(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 March 2016 (31.03.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/048314 A1

(51) **International Patent Classification**:
G06F 12/06 (2006.01)

(21) **International Application Number**:
PCT/US2014/057318

(22) **International Filing Date**:
24 September 2014 (24.09.2014)

(25) **Filing Language**: English

(26) **Publication Language**: English

(71) **Applicant**: **HEWLETT PACKARD ENTERPRISE DEVELOPMENT LP** [US/US]; 11445 Compaq Center Drive West, Houston, TX 77070 (US).

(72) **Inventors**: **MILLS, Mark**; 3404 E Harmony Rd, Fort Collins, Colorado 50528 (US). **M, Kishore Kumar**; Sy.No.192, Whitefield Road, Mahadevapura Post, Karnataka, Bangalore 560048 (IN). **SRIVILLIPUTTUR MANNARSWAMY, Sandya**; Sy.No.192, Whitefield Road, Mahadevapura Post, Karnataka, Bangalore 560048 (IN).

(74) **Agents**: **SU, Benjamin** et al.; Hewlett Packard Enterprise, 3404 E. Harmony Road, Mail Stop 79, Fort Collins, CO 80528 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: BLOCK PRIORITY INFORMATION

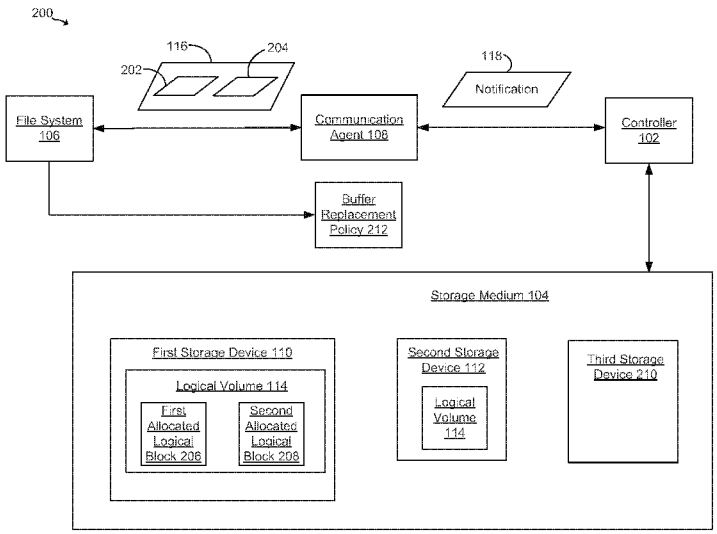

FIG. 2

(57) **Abstract**: Example implementations relate to a storage system. An example implementation includes a storage system including a controller coupled to a storage medium. The controller is to receive block priority information from a file system via a communication agent. The block priority information and the file system are associated with the storage medium. Based on the block priority information, the controller is to selectively perform an operation on the storage medium. The operation includes a reconstruction operation, a data scrubbing operation, and a data migration operation.

WO 2016/048314 A1

BLOCK PRIORITY INFORMATION

BACKGROUND

[0001] Reliability and data availability are important aspects of a storage system. An example mechanism to increase reliability and data availability of a data storage system may be to configure storage devices of the storage system as a redundant array of inexpensive disks (RAID).

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] Some examples of the present application are described with respect to the following figures:

[0003] FIG. 1 is a block diagram of a storage system including a controller to exchange information with a file system, according to an example;

[0004] FIG. 2 is a block diagram of a storage system including a controller to exchange information with a file system, according to an example;

[0005] FIG. 3 is a block diagram of a controller of a storage system, according to an example;

[0006] FIG. 4 is a block diagram of a controller of a storage system, according to an example;

[0007] FIG. 5 is a block diagram of a controller of a storage system, according to an example;

[0008] FIG. 6 is a flowchart illustrating a method of a controller performing a reconstruction operation, according to an example;

[0009] FIG. 7 is a flowchart of illustrating actions of a file system during a reconstruction operation, according to an example;

[0010] FIG. 8 is a flowchart illustrating a method of a controller performing a data scrubbing operation, according to an example;

[0011] FIG. 9 is a flowchart illustrating a method of a file system generating block priority information, according to an example;

[0012] FIG. 10 is a flowchart illustrating a method of a controller performing a data migration operation, according to an example; and

[0013] FIG. 11 is a flowchart illustrating a method of a file system performing a data migration operation, according to an example.

DETAILED DESCRIPTION

[0014] A storage system that implements redundant array of inexpensive disks (RAID) may recover from a storage device, such as a disk drive, failure by reconstructing at least one logical volume of the failed storage device in another storage device. During a reconstruction operation, requests may be sent to the functioning storage devices of a RAID to supply content of the logical volume for the reconstruction operation. Thus, input/output (I/O) loads of the functioning storage devices may be increased. An increased I/O load may increase a likelihood of failure of a storage device.

[0015] Examples described herein provide a storage system that includes a controller to exchange information with a file system. For example, a controller may transmit a reconstruction notification to a file system via a communication agent. The controller may receive block priority information from a file system via the communication agent. The controller may selectively perform a reconstruction operation on a logical volume of a first storage device in a second storage device based on the block priority information.

[0016] As another example, the controller may determine a first scrubbing frequency and a second scrubbing frequency based on the block priority information, where the second scrubbing frequency is lower than the first scrubbing frequency. The controller may perform a first data scrubbing operation on a first logical block associated with a first block number according to the first scrubbing frequency and perform a second data scrubbing operation on a second logical block associated with a second block number according to the second scrubbing frequency.

[0017] As another example, the controller may monitor a reliability parameter of a first storage device of a storage medium. When the reliability parameter exceeds a threshold, the controller may selectively transfer content of the first storage device to a second storage device based on the block priority information. As another example, the controller may transmit the reliability parameter to a file system. The file system may selectively transfer content of the first storage device to a second storage device based on the

reliability parameter. In this manner, examples described herein may increase an efficiency of an operation performed by a controller and/or a file system, such as a reconstruction operation, a data scrubbing operation, and a data migration operation.

[0018] Referring now to the figures, FIG. 1 is a block diagram of a storage system 100 including a controller to exchange information with a file system, according to an example. Storage system 100 may include a controller 102, a storage medium 104, and a file system 106. Controller 102 may be a circuit, an electronic device, a computing device suitable to direct an operation of a RAID. Storage medium 104 may be a storage device or a storage array that implements a RAID. In some examples, storage medium 104 may include a plurality of storage devices, such as disk drives, implementing a RAID with block-level stripping with distributed parity (RAID 5). File system 106 may be an application implemented using processor executable instructions that control how data is stored and retrieved. In some examples, file system 106 may be located in storage medium 104.

[0019] Storage medium 104 may include a plurality of storage devices, such as first storage device 110 and a second storage device 112. In some examples, storage devices 110-112 may be hard disk drives implementing a RAID. First storage device 110 may include a logical volume 114 associated with file system 106. For example, file system 106 may perform block allocation for logical volume 114.

[0020] File system 106 and controller 102 may exchange information via a communication agent 108. Communication agent 108 a set of processor executable instructions suitable to receive information from and transmit information to file system 106 and controller 102. In some examples, communication agent 108 may be located in the user space layer.

[0021] File system 106 and controller 102 may exchange information, such as block priority information 116 and/or a notification 118, to increase efficiencies of particular operations. As described in more detail in FIG. 2, controller 102 may perform a reconstruction operation, a data scrubbing operation, and/or a data migration operation based on information received

from file system 106. File system 106 may also perform a data migration operation based on information received from controller 102.

Reconstruction Operation

[0022] Referring to FIG. 2, FIG. 2 is a block diagram of a storage system 200 including controller 102 to exchange information with file system 106, according to an example. In some examples, controller 102 may reconstruct content of a logical volume of a failed disk based on block priority information 116 received from file system 106 via communication agent 108. To illustrate, First storage device 110 may be a failed disk drive and second storage device 112 may be a functional spare disk drive. The failed disk drive and the functional disk drive may be part of a RAID.

[0023] Controller 102 may selectively reconstruct content of logical volume 114 in second storage device 112 via a reconstruction operation. When controller 102 initiates the reconstruction operation, controller 102 may transmit notification 118 to communication agent 108. Communication agent 108 may forward notification 118 to file system 106. Notification 118 may be a reconstruction notification informing file system 106 that logical volume 114 is to be rebuilt. Based on notification 118, file system 106 may generate block priority information 116.

[0024] Block priority information 116 may include allocated logical block information 202 that is indicative of a distinct allocation status for each logical block of logical volume 114 as identified by file system 106. File system 106 may generate allocated logical block information 202 by recording allocation status of logical blocks of logical volume 114. An allocation status for an allocated logical block may be “allocated” and an allocation status for an unallocated logical block may be “unallocated.” Each logical block in allocated logical block information 202 may be identified by a distinct block number or block address.

