(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
2018 年 3 月 8 日 (08.03.2018)

WIPO | PCT

(10) 国际公布号
WO 2018/040629 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2017/085983

(22) 国际申请日: 2017 年 5 月 25 日 (25.05.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610794448.4 2016年8月31日 (31.08.2016) CN

(71) 申请人: 华 为 技 术 有 限 公 司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 高 峰 (GAO, Feng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 袁 慧 琴 (YUAN, Huiqin); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 北 京 中 博 世 达 专 利 商 标 代 理 有 限 公 司 (BEIJING ZBSD PATENT&TRADEMARK AGENT LTD.); 中国北京市海淀区交大东路31号11号楼8层, Beijing 100044 (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, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR,

WO 2018/040629 A1

(54) **Title:** KEY-VALUE STORAGE METHOD, APPARATUS AND SYSTEM

(54) 发明名称: 键值存储方法、装置及系统

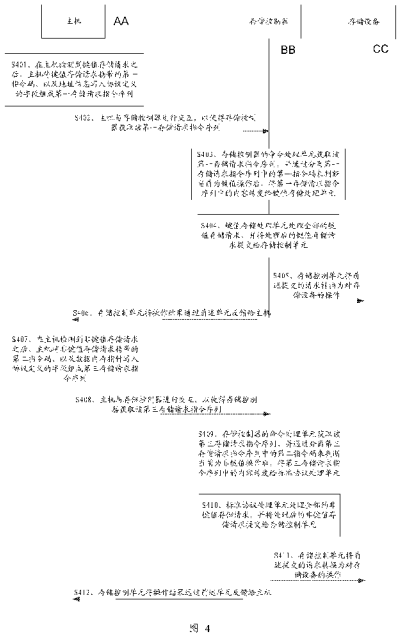

图 4

S401 After a host detects a key-value storage request, the host combining a first instruction code carried in the key-value storage request and a field defined by an address information writing protocol into a first storage request instruction sequence
S402 The host interacting with the storage controller, so that the storage controller acquires the first storage request instruction sequence
S403 A command processing unit of the storage controller acquiring the first storage request instruction sequence, and after determining that a current operation is a key-value operation by separating the first instruction code in the first storage request instruction sequence, forwarding the content in the first storage request instruction sequence to a key-value storage processing unit
S404 The key-value storage processing unit processing all key-value storage requests, and submitting the processed key-value storage requests to a storage control unit
S405, S411 The storage control unit converting the above-mentioned submitted requests into operations on the storage device
S406, S412 The storage control unit feeding back operation results to the host by means of the above-mentioned units
S407 After the host detects a non-key-value storage request, the host combining a second instruction code carried in the non-key-value storage request and a field defined by a data memory pointer writing protocol into a third storage request instruction sequence
S408 The host interacting with the storage controller, so that the storage controller acquires the third storage request instruction sequence
S409 The command processing unit of the storage controller acquiring the third storage request instruction sequence, and after determining that a current operation is a non-key-value operation by separating the second instruction code in the third storage request instruction sequence, forwarding the content in the third storage request instruction sequence to a standard protocol processing unit
S410 The standard protocol processing unit processing all non-key-value storage requests, and submitting the processed non-key-value storage requests to the storage control unit
AA Host
BB Storage controller
CC Storage device

(57) **Abstract:** Provided are a key-value storage method, apparatus and system for at least solving the problem that there is no relevant solution capable of realizing the combination of key-value storage and an efficient storage protocol such as an NVMe protocol at present. The method comprises: after a host detects a key-value storage request, the host combining a first instruction code carried in the key-value storage request and a field defined by an address information writing protocol into a first storage request instruction sequence, wherein the first instruction code is an instruction code defined according to a reserved extension field of the protocol; and

[见续页]

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))。

the host interacting with the storage controller, so that the storage controller acquires the first storage request instruction sequence. The present application is applied to the technical field of storage.

(57) 摘要：本申请实施例提供键值存储方法、装置及系统，以至少解决目前没有相关解决方案能够实现键值存储和NVMe协议这类高效存储协议的结合的问题。方法包括：在主机检测到键值存储请求之后，所述主机将所述键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列，其中，所述第一指令码为根据所述协议的预留扩展字段定义的指令码；所述主机与所述存储控制器进行交互，以使得所述存储控制器获取所述第一存储请求指令序列。本申请适用于存储技术领域。

键值存储方法、装置及系统

本申请要求于 2016 年 08 月 31 日提交中国专利局、申请号为 201610794448.4、发明名称为“键值存储方法、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。

技术领域

本申请涉及存储技术领域，尤其涉及键值存储方法、装置及系统。

背景技术

键值（英文：key-value，缩写：KV）存储，是非关系型数据库（英文：no structured query language，缩写：NoSQL）存储的一种方式，其数据按照键值对的形式进行组织，索引和存储，具有存储语义简单、存储系统扩展性好、数据查询速度快、数据存储量大的特点。

而非易失性存储标准（英文：non-volatile memory express，缩写：NVMe）协议，是目前存储系统的多种存储协议中的一种高效存储协议，使用 NVMe 协议的存储系统和使用传统的小型计算机系统接口（英文：small computer system interface，缩写：SCSI）协议的存储系统相比，由于减少了通用输入输出（英文：input-output，缩写：IO）调度层、SCSI 上层和 SCSI 中间层，因此具有 IO 路径短、时延低、并发处理能力强的特点。

若将键值存储和 NVMe 协议为代表的这类高效存储协议结合起来，将使得存储系统同时具备二者的优势。然而，目前并没有相关解决方案能够实现键值存储和 NVMe 协议这类高效存储协议的结合。

发明内容

本申请实施例提供键值存储方法、装置及系统，用于解决现有技术中无法实现键值存储和一些高效存储协议（如 NVMe 协议）结合的问题。

为解决上述问题，本申请实施例提供如下技术方案：

一方面，本申请实施例提供一种键值存储方法，该方法包括：在主机检测到键值存储请求之后，该主机将该键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列，其中，该第一指令码为根据该协议的预留扩展字段定义的指令码；进而，该主机与该存储控制器进行交互，以使得该存储控制器获取该第一存储请求指令序列。

由于本申请实施例提供的键值存储方法将键值存储操作扩展到存储协议之上，使得主机可以借助存储协议和存储控制器交互以实现键值存储，无需在块层或者文件系统之上进行转换以实现键值存储，因此，还降低了存储系统的 IO 路径时延。

在一种可能的设计中，该主机将该键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列，包括：该主机按照协议定义的字段为第一存储请求指令序列分配内存；该主机将该键值存

储请求携带的第一指令码、以及地址信息写入该内存；进而，该主机与该存储控制器进行交互（如发送通知消息等），以使得该存储控制器获取该第一存储请求指令序列，包括：该主机通知该存储控制器从该内存中读取该第一存储请求指令序列。除了上述让存储控制器获取第一存储请求指令序列该主机也可以向该存储控制器直接发送该第一存储请求指令序列。

其中，主机通知存储控制器从内存中读取第一存储请求指令序列的方式可以节省存储控制器的内存空间。

在一种可能的设计中，该键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；其中，该写数据请求携带的地址信息包括键和值存放的地址信息；该获取数据请求携带的地址信息包括键存放的地址信息；该删除数据请求携带的地址信息包括键存放的地址信息；该废弃数据请求携带的地址信息包括键存放的地址信息。

当然，上述仅是示例性的列举了一些键值存储请求操作，还可能存在其它的键值存储请求操作，本申请实施例对此不作具体限定。

在一些可能的设计中，若该键值存储请求为获取数据请求，则在主机检测到键值存储请求之后，在该主机将该键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列之前，还包括：该主机获取该获取数据请求所请求的值的长度；该主机根据该值的长度为该值分配内存，以使得在该主机与该存储控制器进行交互，以使得该存储控制器获取该第一存储请求指令序列之后，该存储控制器读数据到该主机为该值分配的内存。

也就是说，若该键值存储请求为获取数据请求，则主机首先需要为值分配内存。这样，存储控制器从存储设备中读取的数据才有相应的存储空间。

在一种可能的设计中，该主机获取该获取数据请求所请求的值的长度，包括：该主机将获取键对应的值的长度的指令码、以及该键存放的地址信息写入该协议定义的字段组成第二存储请求指令序列，其中，该获取键对应的值的长度的指令码为根据该协议的预留扩展字段定义的指令码；该主机与该存储控制器进行交互，以使得该存储控制器获取该第二存储请求指令序列；该主机接收该存储控制器发送的该值的长度。

通过上述方式，主机可以获取到该获取数据请求所请求的值的长度。

在一种可能的设计中，该键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，该键值存储请求携带的地址信息通过聚散表的地址信息进行索引，该聚散表中包含该多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

通过将多个单次键值存储请求合并成聚合键值存储请求，并将多个单次键值存储请求中每个单次键值存储请求携带的地址信息通过聚散表进行索引，可以使得在一次键值存储操作流程中同时完成多个单次键值存储操作，提高了键值存储的效率。

在一种可能的设计中，该方法还包括：在该主机检测到非键值存储请求

之后，该主机将该非键值存储请求携带的第二指令码、以及数据内存指针写入该协议定义的字段组成第三存储请求指令序列，其中，该第二指令码为该协议的标准指令码；该主机与该存储控制器进行交互，以使得该存储控制器获取该第三存储请求指令序列。

本申请实施例提供的键值存储方法中，由于主机可以借助存储协议和存储控制器交互以实现非键值存储，因此可以在支持键值存储的同时支持传统块设备存储。

另一方面，本申请实施例提供一种键值存储方法，该方法包括：存储控制器获取第一存储请求指令序列，该第一存储请求指令序列由该主机将键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成，其中，该第一指令码为根据该协议的预留扩展字段定义的指令码；该存储控制器从该第一存储请求指令序列中分离出该第一指令码和该地址信息；该存储控制器根据该第一指令码和该地址信息，对存储设备进行该第一指令码对应的操作。

由于本申请实施例提供的键值存储方法将键值存储操作扩展到存储协议之上，使得主机可以借助存储协议和存储控制器交互以实现键值存储，无需在块层或者文件系统之上进行转换以实现键值存储，因此降低了存储系统的IO路径时延。

在一种可能的设计中，该存储控制器获取第一存储请求指令序列，包括：该存储控制器从主机按照协议定义的字段为该第一存储请求指令序列分配的内存中读取第一存储请求指令序列。或者，除了上述方法外，存储控制器也可以直接接收主机发送的第一存储请求指令序列。

其中，存储控制器从主机按照协议定义的字段为该第一存储请求指令序列分配的内存中读取第一存储请求指令序列的方式可以节省存储控制器的内存空间。

在一种可能的设计中，该键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；其中，该写数据请求携带的地址信息包括键和值存放的地址；该获取数据请求携带的地址信息包括键存放的地址信息；该删除数据请求携带的地址信息包括键存放的地址信息；该废弃数据请求携带的地址信息包括键存放的地址信息。

当然，上述仅是示例性的列举了一些键值存储请求操作，还可能存在其它的键值存储请求操作，本申请实施例对此不作具体限定。

在一种可能的设计中，若该键值存储请求为获取数据请求，则在该存储控制器获取第一存储请求指令序列之前，还包括：该存储控制器获取该获取数据请求所请求的值的长度；该存储控制器向该主机发送该值的长度，以使得该主机根据该值的长度为该值分配内存；进而，该存储控制器根据该第一指令码和该地址信息，对存储设备进行该第一指令码对应的操作，包括：该存储控制器根据该第一指令码和该地址信息，读数据到该主机为该值分配的内存。

也就是说，若该键值存储请求为获取数据请求，则主机首先需要为值分配内存。这样，存储控制器从存储设备中读取的数据才有相应的存储空间。

在一种可能的设计中，该存储控制器获取该获取数据请求所请求的值的长度，包括：该存储控制器获取第二存储请求指令序列，该第二存储请求指令序列由该主机将获取键对应的值的长度的指令码、以及该键存放的地址信息写入该协议定义的字段组成，其中，该获取键对应的值的长度的指令码为根据该协议的预留扩展字段定义的指令码；该存储控制器从该第二存储请求指令序列中分离出该获取键对应的值的长度的指令码、以及该键存放的地址信息；该存储控制器根据该获取键对应的值的长度的指令码、以及该键存放的地址信息，从该存储设备中获取该值的长度。

其中，存储控制器获取第二存储请求指令序列的方式可参考上述存储控制器获取第一存储请求指令序列的方式，此处不再赘述。

通过上述方式，主机可以获取到该获取数据请求所请求的值的长度。

在一种可能的设计中，该键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，该键值存储请求携带的地址信息通过聚散表的地址信息进行索引，该聚散表中包含该多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

通过将多个单次键值存储请求合并成聚合键值存储请求，并将多个单次键值存储请求中每个单次键值存储请求携带的地址信息通过聚散表进行索引，可以使得在一次键值存储操作流程中同时完成多个单次键值存储操作，提高了键值存储的效率。

在一种可能的设计中，该方法还包括：该存储控制器获取第三存储请求指令序列，该第三存储请求指令序列由该主机将非键值存储请求携带的第二指令码、以及数据内存指针写入该协议定义的字段组成，其中，该第二指令码为该协议的标准指令码；该存储控制器从该第三存储请求指令序列中分离出该第二指令码和该数据内存指针；该存储控制器根据该第二指令码和该数据内存指针，对该存储设备进行该第二指令码对应的操作。

其中，存储控制器获取第三存储请求指令序列的方式可参考上述存储控制器获取第一存储请求指令序列的方式，此处不再赘述。

本申请实施例提供的键值存储方法中，由于主机可以借助存储协议和存储控制器交互以实现非键值存储，因此可以在支持键值存储的同时支持传统块设备存储。

又一方面，本申请实施例提供一种主机，该主机包括：处理模块和通信模块；该处理模块，用于在检测到键值存储请求之后，将该键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列，其中，该第一指令码为根据该协议的预留扩展字段定义的指令码；该通信模块，用于与该存储控制器进行交互，以使得该存储控制器获取该第一存储请求指令序列。

在一种可能的设计中，该处理模块具体用于：按照协议定义的字段为第

一存储请求指令序列分配内存；将该键值存储请求携带的第一指令码、以及地址信息写入该内存；该通信模块具体用于：通知该存储控制器从该内存中读取该第一存储请求指令序列；或者，向该存储控制器发送该第一存储请求指令序列。

在一种可能的设计中，该键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；其中，该写数据请求携带的地址信息包括键和值存放的地址信息；该获取数据请求携带的地址信息包括键存放的地址信息；该删除数据请求携带的地址信息包括键存放的地址信息；该废弃数据请求携带的地址信息包括键存放的地址信息。

在一种可能的设计中，若该键值存储请求为获取数据请求，则该处理模块，还用于在检测到键值存储请求之后，将该键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列之前，获取该获取数据请求所请求的值的长度；根据该值的长度为该值分配内存，以使得在该通信模块与该存储控制器进行交互，以使得该存储控制器获取该第一存储请求指令序列之后，该存储控制器读数据到该处理模块为该值分配的内存。

在一种可能的设计中，该处理模块具体用于：将获取键对应的值的长度的指令码、以及该键存放的地址信息写入该协议定义的字段组成第二存储请求指令序列，其中，该获取键对应的值的长度的指令码为根据该协议的预留扩展字段定义的指令码；通过该通信模块与该存储控制器进行交互，以使得该存储控制器获取该第二存储请求指令序列；通过该通信模块接收该存储控制器发送的该值的长度。

在一种可能的设计中，该键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，该键值存储请求携带的地址信息通过聚散表的地址信息进行索引，该聚散表中包含该多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

在一种可能的设计中，该处理模块，还用于在检测到非键值存储请求之后，将该非键值存储请求携带的第二指令码、以及数据内存指针写入该协议定义的字段组成第三存储请求指令序列，其中，该第二指令码为该协议的标准指令码；该通信模块，还用于与该存储控制器进行交互，以使得该存储控制器获取该第三存储请求指令序列。

由于本申请实施例提供的主机可用于执行上述方法实施例中主机所执行的功能，因此其所能获得的技术效果可参考上述方法实施例中的相关描述，此处不再赘述。

又一方面，本申请实施例提供一种存储控制器，该存储控制器包括：前端通信模块、后端通信模块、处理模块和控制模块；该前端通信模块，用于从主机获取第一存储请求指令序列，该第一存储请求指令序列由该主机将键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成，其中，该第一指令码为根据该协议的预留扩展字段定义的指令码；该处理模

块，用于从该第一存储请求指令序列中分离出该第一指令码和该地址信息；该控制模块，用于根据该第一指令码和该地址信息，通过该后端通信模块对存储设备进行该第一指令码对应的操作。

在一种可能的设计中，该前端通信模块具体用于：从该主机按照协议定义的字段为该第一存储请求指令序列分配的内存中读取该第一存储请求指令序列；或者，接收该主机发送的第一存储请求指令序列。

在一种可能的设计中，该键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；其中，该写数据请求携带的地址信息包括键和值存放的地址；该获取数据请求携带的地址信息包括键存放的地址信息；该删除数据请求携带的地址信息包括键存放的地址信息；该废弃数据请求携带的地址信息包括键存放的地址信息。

在一种可能的设计中，若该键值存储请求为获取数据请求，则该处理模块，还用于在该前端通信模块获取第一存储请求指令序列之前，获取该获取数据请求所请求的值的长度；该通信模块，还用于向该主机发送该值的长度，以使得该主机根据该值的长度为该值分配内存；该控制模块具体用于：根据该第一指令码和该地址信息，读数据到该主机为该值分配的内存。

