

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 9/46 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200680001584.8

[43] 公开日 2007 年 12 月 19 日

[11] 公开号 CN 101091165A

[22] 申请日 2006.1.17

[21] 申请号 200680001584.8

[30] 优先权

[32] 2005.1.18 [33] US [31] 11/037,421

[86] 国际申请 PCT/EP2006/050260 2006.1.17

[87] 国际公布 WO2006/077215 英 2006.7.27

[85] 进入国家阶段日期 2007.6.26

[71] 申请人 国际商业机器公司

地址 美国纽约

[72] 发明人 约汉·阿绍夫 尼娜·切里安
布赖恩·斯密斯

[74] 专利代理机构 中国国际贸易促进委员会专利商
标事务所
代理人 李德山

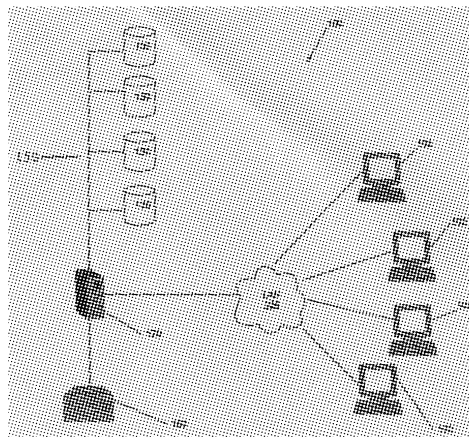
权利要求书 4 页 说明书 16 页 附图 6 页

[54] 发明名称

提供虚拟存储子系统内虚拟存储分配的自动
性能优化的方法、装置和程序存储设备

[57] 摘要

揭示了用于在多个虚拟存储子系统中管理存储分配的卷供应顾问装置。从用户接收到针对数据存储和工作负载要求的请求。接收虚拟化引擎内的存储控制器和被管理的盘的性能信息。基于所接收的性能信息和可用的被管理盘定义资源组，并且它们的利用率被跟踪。基于用户工作负载要求和现有或历史的资源组利用率预测虚拟化引擎内的资源组性能。基于预测的资源组性能，存储被分配给虚拟化引擎内的一个或多个被管理的盘。



1.一种程序存储设备, 包括:

可由处理设备运行以执行用于管理虚拟存储系统中的存储分配的操作的程序指令, 所述操作包括:

评估工作负载简档;

确定被管理的盘的性能特征;

确定被管理的盘与资源组之间的关系; 以及

基于被管理的盘被分配给的资源组, 推荐包括一组被管理的盘的至少一个虚拟盘的分配。

2.如权利要求1所述的程序存储设备, 其中操作进一步包括:

定义工作负载简档。

3.如权利要求1或2所述的程序存储设备, 其中操作进一步包括:

基于被管理的盘被分配到的资源组, 创建包括一组被管理的盘的虚拟盘。

4.如权利要求1、2或3所述的程序存储设备, 其中确定关系的操作基于用户输入。

5.如权利要求1到3的任何一个所述的程序存储设备, 其中确定关系的操作基于自动提供的输入。

6.如权利要求3所述的程序存储设备, 其中创建虚拟盘进一步包括基于被管理的盘被分配到的资源组, 在该组被管理的盘内创建多个虚拟盘。

7.如前面权利要求的任何一个所述的程序存储设备, 进一步包括:

通过识别与一组被管理的盘和资源组相关、包括针对特定工作负载的吞吐能力的一组性能特性, 选择虚拟盘的位置; 以及

基于与一组被管理的盘和资源组相关的性能特性, 估计最优资源利用率。

8.如权利要求2所述的程序存储设备, 其中定义工作负载简档进

一步包括从一个组中选择至少一个，该组包括期望存储量、工作负载和期望吞吐水平的特征的标识、以及分配提示。

9.如前面权利要求的任何一个所述的程序存储设备，进一步包括获得被管理的盘的性能能力。

10.如权利要求9所述的程序存储设备，其中获得被管理的盘的性能能力进一步包括提供表明被管理的盘的性能能力的人工输入。

11.如权利要求9所述的程序存储设备，其中获得被管理的盘的性能能力进一步包括从被管理盘配置器接收性能能力。

12.如权利要求11所述的程序存储设备，其中从被管理盘配置器接收性能能力进一步包括获得有关被管理的盘及其相关的资源组的性能特征的信息，其中所述资源组识别被管理的盘之间的潜在关系。

13.如权利要求9所述的程序存储设备，其中获得被管理的盘的性能能力进一步包括执行受控校准和利用来自受控校准的结果。

14.如权利要求13所述的程序存储设备，其中执行受控校准包括针对被管理的盘运行指定的I/O负载以识别被管理的盘的行为。

15.如权利要求9所述的程序存储设备，其中获得被管理的盘的性能能力进一步包括执行不受控校准和利用来自不受控校准的结果。

16.如权利要求15所述的程序存储设备，其中执行不受控校准包括随着应用程序作为非控制负载运行而分析性能行为，以识别被管理的盘的行为。

17.如权利要求9所述的程序存储设备，其中获得被管理的盘的性能能力进一步包括通过接口接收来自用户的输入。

18.如前面权利要求的任何一个所述的程序存储设备，其中确定被管理的盘的性能特征进一步包括识别被管理的盘所属于的资源组。

19.如权利要求17所述的程序存储设备，其中用户输入包括从被管理的盘的性能简档中进行的選擇。

20.如权利要求17所述的程序存储设备，其中用户输入包括从资源组的性能简档中进行的選擇。

21.一种方法，包括：

评估工作负载简档;
确定被管理的盘的性能特征;
确定被管理的盘与资源组之间的关系; 以及
基于被管理的盘被分配给的资源组, 推荐包括一组被管理的盘的一个或多个虚拟盘的分配。

22.如权利要求21所述的方法, 进一步包括:

定义工作负载简档。

23.如权利要求21或22所述的方法, 进一步包括:

基于被管理的盘被分配到的资源组, 创建包括一组被管理的盘的虚拟盘。

24.如权利要求21、22或23所述的方法, 其中确定关系的操作基于用户输入。

25.如权利要求21到23中的任何一个所述的方法, 其中确定关系的操作基于自动提供的输入。

26.如权利要求23所述的方法, 其中创建虚拟盘进一步包括基于被管理的盘被分配到的资源组, 在该组被管理的盘内创建多个虚拟盘。

27.如权利要求21到26中的任何一个所述的方法, 进一步包括:

通过识别与一组被管理的盘和资源组相关、包括针对特定工作负载的吞吐能力的一组性能特性, 选择虚拟盘的位置; 以及

基于与一组被管理的盘和资源组相关的性能特性, 估计最优资源利用率。

28.如权利要求22所述的方法, 其中定义工作负载简档进一步包括从一个组中选择至少一个, 该组包括期望存储量、工作负载和期望吞吐水平的特征的标识、以及分配提示。

29.如权利要求21到28中的任何一个所述的方法, 进一步包括获得被管理的盘的性能能力。

30.如权利要求29所述的方法, 其中获得被管理的盘的性能能力进一步包括提供表明被管理的盘的性能能力的人工输入。

31.如权利要求29所述的方法, 其中获得被管理的盘的性能能力进

一步包括从被管理盘配置器接收性能能力。

32.如权利要求31所述的方法,其中从被管理盘配置器接收性能能力进一步包括获得有关被管理的盘及其相关的资源组的性能特征的信息,其中所述资源组识别被管理的盘之间的潜在关系。