[0025] In some examples, block priority information 116 may also include prioritized content location information 204. Prioritized content location information 204 may be indicative of a distinct priority associated with each allocated logical block of logical volume 114. File system 106 may generate prioritized content location information 204 by recording a distinct priority

assigned to each allocated logical block of logical volume 114. For example, first content may be stored to a first allocated logical block 206 of logical volume 114 and second content may be stored to a second allocated logical block 208 of logical volume 114. First allocated logical block 206 may be assigned a first priority based on the first content. Second allocated logical block 208 may be assigned a second priority that is lower than the first priority based on the second content. In prioritized content location information 204, first allocated logical block 206 may be identified by a first block number and second allocated logical block 208 may be identified by a second block number.

[0026] File system 106 or an application (i.e., implemented using a set of processor executable instructions) associated with the allocated logical blocks may assign a distinct priority to each allocated logical block based on content. In some examples, the first content may correspond to metadata of file system 106, such as system data. The second content may correspond to non-metadata of file system 106, such as user data. Thus, file system 106 may assign a higher priority to first allocated logical block 206 than second allocated logical block 208 since metadata is more important to the operation of file system 106 than non-metadata. For example, file system 106 may assign a first priority to first allocated logical block 206 and a second priority to second allocated logical block 208 that is lower than the first priority. In some examples, an application may assign a higher priority to an allocated logical block that stores more important data than another allocated logical block that stores less important data.

[0027] Thus, after file system 106 generating allocated logical block information 202 based on the allocation status of the allocated logical block and/or content location information 204 based on the priorities assigned to the allocated logical blocks, file system 106 may transmit block priority information 116 to controller 102 via communication agent 108. Controller 102 may use block priority information 116 to selectively perform the reconstruction operation. For example, controller 102 may perform the reconstruction operation on logical blocks of logical volume 114 having the “allocated” allocation status. Controller 102 may skip performing the reconstruction operation on logical blocks of logical volume 114 having the “unallocated”

allocation status since the logical blocks having the "unallocated" allocation status hold no meaningful content.

[0028] Among the allocated logical blocks having the "allocated" allocation status, controller 102 may perform the reconstruction operation based on the priorities of the allocated logical blocks. For example, controller 102 may perform the reconstruction operation on first allocated logical block 206 having the first priority prior to second allocated logical block 208 since the first priority is higher than the second priority. Thus, when the reconstruction operation is completed, logical volume 114 may be located in second storage device 112.

[0029] By reconstructing allocated blocks but not the unallocated blocks during the reconstruction operation, the amount of blocks to be reconstructed may be reduced. Thus, an amount of time associated with a reconstruction operation may be reduced and an I/O load of storage devices supplying content of the logical volume for the reconstruction operation may also be reduced. Further, by prioritizing the reconstruction operation based on priorities of the allocated logical blocks, the reconstruction operation may be performed in a more intelligent manner.

[0030] In some examples, in response to receiving notification 118, file system 106 may assign an under reconstruction status to logical volume 114. File system 106 may avoid storing new data to logical volume 114 when logical volume 114 has the under reconstruction status. When controller 102 finishes the reconstruction operation, controller 102 may transmit notify file system 106 of the completion of the reconstruction operation via communication agent 108. File system 106 may change the under reconstruction status of logical volume 114 to a normal status. Thus, file system 106 may resume storing data to logical volume 114.

[0031] In some examples, file system 106 may store data to logical volume 114 during the reconstruction operation for a write request initiated by an application, but not other types of I/O requests. In some examples, when logical volume 114 is the single logical volume associated with file system 106, file system 106 may forward a read request or a write request to controller 102, controller 102 may perform the reconstruction operation on a

logical block associated with the read request or the write request irrespective of the allocation status of the logical block.

[0032] In some examples, controller 102 may notify, via communication agent 108, file system 106 storage devices that are involved in the reconstruction operation. For example, controller 102 may notify file system 106 that second storage device 112 is involved in the reconstruction operation. When file system 106 receives an I/O request, such as a write request, targeting second storage device 112, file system 106 may redirect the I/O request to another storage device that is not involved in the reconstruction operation, such as a third storage device 210. File system 106 may also modify a buffer replacement policy 212 to increase a likelihood of second storage device 112 maintaining an association with a buffer pool. Thus, an I/O load of second storage device 112 during the reconstruction operation may be reduced. In some examples, when file system 106 is notified by controller 102 that second storage device 112 is involved in the reconstruction operation, file system 106 may defer maintenance tasks associated with second storage device 112 until the completion of the reconstruction operation. Thus, an I/O load of second storage device 112 during the reconstruction operation may be reduced.

Data Scrubbing Operation

[0033] Still referring to FIG. 2, in some examples, controller 102 may selectively perform a data scrubbing operation based on block priority information 116. A data scrubbing operation may be an operation where an error in content stored in a logical block is identified and fixed. To illustrate, file system 106 may periodically transmit block priority information 116 to controller 102 via communication agent 108. Alternatively, file system 106 may transmit block priority information 116 in response to receiving notification 118. Notification 118 may be a data scrubbing operation notification informing file system 106 that controller 102 is to perform the data scrubbing operation on logical volume 114.

[0034] Based on block priority information 116, controller 102 may scrub the allocated logical blocks of logical volume 114 based on priorities of the allocated logical blocks. For example, based on block priority information

116, controller 102 may determine a first scrubbing frequency associated with first allocated logical block 206 and a second scrubbing frequency associated with second allocated logical block 208. The second scrubbing frequency may be lower than the first scrubbing frequency. Controller 102 may perform a data scrubbing operation on first allocated logical block 206 according to the first scrubbing frequency. Controller 102 may perform a data scrubbing operation on second allocated logical block 208 according to the second scrubbing frequency. Thus, a logical block with a higher priority may be scrubbed more often than a logical block with a lower priority. By performing a data scrubbing operation on allocated logical blocks based on priority, the data scrubbing operation may be performed in a more intelligent manner.

Data Migration Operation

[0035] Still referring to FIG. 2, in some examples, controller 102 may selectively perform a data migration operation content of a storage device, such as first storage device 110 and second storage device 112, based on block priority information 116. To illustrate, storage devices 110-112 may be hard disk drives of a RAID. Controller 102 may monitor a reliability parameter of first storage device 110 and/or second storage device 112. In some examples, the reliability parameter may include a self-monitoring, analysis and reporting technology (SMART) indicator. For example, the SMART indicator may indicate how likely a storage device is about to fail or malfunction. When a value of the SMART indicator is "threshold exceeded," controller 102 may determine that the storage device has a high likelihood of failure. When the value of the SMART indicator is "threshold not exceeded," controller 102 may determine that the storage device has a low likelihood of failure.

[0036] In some examples, the reliability parameter may include an input/output (I/O) request failure counter associated with first storage device 110 and/or second storage device 112. For example, every time a storage device fails to complete an I/O request, a value of the I/O request failure counter associated with the storage device may be increased by 1. Thus, when a reliability parameter of a storage device exceeds a threshold, such as when the value of the SMART indicator is "threshold exceeded" or when the

value of the I/O request failure counter exceeds a particular amount, controller 102 may perform a data migration operation on the storage device.

[0037] As an example, controller 102 may determine that a reliability parameter of first storage device 110 exceeds a threshold. Thus, controller 102 may perform a data migration operation on content of first storage device 110, such as logical volume 114, based on block priority information 116.

[0038] Based on allocated logical block information 202, controller 102 may transfer or migrate logical blocks of logical volume 114 that have the "allocated" allocation status, such as first allocated logical block 206 and second allocated logical block 208, to second storage device 112. When a logical block of logical volume 114 has the "unallocated" allocation status, controller 102 may skip performing the data migration on the logical block.

[0039] In some examples, based on prioritized content location information 204 and based on allocated logical block information 202, controller 102 may selectively transfer the allocated logical blocks. For example, controller 102 may transfer first allocated logical block 206 to second storage device 112 prior to second allocated logical block 208. Thus, by transferring allocated blocks based on priority and not transferring the unallocated blocks during a data migration operation, the data migration operation may be performed in a more intelligent manner.

[0040] In some examples, file system 106 may perform the data scrubbing operation instead of controller 102. For example, when the reliability parameter of first storage device 110 exceeds the threshold, controller 102 may include identification information of first storage device 110 in notification 118 and transmit notification 118 to file system 106 via communication agent 108. Based on the identification information in notification 118 and/or based on block priority information 116, file system 106 may perform the data migration operation on logical volume 114.

[0041] FIG. 3 is a block diagram of a controller 300 of a storage system, according to an example. Controller 300 may implement controller 102 of FIGs. 1-2. Controller 300 may include a processor 302 and a computer-readable storage medium 304.

