



(12)发明专利申请

(10)申请公布号 CN 106980468 A

(43)申请公布日 2017.07.25

(21)申请号 201710124832.8

(22)申请日 2017.03.03

(71)申请人 杭州宏杉科技股份有限公司
地址 310053 浙江省杭州市滨江区西兴街
道阡陌路482号A楼第11层至12层

(72)发明人 上官应兰 张学东

(74)专利代理机构 北京博思佳知识产权代理有
限公司 11415
代理人 林祥

(51) Int. Cl.
G06F 3/06(2006.01)

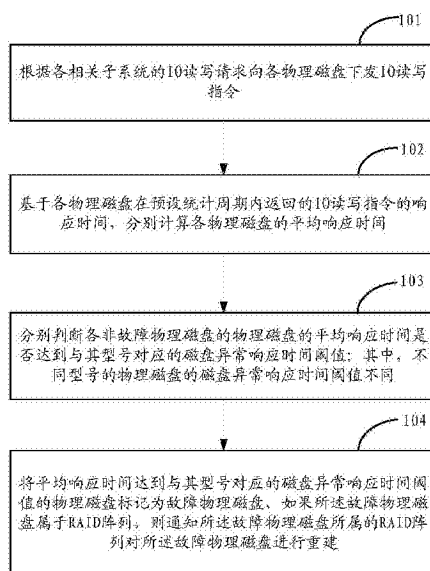
权利要求书3页 说明书9页 附图2页

(54)发明名称

触发RAID阵列重建的方法及装置

(57)摘要

本申请提供一种触发RAID阵列重建的方法及装置,所述方法应用于存储设备的磁盘子系统,可包括:向各物理磁盘下发IO读写指令;基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间,并分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值;将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。使用本方案可以有效地提高判断故障物理磁盘的准确性。



1. 一种触发RAID阵列重建的方法,其特征在于,所述方法应用于存储设备的磁盘子系统,所述存储设备包括至少一个RAID阵列,所述RAID阵列包括若干个物理磁盘,所述方法包括:

根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令;

基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间;

分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值;其中,不同型号的物理磁盘的磁盘异常响应时间阈值不同;

将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

2. 根据权利要求1所述的方法,其特征在于,所述与其型号对应的磁盘异常响应时间阈值为该型号的物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积;

所述分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值,包括:

分别计算各型号的平均响应时间不为零的非故障物理磁盘的物理磁盘数目;

分别累加各型号的若干个非故障物理磁盘的物理磁盘的平均响应时间;

分别将各型号的若干个非故障物理磁盘的物理磁盘累加得到的平均响应时间除以与其型号对应的平均响应时间不为零的非故障物理磁盘的物理磁盘数目,得到各型号物理磁盘的平均响应时间;

分别计算各型号的物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积,得到各型号物理磁盘的磁盘异常响应时间阈值;

判断各非故障物理磁盘的物理磁盘的平均响应时间是否到达与其型号对应的磁盘异常响应时间阈值。

3. 根据权利要求1所述的方法,其特征在于,所述基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间,包括:

累加各物理磁盘针对所述预设统计周期的已完成的IO读写指令的响应时间;

统计各物理磁盘针对所述预设统计周期的已完成的IO读写指令的个数;

将各物理磁盘分别对应的累加的响应时间和统计的IO读写指令的个数相除,分别获得各物理磁盘的平均响应时间。

4. 根据权利要求1所述的方法,其特征在于,所述将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,包括:

分别记录平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘的持续周期数;

如果在若干个统计周期后,所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的任一物理磁盘的持续周期数达到预设的持续周期阈值,则将该物理磁盘标记为故障物理磁盘。

5. 根据权利要求4所述的方法,其特征在于,所述分别记录平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘的持续周期数,包括:

针对所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的每个物理磁盘,在下一个统计周期结束时,如果该物理磁盘再次被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则增加该物理磁盘的持续周期数并记录;如果该物理磁盘未被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则减少该物理磁盘的持续周期数并记录;其中,物理磁盘的持续周期数的初始值为零。

6. 一种触发RAID阵列重建的装置,其特征在于,所述装置应用于存储设备的磁盘子系统,所述存储设备包括至少一个RAID阵列,所述RAID阵列包括若干个物理磁盘,所述装置包括:

下发单元,用于根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令;

计算单元,用于基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间;

判断单元,用于分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值;其中,不同型号的物理磁盘的磁盘异常响应时间阈值不同;

标记单元,用于将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

7. 根据权利要求6所述的装置,其特征在于,所述与其型号对应的磁盘异常响应时间阈值为该型号的物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积;

