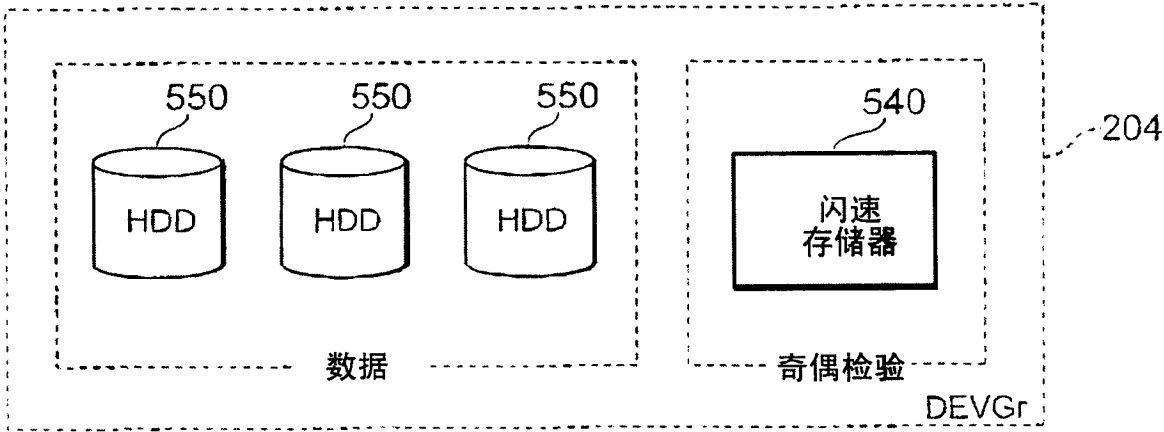


图 20

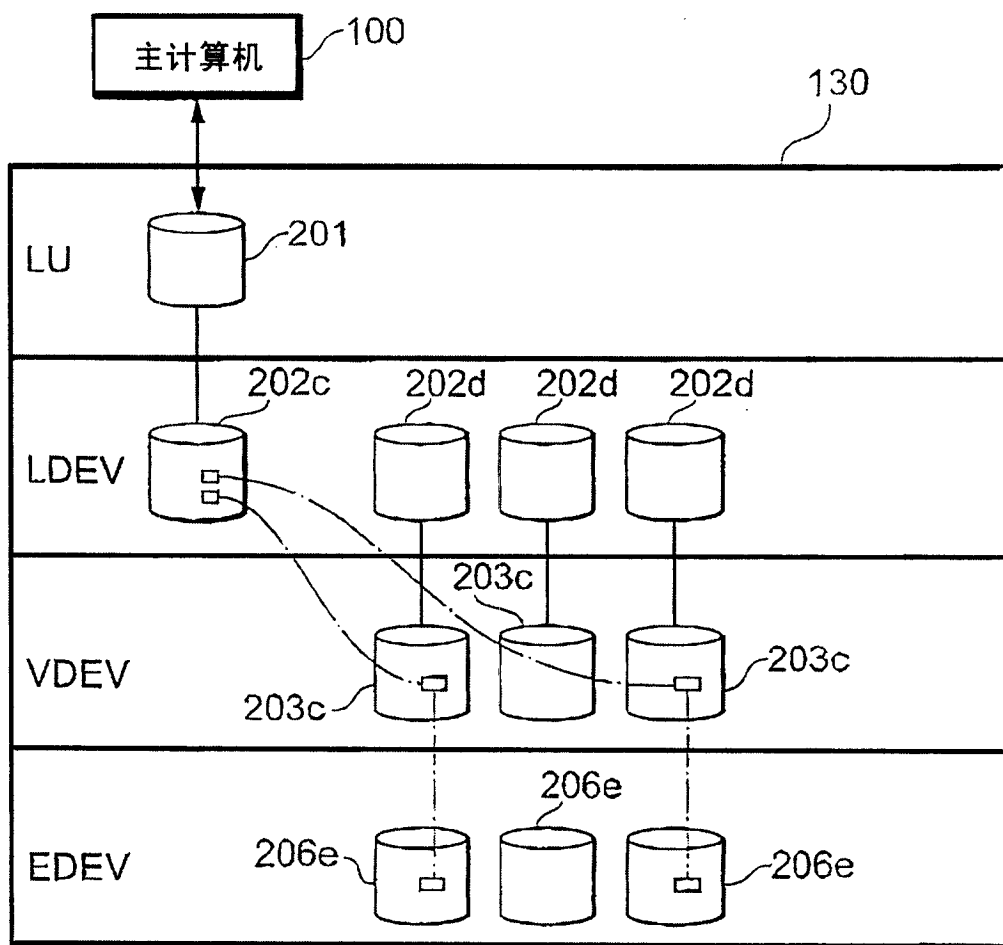


图 21

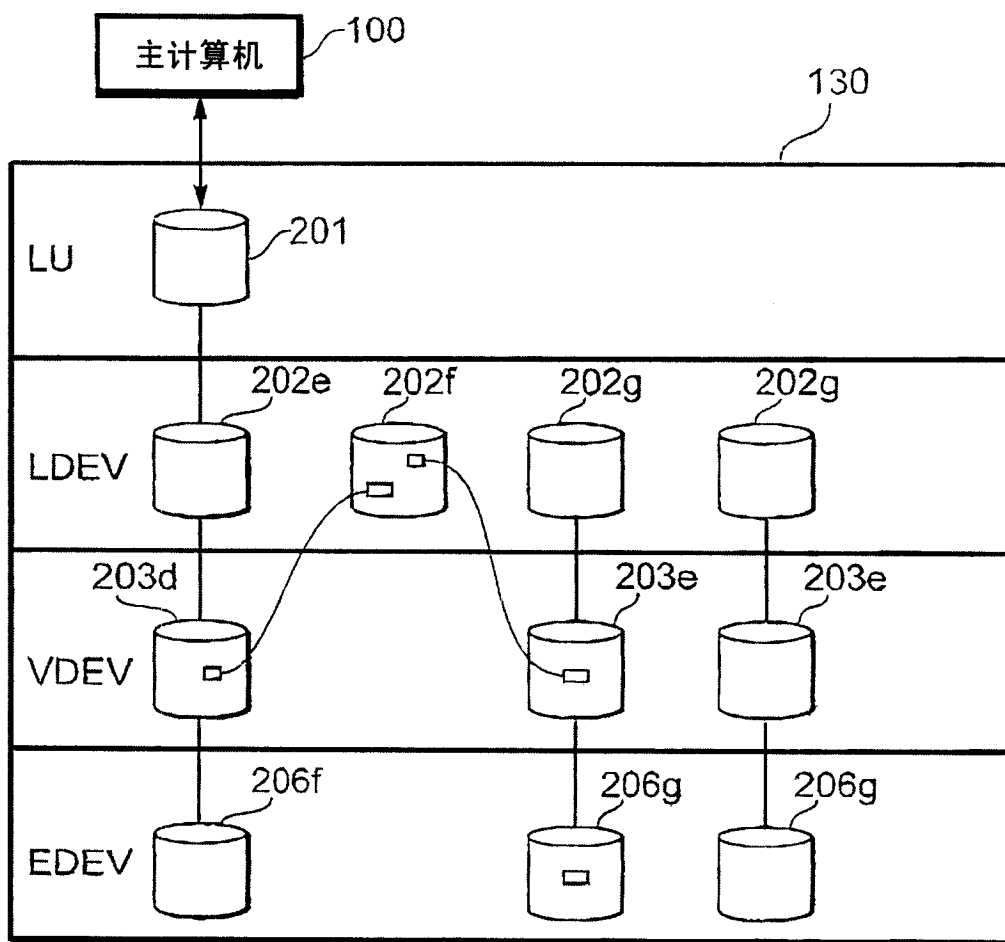


图 22

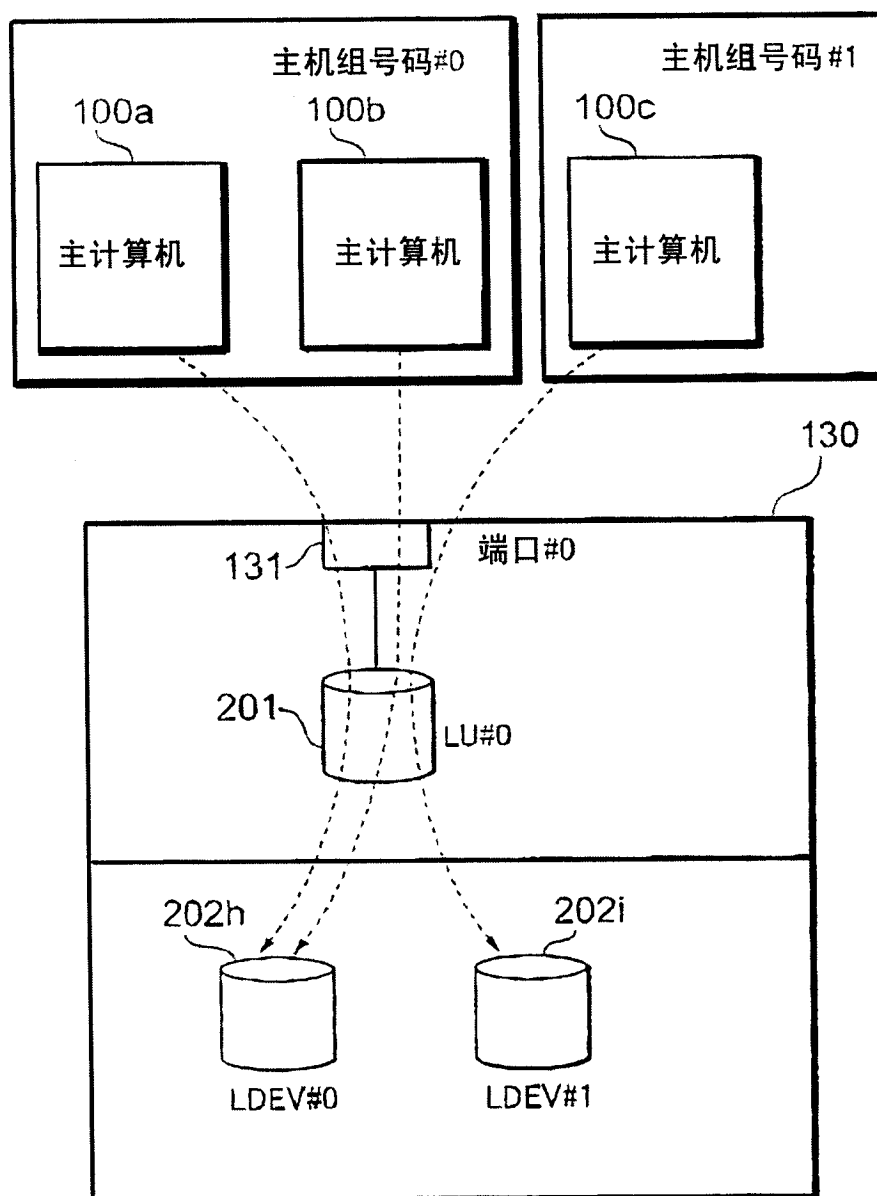


图 23

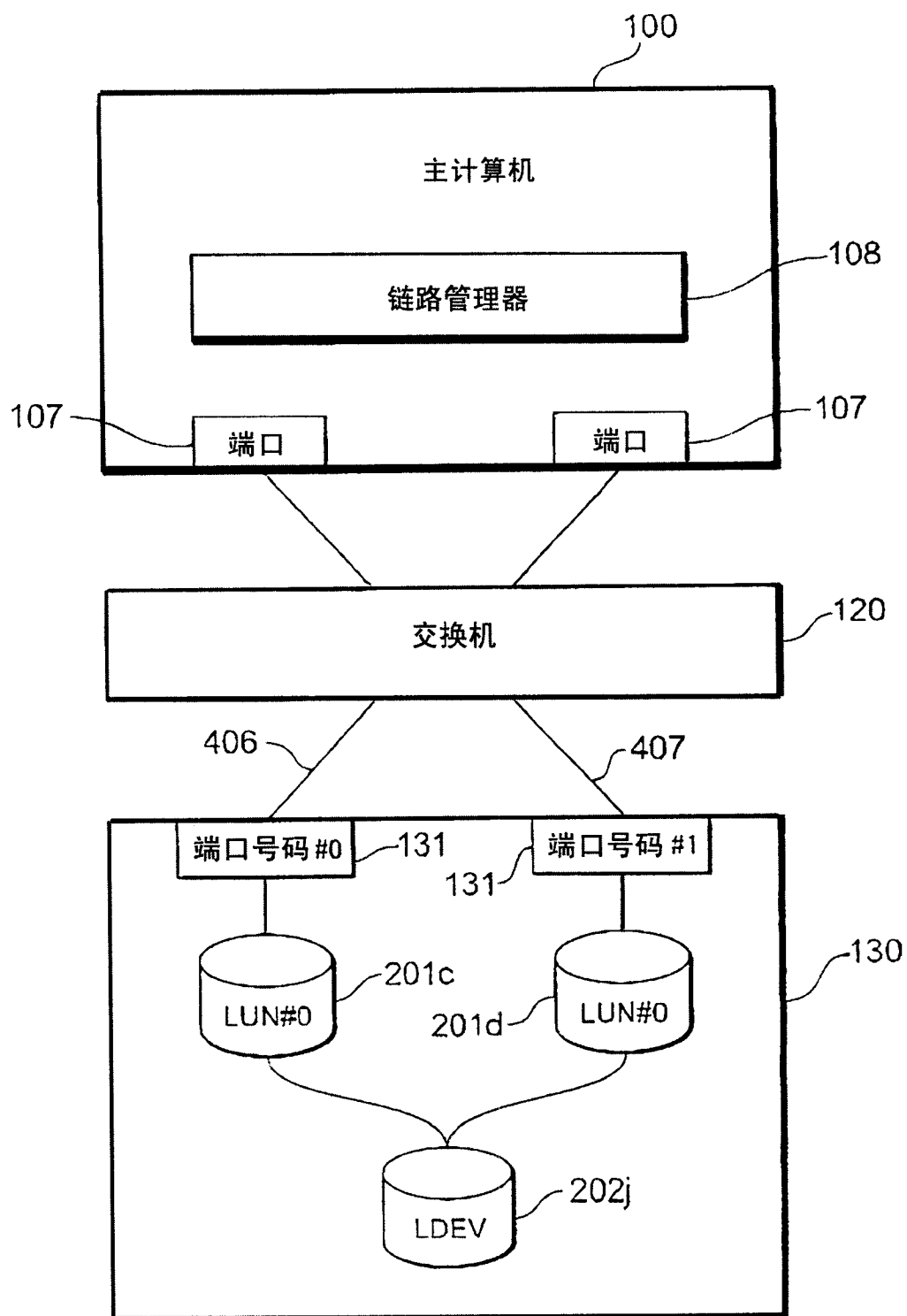


图 24

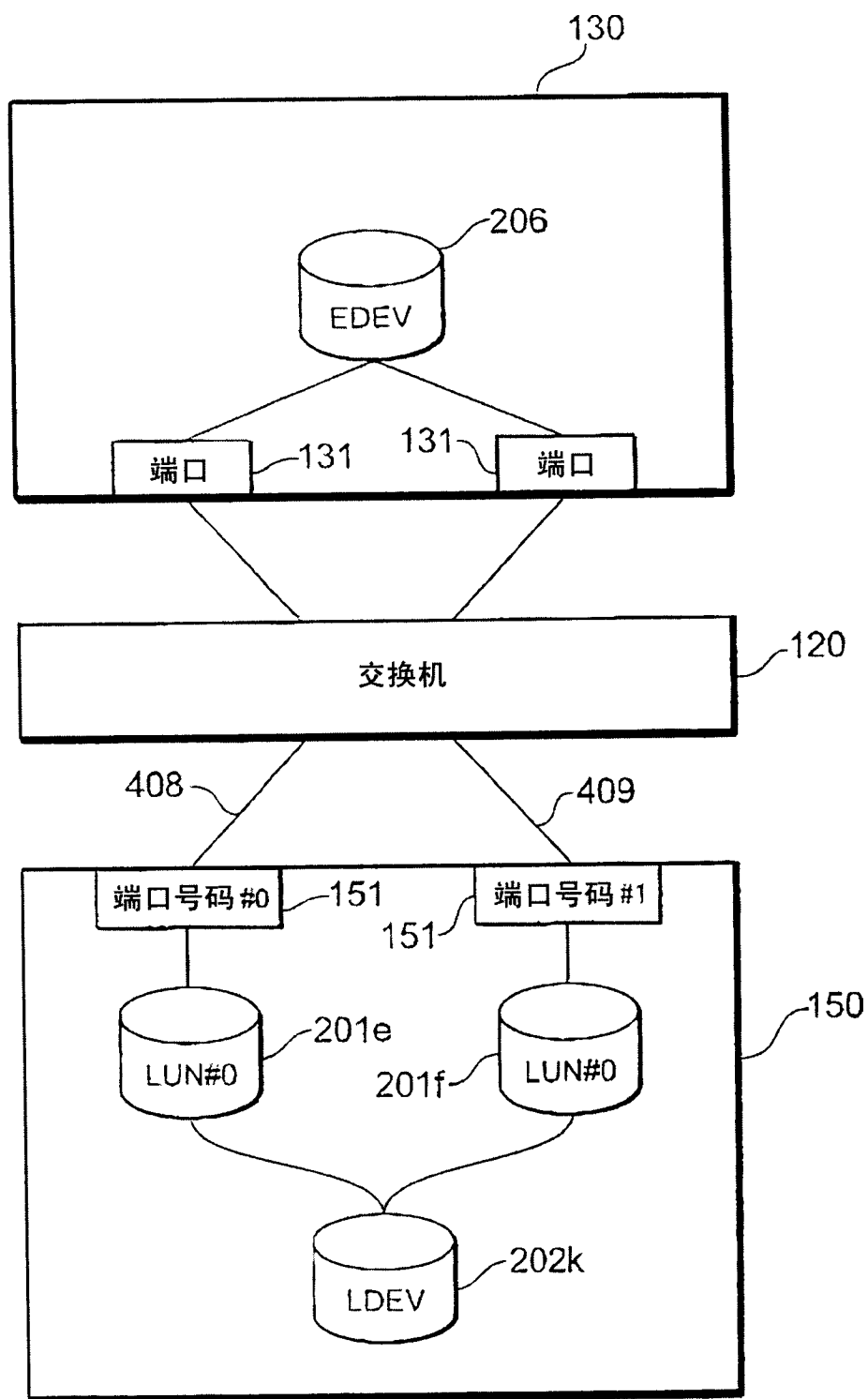


图 25

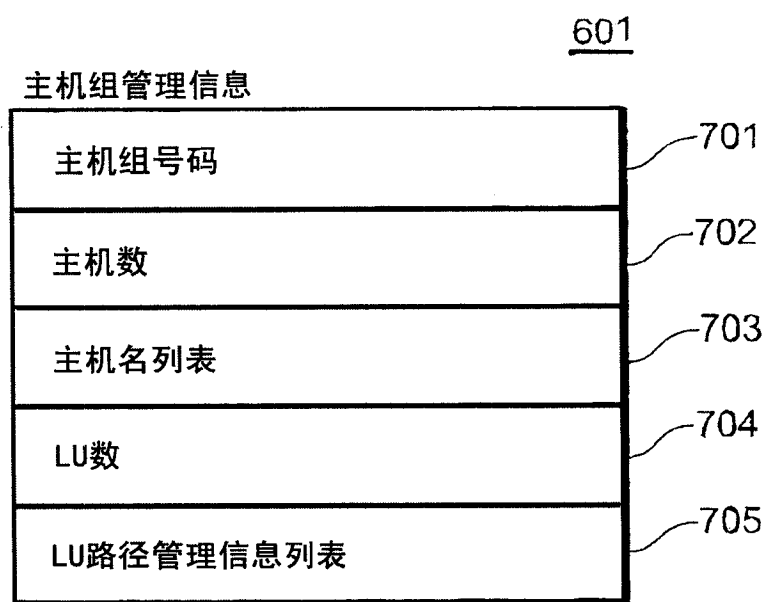


图 26

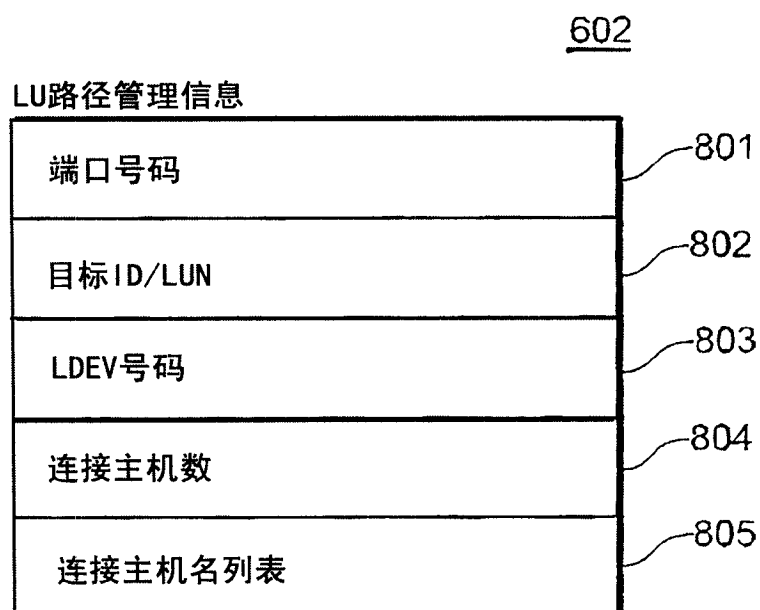


图 27

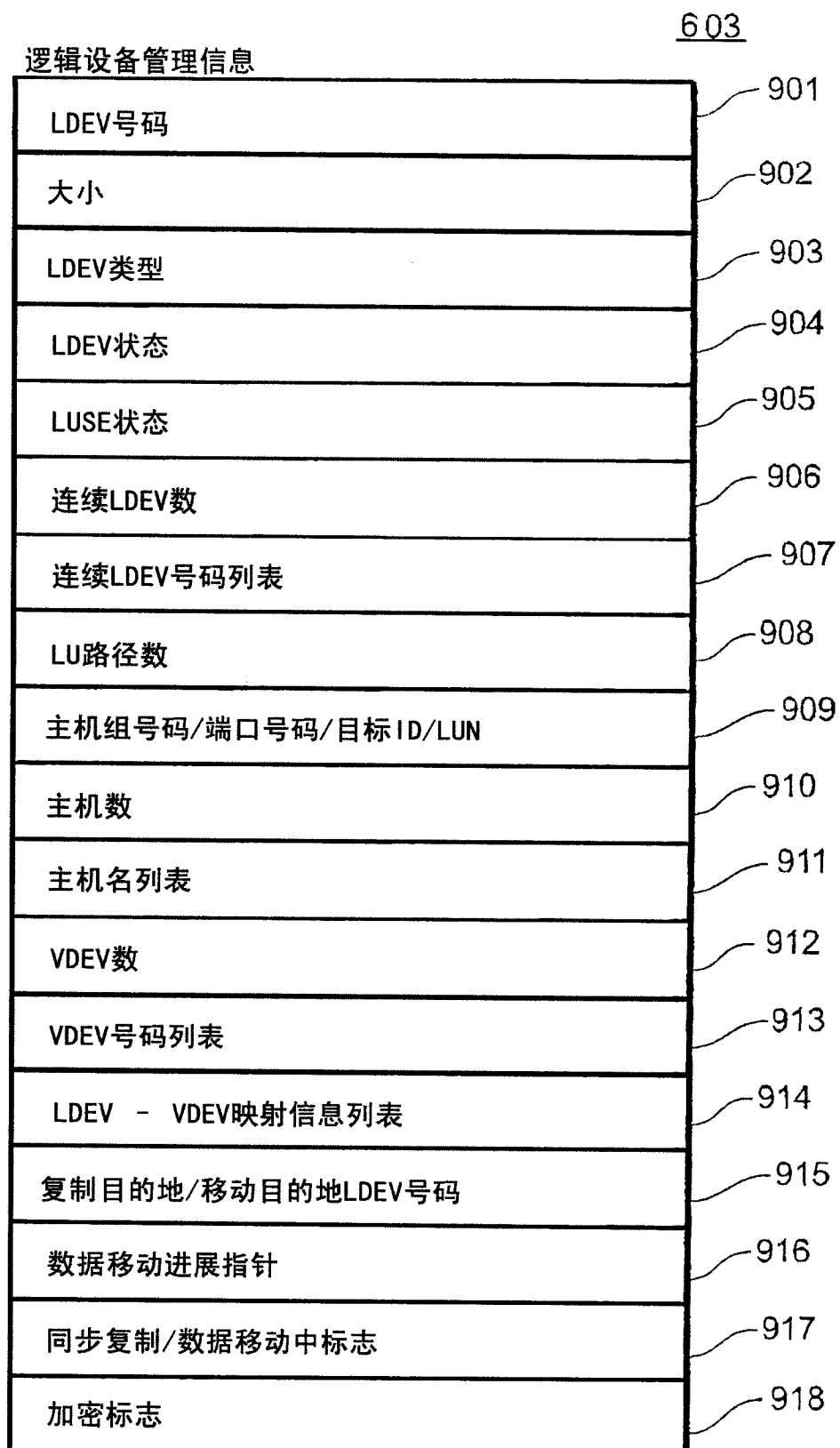


图 28

604

虚拟设备管理信息

VDEV号码	1001
大小	1002
VDEV状态	1003
RAID结构	1004
LDEV数	1005
LDEV号码列表	1006
VEDV - LDEV映射信息列表	1007
设备组数	1008
设备组号码列表	1009
VDEV - DEVGr映射信息列表	1010

