

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2019 年 5 月 16 日 (16.05.2019)



(10) 国际公布号
WO 2019/091322 A1

(51) 国际专利分类号:
G06F 9/445 (2018.01)

(21) 国际申请号: PCT/CN2018/113335

(22) 国际申请日: 2018 年 11 月 1 日 (01.11.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711091731.1 2017年11月8日 (08.11.2017) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 余海斌(SHE, Haibin); 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 鲁振伟(LU, Zhenwei); 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) Title: VIRTUAL MACHINE SNAPSHOT PROCESSING METHOD, APPARATUS AND SYSTEM

(54) 发明名称: 虚拟机快照处理方法、装置及系统

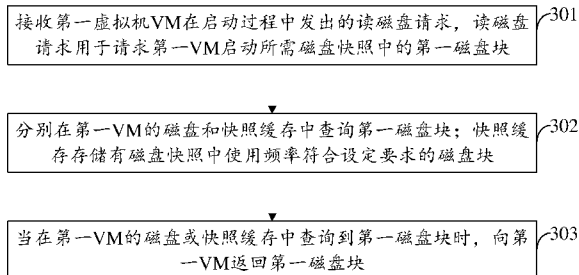


图 3a

- 301 Receive a disk reading request sent by a first virtual machine (VM) during a starting process, the disk reading request being used for requesting a first disk block in a disk snapshot required by the starting of the first VM
- 302 Separately search the disks of the first VM and a snapshot cache for the first disk block, disk blocks, of which the use frequency satisfies set requirements, in the disk snapshot being stored in the snapshot cache
- 303 When the first disk block is found in the disks of the first VM or the snapshot cache, return the first disk block to the first VM

(57) Abstract: Embodiments of the present application provide a virtual machine (VM) snapshot processing method, apparatus and system. In the embodiments of the present application, a snapshot cache is added for a VM to store disk blocks, of which the use frequency satisfies set requirements, in a disk snapshot required during the process of starting the VM, on this basis, when a disk reading request sent by the VM during the starting process is received, the disks of the VM and the snapshot cache are separately searched for a disk block requested by the VM, and when the disk block requested by the VM is found in the disks of the VM or the snapshot

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

cache, said disk block is returned to the VM. Because the disk blocks, of which the use frequency satisfies the set requirements, in the disk snapshot are stored in the snapshot cache, the probability in the disk reading request of hitting the disk block required can be increased, and the probability of requesting a snapshot center for the disk block can be lowered, thereby reducing the overall concurrent pressure of the snapshot center.

(57) 摘要: 本申请实施例提供一种虚拟机快照处理方法、装置及系统。在本申请实施例中, 通过为VM增加快照缓存, 存储VM启动过程中所需磁盘快照中使用频率符合设定要求的磁盘块, 基于此, 当接收到VM在启动过程中发出的读磁盘请求时, 分别在VM的磁盘和快照缓存中查询VM请求的磁盘块, 当在VM的磁盘或快照缓存中查询到VM请求的磁盘块时返回给VM。由于快照缓存中存储的是磁盘快照中使用频率符合设定要求的磁盘块, 可以增大读磁盘请求命中所需磁盘块的概率, 降低向快照中心请求磁盘块的概率, 从而减轻快照中心的整体并发压力。

虚拟机快照处理方法、装置及系统

本申请要求 2017 年 11 月 08 日递交的申请号为 201711091731.1、发明名称为“虚拟机快照处理方法、装置及系统”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及存储技术领域，尤其涉及一种虚拟机快照处理方法、装置及系统。

背景技术

10 虚拟机（Virtual Machine，VM）是指通过软件模拟的具有完整硬件系统功能的、运行在一个完全隔离环境中的完整计算机系统。VM 运行在主机上，具有独立的运行环境，有利于提高主机的安全性，而且使得单个主机上可以运行多个操作系统，因此越来越受欢迎。

基于 VM 的一系列优势，越来越多的互联网应用开始运行在 VM 上。互联网应用的业务流量波动较大，当互联网应用的业务高峰到来时，需要在短时间内启动大量 VM。VM 启动时依赖于某个磁盘快照，磁盘快照包含 VM 运行所需的操作系统和其它数据。磁盘快照一般存储在快照中心中。

对于 VM 启动过程中需要读磁盘块时，首先要在 VM 的磁盘中读取，当未在 VM 的磁盘中读取到时，向快照中心请求。当短时间内启动大量 VM 时，快照中心可能需要并发处理大量请求，快照中心的并发压力较大。

20

发明内容

本申请的多个方面提供一种虚拟机快照处理方法、装置及系统，用以减轻快照中心的并发压力。

25 本申请实施例提供一种虚拟机快照处理方法，包括：

接收第一虚拟机在启动过程中发出的读磁盘请求，所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块；

分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块；所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块；

30 当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时，向所述第

一虚拟机返回所述第一磁盘块。

本申请实施例还提供一种虚拟机快照处理装置，包括：

接收第一虚拟机在启动过程中发出的读磁盘请求，所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块；

5 分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块；所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块；

当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时，向所述第一虚拟机返回所述第一磁盘块。

本申请实施例还提供一种电子设备，包括：存储器和处理器；

10 所述存储器，用于存储程序；

所述处理器，耦合至所述存储器，用于执行所述程序以用于：

接收第一虚拟机在启动过程中发出的读磁盘请求，所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块；

15 分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块；所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块；

当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时，向所述第一虚拟机返回所述第一磁盘块。

本申请还提供一种云计算系统，包括：计算集群、存储集群和快照中心；

20 所述计算集群，用于提供第一虚拟机的计算资源，所述第一虚拟机运行于所述计算集群中；

所述存储集群，用于提供所述第一虚拟机的磁盘以及快照缓存，所述快照缓存存储有所述第一虚拟机启动所需磁盘快照中使用频率符合设定要求的磁盘块；

所述快照中心，用于存储所述磁盘快照；

所述存储集群包括存储管理设备，所述存储管理设备用于：

25 接收第一虚拟机在启动过程中发出的读磁盘请求，所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块；

分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块；所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块；

30 当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时，向所述第一虚拟机返回所述第一磁盘块。

在本申请实施例中，为虚拟机增加快照缓存，存储虚拟机启动过程中所需磁盘快照中使用频率符合设定要求的磁盘块，基于此，当接收到虚拟机在启动过程中发出的读磁盘请求时，分别在虚拟机的磁盘和快照缓存中查询虚拟机请求的磁盘块，当在虚拟机的磁盘或快照缓存中查询到虚拟机请求的磁盘块时返回给虚拟机。由于快照缓存中存储的是磁盘快照中使用频率符合设定要求的磁盘块，可以增大读磁盘请求命中所需磁盘块的概率，降低向快照中心请求磁盘块的概率，从而减轻快照中心的整体并发压力。

附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

图 1 为本申请一示例性实施例提供的示例性存储计算分离的云计算架构的示意图；

图 2a 为本申请另一示例性实施例提供的 IO 线程和 Lazyload 线程在第一 VM 启动过程中处理磁盘块的一过程示意图；

图 2b 为本申请又一示例性实施例提供的 IO 线程和 Lazyload 线程在第一 VM 启动过程中处理磁盘块的另一过程示意图；

图 3a 为本申请又一示例性实施例提供的虚拟机快照处理方法的流程示意图；

图 3b 为本申请又一示例性实施例提供的虚拟机快照处理方法的流程示意图；

图 4 为本申请又一实施例提供的一种虚拟机快照处理装置的结构示意图；

图 5 为本申请又一实施例提供的一种电子设备的结构示意图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合本申请具体实施例及相应的附图对本申请技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

针对现有技术存在的快照中心并发压力较大的问题，本申请实施例提供一种解决方案，主要原理是：为 VM 增加快照缓存，存储 VM 启动过程中所需磁盘快照中使用频率符合设定要求的磁盘块，基于此，当接收到 VM 在启动过程中发出的读磁盘请求时，分别在 VM 的磁盘和快照缓存中查询 VM 请求的磁盘块，当在 VM 的磁盘或快照缓存中查询到 VM 请求的磁盘块时返回给 VM。由于快照缓存中存储的是磁盘快照中使用频率符合设定

要求的磁盘块，可以增大读磁盘请求命中所需磁盘块的概率，降低向快照中心请求磁盘块的概率，从而减轻快照中心的整体并发压力。

以下结合附图，详细说明本申请各实施例提供的技术方案。

图 1 为本申请一示例性实施例提供的存储计算分离的云计算架构的示意图。如图 1 所示，该架构 10 包括：计算集群 101、存储集群（EBS）102 以及快照中心 103。计算集群 101 可以是一个或多个。存储集群 102 也可以是一个或多个。快照中心 103 一般为一个，但不限于一个。

计算集群 101 包括多个计算节点。计算节点是具有一定计算能力的设备，一般包括处理器、内存和以太网控制器等组件。举例说明，计算节点可以是个人计算机、台式计算机、服务器等。计算集群 101 中的计算节点可以由同类型设备实现，也可以由不同类型设备实现。计算集群 101 主要负责向 VM 提供计算资源，VM 可运行在计算集群 101 中的计算节点上。

存储集群 102 包括多台存储设备。存储设备是具有一定存储能力的设备，一般包括处理器、系统总线以及硬盘、内存等至少一块物理存储介质。举例说明，存储设备可以是服务器、台式计算机、个人计算机、手机、平板电脑、数据库等。服务器可以是常规服务器、云服务器、云主机、虚拟中心等。存储集群 102 中的存储设备可以是相同类型，也可以是不同类型。

