



(12)发明专利申请

(10)申请公布号 CN 110795273 A

(43)申请公布日 2020.02.14

(21)申请号 201910995233.2

(22)申请日 2019.10.18

(71)申请人 苏州浪潮智能科技有限公司
地址 215100 江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢

(72)发明人 施培任

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 牛亭亭

(51)Int.Cl.
G06F 11/10(2006.01)
G06F 11/14(2006.01)

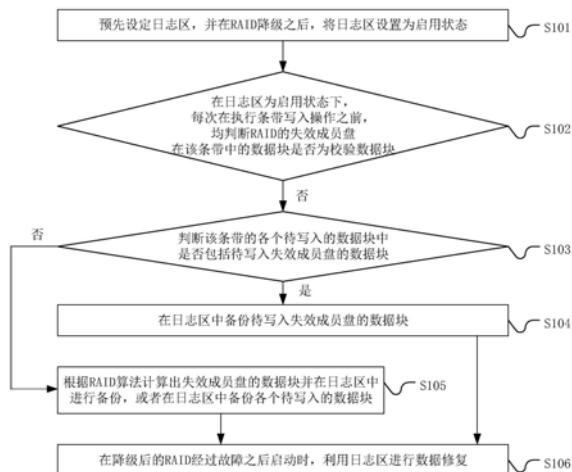
权利要求书2页 说明书11页 附图2页

(54)发明名称

一种RAID的写洞保护方法、系统及存储介质

(57)摘要

本发明公开了一种RAID的写洞保护方法,包括:预先设定日志区,并在RAID降级后设为启用状态;在日志区为启用状态下,每次执行条带写入操作前,判断RAID的失效成员盘在该条带中的数据块是否为校验数据块;如果不是校验数据块,则判断该条带的各个待写入的数据块中是否包括待写入失效成员盘的数据块;若是则在日志区中备份待写入失效成员盘的数据块;若否则计算出失效成员盘的数据块并在日志区中进行备份,或在日志区中备份各个待写入的数据块;在降级后的RAID经过故障之后启动时,利用日志区进行数据修复。应用本申请的方案,避免了RAID的写洞问题。本申请还提供了RAID的写洞保护系统及存储介质,具有相应技术效果。



1. 一种RAID的写洞保护方法,其特征在于,包括:

预先设定日志区,并在RAID降级之后,将所述日志区设置为启用状态;

在所述日志区为启用状态下,每次在执行条带写入操作之前,均判断所述RAID的失效成员盘在该条带中的数据块是否为校验数据块;

如果不是校验数据块,则判断该条带的各个待写入的数据块中是否包括待写入所述失效成员盘的数据块;

如果包括,则在所述日志区中备份待写入所述失效成员盘的数据块;

如果不包括,则根据RAID算法计算出所述失效成员盘的数据块并在所述日志区中进行备份,或者在所述日志区中备份各个待写入的数据块;

在降级后的所述RAID经过故障之后启动时,利用所述日志区进行数据修复。

2. 根据权利要求1所述的RAID的写洞保护方法,其特征在于,每次在所述日志区中进行数据块的备份时,还包括:保存该次备份操作的关联信息以构成一个条带日志;

相应的,所述利用所述日志区进行数据修复,包括:

通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,并基于确定出的各个有效的条带日志进行数据修复。

3. 根据权利要求2所述的RAID的写洞保护方法,其特征在于,保存的关联信息中至少包括:该条带日志的条带序号以及日志编号;

相应的,所述通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,包括:

遍历所述日志区中的各个条带日志的关联信息,并且针对具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

4. 根据权利要求2所述的RAID的写洞保护方法,其特征在于,保存的关联信息中至少包括:该条带日志的条带序号、日志编号以及条带日志校验值;

相应的,所述通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,包括:

遍历所述日志区中的各个条带日志的关联信息,并根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;

针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

5. 根据权利要求2所述的RAID的写洞保护方法,其特征在于,保存的关联信息中至少包括:该条带日志的条带序号、日志编号、条带日志校验值以及日志版本号;

在利用所述日志区进行数据修复之后,还包括:将所述日志区中记录的总体信息中的日志区版本号进行更新;

相应的,所述通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,包括:

遍历所述日志区中的各个条带日志的关联信息,将日志版本号不符合当前的所述日志区中的日志区版本号的各个条带日志进行过滤;

根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;

针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为

有效的条带日志。

6. 根据权利要求2至5任一项所述的RAID的写洞保护方法,其特征在于,所述基于确定出的各个有效的条带日志进行数据修复,包括:

针对日志区中的任意一个有效的条带日志,当该条带日志备份的是失效成员盘的数据块时,基于该条带日志中的备份数据以及读取到的各个成员盘的数据计算出校验块数据,并利用计算出的所述校验块数据进行校验盘的数据修复;

针对日志区中的任意一个有效的条带日志,当该条带日志备份的是各个待写入的数据块时,将各个待写入的数据块写入相应的健康成员盘中完成数据修复。

7. 根据权利要求1所述的RAID的写洞保护方法,其特征在于,所述预先设定日志区,包括:

从所述RAID的每块成员盘中均划分出预设大小的空间,并通过划分出的各个空间构成设定的所述日志区。

8. 根据权利要求1所述的RAID的写洞保护方法,其特征在于,所述预先设定日志区,包括:

利用所述RAID成员盘之外的固态盘设定所述日志区。

9. 一种RAID的写洞保护系统,其特征在于,包括:

日志区设定模块,用于预先设定日志区,并在RAID降级之后,将所述日志区设置为启用状态;

第一判断模块,用于在所述日志区为启用状态下,每次在执行条带写入操作之前,均判断所述RAID的失效成员盘在该条带中的数据块是否为校验数据块;

如果不是校验数据块,则触发第二判断模块;

所述第二判断模块,用于判断该条带的各个待写入的数据块中是否包括待写入所述失效成员盘的数据块;

如果包括,则触发第一备份模块,如果不包括,则触发第二备份模块;

所述第一备份模块,用于在所述日志区中备份待写入所述失效成员盘的数据块;

所述第二备份模块,用于根据RAID算法计算出所述失效成员盘的数据块并在所述日志区中进行备份,或者在所述日志区中备份各个待写入的数据块;

数据修复模块,用于在降级后的所述RAID经过故障之后启动时,利用所述日志区进行数据修复。

10. 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上存储有计算机程序,所述计算机程序被处理器执行时实现如权利要求1至8任一项所述的RAID的写洞保护方法的步骤。

一种RAID的写洞保护方法、系统及存储介质

技术领域

[0001] 本发明涉及计算机数据存储技术领域,特别是涉及一种RAID的写洞保护方法、系统及存储介质。

背景技术

[0002] 机械硬盘、固态硬盘是目前计算机系统中最常用的数据存储设备,现代计算机服务器、存储系统一般配置多块盘,为了提高单个逻辑盘的容量以及性能,可以使用RAID (Redundant Array Of Independent Disks,磁盘冗余阵列) 技术将多块盘组成阵列。主流的RAID类型包括RAID0,RAID1,RAID5,RAID6等。其中,RAID5和RAID6在可用容量、性能方面的综合价值较高,应用最广泛。

[0003] 但是,对于RAID5和RAID6而言,在由于故障而降级之后,如果在降级后又发生了系统故障,导致读写的中断,就会存在数据错误的可能,也即RAID5和RAID6的“写洞”问题。

[0004] 以RAID5为例进行说明,对于一个条带而言,该条带中包括多个常规的数据块以及一个校验数据块,每个数据块存入不同的RAID成员盘中,例如一个条带中包括了1至5号数据块,其中的1号至4号均为常规数据块,依次存储在1号至4号成员盘中,5号为校验数据块存储在5号成员盘中。当任意一个成员盘故障时,均可以基于其他成员盘计算出该故障成员盘的数据块,例如1号成员盘故障,则可以根据2至4号成员盘中的常规数据块以及5号成员盘中的校验数据块对1号成员盘中的数据块进行恢复。出现1个成员盘故障之后,且在未修复之前,RAID5便处于降级状态。

[0005] 由于RAID的条带写入过程中,不是所有成员盘的写入情况和写入进度都是严格同步的,因此RAID5在降级之后,如果又发生了故障,例如断电等原因导致读写的中断,则对于某个条带而言,各个健康成员盘写入的数据量和起止位置会不同,之后在系统恢复后,根据健康成员盘计算故障成员盘的数据就可能出现错误。这也就是RAID5的“写洞”问题。

[0006] 相较于RAID5允许一个成员盘故障,RAID6的一个条带中具有两个校验数据块,最多允许两个成员盘故障,出现“写洞”问题的原因则与RAID5同理,就不再赘述。

[0007] 针对RAID5和RAID6的写洞问题,有方案提出了具有双控制器的存储系统,通过在双控制器之间同步备份条带写入数据的方法来解决写洞问题,但是,随着固态硬盘的普及和性能提高,多块盘的性能已高于控制器系统间的通信性能,这样的方案中,双控制器之间的通信性能成为瓶颈,并且这样的方案对于单控制器系统也不适用。

[0008] 综上所述,如何有效地解决RAID5和RAID6的写洞问题,是目前本领域技术人员急需解决的技术问题。

发明内容

[0009] 本发明的目的是提供一种RAID的写洞保护方法、系统及存储介质,以避免RAID的写洞问题。

[0010] 为解决上述技术问题,本发明提供如下技术方案:

- [0011] 一种RAID的写洞保护方法,包括:
- [0012] 预先设定日志区,并在RAID降级之后,将所述日志区设置为启用状态;
- [0013] 在所述日志区为启用状态下,每次在执行条带写入操作之前,均判断所述RAID的失效成员盘在该条带中的数据块是否为校验数据块;
- [0014] 如果不是校验数据块,则判断该条带的各个待写入的数据块中是否包括待写入所述失效成员盘的数据块;
- [0015] 如果包括,则在所述日志区中备份待写入所述失效成员盘的数据块;
- [0016] 如果不包括,则根据RAID算法计算出所述失效成员盘的数据块并在所述日志区中进行备份,或者在所述日志区中备份各个待写入的数据块;
- [0017] 在降级后的所述RAID经过故障之后启动时,利用所述日志区进行数据修复。
- [0018] 优选的,每次在所述日志区中进行数据块的备份时,还包括:保存该次备份操作的关联信息以构成一个条带日志;
- [0019] 相应的,所述利用所述日志区进行数据修复,包括:
- [0020] 通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,并基于确定出的各个有效的条带日志进行数据修复。
- [0021] 优选的,保存的关联信息中至少包括:该条带日志的条带序号以及日志编号;
- [0022] 相应的,所述通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,包括:
- [0023] 遍历所述日志区中的各个条带日志的关联信息,并且针对具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。
- [0024] 优选的,保存的关联信息中至少包括:该条带日志的条带序号、日志编号以及条带日志校验值;
- [0025] 相应的,所述通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,包括:
- [0026] 遍历所述日志区中的各个条带日志的关联信息,并根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;
- [0027] 针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。
- [0028] 优选的,保存的关联信息中至少包括:该条带日志的条带序号、日志编号、条带日志校验值以及日志版本号;
- [0029] 在利用所述日志区进行数据修复之后,还包括:将所述日志区中记录的总体信息中的日志区版本号进行更新;
- [0030] 相应的,所述通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,包括:
- [0031] 遍历所述日志区中的各个条带日志的关联信息,将日志版本号不符合当前的所述日志区中的日志区版本号的各个条带日志进行过滤;
- [0032] 根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;
- [0033] 针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