所述判断单元,具体用于分别计算各型号的平均响应时间不为零的非故障物理磁盘的物理磁盘数目;分别累加各型号的若干个非故障物理磁盘的物理磁盘的平均响应时间;分别将各型号的若干个非故障物理磁盘的物理磁盘累加得到的平均响应时间除以与其型号对应的平均响应时间不为零的非故障物理磁盘的物理磁盘数目,得到各型号物理磁盘的平均响应时间;分别计算各型号的物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积,得到各型号物理磁盘的磁盘异常响应时间阈值;判断各非故障物理磁盘的物理磁盘的平均响应时间是否到达与其型号对应的磁盘异常响应时间阈值。

8. 根据权利要求6所述的装置,其特征在于,所述计算单元,具体用于累加各物理磁盘针对所述预设统计周期的已完成的IO读写指令的响应时间;统计各物理磁盘针对所述预设统计周期的已完成的IO读写指令的个数;将各物理磁盘分别对应的累加的响应时间和统计的IO读写指令的个数相除,分别获得各物理磁盘的平均响应时间。

9. 根据权利要求6所述的装置,其特征在于,所述标记单元,具体用于分别记录平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘的持续周期数;如果在若干个统计周期后,所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的任一物理磁盘的持续周期数达到预设的持续周期阈值,则将该物理磁盘标记为故障物理磁盘。

10. 根据权利要求9所述的装置,其特征在于,所述标记单元,进一步用于针对所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的每个物理磁盘,在下一个统计周期结束时,如果该物理磁盘再次被确定为所述平均响应时间达到与其型号对应

的磁盘异常响应时间阈值的物理磁盘,则增加该物理磁盘的持续周期数并记录;如果该物理磁盘未被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则减少该物理磁盘的持续周期数并记录;其中,物理磁盘的持续周期数的初始值为零。

触发RAID阵列重建的方法及装置

技术领域

[0001] 本申请涉及计算机通信领域,尤其涉及触发RAID阵列重建的方法及装置。

背景技术

[0002] RAID阵列(Redundant Array of Independent Disks,独立磁盘冗余阵列)是一种把多块独立的磁盘(物理磁盘)按不同的方式组合起来形成一个磁盘组(逻辑磁盘),从而提供比单个磁盘更高的存储性能和数据可靠性的技术。

[0003] 在计算机通信领域,通常会使用RAID阵列技术对磁盘中数据进行冗余保护,当有数据写入时,根据RAID阵列算法把数据拆分到多个成员磁盘中。根据RAID阵列级别不同,可容忍1块或多块磁盘故障或者离线,当检测到磁盘IO错误或者磁盘离线时,可使用专用热备盘或者全局热备盘进行重建,恢复RAID阵列数据冗余性。

[0004] 然而,在现有的触发RAID阵列进行重建的方法中,仅考虑了磁盘IO错误和磁盘离线的情况,没有考虑磁盘老化后响应时间变慢导致业务中断的情况,因此如何在磁盘响应慢的情况下触发RAID阵列重建成为亟待解决的问题。

发明内容

[0005] 有鉴于此,本申请提供一种触发RAID阵列重建的方法及装置,用以提高判断故障物理磁盘的准确性。

[0006] 具体地,本申请是通过如下技术方案实现的:

[0007] 根据本申请的第一方面,提供一种触发RAID阵列重建的方法,所述方法应用于存储设备的磁盘子系统,所述存储设备包括至少一个RAID阵列,所述RAID阵列包括若干个物理磁盘,所述方法包括:

[0008] 根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令;

[0009] 基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间;

[0010] 分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值;其中,不同型号的物理磁盘的磁盘异常响应时间阈值不同;

[0011] 将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

[0012] 根据本申请的第二方面,一种触发RAID阵列重建的装置,所述装置应用于存储设备的磁盘子系统,所述存储设备包括至少一个RAID阵列,所述RAID阵列包括若干个物理磁盘,所述装置包括:

[0013] 下发单元,用于根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令;

[0014] 计算单元,用于基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间;

[0015] 判断单元,用于分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值;其中,不同型号的物理磁盘的磁盘异常响应时间阈值不同;

[0016] 标记单元,用于将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

[0017] 在本申请提出一种触发RAID阵列重建的方法中,一方面,由于磁盘子系统可以基于各物理磁盘的平均响应时间,将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的非故障物理磁盘的物理磁盘标记为故障物理磁盘,并通知该故障物理磁盘所属的RAID阵列进行重建,从而实现了基于物理磁盘的IO读写指令的响应时间来触发对该物理磁盘所属的RAID阵列的重建。

[0018] 另一方面,由于各物理磁盘的响应时间可以与其型号对应的磁盘异常响应时间阈值进行比较,从而使得在判断物理磁盘的平均响应时间是否异常时,综合考虑该磁盘上所有业务下发的IO,比如RAID下发的IO、磁盘检测任务下发的IO等,从而有效地提高磁盘子系统标记出的故障物理磁盘的准确率。

附图说明

[0019] 图1是本申请一示例性实施例示出的一种触发RAID阵列重建的方法的流程图;

