



(12) 发明专利

(10) 授权公告号 CN 115168366 B

(45) 授权公告日 2023. 01. 20

(21) 申请号 202211086860.2

G06F 16/23 (2019.01)

(22) 申请日 2022.09.07

G06F 16/27 (2019.01)

(65) 同一申请的已公布的文献号

审查员 孙思远

申请公布号 CN 115168366 A

(43) 申请公布日 2022.10.11

(73) 专利权人 北京达佳互联信息技术有限公司

地址 100085 北京市海淀区上地西路6号1

幢1层101D1-7

(72) 发明人 班军成 纪艳辉 谢纳政 杨雨

郑涛

(74) 专利代理机构 北京中博世达专利商标代理

有限公司 11274

专利代理师 杨旭

(51) Int. Cl.

G06F 16/22 (2019.01)

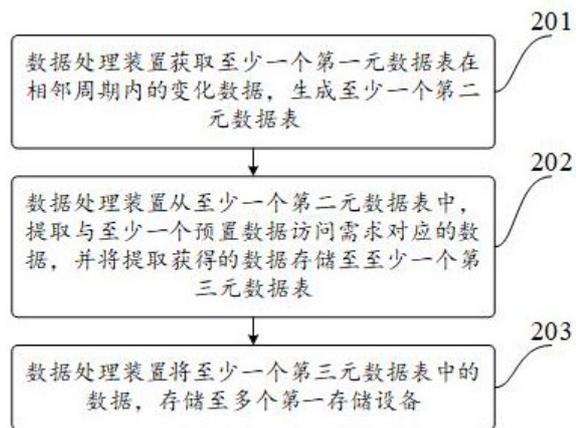
权利要求书3页 说明书15页 附图7页

(54) 发明名称

数据处理方法、装置、电子设备以及存储介质

(57) 摘要

本公开关于一种数据处理方法、装置、电子设备以及存储介质,涉及数据处理技术领域。本公开至少解决相关技术中,同步存储效率较低的问题。该方法包括:获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表;从至少一个第二元数据表中,提取与至少一个预置数据访问需求对应的数据,并将提取获得的数据存储至至少一个第三元数据表;将至少一个第三元数据表中的数据,存储至多个第一存储设备。如此,将变化数据分散存储至多个第一存储设备中,能够横向提高存储变化数据的吞吐量,进而提高同步存储速度和效率、减少同步存储的耗时,最终提高对目标数据进行访问等处理的实时性和处理效率。



1. 一种数据处理方法,其特征在于,包括:

获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表,所述至少一个第一元数据表与所述至少一个第二元数据表一一对应;

获取至少一个预置数据访问需求中的每个预置数据访问需求;

从所述至少一个第二元数据表中,依次提取与每个所述预置数据访问需求对应的列数据,一个所述列数据对应一个数据标识;

按照所述数据标识重新连接提取获得的所述列数据,并将所述连接的结果对应的数据,依次存储至所述至少一个第三元数据表中的每个第三元数据表;

其中,所述至少一个第二元数据表的数量与所述至少一个第三元数据表的数量对应关系包括以下任一项:一对多关系、多对一关系,以及多对多关系,所述至少一个第三元数据表与所述至少一个预置数据访问需求一一对应;

将所述至少一个第三元数据表中的数据,存储至多个第一存储设备,每个所述第一存储设备用于存储至少一个所述第三元数据表中的数据。

2. 根据权利要求1所述的数据处理方法,其特征在于,所述将所述至少一个第三元数据表中的数据,存储至多个第一存储设备之后,所述方法还包括:

响应于客户端发送的数据访问指令,确定目标第一存储设备,所述数据访问指令用于请求访问目标数据,所述目标第一存储设备为所述多个第一存储设备中存储所述目标数据的第一存储设备;

从所述目标第一存储设备中获取所述目标数据,并将所述目标数据发送给所述客户端。

3. 根据权利要求2所述的数据处理方法,其特征在于,所述将所述至少一个第三元数据表中的数据,存储至多个第一存储设备之前,所述方法还包括:

在所述预置数据访问需求包括预置时间标识的情况下,为所述预置数据访问需求对应的第三元数据表中的数据增加时间戳信息;

所述数据访问指令还包括所述目标数据对应的数据更新时间;所述从所述目标第一存储设备中获取所述目标数据,并将所述目标数据发送给所述客户端,包括:

在所述目标第一存储设备中存储第一异常数据的情况下,如果所述数据更新时间在所述第一异常数据对应的时间戳信息之前,则从所述目标第一存储设备中获取所述目标数据,并将所述目标数据发送给所述客户端;

在所述目标第一存储设备中存储第一异常数据的情况下,如果所述数据更新时间与所述第一异常数据对应的时间戳信息相同,或者所述数据更新时间在所述第一异常数据对应的时间戳信息之后,则获取与所述第一异常数据对应的异常数据标识信息,并将所述异常数据标识信息发送给所述客户端。

4. 根据权利要求3所述的数据处理方法,其特征在于,所述方法还包括:

将所述至少一个第二元数据表中的数据,备份存储至第二存储设备;

在所述目标第一存储设备发生故障的情况下,从所述第二存储设备中获取所述目标数据,并将所述目标数据发送给所述客户端;

在所述目标第一存储设备的故障已解除的情况下,从所述目标第一存储设备中获取所述目标数据,并将所述目标数据发送给所述客户端。

5. 根据权利要求4所述的数据处理方法,其特征在于,所述方法还包括:

在所述第一存储设备中的第二异常数据满足预置条件的情况下,根据所述第二异常数据的存储时间,以及当前时间,在所述第二存储设备中查找与所述第二异常数据对应的异常恢复数据;

将所述异常恢复数据作为所述第三元数据表中的数据,存储至所述多个第一存储设备;

其中,所述预置条件为所述第二异常数据对应的数据列数大于预置数量。

6. 根据权利要求2所述的数据处理方法,其特征在于,所述方法还包括:

在所述目标第一存储设备的负载状态大于预置阈值的情况下,响应于客户端发送的数据访问指令,确定所述多个第一存储设备中的协同第一存储设备,所述协同第一存储设备用于存储于所述目标数据中的全部或部分数据;

根据数据包含关系,从至少一个动态第一存储设备中获取所述目标数据,并将所述目标数据发送给所述客户端;

其中,所述数据包含关系为:所述协同第一存储设备中的存储数据,与所述目标数据之间的包含关系;

所述至少一个动态第一存储设备包括:所述协同第一存储设备,或者,所述协同第一存储设备和所述目标第一存储设备。

7. 根据权利要求6所述的数据处理方法,其特征在于,所述根据数据包含关系,从至少一个动态第一存储设备中获取所述目标数据,包括:

在所述协同第一存储设备存储所述目标数据中的全部数据的情况下,从所述协同第一存储设备中获取所述目标数据;

在所述协同第一存储设备存储所述目标数据中的全部数据的情况下,根据所述目标第一存储设备的负载状态,从所述目标第一存储设备和所述协同第一存储设备中分别获取部分所述目标数据;

在所述协同第一存储设备存储所述目标数据中的部分数据的情况下,从所述协同第一存储设备和所述目标第一存储设备中分别获取部分所述目标数据。

8. 一种数据处理装置,其特征在于,包括:获取单元、处理单元和存储单元;

所述获取单元,用于获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表,所述至少一个第一元数据表与所述至少一个第二元数据表一一对应;

所述处理单元,用于获取至少一个预置数据访问需求中的每个预置数据访问需求;从所述至少一个第二元数据表中,依次提取与每个所述预置数据访问需求对应的列数据,一个所述列数据对应一个数据标识;按照所述数据标识重新连接提取获得的所述列数据,并将所述连接的结果对应的数据,依次存储至所述至少一个第三元数据表中的每个第三元数据表;其中,所述至少一个第二元数据表的数量与所述至少一个第三元数据表的数量对应关系包括以下任一项:一对多关系、多对一关系,以及多对多关系,所述至少一个第三元数据表与所述至少一个预置数据访问需求一一对应;

所述存储单元,用于将所述至少一个第三元数据表中的数据,存储至多个第一存储设备,每个所述第一存储设备用于存储至少一个所述第三元数据表中的数据。

9. 一种电子设备,其特征在于,包括:

处理器；

用于存储所述处理器可执行指令的存储器；

其中,所述处理器被配置为执行所述指令,以实现如权利要求1-7中任一项所述的数据处理方法。

10.一种计算机可读存储介质,其特征在于,当所述计算机可读存储介质中的指令被处理器执行时,能够实现如权利要求1-7中任一项所述的数据处理方法。

数据处理方法、装置、电子设备以及存储介质

技术领域

[0001] 本公开涉及数据处理技术领域,尤其涉及一种数据处理方法、装置、电子设备以及存储介质。

背景技术

[0002] 一般的,在数据处理过程中,对于元数据表如分区Hive表中存储的场景数据,通常会同步至数据存储设备中,以便通过数据存储设备来实现低延迟(毫秒级)查询上述场景数据的明细信息。

[0003] 在相关技术中,通过表连接join方式,合并Hive表中存储的场景数据,生成一张大宽表,然后,将大宽表中的场景数据同步存储至数据存储设备。

[0004] 由于数据存储设备的吞吐量有限,在单位时间内存储的数据量有限,而将场景数据同步至数据存储设备的过程效率较低,需要花费的时间较长,可能需要几十分钟,甚至几个小时,导致对场景数据进行访问等处理的实时性和处理效率较低。

发明内容

[0005] 本公开提供一种数据处理方法、装置、电子设备以及存储介质,以至少解决相关技术中,数据同步存储效率低、耗时较长的问题。本公开的技术方案如下:

[0006] 根据本公开实施例的第一方面,提供一种数据处理方法,包括:获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表,至少一个第一元数据表与至少一个第二元数据表一一对应;从至少一个第二元数据表中,提取与至少一个预置数据访问需求对应的数据,并将提取获得的数据存储至至少一个第三元数据表,至少一个第二元数据表的数量与至少一个第三元数据表的数量对应关系包括以下任一项:一对多关系、多对一关系,以及多对多关系,至少一个第三元数据表与至少一个预置数据访问需求一一对应;将至少一个第三元数据表中的数据,存储至多个第一存储设备,每个第一存储设备用于存储至少一个第三元数据表中的数据。

[0007] 可选的,上述数据处理方法还包括:上述将至少一个第三元数据表中的数据,存储至多个第一存储设备之后,响应于客户端发送的数据访问指令,确定目标第一存储设备,数据访问指令用于请求访问目标数据,目标第一存储设备为多个第一存储设备中存储目标数据的第一存储设备;从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。

[0008] 可选的,“从至少一个第二元数据表中,提取与至少一个预置数据访问需求对应的数据,并将提取获得的数据存储至至少一个第三元数据表”,包括:获取至少一个预置数据访问需求中的每个预置数据访问需求;从第二元数据表中,依次提取与每个预置数据访问需求对应的列数据,一个列数据对应一个数据标识;按照数据标识重新连接提取获得的列数据,并将连接的结果对应的数据,依次存储至至少一个第三元数据表中的每个第三元数据表。

[0009] 可选的,上述数据处理方法还包括:上述将至少一个第三元数据表中的数据,存储

至多个第一存储设备之前,在预置数据访问需求包括预置时间标识的情况下,为预置数据访问需求对应的三元数据表中的数据增加时间戳信息;数据访问指令还包括目标数据对应的数据更新时间;从目标第一存储设备中获取目标数据,并将目标数据发送给客户端,包括:在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间在第一异常数据对应的时间戳信息之前,则从目标第一存储设备中获取目标数据,并将目标数据发送给客户端;在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间与第一异常数据对应的时间戳信息相同,或者数据更新时间在第一异常数据对应的时间戳信息之后,则获取与第一异常数据对应的异常数据标识信息,并将异常数据标识信息发送给客户端。

[0010] 可选的,上述数据处理方法还包括:将至少一个第二元数据表中的数据,备份存储至第二存储设备中;在目标第一存储设备发生故障的情况下,从第二存储设备中获取目标数据,并将目标数据发送给客户端;在目标第一存储设备的故障已解除的情况下,从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。

[0011] 可选的,上述数据处理方法还包括:在第一存储设备中的第二异常数据满足预置条件的情况下,根据第二异常数据的存储时间,以及当前时间,在第二存储设备中查找与第二异常数据对应的异常恢复数据;将异常恢复数据作为三元数据表中的数据,存储至多个第一存储设备;其中,预置条件为第二异常数据对应的数据列数大于预置数量。

[0012] 可选的,上述数据处理方法还包括:在目标第一存储设备的负载状态大于预置阈值的情况下,响应于客户端发送的数据访问指令,确定多个第一存储设备中的协同第一存储设备,协同第一存储设备用于存储于目标数据中的全部或部分数据;根据数据包含关系,从至少一个动态第一存储设备中获取目标数据,并将目标数据发送给客户端;其中,数据包含关系为:协同第一存储设备中的存储数据,与目标数据之间的包含关系;至少一个动态第一存储设备包括:协同第一存储设备,或者,协同第一存储设备和目标第一存储设备。

[0013] 可选的,上述“根据数据包含关系,从至少一个动态第一存储设备中获取目标数据”,包括:在协同第一存储设备存储目标数据中的全部数据的情况下,从协同第一存储设备中获取目标数据;在协同第一存储设备存储目标数据中的全部数据的情况下,根据目标第一存储设备的负载状态,从目标第一存储设备和协同第一存储设备中分别获取部分目标数据;在协同第一存储设备存储目标数据中的部分数据的情况下,从协同第一存储设备和目标第一存储设备中分别获取部分目标数据。

[0014] 根据本公开实施例的第二方面,提供一种数据处理装置,包括:获取单元、处理单元和存储单元;获取单元,用于获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表,至少一个第一元数据表与至少一个第二元数据表一一对应;处理单元,用于从至少一个第二元数据表中,提取与至少一个预置数据访问需求对应的数据,并将提取获得的数据存储至至少一个三元数据表,至少一个第二元数据表的数量与至少一个三元数据表的数量对应关系包括以下任一项:一对多关系、多对一关系,以及多对多关系,至少一个三元数据表与至少一个预置数据访问需求一一对应;存储单元,用于将至少一个三元数据表中的数据,存储至多个第一存储设备,每个第一存储设备用于存储至少一个三元数据表中的数据。

[0015] 可选的,上述数据处理装置还包括:确定单元和反馈单元;确定单元,用于上述将

至少一个第三元数据表中的数据,存储至多个第一存储设备之后,响应于客户端发送的数据访问指令,确定目标第一存储设备,数据访问指令用于请求访问目标数据,目标第一存储设备为多个第一存储设备中存储目标数据的第一存储设备;反馈单元,用于从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。

[0016] 可选的,处理单元,具体用于:获取至少一个预置数据访问需求中的每个预置数据访问需求;从第二元数据表中,依次提取与每个预置数据访问需求对应的列数据,一个列数据对应一个数据标识;按照数据标识重新连接提取获得的列数据,并将连接的结果对应的数据,依次存储至至少一个第三元数据表中的每个第三元数据表。

[0017] 可选的,存储单元还用于上述将至少一个第三元数据表中的数据,存储至多个第一存储设备之前,在预置数据访问需求包括预置时间标识的情况下,为预置数据访问需求对应的第三元数据表中的数据增加时间戳信息;在上述数据处理装置中,数据访问指令还包括目标数据对应的数据更新时间;反馈单元,具体用于:在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间在第一异常数据对应的时间戳信息之前,则从目标第一存储设备中获取目标数据,并将目标数据发送给客户端;在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间与第一异常数据对应的时间戳信息相同,或者数据更新时间在第一异常数据对应的时间戳信息之后,则获取与第一异常数据对应的异常数据标识信息,并将异常数据标识信息发送给客户端。

[0018] 可选的,上述数据处理装置还包括:存储单元,还用于将至少一个第二元数据表中的数据,备份存储至第二存储设备中;反馈单元,还用于在目标第一存储设备发生故障的情况下,从第二存储设备中获取目标数据,并将目标数据发送给客户端;反馈单元,还用于在目标第一存储设备的故障已解除的情况下,从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。

[0019] 可选的,上述数据处理装置还包括:存储单元,还用于在第一存储设备中的第二异常数据满足预置条件的情况下,根据第二异常数据的存储时间,以及当前时间,在第二存储设备中查找与第二异常数据对应的异常恢复数据;将异常恢复数据作为第三元数据表中的数据,存储至多个第一存储设备;其中,预置条件为第二异常数据对应的数据列数大于预置数量。

[0020] 可选的,上述数据处理装置还包括:确定单元,还用于在目标第一存储设备的负载状态大于预置阈值的情况下,响应于客户端发送的数据访问指令,确定多个第一存储设备中的协同第一存储设备,协同第一存储设备用于存储于目标数据中的全部或部分数据;反馈单元,还用于根据数据包含关系,从至少一个动态第一存储设备中获取目标数据,并将目标数据发送给客户端;其中,数据包含关系为:协同第一存储设备中的存储数据,与目标数据之间的包含关系;至少一个动态第一存储设备包括:协同第一存储设备,或者,协同第一存储设备和目标第一存储设备。

[0021] 可选的,反馈单元,具体用于:在协同第一存储设备存储目标数据中的全部数据的情况下,从协同第一存储设备中获取目标数据;在协同第一存储设备存储目标数据中的全部数据的情况下,根据目标第一存储设备的负载状态,从目标第一存储设备和协同第一存储设备中分别获取部分目标数据;在协同第一存储设备存储目标数据中的部分数据的情况下,从协同第一存储设备和目标第一存储设备中分别获取部分目标数据。

[0022] 根据本公开实施例的第三方面,提供一种电子设备,包括:处理器;用于存储处理器可执行指令的存储器;其中,处理器被配置为执行上述指令,以实现如第一方面及其任一种可能的设计方式所提供的数据处理方法。

[0023] 根据本公开实施例的第四方面,提供一种计算机可读存储介质,当计算机可读存储介质中的指令被处理器执行时,能够实现如第一方面及其任一种可能的设计方式所提供的数据处理方法。

[0024] 根据本公开实施例的第五方面,提供一种计算机程序产品,包括计算机程序/指令,计算机程序/指令被处理器执行时实现如第一方面及其任一种可能的设计方式所提供的数据处理方法。

[0025] 本公开提供的技术方案至少带来以下有益效果:通过获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表;从至少一个第二元数据表中,提取与至少一个预置数据访问需求对应的数据,并将提取获得的数据存储至至少一个第三元数据表;将至少一个第三元数据表中的数据,存储至多个第一存储设备。如此,在同步存储过程中存储变化数据,减少单位时间内同步存储的数据量,通过预置数据访问需求,将第二元数据表进行拆分重组,生成第三元数据表,以使得第三元数据表能够满足客户端的数据访问需求,同时,将变化数据分散存储至多个第一存储设备中,通过多个存储设备能够横向提高存储变化数据的吞吐量,从而能够提高单位时间内数据存储速度,进而提高同步存储速度和效率、减少同步存储的耗时。再者,仅访问第三元数据表对应的一个第一存储设备,可以得到客户端所需的全部数据,使得客户端可以访问尽可能少的第一存储设备,即可访问或获取目标数据,提高访问或获取数据的效率。