在一种可能的设计中，该处理模块具体用于：

通过该前端通信模块获取第二存储请求指令序列，该第二存储请求指令序列由该主机将获取键对应的值的长度的指令码、以及该键存放的地址信息写入该协议定义的字段组成，其中，该获取键对应的值的长度的指令码为根据该协议的预留扩展字段定义的指令码；从该第二存储请求指令序列中分离出该获取键对应的值的长度的指令码、以及该键存放的地址信息；根据该获取键对应的值的长度的指令码、以及该键存放的地址信息，通过该控制模块和该后端通信模块从该存储设备中获取该值的长度。

在一种可能的设计中，该键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，该键值存储请求携带的地址信息通过聚散表的地址信息进行索引，其中，该聚散表中包含该多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

在一种可能的设计中，该前端通信模块，还用于从该主机获取第三存储请求指令序列，该第三存储请求指令序列由该主机将非键值存储请求携带的第二指令码、以及数据内存指针写入该协议定义的字段组成，其中，该第二指令码为该协议的标准指令码；该处理模块，还用于从该第三存储请求指令序列中分离出该第二指令码和该数据内存指针；该控制模块，还用于根据该第二指令码和该数据内存指针，通过该后端通信模块对该存储设备进行该第二指令码对应的操作。

由于本申请实施例提供的存储控制器可用于执行上述方法实施例中存储控制器所执行的功能，因此其所能获得的技术效果可参考上述方法实施例中的相关描述，此处不再赘述。

基于上述各方面，一种可能的设计中，上述各方面任一方面中所述的协

议包括非易失性存储标准NVMe协议；其中，该NVMe协议定义0-63字节为存储请求指令序列的字段。

由于本申请实施例可以将键值存储和NVMe协议这类高效存储协议结合起来，因此可以使得存储系统同时具备键值存储存储语义简单、存储系统扩展性好、数据查询速度快、数据存储量大的优势，以及NVMe协议IO路径短、时延低、并发处理能力强的优势。

基于上述各方面，一种可能的设计中，上述各方面任一方面中所述的协议包括小型计算机系统接口SCSI协议。

由于SCSI协议为通用的存储协议，因此将键值存储与SCSI协议结合起来，更具有通用性。

当然，上述的协议还可以为其它存储协议，本申请实施例对此不作具体限定。

基于上述各方面，一种可能的设计中，上述各方面任一方面中所述的指令序列的形式可以是队列，也可以是数据包等其它形式，本申请实施例对此不作具体限定。

又一方面，本申请实施例提供了一种主机，该主机可以实现上述方法实施例中主机所执行的功能，该功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。该硬件或软件包括一个或多个上述功能相应的模块。

在一种可能的设计中，该主机的结构中包括处理器和通信接口，该处理器被配置为支持该主机执行上述方法中相应的功能。该通信接口用于支持该主机与其他网元之间的通信。该主机还可以包括存储器，该存储器用于与处理器耦合，其保存该主机必要的程序指令和数据。

又一方面，本申请实施例提供了一种存储控制器，该存储控制器可以实现上述方法实施例中存储控制器所执行的功能，该功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。该硬件或软件包括一个或多个上述功能相应的模块。

在一种可能的设计中，该存储控制器的结构中包括处理器和通信接口，该处理器被配置为支持该存储控制器执行上述方法中相应的功能。该通信接口用于支持该存储控制器与其他网元之间的通信。该存储控制器还可以包括存储器，该存储器用于与处理器耦合，其保存该存储控制器必要的程序指令和数据。

又一方面，本申请实施例提供了一种键值存储系统，该键值存储系统包括存储设备，以及上述方面所述的主机和存储控制器。

再一方面，本申请实施例提供了一种计算机存储介质，用于储存上述主机所用的计算机软件指令，其包含用于执行上述方面所设计的程序。

再一方面，本申请实施例提供了一种计算机存储介质，用于储存为上述存储控制器所用的计算机软件指令，其包含用于执行上述方面所设计的程序。

附图说明

图 1 为现有键值存储的两种主要实现方式；

图 2 为本申请实施例提供的键值存储系统的架构示意图；

图 3 为本申请实施例提供的本申请实施例提供的键值存储方法的操作模块示意图；

图 4 为本申请实施例提供的键值存储方法的流程示意图；

图 5 为本申请实施例提供的键值存储与 NVMe 协议相结合以实现键值存储的操作流程示意图；

图 6 为本申请实施例提供的写数据请求时的键值存储流程示意图；

图 7 为本申请实施例提供的获取数据请求时的键值存储流程示意图；

图 8 为本申请实施例提供的删除数据请求时的键值存储流程示意图；

图 9 为本申请实施例提供的废弃数据请求时的键值存储流程示意图；

图 10 为本申请实施例提供的聚合删除数据请求时的键值存储流程示意图；

图 11 为本申请实施例提供的非键值存储相关的标准 NVMe 设备操作流程示意图；

图 12 为本申请实施例提供的键值存储与 SCSI 协议相结合以实现键值存储的操作流程示意图；

图 13 为本申请实施例提供的主机的结构示意图；

图 14 为本申请实施例提供的存储控制器的结构示意图。

具体实施方式

如图 1 所示，为现有键值存储的两种主要实现方式。

其中，在方式一中，硬件使用 SCSI 设备，SCSI 设备通过 SCSI 底层驱动、SCSI 中层、SCSI 上层、IO 调度层和块设备层对外提供块设备服务。软件在块设备层或者文件系统之上建立中间件，在中间件中完成键值操作和块设备或文件系统操作的转换，从而对用户空间的应用提供键值存储的服务。然而，在该方式中，存在多层存储协议栈转换，因此 IO 路径时延较大。

在方式二中，硬件使用支持键值存储的设备，键值存储设备通过键值存储驱动层与中间件进行通信，中间件将用户存储转换为键值存储操作，从而对用户空间的应用提供键值存储服务。然而，在该方式中，专用的键值存储设备无法支持传统块设备存储，应用范围具有局限性。

本申请实施例提供一种键值存储方法，能够将键值存储操作扩展到存储协议之上，进而，不仅可以降低存储系统的 IO 路径时延，还可以同时支持传统块设备存储和键值存储。

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行描述。

需要说明的是，为了便于清楚描述本申请实施例的技术方案，在本申请的实施例中，采用了“第一”、“第二”等字样对功能和作用基本相同的相同项或相似项进行区分，本领域技术人员可以理解“第一”、“第二”等字样并不对数量和执行

次序进行限定。

需要说明的是，本文中的“/”表示或的意思，例如，A/B 可以表示 A 或 B；本文中的“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。“多个”是指两个或多于两个。

如本申请所使用的术语“组件”、“模块”、“系统”等等旨在指代计算机相关实体，该计算机相关实体可以是硬件、固件、硬件和软件的结合、软件或者运行中的软件。例如，组件可以是，但不限于是：在处理器上运行的处理、处理器、对象、可执行文件、执行中的线程、程序和/或计算机。作为示例，在计算设备上运行的应用和该计算设备都可以是组件。一个或多个组件可以存在于执行中的过程和/或线程中，并且组件可以位于一个计算机中以及/或者分布在两个或更多个计算机之间。此外，这些组件能够从在其上具有各种数据结构的各种计算机可读介质中执行。这些组件可以通过诸如根据具有一个或多个数据分组(例如，来自一个组件的数据，该组件与本地系统、分布式系统中的另一个组件进行交互和/或以信号的方式通过诸如互联网之类的网络与其它系统进行交互)的信号，以本地和/或远程过程的方式进行通信。

需要说明的是，本申请实施例中，“示例性的”或者“例如”等词用于表示作例子、例证或说明。本申请实施例中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言，使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

需要说明的是，本申请实施例中，除非另有说明，“多个”的含义是指两个或两个以上。例如，多个数据包是指两个或两个以上的数据包。

需要说明的是，本申请实施例中，“的(英文：of)”，“相应的（英文：corresponding，relevant）”和“对应的(英文：corresponding)”有时可以混用，应当指出的是，在不强调其区别时，其所要表达的含义是一致的。

需要说明的是，本申请实施例描述的网络架构以及业务场景是为了更加清楚的说明本申请实施例的技术方案，并不构成对于本申请实施例提供的技术方案的限定，本领域普通技术人员可知，随着网络架构的演变和新业务场景的出现，本申请实施例提供的技术方案对于类似的技术问题，同样适用。

如图 2 所示，为本申请实施例提供的键值存储系统的架构示意图。该键值存储系统由主机 20、存储控制器 21 以及存储设备 22 组成。其中，主机 20 通过存储控制器 21 对存储设备 22 中的存储介质进行键值存储操作；存储控制器 21 用于处理键值存储请求并对存储设备 22 进行数据存取操作。

具体的，如图 2 所示，主机 20 可以包括：主机底板 201、以及部署在主机底板 201 上的中央处理单元（英文：central processing unit，缩写：CPU）202、内存 203、桥片 204 和主机通信接口 205。

其中，内存 203 中存储了该主机 20 运行时必要的程序指令和数据；CPU202 用于处理该主机 20 的程序指令；桥片 204 用于连接主机底板 201 上的各类外接设备（未画出）；主机通信接口 205 是主机 20 用于和存储控制器 21 之间连接的总线接口，

用于完成主机 20 与存储控制器 21 之间的通信。

存储控制器 21 可以包括：前端通信接口 211、CPU212、内存 213、以及后端通信接口 214。

其中，内存 213 用于存储存储控制器 21 运行时必要的程序指令和数据；CPU212 用于处理存储控制器 21 运行的指令和运算；前端通信接口 211 是存储控制器 21 用于和主机 20 之间连接的总线接口，用于完成存储控制器 21 与主机 20 之间的通信；后端通信接口 214 是存储控制器 21 用于和存储设备 22 之间连接的总线接口，用于完成存储控制器 21 与存储设备 22 之间的通信。

需要说明的是，本申请实施例中的存储控制器 21 以及存储设备 22 可能独立部署，也可能集成在同一设备上，本申请对此不作具体限定。

下面将基于该键值存储系统，对本申请实施例提供的键值存储方法进行详细介绍。首先，图 3 和图 4 分别为本申请实施例提供的键值存储方法的操作模块示意图和相应的流程示意图，该键值存储方法具体可以包括：

S401、在主机检测到键值存储请求之后，主机将键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列（如表一）。

其中，该第一指令码为根据协议的预留扩展字段定义的指令码。协议的预留扩展字段具体是指协议预留的用于协议扩展或用于用户自定义的字段。

S402、主机与存储控制器进行交互，以使得存储控制器获取该第一存储请求指令序列。

S403、存储控制器的命令处理单元获取该第一存储请求指令序列，并通过分离第一存储请求指令序列中的第一指令码来判断当前为键值操作后，将第一存储请求指令序列中的内容转发给键值存储处理单元。

S404、键值存储处理单元处理全部的键值存储请求，并将处理后的键值存储请求提交给存储控制单元。

S405、存储控制单元将前述提交的请求转换为对存储设备的操作。

S406、存储控制单元将操作结果通过前述单元反馈给主机。

表一

指令其它字段	第一指令码	地址信息	指令其它字段

S407、在主机检测到非键值存储请求之后，主机将非键值存储请求携带的第二指令码、以及数据内存指针写入协议定义的字段组成第三存储请求指令序列（如表二）。

其中，该第二指令码为协议的标准指令码。

S408、主机与存储控制器进行交互，以使得存储控制器获取该第三存储请求指令序列。

S409、存储控制器的命令处理单元获取该第三存储请求指令序列，并通过分离第三存储请求指令序列中的第二指令码来判断当前为非键值操作后，将第三存储请求指令序列中的内容转发给标准协议处理单元。

S410、标准协议处理单元处理全部的非键值存储请求，并将处理后的非键值存

储请求提交给存储控制单元。

S411、存储控制单元将前述提交的请求转换为对存储设备的操作。

S412、存储控制单元将操作结果通过前述单元反馈给主机。

表二

指令其它字段	第二指令码	数据内存指针	指令其它字段

具体的，上述的协议可以是 NVMe 协议，也可以是 SCSI 协议等其它存储协议，本申请实施例对此不作具体限定。

具体的，上述的指令序列的形式可以是队列，也可以是数据包等其它形式，本申请实施例对此不作具体限定。

具体的，上述的键值存储操作可以是写数据、获取数据、删除数据或者废弃数据等操作，本申请实施例对此不作具体限定。

本申请实施例提供的键值存储方法将键值存储操作扩展到存储协议之上，进而，一方面，由于主机可以借助存储协议和存储控制器交互以实现键值存储，无需在块层或者文件系统之上进行转换以实现键值存储，因此降低了存储系统的 IO 路径时延；另一方面，由于主机可以借助存储协议和存储控制器交互以实现非键值存储，因此可以同时支持传统块设备存储。

可选的，步骤 S401 具体可以包括：

在主机检测到键值存储请求之后，主机按照协议定义的字段为第一存储请求指令序列分配内存；

主机将该键值存储请求携带的第一指令码、以及地址信息写入该内存。

进而，步骤 S402 具体可以包括：

主机通知存储控制器从该内存中读取该第一存储请求指令序列；

或者，主机向存储控制器发送该第一存储请求指令序列。

其中，主机通知存储控制器从内存中读取第一存储请求指令序列的方式可以节省存储控制器的内存空间。

需要说明的是，本申请实施例仅是示例性的提供两种主机与存储控制器进行交互，以使得存储控制器获取该第一存储请求指令序列的方式，当然，主机与存储控制器还可能通过其它交互方式以使得存储控制器获取该第一存储请求指令序列，本申请实施例对此不作具体限定。

下面将结合具体的协议以及具体的键值存储操作或者非键值存储操作对本申请实施例提供的键值存储方法进一步说明。

如图 5 所示，为键值存储与 NVMe 协议相结合以实现键值存储的操作流程示意图。

首先，对图 5 中的相关单元进行简要介绍：

一、主机接口：

主机与存储控制器通过主机接口进行连接，主机通过主机接口与存储控制器进行指令、地址和数据的交互，本申请实施例中主机接口为总线和接口标准（英文：peripheral component interface express，缩写：PCIe）接口；

二、主机软件模块：

1）应用层：主机应用程序或者是存储客户端软件。

2）传统应用层：基于传统文件系统或者块设备接口的主机应用程序。

3）中间件：键值存储中间件，对主机应用提供键值存储接口，并将存储请求传递给 NVMe 驱动层。

4）文件系统：例如 EXT3、EXT4、FAT32 等。

5）块层：操作系统对块存储设备的抽象层，一般文件系统都构建于这层之上。

6）NVMe 驱动层：主机操作系统通过驱动软件和 NVMe 存储控制器进行数据传输和命令交互。

7）NVMe 命令转换单元：位于 NVMe 驱动层，用于将键值（Key-Value）存储的扩展指令、Key 和/或 Value 存放的地址等信息填入 NVMe 发送队列，并将 NVMe 发送队列提交给 NVMe 驱动层，由 NVMe 驱动层下发给 NVMe 存储控制器。

三、NVMe 存储控制器：

1）NVMe 命令处理单元：分析 NVMe 控制器接收到的 NVMe 发送队列中的指令码，并根据不同的指令将 NVMe 发送队列分发给键值存储处理单元或者 NVMe 操作处理单元。

2）键值存储处理单元：处理全部的键值操作请求，并将处理后的请求提交给存储控制单元。

3）NVMe 操作处理单元：处理标准 NVMe 协议操作请求。

4）存储控制单元：将前述提交的请求转换为对存储设备的操作，并将操作结果反馈给对存储设备执行数据的存储操作。

四、存储设备：

存储设备中的存储介质包括动态随机存取存储器（英文：dynamic random access memory，缩写：DRAM）、非易失性随机访问存储器（英文：non-volatile random access memory，缩写：NVRAM）、NAND 或者其它存储器件。

其次，给出 NVMe 协议中 NVMe IO 队列操作定义的指令码，如表三所示：

表三

操作码（07）	操作码（06:02）	操作码（01:00）	操作码[2]	O/M[1]	命令[3]
标准命令	功能	数据传送			
0b	000 00b	00b	00h	M	刷新:Flush
0b	000 00b	01b	01h	M	写：Write
0b	000 00b	10b	02h	M	读：Read
0b	000 01b	00b	04h	O	不可纠正的写：Write Uncorrectable
0b	000 01b	01b	05h	O	比较

					Compare
0b	000 10b	00b	08h	O	写零：Write Zeroes
0b	000 10b	01b	09h	O	数据管理：Dataset Management
0b	000 11b	01b	0Dh	O4	预先注册：Reservation Register
0b	000 11b	10b	0Eh	O4	预先报告：Reservation Report
0b	001 00b	01b	11h	O4	预先获得：Reservation Acquire
0b	001 01b	01b	15h	O4	预留释放：Reservation Release
	特定供应商				
1b	na	na	80h-FFh	O	特定供应商：

本申请实施例中，根据 NVMe 协议的预留扩展字段定义键值存储请求携带的第一指令码，如表四所示：

表四

操作码（07）	操作码（06:02）	操作码（01:00）	操作码[2]	命令[3]
标准命令	功能	数据传送		
1b	na	na	80h	单次写数据：Write
1b	na	na	81h	单次获取Key对应Value的长度：GetLength
1b	na	na	82h	单次获取

				数据：Read
1b	na	na	83h	单次删除数据：Delete
1b	na	na	84h	单次废弃数据：TRIM
1b	na	na	85h	聚合写数据：Write
1b	na	na	86h	聚合获取Key对应Value的长度：GetLength
1b	na	na	87h	聚合获取数据：Read
1b	na	na	88h	聚合删除数据：Delete
1b	na	na	89h	聚合废弃数据：TRIM

