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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 January 2017 (26.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/014809 A1

(51) **International Patent Classification**:
G06F 17/30 (2006.01)

(21) **International Application Number**:
PCT/US2016/020260

(22) **International Filing Date**:
1 March 2016 (01.03.2016)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
3697/CHE/2015 18 July 2015 (18.07.2015) IN

(71) **Applicant**: **HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.** [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77070 (US).

(72) **Inventors**: **CHIRUMAMILA, Narendra**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN). **MAHESH, Keshetti**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN). **NAYAKA B, Govindaraja**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN). **BASIREDDY, Ranjith Reddy**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN). **MOHANTA, Taranisen**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN).

(74) **Agents**: **ORTEGA, Arthur S.** et al.; Hewlett Packard Enterprise, 3404 E. Harmony Road, Mail Stop 79, Fort Collins, Colorado 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 the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

[Continued on next page]

(54) **Title**: DATA VOLUME MIGRATION IN DISTRIBUTED STORAGE SYSTEMS

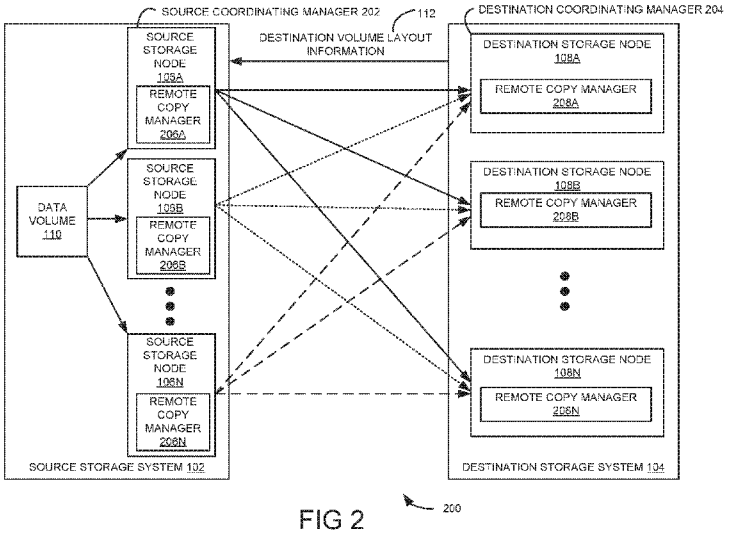

FIG 2

(57) **Abstract**: In one example, a system for migrating data volume in distributed storage systems is described which includes a source storage system and a destination storage system. The source storage system includes source storage nodes. The destination storage system is communicatively connected to the source storage system. The destination storage system includes destination storage nodes. The source storage system to obtain volume layout information associated with the destination storage system, map data volume associated with the source storage system with at least one destination storage node based on the volume layout information, and to migrate the data volume directly from the source storage nodes associated with the data volume to the at least one destination storage node based on the mapped data volume. The data volume is distributed across the source storage nodes in the source storage system.

WO 2017/014809 A1

Published:

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

DATA VOLUME MIGRATION IN DISTRIBUTED STORAGE SYSTEMS

BACKGROUND

[0001] Computer systems may include storage systems with multiple storage nodes/storage controllers communicatively connected with each other. Data may be distributed across the storage nodes.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] Examples are described in the following detailed description and in reference to the drawings, in which:

[0003] Fig. 1 depicts an example system for copying data volume in distributed storage systems;

[0004] Fig. 2 depicts an example block diagram illustrating additional components of the example system of Fig. 1;

[0005] Fig. 3 depicts a flow chart of an example method for copying data volume in distributed storage systems; and

[0006] Fig. 4 is an example block diagram showing a non-transitory computer-readable medium for copying data volume in distributed storage systems.

DETAILED DESCRIPTION

[0007] A distributed storage system may include multiple storage nodes, and data volume(s) distributed across the multiple storage nodes (i.e., at least a part of the data volume is stored in each of the multiple storage nodes). For example, a data volume may refer to an amount of data in the distributed storage system. Further, the amount of data may be distributed across the multiple

storage nodes in the distributed storage system. Further, the data volume may need to be copied or migrated from one distributed storage system (e.g., a source storage system) to another distributed storage system (e.g., a destination storage system). For example, data volume may need to be copied or migrated from the source storage system to the destination storage system for upgrading the source storage system infrastructure.

[0008] The data volume may be copied or migrated using one source storage node of the source storage system. For example, a source storage node acting as a source remote copy manager may communicate with other source storage nodes of the source storage system to determine data volume and volume layout information associated with the other source storage nodes of the source storage system. The source remote copy manager may then migrate the data volume to one of the destination storage nodes of the destination storage system that acts as a destination remote copy manager. The destination remote copy manager may then transfer the data volume to other destination storage nodes of the destination storage system based on volume layout information associated with the destination storage system. Example methods of using a single source storage node of the source storage system for migrating the data volume may lead to underutilization of source storage system resources. Also, utilizing a single node in the destination storage system for coordinating the copy process may result in a single node bottleneck.

[0009] The present application discloses techniques to provide distributed storage architecture in which source storage nodes directly connect to respective destination storage nodes and transfer the data volume(s) based on the volume layout information of the destination storage system. The volume layout information may include information such as a number of destination storage nodes, unique identifiers associated with the destination storage nodes, details of volume data blocks associated with the destination storage nodes, page size, stride size, available free space associated with the destination storage nodes, and the like.

[00010] In one example, the source storage system may obtain volume layout information associated with the destination storage system, map data volume associated with the source storage system with at least one destination storage node based on the volume layout information, and migrate the data volume directly from the source storage nodes associated with the data volume to the destination storage nodes based on the mapped data volume. The data volume may be distributed across the plurality of source storage nodes in the source storage system. For example, at least a part of the data volume is stored in each of the source storage nodes. In one example, the data volume is migrated directly from the source storage nodes to the destination storage nodes by establishing a connection between the source storage nodes associated with the data volume and the destination storage nodes based on the mapped data volume.

