(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 102257481 B
(45)授权公告日 2014.02.19

(21)申请号 200980151636.3
(22)申请日 2009.11.30
(30)优先权数据
12/342,796 2008.12.23 US
(85)PCT国际申请进入国家阶段日
2011.06.21
(86)PCT国际申请的申请数据
PCT/US2009/066065 2009.11.30
(87)PCT国际申请的公布数据
WO2010/074885 EN 2010.07.01
(73)专利权人 思杰系统有限公司
地址 美国佛罗里达州
(72)发明人 A·K·沃菲尔德
(74)专利代理机构 北京泛华伟业知识产权代理有限公司 11280
代理人 王勇
(51)Int.Cl.
G06F 9/50(2006.01)
G06F 3/06(2006.01)
(56)对比文件
CN 101115246 A,2008.01.30, 说明书第1－5页,图1－4.
审查员 刘娜

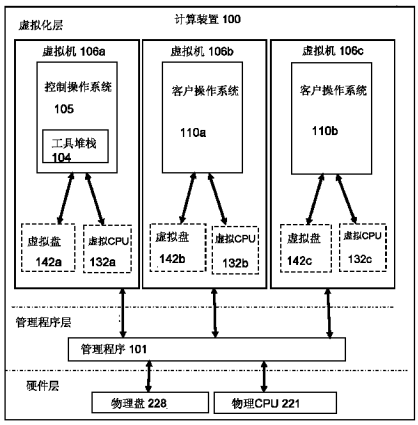
权利要求书2页 说明书12页 附图4页

(54)发明名称

用于由管理程序控制对物理资源的访问的系统和方法

(57)摘要

用于在虚拟机执行期间由管理程序控制对物理资源的访问的系统包括物理盘和管理程序。物理盘是由计算装置提供的并且存储虚拟盘的至少一部分。管理程序在计算装置上执行。管理程序给虚拟盘分配对物理盘的访问量。管理程序确定物理盘的利用程度已超过阈值。管理程序响应于所述确定来限制虚拟盘对物理盘的访问。



CN 102257481 B

1. 一种用于在虚拟机执行期间由管理程序控制对物理资源的访问的方法，所述方法包括：

由管理程序给虚拟机中的虚拟盘分配对至少一个物理盘的访问量；

由管理程序确定所述至少一个物理盘的利用程度超过阀值；

由管理程序至少部分地基于所述确定来从多个预定的行为中选择一个行为，用于限制对所述至少一个物理盘的访问，其中，在文件中定义所述预定的行为，并且其中，所述多个预定的行为中的至少一个预定的行为允许虚拟盘对所述至少一个物理盘的继续访问；以及

由管理程序响应于所述确定来限制虚拟盘对所述至少一个物理盘的访问。

2. 根据权利要求 1 所述的方法，其中确定还包括由管理程序确定已经达到所述至少一个物理盘的容量限制。

3. 根据权利要求 1 所述的方法，其中确定还包括由管理程序确定虚拟盘访问盘的速度已经超过阀值。

4. 根据权利要求 1 所述的方法，其中确定还包括由管理程序确定物理盘已经发生错误。

5. 根据权利要求 4 所述的方法，还包括从虚拟机屏蔽所述错误。

6. 根据权利要求 1 所述的方法，其中限制还包括对虚拟盘所请求的输入 / 输出（I/O）操作进行速度限制。

7. 根据权利要求 1 所述的方法，还包括挂起虚拟机执行的至少一个进程。

8. 根据权利要求 1 所述的方法，还包括暂停虚拟机的执行。

9. 根据权利要求 1 所述的方法，还包括降低虚拟机的执行速度。

10. 一种用于在虚拟机执行期间由管理程序控制对物理资源的访问的系统，所述系统包括：

由计算装置提供的物理盘，其存储虚拟机中虚拟盘的至少一部分；以及

在计算装置上执行的管理程序，所述管理程序用于 i）给虚拟盘分配对物理盘的访问量，ii）确定物理盘的利用程度超过阀值，iii）至少部分地基于所述确定来从多个预定的行为中选择一个行为，用于限制对所述至少一个物理盘的访问，其中，在文件中定义所述预定的行为，并且其中，所述多个预定的行为中的至少一个预定的行为允许虚拟盘对所述至少一个物理盘的继续访问；以及 iv）响应于所述确定来限制虚拟盘对物理盘的访问。

11. 根据权利要求 10 所述的系统，其中所述管理程序还包括用于对虚拟盘所请求的输入 / 输出（I/O）操作进行速度限制的装置。

12. 根据权利要求 10 所述的系统，其中所述管理程序还包括用于挂起虚拟机执行的至少一个进程的装置。

13. 根据权利要求 10 所述的系统，其中所述管理程序还包括用于暂停虚拟机执行的装置。

14. 根据权利要求 10 所述的系统，其中所述管理程序还包括用于降低虚拟机的执行速度的装置。

15. 根据权利要求 10 所述的系统，其中所述管理程序还包括用于确定已经达到所述至少一个物理盘的容量限制的装置。

16. 根据权利要求 10 所述的系统，其中所述管理程序还包括用于确定虚拟盘访问盘的

速度已经超过阀值的装置。

17. 根据权利要求 10 所述的系统，其中所述管理程序还包括用于确定物理盘上已经发生错误的装置。

18. 根据权利要求 17 所述的系统，其中所述管理程序还包括用于从虚拟机屏蔽所述错误的装置。

用于由管理程序控制对物理资源的访问的系统和方法

[0001] 相关申请

[0002] 本申请要求2008年12月23日提交的美国专利申请序列号12/342,796的优先权，将其公开内容视为本申请公开的一部分并且通过引用将其整体包含于此。

技术领域

[0003] 本发明总的涉及用于控制对资源访问的系统和方法。本发明尤其涉及用于由管理程序控制虚拟资源对物理资源的访问的系统和方法。

背景技术

[0004] 在实现管理程序以在宿主计算装置上执行虚拟机的传统计算环境中，管理程序通常向虚拟机提供对宿主计算装置所提供的硬件资源的访问。这样的环境中，传统的管理程序通常不提供用于防止由过度使用所分配的资源引起的错误影响虚拟机用户的功能。例如，如果资源使用了超过其所分配的处理器时间或所分配的磁盘访问，该资源可能引起与其他资源的冲突或导致该虚拟机或共享对物理盘访问的其他虚拟机的不适宜的终止。传统的管理程序可在事后接收已发生错误或冲突的指示，或者在虚拟机执行之前确定可能会发生错误或冲突。然而，传统的管理程序通常不提供灵活的选项用于在这些错误发生后继续执行虚拟机。

发明内容

[0005] 一方面，用于在虚拟机执行期间由管理程序控制对物理资源的访问的方法包括由管理程序给虚拟机中虚拟盘分配对至少一个物理盘的访问量。该方法包括由管理程序确定物理盘的利用程度超过阀值。该方法包括由管理程序响应于所述确定，限制虚拟盘对物理盘的访问。

[0006] 一个实施例中，该方法包括对虚拟盘所请求的输入/输出(I/O)操作进行速度限制。又一个实施例中，该方法包括挂起虚拟机执行的至少一个进程。又一个实施例中，该方法包括由管理程序确定已经达到至少一个物理盘的容量限制。又一个实施例中，该方法包括从虚拟机屏蔽物理盘的错误。

[0007] 又一方面，用于在虚拟机执行期间由管理程序控制对物理资源的访问的系统包括物理盘和管理程序。所述物理盘是由计算装置提供的并且存储虚拟机中虚拟盘的至少一部分。管理程序在计算装置上执行。管理程序给虚拟盘分配对物理盘的访问量。管理程序确定在虚拟机内执行的客户操作系统对物理盘的利用程度超过所分配量的阀值。响应于所述确定，管理程序限制虚拟盘对物理盘的访问。