图 29

605

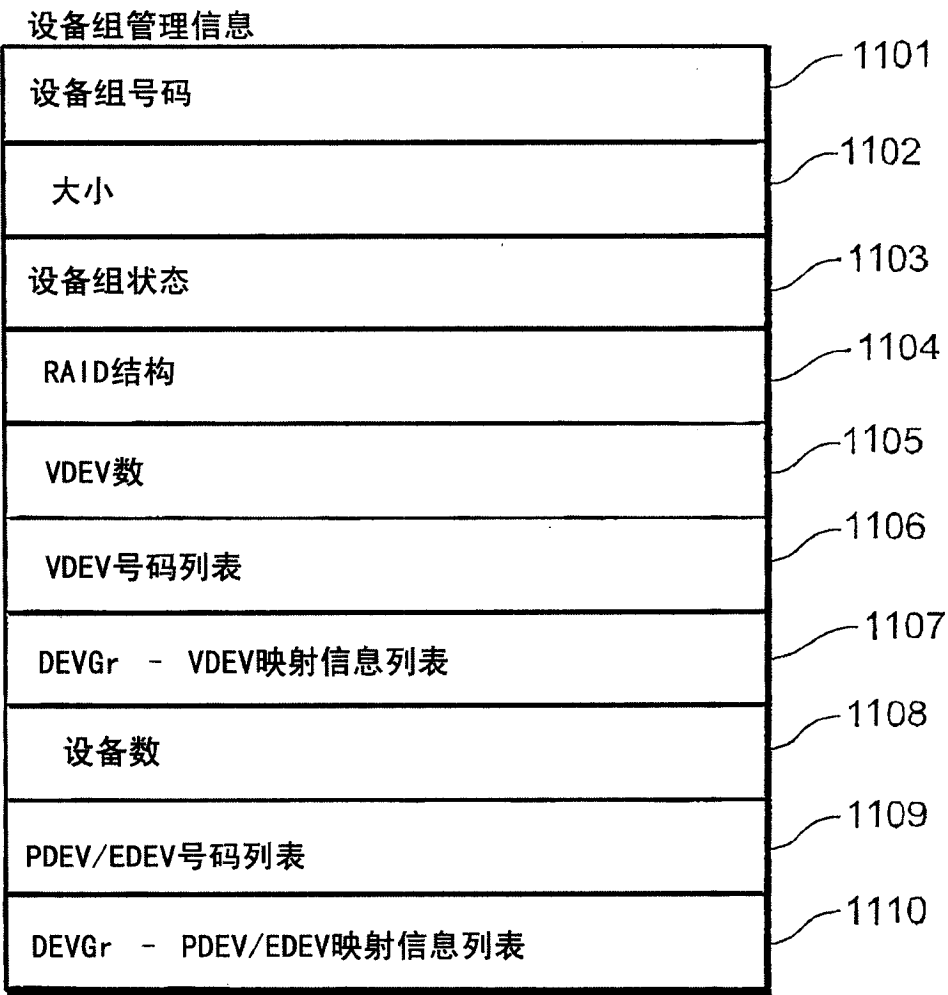


图 30

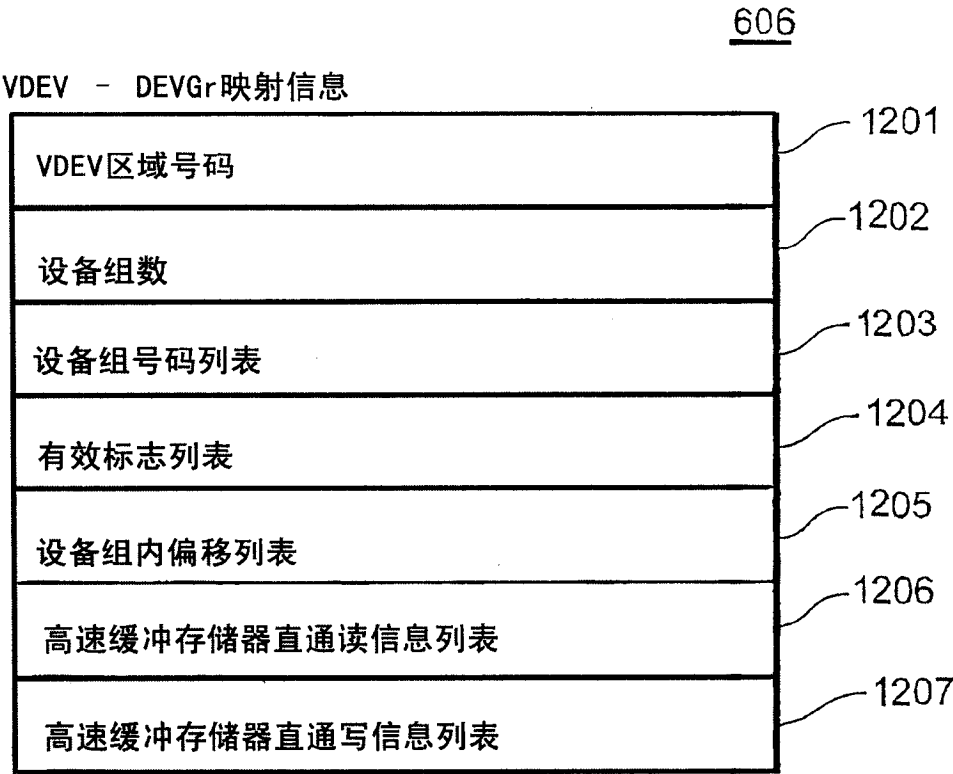


图 31

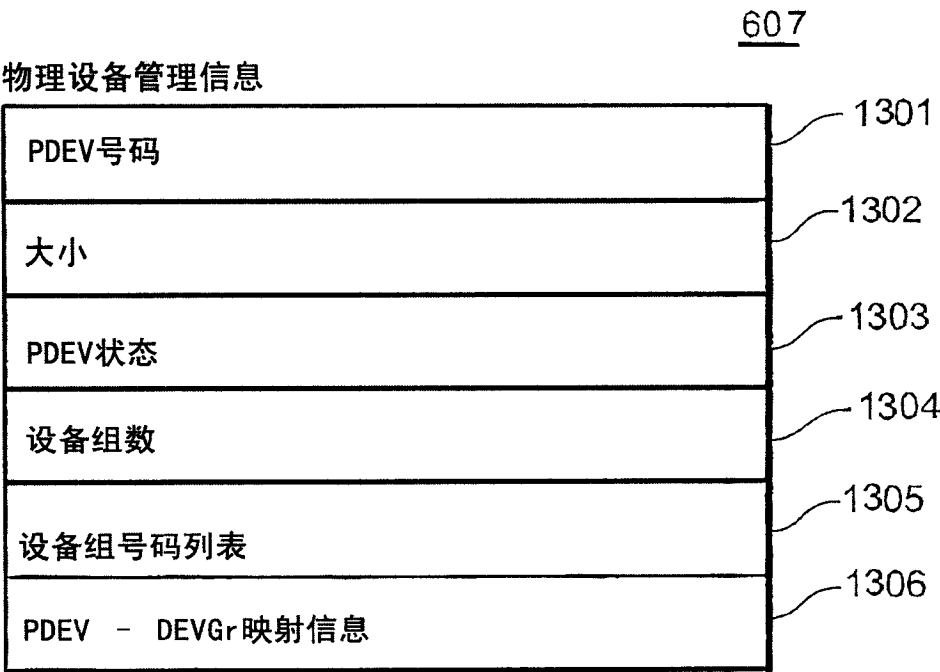


图 32

608

扩展设备管理信息

EDEV号码	1401
大小	1402
EDEV状态	1403
设备组数	1404
设备组号码列表	1405
EDEV - DEVGr映射信息	1406
存储系统识别信息	1407
外部存储系统内LDEV号码	1408
启动器端口号码列表	1409
目标端口ID/LUN	1410
访问控制信息	1411