[0026] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0027] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理,并不构成对本公开的不当限定。

[0028] 图1是根据一示例性实施例示出的一种数据处理系统的结构示意图;

[0029] 图2是根据一示例性实施例示出的一种数据处理方法的流程示意图之一;

[0030] 图3是根据一示例性实施例示出的一种数据处理方法的流程示意图之二;

[0031] 图4是根据一示例性实施例示出的一种数据存储的流程示意图;

[0032] 图5是根据一示例性实施例示出的一种数据处理方法的流程示意图之三;

[0033] 图6是根据一示例性实施例示出的一种数据处理方法的流程示意图之四;

[0034] 图7是根据一示例性实施例示出的一种数据处理方法的流程示意图之五;

[0035] 图8是根据一示例性实施例示出的一种数据处理方法的流程示意图之六;

[0036] 图9是根据一示例性实施例示出的数据存储的数据流向示意图;

[0037] 图10是根据一示例性实施例示出的一种数据处理装置的框图;

[0038] 图11是根据一示例性实施例示出的一种电子设备的结构示意图。

具体实施方式

[0039] 为了使本领域普通人员更好地理解本公开的技术方案,下面将结合附图,对本公开实施例中的技术方案进行清楚、完整地描述。

[0040] 需要说明的是,本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0041] 另外,在本公开实施例的描述中,除非另有说明,“/”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。另外,在本公开实施例的描述中,“多个”是指两个或多于两个。

[0042] 对于待处理的数据(如视频相关数据),可以包括是发布方的用户身份标识(User Identification,UID)、视频时长、视频播放量、视频点赞数量、视频转发数量、视频类型等等。可以周期性地统计视频相关数据,并将一个周期内的视频相关数据存储在一个或多个Hive表中。为了对视频相关数据进行再处理,即分析视频相关数据,生成视频推荐方案,以便于提高根据视频推荐方案为用户推荐的视频,符合用户需求的概率。具体的,通过表连接join方式,合并Hive表中存储的视频相关数据,生成一张大宽表,即,在一张大宽表中存储一个周期内的视频相关数据,然后,将大宽表中的视频相关数据,以全量覆盖方式,替换数据存储设备中上一周期的视频相关数据,以使得将本周期的视频相关数据同步存储与数据存储设备中。采用上述数据同步存储方法,由于单个的数据存储设备的吞吐量有限,在单位时间内能够存储的数据量有限,导致同步存储过程中耗时较长,不利于对场景数据进行处理。

[0043] 因此,本公开实施例提供了一种数据处理方法,通过采用使用增量同步存储替代全量同步存储,并设置多个数据存储设备,以解决吞吐量和单位时间内能够同步存储的数据量有限的问题,从而提高对数据进行访问等处理的实时性和处理效率。

[0044] 本公开实施例提供的一种数据处理方法可以适用于数据处理系统。图1示出了该数据处理系统的结构示意图。如图1所示,数据处理系统包括:业务应用层、在线服务层、在线存储层和原始数据层。

[0045] 业务应用层,用于通过查询目标数据,对目标数据进行再次分析处理。

[0046] 在线服务层,用于通过动态路由以及降级操作,实现了存储集群的横向扩容和及时止损;其中,动态路由即存储目标数据的数据存储设备对应的访问路径,降级操作即查找存储目标数据的数据存储设备对应的协同第一存储设备。其中,动态路由配置在存储层,与业务查询相隔离,能够提升系统的稳定性。

[0047] 在线存储层,用于同步存储原始数据,以便于业务应用层实现毫秒级延迟查询目标数据;

[0048] 原始数据层,用于提供原始数据,是业务应用层可以查询到的数据的原始来源;

[0049] 需要说明的是,数据处理系统可以应用于电子设备,具体可以为是手机、平板电脑

等个人智能设备,或者,也可以是笔记本电脑、手持计算机、台式计算机、超级移动个人计算机(ultra-mobile personal computer,UMPC)、服务器等电子设备,又或者,还可以是其他可以存储并处理大数据的电子设备,在此对电子设备的设备形态不做限制。

[0050] 在实际应用中,本公开实施例提供的数据处理方法应用于上述数据处理系统的在线存储层,下面结合附图,对本公开实施例提供的数据处理方法进行描述,下面以执行主体为数据处理装置为例对该方法进行示例性说明。该数据处理装置具体可以为上述各种形态的电子设备。

[0051] 图2为本公开实施例提供的数据处理方法的流程示意图。如图2所示,本公开实施例提供的数据处理方法包括下述步骤201至步骤203。

[0052] 步骤201、数据处理装置获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表。

[0053] 在本公开实施例中,至少一个第一元数据表与至少一个第二元数据表一一对应。也就是,根据至少一个第一元数据表中的任一个第一元数据表在相邻周期内的变化数据,对应生成一个第二元数据表。

[0054] 在本公开实施例中,第一元数据表用于存储原始数据。该原始数据可以是视频时长、视频播放量、视频点赞数量、视频转发数量、视频类型等等视频相关数据,还可以是广告类型、广告产品、广告投放比例等广告投放相关数据,还可以是出售时间、出售渠道、出售商品数量、商品价格等销售相关数据。

[0055] 需要说明的是,第一元数据表中的数据,根据预置周期进行更新。第一元数据表可以为Hive表。由于第一元数据表在相邻周期(即当前周期与前一周期)内的变化数据的数据量,往往小于第一元数据表在当前周期的数据量,而且,根据变化数据和前一周期的原始数据能够得到当前周期的数据,因此,可以采用存储变化数据的方式,对数据进行同步存储。可以理解的是,预置周期可以为10分钟、1小时、24小时、一星期或一个月,在本公开实施例中对预置周期的周期时间长度不做限定。

[0056] 示例性的,假设第一元数据表在前一周记录10个用户及其对应的用户信息,并且上述10个用户都存在购买行为数据,在将上述用户信息和购买行为数据进行同步存储之后,其中3个用户生成二次购买行为数据。由于在前一周已经记录上述3个用户的用户信息,那么在前一周记录10个用户及其对应的用户信息,并且上述10个用户都存在购买行为数据同步存储上述二次购买行为数据即可。

[0057] 在本公开实施例中,为了便于管理变化数据,整合第一元数据表在相邻周期内的变化数据,生成第二元数据表,第二元数据表用于存储变化数据。第二元数据表中的数据量,在正常情况下远小于第一元数据表中的数据量。

[0058] 可以理解的是,相对于第一元数据表的前一周期的数据而言,变化数据可能是增量也可能是减量。

[0059] 步骤202、数据处理装置从至少一个第二元数据表中,提取与至少一个预置数据访问需求对应的数据,并将提取获得的数据存储至至少一个第三元数据表。

[0060] 在本公开实施例中,至少一个第二元数据表的数量与至少一个第三元数据表的数量对应关系包括以下任一项:一对多关系、多对一关系,以及多对多关系,至少一个第三元数据表与至少一个预置数据访问需求一一对应。

[0061] 在本公开实施例中,第二元数据表可以为Hive表,第二元数据表可以存在多个列数据,根据预置数据访问需求,基于列方向进行拆分重组,生成至少一个第三元数据表。

[0062] 可选的,在本公开实施例中,如图3所示,上述步骤202还可以通过下述步骤301至步骤303实现。

[0063] 步骤301、数据处理装置获取至少一个预置数据访问需求中的每个预置数据访问需求。

[0064] 步骤302、数据处理装置从第二元数据表中,依次提取与每个预置数据访问需求对应的列数据。

[0065] 步骤303、数据处理装置按照数据标识重新连接提取获得的列数据,并将连接的结果对应的数据,依次存储至至少一个第三元数据表中的每个第三元数据表。

[0066] 在本公开实施例中,一个列数据对应一个数据标识。数据标识可以用于标识列数据,数据标识可以是随机数值、可以是用户名、还可以为根据用户信息生成的加密序列。可以理解的是,对于相同数据来源的不同列对应的列数据,对应相同的数据标识;对于不同数据来源的相同列对应的列数据,对应不同的数据标识。

[0067] 在本公开实施例中,每个预置数据访问需求,包括客户端所需查询第二元数据表中部分列数据的集合。例如,在第二元数据表中包括视频时长、视频播放量、视频点赞数量、视频转发数量和视频类型,那么预置数据访问需求可以包括视频时长和视频播放量,预置数据访问需求还可以包括视频点赞数量、视频转发数量和视频类型,预置数据访问需求还可以包括视频时长、视频播放量和视频点赞数量。预置数据访问需求可以为多个,不同的预置数据访问需求可以对应不完全相同的列数据的集合。

[0068] 需要说明的是,预置数据访问需求对应的数据可以为根据数据处理经验得到的第二元数据表中的列数据,还可以为通过统计客户端接收到的用户输入得到的列数据,还可以在第二元数据表中随机选取的列数据,在此,预置数据访问需求对应的数据的获得方式不做限定。

[0069] 在本公开实施例中,以数据标识作为维度字段,第二元数据表中与预置数据访问需求对应的列数据,依据数据标识进行连接,并将连接的结果对应的数据存储至第三元数据表中。需要说明的是,其连接方式可以为连接join方式。

[0070] 示例性的,如图4所示,假设第一元数据表中包括表A和表B,其中,表A包括列A1 和列A2,表B包括列B1 和列B2。获取第一元数据表在相邻周期内的变化数据,生成第二元数据表表A'和表B',表A'包括列A1' 和列A2',表B'包括列B1' 和列B2'。第一个预置数据访问需求对应的列数据包括列A2' 和列B1',第二预置数据访问需求对应的列数据包括列A2' 和列B2',据此,第三元数据表包括表1和表2,并且,将与第一预置数据访问需求对应表A'和表B'中的列数据进行融合生成表1,将与第二预置数据访问需求对应表A'和表B'中的列数据进行融合生成表2。最终,将表1中的数据存储至第一存储设备1中,将表2中的数据存储至第一存储设备2中。