附图说明

[0008] 本发明的前述和其它目的、方面、特征和优点通过参考下述结合附图的描述将会更加明显并更易于理解，其中：

[0009] 图 1A 是描述包括管理程序层、虚拟化层和硬件层的计算环境的实施例的框图；

[0010] 图 1B 和 1C 是描述有益于与此处所描述的方法和系统的相关的计算装置的实施例的框图；

[0011] 图 2 是描述用于由管理程序控制对物理资源访问的系统的实施例的框图；以及

[0012] 图 3 是描述用于由管理程序控制对物理资源访问的方法的实施例的流程图。

具体实施方式

[0013] 现参考图 1A，框图描述了虚拟化环境的一个实施例。简而言之，计算装置 100 包括管理程序层、虚拟化层和硬件层。管理程序层包括管理程序 101（也称为虚拟化管理器），其通过在虚拟化层中执行的至少一个虚拟机来分配和管理对硬件层中的多个物理资源的访问（例如处理器 221 和盘 228）。虚拟化层包括至少一个操作系统 110 和分配给至少一个操作系统 110 的多个虚拟资源。虚拟资源可包括而不限于多个虚拟处理器 132a、132b、132c（总称为 132）和虚拟盘 142a、142b、142c（总称为 142）、以及例如虚拟存储器和虚拟网络接口的虚拟资源。可将多个虚拟资源和操作系统 110 称为虚拟机 106。虚拟机 106 可包括控制操作系统 105，该控制操作系统 105 与管理程序 101 通信，并用于执行应用以管理并配置计算装置 100 上的其他虚拟机。

[0014] 现参考图 1A，具体而言，管理程序 101 可以以模拟可访问物理设备的操作系统的任何方式向操作系统提供虚拟资源。管理程序 101 可以向任一数量的客户操作系统 110a、110b（总称为 110）提供虚拟资源。一些实施例中，计算装置 100 执行一种或多种管理程序。这些实施例中，管理程序可用于模拟虚拟硬件、划分物理硬件、虚拟化物理硬件并执行提供对计算环境的访问的虚拟机。管理程序可包括由位于美国加州的 Palo Alto 的 VMWare 制造的这些程序、XEN 管理程序（一种开源产品，其开发由开源 Xen.org 协会监管）、由微软公司提供的 HyperV、VirtualServer 或虚拟 PC 管理程序或其他。一些实施例中，计算装置 100 执行创建客户操作系统可在其上执行的虚拟机平台的管理程序，该计算装置 100 被称为宿主服务器。在一个这样的实施例中，例如，计算装置 100 是由位于美国佛罗里达州 Fort Lauderdale 的 CitrixSystems 公司提供的 XEN SERVER。

[0015] 一些实施例中，管理程序 101 在计算装置上执行的操作系统之内执行。一个这样的实施例中，执行操作系统和管理程序 101 的计算装置可被视为具有宿主操作系统（在计算装置上执行的操作系统）和客户操作系统（在由管理程序 101 提供的计算资源分区内执行的操作系统）。其他实施例中，管理程序 101 和计算装置上的硬件直接交互而不是在宿主操作系统上执行。一个这样的实施例中，管理程序 101 可被视为是在“裸金属 (bare metal)”上执行，所述“裸金属”指包括计算装置的硬件。

[0016] 一些实施例中，管理程序 101 可以创建操作系统 110 在其中执行的虚拟机 106a-c（总称为 106）。在一个这样的实施例中，管理程序 101 加载虚拟机映像以创建虚拟机 106。另一个这样的实施例中，管理程序 101 在虚拟机 106 内执行操作系统 110。在又一个这样的实施例中，虚拟机 106 执行操作系统 110。

[0017] 一些实施例中，管理程序 101 控制对在计算装置 100 上执行的虚拟机 106 的处理器调度和内存划分。一个这样的实施例中，管理程序 101 控制至少一个虚拟机 106 的执行。另一个这样的实施例中，管理程序 101 向至少一个虚拟机 106 呈现对由计算装置 100 提供

的至少一个硬件资源的抽象。其他实施例中，管理程序 101 控制是否以及如何把物理处理器能力呈现给虚拟机 106。

[0018] 控制操作系统 105 可以执行用于管理和配置客户操作系统的至少一个应用。一个实施例中，控制操作系统 105 可以执行管理应用，如包括如下用户接口的应用，该用户接口为管理员提供对用于管理虚拟机执行的功能的访问，这些功能包括用于执行虚拟机、中止虚拟机执行或者识别要分配给虚拟机的物理资源类型的功能。另一个实施例中，管理程序 101 在由管理程序 101 创建的虚拟机 106 内执行控制操作系统 105。在又一个实施例中，控制操作系统 105 在被授权直接访问计算装置 100 上的物理资源的虚拟机 106 上执行。一些实施例中，计算装置 100a 上的控制操作系统 105a 可经由管理程序 101a 和管理程序 101b 之间的通信与计算装置 100b 上的控制操作系统 105b 交换数据。这样，一个或多个计算装置 100 可以与一个或多个其他计算装置 100 交换关于资源池中可用的处理器和其他物理资源的数据。在这些实施例的一个中，该功能允许管理程序管理分布在多个物理计算装置上的资源池。在这些实施例的又一个中，多个管理程序管理在其中一个计算装置 100 上执行的一个或多个客户操作系统。

[0019] 一个实施例中，控制操作系统 105 在被授权与至少一个客户操作系统 110 交互的虚拟机 106 上执行。另一个实施例中，客户操作系统 110 通过管理程序 101 与控制操作系统 105 通信，以请求访问盘或网络。在又一个实施例中，客户操作系统 110 和控制操作系统 105 可通过由管理程序 101 建立的通信信道通信，例如，通过由管理程序 101 提供的多个共享存储器页面通信。

[0020] 一些实施例中，控制操作系统 105 包括用于直接与由计算装置 100 提供的网络硬件通信的网络后端驱动器。一个这样的实施例中，网络后端驱动器处理来自至少一个客户操作系统 110 的至少一个虚拟机请求。其他实施例中，控制操作系统 105 包括用于与计算装置 100 上的存储单元通信的块后端驱动器。一个这样的实施例中，块后端驱动器根据从客户操作系统 110 接收的至少一个请求从存储单元读写数据。

[0021] 一个实施例，控制操作系统 105 包括工具堆栈 104。其他实施例中，工具堆栈 104 提供如下功能：和管理程序 101 交互、和其他控制操作系统 105（例如位于第二计算装置 100b 上）通信，或者管理计算装置 100 上的虚拟机 106b、106c。另一个实施例中，工具堆栈 104 包括自定义应用，其用于向虚拟机群的管理员提供改进的管理功能。一些实施例中，工具堆栈 104 和控制操作系统 105 中的至少一个包括管理 API，其提供用于远程配置并控制计算装置 100 上运行的虚拟机 106 的接口。其他实施例中，控制操作系统 105 通过工具堆栈 104 和管理程序 101 通信。

[0022] 一个实施例中，管理程序 101 在由管理程序 101 创建的虚拟机 106 内执行客户操作系统 110。另一个实施例中，客户操作系统 110 为计算装置 100 的用户提供对计算环境中的资源的访问。另一个实施例中，资源包括程序、应用、文档、文件、多个应用、多个文件、可执行程序文件、桌面环境、计算环境或对计算装置 100 的用户可用的其他资源。另一个实施例中，可通过多个访问方法将资源传送给计算装置 100，这些方法包括但不限于：常规的直接安装在计算装置 100 上、通过应用流的方法传送给计算装置 100、将由在第二计算装置 100′ 上执行资源产生的并通过表示层协议传送给计算装置 100 的输出数据传送给计算装置 100、将由在第二计算装置 100′ 上执行的虚拟机执行资源所产生的输出数据传送给计

算装置 100、或者从连接到计算装置 100 的移动存储装置（例如 USB 设备）执行或者通过在计算装置 100 上执行的虚拟机执行并且产生输出数据。一些实施例中，计算装置 100 把执行资源所产生的输出数据传输给另一个计算装置 100′ 。

[0023] 一个实施例中，客户操作系统 110 和该客户操作系统 110 在其上执行的虚拟机结合形成完全虚拟化虚拟机，该完全虚拟化虚拟机并不知道自己是虚拟机，这样的机器可称为“Domain U HVM（硬件虚拟机）虚拟机”。另一个实施例中，完全虚拟化机包括模拟基本输入 / 输出系统 (BIOS) 的软件以便在完全虚拟化机中执行操作系统。在另一个实施例中，完全虚拟化机可包括驱动器，其通过和管理程序 101 通信提供功能。这样的实施例中，驱动器通常意识到自己在虚拟化环境中执行。