[0034] 优选的,所述基于确定出的各个有效的条带日志进行数据修复,包括:

[0035] 针对日志区中的任意一个有效的条带日志,当该条带日志备份的是失效成员盘的数据块时,基于该条带日志中的备份数据以及读取到的各个成员盘的数据计算出校验块数据,并利用计算出的所述校验块数据进行校验盘的数据修复;

[0036] 针对日志区中的任意一个有效的条带日志,当该条带日志备份的是各个待写入的数据块时,将各个待写入的数据块写入相应的健康成员盘中完成数据修复。

[0037] 优选的,所述预先设定日志区,包括:

[0038] 从所述RAID的每块成员盘中均划分出预设大小的空间,并通过划分出的各个空间构成设定的所述日志区。

[0039] 优选的,所述预先设定日志区,包括:

[0040] 利用所述RAID成员盘之外的固态盘设定所述日志区。

[0041] 一种RAID的写洞保护系统,包括:

[0042] 日志区设定模块,用于预先设定日志区,并在RAID降级之后,将所述日志区设置为启用状态;

[0043] 第一判断模块,用于在所述日志区为启用状态下,每次在执行条带写入操作之前,均判断所述RAID的失效成员盘在该条带中的数据块是否为校验数据块;

[0044] 如果不是校验数据块,则触发第二判断模块;

[0045] 所述第二判断模块,用于判断该条带的各个待写入的数据块中是否包括待写入所述失效成员盘的数据块;

[0046] 如果包括,则触发第一备份模块,如果不包括,则触发第二备份模块;

[0047] 所述第一备份模块,用于在所述日志区中备份待写入所述失效成员盘的数据块;

[0048] 所述第二备份模块,用于根据RAID算法计算出所述失效成员盘的数据块并在所述日志区中进行备份,或者在所述日志区中备份各个待写入的数据块;

[0049] 数据修复模块,用于在降级后的所述RAID经过故障之后启动时,利用所述日志区进行数据修复。

[0050] 一种计算机可读存储介质,所述计算机可读存储介质上存储有计算机程序,所述计算机程序被处理器执行时实现上述任一项所述的RAID的写洞保护方法的步骤。

[0051] 应用本发明实施例所提供的技术方案,并不涉及多控制器之间的通信,即本申请的方案可以应用在单控制器的场合中,也可以应用在多控制器的场合中。而本申请的方案中预先设定了日志区,在RAID降级之后,将日志区设置为启用状态,每次在执行条带写入操作之前,进行相应的备份操作,从而避免了写洞问题。具体的,当备份的是失效成员盘的数据块时,由于备份时是整体备份,即直接通过日志区便可以获得失效成员盘的数据块,且是正确数据。此外还可以读取其他各个健康成员盘的数据块,结合备份的失效成员盘的数据块,计算出正确的校验数据块,避免校验数据块出错的情况。如果备份的是各个待写入的数据块,即备份的是待写入健康成员盘的数据块,说明该次写操作不涉及失效成员盘,则可以直接将备份的各个待写入的数据块写入相应的健康成员盘中,使得每一个健康成员盘的数据均是正确的数据。因此本申请的方案避免了RAID的写洞问题。

附图说明

[0052] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0053] 图1为本发明中一种RAID的写洞保护方法的实施流程图;

[0054] 图2为本发明中一种RAID的写洞保护系统的结构示意图。

具体实施方式

[0055] 本发明的核心是提供一种RAID的写洞保护方法,避免了RAID的写洞问题。

[0056] 为了使本技术领域的人员更好地理解本发明方案,下面结合附图和具体实施方式对本发明作进一步的详细说明。显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0057] 请参考图1,图1为本发明中一种RAID的写洞保护方法的实施流程图,该RAID的写洞保护方法可以包括以下步骤:

[0058] 步骤S101:预先设定日志区,并在RAID降级之后,将日志区设置为启用状态。

[0059] 在设定日志区时,可以基于RAID的各个成员盘来设定出日志区,也可以利用RAID成员盘之外的其他硬盘作为日志空间。

[0060] 例如在一种具体实施方式中,预先设定日志区可以具体为:从RAID的每块成员盘中均划分出预设大小的空间,并通过划分出的各个空间构成设定的日志区。

[0061] 该种实施方式中,从RAID中的每一块硬盘中划分出一个空间,这些空间便可以构成设定的日志区。在RAID未降级时,设定的日志区为未启用状态,相应的,RAID降级之后,需要将日志区设置为启用状态。可以理解的是,RAID降级之后,由各个健康盘的相应空间组成可被使用的日志区,以基于该日志区进行后续步骤中的数据备份。