[00011] The terms "storage system" and "distributed storage system" are used interchangeably throughout the document. Further, the term "storage node" and "storage controller" are used interchangeably throughout the document. Further, the terms "copy," "migrate,", and "transfer" are used interchangeably throughout the document.

[00012] Fig. 1 depicts an example system 100 for copying data volume 110 in distributed storage systems. The system 100 may include a source storage system 102 and a destination storage system 104 communicatively connected to the source storage system 102. Each of the source storage system 102 and the destination storage system 104 may include multiple storage nodes and data volume 110 distributed across these storage nodes. Further, the source storage system 102 may include source storage nodes 106A to 106N and the destination storage system 104 may include destination storage nodes 108A to 108N. Furthermore, the source storage system 102 includes data volume 110 distributed across the source storage nodes 106A to 106N. In the example of Fig. 1, the source storage system 102 is shown as including a single data volume

110, however, the source storage system 102 can include any number of data volumes that can be distributed across the source storage nodes 106A to 106N. Each storage node may include a physical storage unit that may be used to store at least a portion of the data volume 110. Each storage node may include hardware, software, or embedded logic components or a combination of two or more such components for carrying out the appropriate functionalities implemented or supported by the storage node. For example, each of the storage nodes may include a storage controller, a virtual storage controller, a storage appliance, a virtual storage appliance, and/or associated storage devices. Example storage devices may include Random Access Memory (RAM), volatile memory, non-volatile memory, flash memory, a storage drive (e.g., a hard drive), a solid state drive, and the like, or a combination thereof. Example storage controllers may include virtual storage controllers, disk array controllers, internet small computer system interface (ISCSI) controllers, fiber channel controllers, distributed storage controllers, and the like.

[00013] In operation, the source storage system 102 may obtain volume layout information 112 associated with the destination storage system 104. The volume layout information may include information such as a number of destination storage nodes, unique identifiers associated with the destination storage nodes, details of volume data blocks associated with the destination storage nodes, page size, stride size, and available free space associated with each of the destination storage nodes. For example, the volume layout information may be schema of a distributed storage system. In one example, the volume layout information may be in the form a configuration data structure and/or a file.

[00014] The source storage system 102 may then map data volume/volumes 110 associated with the source storage system 102 with at least one destination storage node using the volume layout information 112. For example, each data volume may be uniquely identified within the distributed storage system by a volume identifier. That is, each data volume is

associated with a volume identifier that is unique within distributed storage system. In one example, data blocks of the data volume 110 (e.g., having unique identifiers) are mapped to the unique identifiers of the destination storage nodes using the volume layout information 112. Upon mapping the data volume 110, the source storage system 102 may migrate the data volume 110 directly from the source storage nodes 106A-N associated with the data volume 110 to the destination storage nodes 108A-N based on the mapped data volume 110. In one example, the data volume 110 may be migrated directly from the source storage nodes 106A-N to the destination storage nodes 108A-N by establishing a connection between the source storage nodes 106A-N associated with the data volume 110 and the destination storage nodes 108A-N based on the mapped data volume 110. For example, the term "directly migrating" may refer to a source storage node migrating associated portion of the data volume 110 to a destination storage node without sending the associated portion of the data volume 110 to any intermediate storage nodes in the source storage system and the destination storage system.

[00015] Referring now to Fig. 2 which is a block diagram 200, illustrating additional components of the example system 100 of Fig. 1. As shown in Fig. 2, the source storage nodes 106A-N may include remote copy managers 206A-N, respectively. Also as shown in Fig. 2, the destination storage nodes 108A-N may include remote copy managers 208A-N, respectively.

[00016] In the example shown in Fig. 2, the source storage node 106A acts as a source coordinating manager 202 and the destination storage node 108A acts as a destination coordinating manager 204. The source coordinating manager 202 and the destination coordinating manager 204 may be communicatively connected to each other.

[00017] The source coordinating manager 202 may obtain the source volume layout information of the source storage nodes 106A-N using the associated remote copy managers 206A-N. For example, the source volume

layout information of the source storage nodes 106A-N may include information such as unique identifiers associated with the source storage nodes 106A-N, details of volume data blocks associated with the source storage nodes 106A-N, and available free space associated with each of the source storage nodes 106A-N. Similarly, the destination coordinating manager 204 may obtain the destination volume layout information of the destination storage nodes 108A-N using the associated remote copy managers 208A-N. For example, the destination volume layout information of the destination storage nodes 108A-N may include information such as unique identifiers associated with the destination storage nodes 108A-N, details of volume data blocks associated with the destination storage nodes 108A-N, and available free space associated with each of the destination storage nodes 108A-N. The remote copy managers 206A-N and remote copy managers 208A-N may represent any combination of circuitry and executable instructions to perform the above described examples.

[00018] Further, the source coordinating manager 202 obtains the destination volume layout information 112 from the destination coordinating manager 204. In one example, the source coordinating manager 202 may establish a session with the destination coordinating manager 204 and fetch the destination volume layout information 112 from the destination storage system 104. Further, the source coordinating manager 202 generates mapping information by mapping data blocks of data volume (e.g., 110) associated with the source storage system 102 with the destination storage nodes 108A-N using the obtained destination volume layout information 112. For example, the mapping information may include information that how the data blocks associated with the source storage nodes 106A-N are mapped with the destination storage nodes 108A-N. Upon mapping, the source coordinating manager 202 propagates the mapping information to the source storage nodes 106A-N. Upon receiving the mapping information, the source storage nodes 106A-N may communicate directly with the destination storage nodes 108A-N to copy the data volume 110. In one example, each source storage node may then establish a connection with corresponding destination storage node(s) using the mapping information for

migrating the data volume 110. This may enable copying of the data volume between the source and destination storage nodes independently and in parallel.

[00019] In one example, each of the source storage nodes 106A-N may communicate directly with at least one of the destination storage nodes 108A-N to copy the data volume 110 as follows. At a source storage node in the source storage system:

a. Create a list of input/output (IO) requests corresponding to the data blocks of the data volume which are allocated on the source storage node.

b. Segregate the list of IO requests into multiple subsets based on the volume layout information associated with the destination storage nodes. Each subset corresponds to a data volume to be copied from the source storage node to a destination storage node. For example, the IO requests in one subset may be targeted to one destination storage node.

c. In a loop (i.e., till the IO requests in the list are completed),

1. Obtain network utilization information from the destination storage nodes. For example, the network utilization information may include network utilization percentage on each destination storage node.