[0024] 另一个实施例中，客户操作系统 110 和该客户操作系统 110 在其上执行的虚拟机结合形成超虚拟化虚拟机 (paravirtualized)，该超虚拟化虚拟机意识到自己是虚拟机，这样的机器可称为“Domain U PV 虚拟机”。另一个实施例中，超虚拟化机包括完全虚拟化机不包括的额外驱动器。另一个实施例中，超虚拟化机包括如上所述的被包含在控制操作系统 105 中的网络后端驱动器和块后端驱动器。

[0025] 计算装置 100 可以被部署为任意类型和形式的计算装置和 / 或在其上执行，例如能够与任意类型和形式的网络通信并且执行此处所描述的操作的计算机、网络装置或设备。图 1B 和 1C 描述了用于实施此处所描述的方法和系统的实施例的计算装置 100 的框图。如图 1B 和 1C 所示，计算装置 100 包括中央处理单元 121 和主存储器单元 122。如图 1B 所示，计算装置 100 可包括存储装置 128、安装装置 116、网络接口 118、I/O 控制器 123、显示装置 124a-124n、键盘 126 和诸如鼠标的指示装置 127。存储装置 128 可包括但不限于操作系统、软件和客户机代理 120。如图 1C 所示，每个计算装置 100 也可包括另外的可选部件，例如存储器端口 103、桥 170、一个或多个输入 / 输出装置 130a-130n（总的使用标号 130 表示）以及与中央处理单元 121 通信的高速缓存存储器 140。

[0026] 中央处理单元 121 是响应并处理从主存储器单元 122 取出的指令的任意逻辑电路。在许多实施例中，中央处理单元 121 由微处理器单元提供，例如 ：由位于美国加州 Mountain View 的 Intel 公司制造的微处理器单元 ；由位于美国伊利诺伊州 Schaumburg 的 Motorola 公司制造的微处理器单元 ；由位于美国加州 Santa Clara 的 Transmeta 公司制造的微处理器单元 ；由位于美国纽约州 White Plains 的 International Business Machines 公司制造的 RS/6000 处理器 ；或者由位于美国加州 Sunnyvale 的 Advanced Micro Devices 公司制造的微处理器单元。计算装置 100 可以基于这些处理器中的任意一种，或者能够按照这里所说明的那样运行的任意其它处理器。

[0027] 主存储器单元 122 可以是能够存储数据并允许微处理器 121 直接访问任意存储位置的一个或多个存储器芯片，例如静态随机存取存储器 (SRAM)、突发 SRAM 或同步突发 SRAM(BSRAM)、动态随机存取存储器 (DRAM)、快速页模式 DRAM(FPM DRAM)、增强 DRAM(EDRAM)、扩展数据输出 DRAM(EDO DRAM)、突发扩展数据输出 DRAM(BEDO DRAM)、同步 DRAM(SDRAM)、JEDEC SRAM、PC100 SDRAM、双倍数据速率 SDRAM(DDR SDRAM)、增强 SDRAM(ESDRAM)、同步链路 DRAM(SLDRAM)、直接 RambusDRAM(DRDRAM) 或铁电 RAM(FRAM)。主存储器 122 可以基于上述存储器芯片的任意一种，或者能够像这里所说明的那样运行的任意其它可用存储器芯片。在图 1B 所示的实施例中处理器 121 通过系统总线 150（在下面进

行更详细的描述）与主存储器 122 进行通信。图 1C 描述了在其中处理器通过存储器端口 103 直接与主存储器 122 通信的计算装置 100 的实施例。例如，在图 1C 中主存储器 122 可以是 DRDRAM。

[0028] 图 1C 描述在其中主处理器 121 通过有时被称为背侧总线的次级总线直接与高速缓存存储器 140 通信的实施例。其他实施例中，主处理器 121 使用系统总线 150 与高速缓存存储器 140 通信。高速缓存存储器 140 通常有比主存储器 122 更快的响应时间，并且通常由 SRAM、BSRAM 或 EDRAM 提供。在图 1C 所示的实施例中，处理器 121 通过本地系统总线 150 与各种 I/O 装置 130 进行通信。可以使用各种不同的总线将中央处理单元 121 连接到任意 I/O 装置 130，包括 VESA VL 总线、ISA 总线、EISA 总线、微通道体系结构 (MCA) 总线、PCI 总线、PCI-X 总线、PCI-Express 总线或 NuBus。对于 I/O 装置是视频显示器 124 的实施例，处理器 121 可以使用高级图形端口 (AGP) 与显示装置 124 通信。图 1C 描述了在其中主处理器 121 通过 HYPERTRANSPORT、RAPIDIO、或 INFINIBAND 通信技术直接与 I/O 装置 130b 通信的计算机 100 的实施例。图 1C 也描述了在其中本地总线和直接通信相混合的实施例：处理器 121 使用本地互连总线与 I/O 装置 130a 通信同时直接与 I/O 装置 130b 通信。

[0029] 各式各样的 I/O 装置 130a-130n 可以存在于计算装置 100 中。输入装置包括键盘、鼠标、触控板、轨迹球、话筒、拨号盘和绘图板。输出装置包括视频显示器、扬声器、喷墨打印机、激光打印机和热升华打印机。如图 1B 所示，I/O 装置可以由 I/O 控制器 123 控制。I/O 控制器可以控制一个或多个 I/O 装置，例如键盘 126 和指示装置 127（如鼠标或光笔）。此外，I/O 装置还可以为计算装置 100 提供存储和 / 或安装介质 116。在其它实施例中，计算装置 100 可以提供 USB 连接（未示出）以接收手持 USB 存储装置，例如由位于美国加州 Los Alamitos 的 Twintech Industry 公司生产的设备的 USB 闪存驱动线。

[0030] 再次参考图 1B，计算装置 100 可以支持任意适当的安装装置 116，例如用于接收像 3.5 英寸、5.25 英寸磁盘或 ZIP 磁盘这样的软盘的软盘驱动器、CD-ROM 驱动器、CD-R/RW 驱动器、DVD-ROM 驱动器、闪存驱动器、多种格式的磁带驱动器、USB 设备、硬盘驱动器或适于安装软件和程序的任意其它设备。计算装置 100 还可以包括存储装置，诸如一个或者多个硬盘驱动器或者独立磁盘冗余阵列，用于存储操作系统和其它相关软件，以及用于存储诸如涉及客户机代理 120 的任意程序的应用软件程序。或者，可以使用安装装置 116 的任意一种作为存储装置。此外，操作系统和软件可从可引导介质运行，可引导介质例如是可引导 CD，诸如 KNOPPIX，其为一种用于 GNU/Linux 的可引导 CD，可自 knoppix.net 作为 GNU/Linux 分发获得。

[0031] 此外，计算装置 100 可以包括网络接口 118 以通过多种连接来与网络 104 对接，所述连接包括但不限于标准电话线路、LAN 或 WAN 链路（例如 802.11、T1、T3、56kb、X.25、SNA、DECNET）、宽带连接（如 ISDN、帧中继、ATM、千兆以太网、SONET 上以太网 (Ethernet-over-SONET)）、无线连接、或上述任意或全部的某种组合。使用多种通信协议（例如 TCP/IP、IPX、SPX、NetBIOS、Ethernet、ARCNET、SONET、SDH、光纤分布数据接口 (FDDI)、RS232、RS485、IEEE802.11、IEEE802.11a、IEEE802.11b、IEEE802.11g、CDMA、GSM、WiMax 和直接异步连接）可建立连接。在一个实施例中，计算装置 100 可以通过任意类型和 / 或形式的网关或诸如安全套接字层 (SSL) 或传输层安全 (TLS) 协议，或由位于美国佛罗里达州 Ft.Lauderdale 的 Citrix Systems 公司制造的 Citrix 网关协议的隧道协议来与