除存储设备之外，存储集群 102 还包括存储管理设备。存储管理设备主要负责存储集群 102 的存储管理逻辑。在部署实现上，存储管理设备可独立于各存储设备，单独部署，或者，也可以部署于某个或某些存储设备上。存储集群 102 主要负责向 VM 提供存储资源，例如为各 VM 提供磁盘（VM disk），即运行于计算集群 101 中的 VM 所需的磁盘位于存储集群 102 中，实现存储计算分离的架构。

快照中心 103 主要负责存储 VM 启动所需的磁盘快照，该磁盘快照包括 VM 启动过程中所需的操作系统和其它数据，但并不限于此。可选地，快照中心 103 可以采用存储代价较低的对象存储（object storage）方式。

计算集群 101 的计算节点上可以运行多个 VM。在这些 VM 中，有些 VM 使用相同的磁盘快照，有些 VM 使用不同的磁盘快照。快照中心 103 中存储有各 VM 需要使用的磁盘快照。对不同 VM 来说，无论使用的磁盘快照是否相同，其依赖于磁盘快照的启动过程均类似。为便于描述和区分，本实施例以第一 VM 为例进行说明。其中，第一 VM 可以是计算节点上运行的任一 VM。

在用户需要运行第一 VM 的情况下，可在存储集群 102 中为第一 VM 创建磁盘。例如，用户可以通过存储管理设备提供的命令界面，向存储管理设备输入磁盘创建指令。磁盘创建指令会因存储系统的不同而不同。存储管理设备可根据磁盘创建指令，在某个或某些存储设备中为第一 VM 创建一磁盘。其中，在第一 VM 启动过程中，第一 VM 的磁盘主要

5 要用于存储第一 VM 启动过程中所需的磁盘快照。

在本实施例中，磁盘快照包括多个磁盘块（block），每个磁盘块按照顺序被分配一个索引号（Index）。其中，磁盘块是磁盘快照的最小存储单位，也是最小加载单位。其中，不同磁盘块的大小可以相同，也可以不相同。在磁盘块大小相同的情况下，并不限定磁盘块的大小，例如可以一个磁盘块是 200MB、300MB 或 500MB 等。

10 在第一 VM 启动过程中，第一 VM 需要使用某个磁盘块（即需要读磁盘）时，可以向存储管理设备发出读磁盘请求。存储管理设备根据该读磁盘请求，在第一 VM 的磁盘中查询第一 VM 请求的磁盘块。当在第一 VM 的磁盘中查询到第一 VM 请求的磁盘块时，向第一 VM 返回查询到的磁盘块；当未在第一 VM 的磁盘中查询到第一 VM 请求的磁盘块时，向快照中心 103 请求该磁盘块。

15 在短时间内启动大量 VM 的情况下，若在各 VM 的磁盘中均未查询到各 VM 请求的磁盘块，此时存储管理设备需要集中向快照中心 103 为各 VM 请求磁盘块，从而造成快照中心 103 的并发压力较大。

为解决快照中心 103 并发压力较大的问题，存储管理设备在存储集群 102 中选择一块存储空间作为快照缓存（cache）。该快照缓存主要用于存储磁盘快中使用频率符合

20 设定要求的磁盘块。可选地，使用频率符合设定要求的磁盘块可以是使用频率最高的若干个磁盘块，或者可以是使用频率大于设定频率阈值的若干个磁盘块，或者可以是使用频率在某个区间内的若干个磁盘块。其中，磁盘块的使用频率可以是磁盘块在一定时间内的使用频率，例如最近一周内的使用频率，或者最近一个月内的使用频率等。快照缓存中存储的磁盘块在第一 VM 读磁盘过程中被命中的概率较高，例如高于设定概率阈值。

25 基于此，在第一 VM 启动过程中，第一 VM 需要使用某个磁盘块（即需要读磁盘）时，可以向存储管理设备发出读磁盘请求，该请求携带有磁盘块的索引号，指向读磁盘请求所请求的磁盘块。为便于描述和区分，将读磁盘请求所请求的磁盘块称为第一磁盘块，第一磁盘块是第一 VM 启动所需磁盘快照中的某个磁盘块。存储管理设备接收第一 VM 在启动过程中发出的读磁盘请求，可以分别在第一 VM 的磁盘和快照缓存中查询第一磁盘

30 块。当在第一 VM 的磁盘或快照缓存中查询到第一磁盘块时，向第一 VM 返回查询到的第

一磁盘块，而不用向快照中心 103 请求第一磁盘块。当在第一 VM 的磁盘和快照缓存中
5 都未查询到第一磁盘块时，向快照中心 103 请求第一磁盘块，并将快照中心 103 返回的
第一磁盘块存储至第一 VM 的磁盘中，以供第一 VM 使用。

在本实施例中，通过快照缓存中存储磁盘快照中使用频率较高的磁盘块，一定程度
5 上可增加第一 VM 在读磁盘时直接在存储集群 102 中命中所需磁盘块的概率，进而降低
向快照中心 103 请求磁盘块的概率，减轻快照中心 103 的并发压力。

其中，快照缓存和第一 VM 的磁盘都是存储集群 102 中的存储空间上，存储管理设
备查询快照缓存和第一 VM 的磁盘的性能基本相同，因此可以不对查询快照缓存和第一
10 VM 的磁盘的顺序做限定。例如，可以先查询快照缓存，当在快照缓存中未查询到第一磁
盘块时，再查询第一 VM 的磁盘。或者，也可以先查询第一 VM 的磁盘，当在第一 VM 的
磁盘中未查询到第一磁盘块时，再查询快照缓存。

进一步，考虑到第一 VM 的磁盘是第一 VM 专享的，而快照缓存是所有 VM 共享的，
因此可以先查询第一 VM 的磁盘，当在第一 VM 的磁盘中查询到第一磁盘块时，无需查询
快照缓存，这样可以减轻快照缓存的压力。

15 在上述实施例或下述实施例中，存储管理设备可以在第一 VM 启动之前，根据其它
VM 在启动过程中对磁盘快照中各磁盘块的使用情况，获得磁盘块中各磁盘块的使用频
率；将各磁盘块中使用频率符合设定要求的磁盘块存储至快照缓存中。其它 VM 是指与
第一 VM 使用相同磁盘快照且在第一 VM 之前启动的 VM。当然，在第一 VM 启动过程中，
存储管理设备也可以根据第一 VM、除第一 VM 之外当前启动的其它 VM 以及在第一 VM 之
20 前启动的其它 VM 在启动过程中对磁盘快照中各磁盘块的使用情况，获得磁盘块中各磁
盘块的使用频率，根据各磁盘块的使用频率实时更新快照缓存中的磁盘块。快照缓存中
存储的磁盘块是动态变化的。例如，存储管理设备可以统计各 VM 在最近一段时间内对
磁盘快照中各磁盘块的使用情况，将使用频率最高的 N 个磁盘块存储至快照缓存中。N
是预先设置的数值，是一正整数。

25 在一些实施例中，在第一 VM 启动过程中，存储管理设备一方面可以响应第一 VM 的
读磁盘请求向第一 VM 提供磁盘块，另一方面可以以磁盘块为粒度采用懒加载(Lazyload)
向第一 VM 的磁盘中加载磁盘快照。这两方面的操作可以并行执行。

在一实现方式中，存储管理设备可以启动两个线程：Lazyload 线程和 IO 线程。其
中，Lazyload 线程主要负责以 Lazyload 方式按照磁盘块加载顺序向第一 VM 的磁盘中加
30 载磁盘快照中的磁盘块。IO 线程主要负责处理第一 VM 发出的读磁盘请求。

参见图 2a, 当第一 VM 需要读磁盘时, 向 I/O 线程发出读磁盘请求 (如图 2a 中的读磁盘请求①), 该读磁盘请求携带有需要读取的磁盘块的索引号。I/O 线程根据该读磁盘请求, 依次查询第一 VM 的磁盘 (如图 2a 中的读磁盘块②) 和快照缓存 (如图 2a 中的读磁盘块③); 若在第一 VM 的磁盘或快照缓存查询到磁盘块索引号所标识的磁盘块, 则返回给第一 VM (如图 2a 中的返回磁盘块④); 若未在第一 VM 的磁盘和快照缓存中查询到磁盘块索引号所标识的磁盘块, 则指示 Lazyload 线程向快照中心 103 请求磁盘块索引号所标识的磁盘块 (如图 2a 中的读磁盘块⑤)。对 Lazyload 线程来说, 在接收到指示之前, 以 Lazyload 方式按照磁盘块加载顺序向第一 VM 的磁盘中加载磁盘快照中的磁盘块。当接收到指示时, Lazyload 线程可以优先从快照中心 103 读取该磁盘块索引号所标识的磁盘块并存储至第一 VM 的磁盘中 (如图 2a 中的写磁盘块⑥), 以便第一 VM 尽快使用到该磁盘块。在从快照中心 103 中读取到该磁盘块索引号所标识的磁盘块之后, Lazyload 线程可以继续以 Lazyload 方式按照磁盘块加载顺序向第一 VM 的磁盘中加载磁盘快照中的磁盘块。

值得说明的是, 在图 2a 中, 假设快照缓存包括磁盘块 d1、磁盘块 d2、... 磁盘块 dk, 快照中心包括磁盘块 d1、磁盘块 d2、... 磁盘块 dn, 这仅作为示例并不限于此。