需要说明的是，上述表四仅是示例性的提供一种根据 NVMe 协议的预留扩展字段定义键值存储请求携带的第一指令码的方式，当然，根据 NVMe 协议的预留扩展字段定义键值存储请求携带的第一指令码不限于上述方式，比如，可以在 90h-99h 字段上定义上述键值存储操作，本申请实施例对此不作具体限定。

如上所述，本实施例中键值（Key-Value）存储操作包括但不限于：写数据 Put (String Key, String Value)、获取数据：Get (String Key)、删除数据：Delete (String Key)、废弃数据：TRIM (String Key)，本申请实施例对此不作具体限定。

下面将结合图 5 对上述键值存储操作进行详细的描述。

示例一：

当键值存储请求为单次写数据请求时，如图 6 所示，本申请实施例提供的键值存储方法包括步骤 S601-S611：

S601、应用层调用中间件写接口：Put (String Key, String Value)。

S602、中间件提交写数据请求、Key 和 Value 存放的地址信息到 NVMe 驱动层。

S603、NVMe 驱动层的 NVMe 命令转换单元写写数据指令 80h、Key 和 Value 存放的地址信息到 NVMe 发送队列。

具体的，在 NVMe 协议中，定义 0-63 字节为 NVMe 发送队列对应的字段。该步骤中 NVMe 发送队列的命令格式如图 6 中的发送队列命令格式。其中，字节 03:00（即第 0 个字节到第 3 个字节）为指令码对应的字节；字节 23：04（即第 4 个字节到第 23 个字节）为 Command Dword 1-6（即命令字的第 1-6 个 4 字节）；字节 39：24（即第 24 个字节到第 39 个字节）为 Key 和 Value 存放的地址信息对应的字节；字节 63：40（即第 40 个字节到第 63 个字节）为 Command Dword 10-15（即命令字

的第 10-15 个 4 字节）。

需要说明的是，本申请实施例中的 NVMe 发送队列是上述指令序列的一种具体形式，当然，如上所述，本申请实施例中的指令序列的形式也可以是数据包等其它形式，本申请实施例对此不作具体限定。

S604、NVMe 驱动层通知 NVMe 存储控制器读 NVMe 发送队列。

S605、NVMe 存储控制器的 NVMe 命令处理单元通过直接数据存取（英文：direct memory access ，缩写：DMA）读 NVMe 发送队列。

S606、NVMe 命令处理单元分离 NVMe 队列中的写数据指令、key 和 Value 存放的地址信息，提交写数据请求到键值存储处理单元。

S607、键值存储处理单元处理 Key 和 Value 存放的地址信息，转换写数据请求为对应存储设备的写数据请求,并提交写数据请求到存储控制单元。

S608、存储控制单元根据 Key 和 Value 存放的地址信息，写数据到存储设备。

S609、存储控制单元返回状态信息到键值存储处理单元。

具体的，这里的状态信息是指是否写数据成功的信息。

S610、键值存储处理单元及前述单元依次传递状态信息到中间件。

具体的，结合图 5 可知，此处的前述单元具体包括：NVMe 存储控制器的 NVMe 命令处理单元、主机的 NVMe 驱动层的 NVMe 命令转换单元。

S611、中间件返回状态信息到应用层。

至此，当键值存储请求为单次写数据请求时，键值存储过程结束。

示例二：

当键值存储请求为单次获取数据请求时，如图 7 所示，本申请实施例提供的键值存储方法包括步骤 S701-S721：

S701、应用层调用中间件读操作接口：Get (String Key)。

S702、中间件提交获取 Key 对应 Value 长度的请求到 NVMe 驱动层。

S703、NVMe 驱动层的 NVMe 命令转换单元写获取 Value 长度的指令 81h、Key 存放的地址信息到 NVMe 发送队列。

具体的，在 NVMe 协议中，定义 0-64 字节为 NVMe 发送队列对应的字段。该步骤中 NVMe 发送队列的命令格式如图 7 中的发送队列命令格式 1。其中，字节 03:00（即第 0 个字节到第 3 个字节）为指令码对应的字节；字节 23：04（即第 4 个字节到第 23 个字节）为 Command Dword 1-6（即命令字的第 1-6 个 4 字节）；字节 39：24（即第 24 个字节到第 39 个字节）为 Key 存放的地址信息对应的字节；字节 63：40（即第 40 个字节到第 63 个字节）为 Command Dword 10-15（即命令字的第 10-15 个 4 字节）。

需要说明的是，本申请实施例中的 NVMe 发送队列是上述指令序列的一种具体形式，当然，如上所述，本申请实施例中的指令序列的形式也可以是数据包等其它形式，本申请实施例对此不作具体限定。

S704、NVMe 驱动层通知 NVMe 存储控制器读 NVMe 发送队列。

S705、NVMe 存储控制器的 NVMe 命令处理单元通过 DMA 读 NVMe 发送队列。

S706、NVMe 命令处理单元分离 NVMe 队列中的获取 Value 长度的指令、key 存

放的地址信息，提交读操作请求到键值存储处理单元。

S707、键值存储处理单元处理 Key 存放的地址信息，转换读操作请求为对应存储设备的读操作请求,并提交读操作请求到存储控制单元。

S708、存储控制单元根据 Key 存放的地址信息，从存储设备获取对应 Value 的长度。

S709、存储控制单元提交 Value 的长度信息到键值存储处理单元。

S710、键值存储处理单元及前述单元依次 Value 的长度信息到中间件。

具体的，结合图 5 可知，此处的前述单元具体包括：NVMe 存储控制器的 NVMe 命令处理单元、主机的 NVMe 驱动层的 NVMe 命令转换单元。

S711、中间件根据 Value 的长度分配 Value 在主机端存放的内存空间。

S712、中间件提交获取 Value 的请求、Value 存放的地址信息到 NVMe 驱动层。

S713、NVMe 驱动层的 NVMe 命令转换单元写获取 Value 的指令 82h、key 存放的地址信息到 NVMe 发送队列。

具体的，在 NVMe 协议中，定义 0-63 字节为 NVMe 发送队列对应的字段。该步骤中 NVMe 发送队列的命令格式如图 7 中的发送队列命令格式 2。其中，字节 03:00（即第 0 个字节到第 3 个字节）为指令码对应的字节；字节 23：04（即第 4 个字节到第 23 个字节）为 Command Dword 1-6（即命令字的第 1-6 个 4 字节）；字节 39：24（即第 24 个字节到第 39 个字节）为 key 存放的地址信息对应的字节；字节 63：40（即第 40 个字节到第 63 个字节）为 Command Dword 10-15（即命令字的第 10-15 个 4 字节）。

需要说明的是，本申请实施例中的 NVMe 发送队列是上述指令序列的一种具体形式，当然，如上所述，本申请实施例中的指令序列的形式也可以是数据包等其它形式，本申请实施例对此不作具体限定。

S714、NVMe 驱动层通知 NVMe 存储控制器读 NVMe 发送队列。

S715、NVMe 存储控制器的 NVMe 命令处理单元通过 DMA 读 NVMe 发送队列。

S716、NVMe 命令处理单元分离 NVMe 队列中的获取 Value 的指令、key 存放的地址信息，提交获取 Value 的请求到键值存储处理单元。

S717、键值存储处理单元处理 Value 存放的地址信息，转换获取 Value 的请求为对应存储设备的获取 Value 的请求,并提交获取 Value 的请求到存储控制单元。

S718、存储控制单元通过 DMA 方式读数据到主机分配给 Value 存放的内存地址。

可选的，主机和存储控制器之间指令和数据的传输方式还可以是远程直接数据存取（英文：remote direct memory access，缩写：RDMA）等其它传输方式，本申请实施例对此不作具体限定。

S719、存储控制单元返回状态信息到键值存储处理单元。

具体的，这里的状态信息是指是否获取数据成功的信息。

S720、键值存储处理单元及前述单元依次传递状态信息到中间件。

具体的，结合图 5 可知，此处的前述单元具体包括：NVMe 存储控制器的 NVMe 命令处理单元、主机的 NVMe 驱动层的 NVMe 命令转换单元。

S721、中间件返回状态信息到应用层。

至此，当键值存储请求为单次获取数据请求时，键值存储过程结束。

示例三、

当键值存储请求为单次删除数据请求时，如图 8 所示，本申请实施例提供的键值存储方法包括步骤 S801-S811:

S801、应用层调用中间件删除接口：Delete (String Key)。

S802、中间件提交删除数据请求、Key 存放的地址信息到 NVMe 驱动层。

S803、NVMe 驱动层的 NVMe 命令转换单元写删除数据指令 83h、Key 存放的地址信息到 NVMe 发送队列。

具体的，在 NVMe 协议中，定义 0-63 字节为 NVMe 发送队列对应的字段。该步骤中 NVMe 发送队列的命令格式如图 8 中的发送队列命令格式。其中，字节 03:00（即第 0 个字节到第 3 个字节）为指令码对应的字节；字节 23：04（即第 4 个字节到第 23 个字节）为 Command Dword 1-6（即命令字的第 1-6 个 4 字节）；字节 39：24（即第 24 个字节到第 39 个字节）为 key 存放的地址信息对应的字节；字节 63：40（即第 40 个字节到第 63 个字节）为 Command Dword 10-15（即命令字的第 10-15 个 4 字节）。

需要说明的是，本申请实施例中的 NVMe 发送队列是上述指令序列的一种具体形式，当然，如上所述，本申请实施例中的指令序列的形式也可以是数据包等其它形式，本申请实施例对此不作具体限定。

S804、NVMe 驱动层通知 NVMe 存储控制器读 NVMe 发送队列。

S805、NVMe 存储控制器的 NVMe 命令处理单元通过 DMA 读 NVMe 发送队列。

S806、NVMe 命令处理单元分离 NVMe 队列中的删除数据指令、key 存放的地址信息，提交删除数据请求到键值存储处理单元。

S807、键值存储处理单元处理 Key 存放的地址信息，转换删除数据请求为对应存储设备的删除数据请求,并提交删除数据请求到存储控制单元。

S808、存储控制单元根据 Key 存放的地址信息，执行对存储设备中数据的删除操作。

S809、存储控制单元返回状态信息到键值存储处理单元。

具体的，这里的状态信息是指是否删除数据成功的信息。

S810、键值存储处理单元及前述单元依次传递状态信息到中间件。

具体的，结合图 5 可知，此处的前述单元具体包括：NVMe 存储控制器的 NVMe 命令处理单元、主机的 NVMe 驱动层的 NVMe 命令转换单元。

S811、中间件返回状态信息到应用层。

至此，当键值存储请求为单次删除数据请求时，键值存储过程结束。

示例四、

当键值存储请求为单次废弃数据请求时，如图 9 所示，本申请实施例提供的键值存储方法包括步骤 S901-S911:

S901、应用层调用中间件废弃接口：TRIM (String Key)。

S902、中间件提交废弃数据请求、Key 存放的地址信息到 NVMe 驱动层。

S903、NVMe 驱动层的 NVMe 命令转换单元写废弃数据指令 84h、Key 存放的地

址信息到 NVMe 发送队列。

具体的，在 NVMe 协议中，定义 0-63 字节为 NVMe 发送队列对应的字段。该步骤中 NVMe 发送队列的命令格式如图 9 中的发送队列命令格式。其中，字节 03:00（即第 0 个字节到第 3 个字节）为指令码对应的字节；字节 23: 04（即第 4 个字节到第 23 个字节）为 Command Dword 1-6（即命令字的第 1-6 个 4 字节）；字节 39: 24（即第 24 个字节到第 39 个字节）为 key 存放的地址信息对应的字节；字节 63: 40（即第 40 个字节到第 63 个字节）为 Command Dword 10-15（即命令字的第 10-15 个 4 字节）。

需要说明的是，本申请实施例中的 NVMe 发送队列是上述指令序列的一种具体形式，当然，如上所述，本申请实施例中的指令序列的形式也可以是数据包等其它形式，本申请实施例对此不作具体限定。

S904、NVMe 驱动层通知 NVMe 存储控制器读 NVMe 发送队列。

S905、NVMe 存储控制器的 NVMe 命令处理单元通过 DMA 读 NVMe 发送队列。

S906、NVMe 命令处理单元分离 NVMe 队列中的废弃数据指令、key 存放的地址信息，提交废弃数据请求到键值存储处理单元。

S907、键值存储处理单元处理 Key 存放的地址信息，转换废弃数据请求为对应存储设备的废弃数据请求,并提交废弃数据请求到存储控制单元。

S908、存储控制单元根据 Key 存放的地址信息，执行对存储设备中数据的废弃操作。

S909、存储控制单元返回状态信息到键值存储处理单元。

具体的，这里的状态信息是指是否废弃数据成功的信息。

S910、键值存储处理单元及前述单元依次传递状态信息到中间件。

具体的，结合图 5 可知，此处的前述单元具体包括：NVMe 存储控制器的 NVMe 命令处理单元、主机的 NVMe 驱动层的 NVMe 命令转换单元。

S911、中间件返回状态信息到应用层。

至此，当键值存储请求为单次废弃数据请求时，键值存储过程结束。

其中，上述图 6-9 所示的实施例均是针对单次键值存储请求时的键值存储。由表二可知，在指令定义的过程中，还可以定义聚合操作。所谓聚合操作，是指一次聚合存储操作的请求可以同时完成多个单次存储操作的请求，流程和单次请求操作的流程基本一致；不同的是，多个键（Key）和/或值（Value）的地址信息需要借助 SGL（聚散表）通过 NVMe 发送队列传递给存储控制器，也就是说，多个键（Key）和/或值（Value）的地址信息可以通过聚散表地址信息进行索引，该聚散表中包含多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

示例五、

下面以当键值存储请求为聚合删除数据请求为例，对本申请实施例提供的键值存储方法进行说明。如图 10 所示，本申请实施例提供的键值存储方法包括步骤 S1001-S1011:

S1001、应用层调用中间件聚合删除接口 Delete_Group (String Key group)。

S1002、中间件提交聚合删除数据请求、指示 Key 数据组存放的地址的聚散表

（SGL）的地址信息到 NVMe 驱动层。

S1003、NVMe 驱动层的 NVMe 命令转换单元写聚合删除数据指令 88h、SGL 的地址信息到 NVMe 发送队列。

具体的，在 NVMe 协议中，定义 0-63 字节为 NVMe 发送队列对应的字段。该步骤中 NVMe 发送队列的命令格式如图 10 中的发送队列命令格式。其中，字节 03:00（即第 0 个字节到第 3 个字节）为指令码对应的字节；字节 23: 04（即第 4 个字节到第 23 个字节）为 Command Dword 1-6（即命令字的第 1-6 个 4 字节）；字节 39: 24（即第 24 个字节到第 39 个字节）为 SGL 的地址信息对应的字节；字节 63: 40（即第 40 个字节到第 63 个字节）为 Command Dword 10-15（即命令字的第 10-15 个 4 字节）。

需要说明的是，聚合操作可以包含任意数量个单次操作，图 10 仅是以聚合删除数据请求包括 5 个单次删除数据请求为例，给出了 SGL 中包含 key1- key5 的地址信息的示意，不构成对本申请技术方案的限定。

需要说明的是，本申请实施例中的 NVMe 发送队列是上述指令序列的一种具体形式，当然，如上所述，本申请实施例中的指令序列的形式也可以是数据包等其它形式，本申请实施例对此不作具体限定。

S1004、 NVMe 驱动层通知 NVMe 存储控制器读 NVMe 发送队列。

S1005、NVMe 存储控制器的 NVMe 命令处理单元通过 DMA 读 NVMe 发送队列。

S1006、NVMe 命令处理单元分离 NVMe 队列中的聚合删除数据指令、SGL 的地址信息，提交聚合删除数据请求到键值存储处理单元。

S1007、键值存储处理单元对 SGL 中 Key 数据组存放的地址信息逐一进行处理，转换聚合删除数据请求为每个 Key 对应存储设备中数据的删除数据请求，并提交请求到存储控制单元。

S1008、存储控制单元逐一执行数据的删除操作。

S1009、全部操作完成后，存储控制单元返回状态信息到键值存储处理单元。

具体的，这里的状态信息是指是否全部删除数据成功的信息。

S1010、键值存储处理单元及前述单元依次传递状态信息到中间件。

具体的，结合图 5 可知，此处的前述单元具体包括：NVMe 存储控制器的 NVMe 命令处理单元、主机的 NVMe 驱动层的 NVMe 命令转换单元。

S1011、中间件返回状态信息到应用层。

至此，当键值存储请求为聚合删除数据请求时，键值存储过程结束。

其中，上述图 6-10 所示的实施例均是针对主机检测到键值存储请求时的操作。当然，如上所述，本申请实施例中，主机还可以借助存储协议和存储控制器交互以实现非键值存储，比如同时支持传统块设备存储，具体参见示例六。

示例六、

当主机检测到非键值存储请求时，如图 11 所示，非键值存储相关的标准 NVMe 设备操作包括步骤 S1101-S1111:

