



[12] 发明专利申请公开说明书

[21] 申请号 03119804.X

[43] 公开日 2003 年 11 月 5 日

[11] 公开号 CN 1453715A

[22] 申请日 2003.2.28 [21] 申请号 03119804.X

[30] 优先权

[32] 2002.4.26 [33] JP [31] 125168/2002

[71] 申请人 株式会社日立制作所

地址 日本东京都

[72] 发明人 荒川敬史 大枝高 松井直人

伊东隆介

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 吴立明 梁永

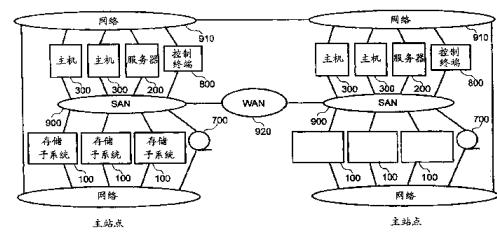
权利要求书 4 页 说明书 37 页 附图 9 页

[54] 发明名称 拥有虚拟资源的存储系统

备中的至少一个提供指令以执行请求。

[57] 摘要

配置用来向与存储系统耦合的主机提供存储区域的存储系统包括第一类型的第一存储设备，第一存储设备配置用来存储数据并提供一种或多种数据 - 存储 - 有关功能。第二类型的第二存储设备配置用来存储数据并提供一种或多种数据 - 存储 - 有关功能。虚拟卷提供存储位置并与第一和第二存储设备中的至少一个相关联。数据库包括涉及第一和第二存储设备配置用来执行的一种或多种数据 - 存储 - 有关功能的信息。第一服务器与第一和第二存储设备相耦合，并且访问数据库，第一服务器配置用来响应来自主机的数据 - 存储 - 有关功能请求访问数据库，请求涉及第一和第二存储设备中的至少一个。访问数据库是为了确定第一和第二存储设备中是否至少有一个能够执行来自主机的请求。第一服务器还被配置用来在判断结果为负时执行请求或在判断结果为正时向第一和第二存储设



1、一种配置用来向与其耦合的主机提供存储区域的存储系统，该存储系统包括：

5 配置用来存储数据并提供一种或多种数据-存储-有关功能的第一类型的第一存储设备；

配置用来存储数据并提供一种或多种数据-存储-有关功能的第二类型的第二存储设备；

10 提供存储位置并与第一和第二存储设备中的至少一个相关联的虚拟卷；

包括涉及第一和第二存储设备被配置用来执行的一个或多个数据-存储-有关功能的信息的数据库；

15 与第一和第二存储设备相耦合并且访问数据库的第一服务器，第一服务器配置用来响应来自主机的数据-存储-有关请求而访问数据库，访问请求涉及第一和第二存储设备中的至少一个，访问数据库是为了确定第一和第二存储设备中是否至少有一个能够执行来自主机的请求，第一服务器还配置用来在判断结果为负时执行请求，或者在判断结果为正时向第一和第二存储设备中的至少一个提供指令。

20 2、权利要求 1 的存储系统，其中数据库是一个信息库，并在第一服务器中提供。

3、权利要求 2 的存储系统，进一步包括：与第一和第二存储系统耦合以传输读写请求的第二服务器，其中第一服务顺是管理服务器。

25 4、权利要求 1 的存储系统，其中第一类型的第一存储设备是有存储区域和存储控制器的存储子系统，第一存储设备由第一制造商提供，第二类型的第二存储设备是有存储区域和存储控制器的存储子系统，第二存储设备由不同于第一制造商的第二制造商提供。

5、权利要求 1 的存储系统，其中第一存储设备和第二存储设备都被配置用来执行请求的功能，第一存储设备配置用来以和第二存储设备不同的方式执行请求的功能。

30 6、权利要求 5 的存储系统，其中所请求的功能是从下列功能中所选的一种：卷集中控制、快照、远程复制、复制和备份/归档，其中虚拟卷与第一和第二存储设备相关联，所请求的功能涉及第一和第二

存储设备。

7、权利要求 6 的存储系统，进一步包括：第三类型的第三存储设备，配置用来存储数据，第三存储设备能够执行下列功能：卷集中控制、快照、远程复制、复制和备份/归档。

5 8、权利要求 1 的存储系统，其中第一存储设备配置用来执行所请求的功能，第二存储设备不配置用来执行所请求的功能，其中虚拟卷与第一和第二存储设备相关联，所请求的功能涉及第一和第二存储设备。

10 9、用于管理有第一和第二存储设备以及第一服务器的存储系统的一种方法，该方法包括：

提供一个数据库，该数据库包括涉及第一和第二存储设备被配置用来提供的一种或多种数据-存储-有关功能的信息；

在第一服务器接收来自主机涉及数据-存储-有关功能的请求，请求涉及第一和第二存储设备中的至少一个；

15 确定第一和第二存储设备是否至少有一个能够通过访问数据库执行请求；以及

如果判断结果为正指令第一和第二存储设备中的至少执行请求

10、权利要求 9 的方法，进一步包括：

如果判断结果为负使第一服务器能够执行请求。

20 11、权利要求 10 的方法，其中第一和第二存储设备是第一和第二存储子系统，该方法还包括：

提供与第一和第二存储设备相关联的逻辑卷；

25 提供与逻辑卷关联的虚拟卷，以使虚拟卷与第一和第二存储设备间接关联，其中所请求的功能涉及第一和第二存储设备，其中第一存储设备被配置用来执行所请求的功能，第二存储设备不被配置用来执行所请求的功能。

12、权利要求 10 的方法，其中第一和第二存储设备是由不同制造商生产的第一和第二存储子系统。

30 13、权利要求 10 的方法，其中数据库是一个信息库，是在第一服务器中提供的。

14、权利要求 10 的方法，其中第一和第二存储设备都可配置用来执行请求的功能，第一存储设备被配置用来以和第二存储设备不同

的方式执行请求的功能。

15、权利要求 14 的方法，其中所请求的功能是从下列功能中选出的一种：卷集中控制、快照、远程复制、复制和备份/归档。

5 16、权利要求 10 的方法，其中第一存储设备可配置用来执行所请求的功能，第二存储设备不被配置用来执行所请求的功能。

17、权利要求 10 的方法，其中确定步骤进一步包括：

确定第一和第二存储设备中是否至少有一个能够执行所请求的功能来满足该请求；并且

确定所请求的功能是否对第一和第二存储设备的至少一个可用。

10 18、用于管理有第一和第二存储设备及第一服务器的一种计算机可读介质，该介质包括：

在第一服务器接收来自主机涉及数据-存储-有关功能的请求的代码，请求涉及第一和第二存储设备中的至少一个；

15 确定第一和第二存储设备中是否有至少一个能够通过访问数据库执行请求的代码，数据库包括涉及第一和第二存储设备被配置用来提供的一种或多种数据-存储-有关功能的信息；

如果判断结果为正指令第一和第二存储设备中的至少一个执行请求的代码；和

如果判断为负使第一服务器执行请求的代码。

20 19、用于管理存储系统的一种服务器，其中服务器耦合到第一和第二存储设备以及主机，该服务器访问一个数据库，数据库包括涉及第一和第二存储设备被配置用来提供的一种或多种数据-存储-有关功能的信息，该服务器包括：

处理信息的处理器；

25 与网络耦合的接口，网络与第一和第二存储设备以及主机相耦合；和

一种计算机可读介质，包括：

在第一服务器接收来自主机涉及数据-存储-有关功能的请求的代码，请求涉及第一和第二存储设备中的至少一个；

30 确定第一和第二存储设备中是否有至少一个能够通过访问数据库执行请求的代码，以及

如果判断结果为正指令第一和第二存储设备中的至少一个执行请

求的代码，和

如果判断为负使第一服务器执行请求的代码。

20、权利要求 19 的服务器，进一步包括：

存储信息的存储位置，其中数据库存储在存储位置中，其中存储系统包括与第一和第二存储设备关联的逻辑卷以及与逻辑卷关联的虚拟卷，以使虚拟卷与第一和第二存储设备间接关闻，其中所请求的功能涉及第一和第二存储设备，其中第一存储设备可配置用来执行所请求的功能，第二存储设备不被配置用来执行请求的功能。

拥有虚拟资源的存储系统

5 技术领域

本发明涉及一种控制存储系统和存储子系统的方法，尤其涉及控制虚拟化资源的存储系统及其中设备的方法。

背景技术

在信息处理系统中，计算机中的处理可能会由于计算机中执行的程序缺陷（bug）而导致异常中断，或者类似的情况，使得信息处理系统中的数据处于一种冲突状态。另外，信息处理系统所管理的数据也可能由于人为的错误而被毁掉。

为了可靠地存储处于这样一种状态的信息处理系统的数据，信息处理系统通常在程序基础上获得数据的备份。日本专利申请公开的 15 2000-132343 号描述了一个创建静态映像副本（快照卷 snapshot volume）的存储子系统（维持它的数据一致性）以便在不妨碍备份对数据的访问的情况下获得备份。

根据该文档中描述的快照创建方法，数据在由主计算机指定的源卷和目标卷之间进行复制，像使数据静止的复制停止（一个卷对的分离）这样的控制和数据再同步是以卷为基础执行的。这样的快照卷不仅用于获得备份，还用于数据移动和数据共享的实现，并在其中维持数据库之间的一致性。

另外，日本专利公开的 Hei 9-274544 号描述了一个把主机访问的逻辑卷之间的存储区域（地址变换）和实际存储逻辑卷中的数据的物理存储设备关联起来处理访问，在该存储子系统中有所获得的与从主机到逻辑卷的访问有关的信息，还可以改变这种关联来执行物理再定位以便实现对所存储的数据的最优化分配。

此外，在日本专利公开的 Hei 10-333839 号和 2000-276406 号中，描述了具有下列特征的存储子系统：当从主机访问存储子系统时，它使用 WWN（WORLD WIDE NAME）来唯一标识主机的一个光纤信道接口（以下称为端口）来判断是否允许从主机对存储子系统的逻辑卷的访问；如果端口没有获得许可，就拒绝对该端口的访问。

近些年来，信息处理系统的问题是因为存储设备和存储区域中所用的数据大幅增加对存储设备和存储区域的管理开销也增加了。根据这一点，由研究公司 Morgan Keegan 报告的“Virtualizing the SAN”（2000 年 7 月 5 日）描述了一个根据来自主机的请求虚拟化逻辑卷的系统，逻辑卷由存储子系统提供，该系统还向主机提供已经被虚拟化的虚拟卷。
5

在这个报告中，描述了几种实现虚拟化（virtualization）的系统配置。

一种具有如下特征的系统：为在主机和存储子系统之间实现虚拟化而连接计算机；通过使用这个计算机向主机提供由一个或多个卷组成的区域作为虚拟卷，用于转换从主机到虚拟卷的访问为到对应的逻辑卷的访问。
10

就另一个实例来说，主机与存储子系统相连；把一个存储区域作为一个虚拟卷来管理存储区域的联合为一个逻辑卷的计算机连接到主机。
15 当主机访问虚拟卷时，主机向该计算机发送与虚拟卷上要访问的数据的位置有关的查询。计算机返回与虚拟卷对应的逻辑卷的位置给主机。主机用接收到的与逻辑卷的位置有关的信息访问由存储子系统提供的卷。

像这里所用的那样，术语“存储子系统”指包括配置用来处理信息的一个或多个存储控制器以及配置用来存储数据的一个或多个存储部件（例如，磁盘、磁带等）的存储装置或设备。存储子系统的实例包括磁盘阵列单元。
20

像这里所用的那样，术语“存储设备”或“存储单元”指配置用来在其中存储数据的任意存储装置。存储设备的实例包括存储子系统、硬盘驱动器、CD-ROM 和磁带。
25

像这里所用的那样，术语“存储系统”指包括与一个或多个信息处理单元相连的一个或多个存储设备。存储系统的实例包括存储区域网络系统、网络区域存储系统和存储子系统。

30 发明内容

当主机使用大量存储区域和大量逻辑卷时，在内部控制大量存储区域和大量逻辑卷，以使各个控制与其它控制相关联，这样的配置能

够激活高级系统操作和管理，导致处理所需的时间减少，主机的处理负载减轻，而且易于管理。但是，传统的控制方法和管理方法中存储子系统的功能是用主机和管理服务器上的软件（管理软件）一个个单独管理的，因而无法实现存储子系统间功能的联合。

5 特别地，如果在多个功能之间存在差异（即，如果存储子系统中所包括的功能的种类与其它存储子系统中所包括的功能的种类不同）并且即使一个功能的种类与其它存储系统中的一种功能的种类相同或相近其功能规范也不同，以及如果一个特定的存储子系统没有某个特定功能或相似的功能，主机就无法用一个统一的模式向每个存储子系统中的存储区域和卷施加该功能而且也无法实现存储子系统间功能的联合。

10 另外，如果虚拟化是在信息处理系统中执行的，例如主机所用的一个虚拟卷可以扩展到多个存储子系统或存储设备。但是，在现有技术中，主机不能用虚拟卷这样的统一和联合的方式使用存储子系统的功能。

15 此外，至于涉及到传统存储子系统中所用的控制方法和管理方法，不可能实现像性能调节这样的操作和管理，在性能调节中由存储子系统提供的功能与不同的功能相关联。

20 此外，在现有技术中，以主机上和主机设备配置有关的信息、与应用软件有关的信息、数据库管理系统、中间件、文件系统或类似系统为基础无法控制大量存储子系统间功能的联合以及每个功能与其它功能的联合。

25 另一方面，在执行虚拟化的信息处理中，可以考虑下列方法：由执行虚拟化处理的计算机（以下称为虚拟化服务器）创建存储子系统的快照或副本；向主机提供上面所描述的功能，例如减轻主机和网络负载、加强系统管理以及确保安全。但是，尽管这种方法允许用统一的方式提供虚拟卷的功能，但虚拟化服务器的处理增加了。这种情况下，虚拟化服务器挪存储区域的虚拟化处理和从主机到虚拟卷的访问的处理。因此，由功能处理引起的负载增加妨碍了对来自主机的访问的处理，导致 I/O 性能下降，I/O 性能下降又引起信息处理系统的处理系统的下降。

另外，由虚拟化服务器创建存储子系统的快照和副本增加了连接

虚拟化服务器和存储子系统的网络上的传输负载和处理负载，还增加了主要几存储子系统之间的 I/O 接口单元（连接到网络的接口单元）上的传输负载和处理负载。如果网络和接口单元还用于处理来自主机的访问，同样也会妨碍来自主机的访问，导致 I/O 性能的下降，I/O 性能下降又引起信息处理系统的处理性能的下降。
5

本发明的一种实施方案涉及一种计算机系统，该计算机系统能够调整拥有多种功能的存储子系统的功能之间的差异和功能规格之间的差异并且能够以一种使每种功能都能与其它功能相关联的方式从整体上使用多个存储子系统的功能。

10 本发明的实施方案与一种计算机系统有关，该计算机系统操作和管理存储子系统使得存储子系统中的大量功能都从整体上相互关联；该计算机系统能够以与应用软件有关的信息、中间件和主机为基础控制大量存储子系统间功能的联合以及各个功能与其它功能的联合。该计算机系统能够减轻信息处理系统中执行存储区域的虚拟化的虚拟化
15 服务器的负载；该计算机系统能够以虚拟卷为基础控制存储子系统中包括的功能。

