[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G06F 9/06

[12] 发明专利说明书

[21] ZL 专利号 01104532.9

[45] 授权公告日 2005 年 1 月 26 日

[11] 授权公告号 CN 1186715C

[22] 申请日 2001.2.15 [21] 申请号 01104532.9

[30] 优先权

[32] 2000. 2. 7 [33] US [31] 09/506,232

[71] 专利权人 国际商业机器公司

地址 美国纽约

[72] 发明人 马修·S·克劳克

杰拉尔德·F·迈克布莱蒂

肖恩·P·穆伦

约翰尼·M－H·施

审查员 刘春霞

[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

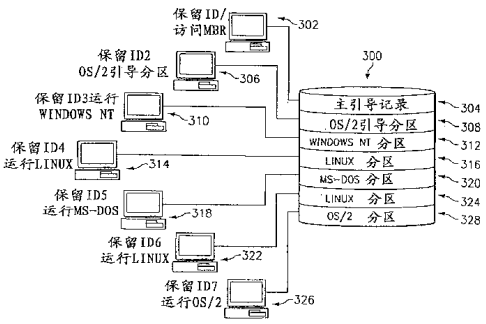
代理人 于 静

权利要求书 3 页 说明书 9 页 附图 3 页

[54] 发明名称 支持多个操作系统的方法和系统

[57] 摘要

提供一种用于同时在不同的计算机上运行来自同一共享系统资源的多个操作系统的方法和系统。这是例如通过采用持久基本盘保留实现的。 每个机器读取不带保留的主引导记录以确定要引导的操作系统的分区。 接着每个机器做出基本排它写持久保留以访问该操作系统引导分区。 接着每个机器做出另一个基本排它写持久保留以访问该操作系统分区本身。 对每个机器分配一个不同的操作系统分区，即使它们正在运行相同的操作系统。



ISSN1008-4274

1. 一种用于在多个处理器上引导来自一个共享系统资源的多个操作系统的方法，该方法包括:

通过该多个处理器中的一个处理器从该共享系统资源读出主引导记录以确定要使用的该多个操作系统中的一个操作系统的分区;

通过该处理器利用一个唯一的保留关键字保留该共享系统资源的一部分以访问一个操作系统引导分区; 以及

通过该处理器利用该唯一保留关键字保留该共享系统资源的一部分以访问一个操作系统分区，其中所述保留该共享系统资源的一部分是通过在该共享系统资源上的一序列连续扇区上进行基本排它写持久保留实现的。

2. 权利要求 1 的方法，其中在引导过程开始时，该处理器以不带保留的方式从共享系统资源读主引导记录。

3. 权利要求 1 的方法，其中从下述: 处理器 ID、群集器 ID、多处理器分区 ID、非均匀存储器存取复合体 ID 和非均匀存储器存取节点 ID 中至少之一建立用于为该操作系统引导分区以及该操作系统分区保留该共享系统资源的一部分的该唯一保留关键字。

4. 权利要求 1 的方法，其中该共享系统资源是下述之一: 系统硬盘机、ZIP 盘机、JAZ 盘机、可重写 CD 盘、软盘机或磁带机。

5. 一种在分布式数据处理系统中用于在该分布式数据处理系统中的多个处理器上执行来自一个共享系统资源的多个操作系统的方法，该方法包括处理器实现的下述步骤:

通过该多个处理器中的每个处理器从该共享系统资源读主引导记录以确定用于该多个处理器的各操作系统;

通过该多个处理器中的每个处理器利用对该多个处理器中的每个为唯一的一个保留关键字保留该共享系统资源的一部分，其中利用该部分访问一个操作系统引导分区; 以及

通过该多个处理器中的每个处理器利用该保留关键字保留该共

享系统资源的一部分以访问一个操作系统分区，其中该多个处理器并发地执行来自该共享系统资源的该多个操作系统，其中所述保留该共享系统资源的一部分是通过在该共享系统资源上的一序列连续扇区上进行基本排它写持久保留实现的。

6. 一种在数据处理系统中用于从多个处理器在一个共享系统资源上安装多个操作系统的方法，该方法包括该数据处理系统实现的下述步骤：

多个处理器中的一个处理器安装该多个操作系统中的一个操作系统，保留对该共享系统资源上的一个主引导记录的排它访问，其中所述保留对该共享系统资源上的一个主引导记录的排它访问是通过在该共享系统资源包含该主引导记录的部分上进行基本排它写持久保留实现的；

该处理器把该操作系统引导信息添加到该主引导记录上的一个分区表上；以及

该处理器释放对该主引导记录的排它访问。

7. 权利要求 6 的方法，其中若该主引导记录上的该基本排它写持久保留失败，向操作系统安装程序发送出错消息。

8. 一种用于在多个处理器上引导来自一个共享系统资源的多个操作系统的分布式数据处理系统，该数据处理系统包括：

确定装置，用于为该多个处理器中的一个处理器从该共享系统资源读一个主引导记录以确定要引导的该多个操作系统中的一个操作系统；

第一保留装置，用于由该处理器利用一个唯一的保留关键字保留该共享系统资源的一部分以访问一个操作系统引导分区；以及

第二保留装置，用于利用该唯一保留关键字保留该共享系统资源的一部分以访问一个操作系统分区，其中所述保留该共享系统资源的一部分是通过在该共享系统资源上连续扇区块上进行基本排它写持久保留实现的。