33.如权利要求29所述的方法,其中获得被管理的盘的性能能力进一步包括执行受控校准和利用来自受控校准的结果。

34.如权利要求33所述的方法,其中执行受控校准包括针对被管理的盘运行指定的I/O负载以识别被管理的盘的行为。

35.如权利要求29所述的方法,其中获得被管理的盘的性能能力进一步包括执行不受控校准和利用来自不受控校准的结果。

36.如权利要求35所述的方法,其中执行不受控校准包括随着应用程序作为非控制负载运行而分析性能行为,以识别被管理的盘的行为。

37.如权利要求29所述的方法,其中获得被管理的盘的性能能力进一步包括通过接口接收来自用户的输入。

38.如权利要求21到37中的任何一个所述的方法,其中确定被管理的盘的性能特征进一步包括识别被管理的盘所属于的资源组。

39.如权利要求37所述的方法,其中用户输入包括从被管理的盘的性能简档中进行的選擇。

40.如权利要求37所述的方法,其中用户输入包括从资源组性能简档中进行的選擇。

41. 一种用于管理虚拟存储系统中的存储分配的卷供应顾问装置,该卷供应顾问装置能够执行如权利要求21到40中的任何一个所述的方法。

42. 如权利要求41所述的卷供应顾问装置,其中资源组包括名字、性能属性、以及属于该资源组的一组被管理的盘。

43. 如权利要求41或42所述的卷供应顾问装置,其中被管理的盘属于一个或多个资源组,其中资源组识别被管理的盘之间的潜在关系。

44. 一种计算机程序,包括当所述程序在计算机上运行时,适于执行如权利要求21到40中的任何一个所述的方法的程序代码装置。

提供虚拟存储子系统内虚拟存储分配的 自动性能优化的方法、装置和程序存储设备

技术领域

本发明通常涉及网络存储系统，并且尤其涉及提供虚拟存储子系统内虚拟存储分配的自动性能优化的方法、装置和程序存储设备。

背景技术

在诸如可被用于公司、政府机构或其它实体的企业数据处理系统中，信息通常被存储在服务器上并且被用户通过例如网络访问。信息可包括将被处理的程序和/或数据的任何类型信息。利用其个人计算机、工作站等（通常是“计算机”）的用户将允许其计算机检索将被处理的信息，并且将信息存储在例如远程服务器上。

通常，服务器将数据存储在通常包括若干盘存储单元的大容量存储子系统中。数据被存储在诸如文件的单元中。在服务器中，文件可被存储在一个盘存储单元上，或者文件的各部分可被存储在若干盘存储单元上。服务器可并行地服务来自若干用户的访问请求，并且可以理解，并行服务的访问操作优选地与分布在多个盘存储单元上的信息相关，使得它们可被并行地服务。除非另有说明，通常期望以这种方式将信息存储在盘存储单元中，即使得一个盘驱动单元不会负载过重或忙于为访问提供服务，而其它的则负载较轻或空闲。

企业的计算机网络可以具有多个彼此远离并且远离企业用户的存储网络。存储网络也可以在不同类型的系统上驻留。为正确地执行任务，企业用户可要求对包含于所有存储网络中的数据进行快速并且可靠的访问。信息技术（IT）雇员必须能够给企业用户提供高速、可靠的访问。

存储区网络(SAN)是将数据存储设备逻辑连接到服务器的高速、

高带宽存储网络。企业用户通常通过服务器被连接到数据存储设备。SAN延伸了由常规服务器/存储设备连接提供的概念，并且提供了更多灵活性、可用性、集成管理和性能。SAN是允许用户随时访问企业中的任何信息的第一IT解决方案。通常，SAN包含用于定义诸如主机、互连设备、存储设备和网络附连服务器(NAS)设备的网络设备的管理软件。SAN管理软件也允许设备之间定义链接。SAN软件内可以定义虚拟存储，其中在若干存储盘上存储数据，而存储盘表现为单个虚拟盘的特征。

达到提供高速、可靠访问给企业用户的这个目标的一个重要因素是允许SAN被那些设计和维护SAN的人完全理解。由于其复杂度，通常难以迅速理解SAN。允许迅速理解和改变SAN和SAN内的虚拟系统的配置的工具是有好处的。

SAN的优点之一是消除了可能发生在服务器上的瓶颈，所述服务器管理若干客户端的存储访问。通过允许对存储的共享访问，SAN可提供较低的数据访问时延和改进的性能。然而，在诸如SAN附连存储的大型存储网络中，由于能够包含若干虚拟存储子系统的网络的复杂度、分析工作负载的复杂度，对于存储管理器来说很难知道将存储的增量分配到何处，使得新分配的空间达到可能最好的性能，而且物理存储属性可对应用程序隐藏。

过去，大型存储环境的存储分配被人工执行。已知有能够基于若干算法分配或推荐将存储分配给何处的存储管理软件。然而，这些算法实际上没有尝试满足包括SAN内的虚拟存储系统的可用存储的约束内的产品性能要求。

发明内容

本发明揭示了用于提供在虚拟存储子系统内虚拟存储分配的自动性能优化的方法、装置和程序存储设备。

根据第一个方面，提供了程序存储设备，包括：可由处理设备运行以执行用于管理虚拟存储系统中的存储分配的操作的程序指令，所

述操作包括：评估工作负载简档；确定被管理的盘的性能特征；确定被管理的盘与资源组之间的关系；以及基于被管理的盘被分配给的资源组，推荐包括一组被管理的盘的至少一个虚拟盘的分配。

在一个实施例中，操作进一步包括定义工作负载简档。

在一个实施例中，操作进一步包括基于被管理的盘被分配到的资源组，建立包括一组被管理的盘的虚拟盘。

在一个实施例中，确定关系的操作基于用户输入。

在一个实施例中，确定关系的操作基于自动提供的输入。

在一个实施例中，建立虚拟盘进一步包括基于被管理的盘被分配到的资源组，在该组被管理的盘内建立多个虚拟盘。

在一个实施例中，通过识别与一组被管理的盘和资源组相关、包括针对特定工作负载的吞吐能力的一组性能特性，选择虚拟盘的位置；以及基于与一组被管理的盘和资源组相关的性能特性，估计最优资源利用率。

在一个实施例中，定义工作负载简档进一步包括从一个组中选择至少一个，该组包括期望存储量、工作负载和期望吞吐水平的特征的标识、以及分配提示。

在一个实施例中，获得被管理的盘的性能能力。

在一个实施例中，获得被管理的盘的性能能力进一步包括提供表明被管理的盘的性能能力的人工输入。

在一个实施例中，获得被管理的盘的性能能力进一步包括从被管理盘配置器接收性能能力。

在一个实施例中，从被管理盘配置器接收性能能力进一步包括获得有关被管理的盘及其相关的资源组的性能特征的信息，其中所述资源组识别被管理的盘之间的潜在关系。

在一个实施例中，获得被管理的盘的性能能力进一步包括执行受控校准，以及利用来自受控校准的结果。

在一个实施例中，执行受控校准包括针对被管理的盘运行指定的I/O负载以识别被管理的盘的行为。

在一个实施例中，获得被管理的盘的性能能力进一步包括执行不受控校准，以及利用来自不受控校准的结果。

在一个实施例中，执行不受控校准包括随着应用程序作为非控制负载运行而分析性能行为，以识别被管理的盘的行为。

在一个实施例中，获得被管理的盘的性能能力进一步包括通过接口接收来自用户的输入。

在一个实施例中，确定被管理的盘的性能特征进一步包括识别被管理的盘所属于的资源组。

根据一个实施例，提供了一种程序存储设备，包括：可由处理设备运行以执行用于管理虚拟存储系统中的存储分配的操作的程序指令，所述操作包括：定义工作负载简档；确定被管理的盘的性能特征；确定被管理的盘与资源组之间的关系；以及基于被管理的盘被分配到的资源组，建立包括一组被管理的盘的虚拟盘。

