



(12) 发明专利申请

(10) 申请公布号 CN 117687564 A

(43) 申请公布日 2024. 03. 12

(21) 申请号 202211078657.0

(22) 申请日 2022.09.05

(71) 申请人 华为云计算技术有限公司

地址 550025 贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心

(72) 发明人 吴延赞 魏明昌

(74) 专利代理机构 深圳市深佳知识产权代理事

务所(普通合伙) 44285

专利代理师 陈松浩

(51) Int. Cl.

G06F 3/06 (2006.01)

G06F 16/182 (2019.01)

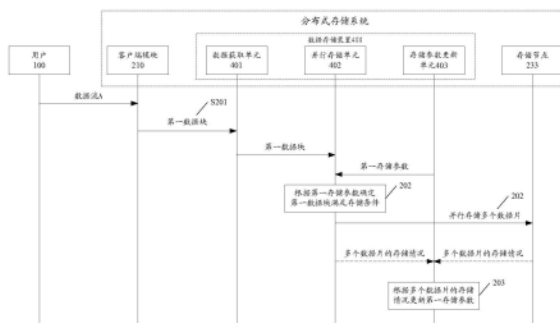
权利要求书3页 说明书19页 附图4页

(54) 发明名称

一种数据存储方法、装置及相关设备

(57) 摘要

本申请提供了一种数据存储方法,包括:获取多个数据片,所述多个数据片属于第一数据流中的第一数据块;响应于第一存储参数指示所述第一数据块满足存储条件,将所述多个数据片并行存储到分布式存储系统,所述第一存储参数根据所述第一数据流中位于所述第一数据块之前的数据块的存储情况确定,所述分布式存储系统包括多个存储节点;根据所述多个数据片被存储的情况更新所述第一存储参数。如此,并行存储接收到的数据流,可以提高分布式存储系统的吞吐率,实现数据的高效率存储。本申请还提供了对应的装置、计算设备集群、计算机可读存储介质以及计算机程序产品。



1. 一种数据存储方法,其特征在于,所述方法应用于分布式存储系统中的计算代理模块,所述方法包括:

获取多个数据片,所述多个数据片属于第一数据流中的第一数据块;

响应于第一存储参数指示所述第一数据块满足存储条件,将所述多个数据片并行存储到所述分布式存储系统,所述第一存储参数根据所述第一数据流中位于所述第一数据块之前的数据块的存储情况确定,所述分布式存储系统包括多个存储节点;

根据所述多个数据片被存储的情况更新所述第一存储参数。

2. 根据权利要求1所述的方法,其特征在于,所述多个数据片包括第一数据片和第二数据片,所述分布式存储系统包括第一存储节点和第二存储节点,所述将所述多个数据片并行存储到分布式存储系统包括:

向所述第一存储节点发送所述第一数据片和所述第二数据片,所述第一存储节点上用于存储所述第一数据片的线程和用于存储所述第二数据片的线程不同且并行执行;或,

向所述第一存储节点发送所述第一数据片,向所述第二存储节点发送所述第二数据片。

3. 根据权利要求1或2所述的方法,其特征在于,所述方法还包括:

响应于所述第一存储参数指示所述第一数据流中位于第二数据块之前的数据块已被存储到所述分布式存储系统,确定所述第一数据块满足所述存储条件;

其中,所述第二数据块为所述第一数据流中位于所述第一数据块之前的数据块,所述第一数据流中所述第二数据块与所述第一数据块之间间隔的数据块的数量小于预设数量。

4. 根据权利要求3所述的方法,其特征在于,所述根据所述第一数据块中数据片被存储的情况更新所述第一存储参数包括:

响应于所述第一数据流中位于所述第一数据块之前的数据块已被存储到所述分布式存储系统,且所述第一数据块中任一数据片已被存储到所述分布式存储系统,调整所述第一存储参数,以使所述第一存储参数指示第三数据块满足所述存储条件;

其中,所述第三数据块为所述第一数据流中位于所述第一数据块之后的数据块,所述第一数据流中所述第一数据块与所述第三数据块之间间隔的数据块的数量小于所述预设数量。

5. 根据权利要求4所述的方法,其特征在于,所述第一数据流是客户端模块发送的,在调整所述第一存储参数之后,所述方法还包括:

根据所述第一存储参数生成通知消息,所述通知消息用于指示所述第一数据块中数据片已被存储到所述分布式存储系统;

向所述客户端模块发送所述通知消息。

6. 根据权利要求1-5任一项所述的方法,其特征在于,所述方法还包括:

响应于第二计算代理模块出现故障,确定第二存储参数,所述第二存储参数用于指示第二数据流中数据块被存储到所述分布式存储系统的情况,所述第二计算代理模块用于将所述第二数据流存储到所述分布式存储系统;

根据所述第二存储参数确定第三数据片的标识,所述第三数据片属于所述第二数据流,所述第二数据流中位于所述第三数据片之前的数据片已被所述第二计算代理模块存储到所述存储系统;

根据所述第二存储参数,将所述第三数据片以及所述第二数据流中位于所述第三数据片之后的数据片存储到所述存储系统。

7. 根据权利要求6所述的方法,其特征在于,所述获取第二存储参数包括:

向所述多个存储节点发送查询请求,所述查询请求包括所述第二数据流的标识;

获取所述多个存储节点上报的标识集合,所述多个存储节点包括第三存储节点,所述第三存储节点上报的标识集合包括所述第三存储节点所存储的、所述第二数据流中数据片的标识;

根据所述多个存储节点上报的标识集合确定所述第二存储参数。

8. 根据权利要求7所述的方法,其特征在于,所述查询请求包括第一查询请求,所述方法还包括:

获取第四数据块的标识,所述第四数据块属于所述第二数据流,所述第四数据块包括多个数据片;

根据所述第四数据块的标识所述生成所述第一查询请求,所述第一查询请求用于查询所述第四数据块中数据片被存储到所述分布式存储系统的情况。

9. 一种数据存储装置,其特征在于,所述装置应用于分布式存储系统中的计算代理模块,所述装置包括:

数据获取单元,用于获取多个数据片,所述多个数据片属于第一数据流中的第一数据块;

并行存储单元,用于响应于第一存储参数指示所述第一数据块满足存储条件,将所述多个数据片并行存储到所述分布式存储系统,所述第一存储参数根据所述第一数据流中位于所述第一数据块之前的数据块的存储情况确定,所述分布式存储系统包括多个存储节点;

存储参数更新单元,用于根据所述多个数据片被存储的情况更新所述第一存储参数。

10. 根据权利要求9所述的装置,其特征在于,所述多个数据片包括第一数据片和第二数据片,所述分布式存储系统包括第一存储节点和第二存储节点;

所述并行存储单元,具体用于向所述第一存储节点发送所述第一数据片和所述第二数据片,所述第一存储节点上用于存储所述第一数据片的线程和用于存储所述第二数据片的线程不同且并行执行;或,向所述第一存储节点发送所述第一数据片,向所述第二存储节点发送所述第二数据片。

11. 根据权利要求9或10所述的装置,其特征在于,

所述并行存储单元,还用于响应于所述第一存储参数指示所述第一数据流中位于第二数据块之前的数据块已被存储到所述分布式存储系统,确定所述第一数据块满足所述存储条件;其中,所述第二数据块为所述第一数据流中位于所述第一数据块之前的数据块,所述第一数据流中所述第二数据块与所述第一数据块之间间隔的数据块的数量小于预设数量。

12. 根据权利要求11所述的装置,其特征在于,

所述存储参数更新单元,具体用于响应于所述第一数据流中位于所述第一数据块之前的数据块已被存储到所述分布式存储系统,且所述第一数据块中任一数据片已被存储到所述分布式存储系统,调整所述第一存储参数,以使所述第一存储参数指示第三数据块满足所述存储条件;其中,所述第三数据块为所述第一数据流中位于所述第一数据块之后的数

据块,所述第一数据流中所述第一数据块与所述第三数据块之间间隔的数据块的数量小于所述预设数量。

13. 根据权利要求12所述的装置,其特征在于,所述第一数据流是客户端模块发送的,在调整所述第一存储参数之后,所述装置还包括通知单元;

所述通知单元,用于根据所述第一存储参数生成通知消息,所述通知消息用于指示所述第一数据块中数据片已被存储到所述分布式存储系统;向所述客户端模块发送所述通知消息。

14. 根据权利要求9-13任一项所述的装置,其特征在于,所述装置还包括存储参数确定单元和数据片确定单元;

所述存储参数确定单元,用于响应于第二计算代理模块出现故障,确定第二存储参数,所述第二存储参数用于指示第二数据流中数据块被存储到所述分布式存储系统的情况,所述第二计算代理模块用于将所述第二数据流存储到所述分布式存储系统;

所述数据片确定单元,用于根据所述第二存储参数确定第三数据片的标识,所述第三数据片属于所述第二数据流,所述第二数据流中位于所述第三数据片之前的数据片已被所述第二计算代理模块存储到所述分布式存储系统;

所述并行存储单元,还用于根据第二存储参数,将所述第三数据片以及所述第四数据流中位于所述第三数据片之后的数据片存储到所述分布式存储系统。

15. 根据权利要求14所述的装置,其特征在于,

所述存储参数确定单元,具体用于向所述多个存储节点发送查询请求,所述查询请求包括所述第二数据流的标识;获取所述多个存储节点上报的标识集合,所述多个存储节点包括第三存储节点,所述第三存储节点上报的标识集合包括所述第一存储节点所存储的、所述第二数据流中数据片的标识;根据所述多个存储节点上报的标识集合确定所述第二存储参数。

16. 根据权利要求15所述的装置,其特征在于,

所述存储参数确定单元,还用于获取第四数据块的标识,所述第四数据块属于所述第二数据流,所述第四数据块包括多个数据片;根据所述第四数据块的标识所述生成所述第一查询请求,所述第一查询请求用于查询所述第四数据块中数据片被存储到所述分布式存储系统的情况。

17. 一种计算设备集群,其特征在于,所述计算设备集群包括至少一个计算设备,每个计算设备包括处理器以及存储器:

所述存储器用于存储指令;

所述处理器用于根据所述指令,以使得所述计算设备集群执行权利要求1-8中任一项所述的方法。

18. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有指令,当其在计算设备上运行时,使得所述计算设备执行如权利要求1至8任一项所述的方法。

19. 一种包含指令的计算机程序产品,当其在计算设备上运行时,使得所述计算设备执行如权利要求1至8任一项所述的方法。

一种数据存储方法、装置及相关设备

技术领域

[0001] 本申请涉及数据处理技术领域,尤其涉及一种数据存储方法、装置及相关设备。

背景技术

[0002] 随着分布式存储系统规模的扩大,基于存储计算分离架构的分布式存储系统逐渐得到了应用。基于存储计算分离架构的分布式存储系统可以包括多个客户端模块、多个计算代理模块和多个存储模块。客户端模块接收到的数据流可以经过计算代理模块的计算处理,并存储到一个或多个存储模块中。存储模块又可以被称为存储节点。