9. 权利要求 8 的系统，其中在引导过程开始时，该处理器以不

带保留的方式从共享系统资源读该主引导记录。

10. 权利要求 8 的系统，其中从下述：处理器 ID、群集器 ID、多处理器分区 ID、非均匀存储器存取复合体 ID 和非均匀存储器存取节点 ID 中至少之一建立用于为该共享系统资源上的该操作系统引导分区以及该操作系统分区保留该共享系统资源的一部分的该唯一保留关键字。

11. 一种分布式数据处理系统，用于在该分布式数据处理系统中的多个处理器上执行来自一个共享系统资源的多个操作系统，该分布式数据处理系统包括：

读装置，用于通过该多个处理器中的每个处理器从该共享系统资源读主引导记录确定用于该多个处理器的各操作系统；

第一保留装置，用于通过该多个处理器中的每个处理器利用对该多个处理器中的每个为唯一的一个保留关键字保留该共享系统资源的一部分，其中利用该部分访问一个操作系统引导分区；以及

第二保留装置，用于通过该多个处理器中的每个处理器利用该保留关键字保留该共享系统资源的一部分以访问一个操作系统分区，其中该多个处理器并发地执行来自该共享系统资源的该多个操作系统，其中所述保留该共享系统资源的一部分是通过在该共享系统资源上的一序列连续扇区上进行基本排它写持久保留实现的。

支持多个操作系统的方法和系统

技术领域

本发明一般地涉及一种改进型的数据处理系统并且尤其涉及一种支持多个操作系统的方法和设备。更具体地，本发明提供一种方法，其允许驻留在同一物理盘上的多个操作系统同时在不同的计算机上运行。

背景技术

通过使用引导程序实现操作系统（OS）的装入和运行。通常，启动操作系统是一个两步骤过程，该过程涉及一个确定要装入哪个操作系统的“简单”引导程序和一个实际装入选定操作系统的较复杂的引导程序。该一般存储在非易失性系统 RAM（NVRAM）中的简单引导程序用于进行系统资源初始化。具体地，该简单引导程序初始化中央处理单元（CPU）中的寄存器并且初始化设备控制器，例如用于系统盘和存储器的控制器。该简单引导程序可读写存储器并且可从系统盘上的装入引导块。该引导块含有主引导记录（MBR）并位于系统盘驱动器的扇区 0 中。

主引导记录（MBR）是从系统盘装入的，它含有一个分区表和一些可执行码。主引导记录可执行码为单个现用分区扫描该分区表、把第一扇区从该现用分区装入到存储器中并且执行该代码，该代码是用于选定操作系统的引导码。该操作系统引导码装入正被引导的操作系统并且以规定方式启动该操作系统。

若硬盘例如包括 MS－DOS 分区、LINUX 分区、Windows NT 分区和 IBM OS/2 分区，通过改变现用分区用户可改变将启动这些系统中的哪一个系统。通过在非易失系统 RAM（NVRAM）中存储有关信息可设定现用分区。通常最近安装在系统盘驱动器上的操作系统更新 NVRAM，从而会引导该操作系统。但是操作系统可以提供一个允许把某不同的操作系统指定成是 NVRAM 中的现用分区的实用程序。从而这允许下次再

引导引导一个不同的操作系统。

尽管上面的方法允许用户选择系统启动时要引导的操作系统，该方法不允许同时在不同的机器上运行来自同一个系统盘的二个或更多的操作系统。这样，任何使用一个共享系统盘例如其具有 OS/2 分区作为现用分区的机器本身必须在该 OS/2 操作系统下运行。这限制了系统的通用性并且对系统资源加以限制，例如可访问的文件系统的类型。因此，具有一种用于访问共享的系统资源，例如系统盘驱动器，从因不同的机器可在同时运行不同的操作系统并访问适当的系统资源的方法和系统是有好处的。

发明内容

本发明提供一种用于同时在不同的计算机上运行来自同一共享系统资源的多个操作系统的方法和系统。这是例如通过采用持久基本盘保留实现的。每个机器在不带保留下读主引导记录以确定要引导的操作系统的分区。然后每个机器做出一个基本排它写持久保留，以访问操作系统引导分区。接着每个机器做出另一个基本排它写持久保留以访问该操作系统分区本身。对每个机器分配一个不同的操作系统分区，即使它们正在运行相同的操作系统。从处理器 ID、群集器 ID、多处理器分区 ID、非均匀存储器存取复合体 ID 和/或非均匀存储器存取节点 ID 中的至少一个建立用于这些保留的唯一保留关键字。

附图说明

在附属权利要求书中陈述确信为表征本发明的各新颖特性。然而，通过阅读时连带着各附图参照下述对一示例实施例的详细说明会最佳地理解本发明本身以及优选使用方式、它的其它目的和优点，其中附图是：

