



(12) 发明专利申请

(10) 申请公布号 CN 114222975 A

(43) 申请公布日 2022.03.22

(21) 申请号 202080056937.4

(74) 专利代理机构 北京市金杜律师事务所  
11256

(22) 申请日 2020.06.15

代理人 鄧迅

(30) 优先权数据

16/546,337 2019.08.21 US

(51) Int.Cl.

G06F 9/455 (2006.01)

(85) PCT国际申请进入国家阶段日

2022.02.11

G06F 9/50 (2006.01)

G06F 12/02 (2006.01)

(86) PCT国际申请的申请数据

PCT/US2020/037681 2020.06.15

G06F 12/0868 (2016.01)

G06F 3/06 (2006.01)

(87) PCT国际申请的公布数据

WO2021/034391 EN 2021.02.25

(71) 申请人 微软技术许可有限责任公司

地址 美国华盛顿州

(72) 发明人 M·布鲁苏 T·L·阮

N·拉德卡尼

R·米索雷·尚塔穆蒂

权利要求书3页 说明书25页 附图6页

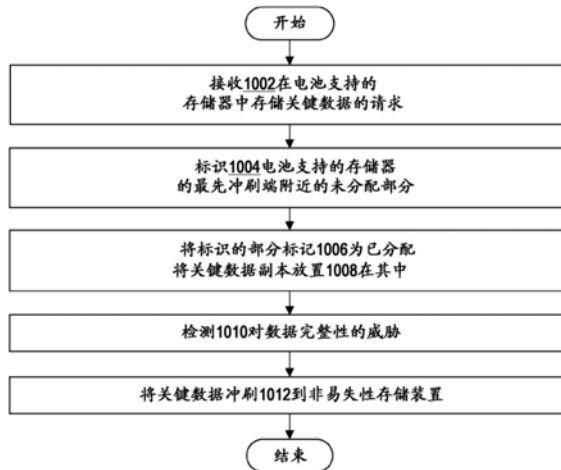
(54) 发明名称

使用存储器孔径冲刷顺序的数据保存

(57) 摘要

组合的操作性步骤和设备特性有助于保存数据免受完整性威胁。基于诸如客户要求、工作负载关键性或虚拟机关键性的标准,数据被分为关键数据和非关键数据。例如,数据可以在计算节点中被生成以存储在存储节点中。关键数据被存储在电池支持的存储器孔径中的物理地址处,其中由于由电池支持的存储器施加或在其上施加的冲刷顺序(例如,自下而上的NVDIMM冲刷顺序),关键数据将在非关键数据之前被冲刷。数据(特别是非关键数据)的冗余副本还可以被保留,以防其未被及时冲刷。根据其电池的特性,电池支持的存储器孔径被调整大小和定位,并且可以随着条件的改变而重新定位或调整大小。执行冲刷碎片整理是为了优化孔径的使用,特别是在保持关键数据的部分内。

示例数据保存方法 1000



1. 一种计算系统中的数据保存子系统,所述数据保存子系统包括:

具有孔径的电池支持的存储器,所述孔径具有冲刷顺序,所述冲刷顺序是响应于数据完整性威胁而将数据从所述孔径复制到与孔径相关联的非易失性存储装置的顺序,所述冲刷顺序定义所述孔径的最先冲刷端和所述孔径的最后冲刷端;

数据保存电路,所述数据保存电路与所述电池支持的存储器可操作地通信,所述数据保存电路被配置为执行数据保存步骤,所述数据保存步骤包括 (a) 接收将数据集A-数据存储在所述孔径中的请求,A-数据包括被指定为关键数据的数据,(b) 标识所述孔径的未分配存储器的一部分A-存储器,A-存储器足够大以保持A-数据,A-存储器的地址比所述孔径的任何其他未分配存储器的任何其他地址更接近所述孔径最先冲刷端,(c) 将A-存储器标记为已分配,并且将A-数据的副本放置在A-存储器中,(d) 接收将数据集Z-数据存储在所述孔径中的请求,Z-数据不包括被指定为关键数据的任何数据,(e) 标识所述孔径的未分配存储器的一部分Z-存储器,Z-存储器足够大以保持Z-数据,Z-存储器的地址比所述孔径的任何其他未分配存储器的任何其他地址更接近所述孔径最后冲刷端,(f) 将Z-存储器标记为已分配,并且将Z-数据的副本放置在Z-存储器中;

其中与通过在所述电池支持的存储器中存储数据而不考虑数据关键性相比,所述数据保存子系统提供成功冲刷关键数据并且从而保存它的更高的可能性。

2. 根据权利要求1所述的数据保存子系统,其中所述电池支持的存储器孔径驻留在网络节点X中或由网络节点X控制,并且将A-数据存储在所述孔径中的所述请求从不同的网络节点Y被发送。

3. 根据权利要求2所述的数据保存子系统,其中所述网络节点X包括云中的存储节点,并且所述网络节点Y包括所述云中的计算节点。

4. 根据权利要求2所述的数据保存子系统,其中所述网络节点X和所述网络节点Y通过存储关系R关联,其中R被定义为使得两个节点M和N在M代表N存储数据时通过R关联,并且其中R不是网络节点X和网络节点Y的一对一关系。

5. 根据权利要求1所述的数据保存子系统,其中所述电池支持的存储器包括NVDIMM存储器。

6. 根据权利要求1所述的数据保存子系统,其中所述计算系统包括虚拟机,所述数据集A-数据从虚拟机VM-A被发送到所述数据保存子系统,并且所述数据集Z-数据从虚拟机VM-Z被发送到所述数据保存子系统。

7. 一种用于在计算系统中的数据保存的方法,所述方法包括:

接收将数据集A-数据存储在电池支持的存储器的孔径中的请求,所述孔径具有冲刷顺序,所述冲刷顺序是响应于数据完整性威胁而将数据从所述孔径复制到与孔径相关联的非易失性存储装置的顺序,所述冲刷顺序定义所述孔径的最先冲刷端和所述孔径的最后冲刷端,A-数据包括被指定为关键数据的数据;

标识所述孔径的未分配存储器的一部分A-存储器,A-存储器足够大以保持A-数据,A-存储器的地址比所述孔径的任何其他未分配存储器的任何其他地址更接近所述孔径最先冲刷端;

将A-存储器标记为已分配,并且将A-数据的副本放置在A-存储器中;

接收将数据集Z-数据存储在所述孔径中的请求,Z-数据不包括被指定为关键数据的任

何数据；

标识所述孔径的未分配存储器的一部分Z-存储器，Z-存储器足够大以保持Z-数据，Z-存储器的地址比所述孔径的任何其他未分配存储器的任何其他地址更接近所述孔径最后冲刷端；以及

将Z-存储器标记为已分配，并且将Z-数据的副本放置在Z-存储器中；

从而通过在所述电池支持的存储器中，以所述冲刷顺序将所述关键数据放在在非关键数据之前，成功冲刷所述关键数据并且从而保存它的可能性被增加。

8. 根据权利要求7所述的方法，还包括以下各项中的至少一项：

至少部分地基于所述电池支持的存储器的电池特性来确定所述孔径的大小；

至少部分地基于所述电池支持的存储器的电池特性来确定所述孔径的位置；

至少部分地基于所述电池支持的存储器的电池特性的改变来调整所述孔径的大小；或者

至少部分地基于所述电池支持的存储器的电池特性的改变来移动所述孔径。

9. 根据权利要求7所述的方法，还包括对所述孔径的至少一个碎片整理部分进行冲刷碎片整理，所述碎片整理部分包括关键数据。

10. 根据权利要求7所述的方法，还包括基于以下标准中的至少一个标准将A-数据的至少一部分指定为关键数据：工作负载关键性、虚拟机优先级或客户状态。

11. 根据权利要求7所述的方法，还包括检测所述数据完整性威胁，将所有所述关键数据从所述孔径冲刷到所述非易失性存储装置，以及将至少一些非关键数据从所述孔径冲刷到所述非易失性存储装置失败。

12. 根据权利要求7所述的方法，还包括检测所述数据完整性威胁，将至少所有所述关键数据从所述孔径冲刷到所述非易失性存储装置，然后将所有冲刷的数据从所述非易失性存储装置恢复到以下各项中的至少一项：易失性存储器、所述数据从其中被冲刷的电池支持的存储器或者与所述数据从其中被冲刷的所述电池支持的存储器不同的电池支持的存储器。

13. 根据权利要求7所述的方法，还包括以下各项中的至少一项：

将Z-数据的至少一个副本保留在所述孔径之外和所述与孔径相关联的非易失性存储装置之外；

指定孔径大小或孔径地址中的至少一项，并且其中所述指定基于以下各项中的至少一项：电池容量、电池寿命或电池可靠性值；

将关键数据从所述孔径冲刷到所述非易失性存储装置，其中所述冲刷包括执行具有统一可扩展固件接口的固件；

保留所述孔径的至少百分之十，以保持未被指定为关键数据的数据；

将未被指定为关键数据的数据移动到所述孔径之外的存储位置，以便为被指定为关键数据的数据腾出空间；或者

对所述孔径的至少一部分进行碎片整理，从而减少或消除先前位于所述孔径的两个已分配部分之间的未分配部分。

14. 根据权利要求7所述的方法，其中所述电池支持的存储器孔径驻留在这里表示为M的物理机中，并且将A-数据存储在该孔径中的请求从这里表示为N的不同的物理机被发

送。

15. 根据权利要求7所述的方法,包括将所述A-数据仅存储在一个存储节点中,并且将所述Z-数据的副本存储在多个存储节点中。

## 使用存储器孔径冲刷顺序的数据保存

### 背景技术

[0001] 设计、实现、修改或优化计算系统的系统架构师和其他人员认识到,数据存储选择通常涉及权衡。例如,一些能够存储数字数据的设备,诸如一些动态随机存取存储器设备,支持相对较快的数据存储操作,但是在断电的情况下不能可靠地保留数据。这些设备通常被称为“易失性”存储设备。存在许多易失性存储设备,它们具有不同的技术特性,诸如容量、成本、预期工作寿命、操作速度、更换难度、电气要求以及与硬件或软件标准或最佳实践的兼容性。其他数据存储设备,诸如机电硬盘驱动器和固态驱动器,即使在断电后也能可靠地保留数据值。这些设备通常被称为“非易失性”存储设备。也存在许多非易失性存储设备,同样具有不同的技术特性。

[0002] 因此,必须在可用的存储机制和过程之间做出选择的系统架构师和其他人面临着极其大量的存储设备选择,在技术选择和系统性能、成本、便利性、可靠性和其他特性之间有许多相互依赖的折衷。为了降低复杂性并增强可预测性,因此聚焦在特定的性能假设、目标或洞察力上可能会有所帮助,以便缩小将认真考虑的存储架构选择的范围并确定其优先级。

### 发明内容

[0003] 本文档中描述的一些实施例提供了改进的数据保存工具和技术,尤其是在联网的多节点环境中,诸如运行虚拟机的云。在一些实施例中,通过基于电池支持的存储器孔径冲刷顺序和关于电池特性的相关信息来进行存储器分配,丢失关键数据的风险被降低。这样的存储分配还可以在不损害数据完整性的情况下降低数据冗余要求。

[0004] 在一些实施例中,计算系统具有数据保存子系统,该数据保存子系统包括具有孔径的电池支持的存储器。该孔径具有冲刷顺序,该冲刷顺序是响应于诸如断电或重启的数据完整性威胁而将数据从该孔径复制到与孔径相关的非易失性存储设备的顺序。冲刷顺序定义了孔径的最先冲刷端和最后冲刷端。

[0005] 在该示例中,数据保存电路与电池支持的存储器可操作地通信。数据保存电路,其可以包括处理器和固件,被配置为执行数据保存步骤。在该示例中,这些数据保存步骤可以包括(a)接收将数据集A-数据存储在孔径中的请求,A-数据包括被指定为关键数据的数据,(b)标识孔径的未分配存储器的一部分A-存储器,A-存储器足够大以保持A-数据,A-存储器的地址比孔径的任何其他未分配存储器的任何其他地址更接近孔径最先冲刷端,以及(c)将A-存储器标记为已分配,并将A-数据的副本放置在A-存储器中。这些步骤(a)-(c)针对包含关键数据的不同的相应数据集可以重复多次。

[0006] 在该示例中,数据保存步骤可以包括(d)接收将数据集Z-数据存储在孔径中的请求,Z-数据不包括被指定为关键数据的任何数据,(e)标识孔径的未分配存储器的一部分Z-存储器,Z-存储器足够大以保持Z-数据,Z-存储器的地址比孔径的任何其他未分配存储器的任何其他地址更接近孔径最后冲刷端,(f)将Z-存储器标记为已分配,并将Z-数据的副本放置在Z-存储器中。这些步骤(d)-(f)针对不包含关键数据的不同的相应数据集可以重复,

并且当仅关键数据被存储时可以被省略。

[0007] 标签 (a) - (f) 在此仅用作标识符, 不必指定操作顺序。例如, 在将任何A-数据存储在孔径中之前,Z-数据可以被存储在孔径中。

[0008] 在该上下文中, 示例数据保存子系统通过在电池支持的存储器中, 以冲刷顺序将关键数据放在非关键数据之前, 与将数据存储在电池支持的存储器而不管数据关键性相比, 提供了成功冲刷关键数据并由此保存它的更高的可能性。考虑到本文描述的数据保存子系统能力, 可以减少或消除备选形式的数据保存, 诸如在受威胁的计算系统之外维护关键数据的复制或其他冗余副本。

