(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113687788 B
(45)授权公告日 2025.04.11

(21)申请号 202111005445.5
(22)申请日 2021.08.30
(65)同一申请的已公布的文献号
申请公布号 CN 113687788 A
(43)申请公布日 2021.11.23
(73)专利权人 超越科技股份有限公司
地址 250104 山东省济南市高新区孙村镇科航路2877号
(72)发明人 刘传刚 梁记斌 夏伟强
(74)专利代理机构 北京连和连知识产权代理有限公司 11278
专利代理师 刘小峰 张涛
(51)Int.Cl.
G06F 3/06(2006.01)
G06F 9/50(2006.01)

(56)对比文件
CN 104866244 A,2015.08.26

审查员 刘梦影

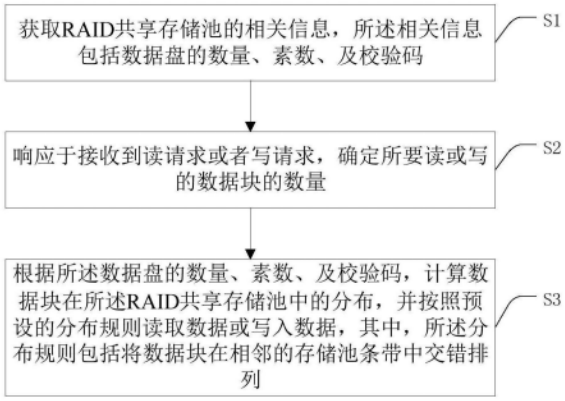
权利要求书2页 说明书6页 附图3页

(54)发明名称

国产平台的数据读写优化方法、装置、计算机及存储介质

(57)摘要

本发明提出了一种适用于国产平台的数据读写优化方法、装置、计算机及存储介质;其中,方法包括:获取RAID共享存储池的相关信息,所述相关信息包括数据盘的数量、素数、及校验码;响应于接收到读请求或者写请求,确定所要读或写的数据块的数量;根据所述数据盘的数量、素数、及校验码,计算数据块在所述RAID共享存储池中的分布,并按照预设的分布规则读取数据或写入数据,其中,所述分布规则包括将数据块在相邻的存储池条带中交错排列。本发明通过重新排布数据块在存储池条带中的分布,让数据块之间共享部分校验块,从而减少校验块的修改,削减操作数目,提升整体的读写性能。



CN 113687788 B

1.一种适用于国产平台的数据读写优化方法,其特征在于,所述方法包括:

获取RAID共享存储池的相关信息,所述相关信息包括数据盘的数量、素数、及校验码;

响应于接收到读请求或者写请求,确定所要读或写的数据块的数量;

根据所述数据盘的数量、素数、及校验码,计算数据块在所述RAID共享存储池中的分布,并按照预设的分布规则读取数据或写入数据,其中,所述分布规则包括将数据块在相邻的存储池条带中交错排列以形成连续的W形分布。

2.如权利要求1所述的适用于国产平台的数据读写优化方法,其特征在于,所述方法还包括:

响应于接收的到读请求或者写请求所要读或写的数据量小于一个数据块的数据量,将所要读或写的数据量小于一个数据块的数据量的读请求或者写请求与其它的读请求或写请求相聚合,并重新确定聚合后的读请求或写请求所要读或写的数据块的数量,直至完成对数据的读取或写入操作。

3.如权利要求1所述的适用于国产平台的数据读写优化方法,其特征在于,所述读请求或者写请求由多个CPU核同步执行,相应的,所述方法还包括:

修改读写处理协议、RAID算法和电子盘读写调度的接口部分,将异步调用转换为同步调用。

4.如权利要求3所述的适用于国产平台的数据读写优化方法,其特征在于,所述方法还包括:

将不同的数据I/O分散到不同的CPU核心和对应的内存。

5.一种适用于国产平台的数据读写优化装置,其特征在于,包括:

初始化模块,配置用于获取RAID共享存储池的相关信息,所述相关信息包括数据盘的数量、素数、及校验码;