图 1 是一个示例图，示意在其中可实现本发明的网络数据处理环境；

图 2 是一个示例方块图，示意在其中可实现本发明的数据处理系统；

图 3 是一个示例图，示意依据本发明的一实施例的在相同系统资源

上同时运行的多个操作系统;

图 4 是一个流程图，其概括依据本发明的一实施例的用于在向系统添加一新操作系统时更新主引导记录的示例操作；以及

图 5 是一个流程图，其概括依据本发明的一实施例的示范性引导处理操作。

具体实施方式

在典型的计算机系统中，系统盘机划分成在制造盘时确定的尺寸和数量都固定的物理扇区。盘机上的分区是彼此相连的扇区的逻辑序列。为了使多个操作系统驻留在同一个盘机上，对操作系统的每个拷贝分配它自己的盘上的分区。这种保留可以是排它的，意味着仅可由带有特定的唯一关键字的计算机访问该指定的分区。分区保留还可以是持久的，即，它在启动程序失败下，例如系统的硬复位时，是受到保护的。持久保留由盘机保持直到被释放为止。

持久保留还可以是基本的，意味着它们可以保留硬盘的一组相邻的扇区。若要改变盘的保留部分中的数据，则该基本持久保留是一个写保留。向某计算机授予对盘上一部分的基本排它写持久保留意味着只有具有其唯一保留关键字的计算机才可以访问该盘机的该部分。对于单处理器机器，该关键字可以是处理器 ID。对于多处理器机器，例如非均匀存储器存取（NUMA）机器，该关键字可包含群集器 ID、多处理器分区 ID、NUMA 复合体 ID 或 NUMA 节点 ID。该关键字将简单地称为保留关键字。

现参照各图，图 1 是一个示例图，示出其中可实现本发明的数据处理系统。分布式数据处理系统 100 例如可以是一个计算机网络，诸如局域网（LAN）、广域网（WAN）、因特网、内联网等。分布式数据处理系统 100 包括至少一个网络 102，其是用于向在分布式数据处理系统 100 内连接在一起的各种部件和计算机之间提供通信链路的媒体。网络 102 可包括永久性连接，例如缆线或光纤线，或者通过电话连接的暂时性连接。

在该示出的例子中，服务器 104 以及存储单元 106 连接到网络 102。此外，客户机 108、110、112 以及其它服务器 114、116、118 也连接到网络 102。这些客户机 108、110 和 112 例如可以是个人计算机或网络计算机。对于本申请来说，网络计算机被认为是任何和网络连接的并从和该网络连接的另一个计算机接收程序或其它应用的计算

机。在示出的该例子中，服务器 104 可向客户机 108－112 提供数据和应用。客户机 108、110 和 112 是服务器 104 的客户。分布式数据处理系统 100 可包括其它服务器、客户机、共享的盘机例如盘 106 和 122 以及其它未示出的设备。

在该示出的例子中，分布式数据处理系统 100 是因特网，其中网络 102 代表彼此利用 TCP/IP 协议组进行通信的网络和网关的全球性集合。因特网的核心是一个由成千上万个商业、政府、教育等计算机系统（它们路由数据和消息）组成的主节点或主计算机间的高速数据通信线路的基干。当然，分布式数据处理系统 100 也可实现成一些不同类型的网络，例如内联网、局域网（LAN）或广域网（WAN）。例如，在 LAN 下，可在客户机 108、110、112 之间共享盘机 106，并且这些客户机中的每个可引导一个驻留在盘 106 上的不同的操作系统。

图 1 的意图是作为一个例子，而不是作为对本发明的体系结构限制。例如，对于本发明，"计算机"可以是具有多个处理器的单个机器中的各处理器，每个处理器具有它自己的存储器分区。在该情况下，只要对每个操作系统分配它自己的处理器以及它自己的存储器分区，就可以同时在该 "同一个机器"上运行多个操作系统。

现参照图 2，一个示例方块图示意可在其中实现本发明的数据处理系统。数据处理系统 200 是计算机，例如图 1 中的计算机 108，的一个例子，在其中可设置实现本发明的各过程的代码或指令。数据处理系统 200 采用外围部件互连（PCI）局部总线体系结构。虽然该示出的例子采用 PCI 总线，也可采用其它总线体系结构例如微通道和工业标准体系结构（ISA）。处理器 202 和主存储器 204 通过 PCI 桥 208 和 PCI 局部总线 206 连接。PCI 桥 208 还可以包括一个用于处理器 202 的集成的存储器控制器和高速缓存。

通过直接部件互连或者通过扩充板可对 PCI 局部总线 206 做出附加的连接。在该示出的例子中，局域网（LAN）适配器 210、小型计算机系统接口（SCSI）主总线适配器 212 以及扩充总线接口 214 通过直接部件连接和 PCI 局部总线 206 连接。相反，声频适配器 216、图