图 33

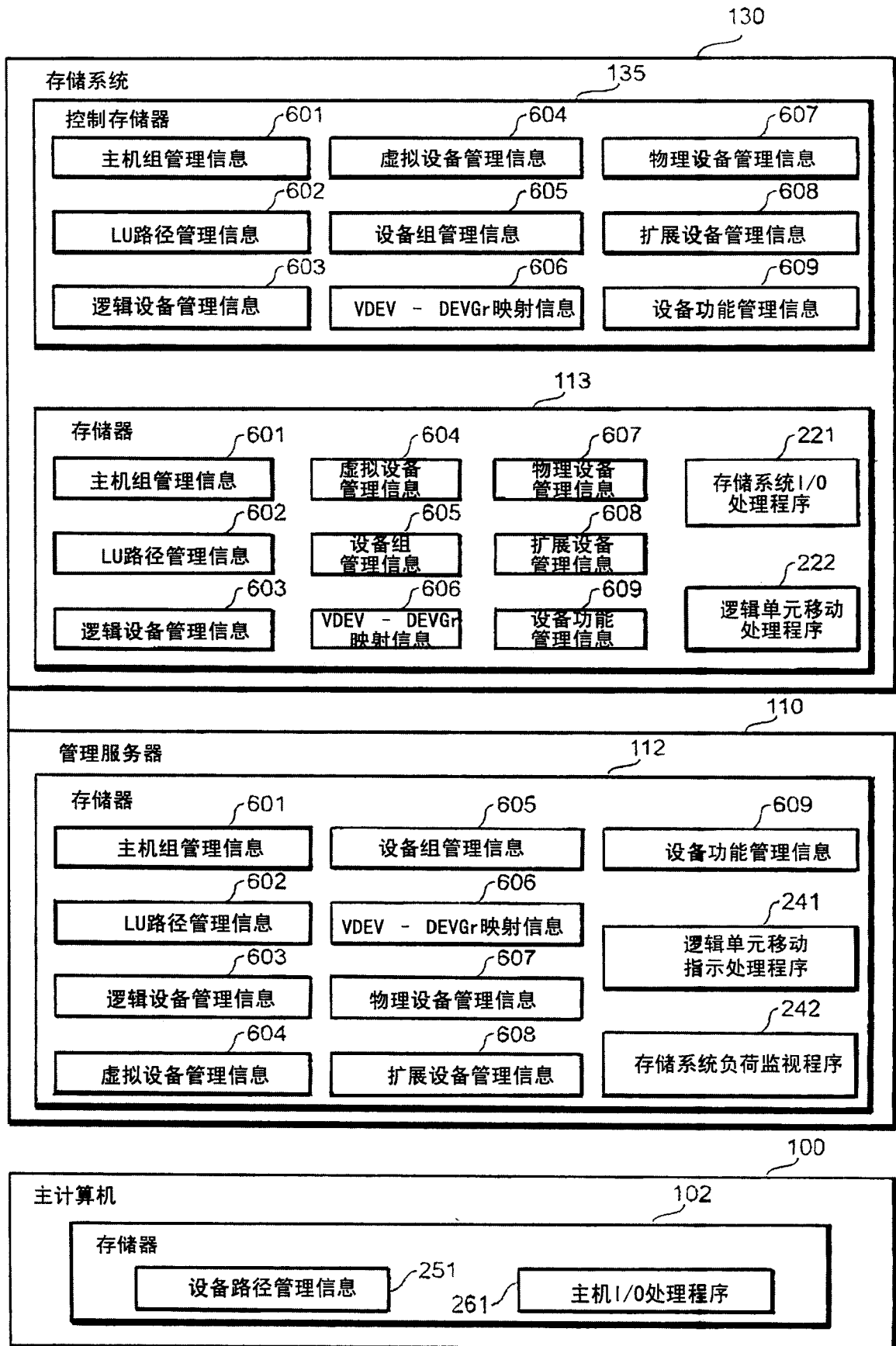


图 34

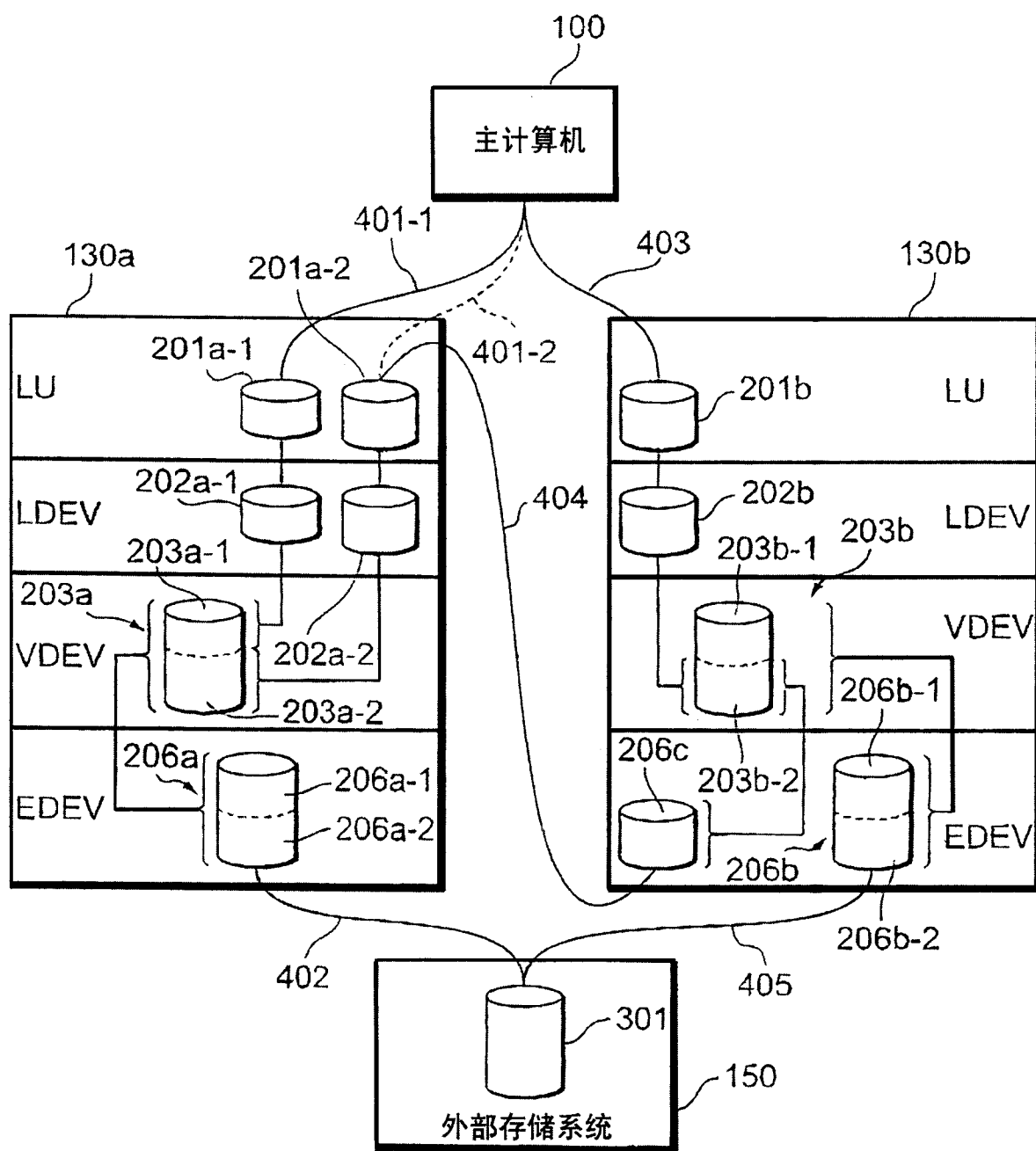


图 35

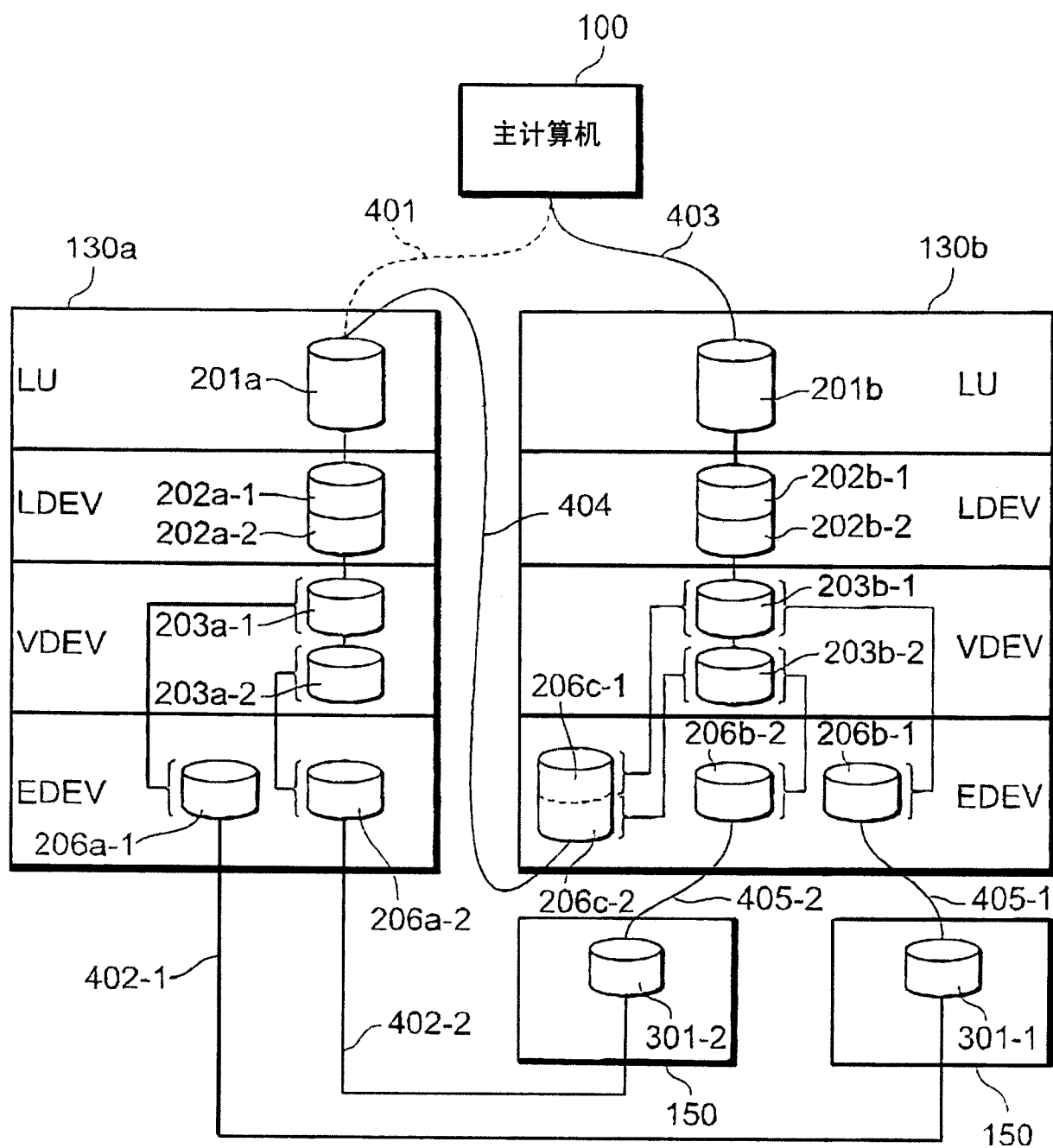


图 36

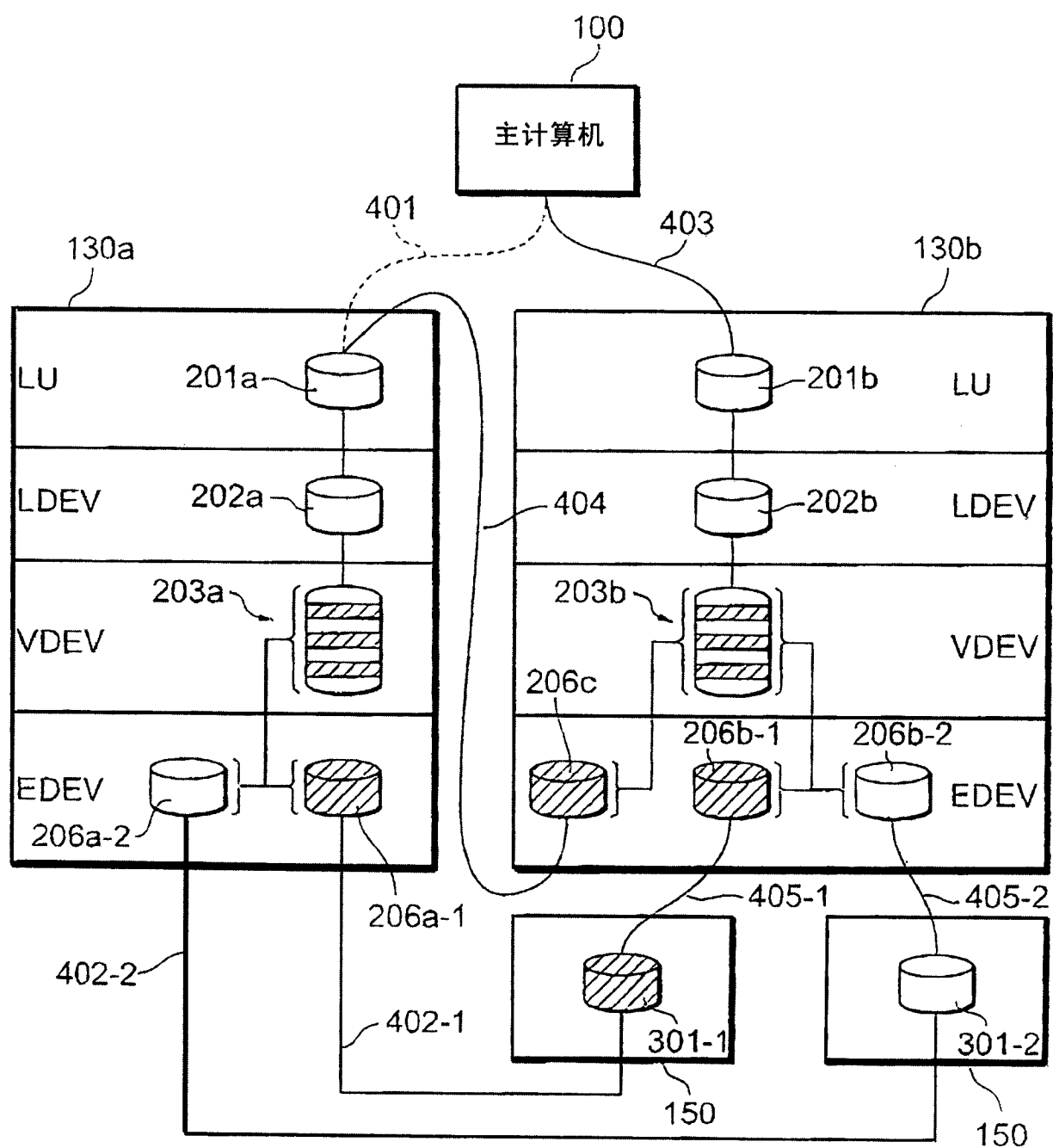


图 37