[0009] 在操作中, 这里描述的一些数据保存实施例接收多个请求, 每个请求寻求在电池支持的存储器的孔径中存储相应的数据集。该孔径具有冲刷顺序, 该冲刷顺序是响应于数据完整性威胁而将数据从该孔径复制到与孔径相关的非易失性存储设备的顺序。冲刷顺序定义了孔径的最先冲刷端和最后冲刷端。在该示例中, 每个相应数据集包括被指定为关键数据的数据。对于请求中的至少两个请求, 该实施例标识孔径的未分配存储器的相应部分, 未分配存储器的每个部分足够大以保持相应的数据集。所标识的未分配存储器的相应部分的地址比该孔径的任何其他未分配存储器的任何其他地址更接近孔径的最先冲刷端。对于请求中的至少一个请求, 该实施例将所标识的未分配存储器的相应部分标记为已分配, 并在其中放置相应数据集的副本。

[0010] 继续该示例, 在存储至少一条关键数据之后的某一时刻, 该实施例检测数据完整性威胁, 例如, 检测即将到来的重启或即将失去外部电源并随后切换到电池电源。响应于威胁检测, 该实施例冲刷被复制到孔径中的所有关键数据, 从而将该关键数据从孔径复制到非易失性存储设备。该冲刷保存了被复制到孔径中的所有关键数据, 并且因此保存了关键数据(并且可能还保存了一些或全部非关键数据), 而不依赖于具有在孔径及其非易失性备份存储器范围之外的数据的冗余副本。

[0011] 对于本领域技术人员来说, 与本文的教导相关的其他技术活动和特征也将变得明显。所举的示例仅仅是说明性的。本发明内容不旨在标识所要求保护的的主题的关键特征或基本特征, 也不旨在用于限制所要求保护的的主题的范围。相反, 提供本发明内容是为了介绍——以简化的形式——一些技术概念, 这些技术概念将在下面的具体实施方式中进一步描述。根据正确理解的权利要求来定义本发明, 如果本发明内容与权利要求相冲突, 则应以权利要求为准。

## 附图说明

[0012] 将参考附图给出更详细的描述。这些附图仅图示了选定的方面, 因此不能完全确定覆盖范围或范围。

[0013] 图1是总体上示出了计算机系统的框图, 也是总体上示出了配置的存储介质的框图;

[0014] 图2是示出了包括计算节点和至少一个存储节点的环境的框图;

[0015] 图3是示出了某些计算环境的某些方面的框图;

[0016] 图4是示出了配置有数据保存功能性的系统的各方面的框图;

[0017] 图5是示出了电池特性的某些示例的框图;

- [0018] 图6是示出了数据关键性标准的某些示例的框图；
- [0019] 图7是示出了某些计算系统的某些示例的框图；
- [0020] 图8是示出了承载若干虚拟机并与具有电池支持的存储器孔径的若干存储节点联网的计算节点的架构图；
- [0021] 图9是示出了运行若干虚拟机并使用存储装置支持的存储器孔径的操作系统的架构图；
- [0022] 图10是示出了某些数据保存方法中的步骤的流程图；以及
- [0023] 图11是进一步示出了某些数据保存方法中的步骤的流程图。

## 具体实施方式

### [0024] 概述

[0025] 本发明可能会超越它们的起源,但了解本发明的起源可以帮助人们更充分地理解本发明。在目前的情况下,本文描述的一些教导的动机是存储用于机器学习的大数据的技术挑战。物联网设备可以快速、大量地输出数据,而机器学习工具和技术有时可以以有用的方式处理这些海量数据。一些机器学习数据库在活动数据库中包含数百万条记录,因此快速读取速度非常重要。尽管涉及大量数据,但数据完整性也很重要,这既体现在不期望对个别数据进行未经授权的故意改变,也体现在不期望重新启动或意外断电导致的数据改变。

[0026] 其他因素也在这种有关动机的情况中发挥了作用。具体地,成本方面的考虑阻止或不鼓励人们简单地将所有数据放在NVDIMM或类似的电池支持的快速随机存取存储器中。即使可以将所有数据放入NVDIMM中,也不能解决电池故障对数据完整性造成的风险。比方说,电池备份不如硬盘或DVD或磁带或一些类似的非易失性介质上的存储设备可靠。电池寿命随时间变化。可以考虑跨多个NVDIMM设备复制数据以提高安全性,但成本也是一个重要因素。此外,数据完整性的可接受风险水平并不总是恒定的。例如,不同的客户可能有不同的服务级别协议(SLA)或其他服务目标或保证。

[0027] 简而言之,本文描述的一些数据保存工具和技术在一定程度上受到保存大量机器学习数据的努力中存在的技术挑战的激励。然而,技术人员将认识到,本文提供的教导也有益地适用于许多其他技术场景。

[0028] 本文描述的一些实施例实现操作步骤和设备特性的组合,以保存数据免受完整性威胁。基于诸如客户要求、工作负载关键性或虚拟机关键性的标准,将数据分为关键数据和非关键数据。关键数据存储在与电池支持的存储器中的物理地址上,在该物理地址中,关键数据将在非关键数据之前被冲刷。一些实施例认识到并解决了电池的功率不足以将关键数据和非关键数据两者冲刷到安全的风险,例如,通过保留非关键数据的冗余副本以防其未被及时冲刷。电池支持的存储器孔径的大小根据电池的特性进行设置,并且可以随着条件的变化而调整大小。执行碎片整理是为了优化孔径的使用,特别是在保持关键数据的部分内。本文还讨论了本发明数据保存的其他方面。

[0029] 本文描述的一些实施例可以由一些人员在更广泛的上下文中来看待。例如,诸如分配、数据、顺序、电源、保存和请求等概念可以被认为与特定实施例相关。然而,可获得广泛上下文并不意味着本文寻求抽象概念的专有权;它们并非如此。相反,本公开集中于提供其技术效果完全或部分解决特定技术问题的适当特定实施例,诸如如何平衡NVDIMM成本与

服务级别协议数据可用性要求。涉及分配、数据、顺序、电源、保存或请求的其他配置存储介质、系统和过程不在本范围内。因此,在正确理解本公开的情况下,也避免了模糊性、纯粹的抽象性、缺乏技术特征以及随之而来的证明问题。

[0030] 更一般地,技术人员将认识到,不一定需要本公开的每个部分或其中的任何特定细节来满足法律标准,诸如可实施性、书面描述或最佳模式。而且,实施例不限于在此描述的特定激励示例、请求、响应、操作系统或其他内核、软件开发环境、接口标准、软件过程、开发工具、标识符、文件、数据结构、符号、控制流、伪代码、命名约定、节点架构或其他实现选择。与任何其他专利公开的任何明显冲突,即使是来自本发明的所有者,都不会在解释本专利公开中提出的权利要求方面扮演任何角色。

[0031] 技术特征

[0032] 本文描述的实施例的技术特征对于本领域普通技术人员来说将是显而易见的,并且对于广泛关注的读者也将以几种方式显而易见。一些实施例解决诸如存储器分配、将数据从易失性存储器冲刷到非易失性存储器、基于电池特性调整设置或调整存储器孔径的大小以及对存储器进行碎片整理等技术活动,这些都是深深植根于计算技术的活动。所讨论的一些技术机制包括例如,NVDIMM或其他电池支持的存储器、统一可扩展固件接口、计算节点、存储节点、存储器映射、冲刷顺序和虚拟机。所讨论的一些技术效果包括,例如,保存关键数据免受完整性威胁的可能性增加,减少对数据冗余的依赖以提供数据保存,以及考虑到诸如电池特性和为非关键数据保留的存储器量的标准来有效地设置电池支持的存储器孔径的大小。因此,纯粹的思维过程显然被排除在外。一些实施例通过在平衡诸如NVDIMM成本、客户服务级别要求和用于数据冗余的节点存储的可用性的标准的同时,保存数据免受完整性威胁,来改进计算系统和服务的功能。根据所提供的描述,基于教义的技术特性的其他优点对于技术人员来说也是显而易见的。

[0033] 缩略词、缩写、名称和符号

[0034] 下面定义了一些缩略词、缩写、名称和符号。其他的在本文的其他地方定义,或者不需要在本文定义以便技术人员理解。

[0035] ACPI:高级配置和电源接口

[0036] ALU:算术逻辑单元

[0037] API:应用程序接口

[0038] BIOS:基本输入/输出系统

[0039] BMC:底板管理控制器

[0040] CD:压缩盘

[0041] CPU:中央处理单元

[0042] EFI:可扩展固件接口

[0043] DRAM:动态随机存取存储器

[0044] DVD:数字多功能盘或数字视频盘

[0045] FPGA:现场可编程门阵列

[0046] FPU:浮点处理单元

[0047] GPU:图形处理单元

[0048] GUI:图形用户界面



- [0049] HDD:硬盘驱动器(例如,固态的、机电的、光的)
- [0050] IaaS或IAAS:基础设施即服务
- [0051] ID:标识或身份
- [0052] IoT:物联网
- [0053] LAN:局域网
- [0054] NVDIMM:非易失性双列直插式存储器模块
- [0055] NVMe:非易失性存储器高速
- [0056] OS:操作系统
- [0057] PaaS或PAAS:平台即服务
- [0058] RAM:随机存取存储器
- [0059] ROM:只读存储器
- [0060] SATA:串行ATA(计算机总线接口)
- [0061] SLA:服务级别协议
- [0062] SMM:系统管理模式
- [0063] TCP/IP:传输控制协议/互联网协议
- [0064] UEFI:统一可扩展固件接口
- [0065] VM:虚拟机
- [0066] WAN:广域网
- [0067] 一些附加术语

[0068] 本文参考诸如附图中所示的那些示例性实施例,并且本文使用特定语言来描述它们。但是,对本文所示的特征的更改和进一步修改,以及这里的特定实施例所示的抽象原理的附加技术应用,应当考虑在权利要求的范围内,它们对于拥有本公开的(多个)相关领域的技术人员来说是可以想到的。

[0069] 本公开中明确了术语的含义,因此阅读权利要求书时应仔细注意这些澄清。给出了具体的示例,但是(多个)相关领域的技术人员将理解,其他示例也可以属于所使用的术语的含义,并且在一项或多项权利要求的范围内。术语在这里的含义不一定与它们在一般用法(特别是在非技术用法中)、在特定行业的用法、或在特定词典或词典集合中的含义相同。附图标记可以与各种短语一起使用,以帮助显示术语的广度。从给定的文本中省略附图标记并不一定意味着文本没有讨论附图的内容。发明人主张并行行使特定和选定词典编纂的权利。引用的术语是显式定义的,但也可以不使用引号隐式定义术语。术语可以在这里的详细描述中和/或应用文件中的其他地方显式或隐式地定义。

[0070] 如本文所使用的,“计算机系统”(又称“计算系统”)可以包括例如,一个或多个服务器、主板、处理节点、膝上型计算机、平板计算机、个人计算机(便携式或非便携式)、个人数字助理、智能电话、智能手表、智能带、蜂窝或移动电话、具有至少一个处理器和存储器的其他移动设备、视频游戏系统、增强现实系统、全息投影系统、电视机、可穿戴计算系统和/或提供至少部分由指令控制的一个或多个处理器的(多个)其他设备。指令可以是存储器和/或专用电路中的固件或其他软件的形式。

[0071] “多线程”计算机系统是支持多个执行线程的计算机系统。术语“线程”应该理解为包括能够或服从于调度的代码,并且可能服从于同步。在本公开之外,线程也可以通过另一

名称被公知,例如,诸如“任务”、“进程”或“协程”。然而,本文在线程和进程之间进行了区分,因为线程定义了进程内的执行路径。此外,进程的线程共享给定的地址空间,而不同的进程具有不同的各自地址空间。进程的线程可以并行、顺序或并行执行和顺序执行(例如,时间分片)的组合来运行。

[0072] “处理器”是线程处理单元,诸如同时多线程实现中的核。处理器包括硬件。给定的芯片可以容纳一个或多个处理器。处理器可以是通用的,也可以是为特定用途定制的,诸如矢量处理、图形处理、信号处理、浮点算术处理、加密、I/O处理、机器学习等。

[0073] “内核”包括操作系统、管理程序、虚拟机、BIOS或UEFI代码以及类似的硬件接口软件。

[0074] “代码”是指处理器指令、数据(包括常量、变量和数据结构)或指令和数据两者。“代码”和“软件”在本文可互换使用。可执行代码、解释代码和固件是代码的一些示例。

[0075] 本文广泛使用的“程序”包括由程序员(也称为开发人员)编写和/或自动生成的应用、内核、驱动程序、中断处理程序、固件、状态机、库和其他代码。

[0076] “服务”是指在云计算环境或其他网络或计算系统环境中提供对多个程序的资源或资源访问的可消费程序。

[0077] “云”是指用于计算、存储和联网的池化资源,这些资源可灵活地用于计量的按需服务。云可以是私有的、公共的、社区的或混合的,并且云服务可以以基础设施即服务(IaaS)、平台即服务(PaaS)、软件即服务(SaaS)或其他服务的形式提供。除非另有说明,否则对从文件读取或向文件写入的任何讨论都包括读取/写入本地文件或通过网络读取/写入,该网络可以是云网络或其他网络,或两者(本地和联网读取/写入)。