[0042] Processor 302 may be a central processing unit (CPU), a semiconductor-based microprocessor, and/or other hardware devices suitable

for retrieval and execution of instructions stored in computer-readable storage medium 304. Processor 302 may fetch, decode, and execute instructions 306-310 to control a process of performing a reconstruction operation based on block priority information, such as block priority information 116 of FIGs. 1-2. As an alternative or in addition to retrieving and executing instructions, processor 302 may include at least one electronic circuit that includes electronic components for performing the functionality of instructions encoded in computer-readable storage medium 304, such as instructions 306, 308, 310, or a combination thereof.

[0043] Computer-readable storage medium 304 may be any electronic, magnetic, optical, or other physical storage device that contains or stores executable instructions. Thus, computer-readable storage medium 304 may be, for example, Random Access Memory (RAM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), a storage device, an optical disc, etc. In some examples, computer-readable storage medium 304 may be a non-transitory storage medium, where the term "non-transitory" does not encompass transitory propagating signals. As described in detail below, computer-readable storage medium 304 may be encoded with a series of processor executable instructions, such as instructions 306-310 for performing a reconstruction operation based on block priority information.

[0044] Reconstruction notification transmission instructions 306 may transmit notification 108 from controller 102 to file system 106 via communication agent 108. Block priority reception instructions 308 may receive block priority information 116 from file system 106 via communication agent 108. Reconstruction operation instructions 310 may perform a reconstruction operation on logical volume 114 based on block priority information 116.

[0045] FIG. 4 is a block diagram of a controller 400 of a storage system, according to an example. Controller 400 may implement controller 102 of FIGs. 1-2. Controller 400 may include processor 302 and computer-readable storage medium 304. Computer-readable storage medium 304 may be encoded with a series of instructions 308 and 402-404.

[0046] Scrubbing frequency determination instructions 402 may determine a scrubbing frequency associated with an allocated logical block, such as first

allocated logical block 206, based on a priority of the allocated logical block. Data scrubbing operations 404 may perform a data scrubbing operation on an allocated logical block, such as first allocated logical block 206, based on a scrubbing frequency.

[0047] FIG. 5 is a block diagram of a controller 500 of a storage system, according to an example. Controller 500 may include processor 302 and computer-readable storage medium 304. Computer-readable storage medium 304 may be encoded with a series of instructions 308 and 502-504. Reliability parameter monitoring instructions 502 may monitor a reliability parameter of a storage device, such as first storage device 110. Data migration operation instructions 504 may perform a data migration operation on a storage device having a reliability parameter that exceeds a threshold, such as first storage device 110.

[0048] FIG. 6 is a flowchart illustrating a method 600 of a controller performing a reconstruction operation, according to an example. Method 600 may be implemented using controller 102 of FIGs. 1-2 and controller 300 of FIG. 3. Method 600 includes transmitting a reconstruction notification to a file system via a communication agent, at 602. For example, referring to FIG. 2, controller 102 may transmit notification 118 to file system 106 via communication agent 108.

[0049] Method 600 also includes receiving block priority information from a file system via the communication agent, where the block priority information and the file system are associated with a logical volume of the first storage device, and where the block priority information includes allocated logical block information and prioritized content location information, at 604. For example, referring to FIG. 2, controller 102 may receive block priority information 116 from file system 106 via communication agent 108.

[0050] Method 600 further includes selectively performing a reconstruction operation on a logical volume of the first storage device in the second storage device based on the block priority information, at 606. For example, referring to FIG. 2, controller 102 may perform the reconstruction operation on logical blocks of logical volume 114 having the “allocated” allocation status. Controller 102 may skip performing the reconstruction operation on logical blocks of logical volume 114 having the “unallocated” allocation status since

the logical blocks having the "unallocated" allocation status hold no meaningful content.

[0051] FIG. 7 is a flowchart illustrating actions of a file system during a reconstruction operation, according to an example. The actions may be implemented by file system 106 of FIGs. 1-2. At 702, file system 106 may generate block priority information based on a reconstruction notification. For example, referring to FIG. 2, based on notification 118, file system 106 may generate block priority information 116.

[0052] At 704, file system 106 may transmit the block priority information to a controller via a communication agent. For example, referring to FIG. 2, file system 106 may transmit block priority information 116 to controller 102 via communication agent 108. At 706, file system 106 may redirect an input/output request associated with a particular storage device to another storage device, where the particular storage device is involved in a reconstruction operation. For example referring to FIG. 2, when file system 106 receives an I/O request, such as a write request, targeting second storage device 112, file system 106 may redirect the I/O request to another storage device that is not involved in the reconstruction operation, such as a third storage device 210.

[0053] At 708, file system 106 may defer a maintenance task of the particular storage device during the reconstruction operation. For example, referring to FIG. 2, file system 106 may defer maintenance tasks associated with second storage device 112 until the completion of the reconstruction operation. At 710, file system 106 may modify a buffer replacement policy to increase a likelihood of the particular storage device maintaining an association with a buffer pool. For example, referring to FIG. 2, file system 106 may modify a buffer replacement policy 212 to increase a likelihood of second storage device 112 maintaining an association with a buffer pool.

[0054] FIG. 8 is a flowchart illustrating a method 800 of a controller performing a data scrubbing operation, according to an example. Method 800 may be implemented using controller 102 of FIGs. 1-2 and controller 400 of FIG. 4. Method 800 includes receiving, at a controller associated with a storage medium, block priority information from a file system, where the block priority information and the file system are associated with the storage

medium, where the block priority information includes a first block number having a first priority and a second block number having a second priority, and where the second priority is lower than the first priority, at 802. For example, referring to FIG. 2, controller 102 may receive block priority information 116 from file system 106 via communication agent 108.

[0055] Method 800 also includes determining a first scrubbing frequency and a second scrubbing frequency based on the block priority information, where the second scrubbing frequency is lower than the first scrubbing frequency, at 804. For example, referring to FIG. 2, controller 102 may determine the first scrubbing frequency and the second scrubbing frequency.

[0056] Method 800 further includes performing a first data scrubbing operation on a first logical block associated with the first block number according to the first scrubbing frequency, at 806. For example, referring to FIG. 2, controller 102 may perform a data scrubbing operation on first allocated logical block 206 according to the first scrubbing frequency.

[0057] Method 800 further includes performing a second data scrubbing operation on a second logical block associated with the second block number according to the second scrubbing frequency, at 808. For example, referring to FIG. 2, controller 102 may perform a data scrubbing operation on second allocated logical block 208 according to the second scrubbing frequency.

[0058] FIG. 9 is a flowchart illustrating a method 900 of a file system generating block priority information, according to an example. Method 900 may be implemented using file system 106 of FIGs. 1-2. Method 900 includes assigning a first priority associated with a first allocated logical block having first content, at 902. For example, referring to FIG. 2, file system 106 may assign the first priority to first allocated logical block 206.

[0059] Method 900 also includes assigning a second priority associated with a second allocated logical block having second content different from the first content, at 904. For example, referring to FIG. 2, file system 106 may assign a second priority to second allocated logical block 208 that is lower than the first priority. Method 900 further includes generating block priority information based on the allocated first logical block and based on the allocated second logical block, at 906. For example, file system 106 may

generate prioritized content location information 204 by recording a distinct priority assigned to each allocated logical block of logical volume 114.

[0060] FIG. 10 is a flowchart illustrating a method 1000 of a controller performing a data migration operation, according to an example. Method 1000 may be implemented by controller 102 of FIGs. 1-2 and controller 500 of FIG. 5. Method 1000 includes monitoring a reliability parameter of a first storage device of a storage medium, at 1002. For example, referring to FIG. 2, controller 102 may monitor a reliability parameter of first storage device 110 and/or second storage device 112.

[0061] Method 1000 also includes receiving block priority information from a file system via a communication agent, where the block priority information and the file system are associated with the first storage device, and where the block priority information includes allocated logical block information, at 1004. For example, referring to FIG. 2, controller 102 may receive block priority information 116 from file system 106 via communication agent 108.

[0062] Method 1000 further includes, when the reliability parameter exceeds a threshold, selectively transferring content of the first storage device to a second storage device based on the block priority information, at 1006. For example, referring to FIG. 2, controller 102 may transfer first allocated logical block 206 to second storage device 112 prior to second allocated logical block 208.

[0063] FIG. 11 is a flowchart illustrating a method 1100 of a file system performing a data migration operation, according to an example. Method 1100 may be implemented by file system 106 of FIGs. 1-2. Method 1100 includes receiving a notification from a controller via a communication agent, where the notification identifies a first storage device having a reliability parameter exceeding a threshold, at 1102. For example, referring to FIG. 2, file system 106 may receive notification 118 from controller 102 via communication agent 108. Controller 102 may include identification information of first storage device 110 in notification 118 when the reliability parameter of first storage device 110 exceeds the threshold.

[0064] Method 1100 also includes generating block priority information based on the notification, at 1104. For example, referring to FIG. 2, based on notification 118, file system 106 may generate block priority information 116.