由上述可见, 结合快照缓存中存储的磁盘快照中使用频率较高的磁盘块, I/O 线程的命中率将有极大地提高, 可以降低 VM 对 Lazyload 线程的依赖度, 因此可以控制 Lazyload 线程的速度尽可能的慢, 尽量降低 Lazyload 线程对存储集群 102 和快照中心 103 的影响, 甚至做到基本没有影响。

在一示例性实施例中, 在快照缓存的基础上, Lazyload 线程以 Lazyload 方式按照磁盘块加载顺序加载磁盘块的方式包括: 按照磁盘块加载顺序, 在快照缓存中查询需要向第一 VM 的磁盘中加载的磁盘块 (如图 2a 中的读磁盘块⑦)。当未在快照缓存中查询到所述需要向第一 VM 的磁盘中加载的磁盘块时, 向快照中心 103 请求所述需要向第一 VM 的磁盘中加载的磁盘块 (如图 2a 中的读磁盘块⑤), 并将快照中心 103 返回的所述需要向第一 VM 的磁盘中加载的磁盘块存储至第一 VM 的磁盘中 (如图 2a 中的写磁盘块⑥)。当在快照缓存中查询到所述需要向第一 VM 的磁盘中加载的磁盘块时, 将在快照缓存中查询到所述需要向第一 VM 的磁盘中加载的磁盘块存储至第一 VM 的磁盘中 (如图 2a 中的写磁盘块⑥)。由此可见, 当从快照缓存中读取到所述需要向第一 VM 的磁盘中加载的磁盘块时, 无需向快照中心 103 请求, 可以减少向快照中心 103 请求磁盘块的次数, 进一步减轻快照中心 103 的压力。

进一步, Lazyload 线程在按照磁盘块加载顺序向第一 VM 的磁盘加载磁盘块之前, 可以从快照缓存或快照中心 103 中读取磁盘块优先列表, 该磁盘块优先列表中存储有第一 VM 对磁盘快照中各磁盘块的使用顺序; 将第一 VM 对所述各磁盘块的使用顺序作为磁盘块加载顺序。其中, 将第一 VM 使用各磁盘块时的先后顺序作为 Lazyload 线程加载磁盘块的顺序, 有利于使第一 VM 启动过程中优先使用的磁盘块被优先加载至第一 VM 的磁盘中, 有利于提高第一 VM 的读磁盘请求在磁盘中的命中概率, 降低因第一 VM 读磁盘请求而查询快照缓存的次数, 减轻快照缓存的压力, 尤其是在短时间内大量 VM 启动的场景中, 可减轻各 VM 的读磁盘请求对快照缓存造成的并发压力。

在一示例性实施例中, 一种获得磁盘块优先列表的方式, 包括: 预先对第一 VM 在历史启动过程中对磁盘快照中各磁盘块的使用情况进行统计, 获得第一 VM 对各磁盘块的使用顺序; 根据第一 VM 对各磁盘块的使用顺序, 将各磁盘块的索引号存储至磁盘块优先列表中; 并将磁盘块优先列表存储至快照缓存或快照中心 103 中, 以供 Lazyload 线程使用。

进一步可选地, 一种对第一 VM 在历史启动过程中对磁盘快照中各磁盘块的使用情况进行统计的方式包括: 统计第一 VM 在历史启动过程中发出的读磁盘请求以及读磁盘请求的先后顺序; 将读磁盘请求所请求的磁盘块的索引号收集起来, 然后按照读磁盘请求的先后顺序, 将各读磁盘请求所请求的磁盘块的索引号按序存储至磁盘块优先列表中。

在上述实施例中, 主要描述了第一 VM 在启动过程中读磁盘的情况。除此情况之外, 第一 VM 在启动过程中还会写磁盘。在第一 VM 启动过程中, 可能需要对某个磁盘块进行写操作, 例如修改某些数据。这些写操作可能涉及第一 VM 的磁盘中已经存在的磁盘块, 也可能涉及第一 VM 的磁盘中不存在的磁盘块。对于需要对第一 VM 的磁盘中不存在的磁盘块进行写操作的情况, 可以先将需要写操作的磁盘块加载到第一 VM 的磁盘中, 然后在对该磁盘块执行写操作, 但这种效率较低。对此, 存储管理设备为第一 VM 的磁盘创建一稀疏文件 (shadow file), 该稀疏文件主要用来临时存储第一 VM 磁盘中某些磁盘块的增量数据。可选地, 该稀疏文件的大小与第一 VM 的磁盘的大小相同。

基于上述稀疏文件, 当第一 VM 需要写磁盘时, 可以向存储管理设备发出写磁盘请求, 该请求会携带需要执行写操作的磁盘块的索引号, 指向需要执行写操作的磁盘块。为便于描述, 将写磁盘请求涉及的磁盘块称为第二磁盘块。存储管理设备接收写磁盘请

求，根据该写磁盘请求，判断第一 VM 的磁盘中是否存在第二磁盘块。当在第一 VM 的磁盘中包含第二磁盘块时，直接在该第二磁盘块中执行写操作；当第一 VM 的磁盘中未包含第二磁盘块时，则将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件中。

在该实施例中，在需要对第一 VM 的磁盘中不存在的磁盘块进行写操作的情况下，无需等待磁盘块加载完成即可完成写操作，有利于提高写操作的效率。

进一步，还可以采用一个位图文件（bitmap file）来记录稀疏文件上增量数据的位置。位图文件中每一个位（bit）对应稀疏文件中的一个扇区（sector），用于记录该扇区的使用状态。若一扇区中存储有增量数据，则位图文件中相应位置为有效，例如置为 1。基于此，在将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件之后，还可以在稀疏文件对应的位图文件中记录增量数据在稀疏文件中的位置，即将相应扇区对应的位置为有效。

在上述实施例中，分别描述了基于快照缓存的读磁盘过程和基于稀疏文件的写磁盘过程。快照缓存和稀疏文件可以单独使用，也可以结合使用。在下面实施例中，将重点描述存储管理设备结合使用快照缓存和稀疏文件在第一 VM 启动过程中对磁盘块进行处理的过程。

参考图 2b，当第一 VM 在启动过程中发出读磁盘请求时，存储管理设备（例如 IO 线程）可以分别在第一 VM 的磁盘中查询第一 VM 所请求的第一磁盘块（参见图 2b 中的读磁盘块②）。当在第一 VM 的磁盘中查询到第一磁盘块时，进一步在第一 VM 的磁盘对应的稀疏文件中查询第一磁盘块的增量数据（参见图 2b 中的读增量数据⑤）。若未在稀疏文件中查询到第一磁盘块的增量数据，则向第一 VM 返回查询到的第一磁盘块。若在稀疏文件中查询到第一磁盘块的增量数据，将第一磁盘块与第一磁盘块的增量数据进行合并，以得到合并后的磁盘块数据；将合并后的磁盘块数据返回给第一 VM。

当在第一 VM 的磁盘查询到第一磁盘块时，继续在快照缓存中查询第一磁盘块（参见图 2b 中的读磁盘块①）。当在快照缓存中查询到第一磁盘块时，进一步在第一 VM 的磁盘对应的稀疏文件中查询第一磁盘块的增量数据（参见图 2b 中的读增量数据⑤）。若未在稀疏文件中查询到第一磁盘块的增量数据，则向第一 VM 返回查询到的第一磁盘块。若在稀疏文件中查询到第一磁盘块的增量数据，将第一磁盘块与第一磁盘块的增量数据进行合并，以得到合并后的磁盘块数据；将合并后的磁盘块数据返回给第一 VM。

当在快照缓存中未查询到第一磁盘块时，存储管理设备（例如 Lazyload 线程）向快照中心 103 请求第一磁盘块（参见图 2b 中的读磁盘块⑥），并在第一 VM 的磁盘对应的稀疏文件中查询第一磁盘块的增量数据（参见图 2b 中的读增量数据⑤）。若未在稀疏文件中查询到第一磁盘块的增量数据，则将快照中心 103 返回的第一磁盘块存储至第一 VM 的磁盘中，以供第一 VM 使用。若在稀疏文件中查询到第一磁盘块的增量数据，将第一磁盘块与第一磁盘块的增量数据进行合并，以得到合并后的磁盘块数据；将合并后的磁盘块数据存储至第一 VM 的磁盘中。

另外，存储管理设备（例如 Lazyload 线程）还可以以 Lazyload 方式按照磁盘块加载顺序向第一 VM 的磁盘中加载磁盘块。该过程包括：按照磁盘块加载顺序，在快照缓存中查询需要向第一 VM 的磁盘中加载的磁盘块（参见图 2b 中的读磁盘块①）。当未在快照缓存中查询到所述需要向第一 VM 的磁盘中加载的磁盘块时，向快照中心 103 请求所述需要向第一 VM 的磁盘中加载的磁盘块（参见图 2b 中的读磁盘块⑥），并在稀疏文件中查询所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据（参见图 2b 中的读增量数据⑤）。若未在稀疏文件中查询到所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据，则将快照中心 103 返回的所述需要向第一 VM 的磁盘中加载的磁盘块存储至第一 VM 的磁盘中（参见图 2b 中的读增量数据③）。若在稀疏文件中查询到所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据，将快照中心 103 返回的所述需要向第一 VM 的磁盘中加载的磁盘块和在稀疏文件中查询到的所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据进行合并，形成新的磁盘块数据，将新的磁盘块数据存储至第一 VM 的磁盘中（参见图 2b 中的读增量数据③）。