根据一个实施例，提供了用于管理虚拟存储系统中的存储分配的设备。该设备包括用于存储存储系统信息的存储器，以及连接到存储器的处理器。处理器被配置成提供用于定义工作负载简档的用户接口，以及虚拟盘分配器，用于确定被管理的盘的性能特征，基于用户定义的或自动的输入来确定被管理的盘与资源组之间的关系，以及根据被管理的盘被分配到的资源组来建立包括一组被管理的盘的虚拟盘。

在一个实施例中，根据被管理的盘被分配到的资源组，处理器在该组被管理的盘内建立一个或多个虚拟盘。

在一个实施例中，通过识别与一组被管理的盘和资源组相关、包括针对特定工作负载的吞吐能力的一组性能特性，以及基于与一组被管理的盘和资源组相关的性能特性估计最优资源利用率，处理器分配器推荐虚拟盘的位置。

在一个实施例中，工作负载简档包括从一个组中选择的至少一个，该组包括期望存储量、工作负载和期望吞吐水平的特征的标识、以及分配提示。

在一个实施例中，处理器提供了图形接口。

在一个实施例中，处理器提供了利用命令行接口的脚本应用程序。

在一个实施例中，处理器提供了利用应用编程接口(API)的脚本应用程序。

在一个实施例中，处理器利用来自包括人工输入、被管理盘配置器、受控校准结果、不受控校准结果以及标准接口输入的组的至少一个来确定被管理的盘的性能能力。

在一个实施例中，人工输入包括从被管理的盘的性能简档中进行的选择。

在一个实施例中，人工输入包括从资源组的性能简档中进行的选择。

在本发明一个实施例中，提供了一种用于管理虚拟存储系统中的存储分配的方法。该方法包括获得用户定义的工作负载简档，确定被管理的盘的性能特征，基于用户定义的或自动的输入确定被管理的盘与资源组之间的关系，以及根据被管理的盘被分配到的资源组来建立包括一组被管理的盘的虚拟盘。

根据另一个方面，提供了一种方法，包括：评估工作负载简档；确定被管理的盘的性能特征；确定被管理的盘与资源组之间的关系；以及基于被管理的盘被分配给的资源组，推荐包括一组被管理的盘的至少一个虚拟盘的分配。

根据一个实施例，提供了一种用于管理虚拟存储系统中的存储分配的卷供应顾问装置(volume provisioning advisor)。该卷供应顾问装置包括用于定义工作负载简档的用户接口，以及连接到用户接口的虚拟盘分配器，所述虚拟盘分配器用于确定被管理的盘的性能特征，基于用户定义的或自动的输入确定被管理的盘与资源组之间的关系，以及根据被管理的盘被分配到的资源组建立包括一组被管理的盘的虚拟盘。

在一个实施例中，工作负载简档包括从一个组中选择的至少一个，该组包括期望存储量、工作负载和期望吞吐水平的特征的标识、

以及分配提示。

在一个实施例中，资源组包括名字、性能属性以及属于该资源组的一组被管理的盘。

在一个实施例中，被管理的盘属于一个或多个资源组，其中资源组识别被管理的盘之间的潜在关系。

根据另一个方面，提供了用于管理虚拟存储系统中的存储分配的卷供应顾问装置，包括：评估工作负载简档的工作负载评估器；以及连接到用户接口的虚拟盘分配器，所述虚拟盘分配器用于确定被管理的盘的性能特征，确定被管理的盘与资源组之间的关系，以及基于被管理的盘被分配到的资源组推荐包括一组被管理的盘的一个或多个虚拟盘的分配。

根据另一个实施例，提供了用于管理虚拟存储系统中的存储分配的卷供应顾问装置，包括：用于获得用户定义的工作负载简档的装置；用于确定被管理的盘的性能特征的装置；以及用于根据被管理的盘被分配到的资源组，推荐包括一组被管理的盘的一个或多个虚拟盘的分配的装置。

在本发明的另一个实施例中，提供了另一个卷供应顾问装置。卷供应顾问装置的该实施例包括用于获得用户定义的工作负载简档的装置，用于确定被管理的盘的性能特征的装置，用于基于用户定义的或自动的输入确定被管理的盘与资源组之间的关系的装置，以及用于根据被管理的盘被分配到的资源组而建立包括一组被管理的盘的虚拟盘的装置。

根据另一个实施例，提供了另一个卷供应顾问装置。该卷供应顾问装置包括用于获得用户定义的工作负载简档的装置，用于确定被管理的盘的性能特征的装置，以及用于根据被管理的盘被分配到的资源组而建立包括一组被管理的盘的虚拟盘的装置。

根据另一个实施例，提供了一种用于管理虚拟存储系统中的存储分配的方法，包括：获得用户定义的工作负载简档；确定被管理的盘的性能特征；根据被管理的盘被分配到的资源组，推荐包括一组被管

理的盘的虚拟盘的分配。

在一个实施例中，根据被管理的盘被分配到的资源组，建立包括一组被管理的盘的虚拟盘。

在一个实施例中，建立虚拟盘进一步包括识别与一组被管理的盘和资源组相关、包括针对特定工作负载的吞吐能力的一组性能特性，以及基于与一组所管理的盘和资源组相关的性能特性，估计最优资源利用率。

在一个实施例中，获得用户定义的工作负载简档进一步包括从一个组中选择至少一个，该组包括期望存储量、工作负载和期望吞吐水平的特征的标识、以及分配提示。

在一个实施例中，确定被管理的盘的性能能力包括利用一个组中的至少一个，该组包括人工输入、获得配置数据、受控校准结果、不受控校准结果以及标准接口输入。

附图说明

仅通过例子，而且参考以下附图（在整个附图中类似参考号码代表对应部分），现在将描述本发明的实施例：

图1图解了局域网形式的计算机网络；

图2根据本发明的一个实施例示出了存储区网络的一个实施例；

图3根据本发明的一个实施例图解了被引入虚拟盘分配器的属性表；

图4根据本发明的一个实施例图解了用户接口的机制，用于获得工作负载简档以供虚拟盘分配器在将被管理的盘分配给资源组时使用；

图5根据本发明的一个实施例图解了用于确定被管理的盘的性能能力的机制；

图6根据本发明的一个实施例图解了由虚拟盘分配器使用的、用于抽象虚拟存储子系统中的重要性能要素的数据结构；以及

图7根据本发明的一个实施例图解了用于管理虚拟存储系统中的

存储分配的方法的流程图。

具体实施方式

在实施例的以下描述中，针对组成本文的一部分的附图加以参考，而且其中通过图解示出了本发明可在其中被实践的特定的实施例。可以理解，在没有背离本发明的范围的情况下，因为可以做结构变化，所以可使用其它实施例。