Method 1100 further includes selectively transferring content of the first storage device to a second storage device based on the notification and based on the block priority information, at 1106. For example, based on the identification information in notification 118 and/or based on block priority information 116, file system 106 may perform the data migration operation on logical volume 104.

[0065] According to the foregoing, examples disclosed herein provide a storage system that includes a controller to exchange information with a file system. For example, the controller may transmit a reconstruction notification, information on a storage device that is involved in a reconstruction operation, information on a storage device that is about to fail, etc. to the file system via a communication agent. The file system may transmit block priority information to the controller. Controller may perform a reconstruction operation, a data scrubbing operation, and/or a data migration operation based on the block priority information. The file system may perform a data migration operation based on information from the controller, such as the information on a storage device that is about to fail. Thus, by exchanging information between the controller and the file system, an efficiency of an operation performed by the controller and/or the file system, such as the reconstruction operation, the data scrubbing operation, and the data migration operation, may be increased.

[0066] The use of "comprising", "including" or "having" are synonymous and variations thereof herein are meant to be inclusive or open-ended and do not exclude additional unrecited elements or method steps.

Claims

What is claimed is:

1. A storage system comprising:
a controller; and
a storage medium coupled to the controller, wherein the storage medium includes a first storage device and a second storage device, and wherein the controller is to:
transmit a reconstruction notification to a file system via a communication agent;
receive block priority information from a file system via the communication agent, wherein the block priority information and the file system are associated with a logical volume of the first storage device, and wherein the block priority information includes allocated logical block information and prioritized content location information; and
selectively perform a reconstruction operation on a logical volume of the first storage device in the second storage device based on the block priority information.

2. The storage system of claim 1, wherein the allocated logical block information is indicative of an allocation status of a logical block in the logical volume, and wherein the allocation status includes allocated and unallocated.

3. The storage system of claim 2, wherein when the allocation status is allocated, the controller is to perform the reconstruction operation on the logical block, and wherein when the allocation status is unallocated, the controller is to skip the reconstruction operation on the logical block.

4. The storage system of claim 1, wherein the logical volume includes a first allocated logical block and a second allocated logical block, wherein the first allocated logical block includes first content having a first priority, wherein the second allocated logical block includes second content having a second priority that is lower than the first priority, wherein the prioritized content location information includes a first block number associated with the first allocated logical block and a second block number associated with the second allocated logical block.

5. The storage system of claim 4, wherein the first content corresponds to metadata of the file system, and wherein the second content corresponds to non-metadata of the file system.

6. The storage system of claim 5, wherein the controller is to perform the reconstruction operation on the first allocated logical block prior to the second allocated logical block.

7. The storage system of claim 1, wherein the storage medium further includes a third storage device, wherein the controller is to redirect an input/output (I/O) request associated with the second storage device to the third storage device during the reconstruction operation.

8. The storage system of claim 1, wherein a buffer replacement policy is to be modified to increase a likelihood of the second storage device maintaining an association with a buffer pool during the reconstruction operation.

9. A method comprising:

receiving, at a controller associated with a storage medium, block priority information from a file system, wherein the block priority information and the file system are associated with the storage medium, wherein the block priority information includes a first block number having a first priority and a second block number having a second priority, and wherein the second priority is lower than the first priority;

determining a first scrubbing frequency and a second scrubbing frequency based on the block priority information, wherein the second scrubbing frequency is lower than the first scrubbing frequency;

performing a first data scrubbing operation on a first logical block associated with the first block number according to the first scrubbing frequency; and

performing a second data scrubbing operation on a second logical block associated with the second block number according to the second scrubbing frequency.

10. The method of claim 9, wherein the first priority is determined based on content of the first logical block.

11. The method of claim 10, wherein the content corresponds to metadata of the file system.

12. A computer-readable storage medium comprising instructions that when executed cause a controller of a storage system to:

monitor a reliability parameter of a first storage device of a storage medium;

receive block priority information from a file system via a communication agent, wherein the block priority information and the file system are associated with the first storage device, and wherein the block priority information includes allocated logical block information; and

when the reliability parameter exceeds a threshold, selectively transfer content of the first storage device to a second storage device based on the block priority information.

13. The computer-readable storage medium of claim 12, wherein the reliability parameter includes a self-monitoring, analysis and reporting technology (SMART) indicator and an input/output request failure counter.

14. The computer-readable storage medium of claim 12, wherein the allocated logical block information is indicative of an allocation status of a logical block in the first storage device, and wherein the allocation status includes allocated and unallocated.

15. The computer-readable storage medium of claim 14, wherein the instructions when executed further cause the controller to:

when the allocation status is allocated, transfer content of the logical block to the second storage device; and

when the allocation status is unallocated, skip transferring the content of the logical block to the second storage device.

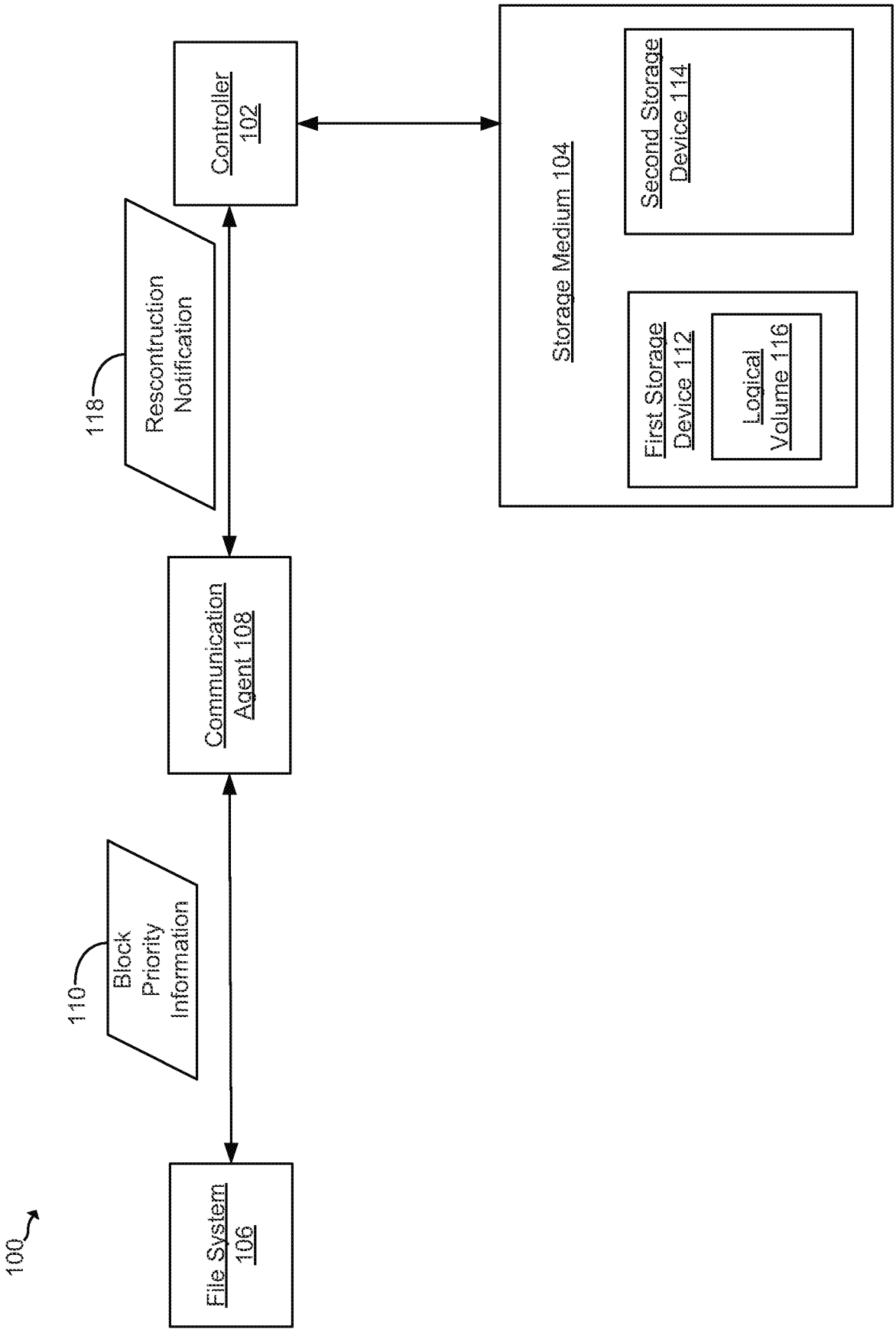
100
File System 106
110
Block Priority Information
Communication Agent 108
118
Reconstruction Notification
Controller 102
Storage Medium 104
First Storage Device 112
Logical Volume 116
Second Storage Device 114