2. Obtain IO workload on the source storage node.

3. Determine a number of IO requests to be issued to the destination storage nodes in each iteration based on the network utilization percentage and the IO workload. For example, when the network utilization on destination storage node is 10% and IO workload on the source storage node is 40%, then 25 IO requests can be issued in one iteration. In another example, when network utilization is 50% and IO workload on the source storage node is 70%, then 5 IO requests can be issued in one iteration.

4. Issue the IO requests to the destination storage nodes by the source storage node.

5. After all the IO requests are completed, inform the status to the source coordinating manager.

[00020] Fig. 3 depicts an example flow chart 300 of an example method for migrating data volume in distributed storage systems. At block 302, destination volume layout information associated with a destination storage system is obtained by a source storage system. Example volume layout information may include a number of destination storage nodes, unique identifiers associated with the destination storage nodes, details of volume data blocks associated with the destination storage nodes, page size, stride size, and/or available free space associated with each of the destination storage nodes.

[00021] At block 304, data volume associated with the source storage system is mapped with at least one destination storage node of the destination storage system based on the volume layout information. In one example, the volume layout information associated with the destination storage system is received by one of the source storage nodes acting as a source coordinating manager. Further, mapping information is generated by mapping the data volume associated with the source storage system with the destination storage nodes based on the volume layout information by the source coordinating manager. The data volume may be distributed across the source storage nodes in the source storage system. Furthermore, the mapping information is propagated to each of the plurality of source storage nodes in the source storage system by the source coordinating manager.

[00022] At block 306, the data volume is copied from the source storage system to the destination storage system by establishing a connection between the source storage nodes associated with the data volume and the destination storage nodes based on the mapped data volume. For example, the connection may be established based on the mapping information.

[00023] In one example, a list of IO requests corresponding to data blocks allocated for each source storage node may be created. Further, the list of IO requests may be segregated into multiple subsets of IO requests based on the volume layout information. Each subset may correspond to at least a portion of

the data volume to be copied to one of the destination storage nodes. Further, the data volume may be copied from the source storage system to the destination storage system based on the multiple subsets of IO requests.

[00024] In another example, in copying the data volume from the source storage system to the destination storage system, a number of IO requests to be issued from a source storage node to a destination storage node may be determined based on network utilization of the destination storage node and IO workload of the source storage node.

[00025] Fig. 3 shows an example method and it should be understood that other configurations can be employed to practice the techniques of the present application. For example, method described above may be employed by copying volume data from a source storage system to multiple destination storage system.

[00026] Fig. 4 is an example block diagram showing a non-transitory computer-readable medium that stores code for operation in accordance with an example of the techniques of the present application. The non-transitory computer-readable medium is generally referred to by the reference number 402 and may be included in a computing system 400 in relation to Fig. 1. The non-transitory computer-readable medium 402 may correspond to any storage device that stores computer-implemented instructions, such as programming code or the like. For example, the non-transitory computer-readable medium 402 may include non-volatile memory, volatile memory, and/or storage devices. Examples of non-volatile memory include, but are not limited to, electrically erasable programmable Read Only Memory (EEPROM) and Read Only Memory (ROM). Examples of volatile memory include, but are not limited to, Static Random Access Memory (SRAM), and dynamic Random Access Memory (DRAM). Examples of storage devices include, but are not limited to, hard disk drives, compact disc drives, digital versatile disc drives, optical drives, and flash memory devices.

[00027] A processor 404 generally retrieves and executes the instructions stored in the non-transitory computer-readable medium 402 to operate the present techniques in accordance with an example. In one example, the tangible, computer-readable medium 402 can be accessed by the processor 404 over a bus.

[00028] The machine-readable storage medium 402 may store instructions 406-412. In an example, instructions 406-412 may be executed by the processor 404 to provide a mechanism for copying data volume in distributed storage systems. Instructions 406 may be executed by the processor 404 to receive an IO request to copy data volume from a source storage system having source storage nodes to a destination storage system having destination storage nodes. Instructions 408 may be executed by the processor 404 to obtain volume layout information associated with the destination storage system (e.g., destination storage nodes). Instructions 410 may be executed by the processor 404 to map data volume associated with the source storage system destination storage nodes of the destination storage system based on the volume layout information. The data volume may be distributed across the source storage nodes in the source storage system. Instructions 412 may be executed by the processor 404 to migrate the data volumes from the source storage system to the destination storage system by establishing a connection between the source storage nodes associated with the data volume and the destination storage nodes based on the mapped data volume.

[00029] Although shown as contiguous blocks, the machine readable instructions can be stored in any order or configuration. For example, if the non-transitory computer- readable medium 402 is a hard drive, the machine readable instructions can be stored in non-contiguous, or even overlapping, sectors.

[00030] As used herein, a "processor" may include processor resources such as at least one of a Central Processing Unit (CPU), a semiconductor-based

microprocessor, a Graphics Processing Unit (GPU), a Field-Programmable Gate Array (FPGA) to retrieve and execute instructions, other electronic circuitry suitable for the retrieval and execution instructions stored on a computer-readable medium, or a combination thereof. The processor fetches, decodes, and executes instructions stored on computer-readable medium 402 to perform the functionalities described below. In other examples, the functionalities of any of the instructions of computer- readable medium 402 may be implemented in the form of electronic circuitry, in the form of executable instructions encoded on a computer-readable storage medium, or a combination thereof.