[0071] 本公开提供的技术方案至少带来以下有益效果:通过数据标识,重新连接第二元数据表中与预置数据访问需求对应的列数据,使得第三元数据表中的数据之间具有关联关系,相应的,能够提高得到第三元数据表的速度,进而能够提高同步存储第一元数据表的速度。

[0072] 步骤203、数据处理装置将至少一个第三元数据表中的数据,存储至多个第一存储设备。

[0073] 在本公开实施例中,每个第一存储设备用于存储至少一个第三元数据表中的数据。

[0074] 可选的,上述数据处理方法中,上述第一存储设备为键值KV存储设备。KV存储是非关系型数据库NoSQL的存储方式,其中的数据按照键值对的形式进行组织,索引和存储。KV存储适合不涉及过多数据关系业务关系的业务数据,同时能有效减少读写磁盘的次数,拥有更好的读写性能,数据查询的延迟较低。基于此,在第一存储设备为键值KV存储设备的情况下,能够使得第一存储设备具有较高的读写性能,以便于后续响应于客户端的请求,低延迟、快速、高效地访问同步存储的数据。

[0075] 在本公开实施例中,每个第一存储设备用于存储至少一个第三元数据表中的数据。第三元数据表与第一存储设备之间的对应关系,可以为一对一,或者多对一。即,一个第三元数据表中的数据存储至任一个第一存储设备,或者多个第三元数据表中的数据存储至任一个第一存储设备。采用多个第一存储设备进行同步存储,能够解决单个数据存储设备或单个数据存储集群,在同步存储过程中产生的吞吐量瓶颈(吞吐量有限),能够提升第一存储设备中同步存储的数据的可用性。

[0076] 在一种示例中,基于图4所示的示例,将表1中的数据存储至第一存储设备1中,将表2中的数据存储至第一存储设备2中。

[0077] 在另一种示例中,基于图4所示的示例,将表1中的数据存储至第一存储设备1中,将表2中的数据也存储至第一存储设备1中。

[0078] 本公开提供的技术方案至少带来以下有益效果:通过获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表;从至少一个第二元数据表中,提取与至少一个预置数据访问需求对应的数据,并将提取获得的数据存储至至少一个第三元数据表;将至少一个第三元数据表中的数据,存储至多个第一存储设备。如此,在同步存储过程中存储变化数据,减少单位时间内同步存储的数据量,通过预置数据访问需求,将第二元数据表进行拆分重组,生成第三元数据表,以使得第三元数据表能够满足客户端的数据访问需求,同时,将变化数据分散存储至多个第一存储设备中,通过多个存储设备能够横向提高存储变化数据的吞吐量,从而能够提高单位时间内数据存储速度,进而提高同步存储速度和效率、减少同步存储的耗时。再者,仅访问第三元数据表对应的一个第一存储设备,可以得到客户端所需的全部数据,使得客户端可以访问尽可能少的第一存储设备,即可访问或获取目标数据,提高访问或获取数据的效率。

[0079] 可选的,在本公开实施例中,上述步骤203之后,如图5所示,上述数据处理方法还包括下述步骤501和步骤502。

[0080] 步骤501、数据处理装置响应于客户端发送的数据访问指令,确定目标第一存储设备。

[0081] 在本公开实施例中,数据访问指令用于请求访问目标数据,目标第一存储设备为多个第一存储设备中存储目标数据的第一存储设备。需要说明的是,用户通过客户端发送数据访问指令,数据访问指令可以是用户在客户端输入的,可以通过客户端接收到的,还可以是依据预置触发条件生成的。

[0082] 可以理解的是,目标数据可以存储在任一个第一存储设备中,还可以存在多个第一存在设备中,据此,目标第一存储设备可以为多个第一存储设备。目标第一存储设备可以通过动态路由来确定。

[0083] 在本公开实施例中,可以根据配置文件记录不同的第一存储设备。上述配置文件可以执行上述步骤203的过程中,根据三元数据表中的数据及其对应的第一存储设备生成的。

[0084] 步骤502、数据处理装置从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。

[0085] 在本公开实施例中,根据动态路由路径,与目标第一存储设备建立连接,以便于从目标第一存储设备中获取目标数据。

[0086] 进一步可选的,用于获取目标数据的第一存储设备,可以根据路由策略并结合存储设备的负载来配置。在本公开实施例中,如图6所示,上述步骤502还包括下述步骤601和步骤602。

[0087] 步骤601、在目标第一存储设备的负载状态大于预置阈值的情况下,数据处理装置响应于客户端发送的数据访问指令,确定多个第一存储设备中的协同第一存储设备。

[0088] 步骤602、数据处理装置根据数据包含关系,从至少一个动态第一存储设备中获取目标数据,并将目标数据发送给客户端。

[0089] 在本公开实施例中,协同第一存储设备用于存储于目标数据中的全部或部分数据。数据包含关系为:协同第一存储设备中的存储数据,与目标数据之间的包含关系。至少一个动态第一存储设备包括:协同第一存储设备,或者,协同第一存储设备和目标第一存储设备。

[0090] 可以理解的是,数据包含关系具体包括:协同第一存储设备中的存储数据包含目标数据中的所有数据;协同第一存储设备中的存储数据包含目标数据中的部分数据。

[0091] 还可以理解的是,在目标第一存储设备的负载状态大于预置阈值的情况下,不能够通过目标第一存储设备及时获取目标数据,因此,通过至少一个动态第一存储设备及时获取目标数据。其中,至少一个动态第一存储设备包括:协同第一存储设备,或者,协同第一存储设备和目标第一存储设备。

[0092] 示例性的,如果地区C的客户端发送的数据访问指令,都用于访问同步存储数据中符合客户访问需求的数据,确定目标第一存储设备为第一存储设备D,那么目标第一存储设备D需要响应地区C的客户端发送的所有数据访问指令,很可能导致负载状态大于预置阈值的情况,即,第一存储设备D不能及时响应数据访问指令。因此,可以确定协同第一存储设备,从目标第一存储设备和协同第一存储设备中获得目标数据。当然,协同第一存储设备的设备数据可以为多个。

[0093] 在本公开实施例中,上述步骤602具体包括:在协同第一存储设备存储目标数据中的全部数据的情况下,数据处理装置从协同第一存储设备中获取目标数据;在协同第一存储设备存储目标数据中的全部数据的情况下,数据处理装置根据目标第一存储设备的负载状态,从目标第一存储设备和协同第一存储设备中分别获取部分目标数据;在协同第一存储设备存储目标数据中的部分数据的情况下,数据处理装置从协同第一存储设备和目标第一存储设备中分别获取部分目标数据。

[0094] 在一种示例中,在协同第一存储设备存储目标数据中的全部数据的情况下,由于目标第一存储设备和协同第一存储设备,都存储目标数据中的全部数据,因此,数据处理装置根据目标第一存储设备的负载状态的变化情况,获取目标第一存储设备对数据访问指令的响应速度,并获取协同第一存储设备对数据访问指令的响应速度,选取对数据访问指令的响应速度快的第一存储设备,获取目标数据。

[0095] 在另一种示例中,在协同第一存储设备存储目标数据中的全部数据的情况下,由于目标第一存储设备和协同第一存储设备,都存储目标数据中的全部数据,因此,数据处理装置根据目标第一存储设备的负载状态的变化情况,确定目标第一存储设备能够处理的数据量,并根据上述数据量,确定目标数据中通过目标第一存储设备获取的部分数据,然后将目标数据中的剩余部分数据通过协同第一存储设备获取。

[0096] 需要说明的是,在协同第一存储设备存储目标数据中的部分数据的情况下,数据处理装置从协同第一存储设备和目标第一存储设备中分别获取部分目标数据,以目标第一存储设备为主,尽可能从目标第一存储设备获取更多的部分目标数据,对于不能从目标第一存储设备获取的部分目标数据,则从协同第一存储设备获取。

[0097] 本公开提供的技术方案至少带来以下有益效果:根据负载状态和数据包含关系,确定目标数据的数据来源,以此避免在目标第一存储设备存在负载状态大于预置阈值的情况下,不能及时响应数据访问指令的问题,能够提高数据处理的效率。

[0098] 本公开提供的技术方案至少带来以下有益效果:在第一存储设备响应客户端的数据访问指令的过程中,通过确定协同第一存储设备,并根据目标第一存储设备和协同第一存储设备与目标数据制水间的数据包含关系,从协同第一存储设备,或者协同第一存储设备和目标第一存储设备中获取目标数据。如此,能够避免在目标第一存储设备存在负载状态大于预置阈值的情况下,不能及时响应数据访问指令的问题,能够提高数据处理的效率。

[0099] 本公开提供的上述技术方案至少能够带来以下有益效果:响应于客户端发送的数据访问指令,确定目标第一存储设备,从目标第一存储设备中获取目标数据,并将目标数据发送给客户端,以使得客户端能够及时准确地得到目标数据,提高了对目标数据进行访问等处理的实时性和处理效率,便于对目标数据进行再次分析或进行其他处理。并且,由于访问三元数据表对应的一个第一存储设备,可以得到客户端所需的全部数据,使得客户端可以访问尽可能少的第一存储设备,即可访问或获取目标数据,能够进一步提高访问或获取数据的效率。