FIG. 1

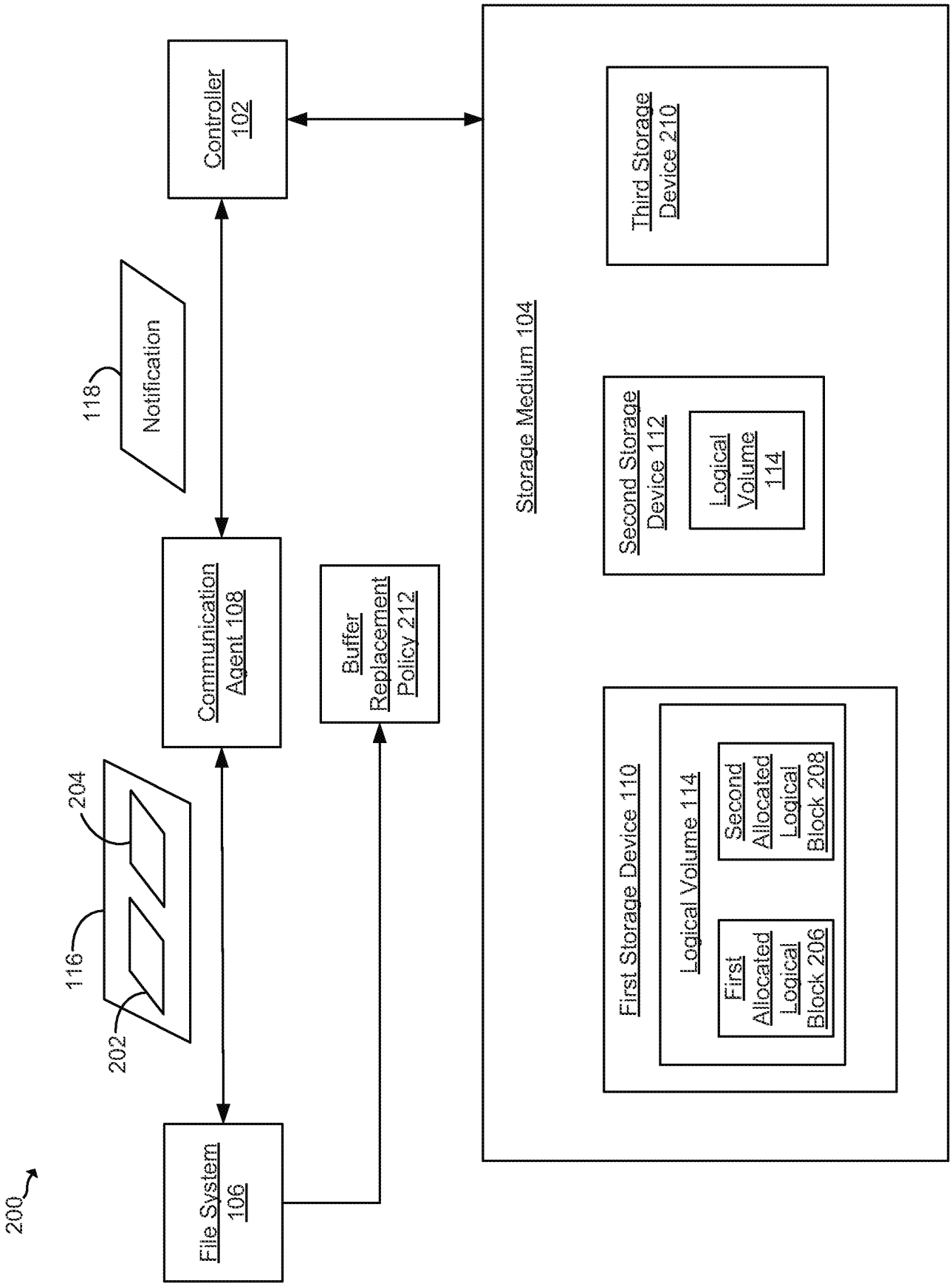
200
File System 106
Communication Agent 108
Controller 102
Buffer Replacement Policy 212
Notification
118
116
202
204
Storage Medium 104
First Storage Device 110
Logical Volume 114
First Allocated Logical Block 206
Second Allocated Logical Block 208
Second Storage Device 112
Logical Volume 114
Third Storage Device 210

FIG. 2

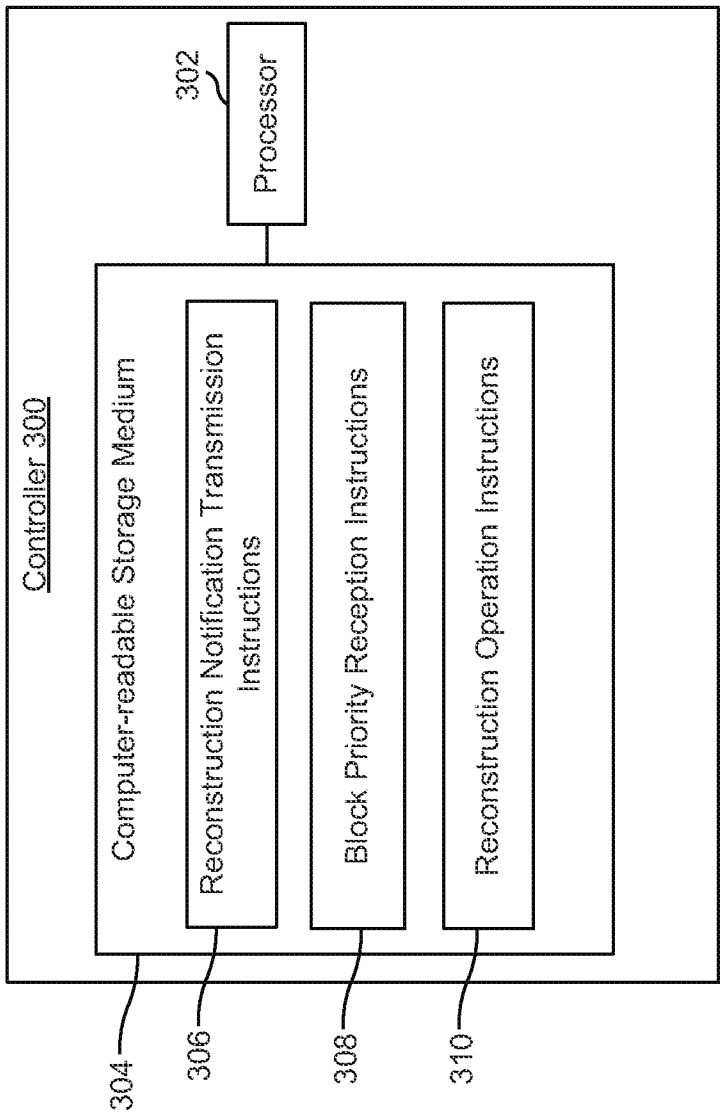
Controller 300
Processor
302
Computer-readable Storage Medium
Reconstruction Notification Transmission Instructions
Block Priority Reception Instructions
Reconstruction Operation Instructions
304
306
308
310

FIG. 3

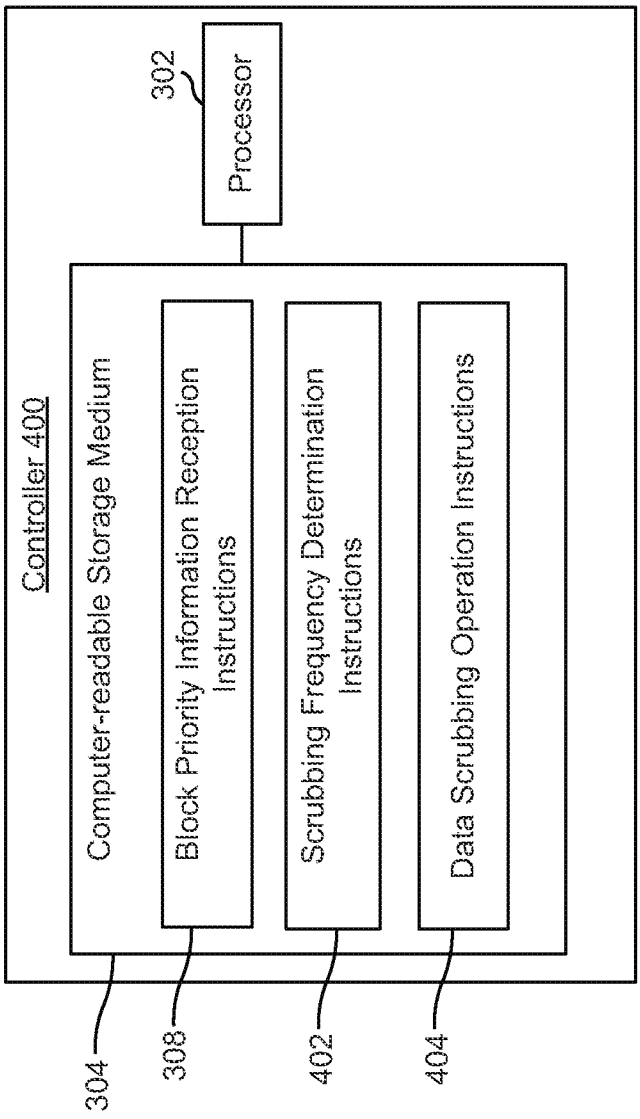
Controller 400
Processor
302
Computer-readable Storage Medium
Block Priority Information Reception Instructions
Scrubbing Frequency Determination Instructions
Data Scrubbing Operation Instructions
304
308
402
404