当在快照缓存中查询到所述需要向第一 VM 的磁盘中加载的磁盘块时，在稀疏文件中查询所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据（参见图 2b 中的读增量数据⑤）。若未在稀疏文件中查询到所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据，则将在快照缓存中查询到的所述需要向第一 VM 的磁盘中加载的磁盘块存储至第一 VM 的磁盘中（参见图 2b 中的读增量数据③）。若在稀疏文件中查询到所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据，将在快照缓存中查询到的所述需要向第一 VM 的磁盘中加载的磁盘块和在稀疏文件中查询到的所述需要向第一 VM 的磁盘中加载的磁盘块的增量数据进行合并，形成新的磁盘块数据，将新的磁盘块数据存储至第一 VM 的磁盘中（参见图 2b 中的读增量数据③）。

其中，在结合使用快照缓存和稀疏文件的场景中，第一 VM 写磁盘的过程与前面基

于稀疏文件写磁盘的过程相同（参见图 2b 中的写增量数据④），在此不再赘述。

值得说明的是，在图 2b 中，假设快照缓存包括磁盘块 d1、磁盘块 d2、... 磁盘块 dk，快照中心包括磁盘块 d1、磁盘块 d2、... 磁盘块 dn，VM 磁盘包括磁盘块 d1、磁盘块 d2、... 磁盘块 dj，稀疏文件包括增量数据 d1、增量数据 d2、... 增量数据 dm，这
5 仅作为示例并不限于此。其中，以相同符号表示的增量数据和磁盘块具有所属关系，例如增量数据 d1 是磁盘块 d1 的增量数据，增量数据 d2 是磁盘块 d2 的增量数据等。

由此可见，将快照缓存和稀疏文件相结合，既可以保证读写磁盘的效率，又可以减轻快照中心 103 的并发压力。

图 3a 为本申请一示例性实施例提供的虚拟机快照处理方法的流程示意图。该方法
10 主要由存储管理设备执行，如图 3a 所示，该方法包括：

301、接收第一虚拟机 VM 在启动过程中发出的读磁盘请求，读磁盘请求用于请求第一 VM 启动所需磁盘快照中的第一磁盘块。

302、分别在第一 VM 的磁盘和快照缓存中查询第一磁盘块；快照缓存存储有磁盘快照中使用频率符合设定要求的磁盘块。

15 303、当在第一 VM 的磁盘或快照缓存中查询到第一磁盘块时，向第一 VM 返回第一磁盘块。

在本实施例中，通过一快照缓存（cache）存储磁盘快照中使用频率符合设定要求的磁盘块。可选地，使用频率符合设定要求的磁盘块可以是使用频率最高的若干个磁盘块，或者可以是使用频率大于设定频率阈值的若干个磁盘块，或者可以是使用频率在某个区
20 间内的若干个磁盘块。其中，磁盘块的使用频率可以是磁盘块在一定时间内的使用频率，例如最近一周内的使用频率，或者最近一个月内的使用频率等。快照缓存中存储的磁盘块在第一 VM 读磁盘过程中被命中的概率较高，例如高于设定概率阈值。

基于快照缓存，当第一 VM 在启动过程中发出读磁盘请求时，可以分别在第一 VM 的磁盘和快照缓存中查询第一 VM 所请求的磁盘块。当在第一 VM 的磁盘或快照缓存中查询
25 到第一 VM 所请求的磁盘块时，向第一 VM 返回查询到的磁盘块，而不用向快照中心 103 请求磁盘块。当在第一 VM 的磁盘和快照缓存中都未查询到第一 VM 所请求的磁盘块时，向快照中心 103 请求磁盘块，并将快照中心 103 返回的磁盘块存储至第一 VM 的磁盘中，以供第一 VM 使用。

在本实施例中，通过快照缓存存储磁盘快照中使用频率符合设定要求的磁盘块，一
30 定程度上可增加第一 VM 在快照缓存中命中所需磁盘块的概率，进而降低向快照中心请

求磁盘块的概率，减轻快照中心的并发压力。

在一可选实施方式中，步骤 302 的一种实现方式为：先查询快照缓存，当在快照缓存中未查询到第一磁盘块时，再查询第一 VM 的磁盘。或者，也可以先查询第一 VM 的磁盘，当在第一 VM 的磁盘中未查询到第一磁盘块时，再查询快照缓存。考虑到第一 VM 的
5 磁盘是第一 VM 专享的，而快照缓存是所有 VM 共享的，因此可以先查询第一 VM 的磁盘，当在第一 VM 的磁盘中查询到第一磁盘块时，无需查询快照缓存，这样可以减轻快照缓存的并发压力。

进一步，当未在第一 VM 的磁盘和快照缓存中查询到第一磁盘块时，可以向存储有磁盘快照的快照中心请求第一磁盘块；将快照中心返回的第一磁盘块存储至第一 VM 的
10 磁盘中。

在一可选实施方式中，在第一 VM 启动之前，可以根据其它 VM 在启动过程中对磁盘快照中各磁盘块的使用情况，获得第一磁盘块中各磁盘块的使用频率；将各磁盘块中使用频率符合设定要求的磁盘块存储至快照缓存中。这样，在第一 VM 启动过程中，可以使用快照缓存中的磁盘块。

15 在一可选实施方式中，除了根据第一 VM 的读磁盘请求，为第一 VM 查询第一磁盘块之外，还可以在 First VM 的启动过程中，以 Lazyload 方式，按照磁盘块加载顺序向第一 VM 的磁盘中加载磁盘快照中的磁盘块。

进一步，可以按照磁盘块加载顺序，在快照缓存中查询需要向第一 VM 的磁盘中加载的磁盘块；当未在快照缓存中查询到需要向第一 VM 的磁盘中加载的磁盘块时，向快
20 照中心请求需要向第一 VM 的磁盘中加载的磁盘块，并将快照中心返回的需要向第一 VM 的磁盘中加载的磁盘块存储至第一 VM 的磁盘中。

更进一步，在按照磁盘块加载顺序，在快照缓存中查询需要向第一 VM 的磁盘中加载的磁盘块之前，可以从快照缓存或快照中心中读取磁盘块优先列表，磁盘块优先列表中存储有第一 VM 对磁盘快照中各磁盘块的使用顺序；将第一 VM 对各磁盘块的使用顺序
25 作为磁盘块加载顺序。进而，以第一 VM 对磁盘快照中各磁盘块的使用顺序，向第一 VM 的磁盘中加载的磁盘块。

在一可选实施方式中，在从快照缓存或快照中心中读取磁盘块优先列表之前，可以获得磁盘块优先列表。一种获得盘块优先列表的方式包括：

对第一 VM 在历史启动过程中对各磁盘块的使用情况进行统计，以获得第一 VM 对各
30 磁盘块的使用顺序；

根据第一 VM 对各磁盘块的使用顺序，将各磁盘块的索引号存储至磁盘块优先列表中；

将磁盘块优先列表存储至快照缓存或快照中心中。

在上述方法实施例中，主要描述了第一 VM 在启动过程中读磁盘的情况。除此情况
5 之外，第一 VM 在启动过程中还会写磁盘。在第一 VM 启动过程中，可能需要对第一 VM 的磁盘中不存在的磁盘块进行写操作。对于这种情况，可以先将需要执行写操作的磁盘块加载到第一 VM 的磁盘中，然后在对该磁盘块执行写操作，但这种效率较低。对此，在本实施例中，为第一 VM 的磁盘创建一稀疏文件，该稀疏文件主要用来临时存储第一 VM 磁盘中某些磁盘块的增量数据。

10 基于上述稀疏文件，一种在写磁盘操作过程中进行快照处理的方法流程如图 3b 所示，包括以下步骤：

304、接收第一 VM 在启动过程中发出的写磁盘请求，写磁盘请求用于请求对磁盘快照中的第二磁盘块进行写操作。

305、判断第一 VM 的磁盘中是否包含第二磁盘块；若判断结果为否，则执行步骤 306；
15 若判断结果为是，则执行步骤 307。

306、将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件中。

307、将写磁盘请求的增量数据写入第二磁盘块中。

在本实施例中，当第一 VM 需要写磁盘时，可以向存储管理设备发出写磁盘请求，该请求会携带需要执行写操作的磁盘块的索引号。存储管理设备根据该写磁盘请求，判
20 断第一 VM 的磁盘中是否存在第一 VM 需要写操作的磁盘块。当在第一 VM 的磁盘中包含第一 VM 需要写操作的磁盘块时，直接在该磁盘块中执行写操作；当第一 VM 的磁盘中未包含第一 VM 需要写操作的磁盘块时，则将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件中。在该实施例中，在需要对第一 VM 的磁盘中不存在的磁盘块进行写操作的情况下，无需等待磁盘块加载完成即可完成写操作，有利于提高写操作的效率。

25 进一步，还可以采用一个位图文件（bitmap file）来记录稀疏文件上增量数据的位置。位图文件中每一个位（bit）对应稀疏文件中的一个扇区（sector），用于记录该扇区的使用状态。若一扇区中存储有增量数据，则位图文件中相应位置为有效，例如置为 1。基于此，在将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件中之后，还可以在稀疏文件对应的位图文件中记录增量数据在稀疏文件中的位置，即将相应扇区
30 对应的位置为有效。