S1101、传统应用层调用文件系统接口。

S1102、文件系统层转换非键值存储请求为块层的操作请求。

S1103、块层提交操作请求到 NVMe 驱动层。

S1104、NVMe 驱动层的 NVMe 命令转换单元转换操作请求为标准的 NVMe 指令，写指令码、数据内存指针等信息到 NVMe 发送队列。

具体的，本实施例中 NVMe 发送队列的命令格式和上述实施例中的 NVMe 发送队列命令格式类似，此处不再赘述。

S1105、NVMe 驱动层通知 NVMe 存储控制器读 NVMe 发送队列。

S1106、NVMe 存储控制器的 NVMe 命令处理单元通过 DMA 读 NVMe 发送队列；

S1107、NVMe 命令处理单元分离 NVMe 发送队列中的指令码、数据内存指针等信息，提交操作请求到 NVMe 操作处理单元。

S1108、NVMe 操作处理单元处理数据内存指针等信息，转换操作请求为对应存储设备的操作请求,并提交操作请求到存储控制单元。

S1109、存储控制单元根据数据内存指针，执行对存储设备的操作请求。

S1110、存储控制单元返回状态信息到 NVMe 操作处理单元。

具体的，这里的状态信息是指是否执行操作请求成功的信息。

S1111、NVMe 操作处理单元及前述单元依次传递状态信息到传统应用层。

具体的，结合图 5 可知，此处的前述单元具体包括：NVMe 存储控制器的 NVMe 命令处理单元、主机的 NVMe 驱动层的 NVMe 命令转换单元、块层、以及文件系统。

至此，非键值存储相关的标准 NVMe 设备操作结束。

其中，上述图 6-11 所示的实施例均是结合图 5 所示的键值存储与 NVMe 协议相结合以实现键值存储的操作流程示意图进行说明。将键值存储和 NVMe 协议这类高效存储协议结合起来，可以使得存储系统同时具备二者的优势。然而，如上所述，本申请实施例中的协议可以是 NVMe 协议，也可以是 SCSI 协议等其它存储协议，本申请实施例对此不作具体限定。

比如，本申请实施例还可以将键值存储与 SCSI 协议相结合，如图 12 所示，为键值存储与 SCSI 协议相结合以实现键值存储的操作流程示意图。

首先，对图 12 中的相关单元进行简要介绍：

一、主机接口：

图 12 所示的主机接口与图 5 所示的主机接口的功能相同，具体可参考图 5 所示的主机接口的描述，此处不再赘述。

二、主机软件模块：

1）图 12 所示的应用、传统应用、中间件、文件系统与块层的功能和图 5 所示的应用、传统应用、中间件、文件系统与块层的功能分别相同，具体可参考图 5 所示的应用、传统应用、中间件、文件系统与块层的描述，此处不再赘述。

2）SCSI 层：处理 SCSI 事务的软件层，包括 SCSI 上层、SCSI 中层和 SCSI 下层。

3）SCSI 驱动层：位于 SCSI 下层，负责将 SCSI 请求提交给串行 SCSI 接口（英文：Serial Attached SCSI，缩写：SAS）存储控制器，完成与 SAS 存储控制器间的控制和数据交互操作。

4）SCSI 命令转换单元：位于 SCSI 驱动层，用于将键值（Key-Value）存储的扩

展指令、Key 和/或 Value 的地址等信息填入 NVMe 发送队列，并将 NVMe 发送队列提交给 SCSI 驱动层下发给 SAS 存储控制器。

三、SAS 存储控制器：

1）图 12 所示的键值存储处理单元与存储控制单元的功能和图 5 所示的键值存储处理单元与存储控制单元的功能分别相同，具体可参考图 5 所示的键值存储处理单元与存储控制单元的描述，此处不再赘述。

2）SCSI 命令处理单元：分析 SCSI 控制器接收到的 SCSI 发送队列中的指令码，并根据不同的指令将 SCSI 发送队列分发给键值存储处理单元或者 SCSI 操作处理单元。

3）SCSI 操作处理单元：处理标准 SCSI 协议操作请求。

四、存储设备：

图 12 所示的存储设备与图 5 所示的存储设备的功能相同，具体可参考图 5 所示的存储设备的描述，此处不再赘述。

其次，给出 SCSI 协议中对块设备操作定义的指令码，如表五所示：

表五

命令（Command）	操作码（Operation code）	类型（Type）	列表编号（Subclause）
CHANGE DEFINITION	40h	O	ANSI NCITS 301 SPC
COMPARE	39h	O	ANSI NCITS 301 SPC
COPY	18h	O	ANSI NCITS 301 SPC
COPY AND VERIFY	3Ah	O	ANSI NCITS 301 SPC
FORMAT UNIT	04h	M	6.1.1
INQUIRY	12h	M	ANSI NCITS 301 SPC
LOCK-UNLOCK CACHE	36h	O	6.1.2
LOG SELECT	4Ch	O	ANSI NCITS 301 SPC
LOG SENSE	4Dh	O	ANSI NCITS 301 SPC
MODE SELECT(6)	15h	O	ANSI NCITS 301 SPC
MODE SELECT(10)	55h	O	ANSI NCITS

			301 SPC
MODE SENSE(6)	1Ah	O	ANSI NCITS 301 SPC
MODE SENSE(10)	5Ah	O	ANSI NCITS 301 SPC
MOVE MEDIUM	A7h	O	SMC
Obsolete	01h	O B	3.3.4
Obsolete	31h	O B	3.3.4
Obsolete	30h	O B	3.3.4
Obsolete	32h	O B	3.3.4
Obsolete	0Bh	O B	3.3.4
PERSISTENT RESERVE IN	5Eh	O 1	ANSI NCITS 301 SPC
PERSISTENT RESERVE OUT	5Fh	O 1	ANSI NCITS 301 SPC
PRE-FETCH	34h	O	6.1.3
PREVENT-ALLOW MEDIUM REMOVAL	1Eh	O	ANSI NCITS 301 SPC
READ(6)	08h	M	6.1.4
READ(10)	28h	M	6.1.5
READ(12)	A8h	O	6.2.4
READ BUFFER	3Ch	O	ANSI NCITS 301 SPC
READ CAPACITY	25h	M	6.1.6
READ DEFECT DATA(10)	37h	O	6.1.7
READ DEFECT DATA(12)	B7h	O	6.2.5
READ ELEMENT STATUS	B4h	O	SMC
READ LONG	3Eh	O	6.1.8
REASSIGN BLOCKS	07h	O	6.1.9

REBUILD	81h	O	6.1.10
RECEIVE DIAGNOSTIC RESULTS	1Ch	O	ANSI NCITS 301 SPC
REGENERATE	82h	O	6.1.11
RELEASE(6)	17h	O 2	ANSI NCITS 301 SPC
RELEASE(10)	57h	M	ANSI NCITS 301 SPC
REPORT LUNS	A0h	O	ANSI NCITS 301 SPC
REQUEST SENSE	03h	M	ANSI NCITS 301 SPC
RESERVE(6)	16h	O 2	ANSI NCITS 301 SPC
RESERVE(10)	56h	M	ANSI NCITS 301 SPC
SEEK(10)	2Bh	O	6.1.12
SEND DIAGNOSTIC	1Dh	M	ANSI NCITS 301 SPC
SET LIMITS(10)	33h	O	6.1.13
SET LIMITS(12)	B3h	O	6.2.8
START STOP UNIT	1Bh	O	6.1.14
SYNCHRONIZE CACHE	35h	O	6.1.15
TEST UNIT READY	00h	M	ANSI NCITS 301 SPC
VERIFY	2Fh	O	6.1.16
WRITE(6)	0Ah	O	6.1.17
WRITE(10)	2Ah	O	6.1.18
WRITE(12)	AAh	O	6.2.13
WRITE AND VERIFY	2Eh	O	6.1.19
WRITE BUFFER	3Bh	O	ANSI NCITS 301 SPC
WRITE LONG	3Fh	O	6.1.20
WRITE SAME	41h	O	6.1.21
XDREAD	52h	O	6.1.22
XDWRITE	50h	O	6.1.23

XDWRITE EXTENGED	80h	O	6.1.24
XPWRITE	51h	O	6.1.25
Key: M =此命令是必须实现的（英文：Command implementation is mandatory）. O =此命令是否实现是可选的（英文：Command implementation is optional）. OB =此命令已经废弃不用了（英文：Obsolete）. SPC = SCSI-3 指令集（英文：SCSI-3 Primary Command Set）. 注释: （1）可选的 PERSISTENT RESERVE 命令会作为一个组来实施（英文：Optional PERSISTENT RESERVE Commands if implemented as a group）. （2）可选的 RELEASE(6)和 RESERVE(6)会作为一个组来实施（英文：Optional RELEASE(6) and RESERVE(6) Commands if implemented shall both be implemented as a group）. （3）02h, 06h, 09h, 0Ch, 0Dh, 0Eh, 0Fh, 10h, 13h, 14h, 19h, 20h, 21h, 22h, 23h, 24h, 26h, 27h, 29h, 2Ch, 2Dh 以及 C0h 到 FFh 之间是用户定义的操作码。所有这些剩余的操作码可能是为将来的块设备协议标准做预留用（英文：The following operation codes are vendor-specific : 02h, 06h, 09h, 0Ch, 0Dh, 0Eh, 0Fh, 10h, 13h, 14h, 19h, 20h, 21h, 22h, 23h, 24h, 26h, 27h, 29h, 2Ch, 2Dh and C0h through FFh. All remaining operation codes for direct-access block devices are reserved for future standardization）.			

本申请实施例中，根据 SCSI 协议的预留扩展字段定义键值存储请求携带的第一指令码，如表六所示:

表六

命令名	指令码
写数据：Write	02h
获取 Key 对应 Value 的长度：GetLength	05h
获取数据：Read	06h
删除数据：Delete	09h
废弃数据：TRIM	0Ch
聚合写数据：Write	0Dh
聚合获取 Key 对应 Value 的长度：GetLength	0Eh
聚合获取数据：Read	0Fh
聚合删除数据：Delete	10h
聚合废弃数据：TRIM	13h

需要说明的是，上述表六仅是示例性的提供一种根据 SCSI 协议的预留扩展字段定义键值存储请求携带的第一指令码的方式，当然，根据 SCSI 协议的预留扩展字段定义键值存储请求携带的第一指令码不限于上述方式，比如，可以在上面表五中注释部分第 3 条提到的其它预留扩展字段上定义上述键值存储操作，本申请实施例对此不作具体限定。

具体的，结合图 12 进行键值存储操作的流程与结合图 5 进行键值存储操作的流程一致，具体可参考图 6-11 所示的实施例，此处不再赘述。

由上述各实施例所示的键值存储的方法可知，本申请实施例提供的键值存储方法将键值存储操作扩展到存储协议之上，进而，一方面，由于主机可以借助存储协议和存储控制器交互以实现键值存储，无需在块层或者文件系统之上进行转换以实现键值存储，因此降低了存储系统的 IO 路径时延；另一方面，由于主机可以借助存储协议和存储控制器交互以实现非键值存储，因此可以同时支持传统块设备存储。

上述主要从各个设备之间交互的角度对本申请实施例提供的方案进行了介绍。可以理解的是，各个设备，例如主机、存储控制器等为了实现上述功能，其包含了执行各个功能相应的硬件结构和/或软件模块。本领域技术人员应该很容易意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，本申请能够以硬件或硬件和计算机软件的结合形式来实现。某个功能究竟以硬件还是计算机软件驱动硬件的方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

本申请实施例可以根据上述方法示例对主机、存储控制器等进行功能模块的划分，例如，可以对应各个功能划分各个功能模块，也可以将两个或两个以上的功能集成在一个处理模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。需要说明的是，本申请实施例中对模块的划分是示意性的，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。

在采用集成的单元的情况，图 13 示出了上述实施例中所涉及的主机 1300 的一种可能的结构示意图，主机 1300 包括：处理模块 1301 和通信模块 1302。通信模块 1302 用于和存储控制器之间进行通信。处理模块 1301 用于支持主机执行图 4 中的过程 S401 和 S407，或者处理模块 1301 可以包括应用层 1301a、中间件 1301b 和驱动层 1301c，用于支持主机执行图 6-10 中应用层、中间件和 NVMe 驱动层所执行的操作，或者处理模块 1301 还可以包括传统应用层 1301d、文件系统 1301e、块层 1301f 和驱动层 1301c，用于支持主机执行图 11 中传统应用层、文件系统、块层和 NVMe 驱动层所执行的操作。其中，上述方法实施例涉及的各步骤的所有相关内容均可以援引到对应功能模块的功能描述，在此不再赘述。此外，主机 1300 还可以包括存储模块，用于存储主机 1300 的程序代码和数据。

其中，处理模块 1301 可以是处理器或控制器，例如可以是图 2 中的 CPU202，也可以是通用处理器，数字信号处理器（英文：digital signal processor，缩写：DSP），专用集成电路（英文：application-specific integrated circuit，缩写：ASIC），现场可

编程门阵列（英文：field programmable gate array，缩写：FPGA）或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。其可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方框，模块和电路。所述处理器也可以是实现计算功能的组合，例如包含一个或多个微处理器组合，DSP 和微处理器的组合等等。通信模块 1302 可以是通信接口，例如可以是图 2 中的主机通信接口 205，也可以是接收器和发送器，或者收发电路等。存储模块 1201 可以是内存或存储器。

当处理模块 1301 为 CPU，通信模块 1302 为通信接口时，本申请实施例所涉及的主机可以为图 2 所示的主机，具体可参见图 2 部分的相关描述，此处不再赘述。

在采用集成的单元的情况，图 14 示出了上述实施例中所涉及的存储控制器 1400 的一种可能的结构示意图，存储控制器 1400 包括：前端通信模块 1401、处理模块 1402、控制模块 1403 和后端通信模块 1404。前端通信模块 1401 用于支持存储控制器 1400 与前端设备之间的通信，例如和图 4、图 6-11 中主机的通信。后端通信模块 1404 用于支持存储控制器 1400 与后端设备之间的通信，例如和图 4、图 6-11 中存储设备的通信。处理模块 1402 可以包括命令处理单元 1402a、键值存储处理单元 1402b 和标准协议处理单元 1402c，用于支持存储控制器 1400 执行图 4 中命令处理单元、键值存储处理单元和标准协议处理单元所执行的操作，或者用于支持存储控制器 1400 执行图 6-11 中 NVMe 命令处理单元、键值存储处理单元和 NVMe 操作处理单元所执行的操作。控制模块 1403 用于支持存储控制器执行图 4 和图 6-11 中存储控制单元所执行的操作。其中，上述方法实施例涉及的各步骤的所有相关内容均可以援引到对应功能模块的功能描述，在此不再赘述。此外，存储控制器 1400 还可以包括存储模块，用于存储存储控制器 1400 的程序代码和数据。

其中，处理模块 1402 和控制模块 1403 可以是处理器或控制器，例如可以是图 2 中的 CPU212，也可以是通用处理器，数字信号处理器（英文：digital signal processor，缩写：DSP），专用集成电路（英文：application-specific integrated circuit，缩写：ASIC），现场可编程门阵列（英文：field programmable gate array，缩写：FPGA）或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。其可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方框，模块和电路。所述处理器也可以是实现计算功能的组合，例如包含一个或多个微处理器组合，DSP 和微处理器的组合等等。前端通信模块 1401 和后端通信模块 1404 可以是通信接口，例如分别可以是图 2 中的前端通信接口 211 后端通信接口 214，也可以是接收器和发送器，或者收发电路或等。存储模块可以是内存或存储器。

当处理模块 1402 和控制模块 1403 为 CPU，前端通信模块 1401 和后端通信模块 1404 为通信接口时，本申请实施例所涉及的存储控制器可以为图 2 所示的存储控制器，具体可参见图 2 部分的相关描述，此处不再赘述。

结合本申请公开内容所描述的方法或者算法的步骤可以硬件的方式来实现，也可以是由处理器执行软件指令的方式来实现。软件指令可以由相应的软件模块组成，软件模块可以被存放于随机存取存储器（英文：random access memory，缩写：RAM）、闪存、只读存储器（英文：read only memory，缩写：ROM）、可擦除可编程只读存储器（英文：erasable programmable ROM，缩写：EPROM）、电可擦可编程只读存

储器（英文：electrically EPROM，缩写：EEPROM）、寄存器、硬盘、移动硬盘、只读光盘（CD-ROM）或者本领域熟知的任何其它形式的存储介质中。一种示例性的存储介质耦合至处理器，从而使处理器能够从该存储介质读取信息，且可向该存储介质写入信息。当然，存储介质也可以是处理器的组成部分。处理器和存储介质可以位于 ASIC 中。另外，该 ASIC 可以位于核心网接口设备中。当然，处理器和存储介质也可以作为分立组件存在于核心网接口设备中。

本领域技术人员应该可以意识到，在上述一个或多个示例中，本申请所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。计算机可读介质包括计算机存储介质和通信介质，其中通信介质包括便于从一个地方向另一个地方传送计算机程序的任何介质。存储介质可以是通用或专用计算机能够存取的任何可用介质。

以上所述的具体实施方式，对本申请的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本申请的具体实施方式而已，并不用于限定本申请的保护范围，凡在本申请的技术方案的基础之上，所做的任何修改、等同替换、改进等，均应包括在本申请的保护范围之内。

权 利 要 求 书

1、一种键值存储方法，其特征在于，所述方法包括：

在主机检测到键值存储请求之后，所述主机将所述键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列，其中，所述第一指令码为根据所述协议的预留扩展字段定义的指令码；

所述主机与所述存储控制器进行交互，以使得所述存储控制器获取所述第一存储请求指令序列。

2、根据权利要求 1 所述的方法，其特征在于，所述主机将所述键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列，包括：

