

(43) International Publication Date  
27 November 2014 (27.11.2014)

- (51) International Patent Classification:  
*G06F 12/00* (2006.01) *G06F 13/00* (2006.01)
- (21) International Application Number:  
PCT/US2014/031256
- (22) International Filing Date:  
19 March 2014 (19.03.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
13/901,333 23 May 2013 (23.05.2013) US
- (71) Applicant: **NETAPP, INC.** [US/US]; 495 East Java Drive, Sunnyvale, California 94089 (US).
- (72) Inventor: **EL-BATAL, Mohamed**; 3664 W. Torreys Peak Drive, Superior, Colorado 80027 (US).
- (74) Agents: **KELTON, Thomas W.** et al.; Haynes and Boone, LLP, 2323 Victory Avenue, Suite 700, Dallas, Texas 75219 (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, LT, 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, 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).

[Continued on next page]

(54) Title: SYSTEMS AND METHODS INCLUDING AN APPLICATION SERVER IN AN ENCLOSURE WITH A COMMUNICATION LINK TO EXTERNAL CONTROLLER

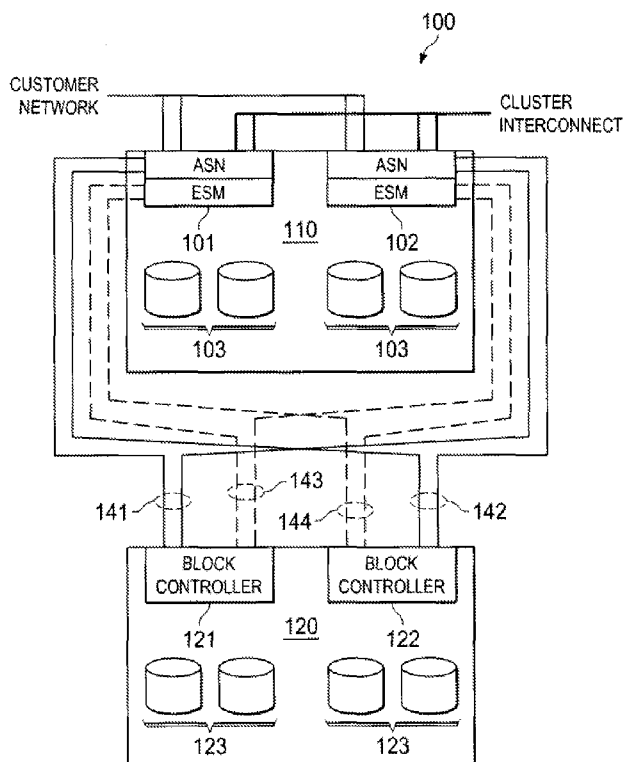


Fig. 1

(57) Abstract: A system including a first array of storage drives in communication with a storage controller, the storage controller providing virtualized storage with the first array of storage drives; an application server running a storage warehouse application and communicatively coupled to a host device and the storage controller; and a communication port expander providing expansion slots to a first enclosure, wherein the first array of storage drives, the application server, and the communication port expander are contained in the first enclosure separate from the controller, further wherein the controller controls the first array of storage drives and is in communication with the first array of storage drives through the communication port expander.



---

**Published:**

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

SYSTEMS AND METHODS INCLUDING AN APPLICATION SERVER  
IN AN ENCLOSURE WITH A COMMUNICATION LINK  
TO AN EXTERNAL CONTROLLER

5

**TECHNICAL FIELD**

The present description relates to network storage systems and, more specifically, to systems and methods in which application server hardware and software are included  
10 with communication links to a storage array controller in a separate enclosure.

**BACKGROUND**

Conventional network storage systems may be implemented using the Storage Bridge Bay standard. The Storage Bridge Bay (SBB) specification is a standard for  
15 storage solution providers who desire to have a standard controller and slot compatibility between respective controllers which can be manufactured by different organizations. The SBB specification describes the physical requirements for storage canisters and the mid-plane connectors of a storage enclosure. The storage enclosure can house one or more storage canisters. Meanwhile, each storage canister houses at least one printed circuit  
20 board having electronics mounted thereon. A current version of the SBB specification, as of this writing, is Storage Bridge Bay Specification, Version 2.0 released Jan. 28, 2008, the entire contents of which are hereby incorporated by reference.

Storage enclosures are conventionally designed to fit in a standard rack defined in EIA-310. A rack unit (or "U") is a measurement of height within the rack, where one rack  
25 unit equals 44.45mm or 1.75 inches. A storage or computing device that is embodied within an enclosure that it is one rack unit tall may be referred to as a "1U" device. Similarly, a storage or computing device that is embodied in an enclosure that is two rack units tall may be referred to as a 2U device. A given rack can be any height, but is usually 19 or 23 inches tall.

30 A conventional network application system includes two application servers in different 2U enclosures, a RAID Bunch of Drives (RBOD) in a 2U enclosure, and an Expansion Bunch of Drives (EBOD) in a 2U enclosure, for a total of 8U taken up by four enclosures. The RBOD includes two storage controllers, and the EBOD includes two

expansion Enclosure Service Modules (ESMs). However, conventional techniques for configuring a network storage system may not be as efficient as could be desired.

## BRIEF DESCRIPTION OF THE DRAWINGS

5           The present disclosure is best understood from the following detailed description when read with the accompanying figures.

Figure 1 is a system diagram of a computing infrastructure including a network storage system according to aspects of the present disclosure.

10           Figure 2 is an illustration of an example circuit board of Figure 1, adapted according to one embodiment.

Figure 3 is an illustration of an example mechanical and cabling diagram for the system of Figure 1.

Figure 4 is a system diagram of a computing infrastructure including a network storage system according to one embodiment.

15           Figure 5 is an illustration of an example circuit board, adapted according to the embodiment of Figure 4 to include a dense server ASN and ESMs.

Figure 6 is a system diagram of a computing infrastructure including a network storage system according to aspects of the present disclosure.

20           Figure 7 is a system diagram of a computing infrastructure including a network storage system according to aspects of the present disclosure.

Figure 8 is an illustration of example rack, adapted according to one embodiment.

Figure 9 is a block diagram of a method of installing a network storage system according to aspects of the present disclosure.

25           Figure 10 is an illustration of an example RBOD, adapted according to one embodiment.

Figures 11 and 12 are illustrations of methods to provide virtual storage to a network storage system, according to one embodiment.

## DETAILED DESCRIPTION

30           All examples and illustrative references are non-limiting and should not be used to limit the claims to specific implementations and embodiments described herein and their equivalents. The headings are solely for formatting and should not be used to limit the subject matter in any way, because text under one heading may cross reference or apply to text under one or more headings. Finally, in view of this disclosure, particular features

described in relation to one aspect or embodiment may be applied to other disclosed aspects or embodiments of the disclosure, even though not specifically shown in the drawings or described in the text.

The present disclosure relates to network storage systems and methods of  
5 implementing network storage systems that provide for more efficient use of space in a computer housing (e.g., a rack). In one example, a network storage system includes an Application Server Node (ASN) in a same enclosure with an ESM, where the ESM includes expansion ports. The enclosure also includes a first array of storage drives. A different enclosure includes a storage controller providing virtualized storage functionality  
10 for the array of storage drives. The storage controller communicates with the array of storage drives via the expansion ports of the ESM.

Thus, in this example, even though the array of storage drives is located in the same enclosure as the ASN, it is not the processing power of the ASN that it used as a storage controller. Rather, the storage controller functionality is provided by a storage  
15 controller in a separate RBOD enclosure that takes advantage of the expansion ports of the ESM to communicate with the storage array.

In one example, the enclosure that includes the storage controller also includes another array of storage drives that are also controlled by the storage controller. The first enclosure and the second enclosure each occupy 2U in the rack, thereby providing a 4U  
20 storage system. By contrast, a conventional network storage system would use separate enclosures for an ASN and an ESM, thereby using more units in the rack.

In another example, an enclosure includes a High Availability (HA) pair of ASNs, each of the ASNs including an ESM with expander ports. The enclosure may be still be embodied as a 2U enclosure. The space savings may be achieved by embodying  
25 the ESMs as Peripheral Component Interconnect Express (PCIe) cards that occupy card slots of the ASN hardware. However, the scope of embodiments is not limited to PCIe cards as ESMs, as any appropriate connection between the ASN hardware and the ESM hardware may be used.

The block controllers may also be embodied as an HA pair, where each block  
30 controller is redundantly connected to each ASN and to each ESM. In fact, the architecture described herein may be adapted for use in clustered systems, as shown in the figures.

In yet another example, the ASN is embodied as a dense server, which has two or more processors and may use more room on a circuit board than a server that has only one

processor. An example of a dense server is one that uses two or four Intel™ Xeon™ processors, though the scope of embodiments is not limited to any particular type or number of processors. In some embodiments employing dense servers as ASNs, there may not be enough space within a 2U enclosure to house two or more redundant ASNs. Thus, some dense server embodiments may include only one ASN in a given enclosure, along with at least one ESM. Redundancy can then be achieved by using an additional ASN/ESM enclosure and making redundant connections between the storage controller and each of the ASNs.

The various embodiments described above are examples, and it is understood that the scope of embodiments includes storage systems with different numbers of controllers, ASNs, and storage arrays. In fact, any of the embodiments described herein may be scaled larger or smaller depending on particular needs of a given implementation.