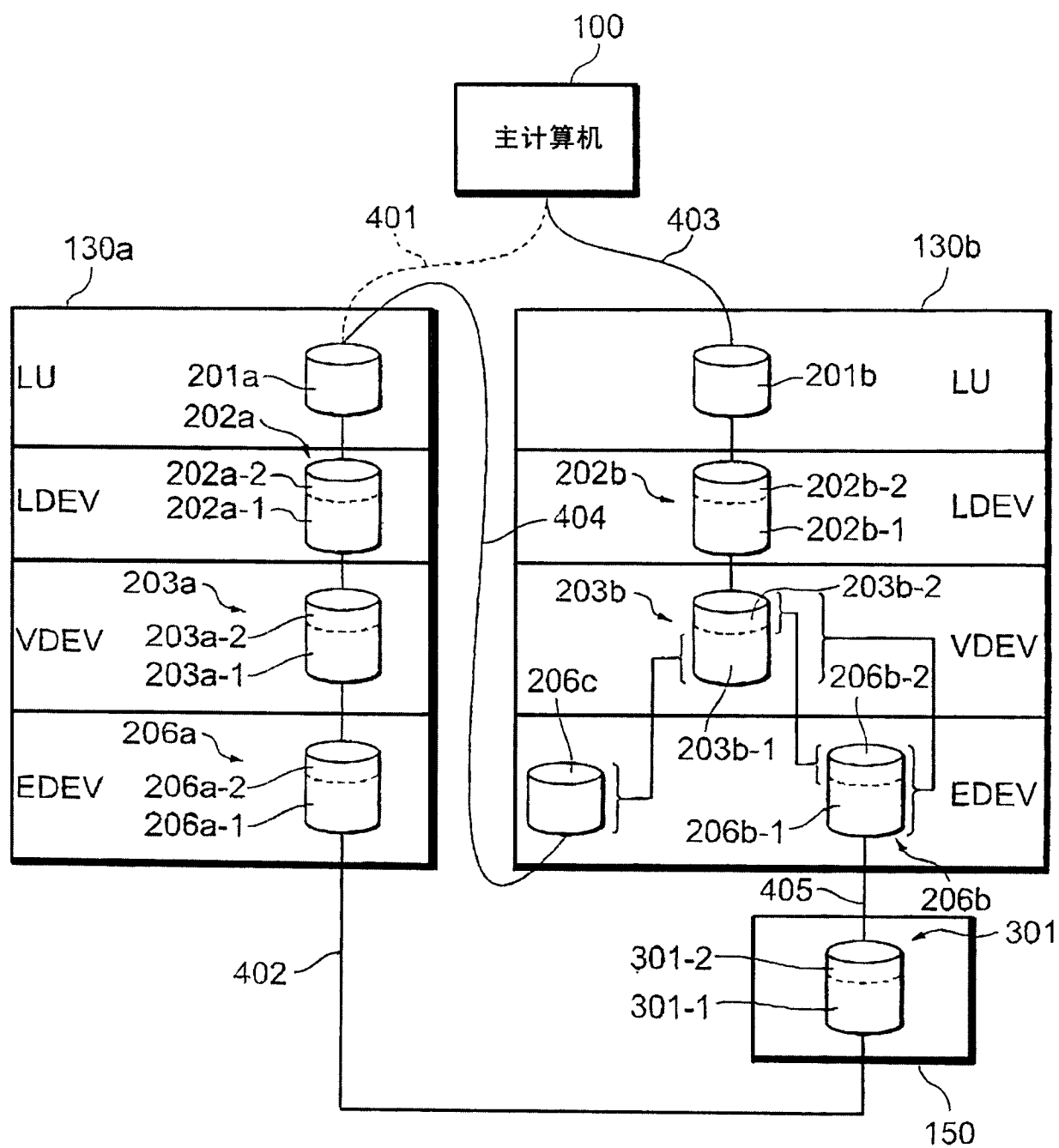


图 38

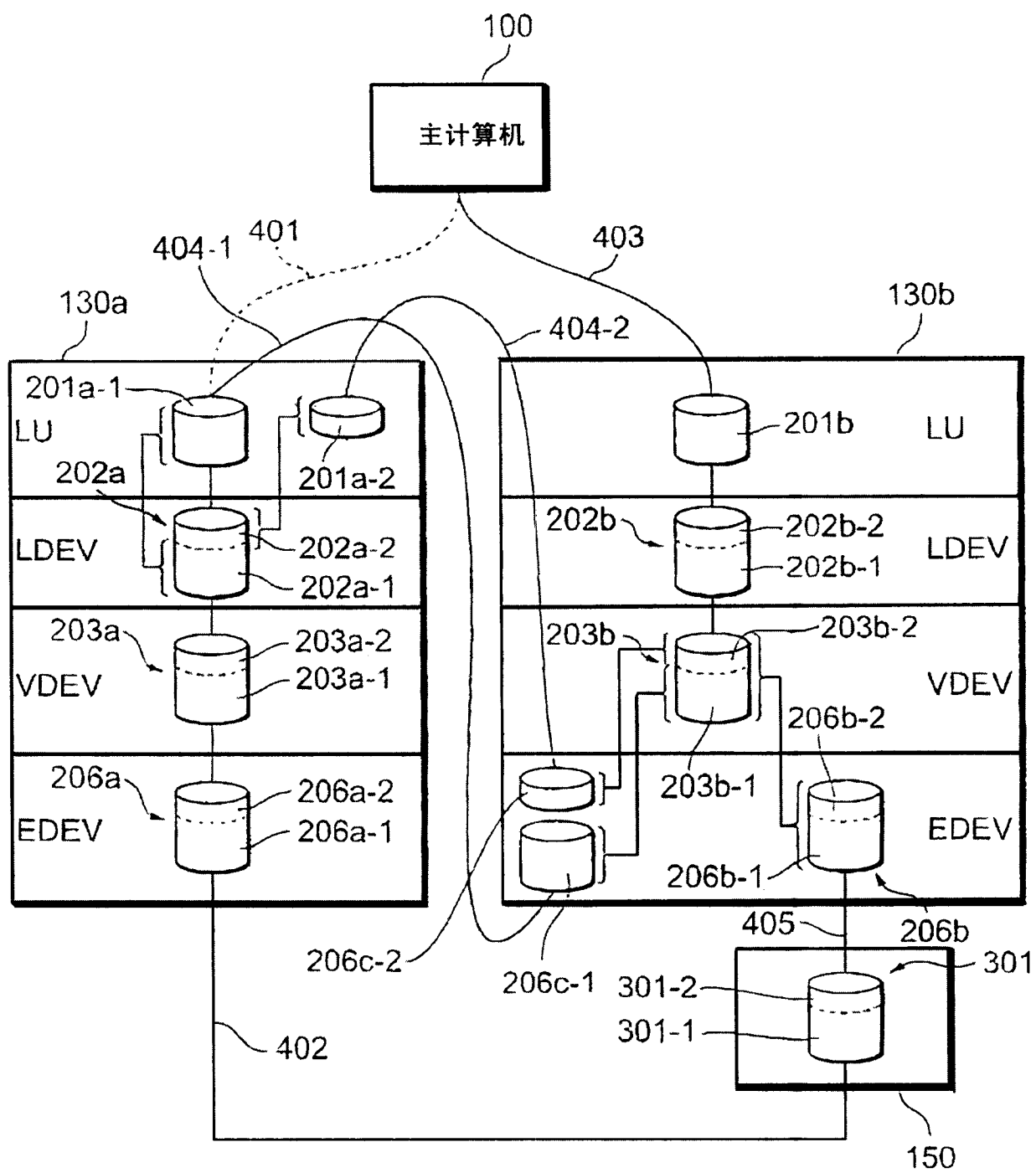


图 39

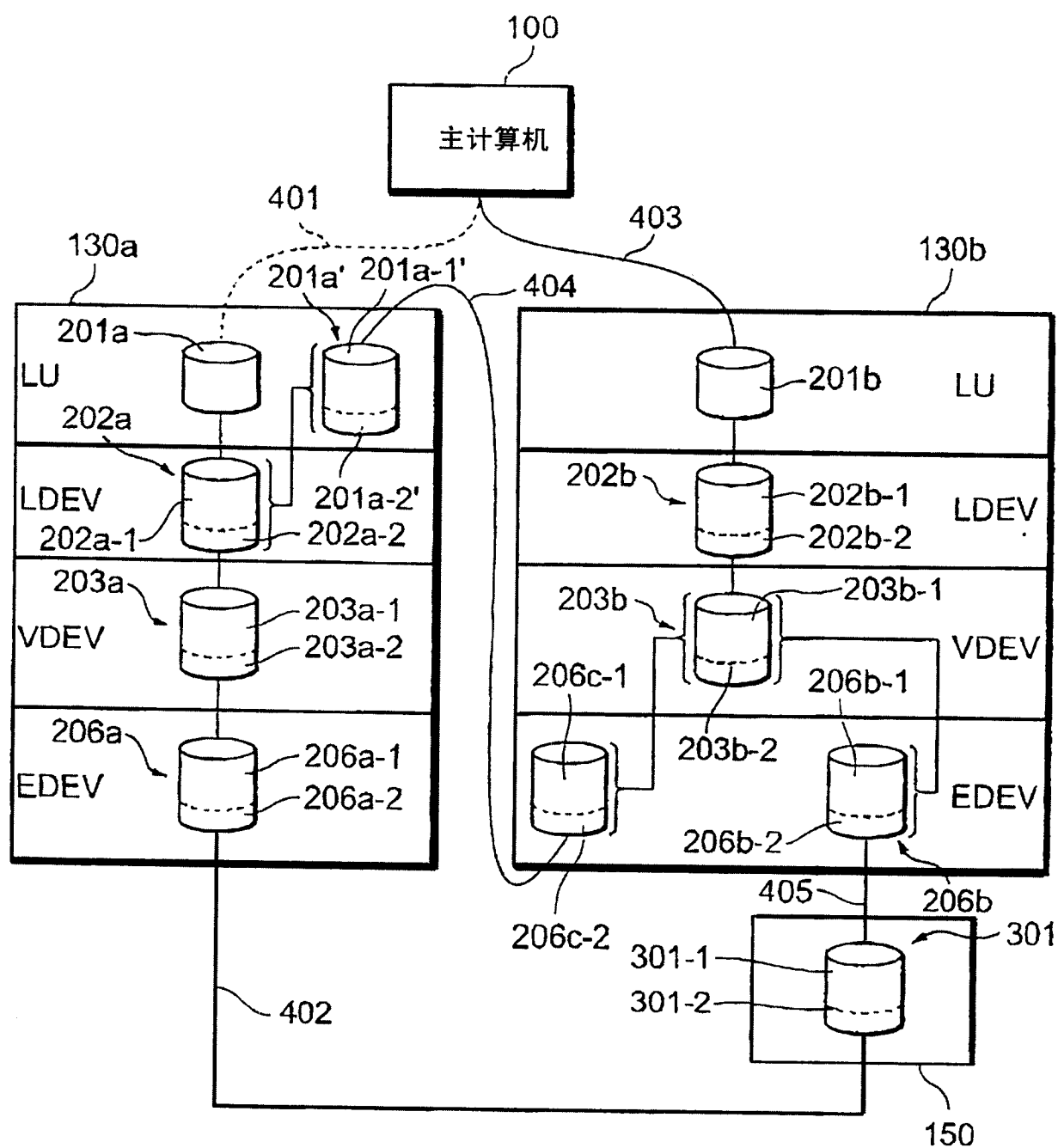


图 40

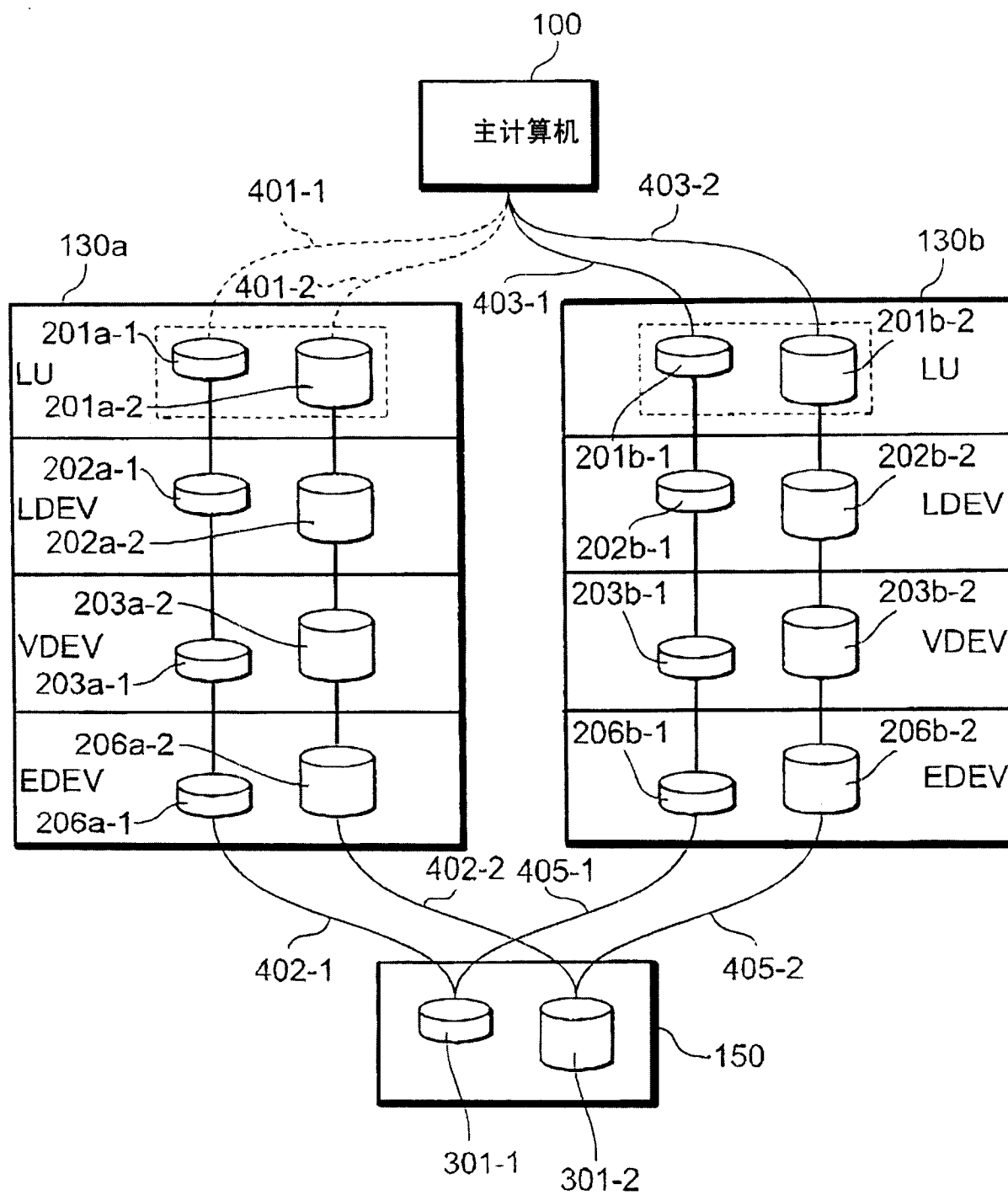


图 41