[0020] 图2是本申请一示例性实施例示出的一种触发RAID阵列重建的装置所在设备的硬件结构图;

[0021] 图3是本申请一示例性实施例示出的一种触发RAID阵列重建的装置的框图。

具体实施方式

[0022] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请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、本申请的一些方面相一致的装置和方法的例子。

[0023] 在本申请使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本申请。在本申请和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。还应当理解,本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0024] 应当理解,尽管在本申请可能采用术语第一、第二、第三等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本申请范围的情况下,第一信息也可以被称为第二信息,类似地,第二信息也可以被称为第一信息。取决于语境,如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0025] RAID阵列是一种把多块独立的磁盘(物理磁盘)按不同的方式组合起来形成一个磁盘组(逻辑磁盘),从而提供比单个磁盘更高的存储性能和数据可靠性的技术。

[0026] 在计算机通信领域,通常会使用RAID阵列技术对磁盘中数据进行冗余保护,当有

数据写入时,根据RAID阵列算法把数据拆分到多个成员磁盘中。根据RAID阵列级别不同,可容忍1块或多块磁盘故障或者离线,当检测到磁盘IO错误或者磁盘离线时,可使用专用热备盘或者全局热备盘进行重建,恢复RAID阵列数据冗余性。

[0027] 在相关的RAID阵列触发重建的方法中,当RAID子系统接收到成员磁盘返回的IO读写错误,并且判断该错误无法恢复时,可以标记该成员磁盘故障,并触发该成员磁盘所属的RAID阵列重建。此外,当RAID子系统接收到成员磁盘离线的通知消息时,也可以触发该离线的成员磁盘所属的RAID阵列重建。

[0028] 在重建时,可以使用热备盘重建故障盘或者离线盘,RAID子系统可以按照RAID阵列算法计算出热备盘中对应条带的的数据,恢复该故障磁盘所属的RAID阵列的冗余性。

[0029] 由于磁盘是机械和电子结合的装置,受到器件老化、环境等因素的影响,在实际应用中可能出现磁盘IO不返错但是响应时间变慢的现象,将导致上层应用读写该磁盘对应的RAID阵列时,响应时间变慢的磁盘上IO返回慢于其他磁盘,上层应用的性能出现波动或IO超时。具体表现为,在将RAID阵列上创建的LUN(逻辑单元号)分配给前端应用服务器进行持续读写时,可能出现LUN的性能有很大的波动甚至IO超时业务中断的情况,但是开发人员在对该LUN性能大波动的现象进行排查时,发现该RAID阵列状态正常,磁盘状态也正常,成员磁盘也未返回IO读写错误。进一步排查,虽然该RAID阵列的成员磁盘的接口相同,转速相同,但是部分成员磁盘上返回的IO读写响应的响应时间明显长于该RAID阵列中其他的成员磁盘。在拔走IO读写响应时间长的成员磁盘,使用热备盘代替IO读写响应时间长的成员磁盘后,该RAID阵列性能和LUN性能恢复正常。

[0030] 综上所述可知,由于成员磁盘的IO读写响应时间长会严重影响该成员磁盘所属的RAID阵列和该RAID阵列上创建的LUN的性能,出现性能波动,极端情况下出现IO超时可能导致业务中断。然而,在现有的触发RAID阵列进行重建的方法中,仅考虑了磁盘IO错误和磁盘离线的情况,没有考虑磁盘老化后响应时间变慢导致业务中断的情况。

[0031] 本申请提出一种触发RAID阵列重建的方法,存储设备的磁盘子系统可以根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令,并可以基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间。磁盘子系统可以分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与该物理磁盘的型号对应的磁盘异常响应时间阈值。磁盘子系统可将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

[0032] 一方面,由于磁盘子系统可以基于各物理磁盘的平均响应时间,将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的非故障物理磁盘的物理磁盘标记为故障物理磁盘,并通知该故障物理磁盘所属的RAID阵列进行重建,从而实现了基于物理磁盘的IO读写指令的响应时间来触发对该物理磁盘所属的RAID阵列的重建。

[0033] 另一方面,由于各物理磁盘的响应时间可以与其型号对应的磁盘异常响应时间阈值进行比较,从而使得在判断物理磁盘的平均响应时间是否异常时,综合考虑该磁盘上所有业务下发的IO,比如RAID下发的IO、磁盘检测任务下发的IO等,从而有效地提高磁盘子系统标记出的故障物理磁盘的准确率。

[0034] 参见图1,图1是本申请一示例性实施例示出的一种触发RAID阵列重建的方法的流

程图,所述方法应用于存储设备的磁盘子系统,所述存储设备包括至少一个RAID阵列,所述RAID阵列包括若干个物理磁盘,所述方法包括:

[0035] 步骤101:根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令;

[0036] 步骤102:基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间;