[0062] 该种实施方式中基于RAID的各个成员盘来设定出日志区,成本较低,便于方案的实施。而在部分实施方式中,也可以利用RAID成员盘之外的固态硬盘来设定日志区,例如使用性能更高的NVME固态硬盘、性能和寿命都更高的3DNAND固态硬盘等。

[0063] 步骤S102:在日志区为启用状态下,每次在执行条带写入操作之前,均判断RAID的失效成员盘在该条带中的数据块是否为校验数据块。如果否,则执行步骤S103。

[0064] 对于任意一次条带写入的操作,如果RAID的失效成员盘在该条带中的数据块为校验数据块,说明失效成员盘在该条带中的数据块不是用户数据,则可以不用进行备份,待写入的数据块中的各个常规数据块均可以被写入相应的健康成员盘中。也就是说,当判断出RAID的失效成员盘在该条带中的数据块为校验数据块时,可以直接开始执行本次条带写入的操作。

[0065] 步骤S103:判断该条带的各个待写入的数据块中是否包括待写入失效成员盘的数据块。如果包括,则执行步骤S104,如果不包括,则执行步骤S105。

[0066] RAID的失效成员盘在该条带中的数据块不为校验数据块,说明是常规数据块,但是,在每次执行条带写入操作时,该条带写入操作涉及到的常规数据块可以是不同的,例如

一个条带中包括了1至5号数据块,其中的1号至4号均为常规数据块,依次存储在1号至4号成员盘中,5号为校验数据块存储在5号成员盘中。例如执行某一次条带写入操作时,该条带写入操作需要将1号,2号以及3号数据块写入相应的成员盘中,并且假设失效成员盘为1号,则此时可以确定各个待写入的数据块中包括了待写入失效成员盘的数据块。

[0067] 又如执行某一次条带写入操作时,该条带写入操作需要将2号和4号数据块写入相应的成员盘中,并且假设失效成员盘为1号,则此时可以确定各个待写入的数据块中不包括待写入失效成员盘的数据块。

[0068] 步骤S104:在日志区中备份待写入失效成员盘的数据块。

[0069] 例如上述例子中,条带写入操作需要将1号,2号以及3号数据块写入相应的成员盘中,即待写入的数据块为1号,2号以及3号,其中1号数据块是待写入失效成员盘的数据块,则在日志区中备份1号数据块。

[0070] 步骤S105:根据RAID算法计算出失效成员盘的数据块并在日志区中进行备份,或者在日志区中备份各个待写入的数据块。

[0071] 步骤S105中公开了触发步骤S105之后的两种备份方式,实际应用中具体选取哪一种可以根据实际需要进行选取,并且也可以随时切换。无论选取哪一种,后续均可以实现数据修复。

[0072] 仍然假设一个条带中包括了1至5号数据块,其中的1号至4号均为常规数据块,依次存储在1号至4号成员盘中,5号为校验数据块存储在5号成员盘中。例如执行某一次条带写入操作时,该条带写入操作需要将1号,2号和5号数据块写入相应的成员盘中,并且假设失效成员盘为4号,则说明各个待写入的数据块中不包括待写入失效成员盘的数据块。

[0073] 此时,一种备份方式是基于RAID算法,即根据待写入的1号,2号和5号数据块,并结合3号数据块,计算出失效成员盘的数据块,即计算出4号成员盘的数据块,进而将计算出的该4号成员盘的数据块备份在日志区中。

[0074] 还有一种备份方式则是直接在日志区中备份各个待写入的数据块,该例子中即直接将待写入的1号,2号以及5号数据块备份在日志区中。

[0075] 可以理解的是,在执行完步骤S104或者步骤S105之后,便可以执行本次条带写入操作。即针对任意一次条带写入操作,判断出对于该条带写入操作而言所需备份的数据块并完成备份之后,便可以执行本次条带写入操作。

[0076] 还需要说明的是,每次在日志区中进行数据块的备份时,还可以包括:保存该次备份操作的关联信息以构成一个条带日志。从而使得后续在进行数据修复时,可以通过日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,并基于确定出的各个有效的条带日志进行数据修复。

[0077] 关联信息指的是进行数据块的备份时产生的相关参数数据,具体包括的项目可以根据需要进行设定和调整。在任意一次进行数据块的备份时,本次备份的各个数据块以及对应的关联信息便构成了一个条带日志。

[0078] 例如在本发明的一种具体实施方式中,保存的关联信息中至少包括:该条带日志的条带序号以及日志编号。

[0079] 则通过日志区中的各个条带日志的关联信息确定出每一个有效的条带日志的操作可以具体为:

[0080] 遍历日志区中的各个条带日志的关联信息,并且针对具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

[0081] 该种实施方式中是考虑到针对同一个条带,可能会执行一次或多次写操作,即可能会有多个条带日志都是针对同一个条带的情况,因此通过日志编号以及条带序号进行区分。

[0082] 对于任意一个条带日志而言,条带序号指的是该条带日志所针对的条带在RAID中的序号,即可以根据条带序号找到对应的条带以及对应的成员盘的起止扇区。该种实施方式中的日志编号是递增的编号,从而用于识别针对同一条带的各个条带日志的先后,即条带序号相同时,日志编号最大的条带日志为有效日志。在其他实施方式中,也可以采用非递增的其他形式的日志编号,能够区分出针对同一条带的各个条带日志的先后即可,并不影响本发明的实施。