形适配器 218 以及声频/视频适配器 219 通过插入到扩充槽中的扩充板和 PCI 局部总线 206 连接。扩充总线接口 214 提供用于键盘和鼠标适配器 220、调制解调器 222 和附加存储器 224 的连接。SCSI 主总线适配器 212 提供用于硬盘机 226、磁带机 228 和 CD－ROM 机 230 的连接。典型的 PCI 局部总线实现可支持三个或四个 PCI 扩充槽或扩充接插件。

一个操作系统在处理器 202 上运行并且用于协调和控制图 2 的数据处理系统 200 内各种部件。该操作系统可以是一个商业上可买到的操作系统，例如可从国际商用机器公司买到的 OS/2、可从微软公司买到的 Windows NT 等。操作系统、应用或程序的指令设置在存储部件例如硬盘机 226 上并且可装入到主存储器 204 中以由处理器 202 执行。还可能如本发明中详述那样，在网络中的共享盘机上如图 1 中的盘 106 上驻留着该操作系统，并且图 2 中描述的计算机系统从该共享的盘机装入操作系统。

业内人士理解可根据实施改变图 2 中的硬件。可以对图 2 中示出的硬件补充地或者替代地使用其它内部硬件或外围设备，例如闪速 ROM（或等同的非易失性存储器）或者光盘机等。另外，本发明的各过程可应用于多处理器数据处理系统。

例如，若选择性地配置成是一个网络计算机，数据处理系统 200 可以不包括 SCSI 主总线适配器 212、硬盘机 226、磁带机 228 和 CD－ROM 230，如图 2 中用虚线 232 表示选用式地包括那样。在该情况下，数据处理系统 200 可包括一个网络通信接口，例如 LAN 适配器 210、调制解调器 222 等。作为另一个例子，数据处理系统 200 可以是一个配置成可不依赖于网络通信接口引导的独立系统，不管数据处理系统 200 是否包括网络通信接口。图 2 中示出的例子以及上面说明的各个例子不意味施加体系结构限制。例如，本发明可在多处理器系统，例如非均匀存储器存取（NUMA）计算机，中实现。

现参照图 3，该示例图示意依据本发明的一优选实施例同时在不同的机器上运行来自同一个盘的多个操作系统。在该示出的例子中，

由多个部件访问的系统资源是硬盘机 300，不过本发明不受限制于这样的实施例。可以使用任何能利用基本排它持久保留分区的系统资源，这不背离本发明的精神和范围。例如，系统资源可以是 ZIP 盘机、JAZ 盘机、可重写 CD 盘等。另外，示出的计算机系统可以是多处理器系统例如非均匀存储器存取计算机中的各个处理器。

盘机 300 可由多个机器访问，出于简明用保留 ID 1 至 7 显出这些机器。对运行盘 300 上某操作系统的每个计算机分配一个独立的用于它的操作系统的分区并且利用它的保留 ID 作为保留该分区的关键字。在该示出的例子中，计算机 302 目前正在引导过程开始时访问主引导记录 304。计算机 306 已经读过主引导记录并已确定要引导 OS/2 操作系统。它现在在执行 OS/2 引导分区 308。计算机 310 目前正运行 Windows NT 操作系统，后者是微软公司的一个产品。它的保留 ID（#3）充当一个关键字以保留盘 300 上的一个用于 Windows NT 分区 312 的基本持久保留。计算机 314 正在运行 Linux 并且利用它的保留 ID（#4）作为关键字保留盘 300 上的 Linux 分区 316。计算机 318 正在运行 MS－DOS 并且利用它的保留 ID（#5）作为关键字保留盘 300 上的 MS－DOS 分区 320。计算机 322 正在运行 Linux，但请注意它在盘 300 上的 Linux 分区 324 和通过计算机 308 保留的 Linux 分区 316 不同。计算机 326 正在运行 OS/2 并且利用它的保留 ID（#7）作为关键字保留盘 300 上的 OS/2 分区 328。这样，通过使用不同的保留 ID 以及共享盘机上的不同分区，不同的计算机可以同时运行来自共享盘机的不同操作系统。

在图 3 中保留 ID 是按简单的整数示出的。业内人士理解，实际值可更复杂。另外，图 3 中盘机 300 上的各分区的示例不表示物理盘空间的相对实际量。例如，主引导记录可能只占盘的一个扇区，而某个操作系统分区可能占数万个扇区。

现参照图 4，图中的流程图概括依据本发明的一实施例的对该系统添加新的操作系统时用来更新主引导记录的示例操作。例如，每次向盘机 300 添加一个操作系统分区，例如图 3 中的分区 312、316、320、

324 和 328，时必须改变主引导记录中的分区表。当安装一个操作系统时，保留系统资源例如盘机 300 上的一个分区并且把该分区信息添加到主引导记录中。重要的是二个操作系统不同时在安装过程中试图修改主引导记录，否则的话主引导记录中的分区表不会包含用于多个操作系统的正确信息。

该操作系统或者代表该操作系统的一个实用程序在主引导记录区上做出基本排它写持久保留（步骤 400）。若该保留不成功（步骤 402: 否），则对主引导记录的修改失败（步骤 404）并向操作系统安装程序发送出错信号（步骤 406）。若在操作系统引导分区上的保留成功（步骤 402: 是），该操作系统或实用程序向主引导记录添加该引导分区信息（步骤 408）。接着该操作系统或实用程序释放主引导记录区上的该基本排它写持久保留（步骤 410）并结束对主引导记录的修改。