[0078] “IoT”或“物联网”指的是可寻址的嵌入式计算节点的任何联网集合。这些节点是这里定义的计算机系统的示例,但它们也具有以下特征中的至少两个:(a)没有本地人类可读显示器;(b)没有本地键盘;(c)主要输入源是追踪非语言数据源的传感器;(d)没有本地旋转磁盘存储设备-RAM芯片或ROM芯片提供唯一的本地存储器;(e)没有CD或DVD驱动器;(f)嵌入家用电器或家用固定装置;(g)嵌入植入式或可穿戴医疗设备;(h)嵌入车辆中;(i)嵌入过程自动化控制系统;或(j)关注于以下其中一项的设计:环境监测、民用基础设施监测、工业设备监测、能源使用监测、人类或动物健康监测、物理安全或物理运输系统监测。IoT存储设备可能成为未经授权访问的目标,无论是经由云、经由另一网络,还是经由直接本地访问尝试。

[0079] 在某些情况下,例如,在机密性-完整性-可用性三位一体的网络安全讨论中,可以在“数据可用性”的含义和“数据完整性”的含义之间进行区分。但是如本文所使用的,“数据完整性”意味着既包括访问数据的能力,也包括访问具有正确值的数据的能力。因此,存储设备断电、值覆写、不可逆加密或编码或压缩、篡改、错位、加扰、不可访问性和数据值不确定性都是数据完整性威胁的示例。

[0080] 除非另有明确说明,否则,如本文所使用的,“碎片整理”指的是冲刷碎片整理(“碎片整理”是“碎片整理程序”的缩写)。冲刷碎片整理是一种操作,它减少孔径最先冲刷端附近未分配的存储器量,或者增加从孔径最先冲刷端的未分配存储器的距离,或者两者。例如,冲刷碎片整理可以涉及复制数据、或移动数据或将数据插入到先前未分配的存储器中。

[0081] “冲刷碎片整理”与“整合碎片整理”的不同之处在于它们有不同的目标。冲刷碎片

整理的目标是优先冲刷关键数据,而不是冲刷非关键数据或冲刷未分配存储器中的垃圾或不相关数据。相比之下,整合碎片整理的目标是减少从存储设备取回数据以用于使用所需的时间,特别是从旋转磁盘或磁带取回数据所需的时间。整合碎片整理尝试将属于特定所有者的所有数据(例如,特定文件中的所有数据)整合在一起。这些都不是等同的操作。

[0082] 例如,设A或B分别表示A或B拥有的关键数据,设下划线\_表示存储器中未分配的空间,并用大括号表示存储器孔径界限。则将{ABA\_}更改为{AAB\_}是整合碎片整理,但不是冲刷碎片整理。此外,将{A\_B\_}改变为{AB\_\_}是冲刷碎片整理(如果最先冲刷端在左侧),但不是整合碎片整理。

[0083] 如本文所使用的,除非另有说明,否则“包括”允许附加的元素(即,包括意味着包含)。

[0084] “优化”的意思是改进,不一定是完美的。例如,可以对已优化的程序或算法进行进一步改进。

[0085] “进程”本文有时用作计算科学领域的术语,并且在该技术意义上包括计算资源用户,例如,其还可以包括或称为协程、线程、任务、中断处理程序、应用进程、内核进程、过程或对象方法。实际上,“进程”是由诸如 Windows® Task Manager、Linux® ps的系统实用程序或其他操作系统环境中的类似实用程序(分别为Microsoft Corporation、Linus Torvalds的商标)标识的计算实体。“过程”本文也用作专利法领域的术语,例如,在描述与系统权利要求或制品(配置的存储介质)权利要求相对的过程权利要求时。类似地,本文中有时使用“方法”作为计算科学领域中的技术术语(一种“例程”),也用作专利法领域的术语(“过程”)。专利法意义上的“过程”和“方法”在本文可互换使用。技术人员将理解在特定情况下的意图是什么,并且还将理解给定的要求保护的过程或方法(在专利法意义上)有时可以使用一个或多个过程或方法(在计算科学意义上)来实现。

[0086] 与没有自动化的情况相反,“自动”意味着通过使用自动化(例如,由软件为本文讨论的特定操作和技术效果配置的通用计算硬件)。具体地,“自动”执行的步骤不是在纸上手动或在人脑中执行的,尽管它们可以由人发起或由人交互地引导。利用机器执行自动步骤,以便获取如果没有这样提供的技术交互就不会实现的一个或多个技术效果。

[0087] 本领域技术人员理解,技术效果是技术实施例的假定目的。例如,在实施例中涉及计算,并且一些计算也可以在没有任何技术组件(例如,通过纸和笔,或者甚至作为思维步骤)的情况下执行,仅此事实并不能消除技术效果的存在或改变实施例的具体和技术性质。可以理解,诸如标识未分配的存储器、将数据复制到存储器、将存储器标记为已分配、将数据从易失性存储器冲刷到非易失性存储器以及这里讨论的许多其他操作的数据保存操作本质上是数字的。人脑不能直接与CPU或其他处理器接口,或与RAM或其他数字存储器接口,以读取和写入必要的数以执行本文教导的数据保存步骤。鉴于本公开,本领域技术人员将很好地理解这一点,但是有时可能需要告知或提醒其他人这一点。

[0088] “以计算方式”同样是指正在使用的计算设备(至少是处理器加存储器),不包括仅仅通过人的想法或仅仅通过人类行为来获得结果。例如,用纸和笔做算术不是这里所理解的算术计算。计算结果更快、更广、更深、更准确、更一致、更全面,和/或以其他方式提供超出人类表现范围的技术效果。“计算步骤”是通过计算执行的步骤。“自动地”和“以计算方式”都不一定意味着“立即”。这里可互换地使用“计算”和“自动”。

[0089] “主动”是指没有来自用户的直接请求。实际上,用户甚至可能不会意识到实施例的主动步骤是可能的,直到该步骤的结果已经呈现给用户。除另有说明外,本文描述的任何计算和/或自动步骤也可以主动完成。

[0090] 在整个文档中,使用可选的复数“(多个)”、“(几个)”或“(若干)”表示存在一个或多个所指示的特征。例如,“(多个)处理器”表示“一个或多个处理器”或等同的“至少一个处理器”。

[0091] 出于美国法律和实践的目的,在权利要求中或在其他地方,此处使用的词“步骤”并不是要援引部件加功能、步骤加功能或35号美国法典第112章第六段/第112(f)节的权利要求解释。任何与此有关的推定在此明确予以反驳。

[0092] 为了美国法律和实践的目的,这些权利要求不意图援引部件加功能解释,除非它们使用“用于…部件”一词。意图解释为部件加功能语言的权利要求语言(如果有的话)将通过使用短语“用于…部件”来明确地陈述该意图。当部件加功能解释适用时,无论是通过使用“用于…部件”和/或通过法院对权利要求语言的合法解释,在说明书中陈述的用于给定名词或给定动词的部件应当被理解为链接到权利要求语言,并且在本文中通过以下任一项链接在一起:出现在附图的框图中的同一框内,由相同或相似的名称表示,由相同的附图标记表示,在任何附图中描述的功能关系,在本公开文本中注明的功能关系。例如,如果权利要求书限制记载了“zac微件”,并且该权利要求限制服从部件加功能的解释,那么至少说明书中任何提及“zac微件”的附图框、段落或示例中的任何地方标识的所有结构,或者由分配给zac微件的任何附图标记捆绑在一起,或者被公开为与zac微件的结构或操作具有功能关系的所有结构,将被认为是用于zac微件的申请中所标识的结构的一部分,并且将有助于定义zac微件结构的等同形式集合。

[0093] 技术人员将认识到,本发明公开讨论了各种数据值和数据结构,并认识到这些项驻留在存储器(RAM、磁盘等)中,从而配置存储器。技术人员还将认识到,本发明公开讨论了在给定实现中要体现在可执行代码中的各种算法步骤,并且这样的代码也驻留在存储器中,并且它有效地配置了执行它的任何通用处理器,从而将其从通用处理器转换为在功能上是专用硬件的专用处理器。

[0094] 因此,本领域技术人员不会犯将(a)权利要求中记载的存储器和(b)权利要求中记载的数据结构或数据值或代码视为非重叠项的错误。数据结构和数据值和代码被理解为驻留在存储器中,即使权利要求没有明确地记载所提及的每个数据结构或数据值或代码段的驻留。因此,不需要对这种驻留经历进行明确记载。然而,它们也不是被禁止的,并且可以呈现一个或两个选择记载以用于强调,而不会因此而将所有其他数据值和数据结构和代码排除在驻留之外。类似地,权利要求中记载的代码功能被理解为配置处理器,而不管权利要求中是否明确记载了该配置质量。

[0095] 在整个文档中,除非另有明确说明,否则对过程中的步骤的任何引用都假定该步骤可以由利害关系方直接执行和/或由该方通过干预机制和/或干预实体间接执行,并且仍然在该步骤的范围内。也就是说,除非直接执行是明确陈述的要求,否则不要求利害关系方直接执行该步骤。例如,关于目的地或其他主题的涉及利害关系方的动作的步骤,诸如分配、复制、碎片整理、特指、检测、确定、消除、执行、失败、冲刷、标识、保持、标记、移动、放置、保存、提供、接收、减少、保留、驻留、调整大小、恢复、存档、发送、指定、存储(以及分配、分配

的、复制、复制的等),可以涉及由某一其他方进行的干预动作,诸如转发、复制、上传、下载、编码、解码、压缩、解压缩、加密、解密、认证、调用等,包括本文档中记载的任何动作,但仍被理解为由利害关系方直接执行。

[0096] 无论何时提及数据或指令,都应该理解,这些项配置计算机可读存储器和/或计算机可读存储介质,从而将其转换成特定物品,而不是例如简单地存在于纸上、人脑中或仅仅作为在电线上传播的信号。出于美国的专利保护的的目的,根据美国专利商标局(USPTO)对In Re Nuijten案的解释,存储器或其他计算机可读存储介质不是传播信号、载波或纯粹超出可专利主体范围的能量。在美国,没有任何权利要求涵盖信号本身或仅仅是能量,任何鉴于本公开而断言不同的权利要求解释从表面上看都是不合理的。除非在美国境外授权的权利要求中另有明确说明,否则权利要求不包括信号本身或仅仅是能量。

[0097] 此外,尽管本文其他地方有任何明显相反的情况,但应理解(a)一方面,计算机可读存储介质和计算机可读存储器,以及(b)另一方面,传输介质(也称为信号介质)之间的明显区别。传输介质是传播信号或载波计算机可读介质。相比之下,计算机可读存储介质和计算机可读存储器不是传播信号或载波的计算机可读介质。除非在权利要求中另有明确说明,否则“计算机可读介质”指的是计算机可读存储介质,而不是传播信号本身,也不仅仅是能量。

[0098] 这里的“实施例”是示例。术语“实施例”不能与“本发明”互换。实施例可以自由地共享或借用各方面以创建其他实施例(只要结果是可操作的),即使在此未明确描述所得到的各方面的组合。对于本领域技术人员而言,要求明确地和单独地描述每个允许的组合是不必要的,并且与承认专利说明书是为本领域技术人员编写的政策相反。关于即使是少量可组合特征产生的可能组合的数目的正式组合计算和非正式共同直觉也将指示对于这里描述的方面存在大量方面组合。因此,要求明确记载每一种组合将与要求专利说明书简明扼要和读者对相关技术领域有一定了解的政策背道而驰。

[0099] 附图标记列表

[0100] 为方便起见,为支持附图,并作为说明书正文的一部分,提供了以下列表,其中通过引用多个项描述了本发明。这里未列出的项仍然可以是给定实施例的一部分。为了提高文本的可读性,在文本中引用的项的某些(但不是全部)记载附近记载了给定的附图标记。对于给定物品的不同示例或不同实例,可以使用相同的附图标记。附图标记列表为:

[0101] 100操作环境,也被称为计算环境

[0102] 102计算机系统,也称为计算式系统或计算系统

[0103] 104用户

[0104] 106外围设备

[0105] 108网络,总体上包括例如,LAN、WAN、软件定义的网络、云和其他有线或无线网络

[0106] 110处理器

[0107] 112计算机可读存储介质,例如,RAM、硬盘

[0108] 114可拆卸配置的计算机可读存储介质

[0109] 116可利用处理器执行的指令;可以在可移除存储介质上或在其他存储器(易失性或非易失性或两者)中

[0110] 118数据

- [0111] 120(多个)内核,例如,(多个)操作系统、BIOS、UEFI、设备驱动器
- [0112] 122工具,例如,反病毒软件、防火墙、分组嗅探器软件、入侵检测系统、入侵防御系统、调试器、剖析器、编译器、解释器、反编译器、汇编器、反汇编器、源代码编辑器、自动完成软件、模拟器、模糊器、存储库访问工具、版本控制工具、优化器、协作工具、软件开发工具和工具套件、硬件开发工具和工具套件,诊断器等。
- [0113] 124应用,例如,文字处理器、网络浏览器、电子表格、游戏、电子邮件工具
- [0114] 126显示屏幕,也称为“显示器”
- [0115] 128计算硬件,不以其他方式与附图标记106、108、110、112、114相关联;在图8中,标记128还指处理器110和存储器112硬件200计算功能性,例如,在计算节点中
- [0116] 202计算节点
- [0117] 204数据源,例如,输出或以其他方式生成数据的程序
- [0118] 206虚拟机,例如,提供硬件虚拟化并包括操作系统的计算构造;可以包括例如,工作存储器资源、CPU资源、IO资源和非易失性存储器
- [0119] 208系统代码,例如,操作系统、管理程序、BIOS、UEFI、固件210存储功能性,例如,在存储节点中
- [0120] 212存储节点
- [0121] 214非易失性存储设备,例如,NVDIMM、磁盘、闪存
- [0122] 216电池
- [0123] 218电池支持的存储器
- [0124] 220存储器孔径;可以指与物理设备中的数字数据存储位置相关联的物理地址空间的一部分,或者可以指数字数据存储位置本身
- [0125] 222数据存储请求
- [0126] 224对数据存储请求的响应
- [0127] 226数据副本
- [0128] 300计算环境方面
- [0129] 302云;也可以称为“云计算环境”
- [0130] 304数据保存电路;包括电子设备和控制这些电子设备或其对数据保存电路的使用的任何固件;不一定包括数据冲刷到的非易失性存储器
- [0131] 306数据保存子系统;包括数据冲刷到的非易失性存储设备
- [0132] 308管理程序
- [0133] 310UEFI;可以指统一可扩展固件接口,或指具有这种接口的固件;例如,可以包括或提供SEC(安全)、PEI(预EFI初始化)、DXE(驱动器执行环境)、BDS(引导设备选择)、SMM(系统管理模式)的功能
- [0134] 312冲刷顺序
- [0135] 314数据完整性威胁
- [0136] 316两个或更多个节点之间的存储关系
- [0137] 318数据关键性;可以指数据是否已被指定为关键,或者指数据是否满足我认为指定为关键的标准,即使还没有这样指定;可以是布尔值或另一指示,例如,指示如何确定数据冲刷优先级的两个或更多个值的范围中的值

- [0138] 402易失性存储器,例如,DRAM本身
- [0139] 404NVDIMM
- [0140] 406电池特性
- [0141] 408碎片整理程序(存储器碎片整理)代码
- [0142] 410存储器的碎片整理部分,即,要进行碎片整理的部分
- [0143] 412包括关键数据(也可以具有一些非关键数据)的数据集
- [0144] 414被分配(或正在被分配)以保存关键数据412的存储器
- [0145] 416不包括任何关键数据的数据集
- [0146] 418被分配(或正被分配)以保存非关键数据416的存储器
- [0147] 420固件
- [0148] 422分配器软件,即标识孔径中未分配空间的软件,向孔径的最先冲刷端放置其中的关键数据,并将该空间标记为已分配;分配器还可以对孔径进行碎片整理;分配器还可以标识用于非关键数据的空间,并向孔径的最后冲刷端放置非关键数据
- [0149] 424存储器112中的物理地址;可以指地址(例如,0x0000)本身或指地址处的存储位置(例如,存储单元)
- [0150] 502电池容量,例如,以毫安时为单位
- [0151] 504可靠性值,例如,平均无故障时间或剩余预期寿命,或枚举值(例如,在10为最好而0为最差的规格中为2)或可靠性等级(例如,中等可靠性)。
- [0152] 506电池寿命,例如,安装后经过的小时,或经历的充电循环次数,或冲刷期间使用的次数
- [0153] 600数据关键性标准示例
- [0154] 602数据关键性标准
- [0155] 604分配给生成数据的虚拟机的优先级
- [0156] 606客户,例如,云租户
- [0157] 608分配给客户的状态;可以是明显的,例如,在SLA中
- [0158] 610工作负载,例如,分配给特定计算节点的处理任务
- [0159] 612分配给工作负载的优先级,该工作负载包括数据或从其执行中生成数据
- [0160] 702物理机,例如,处理器、电路、芯片、电池、显示器、电源、外壳等,包括存储在物理机的非易失性存储器中的任何嵌入式固件或软件;计算系统102包括一个或多个物理机702
- [0161] 704云或其他网络中的节点;计算节点202和存储节点212是节点704的示例
- [0162] 706容器,例如,提供用户空间虚拟化并且本身不包括操作系统但仍然依赖于操作系统来执行的计算构造
- [0163] 800具有与一个或多个存储节点联网的一个或多个计算节点并配备有本文教导的数据保存功能的计算系统
- [0164] 802I/O(输入/输出)硬件和软件,例如,端口、连接器、插座、网络接口卡;也可以称为“I0”
- [0165] 804存储器顶部
- [0166] 806孔径的最先冲刷端

- [0167] 808孔径的最后冲刷端
- [0168] 810孔径内的未分配空间
- [0169] 812在相同存储设备中但在孔径之外的存储器部分
- [0170] 814虚拟地址空间和物理地址空间之间的逻辑映射
- [0171] 900具有运行虚拟机或容器的操作系统并配备有本文教导的数据保存功能的系统的计算系统
- [0172] 902专用于接收冲刷到其上的数据的非易失性存储设备112
- [0173] 904被映射到专用存储设备902的存储器区域;电池支持的的存储器孔径220是区域904的示例
- [0174] 906操作系统
- [0175] 908将数据从易失性存储器复制到后备易失性存储器的专用非易失性存储器的保存操作;可以有意地完成保存,而不会对正被保存的数据的完整性造成任何迫近威胁,或者可以是响应于迫近的或当前的威胁而完成的冲刷
- [0176] 910将数据从专用非易失性存储设备复制到由专用非易失性存储设备后备的易失性存储器的恢复操作
- [0177] 912程序的执行;可以指执行程序的行为,也可以指程序操作的实例
- [0178] 1000流程图;1000也指如图10流程图所示或与其一致的数据保存方法
- [0179] 1002接收存储关键数据的请求,例如,作为malloc()调用、构造器调用、文件或块或blob保存、或寻求存储尚未分配存储位置(字节、页等)的数据的其他操作的结果。
- [0180] 1004标识存储器的未分配部分,例如,使用系统存储器管理软件、空闲列表、显示空闲块的位向量或类似机制
- [0181] 1006将存储器的一部分标记为已分配,例如,通过在追踪存储器分配的数据结构中设置位
- [0182] 1008将数据副本放置在存储器中,例如,通过执行memcpy()、块复制等使用的功能
- [0183] 1010检测对数据完整性的威胁,例如,通过在符合ACPI的系统上接收电源命令,或通过从电压电平监控电路接收信号
- [0184] 1012将数据从易失性存储设备冲刷到专用非易失性后备存储设备,例如,通过将数据从NVDIMM的DRAM部分复制到NVDIMM的闪存部分
- [0185] 1100流程图;1100还指由图11流程图所示或与其一致的数据保存方法(其结合了图9的步骤和图10的步骤)
- [0186] 1102保存数据,即保持数据的至少一个副本的完整性
- [0187] 1104接收数据存储请求
- [0188] 1106确定或指定孔径220的大小或位置
- [0189] 1108孔径220的大小,例如,以字节或块为单位
- [0190] 1110孔径220的位置,例如,孔径的两端相对于包含孔径的存储设备的最低物理地址的偏移量
- [0191] 1112发送数据存储请求
- [0192] 1114调整孔径大小
- [0193] 1115移动孔径



- [0194] 1116保留孔径的至少指定部分以仅保存非关键数据
- [0195] 1118孔径的保留部分;可以由绝对或相对存储器地址指定,或指定为总孔径大小的百分比
- [0196] 1120对孔径的至少一部分进行碎片整理,通过复制数据或移动数据、或将数据插入先前未分配的存储器中,从而减少未分配的存储器量或增加未分配存储器与孔径的最先冲刷端的距离,或两者;这可以被称为“冲刷碎片整理”,以将其与涉及合并属于例如相同用户或相同进程或相同文件的存储器部分的“碎片整理”的其它用途区分开来
- [0197] 1122冲刷来自孔径的部分数据失败;这导致数据至少部分丢失,除非在孔径之外并且在数据的其他部分被冲刷到其中的非易失性存储设备之外的某个地方有数据的另一副本可用
- [0198] 1124驻留在物理机中,例如,存储在作为物理机一部分的存储器(易失性或非易失性)中
- [0199] 1126将数据的副本保存在孔径之外的某个地方,并且也在孔径将被冲刷到的非易失性存储器之外
- [0200] 1128移动数据
- [0201] 1130腾出空间,即,将存储器的给定部分从已分配改变为空闲
- [0202] 1132将数据指定为关键;当数据被视为单元时,则将该单元的任何部分指定为关键也将该单元指定为关键
- [0203] 1134减少已分配部分之间的未分配空间量
- [0204] 1136消除已分配部分之间的未分配空间
- [0205] 1138提供保存关键数据以防威胁的更好的可能性
- [0206] 1140保存关键数据以防威胁的可能性;可以是测量的概率,也可以是合理的评估
- [0207] 1142执行固件;执行912的示例
- [0208] 1144本公开中所讨论的未被赋予某一其它附图标记的任何步骤
- [0209] 操作环境
- [0210] 参考图1,实施例的操作环境100包括至少一个计算机系统102。计算机系统102可以是多处理器计算机系统,也可以不是。操作环境可以包括给定计算机系统中的一个或多个机器,其可以是集群的、客户端-服务器联网的和/或云内的对等联网的。单独的机器是计算机系统,一组协同工作的机器也是计算机系统。给定计算机系统102可以被配置以用于终端用户,例如,利用应用、用于管理员、作为服务器、作为分布式处理节点和/或以其他方式。
- [0211] 人类用户104可以通过使用显示器、键盘和其他外围设备106,经由键入文本、触摸、语音、移动、计算机视觉、手势和/或其他形式的I/O,与计算机系统102交互。屏幕126可以是可移除外围设备106,或者可以是系统102的组成部分。用户界面可以支持实施例与一个或多个人类用户之间的交互。用户界面可以包括命令行界面、图形用户界面(GUI)、自然用户界面(NUI)、语音命令界面和/或其他用户界面(UI)呈现,其可以被呈现为不同的选项或可以被集成。
- [0212] 系统管理员、网络管理员、云管理员、安全分析师和其他安全人员、操作人员、开发人员、测试人员、工程师、审核员、客户和终端用户都是特定类型的用户104。代表一个或多个个人动作的自动代理、脚本、回放软件、设备等也可以是用户104,例如以便于测试系统102。

存储设备和/或联网设备在一些实施例中可以被认为是外围设备,而在其他实施例中可以被认为是系统102的一部分,这取决于它们来自处理器110的可拆卸性。例如,图1中未示出的其他计算机系统可以使用经由网络接口设备到网络108的一个或多个连接,以技术方式与计算机系统102或与另一系统实施例交互。

[0213] 每个计算机系统102包括至少一个处理器110。计算机系统102,与其他适合的系统一样,还包括一个或多个计算机可读存储介质112。存储介质112可以是不同的物理类型。存储介质112可以是易失性存储器、非易失性存储器、固定位置介质、可移除介质、磁介质、光介质、固态介质和/或其他类型的物理耐久存储介质(与仅仅传播的信号或仅仅是能量相反)。具体地,诸如便携式(即,外部)硬盘驱动器、CD、DVD、记忆棒或其他可拆卸非易失性存储介质的配置存储介质114在插入或以其他方式安装时,可以在功能上成为计算机系统的技术部分,使得其内容可访问以用于处理器110交互和使用。可拆卸配置的存储介质114是计算机可读存储介质112的示例。计算机可读存储介质112的一些其他示例包括用户104不容易移除的内置RAM、ROM、硬盘和其他存储器存储设备。为了符合当前的美国专利要求,在美国待决或授权的任何权利要求下,计算机可读介质、计算机可读存储介质或计算机可读存储器都不是信号本身,也不是仅仅能量。

[0214] 存储介质114配置有可由处理器110执行的二进制指令;例如,在此广义上使用的“可执行”包括机器代码、可解释代码、字节码和/或在虚拟机上运行的代码。存储介质114还配置有数据118,该数据118通过执行指令116来创建、修改、引用和/或以其他方式用于技术效果。指令116和数据118配置它们所驻留的存储器或其他存储介质114;当该存储器或其他计算机可读存储介质是给定计算机系统的功能部分时,指令116和数据118也配置该计算机系统。在一些实施例中,数据118的一部分代表诸如产品特性、库存、物理测量值、设置、图像、读数、目标、体积等真实项。这些数据还通过备份、恢复、提交、中止、重新格式化和/或其他技术操作来转换。

[0215] 尽管实施例可以被描述为由计算设备(例如,通用计算机、服务器或集群)中的一个或多个处理器执行的软件指令,但是这样的描述并不意味着穷尽所有可能的实施例。本领域技术人员将理解,相同或相似的功能通常也可以全部或部分直接在硬件逻辑中实现,以提供相同或相似的技术效果。备选地,或者除了软件实现之外,本文描述的技术功能可以至少部分地由一个或多个硬件逻辑组件来执行,这些硬件逻辑组件可以包括固件或由固件控制,或者两者。例如,在不排除其他实现的情况下,实施例可以包括硬件逻辑组件110、128,诸如现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、片上系统组件(SOC)、复杂可编程逻辑器件(CPLD)以及类似组件。例如,实施例的组件可以基于它们的输入、输出和/或它们的技术效果被分组为交互功能模块。

[0216] 除了处理器110(例如,CPU、ALU、FPU和/或GPU)、存储器/存储介质112和显示器126之外,操作环境还可以包括其他硬件128,诸如电池、总线、电源、有线和无线网络接口卡。名词“屏幕”和“显示器”在文本可互换使用。显示器126可以包括一个或多个触摸屏、响应来自笔或写字板的输入屏幕、或仅为输出而操作的屏幕。在一些实施例中,诸如人类用户I/O设备(屏幕、键盘、鼠标、平板电脑、麦克风、扬声器、运动传感器等)的外围设备106将与一个或多个处理器110和存储器进行可操作的通信。