[0083] 进一步地,在本发明的一种具体实施方式中,保存的关联信息中至少包括:该条带日志的条带序号、日志编号以及条带日志校验值;

[0084] 相应的,通过日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,可以具体包括以下两个步骤:

[0085] 第一个步骤:遍历日志区中的各个条带日志的关联信息,并根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;

[0086] 第二个步骤:针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

[0087] 该种实施方式中,关联信息中还包括了条带日志校验值,从而使得可以基于条带日志校验值确定条带日志中的数据块是否完整,采用的校验算法可选择公开的校验算法如CRC32校验算法。由于该种实施方式中在关联信息中加入了条带日志校验值,将不完整的各个条带日志进行了过滤,有利于及时发现条带日志异常的情况,进而也就有利于数据的准确修复。

[0088] 在本发明的一种具体实施方式中,保存的关联信息中至少包括:该条带日志的条带序号、日志编号、条带日志校验值以及日志版本号;

[0089] 该种实施方式中,在步骤S106之后,还需要包括:将日志区中记录的总体信息中的日志区版本号进行更新。即需要在修复完成之后,对日志区版本号进行更新。

[0090] 相应的,该种实施方式中的通过日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,可以具体包括以下三个步骤:

[0091] 第一个步骤:遍历日志区中的各个条带日志的关联信息,将日志版本号不符合当前的日志区中的日志区版本号的各个条带日志进行过滤;

[0092] 第二个步骤:根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;

[0093] 第三个步骤:针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

[0094] 该种实施方式中,是考虑到在部分场合中,在利用日志区进行数据修复之后,RAID可能还是处于降级状态,即工作人员未必能够及时将失效盘恢复为健康盘,因此,RAID需要继续在降级状态下进行写洞保护,即日志区会接着进行数据块的备份,产生相应的条带日

志。因此,为了区分出哪些条带日志是此前已被使用过的,哪些是未被使用的,该种实施方式中,关联信息中便包括了日志版本号,日志版本号需要与总体信息中的日志区版本号配合使用。

[0095] 具体的,该种实施方式中,日志区中需要记录有总体信息,总体信息中的相关数据用来描述日志区的属性。日志区版本号与各个条带日志的日志版本号配合使用,可以确定出哪些条带日志是未被使用的条带日志。此处描述的未被使用,即还没有基于该条带日志执行过数据修复。

[0096] 例如,RAID在某次降级之后,日志区启用,并且执行了16次条带的写入操作,且在日志区相应地备份了16个条带日志,各个条带日志的日志版本号例如均为001,日志区版本号也为001。之后RAID故障,则基于这16个条带日志进行数据修复。修复流程结束之后,日志区版本号更新,例如变更为002,之后例如又执行了20次条带的写入操作,且在日志区相应地备份了20个条带日志,各个条带日志的日志版本号均为002。之后RAID又故障,则需要进行数据修复时,日志区此时有 $16+20=36$ 个条带日志,则基于这36个条带日志进行数据修复时,具体过程可以为:首先可以将日志版本号不符合当前的日志区中的日志区版本号的各个条带日志进行过滤,即过滤掉日志版本号为001的16个条带日志,之后是根据各个关联信息中的条带日志校验值,将剩下的20个条带日志中不完整的各个条带日志进行过滤,例如过滤了3个不完整的条带日志,剩下17个条带日志。最后针对这17个条带日志中,具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志,例如,这17个条带日志中有3个条带日志的条带序号是相同的,其余的14个条带日志的条带序号各不相同,则最终可以确定出15个有效的条带日志。在利用这15个条带日志进行修复之后,可以将日志区版本号调整为003。

[0097] 该种例子中,日志区版本号在每次修复后进行更新,从而与日志版本号配合,以确定出哪些条带日志被使用过,在其他实施方式中,也可以有其他具体实现方式,并不影响本发明的实施。

[0098] 并且,前述实施例中,对日志区版本号进行更新时,是递增更新,即增1。而在其他实施方式中,也可以采用其他的更新方式,只要保证日志区版本号更新之后与各个历史日志区版本号不同即可,例如可以每次将日志区版本号的数值增加2,又如,可以采用一个较大的随机数进行更新等。

[0099] 此外需要说明的是,在实际应用中,关联信息中还可以包括其他参数,例如还可以包括条带日志的日志大小,日志大小表示该条带日志的连续占用空间大小。需要说明的是,由于条带日志是由备份的数据块以及对应的关联信息构成,因此日志大小包含了备份的数据块以及关联信息,即是整个条带日志的大小。又如,关联信息中还可以包括成员盘编号,数据长度等。

[0100] 当然,除了日志区版本号之外,日志区中的总体信息还可以包括其他参数例如总体信息中还包括:日志区的启用状态,日志空间大小,条带日志起始地址,条带日志空间大小,条带日志对齐区域个数以及日志区校验值等,使得工作人员可以基于总体信息快速、方便地获知日志区的状态。

[0101] 日志区的启用状态用于表示日志区是否启用,即在RAID降级之后,日志区启用,不降级则不启用。当然,RAID类型不为RAID5、RAID6时,也是不启用。

[0102] 条带日志空间大小为固定值,条带日志空间大小加上总体信息的占用空间即为日志空间大小。条带日志起始地址指的是首个条带日志的起始地址。

[0103] 条带日志空间大小除以条带日志对齐区域个数,可以得到每一个日志对齐区域的大小。例如条带日志空间大小为2G,预设的条带日志对齐区域个数例如为1024个,则每个日志对齐区域的大小为2M,该2M的含义是指用这2M的空间存放一个或者多个条带日志,但不会有条带日志跨过这2M空间的情况,即每个日志对齐区域的头部均是某一个条带日志的起止位置。这样的方式,将每个日志对齐区域的起始地址与相应的条带日志的起始地址对齐,可以更快速地定位查找条带日志。