选择模块:配置用于响应于接收到读请求或者写请求,确定所要读或写的数据块的数量;

数据块重组及读写模块:配置用于根据所述数据盘的数量、素数、及校验码,计算数据块在所述RAID共享存储池中的分布,并按照预设的分布规则读取数据或写入数据,其中,所述分布规则包括将数据块在相邻的存储池条带中交错排列以形成连续的W形分布。

6.如权利要求5所述的一种适用于国产平台的数据读写优化装置,其特征在于,还包括:

聚合模块,配置用于响应于接收的到读请求或者写请求所要读或写的数据量小于一个数据块的数据量,将所要读或写的数据量小于一个数据块的数据量的读请求或者写请求与其它的读请求或写请求相聚合,并重新确定聚合后的读请求或写请求所要读或写的数据块的数量,直至完成对数据的读取或写入操作。

7.如权利要求5所述的一种适用于国产平台的数据读写优化装置,其特征在于,所述数据块重组及读写模块还配置用于:在进行多个CPU核的同步读或者写请求时,修改存储数据读写处理协议、RAID算法和电子盘读写调度的接口部分,将异步调用转换为同步调用。

8.如权利要求7所述的一种适用于国产平台的数据读写优化装置,其特征在于,还包括:

数据I/O分配模块,配置用于将不同的数据I/O分散到不同的CPU核心和对应的内存。

9.一种计算机,其特征在于,所述计算机包括如权利要求5-8任意一项所述的一种适用于国产平台的数据读写优化装置。

10.一种存储介质,其特征在于,所述存储介质中存储有可运行的计算机程序,所述计算机程序被执行时用于实现如权利要求1-4任意一项所述的一种适用于国产平台的数据读写优化方法的步骤。

国产平台的数据读写优化方法、装置、计算机及存储介质

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种适用于国产平台的数据读写优化方法、装置、计算机及存储介质。

背景技术

[0002] 在当前的国际形势下,自主可控已上升为国家战略,不论是党政军还是民用关键应用场景,推进信息系统的国产自主可控替代已成为共识。自主可控技术的发展对国家信息安全建设起到了核心支撑作用。数据安全是信息安全的核心组成部分,是信息产业体系安全可控的保障。存储作为数据的最终载体,对于信息安全来说是一道必须守住的“底线”。一方面,由于存储系统有层层技术壁垒,核心软硬件具有较高的技术门槛;另外一方面,由于历史原因,存储系统的核心软硬件技术被少数几家国际存储巨头掌控,且存储相关的标准制定也受制于国外。

[0003] FT2000+处理器集成64个飞腾自主研发的高效能处理器内核FTC662,采用乱序四发射超标量流水线,芯片采用偏上并行系统(PSoCb)体系结构,集成高效处理器核心、基于数据亲和的大规模一致性存储结构、层次二维Mesh互联网络,优化存储访问延时,提供业界领先的计算性能、访问带宽和IO扩展能力。该芯片兼容64位ARMv8指令集,适用于高性能、高吞吐率的服务器、存储等领域。

[0004] 因此,为适应上述FT2000+处理器集成64个飞腾自主研发的高效能处理器内核FTC662,采用乱序四发射超标量流水线,亟需一种与之能够匹配的数据读写方法。

发明内容

[0005] 本发明旨在提出一种适合国产服务器多核心处理方式的数据读写方法。并将由每个核心串行的处理读写请求的方式改为由多个核心并行处理读写请求的方式,以弥补国产核心算力不足的问题。为了实现多核心的并行数据读写,需要建立共享存储池RAID,并且重新规划数据块在共享存储池中新的分布以保证多核心并行读写的处理速度。在本发明的一个方面,提出了一种适用于国产平台的数据读写优化方法,所述方法包括:获取RAID共享存储池的相关信息,所述相关信息包括数据盘的数量、素数、及校验码;响应于接收到读请求或者写请求,确定所要读或写的数据块的数量;根据所述数据盘的数量、素数、及校验码,计算数据块在所述RAID共享存储池中的分布,并按照预设的分布规则读取数据或写入数据,其中,所述分布规则包括将数据块在相邻的存储池条带中交错排列。