值得说明的是，上述快照缓存和稀疏文件可以单独使用，也可以结合使用。结合使用的过程可参见图 2b 所示，在此不再赘述。

需要说明的是，上述实施例所提供方法的各步骤的执行主体均可以是同一设备，或者，该方法也由不同设备作为执行主体。比如，步骤 301 至步骤 303 的执行主体可以为设备 A；又比如，步骤 301 和 302 的执行主体可以为设备 A，步骤 303 的执行主体可以为设备 B；等等。

另外，在上述实施例及附图中的描述的一些流程中，包含了按照特定顺序出现的多个操作，但是应该清楚了解，这些操作可以不按照其在本文中出现的顺序来执行或并行执行，操作的序号如 201、202 等，仅仅是用于区分开各个不同的操作，序号本身不代表任何的执行顺序。另外，这些流程可以包括更多或更少的操作，并且这些操作可以按顺序执行或并行执行。需要说明的是，本文中的“第一”、“第二”等描述，是用于区分不同的消息、设备、模块等，不代表先后顺序，也不限定“第一”和“第二”是不同的类型。

图 4 为本申请又一实施例提供的一种虚拟机快照处理装置的结构示意图。如图 4 所示，虚拟机快照处理装置包括：接收模块 401、查询模块 402 以及提供模块 403。

接收模块 401，用于接收第一 VM 在启动过程中发出的读磁盘请求，该读磁盘请求用于请求第一 VM 启动所需磁盘快照中的第一磁盘块。

查询模块 402，用于分别在第一 VM 的磁盘和快照缓存中查询第一磁盘块；其中快照缓存存储有磁盘快照中使用频率符合设定要求的磁盘块。

提供模块 403，用于当在第一 VM 的磁盘或快照缓存中查询到第一磁盘块时，向第一 VM 返回第一磁盘块。

在一可选实施方式中，提供模块 403 还用于：当未在第一 VM 的磁盘和快照缓存中查询到第一磁盘块时，向存储有磁盘快照的快照中心请求第一磁盘块；将快照中心返回的第一磁盘块存储至第一 VM 的磁盘中。

在一可选实施方式中，虚拟机快照处理装置还包括统计模块，用于在第一 VM 启动之前，根据其它 VM 在启动过程中对磁盘快照中各磁盘块的使用情况，获得第一磁盘块中各磁盘块的使用频率；将各磁盘块中使用频率符合设定要求的磁盘块存储至快照缓存中。

在一可选实施方式中，查询模块 402 还用于在第一 VM 的启动过程中，以 Lazyload 方式，按照磁盘块加载顺序向第一 VM 的磁盘中加载磁盘快照中的磁盘块。

进一步，查询模块 402 在加载磁盘快照的磁盘块时，具体用于：

按照磁盘块加载顺序，在快照缓存中查询需要向第一 VM 的磁盘加载的磁盘块；

当未在快照缓存中查询到需要向第一 VM 的磁盘加载的磁盘块时，向快照中心请求需要向第一 VM 的磁盘加载的磁盘块，并将快照中心返回的需要向第一 VM 的磁盘加载的磁盘块存储至第一 VM 的磁盘中。

进一步，查询模块 402 在按照磁盘块加载顺序，在快照缓存中查询需要向第一 VM 的磁盘加载的磁盘块之前，还用于：

从快照缓存或快照中心读取磁盘块优先列表，该磁盘块优先列表中存储有第一 VM 对磁盘快照中各磁盘块的使用顺序；

10 将第一 VM 对各磁盘块的使用顺序作为磁盘块加载顺序。

更进一步，统计模块在查询模块 402 从快照缓存或快照中心读取磁盘块优先列表之前，还用于：

对第一 VM 在历史启动过程中对各磁盘块的使用情况进行统计，以获得第一 VM 对各磁盘块的使用顺序；

15 根据第一 VM 对各磁盘块的使用顺序，将各磁盘块的索引号存储至磁盘块优先列表中；

将磁盘块优先列表存储至快照缓存或快照中心中。

在一可选实施方式中，接收模块 401 还用于：

接收第一 VM 在启动过程中发出的写磁盘请求，且该写磁盘请求用于请求对磁盘快照中的第二磁盘块进行写操作；

当第一 VM 的磁盘中未包含所述第二磁盘块时，将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件中。

进一步，接收模块 401 还用于：在稀疏文件对应的位图文件中记录增量数据在稀疏文件中的位置。

25 相应地，查询模块 402 具体用于：在稀疏文件中查询第一磁盘块的增量数据；当在稀疏文件中查询到第一磁盘块的增量数据时，将第一磁盘块与第一磁盘块的增量数据进行合并，以得到合并后的磁盘块数据。

相应地，基于上述合并后的磁盘块数据，提供模块 403 在向第一 VM 返回第一磁盘块时具体用于：将合并后的磁盘块数据返回给第一 VM。

30 本实施例提供的虚拟机快照处理装置，可用于执行上述快照方法实施例中的流程，

其工作原理不再赘述，详见方法实施例的描述。

本申请实施例提供的虚拟机快照处理装置，为 VM 增加快照缓存，存储 VM 启动过程中所需磁盘快照中使用频率符合设定要求的磁盘块，基于此，当虚拟机快照处理装置收到 VM 在启动过程中发出的读磁盘请求时，分别在 VM 的磁盘和快照缓存中查询 VM 请求的磁盘块，当在 VM 的磁盘或快照缓存中查询到 VM 请求的磁盘块时返回给 VM。由于快照缓存中存储的是磁盘快照中使用频率符合设定要求的磁盘块，该虚拟机快照处理装置可以增大读磁盘请求命中所需磁盘块的概率，降低向快照中心请求磁盘块的概率，从而减轻快照中心的整体并发压力。

以上描述了虚拟机快照处理装置的内部功能和结构，如图 5 所示，实际中，该快照处理装置可实现为电子设备，包括：存储器 500、处理器 501 以及通信组件 502；

通信组件 502，用于接收第一 VM 在启动过程中发出的读磁盘请求，该读磁盘请求用于请求第一 VM 启动所需磁盘快照中的第一磁盘块；

存储器 500，用于存储程序；

处理器 501，耦合至存储器 500，用于执行程序以用于：

分别在第一 VM 的磁盘和快照缓存中查询第一磁盘块；其中快照缓存存储有磁盘快照中使用频率符合设定要求的磁盘块；

当在第一 VM 的磁盘或快照缓存中查询到所述第一磁盘块时，通过通信组件 502 向第一 VM 返回第一磁盘块；

通信组件 502，还用于向第一 VM 返回第一磁盘块。

在一可选实施方式中，处理器 501 还用于：

当未在第一 VM 的磁盘和快照缓存中查询到第一磁盘块时，通过通信组件 502 向存储有磁盘快照的快照中心请求第一磁盘块，并通过通信组件 502 将快照中心返回的第一磁盘块存储至第一 VM 的磁盘中。相应地，通信组件 502 还用于向存储有磁盘快照的快照中心请求第一磁盘块，并将快照中心返回的第一磁盘块存储至第一 VM 的磁盘中。

在一可选实施方式中，在第一 VM 启动之前，处理器 501 还用于：

根据其它 VM 在启动过程中对磁盘快照中各磁盘块的使用情况，获得第一磁盘块中各磁盘块的使用频率；

将各磁盘块中使用频率符合设定要求的磁盘块存储至快照缓存中。

在一可选实施方式中，处理器 501 在第一 VM 的启动过程中，还用于：以 Lazyload 方式，按照磁盘块加载顺序向第一 VM 的磁盘中加载磁盘快照中的磁盘块。

进一步，处理器 501 在加载磁盘快照的磁盘块时，具体用于：

按照磁盘块加载顺序，在快照缓存中查询需要向第一 VM 的磁盘加载的磁盘块；

当未在快照缓存中查询到需要向第一 VM 的磁盘加载的磁盘块时，通过通信组件 502 向快照中心请求需要向第一 VM 的磁盘加载的磁盘块，并通过通信组件 502 将快照中心返回的需要向第一 VM 的磁盘加载的磁盘块存储至第一 VM 的磁盘。相应地，通信组件 502 还用于向快照中心请求需要向第一 VM 的磁盘加载的磁盘块，并将快照中心返回的需要向第一 VM 的磁盘加载的磁盘块存储至第一 VM 的磁盘。

进一步，处理器 501 在按照磁盘块加载顺序，在快照缓存中查询需要向第一 VM 的磁盘加载的磁盘块之前，还用于：

10 从快照缓存或快照中心读取磁盘块优先列表，该磁盘块优先列表中存储有第一 VM 对磁盘快照中各磁盘块的使用顺序；

将第一 VM 对各磁盘块的使用顺序作为磁盘块加载顺序。

更进一步，处理器 501 在从快照缓存或快照中心读取磁盘块优先列表之前，还用于：

15 对第一 VM 在历史启动过程中对各磁盘块的使用情况进行统计，以获得第一 VM 对各磁盘块的使用顺序；

根据第一 VM 对各磁盘块的使用顺序，将各磁盘块的索引号存储至磁盘块优先列表中；

将磁盘块优先列表存储至快照缓存或快照中心。

20 在一可选实施方式中，通信组件 502 还用于接收第一 VM 在启动过程中发出的写磁盘请求，且该写磁盘请求用于请求对磁盘快照中的第二磁盘块进行写操作。相应地，基于通信组件 502 接收的写磁盘请求，处理器 501 还用于：当第一 VM 的磁盘未包含所述第二磁盘块时，将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件中。