其他计算装置 100’通信。网络接口 118 可以包括内置网络适配器、网络接口卡、PCMCIA 网络卡、卡总线网络适配器、无线网络适配器、USB 网络适配器、调制解调器或适用于将计算装置 100 对接到能够通信并执行这里所说明的操作的任意类型的网络的任意其它设备。

[0032] 在一些实施例中，计算装置 100 可以包括多个显示装置 124a-124n 或与其相连，这些显示装置各自可以是相同或不同的类型和 / 或形式。因而，任意一种 I/O 装置 130a-130n 和 / 或 I/O 控制器 123 可以包括任一类型和 / 或形式的适当的硬件、软件或硬件和软件的组合，以支持、允许或提供通过计算装置 100 连接和使用多个显示装置 124a-124n。例如，计算装置 100 可以包括任意类型和 / 或形式的视频适配器、视频卡、驱动器和 / 或库，以与显示装置 124a-124n 对接、通信、连接或以其他方式使用显示装置 124a-124n。在一个实施例中，视频适配器可以包括多个连接器以与多个显示装置 124a-124n 对接。在其它实施例中，计算装置 100 可以包括多个视频适配器，每个视频适配器与显示装置 124a-124n 中的一个或多个连接。在一些实施例中，计算装置 100 的操作系统的任一部分都可以被配置用于使用多个显示器 124a-124n。在其它实施例中，显示装置 124a-124n 中的一个或多个可以由一个或多个其它计算装置提供，诸如（例如通过网络）与计算装置 100 连接的计算装置 100a 和 100b。这些实施例可以包括被设计和构造的任一类型的软件，以使用另一个计算机的显示装置作为计算装置 100 的第二显示装置 124a。本领域的普通技术人员会认识和意识到可以将计算装置 100 配置成拥有多个显示装置 124a-124n 的各种方式和实施例。

[0033] 在进一步的实施例中，I/O 装置 130 可以是系统总线 150 和外部通信总线之间的桥，所述外部通信总线如 USB 总线、Apple 桌面总线、RS-232 串行连接、SCSI 总线、FireWire 总线、FireWire800 总线、以太网总线、AppleTalk 总线、千兆位以太网总线、异步传输模式总线、HIPPI 总线、超级 HIPPI 总线、SerialPlus 总线、SCI/LAMP 总线、光纤信道总线、串行 SCSI 总线或 HDMI 总线。

[0034] 图 1B 和 1C 中所描述类型的计算装置 100 通常地在操作系统的控制下运行，操作系统控制任务的调度和对系统资源的访问。计算装置 100 可运行任一操作系统，诸如任一版本的 MICROSOFT WINDOWS 操作系统；不同发行版本的 Unix 和 Linux 操作系统；用于 Macintosh 计算机的 MAC OS 的任意版本；任意嵌入式操作系统；任意实时操作系统；任意开源操作系统；任意专有操作系统；任意用于移动计算装置的操作系统，或者能够在计算装置上运行且执行此处所描述的操作的任意其它操作系统。典型的操作系统包括但不限于：WINDOWS 3.x、WINDOWS 95、WINDOWS 98、WINDOWS 2000、WINDOWS NT 3.51、WINDOWS NT 4.0、WINDOWS CE、WINDOWS MOBILE、WINDOWS XP 和 WINDOWS VISTA，所有这些均由位于美国华盛顿州 Redmond 的微软公司出品；由位于美国加州 Cupertino 的苹果计算机出品的 Mac OS；由位于美国纽约州 Armonk 的国际商业机器公司出品的 OS/2；以及由位于美国犹他州 Salt Lake City 的 Caldera 公司发布的可免费使用的 Linux 操作系统或者任意类型和 / 或形式的 Unix 操作系统，以及其它。

[0035] 计算机系统 100 可以是任意工作站、电话、台式电脑、膝上型或笔记本电脑、服务器、掌上电脑、移动电话或其他便携电信设备、媒体播放设备、游戏系统、移动计算装置，或能够通信的任意其它类型和 / 或形式的计算、电信或者媒体装置。计算机系统 100 有足够的处理器能力和存储容量以执行此处所述的操作。例如，计算机系统 100 可包括由位于美国加州 Cupertino 的苹果计算机公司出品的 IPOD 系列设备、由位于日本东京的索尼公司出品

的PLAYSTATION2、PLAYSTATION3或PERSONAL PLAYSTATION PORTABLE(PSP)设备，由位于日本Kyoto的Nintendo有限公司出品的NINTENDO DS、NINTENDO GAMEBOY、NINTENDO GAMEBOY ADVANCED或NINTENDO REVOLUTION设备，或者由位于美国华盛顿州Redmond的微软公司出品的XBOX或XBOX 360设备。

[0036] 在一些实施例中，计算装置100可以有不同的处理器、操作系统以及与该装置一致的输入设备。例如，在一个实施例中，计算装置100是由Palm公司出品的TREO180、270、600、650、680、700p、700w/wx、750、750p、800w、Centro、Pro智能电话。在这些实施例的一些中，TERO智能电话是在PalmOS操作系统的控制下操作并且包括笔输入设备以及五向导航设备。

[0037] 在其他实施例中，计算装置100是移动设备，例如，JAVA使能蜂窝电话或个人数字助理(PDA)，诸如i55sr、i58sr、i85s、i88s、i90c、i95c1、i335、i365、i570、I576、i580、i615、i760、i836、i850、i870、i880、i920、i930、ic502、ic602、ic902、i776或im1100，以上这些均由位于美国伊利诺伊州Schaumburg的摩托罗拉公司制造；由位于日本Kyoto的Kyocera公司制造的6035或7135；或者由位于韩国首尔的Samsung电子有限公司制造的i300或i330。在一些实施例中，计算装置100是由位于芬兰的诺基亚公司制造或由位于瑞典Lund的Sony Ericsson MobileCommunications AB公司制造的移动设备。

[0038] 在其他实施例中，计算装置100是黑莓(Blackberry)手持或智能电话，诸如由Research In Motion有限公司制造的设备，包括黑莓7100系列、8700系列、7700系列、7200系列、黑莓7520或黑莓PERAL 8100、8700系列、8800系列、黑莓Storm、黑莓Bold、黑莓Curve 8900、黑莓Pearl Flip。在其他实施例中，计算装置100是智能电话、Pocket PC、Pocket PC电话，或支持Microsft Windows Mobile Software的其它手持移动设备。此外，计算装置100可以是任意工作站、台式电脑、膝上型或笔记本电脑、服务器、掌上电脑、移动电话、任意其它计算机、或能够通信并有足够的处理器能力和存储容量以执行此处所述操作的其它形式的计算或者电信设备。

[0039] 在一些实施例中，计算装置100是数字音频播放器。在这些实施例的一个中，计算装置100是诸如由位于美国加州Cupertino的苹果计算机公司制造的Apple IPOD、IPOD Touch、IPOD NANO和IPOD SHUFFLE系列设备的数字音频播放器。在这些实施例的又一个中，数字音频播放器可作为便携媒体播放器和作为大容量存储装置来工作。在其他实施例中，计算装置100是诸如由位于美国新泽西州Ridgefield Park的Samsung Electronics America公司制造的DigitalAudioPlayer Select MP3播放器，或由位于美国伊利诺伊州Schaumburg的Motorola公司制造的Motorola m500或m25的数字音频播放器。在其他实施例中，计算装置100是便携媒体播放器，诸如由CreativeTechnologies有限公司制造的ZEN VISION W、ZEN VISION系列、ZEN PORTABLEMEDIA CENTER设备或MP3播放器的Digital MP3系列。仍在其他实施例中，计算装置100是支持以下文件格式的便携媒体播放器或数字音频播放器，这些文件格式包括，但不限于MP3、WAV、M4A/AAC、WMA Protected AAC、AIFF、Audible音频书、Apple Lossless音频文件格式和.mov、.m4v以及.mp4MPEG-4(H.264/MPEG-4AVC)视频文件格式。

[0040] 在一些实施例中，计算装置100包括设备的组合，例如与数字音频播放器或便携媒体播放器相组合的移动电话。在这些实施例的一个中，计算装置100是智能电话，例如由