[0104] 还需要指出的是,在往每个日志对齐区域中写入条带日志中,通常可以是顺序写,从而更好地保证硬盘寿命。

[0105] 日志区校验值可以实现对日志区的数据完整性的校验。

[0106] 步骤S106:在降级后的RAID经过故障之后启动时,利用日志区进行数据修复。

[0107] 具体的,可以根据日志区中的各个条带日志进行数据修复,当然,条带日志中的备份的数据块的内容不同时,所采用的数据修复手段可以不同。

[0108] 在利用日志区进行数据修复之前,可以确定日志区是否有效。确定有效之后,在利用各个条带日志进行数据修复时,如前文的描述,通常可以通过日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,并基于确定出的各个有效的条带日志进行数据修复。修复完毕之后,可以重新初始化日志区的总体信息,具体的,可以更新日志区版本号。

[0109] 对于某一个确定为有效的条带日志,如果该条带日志中备份的是各个待写入的数据块,即备份的是待写入健康成员盘的数据块时,说明该条带日志对应的写操作不涉及失效成员盘,则可以直接将备份的各个数据块写入相应的健康成员盘中,使得每一个健康成员盘的数据均是正确的数据,这样基于正确的健康成员盘确定出的失效成员盘的数据块自然也是正确的,因此避免了写洞问题。但需要说明的是,当备份的是待写入健康成员盘的数据块时,对日志区的空间占用较大。

[0110] 而如果条带日志中备份的是失效成员盘的数据块,由于备份时是整体备份,即直接通过条带日志便可以获得失效成员盘的数据块,因此也能够避免写洞问题。具体的,可以基于备份的失效成员盘的数据块以及各个健康成员盘中的常规数据块,计算出正确的校验数据块,将该校验数据块写入校验盘,便可以使得校验盘中的数据为准确的数据。之后,如果要获取失效成员盘的数据块,通过各个健康成员盘康便可以准确计算出。

[0111] 因此,假设出现了写洞问题,即出现了各健康成员盘的写入进度不同步的情况,本申请也能够获得正确的失效成员盘的数据。

[0112] 在本发明的一种具体实施方式中,基于确定出的各个有效的条带日志进行数据修复,可以具体包括:

[0113] 步骤一:针对日志区中的任意一个有效的条带日志,当该条带日志备份的是失效成员盘的数据块时,基于该条带日志中的备份数据以及读取到的各个成员盘的数据计算出检验块数据,并利用计算出的校验块数据进行校验盘的数据修复;

[0114] 步骤二:针对日志区中的任意一个有效的条带日志,当该条带日志备份的是各个待写入的数据块时,将各个待写入的数据块写入相应的健康成员盘中完成数据修复。

[0115] 具体的,针对日志区中的任意一个有效的条带日志,当该条带日志备份的是失效

成员盘的数据块时,该种实施方式可以基于备份数据进行校验盘的数据验证以及修复。即读取其他各个健康成员盘的数据块,结合该条带日志中的备份的失效成员盘的数据块,可以计算出正确的校验数据块。

[0116] 在计算出正确的校验数据块之后,可以直接将计算出的校验块数据直接写入校验盘中,这样后续便可以通过各个健康成员盘的数据块准确确定出失效成员盘的数据块,实现写洞保护。

[0117] 进一步的,在将计算出的校验块数据直接写入校验盘之前,还可以先从校验盘上读取校验数据块,如果计算出的校验数据块与读取的校验数据块不一致,说明校验盘上的校验数据块有误,可以进行事件记录,在记录之后,再将计算出的校验块数据写入校验盘。

[0118] 而针对日志区中的任意一个条带日志,当该条带日志备份的是各个待写入的数据块时,说明条带日志备份的是健康成员盘的数据块。则可以根据条带日志中备份的数据块以及其他成员盘的数据块,计算出失效盘的数据块,记为A。读取该条带在所有健康成员盘上的数据块进而计算出失效盘的数据块,记为B。如果A和B不同,说明健康成员盘上的数据块有误,可以进行事件记录,并将各个待写入的数据块写入相应的健康成员盘中完成数据修复,即将条带日志中的备份的数据块写入相应的健康成员盘中完成数据修复,通常,出现该错误的原因是校验盘的数据错误导致的。当然,也可以不执行该实施例中的比较操作,直接将各个待写入的数据块写入相应的健康成员盘中完成数据修复,因为备份的各个数据块均为正确的数据块。

[0119] 应用本发明实施例所提供的技术方案,并不涉及多控制器之间的通信,即本申请的方案可以应用在单控制器的场合中,也可以应用在多控制器的场合中。而本申请的方案中预先设定了日志区,在RAID降级之后,将日志区设置为启用状态,每次在执行条带写入操作之前,进行相应的备份操作,从而避免了写洞问题。具体的,当备份的是失效成员盘的数据块时,由于备份时是整体备份,即直接通过日志区便可以获得失效成员盘的数据块,且是正确数据。此外还可以读取其他各个健康成员盘的数据块,结合备份的失效成员盘的数据块,计算出正确的校验数据块,避免校验数据块出错的情况。如果备份的是各个待写入的数据块,即备份的是待写入健康成员盘的数据块,说明该次写操作不涉及失效成员盘,则可以直接将备份的各个待写入的数据块写入相应的健康成员盘中,使得每一个健康成员盘的数据均是正确的数据。