[0037] 步骤103:分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值;其中,不同型号的物理磁盘的磁盘异常响应时间阈值不同;

[0038] 步骤104:将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

[0039] 其中,上述RAID子系统,用于管理存储设备中的各RAID阵列。例如,RAID子系统的功能可以包括对接收到的多个IO读写指令进行基于RAID阵列算法的拆分,将拆分后IO读写指令下发给磁盘子系统。RAID子系统的功能还可以包括,在RAID阵列重建时,基于该RAID阵列的算法,计算出热备盘中对应条带的的数据,恢复RAID阵列数据冗余性等功能。当然,RAID子系统还具有多种功能。在这里,不对RAID子系统的功能进行具体的限定。

[0040] 上述磁盘子系统,主要用于管理存储设备中的所有的物理磁盘。磁盘子系统相当于“下层”系统,用于服务“上层”系统如RAID子系统。例如,磁盘子系统可以对存储设备中的物理磁盘进行扫描,并向RAID子系统通知扫描出的各物理磁盘的状态信息等。当然,在实际应用中,磁盘子系统还有其他功能,在这里,不对磁盘子系统的功能进行具体地限定。

[0041] 上述RAID阵列,是指由存储设备中的多个物理磁盘按RAID阵列算法组合成的磁盘组。该RAID阵列中的物理磁盘也可以被称之为成员磁盘。RAID阵列根据其级别的不同以及实现方式的不同,所支持同时重建的物理磁盘个数也不相同。例如,传统RAID5可以支持的同时重建的物理磁盘个数为一个,在某些厂家的实现中,可支持同时重建多个。

[0042] 上述物理磁盘的响应时间,是指磁盘子系统向物理磁盘下发IO读写指令到该物理磁盘返回该IO读写响应所需要的时间。

[0043] 上述物理磁盘的平均响应时间,是指在预设的统计周期内,该物理磁盘累加的已完成的IO读写指令的响应时间除以该物理磁盘累加的已完成的IO读写指令的个数为该物理磁盘的平均响应时间。其中,如果某个物理磁盘对应的累加的已完成的IO读写指令的个数为零,该物理磁盘的平均响应时间按零处理。

[0044] 上述磁盘异常响应时间阈值,用于判断物理磁盘的响应时间是否异常,当物理磁盘的平均响应时间达到(大于或者等于)该异常响应时间阈值时,表示该物理磁盘异常。

[0045] 需要说明的是,上述磁盘异常响应时间阈值可由开发人员根据实际测试数据进行设置,或者可由用户通过交互界面手动输入等。其中,开发人员在设置或者用户手动输入该磁盘异常响应时间阈值时,如果将该磁盘异常响应时间阈值设置的过大,则可能无法精确地检测到IO读写指令的响应时间过长的物理磁盘,使得RAID阵列和其上创建的LUN的性能无法很好地恢复。如果将该磁盘异常响应时间阈值设置地过小,则可能检测到大量性能异常的物理磁盘,误伤平均响应时间正常的物理磁盘,从而造成RAID阵列的频繁重建,影响存储设备的性能。所以在实际应用中,开发人员可以根据实际情况对该异常响应时间阈值进

行设定。例如,开发人员可以将该磁盘异常响应时间阈值设置为同一型号的物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积等。在这里,只是对磁盘异常响应时间阈值的设置进行示例性的说明,不对其进行具体的限定。

[0046] 在本申请实施例中,由于磁盘子系统基于各物理磁盘的平均响应时间,将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的非故障物理磁盘的物理磁盘标记为故障物理磁盘,并通知该故障物理磁盘所属的RAID进行重建,从而实现了基于物理磁盘的IO读写指令的响应时间来触发对该物理磁盘所属的RAID阵列的重建。

[0047] 此外,由于各物理磁盘的响应时间可以与其型号对应的磁盘异常响应时间阈值进行比较,从而使得在判断物理磁盘的平均响应时间是否异常时,综合考虑该磁盘上所有业务下发的IO,比如RAID下发的IO、磁盘检测任务下发的IO等,从而有效地提高磁盘子系统标记出的故障物理磁盘的准确率。

[0048] 下面对上述触发RAID阵列重建的方法进行详细地说明。

[0049] 在实现时,在预设统计周期内,磁盘子系统可以根据各相关子系统的IO读写请求向各个物理磁盘下发IO读写指令。各物理磁盘在接收到IO读写指令并进行相应的处理之后,可以向磁盘子系统返回IO读写响应。

[0050] 磁盘子系统可以基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,对各物理磁盘在预设统计周期内的所有的已完成的IO读写指令的响应时间和已完成的IO读写指令的个数进行统计。在统计周期结束时,分别计算各物理磁盘的平均响应时间。