[0217] 在一些实施例中,该系统包括通过有线和/或无线网络108连接的多个计算机。网

网络接口设备128可以使用例如可能存在于给定计算机系统内的诸如分组交换网络接口卡、无线收发机或电话网络接口的网络组件来提供对网络108的访问。网络接口设备和其他网络组件(诸如交换机或路由器或防火墙)的虚拟化也可以存在于例如软件定义的网络或沙盒或其他安全云计算环境中。给定实施例还可以通过直接存储器访问、可移除非易失性存储介质或其他信息存储-取回和/或传输方法来传送技术数据和/或技术指令。

[0218] 本领域技术人员将认识到,本文的“操作环境”下呈现的前述方面和其他方面可以形成给定实施例的一部分。本文档的标题不旨在将特定严格分类为实施例和非实施例特征集合。

[0219] 一个或多个项在附图中以轮廓形式示出,或在括号内列出,以强调它们不一定是所示操作环境或所有实施例的一部分,但可以与本文讨论的操作环境或某些实施例中的项互操作。这并不意味着,在任何附图或任何实施例中,不是轮廓或括号形式的项都是必要的。特别地,提供图1是为了方便;在图1中包括项并不意味着该项或该项的描述用途在当前发明之前是已知的。

[0220] 有关系统的更多信息

[0221] 参考图1至图11,一些实施例使用或提供功能增强的系统800或900。该功能增强通过实现操作步骤和设备特性的组合来保存数据免受完整性威胁314,从而帮助促进数据保存。数据118被分为关键数据412和非关键数据416,并且关键数据被存储在物理地址424处的一个或多个电池支持的存储器218孔径220中,其中当检测到威胁1010时,关键数据将在非关键数据之前被冲刷1012。电池支持的存储器孔径220根据电池的特性406来设置大小,并且可以随着条件的改变而调整大小。碎片整理程序408优化孔径的使用以存储关键数据412。

[0222] 如图2的示例架构所示,数据118可以由源204生成,诸如在计算节点202上运行912的虚拟机206或容器706。这些数据源调用或依赖系统代码208来将生成的数据118发送到存储节点212。在该图2示例中,存储节点212具有快速易失性存储器402,该快速易失性存储器402的至少一部分是电池支持的存储器218。电池支持的存储器218的至少一部分被指定用作孔径220,以帮助保存关键数据,如本文所教导的。存储节点212和计算节点202经由请求222(例如,这里是要存储的数据,请发回存储在标识符Z下的数据)和响应224(例如,成功存储、孔径内空间不足、这里是所请求的数据、或其他状态或错误代码)进行通信。例如,请求222和响应224可以通过TCP/IP网络108、以太网108、存储区域网络108或其组合传送。

[0223] 通常,计算节点-存储节点关系不一定是一对一的。一个计算节点可以处理从一个存储节点发送或接收的数据,或者与多个存储节点通信;一个存储节点可以代表一个或多个计算节点存储数据。例如,图8示出了使用四个存储节点212的一个计算节点202。

[0224] 如图2和图8所示,在一些实施例中,关键数据412不跨存储节点复制,而非关键数据416跨存储节点复制。因此,可以主要或仅通过将关键数据冲刷到非易失性存储装置来保存关键数据免受威胁,而可以主要或仅通过跨节点复制非关键数据来保存非关键数据免受威胁。

[0225] 在图8中,孔径220中的关键数据412和对应的分配区域414属于虚拟机A、C和D,并且在图中由圆角表示,而孔径220中的非关键数据416和对应的分配区域418属于虚拟机B和X,并且在图中由直角表示。示出了属于虚拟机B和X的数据的数据副本226,其中存储节点

S1、S2、S3具有虚拟机B的非关键数据416的副本,并且存储节点S1、S2、S4具有虚拟机X的非关键数据416的副本。

[0226] 图8和图9示出了系统代码208的一些示例,即内核120(例如,管理程序308、操作系统906)和固件(例如,UEFI固件310)。计算节点和存储节点是计算系统102,因此包括硬件。尽管为了清楚起见,除了存储器112之外的硬件在图2中没有被明确地标注出来,但是在图8中标注了硬件的一些示例,例如,I/O 802和CPU 110,以及存储器112、218。存储器218包括孔径220、孔径外的存储器812、孔径内的已分配存储器418和414以及孔径内的未分配存储器810。

[0227] 电池支持的存储器218或另一包含孔径的映射易失性区域904的物理地址424可以通过映射814映射到虚拟地址。映射814实现一定程度的间接,这允许将相同的保存908和恢复910固件与可以具有彼此不同的物理起始地址806和物理结束地址808的孔径一起使用。孔径220数据被冲刷到的非易失性存储器112可以包括NVDIMM 404或其他专用非易失性存储器902中的闪存或其他非易失性存储器,或者两者。

[0228] 图3示出了一些计算环境100的各个方面300,在本公开中的各个点适当地讨论了这些方面。

[0229] 图4示出了可以将本发明的创新的数据保存功能性添加到计算系统102中的示例数据保存子系统306。子系统306的存储器包括电池支持的存储器218(例如,NVDIMM 404)中的孔径220。如图4和图5所示,电池可以具有被追踪或以其他方式指示的特性406,诸如容量502、可靠性504和寿命506中的一项或多项。这些电池特性406可以彼此重叠或影响,例如,可靠性可能随着寿命的增加而降低。子系统306还可以包括不由电池216支持的易失性存储器402。

[0230] 在操作中,孔径220将具有包含关键数据412的一个或多个区域414,并且孔径220可以具有包含非关键数据416的一个或多个区域418。例如,如图8存储节点S2、S3、S4所示,至少在给定的时间点,一些孔径220可以仅包含非关键数据,但是本文感兴趣的焦点是确实包含关键数据的孔径220。具体地,本公开告知如何将关键数据放置在存储器112中(通过初始分配,或通过碎片整理,或两者),以增加响应于数据威胁而成功冲刷关键数据并因此保存1102关键数据的可能性1140。

[0231] 图6示出了可用于区分关键数据412和非关键数据416的标准602的一些示例600。例如,如果数据是由虚拟机206或容器706或另一数字人工制品生成的,并且该人工制品具有高优先级604,则数据可以被指定为关键数据。如果数据118是由已经付费或以其他方式获取指示关键数据的状态608的特定客户606生成或代表其生成的,则数据118可以被指定1132为关键数据。类似地,如果数据由具有指示关键数据的优先级或状态612的特定工作负载610生成或代表其生成或由其处理,则数据可以被指定1132为关键数据。这些示例标准可以各种方式组合,或与其他标准602一起修改,或两者。例如,(a)当数据由高状态客户的高优先级虚拟机生成但以其他方式被指定为非关键时,数据可以被指定为关键数据;(b)当数据由在指定安全服务器刀片702上运行的高状态客户的任何虚拟机生成但以其他方式被指定为非关键时,数据可以被指定为关键数据;(c)当数据由任何客户在指定时间段内生成但以其他方式被指定为非关键数据时,数据可以被指定为关键数据,等等。

[0232] 图7示出了一些计算系统102的各种示例,在本公开中的各个点适当地讨论了这些

示例。

[0233] 一些实施例提供或使用计算系统102中的数据保存子系统306。数据保存子系统306可以包括具有孔径220的电池支持的存储器218。孔径220具有冲刷顺序312,该冲刷顺序312是响应于数据完整性威胁314将数据从孔径复制到与孔径相关联的非易失性存储器902的顺序。冲刷顺序定义了孔径的最先冲刷端806和孔径的最后冲刷端808。

[0234] 在较高物理地址424位于较低物理地址之上的情况下,也可以使用诸如“自上而下”或“自下而上”的短语来描述冲刷顺序。则自上而下冲刷将从较高地址开始并继续到较低地址,而自下而上冲刷将从较低地址开始并继续到较高地址。按照自上而下的冲刷顺序,该孔径的最先冲刷端806将具有比该孔径的最后冲刷端808高的地址,并且按照自下而上的冲刷顺序,该孔径的最先冲刷端806将具有比该孔径的最后冲刷端808低的地址。在图8中,示出了自下而上(低地址到高地址)的冲刷顺序312,因为该孔径的最先冲刷端806具有比该孔径的最后冲刷端808低的物理地址424。

[0235] 在一些实施例中,数据保存子系统306包括与电池支持的存储器218可操作地通信的数据保存电路304。数据保存电路304被配置为执行数据保存步骤,其可以包括:(a)接收1002将数据集A-数据412存储在孔径中的请求222,A-数据包括被指定为关键数据的数据,(b)标识1004孔径的未分配存储器810的一部分A-存储器414,A-存储器足够大以保持A-数据,A-存储器的地址424比孔径的任何其他未分配存储器的任何其他地址更接近孔径最先冲刷端806,(c)将A-存储器标记1006为已分配,并将A-数据的副本放置1008在A-存储器中,(d)接收将数据集Z-数据416存储在孔径中的请求222,Z-数据不包括被指定为关键数据的任何数据,(e)标识孔径的未分配存储器810的一部分Z-存储器,Z-存储器足够大以保持Z-数据,Z-存储器的地址424比孔径的任何其他未分配存储器的任何其他地址更接近孔径最后冲刷端808,以及(f)将Z-存储器标记为已分配,并将Z-数据的副本放置在Z-存储器中。这里的标签(a)至(f)仅用作标识符,因此其本身不必指定操作顺序。该数据保存子系统306通过在电池支持的存储器中,以冲刷顺序将关键数据放在非关键数据之前,来提供成功冲刷1012关键数据并因此保存它的更高的可能性。关键数据保存可能性比将数据存储在电池支持的存储器中而不考虑数据的关键性或缺乏关键性的架构中的可能性更高。

[0236] 在一些实施例中,电池支持的存储器孔径220驻留在此处指定为X的网络节点704中或由其控制,并且将A-数据存储在孔径中的请求222从此处指定为Y的不同网络节点发送。例如,在图8中,孔径驻留在存储节点212,704S1中,并且从不同于S1的网络节点(即,从计算节点)发送将虚拟机A的A-数据存储在孔径220中的请求222。在一些实施例中,网络节点X包括云302中的存储节点,并且网络节点Y包括在云中的计算节点。然而,在其他实施例中,节点可以驻留在非云网络中。

[0237] 在一些实施例中,网络节点X和网络节点Y通过这里指定为R的存储关系316而被关联,其中,R被定义为使得两个节点M和N在M代表N存储数据时,通过R关联。在一些情况下,R不是特定节点X和Y的一对一关系,例如,计算节点可以将虚拟机VM7的关键数据发送到存储节点S7,并且将虚拟机VM8的关键数据发送到存储节点S8。或者存储节点S12可以保存来自计算节点C10的关键数据,也可以保存来自不同的计算节点C11的关键数据。当然,许多其他关系316示例也是可能的。

[0238] 在一些实施例中,电池支持的存储器218包括NVDIMM存储器。例如,NVDIMM 404可

以包括NVDIMM-F闪存、NVDIMM-N字节可寻址存储器、在相同设备上具有动态RAM和NAND的NVDIMM-P存储器、NVDIMM-SW存储器或NVRAM非易失性RAM存储器。NVDIMM-SW可以包括存储器(例如,DD4 RAM)、磁盘或闪存(例如,SSD分区)、电池和信令机制。

[0239] 在一些实施例中,计算系统102包括虚拟机206,数据集A-数据412从虚拟机VM-A被发送到数据保存子系统306,并且数据集Z-数据416从虚拟机VM-Z被发送到数据保存子系统306。例如,在图8的示例中,关键数据集412从虚拟机A被发送到增强存储节点S1(并由此被发送到其数据保存子系统306),并且非关键数据集416从虚拟机B被发送到S1。

[0240] 通过本文对计算硬件的广泛讨论,本文还描述了其他系统实施例,这些系统实施例直接或可派生为所描述的过程或配置的介质的系统版本。

[0241] 尽管图中示出了具体的架构示例,但是实施例可以不同于这些示例。例如,在实施例中,不同附图中所示的项可以一起包括,图中所示的项可以省略,不同项中所示的功能可以组合成更少的项或单个项,项可以被重命名,或者项可以彼此不同地连接。

[0242] 这里提供示例以帮助说明该技术的各方面,但是本文档内给出的示例并不描述所有可能的实施例。实施例不限于本文提供的特定示例、组件名称、优化、算法选择、数据、数据类型、配置、实现、布置、显示、特征、方法或场景。例如,给定实施例可以包括附加或不同的技术特征、机制、序列、数据结构或功能,并且可以其他方式偏离本文提供的示例。

[0243] 过程(也称为方法)

[0244] 图10示出了方法1000,该方法是可以由本文教导的具有数据保存功能的增强型系统执行或辅助的方法的示例,并由图1至图9中的一个或多个图来图示。图10中所示的接收1002、标识1004、标记1006、放置1008、检测1010和冲刷1012的步骤在整个本公开中被讨论,无论是否明确记载那些附图标记。具体地,图2、图8和图9的讨论涉及图10所示的步骤1002到步骤1012。