[0003] 目前,计算代理模块可以对通过客户端模块接收到的数据流进行分块,将数据流拆分为多个数据块(segment)。拆分出的多个数据块可以被存储到一个存储节点,也可以被分别存储到多个存储节点。计算代理模块可以按顺序存储多个数据块。在存储某个数据块时,计算代理模块可以将该数据块拆分为多个数据片(fragment),并按顺序存储多个数据片。但是,目前基于存储计算分离架构的分布式存储系统串行存储数据,数据的吞吐率低。

发明内容

[0004] 有鉴于此,本申请提供了一种数据存储方法,用于并行存储接收到的数据流,从而提高分布式存储系统的吞吐率,实现数据的高效率存储。本申请还提供了对应的装置、计算设备集群、计算机可读存储介质以及计算机程序产品。

[0005] 第一方面,本申请提供了一种数据存储方法,该方法由数据存储装置执行。数据存储装置例如可以是运行于分布式存储系统的计算代理模块的软件装置。具体实现时,数据存储装置获取多个数据片,多个数据片属于待存储的第一数据流中的第一数据块。接着,数据存储装置根据第一存储参数判断第一数据块是否满足存储条件。其中,第一存储参数根据第一数据流中位于第一数据块之前的数据块的存储情况确定。如果第一数据块满足存储条件,数据存储装置将多个数据片并行存储到分布式存储系统,并根据多个数据片被存储的情况更新第一存储参数。

[0006] 如此,在第一数据块满足存储条件时,数据存储装置可以并行存储第一数据块所包括的多个数据片。如果第一数据流中的多个数据块均满足存储条件,数据存储装置可以并行存储多个数据块中的多个数据片。并行存储多个数据片可以有效地提高分布式存储系统的吞吐率。并且,通过数据流中位于第一数据块之前的其他数据块的存储情况判断第一数据块是否满足存储条件,可以避免出现数据空洞。并且,通过第一存储参数记录第一数据流中已被存储的数据块的情况,如果计算代理单元出现故障,可以根据第一存储参数进行恢复。

[0007] 在一种可能的实施方式中,在并行存储多个数据片时,数据存储装置可以将多个数据片并行存储到一个或多个存储节点。例如,如果数据存储装置获取的多个数据片包括第一数据片和第二数据片,且分布式存储系统包括第一存储节点和第二存储节点。那么在存储第一数据片和第二数据片时,数据存储装置可以同时向第一存储节点发送第一数据片

和第二数据片。第一存储节点通过两个并行执行的线程(或进程)并行存储第一数据片和第二数据片。或者,数据存储装置还可以向第一存储节点发送第一数据片,向第二存储节点发送第二数据片。第一数据片和第二数据片分别被第一存储节点和第二存储节点上不同的线程(或进程)存储。如此,并行将多个数据片存储到一个或多个存储节点,可以提升数据块的存储效率。

[0008] 在一种可能的实施方式中,第一数据块是否满足存储条件根据第一数据流中第二数据块之前的数据块的存储情况确定。其中,第二数据块为第一数据流中位于第一数据块之前的数据块,且第一数据流中位于第二数据块和第一数据块之间的数据块的数量小于预设数量。也就是说,如果第一数据流中首个未被存储的数据块,与第一数据块之间间隔的数据块的数量小于预设数量,那么第一数据块满足存储条件。可选地,第一存储参数根据第一数据流的头部连续存储部分的右边界确定。第一数据流的头部连续存储部分包括位于第一数据流头部的、连续的、一个或多个已被存储的数据片。相应地,在判断第一数据块是否满足存储条件时,可以判断第二数据块是否属于第一数据流的头部连续存储部分。如果第二数据块属于第一数据流的头部连续存储部分,说明第一数据流中第二数据块以及第二数据块之前的所有数据块已被存储,第一数据块满足存储条件。如此,在确定第二数据块以及第一数据流中第二数据块之前的数据块均被存储之后,再存储第一数据块,可以避免存储过程中出现数据空洞。可以理解的是,上文(包括下文)中所述的“存储”是指将数据片或数据块存储到分布式存储系统。

[0009] 在一些可能的实施方式中,随着第一数据块中数据片被存储到分布式存储系统,数据存储装置可以更新第一存储参数。具体地,如果第一数据流中位于第一数据块之前任一数据块已被存储到分布式存储系统,且第一数据块中任一数据片均被存储到分布式存储系统,那么数据存储装置可以调整第一存储参数。也就是说,如果第一数据流的头部连续存储部分的右边界移动到第一数据块,那么数据存储装置根据第一数据块在第一数据流中的位置调整第一存储参数。如果第一数据流的头部连续存储部分的右边界移动到第一数据块,那么调整后的第一存储参数可以指示第三数据块满足存储条件。其中,第三数据块为第一数据流中位于第一数据块之后的数据块,且第一数据流中位于第一数据块和第三数据块之间的数据块的数量小于预设数量。如此,随着第一数据流中数据块的存储逐渐调整第一存储参数,使得位于第一数据流后部的数据块可以得到存储。

[0010] 在一些可能的实施方式中,第一数据流是客户端模块接收并向数据存储装置发送的。相应地,数据存储装置可以在数据流被存储之后通知客户端模块。具体地,数据存储装置可以根据第一存储参数生成通知消息,并向客户端模块发送通知消息。其中,通知消息用于指示第一存储参数对应的头部连续存储部分已被存储。也就是说,如果第一存储参数是在第一数据块以及第一数据流中第一数据块之前的所有数据块被存储之后更新的,那么就第一存储参数生成通知消息可以用于指示第一数据块中数据片已被存储。这样,由于头部连续存储部分不会因故障出现丢失等问题,根据头部连续存储部分的存储情况向客户端模块发送通知消息,可以在避免出现数据空洞的情况下,让用户知晓第一数据流的存储情况。

[0011] 在一些可能的实施方式中,上述数据存储装置为运行于第一计算代理模块的软件装置。如果运行有数据存储装置的第二计算代理模块出现故障,那么第二计算代理模块执行的存储任务可以由第一计算代理模块上的数据存储装置执行,继续数据流的存储。如此,

即使分布式存储系统中某个计算代理模块出现了故障,也可以通过其他计算代理模块实现数据流的存储。

[0012] 在一些可能的实施方式中,故障前的第二计算代理模块用于存储第二数据流。那么在并行存储第二数据流中多个数据片之前,数据存储装置先获取第二存储参数,第二存储参数用于指示第二数据流中数据块被第二计算代理模块存储的情况。接着,计算代理模块根据第二存储参数确定第四数据片的标识。第四数据片为第二数据流中的数据片,且第二数据流中位于第四数据片之前的所有数据片已被第二计算代理模块存储到分布式存储系统。也就是说,第四数据片为第二数据流中头部连续存储部分之后首个数据片。这样,数据存储装置可以根据第二存储参数存储第四数据片,以及第二数据流中位于第四数据片之后的数据片。关于存储片的方法可以参见存储第一数据流中数据片的存储方法,这里不再赘述。这样,根据第二存储参数确定第四数据片并进行存储,可以避免出现数据空洞。

[0013] 在一些可能的实施方式中,数据存储装置还可以存储第三数据片之前的数据片。例如,如果第三数据片属于第五数据块,且第三数据块不为第五数据块中首个数据片,那么在存储第二数据流时,数据存储装置可以从第五数据块开始存储。也就是说,即使第五数据块中存在一个或多个数据片已被存储,数据存储装置仍然从第一数据块中首个数据片开始存储。这样,通过重复存储数据块中的数据片,可以进一步避免出现数据空洞。

[0014] 在一些可能的实施方式中,第二存储参数是数据存储装置通过查询得到的。具体地,在获取第二存储参数时,数据存储装置先向至少一个存储节点发送包括第二数据流的标识的查询请求。接着,数据存储装置可以接收到存储节点发送的标识集合。存储节点发送的标识集合包括本存储节点所存储的、第二数据流中数据片的标识。在获取到标识集合之后,数据存储装置根据标识集合确定第二存储参数。如此,通过向存储节点发送查询请求,可以确定第二数据流中数据片的存储情况,从而恢复存储第二数据流。

[0015] 在一些可能的实施方式中,存储节点可以根据数据块的标识记录数据的存储情况。相应地,在查询第二数据流中数据片的存储情况时,数据存储装置可以根据数据块的标识进行查询。具体地,对于第二数据流中的第四数据块,数据存储装置可以获取第四数据块的标识,并根据第四数据块的标识生成第一查询请求。这样,第一查询请求用于查询第四数据块中数据片被存储的情况。

[0016] 第二方面,本申请还提供了一种数据存储装置,所述装置应用于分布式存储系统中的计算代理模块,所述装置包括:数据获取单元,用于获取多个数据片,所述多个数据片属于第一数据流中的第一数据块;并行存储单元,用于响应于第一存储参数指示所述第一数据块满足存储条件,将所述多个数据片并行存储到所述分布式存储系统,所述第一存储参数根据所述第一数据流中位于所述第一数据块之前的数据块的存储情况确定,所述分布式存储系统包括多个存储节点;存储参数更新单元,用于根据所述多个数据片被存储的情况更新所述第一存储参数。

[0017] 在一些可能的实施方式中,所述多个数据片包括第一数据片和第二数据片,所述分布式存储系统包括第一存储节点和第二存储节点;所述并行存储单元,具体用于向所述第一存储节点发送所述第一数据片和所述第二数据片,所述第一存储节点上用于存储所述第一数据片的线程和用于存储所述第二数据片的线程不同且并行执行;或,向所述第一存储节点发送所述第一数据片,向所述第二存储节点发送所述第二数据片。

[0018] 在一些可能的实施方式中,所述并行存储单元,还用于响应于所述第一存储参数指示所述第一数据流中位于第二数据块之前的数据块已被存储到所述分布式存储系统,确定所述第一数据块满足所述存储条件;其中,所述第二数据块为所述第一数据流中位于所述第一数据块之前的数据块,所述第一数据流中所述第二数据块与所述第一数据块之间间隔的数据块的数量小于预设数量。

[0019] 在一些可能的实施方式中,所述存储参数更新单元,具体用于响应于所述第一数据流中位于所述第一数据块之前的数据块已被存储到所述分布式存储系统,且所述第一数据块中任一数据片已被存储到所述分布式存储系统,调整所述第一存储参数,以使所述第一存储参数指示第三数据块满足所述存储条件;其中,所述第三数据块为所述第一数据流中位于所述第一数据块之后的数据块,所述第一数据流中所述第一数据块与所述第三数据块之间间隔的数据块的数量小于所述预设数量。

[0020] 在一些可能的实施方式中,所述第一数据流是客户端模块发送的,在调整所述第一存储参数之后,所述装置还包括通知单元;所述通知单元,用于根据所述第一存储参数生成通知消息,所述通知消息用于指示所述第一数据块中数据片已被存储到所述分布式存储系统;向所述客户端模块发送所述通知消息。

[0021] 在一些可能的实施方式中,所述装置还包括存储参数确定单元和数据片确定单元;

[0022] 所述存储参数确定单元,用于响应于第二计算代理模块出现故障,确定第二存储参数,所述第二存储参数用于指示第二数据流中数据块被存储到所述分布式存储系统的情况,所述第二计算代理模块用于将所述第二数据流存储到所述分布式存储系统;所述数据片确定单元,用于根据所述第二存储参数确定第三数据片的标识,所述第三数据片属于所述第二数据流,所述第二数据流中位于所述第三数据片之前的数据片已被所述第二计算代理模块存储到所述分布式存储系统;所述并行存储单元,还用于根据第二存储参数,将所述第三数据片以及所述第四数据流中位于所述第三数据片之后的数据片存储到所述分布式存储系统。