[0051] 下面以计算一个物理磁盘的平均响应时间为例,对计算物理磁盘的平均响应时间的方法进行详细地说明。

[0052] 在实现时,磁盘子系统在向某个物理磁盘下发IO读写指令时开始计时,在接收到该物理磁盘返回的该IO读写指令的响应后结束计时,可以计算该物理磁盘上该IO读写指令的响应时间,并可以将此次的IO读写指令的响应时间与当前周期内的该物理磁盘的已完成的IO读写指令的响应时间进行累加,并将该物理磁盘的已完成的IO读写指令的个数加1。

[0053] 在当前统计周期结束时,磁盘子系统可以统计该物理磁盘在当前统计周期的平均响应时间。在实现时,磁盘子系统可以将累加的该物理磁盘的IO读写指令的响应时间除以累加的该物理磁盘的已完成的IO读写指令的个数,得到该物理磁盘的平均响应时间。如果某个物理磁盘在当前统计周期内累加的已完成的IO读写指令的个数为0,该物理磁盘的平均响应时间按照零处理。

[0054] 其他物理磁盘的平均响应时间的计算方法与上述描述的计算方法相同,在这里不再赘述。

[0055] 在当前统计周期结束时,磁盘子系统还可以计算各型号物理磁盘对应的磁盘异常响应时间阈值。

[0056] 在一种可选的实现方式中,磁盘子系统可以计算各型号物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积,作为各型号物理磁盘对应的磁盘异常响应时间阈值。

[0057] 下面对某个型号物理磁盘对应的磁盘异常响应时间阈值的计算方法进行详细地说明。

[0058] 在实现时,磁盘子系统可以对该型号的平均响应时间不为零的非故障物理磁盘的

多个物理磁盘在当前统计周期的IO读写指令的平均响应时间进行累加,并可累加平均响应时间不为零的非故障物理磁盘的物理磁盘的数目。然后用累加得到的该型号的非故障物理磁盘的物理磁盘的IO读写指令的平均响应时间除以累加得到的该型号的平均响应时间不为零的非故障物理磁盘的物理磁盘数目,获得该型号的物理磁盘的平均响应时间。

[0059] 磁盘子系统可以将计算得到的该型号物理磁盘的平均响应时间与预设的异常响应时间加权值相乘,获得该型号物理磁盘对应的磁盘异常响应时间阈值。

[0060] 其中,上述异常响应时间加权值,可用于判断磁盘平均响应时间是否异常,通常由用户根据实际情况进行自行设定,以百分比的形式存在,如200%等,在这里,不对该异常响应时间加权值进行特别地限定。

[0061] 其余型号的物理磁盘的平均响应时间的计算方式与上述该型号物理磁盘的平均响应时间的计算方式相同,在这里不再赘述。

[0062] 在本申请实施例中,磁盘子系统可以通过上述方法获得各物理磁盘的平均响应时间和各型号物理磁盘的磁盘异常响应时间阈值。磁盘子系统可以判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到(大于或者等于)与该型号物理磁盘对应的磁盘异常响应时间阈值。

[0063] 磁盘子系统可以将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对该故障物理磁盘进行重建。

[0064] 为了提高磁盘子系统检测故障物理磁盘的精准性,避免将临时性异常响应的物理磁盘标记为故障物理磁盘,磁盘子系统可以不立即标记该物理磁盘为故障物理磁盘,而是记录该物理磁盘在连续若干个周期内的判断结果,如果该物理磁盘在连续若干个周期内的均被确定为平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则将该物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则触发该物理磁盘所属的RAID阵列进行重建。

[0065] 在一种可选的实现方式中,为增加检测故障物理磁盘的准确性与实用性,上述连续若干个统计周期,可为“相对连续”的若干统计周期。

[0066] 在标记时,上述磁盘子系统可分别记录平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘的持续周期数;如果在若干个统计周期后,平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的任一磁盘的持续周期数达到预设的持续周期阈值,则将该物理磁盘标记为故障物理磁盘。

[0067] 在记录时,磁盘子系统可针对所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的每一个物理磁盘,在下一个统计周期结束时,如果该物理磁盘再次被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则增加该物理磁盘的持续周期数并记录;如果该物理磁盘未被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则减少该物理磁盘的持续周期数并记录;其中,物理磁盘的持续周期数的初始值为零。

[0068] 例如,在上述物理磁盘中,如果某个物理磁盘第一次被确定为平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则可将该物理磁盘的持续周期数设置为1。在下一个统计周期结束时,如果该物理磁盘再次被确定为所述平均响应时间达到与其

型号对应的磁盘异常响应时间阈值的物理磁盘,则将该物理磁盘的持续周期数自加1;如果该物理磁盘没有被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则将该物理磁盘的持续周期数自减1,如果该物理磁盘的持续周期数减到零,不再记录该物理磁盘的持续周期。