[00031] As used herein, a "computer-readable medium" may be any electronic, magnetic, optical, or other physical storage apparatus to contain or store information such as executable instructions, data, and the like. For example, any computer-readable storage medium described herein may be any of Random Access Memory (RAM), volatile memory, non-volatile memory, flash memory, a storage drive (e.g., a hard drive), a solid state drive, any type of storage disc (e.g., a compact disc, a DVD, etc.), and the like, or a combination thereof. Further, any computer-readable medium described herein may be non-transitory. In examples described herein, a computer-readable medium or media is part of an article (or article of manufacture). An article or article of manufacture may refer to any manufactured single component or multiple components. The medium may be located either in the system executing the computer-readable instructions, or remote from but accessible to the system (e.g., via a computer network) for execution. In the example of Fig. 4, computer- readable medium 402 may be implemented by one computer-readable medium, or multiple computer-readable media.

[00032] In examples described herein, the source storage system may communicate with the destination storage system via a network interface device. Further, in examples described herein, the source storage nodes may communicate with each other via a network interface device. Furthermore, the destination storage nodes may communicate with each other via a network

interface device. In examples described herein, a "network interface device" may be a hardware device to communicate over at least one computer network. In some examples, a network interface may be a Network Interface Card (NIC) or the like. As used herein, a computer network may include, for example, a Local Area Network (LAN), a Wireless Local Area Network (WLAN), a Virtual Private Network (VPN), the Internet, or the like, or a combination thereof. In some examples, a computer network may include a telephone network (e.g., a cellular telephone network).

[00033] In some examples, instructions may be part of an installation package that, when installed, may be executed by processor 404 to implement the functionalities described herein in relation to instructions. In such examples, computer- readable medium 402 may be a portable medium, such as a CD, DVD, or flash drive, or a memory maintained by a server from which the installation package can be downloaded and installed. In other examples, instructions may be part of an application, applications, or component(s) already installed on the computing system 400 including processor 404. In such examples, the computer-readable medium 402 may include memory such as a hard drive, solid state drive, or the like. In some examples, functionalities described herein in relation to Figs. 1 through 4 may be provided in combination with functionalities described herein in relation to any of Figs. 1 through 4.

[00034] The example methods and systems described through Figs. 1-4 may enable faster copy of data volumes in distributed storage systems as all the source and destination storage nodes are engaged in the volume copy process. The example methods and systems described through Figs. 1-4 may provide efficient utilization of storage system resources to complete the data volume copy. The example methods and systems described through Figs. 1-4 may also provide regular optimization of data volume copy load on each source storage node based on the IO load associated with corresponding source storage node.

[00035] It may be noted that the above-described examples of the present solution is for the purpose of illustration only. Although the solution has been described in conjunction with a specific embodiment thereof, numerous modifications may be possible without materially departing from the teachings and advantages of the subject matter described herein. Other substitutions, modifications and changes may be made without departing from the spirit of the present solution. All of the features disclosed in this specification (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

[00036] The terms "include," "have," and variations thereof, as used herein, have the same meaning as the term "comprise" or appropriate variation thereof. Furthermore, the term "based on," as used herein, means "based at least in part on." Thus, a feature that is described as based on some stimulus can be based on the stimulus or a combination of stimuli including the stimulus.

[00037] The present description has been shown and described with reference to the foregoing examples. It is understood, however, that other forms, details, and examples can be made without departing from the spirit and scope of the present subject matter that is defined in the following claims.

WHAT IS CLAIMED IS:

1. A system for migrating data volume in distributed storage systems, comprising:

a source storage system comprising:

a plurality of source storage nodes; and

a destination storage system communicatively connected to the source storage system, the destination storage system comprising:

a plurality of destination storage nodes, wherein the source storage system to:

obtain volume layout information associated with the destination storage system;

map data volume associated with the source storage system with at least one destination storage node based on the volume layout information, wherein the data volume is distributed across the plurality of source storage nodes in the source storage system; and

migrate the data volume directly from the plurality of source storage nodes associated with the data volume to the at least one destination storage node based on the mapped data volume.

2. The system of claim 1, wherein the data volume is migrated directly from the plurality of source storage nodes to the at least one destination storage node by establishing a connection between the plurality of source storage nodes associated with the data volume and the at least one destination storage node based on the mapped data volume.

3. The system of claim 1, wherein the source storage system comprising a source storage node acting as a source coordinating manager, wherein the destination storage system comprising a destination storage node acting as a destination coordinating manager, wherein the source coordinating manager communicatively connected to the destination coordinating manager to:

receive the volume layout information associated with the destination storage system;

generate mapping information by mapping the data volume associated with the source storage system with the at least one destination storage node based on the volume layout information; and

propagate the mapping information to each of the plurality of source storage nodes.

4. The system of claim 3, wherein each source storage node comprises a remote copy manager, wherein the source coordinating manager to obtain volume layout information of the plurality of source storage nodes using the remote copy manager associated with each of the plurality of source storage nodes.

5. The system of claim 3, wherein each destination storage node comprises a remote copy manager, wherein the destination coordinating manager to obtain the volume layout information of the plurality of destination storage nodes using the remote copy manager associated with each destination storage node.

6. The system of claim 1, wherein each of the plurality of source storage nodes and each of the plurality of destination storage nodes comprises a storage controller, a virtual storage controller, a storage appliance, and a virtual storage appliance.

7. The system of claim 1, wherein the volume layout information comprises information selected from the group consisting of a number of destination storage nodes, unique identifiers associated with the plurality of destination storage nodes, details of volume data blocks associated with the plurality of destination storage nodes, page size, stride size, and available free space associated with each of the plurality of destination storage nodes.

8. A method for migrating data volume in distributed storage systems, the method comprising:

obtaining volume layout information associated with a destination storage system by a source storage system;

mapping data volume associated with the source storage system with at least one destination storage node of the destination storage system based on the volume layout information, wherein the data volume is distributed across a plurality of source storage nodes in the source storage system; and

migrating the data volume from the source storage system to the destination storage system by establishing a connection between the plurality of source storage nodes associated with the data volume and the at least one destination storage node based on the mapped data volume.

9. The method of claim 8, wherein the volume layout information comprises information selected from the group consisting of a number of destination storage nodes, unique identifiers associated with a plurality of destination storage nodes, details of volume data blocks associated with the plurality of destination storage nodes, page size, stride size, and available free space associated with each of the plurality of destination storage nodes.