[0120] 相应于上面的方法实施例,本发明实施例还提供了一种RAID的写洞保护系统,可与上文相互对应参照。

[0121] 参见图2所示,为本发明中一种RAID的写洞保护系统的结构示意图,包括:

[0122] 日志区设定模块201,用于预先设定日志区,并在RAID降级之后,将日志区设置为启用状态;

[0123] 第一判断模块202,用于在日志区为启用状态下,每次在执行条带写入操作之前,均判断RAID的失效成员盘在该条带中的数据块是否为校验数据块;

[0124] 如果不是校验数据块,则触发第二判断模块203;

[0125] 第二判断模块203,用于判断该条带的各个待写入的数据块中是否包括待写入失效成员盘的数据块;

[0126] 如果包括,则触发第一备份模块204,如果不包括,则触发第二备份模块205;

[0127] 第一备份模块204,用于在日志区中备份待写入失效成员盘的数据块;

[0128] 第二备份模块205,用于根据RAID算法计算出失效成员盘的数据块并在日志区中进行备份,或者在日志区中备份各个待写入的数据块;

[0129] 数据修复模块206,用于在降级后的RAID经过故障之后启动时,利用日志区进行数据修复。

[0130] 在本发明的一种具体实施方式中,还包括条带日志生成模块,用于:每次在所述日志区中进行数据块的备份时,保存该次备份操作的关联信息以构成一个条带日志;

[0131] 相应的,所述数据修复模块206,具体用于:在降级后的RAID经过故障之后启动时,通过所述日志区中的各个条带日志的关联信息确定出每一个有效的条带日志,并基于确定出的各个有效的条带日志进行数据修复。

[0132] 在本发明的一种具体实施方式中,日志生成模块保存的关联信息中至少包括:该条带日志的条带序号以及日志编号;

[0133] 相应的,数据修复模块206,具体用于:

[0134] 在降级后的RAID经过故障之后启动时,遍历所述日志区中的各个条带日志的关联信息,并且针对具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

[0135] 在本发明的一种具体实施方式中,日志生成模块保存的关联信息中至少包括:该条带日志的条带序号、日志编号以及条带日志校验值;

[0136] 相应的,数据修复模块206,具体用于:

[0137] 在降级后的RAID经过故障之后启动时,遍历所述日志区中的各个条带日志的关联信息,并根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;

[0138] 针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

[0139] 在本发明的一种具体实施方式中,日志生成模块保存的关联信息中至少包括:该条带日志的条带序号、日志编号、条带日志校验值以及日志版本号;

[0140] 在利用所述日志区进行数据修复之后,还包括日志区版本号更新模块,用于在数据修复模块206利用所述日志区进行数据修复之后,将所述日志区中记录的总体信息中的日志区版本号进行更新;

[0141] 相应的,数据修复模块206,具体用于:

[0142] 在降级后的RAID经过故障之后启动时,遍历所述日志区中的各个条带日志的关联信息,将日志版本号不符合当前的所述日志区中的日志区版本号的各个条带日志进行过滤;

[0143] 根据各个关联信息中的条带日志校验值,将不完整的各个条带日志进行过滤;

[0144] 针对过滤后的具有相同的条带序号的各个条带日志,将日志编号最大的条带日志作为有效的条带日志。

[0145] 在本发明的一种具体实施方式中,数据修复模块206,具体用于:

[0146] 在降级后的RAID经过故障之后启动时,针对日志区中的任意一个有效的条带日志,当该条带日志备份的是失效成员盘的数据块时,基于该条带日志中的备份数据以及读取到的各个成员盘的数据计算出校验块数据,并利用计算出的所述校验块数据进行校验盘

的数据修复；

[0147] 在降级后的RAID经过故障之后启动时，针对日志区中的任意一个有效的条带日志，当该条带日志备份的是各个待写入的数据块时，将各个待写入的数据块写入相应的健康成员盘中完成数据修复。

[0148] 在本发明的一种具体实施方式中，日志区设定模块201具体用于：

[0149] 从所述RAID的每块成员盘中均划分出预设大小的空间，并通过划分出的各个空间构成设定的所述日志区。

[0150] 在本发明的一种具体实施方式中，日志区设定模块201具体用于：

[0151] 利用所述RAID成员盘之外的固态盘设定所述日志区。

[0152] 相应于上面的方法和系统实施例，本发明实施例还提供了一种计算机可读存储介质，计算机可读存储介质上存储有计算机程序，计算机程序被处理器执行时实现上述任一项的RAID的写洞保护方法的步骤。这里所说的计算机可读存储介质包括随机存储器 (RAM)、内存、只读存储器 (ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质。

[0153] 专业人员还可以进一步意识到，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，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，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发明的范围。

[0154] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的技术方案及其核心思想。应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以对本发明进行若干改进和修饰，这些改进和修饰也落入本发明权利要求的保护范围内。