[0245] 图11还示出了适合于在具有增强的数据保存功能的系统的操作期间使用的访问控制方法(在该词的法律意义上也可以被称为“过程”),包括用于图10中所示步骤的一些改进、补充或上下文动作。图11还结合了图9或图10中所示的步骤。除非另有说明,否则附图中所示或以其他方式公开的技术过程将自动地执行,例如由数据保存子系统306执行。过程也可以部分自动地和部分手动地执行,以达到涉及人类管理员或其他人的动作的程度,例如,在一些实施例中,人可以指定要为非关键数据保留1116的孔径的最小或最大百分比。在本文中被认为是本发明的任何过程都不是完全手动的。在给定实施例中,可以利用要操作的不同参数或数据,可以重复处理的零个或多个所示步骤。实施例中的步骤也可以按照与图10和图11中所示的从上到下的顺序不同的顺序来完成。可以串行、部分重叠的方式或完全并行地执行步骤。具体地,流程图1000动作项或流程图1100动作项被遍历以指示在过程期间执行的步骤的顺序可以从过程的一种表现到过程的另一种表现而不同。流程图遍历顺序也可以从一个过程实施例到另一个过程实施例而不同。步骤还可以省略、组合、重命名、重新分组、在一个或多个机器上执行、或者以其他方式偏离所示流程,只要所执行的过程是可操作的并且符合至少一个权利要求。

[0246] 一些实施例使用或提供一种用于计算系统中的数据保存的方法。该方法可以包括接收1002将数据集A-数据存储在电池支持的存储器的孔径中的请求。该孔径具有冲刷顺序312,该冲刷顺序312是响应于数据完整性威胁314将数据从该孔径复制908到与该孔径相关

联的非易失性存储器的顺序。冲刷顺序定义了孔径的最先冲刷端和最后冲刷端。A-数据包括被指定为关键数据的数据。

[0247] 该方法还可以包括标识1004孔径的未分配存储器的一部分A-存储器,其中A-存储器足够大以保持A-数据,并且A-存储器的地址比孔径的任何其他未分配存储器的任何其他地址更接近孔径的最先冲刷端。该方法还可以包括将A-存储器标记1006为已分配,并将A-数据的副本放置1008在A-存储器中。

[0248] 该方法还可以包括接收1104将数据集Z-数据存储在孔径中请求,其中Z-数据不包括被指定为关键数据的任何数据。

[0249] 该方法还可以包括标识1004孔径的未分配存储器的一部分Z-存储器,其中Z-存储器足够大以保持Z-数据,并且Z-存储器的地址比孔径的任何其他未分配存储器的任何其他地址更接近孔径最后冲刷端。该方法还可以包括将Z-存储器标记1006为已分配,并将Z-数据的副本放置1008在Z-存储器中。

[0250] 通过这样的方法,通过在电池支持的存储器中,以冲刷顺序将关键数据放置在非关键数据之前,与在电池支持的存储器中,以冲刷顺序不将关键数据放置在非关键数据之前相比,增加了成功冲刷1012关键数据并因此保存1102关键数据的可能性1138。

[0251] 在一些实施例中,该方法还可以包括至少部分地基于电池支持的存储器的电池特性来确定1106孔径的大小1108。例如,容量较小的电池、较旧的电池和不太可靠的电池可能会产生较小的孔径。在一些实施例中,该方法还可以包括至少部分地基于电池支持的存储器的电池特性来确定1106孔径的位置1110。例如,可以选择由较新、较大容量或更可靠的电池支持存储器中的孔径,而不是由较小、较旧或较不可靠的电池支持的存储器中的孔径。在一些实施例中,该方法可以包括至少部分地基于电池支持的存储器的电池特性的改变来调整1114孔径大小或移动1115孔径,例如,通过当用更新的、更大的或更可靠的电池替换电池时增加孔径大小。与本文公开的其他步骤一样,这些步骤可以在给定实施例中组合。

[0252] 在一些实施例中,该方法还可以包括对孔径的至少一个碎片整理部分进行碎片整理1120,该碎片整理部分包括关键数据。如本文别处所讨论的,碎片整理1120指的是冲刷碎片整理;整合碎片整理可以是副作用,但本质上不足以充当冲刷碎片整理1120。碎片整理1120可以是主动的。当租户或进程终止、腾出或搬迁时,也可以触发碎片整理1120。

[0253] 在一些实施例中,该方法还可以包括基于以下标准602中的至少一个将A-数据的至少一部分指定为关键数据:工作负载关键性、虚拟机优先级或客户状态。

[0254] 在一些实施例中,该方法可以包括检测1010数据完整性威胁,将所有关键数据从孔径冲刷1012到非易失性存储器,以及将至少一些非关键数据从孔径冲刷到非易失性存储器失败1122。例如,存储在电池中的功率可能不足以冲刷整个孔径,但仍可能足以冲刷所有关键数据。此示例场景说明了放置关键数据的优势,使得在电池电量功率的情况下,在冲刷非关键数据之前对关键数据进行冲刷。

[0255] 在一些实施例中,该方法可以包括检测1010数据完整性威胁,将至少所有关键数据从孔径冲刷1012到非易失性存储器,然后从非易失性存储器恢复910所有冲刷的数据。可以将冲刷的关键数据恢复910到以下各项中的至少一项中:易失性存储器、数据从其中被冲刷的电池支持的存储器、或与数据从其中被冲刷的电池支持的存储器不同的电池支持的存储器。

[0256] 在一些实施例中,该方法可以包括将Z-数据的至少一个副本保持1126在孔径之外和与孔径相关的非易失性存储设备之外。也就是说,当构建用于数据保存的系统时,非关键数据的复制与对关键数据的仔细放置1008和冲刷1012的依赖是兼容的。

[0257] 在一些实施例中,电池支持的存储器孔径驻留1124在这里表示为M的物理机中,并且在该孔径中存储A-数据的请求从这里表示为N的不同物理机发送1112。例如,在图2或图8所示的配置中,计算代码和存储节点可以是不同的物理机。在备选配置中,它们可以是运行在相同底层物理硬件上的虚拟设备。

[0258] 在一些实施例中,该方法可以包括仅将A-数据存储在一个存储节点中,并且将Z-数据的副本存储在多个存储节点中。也就是说,在一些示例中,关键数据保存仅依赖于放置1008和冲刷1012,而非关键数据保存至少部分依赖于复制226。

[0259] 配置的存储介质

[0260] 一些实施例包括配置的计算机可读存储介质。存储介质112可以包括磁盘(磁、光或其他)、RAM、EEPROM或其他ROM、和/或其他可配置存储器,具体包括计算机可读存储介质(其不仅仅是传播信号)。被配置的存储介质尤其可以是诸如CD、DVD或闪存的可移除存储介质114。通用存储器,可以是可移除的或不可移除的,并且可以是易失性的或不可移除的,可以被配置为使用诸如分配器422、碎片整理器408和映射814的项的实施例,以从可移除存储介质114和/或诸如网络连接的另一源读取的数据118和指令116的形式,来形成配置的存储介质。如本文所公开的,配置的存储介质112能够使计算机系统102执行用于数据保存的技术处理步骤。因此,这些图有助于说明配置的存储介质实施例和过程(也称为,方法)实施例,以及系统和过程实施例。具体地,图9、10或11中所示的或以其他方式在此教导的任何过程步骤可用于帮助配置存储介质以形成配置的存储介质实施例。

[0261] 一些实施例使用或提供计算机可读存储介质112、114,其配置有数据118和指令116,其在由处理器110执行时使计算系统或其子系统执行用于数据保存的方法。该方法包括接收1002多个请求,每个请求寻求在电池支持的存储器的孔径220中存储相应的数据集,该孔径具有冲刷顺序312,该冲刷顺序是响应于数据完整性威胁而将数据从该孔径复制到与孔径相关的非易失性存储设备的顺序,该冲刷顺序定义该孔径的最先冲刷端806和该孔径的最后冲刷端808,每个相应数据集包括被指定为关键数据的数据。

[0262] 该方法还包括,针对请求中的至少两个请求,标识1004孔径的未分配存储器的相应部分,未分配存储器的每个部分足够大以保持相应的数据集,所标识的未分配存储器的相应部分的地址比孔径的任何其他未分配存储器的任何其他地址更接近孔径的最先冲刷端。

[0263] 该方法还包括,针对请求中的至少一个请求,将所标识的未分配存储器的相应部分标记1006为已分配,并在其中放置1008相应数据集的副本。

[0264] 该方法还包括检测1010数据完整性威胁,以及将复制到孔径中的所有关键数据从孔径冲刷1012到非易失性存储设备。因此,尽管存在数据完整性威胁,该方法仍保存1102被复制到孔径中的所有关键数据。

[0265] 在一些实施例中,该方法还包括指定1106孔径大小1108或孔径地址1110中的至少一项,并且该指定基于以下各项中的至少一项:电池容量502、电池寿命506或电池可靠性值504。



[0266] 在一些实施例中,冲刷1012包括执行1142具有统一可扩展固件接口的固件310。

[0267] 在一些实施例中,该方法还包括保留1116孔径的至少百分之十1118以保持未被指定为关键数据的数据。例如,保留百分之二十或保留百分之三十将分别符合保留1116至少百分之十。在一些实施例中,保留1116的部分1118至多为百分之十,或者至多百分之二十。保留1116还可以被视为或实现为将孔径或存储器分区划分为子分区。

[0268] 在一些实施例中,该方法还包括将未被指定为关键数据的数据移动1128到孔径外部的存储位置,以便为被指定为关键数据的数据腾出空间。在一些情况下,该方法包括对孔径的至少一部分进行碎片整理1120,从而减少1134或消除1136先前位于孔径的两个已分配部分之间的未分配部分。

[0269] 附加示例和意见

[0270] 技术人员将认识到,不一定需要本公开的每个部分或其中的任何特定细节来满足法律标准,诸如可实施性、书面描述或最佳模式。而且,实施例不限于这里描述的特定网络、协议、工具、标识符、字段、数据结构、功能、秘密或其他证明、或其他实现选择。与任何其他专利公开的任何明显冲突,即使是来自本发明所有者的,都不会在解释本专利公开中提出的权利要求方面扮演任何角色。基于与本公开的所有部分相关的理解,提供了一些附加示例和意见。

[0271] 在一些实施例中,易失性存储器402由电池216备份,并且当数据的完整性受到威胁314时(例如,由于即将断电或重启),可以被快速冲刷1012以在非易失性存储设备902中创建数据118的副本。冲刷1012受冲刷顺序312的约束,因此一些数据可能不会冲刷到安全位置。例如,冲刷可能由于电池故障或耗尽而不完整。冲刷是数据生命周期操作的示例;一些其他示例为分配、使用、复制、释放、移动、碎片整理和恢复。

[0272] 一些实施例的设计和实现具有供应足够大的电池的目标。因此,孔径大小可以取决于架构师或开发者得出的电池可以保护的数据量118。因此,更大的电池、更新的电池、更可靠的电池可以导致更大的孔径,反之,更小/更旧/更不可靠的电池或未知的电池特性可以导致更小的孔径。此外,电池可能会在一段时间内降低级别/降级,因此保存大量数据的能力可能会随着时间的推移而受到影响。可以相应地调整孔径220的大小。

[0273] 一些实施例实现混合放置顺序,由此关键数据在孔径的第一冲刷端806端或其附近分组,而非关键数据在孔径的最后冲刷端808处或其附近分组。非关键数据或关键数据(或两者)也可以复制到其他孔径或任何电池支持的孔径外部的其他存储器。

[0274] 在某些情况下,可以从文档、从存储器映射814中、从与关于数据关键性的信息组合的存储器转储、从对机器的探测或它们之间的通信、从联合测试行动小组(JTAG)调试、从NVDIMM-SW保存/恢复测试、从固件或其他代码列表、从其他证据或从其组合,来推断本文教导的数据保存技术的存在。

[0275] 一些实施例提供或使用NVDIMM-SW保存算法,该算法可以将工作负载的关键性612作为对NVDIMM-SW数据集进行排名和排序的方式。一些实现由虚拟机ACPI地址范围描述的存储类存储器(SCM)分区904,以及在虚拟机之间和到SCM分区的映射222、224。这将指导NVDIMM-SW在系统固件的运行组件(例如,用于Intel/AMD CPU的UEFI/SMM和用于ARM CPU的UEFI/信任区)中保存路由。一些实施例在任何优雅或不优雅的关机中保护908内容118,允许虚拟机内容118在下次启动时被恢复910。在某些情况下,NVDIMM-SW有助于确保快速

访问数据,同时帮助确保数据冲刷到非易失性存储,即使在服务中断(诸如节点脱机)的情况下也是如此。

[0276] 在一些实施例中,例如,可以触发冲刷的事件可以包括有序关机、硬重启或软重启。

[0277] 有序关机使正在运行的软件有机会通过保存目录、分配表和其他描述数据组织的结构的当前版本来防止数据损坏,以及保存数据本身。例如,关机命令命令应用和其他进程关闭,这进而使这些进程有机会将数据冲刷到非易失性存储器,并关闭任何打开的文件,并将分配的存储器释放回OS。关机命令还命令设备驱动程序将IO数据和当前目录信息冲刷到附接的设备。在符合ACPI的系统上,关机命令可以导致发出电源命令,这使得NVDIMM将数据从其易失性部分保存908到其非易失性部分。

[0278] 硬重启在不通电的情况下开始,可能电源按钮除外。电力被给到系统。启动存储器代码(BIOS/EFI/UEFI)执行POST。然后,启动存储器代码将启动加载程序从启动设备加载到RAM中。启动加载程序从非易失性工作存储器或通过网络连接将OS加载到RAM中。硬重启开始时RAM中的任何内容都可能被覆写。