20 在一种实施方案中，一种配置用来向与存储系统耦合的主机提供存储区域的存储系统包括配置用来存储数据并提供与一个或多个数据存储有关的功能的第一类型的第一个存储设备。第二类型的第二个存储设备可配置用来存储数据并提供一个或多个数据-存储-有关功能。
25 虚拟卷提供存储位置并且与第一和第二存储设备中的至少一个相关联。数据库包括与第一和第二存储设备配置用来执行的一个或多个数据-存储-有关功能有关的信息。第一服务器与第一和第二存储设备耦合并且访问数据库，第一服务器配置用来响应来自主机的数据-存储-有关功能请求而访问数据库，主机包含第一和第二存储设备中的至少一个；访问数据库是为了确定第一和第二存储设备中是否至少有一个能够执行来自主机的请求。如果判断为负或第一服务器还可以配置用来处理请求，如果判断为正第一服务器可配置用来向第一和第二设备中的至少一个提供处理请求的指令。

30 在一种实施方案中，一种管理有第一和第二存储设备以及第一服务器的存储系统的方法包括：提供包括涉及一种或多种数据-存储-有关功能的信息的数据库，这些数据-存储有关的功能是由第一和第二存

5 储设备被配置用来提供的；在第一服务器接收来自主机涉及数据-存储-有关功能的请求，请求涉及第一和第二存储设备中的至少一个；确定第一和第二存储设备中是否至少有一个能够通过访问数据库处理请求；并在判断结果为正时指令第一和第二存储设备中的至少一个处理请求。

10 在另一种实施方案中，一种用于管理存储系统（该存储系统包括第一和第二个存储设备和第一服务器）的计算机可读介质包括在第一服务器接收来自主机涉及数据-存储-有关功能的请求，该请求涉及第一和第二存储设备中的至少一个；用于确定第一和第二存储设备中是否至少有一个能够通过访问数据库处理请求的代码，数据库包括涉及一种或多种数据-存储-有关功能的信息，这些功能是由第一和第二存储设备配置提供的；如果判断结果为正时命令第一和第二存储设备中的至少一个处理请求的代码；判断结果为负时激活第一服务器来处理请求的代码。

15 在另一种实施方案中，用于管理存储系统的服务器包括处理信息的处理器，其中服务器与第一和第二存储设备以及主机相耦合，服务器访问包括涉及一种或多种数据-存储-有关功能的信息的数据库，这些功能是由第一和第二存储设备配置提供的；与连接到第一和第二存储设备以及主机的网络相连的接口；以及计算机可读介质。该计算机
20 介质包括在第一服务器接收涉及来自主机涉及数据-存储-有关功能的请求的代码，该请求涉及第一和第二存储设备中的至少一个，确定第一和第二存储设备中是否至少有一个能够通过访问数据库处理请求的代码，判断结果为正时命令第一和第二存储设备中的至少一个处理请求的代码，以及判断结果为负时激活第一服务器来处理请求的代码。

25

附图概述

图 1 依照本发明的一种实施方案描述有主站点和次站点的网络或存储系统的配置；

图 2 描述图 1 的存储系统的主站点或次站点之一的配置；

30 图 3 依照本发明的一种实施方案描述存储系统中的存储子系统所提供的功能的实例以及存储子系统和服务器之间的管理相互关系。

图 4 依照本发明的一种实施方案描述服务器中提供的功能的实例

以及存储子系统、服务器和主机之间的相互关系；

图 5 是依照本发明的一种实施方案执行卷集中控制（pool）功能的方法；

图 6 依照本发明的一种实施方案描述涉及虚拟卷和存储设备的关联的地址信息；

图 7 依照本发明的一种实施方案描述提供给主机和虚拟存储区域的路径地址和逻辑单元号；

图 8 依照本发明的一种实施方案描述用来访问存储设备中的存储区域的路径地址和逻辑单元号的关联；

图 9 依照本发明的一种实施方案描述执行备份和归档的方法；

图 10 依照本发明的一种实施方案描述创建快照的方法；

图 11 依照本发明的一种实施方案描述与复制功能有关的信息；

图 12 依照本发明的一种实施方案描述与快照功能有关的信息；

图 13 依照本发明的一种实施方案描述与远程复制功能有关的信息；

具体实施方式

图 1 是依照本发明的一种实施方案描述存储系统配置的示意图。该存储系统包括主站点和次站点，它们通过 WAN（广域网）920 互连。

每个站点包括一个或多个主计算机（以下称为主机）300、一个或多个计算机（以下称为服务器）、一个或多个备份存储设备 700，以及一个或多个存储子系统 100，它们通过 SAN（存储域网络）900 相互连接，SAN 900 包括交换机 400 和集线器。SAN 900 中所用的协议和标准包括光纤信道（FC）、IP、Infini 和 Band。在下面的描述中将作为实例描述按照 FC 配置的 SAN 900。

主机 300、服务器 200 和存储子系统 100 也通过网络 910 互连。网络 910 中所用的协议包括 IP。

图 2 描述每个站点中服务器和存储子系统的配置。服务器 200、主机 300、存储子系统 100、备份存储设备 700 和控制终端 800 也通过网络 910 互连。另外，服务器 100、主机 200、存储设备 300、备份存储设备 700 以及控制终端 800 以通过 SAN 900 互连。更明确地说，SAN 900 包括通信线路和多个交换机 400。

主机 300 包括中央处理单元 (CPU)、存储器和存储设备 (未显示)。主机 300 存储应用软件 310、OS 320、访问处理软件 330 和存储设备中的代理 340。这些程序被读入存储器，然后由 CPU 执行。

5 应用程序 310 包括备份软件。另外，应用软件 310 也包括像数据库管理软件 (DBMS) 和聚类软件这样的中间件。访问处理软件 330 包括文件系统、设备驱动以及类似的软件。

备份存储设备 700 是存储备份子系统 100 中所存储的数据的备份数据的存储设备。备份存储设备 700 的存储介质包括磁带、磁带库、磁盘、磁-光盘和类似的介质。

10 存储子系统 100 包括控制单元 (存储控制器) 110、存储器 112 和磁盘单元 550。存储子系统 100 根据来自服务器 200 或主机 300 的请求记录并复制主机 300 所用的数据。存储子系统 100 中包括的磁盘单元 550 的数量没有特别限制。为了更明确些，存储子系统的实例包括单一磁盘单元、有多个存储单元 550 的集群磁盘单元 (JBOD)、用多个磁盘单元 550 配置成 RAID 的磁盘阵列以及类似磁盘单元的存储设备。存储子系统 100 的控制单元 110 向服务器 200 和主机 300 提供逻辑卷 500 作为数据存储区域。存储子系统 100 联合服务器 200 或主机 300 所访问的逻辑卷 500 和实际存储逻辑 500 中所存储的数据的磁盘单元 550 的存储区域 (即，地址转换)。一个逻辑卷可以与单个磁盘单元或存储设备关联，或者与多个磁盘单元或存储设备关联。

15

逻辑卷依次与多个虚拟卷 600 相关联。一个虚拟卷可以与单个存储设备或存储子系统或者其中的多个相关联。另外，一个虚拟卷可以与单个或多个逻辑卷相关联。作为选择，多个虚拟卷也可以与单个逻辑卷相关联。

25 控制单元 110 保存上述地址转换处理所需的地址转换信息以及后面要描述的各自功能的处理所需的信息。使用上述地址转换，存储子系统 100 允许外部设备把多个磁盘单元 550 所拥有的存储区域看作一个或多个逻辑卷 500。存储器 112 存储执行不同功能的计算机可读代码，例如卷定义和远程复制，后面将对这些功能进行详细描述。另外，存储子系统 100 有高速缓冲存储器 120。

30

服务器 200、主机 300 和控制终端 800 都有 CPU、存储器以及类似的部件，这些部件用于执行每个计算机中所包括的程序。

5 服务器 200 包括与网络 910 耦合的第一接口 202、与 SAN 200 耦合的第二接口、处理信息的控制单元或操作单元 206、提供数据库或信息库 210 的存储器 220 以及存储数据的存储设备 208。在一种实施方案中，信息库 210 存储在存储设备 208 中。服务器 200 虚拟化由存储子系统 100 提供的逻辑卷 500 以创建虚拟卷 600，然后向主机 300 提供这个虚拟卷 600 作为存储区域。虚拟化是在服务器 200 的 CPU 执行控制器程序或元文件系统时实现的，后面将描述控制器程序和元文件系统。

10 存储子系统所执行的功能

15 图 3 描述存储子系统 100 中所包括的功能的实例以及存储子系统 100 和服务器 200 之间的管理相互关系。服务器 200 管理多个存储子系统 100。此外，服务器 200 管理多个存储子系统 100 的逻辑卷 500 和存储子系统中所包括的功能并在虚拟化这些功能后向主机 300 提供这些功能。

20 站点通常包括多个存储子系统 100，存储子系统 100 可以由多个不同的制造商生产。因此，不同的存储子系统可以提供不同的功能并提供执行相同功能（例如远程复制和快照）的不同方式。站点还可以包括原始存储设备（未显示），这种存储设备限制信息处理能力为根据更复杂设备（例如，服务器或主机）的命令读出或写入数据。

下面描述存储子系统 100 中所包括的功能。注意每种功能都是由存储子系统 100 的控制单元 110 执行与该功能相关的程序来实现的。

(A) 卷的定义

25 卷定义功能是在存储子系统 100 中创建逻辑卷 500 的功能。为了更明确些，逻辑卷 500 与磁盘单元 600 中的存储区域相关联。卷定义功能包括把逻辑卷 500 分成几个部分来定义新的逻辑卷 500 的功能，每个部分大小任意。此外，卷定义功能还包括组合多个逻辑卷 500 以定义更大的逻辑卷 500 的功能。

30 (B) 卷管理和安全协议

存储子系统 100 有一个或多个接口（物理端口）通过一个端口与 SAN 900 相连。卷管理功能是分配存储子系统 100 中所包括的物理端

口和逻辑单元号 (LUN) 给由卷定义功能创建的逻辑卷 500 的功能。服务器 200 和主机 300 指定 WWN 和 LUN 来访问逻辑卷 500, WWN 和 LUN 被用作分配给 SAN 900 上的一个物理端口的地址。

5 有多种分配地址的方法, 地址指示物理端口的位置。无论如何, 分配地址是使物理端口能够被唯一标识。

如图 4 所示, 这个功能包括把主机 300 和服务器 200 分开并以群组为基础独立地分配各个主机 300 和各个服务器 200 所访问的逻辑卷 500 的 LUN 的功能, 主机 300 和服务器 200 使用一个端口进入多个组群。说得更明确些, 一个物理端口的相同 LUN 可以被分配给不同群组 10 的不同逻辑卷 500。

因此, 存储子系统 100 允许主机 300 等把一个物理端口看作多个逻辑端口, 并且能够给每个群组提供一个逻辑端口。为了实现这个功能, 存储子系统 100 保存与属于每个群组的主机 300 和服务器 200 所拥有的 FC 接口 (以下称为主机端口) 的地址 (WWN) 有关的信息。说得更明确些, 存储子系统 100 保存下列信息: 与主机 WWN (或与 WWN 15 相关联的 ID) 和群组的关联有关的信息; 以及与每个物理端口中以群组为基础的 LUN 分配有关的信息。

存储子系统 100 获得访问一个物理端口的主机 300 和服务器 200 的主机端口 WWN, 参考获得的 WWN 所属的群组的 LUN 分配信息以便根据所指定的要访问的 LUN 来标识要访问的逻辑卷 500, 然后访问逻辑 20 卷 500。

此外, 存储子系统 100 管理主机 300 或服务器 200 的主机端口并控制从主机 300 或服务器 200 对逻辑卷 500 的访问, 主机 300 和服务器 200 用上面提到的信息可以访问每个逻辑卷 500。更明确地说, 存储子系统 100 保存与主机 300 或服务器 200 的主机端口有关的信息并且拒绝从没有保存其相关信息的主机端口对逻辑卷 500 的访问, 主机 25 300 和服务器 200 能够访问特定的逻辑卷 500。其结果是, 即使不被允许访问存储子系统 100 所拥有的逻辑卷的主机 300 和服务器 200 和存储子系统 100 之间有物理连接, 也能保证逻辑卷 500 的安全。

30 (C) 快照

快照功能是创建存储子系统 100 中逻辑卷 500 的快照卷的功能以及管理所创建的快照的功能。

快照功能用于存储子系统 100 的正常的输入输出。例如，快照功能用于以下目的：在连续的日常操作期间获得备份（快照备份），在数据库之间传输数据并在数据库之间共享数据。

存储器子系统 100 一从服务器 200 接到复制的指令，就复制主卷的内容到副卷，复制指令包括对源逻辑卷 500（主卷）和目标卷 500

（副卷）的说明。在复制处理开始之后，如果在主卷中发出了更新数据的指令，存储子系统 100 也在副卷中更新数据以使数据更新保持同步。其结果是，在复制处理完成之后，主卷的内容保持同步以使主卷的内容总是与副卷的内容相同。

另外，如果服务器 200 命令对分离，存储子系统 100 就停止与副卷的数据更新同步，并在对分离指令的同时获得副卷中的快照卷。在这个连接中，如果在对分离之后更新了主卷中的数据，存储子系统 100 记录主卷中更新的数据的位置作为更新信息。

在快照卷用于像备份处理这样的处理之后，如果服务器 200 指令另一个复制（再同步），存储子系统 100 就以主卷的更新信息为基础为同步而复制主卷的更新过的数据到副卷以保持主卷的内容同步，从而使主卷的内容总是与副卷的内容相同。注意在这个处理实例中，在通过复制主卷的内容到副卷而使两个卷的内容相同之前不能执行快照获取（对分离），但在下面的处理实例中可以在任意时间获取快照。

当从服务器 200 接收复制指令（指令包括指令主卷和副卷的信息）时，存储子系统 100 在开始主卷和副卷之间的复制之前准备与主卷和副卷之间的差异有关的信息。当开始复制时，差异信息指示主卷的所有区域。这种情况下，指示复制已经完成的区域（包括已经用其对主卷的更新进行同步的区域）的信息在复制执行之后就被排除在外。

服务器 200 一指令对分离就能够从副卷读取数据。存储子系统 100 已经从服务器 200 或类似装置接收到从副卷读数据的请求，存储子系统 100 检查上面提到的差异信息。如果要读取的目标存储区域已经被复制完毕，存储子系统 100 就读取副卷中存储的数据，然后传输读取的数据到发出请求的服务器 200 或类似的装置。另一方面，如果要读取的目标数据还没有被复制到副卷，存储子系统 100 就从主卷读取目标数据，然后传输读取的数据到发出请求的服务器 200 或类似的装置，并在把读目标区域排除出差异信息之前复制数据到副卷。

5 当在指令对分离之后请求主卷的数据更新时，存储子系统 100 检查差异信息。如果写目标区域中存储的数据还没有被复制到副卷，存储子系统 100 先复制写目标区域中存储的数据到副卷，然后把该区域排除出差异信息。在此之后，存储子系统 100 更新主卷的数据，然后在它的更新信息中记录更新区域。