苹果计算机公司制造的 iPhone，或由 Research InMotion 有限公司制造的黑莓设备。又一个实施例中，计算装置 100 是装有 web 浏览器和麦克风与扬声器系统（如电话耳机）的笔记本电脑或台式电脑。在这样的实施例中，计算装置 100 是 web 使能的 (web-enabled) 的并且能接听和启动电话呼叫。在其他实施例中，计算装置 100 是组合数字音频播放器和移动电话的 Motorola RAZR 或 Motorola ROKR 系列。

[0041] 计算装置 100 可以是文件服务器、应用服务器、web 服务器、代理服务器、设备、网络设备、网关、应用网关、网关服务器、虚拟化服务器、部署服务器、SSL VPN 服务器或防火墙。在一些实施例中，计算装置 100 提供远程认证拨入用户服务，并被称为 RADIUS 服务器。在其他实施例中，计算装置 100 可以有作为应用服务器或者作为主应用服务器工作的能力。在其他实施例中，计算装置 100 是刀片服务器。

[0042] 在一个实施例中，计算装置 100 可包括活动目录。计算装置 100 可以是应用加速设备。对于其中计算装置 100 是应用加速设备的实施例，计算装置 100 可提供包括防火墙功能、应用防火墙功能，或负载平衡功能的功能。在一些实施例中，计算装置 100 包括诸如由位于美国加州 San Jose 的 Citrix 应用网络组、位于美国加州 Mountain View 的 Silver Peak Systems 公司、位于美国加州 San Francisco 的 Riverbed Technology 公司、位于美国华盛顿州 Seattle 的 F5 Networks 公司或位于美国加州 Sunnyvale 的 JuniperNetworks 公司制造的系列设备中的一个设备。

[0043] 在其他实施例中，计算装置 100 可被称为客户机节点、客户机器、端节点或端点。在一些实施例中，客户机 100 具有作为客户机节点寻求访问服务器所提供的资源的能力以及作为服务器节点为其它客户机提供对所寄载的资源访问的能力。

[0044] 在一些实施例中，第一客户机计算装置 100a 与第二服务器计算装置 100b 通信。在一个实施例中，客户机与服务器群中的计算装置 100 之一进行通信。例如，通过网络，客户机可以请求执行由服务器群中的计算装置 100 所寄载的各种应用以及接收应用执行结果的输出数据以用于显示。在一个实施例中，客户机执行程序邻近应用 (program neighborhood application) 以与服务器群中的计算装置 100 通信。

[0045] 计算装置 100 可以执行、操作或者以其它方式提供应用，该应用可以是任何类型和 / 或形式的软件、程序或可执行指令，诸如任何类型和 / 或形式的 web 浏览器、基于 web 的客户机、客户机 - 服务器应用、瘦客户端计算客户机、ActiveX 控件、或 Java 小程序，或者能够在计算装置 100 上执行的任何其他类型和 / 或形式的可执行指令。在一些实施例中，应用可以是由第二计算装置代表第一计算装置的用户执行的基于服务器的或者基于远程的应用。在其他实施例中，第二计算装置可使用任何瘦客户端或远程显示协议向第一客户计算装置显示输出的数据，所述协议如：由位于美国佛罗里达州 Ft.Lauderdale 的 Citrix Systems 公司制造的独立计算架构 (ICA) 协议；或由位于美国华盛顿州 Redmond 的 Microsoft 公司制造的远程桌面协议 (RDP)；X11 协议；由 AT&T 贝尔实验室制造的虚拟网络计算 (VNC) 协议；由位于美国加州 Sunnyvale 和以色列 Raanana 的 Qumranet 公司制造的 SPICE 协议；由位于美国加州 Milpitas 的 VESA 公司制造的 Net2Display 协议；由位于加拿大不列颠哥伦比亚省 Burnaby 的 Teradici 公司制造的 PC-over-IP 协议；由位于美国加州 San Jose 的 Wyse Technology 公司制造的 TCX 协议；由位于美国纽约州 NewYork 的纽约市哥伦比亚大学开发的 THINC 协议；或者由位于美国马萨诸塞州 Chelmsford 的 Desktone 公

司制造的Virtual-D协议。应用可以使用任何类型的协议，并且其可以是例如HTTP客户机、FTP客户机、Oscar客户机或Telnet客户机。在其他实施例中，应用包括与诸如软IP电话的网络电话(VoIP)通信有关的任意类型的软件。在进一步的实施例中，应用包括与实时数据通信有关的任何应用例如用于流式传输视频和/或音频的应用。

[0046] 一些实施例中，第一计算装置100a代表客户机计算装置100b的用户执行应用。其他实施例中，计算装置100a执行虚拟机，其提供执行会话，在该会话中，代表用户或客户机计算装置100b执行应用。一个这样的实施例中，执行会话是寄载的桌面会话。另一个这样的实施例中，计算装置100执行终端服务会话。终端服务会话可以提供寄载的桌面环境。在另一个这样的实施例中，执行会话提供对计算环境的访问，计算环境可包括以下的一个或多个：应用、多个应用、桌面应用以及在其中可执行一个或多个应用的桌面会话。

[0047] 现参考图2，框图描述了用于由管理程序控制对物理资源的访问的系统的一个实施例。总的来说，该系统包括虚拟机250、管理程序101和物理盘228。管理程序101在计算装置100上执行，该计算装置100还提供物理盘228。虚拟机250包括客户操作系统110和虚拟盘142。物理盘228存储虚拟盘142的至少一部分。管理程序101在计算装置100上执行；例如，管理程序可在和图1A-1C有关的上文所描述的物理CPU221上执行。管理程序101给虚拟盘142分配对物理盘228的访问。管理程序101确定物理盘228的利用程度超过阀值。响应于所述确定，管理程序101限制虚拟盘142对物理盘228的访问。

[0048] 现参考图2，更具体地，物理盘228是由计算装置100提供的并且存储虚拟盘142的至少一部分。一些实施例中，虚拟盘142与多个物理盘228相关联。在这些实施例的一个中，正如与图1A-1C有关的上文所描述的，一个或多个计算装置100可以与一个或多个其他计算装置100交换关于资源池中可用的处理器和其他物理资源的数据，这允许管理程序管理分布在多个物理计算装置上的资源池。这样的实施例中，管理程序101可给虚拟盘142分配对多个计算装置100的每一个所提供的多个物理盘228的访问。

[0049] 管理程序在计算装置100的处理器上执行。管理程序给虚拟盘分配对物理盘的访问量。一个实施例中，管理程序101分配物理盘的一定量的空间。又一个实施例中，管理程序101分配物理盘的多个页面。一些实施例中，管理程序101将虚拟盘142提供为初始化和执行虚拟机250的进程的一部分。

[0050] 一个实施例中，管理程序101包括用于确定物理盘228的利用程度超过阀值的功能以及用于响应于所述确定来限制虚拟盘142对物理盘228的访问的功能。一些实施例中，管理程序101可包括用于沿虚拟化的输入/输出(I/O)路径拦截虚拟盘142对物理盘228的调用的功能。在这些实施例的一个中，管理程序101包括拦截和转换功能，其允许管理程序101拦截来自虚拟盘142的用于访问物理盘228的访问请求(例如拦截读或写物理盘228的请求)并且将来自虚拟盘142的请求中所包含的地址转换成物理盘228能够处理的格式。在这些实施例的又一个中，修改拦截和转换功能以便包括允许管理程序确定物理盘228的利用程度超过阀值的功能。在这些实施例的又一个中，管理程序101包括从物理盘228接收错误消息并从该错误确定物理盘228的利用程度超过阀值的组件。在这些实施例的又一个中，管理程序101包括用于响应于所述确定来限制虚拟盘142对物理盘228的访问的功能。一些实施例中，拦截和转换功能也被称为块抽头模块(block tap module)。