Figure 1 is a system diagram of a computing infrastructure 100 including a network storage system according to aspects of the present disclosure. Infrastructure 100 includes enclosure 110 and enclosure 120, which may be adapted according to the EIA-310 standard to fit into standard racks. However, the scope of embodiments is not limited to standard enclosures, and any type of enclosure, whether standard or otherwise, may be adapted for use in some embodiments.

Enclosure 100 includes circuit boards 101 and 102, as well as storage array 103. In this example, circuit board 101 includes hardware and software to provide ASN and ESM functionality, as does circuit board 102. The ASNs of circuit boards 101 and 102 are configured as a HA pair.

The ASNs of circuit boards 101 and 102 are in communication over the customer network with one or more host devices (not shown). The ASNs of circuit boards 101 and 102 are also in communication with a cluster interconnect, which may provide communication with other devices in a cluster (e.g., additional ASNs, not shown). Connections to the customer network and cluster interconnect may include Ethernet, Infiniband, IEEE 802.11, and/or any other appropriate communication link.

The ASNs of Figure 1 host storage applications that are made available to host systems that access to the storage system. For example, some ASNs run applications to provide: continuous data protection through automated backup procedures, database management application processes, snapshot management processes, the de-duplication management processes, and the like. Other examples of applications include a cluster-based file system program (e.g., a Lustre™ application), a high-performance computing

application (a data warehousing application, a video delivery and processing application, an Oracle™ database application, etc.), a big data application (e.g., Hadoop™ or other map-reduce application), and the like.

In some examples, the ASNs utilize Virtual Machine Managers (commonly referred to as "hypervisors") to provide a virtual machine for each of the multiple application processes. In general, a hypervisor controls the overall operation of each of a plurality of virtual machines. Each virtual machine may include its own specific operating system kernel and associated application processes such that the hypervisor hides the underlying physical hardware circuitry interfaces from the operating system and application processes operating within a virtual machine.

The ASNs of Figure 1 may use any appropriate operating system or hypervisor, but in this example, the ASN may include one or more general purpose processors (e.g., Intel Xeon or other processors) running a data warehousing program (e.g., a map reduce program) on top of Linux.

The ESMs of Figure 1 include hardware and software functionality to manage the physical operation of hardware within the storage system. Examples of functionality include environment reporting and setting via SCSI enclosure services, such as temperature control, power management, and the like. The ESMs may include their own processing power and memory suitable to provide such functionality (e.g., by an Application Specific Integrated Circuit or Field Programmable Gate Array), and in many instances, the amount of processing power and memory used in an ESM is less than that used in an application server.

Further, in this example, the ESMs of Figure 1 include expander ports for connecting further devices within the cluster. Before discussing the expander ports further, a discussion of the data transfer links may be instructive. System 100 may be implemented with serial attached SCSI (SAS) architecture or a like architecture suitable for supporting full duplex data transfer. SAS may refer to a point to point serial interface suitable for providing full duplex communication at three, six, or twelve gigabits per second (Gb/s) or higher per direction per channel. Protocols that may communicate via a SAS interface may be small computer system interface (SCSI), advanced technology attachment (ATA), SAS management or like protocols.

Expander devices included in the ESMs of Figure 1 may include edge expander devices, edge expander device sets, and fanout expanders. SAS devices, such as end devices and expander devices, may be interconnected to deliver scalable storage in a SAS

environment. SAS devices are typically interconnected with a set of point-to-point links in the SAS domain. SAS devices may also include at least two connections for enhanced performance and availability. In an interconnection of SAS devices or SAS domain, each logical connection to a node is made via a port. A port may include a plurality of point-to-point links, which may be denoted as phys. For instance, a typical media device, such as a storage drive, may comprise two ports with each port including at least one phy. Additionally, multiple phys may be arranged together to make up a port, allowing for multiple concurrent connections to be established. In some embodiments, the expanders included in the ESMs of Figure 1 include edge expander devices or like expander devices suitable for communication with up to 128 SAS addresses. An edge expander device enables communication to be established between nodes that are directly connected to the edge expander device.

Nevertheless, as mentioned above, the scope of embodiments is not limited to SAS. Other suitable data transfer links may be appropriate, such as Fibre Channel (FC), Ethernet, Infiniband, IEEE 802.11, and the like.

Moving to enclosure 120, block controllers 121 and 122 are embodied in circuit boards housed within enclosure 120 along with storage drive array 123. Block controllers 121 and 122 in this example are configured as a HA pair, though the scope of embodiments is not limited just to HA pairs.

Each controller 121, 122 includes one or more processors or application specific logic that causes connected storage arrays to read or write data to or from persistent storage in response to requests received from an external host. Each controller 121, 122 may also include memory that acts as a buffer or temporary storage for data being transferred between the host and persistent storage. Each controller 121, 122 may further include host input/output (I/O) controllers (HICs) that interface the storage array controller with the host, and target I/O controllers (TICs) that interface the storage array controller with the storage drives. The block controllers 121, 122 may include processing power and memory suitable to virtualize storage and control reads and writes (e.g., by an Application Specific Integrated Circuit or Field Programmable Gate Array), and in many instances, the amount of processing power and memory used in a block controller 121, 122 is less than that used in an application server.

In one example, each block controller 121, 122 virtualizes the storage within storage arrays 103, 123 to provide RAID storage. The acronym "RAID" is an umbrella term for computer data-storage schemes that can divide and replicate data among multiple

physical storage drives. When several physical storage drives are set up to use RAID technology, the storage drives are said to be in a RAID array. The RAID array distributes data across several storage drives, but the array is exposed to the operating system at the application server as a single logical storage drive or multiple logical storage drives, where the logical storage drives do not correspond one-to-one with the physical storage drives. Examples of storage drives that may be used in various embodiments include Solid State Drives (SSDs), Hard Disk Drives (HDDs), and the like.

Enclosure 120 may be referred to as a RAID Bunch of Drives or (RBOD). In one example, block controllers 121, 122 are essentially identical storage controllers coupled with one another to provide redundancy in case of failure of one. Figure 10 provides a more detailed illustration of RBOD 120 according to one embodiment. In Figure 10, controller 121 has a circuit board 191 with a CPU, an SAS I/O controller, and an SAS expander. Similarly, controller 122 has a circuit board 192 with a CPU, an SAS I/O controller, and a SAS expander. The controllers 121, 122 are cross-coupled via their SAS I/O controllers. Further, each of the storage drives in array 123 is communicatively coupled to each controller 121, 122. Each of the RBODs shown in the various embodiments herein may include cross-coupled controllers, as shown in Fig. 10. In the exemplary embodiment of FIG. 2, block controllers 121, 122 are interconnected via a path (not shown), which may provide any of a variety of well-known communication protocols and media including, for example, PCI (e.g., PCI Express), SAS, Fibre Channel, Infiniband, Ethernet, IEEE 802.11, etc. This inter-controller interface and medium is utilized for exchanges between the controllers 121, 122 within the storage enclosure 120.

The various components of system 100 are communicatively coupled as follows. Block controller 121 is connected to the ASNs of circuit boards 101 and 102 via communication link 141. Thus, block controller 121 is connected to both ASNs. Similarly, block controller 122 is connected to the ASNs via communication links 142. Communication links 141 and 142 may include any appropriate type of communication link, including but not limited to SAS, Fibre Channel, Infiniband, Ethernet, IEEE 802.11, and the like.

Block controller 121 is connected to the ESMs of circuit boards 101 and 102 via communication links 143. Thus, block controller 121 is connected to both of the redundant ESMs in enclosure 110. Similarly, block controller 122 is connected to the ESMs by communication links 144. Communication links 143 and 144 may include any appropriate

type of communication link, including but not limited to SAS, Fibre Channel, Infiniband, Ethernet, IEEE 802.11, and the like.

Block controllers 121 and 122 communicate with, and control, storage drive array 123 through a midplane interface (not shown) or other interface within enclosure 120.

5 Furthermore, block controllers 121, 122 communicate with, and control, storage drive array 103 (in enclosure 110) via communication links 143, 144 and the EMS. In other words, storage drive array 103 in enclosure 110 is in communication with controllers 121, 122, which are external to enclosure 110, through the ESMs that are included within enclosure 110.

10 The embodiment shown in Figure 1 may provide one or more advantages over conventional network storage systems. Combining ASN and ESM hardware into the same enclosure may provide space savings by creating higher density of storage drives. Thus, the ratio of number of rack units used for storage versus the number of rack units used for processing can be increased. Increasing such ratio may be desirable because, as processing  
15 power increases, the ability to control storage drive arrays using a same size circuit board also increases; but as the number of rack units for storage increases, there is less space for processing power in the rack.

Figure 2 is an illustration of an example circuit board 101 of Figure 1, adapted according to one embodiment. It is understood that in various embodiments, system 100  
20 (Figure 1) may include circuit board 102, which is substantially similar to circuit board 101. It should also be noted that the specific implementation shown in Figure 2 is an example, and other embodiments with different types and numbers of I/O ports, processing power, memory, and the like may be adapted for use. In one example, the hardware of a RAID controller may be used as the hardware of circuit board 101. For instance, if a  
25 RAID controller board has sufficient processing hardware, its RAID controller software and firmware may be replaced with application server hardware and firmware to provide application server functionality in any of the embodiments of Figures 1, 6, or 7. In other words, the embodiment of Figure 2 shows a modified RAID controller board used as an ASN. However, any appropriate processing hardware may be used in any of the  
30 embodiments.