作为上述处理的结果，即使在对分离时到副卷的复制没有完成，从读数据的服务器 200 看来，好象是在主卷的对分离时在副卷中创建了一个快照卷。

10 注意尽管上述处理中主卷和副卷之间的关系是一对一的对应关系，但一个主卷可以对应于多个副卷，也可以实现多次复制，在多次复制中副卷是新的主卷。另外，在上述处理中，可以在卷基础上创建副本。但是，如果存储子系统 100 有用于管理任意大小的存储区域之间的关系的信息，可以为任意大小的区域创建快照卷。

(D) 远程复制

15 这个功能是通过 SAN 900 或别的通信线路创建逻辑卷 500 的复制到存储子系统 100 的另一位置的功能。在服务器 200 或主机 300 在拥有主卷的存储子系统 100（以下称为主存储子系统）和拥有副卷的存储子系统 100（以下称为副存储子系统）之间建立一对和数据传输通道（通道），一旦服务器 200 指令复制就从主存储子系统复制数据到副存储子系统。

20 如果在对分离之前，服务器 200 或主机 300 向主卷写入数据作为数据更新，主存储子系统就更新主卷中存储的数据，然后报告对服务器 200 和主机 300 的数据写入完成。此外，主存储子系统不考虑数据更新及时向副存储子系统传输要更新的数据以使数据更新和副卷的内容同步。

25 接收到要更新的数据的副存储子系统使要更新的数据和副卷中的数据同步。但是，在更新数据时，即使在从主存储子系统进行数据传输时确保数据到达命令，副存储子系统以时间序列为基础根据更新命令使数据同步。说得更明确些，如果已经不止一次地向主卷写入了数据，副存储子系统根据已经向主卷写入数据的命令使数据更新与副卷的数据更新同步。这种策略能够始终保持数据库的一致性，例如在事务处理所伴随的操作中。因此，能够保持事务的一致性。

5 服务器 200 一指令对分离主存储子系统就停止向副存储子系统传输要更新的数据。副存储子系统在指令对分离之前更新接收到的数据，然后允许服务器 200 或类似装置读取副卷的数据。为了确保上述更新顺序，包括更新顺序的管理信息被添加到存储子系统 100 间传输的数据包。

10 如上所述，在多个存储子系统 100 之间复制数据并提供使用各个存储子系统 100 的服务器 200 和主机 300，使我们能够构建一种能克服故障的信息处理系统。说得更明确一些，如果在系统中发生了故障，能够通过用服务器 200 或主机 300 上的集群软件或类似的软件把处理移到其它系统上以在不中断操作的情况下继续处理。

15 特别地，能够在发生故障时或类似的时刻通过远距离的主站点和副站点使操作始终能够进行，从而使两个站点不会同时出现故障或类似情况。在上述处理中，副本是以卷为基础创建的。但是，如果存储子系统有用于管理任意大小的区域之间的关联的信息，可以为任意大小的区域创建副本。

(E) 复制

20 复制功能指下列功能：以来自服务器 200 或类似装置的复制指令为基础，存储子系统 100 把逻辑卷 500 中存储的数据复制到另一个存储子系统 100 或备份存储设备 700；或者以来自服务器 200 或类似装置的复制指令为基础，存储子系统 100 从另一个存储子系统 100 或备份存储设备 700 读取数据，然后把数据复制到被指令的存储子系统 100 所拥有的逻辑卷 500。

25 存储子系统 100 在从服务器 200 和主机 300 接收到复制数据到另一个存储子系统 100 或备份存储设备 700 的指令以及存储区域的说明后把逻辑 500 中指定的存储区域中存储的数据传输到上述另一存储子系统 100 或备份存储设备 700。

30 另外，存储子系统 100 在从服务器 200 和主机 300 接收到复制数据到另一个存储子系统 100 或备份存储设备 700 的指令以及存储区域的说明后指令另一存储子系统 100 或备份存储设备 700 读取指定的数据，然后把从上述另一存储子系统传输来的数据存储到逻辑卷 500 的指定存储区域中。

在上述远程复制功能中，所支持的分组格式和处理协议必须必两

一个存储子系统 100 通用；由两个存储子系统 100 执行的处理必须相互一致。但是，这个复制功能不管理主卷和副卷之间的差异，而且不确定更新顺序。因为复制指令仅仅读和写数据，上述限制并没有被施加给全部两个存储子系统 100。

5 如果服务器 200 运行备份软件来管理复制处理和处理序列/定时中的差异，可以在服务器 200 或主机 300 不传输数据的情况下执行处理。上述复制指令的实例是 SCSI 标准中所用的 EXTENDED COPY 命令。

(F) 性能监控

10 性能监控功能是用存储子系统 100 自身收集与存储子系统 100 的性能有关的信息。作为与性能有关的信息的例子，有下列指令性能或影响性能的数值：性能数值包括显示逻辑卷 500 每单位时间接收到的输入输出（以下称为 I/O）数量的数值，以及物理端口每单位时间传输的数据量；数据写和读的比例；高速缓冲存储器命中率；CPU 使用率；内部数据传输线路的使用率；内部交换机的使用率；以及磁盘单元的使用率。

20 存储子系统 100 可以累积上述数值为日志。存储子系统 100 还可以分析并总结累积的数值来作用报告或警告通知它们的外部设备。为了向外部设备提供数值和日志，存储子系统 100 有多种软件接口，例如 MIB/SNMP、CIM/XML 和读/写命令进行的数据传输，还有对这些接口对应的 API 和 CLI。另一方面，服务器 200 能够在对逻辑卷 500 之间的负载平衡的优化、对逻辑卷 500 的物理位置的优化以及对接口之间的负载平衡的基础上用通过软件接口从存储子系统 100 获取的上述多个数值进行分析来方便地执行性能计划和性能调节。

25 (G) 卷再定位

卷再定位功能是通过改变逻辑卷 500 所在的磁盘单元 550 所拥有存储区域的关联来改变逻辑卷 500 的物理位置。如上所述，存储子系统 100 通过关联磁盘单元 550 的存储区域和逻辑卷 500 来向逻辑卷 500 向服务器 200 提供逻辑卷 500。

30 存储子系统 100 复制逻辑卷 500 中所存储的数据到另一磁盘单元 550 所拥有的存储区域。然后，源磁盘单元 550 和逻辑卷 500 的关联被改为目标磁盘单元 550 和逻辑卷 500 的关联，已经向目标磁盘单元

550 复制了数据。这种机制允许存储子系统 100 在不改变逻辑卷 500 的属性（例如，标识符（名字或编号）和端口分配）的情况下能改变与逻辑卷 500 对应的物理位置。说得更明确些，可以在不改变服务器 200 和主机 300 的情况下更改逻辑卷 500 的数据再定位。

5 例如，如果存储子系统 100 拥有小容量高速度的磁盘单元 550（磁盘单元 A）和大容量低性能的磁盘单元 550（磁盘单元 B），存储子系统 100 把存储高访问频率数据的逻辑卷 500 重定位到磁盘单元 A，把存储低访问频率数据的逻辑卷 500 重定位到磁盘单元 B。这使得存储子系统 100 能够以与性能相适合的方式存储数据。

10 另外，如果一个（或一组）磁盘单元 550 与大量存储高访问频率数据的逻辑卷 500 相关联，并由此导致负载集中在磁盘单元 550 上，引起存储子系统 100 的性能下降，存储子系统 100 可以通过分配负载来提高性能，尤其是通过为再定位而在磁盘单元 550 之间分配存储高访问频率数据的逻辑卷 500。此外，存储子系统 100 能够通过从服务器 200 获取负载阈值作为大致标准、获取逻辑卷 500 作为候选、并获取目录和其它自满参数来自动重定位逻辑卷 500。

15 (H) 端口控制

20 端口控制功能是控制存储子系统 100 的每个物理端口或服务器 200 或主机 300 的每个主机端口的 I/O 功能。如果是为存储子系统的每个物理端口控制 I/O，存储子系统 100 就控制来自服务器 200 和主机 300 的访问，这些访问在它自己的各个物理端口中进行处理，以使 I/O 性能不超过物理端口基础上能够提供的阈值。

25 如果是为服务器 200 或主机 300 控制 I/O，存储子系统 100 就获取主机 300 和服务器 200 的主机端口的 WWN，主机 300 和服务器 200 访问存储子系统所拥有的物理端口，然后存储子系统 100 控制访问以使 I/O 性能不超过 WWN 基础上所能提供的阈值。作为上述处理的结果，能够确保使用存储子系统 100 中所包括的特定物理端口或者使用服务器 200 或主机 300 的特定主机端口的操作的性能。例如，如果存储子系统 100 用于一个实时系统和一个开发系统，即使来自开发系统的访问请求增加，也能避免由于处理负载增加而引起的实时系统的处理性能下降以及开发系统带宽负载的上升。

30 因为可以指定涉及 I/O 性能（例如，每单位时间的 I/O 次数，以

及每单位时间的数据传输量) 的阈值。另外, 还可以考虑下列处理:
5 为涉及给予了更高优先级的端口或 WWN 的处理设置确定是否执行端口
控制功能的阈值; 如果涉及给予了更高优先级的端口或 WWN 的处理性
能小于或等于所设置的阈值, 存储子系统 100 就不向该处理施加端口
控制功能。当存储子系统 100 的处理系统足够时, 这种处理能够避免
涉及给予了更高优先级的端口或 WWN 的处理免受不必要的限制。

(I) 同速缓冲存储器控制

10 高速缓冲存储器功能是存储子系统 100 用来根据服务器 200 的指
令把指定的逻辑卷 500 的存储区域部分或全部保存在存储子系统 100
的高速缓冲存储器 120 中的功能。在比磁盘单元 550 速度更快的高速
缓冲存储器 120 (例如, 高速半导体存储器) 中临时存储需要高速 I/O
性能的数据能够提高信息处理系统的处理性能。

15 至于通用高速缓存器的处理情况, 在高速缓冲存储器 120 中临时
存储或保存数据不影响逻辑卷和存储数据的存储区域的属性, 例如标
识符 (名字、编号和值) 和端口分配。因此, 能够在不改变服务器 200
和主机 300 的设置的情况下继续对操作的处理。至于通用高速缓存器
的处理情况, 开始高速缓冲存储器 120 中的数据保存的时间选择可以
是访问数据时, 或者预先指定的与从服务器 200 读数据根本无关的时
间。像后面所描述的情况那样在高速缓冲存储器 120 中预先保存数据
20 使得存储子系统 100 能够给第一访问请求提供高速访问。

(J) 管理接口

存储子系统 100 不是总有上述所有功能。另外, 即使单个存储子
25 系统 100 有类似功能, 在制造商之间也存在功能规范和界限上的差异。

因此, 存储子系统 100 有向外部设备提供与存储子系统 100 所支
持的功能有关的信息的软件接口 (管理接口), 以及这些功能的规范
和界限 (功能信息)。服务器 200 和主机 300 能够通过管理接品从存
储子系统 100 获取功能信息。另一方面, 当服务器 200 和主机 300 使
用或控制存储子系统 100 所拥有的功能时存储子系统 100 提供信息并
从服务器 200 接收必要的信息。
30

所提供的和接收的信息包括下列信息: 逻辑卷的创建 (定义、大小
设置、属性设置)、删除 (释放); 逻辑卷 500 的物理端口, 逻辑单
元号的分配、收回; 涉及主机 300 和服务器 200 的主机端口 (WWN) 的

群组的定义；能够访问逻辑卷 500 的主机 300 和服务器 200 的主机端口（WWN）的设置；用于创建快照的逻辑卷 500 的对定义，复制指令、分离指令、再同步指令，对释放；在存储子系统 100 之间创建逻辑卷 500 的副本的对定义，复制指令、分离指令、再同步指令，对分离；复制指令（复制目标、方向、区域规格等），性能信息的获取（单位时间的 I/O 次数、单位时间的数据传输量、读和写的比率、高速缓冲存储器命中率、CPU 使用率、内部数据传输线路的使用率、内部交换机的使用率和磁盘单元的使用率等）；性能报告的获取；性能警告阈值的设备和获取；与性能有关的警报通知；卷再定位指令（目标和源的规格）；卷自动再定位参数的设备和获取（负载阈值、目标候选、目录和算法参数等）；端口控制指令和取消（优先和非优先端口、优先和非优先 WWN、上限性能、可用标准阈值的设置和获取）；高速缓冲存储器控制指令或取消（目标逻辑卷和存储区域的规格、装载和预取方法的规格）；及类似信息。

另外，存储子系统 100 向服务器或类似装置提供与存储子系统 100 所拥有的存储区域有关的资源信息。例如，资源信息包括下列信息：存储子系统 100 所拥有的逻辑卷的标识符（名字和编号）；逻辑卷 500 的大小、属性和种类；与是否分配逻辑卷 500 到一个通路有关的信息；向其分配逻辑卷 500 的通路，和逻辑单元号。

此外，存储子系统 100 向外部设备提供用于从与 SAN 相连的设备中标识出存储子系统 100 的设备信息。例如，设备信息包括存储子系统 100 的制造商名、模式名、版本、产品号。

此外，存储子系统 100 向外部设备提供指示组成存储子系统 100 的模块和零件的信息（配置信息）以及编号。例如，配置信息包括各个部件的种类、编号、标识符（名字和编号）、位置、属性、状态（该部件是否能够使用、是否发生故障），这些部件例如物理端口、CPU、CPU 主板、高速缓存器、逻辑存储器、共享存储器、磁盘单元、内部数据传输线缆、内部交换机以及机箱。

另外，存储子系统 100 检测存储子系统 100 的配置过的模块或部件中发生的故障或者其中的异常处理状况以在向服务器 200 或主机 300 通知故障和异常状况之前识别故障或异常状况的位置、然后识别故障或异常状况的种类和状态。存储子系统 100 可以有用于根据故障和异

常状况的位置、种类和状态来假设起因和措施以便向服务器 200 和主机 300 一起通知识别出的原因和措施。此外，服务器 200 和主机 300 还可以在任意时间向存储子系统发送与上述故障信息有关的查询以获取故障信息。

5 存储子系统 100 用于提供上述每种信息的软件包括：使用网络 910 的 MIB/SNMP 和 CIM/XML；根据 SCSI 标准由读/写/INQUIRY/MODE SENSE/MODE SELECT 用 SAN 进行的数据传输；确定的协议；使用它们的 API 和 CLI；以及类似软件。

10 如图 3 所示，服务器 200 检测（发现）通过网络 910 和 SAN 900 与 SAN 900 连接的设备，例如存储子系统 100、备份存储设备 700、交换机 400、集线器、主机 300 等，然后从设备收集不种类的信息来创建数据库 210（称为信息库）。信息库 210 中所收集和存储的信息涉及设备类型、设备能够执行的功能、设备的关联信息以及在其中提供虚拟化资源和管理所需的其它信息。图 6-8 和 11-13 中提供了这样的 15 信息的一些实例。在一种实施方案中，用来采集这种信息的方法类似于和传统 SAN 系统耦合的管理服务器用来收集管理信息的方法。可以在服务器或服务器可以访问的远程位置中提供信息库 210。

20 服务器 200 使用不同种类的协议从存储子系统 100 获取不同种类的信息，然后把这些信息存储在信息库 210 中。特别地，信息库 210 包括与上面举例说明的每个存储子系统 100 有关的不同种类的信息、与制造商名、模式名、版本、产品号及类似参数有关的信息作为设备信息，这些信息用于识别与 SAN 900 相连的各个设备。