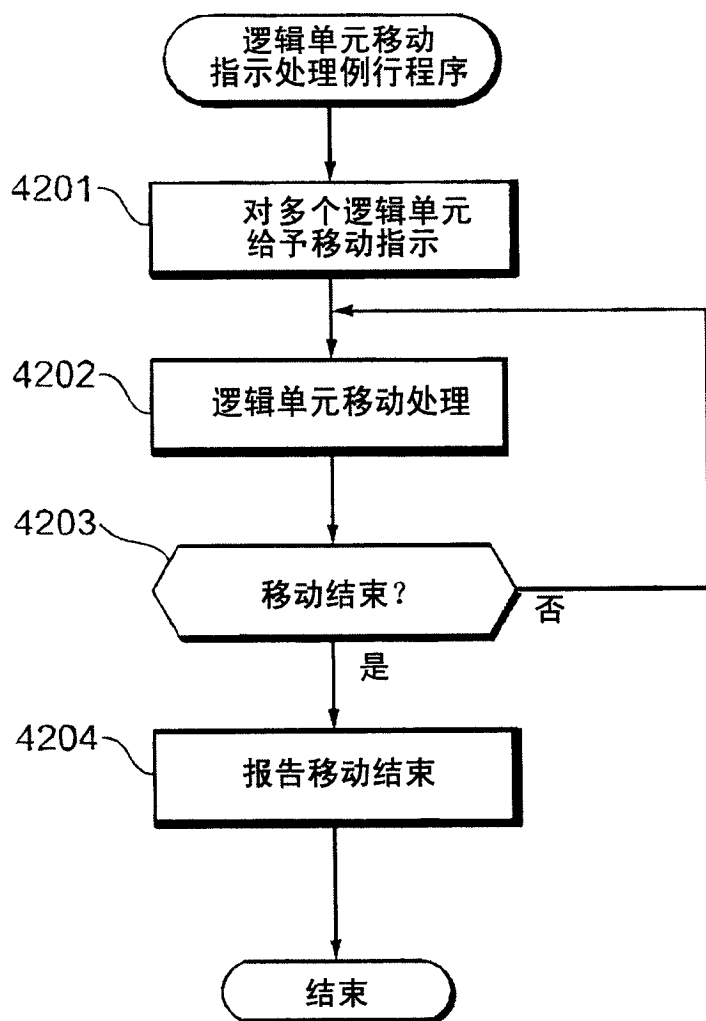


图 42

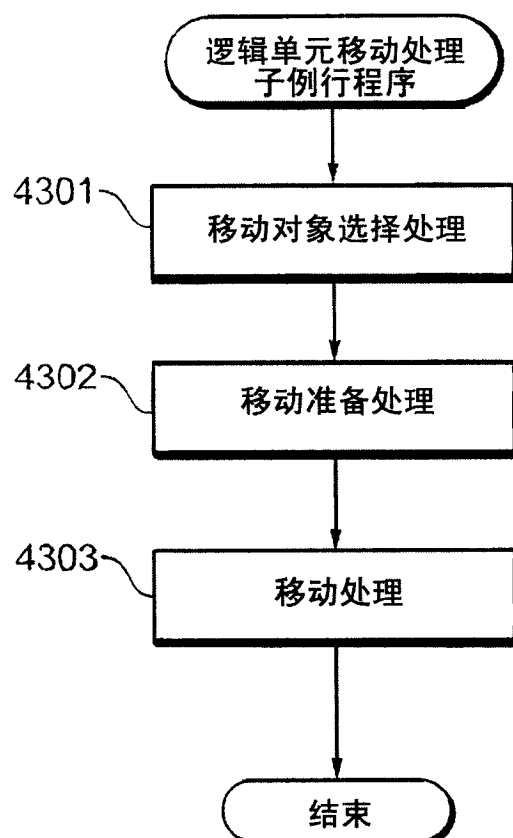


图 43

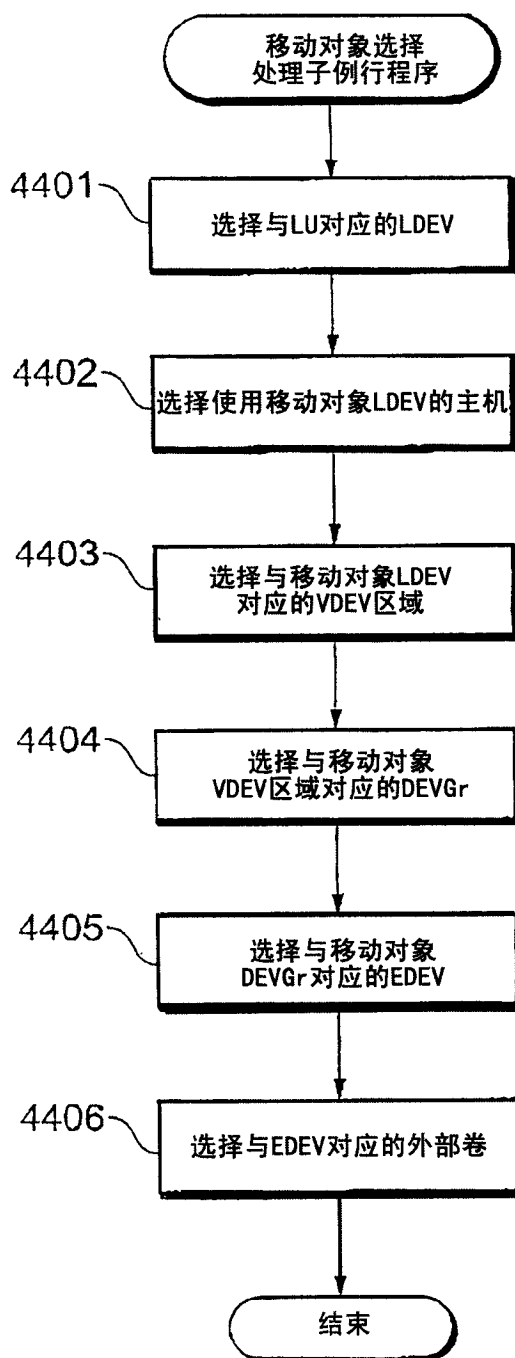


图 44



图 45

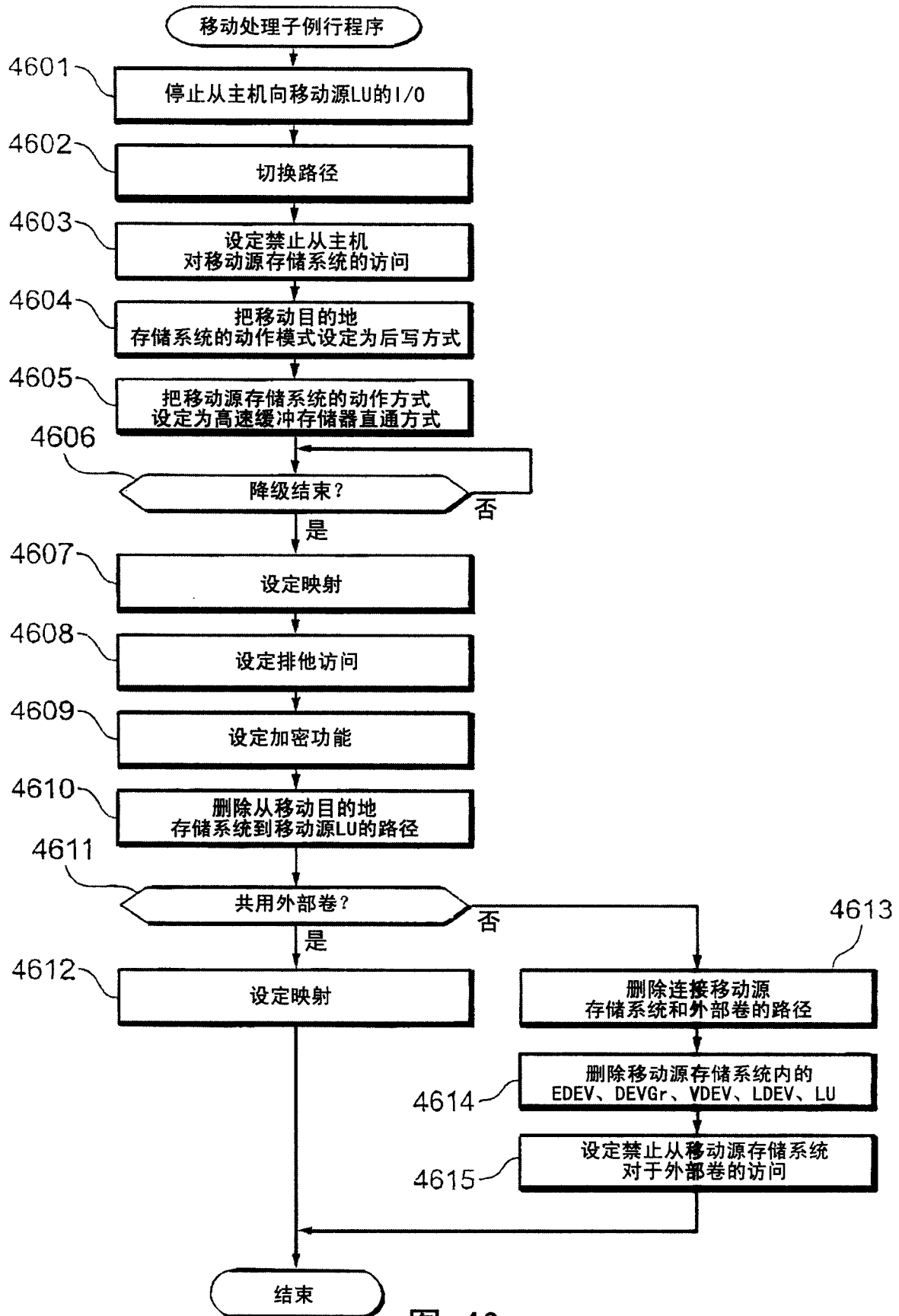


图 46