[0023] 在一些可能的实施方式中,所述存储参数确定单元,具体用于向所述多个存储节点发送查询请求,所述查询请求包括所述第二数据流的标识;获取所述多个存储节点上报的标识集合,所述多个存储节点包括第三存储节点,所述第三存储节点上报的标识集合包括所述第一存储节点所存储的、所述第二数据流中数据片的标识;根据所述至少一个存储节点上报的标识集合确定所述第二存储参数。

[0024] 在一些可能的实施方式中,所述存储参数确定单元,还用于获取第四数据块的标识,所述第四数据块属于所述第二数据流,所述第四数据块包括多个数据片;根据所述第四数据块的标识所述生成所述第一查询请求,所述第一查询请求用于查询所述第四数据块中数据片被存储到所述分布式存储系统的情况。

[0025] 第三方面,本申请提供一种计算设备集群,所述计算设备包括至少一个计算设备,所述至少一个计算设备包括至少一个处理器和至少一个存储器;所述至少一个存储器用于存储指令,所述至少一个处理器执行所述至少一个存储器存储的该指令,以使所述计算设备集群执行上述第一方面或第一方面任一种可能实现方式中的数据存储方法。需要说明的是,该存储器可以集成于处理器中,也可以是独立于处理器之外。所述至少一个计算设备还

可以包括总线。其中,处理器通过总线连接存储器。其中,存储器可以包括可读存储器以及随机存取存储器。

[0026] 第四方面,本申请提供一种计算机可读存储介质,所述计算机可读存储介质中存储有指令,当其在至少一个计算设备上运行时,使得所述至少一个计算设备执行上述第一方面或第一方面的任一种实现方式所述的方法。

[0027] 第五方面,本申请提供了一种包含指令的计算机程序产品,当其在至少一个计算设备上运行时,使得所述至少一个计算设备执行上述第一方面或第一方面的任一种实现方式所述的方法。

[0028] 本申请在上述各方面提供的实现方式的基础上,还可以进行进一步组合以提供更多实现方式。

附图说明

[0029] 图1为本申请实施例提供的一示例性应用场景的示意图;

[0030] 图2为本申请实施例提供的另一示例性应用场景示意图;

[0031] 图3为本申请实施例提供的数据存储方法的一种流程示意图;

[0032] 图4为本申请实施例提供的又一示例性应用场景示意图;

[0033] 图5为本申请实施例提供的数据存储方法的另一种流程示意图;

[0034] 图6为本申请实施例提供的一种计算设备的结构示意图;

[0035] 图7为本申请实施例提供的一种计算设备集群的结构示意图。

具体实施方式

[0036] 下面将结合本申请中的附图,对本申请提供的实施例中的方案进行描述。

[0037] 本申请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术语在适当情况下可以互换,这仅仅是描述本申请的实施例中对相同属性的对象在描述时所采用的区分方式。

[0038] 分布式存储系统可以包括一个或多个存储节点。每个存储节点都可以用于存储数据。通过多个存储节点存储数据,可以扩大存储系统的存储容量。为了提高分布式存储系统的动态扩展能力,目前可以将分布式存储系统的存储功能与计算功能分离。比如,可以通过客户端(client)模块接收待存储的数据流,通过计算代理(broker)模块实现分布式存储系统的计算功能,通过存储(store)节点实现分布式存储系统的存储功能。

[0039] 例如,参见图1给出的示例性应用场景。在图1所示的应用场景中,分布式存储系统200包括客户端模块210、计算代理模块221、计算代理模块222、存储节点231、存储节点232、存储节点233和存储节点234。用户100可以(通过客户端或用户终端)向分布式存储系统200发送数据,以使分布式存储系统200存储数据。其中,数据可以以数据流的形式被分布式存储系统200中的客户端模块210接收。客户端模块210在接收到数据流后,可以将数据流转发给计算代理模块221。计算代理模块221可以将数据流中的数据拆分为多个数据块,并将多个数据块存储到一个或多个存储节点。

[0040] 例如,假设用户100将待存储的数据以数据流A的形式发送给分布式存储系统200

中的客户端2模块210。客户端模块210可以接收数据流A,并将数据流A转发给计算代理模块221。计算代理模块221根据数据流A中的顺序,先将数据流A拆分为多个数据块(图1所示场景中为3个)。接着,计算代理模块221再将每个数据块拆分为多个(图1所示场景中为3个)数据片,并将多个数据片存储到存储节点中。

[0041] 为便于说明,在本申请实施例中,假设数据块1与存储节点231对应,数据块2与存储节点232对应,数据块3与存储节点233对应。

[0042] 如果数据流中部分数据未被存储,而数据流中位于这部分数据之前的部分数据已被存储,且数据流中位于这部分数据之后的部分数据已被存储,那么可以称该数据流存在数据空洞。也就是说,如果数据块1和数据块3已存储完成,但是数据块2或数据块2中任一数据片并未被存储,则数据流A中存在数据空洞。如果数据片1和数据片3已存储完成,但是数据片2并未被存储,数据流A中同样存在数据空洞。

[0043] 数据空洞可能导致被存储的数据无法恢复。因此,为了避免出现数据空洞,计算代理模块221按顺序存储数据流中的数据块。其中,计算代理模块221存储数据块所遵循的顺序为数据流中数据块被客户端模块210接收的顺序,即用户100需要存储的数据中各个数据块的实际顺序。也就是说,计算代理模块221可以先向存储节点231发送数据块1。在数据块1被写入存储节点231之后,计算代理模块221向存储节点232发送数据块2。相似地,在数据块2被写入存储节点232之后,计算代理模块221向存储节点233发送数据块3。在将数据块1写入存储节点231时,计算代理模块221先向存储节点231发送数据块1中的数据片1。在数据块1被写入存储节点231之后,计算代理模块221再向存储节点发送数据块1中的数据片1。可以理解的是,如果数据流A中存在三个以上的数据块,计算代理模块221按照接收数据块的先后顺序将数据块写入存储节点;如果数据流A中每个数据块包括三个以上的数据片,计算代理模块221按照数据片在数据块中的顺序将数据片写入对应的存储节点。

[0044] 通过上述介绍可以看出,由于数据流中的数据块被计算代理模块串行写入存储节点,而对于每个数据块,计算代理模块又是串行写入数据块中的数据片。因此,对于客户端接收到的一条数据流,计算代理模块在同一时间向存储节点写入一个数据片。这样,单位时间内对该数据流的吞吐量,取决于计算代理模块与单个存储节点之间的带宽。数据流的吞吐率低,无法实现数据流的高效率存储。

[0045] 基于此,本申请实施例提供了一种数据存储方法,该方法可以由数据存储装置执行,实现数据流的并行存储,以提高数据流的吞吐率。具体地,数据存储装置先获取第一数据流中的第一数据块,第一数据流还包括多个数据块。然后,数据存储装置根据第一存储参数判断第一数据块是否满足存储条件。如果第一存储参数指示第一数据块满足存储条件,数据存储装置并行存储第一数据块所包括的多个数据片。其中,数据存储装置可以运行于图1所示实施例中数据代理模块221(或数据代理模块222)的软件模块,第一数据流可以是图1所示实施例中数据流A。第一存储参数根据第一数据流中位于第一数据块之前的数据块被存储的情况确定。在存储第一数据块所包括的数据片时,数据存储装置可以根据第一数据块中数据片被存储的情况更新第一存储参数。比如,数据存储装置(即上述数据存储装置)在接收到数据块1之后,可以同时将其中的数据片1、数据片2和数据片3并行写入存储节点231,并根据数据块1中数据片被存储的情况调整第一存储参数。同理,数据存储装置在接收到数据块2之后,可以同时将其中的数据片4、数据片5和数据片6并行写入存

储节点232,并根据数据块2中数据片被存储的情况调整第一存储参数。数据存储装置在接收到数据块3之后,可以判断第一存储参数是否允许数据存储装置并行存储数据块3。

[0046] 比如,存储条件包括数据流中位于第一数据块前的数据块已被写入存储节点。那么如果第一存储参数指示数据块1已被写入存储节点231,数据存储装置确定数据块2满足存储条件。数据存储装置同时将数据块2中的数据片4、数据片5和数据片6并行写入存储节点232,并根据数据块3中数据片被存储的情况调整第一存储参数。如果第一存储参数指示数据块1中任意一个或多个数据片尚未被写入存储节点231,数据存储装置确定数据块2不满足存储条件,暂停写入数据块2。

[0047] 或者,存储条件包括数据流中位于第一数据块之前、且与第一数据块之间间隔至少一个数据块的所有数据块已被写入存储节点。那么如果第一存储参数指示数据块1已被写入存储节点231,数据存储装置确定数据块3满足存储条件。数据存储装置同时将数据块3中的数据片7、数据片8和数据片9并行写入存储节点233,并根据数据块3中数据片被存储的情况调整第一存储参数。如果第一存储参数指示数据块1中任意一个或多个数据片尚未被写入存储节点231,数据存储装置确定数据块3不满足存储条件,暂停写入数据块3。

[0048] 如此,在数据块满足存储条件时,数据存储装置可以并行写入数据块所包括的多个数据片。如果多个数据块均满足存储条件,数据存储装置可以并行存储多个数据块中的多个数据片。并行存储多个数据片可以有效地提高存储系统的吞吐率。并且,通过数据流中位于第一数据块之前的其他数据块的存储情况判断第一数据块是否满足存储条件,可以避免出现数据空洞。并且,通过第一存储参数记录第一数据流中已被存储的数据块的情况,如果计算代理单元出现故障,可以根据第一存储参数进行恢复。

[0049] 作为一种示例,上述数据存储装置可以被部署于云端,用于为用户100提供数据存储的云服务。例如,在图1所示的应用场景中,分布式存储系统200具体可以是位于云端的计算设备(或者计算设备集群)。并且,该计算设备可以对外提供客户端模块210,用于实现与用户100的交互,如获取用户100提供的数据流等。实际应用时,该客户端210例如可以是运行在用户侧设备上的应用程序,或者可以是计算设备对外提供的网络浏览器等。并且,该计算机设备可以提供计算代理模块221和计算代理模块222,以实现本申请实施例中数据存储装置的功能,用于将客户端模块210接收到的数据流分块分片,并确定数据片对应的存储节点。并且,该计算设备可以存储系统,存储系统包括存储节点231、存储节点232、存储节点233和存储节点234,用于存储计算代理模块221或计算代理模块222发送的数据。