[0100] 可选的,在本公开实施例中,上述步骤203之前,如图7所示,上述数据处理方法还包括下述步骤701。

[0101] 步骤701、在预置数据访问需求包括预置时间标识的情况下,数据处理装置为预置数据访问需求对应的三元数据表中的数据增加时间戳信息。

[0102] 相应的,在本公开实施例中,第一存储设备中也包括时间戳信息。时间戳信息用于标识三元数据表中的数据的生成时间。

[0103] 进一步可选的,在本公开实施例中,数据访问指令还包括目标数据对应的数据更新时间。如图7所示,上述步骤502还可以通过下述步骤702或步骤703实现。

[0104] 步骤702、在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间在第一异常数据对应的时间戳信息之前,则数据处理装置从目标第一存储设备中获取目标

数据,并将目标数据发送给客户端。

[0105] 步骤703、在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间与第一异常数据对应的时间戳信息相同,或者数据更新时间在第一异常数据对应的时间戳信息之后,则数据处理装置获取与第一异常数据对应的异常数据标识信息,并将异常数据标识信息发送给客户端。

[0106] 在本公开实施例中,通过时间戳信息记录第三元数据表中数据的生成时间,即存储在第一存储设备中的数据的生成时间。在目标第一存储设备中存储第一异常数据的情况下,为了避免获取的目标数据是第一异常数据,可以将目标数据对应数据更新时间,与时间戳信息进行比较,如果数据更新时间在第一异常数据对应的时间戳信息之前,则数据处理装置获取目标数据。如果数据更新时间与第一异常数据对应的时间戳信息相同,或者数据更新时间在第一异常数据对应的时间戳信息之后,则数据处理装置获取与第一异常数据对应的异常数据标识信息,并将异常数据标识信息发送给客户端。

[0107] 可以理解的是,第一异常数据与目标数据,属于同一列。

[0108] 本公开提供的技术方案至少带来以下有益效果:在预置数据访问需求包括预置时间标识的情况下,为第三元数据表中的数据增加时间戳信息,然后根据时间戳信息判断目标第一存储设备中记录的数据是否为第一异常数据。如果数据更新时间对应的在第一异常数据对应的时间戳信息之前,则数据访问指令对应的目标数据是正常数据,可以获取目标数据。如果数据更新时间与第一异常数据对应的时间戳信息相同,或者数据更新时间在第一异常数据对应的时间戳信息之后,则确定数据访问指令对应的目标数据是第一异常数据,获取与第一异常数据对应的异常数据标识信息,并将异常数据标识信息发送给客户端。如此,确保目标数据的准确性,以便于提供根据目标数据得到的数据统计结论的准确性。

[0109] 可选的,在本公开实施例中,如图8所示,上述数据处理方法还可以包括下述步骤801至步骤803。

[0110] 步骤801、数据处理装置将至少一个第二元数据表中的数据,备份存储至第二存储设备。

[0111] 步骤802、在目标第一存储设备发生故障的情况下,数据处理装置从第二存储设备中获取目标数据,并将目标数据发送给客户端。

[0112] 步骤803、在目标第一存储设备的故障已解除的情况下,数据处理装置从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。

[0113] 在本公开实施例中,为了避免数据生产、同步存储出现异常时,在漫长的数据恢复过程中,数据处于不可用状态,如图9所示,与图1和图4相对应的,数据处理装置可以将至少一个第二元数据表中的数据,备份存储至第二存储设备中,进行全量备份存储,以缩短数据同步周期、止损异常数据带来的影响,提高同步存储系统的健壮性。

[0114] 可以理解的是,对于备份存储至第二存储设备而言,正常情况下不会对进行数据查询,因此,可以将所有的资源用来同步数据,从而大大提升数据同步速度,如果同步的数据是要恢复异常数据的数据,那么可以大大提升数据恢复速度,快速完成异常数据止损。

[0115] 需要说明的是,将至少一个第二元数据表中的数据,存储至多个第一存储设备,同时,备份存储至第二存储设备。第二存储设备为热备。即,多个第一存储设备和第二存储设备共同运行,在第一存储设备发生故障的情况下,热备(第二存设备)立即承担故障设备(目

标第一存储设备)的任务。

[0116] 本公开提供的技术方案至少带来以下有益效果:对于备份存储至第二存储设备而言,正常情况下不会对进行数据查询,如果正常提供查询的目标第一存储设备发生故障的情况下,可以将查询动态路由到第二存储设备,在目标第一存储设备的故障已解除的情况下,再将查询切回目标第一存储设备即可。如此,能够确保同步存储的数据的稳定性,提高同步存储数据方法的健壮性。并且,在数据备份过程中,数据同步存储时间短,占用资源少,即使存储大量数据也不影响存储性能。

[0117] 可选的,上述数据处理方法中,在第一存储设备中的第二异常数据满足预置条件的情况下,根据第二异常数据的存储时间,以及当前时间,在第二存储设备中查找与第二异常数据对应的异常恢复数据;将异常恢复数据作为第三元数据表中的数据,存储至多个第一存储设备;其中,预置条件为第二异常数据对应的数据列数大于预置数量。

[0118] 在本公开实施例中,在第二异常数据对应的数据列数大于预置数量的情况下,根据异常数据的时效性和系统对异常发现的及时性,确定第二存储设备中的哪些数据需要同步至第一存储设备。

[0119] 本公开提供的技术方案至少带来以下有益效果:在第二异常数据对应的数据列数大于预置数量的情况下,根据第二异常数据的存储时间,以及当前时间,在第二存储设备中查找与第二异常数据对应的异常恢复数据,然后将异常恢复数据作为第三元数据表中的数据,复用上述数据同步存储方案,能够提高实现数据同步存储方案所占用的硬件资源,提高数据存储、数据发送等数据处理效率。

[0120] 上述主要从方法的角度对本公开实施例的方案进行了介绍。上述方法所需的数据同步存储耗时短,占用资源少,对存储性能的影响较小。可以理解的是,数据处理装置为了实现上述功能,其包含了执行各个功能相应的硬件结构和软件模块中的至少一个。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,本公开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公开的范围。

[0121] 本公开实施例可以根据上述方法示例对数据处理装置进行功能单元的划分,例如,可以对应各个功能划分各个功能单元,也可以将两个或两个以上的功能集成在一个处理单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。需要说明的是,本公开实施例中对单元的划分是示意性的,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。

[0122] 图10是根据一示例性实施例示出的一种数据处理装置的结构示意图。参照图10所示,本公开实施例提供的数据处理装置,包括获取单元1001、处理单元1002和存储单元1003。

[0123] 获取单元1001,用于获取至少一个第一元数据表在相邻周期内的变化数据,生成至少一个第二元数据表,至少一个第一元数据表与至少一个第二元数据表一一对应;例如,如图2所示,获取单元1001可以用于执行步骤201。

[0124] 处理单元1002,用于从至少一个第二元数据表中,提取与至少一个预置数据访问

需求对应的数据,并将提取获得的数据存储至至少一个第三元数据表,至少一个第二元数据表的数量与至少一个第三元数据表的数量对应关系包括以下任一项:一对多关系、多对一关系,以及多对多关系,至少一个第三元数据表与至少一个预置数据访问需求一一对应;例如,如图2所示,处理单元1002可以用于执行步骤202。

[0125] 存储单元1003,用于将至少一个第三元数据表中的数据,存储至多个第一存储设备,每个第一存储设备用于存储至少一个第三元数据表中的数据。例如,如图2所示,存储单元1003可以用于执行步骤203。

[0126] 可选的,如图10所示,上述数据处理装置还包括:确定单元1004和反馈单元1005。

[0127] 确定单元1004,用于上述将至少一个第三元数据表中的数据,存储至多个第一存储设备之后,响应于客户端发送的数据访问指令,确定目标第一存储设备,数据访问指令用于请求访问目标数据,目标第一存储设备为多个第一存储设备中存储目标数据的第一存储设备;例如,如图5所示,确定单元1004可以用于执行步骤501。

[0128] 反馈单元1005,用于从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。例如,如图5所示,反馈单元1005可以用于执行步骤502。

[0129] 可选的,如图10所示,处理单元1002,具体用于:获取至少一个预置数据访问需求中的每个预置数据访问需求;从第二元数据表中,依次提取与每个预置数据访问需求对应的列数据,一个列数据对应一个数据标识;按照数据标识重新连接提取获得的列数据,并将连接的结果对应的数据,依次存储至至少一个第三元数据表中的每个第三元数据表中。例如,如图3所示,处理单元1002可以用于执行步骤301至步骤303。

[0130] 可选的,如图10所示,存储单元1003还用于上述将至少一个第三元数据表中的数据,存储至多个第一存储设备之前,在预置数据访问需求包括预置时间标识的情况下,为预置数据访问需求对应的第三元数据表中的数据增加时间戳信息;在上述数据处理装置中,数据访问指令还包括目标数据对应的数据更新时间;例如,如图7所示,存储单元1003可以用于执行步骤701。

[0131] 反馈单元1005,具体用于:在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间在第一异常数据对应的时间戳信息之前,则从目标第一存储设备中获取目标数据,并将目标数据发送给客户端;在目标第一存储设备中存储第一异常数据的情况下,如果数据更新时间与第一异常数据对应的时间戳信息相同,或者数据更新时间在第一异常数据对应的时间戳信息之后,则获取与第一异常数据对应的异常数据标识信息,并将异常数据标识信息发送给客户端。例如,如图7所示,反馈单元1005可以用于执行步骤702和步骤703。

[0132] 可选的,如图10所示,上述数据处理装置还包括:

[0133] 存储单元1003,还用于将至少一个第二元数据表中的数据,备份存储至第二存储设备中;例如,如图8所示,存储单元1003可以用于执行步骤801。

[0134] 反馈单元1005,还用于在目标第一存储设备发生故障的情况下,从第二存储设备中获取目标数据,并将目标数据发送给客户端;例如,如图8所示,反馈单元1005可以用于执行步骤802。