10. The method of claim 8, wherein mapping data volume associated with the source storage system with the at least one destination storage node of the destination storage system based on the volume layout information, comprises:

receiving the volume layout information associated with the destination storage system by one of the plurality of source storage nodes acting as a source coordinating manager;

generating mapping information by mapping the data volume associated with the source storage system with the at least one destination storage node based on the volume layout information by the source coordinating manager; and

propagating the mapping information to each of the plurality of source storage nodes by the source coordinating manager.

11. The method of claim 8, wherein migrating the data volume from the source storage system to the destination storage system, comprises:

creating a list of input/output (IO) requests corresponding to data blocks allocated for each source storage node;

segregating the list of IO requests into multiple subsets of IO requests based on the volume layout information, wherein each subset corresponding to at least a portion of the data volume to be copied to one of the destination storage nodes; and

migrating the data volume from the source storage system to the destination storage system based on the multiple subsets of IO requests.

12. The method of claim 8, wherein, in migrating the data volume from the source storage system to the destination storage system, a number of IO requests to be issued from a source storage node to a destination storage node is determined based on network utilization of the destination storage node and IO workload of the source storage node.

13. A non-transitory computer-readable medium having computer executable instructions stored thereon for migrating data volume in distributed storage systems, the instructions are executable by a processor to:

receive an input/output (IO) request to migrate data volume from a source storage system having a plurality of source storage nodes to a destination storage system having a plurality of destination storage nodes;

obtain volume layout information associated with the destination storage system;

map data volume associated with the source storage system with at least one destination storage node of the destination storage system based on the volume layout information, wherein the data volume is distributed across the plurality of source storage nodes in the source storage system; and

migrate the data volume from the source storage system to the destination storage system by establishing a connection between the plurality of source storage nodes associated with the data volume and the at least one destination storage node based on the mapped data volume.

14. The non-transitory computer-readable medium of claim 13, wherein the volume layout information comprises information selected from the group consisting of a number of destination storage nodes, unique identifiers associated with the plurality of destination storage nodes, details of volume data blocks associated with the plurality of destination storage nodes, page size, stride size, and available free space associated with each of the plurality of destination storage nodes.

15. The non-transitory computer-readable medium of claim 13, wherein migrating the data volume from the source storage system to the destination storage system, comprises instructions executable by the processor to:

create a list of IO requests corresponding to data blocks allocated for each source storage node;

segregate the list of IO requests into multiple subsets of IO requests based on the volume layout information, wherein each subset corresponding to data to be copied to one of the destination storage nodes; and

migrate the data volume from the source storage system to the destination storage system based on the multiple subsets of IO requests.

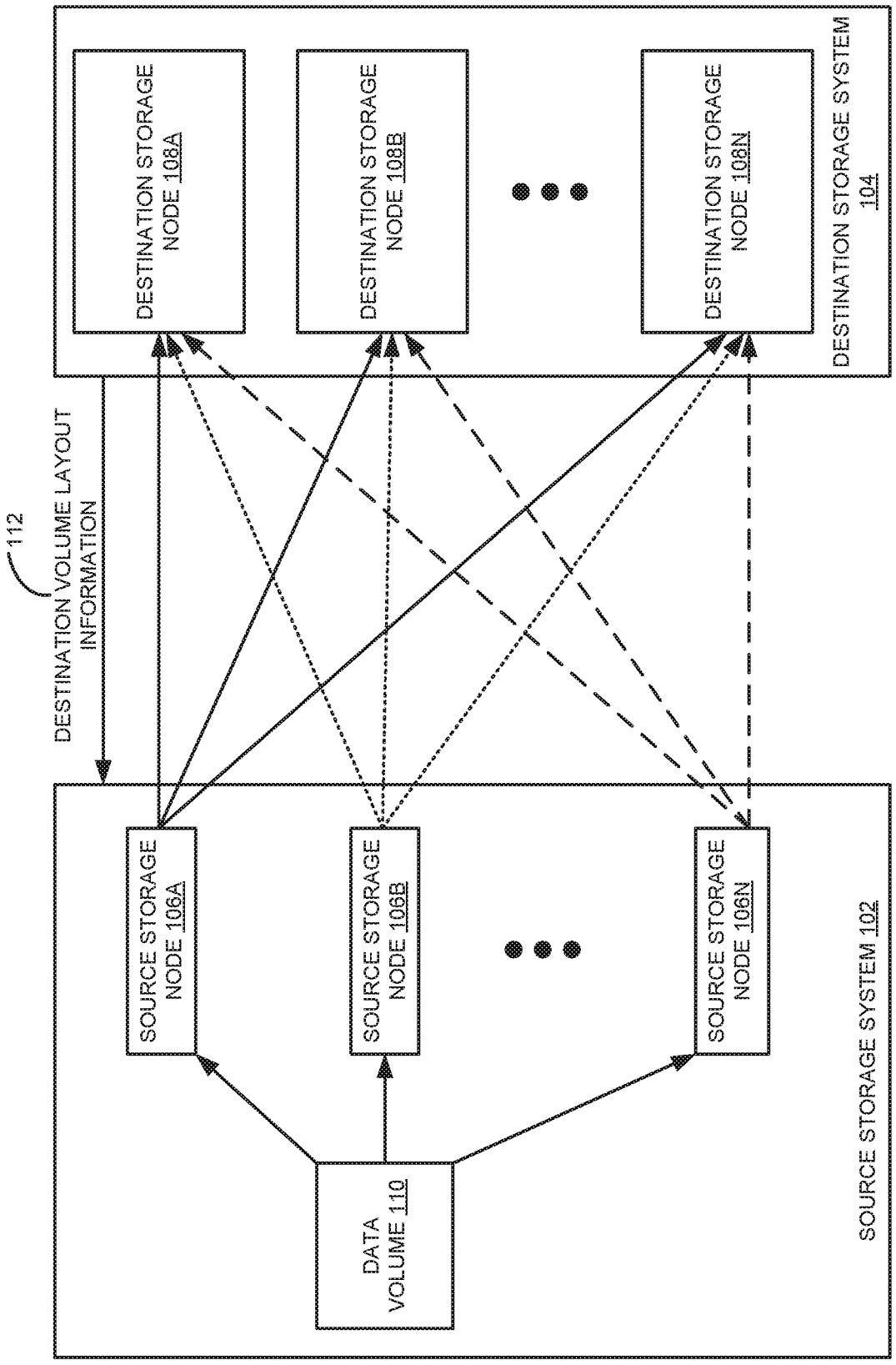
DESTINATION STORAGE NODE 108A
DESTINATION STORAGE NODE 108B
DESTINATION STORAGE NODE 108N
DESTINATION STORAGE SYSTEM 104
112
DESTINATION VOLUME LAYOUT INFORMATION
SOURCE STORAGE NODE 106A
SOURCE STORAGE NODE 106B
SOURCE STORAGE NODE 106N
DATA VOLUME 110
SOURCE STORAGE SYSTEM 102
100