[0050] 如图2所示,数据存储装置400可以包括数据获取单元401、并行存储单元402和存储参数更新单元403。其中,如果数据存储装置400通过图1所示实施例中的计算代理模块211或计算代理模块222实现,数据获取单元401用于接收用户100通过客户端模块210发送的数据流A中的第一数据块,并将第一数据块提供给并行存储单元402。并行存储单元402可以从存储参数更新单元403处获取第一存储参数,并根据第一存储参数判断第一数据块是否满足存储条件。如果第一数据块满足存储条件,并行存储单元402将第一数据块所包括的多个数据片并行存储到存储系统。例如,在确定数据块1满足存储条件之后,并行存储单元402同时将数据块1中的数据片1、数据片2和数据片3写入存储节点231。在确定数据块2满足存储条件之后,并行存储单元402同时将数据块2中的数据片4、数据片5和数据片6写入存储节点232。在确定数据块3满足存储条件之后,并行存储单元402同时将数据块1中的数据片

7、数据片8和数据片9写入存储节点233。

[0051] 实际应用时,上述数据存储装置400可以通过软件实现,或者可以通过硬件实现。

[0052] 数据存储装置作为软件功能单元的一种举例,可以包括运行在计算实例上的代码。其中,计算实例可以包括物理主机(计算设备)、虚拟机、容器中的至少一种。进一步地,上述计算实例可以是一台或者多台。例如,数据存储装置可以包括运行在多个主机/虚拟机/容器上的代码。需要说明的是,用于运行该代码的多个主机/虚拟机/容器可以分布在相同的区域(region)中,也可以分布在不同的region中。进一步地,用于运行该代码的多个主机/虚拟机/容器可以分布在相同的可用区(availability zone,AZ)中,也可以分布在不同的AZ中,每个AZ包括一个数据中心或多个地理位置相近的数据中心。其中,通常一个region可以包括多个AZ。

[0053] 同样,用于运行该代码的多个主机/虚拟机/容器可以分布在同一个虚拟私有云(virtual private cloud,VPC)中,也可以分布在多个VPC中。其中,通常一个VPC设置在一个region内,同一region内两个VPC之间,以及不同region的VPC之间跨区通信需在每个VPC内设置通信网关,经通信网关实现VPC之间的互连。

[0054] 数据存储装置作为硬件功能单元的一种举例,数据存储装置可以包括至少一个计算设备,如服务器等。或者,数据存储装置也可以是利用专用集成电路(application-specific integrated circuit,ASIC)实现、或可编程逻辑器件(programmable logic device,PLD)实现的设备等。其中,上述PLD可以是复杂程序逻辑器件(complex programmable logical device,CPLD)、现场可编程门阵列(field-programmable gate array,FPGA)、通用阵列逻辑(generic array logic,GAL)或其任意组合实现。

[0055] 数据存储装置包括的多个计算设备可以分布在相同的region中,也可以分布在不同的region中。数据存储装置包括的多个计算设备可以分布在相同的AZ中,也可以分布在不同的AZ中。同样,数据存储装置包括的多个计算设备可以分布在同一个VPC中,也可以分布在多个VPC中。其中,所述多个计算设备可以是服务器、ASIC、PLD、CPLD、FPGA和GAL等计算设备的任意组合。

[0056] 接下来,对数据存储过程的各种非限定性的具体实施方式进行详细描述。

[0057] 参阅图3,为本申请实施例中数据存储方法的一种流程示意图。该方法可以应用于上述图2所示的应用场景中,或者也可以是应用于其它可适用的应用场景中。下面以应用于图2所示的应用场景,且数据存储装置400由计算代理模块221实现为例进行说明。需要说明的是,图2所示的应用场景中,数据流A又可以被称为第一数据流,数据存储装置400不仅包括数据获取单元401、并行存储单元402和存储参数更新单元403,还可以进一步包括通知单元404。并且,各个单元的功能,具体参见下述实施例的相关描述。

[0058] 图3所示的数据存储方法具体可以包括:

[0059] S301:数据获取单元401获取多个数据片。

[0060] 数据获取单元401获取的多个数据片属于数据流A中的第一数据块。其中,数据流A是用户100向分布式存储系统200中存储的数据形成的数据流。第一数据块属于数据流A,可以是数据流A中首个数据块,也可以是数据流A中其他位置的数据块。即,第一数据块可以是图2中的数据块1,也可以是图2中的数据块2或数据块3。第一数据块包括多个数据片。可选地,数据获取单元401获取的多个数据片可以是第一数据块所包括的全部数据片,也可以是

第一数据块所包括的部分数据片。即,数据获取单元401可以获取第一数据块的全部数据片,也可以获取第一数据片的部分数据片。

[0061] 其中,在获取多个数据片时,数据获取单元401可以通过客户端模块210从用户100出获取第一数据块所包括的多个数据片。例如,客户端模块210可以通过网络连接接收用户100发送的待存储数据。受限于网络带宽,用户100发送的待存储数据以数据流的形式被客户端模块210接收。相应地,客户端模块210以数据流的形式向数据存储装置400的数据获取单元401发送数据流。

[0062] 可选地,客户端模块210可以每接收到一个数据片,就向数据获取单元401转发接收到的数据片。或者,客户端模块210可以在接收到预设数量个数据片之后,向数据获取单元401转发接收到的预设数量个数据片。

[0063] 数据获取单元401在获取到多个数据片之后,可以将多个数据片转发给并行存储单元402。相似地,数据获取单元401可以每接收到一个数据片,就向并行存储单元402转发接收到的数据片。或者,数据获取单元401可以在接收到第一数量个数据片之后,向并行存储单元402转发接收到的第一数量个数据片。

[0064] S302:并行存储单元402根据第一存储参数判断所述第一数据块是否满足存储条件,若满足,并行存储第一数据块中的多个数据片。

[0065] 为了提高分布式存储系统200的吞吐率,并行存储单元402可以并行存储接收到的数据片。但是,如果并行存储单元402直接并行存储接收到的多个数据片,可能产生数据空洞,影响数据的正常存储。

[0066] 为此,在本申请实施例中,并行存储单元402可以先判断第一数据块是否满足存储条件,并在确定第一数据块满足存储条件之后再并行存储第一数据块中的数据片。其中,第一存储参数可以是并行存储单元402从存储参数更新单元403处获取的。存储参数更新单元403用于根据数据流A中数据块的存储情况更新第一存储参数。

[0067] 下面首先对第一存储参数进行介绍。

[0068] 第一存储参数用于指示位于数据流A前部的数据块被存储的情况。其中,所述数据流A前部的数据块,是指数据流A中位置先于当前待存储的数据块的数据块。其中,数据块在数据流A中的顺序,可以是数据获取单元401接收数据流A中数据块的顺序。也就是说,在并行存储单元402接收到第一数据块中的数据片时,第一数据块为待存储的数据块,那么第一存储参数根据数据流A中位于第一数据块之前的数据块的存储情况确定。

[0069] 可以理解的是,如果并行存储单元402接收到的多个数据片来自不同的数据块,并行存储单元402可以判断分别多个数据片对应的多个数据块是否满足存储条件,然后并行存储满足存储条件的数据块中的数据片。

[0070] 在一些可能的实施方式中,第一存储参数根据数据流A的头部连续存储部分的右边界确定。其中,数据流A的头部连续存储部分包括位于数据流A头部的、连续的、一个或多个已被存储的数据片。数据流A的头部连续存储部分的右边界是指数据流A的头部连续存储部分中,在数据流A中的位置最靠后的数据片。

[0071] 以图2所示场景举例,假设数据片1、数据片2、数据片3、数据片5和数据片6已被存储。因为数据片4未被存储,被存储的数据片3和被存储的数据片5之间不连续,那么数据流A中头部连续部分包括数据片1、数据片2和数据片3,数据流A的头部连续存储部分的右边界

为数据片3。第一存储参数根据数据片3确定,用于指示数据片3已被存储。具体实现时,第一存储参数可以为数据流A的最后添加确认(last add confirmed,LAC)参数。

[0072] 可以理解的是,如果数据流A中首个数据片未被存储,第一存储参数的值可以为初始值,表示数据流A中任一数据片均未被存储。

[0073] 下面介绍并行存储单元402根据第一存储参数判断第一数据块是否满足存储条件的过程。

[0074] 根据前文介绍可知,数据空洞是指部分数据片未被存储,但是数据流中位于这部分数据片之后的数据片已被存储。因此,为了避免出现数据空洞,并行存储单元402可以控制同一时间被并行存储的数据片来自数据流A中有限范围内的一个或多个数据块,且数据流A中位于被存储的数据片所属的数据块之前的数据块已被存储。这样,因为数据流A中位于待存储的数据片之前的数据块已被存储,所以将待存储的数据片全部存入存储节点之后,从最后一个待存储的数据片,到数据流A中的首个数据片,均已被存储到存储节点。如此,可以保证数据流A的正常存储,避免出现数据空洞。

[0075] 也就是说,在判断第一数据块是否满足存储条件时,并行存储单元402可以判断第一数据块是否属于第一数据流所指示的能够存储的范围。若属于,则说明第一数据块满足存储条件。若不属于,说明如果将第一数据存储到存储节点,数据流A会在第一数据块之前出现数据空洞。

[0076] 实际应用中,并行存储单元402可以具有并行存储来自不同数据块的多个数据片的能力,也可以不具有并行存储来自不同数据块的多个数据片的能力。相应地,并行存储单元402所存储的数据片来源的数据块的数量不同,第一数据块所满足的存储条件也不同。下面分别进行介绍。

[0077] 在第一种可能的实现中,并行存储单元402并行存储来自同一个数据块的多个数据片,不具有并行存储来自不同数据块的多个数据片的能力。

[0078] 这样,在接收到来自第一数据块的多个数据片之后,如果第一数据块之前的数据块未被存储,而并行存储单元402将第一数据块中的任一数据片写入存储节点,那么数据流A在第一数据块之前的位置出现数据空洞。

[0079] 因此,为了避免数据空洞的出现,存储第一数据块所需满足的存储条件包括:数据流A中位于第一数据块之前的数据块已被存储。这样,如果第一数据块满足存储条件,并行存储单元402将第一数据块所包括的多个数据片写入存储节点之后,数据流A中位于第一数据块之前的所有数据块已被写入存储节点,且第一数据块也被写入存储节点,数据流A中不存在数据空洞。

[0080] 在根据第一存储参数判断第一数据块是否满足存储条件时,并行存储单元402可以根据第一存储参数判断数据流A中位于第一数据块之前的所有数据块是否已被存储。根据前文介绍可知,第一存储参数可以根据数据流A的头部连续存储部分的右边界确定。相应地,并行存储单元402可以根据第一存储参数判断头部连续存储部分的右边界是否与第一数据块相邻,或右边界是否位于第一数据块内。

[0081] 如果数据流A的头部连续存储部分的右边界与第一数据块相邻,说明数据流A中位于第一数据块之前的所有数据片已被存储,那么第一数据块满足存储条件。如果数据流A的头部连续存储部分的右边界位于第一数据块内,说明数据流A中位于第一数据块之前的所

有数据片已被存储,且第一数据块内部分数据片也已被存储,那么第一数据块满足存储条件。如果数据流A的头部连续存储部分的右边界不与第一数据块相邻且不属于第一数据块,说明存在未被存储的数据片位于第一数据块之前,第一数据块不满足存储条件。

[0082] 可以理解的是,在并行存储第一数据块中的数据片时,如果并行存储单元402又接收到了来自其他数据块的数据片,并行存储单元402可以将这些数据片挂起,待存储完第一数据块中的所有数据片之后再存储。