业内人士理解，可以通过由该操作系统调用的一个实用程序执行图 4 中说明的过程。通过采用这种方法，对该实用程序的改变不需要改变该操作系统本身。

现参照图 5，图中的流程图概括依据本发明的一实施例的引导过程的一示例操作。此刻假定已经在该同一系统资源上成功地安装二个或更多的操作系统。图 3 中的各机器示成处于引导某操作系统的不同阶段。计算机 302 在读主引导记录，计算机 306 在读 OS/2 引导分区 308，而计算机 310、314、318、322 和 326 已完成引导过程。

最初，引导码读不带保留的系统资源（步骤 500）。接着，引导码为系统资源登记持久保留关键字（步骤 502）。引导码读主引导记录（MBR）并且根据 NVRAM 中存储的现用操作系统分区信息确定用于要引导的操作系统的分区（步骤 504）。接着引导码做出该操作系统引导分区上的基本排它写持久保留（步骤 506）。若该保留不成功（步骤 508: 否），则该引导失败（步骤 510）。

若该操作系统引导分区上的保留成功（步骤 508: 是），则执行该操作系统引导分区中的代码（步骤 512）。该操作系统引导分区在该操作系统分区上进行基本排它写持久保留（步骤 514）。若该保留不成功

（步骤 516: 否），则引导失败（步骤 518）。若该操作系统引导分区上的保留成功（步骤 516: 是），则执行该操作系统分区中的代码（步骤）。

为了避免多个操作系统彼此干扰，每个操作系统具有一个唯一的保留关键字。另外，通过可能在其它机器上存在着对来自相同的盘机的相同操作系统进行引导的多个引导过程的环境中使操作系统引导码设置操作系统分区上的基本持久保留，保护操作系统引导过程的完整性。重要的是请注意，用于不同操作系统分区的文件系统组织不需要是兼容的。所需要的只是每个引导程序能正确地读和解释主引导记录中的信息。

概言之，在现有技术中，可以从同一个盘引导二个或多个操作系统但不能同时运行，即使这些操作系统是在不同的机器上运行时。通过以本发明中表示的按上述说明的方式使用持久基本盘保留，有可能同时在不同的机器上运行来自同一系统资源的多个操作系统。

虽然本发明的上述说明假定多个计算机共享一个系统盘机，本发是不受限于这样的实施例。尤其，本发明中开发的方法可扩充到带有多个处理器的机器，其中每个处理器具有自己的存储器分区并且在自己的操作系统下运行。在这种情况下，只要对每个操作系统分配它自己的处理器以及它自己的存储器分区，可以同时在“同一机器”上运行多个操作系统。业内人士会理解，对于 NUMA（非均匀存储器存取）复合体类似的启动是可能的。

重要的是要注意，尽管本发明是在全功能数据处理系统的环境下说明的，业内人士理解可以以指令的计算机可读媒体的形式在各种形式下分布本发明，并且和为实现这种分布实际上使用的信号承载媒体的具体类型无关地可以等同地应用本发明。计算机可读媒体的例子包括可记录型媒体例如软盘、硬盘机、RAM 和 CD－ROM 以及传输型媒体如数字式和模拟式通信链路。

出于示意和描述的目的已对本发明提出说明，但本发明在所公开的形式上不是排它的或是受限制的。对于业内人士许多修改和变型是明显的。例如，随着存储区网络（SAN）变得更为流行，如本发明中