所述主机按照协议定义的字段为第一存储请求指令序列分配内存；

所述主机将所述键值存储请求携带的第一指令码、以及地址信息写入所述内存；

所述主机与所述存储控制器进行交互，以使得所述存储控制器获取所述第一存储请求指令序列，包括：

所述主机通知所述存储控制器从所述内存中读取所述第一存储请求指令序列。

3、根据权利要求 1 或 2 所述的方法，其特征在于，所述键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；

其中，所述写数据请求携带的地址信息包括键和值存放的地址信息；

所述获取数据请求携带的地址信息包括键存放的地址信息；

所述删除数据请求携带的地址信息包括键存放的地址信息；

所述废弃数据请求携带的地址信息包括键存放的地址信息。

4、根据权利要求 3 所述的方法，其特征在于，若所述键值存储请求为获取数据请求，则在主机检测到键值存储请求之后，在所述主机将所述键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列之前，还包括：

所述主机获取所述获取数据请求所请求的值的长度；

所述主机根据所述值的长度为所述值分配内存，以使得在所述主机与所述存储控制器进行交互，以使得所述存储控制器获取所述第一存储请求指令序列之后，所述存储控制器读数据到所述主机为所述值分配的内存。

5、根据权利要求 4 所述的方法，其特征在于，所述主机获取所述获取数据请求所请求的值的长度，包括：

所述主机将获取键对应的值的长度的指令码、以及所述键存放的地址信息写入所述协议定义的字段组成第二存储请求指令序列，其中，所述获取键对应的值的长度的指令码为根据所述协议的预留扩展字段定义的指令码；

所述主机与所述存储控制器进行交互，以使得所述存储控制器获取所述第二存储请求指令序列；

所述主机接收所述存储控制器发送的所述值的长度。

6、根据权利要求 1-5 任一项所述的方法，其特征在于，所述键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，

所述键值存储请求携带的地址信息通过聚散表的地址信息进行索引，所述聚散表中包含所述多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

7、根据权利要求 1-6 任一项所述的方法，其特征在于，所述方法还包括：

在所述主机检测到非键值存储请求之后，所述主机将所述非键值存储请求携带的第二指令码、以及数据内存指针写入所述协议定义的字段组成第三存储请求指令序列，其中，所述第二指令码为所述协议的标准指令码；

所述主机与所述存储控制器进行交互，以使得所述存储控制器获取所述第三存储请求指令序列。

8、根据权利要求 1-7 任一项所述的方法，其特征在于，所述协议为非易失性存储标准 NVMe 协议；

其中，所述 NVMe 协议定义 0-63 字节为存储请求指令序列的字段。

9、一种键值存储方法，其特征在于，所述方法包括：

存储控制器获取第一存储请求指令序列，所述第一存储请求指令序列由所述主机将键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成，其中，所述第一指令码为根据所述协议的预留扩展字段定义的指令码；

所述存储控制器从所述第一存储请求指令序列中分离出所述第一指令码和所述地址信息；

所述存储控制器根据所述第一指令码和所述地址信息，对存储设备进行所述第一指令码对应的操作。

10、根据权利要求 9 所述的方法，其特征在于，所述存储控制器获取第一存储请求指令序列，包括：

所述存储控制器从主机按照协议定义的字段为所述第一存储请求指令序列分配的内存中读取第一存储请求指令序列。

11、根据权利要求 9 或 10 所述的方法，其特征在于，所述键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；

其中，所述写数据请求携带的地址信息包括键和值存放的地址；

所述获取数据请求携带的地址信息包括键存放的地址信息；

所述删除数据请求携带的地址信息包括键存放的地址信息；

所述废弃数据请求携带的地址信息包括键存放的地址信息。

12、根据权利要求 11 所述的方法，其特征在于，若所述键值存储请求为获取数据请求，则在所述存储控制器获取第一存储请求指令序列之前，还包括：

所述存储控制器获取所述获取数据请求所请求的值的长度；

所述存储控制器向所述主机发送所述值的长度，以使得所述主机根据所述值的长度为所述值分配内存；

所述存储控制器根据所述第一指令码和所述地址信息，对存储设备进行所述第一指令码对应的操作，包括：

所述存储控制器根据所述第一指令码和所述地址信息，读数据到所述主机为所述值分配的内存。

13、根据权利要求 12 所述的方法，其特征在于，所述存储控制器获取所述获取数据请求所请求的值的长度，包括：

所述存储控制器获取第二存储请求指令序列，所述第二存储请求指令序列由所述主机将获取键对应的值的长度的指令码、以及所述键存放的地址信息写入所述协议定义的字段组成，其中，所述获取键对应的值的长度的指令码为根据所述协议的预留扩展字段定义的指令码；

所述存储控制器从所述第二存储请求指令序列中分离出所述获取键对应的值的长度的指令码、以及所述键存放的地址信息；

所述存储控制器根据所述获取键对应的值的长度的指令码、以及所述键存放的地址信息，从所述存储设备中获取所述值的长度。

14、根据权利要求 9-13 任一项所述的方法，其特征在于，所述键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，

所述键值存储请求携带的地址信息通过聚散表的地址信息进行索引，所述聚散表中包含所述多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

15、根据权利要求 9-14 任一项所述的方法，其特征在于，所述方法还包括：

所述存储控制器获取第三存储请求指令序列，所述第三存储请求指令序列由所述主机将非键值存储请求携带的第二指令码、以及数据内存指针写入所述协议定义的字段组成，其中，所述第二指令码为所述协议的标准指令码；

所述存储控制器从所述第三存储请求指令序列中分离出所述第二指令码和所述数据内存指针；

所述存储控制器根据所述第二指令码和所述数据内存指针，对所述存储设备进行所述第二指令码对应的操作。

16、根据权利要求 9-15 任一项所述的方法，其特征在于， 所述协议包括非易失性存储标准 NVMe 协议；

其中，所述 NVMe 协议定义 0-63 字节为存储请求指令序列的字段。

17、一种主机，其特征在于，所述主机包括：处理模块和通信模块；

所述处理模块，用于在检测到键值存储请求之后，将所述键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列，其中，所述第一指令码为根据所述协议的预留扩展字段定义的指令

码；

所述通信模块，用于与所述存储控制器进行交互，以使得所述存储控制器获取所述第一存储请求指令序列。

18、根据权利要求 17 所述的主机，其特征在于，所述处理模块具体用于：

按照协议定义的字段为第一存储请求指令序列分配内存；

将所述键值存储请求携带的第一指令码、以及地址信息写入所述内存；

所述通信模块具体用于：

通知所述存储控制器从所述内存中读取所述第一存储请求指令序列。

19、根据权利要求 17 或 18 所述的主机，其特征在于，所述键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；

其中，所述写数据请求携带的地址信息包括键和值存放的地址信息；

所述获取数据请求携带的地址信息包括键存放的地址信息；

所述删除数据请求携带的地址信息包括键存放的地址信息；

所述废弃数据请求携带的地址信息包括键存放的地址信息。

20、根据权利要求 19 所述的主机，其特征在于，若所述键值存储请求为获取数据请求，则

所述处理模块，还用于在检测到键值存储请求之后，将所述键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列之前，获取所述获取数据请求所请求的值的长度；

根据所述值的长度为所述值分配内存，以使得在所述通信模块与所述存储控制器进行交互，以使得所述存储控制器获取所述第一存储请求指令序列之后，所述存储控制器读数据到所述处理模块为所述值分配的内存。

21、根据权利要求 20 所述的主机，其特征在于，所述处理模块具体用于：

将获取键对应的值的长度的指令码、以及所述键存放的地址信息写入所述协议定义的字段组成第二存储请求指令序列，其中，所述获取键对应的值的长度的指令码为根据所述协议的预留扩展字段定义的指令码；

通过所述通信模块与所述存储控制器进行交互，以使得所述存储控制器获取所述第二存储请求指令序列；

通过所述通信模块接收所述存储控制器发送的所述值的长度。

22、根据权利要求 17-21 任一项所述的主机，其特征在于，所述键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，

所述键值存储请求携带的地址信息通过聚散表的地址信息进行索引，所述聚散表中包含所述多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

23、根据权利要求 17-22 任一项所述的主机，其特征在于，

所述处理模块，还用于在检测到非键值存储请求之后，将所述非键值存

储请求携带的第二指令码、以及数据内存指针写入所述协议定义的字段组成第三存储请求指令序列，其中，所述第二指令码为所述协议的标准指令码；

所述通信模块，还用于与所述存储控制器进行交互，以使得所述存储控制器获取所述第三存储请求指令序列。

24、根据权利要求 17-23 任一项所述的主机，其特征在于，所述协议包括非易失性存储标准 NVMe 协议；

其中，所述 NVMe 协议定义 0-63 字节为存储请求指令序列的字段。

25、一种存储控制器，其特征在于，所述存储控制器包括：前端通信模块、后端通信模块、处理模块和控制模块；

所述前端通信模块，用于从主机获取第一存储请求指令序列，所述第一存储请求指令序列由所述主机将键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成，其中，所述第一指令码为根据所述协议的预留扩展字段定义的指令码；

所述处理模块，用于从所述第一存储请求指令序列中分离出所述第一指令码和所述地址信息；

所述控制模块，用于根据所述第一指令码和所述地址信息，通过所述后端通信模块对存储设备进行所述第一指令码对应的操作。

26、根据权利要求 25 所述的存储控制器，其特征在于，所述前端通信模块具体用于：

从所述主机按照协议定义的字段为所述第一存储请求指令序列分配的内存中读取所述第一存储请求指令序列。

27、根据权利要求 25 或 26 所述的存储控制器，其特征在于，所述键值存储请求包括：写数据请求、或者获取数据请求、或者删除数据请求、或者废弃数据请求；

其中，所述写数据请求携带的地址信息包括键和值存放的地址；

所述获取数据请求携带的地址信息包括键存放的地址信息；

所述删除数据请求携带的地址信息包括键存放的地址信息；

所述废弃数据请求携带的地址信息包括键存放的地址信息。

28、根据权利要求 27 所述的存储控制器，其特征在于，若所述键值存储请求为获取数据请求，则

所述处理模块，还用于在所述前端通信模块获取第一存储请求指令序列之前，获取所述获取数据请求所请求的值的长度；

所述通信模块，还用于向所述主机发送所述值的长度，以使得所述主机根据所述值的长度为所述值分配内存；

所述控制模块具体用于：

根据所述第一指令码和所述地址信息，读数据到所述主机为所述值分配的内存。

29、根据权利要求 28 所述的存储控制器，其特征在于，所述处理模块具体用于：

通过所述前端通信模块获取第二存储请求指令序列，所述第二存储请求指令序列由所述主机将获取键对应的值的长度的指令码、以及所述键存放的地址信息写入所述协议定义的字段组成，其中，所述获取键对应的值的长度的指令码为根据所述协议的预留扩展字段定义的指令码；

从所述第二存储请求指令序列中分离出所述获取键对应的值的长度的指令码、以及所述键存放的地址信息；

根据所述获取键对应的值的长度的指令码、以及所述键存放的地址信息，通过所述控制模块和所述后端通信模块从所述存储设备中获取所述值的长度。

30、根据权利要求 25-29 任一项所述的存储控制器，其特征在于，所述键值存储请求为包含多个单次键值存储请求的聚合键值存储请求；其中，

所述键值存储请求携带的地址信息通过聚散表的地址信息进行索引，所述聚散表中包含所述多个单次键值存储请求中每个单次键值存储请求携带的地址信息。

31、根据权利要求 25-30 任一项所述的存储控制器，其特征在于，

所述前端通信模块，还用于从所述主机获取第三存储请求指令序列，所述第三存储请求指令序列由所述主机将非键值存储请求携带的第二指令码、以及数据内存指针写入所述协议定义的字段组成，其中，所述第二指令码为所述协议的标准指令码；

所述处理模块，还用于从所述第三存储请求指令序列中分离出所述第二指令码和所述数据内存指针；

所述控制模块，还用于根据所述第二指令码和所述数据内存指针，通过所述后端通信模块对所述存储设备进行所述第二指令码对应的操作。

32、根据权利要求 25-31 任一项所述的存储控制器，其特征在于，

所述协议包括非易失性存储标准 NVMe 协议；

其中，所述 NVMe 协议定义 0-63 字节为存储请求指令序列的字段。

33、一种键值存储系统，其特征在于，所述键值存储系统包括存储设备、如权利要求 17-24 任一项所述的主机、以及如权利要求 25-32 任一项所述的存储控制器。

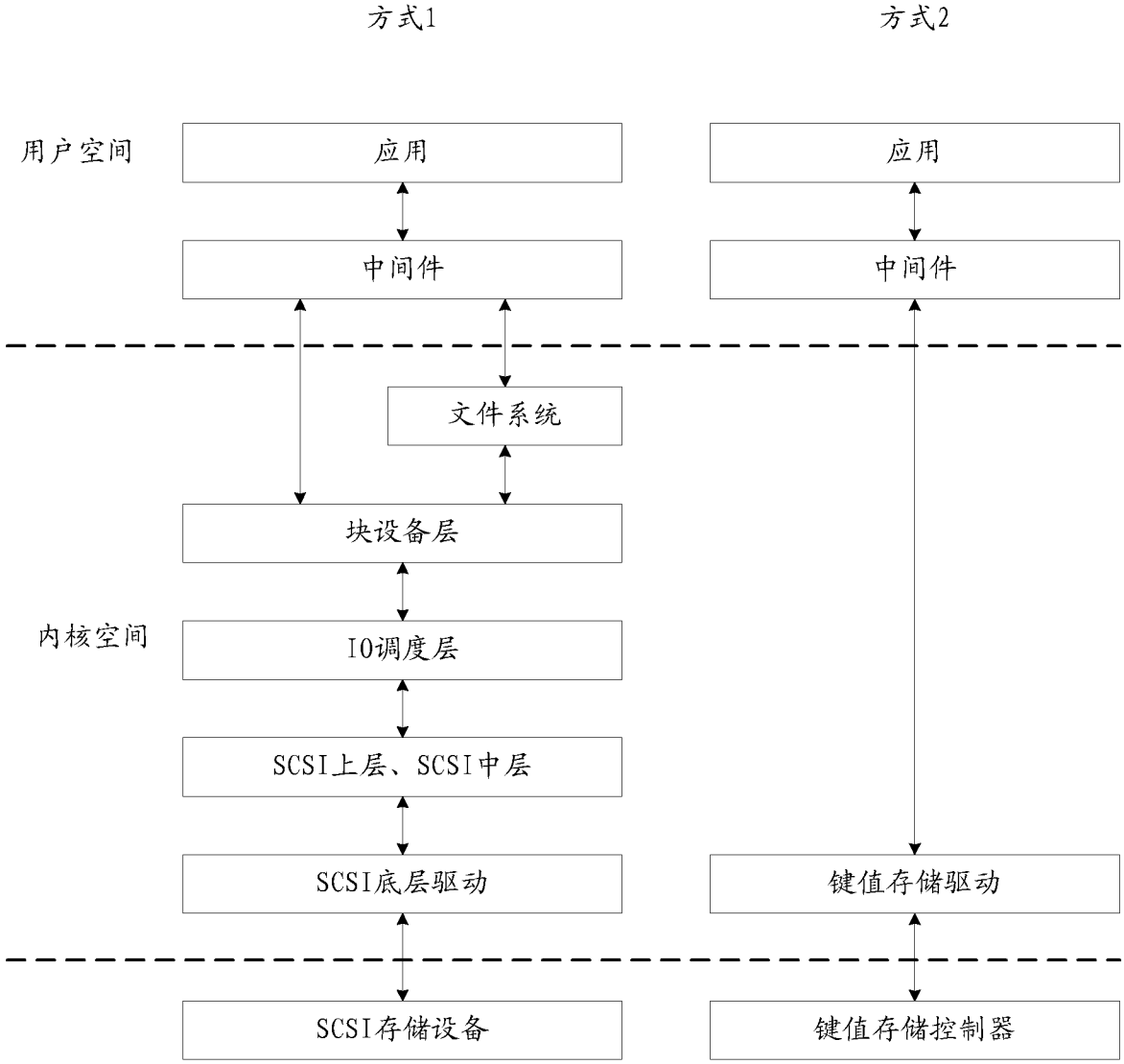
方式1
方式2
用户空间
应用
应用
中间件
中间件
文件系统
块设备层
内核空间
IO调度层
SCSI上层、SCSI中层
SCSI底层驱动
键值存储驱动
SCSI存储设备
键值存储控制器