[0083] 在第二种可能的实现中,并行存储单元402并行存储来自不同数据块的多个数据片的能力。

[0084] 这样,如果数据流A中位于第一数据块之前的数据块未被存储,但是该未被存储的数据块为并行存储单元402正在存储或能够存储的数据块,那么即使并行存储单元402存储了第一数据块中的数据片,由于未被存储的数据块仍然可以被并行存储单元402存储,数据流A不会出现数据空洞,不会影响数据流A的正常存储。

[0085] 在本申请实施例中,可以将并行存储单元402可以并行存储的数据块的最大数量称为预设数量。即,并行存储单元402同时可以并行存储的多个数据片所属的数据块的数量不超过预设数量。

[0086] 相应地,在根据第一存储参数判断第一数据块是否满足存储条件时,并行存储单元402可以先根据第一数据块和预设数量确定第二数据块,然后判断第一存储参数所指示的右边界是否为第二数据块中最后一个数据块,或者位于第二数据块之后。

[0087] 其中,第二数据块为数据流A中位于第一数据块之前,且数据流A中位于第一数据块和第二数据块之间的数据块的数量小于预设数量。举例说明。假设并行存储单元402最多可以并行存储来自2个数据块的数据片(即预设数量为2),那么数据流A中第一数据块与第二数据块之间间隔的数据块的数量小于2。即数据流A中第一数据块与第二数据块之间最多间隔有一个数据块。结合图2所示的应用场景为例说明,如果第一数据块为图2所示实施例中的数据块3,且预设数量为2,那么第二数据块为图2所示实施例中的数据块1,数据流A中第一数据块和第二数据块之间存在1个数据块(即数据块2),小于预设数量。

[0088] 如果第一存储参数所指示的右边界为第二数据块中最后一个数据块,说明第二数据块属于数据流A的头部连续存储部分,第二数据块以及第二数据块之前的数据块已经被存储,那么第一数据块满足存储条件。如果第一存储参数所指示的右边界为位于第二数据块之后,说明第二数据块以及第二数据块之前的数据块已经被存储,那么第一数据块满足存储条件。如果数据流A的头部连续存储部分的右边界位于第二数据块中最后一个数据块之前,说明第二数据块中存在未被存储的数据片或存储未被存储的数据片位于第二数据块之前,第一数据块不满足存储条件。

[0089] 如果第一数据块不满足存储条件,并行存储单元402暂不存储第一数据块中的数据片。如果第一数据块满足存储条件,并行存储单元402并行存储第一数据块中多个数据片。

[0090] 下面介绍并行存储单元402并行存储多个数据片的方法。可选地,所述多个数据片包括上述步骤S301中获取的多个数据片,也可以包括第一数据块以外其他数据块的数据片。

[0091] 本实施例中,并行存储是指在同一时间存储多个数据片。本申请实施例提供了以

下两种可能的实施方式。

[0092] 在第一种实施方式中,并行存储单元402可以将多个数据块并行存储到一个存储节点。具体实现时,并行存储单元402可以先确定第一数据块对应的存储节点,然后在该存储节点上创建多个存储线程。每个存储线程用于将一个数据片写入存储节点。这样,多个存储线程并行执行,实现了多个数据片的并行存储。

[0093] 以图2所示的应用场景为例进行说明,假设第一数据块为数据流A中的数据块1。在确定数据块1满足存储条件之后,并行存储单元402可以在存储节点231上创建存储线程1、存储线程2和存储线程3。其中,存储线程1用于将数据片1写入存储节点231,存储线程2用于将数据片2写入存储节点231,存储线程3用于将数据片3写入存储节点231。这样,存储线程1、存储线程2和存储线程3并行执行,实现了数据片1、数据片2和数据片3的并行存储。

[0094] 在第二种实施方式中,并行存储单元403可以将多个数据块并行存储到多个存储节点。具体实现时,并行存储单元可以先确定多个数据片中每个数据片对应的存储节点,然后分别在每个数据片对应的存储节点上创建存储线程,通过多个存储线程分别存储多个数据片,实现数据片的并行存储。

[0095] 以图2所示的网络架构为例进行说明。假设并行存储单元确定数据块2满足存储条件,且数据片4对应存储节点231,数据片5对应存储节点232,数据片6对应存储节点233。那么并行存储单元402可以在存储节点231上创建存储线程4,在存储节点232上创建存储线程5,在存储节点233上创建存储线程6。其中,存储线程4用于将数据片5写入存储节点231,存储线程5用于将数据片5写入存储节点232,存储线程6用于将数据片6写入存储节点233。这样,存储线程4、存储线程5和存储线程6并行执行,实现了数据片4、数据片6和数据片7的并行存储。

[0096] 可以理解的是,上述两种实施方式可以混合使用。例如,在一些可能的实施方式中,并行存储单元402可以通过两个并行运行的存储线程将数据片1和数据片2存储到存储节点231,并将数据片3存储到存储节点232。

[0097] 在一些可能的实施方式中,并行存储单元402在存储数据片之前可以先生成数据片的标识。其中,数据片的标识用于唯一标识数据片。存储节点在完成数据片的写入之后,可以记录数据片的标识,以便在计算代理单元402出现故障时恢复数据。可选地,数据片的标识包括数据片所属的数据流的标识、数据片所属的数据块的标识和数据片在数据块内部的标识。例如,第一数据块中的数据片的标识可以包括数据流A的标识、第一数据块在数据流A中的序号以及该数据片在第一数据块中的序号。关于这部分内容的介绍可以参见下文,这里不再赘述。

[0098] S303:存储参数更新单元403根据多个数据片的存储情况更新第一存储参数。

[0099] 具体实现时,并行存储单元402(或存储节点)在写入数据片之后,可以将数据片的存储情况发送给存储参数更新单元403。存储参数更新单元403根据数据片被存储的情况更新第一存储参数。

[0100] 其中,数据片的存储情况可以通过数据片的标识表示。具体实现时,并行存储单元402(或存储数据片的存储节点)在完成数据片的写入之后,向存储参数更新单元403发送数据片的标识。存储参数更新单元403在接收到数据片的标识之后,可以确定该标识对应的数据片已被存储,并以此更新第一存储参数。

[0101] 根据前文介绍可知,第一存储参数可以根据数据流A的头部连续存储部分的右边界确定。相应地,在更新第一存储参数时,存储参数更新单元403可以根据数据片的存储情况判断头部连续存储部分的右边界是否发生变化。如果头部连续存储部分的右边界发生变化,存储参数更新单元403根据变化后的头部连续存储部分的右边界调整第一存储参数。如果头部连续存储部分的右边界没有发生变化,存储参数更新单元403可以不调整第一存储参数。

[0102] 以图2所示的应用场景为例进行说明,假设第一数据块为数据流A中的数据块3,且数据块1和数据块2已被存储,并行存储单元402通过存储线程7、存储线程8和存储线程9分别将数据片7、数据片8被存储的情况,判断是否更新第一存储参数。由于数据片7和数据片8未被存储,数据流A和数据片9存储到存储节点233。如果并行存储单元402监测到存储线程9执行完毕,说明数据片9已经写入完成,并行存储单元402向存储参数更新单元403发送数据片9的标识。存储参数更新单元403根据数据片9的标识确定数据片9已被存储。接着,存储参数更新单元403根据数据流A中位于数据片9之前的数据片的头部连续存储部分的右边界没有扩展到数据片9,存储参数更新单元403不更新第一存储参数。

[0103] 如果并行存储单元402监测到存储线程7执行完毕,说明数据片7已经写入完成,并行存储单元402向存储参数更新单元403发送数据片7的标识。存储参数更新单元403根据数据片7的标识确定数据片7已被存储。接着,存储参数更新单元403根据数据流A中位于数据片7之前的数据片被存储的情况,判断是否更新第一存储参数。由于数据块1和数据块2中的数据片已被存储,数据流A的头部连续存储部分的右边界扩展到数据片7,存储参数更新单元403根据数据片7的标识更新第一存储参数。更新后的第一存储参数表示数据流A中数据片7以及数据片7之前的数据片已被存储。

[0104] 在更新第一存储参数之后,数据存储装置400中的通知单元404可以根据更新后的第一存储参数生成通知消息,并向客户端模块210发送通知消息,以使客户端模块210向用户100展示通知消息。其中,通知消息用于指示第一数据块中的数据片已被存储。具体实现时,通知单元404可以在第一存储参数指示第一数据块中全部数据片均被存储之后生成通知消息。通知消息用于通知用户数据流A中第一数据块以及位于第一数据块之前的数据已被存储。

[0105] 在上述介绍中,第一存储参数用于第一数据块是否满足存储条件的判断。具体实现中,第一存储参数可以用于判断数据流A中任一数据块是否满足存储条件。例如,对于并行存储单元402可以并行存储预设数量个数据块中的数据片的实施方式,如果第一数据块中每个数据片的存储均已完成,并行存储单元402可以根据更新后的第一存储参数确定第三数据块第三数据块满足存储条件。第三数据块在数据流A中位于第一数据块之后,且第一数据块与第三数据块之间间隔的数据块的数量小于预设数量。

[0106] 实际应用时,数据存储装置400或者运行数据存储装置400的计算代理单元221可能出现故障。如果数据存储装置400或计算代理单元221出现故障,那么数据流A的存储无法继续进行。这样,为了保证数据流A的正常存储,可以通过分布式存储系统200中其他的计算代理单元(例如计算代理单元222)完成数据流A的存储。

[0107] 新的用于存储数据流A的计算代理单元也可以从头开始存储数据流A。或者,为了提高存储效率,如果数据流A中部分数据已经被数据存储装置400存储到存储节点,新的用

于存储数据流A的计算代理单元可以根据数据流A中各个数据片的存储情况恢复第一存储参数,并在此基础上重新存储数据流A。

[0108] 如图4所示,在图2所示应用场景的基础上,假设计算代理单元222用于将数据流B存储到多个存储节点(多个存储节点可以包括前述存储节点231、存储节点232、存储节点233和存储节点234)。在计算代理模块222出现故障之后,计算代理模块221中的数据存储装置400负责存储数据流B。其中,数据流B又可以被称为第二数据流,包括数据块4、数据块5和数据块6。数据块4包括数据片10、数据片11和数据片12,数据片5包括数据片13、数据片14和数据片15,数据块6包括数据片16、数据片17和数据片18。

[0109] 在图2所示结构的基础上,数据存储装置400还包括存储参数确定单元405和数据片确定单元406。存储参数确定单元405可以从多个存储节点中获取数据流B中数据片的存储情况,并据此确定第二存储参数。接着,存储参数确定单元405将第二存储参数发送给数据片确定单元406。数据片确定单元406根据第二存储参数确定第三数据片,并将第三数据片的标识发送给并行存储单元402。并行存储单元402根据第三数据片的标识,和数据获取单元401发送的数据流B中的多个数据片,将数据流B中未被存储的多个数据片并行存储到多个存储节点。其中,第二存储参数用于指示位于数据流B前部的数据块被存储的情况。第三数据片为数据流B中首个未被存储的数据片。并行存储单元402所并行存储的数据片包括第三数据片,以及数据流B中位于第三数据片后的一个或多个数据片。