本发明提供了用于在虚拟存储子系统内提供虚拟存储分配的自动性能优化的方法、装置和程序存储设备。

图1图解了具有局域网（LAN）的形式的计算机网络100。在图1中，工作站节点102经由LAN互连104被连接到服务器120。数据存储130经由数据总线150被连接到服务器120。LAN互连100可以是任何数量的网络拓扑，例如以太网。

图1中所示的网络被称作网络的客户-服务器模型。客户端是连接到网络、共享服务或其它资源的设备。服务器120管理这些服务或资源。服务器120是一个计算机或软件程序，其给客户端102提供服务。可由服务器管理的服务包括访问数据存储130、由服务器120提供的应用程序或其它连接的节点（未示出），或打印机共享160。

在图1中，工作站102是服务器120的客户端，并且共享访问由服务器120管理的数据存储130。当工作站102中的一个需要访问数据存储130时，工作站102经由LAN互连100向服务器120提交请求。服务器120对从工作站102访问数据存储130的请求提供服务。服务器与存储设备之间的可能的互连技术是光纤通道协议（FCP）或小型计算机系统接口（SCSI）。

随着诸如图1中所示的网络的增长，可以增加新的客户端102，可以增加更多的存储130，并且提供服务的要求可能增加。如上所述，服务器120将为访问存储130的所有请求提供服务。因此，存储130上的I/O负载可显著增加，并且服务器120上的工作负载可显著增加，其导致执行拒绝的可能。为帮助减少传统客户-服务器模型的带宽限制，存储区

网络(SAN)近年来已变得越来越普及。存储区网络以高速互连服务器和存储。通过将诸如LAN的现有网络模型与存储区网络组合,可以改进总的计算机网络的性能。

图2根据本发明的一个实施例示出了SAN 200的一个实施例。在图2中,服务器202经由SAN互连204被连接到数据存储设备230。每个服务器202和每个存储设备230都被连接到SAN互连204。服务器202直接访问连接到SAN互连的任何存储设备230。SAN互连204可以是诸如光纤通道或SCSI的高速互连。如图2所示,服务器202和存储设备230包括自身中的和自身的网络。

在图2的SAN 200中,不是像在LAN中那样,没有服务器202专用于特定的存储设备230。任何一个服务器202可访问图2中的SAN 200上的任何一个存储设备230。SAN 200的典型特征可包括高带宽、从服务器到存储节点的众多路径、较大的连接距离以及非常大的存储容量。因此,由于SAN的复杂度,基于光纤通道的SAN 200的性能、灵活性和可伸缩性可以比典型的基于SCSI的系统有显著提高。

图2还示出了连接到SAN互连204的网络管理器270。以从盘到数据路径充分利用所有可用SAN资源,以及提供充分的数据保护和可恢复性的方式有效地分配SAN 200中的存储设备230具有特别的重要性。管理器270可以被配置成协助在存储单元的大型网络内选择存储位置。管理器270包括虚拟盘分配器(VDA)272,根据本发明的一个实施例,其根据用户指定的性能和空间要求、指定的期望性能的等级、用户的工作负载的属性、存储的变化性能属性及其针对不同类型工作负载的响应、以及网络内竞争工作负载的存在,来处理输入/输出存储分配。

虚拟盘分配器272按满足请求所指定的性能要求的方式,或通过存储策略机制,满足对存储单元的网络内的存储的请求。虚拟盘分配器272可以在诸如IBM(R) 2145 SAN卷控制器(SAN VC)(其为虚拟存储子系统)的环境中运行。(IBM为国际商用机器公司的注册商标)。虚拟盘分配器272确定被管理的盘的性能特征。虚拟盘分配器272基于用户定义的或自动的输入确定被管理的盘与资源组之间的关系,并且

创建包含一组被管理的盘的虚拟盘，其中考虑到被管理的盘被分配到的资源组，以及所述资源组的诸如高速缓存和数据路径的存储资源。

虚拟盘分配器272将基于策略的方面扩展到开放系统环境，并且自动选择虚拟存储子系统内的存储单元以满足性能要求。推荐虚拟存储系统内所选择的存储单元允许优化使用由OS或卷管理器（volume manager）软件、或支持分条（striped）卷概念的诸如DB2®及其他数据库产品的应用程序（例如数据库应用程序）所支持的分条或合成卷。（DB2为国际商用机器公司的注册商标）。虚拟盘分配器272在考虑到长期数据使用模式的情况下，也扩展了分配存储的概念。虚拟盘分配器272引入做出数据安排的智能选择所需的各种算法。

虚拟盘分配器272可确定哪些节点，即诸如虚拟化引擎274的引擎可访问数据，以及哪些被管理盘组（MDG），即支持虚拟盘的盘的组会组成要被选择的LUN。MDG内的是至少一个被管理的盘，其被虚拟化引擎274和卷管理器276使用，以将可与企业存储系统（ESS）中的逻辑盘相比的虚拟盘中的数据分条。虚拟盘分配器272因而可以在多个存储单元上的多个资源组中选择LUN或多个LUN，以便满足用户所期望的性能等级。

管理器可执行校准过程278以发现底层盘的性能能力。与仅仅利用有关在资源组中以特定配置组织的盘的性能能力的特定知识相反，这会要求运行特定的测试以发现那些盘组的性能参数。

图3根据本发明的一个实施例图解了被引入虚拟盘分配器272的属性表300。这包括理解用户工作负载属性和所期望的性能等级310，确定被管理的盘或任意资源组的性能特征312，基于用户定义的或自动的输入确定被管理的盘与资源组之间的关系314，以及考虑到被管理的盘被分配到的资源组，创建具有一组被管理的盘的虚拟盘316。

在没有基本理解应用程序工作负载要求，或至少做出关于那些工作负载的合理假设的情况下，做出智能数据布置的决定几乎是不可能的。例如，如果用户请求100GB存储，低性能要求可允许分配单个100GB逻辑盘，然而高性能应用程序可能要求分配10个盘阵列上的10

个10GB逻辑盘，并且将那些逻辑盘上的数据分条。不幸的是，当问到多数用户其工作负载怎样时，他们通常不知道。

图4根据本发明的一个实施例图解了用户接口的机制400，用于获得工作负载简档，以供虚拟盘分配器272在将被管理的盘分配给资源组时使用。工作负载简档描述了工作负载的特征和期望的吞吐水平，并且可包括诸如所期望的存储量和分配提示的信息。

首先，可以提供封装（canned）工作负载简档410。参考图2，虚拟盘分配器272可在存储器292中提供封装工作负载简档。封装工作负载简档410可以基于各行业和应用上的用户环境的特征。例如，可以提供一组所谓的封装工作负载，例如SAP_OLTP、DB2企业智能等。听从应用专家的某些建议，用户最初选择这些封装工作负载中的一个410。

也可以基于对用户的工作负载的观察自动地创建工作负载简档412。因为每个用户的工作负载具有唯一属性，所以通过观察用户的环境中的存储访问模式，可以获得更好的工作负载假设。参考图2，虚拟盘分配器272可将其许多决定基于观察的盘访问行为，所述行为在存储器292中以数据库的形式保存。协同地连接到虚拟盘分配器272的诸如图形接口的用户接口允许用户指向卷的分组和特定时间窗口，然后基于所观察的那些卷的行为创建工作负载简档。通过这种方式，虚拟盘分配器272了解了用户的工作负载，并且增强了其随时决策。或者，用户接口可以是利用命令行接口（CLI）的脚本应用程序。

工作负载简档也可以由智能软件部件提供414。参考图2，虚拟盘分配器272也可以包括智能软件部件以提供工作负载简档。这些工作负载简档可基于应用程序中固有的特殊知识。