FIG 1

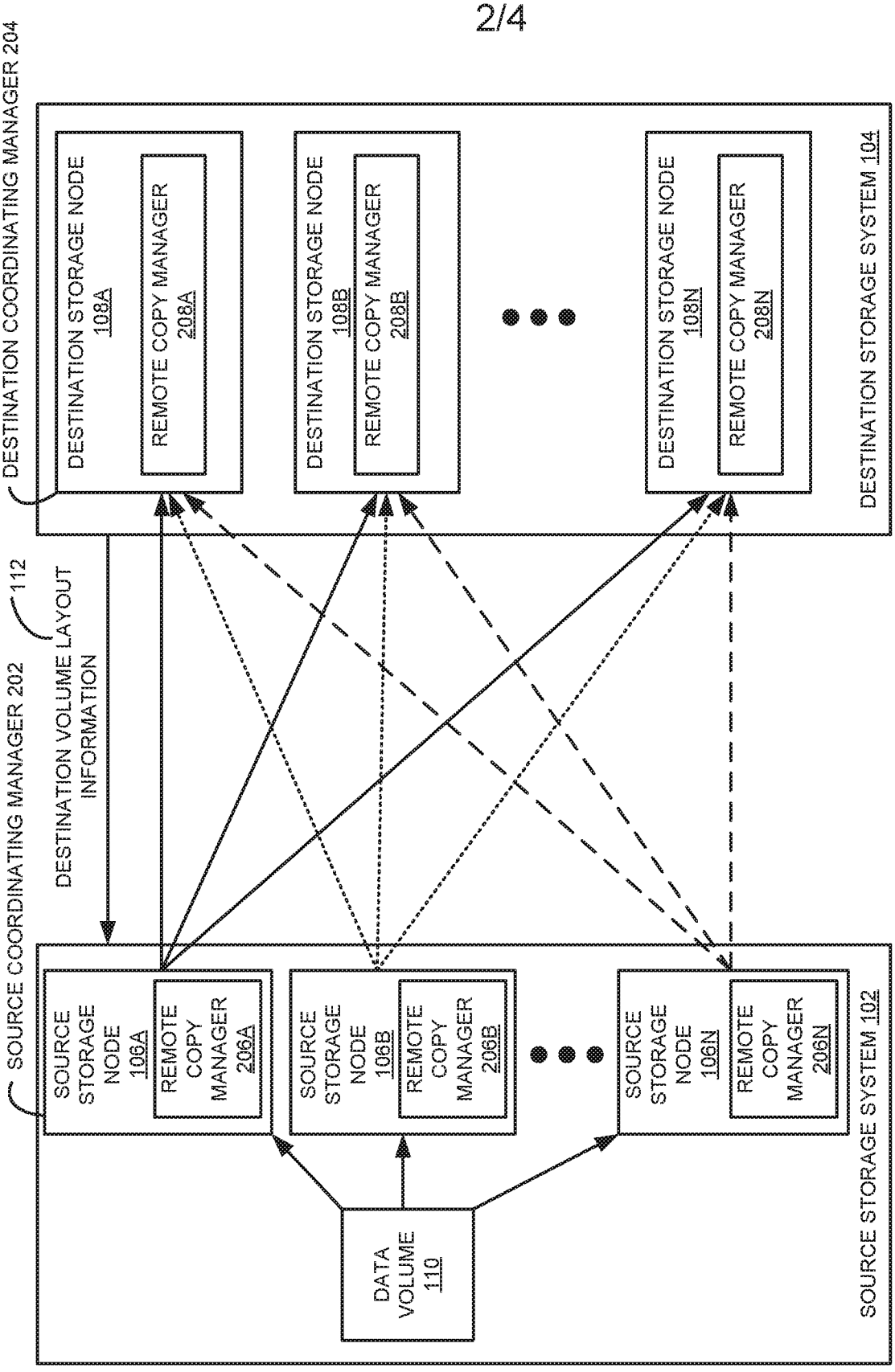
DESTINATION COORDINATING MANAGER 204
DESTINATION STORAGE NODE 108A
REMOTE COPY MANAGER 208A
DESTINATION STORAGE NODE 108B
REMOTE COPY MANAGER 208B
DESTINATION STORAGE NODE 108N
REMOTE COPY MANAGER 208N
DESTINATION STORAGE SYSTEM 104
112
DESTINATION VOLUME LAYOUT INFORMATION
SOURCE COORDINATING MANAGER 202
SOURCE STORAGE NODE 106A
REMOTE COPY MANAGER 206A
SOURCE STORAGE NODE 106B
REMOTE COPY MANAGER 206B
SOURCE STORAGE NODE 106N
REMOTE COPY MANAGER 206N
DATA VOLUME 110
SOURCE STORAGE SYSTEM 102
200