FIG. 4

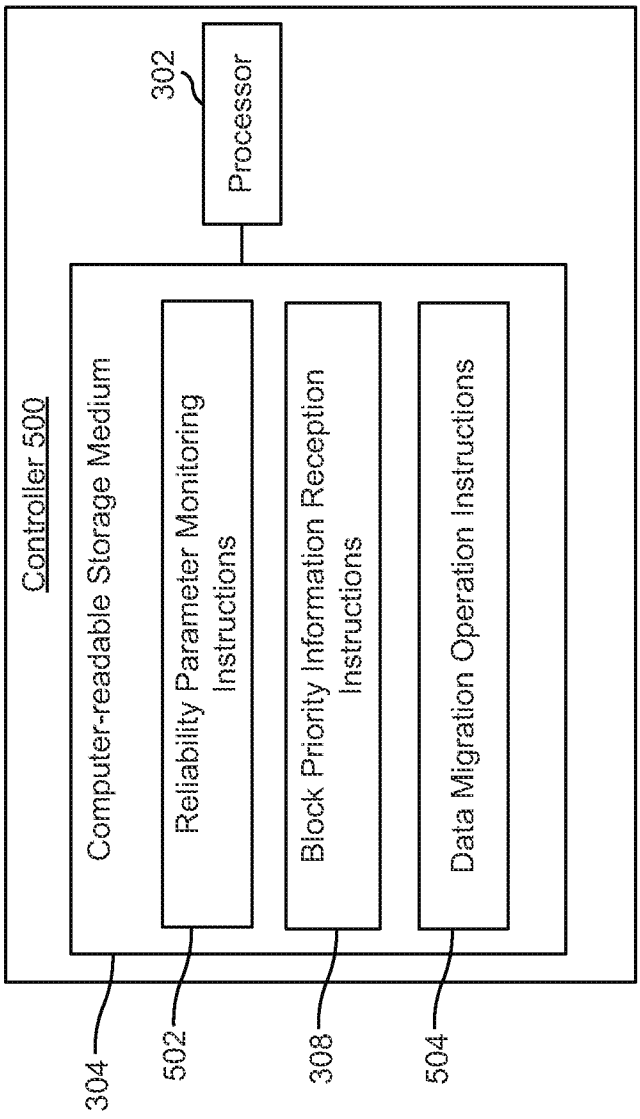
Controller 500
Computer-readable Storage Medium
Processor
302
Reliability Parameter Monitoring Instructions
Block Priority Information Reception Instructions
Data Migration Operation Instructions
304
502
308
504

FIG. 5

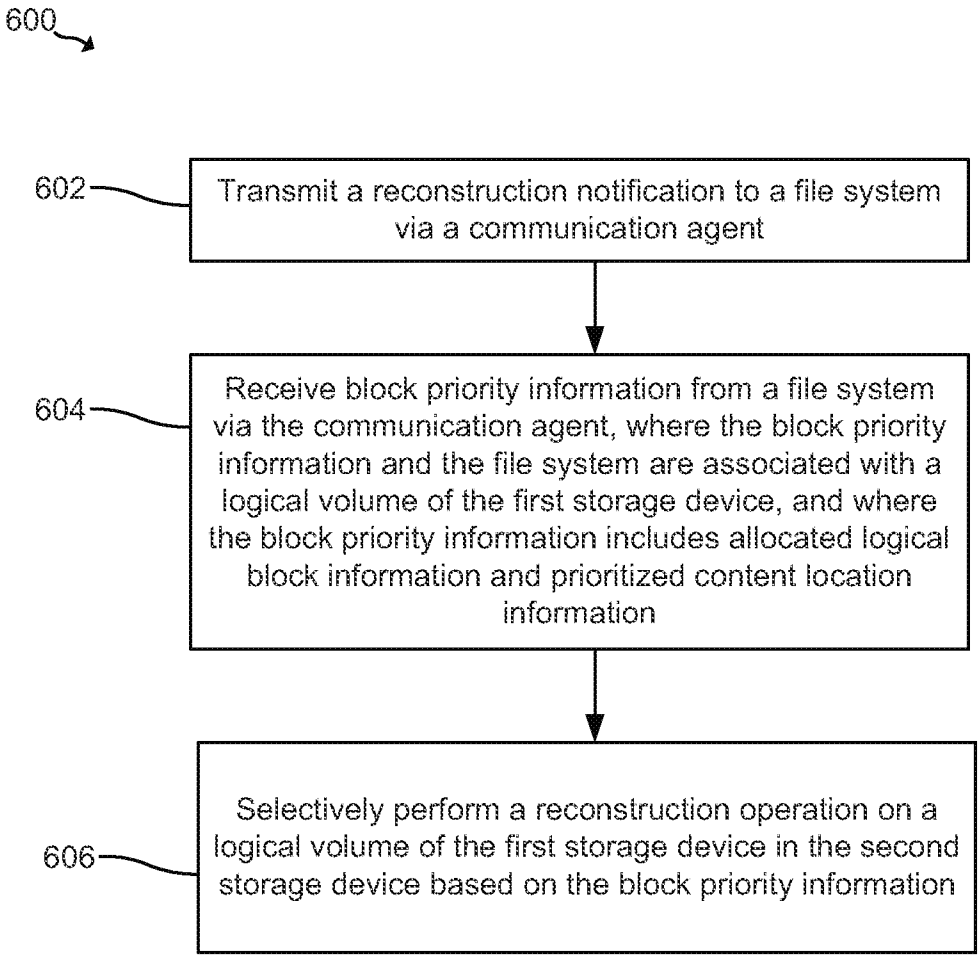
600
602
Transmit a reconstruction notification to a file system via a communication agent
604
Receive block priority information from a file system via the communication agent, where the block priority information and the file system are associated with a logical volume of the first storage device, and where the block priority information includes allocated logical block information and prioritized content location information
606
Selectively perform a reconstruction operation on a logical volume of the first storage device in the second storage device based on the block priority information

FIG. 6

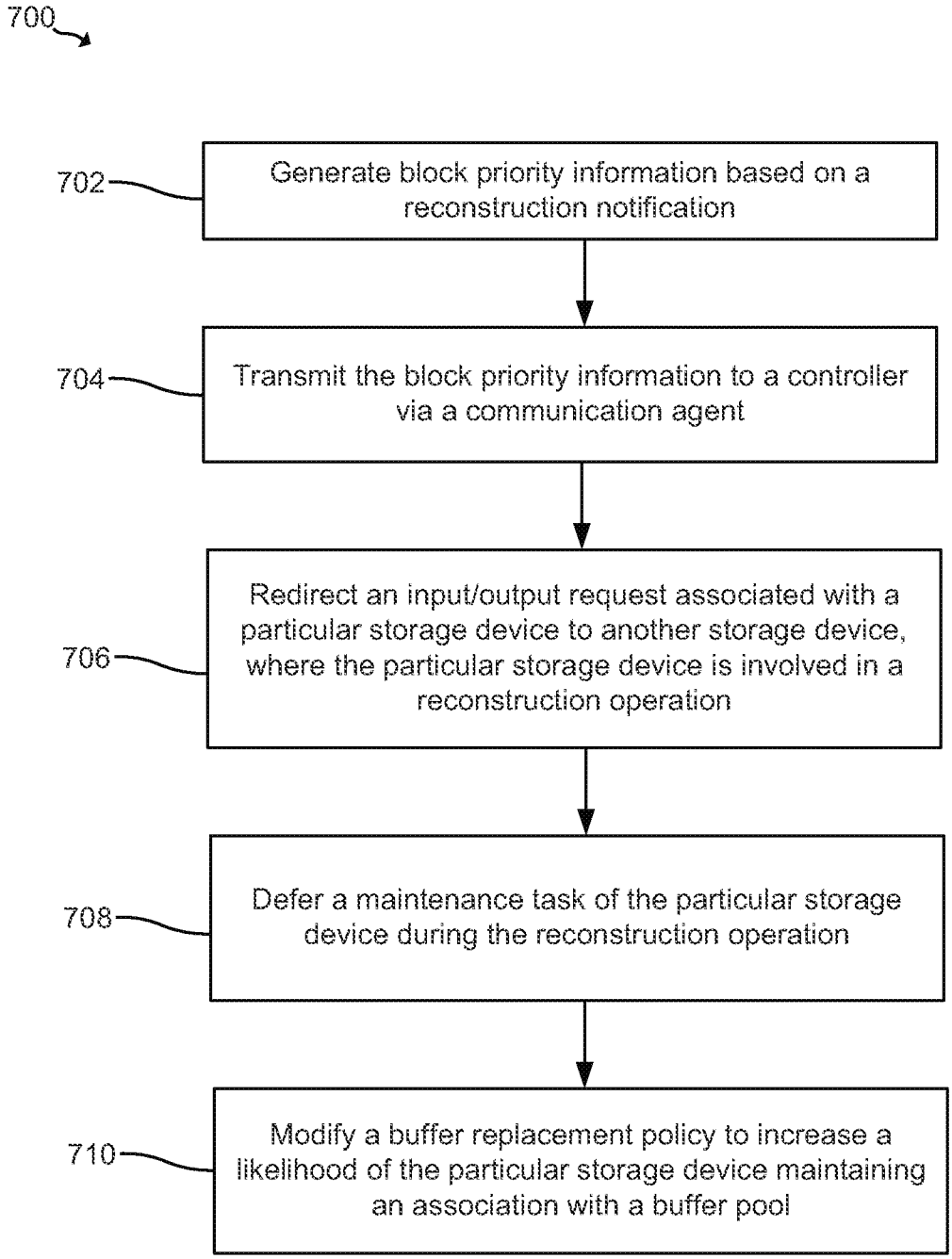
700
702
Generate block priority information based on a reconstruction notification
704
Transmit the block priority information to a controller via a communication agent
706
Redirect an input/output request associated with a particular storage device to another storage device, where the particular storage device is involved in a reconstruction operation
708
Defer a maintenance task of the particular storage device during the reconstruction operation
710
Modify a buffer replacement policy to increase a likelihood of the particular storage device maintaining an association with a buffer pool