概述那样会存在更多的共享。该实施例选择和说明的目的是最佳地解释本发明的原理、实际应用并且使业内人士理解本发明的各实施相对于各种预期的特定应用可存在各种修改。

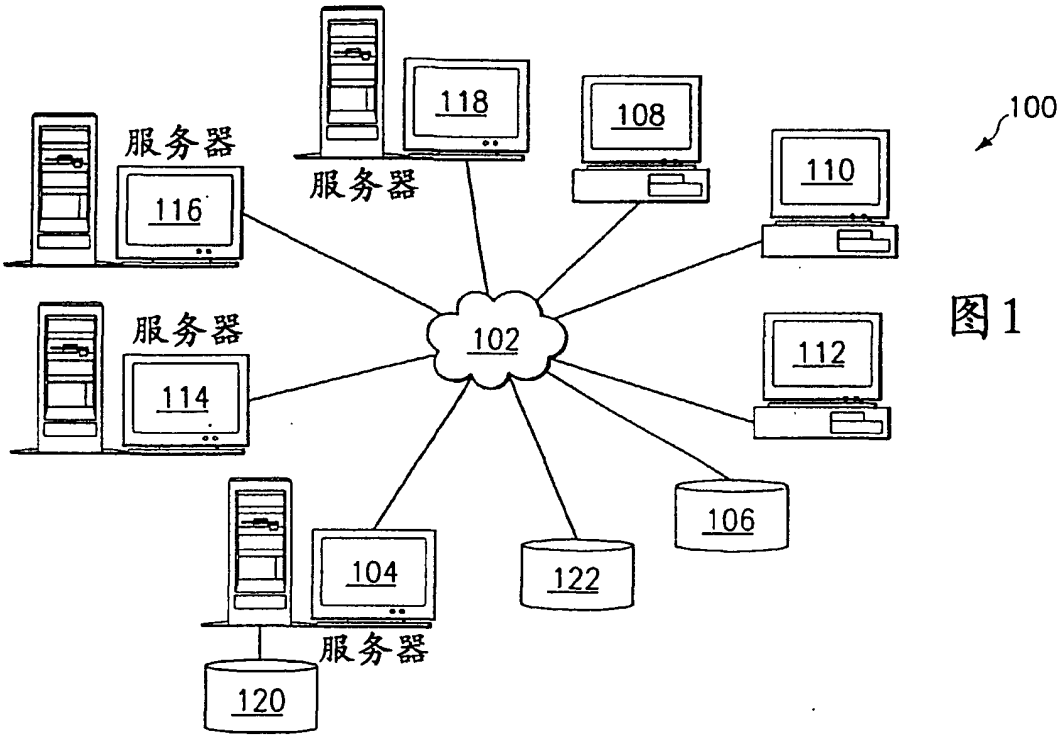

100
服务器
116
服务器
118
108
110
服务器
114
102
112
104
122
106
服务器
120

图1

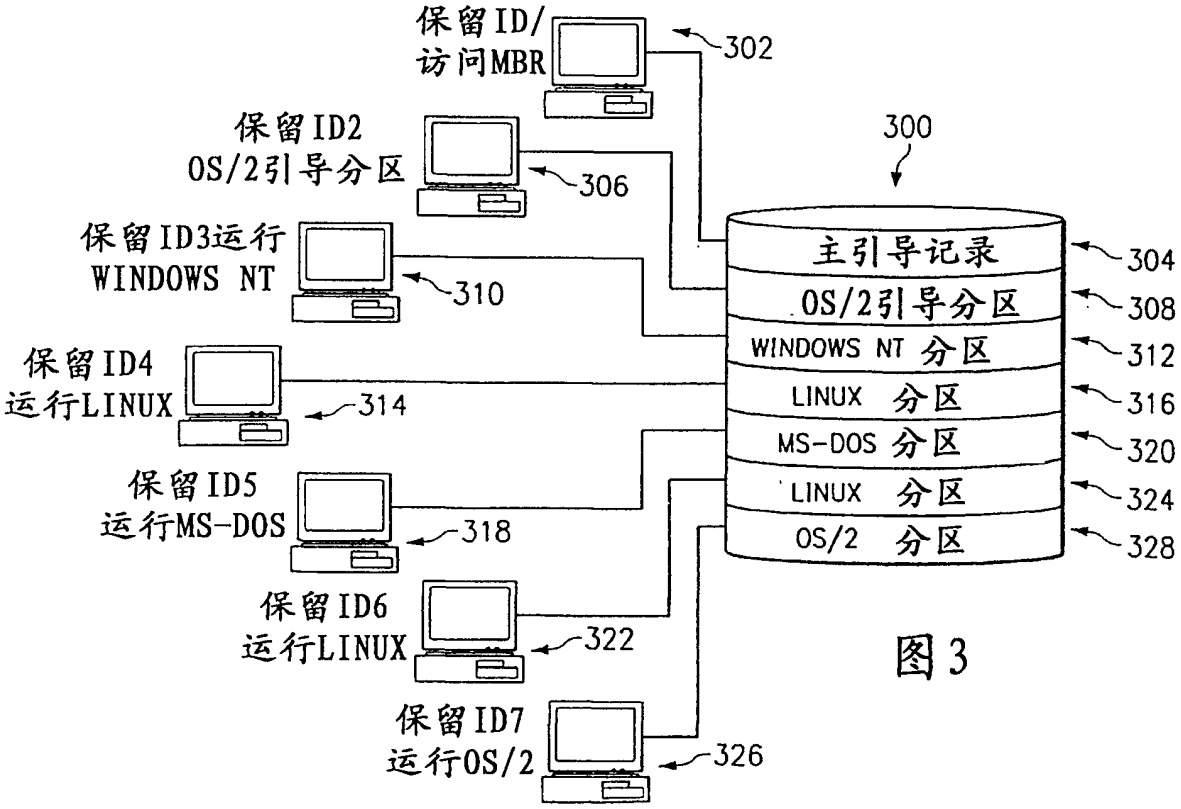

保留ID/
访问MBR
302
保留ID2
OS/2引导分区
306
300
保留ID3运行
WINDOWS NT
310
主引导记录
304
OS/2引导分区
308
WINDOWS NT 分区
312
保留ID4
运行LINUX
314
LINUX 分区
316
MS-DOS 分区
320
保留ID5
运行MS-DOS
318
LINUX 分区
324
OS/2 分区
328
保留ID6
运行LINUX
322
保留ID7
运行OS/2
326