图 1

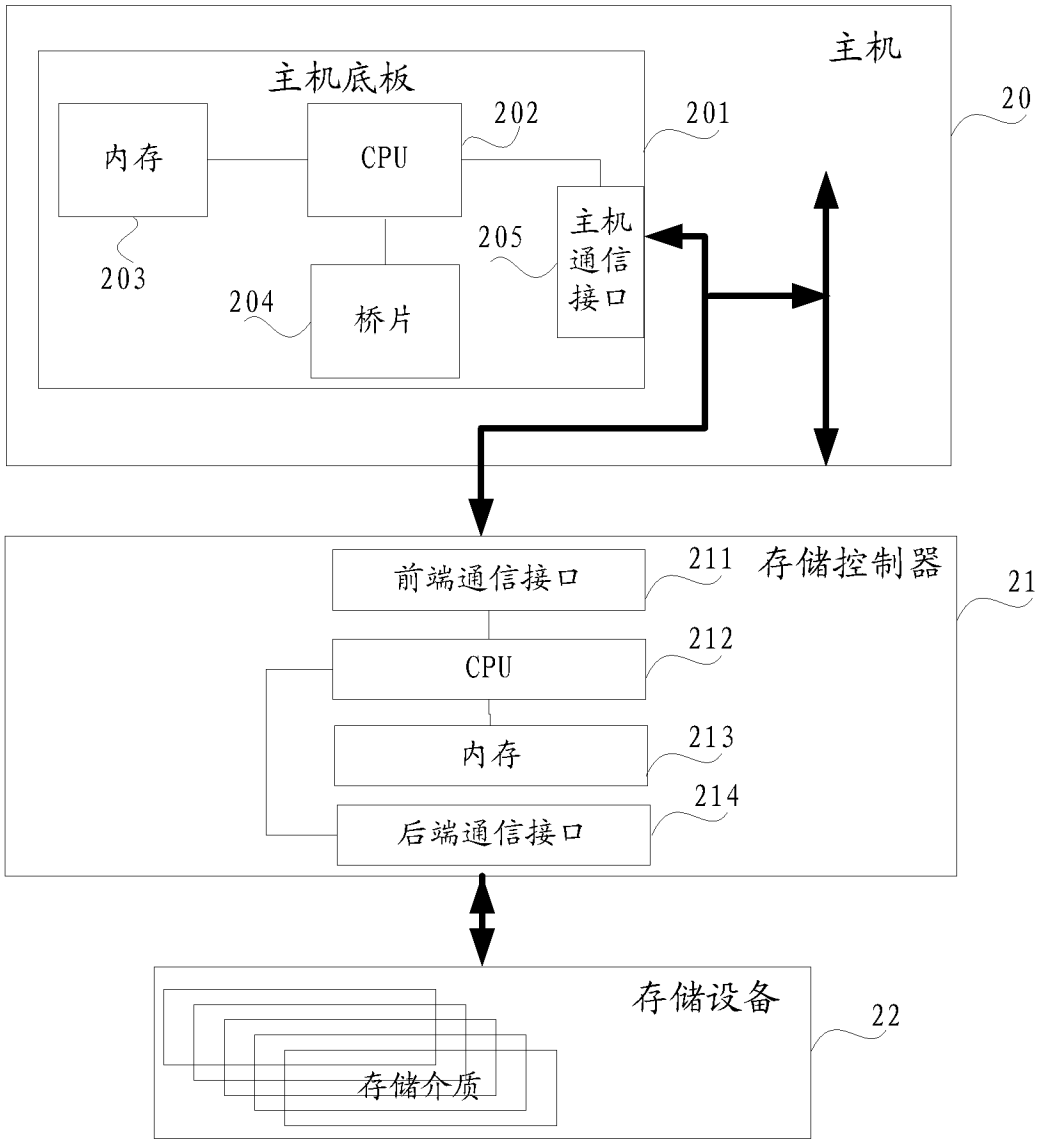
主机
20
主机底板
内存
CPU
202
201
203
205
主机通信接口
204
桥片
前端通信接口
211
存储控制器
21
CPU
212
内存
213
后端通信接口
214
存储设备
22
存储介质

图 2

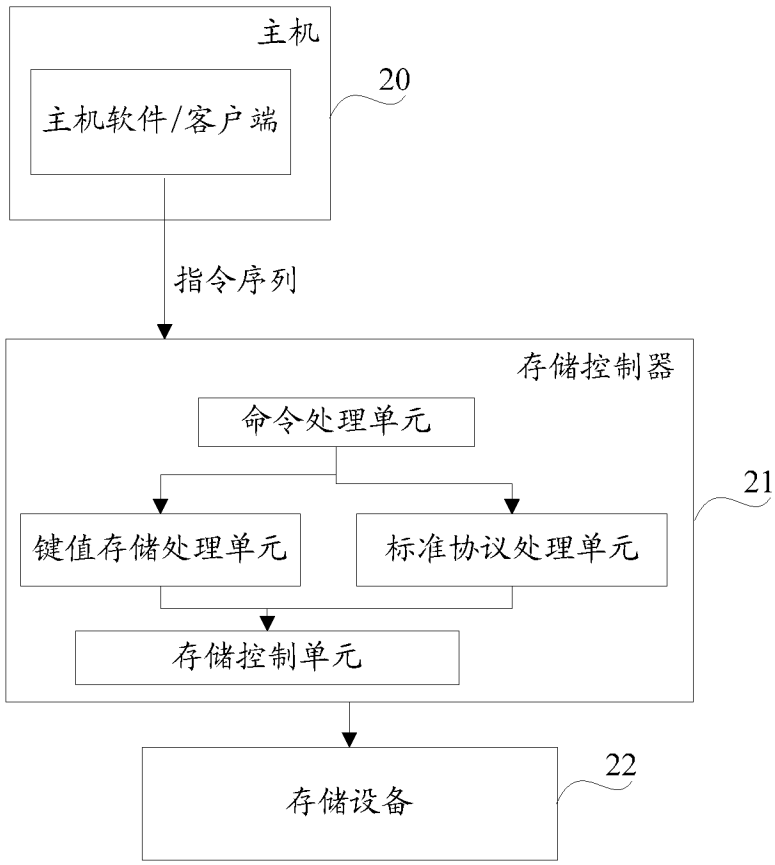
主机
主机软件/客户端
20
指令序列
存储控制器
命令处理单元
21
键值存储处理单元
标准协议处理单元
存储控制单元
存储设备
22

图 3

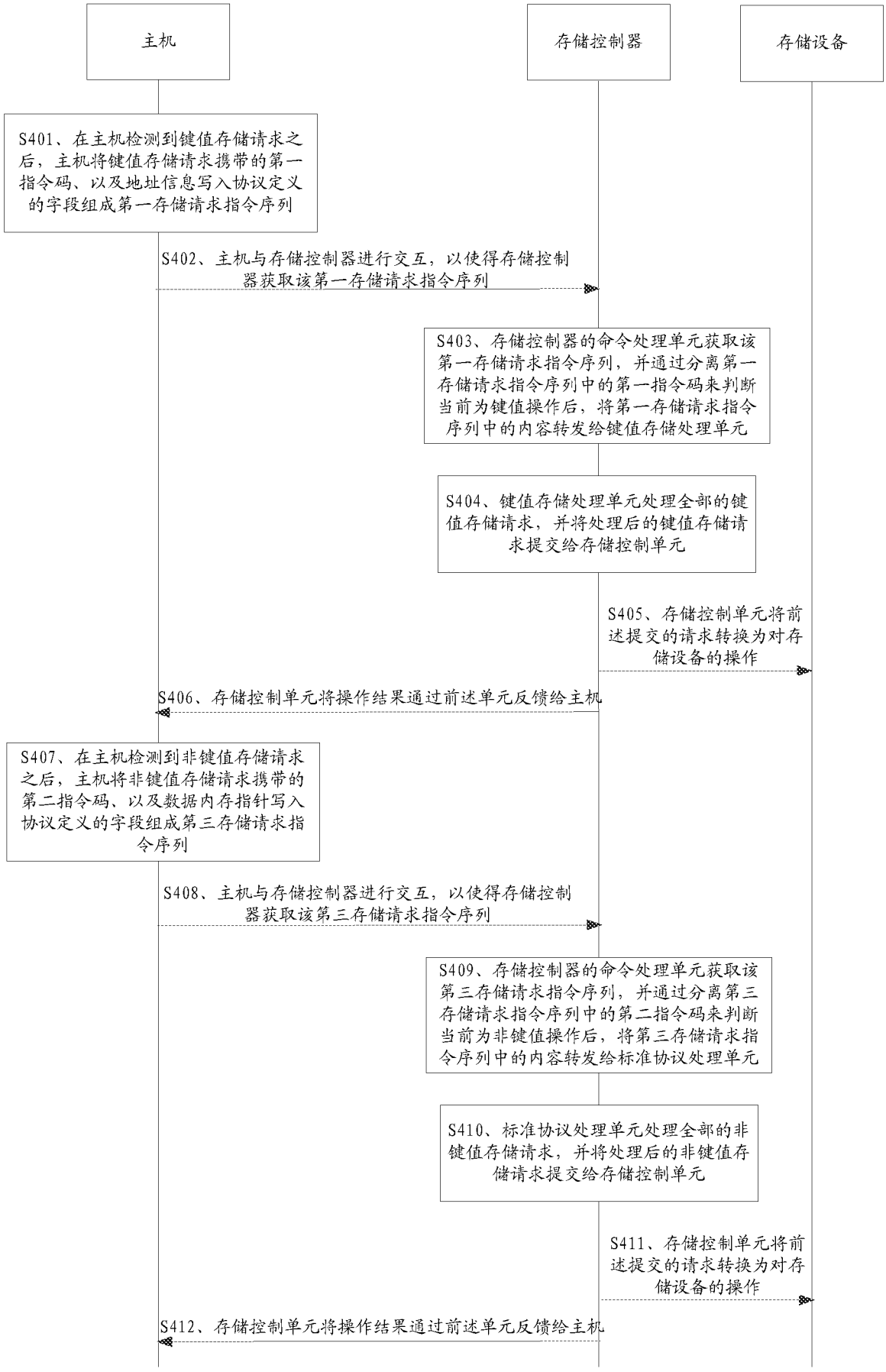
主机
存储控制器
存储设备
S401、在主机检测到键值存储请求之后，主机将键值存储请求携带的第一指令码、以及地址信息写入协议定义的字段组成第一存储请求指令序列
S402、主机与存储控制器进行交互，以使得存储控制器获取该第一存储请求指令序列
S403、存储控制器的命令处理单元获取该第一存储请求指令序列，并通过分离第一存储请求指令序列中的第一指令码来判断当前为键值操作后，将第一存储请求指令序列中的内容转发给键值存储处理单元
S404、键值存储处理单元处理全部的键值存储请求，并将处理后的键值存储请求提交给存储控制单元
S405、存储控制单元将前述提交的请求转换为对存储设备的操作
S406、存储控制单元将操作结果通过前述单元反馈给主机
S407、在主机检测到非键值存储请求之后，主机将非键值存储请求携带的第二指令码、以及数据内存指针写入协议定义的字段组成第三存储请求指令序列
S408、主机与存储控制器进行交互，以使得存储控制器获取该第三存储请求指令序列
S409、存储控制器的命令处理单元获取该第三存储请求指令序列，并通过分离第三存储请求指令序列中的第二指令码来判断当前为非键值操作后，将第三存储请求指令序列中的内容转发给标准协议处理单元
S410、标准协议处理单元处理全部的非键值存储请求，并将处理后的非键值存储请求提交给存储控制单元
S411、存储控制单元将前述提交的请求转换为对存储设备的操作
S412、存储控制单元将操作结果通过前述单元反馈给主机

图 4

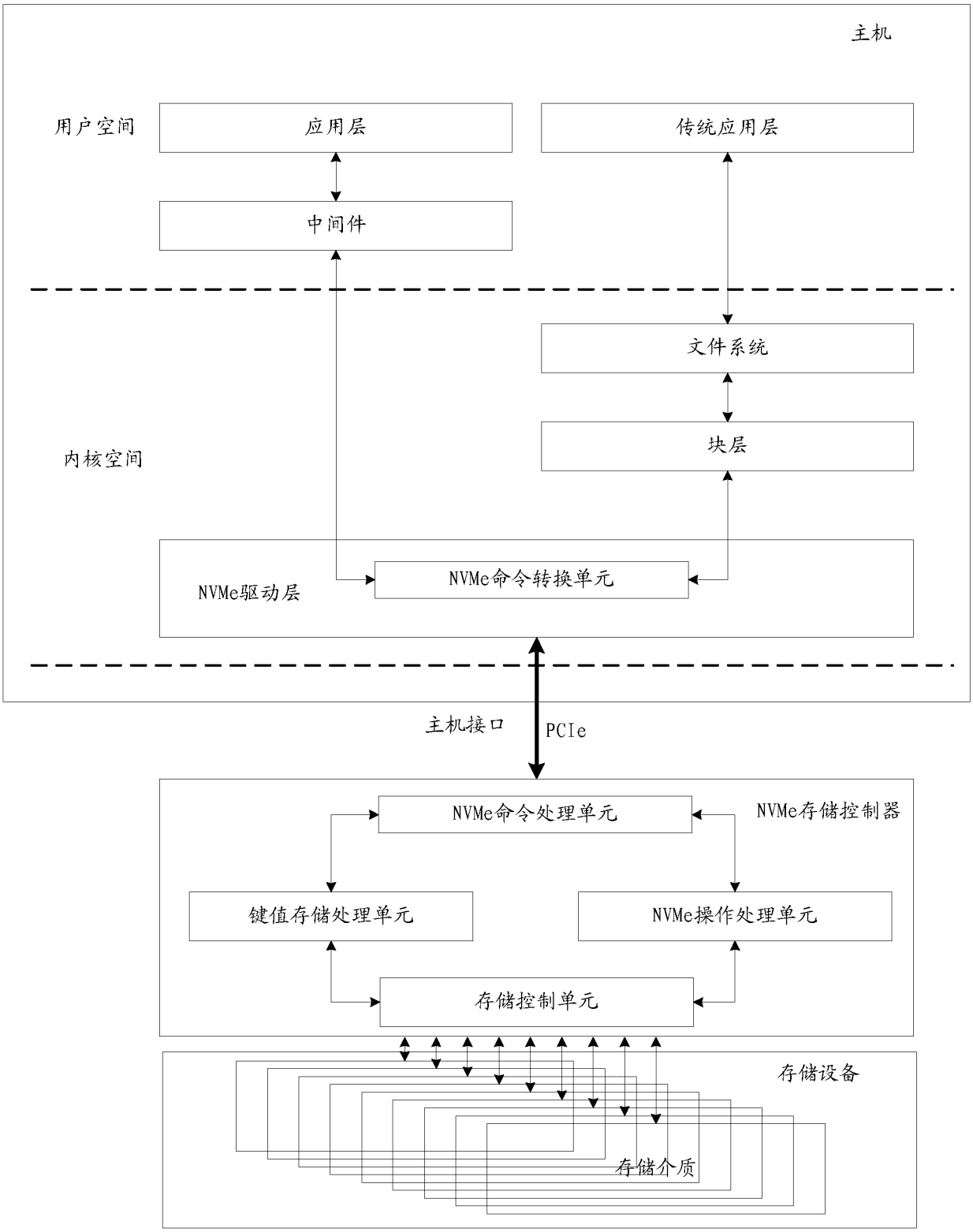
主机
用户空间
应用层
传统应用层
中间件
文件系统
内核空间
块层
NVMe驱动层
NVMe命令转换单元
主机接口
PCIe
NVMe命令处理单元
NVMe存储控制器
键值存储处理单元
NVMe操作处理单元
存储控制单元
存储设备
存储介质

图 5

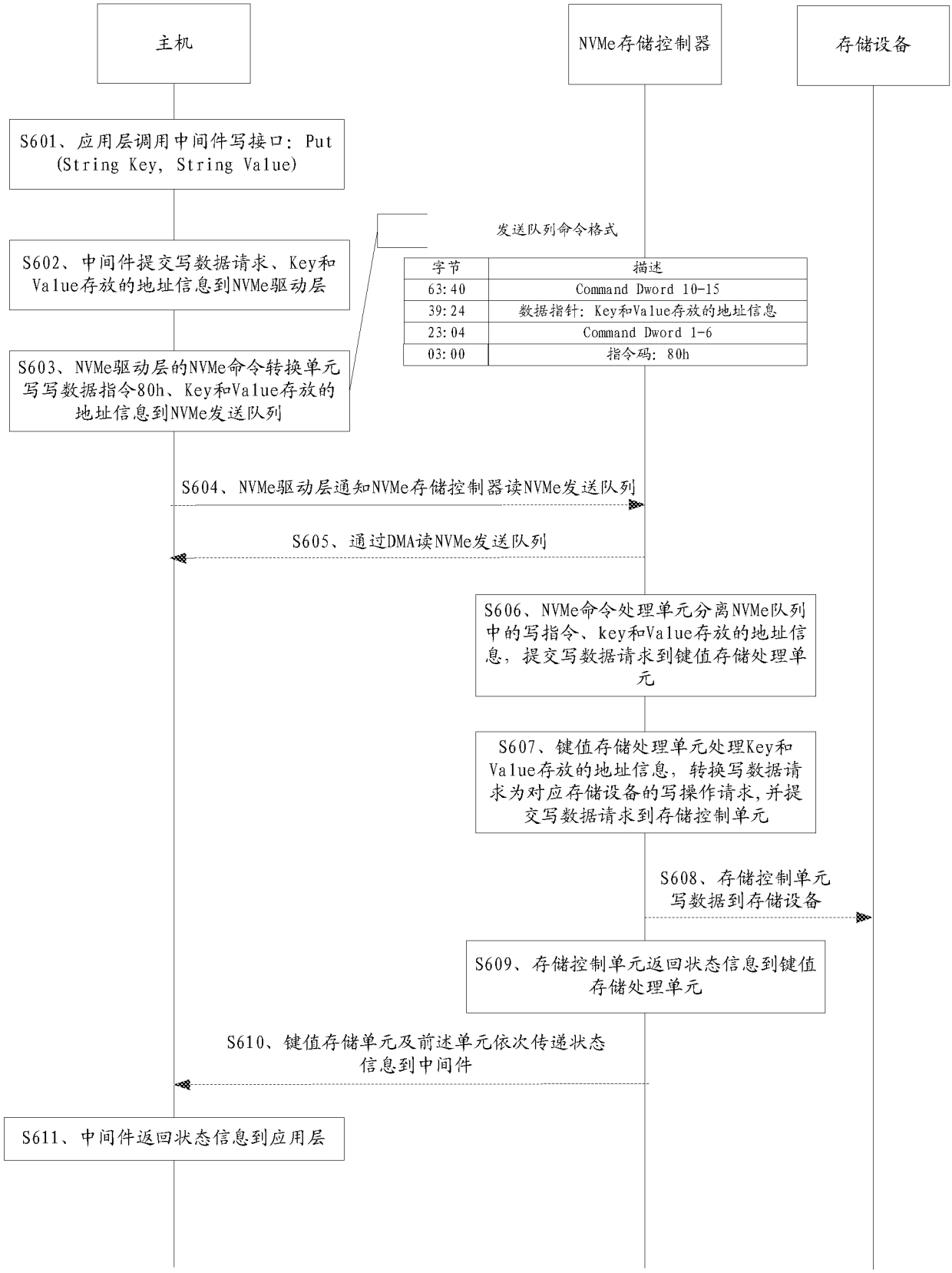
主机
NVMe存储控制器
存储设备
S601、应用层调用中间件写接口：Put (String Key, String Value)
S602、中间件提交写数据请求、Key和Value存放的地址信息到NVMe驱动层
S603、NVMe驱动层的NVMe命令转换单元写写数据指令80h、Key和Value存放的地址信息到NVMe发送队列
发送队列命令格式
字节 描述
63:40 Command Dword 10-15
39:24 数据指针：Key和Value存放的地址信息
23:04 Command Dword 1-6
03:00 指令码：80h
S604、NVMe驱动层通知NVMe存储控制器读NVMe发送队列
S605、通过DMA读NVMe发送队列
S606、NVMe命令处理单元分离NVMe队列中的写指令、key和Value存放的地址信息，提交写数据请求到键值存储处理单元
S607、键值存储处理单元处理Key和Value存放的地址信息，转换写数据请求为对应存储设备的写操作请求，并提交写数据请求到存储控制单元
S608、存储控制单元写数据到存储设备
S609、存储控制单元返回状态信息到键值存储处理单元
S610、键值存储单元及前述单元依次传递状态信息到中间件
S611、中间件返回状态信息到应用层