[0135] 反馈单元1005,还用于在目标第一存储设备的故障已解除的情况下,从目标第一存储设备中获取目标数据,并将目标数据发送给客户端。例如,如图8所示,反馈单元1005可

以用于执行步骤803。

[0136] 可选的,如图10所示,上述数据处理装置还包括:

[0137] 存储单元1003,还用于在第一存储设备中的第二异常数据满足预置条件的情况下,根据第二异常数据的存储时间,以及当前时间,在第二存储设备中查找与第二异常数据对应的异常恢复数据;将异常恢复数据作为第三元数据表中的数据,存储至多个第一存储设备;其中,预置条件为第二异常数据对应的数据列数大于预置数量。

[0138] 可选的,如图10所示,上述数据处理装置还包括:

[0139] 确定单元1004,还用于在目标第一存储设备的负载状态大于预置阈值的情况下,响应于客户端发送的数据访问指令,确定多个第一存储设备中的协同第一存储设备,协同第一存储设备用于存储于目标数据中的全部或部分数据;例如,如图6所示,确定单元1004可以用于执行步骤601。

[0140] 反馈单元1005,还用于根据数据包含关系,从至少一个动态第一存储设备中获取目标数据,并将目标数据发送给客户端;其中,数据包含关系为:协同第一存储设备中的存储数据,与目标数据之间的包含关系;至少一个动态第一存储设备包括:协同第一存储设备,或者,协同第一存储设备和目标第一存储设备。例如,如图6所示,反馈单元1005可以用于执行步骤602。

[0141] 可选的,如图10所示,反馈单元1005,具体用于:在协同第一存储设备存储目标数据中的全部数据的情况下,从协同第一存储设备中获取目标数据;在协同第一存储设备存储目标数据中的全部数据的情况下,根据目标第一存储设备的负载状态,从目标第一存储设备和协同第一存储设备中分别获取部分目标数据;在协同第一存储设备存储目标数据中的部分数据的情况下,从协同第一存储设备和目标第一存储设备中分别获取部分目标数据。

[0142] 关于上述实施例中的装置,其中各个单元执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0143] 图11是本公开提供的一种电子设备的结构示意图,该电子设备可以为上述数据处理装置。如图11所示,该电子设备可以包括处理器1101、用于存储处理器1101可执行指令的存储器1102;其中,上述处理器1101被配置为执行上述指令,以实现上述实施例中的数据处理方法。

[0144] 另外,电子设备还可以包括通信总线1103以及至少一个通信接口1104。

[0145] 处理器1101可以是一个中央处理器(central processing units,CPU),微处理单元,专用集成电路(Application Specific Integrated Circuit,ASIC),或一个或多个用于控制本公开方案程序执行的集成电路。

[0146] 通信总线1103为信号通路,用于在上述组件之间传送信息。

[0147] 通信接口1104,使用任何收发器一类的装置,用于与其他设备或通信网络通信,如以太网,无线接入网(radio access network,RAN),无线局域网(wireless local area networks,WLAN)等。

[0148] 存储器1102可以是只读存储器(read-only memory,ROM)或可存储静态信息和指令的其他类型的静态存储设备,随机存取存储器(random access memory,RAM)或者可存储信息和指令的其他类型的动态存储设备,也可以是电可擦可编程只读存储器

(electrically erasable programmable read-only memory,EEPROM)、只读光盘(compact disc read-only memory,CD-ROM)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质,但不限于此。存储器1102可以是独立存在,通过通信总线1103与处理器1101相连接。存储器1102也可以和处理器1101集成在一起。

[0149] 其中,存储器1102用于存储执行本公开方案的指令,并由处理器1101来控制执行。处理器1101用于执行存储器1102中存储的程序或指令,从而实现本公开方法中的功能。

[0150] 作为一个示例,结合图10,数据处理装置中的获取单元1001、处理单元1002和存储单元1003实现的功能与图11中的处理器1101的功能相同。

[0151] 在具体实现中,作为一种实施例,处理器1101可以包括一个或多个CPU,例如图11中的CPU0和CPU1。

[0152] 在具体实现中,作为一种实施例,电子设备可以包括多个处理器1101,这些处理器1101中的每一个可以是一个单核(single-CPU)处理器,也可以是一个多核(multi-CPU)处理器。这里的处理器1101可以指一个或多个设备、电路、和/或用于处理数据(例如计算机程序指令)的处理核。

[0153] 在具体实现中,作为一种实施例,电子设备还可以包括输出设备1105和输入设备1106。输出设备1105和处理器1101通信,可以以多种方式来显示信息。例如,输出设备1105可以是液晶显示器(liquid crystal display,LCD),发光二极管(light emitting diode,LED)显示设备,阴极射线管(cathode ray tube,CRT)显示设备,或投影仪(projector)等。输入设备1106和处理器1101通信,可以以多种方式接受用户的输入。例如,输入设备1106可以是鼠标、键盘、触摸屏设备或传感设备等。

[0154] 本领域技术人员可以理解,图11中示出的结构并不构成对电子设备的限定,可以包括比图示更多或更少的组件,或者组合某些组件,或者采用不同的组件布置。图11中的电子设备可以是服务器,也可以是客户端或其他设备。

[0155] 另外,本公开还提供一种计算机可读存储介质,计算机可读存储介质上存储程序或指令,当上述可读存储介质中的指令被处理器执行时,使得电子设备能够执行如上述实施例所提供的数据处理方法。可选的,可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0156] 另外,本公开还提供一种计算机程序产品,包括计算机程序/指令,该计算机程序产品被存储在非易失的可读存储介质中,计算机程序产品被至少一个处理器执行时,使得电子设备执行如上述实施例所提供的数据处理方法。

[0157] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其他实施方案。本公开旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由权利要求指出。