[0069] 在另一种可选的实现方式中,上述磁盘子系统还可基于“绝对连续”的若干个统计周期对物理磁盘进行故障物理磁盘标记。

[0070] 在实现时,上述磁盘子系统可分别记录上述查找到的平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘的持续周期数;如果在若干个统计周期后,该平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的任一成员磁盘的持续周期数达到预设的持续周期阈值,则将该物理磁盘标记为故障物理磁盘。

[0071] 在记录时,磁盘子系统可针对确定的平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的每个成员磁盘,在下一个统计周期结束时,如果该物理磁盘再次被确定为平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则增加该物理磁盘的持续周期数并记录;如果该物理磁盘没有被查找为平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则将该物理磁盘的持续周期数置为零。其中,成员磁盘的持续周期数的初始值为零。

[0072] 此外,需要说明的是,在RAID阵列在接收到上述磁盘子系统针对上述故障物理磁盘重建的通知后,可以判断该RAID阵列当前是否满足重建条件,如果满足,则对该故障物理磁盘进行重建;如果不满足,则不对该故障物理磁盘进行重建,直到该RAID阵列满足重建条件时再触发对该故障物理磁盘进行重建。

[0073] 其中,上述重建条件可包括,RAID阵列健康状态满足重建要求,有可用热备盘,重建上述故障物理磁盘不会超过该故障物理磁盘所属的RAID阵列支持的同时重建的物理磁盘的个数等。在实际应用中,开发人员可以根据实际情况,设定上述重建条件,这里只是对重建条件进行示例性说明,不对其进行特别地限定。

[0074] 在磁盘子系统完成本周期的相关判断和处理后,磁盘子系统可以将统计的各物理磁盘在当前统计周期的已完成IO读写指令的响应时间和已完成IO读写指令的个数清空,以使得磁盘子系统可以在下一个统计周期对这两个参数进行统计。

[0075] 本申请提出一种触发RAID阵列重建的方法,存储设备的磁盘子系统可以根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令,并可以基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间。磁盘子系统可以分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与该物理磁盘的型号对应的磁盘异常响应时间阈值。磁盘子系统可将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

[0076] 一方面,由于磁盘子系统可以基于各物理磁盘的平均响应时间,将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的非故障物理磁盘的物理磁盘标记为故障物理磁盘,并通知该故障物理磁盘所属的RAID进行重建,从而实现了基于物理磁盘的IO读写指令的响应时间来触发对该物理磁盘所属的RAID阵列的重建。

[0077] 另一方面,由于各物理磁盘的响应时间可以与其型号对应的磁盘异常响应时间阈

值进行比较,从而使得在判断物理磁盘的平均响应时间是否异常时,综合考虑该磁盘上所有业务下发的IO,比如RAID下发的IO、磁盘检测任务下发的IO等,从而有效地提高磁盘子系统标记出的故障物理磁盘的准确率。

[0078] 此外,由于磁盘子系统可以在连续若干个周期检测到同一物理磁盘为故障物理磁盘时才触发该物理磁盘所属的RAID阵列重建,因此可以有效地提高RAID子系统检测故障成员磁盘的精准性,避免出现成员磁盘临时性异常响应。

[0079] 与前述触发RAID阵列重建的方法的实施例相对应,本申请还提供了触发RAID阵列重建的装置的实施例。

[0080] 本申请触发RAID阵列重建的装置的实施例可以应用在存储设备上。装置实施例可以通过软件实现,也可以通过硬件或者软硬件结合的方式实现。以软件实现为例,作为一个逻辑意义上的装置,是通过其所在存储设备的处理器将非易失性存储器中对应的计算机程序指令读取到内存中运行形成的。从硬件层面而言,如图2所示,为本申请触发RAID阵列重建的装置所在存储设备的一种硬件结构图,除了图2所示的处理器、内存、网络出接口、以及非易失性存储器之外,实施例中装置所在的存储设备通常根据该存储的实际功能,还可以包括其他硬件,对此不再赘述。

[0081] 请参考图3,图3是本申请一示例性实施例示出的一种触发RAID阵列重建的装置的框图。所述装置应用于存储设备的磁盘子系统,所述存储设备包括至少一个RAID阵列,所述RAID阵列包括若干个物理磁盘,所述装置包括:

[0082] 下发单元310,用于根据各相关子系统的IO读写请求向各物理磁盘下发IO读写指令;

[0083] 计算单元320,用于基于各物理磁盘在预设统计周期内返回的IO读写指令的响应时间,分别计算各物理磁盘的平均响应时间;

[0084] 判断单元330,用于分别判断各非故障物理磁盘的物理磁盘的平均响应时间是否达到与其型号对应的磁盘异常响应时间阈值;其中,不同型号的物理磁盘的磁盘异常响应时间阈值不同;