25 另外，服务器 200 从主机 300 获得正在使用的与主机 300 的主机总线适配器、主机端口的配置以及虚拟卷 600 有关的信息，然后在信息库 210 中注册这些信息。此外，服务器 200 从交换机 400 获取与交换机 400 的物理端口的配置有关的信息和类似信息，然后在信息库 210 中注册这些信息。此外，服务器 200 分析连接到 SAN 900 的设备的布局，然后在信息库 210 记录这些信息为布局信息。

30 服务器 200 检测连接到 SAN 900 的设备的配置中的变化、布局中的变化等，然后更新信息库 210 中存储的信息以保存新信息。此外，服务器 200 能够通过保存信息库 210 的更新历史发现连接到 SAN 900 的设备的配置中的时间序列变化、布局中的时间-序列变化等。另外，

服务器 200 还从主机 300 的代理 340 获取与应用软件 310 所用的虚拟卷 600 有关的信息、虚拟卷 600 的区域、使用趋势、性能以及类似信息，然后把这些信息作为主机信息和上面提到的信息一起记录在信息库 210 中。

5

服务器提供的功能

图 4 描述由服务器 200 提供给主机 300 的功能的实例。服务器 200 根据来自主机 300 的指令或请求向主机 300 提供虚拟卷 600 和与虚拟卷 600 相关联的功能。在提供与虚拟卷 600 相关联的功能时，服务器 300 虚拟化与存储子系统 100 所提供的逻辑卷 500 相关联的功能以提供该功能。
10

15

另外，服务器 200 检查存储子系统 100 所拥有的功能是否存在。然后，根据检查结果服务器 200 用自身执行像复制这样的处理或者指令存储子系统 100 执行处理以便处理单个设备中所分发的处理。下面将描述图 4 中所示的服务器 200 的各种功能。这些功能从服务器 200 的存储设备读取，然后通过执行存储在存储器 200 中的与每种功能对应的程序来实现。

(A) 卷 集中控制

图 5 依照本发明的一种实施方案描述卷 集中控制方法 1000。服务器 200 通过执行控制器功能和元文件系统向主机 300 提供逻辑卷 500 的一个存储区域作用虚拟卷 600，后面将描述控制器功能和元文件系统。主机 300 执行代理 340、应用软件 310、OS 320 或访问处理软件 330，并指定要送往服务器 200 作为附加存储空间请求的条件（步骤 1001），例如大小和属性。服务器 200 接收请求并查找信息库 210 以找到逻辑卷 500 的一个存储区域，该区域尚未使用并且符合条件（步骤 1002）。在步骤 1003，如果在逻辑卷 500 中找到合适的存储区域，服务器 200 继续用已经在逻辑卷中定位的存储区域创建主机 300 所请求的虚拟卷 600（步骤 1011），然后分配虚拟卷 600 到主机 300（步骤 1012）。
20
25

在步骤 1003，如果在逻辑卷 500 中没有找到合适的存储区域，服务器 200 就查找信息库 210 以找到有未使用存储区域的存储子系统 100（步骤 1004），即查找物理存储区域。如果找到了存储子系统 100 中的可用存储区域（步骤 1005），服务器 200 用信息库 210 中的功能信
30

息检查是否有找到的存储子系统 100 的卷定义功能和管理功能以及其中的规范（步骤 1007）。

如果找到的存储子系统 100 有这些功能（步骤 1008），服务器 200 指令存储子系统 100 创建逻辑卷 500 并分配端口和逻辑单元号（步骤 1010）。如果在单个存储子系统中没有找到足够的存储区域，就用多个存储子系统 100 创建逻辑卷 502。这种情况下，服务器 200 指令每个存储子系统 100 创建最终满足主机 300 所设置的条件的逻辑卷 502。每个存储子系统可以有不同的方式创建和管理包括在多个存储子系统 100 之间的接口规范的逻辑卷 502。

在步骤 1008，如果确定已经定位的存储子系统没有合适的能力和功能来创建新的逻辑卷（例如逻辑卷 502），服务器 200 代表那个存储子系统创建逻辑卷 502（步骤 1009）。这样没有合适能力的存储子系统的实例包括单个磁盘驱动器单元或 JBOD。这种情况下，服务器 200 管理逻辑卷 502 的存储区域、端口和逻辑单元号。

此后，服务器 200 根据主机 300 的请求把新形成的逻辑卷 502 的存储区域和悍然卷 600 关联在一起（步骤 1011）。逻辑卷 502 和虚拟卷 600 的关闻包括下列情况：一个虚拟卷 600 与一个逻辑卷关联；一个虚拟卷 600 与多个逻辑卷相关联；多个虚拟卷 600 与一个逻辑卷相关联。此外，还有一种情况，与一个虚拟卷关联的多个逻辑卷中的每一个都属于彼此不同的存储子系统 100。然后相关的存储区域被分配给主机 300（步骤 1012）。

图 6-8 描述在上述卷集中控制操作中创建的表 50、52 和 54。表 50 提供与虚拟卷和逻辑卷以及物理存储单元的关联有关的信息。列 56 提供虚拟卷编号；列 58 提供虚拟卷的大小；列 60 提供存储子系统编号；列 62 提供逻辑卷编号；列 64 提供存储区域地址；列 66 提供存储区域的大小。例如，行或记录 68 指示虚拟卷 0 的大小为 4000，与存储子系统 1 和逻辑卷单元 2 相关联。逻辑卷 2 的存储区域地址从 10000 开始，逻辑卷 2 的存储区域大小为 4000。

表 52 提供提供给主机 300 的与通路地址和逻辑单元号的关联有关的信息，以使主机能够提供合适的通路信息给服务器以访问虚拟卷。列 68 提供已经由服务器提供给主机的通路地址；列 70 提供逻辑单元号；列 72 提供虚拟卷号。例如，行或记录 74 提供主机访问与逻辑单

元 1 相关联的虚拟卷 0 所需要的信息。

表 54 提供提供给服务器 200 与通路地址和逻辑单元号的关联有关的信息，以使服务器能够访问与主机指定的虚拟卷相关联的逻辑卷对应的适当的存储子系统。更 76 提供通路地址信息给存储子系统；列 5 78 提供与逻辑单元编号有关的信息；列 80 提供与存储子系统编号有关的信息；列 82 提供与虚拟卷编号有关的信息。例如，行或记录 84 提供访问虚拟卷 0 所需要的信息。

(B) 备份和归档

图 9 描述依照本发明的一种实施方案所执行的备份和归档方法。
10 备份虚拟卷 600 中所存储的数据使得主机 300 即使在发生数据冲突或
数据丢失的情况下也能够恢复数据并重新启动正常操作。

15 服务器 200 检查虚拟卷的备份数据的目标的位置（步骤 1101）。如果目标和源位置相同，即，仅仅希望虚拟卷的快照，服务器 200 创建虚拟卷中的数据的快照（步骤 1103）。后面描述执行快照操作所涉及的步骤。服务器或别的备份软件记录与备份数据和介质有关的管理信息（步骤 1109）。

20 如果目标是一个远程位置，即目标位置与源位置不同时，也创建数据的一个快照，以便初始化备份操作（步骤 1102）。快照功能确保已经备份的数据与源数据一致。为此，服务器 200 通过从由主机 300 执行的代理 340 获取与应用软件 310 和 OS 320 的处理有关的信息，并通过执行代理 340 和访问处理软件 330 来标识快照创建的时间选择以控制来自应用软件 310 和 OS 320 的访问请求和从主机 300 到虚拟卷 600 的访问。

25 服务器 200 检查信息库 210 的功能信息以确定是否有相应的存储设备（即目标设备）拥有复制功能和其中的能力（步骤 1104）。目标设备可以是复杂的存储设备，例如有控制器的存储子系统，或者是信息处理能力有限的原始的磁盘驱动器。通常，存储子系统有复制功能而原始磁盘驱动器没有。如果目标存储设备有复制功能（步骤 1105）并且该复制功能可以用于正在讨论的具体的备份操作（步骤 1106），
30 服务器指令存储设备从快照卷复制数据到目标（步骤 1107）。此后方法 1100 继续到步骤 1109。在一种实施方案中，在步骤 1109 服务器 200 存储与备份数据有关的信息到信息库 210 中，与备份数据有关的信息

包括备份目标、备份日期和时间、涉及备份和恢复的版本以及与记录备份数据的介质有关的管理信息。作为选择，可以由备份软件管理这样的信息。

在步骤 1105 和 1106，如果目标存储设备没有复制能力或这样的
5 复制能力不适用于正在讨论的备份操作，服务器复制快照数据到目标
存储设备（步骤 1108）。此后，方法继续到步骤 1109。但是，服务器
进行这样的操作增加它的负载，而这是不希望看到的。

图 11 依照本发明的一种实施方案描述涉及存储系统中不同存储设备的复制能力的表 150。在步骤 1105 和 1106 访问并使用表 150 来确定复制功能是否由目标存储设备或服务器执行。列 512 标识存储系统中的存储设备。列 154 指示给定的存储设备是否能执行复制功能。列 156 指示的存储设备在给定时间是否有复制功能。列 158 指示存储设备类型。列 160 提供与给定存储设备执行什么类型的复制功能有关的信息。例如，行或记录 166 指示存储设备 0 可以从磁盘设备复制数据到磁带设备，或从磁带设备复制数据到磁盘设备。列 162 提供与给定存储设备能够充当的最大目标数有关的信息。例如，存储设备 0 能够充当的最当目标数被标示为 128。列 164 提供与给定存储设备能够执行的最大的并发操作数有关的信息。
10
15

(C) 快照

图 10 依照本发明的一种实施方案描述执行快照功能的方法 1200。
20 服务器 200 的快照功能是在给定时间创建并管理虚拟卷 600 或逻辑卷 500 的快照的功能。首先，服务器 200 确定组成虚拟卷 600 的一个或多个逻辑卷，然后确定逻辑卷 500 要被复制到的一个或多个存储子系统 100（存储 1201）。

如果所复制的数据要存储到相同的存储子系统中，方法继续进行
25 到步骤 1202。在那个步骤，服务器 200 参考信息库 210 的功能信息来
检查正在讨论的存储子系统 100 是否有快照功能及其中的规范。如果
正在讨论的存储子系统 100 有快照功能（步骤 1203），而且从功能规
范的观点可以使用该快照功能（步骤 1204），服务器 200 就使用该存
储子系统 100 的快照功能（步骤 1205）。然后服务器 200 创为复制的
30 数据创建一个新的虚拟卷并关联该新虚拟卷到原始虚拟卷（步骤
1211）。

如果在步骤 1203 确定每个存储系统都没有快照功能或在步骤 1204 确定该功能不可用，方法 1200 就进行到步骤 1210，以使服务器 200 创建快照卷。

图 12 描述用来确定存储子系统是否有快照功能及该功能是否可用的信息库 210 中所包括的表 250。列 252 列出了存储子系统编号；列 254 指示给定的存储子系统是否有快照功能；列 256 指示该快照功能是否可用；更 258 提供与存储子系统类型有关的信息；列 260 提供与给定存储子系统可以有的最大（源和目标卷）对数有关的信息；列 262 提供与给定存储子系统中可以有的最大目标数有关的信息；列 264 提供与给定存储子系统可以有的并发操作数量有关的信息。如果给定存储子系统达到了列 260、262 和 264 中提供的限制，就取消该存储子系统的快照功能，其快照功能不可用。

再回到步骤 1201，如果从源卷在远程位置提供的目标卷，就执行远程复制。服务器 200 参考信息库 210 的功能信息检查源存储子系统 100 和目标存储子系统 100 是否有远程复制功能和其中的规范（步骤 1206）。如果两个存储子系统 100 都有远程复制功能（步骤 1207）并且该功能对两个存储子系统都可用（步骤 1208），服务器 200 就通知在存储子系统和数据传输通路的设置之间包含一对主卷和副卷的两个存储子系统 100 以启动数据复制（步骤 1209）。此外，当使用虚拟卷 600 的复制时，服务器 200 指令存储子系统 100 分离对。此后，方法 1200 进行到步骤 1211。

图 13 依照本发明的一种实施方案描述服务器的信息库中所提供用来提供涉及存储系统中的存储设备的远程复制功能的表 350。在方法 1200 中，至少在步骤 1207 和 1208 访问和使用表 350 来确定服务器或存储设备是否需要执行远程复制。列 352 标识存储设备。列 354 指示给定的存储设备是否有远程复制功能。列 356 指示是否已经激活远程复制功能。列 358 提供与存储设备类型有关的信息。列 360 指示给定存储设备可以拥有的最大对数。列 362 指定给定存储设备中的给定卷可以充当的最大目标数。列 364 提供与给定存储设备能够执行的最大并发操作数有关的信息。

如果存储子系统 100 的一个或全部没有远程复制功能（步骤 1207），或者因为功能规范之间的差异而无法使用远程复制功能（步

骤 1208），就由服务器 200 执行数据复制。说得更明确些，服务器 200 通过从主卷读数据、传输数据到副卷并检测从主机 300 到主卷的数据更新来使更新和副卷的内容同步。这种情况下，服务器 200 也分离卷对。

但是，优先使用存储子系统 100 来执行快照功能，因为这能减轻服务器 200 和 SAN 900 的负载。另外，如果虚拟卷 600 被包括在多个逻辑卷 500 中，服务器 200 在创建虚拟卷 600 的快照时管理多个对，并在执行每个对的处理之前选择创建快照的方法。特别地，在分离对时，即使有由存储子系统 100 根据指令分离的对以及由服务器 200 分离的对，服务器 200 也执行处理以使在相同时间分离对，而且由此服务器能够在不发生冲突的情况下创建虚拟卷 600 的快照。

另外，在上述处理中，处理是以卷为基础执行的。但是，如果服务器 200 有用于管理任意大小的区域之间的关联的信息，就能创建任意大小的区域的副本。此外，服务器 200 能够用与元文件系统有关的信息以文件为基础创建副本。

为了在创建快照时选择副卷，服务器 200 用服务器 200、主机 300 以及控制终端 800 中包括的显示屏幕提供给用户和主机 300 的系统管理员一个逻辑卷 500 和一个存储区域，所提供的逻辑卷和存储区域没有用在主卷所在的存储子系统 100 中。用户和系统管理员从所提供的内容确定要用的副卷。

服务器 200 用创建新的虚拟卷 600 的方法向主机 300 提供指定的副卷。由服务器 200 自动提供副卷能够减少用户或系统管理员确定副卷所需的工作。作为上述处理的结果，使用与 SAN 900 相连的一个或多个存储子系统 100 的快照创建功能或由服务器 200 自身创建快照卷使得服务器 200 能够调节或隐藏快照创建功能的缺乏以及其中规范中的差异，并通过链接存储子系统 100 中逻辑卷 500 的处理来创建和管理虚拟卷 600 的快照。