图3

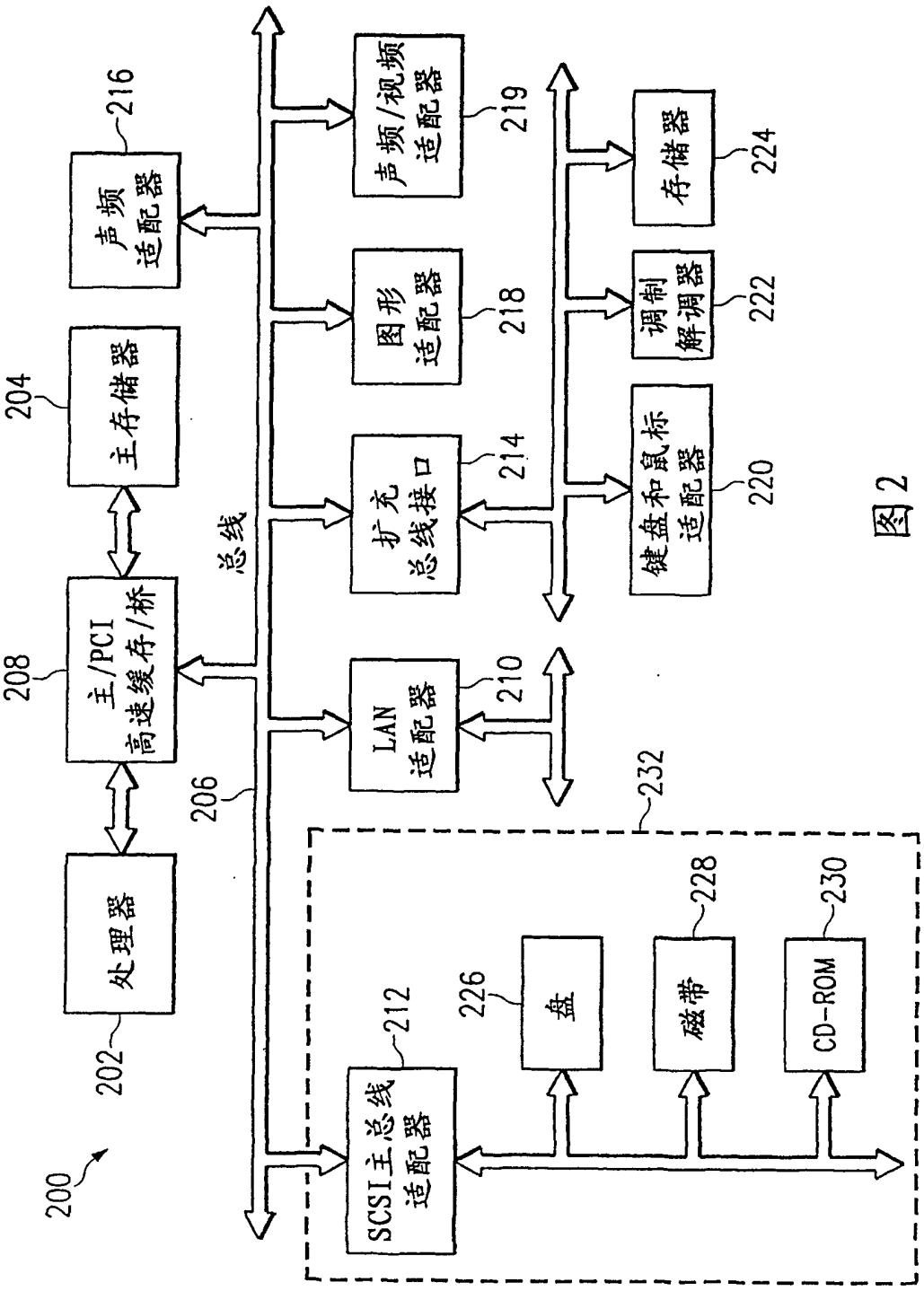

200
202
处理器
208
主/PCI
高速缓存/桥
204
主存储器
206
总线
216
声频
适配器
LAN
适配器
210
扩充
总线接口
214
图形
适配器
218
声频/视频
适配器
219
232
SCSI主总线
适配器
212
226
盘
228
磁带
230
CD-ROM
键盘和鼠标
适配器
220
调制
解调器
222
存储器
224