[0085] 标记单元340,用于将平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘标记为故障物理磁盘,如果所述故障物理磁盘属于RAID阵列,则通知所述故障物理磁盘所属的RAID阵列对所述故障物理磁盘进行重建。

[0086] 在一种可选的实现方式中,所述与其型号对应的磁盘异常响应时间阈值为该型号的物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积;

[0087] 所述判断单元330,具体用于分别计算各型号的平均响应时间不为零的非故障物理磁盘的物理磁盘数目;分别累加各型号的若干个非故障物理磁盘的物理磁盘的平均响应时间;分别将各型号的若干个非故障物理磁盘的物理磁盘累加得到的平均响应时间除以累加的与其型号对应的平均响应时间不为零的非故障物理磁盘的物理磁盘数目,得到各型号物理磁盘的平均响应时间;分别计算各型号的物理磁盘的平均响应时间与预设的异常响应时间加权值的乘积,得到各型号物理磁盘的磁盘异常响应时间阈值;判断各非故障物理磁盘的物理磁盘的平均响应时间是否到达与其型号对应的磁盘异常响应时间阈值。

[0088] 在另一种可选的实现方式中,所述计算单元320,具体用于累加各物理磁盘针对所述预设统计周期的已完成的IO读写指令的响应时间;统计各物理磁盘针对所述预设统计周

期的已完成的IO读写指令的个数;将各物理磁盘分别对应的累加的响应时间和统计的IO读写指令的个数相除,分别获得各物理磁盘的平均响应时间。

[0089] 在另一种可选的实现方式中,所述标记单元340,具体用于分别记录平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘的持续周期数;如果在若干个统计周期后,所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的任意一物理磁盘的持续周期数达到预设的持续周期阈值,则将该物理磁盘标记为故障物理磁盘。

[0090] 在另一种可选的实现方式中,所述标记单元340,进一步用于针对所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘中的每个物理磁盘,在下一个统计周期结束时,如果该物理磁盘再次被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则增加该物理磁盘的持续周期数并记录;如果该物理磁盘未被确定为所述平均响应时间达到与其型号对应的磁盘异常响应时间阈值的物理磁盘,则减少该物理磁盘的持续周期数并记录;其中,物理磁盘的持续周期数的初始值为零。

[0091] 上述装置中各个单元的功能和作用的实现过程具体详见上述方法中对应步骤的实现过程,在此不再赘述。

[0092] 对于装置实施例而言,由于其基本对应于方法实施例,所以相关之处参见方法实施例的部分说明即可。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模块来实现本申请方案的目的。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0093] 以上所述仅为本申请的较佳实施例而已,并不用以限制本申请,凡在本申请的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本申请保护的范围之内。