注意，就在另一存储子系统 100 中创建副本的情况来说，也可以创建任意大小区域的快照。此外，服务器 200 能用与元文件系统有关的信息在文件基础上创建快照。

此外，同时使用远程复制功能和快照处理使得通过对任意存储子系统 100 中任意逻辑卷 500 的配置能够创建虚拟卷 600 的快照卷。快

照功能用于在继续操作的同时（非中止备份）获得备份，还用于数据移动和数据共享，它们维持数据库间的数据一致性。

(D) 控制器

控制器功能是服务器 200 用来通过地址转换向主机 300 提供一个或多个存储子系统 100 的一个或多个逻辑卷 500 中所包括的一个或多个存储区域作为虚拟卷 600。说得更明确些，使用这个功能，服务器 200 管理逻辑卷 500 的区域与虚拟卷 600 的区域之间的关联，并在处理主机 300 的访问请求之前转换主机 300 对虚拟卷 600 的访问为对逻辑卷 500 的访问。

与逻辑卷 500 的存储区域和虚拟卷 600 的存储区域有关的关联信息存储在信息库 210 中。这个关联信息包括与端口标识符（名字、编号、WWN 和 ID）和 LUN（由服务器 200 提供给主机 300 作为虚拟卷的端口和 LUN）与逻辑卷 500 的端口标识符和 LUN（由存储子系统 100 提供给服务器 200）的关联有关的信息。

当改变（例如，创建、删除或扩充）虚拟卷 600 时更新信息库 210 中存储的关联信息。此外，当建立关联时，产生冗余数据并分配数据来配置 RAID 能够提高数据的可靠性、可用性和性能。

(E) 协议转换

协议转换功能是在用控制器功能向主机 300 提供虚拟卷 600 时，不考虑与另一 SAN 900 相连的设备（例如存储子系统 100 和备份存储设备 700）所用的协议，服务器 200 提供虚拟卷 600 与主机 300 所用协议对应的设备。

更具体的，服务器 200 用和存储子系统 100 所对应的协议与存储子系统 100 通信，用和备份存储设备 700 对应的协议与备份存储设备 700 通信，用和主机 300 对应的协议与主机 300 通信。存储子系统 100、备份存储设备 700 及主机 300 所用的协议可以不同，即使它们是相同类型的设备。当服务器检测与 SAN 900 相连的每个设备时，服务器 200 确定与该设备对应的协议类型，然后在信息库 210 中记录该协议类型作为设备和配置信息。

(F) 元文件系统

服务器 200 执行元文件系统，然后向多个主机 300 的应用软件 310、OS 320 提供由多个逻辑卷 500 组成的存储区域作为一个文件系

统。主机 300 用访问处理软件 330 接收由应用软件 310 或 OS 320 发出的访问虚拟卷 600 的请求，然后执行访问处理软件向服务器 200 发送与逻辑卷 500 中要访问的数据的位置有关的查询。

接收到查询的服务器 200 用元文件系统向主机 300 发送与逻辑卷 500 中的位置有关的应答。主机 300 执行访问处理软件 330 来从服务器 200 获取的位置访问逻辑卷 500，逻辑卷 500 由存储子系统 100 提供。

读取和写入的数据在主机 300 和存储子系统 100 之间直接传输。由服务器 200 和主机 300 的访问处理文件执行上述处理使得应用软件 310 和 OS 320 能够访问元文件系统上的文件并访问虚拟卷 600 上的任意区域（块）。

(G) 通路管理

通路管理功能是管理存储子系统 100 的物理端口和服务器 200 的主机端口之间的通路、服务器 200 的主机端口和主机 300 的主机端口之间的通路以及存储子系统 100 的物理端口和主机 300 的主机端口之间的通路来实现存储区域的虚拟化并在故障发生时克服故障的功能。

当服务器 200 向主机 300 提供虚拟卷 300 时，如果服务器 200 重新访问存储子系统 100 的端口和 LUN 来创建逻辑卷 500，服务器 200 使用通路管理功能来建立服务器 200 的一个端口以使服务器 200 能够访问存储子系统 100 的对应端口/LUN 的逻辑卷 500。

另外，如果主机 300 重新访问存储子系统 100 的端口和 LUN，服务器 200 执行通路管理功能来指令主机 300 的代理 340 建立端口。接收到指令的代理 340 建立存储子系统 100 的端口以使主机 300 能够访问存储子系统 100 的端口/LUN 的逻辑卷 500。

此外，如果主机 300 访问由服务器的控制器功能提供的虚拟卷 600，服务器 200 用通路管理功能指令主机 300 的代理 340 建立一个端口以使主机 300 能够访问虚拟卷 600。另外，至于有分配逻辑卷 500 的端口/LUN 功能的存储子系统 100，如果根据 SAN 900 中主机 300、服务器 200 和存储子系统 100 的布局对主机 300 和服务器 200 能够访问的存储子系统的端口/LUN 有限制，服务器 200 就使用通路管理功能来指令存储子系统 100 分配主机 300 能够访问的端口/LUN。

另外，当主机 300 访问由服务器 200 的控制器功能提供的虚拟卷

600 时，如果对主机 300 能够访问的服务器 200 的端口/逻辑单元号有
限制，服务器 200 用通路管理功能给主机 300 分配它能够访问的端口
/LUN。另外，根据 SAN 900 中主机 300、服务器 200 和存储子系统 100
的布局，如果有必要建立交换机 400 等的分区功能以便实现来自主机
5 300 的访问，服务器 200 使用通路管理功能来执行交换机 400 所需的
设置。

此外，在创建端口之间的通路时使通路加倍，并在出现故障时切
换通路以使其能够使用，使得即使在故障发生时操作也能继续。例如，
10 当为逻辑卷 500 使通路加倍时，服务器 200 使用通路管理功能来指令
有分配逻辑卷 500 的端口/LUN 功能的存储子系统同时分配不同的端口
给目标逻辑卷 500。

15 使用通路管理功能，服务器 200 能够参考信息库 210 的功能信息
来检查存储子系统 100 是否有端口/LUN 分配功能。服务器 200 把上面
提到的与端口/LUN、通路和分区有关的信息记录在信息库 210 中，并
放置、更新并参考这些信息。

(H) 高可用性和灾难恢复

在这个系统中，主机 300、服务器 200 和存储子系统 100 以及它
们之间的通路都被加倍；在发生故障时改变通路，这使得在故障发生
20 时操作仍能继续。当主机 300 和服务器 200 被加倍时，集群软件用于
另一主机 300 和另一服务器 200 以使它们加倍。

另外，使用下面所描述的服务器 200 的远程复制功能来创建存储
子系统 100 中逻辑卷 500 的副本能够使存储子系统 100 加倍，逻辑卷
500 由任意虚拟卷 600 组成。如果在一个系统中发生故障，其它系统
25 能够使系统不需中断即可继续执行。如果在远距离提供主站点和副站
点以使这两个站点不会同时遭受灾难，在每个站点中建立系统，在一
个站点遭遇灾难时，处理被快速移动到另一站点用于恢复，这使得操
作能够在出现灾难时不需中断即可继续执行。

(I) 远程复制

30 服务器 200 的远程复制功能是在另一存储子系统 100 中创建虚拟
卷 600 和逻辑卷 500 的副本并管理副本的功能。说得更明确些，服务
器 200 在与逻辑卷 500 所在的存储子系统 100 不同的存储子系统 100
中创建逻辑卷 500 的副本，逻辑卷 500 由虚拟卷 200 组成。

首先，服务器 200 参考信息库 210 的功能信息检查源存储系统 100 和目标存储系统 100 是否有用在存储系统 100 之间的远程复制功能，还检查远程复制功能规范中的差异。如果源和目标存储系统 100 都有远程复制功能并且从功能规范的观点来看能够执行该功能，服务器 200 就通知源和目标存储系统 100 在存储系统之间的一对主卷和副卷以及数据传输通路的设置来指令数据的复制。此外，当使用虚拟卷 600 的副本时，服务器 200 指令存储系统 100 分离源和目标对。

另外，如果一个存储系统 100 没有远程复制功能（或者两个存储系统 100 都没有远程复制功能），或者因为功能规范中的差异而无法使用远程复制功能，就由服务器 200 执行数据复制。说得更明确些，服务器 200 通过从主卷读取数据并传输数据到副卷来复制数据，并且检查从主机 300 到主卷的数据更新以便使更新和副卷的更新同步。这种情况下，服务器 200 也分离源和目标对。

如上所述，如果使用存储系统 100 的功能，使用存储系统 100 的功能能够减轻服务器 200 和 SAN 900 的负载。另外，如果虚拟卷有多个逻辑卷 500，在创建虚拟卷 600 的副本时服务器 200 管理多个对，并在为每个对执行实际的处理之前选择创建数据副本的方法。特别地，当分离对时，即使有由存储系统 100 根据指令分离的对和由服务器 200 分离的对，服务器 200 执行处理以使这些对被同时分离，并且在无冲突情况下创建虚拟卷 600 的副本。

为了在创建虚拟卷的副本时选择副卷，服务器 200 用服务器 200、主机 300 和控制终端 800 中所包括的显示屏幕向用户和主机 300 的系统管理员提供与逻辑卷 500 和存储区域有关的信息，所提供的逻辑卷 500 和存储区域在存储系统 100 中没有使用。用户和系统管理员从所提供的内容确定要用的副卷。

服务器 200 用创建新的虚拟卷 600 的方法向主机 300 提供由用户或类似人员指定的副卷。此外，由服务器 200 自动提供副卷能够减少由用户或系统管理员选择副卷所需要的工作量。

作为上述处理的结果，使用与 SAN 900 相连的多个存储系统 100 的副本创建功能，或者由服务器 200 复制数据，使得服务器 200 能够调节并隐藏特定功能的缺乏以及特定功能的规范中的差异，并且通过关联多个存储系统 100 之间对逻辑卷的处理在另一存储系统 100 中创

建和管理虚拟卷 600 的副本。

另外，在上面提到的处理中，处理是以卷为基础执行的。但是，如果服务器 200 有用于管理任意大小的区域之间的关联的信息，就能够为任意大小的区域创建副本。此外，服务器 200 能够用与元文件系统有关的信息在文件基础上创建副本。
5

(J) 复制

服务器 200 根据来自主机 300 的复制指令或类似设备的复制虚拟卷 600 的存储区域中所存储的数据到另一个虚拟卷 600 或备份存储设备 700；服务器 200 从备份存储设备 700 读取指定的数据并把它们复
10 制到虚拟卷 600。

接收到复制指令的服务器 200 检查信息库 210 中的功能信息。如果存储系统 100、备份存储设备 700 和其它目标设备的复制功能可以使用，服务器 200 就指令每个设备都执行复制。如果每个设备都不能使用复制功能，服务器 200 从指定的源地址读取数据，然后把数据传
15 输到指定的目标地址以复制数据。如上所述，如果存储系统 100 的复制功能可以使用，使用复制功能可以减轻服务器 200 执行复制处理的负载。

(K) 性能管理

服务器 200 从与 SAN 900 相连的存储子系统 100 获取与性能有关的数据，然后把这些信息累积在信息库 210 中，性能相关信息是由存储子系统 100 测量的。另外，服务器 200 获取与从主机 300 到虚拟卷 600 的访问有关的下列信息：例如，虚拟卷 600 每单位时间接收到的 I/O 次数；读写比率；I/O 传输长度；服务器 200 每单位时间传输的数据量；高速缓冲存储器命中率；内部数据传输线路的使用率；以及类似的数据。然后，服务器 200 把获得的信息注册在信息库 210 中。
20
25

此外，服务器 200 从代理 340 和访问管理软件 330 获得下列信息：例如，用于主机 300 的单个应用程序 310 和 OS 320 的虚拟卷的访问区域；每单位时间的 I/O 请求次数；读写请求比率；I/O 传输长度；每单位时间在主机 300 的一个端口中的 I/O 请求次数；读写比率；I/O 传输长度；每单位时间传输的数据量；主机 300 的 CPU 使用率；高速缓冲存储器命中率；以及类似的数据。然后，服务 200 把获得的信息记录在信息库 210 中。
30

5 此外，服务器 200 还从像备份存储设备 700、交换机 400、集线器和主机 300 这样与 SAN 相连的设备获取与性能有关的信息，然后把这些信息记录在信息库 210 中。服务器 200 能够通过把信息库 210 中记录的性能信息累积成日志来确定性能中的时间-序列变化以及变化的趋势。

10 另外，服务器 200 把已经累积并记录在信息库 210 中的性能信息提供给主机 300 的外部设备或类似设备。此外，服务器 200 还有像显示屏幕这样的设备用于把信息库 210 中所存储的性能信息展现给用户或系统管理员。不仅服务器 200 有显示屏幕，主机 300 和控制终端 800 也有。

15 此外，服务器 200 分析在信息库 210 中积累和记录的性能信息以检测整个系统的性能问题和性能瓶颈。之后，服务器 200 进行以下性能调试：通过性能规划和负载分布保证 QoS；通过负载分布改进性能：优化逻辑卷 500 的物理位置等等。负载分布和分配处理的一个例子将在以下进行说明。

20 如果检测到负载集中在主机 300、交换机 400、SAN 900、服务器 200、存储子系统 100（这些设备用于处理对虚拟卷 600 和逻辑卷 500 的访问）等所拥有的一些资源或部件中，服务器 200 参考并分析信息库 210 中的不同种类的性能信息、资源信息和配置信息来计划负载分发和分配。负载分发和分配的计划包括移动逻辑卷 500 的物理位置的分发和分配（再定位），以及由端口控制控制的负载分发和分配。

25 如果执行逻辑卷 500 的再定位，服务器 200 确定上面提到的计划中的要再定位的逻辑卷 500（源卷）以及目标存储区域。如果源卷和目标存储区域存在于相同存储子系统 100 中，服务器 200 就参考信息库 210 中的功能信息检查存储子系统 100 是否有卷再定位功能，并且检查功能的规范。

30 如果目标存储子系统 100 有卷再定位功能并且从功能规范的观点来看卷再定位功能可以使用，服务器 200 就指定源卷和目标存储区域以指令存储子系统 100 执行再定位。另一方面，如果目标存储子系统 100 没有卷再定位功能，或者如果从功能规范的观点来看卷再定位功能不可用，或者如果源卷所在的存储系统 100 与目标存储区域所在的存储系统 100 不同，服务器 200 就执行再定位处理。

更具体的，首先，如果必要的话服务器 200 指令存储子系统 100 创建目标存储区域的逻辑卷 500（目标卷）以便创建目标卷。接着，
5 服务器 200 从源卷读取数据，然后把数据传输到目标卷以复制数据。
另外，服务器 200 检测从主机 300 到源卷的数据更新，并使该更新和
目标卷的更新同步。然后，服务器 200 更改与逻辑卷 500 的存储区域
以及元文件系统的数据定位信息以改变提供给主机 300 的虚拟卷 600
的逻辑卷 500 上的存储区域。

10 作为上述处理的结果，服务器 200 在不改变像虚拟卷 600 的标识符（名字和编号）的情况下更改存储数据的物理位置。

另外，当执行端口控制时，服务器 200 参考信息库 210 中的功能信息检查存储子系统 100 是否有端口控制功能，并检查端口控制功能的规范。如果存储子系统 100 有端口控制功能，并且从功能规范的观点来看该端口控制功能可用，服务器 200 就指令存储子系统 100 以计划为基础执行端口控制。
15