FIG 2

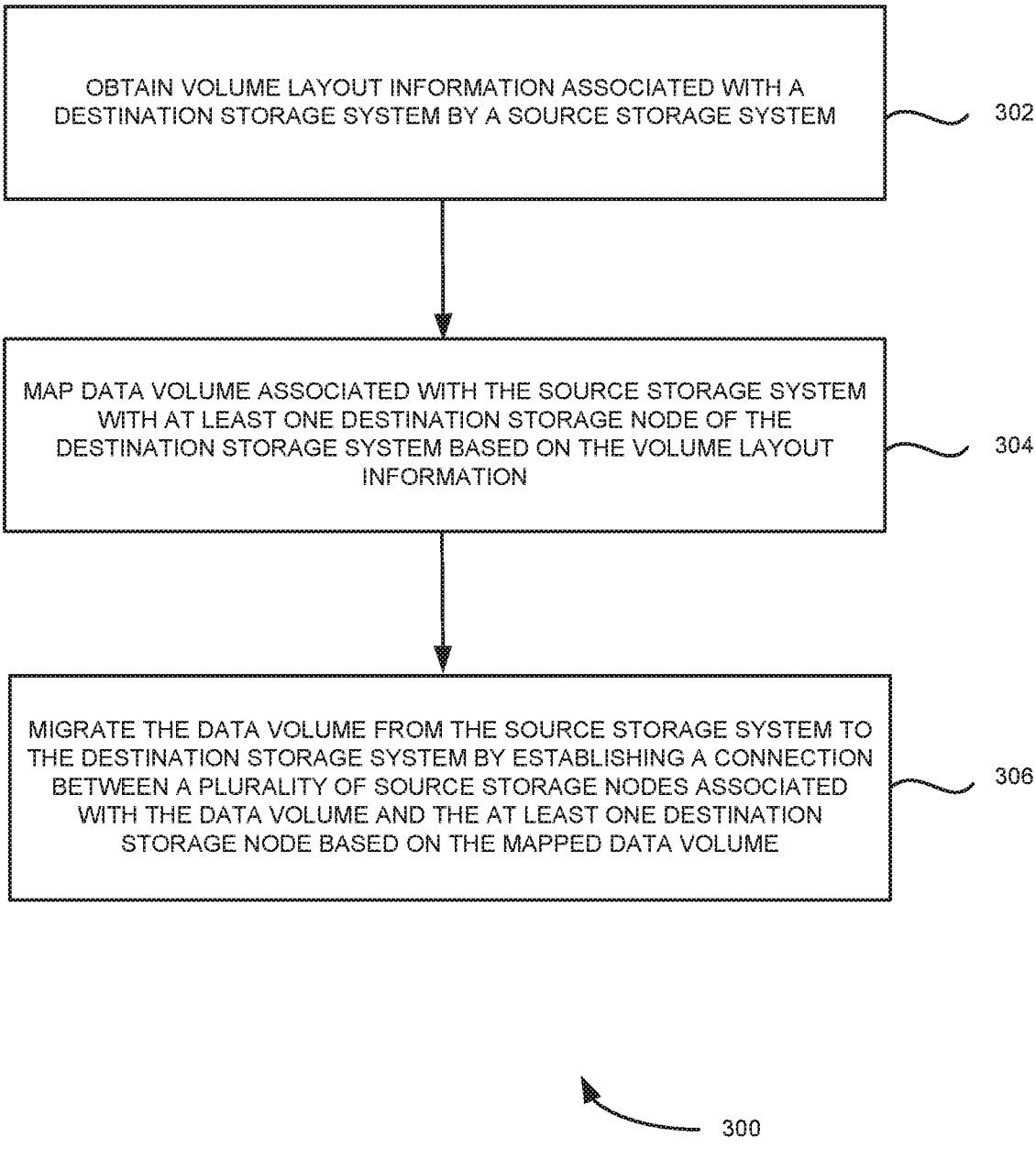
OBTAIN VOLUME LAYOUT INFORMATION ASSOCIATED WITH A DESTINATION STORAGE SYSTEM BY A SOURCE STORAGE SYSTEM
302
MAP DATA VOLUME ASSOCIATED WITH THE SOURCE STORAGE SYSTEM WITH AT LEAST ONE DESTINATION STORAGE NODE OF THE DESTINATION STORAGE SYSTEM BASED ON THE VOLUME LAYOUT INFORMATION
304
MIGRATE THE DATA VOLUME FROM THE SOURCE STORAGE SYSTEM TO THE DESTINATION STORAGE SYSTEM BY ESTABLISHING A CONNECTION BETWEEN A PLURALITY OF SOURCE STORAGE NODES ASSOCIATED WITH THE DATA VOLUME AND THE AT LEAST ONE DESTINATION STORAGE NODE BASED ON THE MAPPED DATA VOLUME
306
300