[0158] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限制。

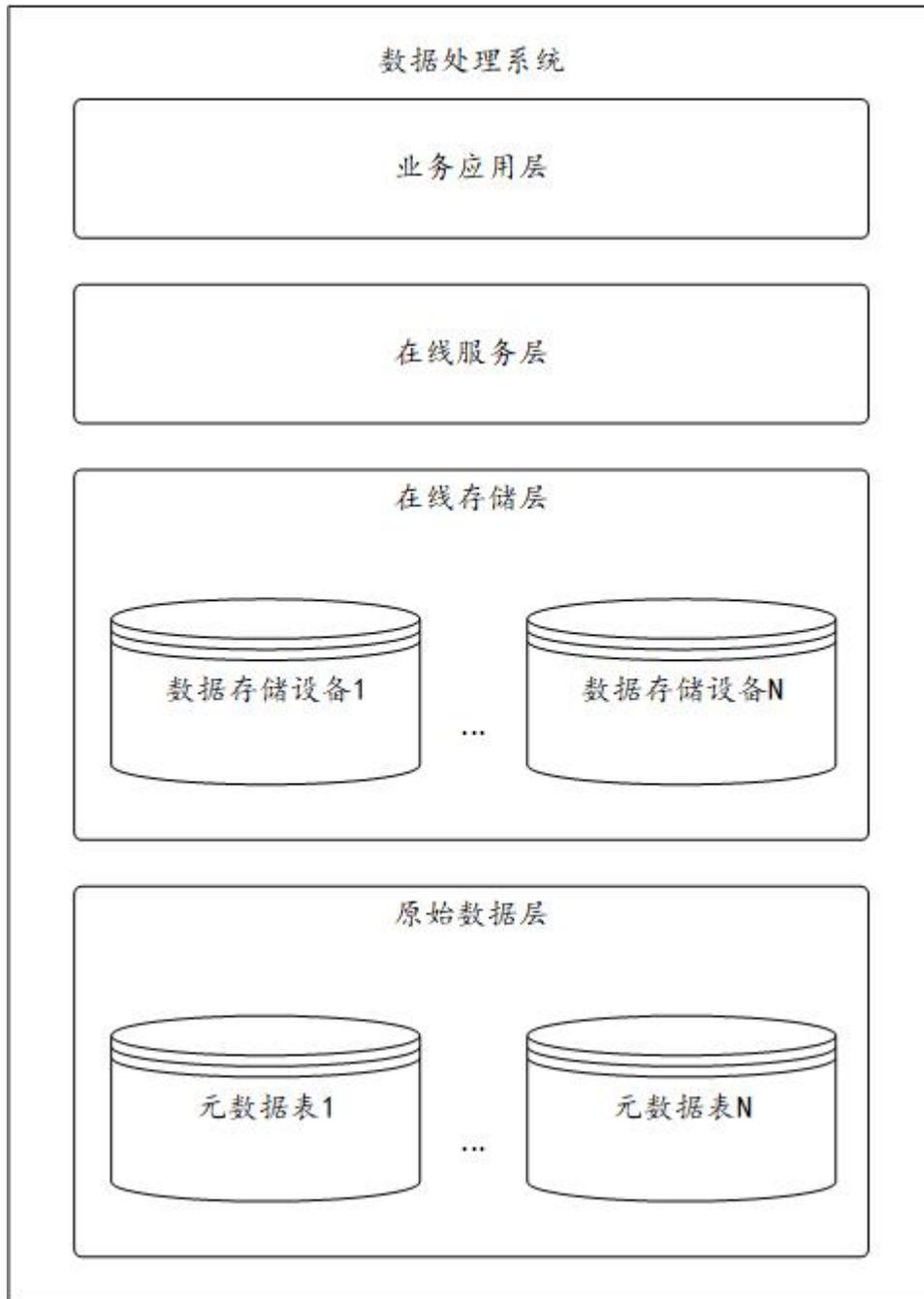


图1

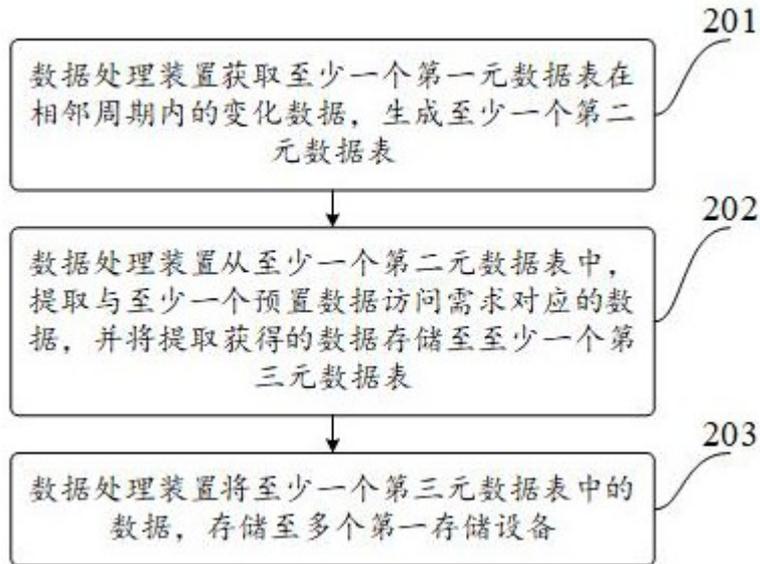


图2

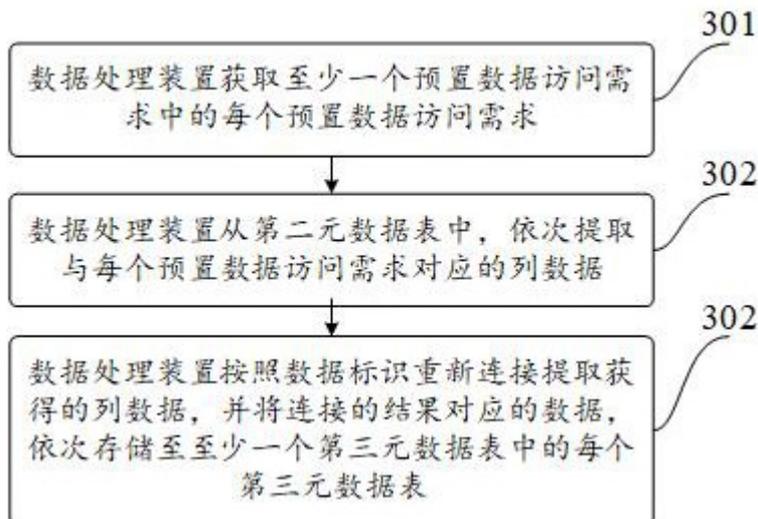


图3

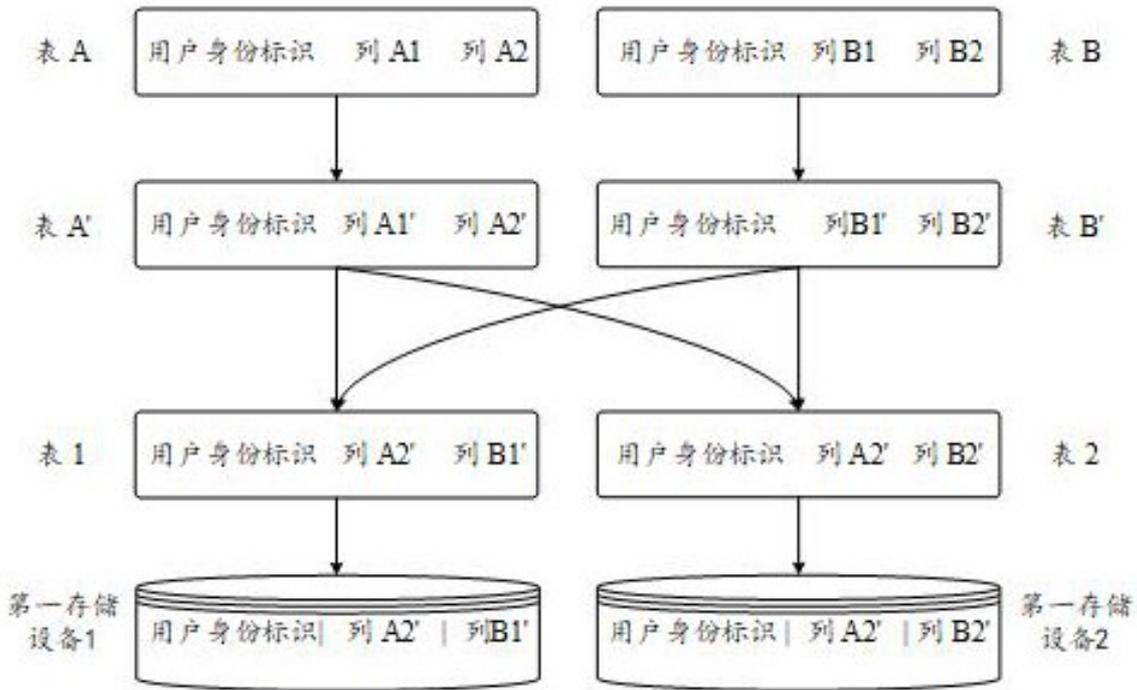


图4

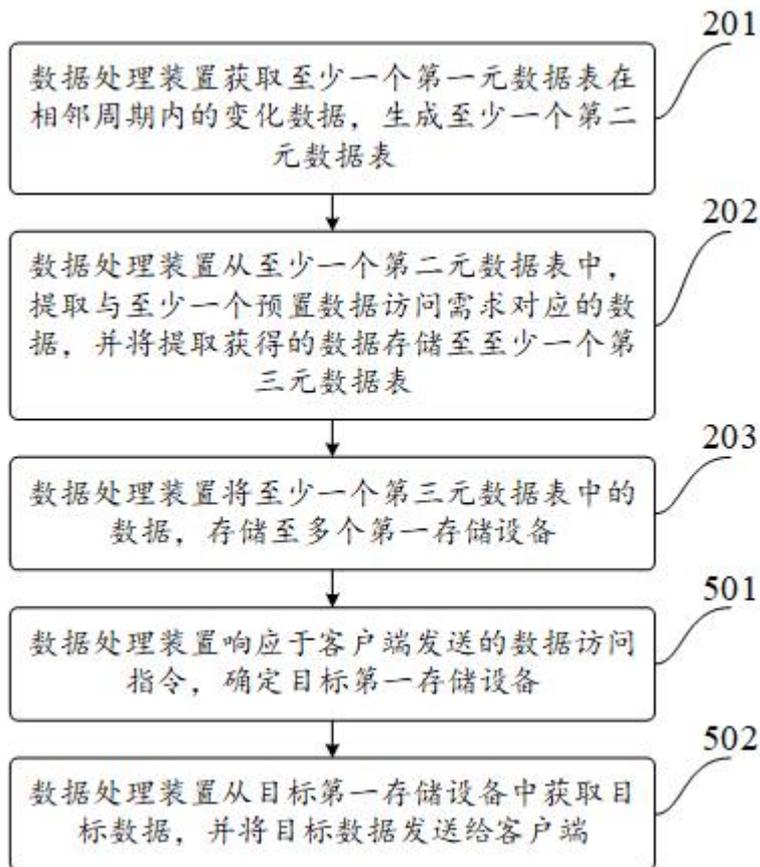