图5图解了用于确定被管理的盘或任意资源组（下文中的被管理盘）的性能能力的机制500。被管理盘性能能力连同工作负载简档一起由虚拟盘分配器272使用，以将被管理的盘分配给资源组。这样的机制涉及从来自管理员的人工输入到用于导出性能能力的自动探索式算法的范围。该范围内包括在受控或不受控的存储环境中，利用通过校准

工作负载进行的能力校准来确定性能的机制。

用于确定被管理的盘的性能能力的一个机制包括人工输入方法510。该方法虽然简单，但是对于理解被管理盘性能能力而言是较不期望的方法。例如，可以定义少量被管理的盘（例如，“镜像盘”、“具有8个盘的RAID-5阵列”）性能简档。这些简档中的每个都会具有特定的缺省性能属性。管理器于是将选择与被管理盘配置匹配的合适简档。该方法将缩小被管理的盘的性能能力方面的一些非常大的差异。

理解被管理的盘的性能能力的另一个方法涉及被管理盘配置器512。特定于设备的配置器可提供有关“所理解的”盘控制器的有限集的性能特征以及被管理的盘与存储控制器之间的关系的输入。例如，FASTT900配置器可提供创建RAID阵列、LUN以及子系统上的被管理的盘的统一技术，然后为那些被管理的盘提供被管理盘性能假设。配置器也可以动态地识别被管理的盘与存储控制器之间的关系。

在另一方法中，受控校准514被用来在观察被管理盘的行为的同时针对被管理盘运行指定的I/O负载。通常，受控校准会在被管理盘上分配空间之前执行，尽管在保留空间的情况下可在之后执行。

在不受控校准方法516中，随着应用程序对被管理的盘运行正常工作负载，分析被管理的盘的性能行为。如果被管理的盘已被分配（并且没有空闲空间保留给校准），则会使用该方法。该方法还可用于确定自从受控校准步骤开始条件是否已改变（例如，被管理的盘正以降低的条件运行）。作为一个例子，不受控校准可能观察到响应时间在特定负载点处一贯为高，假定该负载点更准确地反映最大吞吐能力。

另一个方法使用标准化接口518，其可允许SMIS规范提供用于标识LUN和LUN与物理资源的关联，以及那些单元的静态性能能力的概念模型。

类似于如上所述那些方法的其它方法可用于做出分配决定。类似地，可以选择以上方法的组合。例如，管理人员可选择针对SAN VC后面配置的特定类型的第一存储控制器运行受控校准步骤，然后为相同类型的后续控制器人工分配属性，例如“这些被管理的盘看起来就

像那些被管理的盘”。

基于其准确预测盘存储性能的能力，以及基于其通过数据收集工具的一般可获得性，选择虚拟盘分配器272使用的工作负载参数。所使用的工作负载参数包括每个被管理的盘的最大吞吐率、每秒最多随机读和写、每秒最多混合读和写、每秒最多顺序读和写、每秒最多混合顺序读和写、以及最低和最大负载处的时延。

对于确定部件是否具有双向能力来说混合读写是很重要的。小块随机操作被用来确定处理能力和费用，而大块顺序度量被用来确定数据传送能力和费用。

然后这些最大吞吐率被用来确定每次读或写操作，以及每次读或写所传送的兆字节的开销。然后将开销应用于计划的工作负载特征以确定被管理的盘的计划利用率。

因为被管理的盘具有单独的性能能力，并且当考虑到与那些被管理的盘相关的其它资源的性能能力时具有不同性能能力，虚拟盘分配器272定义资源组，每个组具有属于该资源组的一组被管理的盘的相关性能属性。通过考虑被管理盘的性能能力和其相关资源的性能能力，被管理的盘的整体性能可被确定。这些其它资源可包括被管理的盘驻留在的物理盘或控制器。然而，不必识别所有与被管理的盘相关的可能资源。相反，当定义资源组时，虚拟盘分配器272将概括与资源的关联。

每个资源组被指定性能属性（与被管理盘的相同）。资源组的性能能力可被人工地指定或通过称作受控校准的受控基准测试程序过程来确定。校准可以针对所识别的和每次由管理人员识别的被管理的盘来执行。校准可以在被管理的盘被分配给虚拟盘之前针对被管理的盘执行。这也可被包括在在指定存储控制器上自动创建被管理的盘的操作过程内。只要各组中所有被管理的盘未被分配，校准将针对各个被管理的盘以及所有相关的资源组执行。用于校准的技术类似于那些针对竞争盘子系统进行工程测量时所使用的技术。

因为被管理的盘可属于不止一个的资源组，所以资源组的校准允

许优化数据分配决定。然而，不必发现被管理的盘与相关资源之间的所有潜在关系，并且当猜测的关系不存在时，不影响结果。虚拟盘分配器272的各种实现可以在关系发现中置入不同水平的复杂度（sophistication）。

图6根据本发明的一个实施例图解了由虚拟盘分配器272所使用的数据结构600，用于抽象虚拟存储子系统中的重要性能要素。数据结构可被表示成代表诸如集群、设备适配器、单个盘或盘阵列以及任何相关资源的存储单元的节点树。然而，本领域技术人员将认识到，本发明不意味被限于图6中所示的结构。而是可使用比树结构更通用的节点网络。

数据结构600被用来提供存储分配的面向目标的方法。作为一个例子，SAN VC提出了与常规存储控制器有某种程度不同的存储模型。以可与IBM企业存储服务器（ESS）中的逻辑盘相比的虚拟盘为目的，提出了应用服务器。通过识别与一组被管理的盘和资源组相关的、包括针对特定工作负载的吞吐能力的一组性能特性，并且估计从所识别的与一组被管理的盘和资源组相关的性能特性组导出的资源利用率，虚拟盘分配器272创建虚拟盘。例如通过SAN VC主节点访问虚拟盘，并且在该节点中缓存虚拟盘的数据，这是可与ESS集群相比的功能。参考图6，虚拟盘605的存储设备被分配于一个或多个被管理的盘610、620、630以及640上。图解的被管理的盘被分组成被管理盘组（MDG）650。虚拟盘605与单个MDG 650相关。通常，为了最好的性能，虚拟盘605会在MDG 650中的所有被管理的盘上被分条。

被管理的盘610、620、630和640中的每个可被作为与其它应用共享的单个物理盘的一部分而被分配，或作为由若干盘构成的整个RAID阵列而被分配。此外，在选择被管理盘的配置方面为存储管理器给予巨大的灵活性。由于与其它应用的关联，所以在一个被管理的盘上的活动可影响另一个被管理的盘的性能，例如当它们驻留在相同的存储设备上或在相同的存储控制器后面时。参考图6，被管理的盘610与三个资源611、612和613相关，而被管理的盘620与资源613、621和622

相关。因为资源613与两个被管理的盘610和620都相关，所以被管理的盘中的每个的性能特征可能是未知和高度可变的。因此，智能数据布置决定的关键是获得和保持关于被管理的盘的能力和行为的合理性能假设。

根据本发明的实施例，每个被管理的盘与一个或多个资源组（RG）相关。RG1 660、RG2 670和RG3 680由虚拟盘分配器272定义。RG1 660包括被管理的盘610和资源611、612和613。RG2 670包括被管理的盘620、630和被管理的盘640的一部分，并且具有资源613、621、622、631、632和641。RG3 680包括被管理的盘640的一部分，以及资源632和641。如所见的，资源613在被管理的盘610与620之间共享，而资源632和641在被管理的盘630与640之间共享。因为被管理的盘和相关资源被分组到允许观察资源组内和之间的相互作用的资源组中，所以有关被管理的盘的能力和行为的增强的性能假设可以由虚拟盘分配器272做出。