另一方面，如果目标存储子系统 100 没有端口控制功能，或者从功能规范的观点来看端口控制功能不可用，或者源卷所在的存储子系统 100 与目标存储区域所在的存储子系统 100 不同，服务器 200 就执行性能控制。特别地，服务器 200 以上述计划为基础处理来自主机 300
20 的访问以使 I/O 性能不超过预先确定的阈值。否则，服务器 200 需要访问处理软件 330 在以上述阈值为基础来控制性能。接收到请求的访问处理软件 330 处理访问以使 I/O 性能不超出上述阈值。该处理允许服务器 200 实现负载分发和分配。

下面将描述优化逻辑卷 500 的物理位置的处理的一个实例。服务器 200 参考并分析信息库 210 中的上述不同种类的性能信息以及资源信息、配置信息等来计划对逻辑卷 500 的物理位置的优化。物理位置优化计划包括逻辑卷 500 的物理位置的移动（再定位），以及由高速缓冲存储器控制在高速缓冲存储器 120 的存储区域中保存数据。
25

在执行对逻辑卷 500 的再定位时，如果可以使用存储子系统 100 所拥有的卷再定位功能，服务器 200 就使用卷再定位功能；如果该功能无法使用，就由服务器 200 再定位逻辑卷。
30

此外，在执行高速缓冲存储器控制时，服务器 200 参考信息库 210

中的功能信息来检查目标存储区域所在的存储子系统 100 是否有高速缓冲存储器控制功能，并检查高速缓冲存储器控制功能的规范。如果存储子系统 100 有高速缓冲存储器控制功能，并且从功能规范的观点来看可以使用高速缓冲存储器控制功能，服务器 200 就选择存储子系统 100 所拥有的高速缓冲存储器 120 或服务器 200 所拥有的高速缓冲存储器作为保存目标存储区域中数据的高速缓冲存储器。
5

如果存储子系统 100 的高速缓冲存储器 120 中保存了存储区域中所存储的数据，服务器 200 就指令目标存储子系统 100 在高速缓冲存储器 120 中保存目标存储区域中所存储的数据。另一方面，如果存储子系统 100 没有高速缓冲存储器控制功能，或者从存储子系统 100 的功能规范的观点来看不能使用该高速缓冲存储器功能，或者如果选择在服务器 200 的高速缓冲存储器中保存数据，服务器 200 就把目标存储区域中存储的数据保存在服务器 200 所拥有的高速缓冲存储器中。
10

作为上述处理的结果，能够通过下列措施优化性能：例如，把访问频率低的数据存储在速度相对较慢的磁盘单元 550 中；把访问频率高的数据存储在快速磁盘单元 550 中；把访问频率更高的数据存储在速度更快的高速缓冲存储器中。此外，如果服务器 200 使用存储子系统 100 的高速缓冲存储器控制功能，以条件（例如，每个高速缓冲存储器的性能、容量使用率）为基础使用存储子系统 100 的高速缓冲存储器 120 和服务器 200 的高速缓冲存储器能够优化每个高速缓冲存储器的使用效率。
15
20

在这个连接中，上述处理的目标是逻辑卷 500。但是，为组成虚拟卷 600 的逻辑卷 500 执行处理还能激活虚拟卷 600 的性能管理和优化处理。特别地，服务器 200 能够通过下列处理用更高的精度实现性能管理和性能调节：参考信息库 210 中记录的主机信息；分析与各个应用软件 310（例如，企业资源计划（ERP）、DBMS 和消息服务器软件）所用的虚拟卷 600 有关的信息，以及像虚拟卷 600 的区域、使用趋势和性能这样的信息；根据获得的分析结果和趋势预测执行处理。
25

此外，服务器 200 能够用元文件系统的信息为文件执行性能管理和优化处理。此外，如果服务器 200 包括与外部（例如，显示屏幕和管理接口）连接的输入输出装置，主机 300 的用户和系统管理员能够指定与要用的虚拟卷 600 的区域有关的信息、当前使用趋势、安排的
30

未来使用趋势、处理操作的安排、所需性能以及类似数据来指令服务器 200 以这些信息为基础执行性能管理性能调节。

服务器 200、主机 300 和控制终端 800 都可以有显示屏幕。另外，由主机 300 的用户和系统管理员按照上述方法把条件定义为策略能够促进对条件的管理。

(L) 安全管理

服务器 200 以下列不同种类的信息为基础管理用户的的数据访问权限和应用软件 310 的进程并确保数据安全：分配给主机 300 的用户的
10 数据访问权限、用户所用的虚拟卷 600 以及对应的虚拟卷 600 所拥有的存储区域之间的关系；分配给应用软件 310 的进程的数据访问权限、
应用软件 310 的进程所用的虚拟卷 600 以及对应的虚拟卷的存储区域之间的关系；由用户和进程使用虚拟卷的调度；以及类似信息。

15 主机 300 的用户和系统管理员通过外部输入装置（例如显示屏幕或管理接口）输入不同种类的信息到服务器 200 中。服务器 200 把这些输入的信息记录在信息库 210 中作为安全信息。服务器 200 用 WWN 等决定主机 300 中访问由控制器功能提供的虚拟卷 600 的主机端口，并拒绝没有访问权限的主机端口对虚拟卷 600 的访问。

20 另外，服务器 200 参考信息库 210 中存储的功能信息检查存储子 100 是否有安全控制功能。如果存储子系统 100 有安全控制功能，服务器 200 根据信息库 210 中记录的信息指令存储子系统 100 执行安全控制，并由此提供主机 300 和存储子系统 100 之间以及服务器 200 和存储子系统 100 之间的安全。

25 此外，如果需要对虚拟卷 600 进行数据加密，服务器 200 参考信息库 210 中的功能信息检查组成目标虚拟卷 600 的逻辑卷 500 所在的存储子系统 100 是否有数据加密功能，并检查数据加密功能的规范，例如加密算法和加密长度。

30 如果目标存储子系统 100 有数据加密功能，并且从功能规范的观点来看可以使用该数据加密功能，服务器 200 就指令目标存储子系统 100 依照加密有关信息对逻辑卷 500 中的数据加密。另一方面，如果目标存储子系统 100 没有数据加密功能，或者从功能规范的观点来看该数据加密功能不可用，就由服务器 200 加密数据；即使需要由服务器 200 进行数据加密，也尽可能多地减轻伴随数据加密处理的服务器

200 的负载。

注意如果服务器 200 用主机 300 对请求访问的用户和进程进行鉴别，服务器能够为用户和进程执行上面提到的处理以在用户基础和进程基础上管理和控制安全。例如，响应来自主机 300 的访问处理软件 330 的与文件和数据定位有关的查询，服务器 200 通知主机 300 访问处理软件用来请求访问的进程和执行该进程的用户以执行鉴别。然后，服务器 200 根据鉴别结果拒绝没有访问权限的用户和进程以提供安全。

此外，服务器 200 能够根据上面提到的调度信息通过改变上述与安全设置有关的信息而以日期和时间为基础上执行安全管理。

(M) 设备供应

在接收到主机 300 对分配虚拟卷 600 的请求时，服务器 200 用信息库 210 中记录的上述不同种类的信息以及从主机 300 通过代理 340 收集的不同种类的信息分析并预测所请求的虚拟卷 600 的使用及使用趋势以及所需要的条件，例如性能、可靠性、安全和功能，并由此分配适合所需条件的虚拟卷 600 给主机 300。

例如，如果所请求的虚拟卷 需要灾难恢复，服务器 200 参考信息库 210 中的功能信息来找到有远程复制功能的存储子系统 100。然后，服务器 200 用有远程复制功能的存储子系统 100 的逻辑卷 500 把虚拟卷 600 分配给主机 300。

另外，如果所请求的虚拟卷 600 是用快照备份的目标，服务器 200 用有快照功能的存储子系统 100 的逻辑卷 500 分配虚拟卷 600。作为上述处理的结果，在分配之后对主机 300 的处理能够减轻服务器 200 和 SAN 900 的负载。

此外，举例来说，如果希望所请求的虚拟卷 600 有低访问频率和低请求性能，虚拟卷 600 被分配给与速度相对较低的磁盘单元 550 相关联的逻辑卷 500。另一方面，如果希望所请求的虚拟卷 600 有高访问频和高请求性能，把与高速磁盘单元 550 相关的逻辑卷 500 分配给虚拟卷 600 能够在分配时激活性能优化。

注意如果主机 300 的用户和系统管理员定义所请求的虚拟卷 600 的使用及使用趋势以及所请求的上述条件（例如性能、可靠性、安全和功能）作为对促进条件管理的策略。

(N) 资源管理

使用信息库 210 中记录的资源信息和其它不同种类的信息，服务器 200 管理连接到 SAN 900 的存储子系统 100 的逻辑卷 500、端口、通路、高速缓冲存储器 120 以及磁盘单元 550，以及像服务器 200 的端口、通路和高速缓冲存储器这样不同种类资源的分类、属性、数量、使用数量和状态。另外，服务器 200 还管理使用这些资源的各个虚拟卷 600 的分类、属性、数量、使用数量和状态。

此外，服务器 200 通过服务器 200、主机 300 和控制终端 800 所拥有的显示屏幕以及服务器 200 提供的管理接口把上述管理信息提供给外部，并接受分配不同资源的指令。

(O) 数据管理

使用信息库 210 中记录的不同种类的信息以及与元文件系统有关的信息，服务器 200 管理与存储子系统 100 和备份存储设备 700 中所存储的数据和文件的归类、属性、创建的日期和时间、变为多余的时间、存储的位置（存储子系统 100、逻辑卷 500、存储区域、物理存储设备等）有关的信息。

分类和属性包括与下列有关的信息：正式数据、备份数据、副本、快照等的分类和生成；所需的性能和可靠性；使用这些数据和文件的主机 300 和应用软件 310（备份软件、数据库管理系统等）；以及类似信息。

此外，服务器 200 通过服务器 200、主机 300 和控制终端 800 所拥有的显示屏幕和服务器 200 提供管理接口向外部提供上述管理信息，并接收属性设置和类似信息。

(P) 故障管理

服务器 200 从与 SAN 900 相连的存储子系统 100 和其它不同种类的设备收集故障信息。另外，服务器 200 还从与 SAN 900 相连的存储子系统 100 等设备接收故障通知并管理它。服务器 200 把收集到和接收到的故障信息记录在信息库 210 中，然后把这些信息通知给主机 300 的用户和系统管理员。

另外，如果服务器 200 有用于根据故障的位置、种类、状态和异常条件假定原因和措施的信息，服务器 200 可以向主机 300 一起通知识别出的原因和措施。服务器 200 执行通知并通过服务器 200、主机 300

和控制终端 800 所拥有的显示屏幕以及服务器 200 所提供的管理接口提供故障信息。

此外，服务器 200 通过显示屏幕和管理接口从主机 300 的显示屏幕和系统管理员接收清单指令。接收到清单指令的服务器 200 诊断被 5 指令的设备为清单目标，然后把结果报告给用户和系统管理员。

(Q) 布局管理

服务器 203 用信息库 210 中记录的布局信息和其它不同种类的信息管理连接到 SAN 900 的设备和端口的布局、状态、分区、范围等。

服务器 200 通过服务器 200、主机 300 和控制终端 800 所拥有的显示 10 屏幕和服务器 200 所提供的管理接口预先设置上述管理信息给外部设备。另外，服务器 200 接收布局、分区和范围的设置中的变化，并指令主机 300 的代理 340、存储子系统 100 和交换机 400 根据改变的内容的设置的内容设置端口和分区。

(R) 主机管理

服务器 200 管理下列不同种类的信息：信息库 210 中记录的主机 15 信息；主机 300 的主机总线适配器和端口的配置；端口和所用虚拟卷 600 的关系；以及类似信息。另外，服务器通过服务器 200、主机 300 或控制终端 800 所拥有的显示屏幕以及服务器 200 所提供的管理接口 提供上述管理信息给外部。

(S) 优点（属性）管理

服务器 200 用信息库 210 中记录的设备信息和其它不同种类的信息管理存储子系统 100、备份存储设备 700、交换机 400 和控制终端 800 的安装位置、大小、安装日期、所有者、获得的数量以及优点管理状态。此外，服务器 200 通过服务器 200、主机 300 和控制终端 800 所 25 拥有的显示屏幕和服务器 200 提供的管理接口把上述管理信息提供给 外部，还从外部接收管理信息的设置、更新和内容。

(T) 许可管理

使用信息库 210 中记录的主机信息和其它不同种类的信息，服务 30 器 200 管理服务器 200 提供的不同种类功能的许可状态、许可范围和 时期等（这些信息被提供给主机 300）以及由存储子系统 100 提供的 不同种类功能的用户、许可状态、许可的范围和时期等信息。另外， 服务器 200 通过服务器 200、主机 300 和控制终端 800 所拥有的显示

屏幕和服务器 200 提供的管理接口把上述管理信息提供给外部。

(U) 报告输出/警告通知

5 服务器 200 收集服务器 200 中保存的不同种类的信息作为报告，
这些信息特别包括资源信息、卷集中控制信息、性能信息及其变化、
性能问题、性能瓶颈分析、负载分发和分配计划、数据物理定位优化
计划以及对涉及应用软件 310 的虚拟卷 600 的使用趋势的预测。然后，
服务器 200 把收集到的信息提供给主机 300 的用户和系统管理员。

10 另外，服务器 200 分析并总结与资源的使用状态、性能等有关的
信息。如果这些信息超出了阈值，服务器 200 就警告主机 300 以及主
机 300 的用户和系统管理员有情况发生。服务器 200 通过服务器 200、
主机 300 和控制终端 800 所拥有的显示屏幕和服务器 200 提供的管理
接口把报告和警告输出到外部。

(V) 管理接口

15 服务器 200 向外部设备提供连接到服务器 200 的设备用来设置并
控制服务器 200 所提供的不同种类功能的协议，以及连接到服务器 200
的设备用来参考和设置服务器 200 所保存的不同种类信息的软件接
口。这些软件接口包括：使用网络 910 的 MIB/SNMP 和 CIM/XML；由
READ/WRITE/INQUIRY/MODE SENSE/MODE SELECT 命令根据 SCSI 标准用
20 SAN 9—进行的数据传输；以及确定的协议。另外，服务器 200 还有用
于使用它们的 API、CLI 等。

25 此外，服务器 200、主机 300 和控制终端 800 能够用软件接口控
制它们自己的显示屏幕。因为服务器 200 向主机 300 提供服务器 200
整体上所拥有的不同种类的功能，还因为服务器 200 在整体上提供上
面提到的协议和管理屏幕，只有由主机 300 的用户和系统管理员对服
务器 200 的管理足以使用这些功能。说得更明确些，能够促进系统管
理，导致系统管理开销下降。

30 另外，因为服务器 200 在整体上包括不同种类的功能并把它们提
供给主机 300，因此可以把提供每种功能所需的信息收集在信息库 210
中，这能够减少信息传输处理、排除处理和类似处理的负担，这是在
提供各种功能时获得并更新信息所需要的。

此外，因为服务器 200 在整体上包括不同种类的功能并把它们提
供给主机 300，因此能够避免用来使用存储子系统 100 的各种功能的

指令冲突以及多个计算机间的使用冲突，存储子系统 100 用于提供服务器 200 的功能。说得更明确些，能够减轻使用存储子系统 100 时排除处理的负担。

作为上述处理的结果，服务器 200 能够调节有相应功能的多个存储子系统 100 之间多个功能的差异以及功能规范的差异，并从整体上关联多个存储子系统 100 的功能以向主机 300 提供这些功能。
5