[0279] 软重启从系统通电开始。已跳过开机自检(POST)。启动存储器代码将启动加载程序从启动设备加载到RAM中。启动加载程序从非易失性工作存储器或通过网络连接将OS加载到RAM中。软重启开始时RAM中的任何内容都可能被覆写。

[0280] 在一些实现中,OS使用系统存储器映射来标识在OS启动之后为UEFI运行时功能预留哪些存储器区域904以及为OS使用回收哪些区域。在某些情况下,存储器映射是在UEFI预启动的所有阶段开发的。当分配存储器用于存储代码和数据时,它会发生改变。存储器分配的类型基于代码和数据的类型而变化,例如启动服务代码、启动服务数据、运行时代码、运行时数据、ACPI NVS存储器、SMM等。

[0281] 在一些实施例中,存在NVDIMM-SW,并提供固件辅助部件,由其通过DIMM和块存储设备(NVMe或SATA等)的耦合来模拟非易失性DIMM。目标是采用标准驱动的方法,从而将NVDIMM-SW设备无缝发布到操作系统。在标准操作模式下,预启动固件会将可用的NVDIMM-SW设备发布到操作系统。来自利用NVDIMM-SW的应用程序的所有写入操作都针对与NVDIMM-SW关联的DIMM区域。SAVE#在正常或非正常关机情境下启动。SAVE#操作在系统关机和后续重启之前完成。RESTORE#操作在控制权交还给操作系统之前由预启动固件启动。NVDIMM-SW可以通过耦合DDR4 DIMM和板载M.2NVMe模块来构建。NVDIMM-SW范例重新调整传统DIMM的用途,以模拟字节可寻址存储器。在某些实现中,系统中填充了大小和存储器类型相同的所有DIMM。NVDIMM-SW可以支持两种操作模式,即NVDIMM非交织或NVDIMM交织。NVDIMM非交织,其中插槽1上的DIMM将被选为非易失性NVDIMM(SW)NVDIMM,取决于经由设置选项选择的非易失性存储器大小。最大非易失性存储器大小取决于非正常保存场景期间的电源持续时间。非交织NVDIMM-SW除了模拟所有I2C(串行协议)\_DSM(设备特定方法)数据外,与NVDIMM-N基本相同。在NVDIMM交织中,插槽内的所有DIMM将交织在一起以支持NUMA。此外,基于所选择的非易失性存储器大小,从系统存储器映射中将每个插槽的顶部存储器切出以作为NVDIMM-SW。事件处理的目标是尝试将数据从易失性存储器保存到非易失性存储器区域。固件堆栈(UEFI、BMC)还负责记录带内监听器和带外监听器两者的错误。

[0282] 一些附加组合和变型

[0283] 代码、数据结构、逻辑、组件、通信和/或它们的功能等同的这些组合中的任何一个也可以与上述任何系统及其变型相结合。过程可以包括在此以可操作的任何子集或组合或序列描述的任何步骤。每个变型可以单独出现,也可以与任何一个或多个其他变型组合出现。每个变型可以与任何过程一起出现,并且每个过程可以与任何一个或多个其他过程组合。包括变型的每个过程或过程组合可以与上述配置的存储介质组合和变型中的任何一个组合。

[0284] 结论

[0285] 简而言之,本文提供的教导可以应用于增强计算系统中的数据保存功能。在一些实施例中,组合的操作步骤和设备特性有助于保存数据118免受完整性威胁314。基于诸如客户606要求608、工作负载610关键性612或虚拟机206关键性604的标准602,数据118被划分为关键数据412和非关键数据416。例如,数据118可以在计算节点202中生成并被发送222以存储在存储节点212中。关键数据412被存储在电池支持的存储器218孔径220中的物理地址424处,其中由于由电池支持的存储器的电路施加的或由外部力(例如,系统固件420)施加在电池支持的存储器上的冲刷顺序312,关键数据412将在非关键数据416之前被冲刷1012。冲刷顺序312可以是例如自下而上的NVDIMM 404冲刷顺序。数据118(尤其是非关键数据416)的冗余副本226也可以被保留1126,以防复制的数据没有被及时冲刷1012以保存1102它。电池支持的存储器孔径220根据它们的电池特性406来设置大小1106和位置1106,并且可以随着条件改变而重新定位1115或调整大小1114。执行冲刷碎片整理1120以优化孔径220的使用,尤其是在保存关键数据412的孔径的部分414内。

[0286] 实施例被理解为还包括或受益于经测试的和适当的安全控制和隐私控制,诸如,通用数据保护条例(GDPR),例如,应当理解应当采取适当的措施来帮助防止通过恶意软件的注入或激活来滥用计算系统,并帮助避免篡改增强系统在程序执行期间可能处理的任何个人或私人信息。这里教导的工具和技术的使用与这种控件的使用是兼容的。

[0287] 尽管特定实施例在此被明确地示出并描述为过程、配置的存储介质或系统,但是可以理解,对一种类型的实施例的讨论通常也扩展到其他实施例类型。例如,结合图9、10和11对过程的描述还有助于描述配置的存储介质,并有助于描述系统和制造商的技术效果和操作,如结合其他附图讨论的那些。这并不意味着必须将来自一个实施例的限制读入另一实施例。具体地,过程不一定限于在讨论系统或制造商(例如,配置的存储器)时呈现的数据结构和布置。

[0288] 本领域技术人员将理解,实现细节可能与特定代码有关,例如特定API、特定字段、特定种类的组件(硬件或软件)和特定样本程序,因此不需要出现在每个实施例中。本领域技术人员还将理解,在讨论细节时使用的程序标识符和一些其他术语是特定于实现的,因此不需要与每个实施例相关。然而,尽管它们不一定要在这里出现,但是这些细节可以通过提供上下文和/或可以说明这里讨论的技术的许多可能实现中的几个来帮助一些读者。

[0289] 通过适当关注这里提供的项,包括技术过程、技术效果、技术机制和技术细节,其对所有权利要求或可主张权利要求的实施例的说明性的但不是全面的,技术人员将理解,本公开和在此描述的实施例不针对技术领域之外的主题,或者针对诸如主要或原始原因或动机的本身的任何想法,或者仅仅针对结果本身,或者针对思维过程或思维步骤,或者针对商业方法或流行的经济实践,或者仅仅针对方法或涉及自然发生的事物或过程,或涉及生

物或生物的一部分,或涉及数学公式本身,或涉及孤立的软件本身,或仅涉及常规计算机,或涉及任何完全不可察觉的或任何抽象概念本身,或涉及微不足道的解决方案后活动,或涉及完全在未指明的装置上实现的任何方法,或涉及未能产生有用和具体结果的任何方法,或涉及对所有使用领域的抢占,或涉及根据寻求或正在许可或实施专利保护的管辖区的法律不符合专利保护资格的任何其他主题。

[0290] 这里引用具有某特征X的实施例,并且在本文其他地方引用具有某特征Y的实施例,这并不排除具有特征X和特征Y两者的实施例,除非在此明确声明这样的排除。所有可能的否定权利要求限制都在本公开的范围之内,也就是说,被声明为实施例的一部分的任何特征也可以明确地从另一个实施例中移除,即使在这里的任何示例中没有给出该特定排除。在此,术语“实施例”仅用作“过程、系统、制品、经配置的计算机可读存储介质、和/或以符合适用法律的方式应用的本文教导的其他示例”的更方便的形式。因此,给定的“实施例”可以包括这里公开的特征的任何组合,只要该实施例与至少一项权利要求一致。

[0291] 并不是图中所示的每一项都需要出现在每个实施例中。相反,实施例可以包含图中未明确示出的(多个)项。尽管这里通过特定示例在文本和附图中图示了一些可能性,但是实施例可以偏离这些示例。例如,示例的特定技术效果或技术特征可以被省略、重命名、不同地分组、重复、以不同方式在硬件和/或软件中实例化,或者是出现在两个或更多个示例中的效果或特征的混合。在一些实施例中,在一个位置所示的功能也可以在不同的位置提供;本领域技术人员认识到,在给定的实现中可以以各种方式定义功能模块,而不必从作为整体来看的交互模块的集合中省略期望的技术效果。由于空间限制或为方便起见,可以在图中的单个框中一起示出不同的步骤,但是仍然可以单独执行,例如,在方法的给定执行中,可以在没有另一个步骤的情况下执行一个步骤。

[0292] 通过附图标记通篇引用附图。在附图或文本中,与给定附图标记相关联的措辞中的任何明显不一致都应理解为简单地扩大了该附图标记所引用的范围。即使使用相同的附图标记,给定附图标记的不同实例也可以指不同的实施例。类似地,给定的附图标记可用于指代动词、名词和/或每一个的对应实例,例如,处理器110可通过执行指令来处理110指令。

[0293] 如本文所使用的,术语诸如“一”、“一个”和“该”包括一个或多个所指示的项或步骤。具体地,在权利要求中,对项的引用通常意味着存在至少一个这样的项,并且对步骤的引用意味着执行该步骤的至少一个实例。类似地,在上下文允许的情况下,“是”和其他单数动词形式应该包含“都是”和其他复数形式的可能性,以避免语法错误或误解。

[0294] 标题仅为方便起见;有关给定主题的信息可以在标题指示该主题的部分之外找到。

[0295] 所有权利要求和提交的摘要都是说明书的一部分。

[0296] 在本文中使用的任何术语暗示或以其他方式提及行业标准的范围内,以及在适用法律要求标识这种标准的特定版本的范围内,本公开应被理解为指的是在适用专利法下本公开的最早优先日期之前,至少以草案形式(如果较新,则以最终形式为准)发布的该标准的最新版本。

[0297] 虽然示例性实施例已经在附图中示出并在上面描述,但是对于本领域的普通技术人员来说显而易见的是,可以在不背离权利要求中提出的原理和概念的情况下进行许多修改,并且这样的修改不需要包含整个抽象概念。尽管该主题是以特定于结构特征和/或过程

动作的语言描述的,但是应当理解,所附权利要求中定义的主题不一定限于上述权利要求中描述的特定技术特征或动作。不必在每个实施例中呈现或使用给定定义或示例中标识的每个部件或方面或技术效果。相反,所描述的具体特征和动作和效果被公开作为实施权利要求时考虑的示例。