[0006] 在一个或多个实施例中,所述方法还包括:响应于接收的到读请求或者写请求所要读或写的数据量小于一个数据块的数据量,将所要读或写的数据量小于一个数据块的数据量的读请求或者写请求与其它的读请求或写请求相聚合,并重新确定聚合后的读请求或写请求所要读或写的数据块的数量,直至完成对数据的读取或写入操作。

[0007] 在一个或多个实施例中,所述读或者写请求由多个CPU核同步执行,相应的,所述方法还包括:修改读写处理协议、RAID算法和电子盘读写调度的接口部分,将异步调用转换

为同步调用。

[0008] 在一个或多个实施例中,所述方法还包括:将不同的数据I/O分散到不同的CPU核心和对应的内存。

[0009] 在本发明的另一个方面,提出了一种适用于国产平台的数据读写优化装置,包括:初始化模块,配置用于获取RAID共享存储池的相关信息,所述相关信息包括数据盘的数量、素数、及校验码;选择模块:配置用于响应于接收到读请求或者写请求,确定所要读或写的数据块的数量;数据块重组及读写模块:配置用于根据所述数据盘的数量、素数、及校验码,计算数据块在所述RAID共享存储池中的分布,并按照预设的分布规则读取数据或写入数据,其中,所述分布规则包括将数据块在相邻的存储池条带中交错排列。

[0010] 在一个或多个实施例中,所述数据读写优化装置还包括:聚合模块,响应于接收的到读请求或者写请求所要读或写的数据量小于一个数据块的数据量,将所要读或写的数据量小于一个数据块的数据量的读请求或者写请求与其它的读请求或写请求相聚合,并重新确定聚合后的读请求或写请求所要读或写的数据块的数量,直至完成对数据的读取或写入操作。

[0011] 在一个或多个实施例中,所述数据块重组及读写模块还配置用于:在进行多个CPU核的同步读或者写请求时,修改存储数据读写处理协议、RAID算法和电子盘读写调度的接口部分,将异步调用转换为同步调用。

[0012] 在一个或多个实施例中,所述数据读写优化装置还包括:数据I/O分配模块,配置用于将不同的数据I/O分散到不同的CPU核心和对应的内存。

[0013] 在本发明的另一个方面,提出了一种计算机,所述计算机包括如上述任意一项实施例所提出的一种数据读写优化装置。

[0014] 在本发明的另一个方面,提出了一种存储介质,所述存储介质中存储有可运行的计算机程序,所述计算机程序被执行时用于实现如上述任意一项实施例所提出的一种数据读写优化方法的步骤。

[0015] 本发明的有益效果包括:本发明提出的一种数据读写优化方法、装置、计算机及存储介质,通过重组读写双控存储共享存储池时数据块的分布,在各个数据块之间共享数据校验块,从而提高存储池整体的读写性能,该方法可有效的减少所需修改的数据校验块的数量 ,从而从整体上提高多核系统的并行数据读写性能。

附图说明

[0016] 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实施例。

[0017] 图1为本发明的一种适用于国产平台的数据读写优化方法的工作流程图;

[0018] 图2为一般数据块在存储池条带中的分布示意图;

[0019] 图3为本发明的数据块在存储池条带中的分布示意图;

[0020] 图4为本发明的第一实施例的一种适用于国产平台的数据读写优化装置的结构示意图;

[0021] 图5为本发明的第二实施例的一种适用于国产平台的数据读写优化装置的结构示意图;

[0022] 图6为本发明实施例的基于国产平台的双控存储硬件架构示意图。

具体实施方式

[0023] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明实施例进一步详细说明。

[0024] 需要说明的是,本发明实施例中所有使用“第一”和“第二”的表述均是为了区分两个相同名称非相同的实体或者非相同的参量,可见“第一”“第二”仅为了表述的方便,不应理解为对本发明实施例的限定,后续实施例对此不再一一说明。