[0051] 一个实施例中，管理程序101中的模块接收指示物理盘228的利用程度已超过阀

值的错误消息。又一个实施例中，管理程序 101 内的模块确定是否以及如何限制虚拟盘 142 对物理盘 228 的访问。又一个实施例中，该模块修改虚拟盘 142 的配置以限制访问。仍在又一个实施例中，管理程序 101 内的模块将要施加于虚拟盘 142 的限制的标识符传输到管理程序 101 内的第二模块。又一个实施例中，管理程序 101 内的模块是块抽头模块。

[0052] 一个实施例中，控制操作系统 105 中的模块从管理程序 101 接收指示物理盘 228 的利用程度已超过阀值的错误消息。又一个实施例中，控制操作系统 105 中的模块确定是否以及如何限制虚拟盘 142 对物理盘 228 的访问。又一个实施例中，控制操作系统 105 中的模块修改虚拟盘 142 的配置以限制访问。又一个实施例中，控制操作系统 105 中的模块将要施加于虚拟盘 142 的限制的标识符传输到管理程序 101。

[0053] 现参考图 3，流程图描述了用于由管理程序控制对物理资源的访问的方法的一个实施例。总的来说，该方法包括由管理程序给虚拟机中的虚拟盘分配对至少一个物理盘的访问量 (302)。该方法包括由管理程序确定至少一个物理盘的利用程度超过阀值 (304)。该方法包括由管理程序响应于所述确定，限制虚拟盘对物理盘的访问 (306)。

[0054] 现参考图 3，更具体地，管理程序给虚拟机中的虚拟盘分配对至少一个物理盘的访问量 (302)。一个实施例中，管理程序 101 在虚拟机的实例化期间分配访问量。又一个实施例中，管理程序 101 检索要分配的访问量的标识符并在虚拟机的初始化期间分配访问；例如，配置文件或虚拟机映像可指定要分配的访问量的值。

[0055] 管理程序确定至少一个物理盘的利用程度超过阀值 (304)。一个实施例中，管理程序 101 确定已经到达至少一个物理盘 228 的容量限制。又一个实施例中，管理程序 101 确定虚拟机访问盘的速度已经超过预定的阀值。又一个实施例中，管理程序 101 确定物理盘上已发生错误。又一个实施例中，管理程序 101 从物理盘 228 接收指示已超过阀值的错误消息。

[0056] 一个实施例中，管理程序 101 确定虚拟盘 142 已经超过所分配的对物理盘 228 的访问量。又一个实施例中，管理程序 101 确定在计算装置 100 上执行的第二虚拟机 106b 已经超过所分配的对物理盘 228 的访问量。又一个实施例中，管理程序 101 确定在第二计算设装置 100b 上执行的并且具有所分配的对计算装置 100 上的物理盘 228 的访问量的进程（例如应用、操作系统或虚拟机）已经超过所分配的对物理盘 228 的访问量。

[0057] 一个实施例中，管理程序 101 响应于接收来自物理盘 228 的错误消息来做出已经超过阀值的确定。又一个实施例中，管理程序 101 将规则应用于来自物理盘 228 的错误消息以便确定是否已经超过阀值。又一个实施例中，管理程序 101 从控制操作系统 105 接收已超过阀值的指示。

[0058] 一些实施例中，在控制操作系统 105 中执行的监控代理将提供物理盘 228 状态的请求直接或经由管理程序 101 传输到物理盘 228。在这些实施例的一个中，监控代理接收来自物理盘 228 的状态的标识符；例如，物理盘 228 可向监控代理传输物理盘 228 上可用的未被使用的盘空间量的标识符。在这些实施例的又一个中，监控代理应用规则以确定物理盘 228 的状态是否超过阀值。在这些实施例的又一个中，监控代理向管理程序 101 传输已超过阀值的确定。在这些实施例的又一个中，监控代理向管理程序 101 传输用于限制虚拟盘 142 对物理盘 228 访问的方法的标识符。

[0059] 一些实施例中，管理程序 101 对虚拟机 250 屏蔽物理盘 228 的错误。在这些实施

例的一个中，该错误指示已经超过阀值或者已经达到物理盘 228 的容量限制。在这些实施例的又一个中，该错误包括物理盘 228 的硬件故障。在这些实施例的又一个中，该错误包括报告故障的超时错误以便给一个时间段内的请求提供响应。在这些实施例的又一个中，该错误是不合适配置的结果，例如不合适的权限设置。

[0060] 响应于所述确定，管理程序限制虚拟盘对至少一个物理盘的访问 (306)。一个实施例中，管理程序 101 对虚拟盘 142 所请求的输入 / 输出 (I/O) 操作进行速度限制。又一个实施例中，管理程序 101 挂起虚拟机 250 到盘，例如物理盘 228 或第二计算装置 100 上的物理盘 228b。又一个实施例中，管理程序 101 临时停止虚拟机 250 的执行 (其也可以被称为“暂停”虚拟机 250)。又一个实施例中，管理程序 101 挂起由虚拟机 250 执行的至少一个进程。又一个实施例中，管理程序 101 降低虚拟机 250 执行速度。一些实施例中，管理程序 101 响应于来自控制操作系统 105 的指令来限制虚拟盘 142 对物理盘 228 的访问。

[0061] 一些实施例中，管理程序 101 应用策略以确定是否以及如何限制虚拟机对物理盘的访问。其他实施例中，管理程序 101 访问配置文件以确定是否以及如何限制虚拟机对物理盘 228 的访问。在这些实施例的一个中，例如，管理程序 101 访问文件，所述文件识别所分配的访问量，定义阀值以及指定对限制访问所采取的行为。在这些实施例的又一个中，管理程序 101 访问包含阀值和要施加的限制的类型之间映射的文件。通过示例但不限于，管理程序 101 可访问包含如下映射的文件 :

[0062]

分配量	阀值	行为
分配的存储量为 2GB	总的可用空间<=物理盘上可利用的可用空间的 1%	终止虚拟机的执行
分配的存储量为 2GB	总的可用空间<=物理盘上可利用的可用空间的 2.5%	对虚拟盘对物理盘的 I/O 请求进行速度限制
分配的存储量为 2GB	15 秒之内服务所有单个请求	通过时间相对变慢效应 (time dilation) 对虚拟机的执行进行速度限制
分配的存储量为 2GB	总的可用空间<=物理盘上可利用的可用空间的 5%	暂时停止虚拟机执行的至少一个进程(“暂停”)

[0063] 在上面的例子中，如果管理程序 101 确定物理盘 228 上可用空间总量小于 1%，则管理程序通过终止虚拟机的执行来限制虚拟盘 142 对物理盘 228 的访问。在这个例子中，如果管理程序 101 确定物理盘 228 上可用空间总量小于 2.5%，则管理程序通过对来自虚拟盘 142 对物理盘 228 的输入 / 输出请求进行速度限制来限制虚拟盘 142 对物理盘 228 的访

问。在上面的例子中，如果管理程序 101 确定物理盘 228 对单个请求的响应多于 15 秒，则管理程序通过对虚拟机的执行进行速度限制来限制虚拟盘 142 对物理盘 228 的访问，例如，通过时间相对变慢效应。在上面的例子中，如果管理程序 101 确定物理盘 228 上可用空间总量小于 5%，则管理程序通过临时停止虚拟机执行的至少一个进程来限制虚拟盘 142 对物理盘 228 的访问；例如，管理程序 101 可指示虚拟机 250 保存进程的执行状态和该进程所访问的数据的状态并暂停该进程的执行。上面所描述的阀值和行为仅是示例，管理程序 101 可应用与上面所描述的那些阀值不同的阀值或限制；另外，管理员可配置与上文所述不同的阀值或行为。

[0064] 在一个实施例中，上文所述的方法和系统允许管理程序确定物理盘上已发生错误并灵活地管理虚拟机的执行以便从虚拟机屏蔽该错误。又一个实施例中，例如，管理程序可基于已发生的错误的类型来确定如果虚拟机挂起其执行的进程的子集，那么该虚拟机可继续执行，而如果发生第二类型的错误，则确定如果管理程序对虚拟机的输入 / 输出操作进行速度限制，那么虚拟机可以继续执行。又一个实施例中，通过确定物理盘上已超过阀值以及识别用于限制虚拟盘对物理盘的访问的多个方法之一，此处所述的方法和系统提供用于控制对物理资源访问的功能的改善。