另外，作为上述处理的结果，使用服务器 200，用户能够执行包括性能管理和安全管理在内的操作和管理，在这些操作和管理中存储子系统 100 中包括的多个功能被从整体上相互关联。
10

另外，作为上述处理的结果，服务器 200 能够根据主机 300 的信息和主机 300 上的应用控制多个存储子系统间功能的关联以及多个功能之间的关联。
15

此外，作为上述处理的结果，当服务器 200 向主机 300 提供用于减轻主机 300 和 SAN 900 的负载的不同种类的功能、促进系统管理、确保安全以及类似操作时，服务器 200 能够减轻虚拟化存储区域的服务器 200 自身的负载。
20

此外，作为上述处理的结果，服务器 200 能够以虚拟卷 600 为基础控制存储子系统 100 中所包括的功能。
25

在上述描述中，服务器 200 被配置为单独的设备或计算机。但是，从配置的观点来看，服务器 200 可以由互相连接的多个设备或计算机组成，上面提到的功能可以在这些设备或计算机之间进行分配。例如，服务器 200 还可以配置为有包含硬件和软件的计算机，硬件有控制器功能，专门用于处理对虚拟卷 600 的访问，软件执行有交换机结构的设备的不同种类的管理和控制。在这种配置情况下，服务器 200 可以采用在每种功能、管理和控制中进行了优化的体系结构。其结果是，可以提高 I/O 性能、每种功能、管理和控制的性能，最后导致每种处理的负载减轻。
30