Dashed line 201 shows a division between the ASN hardware and the ESM hardware. Labels to the left side of circuit board 101 show that some ports communicate with the cluster interconnect, the host network, and the controllers 121 and 122. Among

other things the ASN hardware includes CPU 210, HIC card 211, and SAS input/output control 212 (to communicate with controllers 121 and 122).

The ESM hardware can be embodied as a PCIe card, mounted directly on board 101, or any other appropriate configuration. Among other things, the ESM hardware  
5 includes expander 213, which provides ports for communication with controllers 121 and 122. The communication with controllers 121 and 122 is passed through connection 214 to midplane connector 215 and to storage array 103. Thus, as explained above, controllers 121 and 122 control storage array 103 through the expander 213 provided by the ESM within enclosure 110. Storage array 103 in this example is not implemented on  
10 circuit board 101, but is instead implemented on a different circuit board in communication with midplane connector 215. Glue logic for the ESM is provided in this example by FPGA 216, which provides less processing power than does CPU 210.

Figure 3 is an illustration of an example mechanical and cabling diagram for the system of Figure 1. Enclosures 110 and 120 are shown as standard 2U rack-mounted  
15 enclosures. Fig. 3 also shows communication links 141-144 with respective SAS ports, and this arrangement is consistent with the communication link layout shown in Figure 1. Ports labeled "IB" are Infiniband ports, and ports labeled 10G are Ethernet ports. The boxes labeled "AC" are example power supply units. It is noted that in practice a rack mounted enclosure may have different types and numbers of ports than those shown in Fig.  
20 3. Also, as noted above, the scope of embodiments is not limited to SAS communication links, as other embodiments may use Ethernet, Infiniband, IEEE 802.11, and/or Fiber Channel. The example of Figure 3 is consistent with the example of Figure 2 in that the ASN is built using modified RAID controller hardware. However, the scope of embodiments is not limited to using RAID controller hardware in the ASNs.

Figure 4 is a system diagram of a computing infrastructure 400 including a  
25 network storage system according to aspects of the present disclosure. The embodiment of Figure 4 utilizes dense servers for ASNs and, therefore, utilizes one dense server per enclosure. As explained above, a dense server may include more processors and memory than the ASNs shown in Figure 1 and may utilize more space within a given enclosure.  
30 For instance, enclosure 410 includes a circuit board including a dense server ASN 411 and two ESMs 412, 414. Similarly, enclosure 420 includes a dense server ASN 421 and ESMs 422, 424. ASNs 411 and 421 are configured as a HA pair via the cluster interconnect and are both connected to hosts (not shown) via the customer network. Each enclosure 410, 420 also includes a storage drive array 413 and 423, respectively. ASNs 411 and 421 run a

storage warehousing application or other application on top of an operating system, similar to that described above with respect to the ASNs of Figure 1. However, ASNs 411 and 421 may handle more concurrent processes due to their increased processing power.

Enclosure 430 includes two block controllers 431 and 432 and a storage drive  
5 array 433. Controllers 431 and 432 of Figure 4 operate substantially similarly to controllers 121 and 122 of Figure 1 by providing low-level virtual storage control to storage drives 413, 423, and 433.

ASN 411 has redundant communication links 441 and 442 to each of block  
controllers 431 and 432. Similarly, ASN 421 has redundant communication links 443, 444  
10 to each of block controllers 431 and 432. Such communication links may be substantially similar to links 141 and 142 of Figure 1.

Each ESM 412, 414, 422, 424 has a respective communication link 445-448 to a particular block controller 431, 432, as shown. Communication links 445-448 are substantially similar to the communication links 143, 144 of Figure 1. Controller 431  
15 communicates with, and controls, storage arrays 413 and 423 via respective ESMs 412 and 422. Similarly, block controller 432 communicates with, and controls, storage arrays 413 and 423 via respective ESMs 414 and 424.

Figure 5 is an illustration of two example ASNs 520 and 530, adapted according to the embodiments herein. Each ASN 520, 530 is included in its own circuit board within  
20 its own enclosure and is in communication with midplane 550 through midplane cards 505, 505. ASNs 520 and 530 are provided in contrast to the circuit board of Figure 2, which includes RAID controller hardware. ASN 520 includes CPU 502, and ASN 530 includes CPU 503, though the scope of embodiments may include any number of CPUs in each of the ASNs. Figure 5 illustrates one example configuration to accommodate ESMs 524 and  
25 534 in a same enclosure with respective ASNs 520 and 530.

Each ASN 520, 530 includes its respective ESMs 524 and 534 as PCIe cards, though various embodiments may implement the hardware of ESMs 524 and 534 in any appropriate manner. ESM 524 communicates with block controllers (not shown) and passes that communication to midplane card 504 and to storage array 513. Similarly, ESM  
30 534 communicates with block controllers (not shown) and passes that communication to midplane card 505 and on to storage array 513. Each ESM 524 and 534 has a respective expander 512, 514 to provide communication links to the controllers. In some instances, PCI is used for providing power only, while the data connections between a server and an ESM are provided by technologies other than PCI.

In one example, both ASNs 520, 530 may be implemented side-by-side in 2U so that the total rack unit space occupied thereby is 2U, though the scope of embodiments includes any sizes for enclosures ASNs 520, 530.

With respect to the embodiment of Figure 4, each ASN 411, 421 may be  
5 implemented similarly to ASN 520 but with at least one more CPU and two ESMs in each ASN. With respect to the embodiments of Figures 1, 6, and 7, each ASN enclosure therein (e.g., enclosure 110) may be implemented using the system shown in Figure 5 in which the ASNs 520, 530 are in a HA pair and may be optionally cross-coupled.

Figure 6 is a system diagram of a computing infrastructure 600 including a  
10 network storage system according to aspects of the present disclosure. Figure 6 includes many of the same features shown in Figure 1, and like features are shown with like reference numbers.

System 600 includes enclosure 110 that has ASNs configured as a HA pair. Block controller 121 communicates with, and controls, storage array 103 via  
15 communication link 623 and the ESM of circuit board 101. Similarly, block controller 122 communicates with, and controls, storage array 103 via communication link 624 and the ESM of circuit board 102.

System 600 also includes enclosure 610. Enclosure 610 has ESMs 611 and 612 as well as storage array 613. Each of ESMs 611 and 612 has a respective expander that  
20 provides communication ports to provide communication with block controllers 121 and 122 via communication links 621, 622.

Block controller 121 communicates with, and controls, storage array 613 via communication link 621 and ESM 611. Block controller 122 communicates with, and controls, storage array 613 via communication link 622 and ESM 612. The embodiment of  
25 Figure 6 illustrates that additional storage arrays (e.g., array 613) may be added using expanders as long as the block controllers of the system have adequate communication ports and processing power. Communication links 621-624 may include any appropriate type of communication link, including but not limited to SAS, Fibre Channel, Infiniband, Ethernet, IEEE 802.11, and the like.

Figure 7 is a system diagram of a computing infrastructure 700 including a  
30 network storage system according to aspects of the present disclosure. Figure 7 includes many of the same features shown in Figures 1 and 6, and like features are shown with like reference numbers.

The ASNs of enclosures 110 and 610 are connected to the customer network, as shown, and to each other via the cluster interconnect. In addition to the concepts shown in Figure 6, Figure 7 illustrates that at least four ASNs can be accommodated in two different enclosures and that storage drives in those enclosures can be communicatively coupled to  
5 controllers via ESMs with the ASNs. Put another way, the ESMs of Figure 7 provide communication ports so that storage arrays 103 and 613 can be included in the same enclosures as ASNs all while using external controllers.

The embodiment shown in Figure 7 is similar to the embodiment shown in Figure 6, but the embodiment of Figure 7 adds ASNs to enclosure 610. In fact, in Figure 7, the  
10 components and functionality within enclosure 610 are substantially similar to those within enclosure 110. Thus, ESM 611 is included in circuit board 701, and ESM 612 is included in circuit board 702. Block controllers 121, 121 include redundant cross-coupling to the ASNs of circuit boards 701 and 702 via communication links 741 and 742. Once again, block controllers 121 and 122 communicate with, and control, storage array 613 through  
15 respective ESMs 611 and 612.

In an exemplary embodiment, the various enclosures discussed above may be installed in a standard rack. Figure 8 is an illustration of example rack 801, adapted according to one embodiment. In the example of Figure 8, the system 700 (Figure 7) has been installed in the rack 801, and each of enclosures 110, 120, 610 is embodied as a 2U  
20 unit, thereby occupying 6U total space.

Figure 8 does not show other units in rack 801 for simplicity of illustration, but it is understood that the scope of embodiments includes racks with any number of enclosures therein. It is also understood that any of the embodiments of Figures 1, 3, 4, 6, and 7 may be configured as standard rack-mounted enclosures and installed in a rack, such as rack  
25 810. However, the scope of embodiments is not limited to standard rack-mounted computing systems, as any appropriate configuration of enclosures (even outside of racks) may be adopted.

Figure 9 is a block diagram of a method 900 of installing a network storage system according to aspects of the present disclosure. It is understood that additional steps  
30 can be provided before, during, and after the method 900, and some of the steps described can be replaced or eliminated for other embodiments of the method 900.

Block 902 includes installing a first computing device in a housing with a plurality of slots. An example of a suitable housing includes a EIA-310 rack with slots that are each an integer number of units tall. The first computing device is placed in a first slot

of the housing, and the first computing device includes a first enclosure with an application server, a communication port expander, and an array of storage drives in communication with the expander. In one example, the enclosures of method 900 are standard enclosures that are each an integer number of rack units tall.