FIG. 3

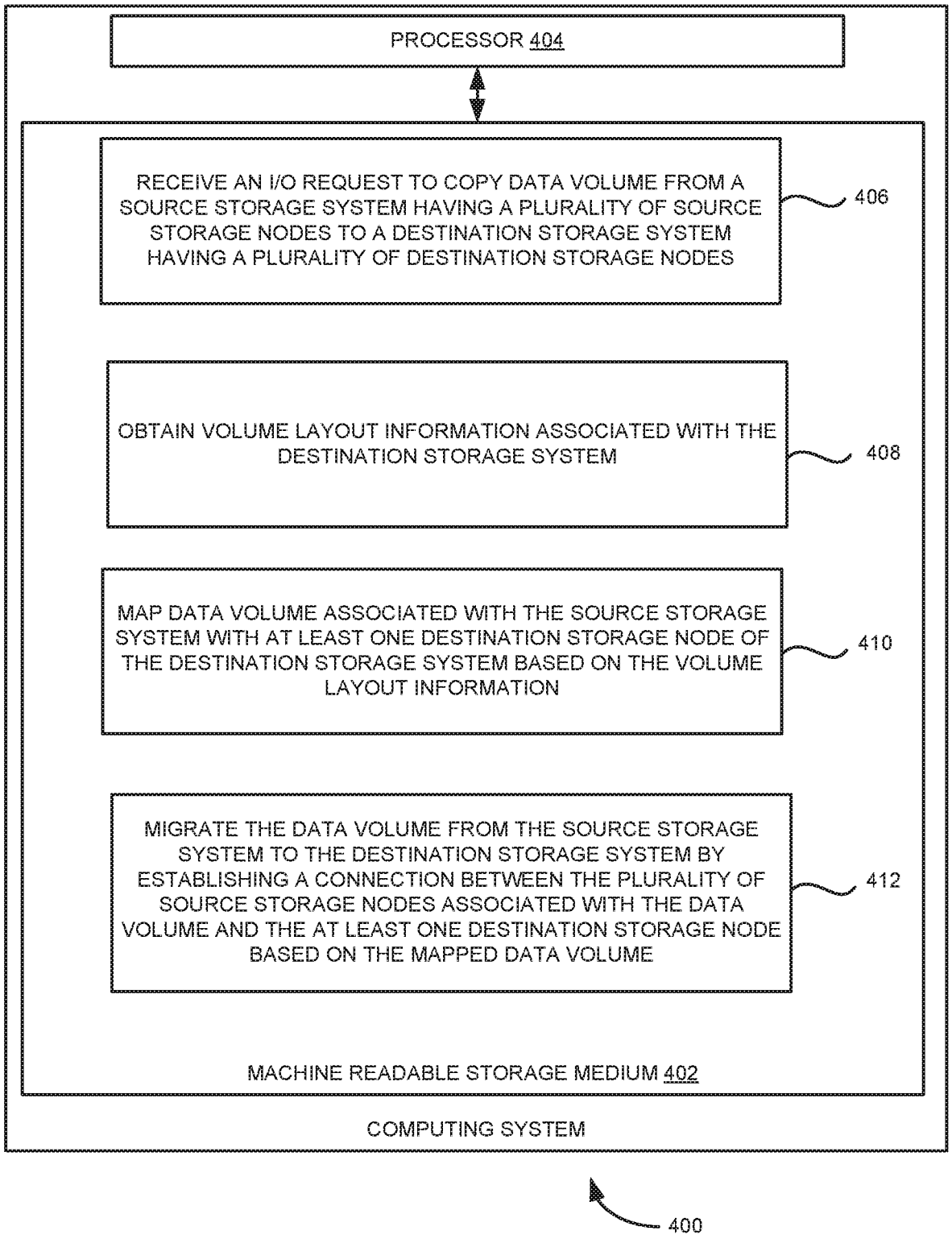
PROCESSOR 404
RECEIVE AN I/O REQUEST TO COPY DATA VOLUME FROM A SOURCE STORAGE SYSTEM HAVING A PLURALITY OF SOURCE STORAGE NODES TO A DESTINATION STORAGE SYSTEM HAVING A PLURALITY OF DESTINATION STORAGE NODES
406
OBTAIN VOLUME LAYOUT INFORMATION ASSOCIATED WITH THE DESTINATION STORAGE SYSTEM
408
MAP DATA VOLUME ASSOCIATED WITH THE SOURCE STORAGE SYSTEM WITH AT LEAST ONE DESTINATION STORAGE NODE OF THE DESTINATION STORAGE SYSTEM BASED ON THE VOLUME LAYOUT INFORMATION
410
MIGRATE THE DATA VOLUME FROM THE SOURCE STORAGE SYSTEM TO THE DESTINATION STORAGE SYSTEM BY ESTABLISHING A CONNECTION BETWEEN THE PLURALITY OF SOURCE STORAGE NODES ASSOCIATED WITH THE DATA VOLUME AND THE AT LEAST ONE DESTINATION STORAGE NODE BASED ON THE MAPPED DATA VOLUME
412
MACHINE READABLE STORAGE MEDIUM 402
COMPUTING SYSTEM
400

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/020260

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(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 17/30; G06F 12/02; G06F 11/14; G06F 9/455; G06F 7/04

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: migrating, storage, plurality, volume layout information, directly, and similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009-0077336 A1 (TSUKASA SHIBAYAMA et al.) 19 March 2009 See paragraphs [0059], [0062], [0071], [0075], [0080]-[0086], [0090]-[0097], [0105], [0114], [0122], [0129], [0139], [0181], and [0204]-[0207]; claims 1 and 12; and figures 1-2, 4, 6, 8, and 10.	1-15
A	US 8,473,463 B1 (TOMASZ WILK) 25 June 2013 See column 3, lines 4-41; and figures 1 and 4.	1-15
A	US 2012-0284707 A1 (VISWESVARAN JANAKIRAMAN) 08 November 2012 See paragraphs [0040]-[0042] and figure 2.	1-15
A	US 2015-0127975 A1 (DATRIUM INC.) 07 May 2015 See paragraphs [0015]-[0016] and figure 2.	1-15
A	WO 2013-179171 A1 (INTERNATIONAL BUSINESS MACHINES CORP.) 05 December 2013 See paragraphs [036]-[037] and [058]; and figures 1 and 4-5.	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
30 June 2016 (30.06.2016)	**30 June 2016 (30.06.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
NHO, Ji Myong
Telephone No. +82-42-481-8528

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2016/020260

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009-0077336 A1	19/03/2009	JP 2007-286709 A	01/11/2007
		JP 4900784 B2	21/03/2012
		US 2007-0245110 A1	18/10/2007
		US 2010-0332784 A1	30/12/2010
		US 2011-0252213 A1	13/10/2011
		US 2012-0137099 A1	31/05/2012
		US 7469325 B2	23/12/2008
		US 7805585 B2	28/09/2010
		US 7996640 B2	09/08/2011
		US 8140802 B2	20/03/2012
		US 8364925 B2	29/01/2013
US 8473463 B1	25/06/2013	None	
US 2012-0284707 A1	08/11/2012	US 555278 B2	08/10/2013
US 2015-0127975 A1	07/05/2015	None	
WO 2013-179171 A1	05/12/2013	AU 2013-269206 A1	13/11/2014
		CN 104335188 A	04/02/2015
		EP 2856322 A1	08/04/2015
		EP 2856322 A4	17/06/2015
		JP 2015-519667 A	09/07/2015
		KR 10-2014-0136473 A	28/11/2014
		SG 11201404021W A	28/08/2014
		US 2013-0326182 A1	05/12/2013