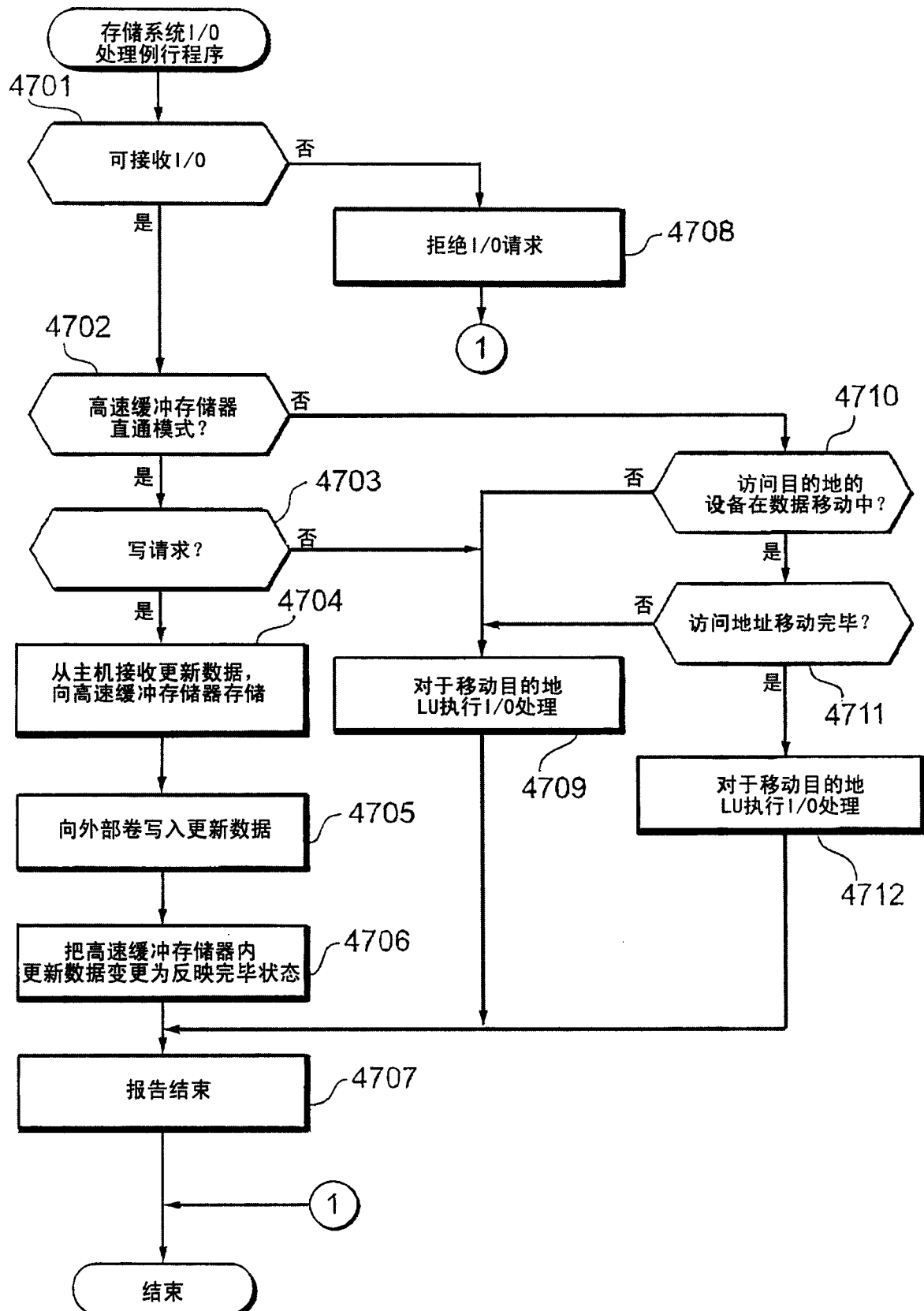


图 47

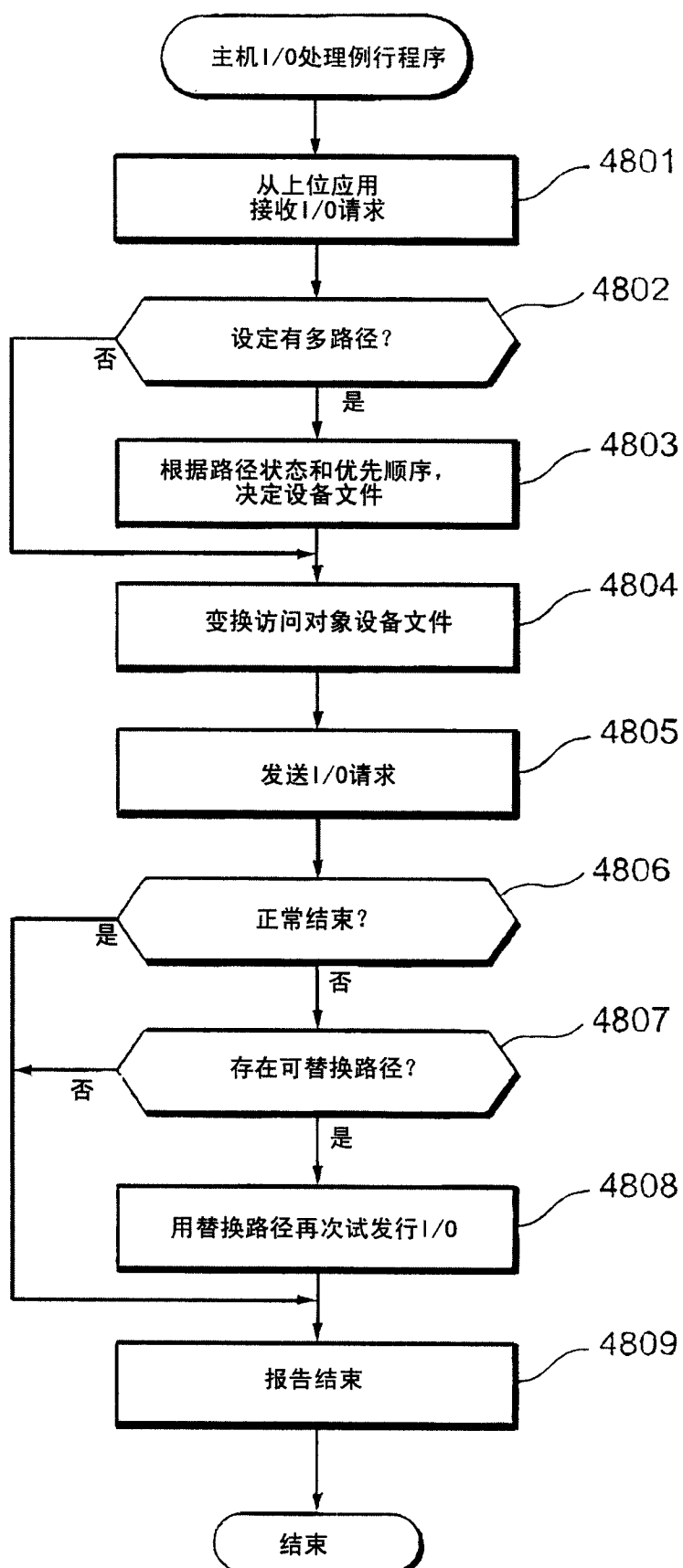


图 48

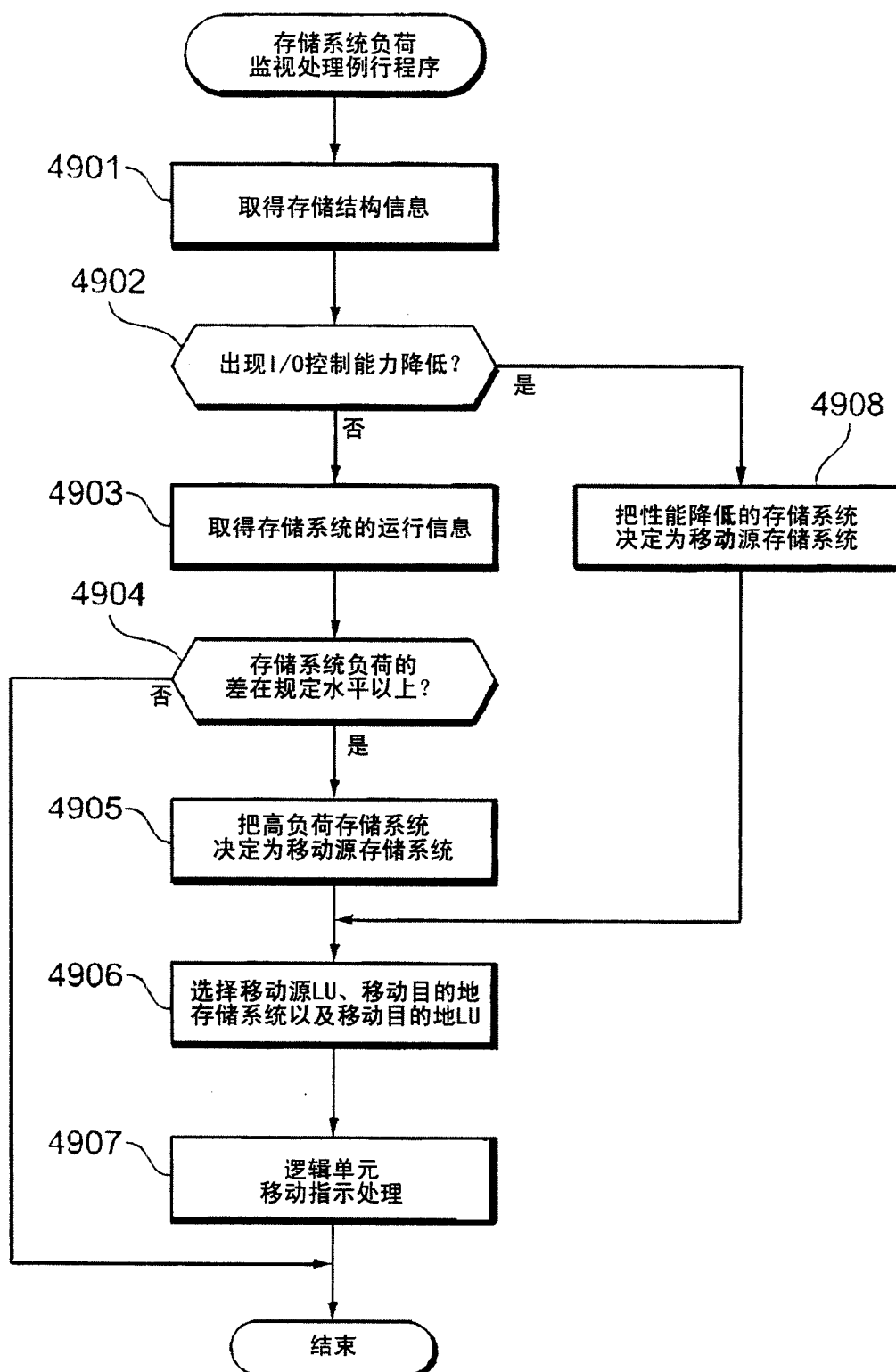


图 49

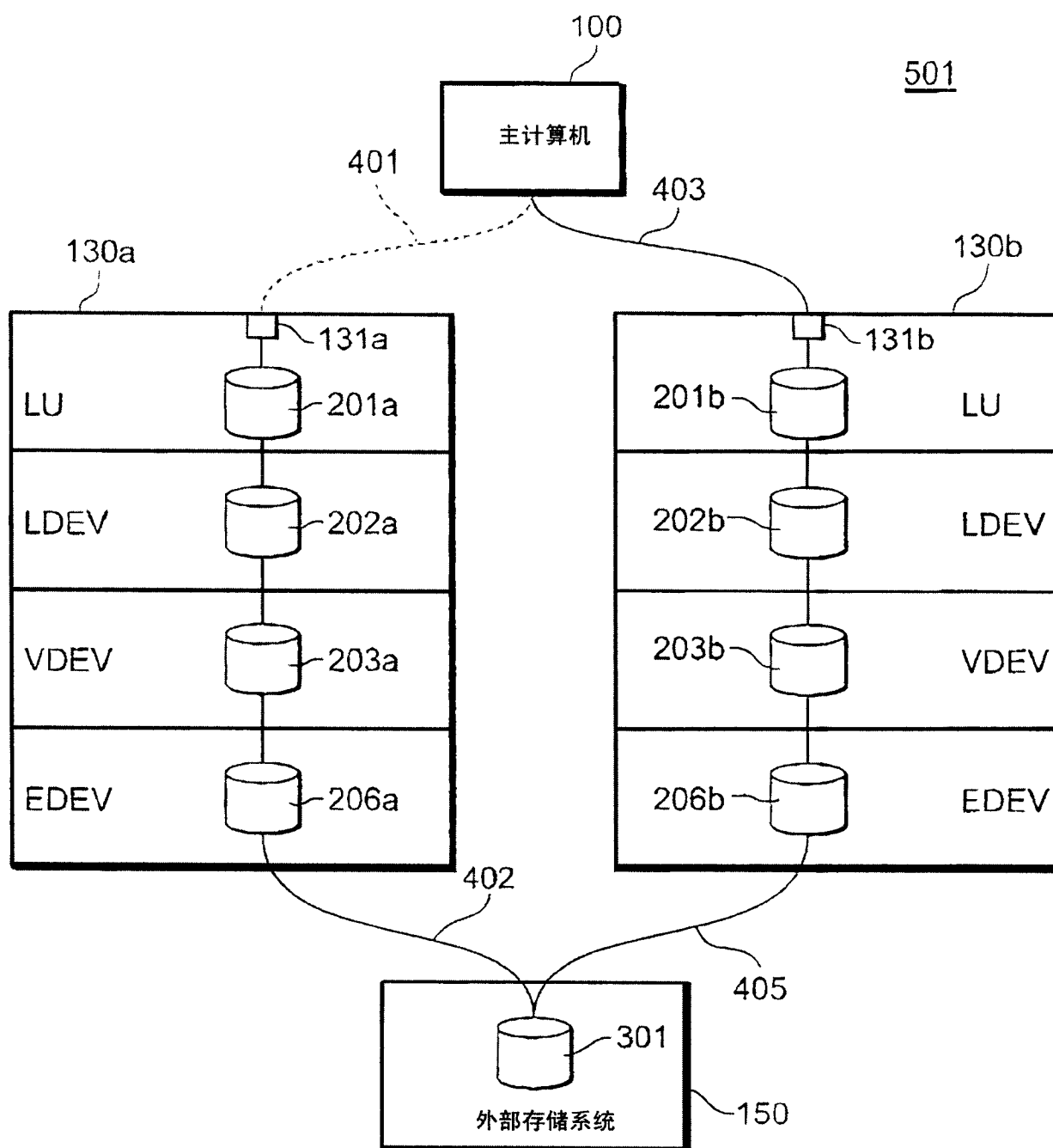


图 50

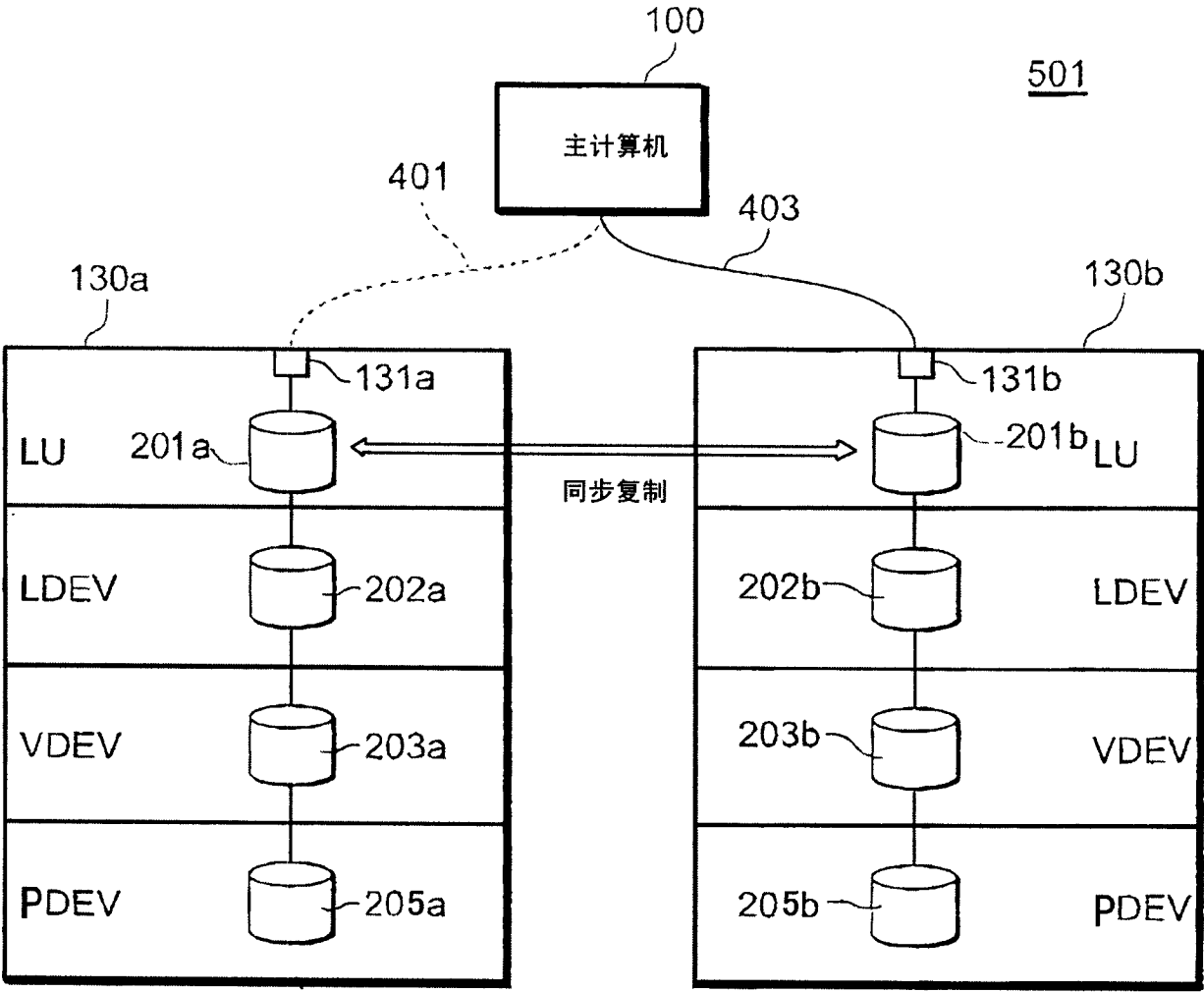