5           Block 904 includes installing a second computing device in the housing by placing the second computing device in a second slot. Further, the second computing device includes a second enclosure with a storage drive controller. In other words, the storage drive controller and the storage drive are included in separate enclosures.

10           Block 906 includes operably coupling the storage drive controller to the array of storage drives through the communication port expander and operably coupling the application server to the storage drive controller. The coupling of block 906 may include installing cables and, additionally or alternatively, setting up wireless communication links.

15           The scope of embodiments is not limited to EIA-310 standard rack housings and enclosures. Rather, various embodiments may include any appropriate housing or enclosure, whether standard or otherwise. Furthermore, the communication links among the various components may include any appropriate hardware interface and protocol. Various communication links that may be used with the embodiments herein include, but are not limited to, Ethernet, SAS, Infiniband, Fiber Channel, IEEE 802.11, and the like.

20           Figure 11 is a block diagram of a method 1100 of providing virtual storage according to aspects of the present disclosure. Method 1100 is from the perspective of the hardware and software within an enclosure, such as enclosure 110 (Figure 1). It is understood that additional steps can be provided before, during, and after the method 1100, and some of the steps described can be replaced or eliminated for other embodiments of  
25           the method 1100.

          Block 1102 includes interacting with a client computer to receive write requests for a virtual storage volume. In this example, the virtual storage volume is implemented on an array of storage drives within a first enclosure that includes the application server node.

30           As described above, an ASN provides higher-level storage functionality, such as by implementing a map reduce system or other system. The ASN interfaces with the client computer to receive write requests and to return responses to read requests.

          Further, as described above, the lower-level storage functionality, such as RAID control, is provided by the storage controller. Thus, in block 1104, the ASN passes the

write request to the storage controller. Further, in this example, the storage controller is provided in an enclosure that is separate from the enclosure that includes the ASN and the storage drives.

5 In block 1106, physical storage drives of the array receive control signals from the storage controller to effect the write request. The control signals are received via a communication port expander located in the enclosure with the ASN and the storage array.

Method 1100 is offered to illustrate a data path that exists in various systems according to one or more embodiments. Specifically, the physical storage drives of the array are included in an enclosure that is separate from the enclosure that includes the storage controller. Therefore, write requests received at the ASN during interaction with the client are passed on to the controller in the second enclosure, and the controller then provides low-level storage control signals to the storage drives in the first enclosure. Block 1108 illustrates that read requests are effected using the same data path. There is no requirement that write requests are performed before read requests; rather, write and read requests are performed in any appropriate order.

Figure 12 is a block diagram of a method 1200 of providing virtual storage according to aspects of the present disclosure. Method 1200 is from the perspective of the hardware and software of the storage controller, such as in enclosure 120 (Figure 1). It is understood that additional steps can be provided before, during, and after the method 1200, and some of the steps described can be replaced or eliminated for other embodiments of the method 1200.

At block 1202, the storage controller receives a write request from an ASN. In this example, the ASN performs data warehousing functionality, and the write request is in accordance with such functionality.

25 The write request is received via a communication port expander in a first enclosure with the ASN and is received by a storage controller in a second enclosure separate from the first enclosure. At block 1204, the block controller sends control signals from the controller to a storage array within the first enclosure to write data to a virtual volume implemented in the storage array. Such control signals are a part of the low-level storage functionality of the storage controller that provides, for example, RAID

30 functionality in the storage drive array.

At block 1206, the block controller processes read requests in the same manner that the write request is processed in blocks 1202 and 1204. Method 1200 illustrates a data path that includes a write or read request coming from a client and being received at an

ASN, then passed to a storage controller in a different enclosure. The storage controller then communicates with the physical disk drives in the ASN enclosure via a communication port expander to effect the read or write request.

For both the methods 1100 and 1200, many embodiments include storage drives  
5 in the controller's enclosure and in other enclosures as well as in the ASN enclosure. In such embodiments, the drives that are included in the same enclosure as the ASN are presented to the RAID controller via the port expander such that they become members of the total set of drives that are managed by the RAID controller. In other words, the RAID controller combines the drives in its enclosure with drives in an Expansion JBODs  
10 (EBOD) (e.g., 610 in figure 6, if applicable) and the drives in the ASN enclosure to create RAID volumes to be presented to the applications running on the ASN.

The foregoing outlines features of several embodiments so that those skilled in the art may better understand the aspects of the present disclosure. Those skilled in the art should appreciate that they may readily use the present disclosure as a basis for designing  
15 or modifying other processes and structures for carrying out the same purposes and/or achieving the same advantages of the embodiments introduced herein. Those skilled in the art should also realize that such equivalent constructions do not depart from the spirit and scope of the present disclosure, and that they may make various changes, substitutions, and alterations herein without departing from the spirit and scope of the present disclosure.

20

**CLAIMS**

1. A system comprising:  
a first array of storage drives in communication with a storage controller, the  
5 storage controller providing virtualized storage with the first array of storage drives;  
an application server running a storage warehouse application and  
communicatively coupled to a host device and the storage controller; and  
a communication port expander providing expansion slots to a first enclosure,  
wherein the first array of storage drives, the application server, and the communication  
10 port expander are contained in the first enclosure separate from the controller, further  
wherein the controller controls the first array of storage drives and is in communication  
with the first array of storage drives through the communication port expander.
2. The system of claim 1, wherein the communication port expander is  
provided by an Enclosure Service Module (ESM) within the first enclosure, the ESM  
15 providing power and temperature control to the application server.
3. The system of claim 2, wherein the application server and the ESM are  
included on a same circuit board within the enclosure, wherein the ESM comprises CPU  
resources separate from CPU resources of the application server.
4. The system of claim 1, wherein the communication port expander is  
20 included on a peripheral interconnect card in communication with the application server.
5. The system of claim 1, wherein the expander comprises an SAS expander.
6. The system of claim 1, wherein the application server is part of a High  
availability (HA) pair of application servers.
7. The system of claim 1, wherein the storage controller is part of a High  
25 Availability (HA) pair of storage controllers.
8. The system of claim 1, wherein the storage controller comprises a RAID  
controller.

9. The system of claim 1, wherein the first enclosure including the first array of storage drives, the communication port expander, and the application server comprises a 2U blade device, further wherein a second enclosure including a second array of storage drives and the controller comprises a 2U blade device installed in a same rack with the first enclosure.

10. A method comprising:  
installing a first computing device in a housing with a plurality of slots, the first computing device being placed in a first slot, wherein the first computing device comprises a first enclosure with an application server, a communication port expander, and an array of storage drives in communication with the expander;  
installing a second computing device in the housing, the second computing device being placed in a second slot, wherein the second computing device comprises a second enclosure with a storage drive controller; and  
operably coupling the storage drive controller to the array of storage drives through the communication port expander and operably coupling the application server to the storage drive controller.

11. The method of claim 10, wherein the housing comprises a EIA-310 standard rack, and wherein the first and second computing devices comprise EIA-310 standard enclosures.

12. The method of claim 10, wherein operably coupling the storage drive controller to the array of storage drives comprising providing a Serial Attached SCSI (SAS) communication link.

13. A system comprising:  
an array of storage drives in communication with a storage controller, the storage controller providing virtualized storage with the array of storage drives;  
an application server running a storage warehouse application and serving the storage controller; and  
an Enclosure Service Module (ESM) providing expansion slots to the array of storage drives, wherein the array of drives, the application server, and the ESM are contained in a same enclosure separate from the controller, further wherein the controller

controls the array of storage drives and is in communication with the array of storage drives through the ESM.

14. The system of claim 13, wherein the application server and the ESM are included on a same circuit board within the enclosure.

5 15. The system of claim 13, wherein the ESM is included as a peripheral interconnect card in communication with the application server.

16. The system of claim 13, wherein the expander comprises an SAS expander.

17. The system of claim 13, wherein the application server is part of a High availability (HA) pair of application servers.

10 18. The system of claim 13, wherein the storage controller is part of a High Availability (HA) pair of storage controllers.

19. The system of claim 13, wherein the storage controller comprises a RAID controller.

15 20. The system of claim 13, wherein the enclosure including the array of storage drives, the ESM, and the application server comprises a 2U blade device, further wherein another enclosure including the controller comprises a 2U blade device installed in a same rack as the enclosure.

21. A method for providing virtual storage for a client, the method comprising:  
at an application server node, receiving write requests from a client computer for a  
20 virtual storage volume, the virtual storage volume being implemented on an array of storage drives accessible by a first computing system implementing the application server node, the storage drives being physically located with the first computing system;

passing a first write request from the application server node to a storage controller, the storage controller implemented by a second computing system physically separate from  
25 the first computing system; and

receiving drive access instructions from the storage controller to the array of storage drives to effect the write request, where the drive access instructions are received via a communication path between the first computing system and the second computing system.

22. The method of claim 21, wherein the storage controller comprises RAID controller hardware, and wherein the application server node comprises a CPU executing a data warehousing program.

23. The method of claim 21 wherein the communication path comprises a  
5 communication port expander that is included in an Enclosure Service Module (ESM), and wherein the communication port expander comprises a Serial Attached SCSI (SAS) expander.