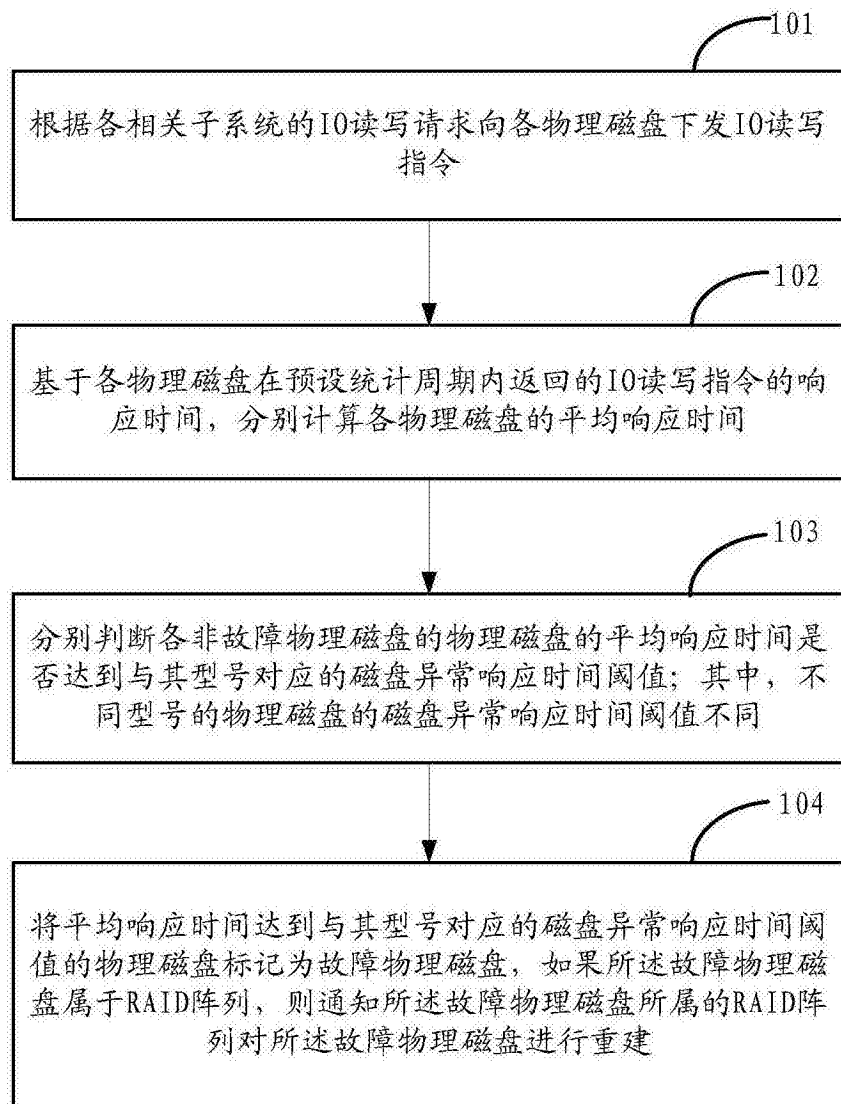


图1

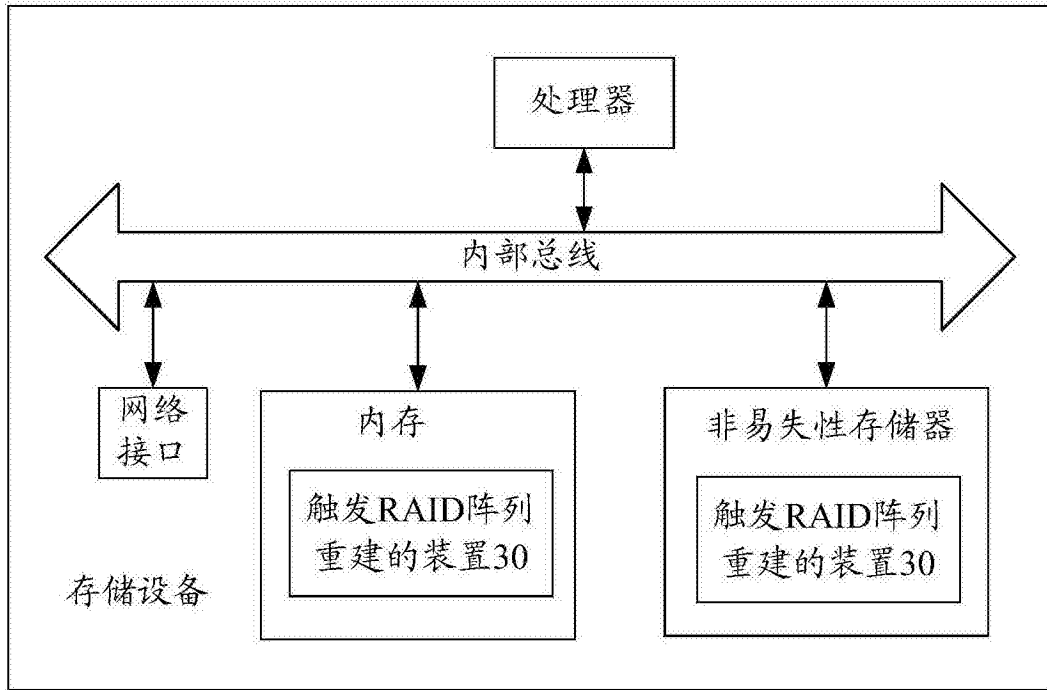


图2

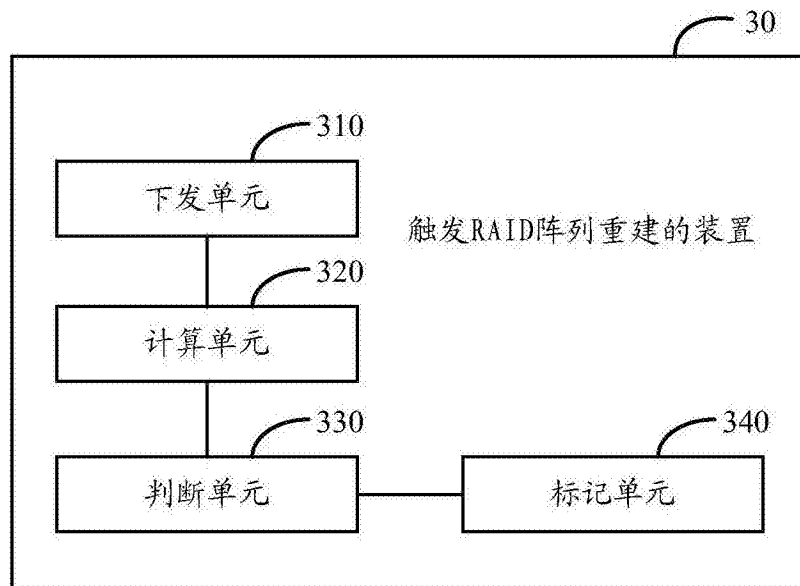


图3