图5

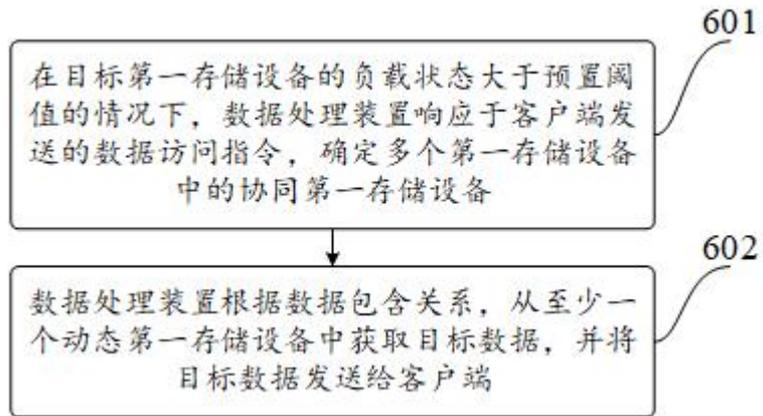


图6

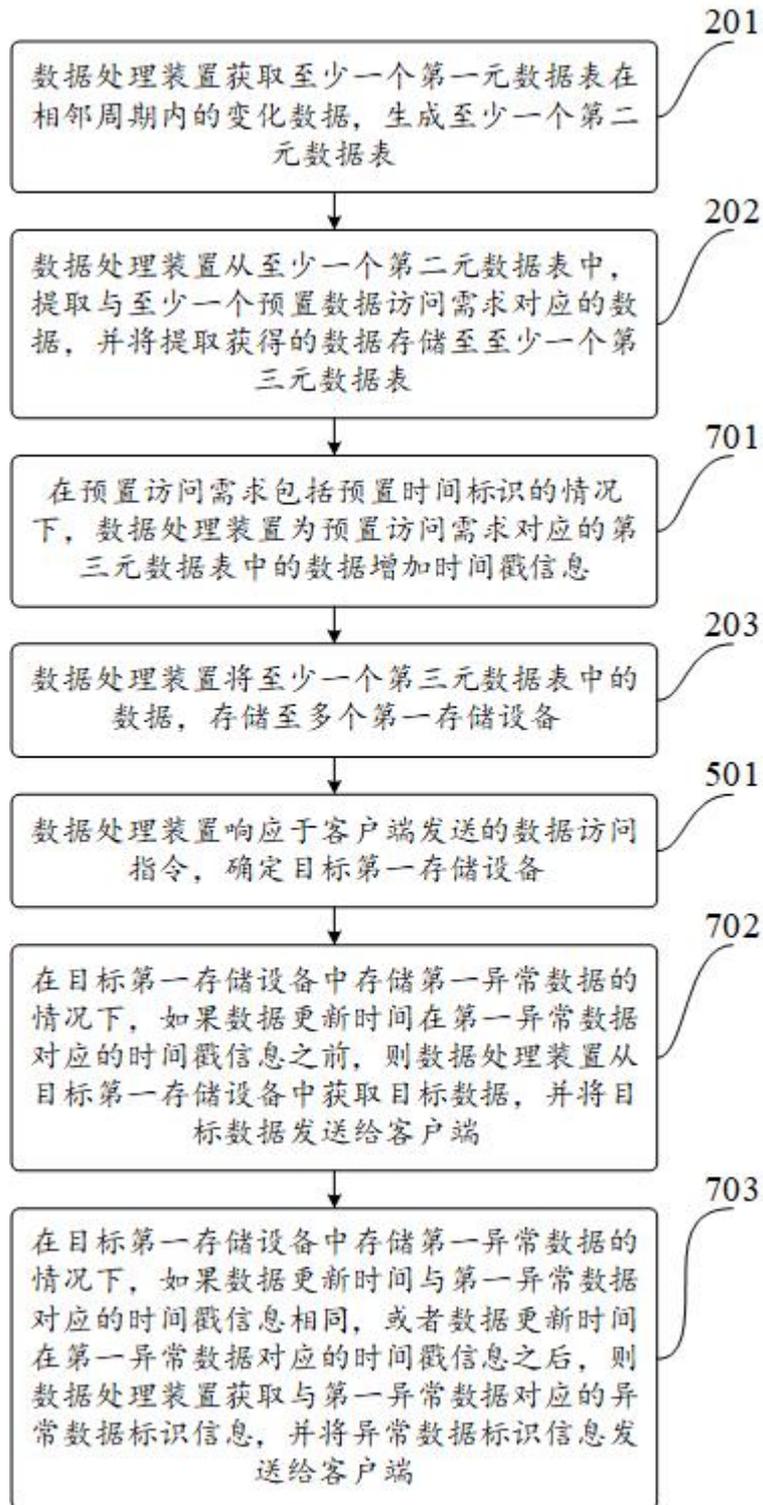


图7

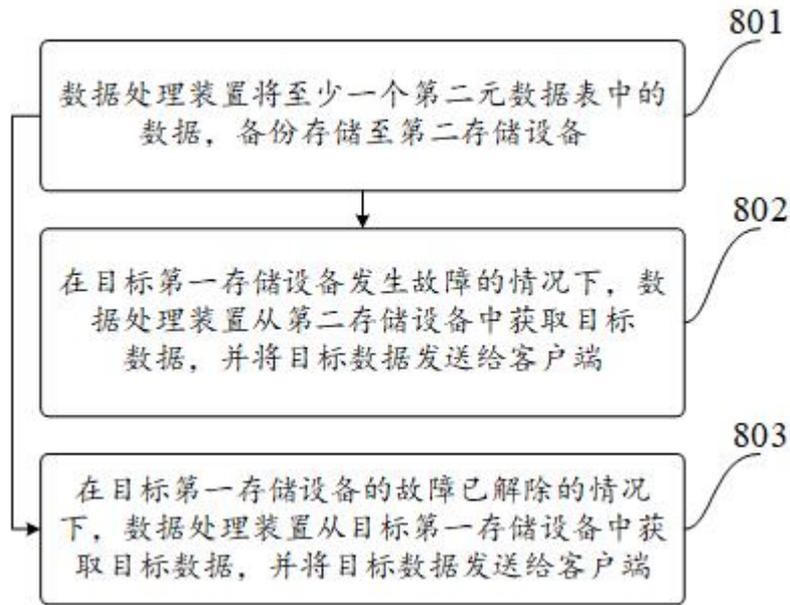


图8

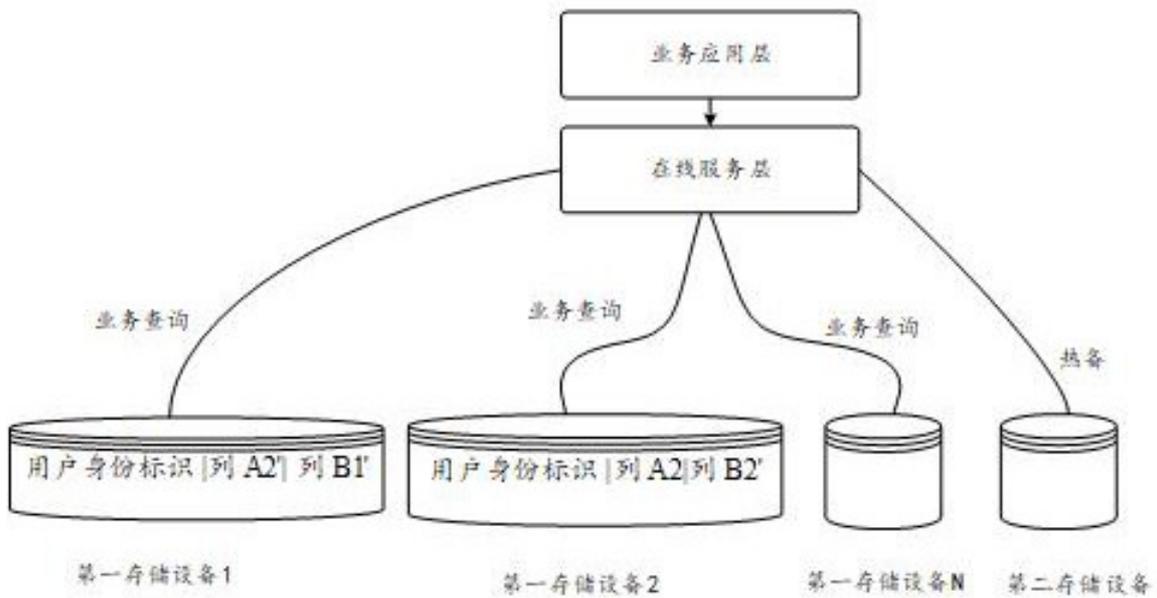


图9

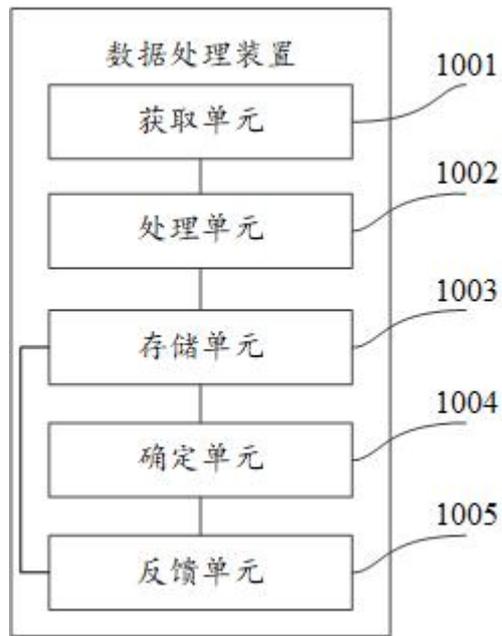


图10

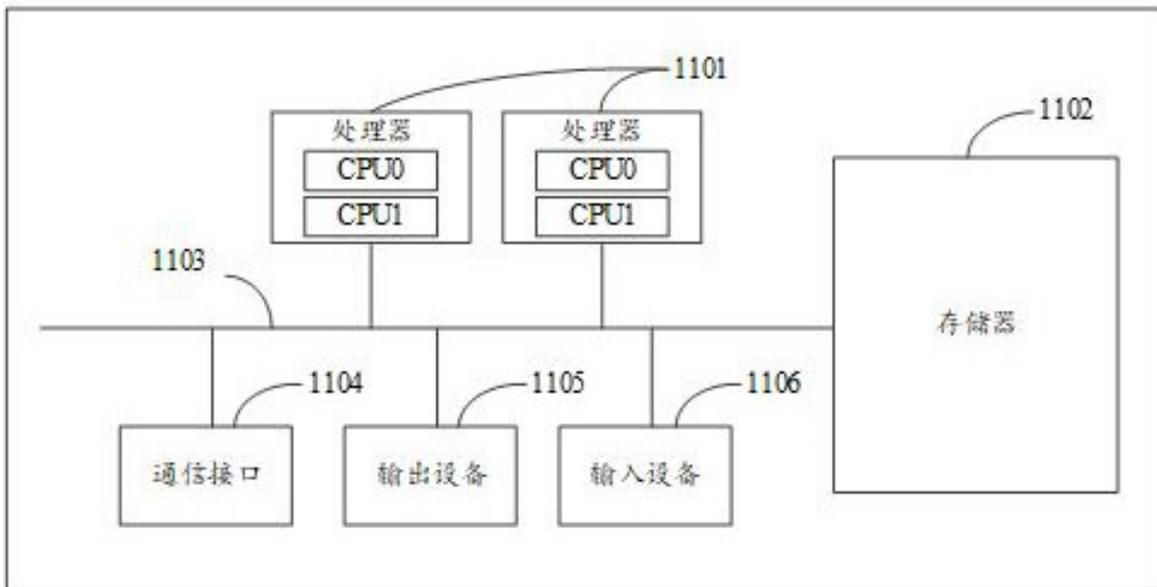


图11