依照本发明能够提供把服务器和存储子系统中所包括的功能相互关联以便实际向外部设备提供功能的计算机系统。

提供上面的详细描述是为了描述本发明的特定实施方案但并不是为了限制本发明。可以有很多在本发明范围内的修改和变体。因此，本发明由所附权利要求来定义。

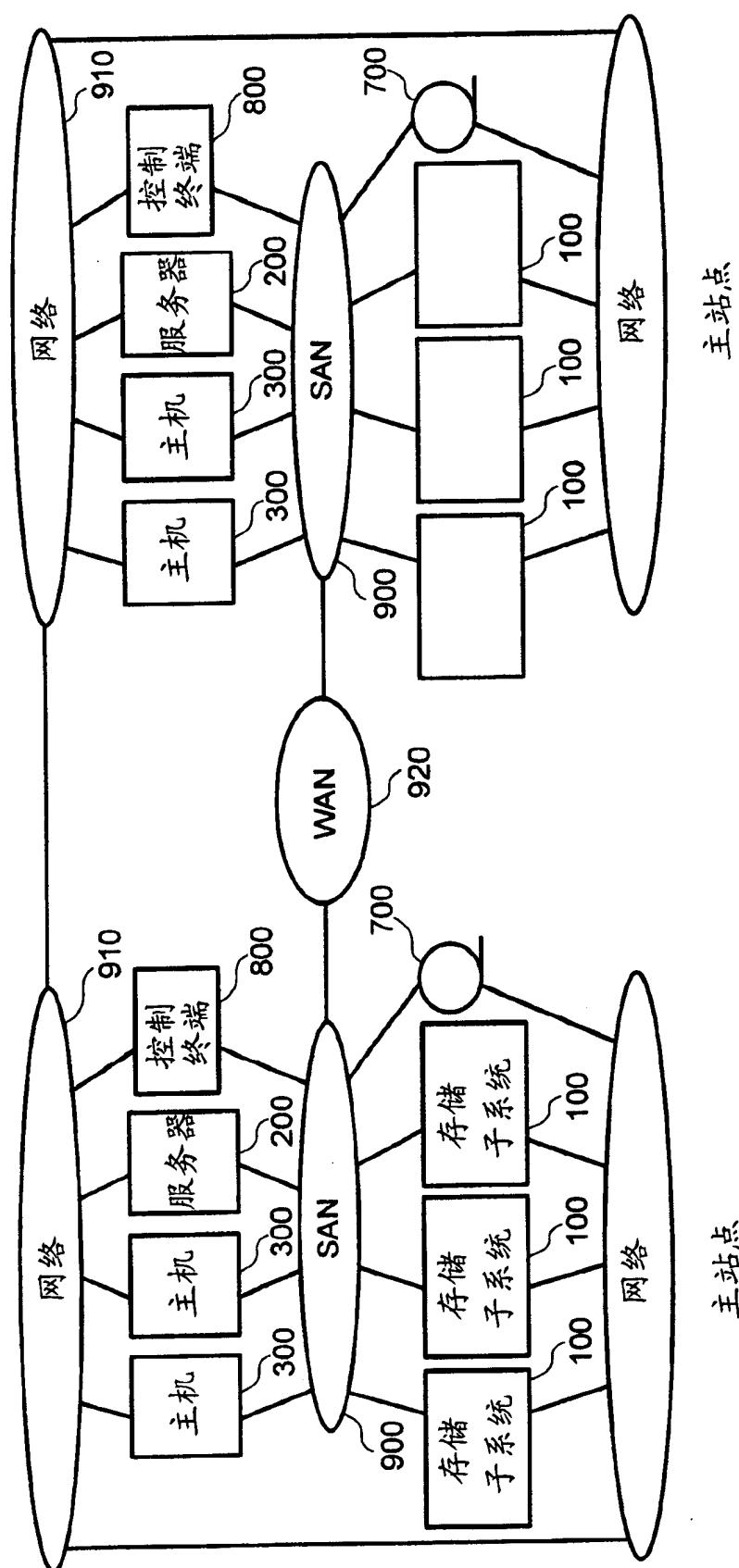


图 1

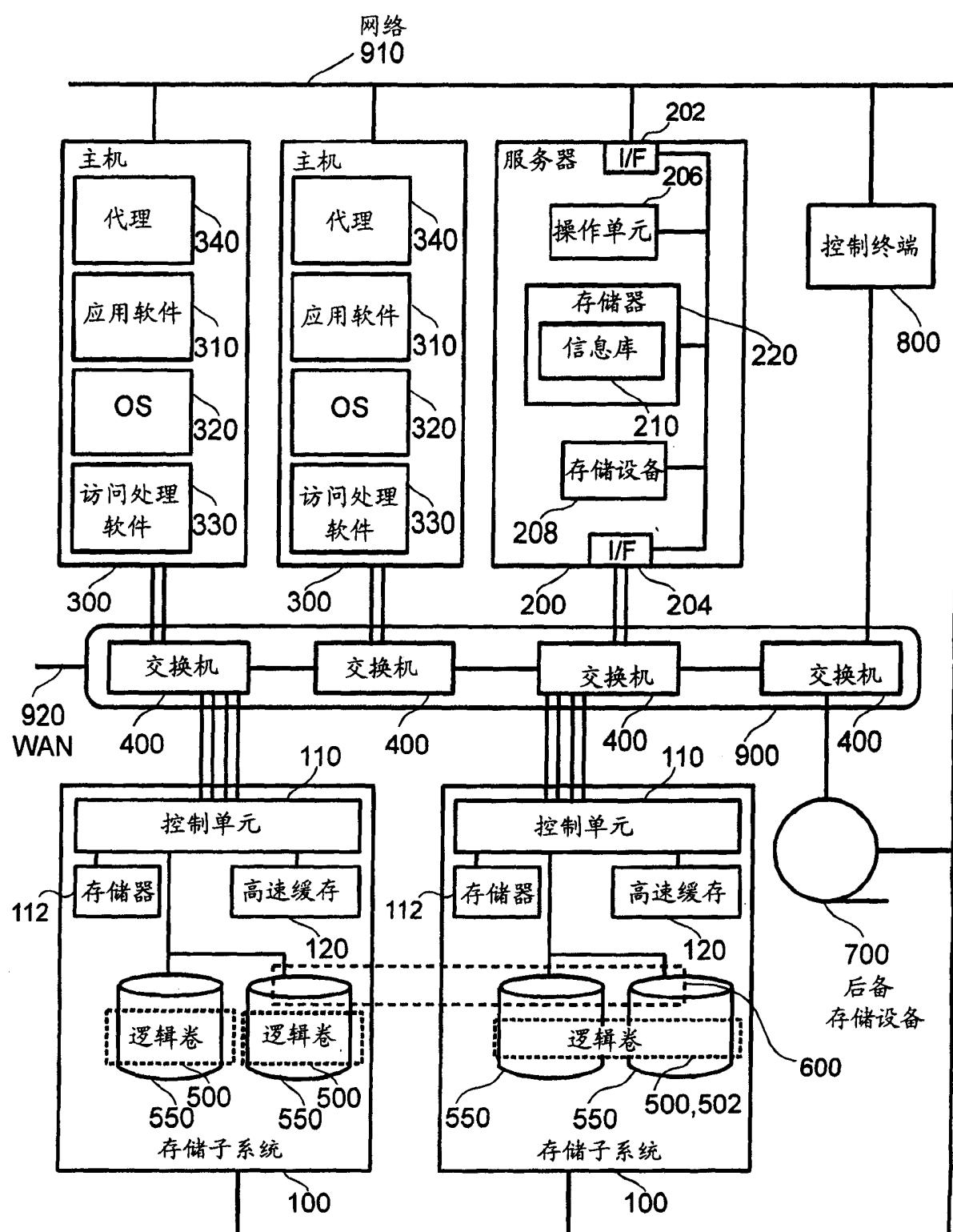


图 2

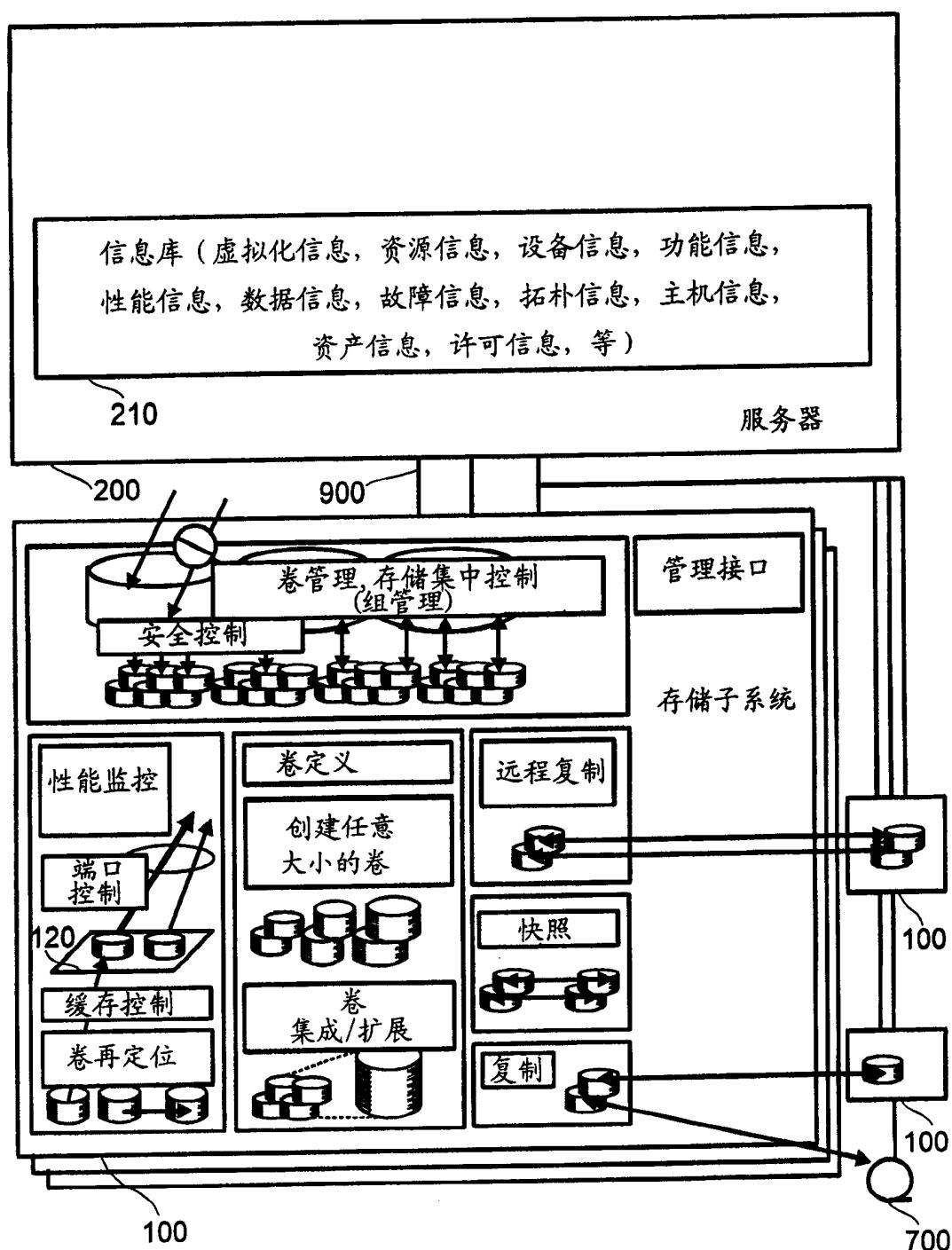


图 3

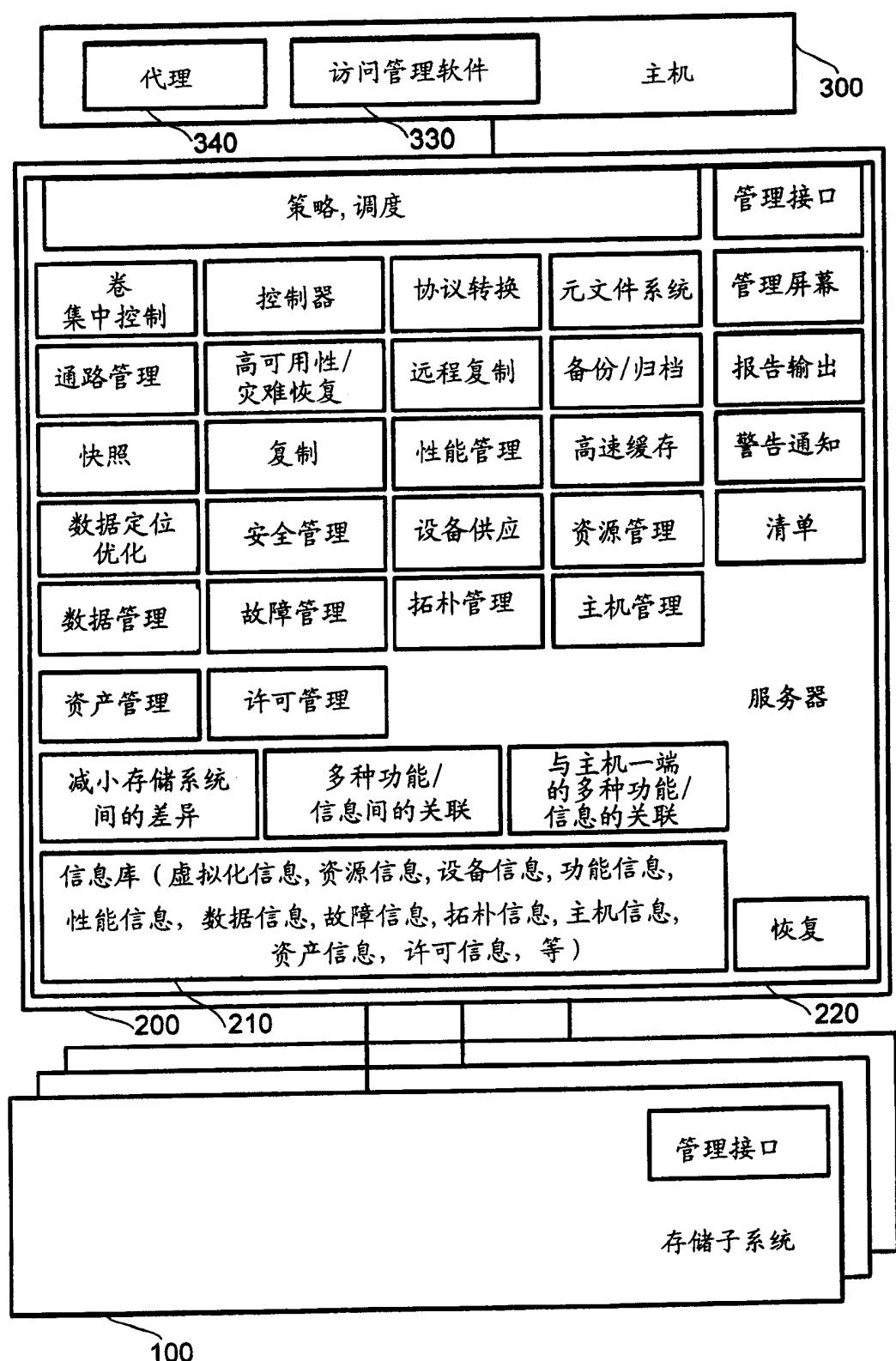


图 4

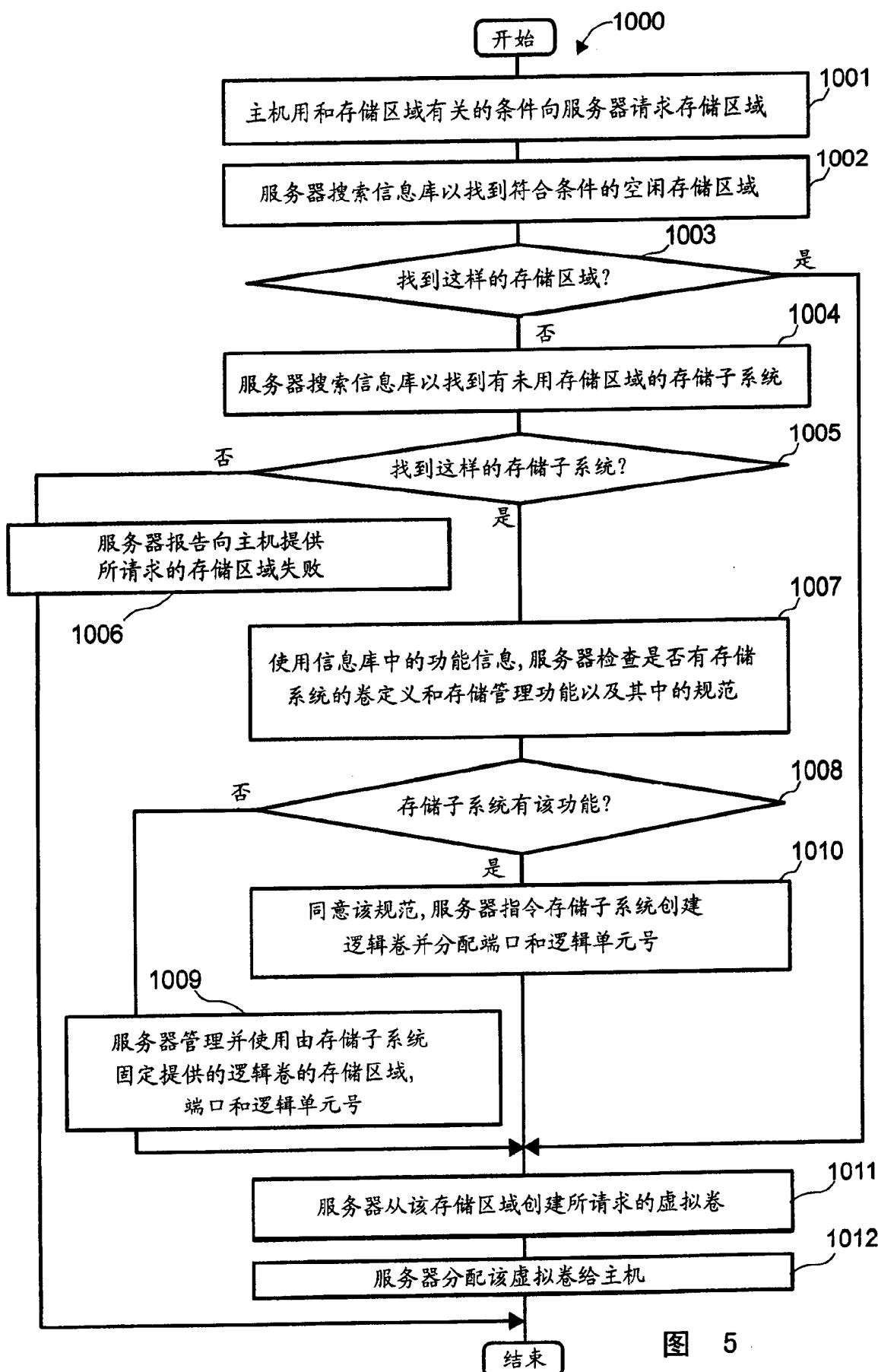


图 5

虚拟逻辑卷	虚拟存储区域大小	存储子系统号	逻辑卷号	存储区域地址	存储区域大小
0	4000	1	2	10000	4000
1	3000	3	1	0	1000
		2	3	5000	2000

图 6

提供给主机的通路地址	逻辑单元号	虚拟逻辑卷号
WWN = 00:11:22:33:44:55:66:77 Port Address = 1	1	0
WWN = 00:10:20:30:40:50:60:70 Port Address = 2	0	1

图 7

通路地址	逻辑单元号	存储子系统号	虚拟逻辑卷号
WWN = 01:02:03:04:05:06:07:08 Port Address = 11	2	1	0
WWN = 11:22:33:44:55:66:77:88 Port Address = 12	3	3	1
WWN = 21:22:23:24:25:26:27:28 Port Address = 13	1	2	1

图 8

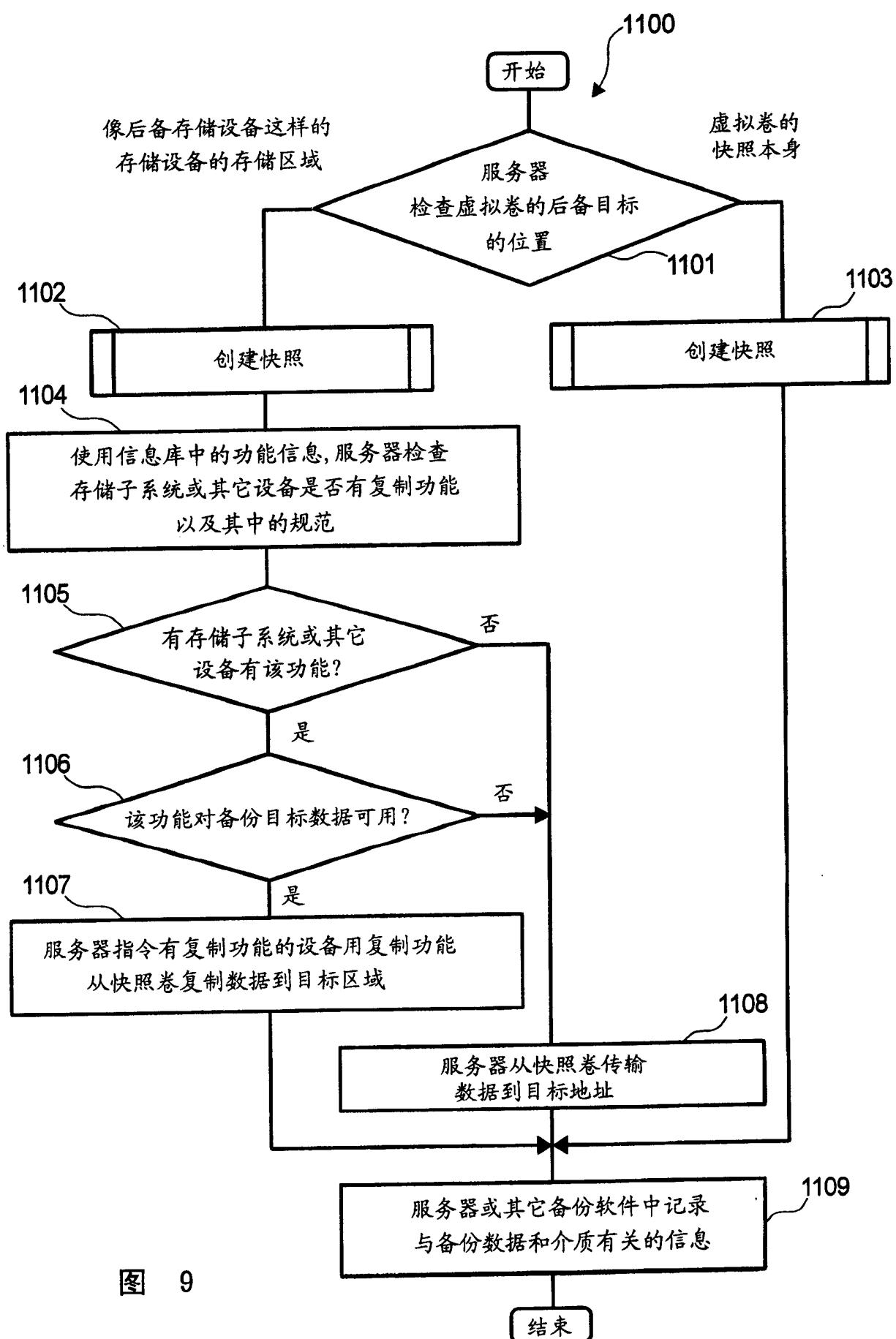


图 9

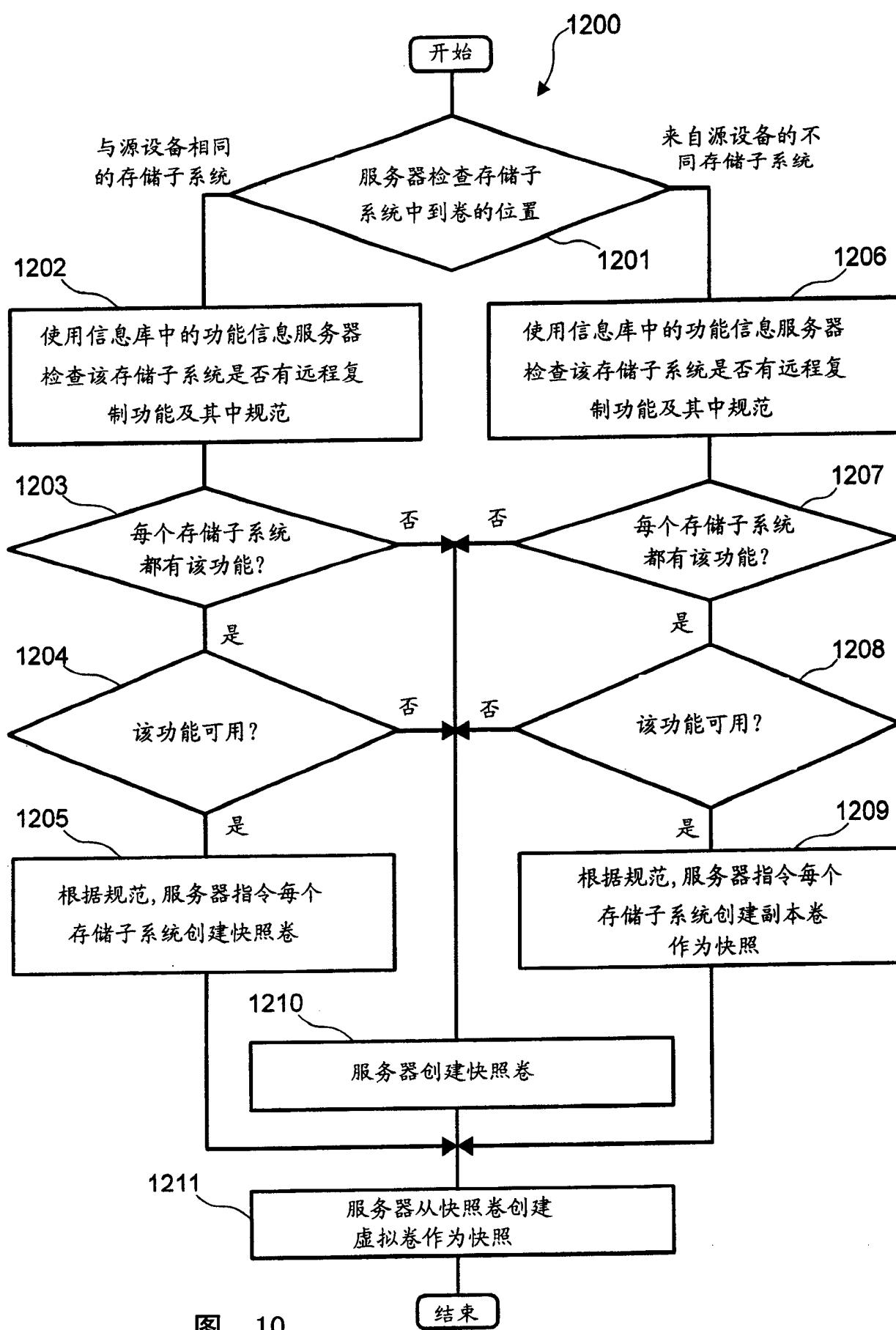


图 10

150

存储子系统编号	复制功能	功能可用性	类型	可用复制	最大目标数	当前操作数
0	Exist	Enable	CP_VENDOR_A_TYPE_A	Disk_to_Tape Tape_to_Disk	128	512
1	Exist	Enable	CP_VENDOR_A_TYPE_B	Disk_to_Tape Tape_to_Disk Disk_to_Disk	256	2048
2	Exist	Disable	CP_VENDOR_B	Disk_to_Tape	64	128
3	Not exist	-	-	-	-	-

166

152 154 156 158 160 162 164

图 11

250

存储子系统编号	快照功能	功能可用性	类型	最大对数	一个卷的最大副卷数	当前操作数
0	Exist	Enable	SS_VENDOR_A_TYPE_A	1024	4	512
1	Exist	Enable	SS_VENDOR_A_TYPE_B	2048	9	2048
2	Exist	Disable	SS_VENDOR_B	512	1	128
3	Not exist	-	-	-	-	-

252 254 256 258 260 262 264

图 12

350

存储子系统编号	远程复制功能	功能可用性	类型	最大对数	一个卷的最大副卷数	当前操作数
0	Exist	Enable	RR_VENDOR_A_TYPE_A	512	3	128
1	Exist	Enable	RR_VENDOR_A_TYPE_B	1024	8	256
2	Exist	Disable	RR_VENDOR_B	256	1	64
3	Not exist	-	-	-	-	-

366

352 354 356 358 360 362 364

图 13