图 6


主机
NVMe存储控制器
存储设备
S701、应用层调用中间件读操作接口：Get (String Key)
S702、中间件提交获取Key对应Value长度的请求到NVMe驱动层
S703、NVMe驱动层的NVMe命令转换单元写获取Value长度的指令81h、Key存放的地址信息到NVMe发送队列
发送队列命令格式1
字节 描述
63:40 Command Dword 10-15
39:24 数据指针：Key存放的地址信息
23:04 Command Dword 1-6
03:00 指令码：81h
S704、NVMe驱动层通知NVMe存储控制器读NVMe发送队列
S705、NVMe存储控制器通过DMA读NVMe发送队列
S706、NVMe存储控制器的NVMe命令处理单元分离NVMe队列中的获取Value长度的指令、key存放的地址信息，提交读操作请求到键值存储处理单元
S707、键值存储处理单元处理Key存放的地址信息，转换读操作请求为对应存储设备的读操作请求，并提交读操作请求到存储控制单元
S708、存储控制单元从存储设备获取对应Value的长度
S709、存储控制单元提交Value的长度信息到键值存储处理单元
S710、键值存储处理单元及前述单元依次Value的长度信息到中间件
S711、中间件分配Value在主机的内存
S713、NVMe驱动层的NVMe命令转换单元写获取Value的指令82h、key存放的地址信息到NVMe发送队列
发送队列命令格式2
字节 描述
63:40 Command Dword 10-15
39:24 数据指针：key存放的地址信息
23:04 Command Dword 1-6
03:00 指令码：82h
S714、NVMe驱动层通知NVMe存储控制器读NVMe发送队列
S715、通过DMA读NVMe发送队列
S716、NVMe命令处理单元分离NVMe队列中的获取Value的指令、key存放的地址信息，提交获取Value的请求到键值存储处理单元
S717、键值存储处理单元处理Value存放的地址信息，转换获取Value的请求为对应存储设备的获取Value的请求，并提交获取Value的请求到存储控制单元
S718、存储控制单元读数据到主机分配给Value的内存
S719、存储控制单元返回状态信息到键值存储处理单元
S720、键值存储处理单元及前述单元依次传递状态信息到中间件
S721、中间件返回状态信息到应用层

图 7

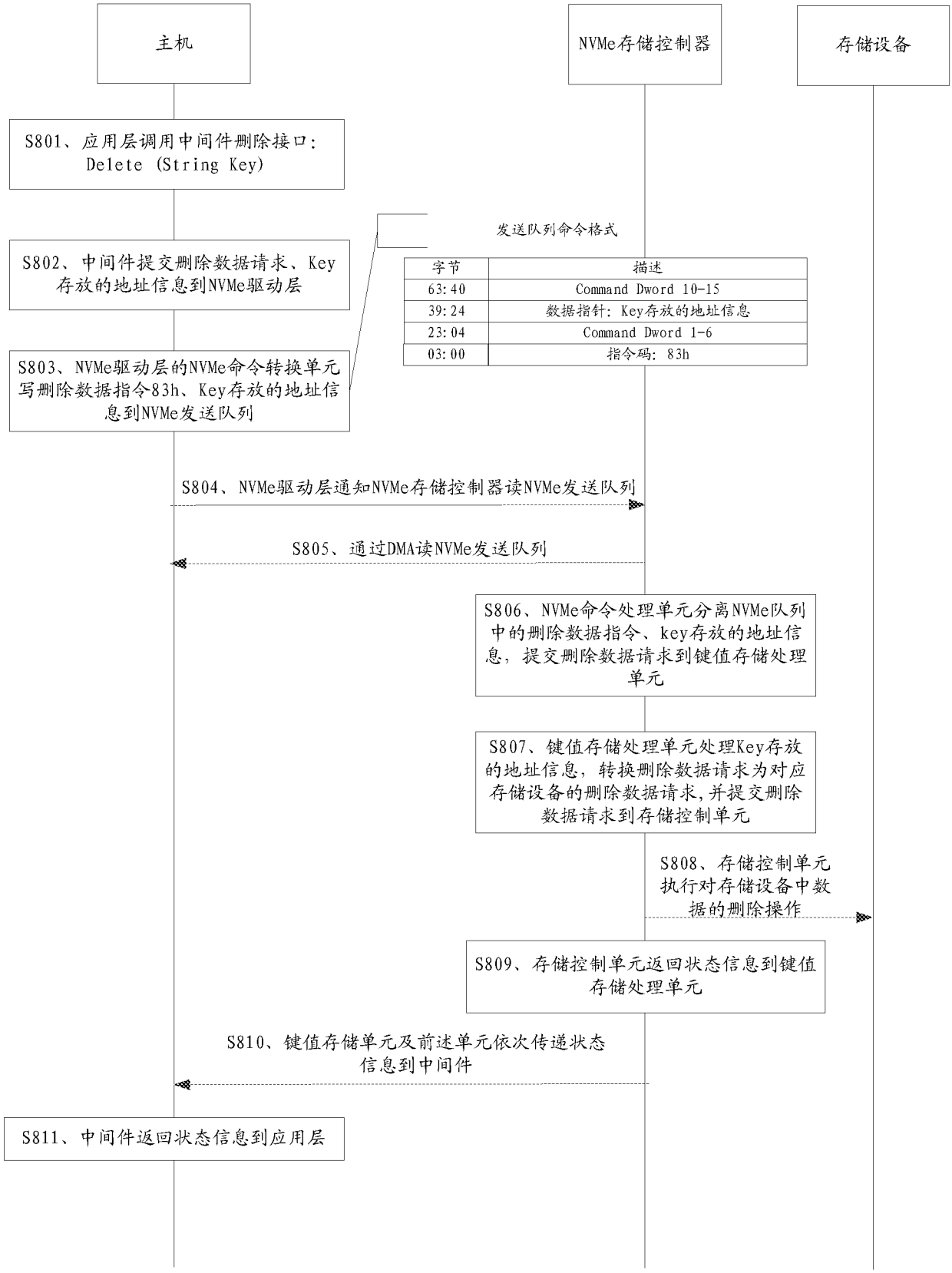
主机
NVMe存储控制器
存储设备
S801、应用层调用中间件删除接口：Delete (String Key)
S802、中间件提交删除数据请求、Key存放的地址信息到NVMe驱动层
S803、NVMe驱动层的NVMe命令转换单元写删除数据指令83h、Key存放的地址信息到NVMe发送队列
发送队列命令格式
字节 描述
63:40 Command Dword 10-15
39:24 数据指针：Key存放的地址信息
23:04 Command Dword 1-6
03:00 指令码：83h
S804、NVMe驱动层通知NVMe存储控制器读NVMe发送队列
S805、通过DMA读NVMe发送队列
S806、NVMe命令处理单元分离NVMe队列中的删除数据指令、key存放的地址信息，提交删除数据请求到键值存储处理单元
S807、键值存储处理单元处理Key存放的地址信息，转换删除数据请求为对应存储设备的删除数据请求，并提交删除数据请求到存储控制单元
S808、存储控制单元执行对存储设备中数据的删除操作
S809、存储控制单元返回状态信息到键值存储处理单元
S810、键值存储单元及前述单元依次传递状态信息到中间件
S811、中间件返回状态信息到应用层

图 8

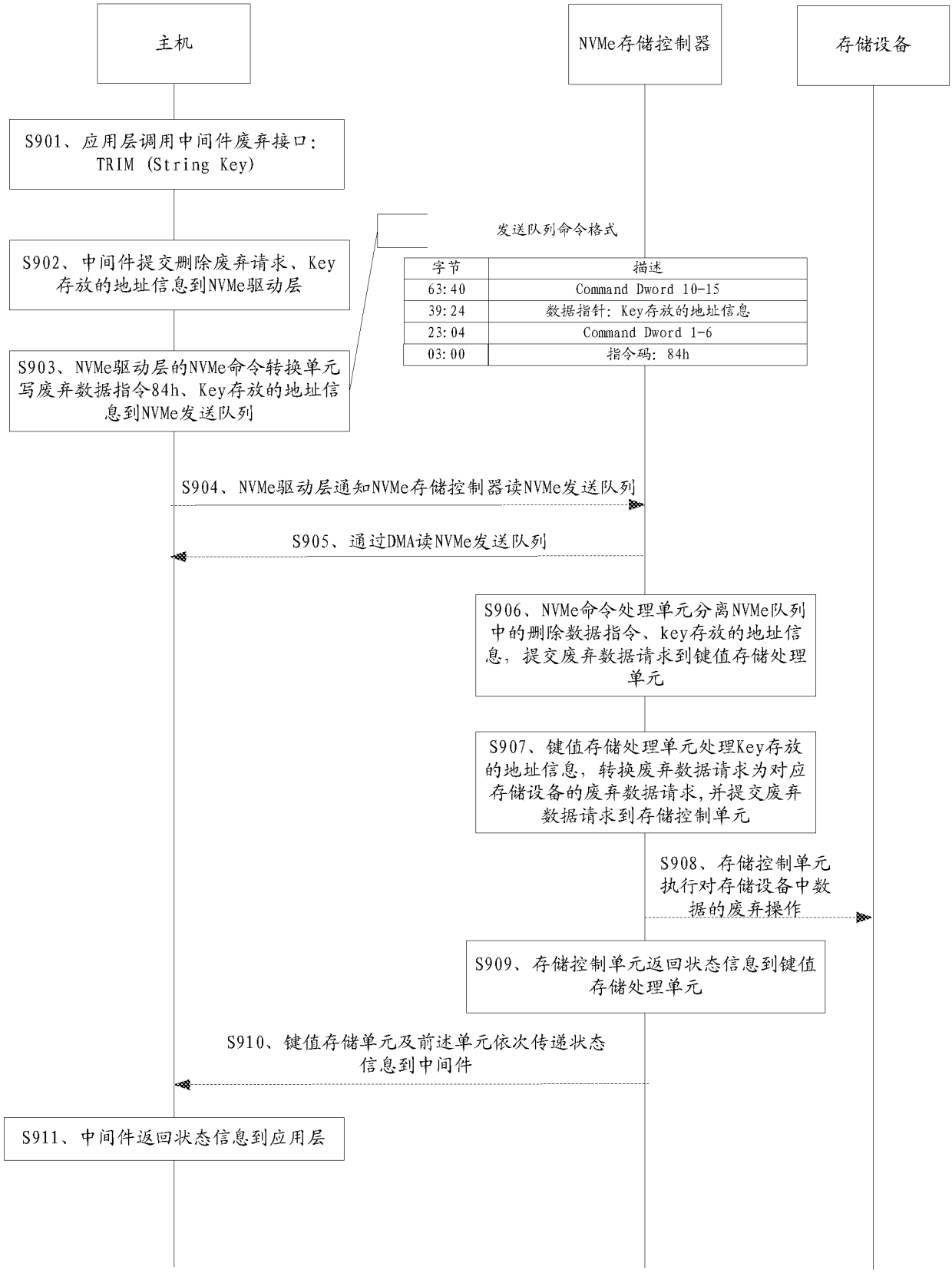
主机
NVMe存储控制器
存储设备
S901、应用层调用中间件废弃接口：TRIM (String Key)
S902、中间件提交删除废弃请求、Key存放的地址信息到NVMe驱动层
S903、NVMe驱动层的NVMe命令转换单元写废弃数据指令84h、Key存放的地址信息到NVMe发送队列
发送队列命令格式
字节 描述
63:40 Command Dword 10-15
39:24 数据指针：Key存放的地址信息
23:04 Command Dword 1-6
03:00 指令码：84h
S904、NVMe驱动层通知NVMe存储控制器读NVMe发送队列
S905、通过DMA读NVMe发送队列
S906、NVMe命令处理单元分离NVMe队列中的删除数据指令、key存放的地址信息，提交废弃数据请求到键值存储处理单元
S907、键值存储处理单元处理Key存放的地址信息，转换废弃数据请求为对应存储设备的废弃数据请求，并提交废弃数据请求到存储控制单元
S908、存储控制单元执行对存储设备中数据的废弃操作
S909、存储控制单元返回状态信息到键值存储处理单元
S910、键值存储单元及前述单元依次传递状态信息到中间件
S911、中间件返回状态信息到应用层

图 9

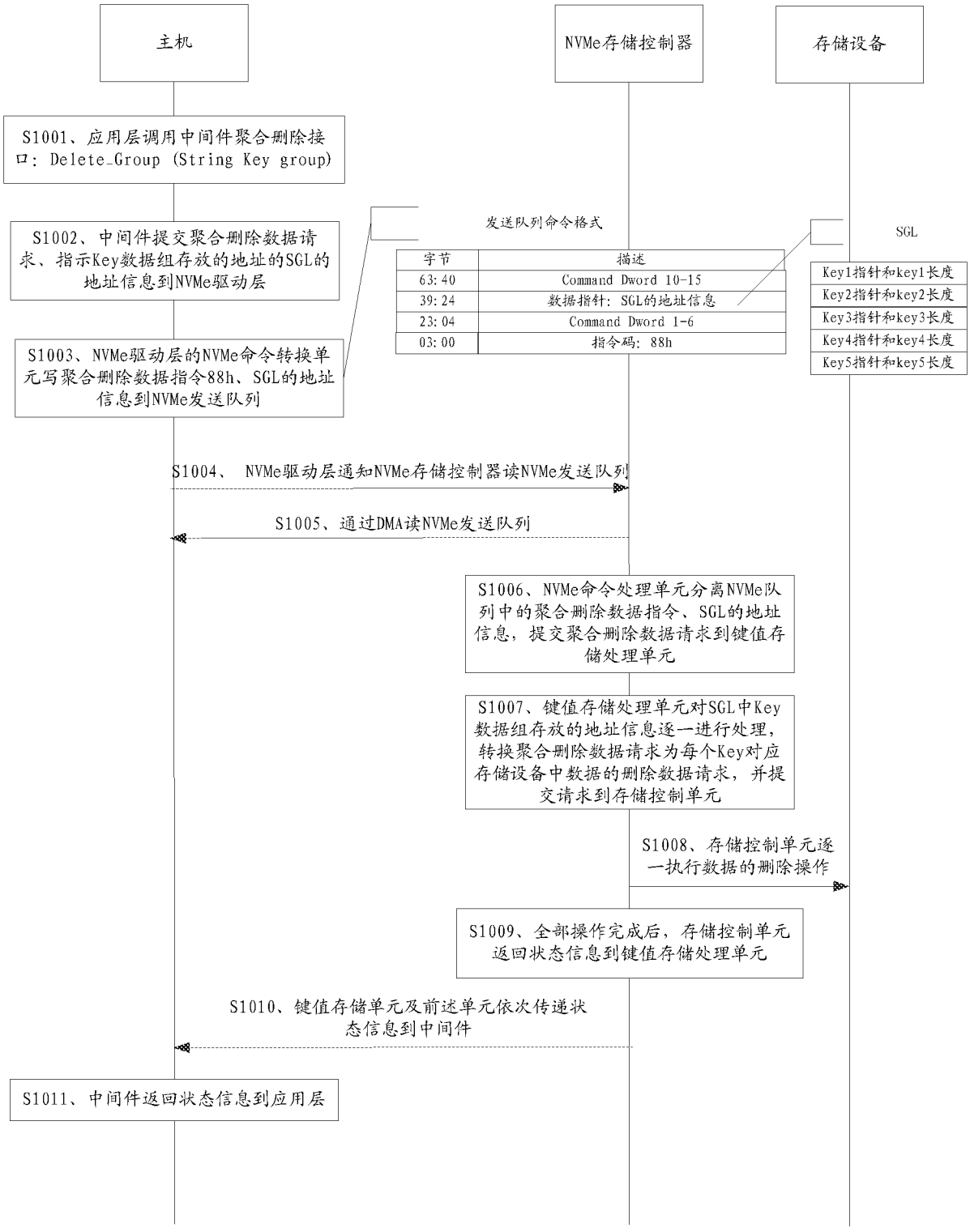
主机
NVMe存储控制器
存储设备
S1001、应用层调用中间件聚合删除接口：Delete_Group (String Key group)
S1002、中间件提交聚合删除数据请求、指示Key数据组存放的地址的SGL的地址信息到NVMe驱动层
S1003、NVMe驱动层的NVMe命令转换单元写聚合删除数据指令88h、SGL的地址信息到NVMe发送队列
发送队列命令格式
字节
描述
63:40
Command Dword 10-15
39:24
数据指针：SGL的地址信息
23:04
Command Dword 1-6
03:00
指令码：88h
SGL
Key1指针和key1长度
Key2指针和key2长度
Key3指针和key3长度
Key4指针和key4长度
Key5指针和key5长度
S1004、 NVMe驱动层通知NVMe存储控制器读NVMe发送队列
S1005、通过DMA读NVMe发送队列
S1006、NVMe命令处理单元分离NVMe队列中的聚合删除数据指令、SGL的地址信息，提交聚合删除数据请求到键值存储处理单元
S1007、键值存储处理单元对SGL中Key数据组存放的地址信息逐一进行处理，转换聚合删除数据请求为每个Key对应存储设备中数据的删除数据请求，并提交请求到存储控制单元
S1008、存储控制单元逐一执行数据的删除操作
S1009、全部操作完成后，存储控制单元返回状态信息到键值存储处理单元
S1010、键值存储单元及前述单元依次传递状态信息到中间件
S1011、中间件返回状态信息到应用层