[0025] 本发明旨在提出一种适合国产服务器多核心处理方式的数据读写方法。并将由每个核心串行的处理读写请求的方式改为由多个核心并行处理读写请求的方式,以弥补国产核心算力不足的问题。为了实现多核心的并行数据读写,需要建立共享存储池RAID,并且重新规划数据块在共享存储池中新的分布以保证多核心并行读写的处理速度。本发明的方案包括:

[0026] 图1为本发明的一种适用于国产平台的数据读写优化方法的工作流程图。如图1所示,本发明的数据读写优化方法的工作流程包括:步骤S1、获取RAID共享存储池的相关信息,相关信息包括数据盘的数量、素数、及校验码;步骤S2、响应于接收到读请求或者写请求,确定所要读或写的数据块的数量;步骤S3、根据数据盘的数量、素数、及校验码,计算数据块在RAID共享存储池中的分布,并按照预设的分布规则读取数据或写入数据,其中,分布规则包括将数据块在相邻的存储池条带中交错排列。

[0027] 在现有技术中,数据块在存储池条带中的分布如图2所示,图2为一般数据块在存储池条带中的分布示意图。如图2所示,按照传统的EVENODD数据编码方式向存储池条带中写入连续5个数据块A、B、C、D、E,行校验码为$a_{0,7}$,列校验码为$a_{0,8}$、$a_{1,8}$、$a_{2,8}$、$a_{3,8}$、$a_{4,8}$,斜校验码为$a_{0,9}$、$a_{6,9}$、$a_{5,9}$、$a_{4,9}$、$a_{3,9}$,因为Adjuster位$a_{6,9}$被修改,所以$a_{1,9}$和$a_{2,9}$也需要修改,总共需要更新1个行校验码、5个列校验码和7个斜校验码,共计13个校验码。

[0028] 而按照本发明上述实施例中提出的分布规则,数据块在存储池条带中的分布如图3所示,图3为本发明的数据块在存储池条带中的分布示意图。为了减少数据写入时所修改的校验码,本发明将连续写入的数据块以在相邻两条存储池条带中交错排列成连续的“W”形,从而使数据块与数据块之间可以共享部分斜向校验码,减少校验码的修改量。按照本发明的分布规则写入连续5个数据块A、B、C、D、E,行校验码为$a_{0,7}$、$a_{1,7}$,列校验码为$a_{0,8}$、$a_{1,8}$、$a_{2,8}$、$a_{3,8}$、$a_{4,8}$,斜校验码为$a_{0,9}$、$a_{2,9}$、$a_{5,9}$,需更新2个行校验码、5个列校验码和3个斜向校验码,共计10个校验码,相比于传统EVENODD方法的13个校验码,可以少更新3个校验码,从而提高存储池的读写性能。

[0029] 因此,本实施例通过重新排布数据块在存储池条带中的分布,让数据块之间共享部分校验块,从而减少校验块的修改,削减操作数目,提升整体的读写性能。

[0030] 在进一步的实施中,本发明的方法还包括:响应于接收的到读请求或者写请求所要读或写的数据量小于一个数据块的数据量,将所要读或写的数据量小于一个数据块的数

据量的读请求或者写请求与其它的读请求或写请求相聚合,并重新确定聚合后的读请求或写请求所要读或写的数据块的数量,直至完成对数据的读取或写入操作。

[0031] 本实施例能够进一步使得保证数据读写的效率,进而实现对读写过程的进一步优化;并且,本实施例方法对于多核系统,如FT2000+多核心,能够充分发挥出其并发能力(并发的进行读请求或写请求)。

[0032] 在进一步的实施例中,读请求或者写请求由多个CPU核(双控)同步执行,相应的,本发明的方法还包括:修改存储数据读写处理协议、RAID算法和电子盘读写调度的接口部分,将异步调用转换为同步调用。