例如，在卷选择过程期间，虚拟盘分配器272跟踪被管理的盘和资源组的利用率，这类似于其维护ESS的VPA中配置树的部件的利用率的方式。当选择潜在的被管理盘用于分配时，针对被管理的盘所属于的所有资源组评估效果。然后，被管理的盘的利用率被假定为在所有相关的RG中是最高的。这类似于在当前版本的VPA使用的配置树中向上导航的技术。虚拟盘分配器272对于满足指定的能力和性能要求的存储分配做出推荐。

再次参考图2，虚拟盘分配器272用有关存储单元如何实际执行的知识做出改进，但不依赖于极其准确的信息，这就是为什么虚拟盘分配器272可以适用于来自不同厂商的异构类型存储的原因。但是准确的实时或历史性能数据可用来将一个厂商的产品区别于其它厂商的产品，以及使存储分配偏离可能在感兴趣的时间段期间竞争的工作负载。

虚拟盘分配器272的一个重要方面涉及利用能力和性能结构来平衡可用资源上的存储分配。如果虚拟盘分配器272中可能有多种选择，则能力和性能结构可被用来通过使用伪随机数使分配偏置到一组资

源。用这种方式可以选择若干样本分配，并且为答案选择样本中间最好的样本。该技术防止算法在相似情况下产生相同的推荐，从而增加在可用资源上工作负载的散布。通过这种方式，存储分配将偏向于最能够处理指定的工作负载的网络中的单元。

图7根据本发明的一个实施例图解了用于管理虚拟存储系统中的存储分配的方法的流程图700。方法包括获得用户定义的工作负载简档710，以及确定被管理的盘的性能特征720。基于用户定义的或自动的输入确定被管理的盘与资源组之间的关系730。考虑到被管理的盘被分配到的资源组，创建包括一组被管理的盘的虚拟盘740。

参考图3-7所说明的过程可被有形地体现在计算机可读介质或载体中，例如图2中图解的一个或多个固定和/或可移动数据存储设备288，或其它数据存储或数据通信设备。数据存储设备288或计算机程序290可被加载到存储器292中，以对管理器270或虚拟盘分配器272进行配置以便执行。此类计算机程序包括指令，当由诸如图2的处理器294的处理器读取和执行时，所述指令使得管理器270或虚拟盘分配器272执行完成本发明的步骤或单元所需的步骤。

为了说明和描述的目的，已提出本发明的示范实施例的上述描述。其不意欲详尽列出或将本发明限制于所揭示的具体形式。按照上述教导，许多修改和变化是可能的。本发明的范围未被限制于该详细说明，而是由此处附加的权利要求书来限制。