FIG. 7

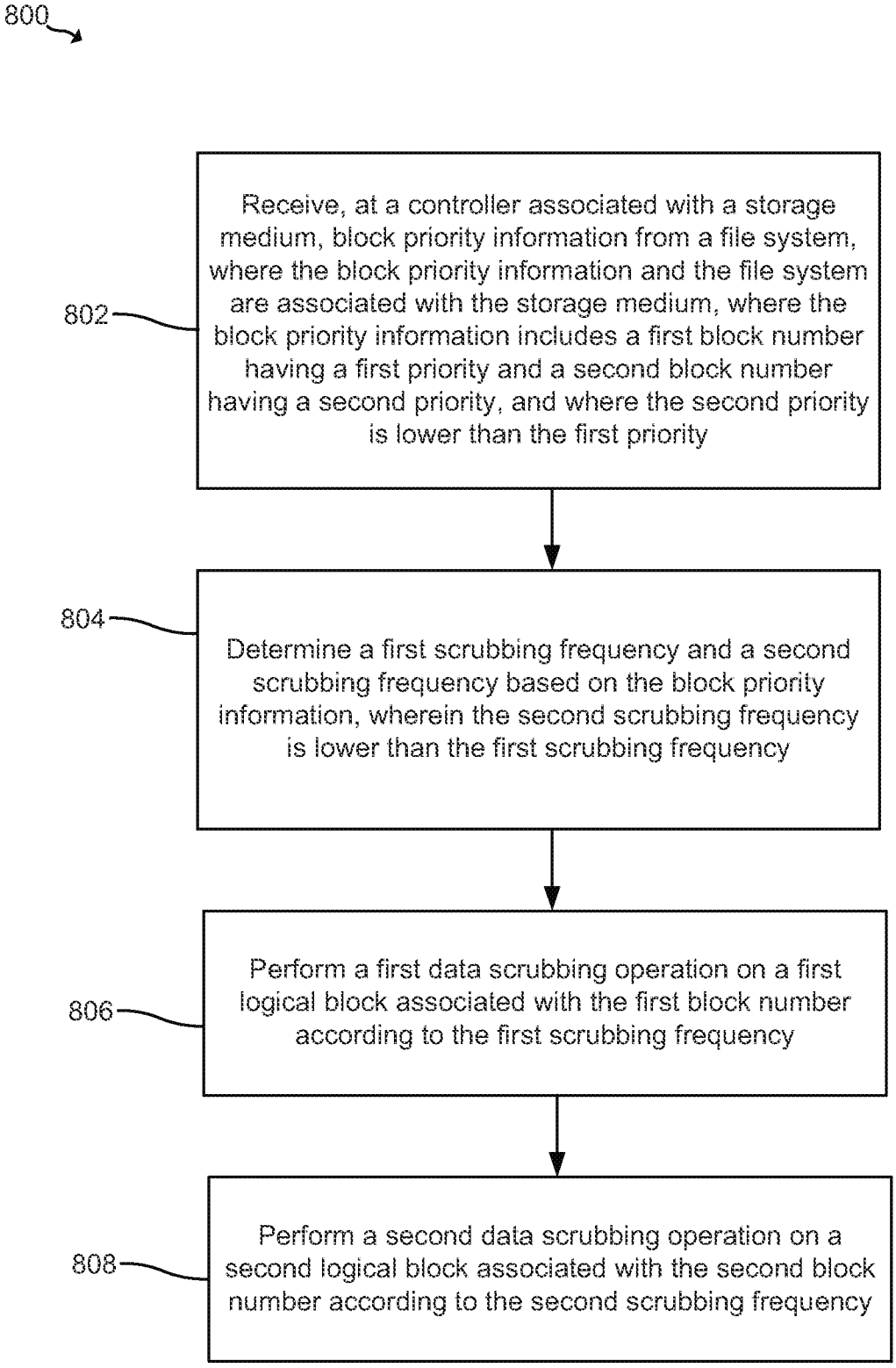
800
802
Receive, at a controller associated with a storage medium, block priority information from a file system, where the block priority information and the file system are associated with the storage medium, where the block priority information includes a first block number having a first priority and a second block number having a second priority, and where the second priority is lower than the first priority
804
Determine a first scrubbing frequency and a second scrubbing frequency based on the block priority information, wherein the second scrubbing frequency is lower than the first scrubbing frequency
806
Perform a first data scrubbing operation on a first logical block associated with the first block number according to the first scrubbing frequency
808
Perform a second data scrubbing operation on a second logical block associated with the second block number according to the second scrubbing frequency

FIG. 8

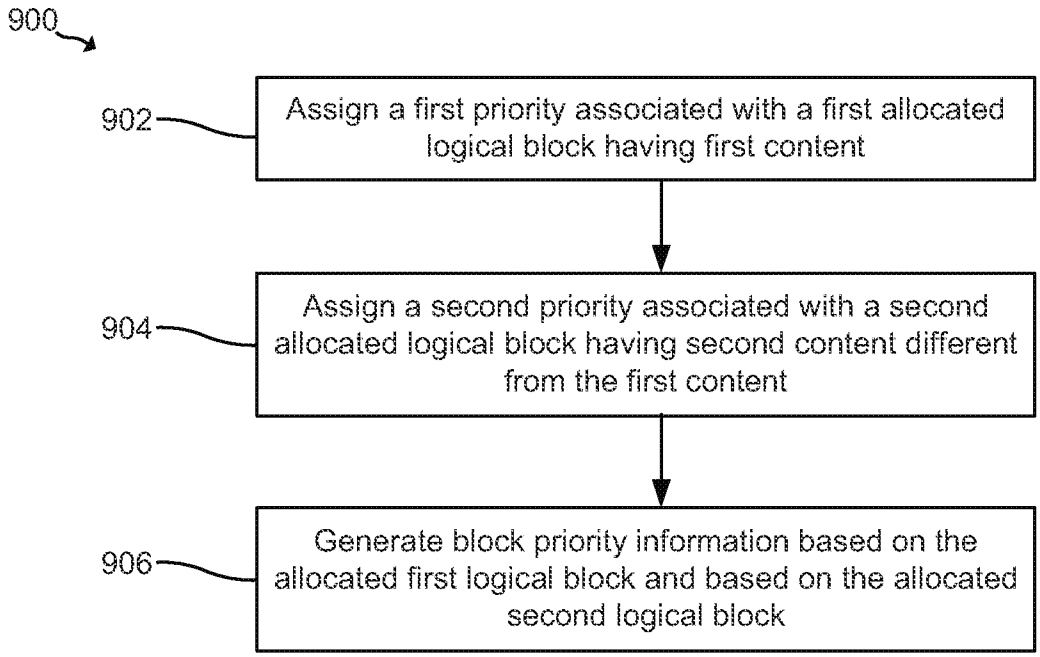
900
902
Assign a first priority associated with a first allocated logical block having first content
904
Assign a second priority associated with a second allocated logical block having second content different from the first content
906
Generate block priority information based on the allocated first logical block and based on the allocated second logical block