25 进一步，处理器 501 在将写磁盘请求的增量数据写入第一 VM 的磁盘对应的稀疏文件中之后，还用于：在稀疏文件对应的位图文件中记录增量数据在稀疏文件中的位置。

相应地，处理器 501 还用于：在稀疏文件中查询第一磁盘块的增量数据；当在稀疏文件中查询到第一磁盘块的增量数据时，将第一磁盘块与第一磁盘块的增量数据进行合并，以得到合并后的磁盘块数据；并通过通信组件 502 将合并后的磁盘块数据返回给第一 VM。相应地，通信组件 502 具体用于将合并后的磁盘块数据返回给第一 VM。

30 进一步，如图 5 所示，电子设备还包括：显示器 503、电源组件 504 以及音频组件

505 等其它组件。图 5 中仅示意性给出部分组件，并不意味着电子设备只包括图 5 所示组件。

在图 5 中的通信组件，可被配置为便于通信组件所属设备和其他设备之间有线或无线方式的通信。通信组件所属设备可以接入基于通信标准的无线网络，如 WiFi，2G 或 3G，或它们的组合。在一个示例性实施例中，通信组件经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，所述通信组件还包括近场通信（NFC）模块，以促进短程通信。例如，在 NFC 模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

在图 5 中的显示器，可以包括屏幕，其屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界，而且还检测与所述触摸或滑动操作相关的持续时间和压力。

在图 5 中的电源组件，为电源组件所属设备的各种组件提供电力。电源组件可以包括电源管理系统，一个或多个电源，及其他与为电源组件所属设备生成、管理和分配电力相关联的组件。

在图 5 中的音频组件，被配置为输出和/或输入音频信号。例如，音频组件包括一个麦克风（MIC），当音频组件所属设备处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器或经由通信组件发送。在一些实施例中，音频组件还包括一个扬声器，用于输出音频信号。

本申请实施例提供的电子设备，用于对虚拟机快照进行处理。为 VM 增加快照缓存，存储 VM 启动过程中所需磁盘快照中使用频率符合设定要求的磁盘块，基于此，当电子设备接收到 VM 在启动过程中发出的读磁盘请求时，分别在 VM 的磁盘和快照缓存中查询 VM 请求的磁盘块，当在 VM 的磁盘或快照缓存中查询到 VM 请求的磁盘块时返回给 VM。由于快照缓存中存储的是磁盘快照中使用频率符合设定要求的磁盘块，可以增大读磁盘请求命中所需磁盘块的概率，降低向快照中心请求磁盘块的概率，从而减轻快照中心的整体并发压力。

相应地，本申请实施例还提供一种存储有计算机程序的计算机可读存储介质，所述

计算机程序被执行时能够实现：

接收第一 VM 在启动过程中发出的读磁盘请求，该读磁盘请求用于请求第一 VM 启动所需磁盘快照中的第一磁盘块；

5 分别在第一 VM 的磁盘和快照缓存中查询第一磁盘块；其中快照缓存存储有磁盘快照中使用频率符合设定要求的磁盘块；

当在第一 VM 的磁盘或快照缓存中查询到第一磁盘块时，向第一 VM 返回第一磁盘块。

上述计算机程序被执行时，除了可以实现上述步骤之外，还可以实现其前述方法实施例中的其它步骤，具体不再赘述。

本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序
10 产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图
15 图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中
20 指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

25 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

在一个典型的配置中，计算设备包括一个或多个处理器（CPU）、输入/输出接口、
30 网络接口和内存。

内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器（RAM）和/或非易失性内存等形式，如只读存储器（ROM）或闪存（flash RAM）。内存是计算机可读介质的示例。

计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存（PRAM）、静态随机存取存储器（SRAM）、动态随机存取存储器（DRAM）、其他类型的随机存取存储器（RAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、快闪记忆体或其他内存技术、只读光盘只读存储器（CD-ROM）、数字多功能光盘（DVD）或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体（transitory media），如调制的数据信号和载波。

还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

以上所述仅为本申请的实施例而已，并不用于限制本申请。对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的权利要求范围之内。

权利要求书

1. 一种虚拟机快照处理方法，其特征在于，包括：

接收第一虚拟机在启动过程中发出的读磁盘请求，所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块；

5 分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块；所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块；

当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时，向所述第一虚拟机返回所述第一磁盘块。

2. 根据权利要求1所述的方法，其特征在于，还包括：

10 当未在所述第一虚拟机的磁盘和所述快照缓存中查询到所述第一磁盘块时，向存储有所述磁盘快照的快照中心请求所述第一磁盘块；

将所述快照中心返回的所述第一磁盘块存储至所述第一虚拟机的磁盘中。

3. 根据权利要求1所述的方法，其特征在于，还包括：

15 在所述第一虚拟机启动之前，根据其它虚拟机在启动过程中对所述磁盘快照中各磁盘块的使用情况，获得所述第一磁盘块中各磁盘块的使用频率；

将所述各磁盘块中使用频率符合设定要求的磁盘块存储至所述快照缓存中。

4. 根据权利要求1所述的方法，其特征在于，还包括：

在所述第一虚拟机的启动过程中，以懒加载方式，按照磁盘块加载顺序向所述第一虚拟机的磁盘中加载所述磁盘快照中的磁盘块。

20 5. 根据权利要求4所述的方法，其特征在于，所述按照磁盘块加载顺序向所述第一虚拟机的磁盘中加载所述磁盘快照中的磁盘块，包括：

按照所述磁盘块加载顺序，在所述快照缓存中查询需要向所述第一虚拟机的磁盘中加载的磁盘块；

25 当未在所述快照缓存中查询到需要向所述第一虚拟机的磁盘中加载的磁盘块时，向所述快照中心请求需要向所述第一虚拟机的磁盘中加载的磁盘块，并将所述快照中心返回的需要向所述第一虚拟机的磁盘中加载的磁盘块存储至所述第一虚拟机的磁盘中。

6. 根据权利要求5所述的方法，其特征在于，按照所述磁盘块加载顺序，在所述快照缓存中查询需要向所述第一虚拟机的磁盘中加载的磁盘块之前，所述方法还包括：

30 从所述快照缓存或所述快照中心中读取磁盘块优先列表，所述磁盘块优先列表中存储有所述第一虚拟机对所述磁盘快照中各磁盘块的使用顺序；

将所述第一虚拟机对所述各磁盘块的使用顺序作为所述磁盘块加载顺序。

7. 根据权利要求 6 所述的方法, 其特征在于, 在从所述快照缓存或所述快照中心中读取磁盘块优先列表之前, 所述方法包括:

对所述第一虚拟机在历史启动过程中对所述各磁盘块的使用情况进行统计, 以获得

5 所述第一虚拟机对所述各磁盘块的使用顺序;

根据所述第一虚拟机对所述各磁盘块的使用顺序, 将所述各磁盘块的索引号存储至所述磁盘块优先列表中;

将所述磁盘块优先列表存储至所述快照缓存或所述快照中心中。

8. 根据权利要求 1-7 任一项所述的方法, 其特征在于, 还包括:

10 接收所述第一虚拟机在启动过程中发出的写磁盘请求, 所述写磁盘请求用于请求对所述磁盘快照中的第二磁盘块进行写操作;

当所述第一虚拟机的磁盘中未包含所述第二磁盘块时, 将所述写磁盘请求的增量数据写入所述第一虚拟机的磁盘对应的稀疏文件中。

9. 根据权利要求 8 所述的方法, 其特征在于, 在将所述写磁盘请求的增量数据写入所述第一虚拟机的磁盘对应的稀疏文件中之后, 所述方法还包括:

在所述稀疏文件对应的位图文件中记录所述增量数据在所述稀疏文件中的位置。

10. 根据权利要求 8 所述的方法, 其特征在于, 所述向所述第一虚拟机返回所述第一磁盘块, 包括:

在所述稀疏文件中查询所述第一磁盘块的增量数据;

20 当在所述稀疏文件中查询到所述第一磁盘块的增量数据时, 将所述第一磁盘块与所述第一磁盘块的增量数据进行合并, 以得到合并后的磁盘块数据;

将所述合并后的磁盘块数据返回给所述第一虚拟机。

11. 一种虚拟机快照处理装置, 其特征在于, 包括:

25 接收第一虚拟机在启动过程中发出的读磁盘请求, 所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块;

分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块; 所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块;

当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时, 向所述第一虚拟机返回所述第一磁盘块。

30 12. 一种电子设备, 其特征在于, 包括: 存储器、处理器以及通信组件;

所述通信组件，用于接收第一虚拟机在启动过程中发出的读磁盘请求，所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块；

所述存储器，用于存储程序；

所述处理器，耦合至所述存储器，用于执行所述程序以用于：

- 5 分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块；所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块；

当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时，通过所述通信组件向所述第一虚拟机返回所述第一磁盘块；