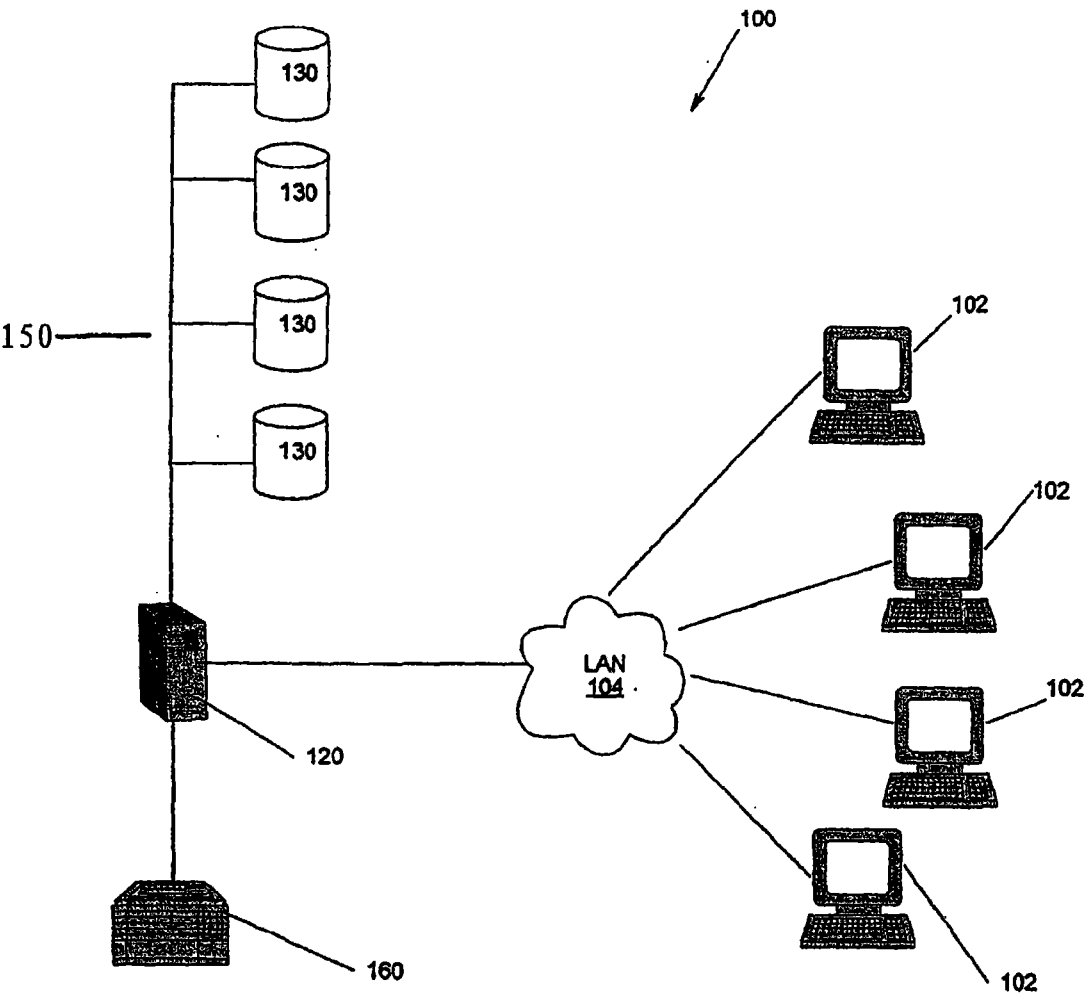


图1

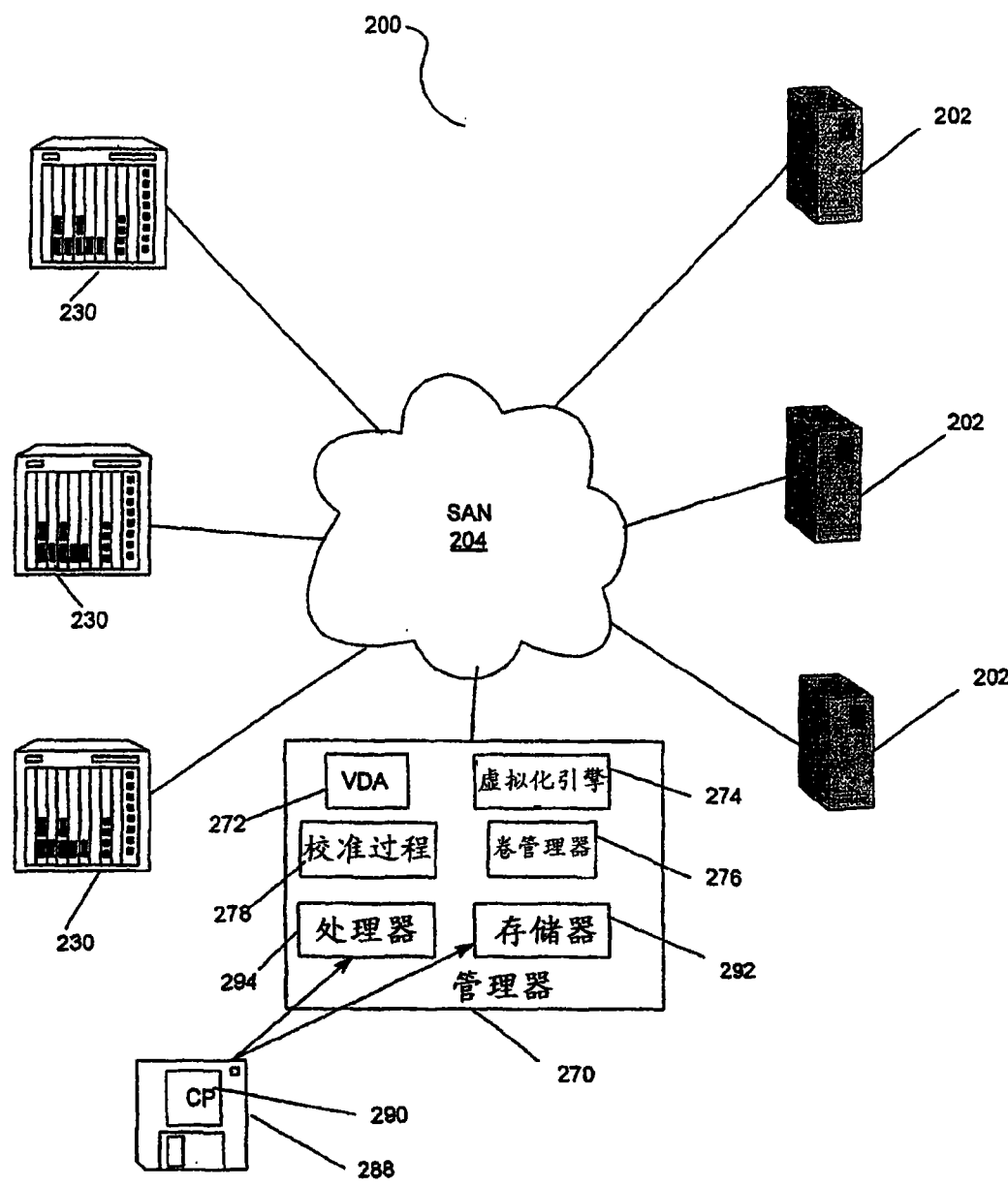


图 2

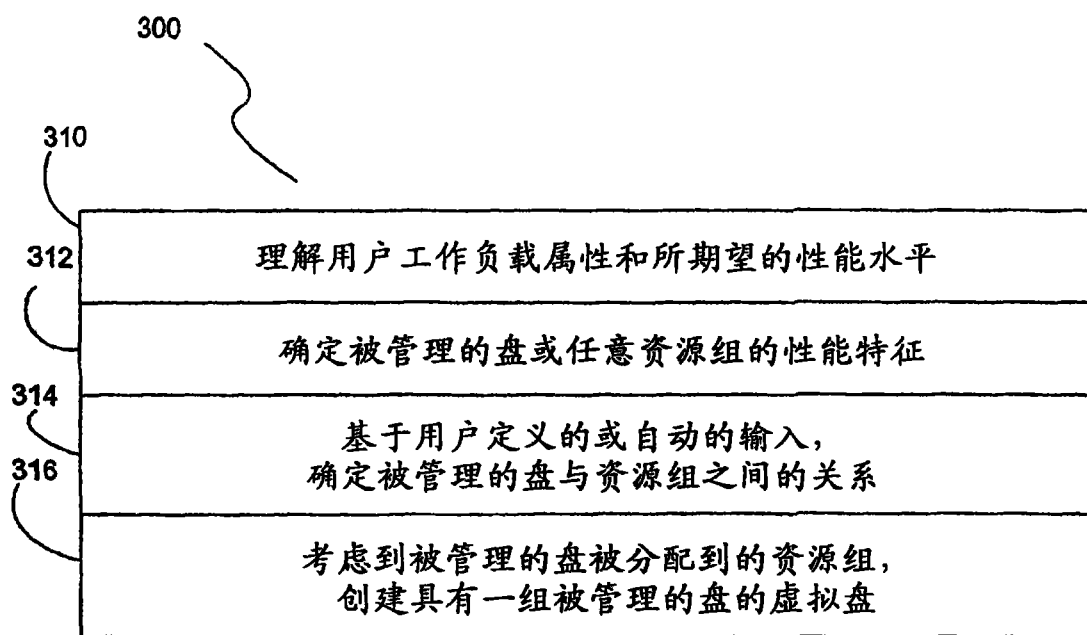


图 3