图 10

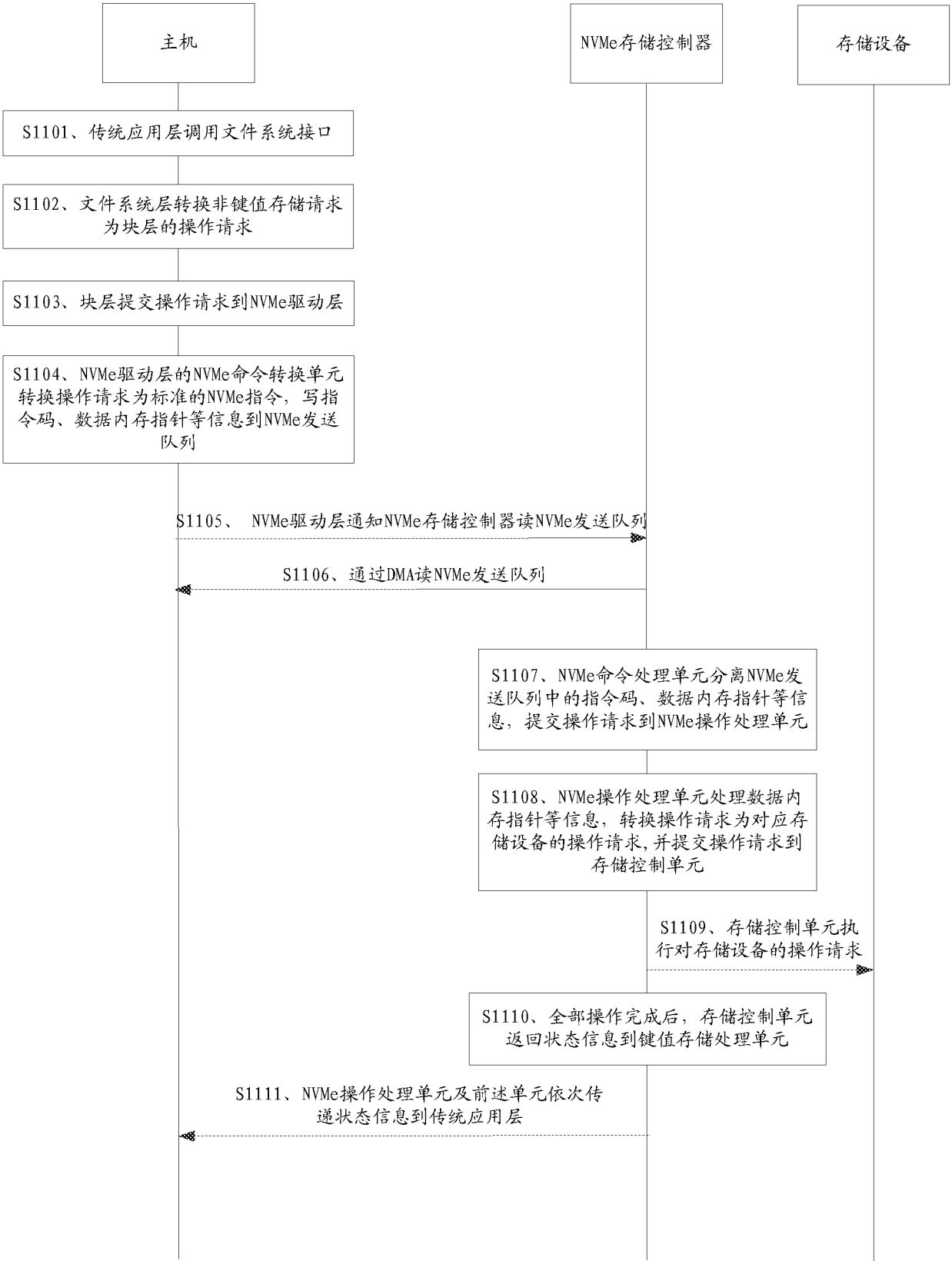
主机
NVMe存储控制器
存储设备
S1101、传统应用层调用文件系统接口
S1102、文件系统层转换非键值存储请求为块层的操作请求
S1103、块层提交操作请求到NVMe驱动层
S1104、NVMe驱动层的NVMe命令转换单元转换操作请求为标准的NVMe指令，写指令码、数据内存指针等信息到NVMe发送队列
S1105、 NVMe驱动层通知NVMe存储控制器读NVMe发送队列
S1106、通过DMA读NVMe发送队列
S1107、NVMe命令处理单元分离NVMe发送队列中的指令码、数据内存指针等信息，提交操作请求到NVMe操作处理单元
S1108、NVMe操作处理单元处理数据内存指针等信息，转换操作请求为对应存储设备的操作请求，并提交操作请求到存储控制单元
S1109、存储控制单元执行对存储设备的操作请求
S1110、全部操作完成后，存储控制单元返回状态信息到键值存储处理单元
S1111、NVMe操作处理单元及前述单元依次传递状态信息到传统应用层

图 11

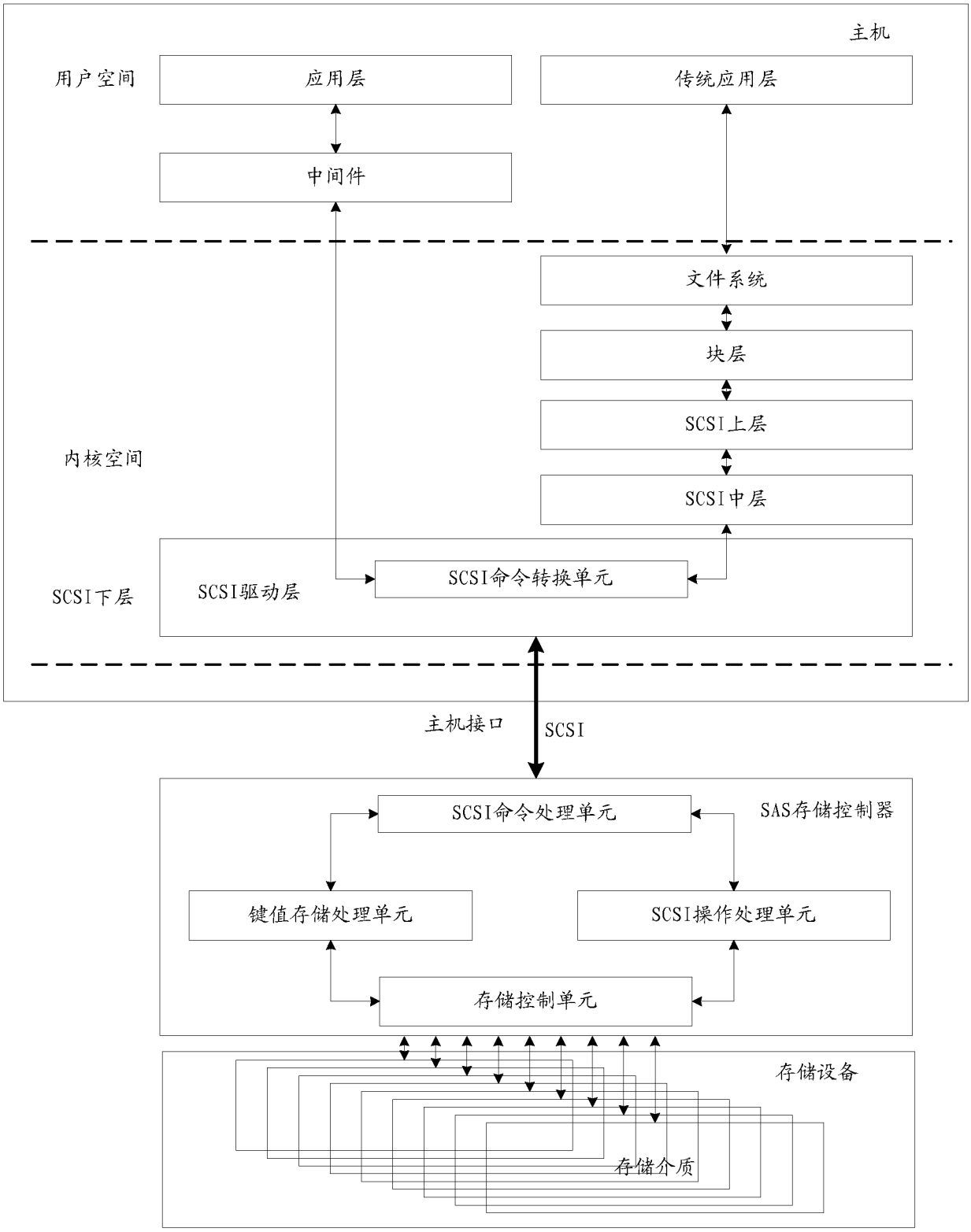
主机
用户空间
应用层
传统应用层
中间件
文件系统
块层
SCSI上层
内核空间
SCSI中层
SCSI下层
SCSI驱动层
SCSI命令转换单元
主机接口
SCSI
SCSI命令处理单元
SAS存储控制器
键值存储处理单元
SCSI操作处理单元
存储控制单元
存储设备
存储介质

图 12

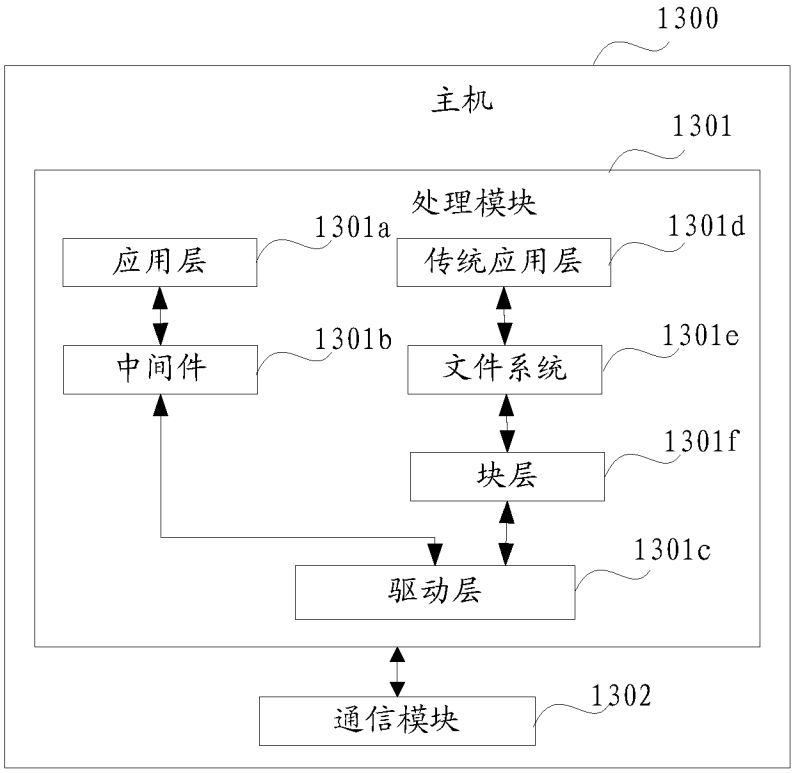
1300
主机
1301
处理模块
1301a
应用层
1301d
传统应用层
1301b
中间件
1301e
文件系统
1301f
块层
1301c
驱动层
1302
通信模块

图 13

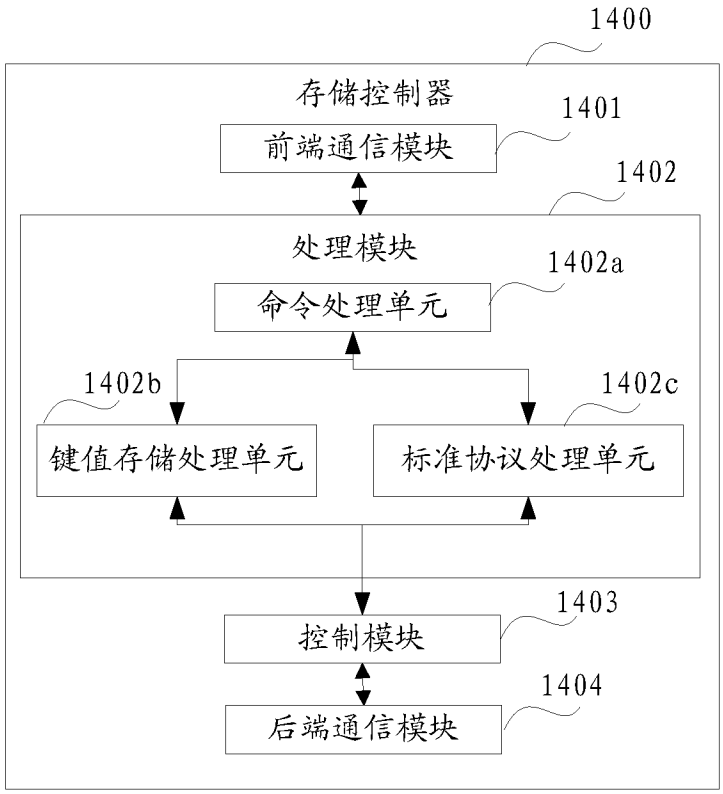
1400
存储控制器
1401
前端通信模块
1402
处理模块
1402a
命令处理单元
1402b
键值存储处理单元
1402c
标准协议处理单元
1403
控制模块
1404
后端通信模块

图 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/085983

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) 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, DWPI, CNKI: 键值, 存储, 指令, 协议, 字段, 地址, NVMe, key-value, memory, instruction, protocol, field, address

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106469198 A (HUAWEI TECHNOLOGIES CO., LTD.) 01 March 2017 (01.03.2017), entire document	1-33
A	CN 103955440 A (RAMAXEL TECHNOLOGY (SHENZHEN) LTD.) 30 July 2014 (30.07.2014), entire document	1-33
A	CN 104111907 A (HUAWEI TECHNOLOGIES CO., LTD.) 22 October 2014 (22.10.2014), entire document	1-33
A	US 2015254003 A1 (FUTUREWEI TECHNOLOGIES INC.) 10 September 2015 (10.09.2015), entire document	1-33

☐ 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	Date of mailing of the international search report
02 July 2017	13 July 2017

Name and mailing address of the ISA
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
KANG, Jian
Telephone No. (86-10) 62411639

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/085983

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106469198 A	01 March 2017	None	
CN 103955440 A	30 July 2014	None	
CN 104111907 A	22 October 2014	EP 3147792 A1	29 March 2017
		WO 2015197027 A1	30 December 2015
US 2015254003 A1	10 September 2015	US 2015255130 A1	10 September 2015

国际检索报告

国际申请号
PCT/CN2017/085983

A. 主题的分类

G06F 17/30(2006.01)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,DWPI,CNKI：键值，存储，指令，协议，字段，地址，NVMe, key-value, memory, instruction, protocol, field, address

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 106469198 A (华为技术有限公司) 2017年 3月 1日 (2017-03-01) 全文	1-33
A	CN 103955440 A (记忆科技深圳有限公司) 2014年 7月 30日 (2014-07-30) 全文	1-33
A	CN 104111907 A (华为技术有限公司) 2014年 10月 22日 (2014-10-22) 全文	1-33
A	US 2015254003 A1 (FUTUREWEI TECHNOLOGIES INC.) 2015年 9月 10日 (2015-09-10) 全文	1-33

☐ 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	国际检索报告邮寄日期
2017年 7月 2日	2017年 7月 13日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	康健
传真号 (86-10)62019451	电话号码 (86-10)62411639

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/085983

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）
CN	106469198	A	2017年 3月 1日	无			
CN	103955440	A	2014年 7月 30日	无			
CN	104111907	A	2014年 10月 22日	EP	3147792	A1	2017年 3月 29日
				WO	2015197027	A1	2015年 12月 30日
US	2015254003	A1	2015年 9月 10日	US	2015255130	A1	2015年 9月 10日