24. A method for providing virtual storage in a storage network, the method comprising:  
10 receiving a write request from an application server node, the write request being in accordance with data warehousing functionality of the application server node, the write request received via a communication port expander in a first enclosure with the application server node, the write request being received by a storage controller in a second enclosure separate from the first enclosure; and  
15 in response to the write request, sending control signals from the controller to a storage array within the first enclosure to write data to a virtual volume implemented in the storage array.

25. The method of claim 24, wherein the storage controller comprises a RAID controller.

20 26. The method of claim 24, further comprising:  
receiving a read request from the application server at the storage controller; and  
in response to the read request, sending control signals from the controller to the storage array to read data to the virtual volume.

1/10

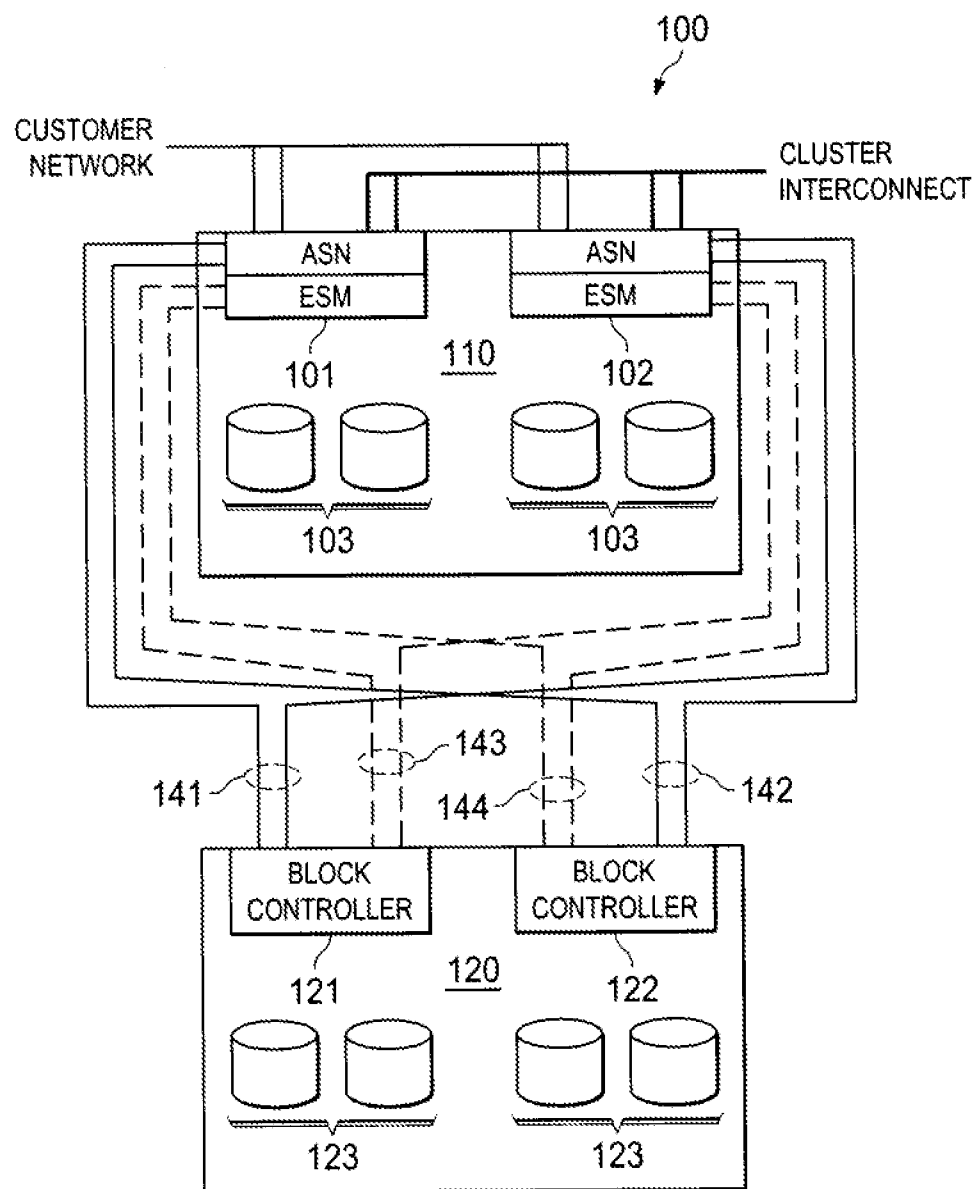
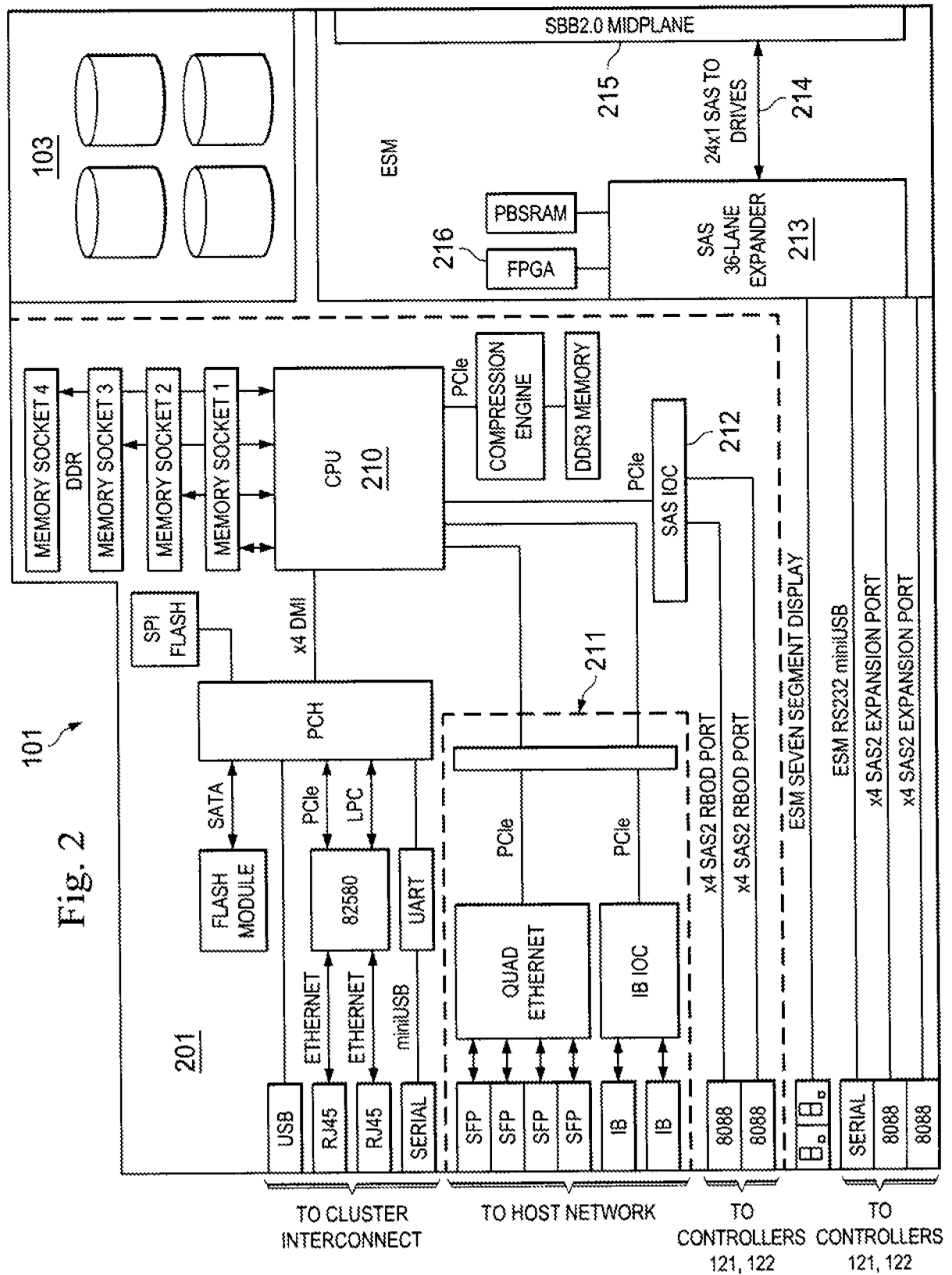