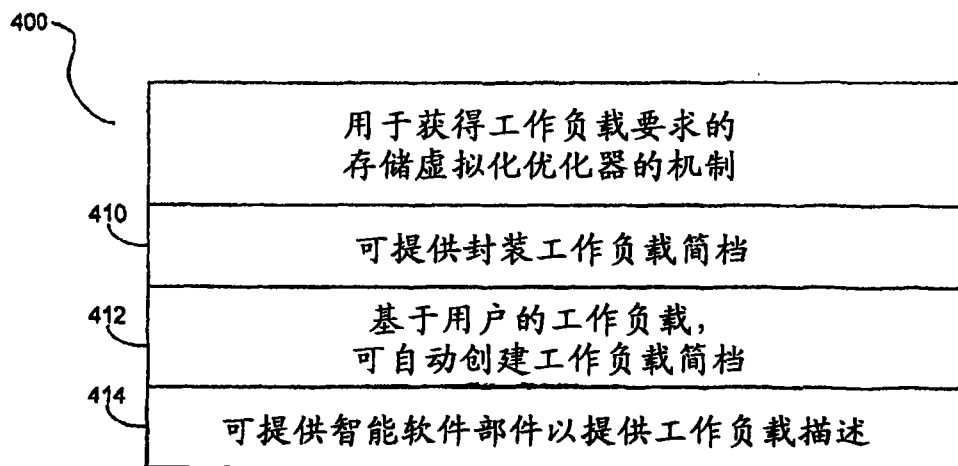


图 4

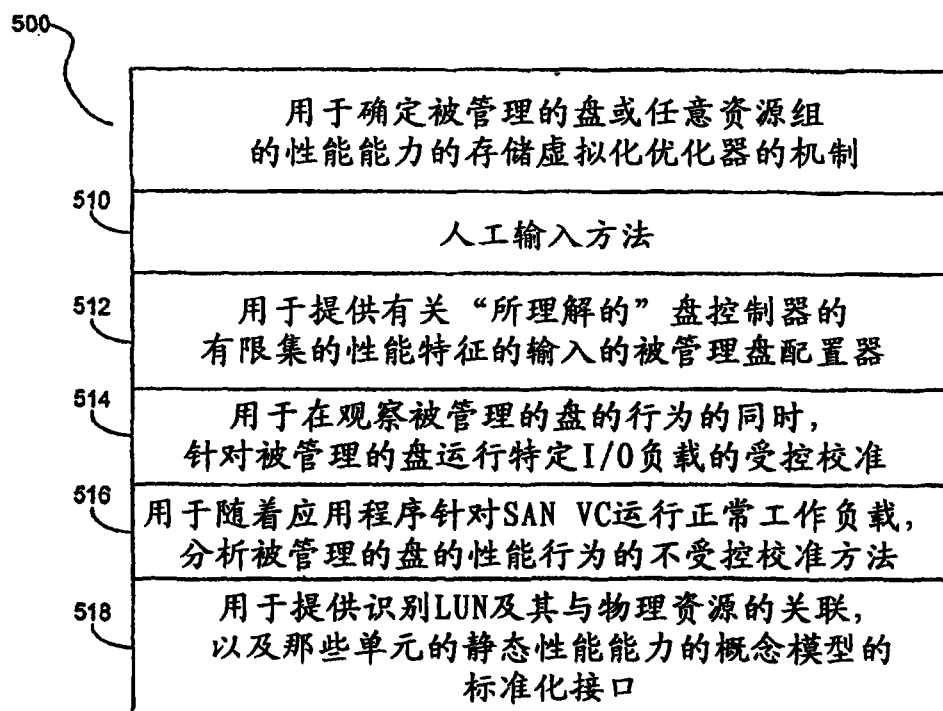
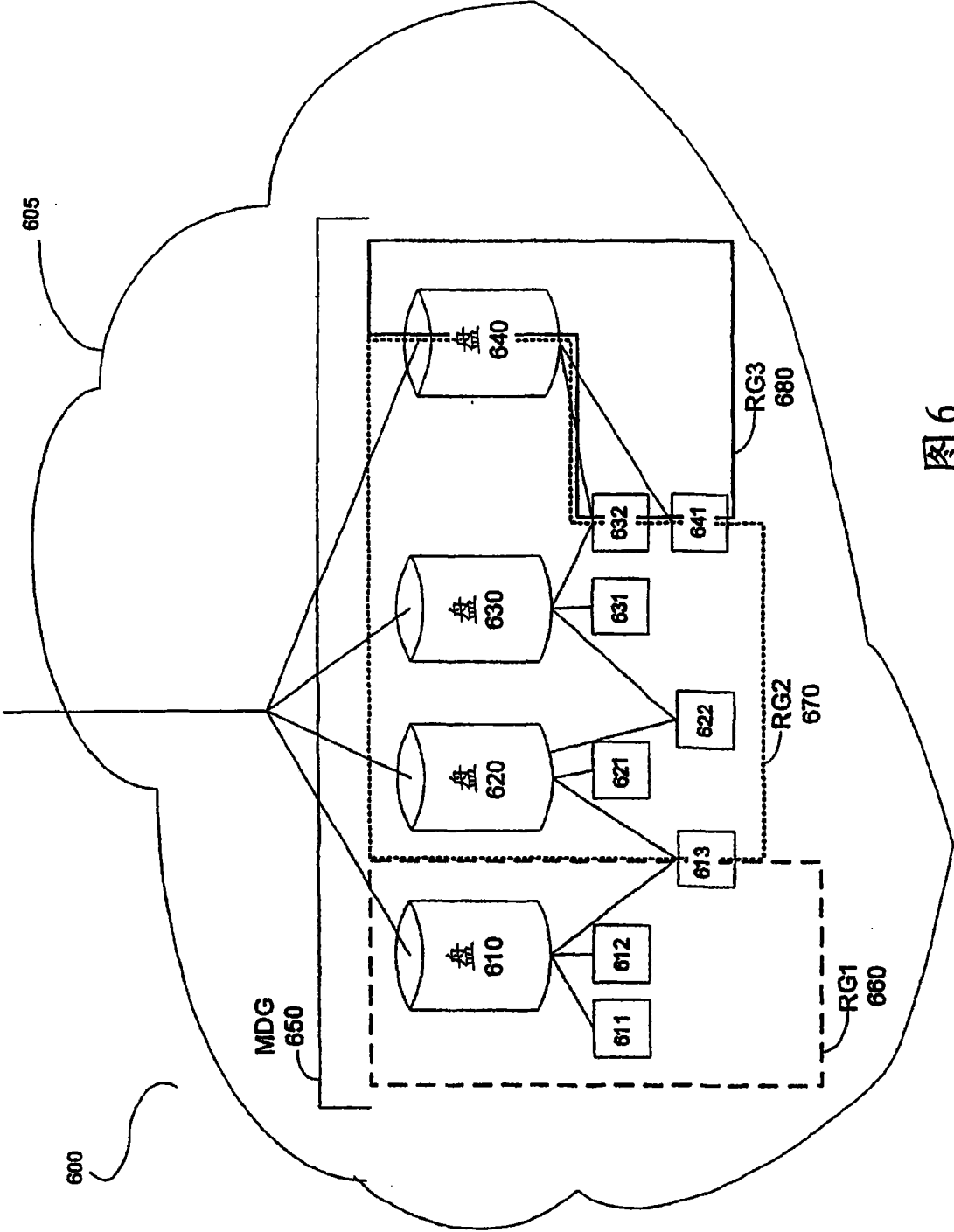


图 5



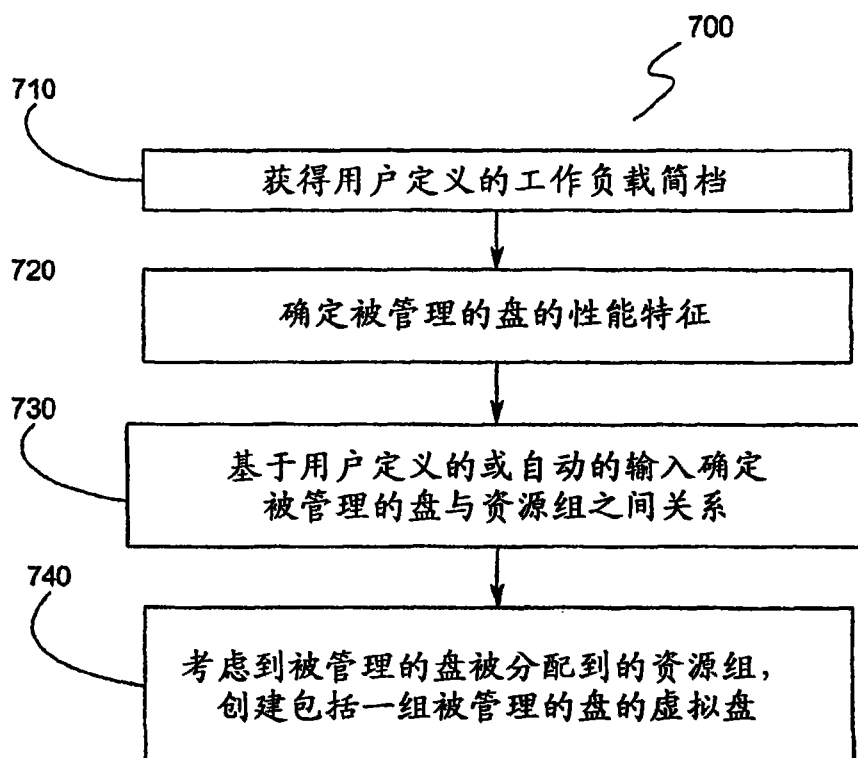


图 7