FIG. 9

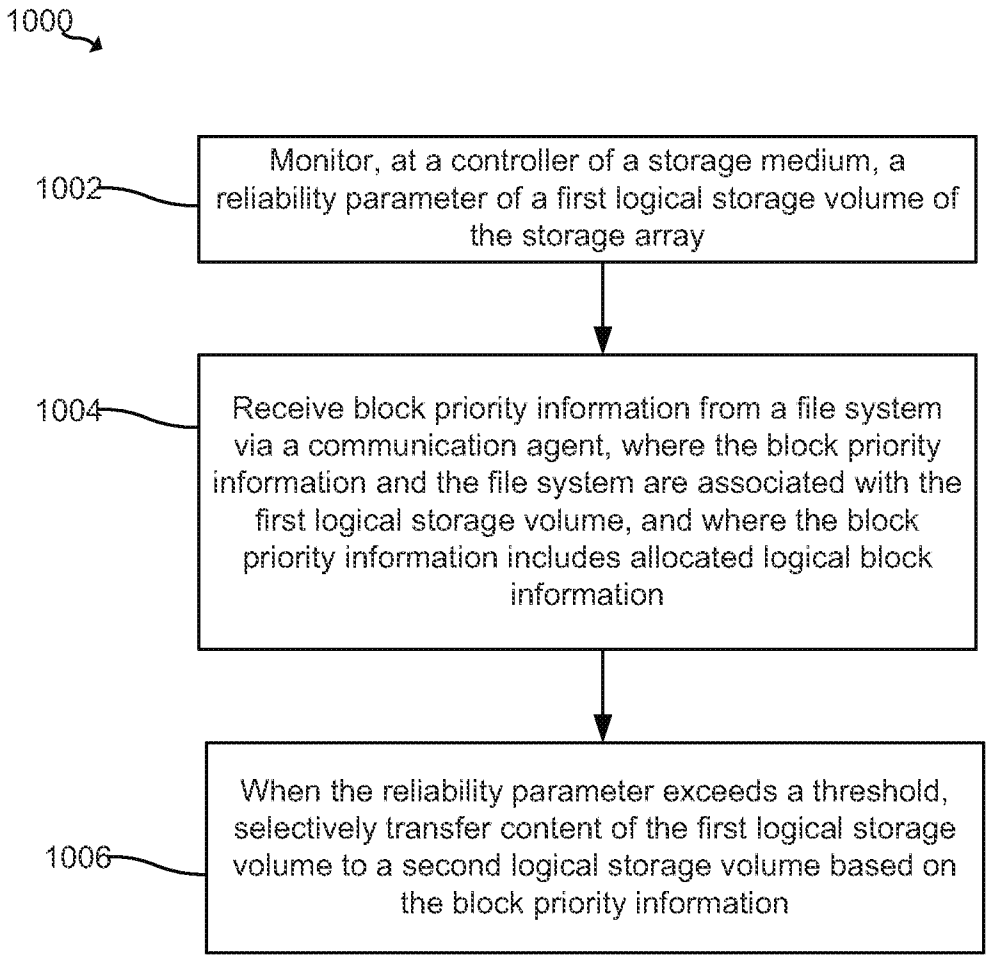
1000
1002
Monitor, at a controller of a storage medium, a reliability parameter of a first logical storage volume of the storage array
1004
Receive block priority information from a file system via a communication agent, where the block priority information and the file system are associated with the first logical storage volume, and where the block priority information includes allocated logical block information
1006
When the reliability parameter exceeds a threshold, selectively transfer content of the first logical storage volume to a second logical storage volume based on the block priority information

FIG. 10

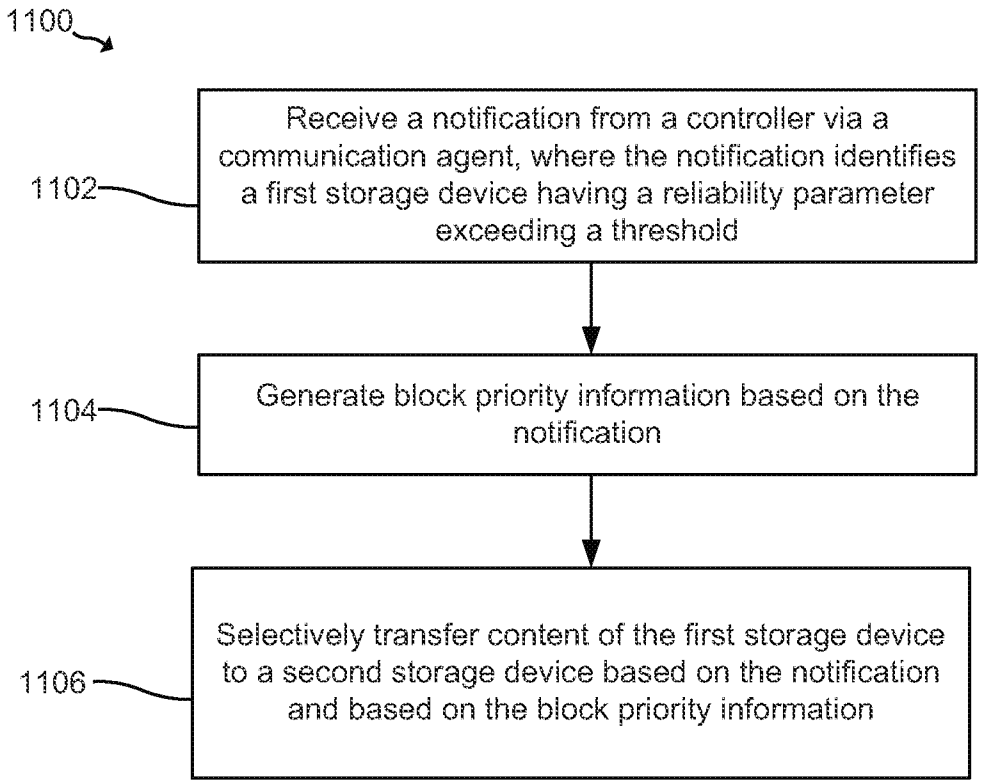
1100
Receive a notification from a controller via a communication agent, where the notification identifies a first storage device having a reliability parameter exceeding a threshold
1102
Generate block priority information based on the notification
1104
Selectively transfer content of the first storage device to a second storage device based on the notification and based on the block priority information
1106

FIG. 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/057318

A. CLASSIFICATION OF SUBJECT MATTER

G06F 12/06(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F 12/06; G06F 11/20; G06F 12/00; G06F 11/00; H02H 3/05; G06F 11/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: storage, system, filesystem, controller, communication, block, priority, reconstruct, scrub, migration, allocation, unallocation, metadata, reliability, SMART, RAID

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2009-0271659 A1 (ULF TROPPENS et al.) 29 October 2009 See paragraphs [0002], [0006], [0009]-[0010], [0016], [0018], [0022]-[0023], [0028]; and figures 1, 4.	1-15
A	US 2006-0085594 A1 (RANDY L. ROBERSON et al.) 20 April 2006 See paragraphs [0002], [0007], [0044], [0063]-[0064], [0077]; and figures 1, 6, 13, 18, 20.	1-15
A	US 8086911 B1 (JAMES TAYLOR) 27 December 2011 See column 5, lines 39-40; column 13, lines 2-3, 29-31, 37-42; and figures 3, 7-8.	1-15
A	US 2003-0237019 A1 (STEVEN R. KLEIMAN et al.) 25 December 2003 See paragraph [0036]; and figure 1.	1-15
A	US 2007-0294570 A1 (SURI B. POLISETTI et al.) 20 December 2007 See paragraph [0020]; and figure 2.	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 May 2015 (19.05.2015)	**26 May 2015 (26.05.2015)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701, Republic of Korea
Facsimile No. ++82 42 472 7140

Authorized officer
BYUN, Sung Cheal
Telephone No. +82-42-481-8262

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/057318

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009-0271659 A1	29/10/2009	None	
US 2006-0085594 A1	20/04/2006	JP 2006-120118 A	11/05/2006
		US 7594075 B2	22/09/2009
US 8086911 B1	27/12/2011	None	
US 2003-0237019 A1	25/12/2003	AU 2003-278864 A1	06/01/2004
		CN 100378672 C	02/04/2008
		CN 101154174 A	02/04/2008
		CN 101154174 B	06/10/2010
		CN 1679000 A	05/10/2005
		CN 1679000 C	02/04/2008
		EP 1535154 A1	01/06/2005
		EP 1535154 A4	10/09/2008
		IL 165831 A	30/12/2010
		IL 165831 D0	15/01/2006
		JP 04846763 B2	28/12/2011
		JP 2005-531071 A	13/10/2005
		JP 2008-269638 A	06/11/2008
		US 7024586 B2	04/04/2006
		WO 04-001600 A1	31/12/2003
US 2007-0294570 A1	20/12/2007	US 7721146 B2	18/05/2010