[0110] 参阅图5为本申请实施例中数据存储方法的另一种流程示意图。该方法可以应用于上述图4所示的应用场景中,或者也可以是应用于其它可适用的应用场景中。下面以应用于图4所示的应用场景,且数据存储装置400由计算代理模块221实现为例进行说明。需要说明的是,图4所示的应用场景中,数据获取单元401、并行存储单元402和存储参数更新单元403可以实现上述图5所示实施例中的功能,也可以实现上述图3所示实施例中的功能。并且,各个单元的功能,具体参见下述实施例的相关描述。

[0111] 图5所示的数据存储方法具体可以包括:

[0112] S501:响应于计算代理模块222出现故障,存储参数确定单元405根据数据流B中数据片的存储情况确定第二存储参数。

[0113] 如果计算代理模块222出现故障,计算代理模块222所承载的存储任务无法继续完成。为了保证数据的正常存储,客户端210或分布式存储系统220中用于分配存储任务的其他模块可以指示计算代理模块221中的数据存储装置400完成数据流B的存储。

[0114] 在本实施例中,为了提高存储空间的利用率,数据存储装置400可以在计算代理模块222已完成部分的基础上存储数据流B。也就是说,对于数据流B中已被存储的部分数据片,数据存储装置400不再重新进行存储。因此,为了确定数据流B中哪些数据片已被存储,数据存储装置400可以确定数据流B对应的第二存储参数。

[0115] 数据存储装置400中的存储参数确定单元405可以从分布式存储节点200中的多个存储节点获取数据流B中数据片的存储情况,并根据数据流B中数据片的存储情况确定第二存储参数。

[0116] 首先介绍存储参数确定单元405获取数据流B中数据片的存储情况的方法。

[0117] 在本实施例中,存储参数确定单元405可以通过存储节点提供的接口,查询数据流B中各个数据片是否被存储节点存储。具体实现时,存储参数确定单元405可以根据数据流B

的标识生成查询请求,然后向存储系统中的一个或多个存储节点发送查询请求。存储节点在接收到查询请求之后,可以根据查询请求中包括的数据流B的标识,判断本地是否存在已被存储的数据流B中的数据片。若存在,存储节点可以根据被存储在本地的、第二数据流中的数据片的标识生成标识集合,并向存储参数确定单元405发送标识集合。这样,存储参数确定单元405可以根据存储节点返回的标识集合,确定数据流B中哪些数据片已被存储,得到数据流B中数据片的存储情况。

[0118] 根据前文介绍可知,数据流B中数据片的标识可以包括数据流B的标识、数据片所属的数据块的标识和数据片在数据块内的标识。相应地,在获取数据流B中数据片的存储情况时,为了更加方便地确定数据片所属的数据块,存储参数确定单元405可以生成多个查询请求,每个查询请求用于查询数据流B中一个数据块中数据片的存储情况。例如,对于数据流B中的数据块4,存储参数确定单元405可以生成第一查询请求,第一查询请求包括数据流B的标识和数据块4的标识,用于查询数据块4中各个数据片在存储节点的存储情况。

[0119] 下面介绍存储参数确定单元405根据数据流B中数据片的存储情况确定第二存储参数的方法。

[0120] 与第一存储参数类似,第二存储参数用于指示位于数据流B前部的数据块被存储的情况,例如可以根据数据流B的头部连续存储部分的右边界确定。实际应用时,存储参数确定单元405在得到数据流B中数据片的存储情况之后,可以据此判断位于数据流B头部的多个数据片是否已被存储,从而确定数据流B的头部连续存储部分,并基于数据流B的头部连续存储部分确定第二存储参数。

[0121] 结合图4所示应用场景进行说明。通过向多个存储节点发送查询请求,存储参数确定单元405确定数据流B中数据片10、数据片11、数据片12、数据片13和数据片15已被存储节点存储,数据流B中其他数据片未被存储。由于数据片14未被存储,存储参数确定单元405确定数据流B的头部连续存储部分为数据片10到数据片13。数据流B的头部连续存储部分的右边界为数据片13。存储参数确定单元405根据数据片13确定第二存储参数。

[0122] 在确定了第二存储参数之后,存储参数确定单元405可以向数据片确定单元406发送第二存储参数。

[0123] S502:数据片确定单元406根据第二存储参数确定第三数据片的标识。

[0124] 在接收到存储参数确定单元405发送的第二存储参数之后,数据片确定单元406可以根据第二存储参数确定第三数据片的标识。其中,第三数据片为数据流B中未被存储节点存储的数据片,且数据流B中位于第三数据片之前的任一数据片已被存储。这样,从第三数据片开始存储,使得数据流B中每个数据片都能够得到存储,避免数据流B中存在数据空洞。

[0125] 沿用上文的例子进行说明,即第二存储参数根据数据片13的标识确定。在接收到存储参数确定单元405发送的第二存储参数之后,数据片确定单元406可以根据第二存储参数,将数据流B中头部连续存储部分之后的首个数据片确定为第三数据片。即第三数据片为数据流B中的数据片14。

[0126] 在确定第三数据片之后,数据片确定单元406可以向并行存储单元402发送第三数据片的标识。

[0127] S503:并行存储单元402将第三数据片,以及数据流B中位于第三数据片之后的数据片并行存储到多个存储节点。

[0128] 在获取到第三数据片的标识之后,并行存储单元402可以将第三数据片以及数据流B中位于第三数据片之后的数据片并行存储到多个存储节点。具体实现时,并行存储单元402可以采用类似S302所述的方法,基于第二存储参数判断数据流B中多个数据块是否满足存储条件,并在确定数据块满足存储条件之后再并行存储数据块所包括的多个数据片。这样,由于第三数据片之前的数据片属于数据流B的头部连续存储部分,采用S302所述的方法并行存储数据流B中剩余的数据片,可以使得每次被存储的数据片所属的数据块都满足存储条件,避免数据流B中出现数据空洞。

[0129] 可以理解的是,即使数据流B中第三数据片之后的部分数据片在计算代理模块222故障之前已被存储到数据节点,并行存储单元402仍可以存储这些数据片,以避免出现数据空洞。

[0130] 需要说明的是,本申请实施例中,对于数据存储装置内的各个单元的划分以及功能描述仅作为一种示例。比如,在其他实施例中,数据获取单元401可以用于执行上述数据存储方法中的任意步骤,类似的,并行存储单元402、存储参数更新单元403、通知单元404、存储参数确定单元405和数据片确定单元406均可以用于执行上述数据存储方法中的任意步骤,并且,并行存储单元402、存储参数更新单元403、通知单元404、存储参数确定单元405和数据片确定单元406负责实现的步骤可根据需要指定,通过并行存储单元402、存储参数更新单元403、通知单元404、存储参数确定单元405和数据片确定单元406分别实现数据存储方法中不同的步骤来实现数据存储装置的功能。

[0131] 上述图3或图5所示实施例中,针对数据存储过程中所涉及到的数据存储装置(包括上述数据获取单元401、并行存储单元402、存储参数更新单元403、通知单元404、存储参数确定单元405、数据片确定单元406)可以是配置于计算设备或者计算设备集群上的软件,并且,通过在计算设备或者计算设备集群上运行该软件,可以使得计算设备或者计算设备集群实现上述数据存储装置所具有的功能。下面,基于硬件设备实现的角度,对数据存储的过程中所涉及的数据存储装置进行详细介绍。

[0132] 图6示出了一种计算设备的结构示意图,上述数据存储装置可以部署在该计算设备上,该计算设备可以是云环境中的计算设备(如服务器),或边缘环境中的计算设备,或终端设备等具体可以用于实现上述图3所示实施例中数据获取单元401、并行存储单元402、存储参数更新单元403和通知单元404的功能,或者用于实现图3所示实施例中数据获取单元401、并行存储单元402、存储参数更新单元403、存储参数确定单元405和数据片确定单元406的功能。

[0133] 如图6所示,计算设备600包括处理器610、存储器620、通信接口630和总线640。处理器610、存储器620和通信接口630之间通过总线640通信。总线640可以是外设部件互连标准(peripheral component interconnect,PCI)总线或扩展工业标准结构(extended industry standard architecture,EISA)总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示,图6中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。通信接口630用于与外部通信,例如接收用户提供的原始数据以及待训练的特征提取网络模型等。

[0134] 其中,处理器610可以为中央处理器(central processing unit,CPU)、专用集成电路(application specific integrated circuit,ASIC)、图形处理器(graphics

processing unit, GPU) 或者一个或多个集成电路。处理器610还可以是一种集成电路芯片, 具有信号的处理能力。在实现过程中, 数据存储装置中各个单元的功能可以通过处理器610中的硬件的集成逻辑电路或者软件形式的指令完成。处理器610还可以是通用处理器、数据信号处理器(digital signal process, DSP)、现场可编程逻辑门阵列(field programmable gate array, FPGA)或者其他可编程逻辑器件, 分立门或者晶体管逻辑器件, 分立硬件组件, 可以实现或者执行本申请实施例中公开的方法、步骤及逻辑框图。其中, 通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等, 结合本申请实施例所公开的方法可以直接体现为硬件译码处理器执行完成, 或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器、闪存、只读存储器, 可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器620, 处理器610读取存储器620中的信息, 结合其硬件完成数据存储装置中的部分或全部功能。

[0135] 存储器620可以包括易失性存储器(volatile memory), 例如随机存取存储器(random access memory, RAM)。存储器620还可以包括非易失性存储器(non-volatile memory), 例如只读存储器(read-only memory, ROM), 快闪存储器, HDD或SSD。

[0136] 存储器620中存储有可执行代码, 处理器610执行该可执行代码以执行前述数据存储装置所执行的方法。

[0137] 具体地, 在实现图3所示实施例的情况下, 且图3所示实施例中所描述的数据获取单元401、并行存储单元402、存储参数更新单元403和通知单元404为通过软件实现的情况下, 执行图3中的数据获取单元401、并行存储单元402、存储参数更新单元403和通知单元404功能所需的软件或程序代码存储在存储器620中, 客户端模块210与其它设备的交互通过通信接口630实现, 处理器用于执行存储器620中的指令, 实现数据存储装置所执行的方法。在实现图5所示实施例的情况下, 且图5所示实施例中所描述的数据获取单元401、并行存储单元402、存储参数更新单元403、存储参数确定单元405和数据片确定单元406为通过软件实现的情况下, 执行图5中的数据获取单元401、并行存储单元402、存储参数更新单元403、存储参数确定单元405和数据片确定单元406功能所需的软件或程序代码存储在存储器620中, 客户端模块210与其它设备的交互通过通信接口630实现, 处理器用于执行存储器620中的指令, 实现数据存储装置所执行的方法。

[0138] 图7示出的一种计算设备集群的结构示意图。其中, 图7所示的计算设备集群70包括多个计算设备, 上述数据存储装置可以分布式地部署在该计算设备集群70中的多个计算设备上。如图7所示, 计算设备集群70包括多个计算设备700, 每个计算设备700包括存储器720、处理器710、通信接口730以及总线740, 其中, 存储器720、处理器710、通信接口730通过总线740实现彼此之间的通信连接。