图 51

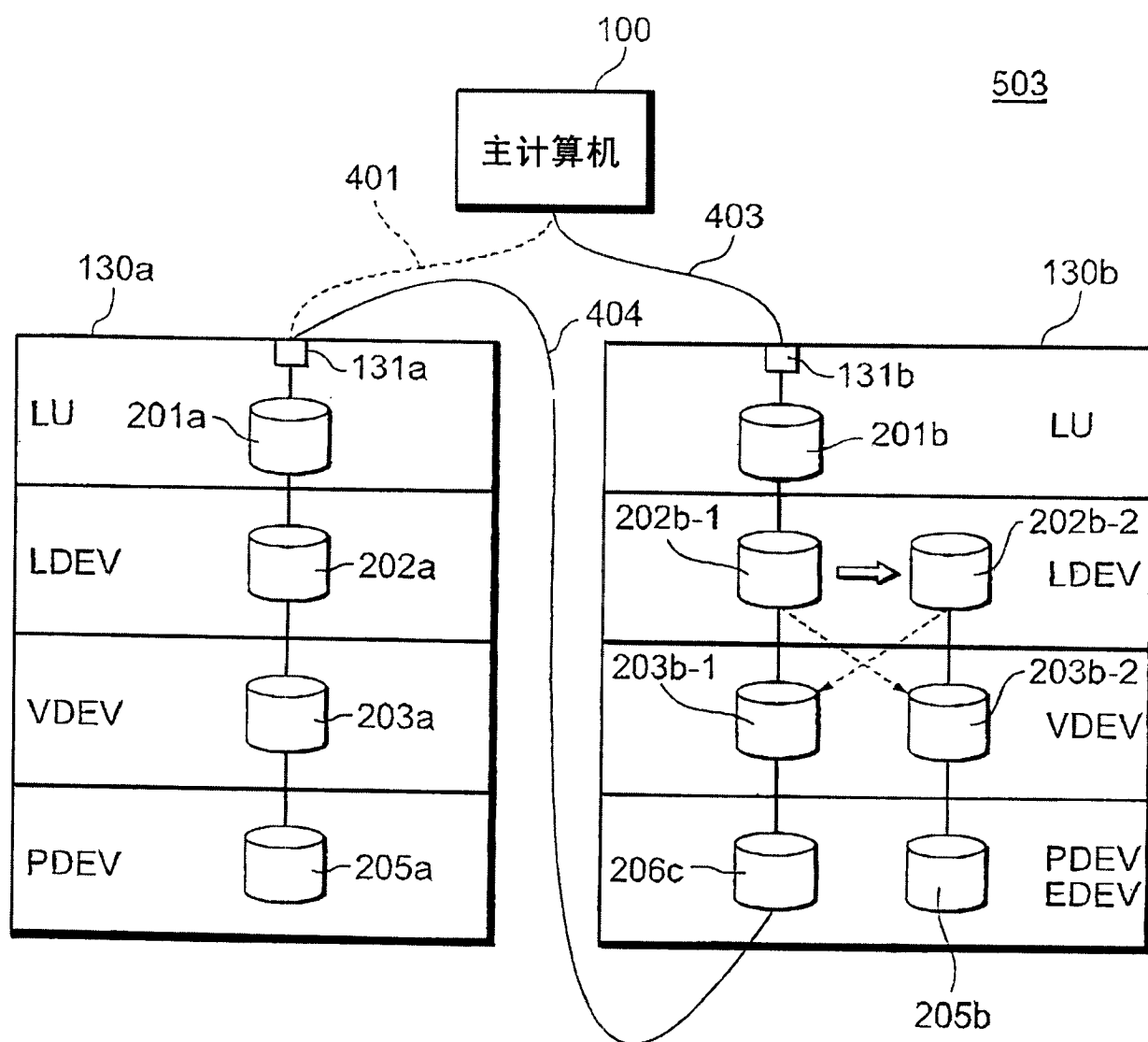


图 52

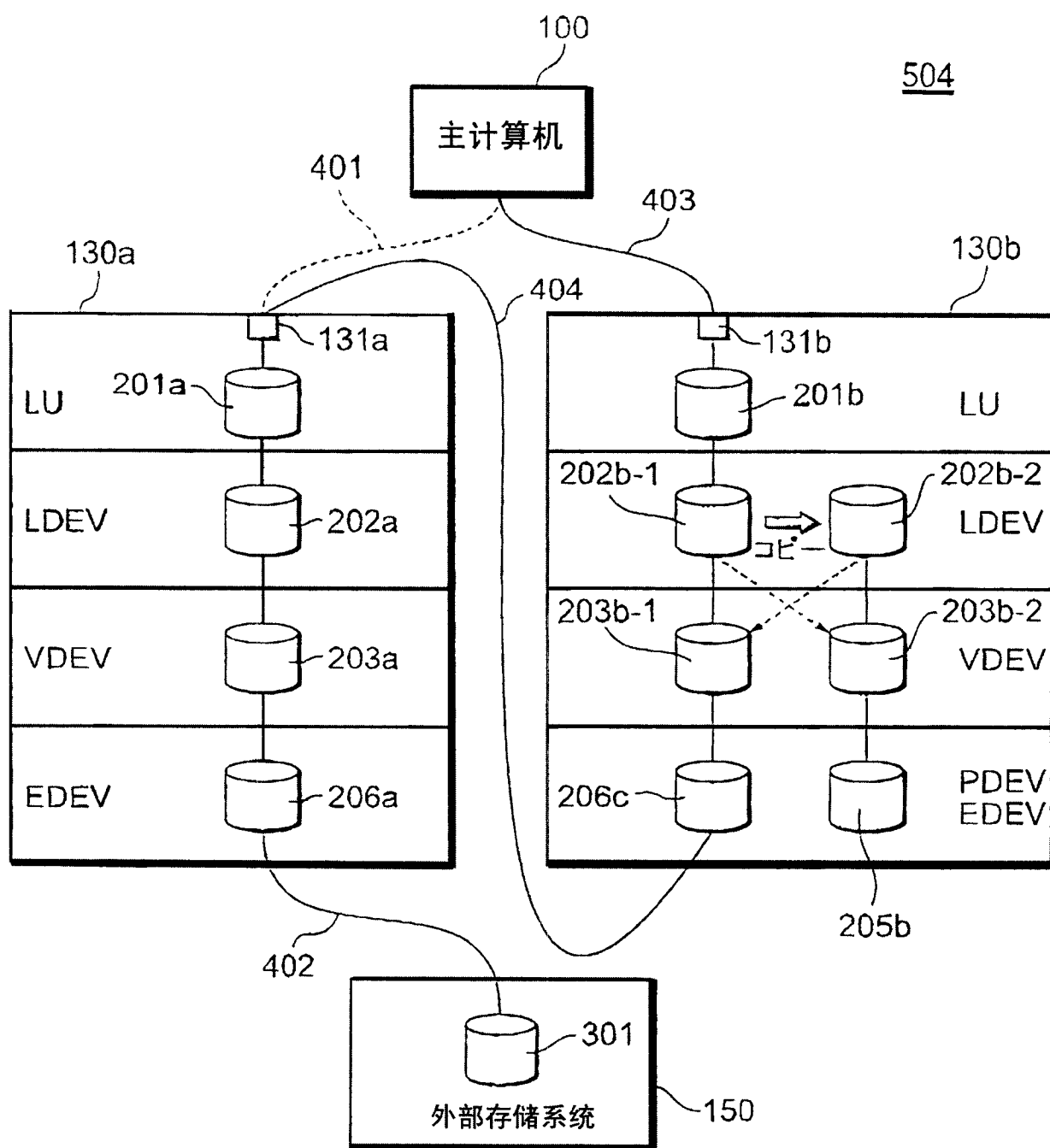


图 53

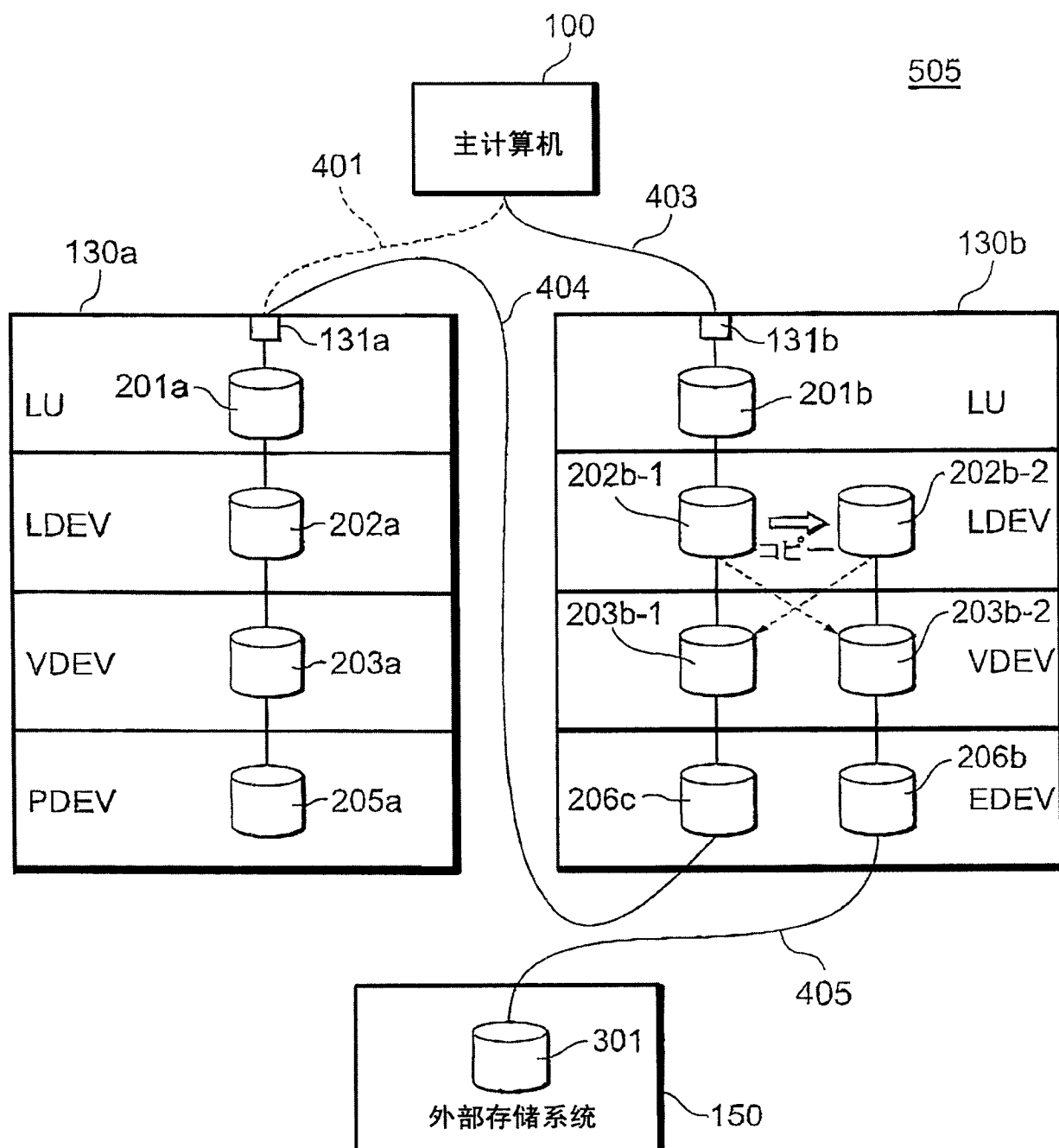


图 54

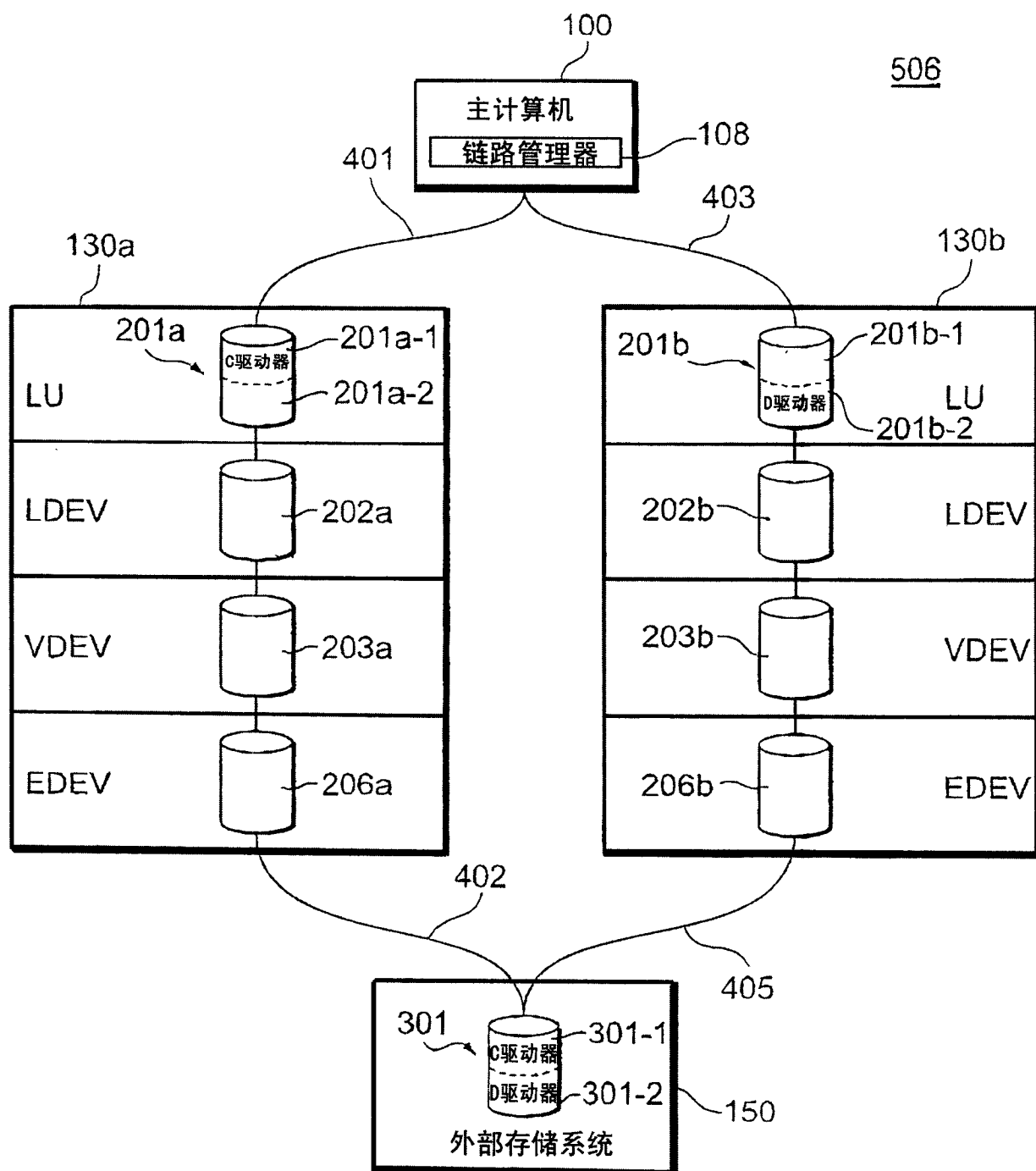