Fig. 1

2/10



3/10

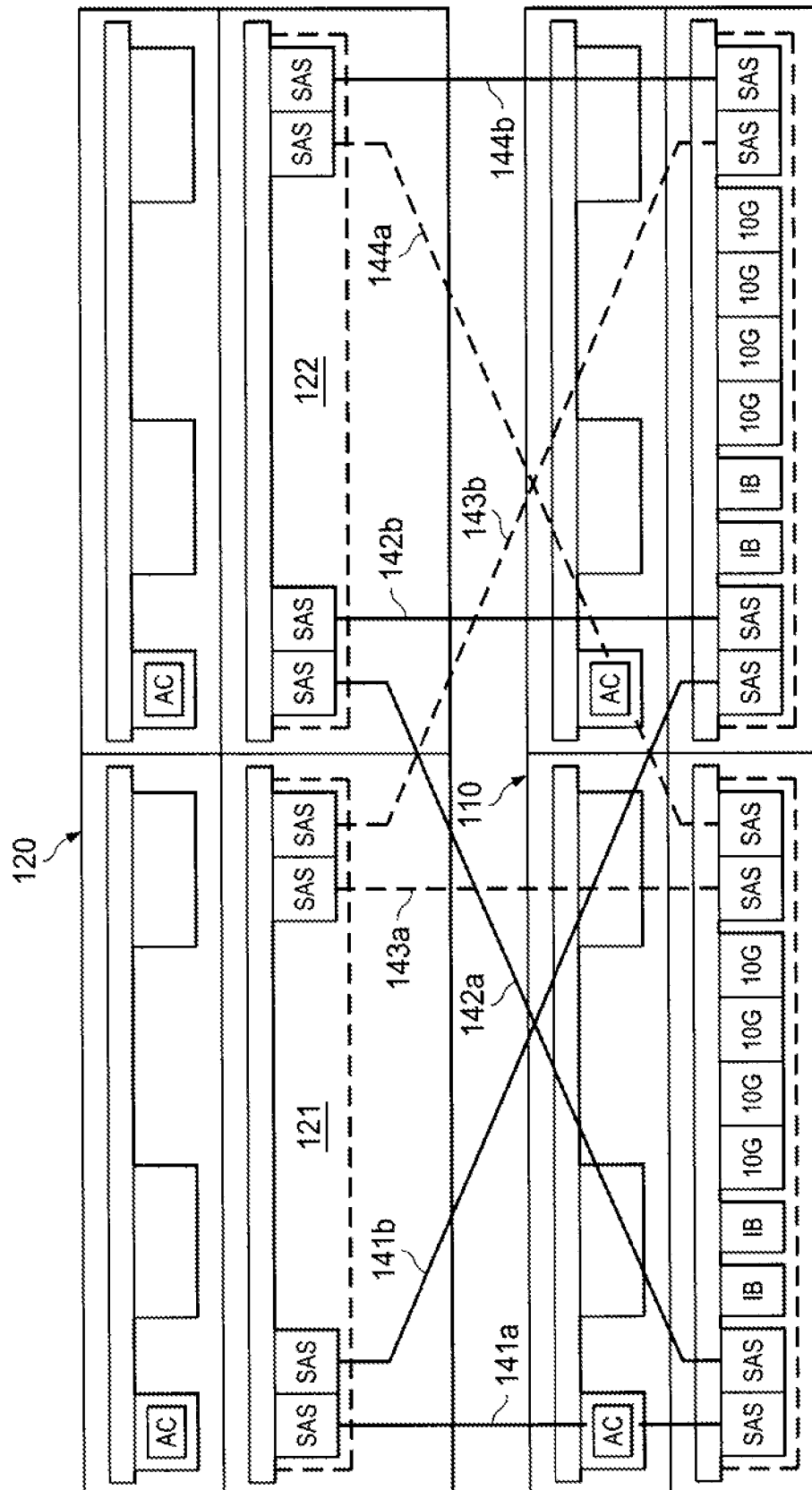


Fig. 3

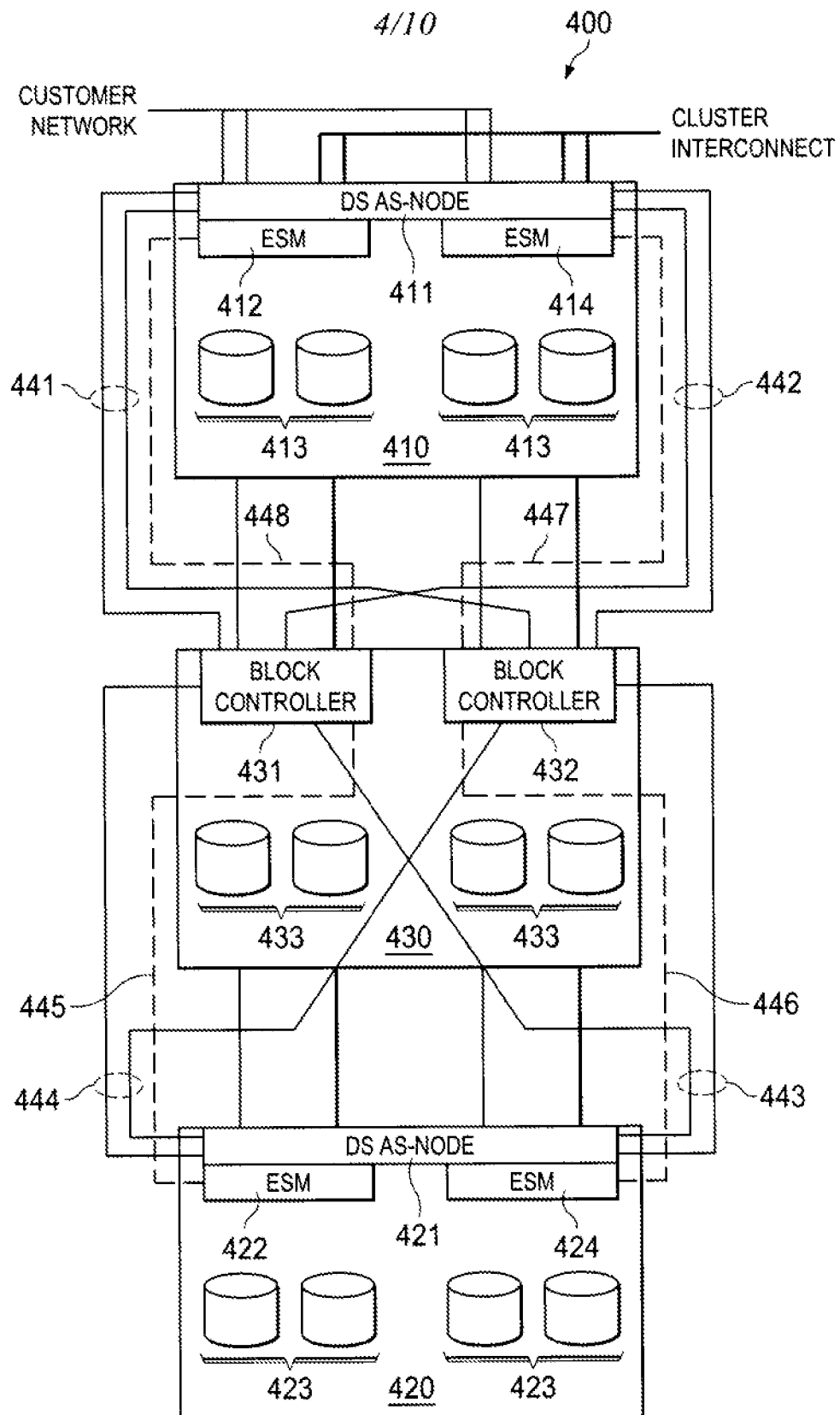


Fig. 4

5/10

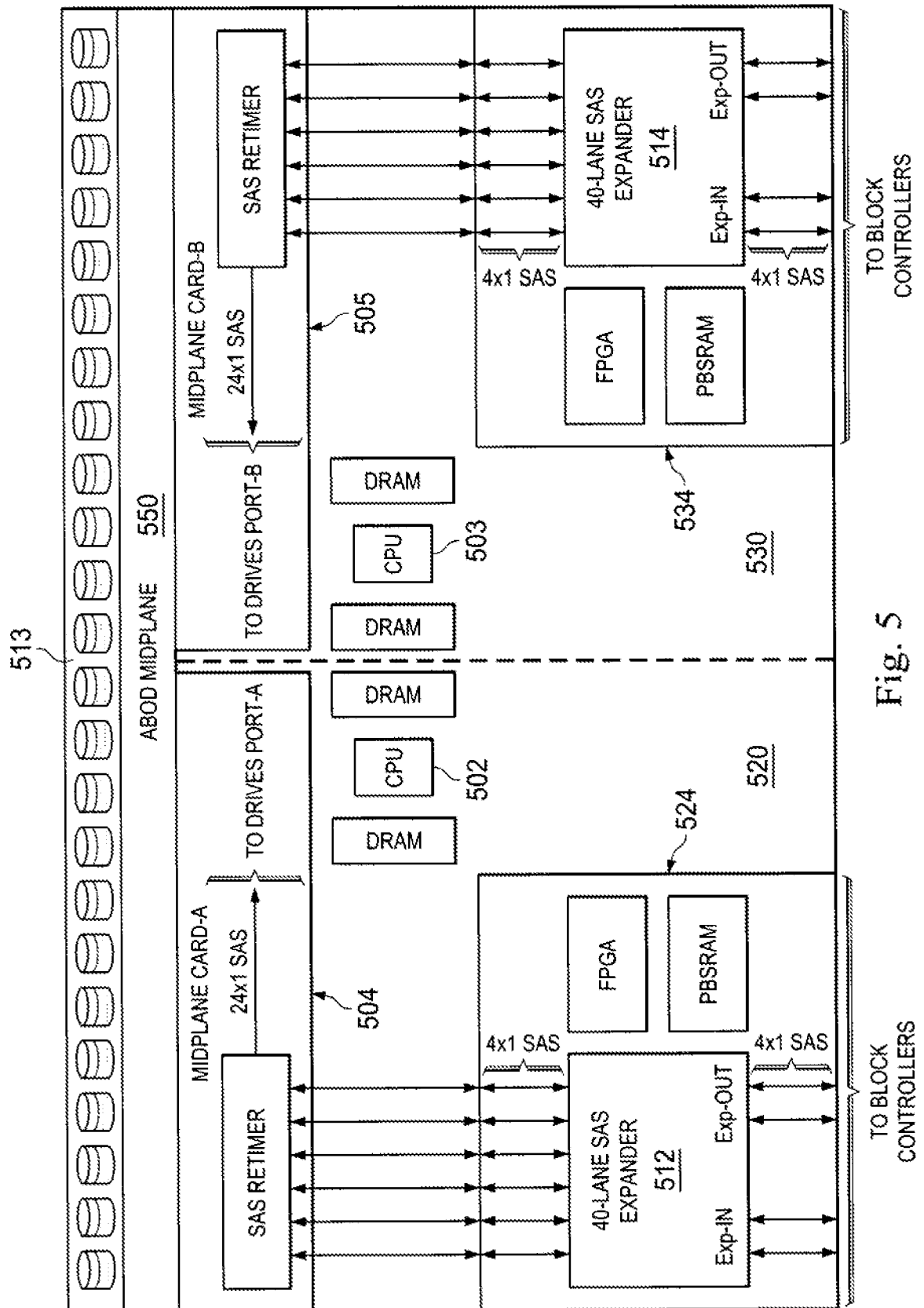


Fig. 5

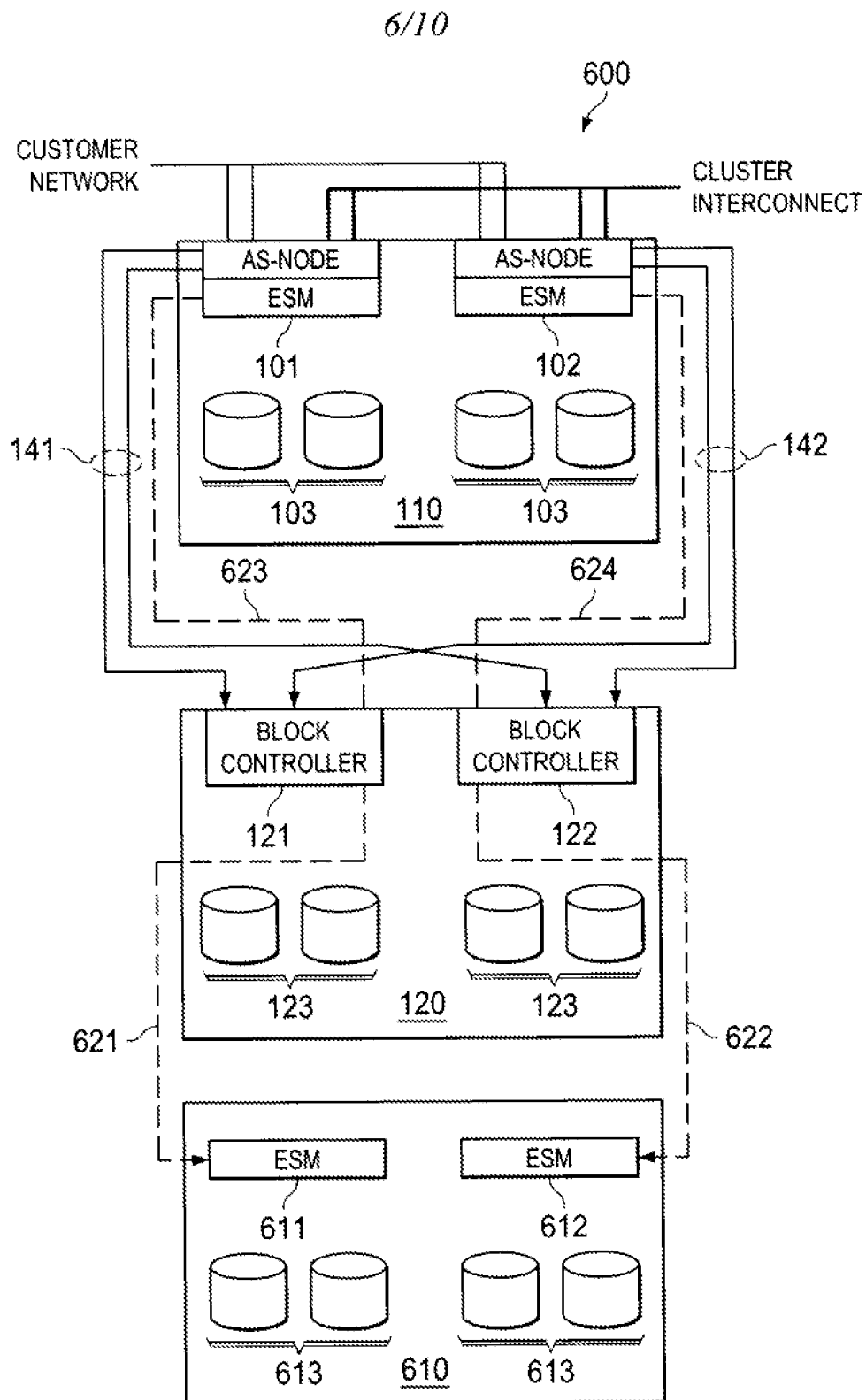


Fig. 6

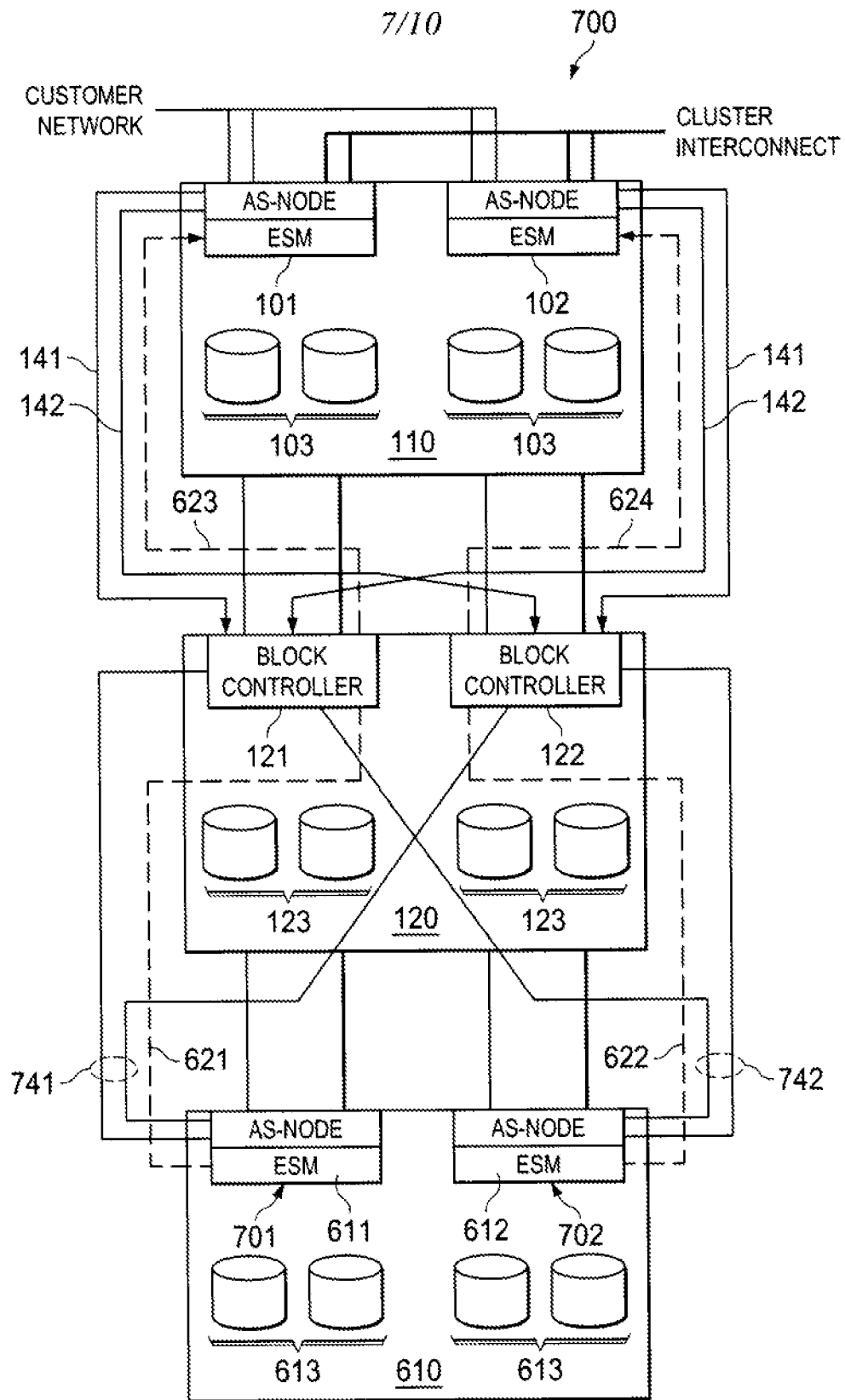


Fig. 7

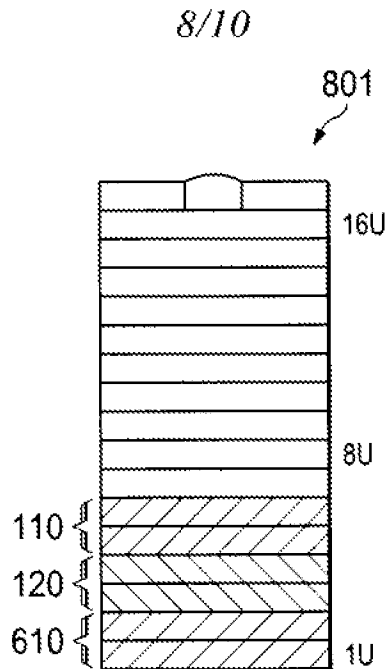


Fig. 8

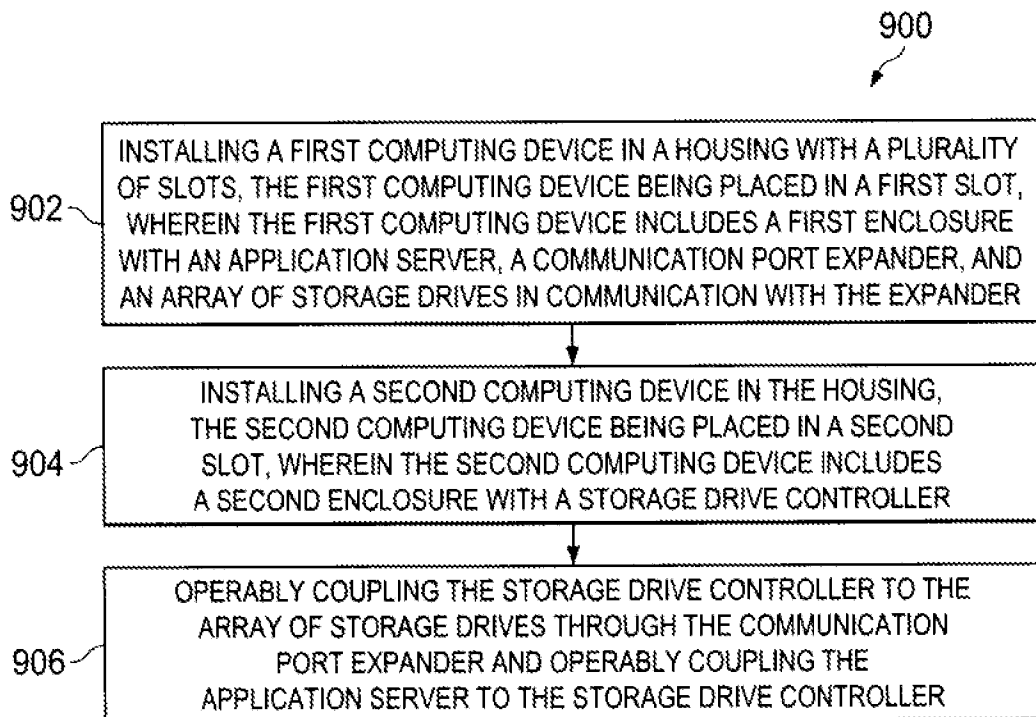


Fig. 9

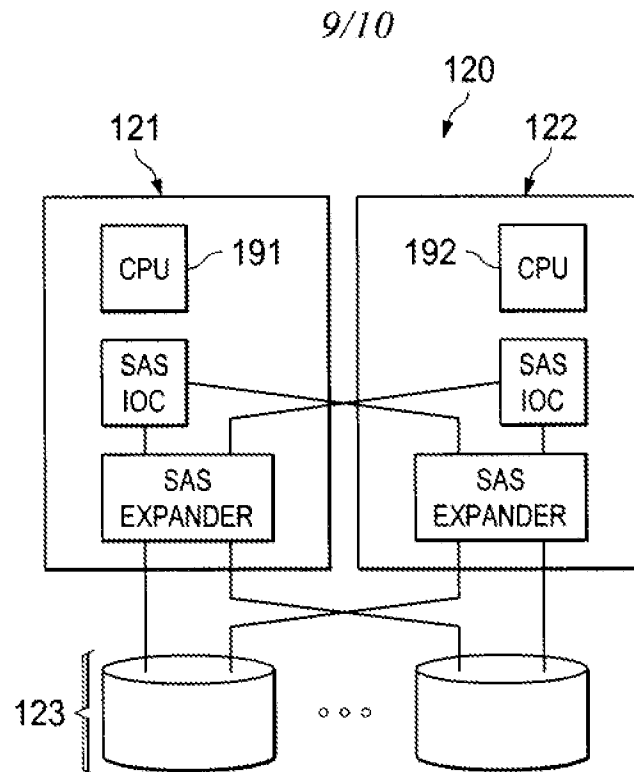


Fig. 10

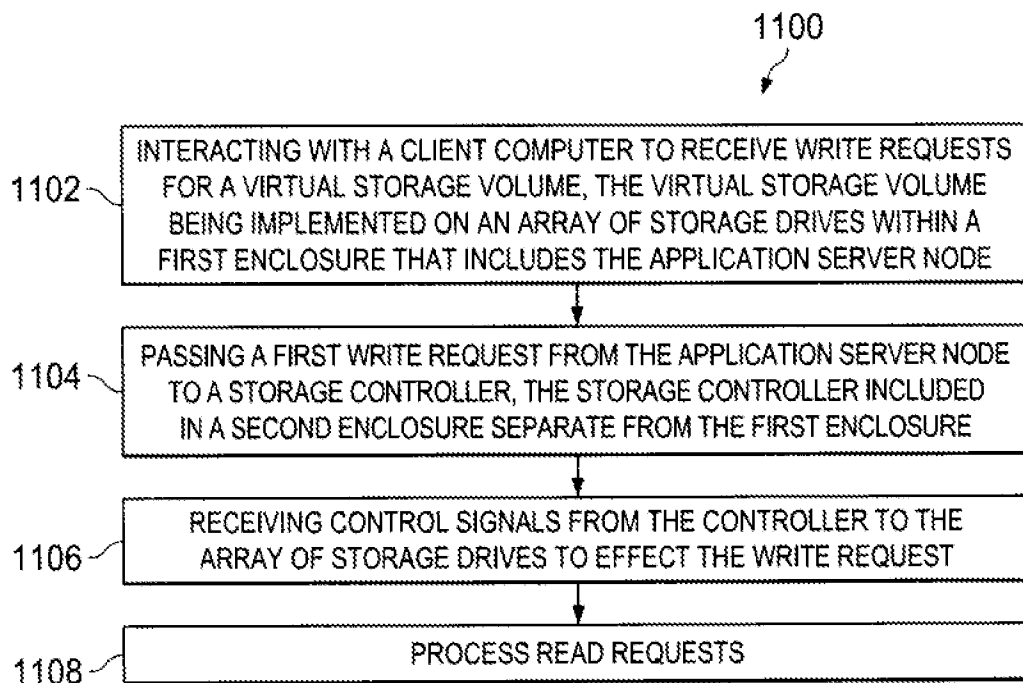


Fig. 11

10/10

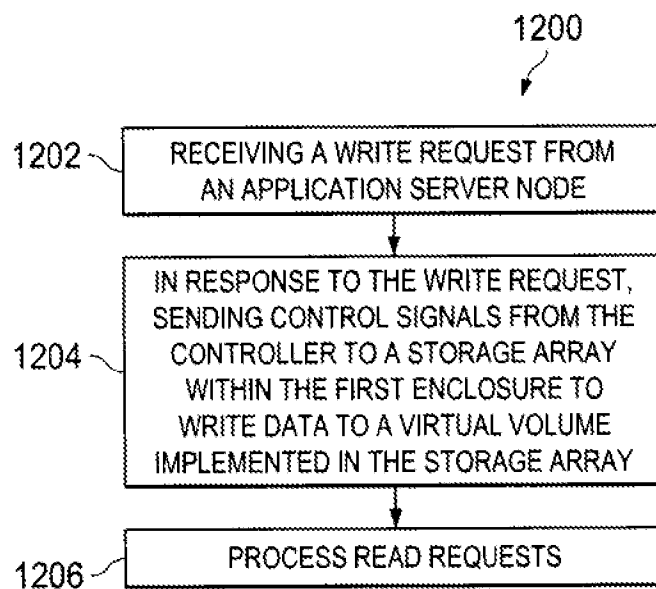


Fig. 12

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 14/31256

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 12/00 (2014.01), G06F 13/00 (2014.01)

USPC - 711/111, 711/114

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)