[0139] 处理器710可以采用CPU、GPU、ASIC或者一个或多个集成电路。处理器710还可以是一种集成电路芯片, 具有信号的处理能力。在实现过程中, 数据存储装置的部分功能可用通过处理器710中的硬件的集成逻辑电路或者软件形式的指令完成。处理器710还可以是DSP、FPGA、通用处理器、其他可编程逻辑器件, 分立门或者晶体管逻辑器件、分立硬件组件, 可以实现或者执行本申请实施例中公开的部分方法、步骤及逻辑框图。其中, 通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等, 结合本申请实施例所公开的方法的

步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器、闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器720,在每个计算设备700中,处理器710读取存储器720中的信息,结合其硬件可以完成数据存储装置的部分功能。

[0140] 存储器720可以包括ROM、RAM、静态存储设备、动态存储设备、硬盘(例如SSD、HDD)等。存储器720可以存储程序代码,例如,用于实现数据片获取单元401的部分或者全部程序代码、用于实现并行存储单元402的部分或者全部程序代码、用于实现存储参数更新单元403的部分或者全部程序代码、用于实现通知单元404的部分或者全部程序代码、用于实现存储参数确定单元405的部分或者全部程序代码、用于实现数据片确定单元406的部分或者全部程序代码等。针对每个计算设备700,当存储器720中存储的程序代码被处理器710执行时,处理器710基于通信接口730执行数据存储装置所执行的部分方法,如其中一部分计算设备700可以用于执行上述数据获取单元401、并行存储单元402、存储参数更新单元403和通知单元所执行的方法,另一部分计算设备700用于执行上述存储参数确定单元405和数据片确定单元406所执行的方法。存储器720还可以存储数据,例如:处理器710在执行过程中产生的中间数据或结果数据,例如,上述第一存储参数等。

[0141] 每个计算设备700中的通信接口730用于与外部通信,例如与其它计算设备700进行交互等。

[0142] 总线740可以是外设部件互连标准总线或扩展工业标准结构总线等。为便于表示,图7中每个计算设备700内的总线740仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0143] 上述多个计算设备700之间通过通信网络建立通信通路,以实现数据存储装置的功能。任一计算设备可以是云环境中的计算设备(例如,服务器),或边缘环境中的计算设备,或终端设备。

[0144] 此外,本申请实施例还提供了一种计算机可读存储介质,该计算机可读存储介质中存储有指令,当其在—一个或者多个计算设备上运行时,使得该—一个或者多个计算设备执行上述实施例数据存储装置的各个单元所执行的方法。

[0145] 此外,本申请实施例还提供了一种计算机程序产品,所述计算机程序产品被—一个或者多个计算设备执行时,所述—一个或者多个计算设备执行前述数据存储方法中的任一方法。该计算机程序产品可以为—一个软件安装包,在需要使用前述数据存储方法的任一方法的情况下,可以下载该计算机程序产品并在计算机上执行该计算机程序产品。

[0146] 另外需说明的是,以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于—一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。另外,本申请提供的装置实施例附图中,单元之间的连接关系表示它们之间具有通信连接,具体可以实现为—一条或多条通信总线或信号线。

[0147] 通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件的方式来实现,当然也可以通过专用硬件包括专用集成电路、专

用CPU、专用存储器、专用元器件等来实现。一般情况下,凡由计算机程序完成的功能都可以很容易地用相应的硬件来实现,而且,用来实现同一功能的具体硬件结构也可以是多种多样的,例如模拟电路、数字电路或专用电路等。但是,对本申请而言更多情况下软件程序实现是更佳实施方式。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在可读取的存储介质中,如计算机的软盘、U盘、移动硬盘、ROM、RAM、磁碟或者光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,训练设备,或者网络设备)执行本申请各个实施例所述的方法。

[0148] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时,可以全部或部分地以计算机程序产品的形式实现。

[0149] 所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时,全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、训练设备或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、训练设备或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存储的任何可用介质或者是包含一个或多个可用介质集成的训练设备、数据中心等数据存储设备。所述可用介质可以是磁性介质,(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质(例如固态硬盘(Solid State Disk,SSD))等。