[0033] 本实施例通过修改存储数据读写处理协议、RAID算法和电子盘读写调度的接口部分异步调用转换为同步调用,减少读写时因异步调用导致的读写线程切换延时,有助于提升多核并行读写速度。

[0034] 在进一步的实施例中,本发明的方法还包括:将不同的数据I/O分散到不同的CPU核心和对应的内存。

[0035] 本实施例的目的是充分利用FT2000+多核心的优势,让不同的数据IO分散到不同的CPU核心和对应的内存,减少CPU核心之间的耦合 ,提高CPU多核心的使用效率。

[0036] 在进一步的实施例中,本发明的方法还包括:使得所述新的分布中的多个数据块共享数据校验块。

[0037] 本实施例通过结合重组读写双控存储共享存储池时数据块的分布的实施例,能够进一步的优化RAID算法,使得各个数据块之间共享数据校验块,以进一步的发挥FT2000+多核心的优势,从而提高双控存储整体的读写性能。

[0038] 在上述各实施例的基础上,本发明还提出了一种数据读写优化装置,如图4所示,图4为本发明的第一实施例的一种适用于国产平台的数据读写优化装置的结构示意图。本发明的数据读写优化装置包括:初始化模块10,配置用于获取RAID共享存储池的相关信息,其中,相关信息包括数据盘的数量、素数、及校验码;选择模块20,配置用于响应于接收到读请求或者写请求,确定所要读或写的数据块的数量;数据块重组及读写模块30,配置用于根据数据盘的数量、素数、及校验码,计算数据块在RAID共享存储池中的分布,并按照预设的分布规则读取数据或写入数据,其中,分布规则包括将数据块在相邻的存储池条带中交错排列。

[0039] 在进一步的实施例中,本发明的数据读写优化装置还包括:聚合模块40,配置用于响应于接收的到读请求或者写请求所要读或写的数据量小于一个数据块的数据量,将所要读或写的数据量小于一个数据块的数据量的读请求或者写请求与其它的读请求或写请求相聚合,并将聚合后的读请求或写请求返回给选择模块20,以通过选择模块20重新确定聚合后的读请求或写请求所要读或写的数据块的数量,直至完成对数据的读取或写入操作。本实施例的数据读写优化装置的结构如图5所示。图5为本发明的第二实施例的一种适用于国产平台的数据读写优化装置的结构示意图。

[0040] 在进一步的实施例中,数据块重组及读写模块还配置用于:在进行多个CPU核的同步读或者写请求时,修改存储数据读写处理协议、RAID算法和电子盘读写调度的接口部分,将异步调用转换为同步调用。

[0041] 在进一步的实施例中,本发明的装置还包括数据I/O分配模块,配置用于将不同的

数据I/O分散到不同的CPU核心和对应的内存。

[0042] 在进一步的实施例中,所述数据块重组及读写模块还配置用于,使得所述新的分布中的多个数据块共享数据校验块。

[0043] 在上述实施例的基础上,本发明还提出了一种计算机,该计算机包括如上述任意一实施例中所提及的一种数据读写优化装置。

[0044] 在一个具体实施例中,所述计算机为FT2000+双控平台,其能够应用本发明上述任一实施例中的一种数据读写优化方法或装置实现快速并行的数据读写操作。

[0045] 图6为本发明实施例的基于国产平台的双控存储硬件架构示意图。如图6所示FT2000+双控存储主要由机械结构层、电路硬件层、支撑软件层和集成应用层组成。

[0046] 1)机械结构层

[0047] 机械结构层包括机箱、散热模组和物理加固件等,双控存储机箱采用19英寸标准上架结构,物理加固采用抗振动冲击、热设计、电磁兼容防护、三防等多种加固技术,实现了高环境适应性和高可靠性。

[0048] 2)电路硬件层

[0049] 电路硬件层主要包括主板模块、电源模块、阵列模块和硬盘模块组成。主板基于国产FT2000+处理器,提供千兆和万兆网络接口,可提供高性能、高可靠的信息处理服务;电源具有双冗余功能,提升可靠性。对外最多可支持24路可拔插硬盘,满足大容量数据存储需求。