[0065] 应该理解，此处描述的系统可提供这些组件中的任意数个或每个组件并且这些组件可以在单独机器上提供，或者在一些实施例中，可在分布式系统的多个机器上提供。此外，上述系统和方法可作为一件或多件产品上所体现的或在其中的一个或多个计算机可读程序而被提供。所述产品可以是软盘、硬盘、CD-ROM，闪存卡、PROM、RAM、ROM 或磁带。通常，计算机可读程序可以任何编程语言来实现，如 LISP、PERL、C、C++、C#、PROLOG，或者诸如 JAVA 的任何字节码语言。软件程序可以作为目标代码被存储在一件或多件产品上或其中。

[0066] 已经描述了用于由管理程序控制对物理资源访问的方法和系统的某些实施例，对本领域技术人员而言，显而易见可以使用包含所公开的概念的其他实施例。因此，本发明应该不限于某些实施例，而应由下列权利要求的精神和范围进行限制。

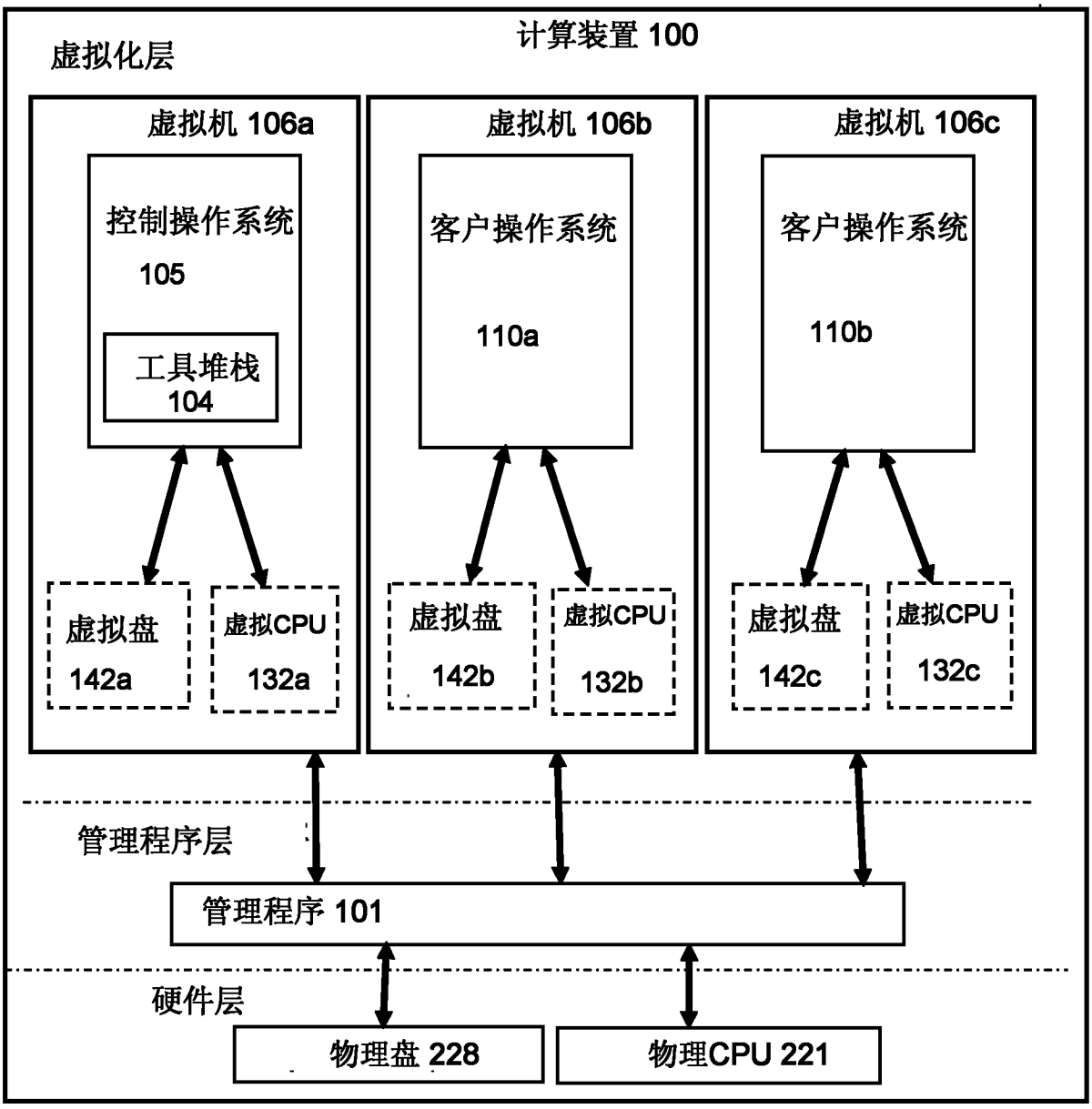

计算装置 100
虚拟化层
虚拟机 106a
虚拟机 106b
虚拟机 106c
控制操作系统
105
工具堆栈
104
客户操作系统
110a
客户操作系统
110b
虚拟盘
142a
虚拟CPU
132a
虚拟盘
142b
虚拟CPU
132b
虚拟盘
142c
虚拟CPU
132c
管理程序层
管理程序 101
硬件层
物理盘 228
物理CPU 221