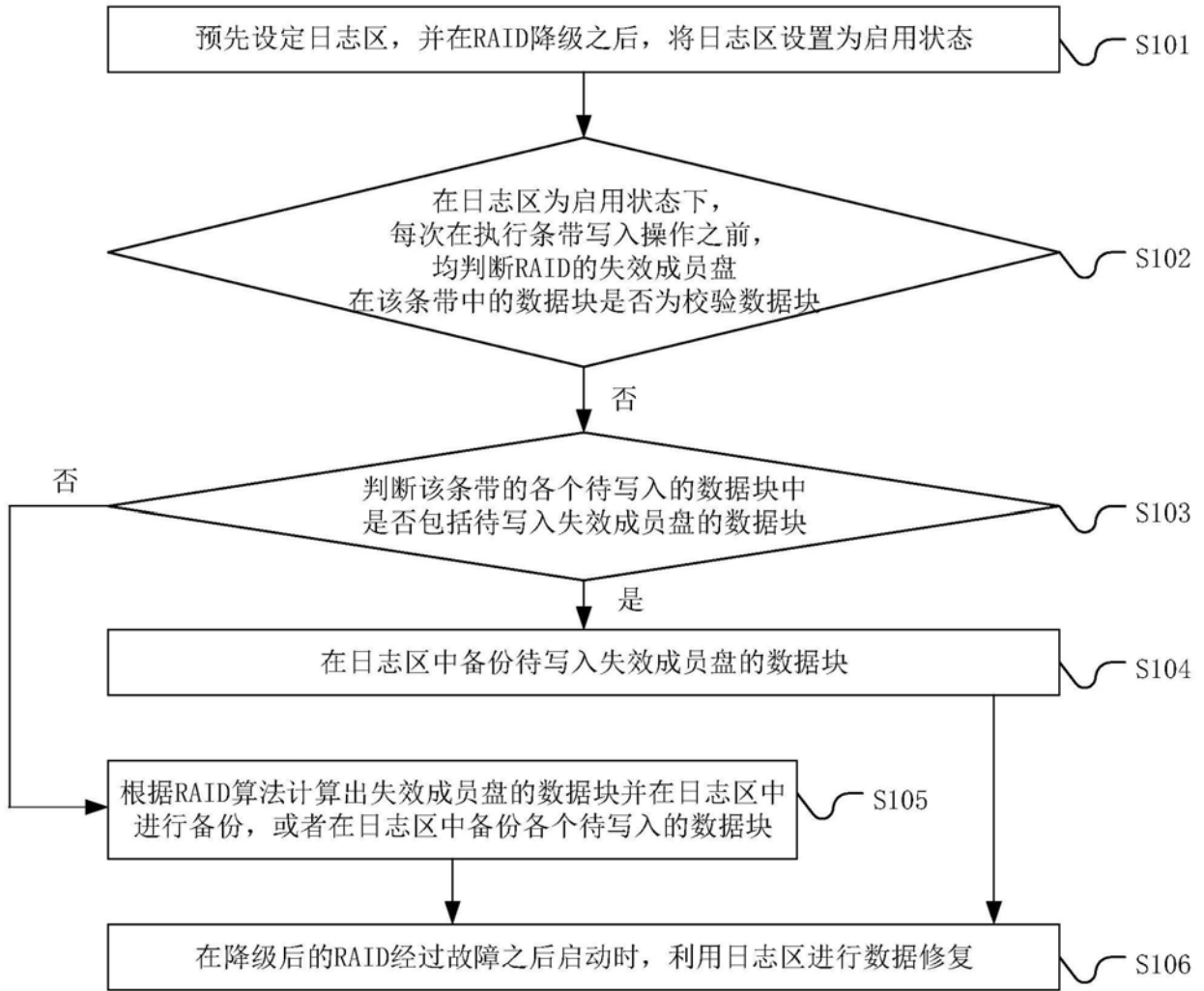


图1

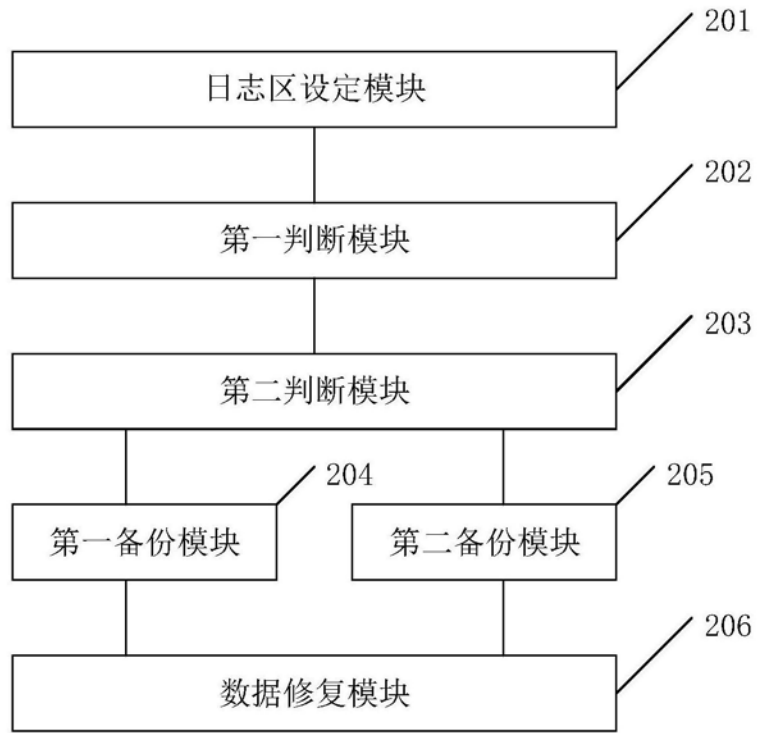


图2