[0050] 3)支撑软件层

[0051] 支撑软件层包括底层BIOS、驱动、内核、Linux文件系统和存储管理软件和系统管理软件。支撑层对应用层提供基本服务,包括故障检测、信息同步、故障接管及故障恢复功能。

[0052] 4)集成应用层

[0053] 集成应用层主要提供基本存储管理功能,包括SMI-S统一存储管理软件和Web服务软件,主要通过存储系统故障链路分析并制定相关策略实现存储管理。

[0054] 基于上述实施例,办发明还提出了一种存储介质,该存储介质中存储有可运行的计算机程序,所述计算机程序被执行时用于实现如任一实施例所提及的一种数据读写优化方法的步骤。

[0055] 以上是本发明公开的示例性实施例,但是应当注意,在不背离权利要求限定的本发明实施例公开的范围的前提下,可以进行多种改变和修改。根据这里描述的公开实施例的方法权利要求的功能、步骤和/或动作不需以任何特定顺序执行。此外,尽管本发明实施例公开的元素可以以个体形式描述或要求,但除非明确限制为单数,也可以理解为多个。

[0056] 应当理解的是,在本文中使用的,除非上下文清楚地支持例外情况,单数形式“一个”旨在也包括复数形式。还应当理解的是,在本文中使用的“和/或”是指包括一个或者一个以上相关联地列出的项目的任意和所有可能组合。

[0057] 上述本发明实施例公开实施例序号仅仅为了描述,不代表实施例的优劣。

[0058] 所属领域的普通技术人员应当理解:以上任何实施例的讨论仅为示例性的,并非旨在暗示本发明实施例公开的范围(包括权利要求)被限于这些例子;在本发明实施例的思路下,以上实施例或者不同实施例中的技术特征之间也可以进行组合,并存在如上的本发

明实施例的不同方面的许多其它变化,为了简明它们没有在细节中提供。因此,凡在本发明实施例的精神和原则之内,所做的任何省略、修改、等同替换、改进等,均应包含在本发明实施例的保护范围之内。

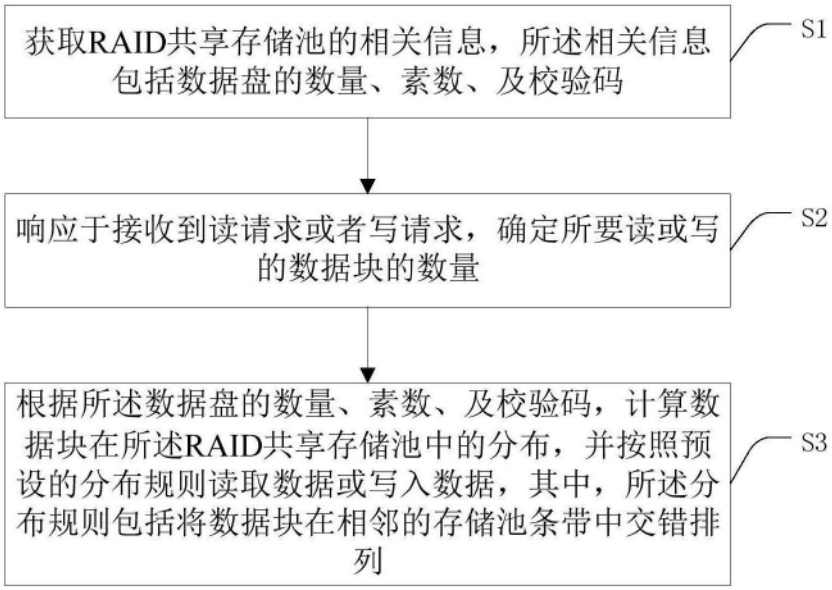

获取RAID共享存储池的相关信息，所述相关信息包括数据盘的数量、素数、及校验码
S1
响应于接收到读请求或者写请求，确定所要读或写的数据块的数量
S2
根据所述数据盘的数量、素数、及校验码，计算数据块在所述RAID共享存储池中的分布，并按照预设的分布规则读取数据或写入数据，其中，所述分布规则包括将数据块在相邻的存储池条带中交错排列
S3