图 1A

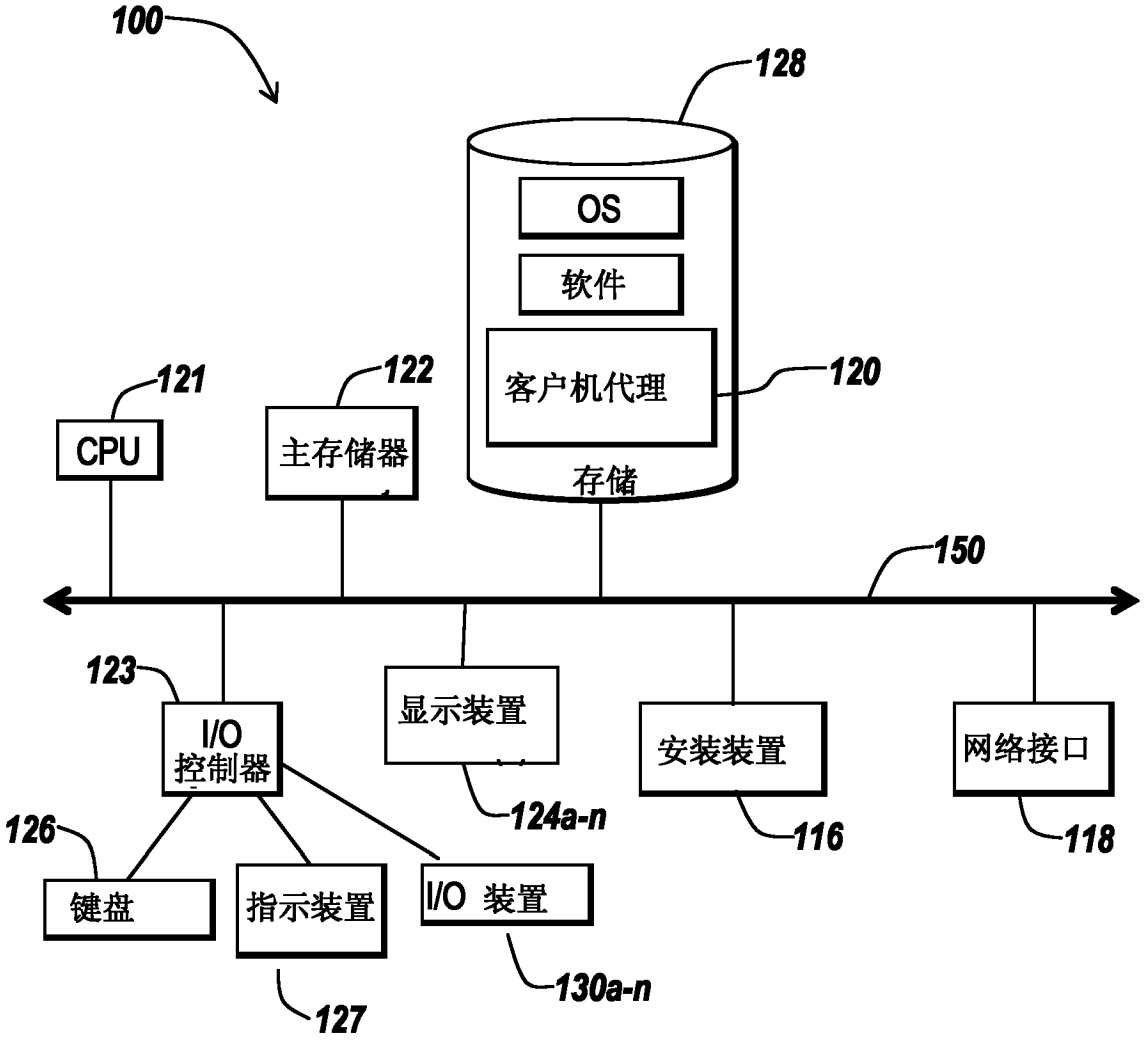

100
128
OS
软件
客户机代理
120
存储
121
CPU
122
主存储器
150
123
I/O
控制器
显示装置
124a-n
安装装置
116
网络接口
118
126
键盘
指示装置
127
I/O 装置
130a-n

图 1B

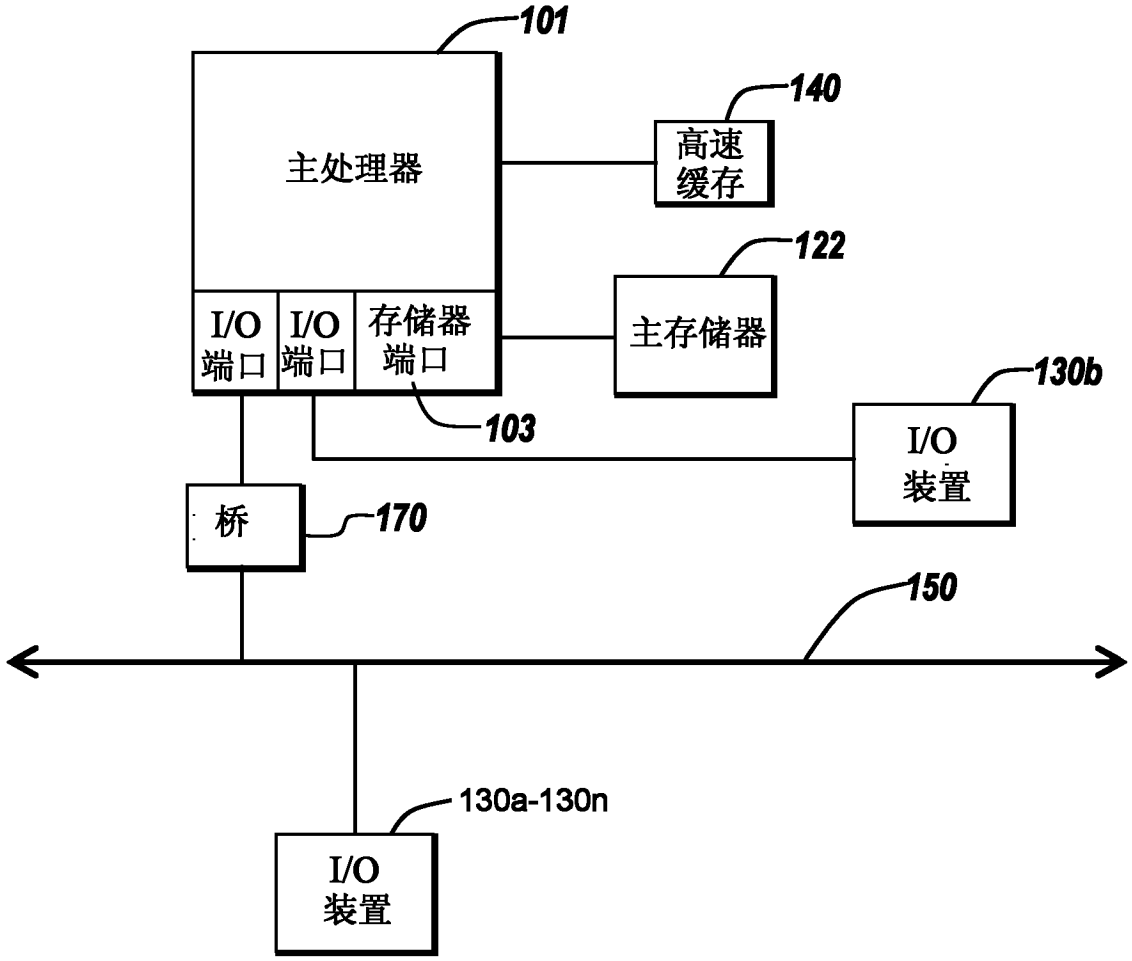

101
主处理器
140
高速
缓存
122
主存储器
I/O
端口
I/O
端口
存储器
端口
103
130b
I/O
装置
桥
170
150
130a-130n
I/O
装置

图 1C

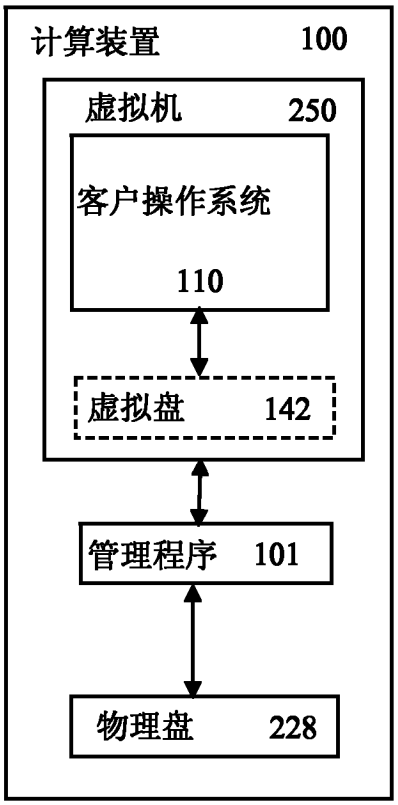

计算装置 100
虚拟机 250
客户操作系统 110
虚拟盘 142
管理程序 101
物理盘 228

图 2

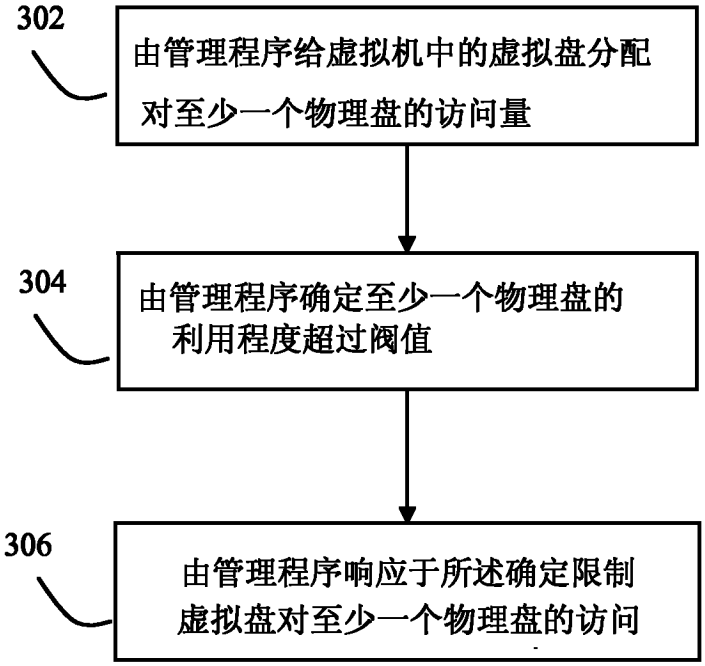

302
由管理程序给虚拟机中的虚拟盘分配对至少一个物理盘的访问量
304
由管理程序确定至少一个物理盘的利用程度超过阀值
306
由管理程序响应于所述确定限制虚拟盘对至少一个物理盘的访问

图 3