所述通信组件，还用于向所述第一虚拟机返回所述第一磁盘块。

- 10 13. 一种云计算系统，其特征在于，包括：计算集群、存储集群和快照中心；

所述计算集群，用于提供第一虚拟机的计算资源，所述第一虚拟机运行于所述计算集群中；

所述存储集群，用于提供所述第一虚拟机的磁盘以及快照缓存，所述快照缓存存储有所述第一虚拟机启动所需磁盘快照中使用频率符合设定要求的磁盘块；

- 15 所述快照中心，用于存储所述磁盘快照；

所述存储集群包括存储管理设备，所述存储管理设备用于：

接收第一虚拟机在启动过程中发出的读磁盘请求，所述读磁盘请求用于请求所述第一虚拟机启动所需磁盘快照中的第一磁盘块；

- 20 分别在所述第一虚拟机的磁盘和快照缓存中查询所述第一磁盘块；所述快照缓存存储有所述磁盘快照中使用频率符合设定要求的磁盘块；

当在所述第一虚拟机的磁盘或所述快照缓存中查询到所述第一磁盘块时，向所述第一虚拟机返回所述第一磁盘块。

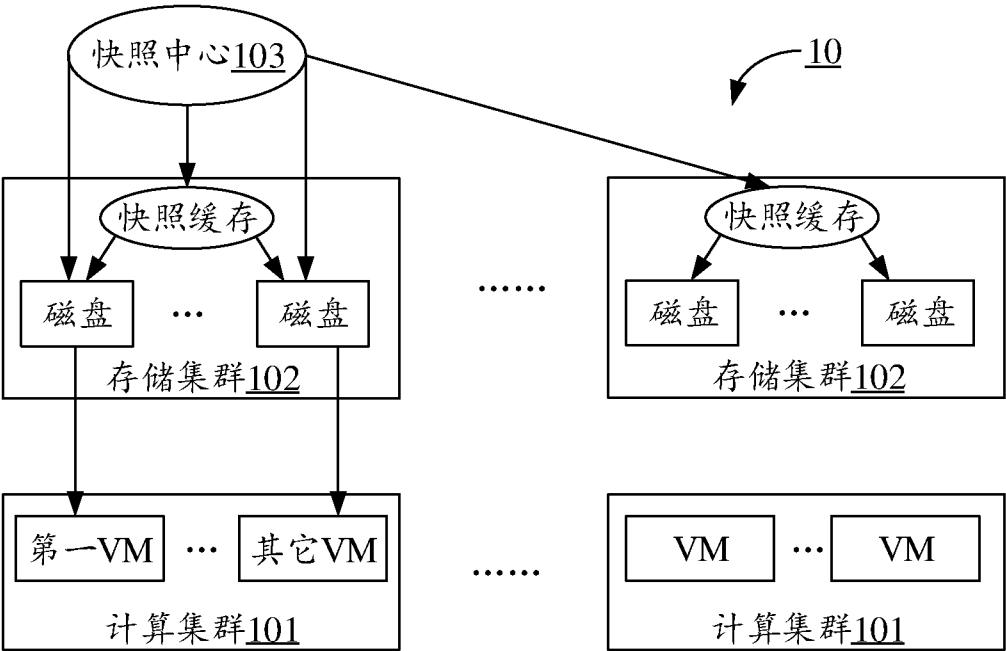


图 1

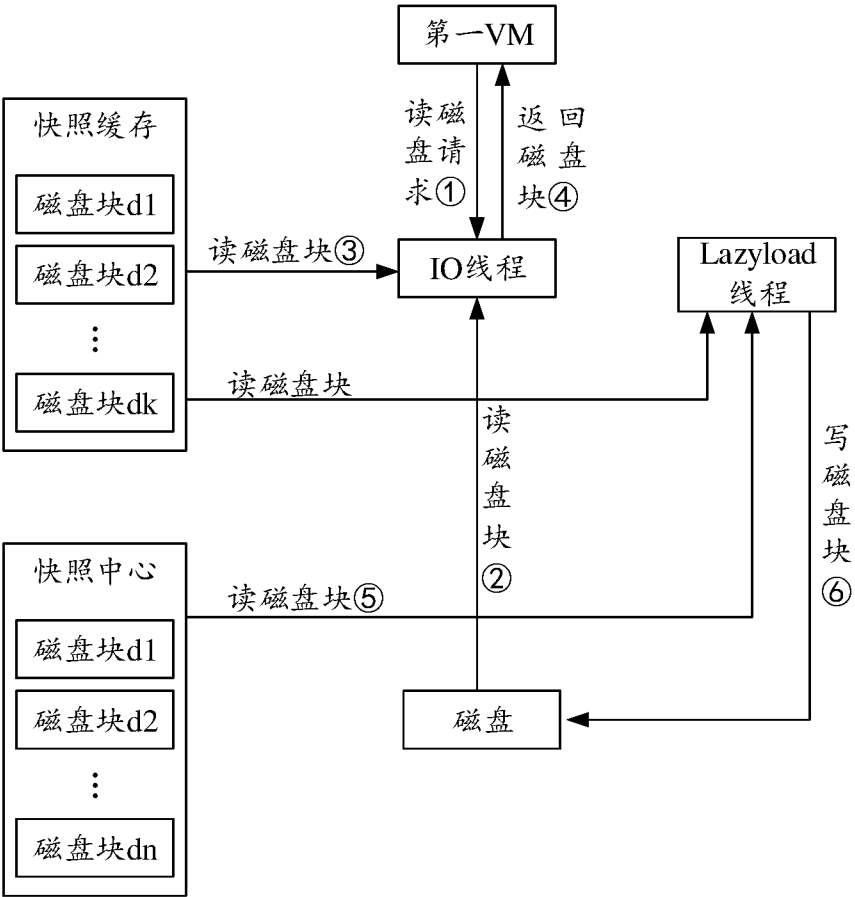


图 2a

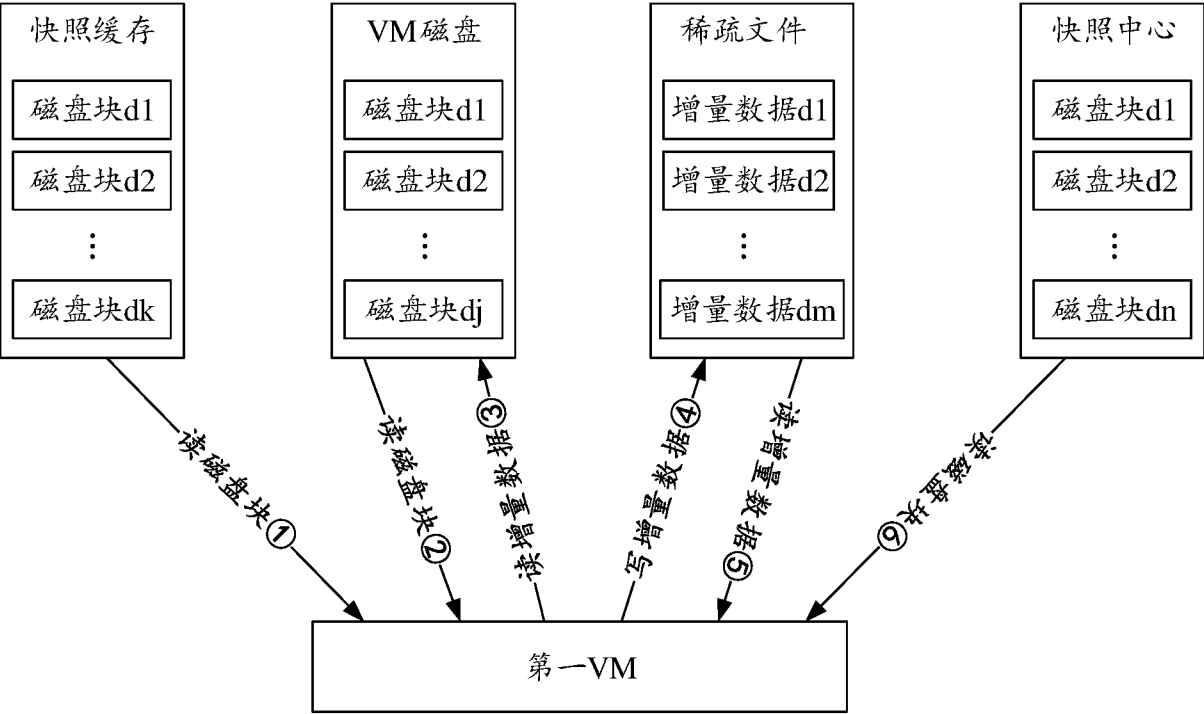


图 2b

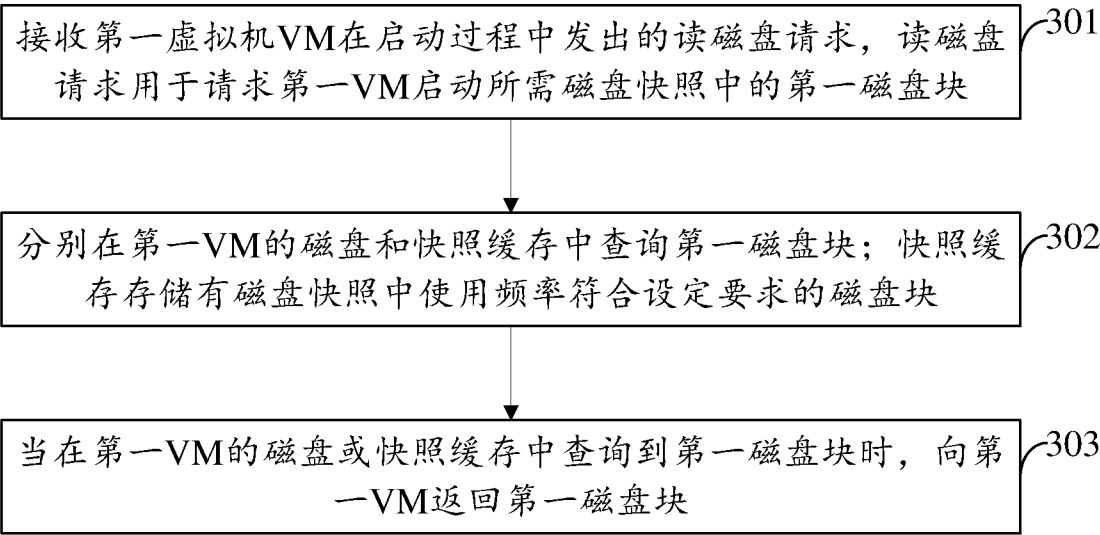


图 3a

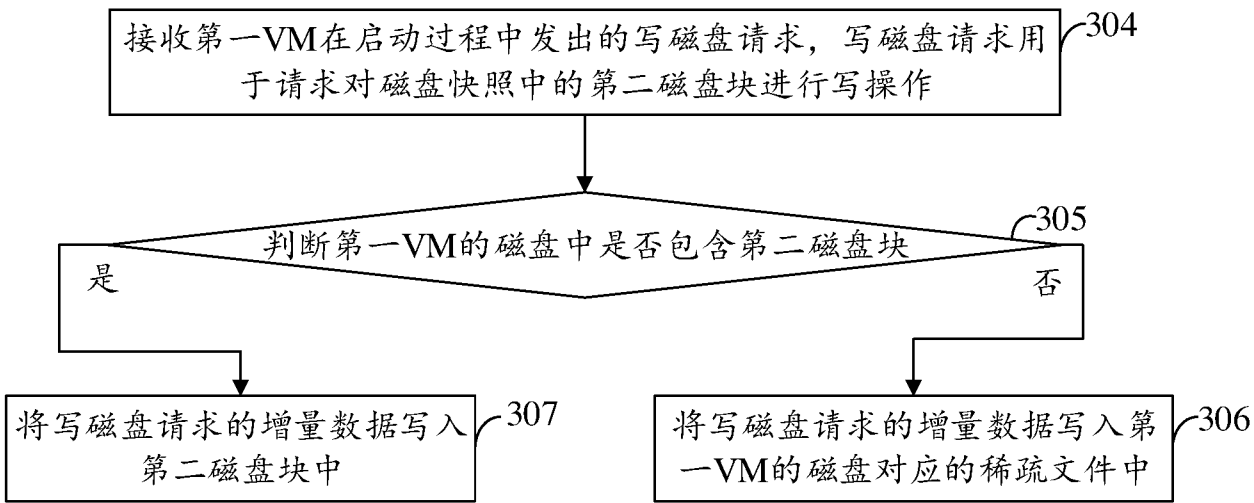


图 3b

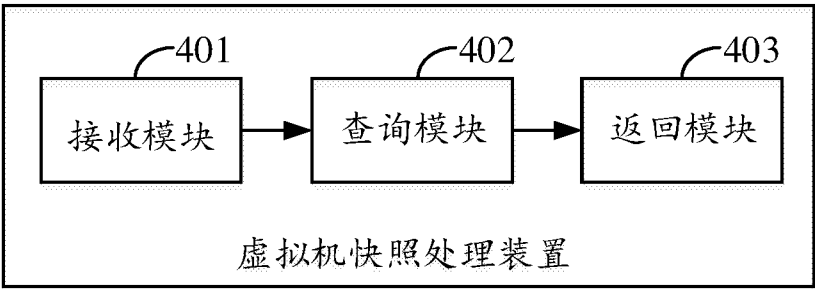


图 4

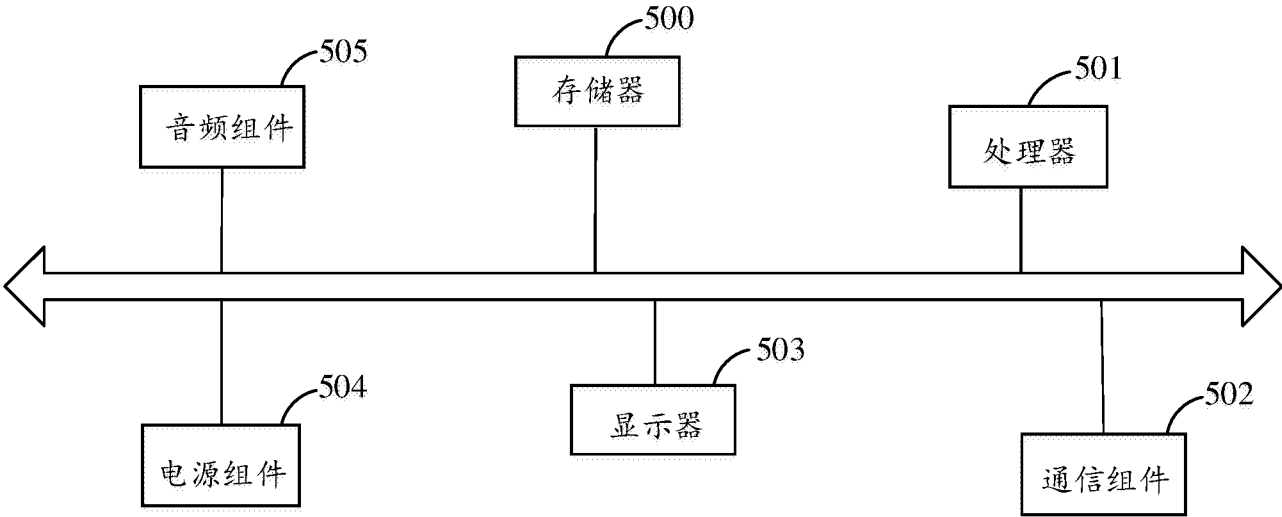


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/113335

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/445(2018.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, VEN, USTXT, EPTXT, WOTXT, CNKI; 虚拟机, 快照, 缓存, 磁盘, 频率, 频次, 次数, 懒加载, 顺序, 次序, 稀疏, 合并, 增量, 读, 写, VM, virtual machine, snapshot, cache, disc, disk, frequency, times, lazyload, order, sparse, merge, increment, read, write

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2013198459 A1 (FUSION-IO, INC.) 01 August 2013 (2013-08-01) description, paragraphs 0005-0007 and 0048-0108, and figures 1A and 4	1-9, 11-13
X	US 2012210068 A1 (FUSION-IO, INC.) 16 August 2012 (2012-08-16) description, paragraphs 0044-0058, and figures 2 and 5	1-9, 11-13
A	CN 106406981 A (SANGFOR INC.) 15 February 2017 (2017-02-15) entire document	1-13



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

24 January 2019

Date of mailing of the international search report

31 January 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/113335

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2013198459	A1	01 August 2013	US	9116812	B2	25 August 2015
US	2012210068	A1	16 August 2012	US	8996807	B2	31 March 2015
CN	106406981	A	15 February 2017	None			

国际检索报告

国际申请号

PCT/CN2018/113335