图1

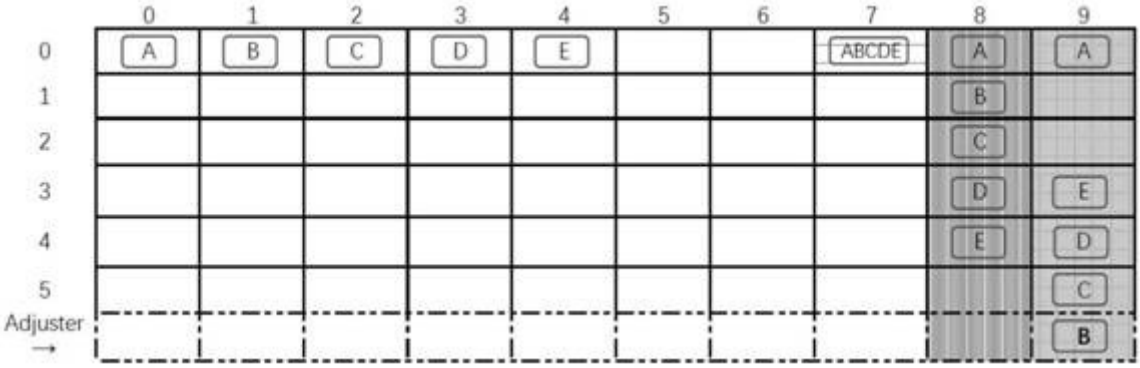

0 1 2 3 4 5 6 7 8 9
0 A B C D E ABCDE A A
1 B
2 C
3 D E
4 E D
5 C
Adjuster
B

图2

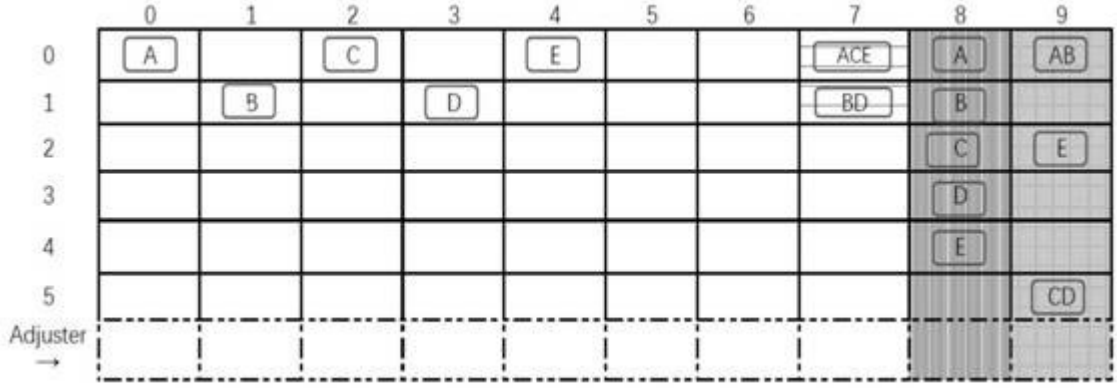

0 1 2 3 4 5 6 7 8 9
0 A C E ACE A AB
1 B D BD B
2 C E
3 D
4 E
5 CD
Adjuster

图3

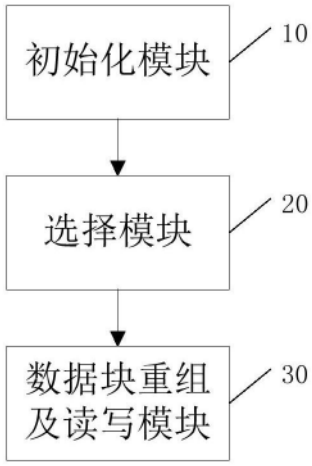