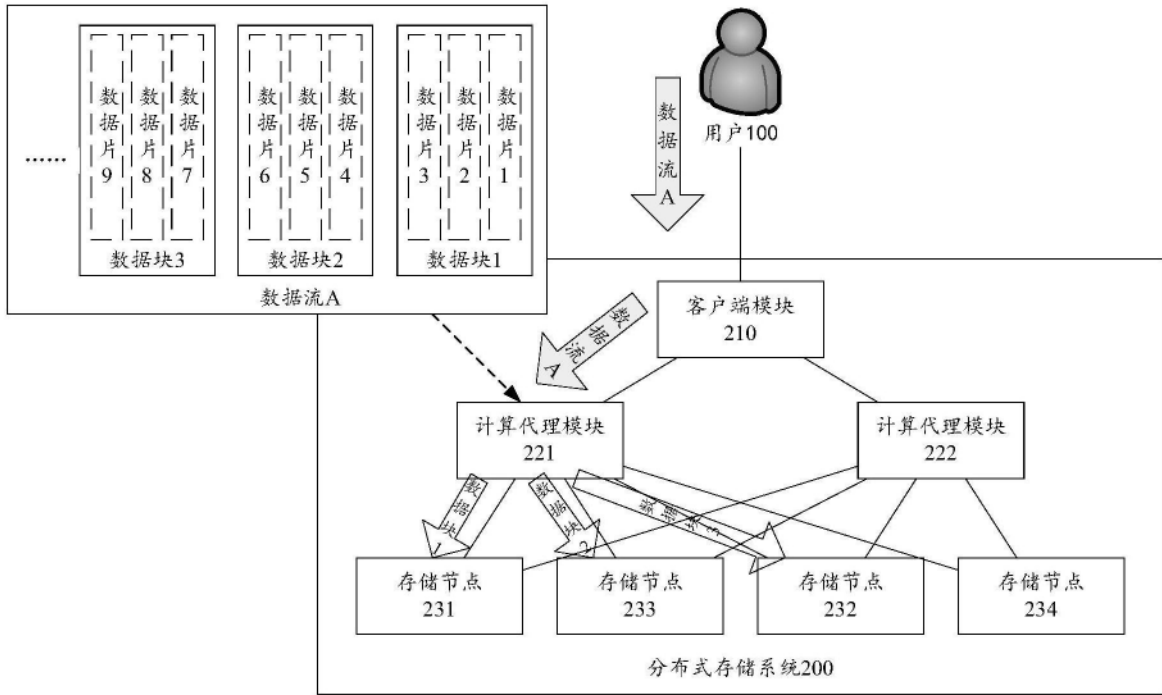


图1

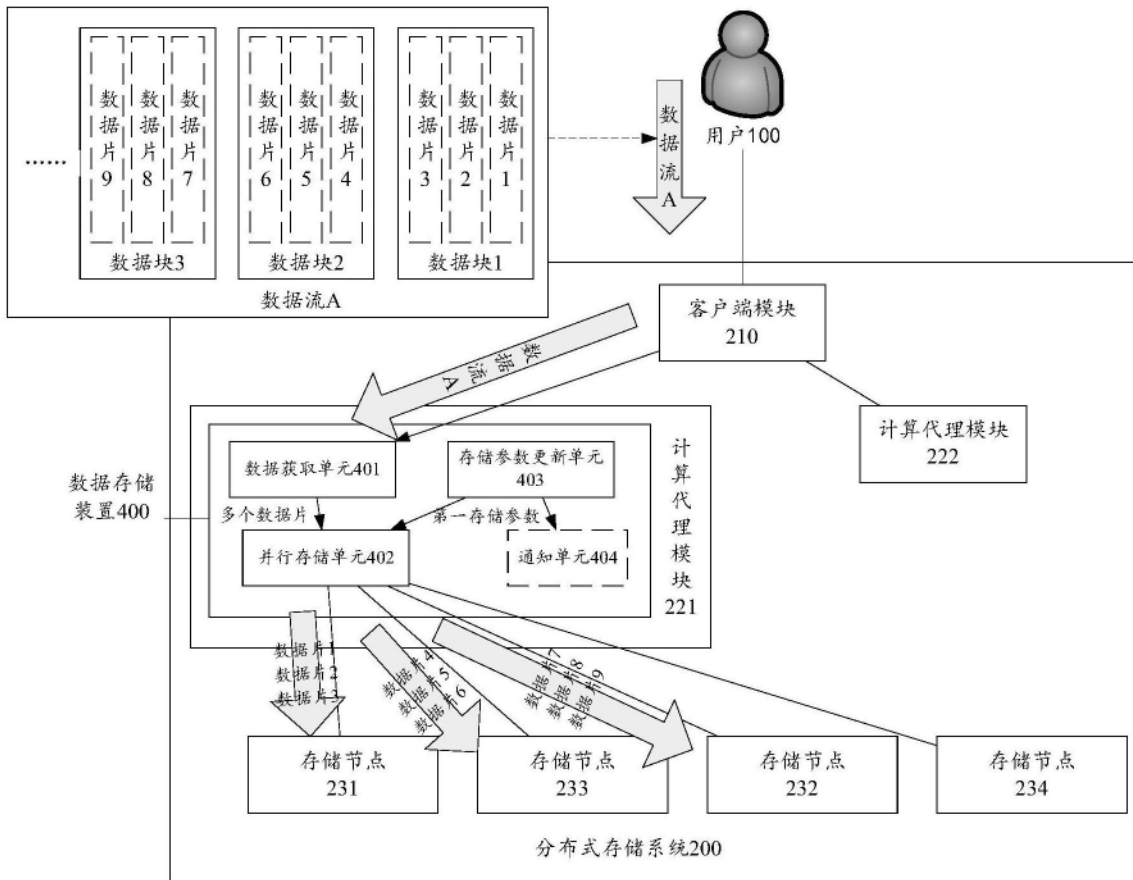


图2

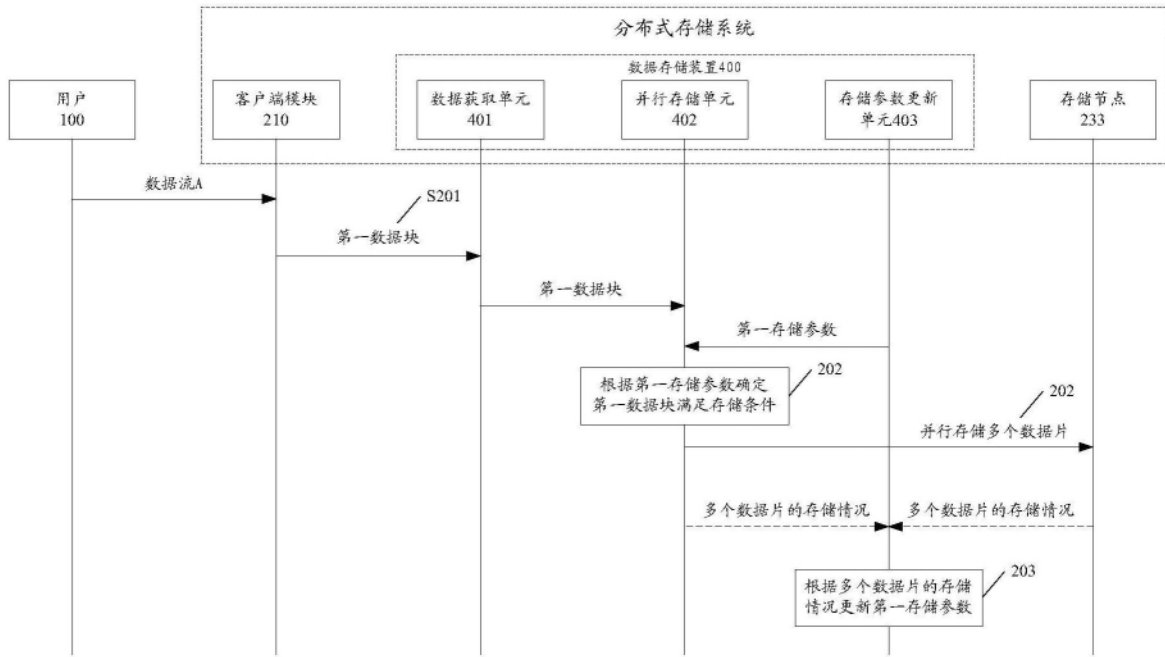


图3

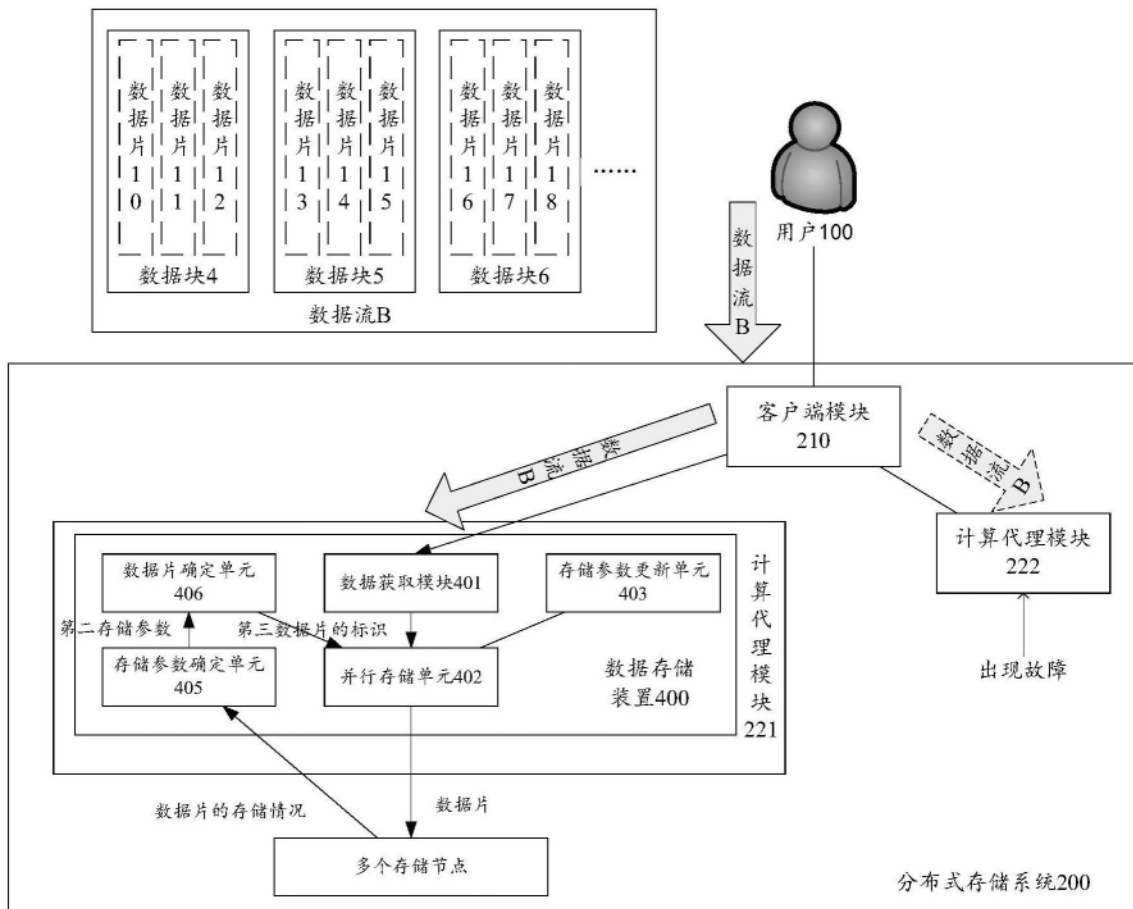


图4

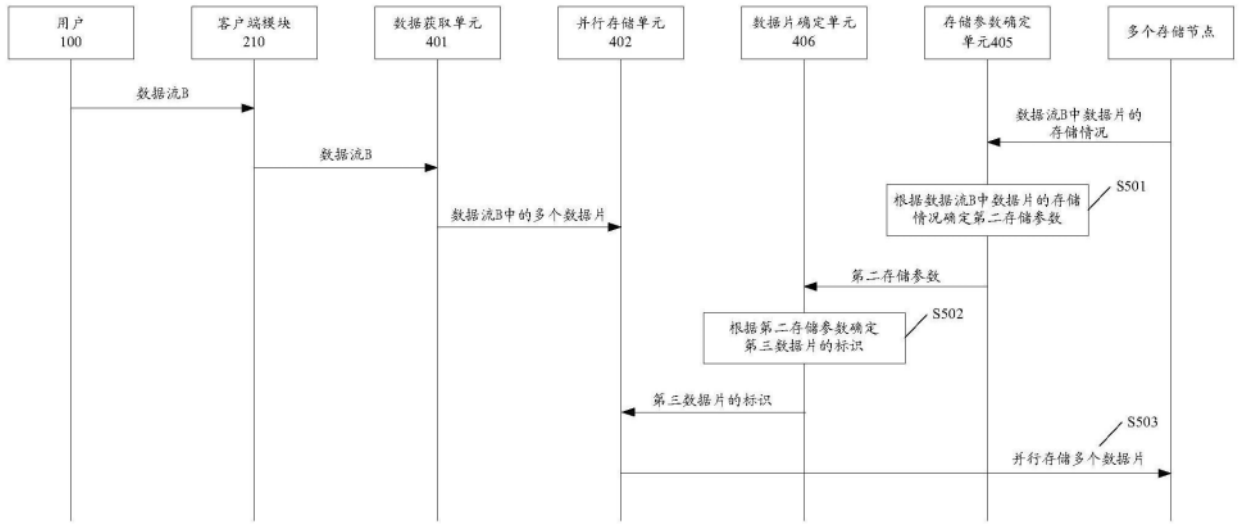


图5

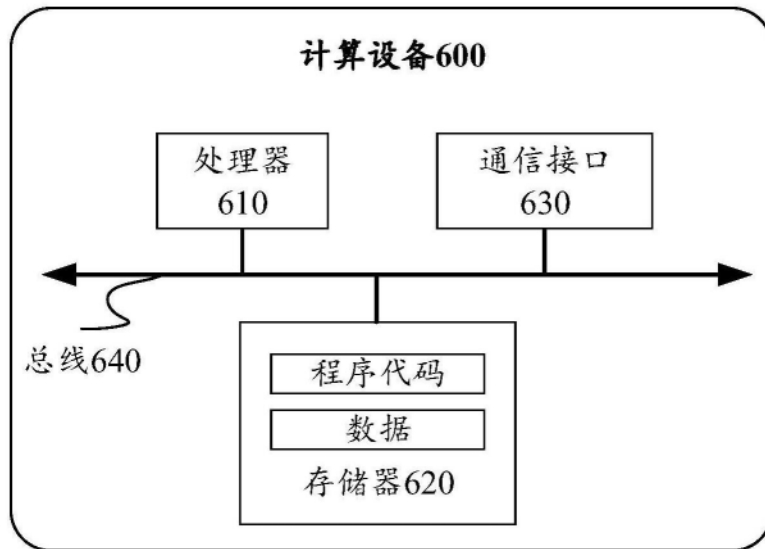


图6

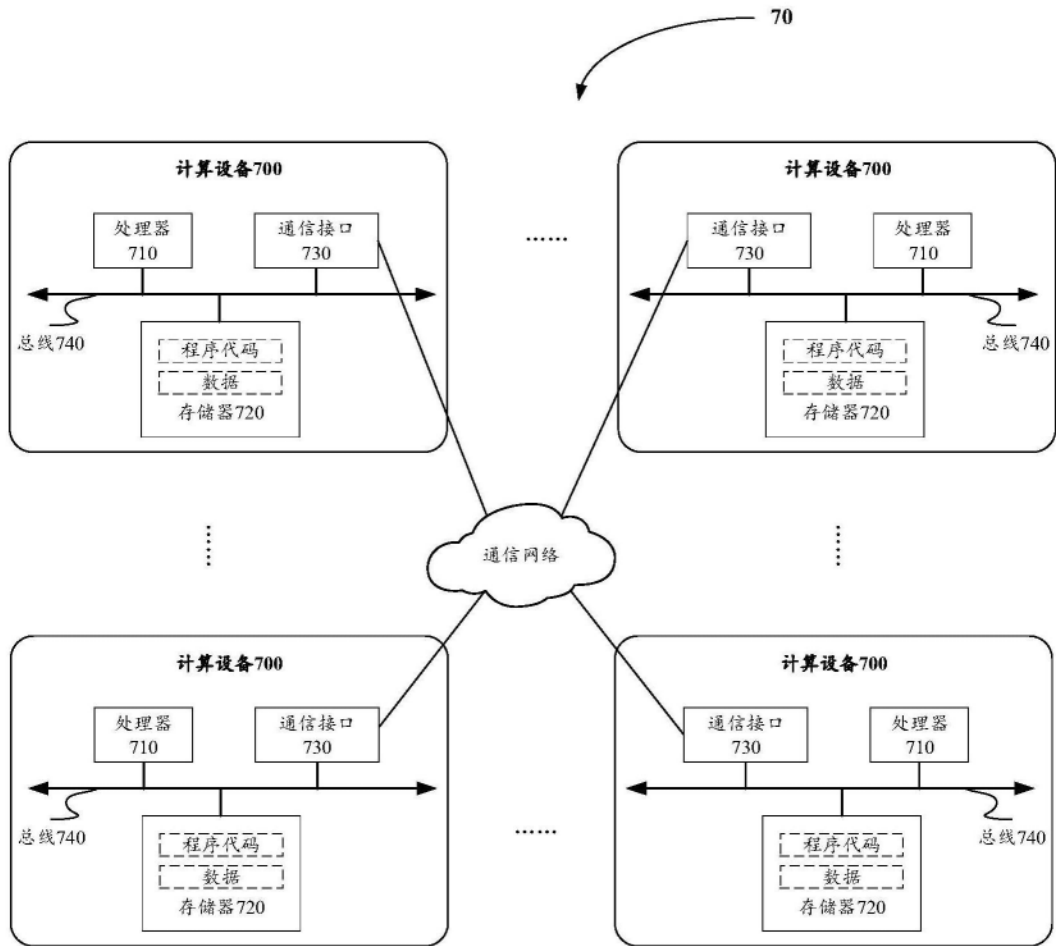


图7