[0298] 不能涵盖整个抽象概念,但在权利要求的含义和等同范围内的所有变更都应被纳入其在法律允许的整个范围内。

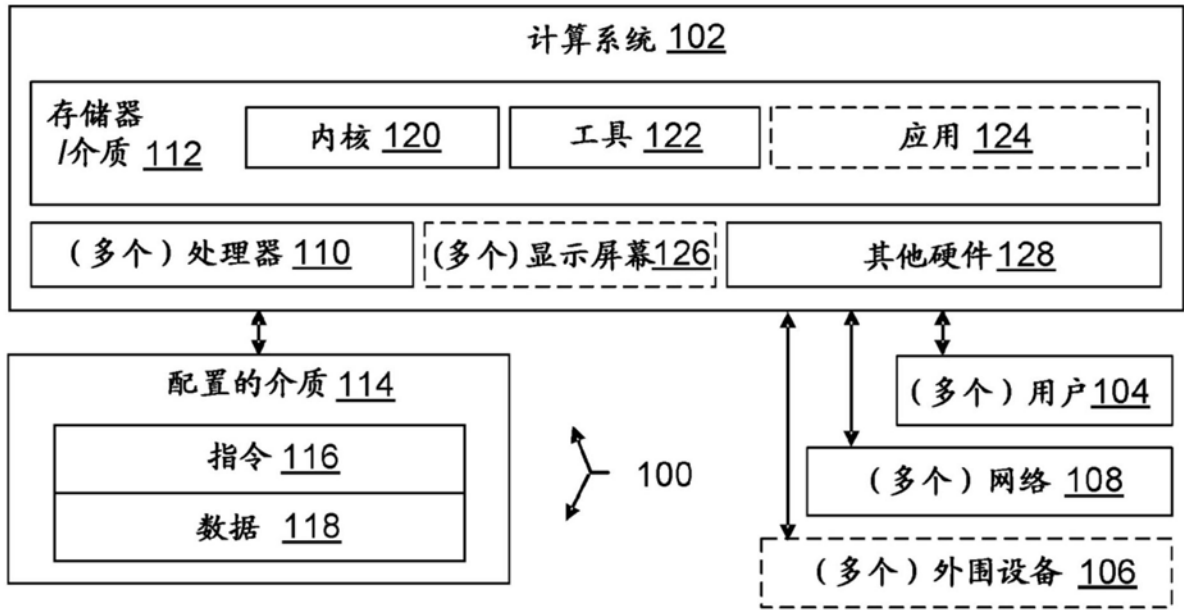


图1

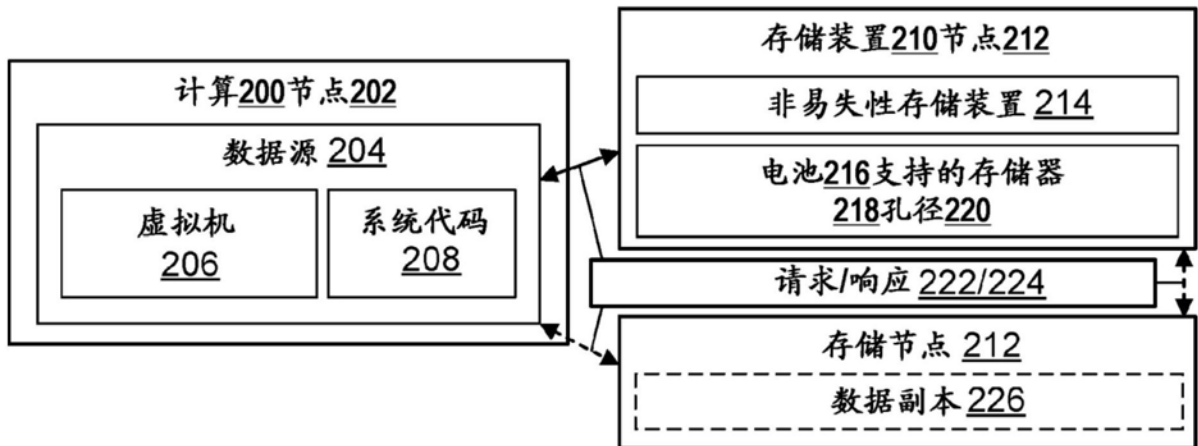


图2

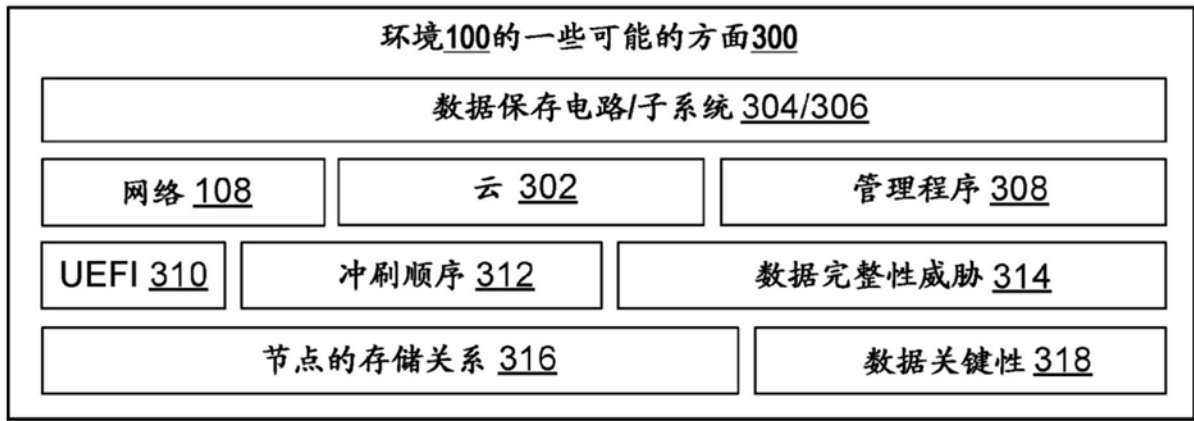


图3

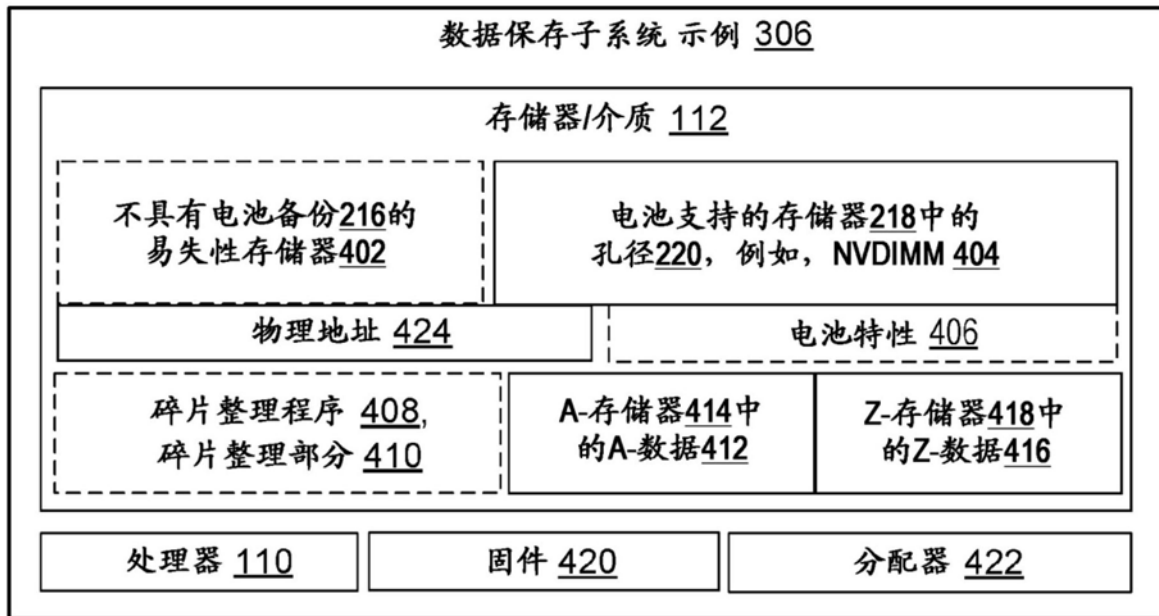


图4

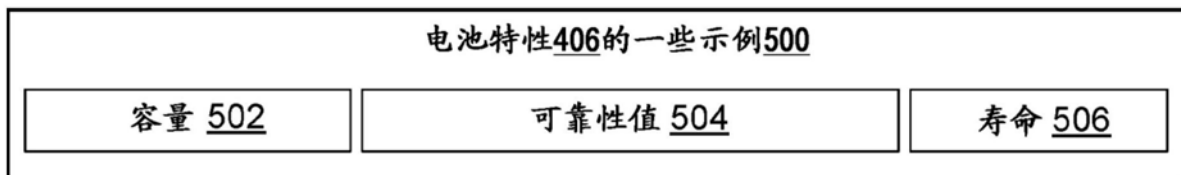


图5

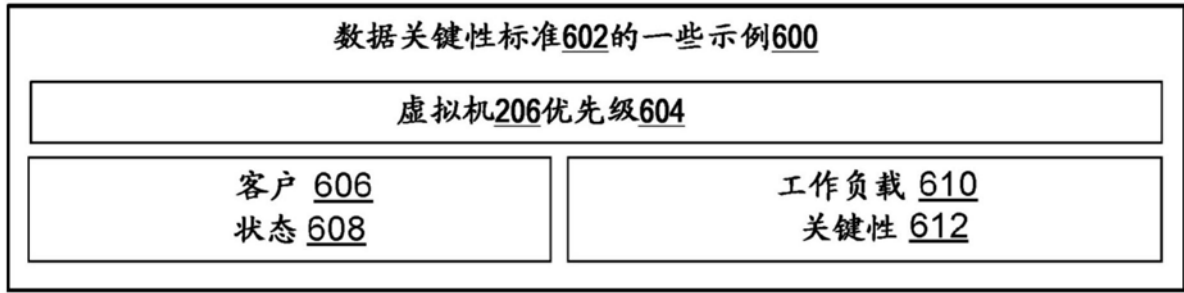


图6

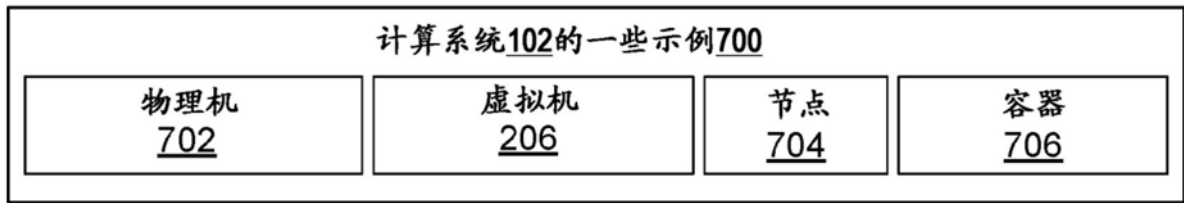


图7



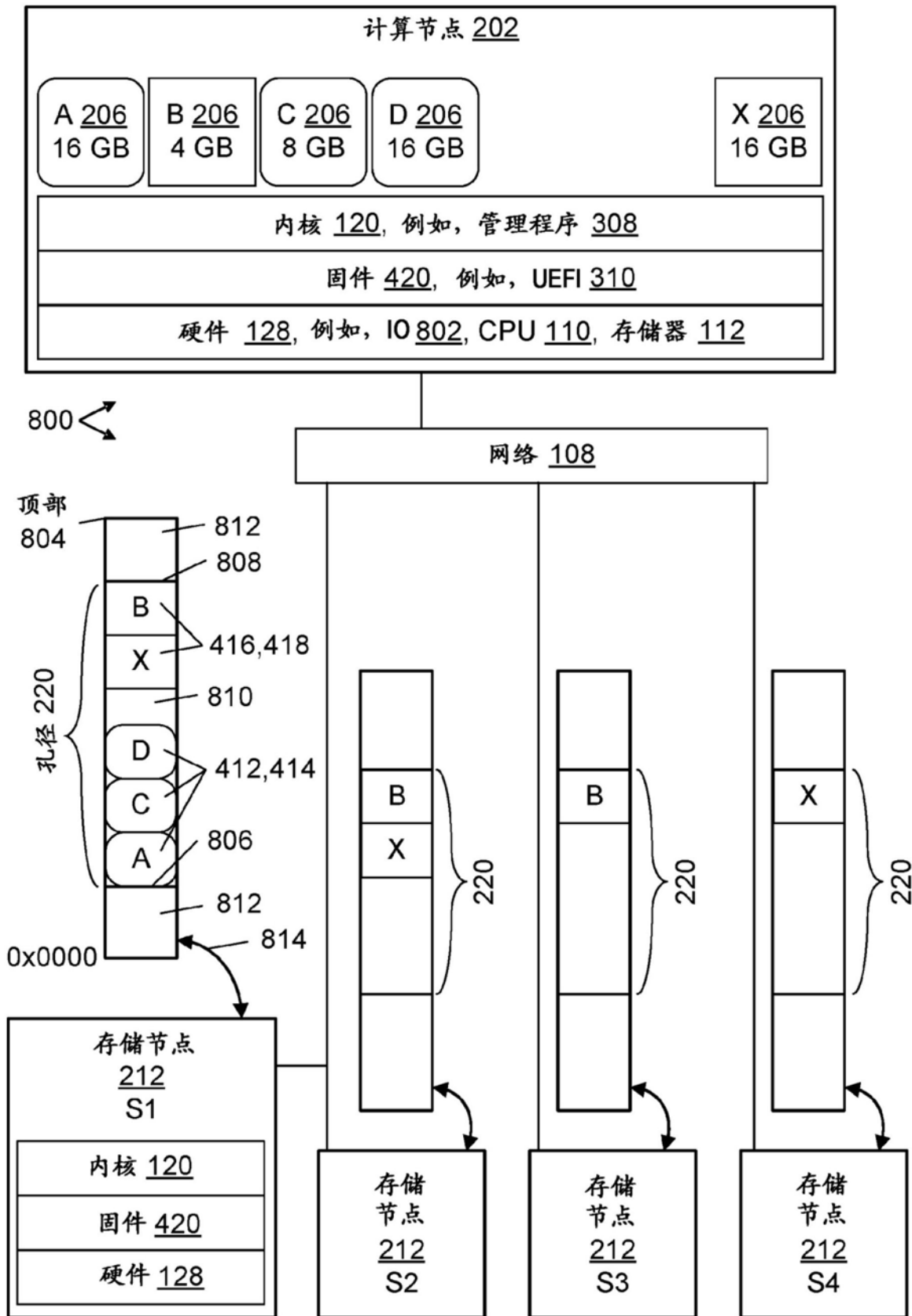


图8

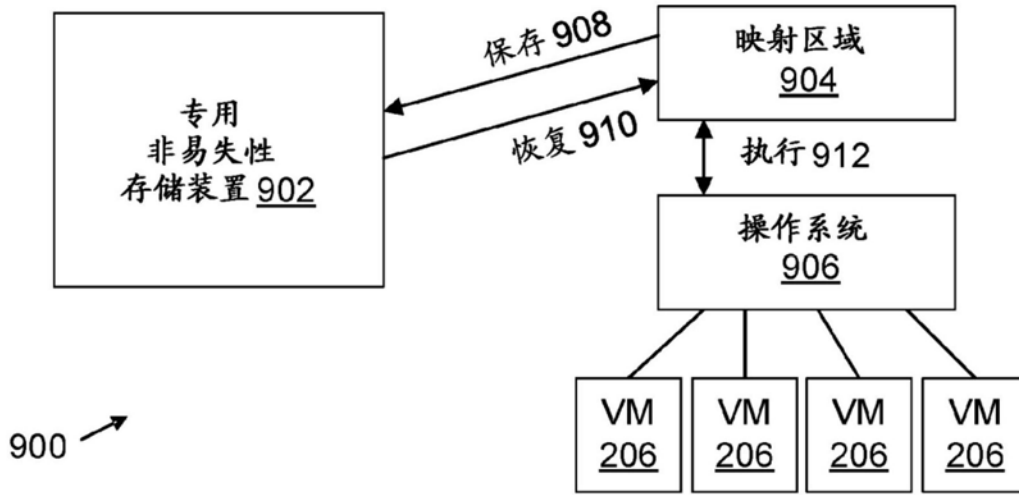


图9

示例数据保存方法 1000



图10



图11