图 2

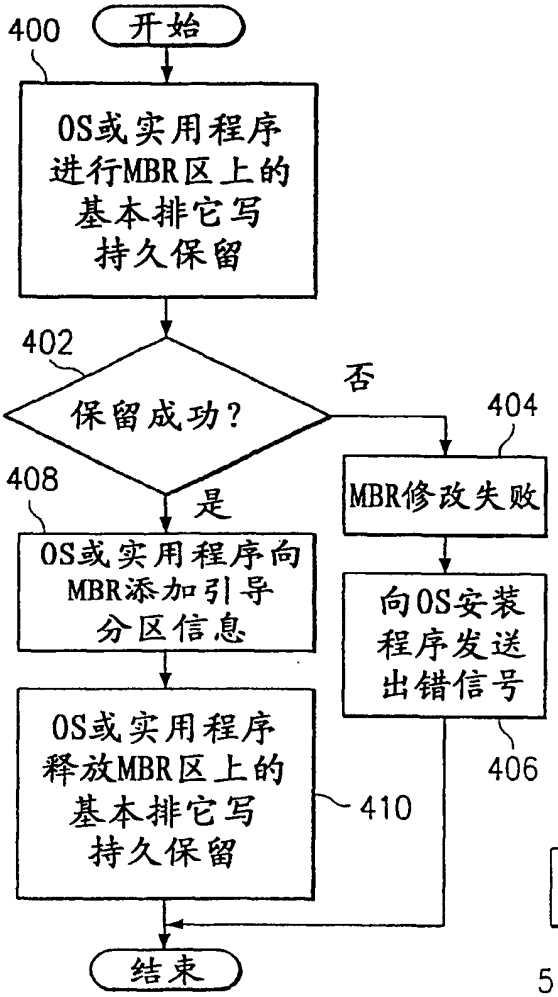

400
开始
OS或实用程序进行MBR区上的基本排它写持久保留
402
保留成功?
否
404
MBR修改失败
向OS安装程序发送出错信号
406
是
408
OS或实用程序向MBR添加引导分区信息
OS或实用程序释放MBR区上的基本排它写持久保留
410
结束

图4

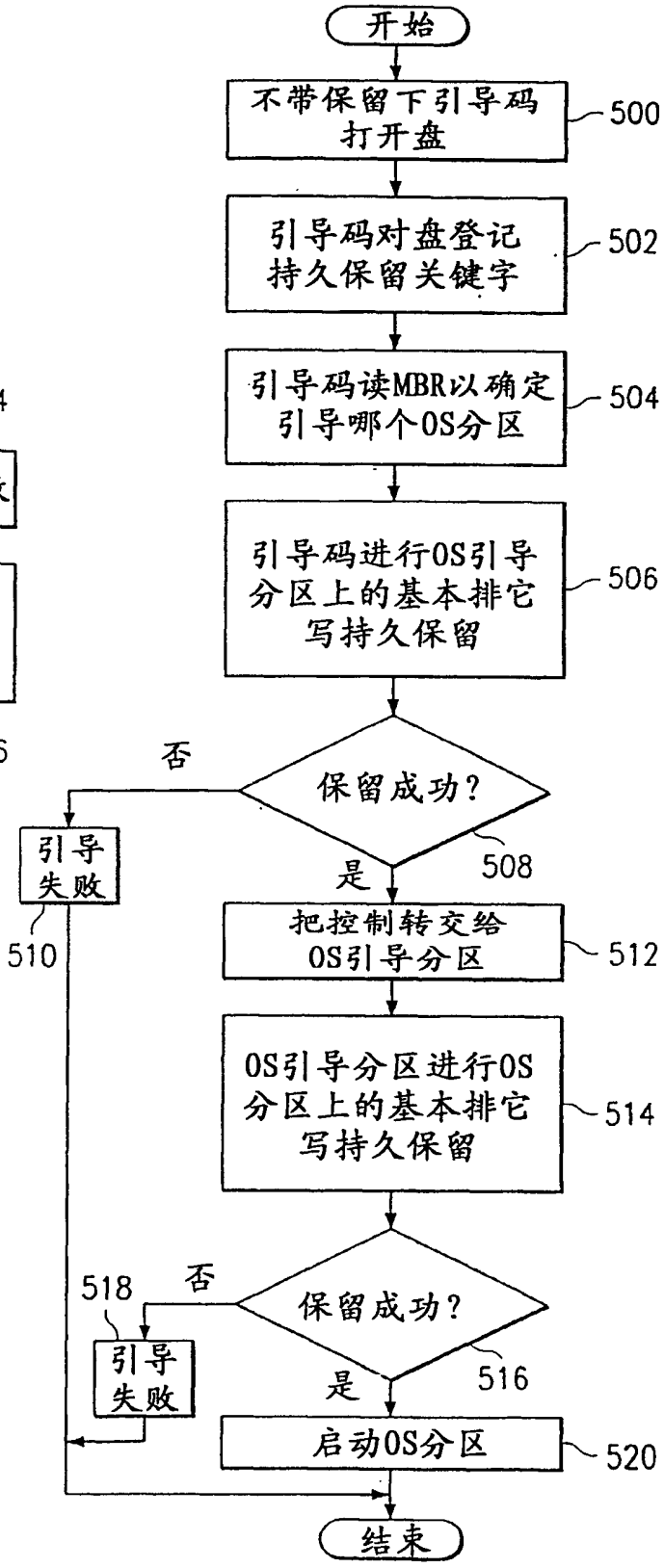

开始
不带保留下引导码打开盘
500
引导码对盘登记持久保留关键字
502
引导码读MBR以确定引导哪个OS分区
504
引导码进行OS引导分区上的基本排它写持久保留
506
保留成功?
否
508
是
引导失败
510
把控制转交给OS引导分区
512
OS引导分区进行OS分区上的基本排它写持久保留
514
保留成功?
否
518
516
是
引导失败
启动OS分区
520
结束

图5