A. 主题的分类 G06F 9/445 (2018.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类														
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNABS, CNTXT, VEN, USTXT, EPTXT, WOTXT, CNKI; 虚拟机, 快照, 缓存, 磁盘, 频率, 频次, 次数, 懒加载, 顺序, 次序, 稀疏, 合并, 增量, 读, 写, VM, virtual machine, snapshot, cache, disc, disk, frequency, times, lazyload, order, sparse, merge, increment, read, write														
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>US 2013198459 A1 (FUSION IO INC) 2013年 8月 1日 (2013 - 08 - 01) 说明书第0005-0007, 0048-0108段, 图1A, 图4</td> <td>1-9、11-13</td> </tr> <tr> <td>X</td> <td>US 2012210068 A1 (FUSION IO INC) 2012年 8月 16日 (2012 - 08 - 16) 说明书第0044-0058段, 图2, 图5</td> <td>1-9、11-13</td> </tr> <tr> <td>A</td> <td>CN 106406981 A (深圳市深信服电子科技有限公司) 2017年 2月 15日 (2017 - 02 - 15) 全文</td> <td>1-13</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	US 2013198459 A1 (FUSION IO INC) 2013年 8月 1日 (2013 - 08 - 01) 说明书第0005-0007, 0048-0108段, 图1A, 图4	1-9、11-13	X	US 2012210068 A1 (FUSION IO INC) 2012年 8月 16日 (2012 - 08 - 16) 说明书第0044-0058段, 图2, 图5	1-9、11-13	A	CN 106406981 A (深圳市深信服电子科技有限公司) 2017年 2月 15日 (2017 - 02 - 15) 全文	1-13
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求												
X	US 2013198459 A1 (FUSION IO INC) 2013年 8月 1日 (2013 - 08 - 01) 说明书第0005-0007, 0048-0108段, 图1A, 图4	1-9、11-13												
X	US 2012210068 A1 (FUSION IO INC) 2012年 8月 16日 (2012 - 08 - 16) 说明书第0044-0058段, 图2, 图5	1-9、11-13												
A	CN 106406981 A (深圳市深信服电子科技有限公司) 2017年 2月 15日 (2017 - 02 - 15) 全文	1-13												
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。														
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件														
国际检索实际完成的日期 2019年 1月 24日		国际检索报告邮寄日期 2019年 1月 31日												
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10) 62019451		受权官员 马婷婷 电话号码 86-(20)-28958507												

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/113335

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
US	2013198459	A1	2013年 8月 1日	US	9116812	B2	2015年 8月 25日
US	2012210068	A1	2012年 8月 16日	US	8996807	B2	2015年 3月 31日
CN	106406981	A	2017年 2月 15日	无			