初始化模块
10
选择模块
20
数据块重组
及读写模块
30

图4

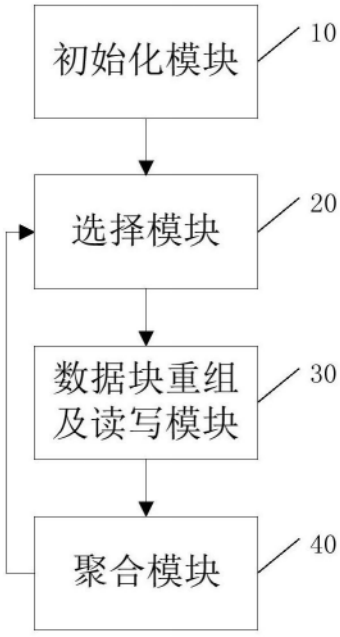

初始化模块
10
选择模块
20
数据块重组
及读写模块
30
聚合模块
40

图5

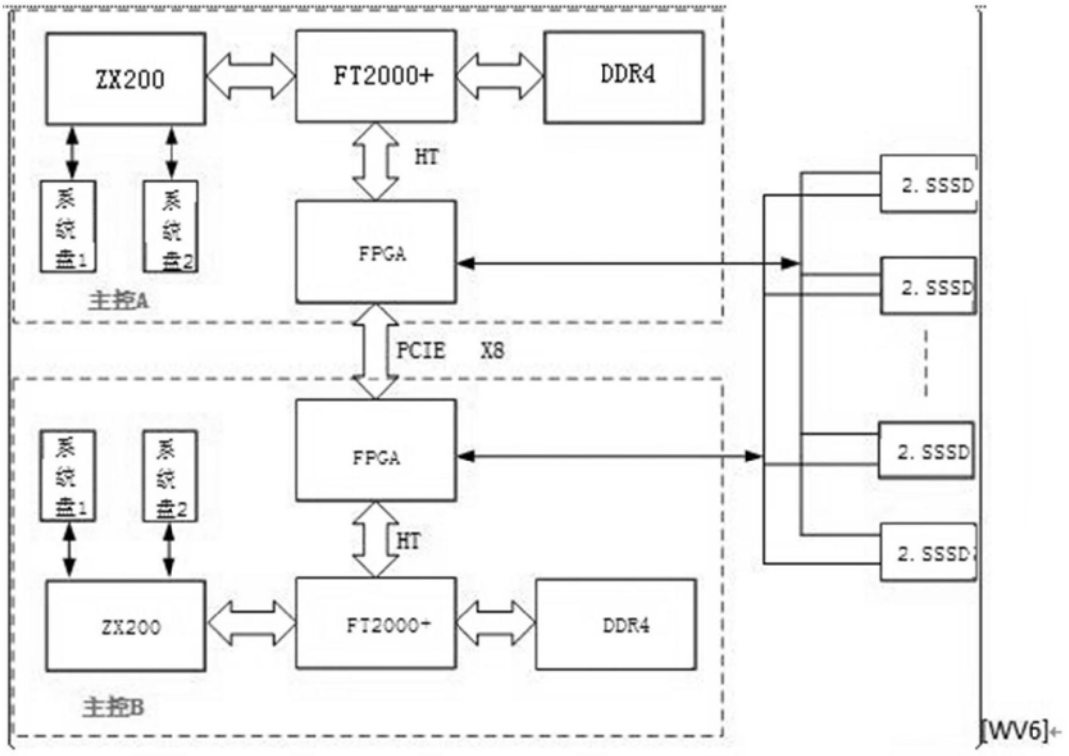

ZX200
FT2000+
DDR4
HT
系统盘1
系统盘2
FPGA
主控A
2.SSSD
2.SSSD
PCIE X8
FPGA
系统盘1
系统盘2
2.SSSD
2.SSSD
HT
ZX200
FT2000+
DDR4
主控B
[WV6]

图6