USPC: 711/111, 711/114; IPC(8): G06F 12/00 (2014.01), G06F 13/00 (2014.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC: 711/111, 711/203, 711/114, 711/E12.019, 711/112, 711/154, 711/202; 717/177; 709/208, 709/206, 709/207, 709/212, 709/250, 709/230, 709/203, 709/229; 705/34, 705/400; 726/26; 718/1; 707/E17.018, 707/999.107, 707/999.104, 707/999.003; 710/62; 710/1; 714/6.1 (keywords)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase; Google Scholar

Search Terms: disk, drive, storage, array, RAID, control, virtual, server, app, application, software, program, warehouse, database, repository, data store, port, expand, slot, separate, external, remote, outside, rack, chassis, enclosure, housing, 2U, power, temperature

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                           | Relevant to claim No.                       |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------|
| X         | US 2010/0064169 A1 (DAVIES et al.) 11 March 2010 (11.03.2010), entire document, especially abstract and para [0008]-[0009], [0012]-[0014], [0041]-[0050], [0052], [0054], [0057], [0060], [0064], [0068], [0071], [0073], [0075]-[0076], [0087]-[0088], [0092]-[0094], [0124]-[0128], [0133], [0170]-[0171]. | 1-8, 10, 12-19, 21-26<br>-----<br>9, 11, 20 |
| Y         | US 2010/0200523 A1 (HENDERSON) 12 August 2010 (12.08.2010), entire document, especially abstract and para [0020]-[0024], Fig. 2, Fig. 3.                                                                                                                                                                     | 9, 11, 20                                   |

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


\* 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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

02 July 2014 (02.07.2014)

Date of mailing of the international search report

11 AUG 2014

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774