图 55

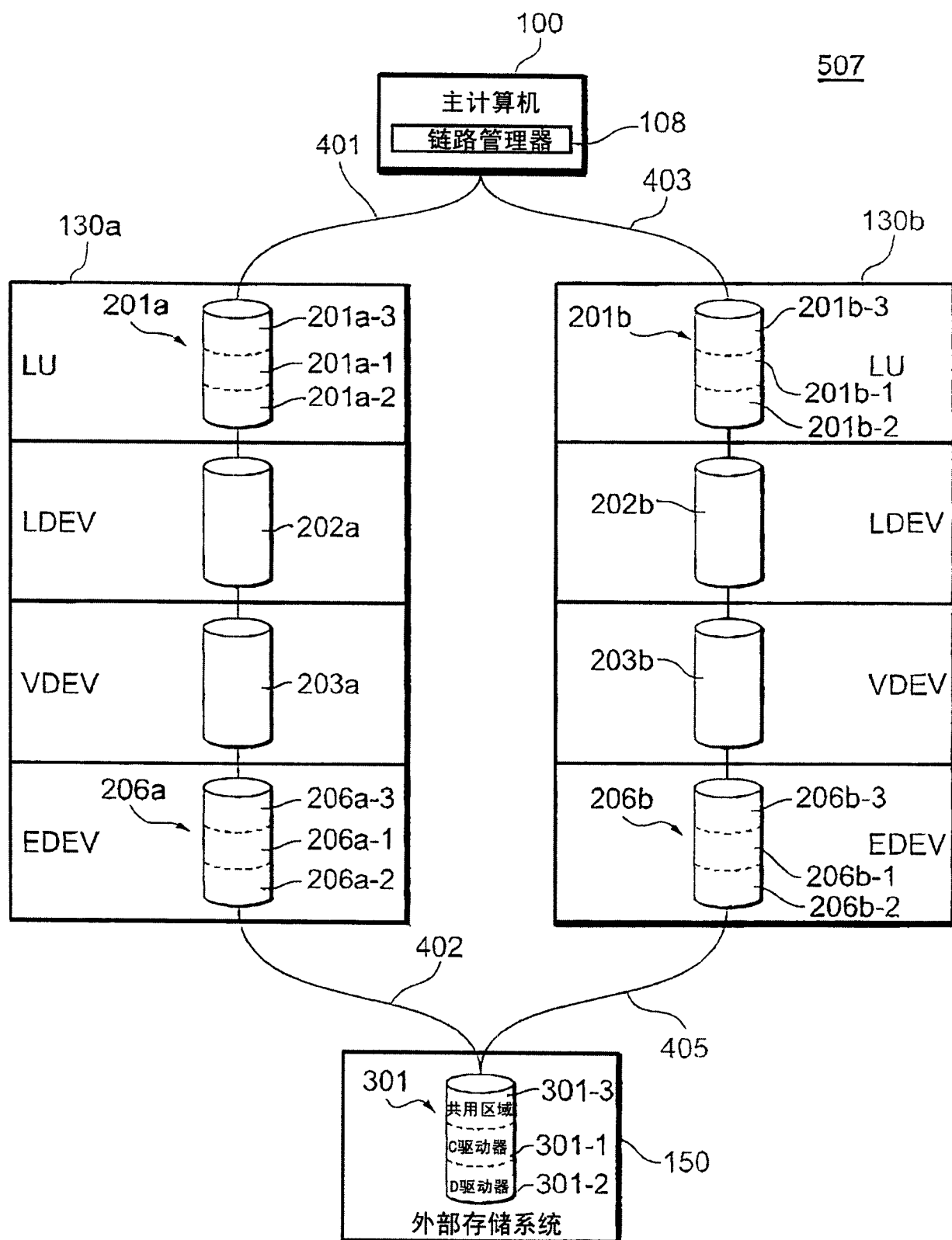


图 56

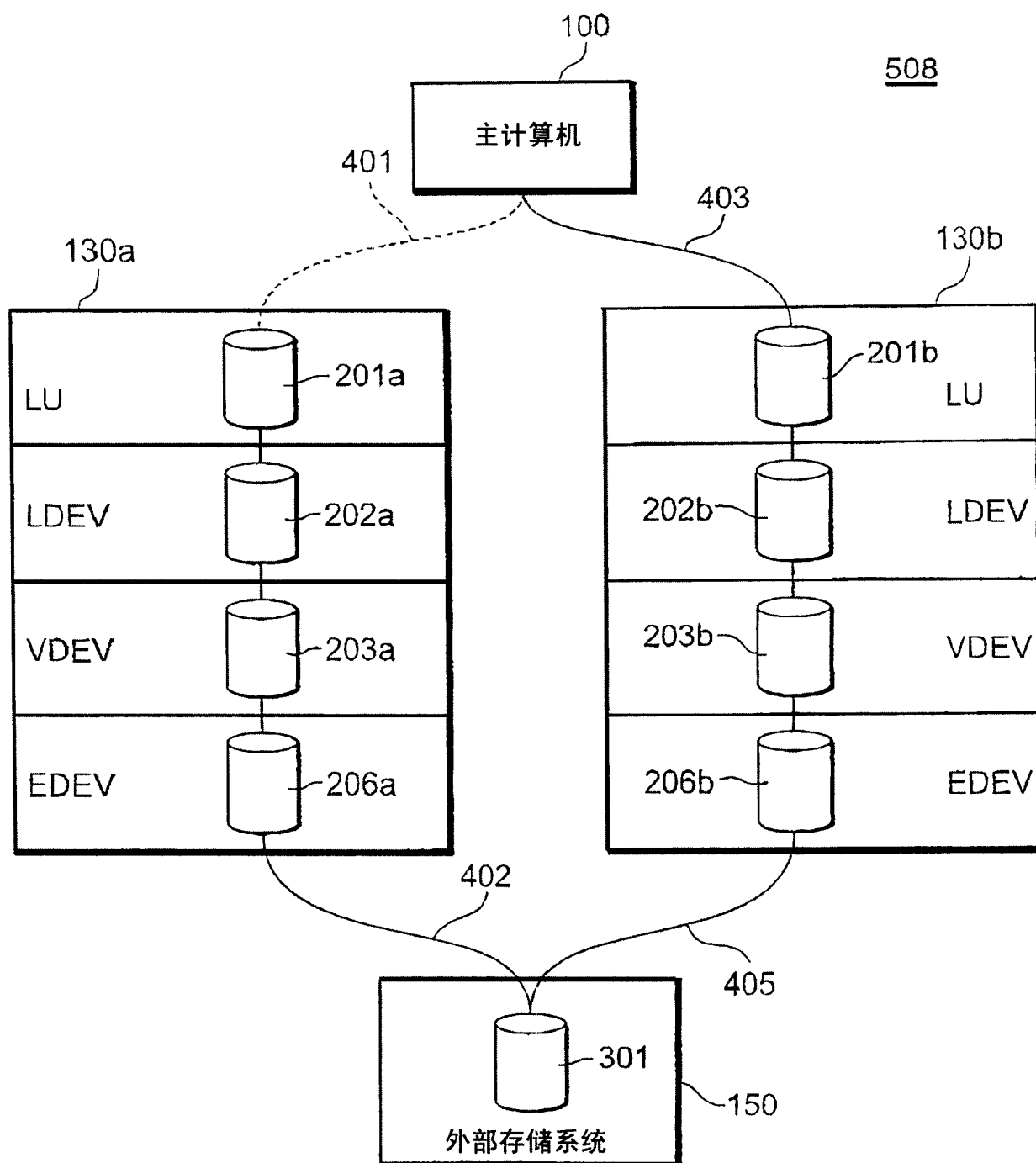


图 57

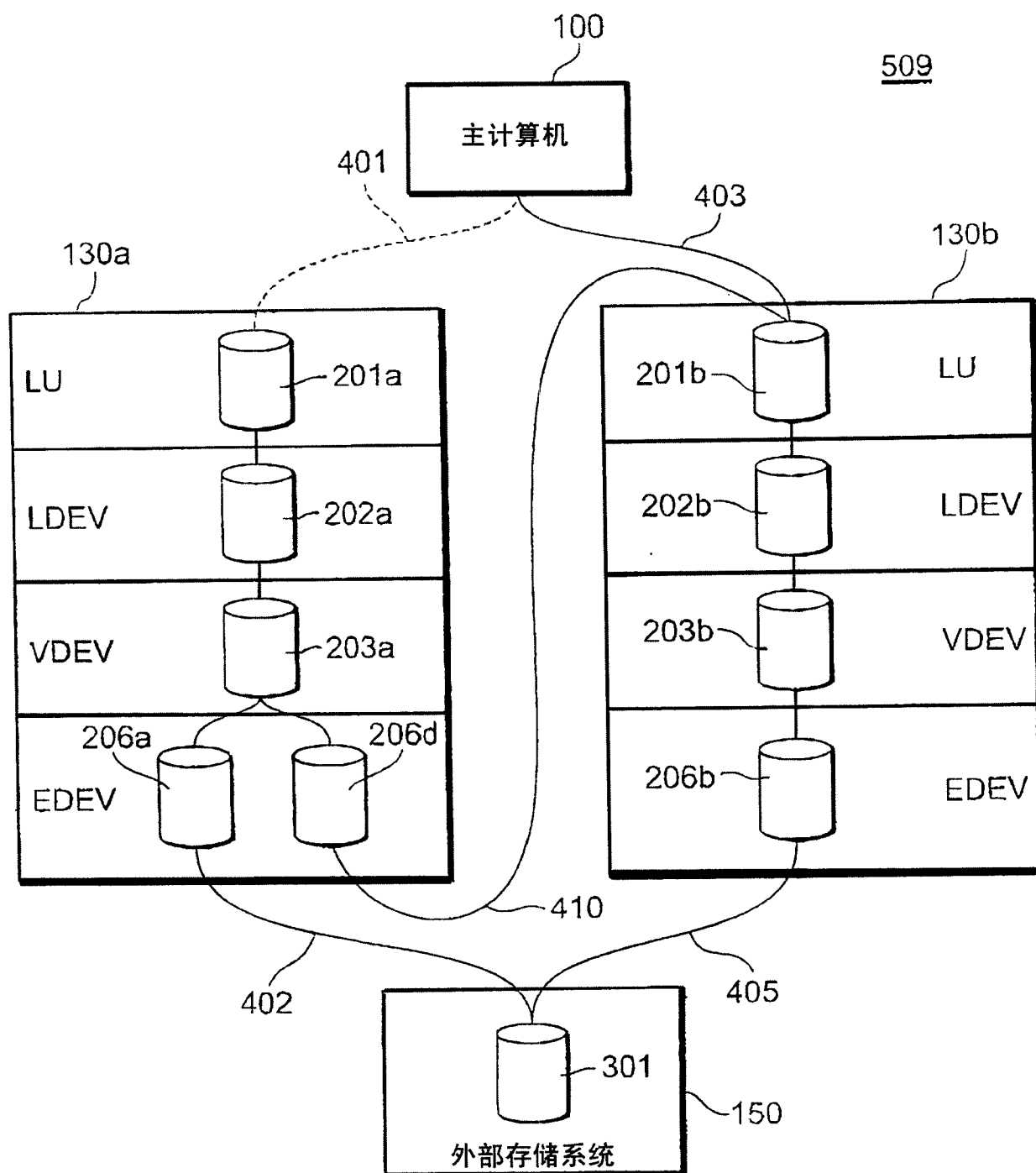


图 58

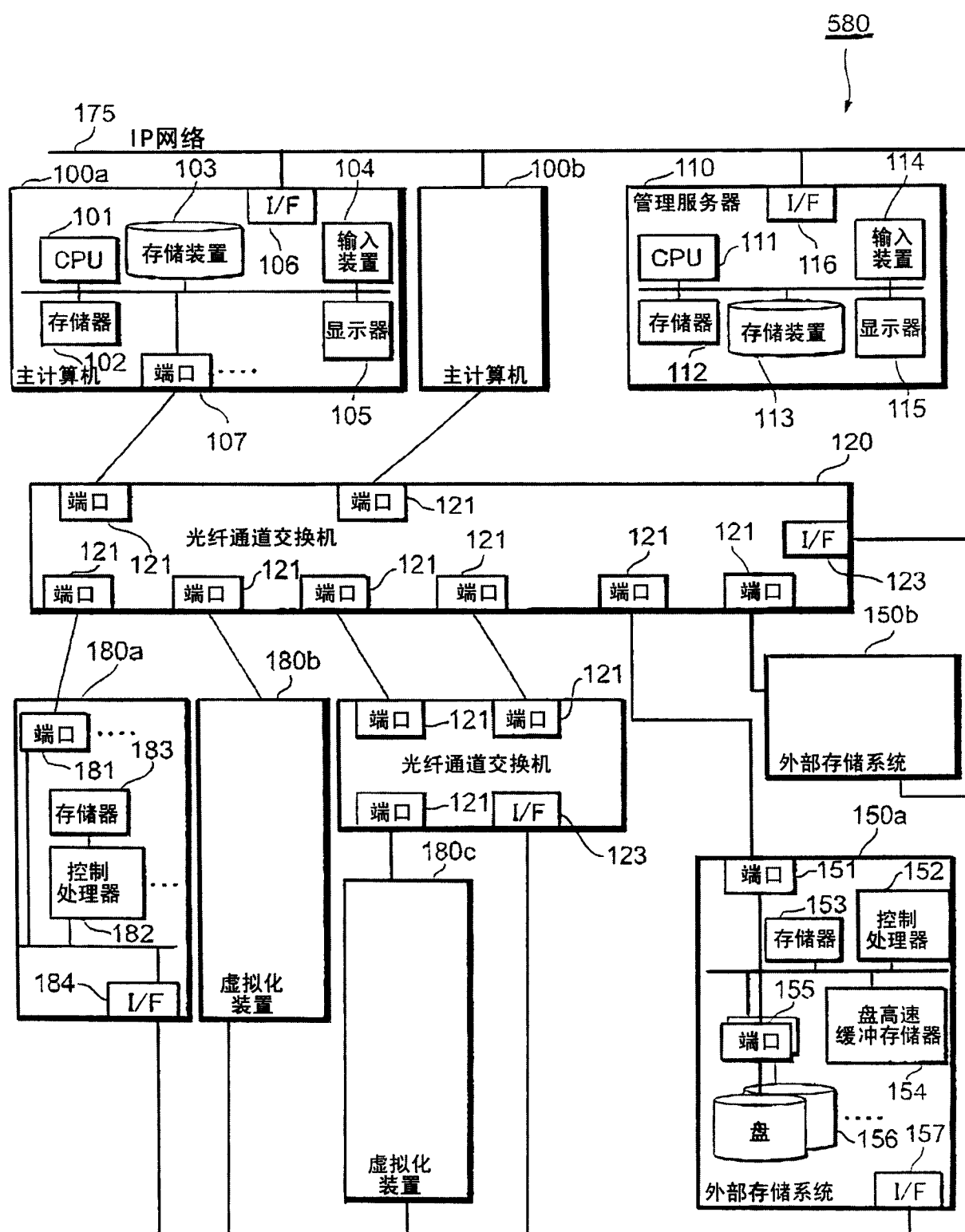


图 59